Università degli Studi Di Udine

Tesi di Dottorato di Ricerca

Per ottenere il titolo di

Dottore di Ricerca dell'Università di Udine
e
Docteur de l'Institut National Polytechnique de Grenoble

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

Institut de Microélectronique, Electromagnétisme et Photonique - LAboratoire d'Hyperfrequences et de Caractérisation
Ecole Doctorale Électronique, Électrotechnique, Automatique, Télécomunications, Signal (EEATS)

Francesco CONZATTI

Udine,...

Numerical Simulation of Advanced CMOS and Beyond CMOS Devices

Relatori:
David ESSENI
Edwige BANO

Commissione esaminatrice
Prof. Roberto RINALDO, Presidente
Prof. Andreas SCHENK, Esaminatore
Prof. Wim MAGNUS, Esaminatore
Prof. Edwige BANO, Tutor (Fr)
Prof. David ESSENI, Tutor (It)
Dr. Marco PALA, Invitato

INSTITUT POLYTECHNIQUE DE GRENOBLE

N° attribué par la bibliothèque
|__|__|__|__|__|__|__|__|__|__|

THESE EN COTUTELLE INTERNATIONALE

pour obtenir le grade de

DOCTEUR DE L'Institut polytechnique de Grenoble
et
de l'Università degli Studi di Udine

Spécialité : **Micro et Nano électronique**

préparée au laboratoire **Institut Microélectronique, Electromagnétisme et Photonique**

dans le cadre de **l'Ecole Doctorale Électronique, Électrotechnique, Automatique et Traitement du Signal**

et au laboratoire **Dipartimento di Ingegneria Elettrica, Gestionale e Meccanica**

présentée et soutenue publiquement

par

Francesco CONZATTI

Le ??/ ?? 2012

NUMERICAL SIMULATION OF ADVANCED CMOS AND BEYOND CMOS DEVICES

***DIRECTEURS DE THESE*: Edwige Bano et David ESSENI**
***ENCADRANT DE THESE*: Marco PALA**

JURY

Pr.	Roberto RINALDO	, Président
Pr.	Andreas SCHENK	, Examinateur
Pr.	Wim MAGNUS	, Examinateur
Dr.	Marco PALA	, Co-encadrant (Fr)
Pr.	Edwige BANO	, Directeur de thèse (Fr)
Pr.	David ESSENI	, Directeur de thèse (It)

Contents

Notation

Abbreviations and acronyms

BTBT	Band-to-band-tunneling
BTE	Boltzmann transport equation
CCS	Crystal coordinate system
DCS	Device coordinate system
DG	Double gate
DOS	Density of states
EMA	Effective mass approximation
EOT	Equivalent oxide thickness
FDM	Finite difference method
GF	Green's Function
HP	High-performance
ITRS	International technology roadmap for semiconductors
LOP	Low-operating power
LSTP	Low stand-by power
MC	Monte Carlo
MOS	Metal-oxide-semiconductor
MOSFET	MOS field effect transistor
MRT	Momentum relaxation time
NEGF	Non-Equilibrium Green's Function
PSM	Pseudo-spectral method
SG	Single gate
SOI	Silicon on insulator
SS	Subthreshold swing
TCAD	Technology computer-aided design
TFET	Tunnel Field-Effect-Transistor
VLSI	Very large scale integration
VS	Virtual source

Notation

x	Scalar
$\mathrm{x}^\dagger$	Complex conjugate of the scalar x
x + (c.c)	A scalar plus the complex conjugate, namely $(\mathrm{x} + \mathrm{x}^\dagger)$
$\mathbf{x}$	Vector, matrix or multi-dimensional tensor
x_{ij}	Element of the matrix $\mathbf{x}$
$\mathbf{x}^\mathrm{t}$	Transpose of the vector or matrix $\mathbf{x}$
$\mathbf{x}^\dagger$	Transpose conjugate (or self-adjoint) of the vector or matrix $\mathbf{x}$
$\mathbf{x}\cdot\mathbf{y}$	Scalar product between vectors $\mathbf{x}$ and $\mathbf{y}$
$\hat{\mathrm{H}}$	Operator: typically consisting of a differential and an algebraic part
∇f	Gradient of the function f
$\nabla^2 f$	Laplacian of the function f
$[hkl]$	Miller indices that specify a crystal direction
(hkl)	Miller indices that specify the crystal plane normal to $[hkl]$

Symbols:

a_0	Direct lattice constant of a crystal	m
n_{sp}	Spin degeneracy factor: can be either 1 or 2	unitless
$\mathbf{F}$	Electric field	V m^{-1}
F_{eff}	Effective electrical field in an inversion layer	V m^{-1}
$\mathbf{v}_{\mathrm{g}}$	Group velocity	m s^{-1}
m_x, m_y, m_z	Effective electron masses in the x, y and z direction	Kg
A	Normalization area	m^2
ϕ	Electrostatic potential	V
U	Potential energy	J
T	Temperature	K
V	Voltage at device terminals	V
V_{GS}	Intrinsic terminal voltage difference from gate to source	V
V_{DS}	Intrinsic terminal voltage difference from drain to source	V
V_{T}	Threshold voltage	V
L_{G}	Gate length	m
I_{on}	Drain current per unit width at V_{GS}=V_{DS}=V_{DD}	A/m
I_{off}	Drain current per unit width at V_{GS}=0, V_{DS}=V_{DD}	A/m
T_{ox}	Physical oxide thickness	m
N_{inv}	Electron inversion layer density	m^{-2}
P_{inv}	Hole inversion layer density	m^{-2}
v^+	Average velocity of carriers moving from source to drain	m/s
r	Back-scattering coefficient	unitless
G	Retarded Green's function	J^{-1}
G^a	Advanced Green's function	J^{-1}
$G^<$	Lesser-than Green's function	J^{-1}
$G^>$	Greater-than Green's function	J^{-1}
Σ	Retarded self-energy	J
$LDOS$	Local density of states	J^{-1}

Physical constants

h	Planck's constant	6.626075×10^{-34} Js
$\hbar$	Reduced Planck's constant	$h/(2\pi)$
k_B	Boltzmann's constant	1.380662×10^{-23} JK^{-1}
e	Positive electron charge	1.602189×10^{-19} C
m_0	Electron rest mass	9.109390×10^{-31} Kg
ε_0	Dielectric constant of vacuum	8.854188×10^{-12} CV^{-1}m^{-1}

Abstract

THE ongoing scaling of nanoscale semiconductor devices demands of predictive simulation tools which can reliably forecast the performance of future devices. In particular, the physics-based modeling should be able to support the most relevant technological options which are being explored in modern nano-scale devices.

The evaluation of these engineering options is a real challenge that demands a reliable physics-based modeling, with a constant validation of the simulation results against experimental data. A realistic modeling approach of nano-scale devices must take into account the complex features of the band-structure, it must include the quantum mechanical effects and the major scattering mechanisms, that play an important role even in extremely short devices.

In this context, the aim of this co-tutored PhD thesis is to investigate the effect of different technology boosters on both conventional and innovative devices by means of numerical simulations. To this purpose, we developed two TCAD simulation tools based on advanced modeling techniques.

We first developed a self-consistent $\mathbf{k}\cdot\mathbf{p}$-Poisson solver for hole inversion layers and with this simulation code we investigated the effect of strain and channel orientation on the hole mobility in both bulk and Double-Gate (DG) MOSFETs. In particular, we investigated the effect of complex strain configurations on the low-field mobility of both n- and p-type FinFETs and finally identified the optimal stress configurations in order to obtain the highest mobility enhancements. Then, we presented a re-examination of the effect of biaxial strain on the mobility enhancements in MOS transistors. Finally, we explored the potential in terms of on-current (I_{on}) for Ge and strained Ge (sGe) n- and p-MOSFETs versus Si and strained Si (sSi) MOSFETs, where we have shown that, according to our simulation model, sGe MOSFETs can outperform sSi MOSFETs in terms of I_{on}.

During the second part of the PhD, we investigated the effect of strain and variability issues in InAs Tunnel-FETs, which are very promising devices for future low-power applications. To do so, we developed a 3-D NEGF-Poisson solver based on a 8-band-$\mathbf{k}\cdot\mathbf{p}$ Hamiltonian, which can deal with arbitrary crystal orientations and strain conditions. In particular, we demonstrated that strain helps open an I_{off} and V_{DD} window where Tunnel-FETs may outperform sSi MOSFETs. Finally, we demonstrated that, for a very low supply voltage of about 0.3V, the variability of Tunnel-FETs is much smaller for I_{on} than for I_{off} and, furthermore, the I_{on} variability is small compared to sSi MOSFETs.

Chapter 1

Introduction

1.1 Trends in modern nanoelectronic industry

In the last 40 years we have observed an extraordinary growth of the electronic industry, mainly enabled by the continued shrinkage of the device dimensions, which is known as *scaling*. The largest industries in the microelectronic field have made huge investments in research and development in order to meet the expectations produced by the *Moore's law* [1], which states that the device density per chip doubles approximately every one and a half year. This has been achieved thanks to the extraordinary scalability of the Complementary-Metal-Oxide-Semiconductor (CMOS) transistor [2], which represents the fundamental building block of all digital integrated circuits.

Despite the end of traditional scaling has been predicted several times, these predictions have been proven wrong [3]. Two main examples of wrong predictions are the lithography limit of 400nm forecasted in the 70s [4] and the apparent lower bound of 3nm for the gate oxide thickness [5]. Nowadays, scaling still continues, but few constraints have been relaxed in order to maintain high performance enhancements when the dimensions are scaled down. As an example, to obtain the expected benefits from strain, the off-current specification has been increased from 1nA/μm to 100nA/μm, thus augmenting the off-state leakage and finally the static power consumption for high-performance transistors.

However, it is widely recognized in the scientific community that we are approaching the fundamental limits of conventional MOSFETs [3]. As an example, the physical thinning of the dielectric oxide has stopped and other solutions have been explored in order to enhance the gate capacitance, such as replacing the standard silicon dioxide with alternative materials offering a high dielectric constant, the so-called high-κ materials [6, 7]. However, transistors with a high-κ material as the gate dielectric suffer from a degraded mobility [8, 9], thus usually an interfacial layer (SiO_2 for Si channel transistors) is integrated between the channel and the high-κ dielectric in order to limit the mobility reduction and obtain a better interface.

Also the choice of the device architecture emerged as an important issue because of the limitations of bulk planar MOSFETs. The most promising alternative device options are the single-gate SOI, the double-gate SOI, the FinFET and Gate-All-Around (GAA) nanowire transistors, which enable a better electrostatic control. In this respect, very recently Intel has announced its new 22nm technology, where for the first time they replaced the standard bulk transistors with the 3-D FinFET architecture, which ensures better performances with a lower power consumption [10].

Another important aspect deals with the mobility of the semiconductor channel. In the past, unstrained silicon devices were able to meet the specifications of the International Technology Roadmap for Semiconductors (ITRS), while more recently we have observed alternative solutions emerging to increase the channel mobility. Among the others, we mention here the introduction of mechanical strain, which is the most mature solution, the replacement of silicon with alternative semiconductor materials, such as germanium or III-V compounds, and the exploitation of different crystal orientations.

Due to the large diffusion of portable applications, also the power budget represents a major concern in modern nanoelectronic technologies. This issue shifted the emphasis of the device design from performance to power, the latter becoming more and more important. Various solutions are being explored in order to reduce the power consumption of integrated circuits. One possibility is to replace the standard MOSFET with the tunnel-field-effect-transistor (Tunnel-FET), which may offer very low values for the inverse sub-threshold slope, far below the 60mV/decade limit of conventional MOSFETs at room temperature. The main drawback of this devices consists in the very low values of the ON-current. Various solutions are being explored in order to enhance their performances, ranging from the device design to the choice of the semiconductor material.

In the course of the PhD program we have explored the potentials of some technology boosters for advanced CMOS transistors and also for Tunnel-FETs. In this respect, this chap-

ter presents the most promising engineering knobs to improve the performance of nano-scale transistors and then describes the organization of the rest of the manuscript.

1.2 Technology boosters and innovative devices

1.2.1 Channel mobility enhancement

From the early 90s it has been known that inducing mechanical strain in the channel of a MOSFET could lead to an increase of the carrier mobility. Thereafter, a lot of work has been carried out in order to obtain a stable fabrication process to exploit strain engineering in CMOS technologies. In 2004 Intel has announced the introduction of strain in an industrial 90nm CMOS technology [11], which represented a milestone in the strained-silicon technology. The strain in the channel of the transistors is still largely exploited in nowadays CMOS technologies.

The on-current enhancements achieved with the strain engineering have been so remarkable that the strain has sometimes replaced rather than simply augmented the geometrical scaling, as can be seen in Fig.1.1 reporting the equivalent oxide thickness (EOT) and the gate length, L_G, corresponding to different technology nodes [12]. In particular, Fig.1.1 shows that from the 90nm to the 45nm technology nodes the physical channel length L_G and the EOT did not scale and the manufacturers obtained the performance improvement thanks to the strain technology. The strain mainly affects the characteristics of the MOSFETs through the band-structure of the

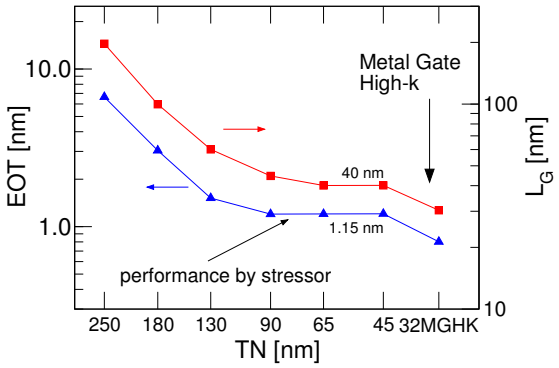


Figure 1.1: Equivalent oxide thickness, EOT, and gate length, L_G, for different technology nodes, TN, as reported in [12]. From the 90nm to the 45nm technology node essentially no geometrical scaling occurred and stress engineering was used to achieve the performance improvements.

carriers. Hence the corresponding physics-based device modelling must account for a credible description of the band-structure, for a consistent treatment of the main scattering mechanisms and for a self-consistent evaluation of the transport and the electrostatics in the device [13]. A reliable modeling framework can be extremely useful for screening of the numerous technology options. In this respect, the Chapter 3 of this manuscript is devoted to the strain technology and describes in detail the models for strain we used throughout the work illustrated in this thesis.

Another possible way to enhance the carrier mobility consists of changing the crystal orientation of the semiconductor substrate [14, 15, 16]. For silicon substrates, the impact of crystal orientation on carrier mobility has been already studied experimentally. It turns out that the most favorable configurations are the (100) and (110) orientations for electrons and holes, respectively. Hence, if we fabricate the devices on the same substrate, the mismatch between the mobilities must be counterbalanced by sizing the n-MOS and the p-MOS transistors with

different widths. Section 4.5 presents the results for the hole mobility simulations with arbitrary channel orientations.

The treatment of arbitrary crystal orientations, strain configurations or even new channel materials demands a large innovative effort in the physics-based and in the TCAD oriented modeling. During this work we will try to give an insightful view of the effect of strain and channel orientation on the performances of nano-scale transistors.

1.2.2 Tunneling-based devices

In order to overcome the limitation of conventional MOSFETs, the Tunnel-FET has been proposed as an alternative device [17]. The main advantage of tunneling-based devices is that they can offer values of the inverse sub-threshold slope (SS) smaller than the theoretical limit of 60mV/decade of conventional MOSFETs (see the sketch in Fig.1.2). However, since the drain current is due to band-to-band tunneling, the on-current (I_{ON}) is typically far below the I_{ON} obtained with conventional MOSFET technologies [18].

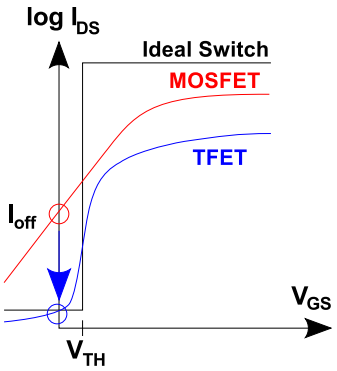


Figure 1.2: Sketch of the drain current I_{DS} versus the gate voltage V_{GS} for an ideal switch (dashed line), a MOSFET (solid line) and a Tunnel-FET (dashed-dotted line). The MOSFET has a far better ON-current with respect to the Tunnel-FET, but at the same time, the Tunnel-FET offers a much steeper turn-on characteristics.

We will see in Chapter 7 that I_{ON} and SS are not the only important metrics for digital applications. In fact, also the switching delay and the switching energy of a digital inverter must be taken into account, and we will present the results of a systematic comparison between InAs Tunnel-FETs and Si MOSFETs in terms of some important metrics.

The basic idea of a Tunnel-FET is to build a gated p-i-n diode. The particular band alignment reported in Fig.1.3 favors the band-to-band tunneling (BTBT) components of the current, while the thermionic part is fully suppressed. In particular, the tunneling occurs from the top of the VB to the bottom of the CB. Fig.1.3(a) and (b) show the bands for a Tunnel-FET in the off-state and in the on-state regime, respectively. We readily notice that the transistor behaves like a band-pass filter, that is it filters out the energies outside of the tunneling window.

From a theoretical point of view, there is a strong difference between BTBT and energy barrier tunneling. The former occurs between valence band and conduction band states and, while being forbidden when perfect translational invariance of the crystal holds, is enabled when strong local electric fields are induced. In direct BTBT, an electron tunnels elastically across the band gap without the assistance of traps or phonons, as sketched in Fig.1.3.

On the other side, energy barrier tunneling refers to particle propagation through an energy barrier. An example where this type of tunneling is non-negligible is represented by the Schottky

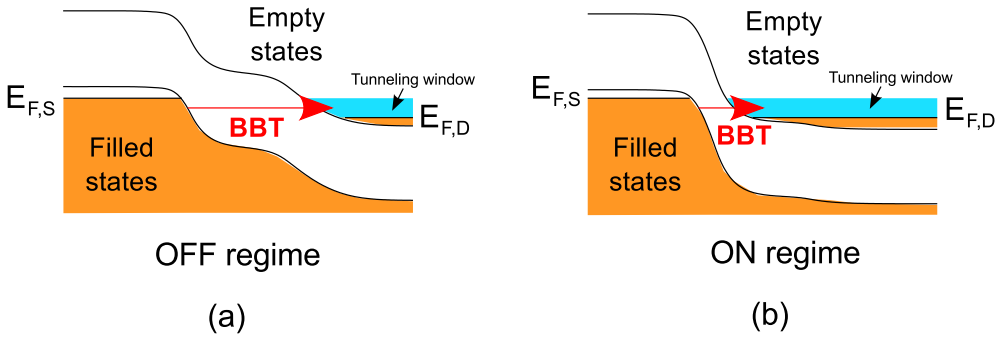


Figure 1.3: Sketch of the VB and the CB along the transport direction for a p-i-n TFET driven in the OFF state (a) and in the ON state (b) regimes. The tunneling occurs from the top of the VB to the bottom of the CB. In the OFF-state regime the CB level in the channel is high, thus the tunneling path is very long. In the ON condition, instead, the CB in the channel is pushed down, so the tunneling path is much shorter.

barrier MOSFETs [19], where the integration of a metal inside the source and the drain of a transistor creates an energy barrier between the reservoir and the channel.

The main drawback of Tunnel-FETs is the very poor value for the I_{ON} experimentally reported so far in literature, especially for silicon-based devices [17]. Hence, many authors have proposed direct-low-bandgap materials as a possible alternative to Si based Tunnel-FETs [20, 21, 22].

The design of the device is of utmost importance in order to obtain good performances and, in particular, it has been demonstrated that the electrostatic control of the gate on the channel potential is necessary to obtain good SS values [20, 21].

In this thesis we have considered Gate-All-Around nanowires Tunnel-FETs. Chapter 7 presents a detailed simulation study of InAs nanowire Tunnel-FETs in the context of a systematic comparison against silicon MOSFETs.

1.3 Purpose of this work

In this context, the aim of this co-tutored PhD thesis is to investigate the effect of different technology boosters on both conventional and innovative devices by means of numerical simulations. To this purpose, we developed two TCAD simulation tools based on advanced modeling techniques.

The first part of the PhD activity has been carried out at the University of Udine, Italy, where we developed a self-consistent **k·p**-Poisson solver for the calculation of the hole mobility in inversion layers, determined according to the Momentum-Relaxation-Time (MRT) approximation. With this simulator we investigated the effect of strain and channel orientation on the hole mobility in either bulk or Double-Gate (DG) MOSFETs.

During the second phase of the PhD, we tackled the quantum transport simulation method, and this work has been carried on at the IMEP-LAHC laboratory in Grenoble, France. We thus developed a 3-D NEGF-Poisson solver which employs a **k·p** Hamiltonian. Subsequently, we deeply investigated the effect of strain and the variability issues in InAs Tunnel-FETs, which are very promising devices for future low-power applications.

The present manuscript is organized as follows. Chapter 2 introduces the 6-band and 8-band **k·p** methods both for bulk semiconductors and in the presence of quantization, where the computational effort increases substantially. Chapter 3 presents the strain in CMOS technologies and how it can be modeled. The **k·p** models and the possible exploitation of strain engineering

are probably the most developed topics of the thesis, in fact are present in all the simulation studies we have carried out.

Then, Chapter 4 introduces the general framework of the semiclassical transport approach and, in particular, the MRT method to calculate the hole mobility by taking into account of several scattering mechanisms. Then, Chapter 5 closes the first part of the thesis showing the main results we obtained with the semiclassical model in the study of the impact of strain and of novel channel materials on the performance of advanced nanoscale MOS devices.

Concerning the second part of the PhD activity, Chapter 6 illustrates the methodologies used to simulate realistic devices by using a quantum transport model and, in particular, reports an introduction to the Non-Equilibrium Green's Function (NEGF) formalism and to the theoretical and implementation techniques we adopted to reduce the computational burden. Then, Chapter 7 presents the simulation results obtained with the quantum transport simulator, thereby focusing on the simulation of homo-junction InAs nanowire Tunnel-FETs.

Finally, Chapter 8 draws the conclusions of this work and provides an outlook towards future developments.

Bibliography

[1] G. Moore, "Cramming More Components onto Integrated Circuits," *Electronics*, vol. 38, no. 8, 1965.

[2] R. H. Dennard, F. H. Gaensslen, L. Kuhn, N. Y. Yu, V. L. Ridout, E. Bassous, and A. LeBlanc, "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE Journal of Solid State Circuits*, vol. SC-9, pp. 256–268, 1974.

[3] W. Haensch, E. Nowak, R. Dennard, P. Solomon, A. Bryant, O. Dokumaci, A. Kumar, X. Wang, J. Johnson, and M. Fischetti, "Silicon CMOS devices beyond scaling," *IBM Journal of Research and Development*, vol. 50, pp. 339–361, 2006.

[4] B. Hoeneisen and C. Mead, "Fundamental limits in microelectronics-I.MOS technology," *Solid State Electronics*, vol. 15, pp. 819–829, 1972.

[5] J. Stathis and D. J. DiMaria, " Reliability Projection for Ultra-Thin Oxides at Low Voltage," in *IEEE IEDM Technical Digest*, 1998, pp. 167–170.

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

[7] J. Robertson, "High dielectric constant oxides," *EPJ Applied Physics*, vol. 28, no. 3, pp. 265–291, 2004.

[8] M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, "Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-κ insulator: The role of remote phonon scattering," *Journal of Applied Physics*, vol. 90, no. 9, pp. 4587–4608, 2001.

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

[10] Intel. 22nm technology, 2011. http://www.intel.com/content/www/us/en/silicon-innovations/intel-22nm-technology.html.

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

[12] M. Horstmann, M. Wiatr, A. Wei, J. Hoentschel, T. Feudel, M. Gerhardt, R. Stephan, S. Krügel, D. Greenlaw, and M. Raab, "Advanced soi cmos transistor technology for high performance microprocessors," in *Proceedings Workshop ULIS*, 2009, pp. 81–84.

[13] D.Esseni, P.Palestri, and L.Selmi, "Transport in deca-nanometric MOSFETs: from band-structure to on-currents," in *IEEE IEDM Technical Digest*, 2006, pp. 933–936.

[14] A. Teramoto, T. Hamada, M. Yamamoto, P. Gaubert, H. Akahori, K. Nii, M. Hirayama, K. Arima, K. Endo, S. Sugawa, and T. Ohmi, "Very High Carrier Mobility for High-Performance CMOS on a Si(110) Surface," *IEEE Trans. Electron Devices*, vol. 54, no. 6, pp. 1438 –1445, june 2007.

[15] G. Tsutsui, M. Saitoh, and T. Hiramoto, "Experimental study on superior mobility in [110]-oriented UTB SOI pMOSFETs," *IEEE Electron Device Lett.*, vol. 26, no. 11, pp. 836 – 838, 2005.

[16] H. Nakamura, T. Ezaki, T. Iwamoto, M. Togo, T. Ikezawa, N. Ikarashi, M. Hane, and T. Yamamoto, "Effects of Selecting Channel Direction in Improving Performance of Sub-100nm MOSFETs Fabricated on (110) Surface Si Substrate," *Japanese Journal of Applied Physics*, vol. 43, no. 4B, p. 17231728, 2004.

[17] W. Y. Choi, B.-G. Park, J. D. Lee, and T.-J. K. Liu, "Tunneling Field-Effect Transistors (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec," *IEEE Electron Device Lett.*, vol. 28, no. 8, pp. 743 –745, aug. 2007.

[18] A. Seabaugh and Q. Zhang, "Low-Voltage Tunnel Transistors for Beyond CMOS Logic," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2095 –2110, dec. 2010.

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

[20] M. Luisier and G. Klimeck, "Atomistic Full-Band Design Study of InAs Band-to-Band Tunneling Field-Effect Transistors," *IEEE Electron Device Lett.*, vol. 30, no. 6, pp. 602 –604, june 2009.

[21] J. Knoch and J. Appenzeller, "Modeling of High-Performance p-Type III-V Heterojunction Tunnel FETs," *IEEE Electron Device Lett.*, vol. 31, no. 4, pp. 305 –307, april 2010.

[22] G. Dewey, B. Chu-Kung, J. Boardman, J. Fastenau, J. Kavalieros, R. Kotlyar, W. Liu, D. Lubyshev, M. Metz, N. Mukherjee, P. Oakey, R. Pillarisetty, M. Radosavljevic, H. Then, and R. Chau, "Fabrication, Characterization and Physics of III-V Heterojunction Tunneling Field Effect Transistors (H-TFET) for Steep Sub-Threshold Swing," in *IEEE IEDM Technical Digest*, Dec. 2011, pp. 33.6.1–33.6.4.

Chapter 2

The k·p Method

2.1 Introduction to the k·p model

THE knowledge of the band-structure is of utmost importance in order to study the physical properties of the most common semiconductor materials. A wide number of methods to calculate band-structure have been developed during the past years. The most popular and widely studied in the electron device community are the tight-binding [1], the pseudopotential [2] and the **k·p** method [3]. The main difference between these approaches is the choice of the basis functions in which the solutions of the time-independent Schrödingerequation are to be expanded: atomic-like, plane-wave and Bloch states, respectively.

The **k·p** method approximates the band-structure of the most common semiconductors starting from the knowledge of a few parameters (a limited number of energy gaps and momentum matrix elements), which are generally used as fitting parameters. Even if the **k·p** method can be used to explore the energy dispersion in the entire Brillouin zone [4, 5], its first and most important application is to approximate with great accuracy the band-structure close to a band edge, namely the top (bottom) of the valence (conduction) band.

The first part of this chapter is devoted to the introduction to the **k·p** method and its application to bulk semiconductors. In the first part we introduce the most common formulation for the **k·p** method. Section 2.2 describes the 6-bands **k·p** formulation, which is well suited for the description of top of the valence band. Section 2.5 presents the 8-bands **k·p** model for the calculation of the band-structure of direct-bandgap materials.

2.1.1 k·p method for a bulk semiconductor

We start by considering the one-electron Schrödinger equation for stationary states written in the Crystal Coordinate System (CCS) by following the approach of [6]:

$$\hat{H}^{(0)}\,\Psi_{n\mathbf{k}}(\mathbf{r}) = E_{n\mathbf{k}}\Psi_{n\mathbf{k}}(\mathbf{r}) \tag{2.1}$$

with

$$\hat{H}^{(0)} = -\frac{\hbar^2}{2m_0}\nabla^2 + U_C(\mathbf{r}) \tag{2.2}$$

where $U_C(\mathbf{r})$ is the periodic potential energy inside the crystal. If we substitute the form of the Bloch wave function for the state $|n\mathbf{k}\rangle$:

$$\Psi_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r})\, e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \tag{2.3}$$

we readily obtain

$$\left[-\frac{\hbar^2}{2m_0}\nabla^2 + U_C(\mathbf{r}) + \frac{\hbar}{m_0}\mathbf{k}\cdot\hat{\mathbf{p}}\right] u_{n\mathbf{k}}(\mathbf{r}) = \left(E_{n\mathbf{k}} - \frac{\hbar^2\mathbf{k}^2}{2m_0}\right) u_{n\mathbf{k}}(\mathbf{r}) \tag{2.4}$$

which is the Schrödinger equation for a relaxed lattice and for a given wavevector $\mathbf{k}$, where $\mathbf{p}$ is the corresponding momentum operator. The index n denotes the band of the bulk semiconductor. We have here implicitly assumed that the band edge corresponds to $\mathbf{k_0}=0$. Eq.2.4 can be re-written as

$$\left[\hat{H}^{(0)} + \hat{H}^{(1)}\right]\, u_{n\mathbf{k}}(\mathbf{r}) = E_{n\mathbf{k}}\, u_{\mathbf{nk}}(\mathbf{r}) \tag{2.5}$$

where

$$\hat{H}^{(1)} = \frac{\hbar}{m_0}\mathbf{k}\cdot\hat{\mathbf{p}} + \frac{\hbar^2\mathbf{k}^2}{2m_0} \tag{2.6}$$

The term $\hat{H}^{(1)}$ can be considered a perturbation term of the unperturbed Hamiltonian $\hat{H}^{(0)}$ because it vanishes for $\mathbf{k}$=0, thus Eq.2.5 can be solved with the standard time-independent second-order perturbation theory around the Γ point ($\mathbf{k}$=0) [7].

We can expand the periodic part of the Bloch wave function as

$$u_{n\mathbf{k}}(\mathbf{r}) = \sum_{J,i} C_{J_i\mathbf{k}} u_{J_i 0}(\mathbf{r}) \tag{2.7}$$

where $u_{J_i 0}(\mathbf{r})$ is the Bloch function at the Γ point. The indexes J refers to the energy and i to the degenerate functions for a given energy [8]. The fundamental branches and the wave functions for $\mathbf{k}\neq 0$ are calculated by solving the eigenvalue problem

$$\mathbf{H}_{\mathbf{k}\cdot\mathbf{p}}\mathbf{C}_{\mathbf{k}} = \left[E_{n\mathbf{k}} - \frac{\hbar^2\mathbf{k}^2}{2m_0}\right]\mathbf{C}_{\mathbf{k}} \tag{2.8}$$

where $\mathbf{C}_{\mathbf{k}}$ is the vector of the expansion coefficients $C_{J_i\mathbf{k}}$. The matrix $\mathbf{H}_{\mathbf{k}\cdot\mathbf{p}}$ can be expressed in the DKK (Dresselhaus, Kip and Kittel) representation as [9, 8]

$$\mathbf{H}_{\mathbf{k}\cdot\mathbf{p}\,(J_i,k)} = \sum_{Lm, E_{L\mathbf{k}}\neq E_{J\mathbf{k}}} \frac{\langle u_{J_i 0}|(\hbar/m_0)\mathbf{k}\cdot\hat{\mathbf{p}}|u_{L_m 0}\rangle\langle u_{L_m 0}|(\hbar/m_0)\mathbf{k}\cdot\hat{\mathbf{p}}|u_{J_k 0}\rangle}{E_{J\mathbf{k}} - E_{L\mathbf{k}}} \tag{2.9}$$

The expansions of Eqs.2.9 is usually truncated to the first terms which correspond to the highest (lowest) branches of the valence (conduction) band. This leads to matrix equations whose rank corresponds to the number of bands we account for. The expression we have derived does not take into account the effect of the spin-orbit interaction, which is usually non negligible and doubles the rank of the **k·p** problem. Interested readers can find further details in [10].

In the next sections we will introduce the 6-bands **k·p** method for hole inversion layers and the 8-bands **k·p** for the calculation of both the valence band and the conduction band of direct-bandgap materials. In particular, we will start from the bulk matrices and then derive the expressions to compute the band-structure in presence of size quantization.

2.2 6-bands k·p method for hole inversion layers

Using the perturbation theory for degenerate states and taking into account the three highest degenerate energies in the valence band at $\mathbf{k}=0$, it can be demonstrated that Eq.2.8 can be written as a 3×3 eigenvalue problem [6]:

$$\mathbf{S}\,\mathbf{b}_n = E'_{n\mathbf{k}}\mathbf{b}_n \tag{2.10}$$

where $\mathbf{b}_n$ is a three component column vector and

$$E'_{n\mathbf{k}} = E_{n\mathbf{k}} - \hbar^2|\mathbf{k}|^2/2m_0 \tag{2.11}$$

is the n-th subband. The vector $\mathbf{b}_n$ expresses the unknown eigenfunction u_{n0} as:

$$u_{n0} = b_{n1}u_{1,0} + b_{n2}u_{2,0} + b_{n3}u_{3,0} \tag{2.12}$$

where the $u_{i,0}$ are the wave functions of the three branches of the valence band at $\mathbf{k}=0$, which are degenerate as long as we neglect the spin-orbit interaction.

The matrix $\mathbf{S}$ is written as:

$$\mathbf{S} = \begin{pmatrix} Lk_x^2 + M(k_y^2+k_z^2) & Nk_xk_y & Nk_xk_z \\ Nk_xk_y & Lk_y^2 + M(k_x^2+k_z^2) & Nk_yk_z \\ Nk_xk_z & Nk_yk_z & Lk_z^2 + M(k_x^2+k_y^2) \end{pmatrix} \tag{2.13}$$

Symmetry arguments are used in order to obtain the parameters L, M, N and their value can be determined from cyclotron resonance measurements [9, 11]. It should be noted that the

	Silicon	Germanium
L [eV Å]	-6.53	-31.53
M [eV Å]	-4.64	-5.64
N [eV Å]	-8.32	-33.64
l [eV]	-2.30	-2.20
m [eV]	4.30	4.10
n [eV]	-9.18	-12.12

Table 2.1: **k·p** parameters for silicon and germanium taken from [13] and [14]. L and M differ by 1 in atomic units ($\hbar^2/2m_0$) with respect to the numerical values of [13] because of a different convention in the definition of the eigenvalues in Eq.2.11.

definition of the eigenvalues in Eq.2.11 is different to what has been reported in Eq.1 of Ref. [12]. For this reason, the parameters we will use in our 6-bands **k·p** model differ by 1 in atomic units ($\hbar^2/2m_0$) with respect to the numerical values of [13]. In Tab.2.1 the **k·p** parameters used for silicon and germanium have been reported.

When the spin-orbit interaction effects are taken into account the size of the Hamiltonian matrix in Eq.2.10 doubles. It can be demonstrated that the overall **k·p** matrix becomes:

$$\mathbf{H} = \mathbf{H}_{k\cdot p} + \mathbf{H}_{so} \tag{2.14}$$

where all matrices are 6×6 and $\mathbf{H}_{k\cdot p}$ is set by:

$$\mathbf{H}_{k\cdot p} = \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \tag{2.15}$$

and the spin-orbit matrix is given by [15]:

$$\mathbf{H}_{so} = -\frac{\Delta}{3}\begin{pmatrix} 0 & \mathrm{i} & 0 & 0 & 0 & -1 \\ -\mathrm{i} & 0 & 0 & 0 & 0 & \mathrm{i} \\ 0 & 0 & 0 & 1 & -\mathrm{i} & 0 \\ 0 & 0 & 1 & 0 & -\mathrm{i} & 0 \\ 0 & 0 & \mathrm{i} & \mathrm{i} & 0 & 0 \\ -1 & -\mathrm{i} & 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.16}$$

where Δ is the gap at the Γ point induced by the spin-orbit interaction, which is 44meV for silicon and 296meV for germanium. The overall eigenvalue problem reads:

$$\mathbf{H}\,\boldsymbol{\psi}_{n\mathbf{k}} = \left[E_{n\mathbf{k}} - \frac{\hbar^2(k_x^2 + k_y^2 + k_z^2)}{2m_0}\right]\boldsymbol{\psi}_{n\mathbf{k}} \tag{2.17}$$

where $\boldsymbol{\psi}$ is a 6-component vector.

2.2.1 Formulation of the k·p model for a hole inversion layer

When a confining potential $U(z)$ along the z direction is present, the formulation of the **k·p** method becomes:

$$\left[\hat{\mathbf{H}}\left(\mathbf{k}, -\mathrm{i}\frac{\partial}{\partial z}\right) + \mathbf{I}\,U(z)\right]\boldsymbol{\psi}_{n\mathbf{k}}(z) = E(\mathbf{k})\,\boldsymbol{\psi}_{n\mathbf{k}}(z) \tag{2.18}$$

where the wavevector $\mathbf{k}$ is normal to the quantization direction and the k_z vector has been substituted with the differential operator $-\mathrm{i}\partial/\partial z$. The kinetic energy operator $\mathbf{H}$ includes the factor $[\hbar^2/2m_0]\,\mathbf{I}$ stemming from the right hand side of Eq.2.17. The 2-dimensional wavevector $\mathbf{k}$ is a parameter in Eq.2.18 hence, due to the strong anisotropy of the valence band, we need

to solve the eigenvalue problem for each point in the **k**-space. This could augment the computational burden in an unacceptable way, so we must find ways to reduce the simulation time. In the work of this thesis we carefully discretized the 2D **k**-space (see Section 2.4) and we used a smart discretization in real space along the quantization direction (see Section 2.3.1), thus reducing the number of grid points in both spaces.

The Hamiltonian operator changes if we consider different quantization directions z: the most interesting cases are those with quantization along [001], [110] and [111], corresponding to the crystallographic orientations (001), (110) and (111), respectively. The matrix **S** of Eq.2.13 is defined in the CCS, thus, in case of crystal orientations different from (001), it is necessary to perform a rotation of the principal axes. The quantization direction corresponds to the z axis of the Device Coordinate System (DCS), which is rotated with respect to the CCS. For a detailed description of the transformations, see Section 3.2.3.

In the following the expression of the **k·p** matrices for the main crystal orientations are reported.

(001) crystal orientation

In this case the CCS and the DCS are the same and so no axes rotation is needed. The **S** matrix of Eq.2.10 can be written as:

$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix} \tag{2.19}$$

where

$$S_{11} = \left(L + \frac{\hbar^2}{2m_0}\right) k_x^2 + \left(M + \frac{\hbar^2}{2m_0}\right) k_y^2 - \left(M + \frac{\hbar^2}{2m_0}\right) k_z^2 \tag{2.20}$$

$$S_{22} = \left(M + \frac{\hbar^2}{2m_0}\right) k_x^2 + \left(L + \frac{\hbar^2}{2m_0}\right) k_y^2 - \left(M + \frac{\hbar^2}{2m_0}\right) k_z^2 \tag{2.21}$$

$$S_{33} = \left(M + \frac{\hbar^2}{2m_0}\right) k_x^2 + \left(M + \frac{\hbar^2}{2m_0}\right) k_y^2 - \left(L + \frac{\hbar^2}{2m_0}\right) k_z^2 \tag{2.22}$$

$$S_{12} = S_{21} = N k_x k_y$$

$$S_{13} = S_{31} = N k_x k_z$$

$$S_{23} = S_{32} = N k_y k_z$$

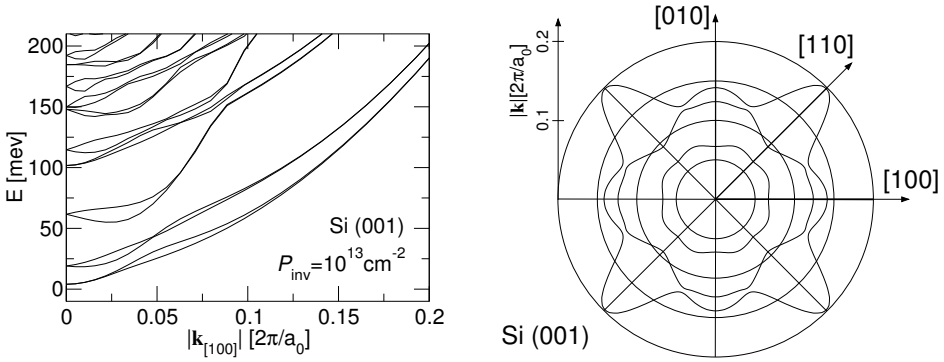
The standard quantum mechanics prescription turns the operator $\hat{k}_z$ into the corresponding derivative $-\mathrm{i}\partial/\partial z$, because z corresponds to the quantization direction. If we then discretize the operators, the matrix **S** of Eq.2.19 becomes a block matrix. An explicit example of **k·p** quantization matrix is given in Section 2.3.1. This is valid for all quantization directions.

Fig.2.1(a) and (b) report the energy dispersion along the [001] direction for a hole inversion density of $10^{13}\mathrm{cm}^{-2}$ and the corresponding equi-energy curves at 100 meV above the minimum.

(110) crystal orientation

In this case the quantization direction (labeled as z in DCS) is along the line k_x=k_y in the CCS. The rotation matrix **R** such that $\mathbf{k}_{CCS} = \mathbf{R}\,\mathbf{k}_{DCS}$ can be taken as

$$\mathbf{R} = \begin{pmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{pmatrix} \tag{2.23}$$



(a) (b)

Figure 2.1: (a) Energy dispersion for a (001) Si hole inversion layer along the [100] crystal direction and for an inversion density of 10^{13}cm^{-2}. The energy reference corresponds to the Fermi energy and is equal to 0eV. **The convention for the hole energy is electron-like**. (b) Equi-energy curves 100 meV above the minimum of the valence band.

The **S** matrix defining $\mathbf{H}_{k\cdot p}$ according to Eq.2.15 can be expressed as a function of (k_x,k_y,k_z) belonging to the DCS as:

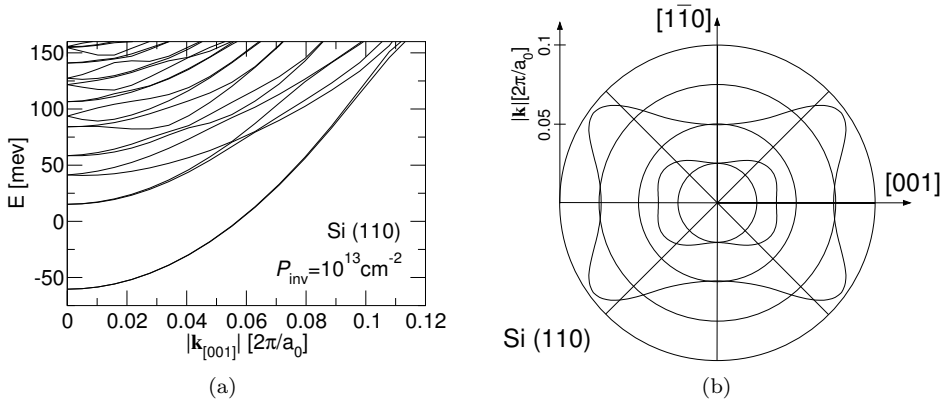
$$
\begin{aligned}
S_{11} &= \left(M + \frac{\hbar^2}{2m_0}\right)k_x^2 + \left(\frac{L+M}{2} + \frac{\hbar^2}{2m_0}\right)(k_y^2 + k_z^2) + (L-M)k_y k_z \\
S_{22} &= \left(M + \frac{\hbar^2}{2m_0}\right)k_x^2 + \left(\frac{L+M}{2} + \frac{\hbar^2}{2m_0}\right)(k_y^2 + k_z^2) + (M-L)k_y k_z \\
S_{33} &= \left(L + \frac{\hbar^2}{2m_0}\right)k_x^2 + \left(M + \frac{\hbar^2}{2m_0}\right)(k_y^2 + k_z^2) \\
S_{12} &= S_{21} = \frac{N}{2}(k_z^2 - k_y^2) \\
S_{13} &= S_{31} = \frac{N}{\sqrt{2}}k_x(k_y + k_z) \\
S_{13} &= S_{31} = \frac{N}{\sqrt{2}}k_x(k_z - k_y)
\end{aligned}
$$

Fig.2.2 reports the energy dispersion along the [001] direction for a hole inversion density of 10^{13}cm^{-2} and the corresponding equi-energy curves at 100 meV above the minimum.

(111) crystal orientation

In this case the quantization direction forms an angle of $\pi/4$ with each axis of the CCS. The rotation matrix **R** such that $\mathbf{k}_{CCS} = \mathbf{R}\,\mathbf{k}_{DCS}$ is equal to:

$$
\mathbf{R} = \begin{pmatrix} -2/\sqrt{6} & 0 & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{pmatrix} \tag{2.24}
$$



(a) (b)

Figure 2.2: (a) Energy dispersion for a (110) Si hole inversion layer along the [001] crystal direction and for an inversion density of 10^{13}cm^{-2}. The energy reference corresponds to the Fermi energy and is equal 0eV. (b) Equi-energy curves 100 meV above the minimum of the valence band.

As done for the previous cases, the **S** matrix can be expressed as a function of the CCS wavevector components:

$$
\begin{aligned}
S_{11} &= \left(\frac{2L+M}{3}+\frac{\hbar^2}{2m_0}\right)k_x^2+\left(M+\frac{\hbar^2}{2m_0}\right)k_y^2+\left(\frac{L+2M}{3}+\frac{\hbar^2}{2m_0}\right)k_z^2+ \\
&\quad \frac{4}{3\sqrt{2}}(M-L)\,k_xk_z \\
S_{22} &= \left(\frac{L+5M}{6}+\frac{\hbar^2}{2m_0}\right)k_x^2+\left(\frac{L+M}{2}+\frac{\hbar^2}{2m_0}\right)k_y^2+\left(\frac{L+2M}{3}+\frac{\hbar^2}{2m_0}\right)k_z^2+ \\
&\quad (M-L)\,k_xk_y/\sqrt{3}+\left[\frac{1}{2}\left(L+\frac{\hbar^2}{2m_0}\right)-2\left(M+\frac{\hbar^2}{2m_0}\right)\right]k_xk_z/3\sqrt{2}+ \\
&\quad (M-L)\,k_yk_z/2\sqrt{6} \\
S_{33} &= \left(\frac{L+5M}{6}+\frac{\hbar^2}{2m_0}\right)k_x^2+\left(\frac{L+M}{2}+\frac{\hbar^2}{2m_0}\right)k_y^2+\left(\frac{L+2M}{3}+\frac{\hbar^2}{2m_0}\right)k_z^2+ \\
&\quad (L-M)\,k_xk_y/\sqrt{3}+\left[\frac{1}{2}\left(L+\frac{\hbar^2}{2m_0}\right)-2\left(M+\frac{\hbar^2}{2m_0}\right)\right]k_xk_z/3\sqrt{2}+ \\
&\quad (L-M)\,k_yk_z/2\sqrt{6} \\
S_{12} &= S_{21}=-N(k_x^2/3-k_z^2/3-k_xk_y/\sqrt{3}+k_xk_z/3\sqrt{2}+k_yk_z/\sqrt{6}) \\
S_{13} &= S_{31}=-N(k_x^2/3-k_z^2/3+k_xk_y/\sqrt{3}+k_xk_z/3\sqrt{2}-k_yk_z/\sqrt{6}) \\
S_{23} &= S_{32}=N(k_x^2/6-k_y^2/2+k_z^2/3+\sqrt{2}k_xk_z/3)
\end{aligned}
$$

Fig.2.2 reports the energy dispersion along the [11$\bar{2}$] direction for a hole inversion density of 10^{13}cm^{-2} and the corresponding equi-energy curves at 100 meV above the minimum.

2.3 Pseudo-spectral methods

2.3.1 Pseudo-spectral methods in the k·p framework

In this section we will describe concisely the numerical technique called *Pseudo-Spectral Methods* (PSM) used hereafter in order to solve in an efficient way the **k·p** problem in the presence of size quantization due to hole confinement.

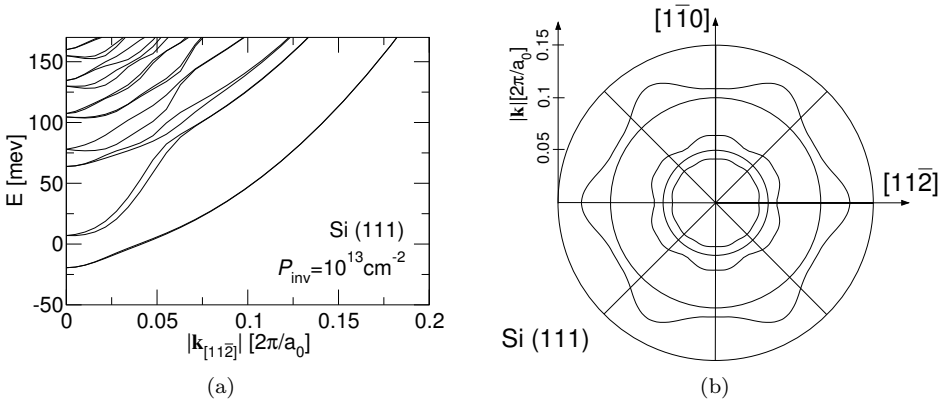


Figure 2.3: (a) Energy dispersion for a (111) Si hole inversion layer along the $[11\bar{2}]$ crystal direction and for an inversion density of 10^{13}cm^{-2}. The energy reference corresponds to the Fermi energy and is equal to 0eV. (b) Equi-energy curves 100 meV above the minimum of the valence band.

Spectral Methods

The PSM for a differential equation or a differential eigenvalue problem approximates the unknown function by using algebraic or trigonometric polynomials. The approximation error decay of this approach with respect to degree N of the polynomial depend on the degree of the smoothness of the unknown function: the more regular the unknown function is, the more rapidly the error decreases. In particular, the PSM can achieve the so-called *spectral accuracy* for analytic functions, which means that the approximation error decreases exponentially as $O(C^N)$, with C being a real number $\in (0,1)$ and N being the degree of the approximating polynomial. This behavior is expected to yield a remarkable reduction of the CPU time with respect to the conventional finite difference and finite element methods that result in a fixed order accuracy (typically second order in the 1D case), since the convergence is attained by decreasing the mesh size while keeping the polynomial degree constant.

Being defined on a closed interval $[0, W]$, any unknown function $u(x)$ can be approximated by Lagrange polynomials as follows:

$$u(x) \simeq p(x) = \sum_{j=0}^{N_x} l_j(x)\, u(x_j) \tag{2.25}$$

with $x_j \in [0, W]\ \forall j \in \{0, \dots N_x\}$, $x_i \neq x_j\ \forall i \neq j$ and

$$l_j(x) = \prod_{i=0, i\neq j}^{N_x} \frac{x - x_i}{x_j - x_i}. \tag{2.26}$$

$l_j(x)$ is the j-th Lagrange basis polynomial, which for $x = x_i$ reads:

$$l_j(x_i) = \begin{cases} 1 & for\ i = j \\ 0 & otherwise \end{cases} \tag{2.27}$$

It is worth noting that the choice of the discretization points is critical. For example, equidistant discretization points (see Fig.2.4) lead to large errors, due to the so-called *Runge-phenomenon* [16]. Indeed, the interpolation error increases exponentially with N_x: $p(x)$ can oscillate near the boundaries of the domain, as we can see in Fig.2.5.

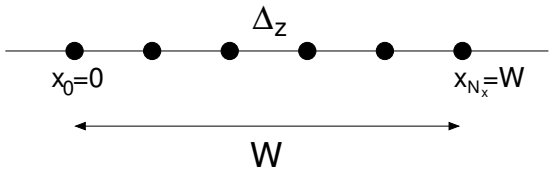


Figure 2.4: Equispaced grid with $N_x + 1$ discretization points.

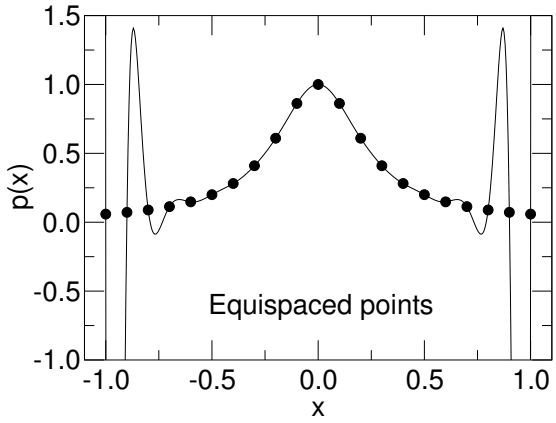


Figure 2.5: Degree N_x interpolation of $u(x) = 1/(1+16x^2)$ on a equispaced grid. The oscillations near the boundaries are called the *Runge phenomenon* [16].

To avoid this unwanted behavior the idea is to use unevenly spaced points. For the considered unknown function $u(x)$, the *Chebyshev extremal nodes* represent a good choice [16]. They are defined as:

$$x_j = W/2 - W/2 \cos(j\pi/N_x) \quad j = 0, 1, \ldots, N_x \tag{2.28}$$

From a geometrical point of view we can picture these points as the projections onto a segment of equally spaced points along a circle, as sketched in Fig.2.6. We notice that the *Chebyshev extremal nodes* accumulate near the boundaries.

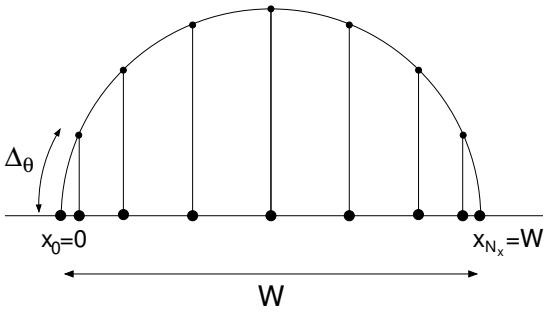


Figure 2.6: PSM grid. The *Chebyshev extremal nodes* are the projections onto a segment of equally spaced points along a circle. The width of the domain is equal to W.

If we perform the same polynomial interpolation as in Fig.2.5 with the *Chebyshev extremal nodes* we obtain an extremely different result in terms of accuracy, as illustrated in Fig.2.7.

The PSM approximate the derivatives of the unknown function with the derivatives of the

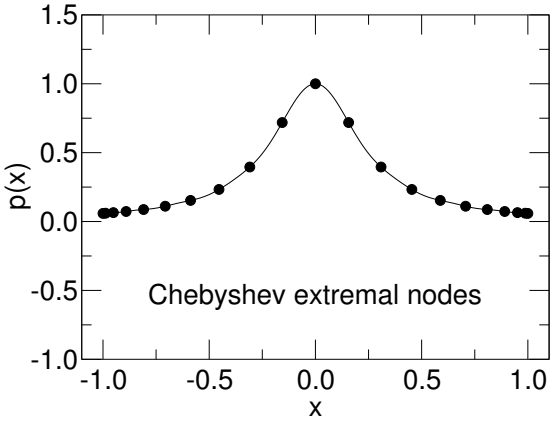


Figure 2.7: Degree N_x interpolation of $u(x) = 1/(1+16x^2)$ on a PSM grid.

approximating polynomial. From Eq.2.25 we thus have

$$\frac{\partial u(x)}{\partial x} \simeq \frac{\partial p(x)}{\partial x} = \sum_{j=0}^{N_x} \frac{\partial l_j(x)}{\partial x}\, u_j, \tag{2.29}$$

$$\frac{\partial^2 u(x)}{\partial x^2} \simeq \frac{\partial^2 p(x)}{\partial x^2} = \sum_{j=0}^{N_x} \frac{\partial^2 l_j(x)}{\partial x^2}\, u_j \tag{2.30}$$

where we have simplified the notation by introducing $u_j = u(x_j)$. It is clear that the first derivative of $p(x)$ at a given point x_j depends on the values u_j of the unknown function in all the points of the discretized domain. This is different with respect to what happens in the 2^{nd} order finite difference case, where the first derivative depends on the value of the function in the two nearest discretization points.

Using Eq.2.29, the first derivative of the unknown function can be approximated in all the discretization points x_j; by doing so we obtain:

$$\mathbf{u'}_{N_x} \simeq \mathbf{D}_{N_x}\, \mathbf{u}_{N_x} \tag{2.31}$$

where $\mathbf{u}_{N_x} = (u(x_0), u(x_1), \ldots, u(x_{N_x}))^t$, $\mathbf{u'}_{N_x} = (u'(x_0), u'(x_1), \ldots, u'(x_{N_x}))^t$ and $\mathbf{D}_{N_x}$ is the so-called *discretization matrix*. $\mathbf{D}_{N_x}$ is a $(N_x+1)\times(N_x+1)$ matrix whose elements are defined as:

$$[\mathbf{D}_{N_x}]_{ij} = \left.\frac{\partial l_i(x)}{\partial x}\right|_{x=x_j}. \tag{2.32}$$

Similar concepts are valid for higher order derivatives. Moreover, it is easy to see that the matrix $\mathbf{D}^{(k)}_{N_x}$ corresponding to the k-th derivative is given by

$$\mathbf{D}^{(k)}_{N_x} = (\mathbf{D}_{N_x})^k \tag{2.33}$$

with $k \leq N_x$.

Application of the PSM to the k·p problem

The matrix formulation of the differentiation rules are very useful for the discretization of a differential problem such as the **k·p** one. If we want to write the Hamiltonian of Eq.2.18 for a given wavevector **k** discretized with the PSM, we first need to define the width of the domain W

and apply Eq.2.28 to find the nodes. Depending on the crystal orientation we will use Eq.2.19, 2.23 or 2.24 and substitute the differential operators with the matrix **D**. Hence, **S** becomes a block matrix. As an example, we re-write the **S** matrix coefficients for the (001) crystal orientation:

$$\begin{aligned}
S_{11} &= \left[\left(L+\frac{\hbar^2}{2m_0}\right)k_x^2+\left(M+\frac{\hbar^2}{2m_0}\right)k_y^2\right]\mathbf{I}-\left(M+\frac{\hbar^2}{2m_0}\right)\mathbf{D}^2\\
S_{22} &= \left[\left(M+\frac{\hbar^2}{2m_0}\right)k_x^2+\left(L+\frac{\hbar^2}{2m_0}\right)k_y^2\right]\mathbf{I}-\left(M+\frac{\hbar^2}{2m_0}\right)\mathbf{D}^2\\
S_{33} &= \left[\left(M+\frac{\hbar^2}{2m_0}\right)k_x^2+\left(M+\frac{\hbar^2}{2m_0}\right)k_y^2\right]\mathbf{I}-\left(L+\frac{\hbar^2}{2m_0}\right)\mathbf{D}^2\\
S_{12} &= S_{21}=Nk_xk_y\mathbf{I}\\
S_{13} &= S_{31}=-i\,Nk_x\mathbf{D}\\
S_{23} &= S_{32}=-i\,Nk_y\mathbf{D}
\end{aligned}$$

where **I** is a $(N_z+1)\times(N_z+1)$ identity matrix. The **S** matrix is a $3(N_z+1)\times 3(N_z+1)$ matrix and, when we account for the spin-orbit interaction, the total Hamiltonian becomes a $6(N_z+1)\times 6(N_z+1)$ matrix. In this respect the wave function components of Eq.2.18 are ordered in the following way:

$$\boldsymbol{\psi}=[\boldsymbol{\psi_1}\ \ \boldsymbol{\psi_2}\ \ \boldsymbol{\psi_3}\ \ \boldsymbol{\psi_4}\ \ \boldsymbol{\psi_5}\ \ \boldsymbol{\psi_6}]^T \tag{2.34}$$

where each $\boldsymbol{\psi_i}$ has N_z+1 elements. We impose Dirichlet boundary conditions at both boundaries of the 1D grid, i.e.

$$\psi_i(0)=\psi_i(W)=0 \tag{2.35}$$

The *Spectral accuracy* allows us to use fewer discretization points while having the same, or even better accuracy with respect to other differentiation schemes, such as the finite difference method (FDM). This has a dramatic impact on the simulation time because the complexity of the eigenvalues calculation increases with the rank of the matrix as $O(N^3)$. After the implementation of PSM we substantially decreased the computational burden of the simulator reducing the number of discretization points N_z with respect to the FDM.

We have solved the $\mathbf{k}\cdot\mathbf{p}$ problem with the PSM and with the 2^{nd} order FDM [17, 18]. Fig.2.8 shows the error of the first eigenvalue at $\mathbf{k}=0$ for a triangular potential $F_z=0.7$MV/cm versus the number of discretization points calculated with both discretization schemes and for two different values of W. The spectral solution with 300 points has been taken as reference for the calculation of the error. It can be seen that, while the finite difference error decreases linearly (in this log-log plot), the PSM related error drops faster reaching the double precision with only 20 points for both values of W. Finally, in Fig.2.9 we reported the squared modulus of the 10^{th} eigenfunction. We easily see that the finite difference solution converges to the spectral one (in this case with only 30 points), but very slowly. The FDM with 1000 points gives a good approximation of the PSM with only 30 points.

2.4 The triangular grid

The valence band is strongly anisotropic in the 2D **k** space, thus we have to solve the eigenvalue problem of Eq.2.18 for each **k** point as sketched in Fig.2.10. This is an issue from a computational point of view, hence we have to find a smart discretization scheme for the **k** space. The **k**·**p** problem has been solved for each point in a triangular grid, as in Fig. 2.11. We employed a non-uniform grid in the $(0\div 0.3)[2\pi/a_0]$ interval. In Fig. 2.11 we used a smaller step size for the modulus of **k** near **k**=0 to limit the total number of points while maintaining a good accuracy [19].

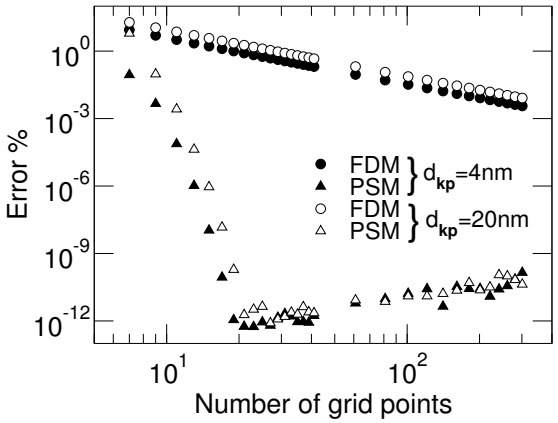


Figure 2.8: Relative error (%) in the calculation of the first ten eigenvalues at $\mathbf{k} = 0$ versus the number of grid points computed with the finite differences (circles) and the spectral methods (triangles) discretization schemes. The spectral solution with 300 points has been taken as reference. The error has been calculated for 2 different widths of the quantum well and for a triangular potential $F_z = 0.7$MV/cm.

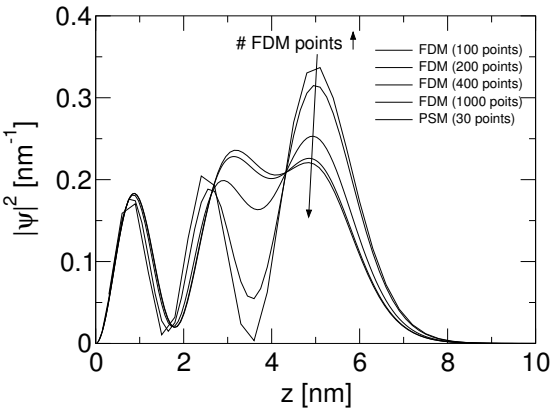


Figure 2.9: Squared magnitude of the tenth wave function calculated with the PSM and with the FDM calculated for a quantum well of W=30nm, k$_x$=k$_y$=0.1$[2\pi/a_0]$ and a triangular confining potential with an electric field F$_z$=1MV/cm. By increasing the number of grid points the FDM solution converges to the PSM one.

Of course the subband structure changes continuously over a single triangle, but we only know the energies at the vertexes. The energy dispersion has been linearly interpolated over each triangle, thus the x and y velocity components are constant within each triangle.

Finally, we exploited the symmetry of the valence band to reduce the region of the $\mathbf{k}$-plane where the band-structure has to be calculated (see Figs.2.1, 2.2 and 2.3): for example, in the (001)-plane and for unstrained silicon we calculated the band-structure in the interval $(0 \div \pi/4)$ as in Fig.2.12(a) (i.e. 1/8 of the entire circumference) and then rotated the bands to obtain the complete energy dispersion. In more general terms, for (001), (110) and (111) crystal orientations the band structure has been calculated between 0 and $2\pi/iw_{coeff}$ (i.e. irreducible wedge coefficient), where iw_{coeff} is equal to 8, 4 and 6, respectively (see Fig.2.12).

The symmetry of the energy dispersion can change not only with the crystallographic orientation but also if we apply a shear stress component in the transport plane. For example, if we consider (001) silicon and we apply a uniaxial strain component along the [110] direction we

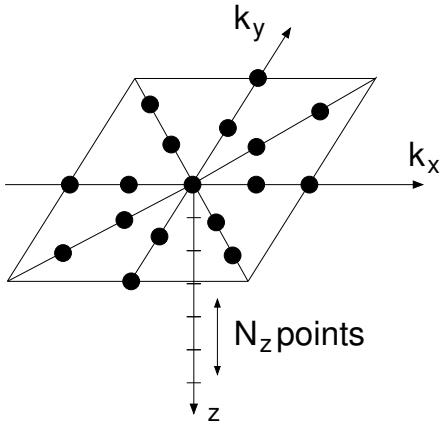


Figure 2.10: Due to the strong anisotropy of the valence band we need to solve Eq.2.18 for each **k** point in a 2D space.

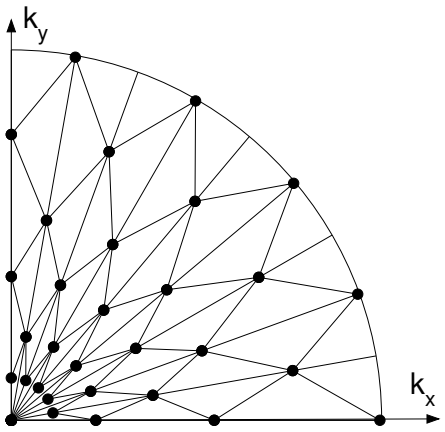


Figure 2.11: Triangular grid used for the discretization of the 2 dimensional **k**-space. For each node a quantized **k**·**p** problem has been solved. The first Brillouin zone has been approximated by a circle and we employed a non-uniform grid in the $[0\ 0.3](2\pi/a_0)$ interval. We used a smaller step size for the modulus of **k** near **k**=0 to limit the total number of points while maintaining a good accuracy

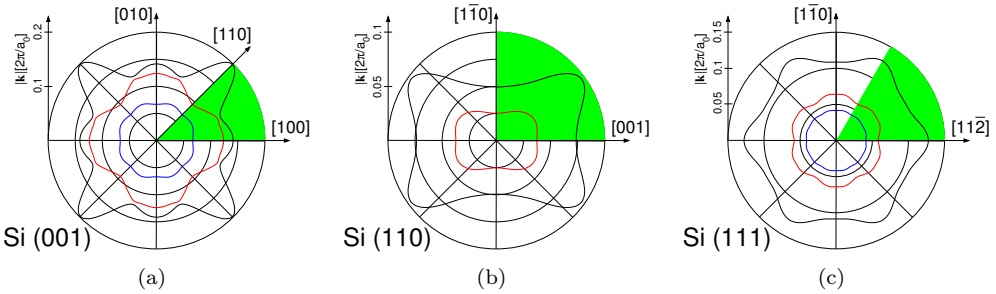


Figure 2.12: Equi-energy curves for the (a) (100), (b) (110) and (c) (111) crystal orientations. The shaded region highlights the portion of the plane where we compute the energy-dispersion. By exploiting the symmetries we generate the band-structure in the remaining part of the plane.

obtain a warped energy dispersion which has $\pi/2$ symmetry (instead of $\pi/4$). In our model, we can account for arbitrary crystal orientations and strain configurations, which are automatically included in the simulation.

2.4.1 Interpolation and integration over k

The bands are calculated on a triangular grid, but we need to interpolate them over each triangle to obtain a continuous function. For simplicity we dropped the subband index in the following equations. The linear interpolation of the energy dispersion over one simplex (i.e. triangle) S can be written as

$$E^S(k_x, k_y) = E_0^S + a_x^S\, k_x + a_y^S\, k_y \tag{2.36}$$

where the coefficients E_0^S, a_x^S and a_y^S are uniquely defined by the vertexes of the simplex S and the associated band energies, as sketched in Fig.2.13.

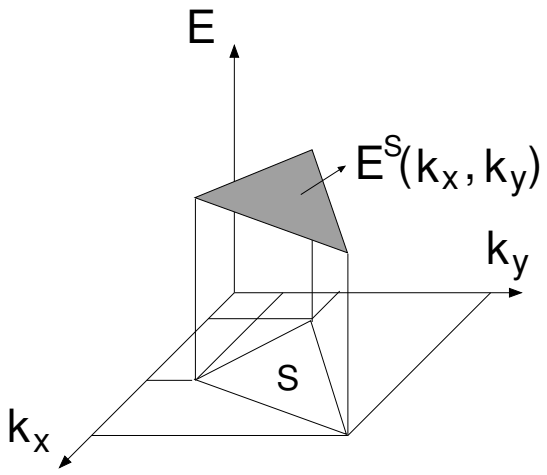


Figure 2.13: Linear interpolation of the energy dispersion over a single triangle S.

Of course, we have one set of coefficients for Eq.2.36 within each simplex and for each subband. Hence we need an ordered data structure that contains the bands and the coefficients for each triangle. From Eq.2.36 we readily see that subband velocity components along x and y are constant for each simplex and equal

$$v_x^S = \frac{a_x^S}{\hbar} \qquad v_y^S = \frac{a_y^S}{\hbar} \tag{2.37}$$

The computation of the hole concentration as well as the implementation of the MRT method require to perform a large number of integrals in the 2 dimensional $\mathbf{k}$ space. The standard approach requires to substitute the sum over all the states with an area integral:

$$\frac{1}{A}\sum_{\mathbf{k}} \longrightarrow \frac{n_{\mathrm{sp}}}{(2\pi)^2}\int \mathrm{d}^2k \tag{2.38}$$

where A is the normalization area and n_{sp} is the spin degeneracy factor. In the case of the $\mathbf{k}\cdot\mathbf{p}$ the states have no spin degeneracy, so n_{sp}=1 and will be dropped in the following of the chapter. In our case we have to integrate over each simplex thus Eq.2.38 becomes

$$\frac{1}{(2\pi)^2}\int \mathrm{d}^2k \longrightarrow \frac{1}{(2\pi)^2}\sum_S \int_S \mathrm{d}^2k \tag{2.39}$$

Since we know the analytical expression for the energy dispersion from Eq.2.36, the integrals can be easily simplified. For example, if we were to integrate a generic function F over a triangle S, Eq.2.39 would read

$$\frac{1}{(2\pi)^2}\sum_S \int_S \mathrm{d}^2k\, F(\mathbf{k}) = \frac{1}{(2\pi)^2}\sum_S \int_{k_{xm}}^{k_{xM}} \mathrm{d}k_x \int_{k_{ym}(k_x)}^{k_{yM}(k_x)} \mathrm{d}k_y\, F(k_x,k_y) \tag{2.40}$$

where k_{xm} and k_{xM} are the minimum and the maximum value of k_x in the S simplex. For each value of k_x we denote the minimum and maximum values of k_y with $k_{ym}(k_x)$ and $k_{yM}(k_x)$, respectively. We are thus transforming an area integral to line integrals and to do so we need to define a discretization for the k_x grid.

Most of the integrals in the $\mathbf{k}$ space are equi-energy, because they contain a delta function that selects only the final states with a given energy value, namely we use

$$\frac{1}{(2\pi)^2}\sum_S \int_S \mathrm{d}^2k\, F(\mathbf{k})\,\delta(E(\mathbf{k})-E) \tag{2.41}$$

In this case it is useful to change one integration variable in order to eliminate one integral. From Eq.2.36 we can write

$$k'_y = \frac{E' - E_0^S - a_x^S k_x}{a_y^S} \tag{2.42}$$

thus Eq.2.41 becomes

$$\begin{aligned}
&\tfrac{1}{(2\pi)^2}\textstyle\sum_S \int_S \mathrm{d}^2k\, F(\mathbf{k})\,\delta(E(\mathbf{k})-E) = \\
&= \tfrac{1}{(2\pi)^2}\textstyle\sum_S \frac{1}{a_y^S}\int \mathrm{d}k_x \int \mathrm{d}E'\, F\Big(k_x, \tfrac{E'-E_0^S-a_x^S k_x^S}{a_y^S}\Big)\,\delta(E'-E) \\
&= \tfrac{1}{(2\pi)^2}\textstyle\sum_S \frac{1}{a_y^S}\int \mathrm{d}k_x\, F\Big(k_x, \tfrac{E-E_0^S-a_x^S k_x}{a_y^S}\Big)
\end{aligned}$$

where we have used the delta function to reduce the second integral.

In the simple case of the density of states DOS the function F is simply 1 and one obtains

$$DOS(E) = \frac{1}{(2\pi)^2}\sum_S \frac{1}{a_y^S}\int_{k_{xm}(E)}^{k_{xM}(E)} \mathrm{d}k_x = \frac{1}{(2\pi)^2}\sum_S \frac{k_{xm}(E)-k_{xM}(E)}{a_y^S} \tag{2.43}$$

In this case we analytically solved the integral, hence we don't need to define a k_x discretization for each simplex.

The eigenfunctions we have calculated with the $\mathbf{k}\cdot\mathbf{p}$ problem are used to compute the hole concentration which can be written as:

$$p(z) = \frac{1}{A}\sum_{\mathbf{k},n} |\psi_{n\mathbf{k}}(z)|^2\, f_0(E_n(\mathbf{k})) \tag{2.44}$$

where f_0 is the Fermi-Dirac occupation function and n is the subband index. Each wave function $\psi_{n\mathbf{k}}(z)$ is normalized to one, i.e.

$$\int_0^W \mathrm{d}z\, |\psi_{n\mathbf{k}}(z)|^2 = 1, \quad \forall n \tag{2.45}$$

Using Eqs.2.38 and 2.40 we can rewrite Eq.2.44 as

$$p(z) = \frac{1}{(2\pi)^2}\sum_{S,n} |\psi_{S,n}(z)|^2 \int_{k_{xm}}^{k_{xM}} \mathrm{d}k_x \int_{k_{ym}(k_x)}^{k_{yM}(k_x)} \mathrm{d}k_y\, \frac{1}{1+\exp\left(\frac{E_n(\mathbf{k})-E_f}{k_B T}\right)} \tag{2.46}$$

Now the wave function depends on the simplex S, in fact we transformed the wave function doing an average of its squared modulus for each simplex and for each subband. Substituting k_y with the corresponding energy as in Eq.2.42 we can calculate the innermost integral obtaining

$$p(z) = \frac{1}{(2\pi)^2} \sum_{S,n} |\psi_{S,n}(z)|^2 \int_{k_{xm}}^{k_{xM}} \frac{\mathrm{d}k_x}{a_y^S} \ln \left[\frac{1 + \exp\left(\frac{E_f - E_{n,m}^S(k_x)}{k_B T} \right)}{1 + \exp\left(\frac{E_f - E_{n,M}^S(k_x)}{k_B T} \right)} \right] \tag{2.47}$$

where

$$\begin{cases} E_{n,m}^S(k_x) = E_{n,0}^S + a_{n,x}^S \, k_x + a_{n,y}^S \, k_{ym} \\ E_{n,M}^S(k_x) = E_{n,0}^S + a_{n,x}^S \, k_x + a_{n,y}^S \, k_{yM} \end{cases} \tag{2.48}$$

2.5 8-bands k·p for direct-bandgap materials

In the last years, we have assisted to a growing interest in the use of direct-bandgap semiconductors such as GaAs, InP or InAs for optoelectronic and logic applications [20, 21]. In order to understand the properties of such materials, it is of utmost importance to correctly model the energy dispersion in low-dimensional structures. In particular, for the applications we are going to discuss in Chapter 6 we need to model the valence and conduction bands by using the same Hamiltonian.

As in the case of the calculation of the valence band of semiconductors, there exist many methods able to correctly reproduce the full band-structure of direct-bandgap materials. Being a good trade-off between accuracy and computation burden, the 8-bands **k·p** represents for us a very convenient choice. Since carrier transport is mainly governed by the modes with eigenenergies close to the conduction and the valence band edges, the **k·p** method is still a very good approximation [22]. The rest of this chapter will be devoted to the formulation we have used to calculate the bands for bulk and low-dimensional systems.

When we include in the summation of Eq.2.9 the bands originating from the top of the valence band and the bottom of the conduction band we can write the 8-bands **k·p** problem for a bulk system as

$$\mathbf{H}_8\,\boldsymbol{\psi}_n = E_n\boldsymbol{\psi}_n \tag{2.49}$$

Following [23], the 8-bands **k·p** Hamiltonian $\mathbf{H}_8$ can be written as

$$\mathbf{H}_8 = \begin{pmatrix} C & 0 & P^+/\sqrt{6} & 0 & P^-/\sqrt{2} & -\sqrt{\frac{2}{3}}P_z & -P_z/\sqrt{3} & P^+/\sqrt{3} \\ & C & -\sqrt{\frac{2}{3}}P_z & -P^+/\sqrt{2} & 0 & -P^-/\sqrt{6} & P^-/\sqrt{3} & P_z/\sqrt{3} \\ & & -P+Q & -S^* & R & 0 & \sqrt{\frac{3}{2}}S & -\sqrt{2}Q \\ & & & -P-Q & 0 & R & -\sqrt{2}R & S/\sqrt{2} \\ & & & & -P-Q & S^* & S^*/\sqrt{2} & \sqrt{2}R^* \\ & & & & & -P+Q & \sqrt{2}Q & \sqrt{\frac{3}{2}}S^* \\ & & & & & & -P-\Delta_{so} & 0 \\ & & & & & & & -P-\Delta_{so} \end{pmatrix} \tag{2.50}$$

where

$$\begin{aligned} C &= E_g + \tfrac{\hbar^2}{2m_0}\,\gamma_c\,\left(k_x^2+k_y^2+k_z^2\right) \\ P &= \tfrac{\hbar^2}{2m_0}\,\gamma_1\,\left(k_x^2+k_y^2+k_z^2\right) \\ Q &= \tfrac{\hbar^2}{2m_0}\,\gamma_2\,\left(k_x^2+k_y^2-2k_z^2\right) \\ R &= -\tfrac{\hbar^2}{2m_0}\sqrt{3}\,\left[\gamma_2(k_x^2-k_y^2)-2i\gamma_3 k_x k_y\right] \\ S &= \tfrac{\hbar^2}{2m_0}\,\gamma_3\,2\sqrt{3}\,(k_x-ik_y)k_z \\ P^{\pm} &= P_0\,(k_x \pm ik_y) \\ P_z &= P_0\,k_z \\ P_0 &= \sqrt{E_p\hbar^2/2m_0} \\ \gamma_c &= \tfrac{1}{m_c^*} - \tfrac{E_p}{3}\left(\tfrac{2}{E_g}+\tfrac{1}{E_g+\Delta_{so}}\right) \end{aligned} \tag{2.51}$$

and the modified Luttinger parameters γ_i are related to the Luttinger parameters γ_i^L as

$$\begin{aligned}\gamma_1 &= \gamma_1^L - \frac{E_p}{3E_g} \\ \gamma_2 &= \gamma_2^L - \frac{E_p}{6E_g} \\ \gamma_3 &= \gamma_3^L - \frac{E_p}{6E_g}\end{aligned} \tag{2.52}$$

It has been recently argued that the spurious solutions of the differential operators obtained by introducing the quantization in the **k·p** Hamiltonians are mainly caused by the possible non-ellipticity of such operators [24]. To eliminate the spurious solutions in our **k·p** problem the parameter E_p controlling the mixing of valence and conduction bands was set to the value recommended in [24], which assures that the Hamiltonian operator be elliptic. All parameters are listed in Tab.2.2 and were taken from [20, 24].

		GaAs	**InP**	**InAs**	**AlAs**
γ_c		2.5634	1.6037	2.2496	1.0188
E_p	[eV]	20.0	16.0	18.0	18.0
γ_1^L		6.98	5.08	20.00	3.76
γ_2^L		2.06	1.60	8.50	0.82
γ_3^L		2.93	2.10	9.20	1.42
m_c^*	$[m_0]$	0.0670	0.0795	0.0260	0.1500
a_0	[nm]	0.5653	0.5870	0.6060	0.5660
E_g	[eV]	1.5190	1.4236	0.4170	3.0990
Δ_{so}	[eV]	0.341	0.108	0.390	0.280

Table 2.2: Band structure parameters for the most common III-V compound semiconductors.

We extensively verified that the model is able to reproduce the results for the most common III-V materials. Fig.2.14 shows the comparison between the bulk bands of GaAs calculated with the 8-bands **k·p** Hamiltonian of Eq.2.50 using the parameters of Tab.2.2 and the results of the Empirical Pseudo-Potential (EPM) method taken from [20]. The agreement between the **k·p** and the EPM results is quite good along different directions in the Brillouin Zone.

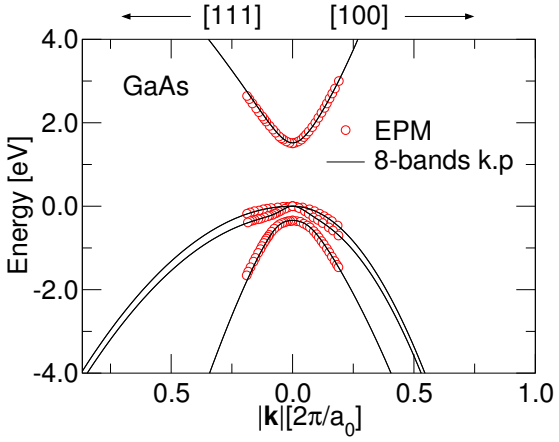


Figure 2.14: Bulk GaAs full-band structure along the [100] and [111] directions. 8-bands **k·p** results (lines) compared with Empirical Pseudo-Potential (EPM) calculations (symbols) taken from [20].

2.5.1 8-bands k·p in presence of quantization

The purpose of our work was to use the 8-bands **k·p** model to simulate electron and hole transport in nano-structures, in particular Gate-All-Around nanowires. The geometrical confinement together with the applied electrostatic potential force us to include the quantization effects in our model. In this section we will focus on the problem of the calculation of the subbands and wave-functions in nanowires with the 8-bands **k·p** model. As in the case of the 6-bands **k·p** model, the quantum mechanical ansatz $k_i \rightarrow -\mathrm{i}\,\partial/\partial i$ leads to a set of eigenvalue differential equation, but in this case the rank of the eigenvalue problem is bigger.

For an infinitely-long nanowire with a potential energy $U(y,z)$ in the section, the 8-bands **k·p** problem reads

$$\left[\hat{\mathbf{H}}_8\left(k_x, -\mathrm{i}\frac{\partial}{\partial y}, -\mathrm{i}\frac{\partial}{\partial z}\right) + \mathbf{I}\,U(y,z)\right]\boldsymbol{\psi}_{nk_x}(y,z) = E_n(k_x)\,\boldsymbol{\psi}_{nk_x}(y,z) \tag{2.53}$$

where $\boldsymbol{\psi}_{nk_x}$ is the n-th 8-component wave-function. When we discretize Eq.2.53 with the method already seen in Section 2.3, we substitute each derivative with the corresponding differentiation matrix. If we denote with N_y and N_z the number of discretization points along the y and z directions, respectively, we thus end up with a $8(N_y+1)(N_z+1) \times 8(N_y+1)(N_z+1)$ matrix. We have to to calculate the eigenvalues and the eigenvectors of the discretized 2-D Hamiltonian for each section of the device, which is too CPU and memory demanding with a standard finite difference discretization. Hence, to speed up the determination of modes, the 2-D Schrödinger equation was solved by using a basis of sine functions; this reduces remarkably the size of the problem in Eq.2.53 and corresponds to a transformation of the Hamiltonian through the unitary matrices given by the basis of sine functions in the squared device section. The details of this calculation are given in App.B.

Before using the discretized Hamiltonians in the transport model described in Chapter 6, we extensively verified the formulations by calculating the subbands and the eigenfunctions for confined systems.

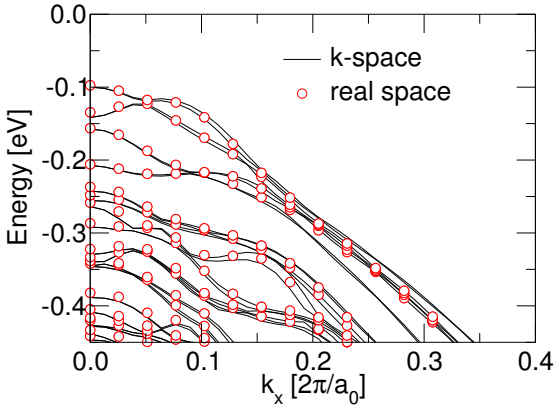
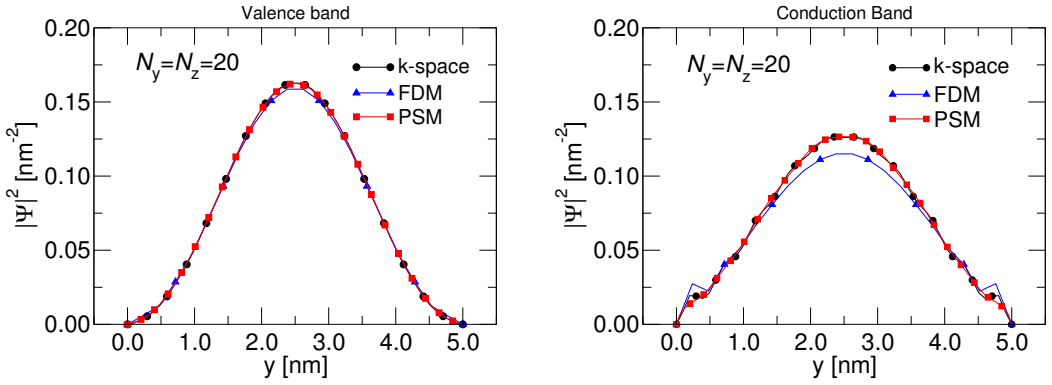


Figure 2.15: Valence subband structure for a 5nm-wide GaAs square nanowire along the [100] direction calculated using the real-space (symbols) and the k-space transformations (lines) for a flat potential. The labels real space and k-space refer to the solution of Eq.2.53 with a standard finite difference discretization scheme or by using the basis of sine functions, respectively.

Fig.2.15 illustrates the valence subband structure for a square nanowire of GaAs, comparing the real-space (symbols) and the k-space solutions (lines). We verified that 10 sine basis functions are enough in order to reproduce the subbands in the presence of quantization, while we need at least 20 points along y and z directions in real space. Fig.2.16 shows the squared magnitude of the fundamental valence and conduction band wave functions for the same GaAs

nanowire. The wave functions have been calculated either in k-space or in real space using the standard finite difference discretization scheme and the pseudo-spectral methods (see Section 2.3). The different approaches used to calculate the band-structure agree quite well but, in general, the real space methods require a larger number of points and thus a larger eigenvalue problem.



(a) (b)

Figure 2.16: Squared magnitude of the first valence band (left) and conduction band (right) wave functions for a 5nm-wide GaAs square nanowire at k_x=0. The wire is infinite along the transport direction and the applied potential is flat. The eigenfunctions have been calculated with the finite difference method (FDM), with the pseudo-spectral method (PSM) and with the k-space transformations.

This analysis has been carried out for different materials and strain conditions. In all cases we found that the k-space solution outperforms the real-space discretization in terms of matrix size and thus of CPU time.

The geometrical quantization plays an important role in extremely narrow nanowires because it increases the energy band gap. Fig.2.17 shows the band edges for a GaAs nanowire as a function of the nanowire width W. It can be seen that the geometric quantization increases the band gap value for W smaller than approximately 10nm. This could represent a problem for tunneling-based devices. We will see in Section 7.2.2 that the drain current is substantially reduced in very narrow nanowires.

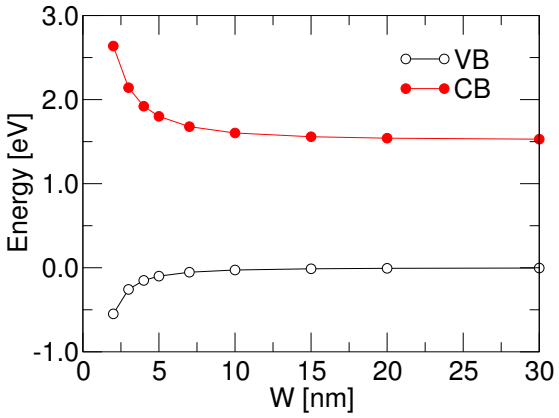


Figure 2.17: 5nm-wide GaAs square nanowire. Highest valence and lowest conduction subbands as a function of nanowire width W.

Bibliography

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

[2] M. M. Rieger and P. Vogl, "Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates," *Phys. Rev. B*, vol. 48, pp. 14 276–14 287, Nov 1993.

[3] G. L. Bir and G. E. Pikus, *Symmetry and strain induced effects in semiconductors.* Wiley, 1974.

[4] S. Richard, F. Aniel, and G. Fishman, "Band diagrams of Si and Ge quantum wells via the 30-band $\mathbf{k} \cdot \mathbf{p}$ method," *Phys. Rev. B*, vol. 72, p. 245316, Dec 2005.

[5] 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 $\mathbf{k} \cdot \mathbf{p}$ method," *Phys. Rev. B*, vol. 74, p. 195208, Nov 2006.

[6] T. Manku and A. Nathan, "Valence energy-band structure for strained group-IV semiconductors," *Journal of Applied Physics*, vol. 73, no. 3, pp. 1205–1213, 1993.

[7] D. Miller, *Quantum Mechanics for Scientists and Engineers.* Cambridge, 2008.

[8] E.O. Kane, "Energy band structure in p-type germanium and silicon," *Journal of Phys. Chem. Solids*, vol. 1, pp. 82–99, 1956.

[9] G. Dresselhaus, A. F. Kip, and C. Kittel, "Cyclotron Resonance of Electrons and Holes in Silicon and Germanium Crystals," *Phys. Rev.*, vol. 98, pp. 368–384, Apr 1955.

[10] P. Marconcini and M. Macucci, "The $\mathbf{k} \cdot \mathbf{p}$ method and its application to graphene, carbon nanotubes and graphene nanoribbons: the Dirac equation," *La Rivista del Nuovo Cimento*, vol. 34, pp. 489–584, 2011.

[11] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors: Physics and Materials Properties.* Berlin: Springer, 2005.

[12] M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, "Six-band kp calculation of the hole mobility in silicon inversion layers: Dependence on surface orientation, strain, and silicon thickness," *Journal of Applied Physics*, vol. 94, no. 2, pp. 1079–1095, 2003.

[13] M. V. Fischetti and S. E. Laux, "Band structure deformation potentials, and carrier mobility in strained Si,Ge and SiGe alloys," *Journal of Applied Physics*, vol. 80, no. 4, pp. 2234–2252, 1996.

[14] J. M. Hinckley and J. Singh, "Monte Carlo studies of ohmic hole mobility in silicon and germanium: Examination of the optical phonon deformation potential," *Journal of Applied Physics*, vol. 76, no. 7, pp. 4192–4200, 1994.

[15] S. Datta, *Quantum Transport - Atom to Transistor*. United Kingdom: Cambridge University Press, 2005.

[16] L. N. Trefethen, *Spectral Methods in MATLAB*. Philadelphia: PA:SIAM, 2000.

[17] A. Paussa, F. Conzatti, D. Breda, R. Vermiglio and D.Esseni, "Pseudo-Spectral Method for the Modelling of Quantization Effects in Nanoscale MOS Transistors," in *Proc.SISPAD*, 2010, pp. 299–302.

[18] A. Paussa, F. Conzatti, D. Breda, R. Vermiglio, D. Esseni and P. Palestri, "Pseudo-Spectral Method for the Efficient Simulation of Quantization Effects in Nanoscale MOS Transistors," *IEEE Trans. Electron Devices*, vol. 57, no. 12, pp. 3239–3249, 2010.

[19] L. Donetti, F. Gamiz, A. Godoy, and N. Rodriguez, "Fully self-consistent $\mathbf{k} \cdot \mathbf{p}$ solver and monte carlo simulator for hole inversion layers," in *Proc. European Solid State Device Res. Conf.*, 2008, pp. 254–257.

[20] 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.

[21] S. Datta, A. Ali, S. Mookerjea, V. Saripalli, L. Liu, S. Eachempati, T. Mayer, and V. Narayanan, "Non-silicon logic elements on silicon for extreme voltage scaling," in *Silicon Nanoelectronics Workshop (SNW), 2010*, june 2010, pp. 1 –2.

[22] M. Shin, "Full-quantum simulation of hole transport and band-to-band tunneling in nanowires using the k·p method," *Journal of Applied Physics*, vol. 106, no. 5, p. 054505, 2009.

[23] T. B. Bahder, "Eight-band k·p model of strained zinc-blende crystals," *Phys. Rev. B*, vol. 41, no. 17, pp. 11 992–12 001, Jun 1990.

[24] R. G. Veprek, S. Steiger, and B. Witzigmann, "Ellipticity and the spurious solution problem of k·p envelope equations," *Phys. Rev. B*, vol. 76, no. 16, p. 165320, Oct 2007.

Chapter 3

Strain Engineering

3.1 Introduction to the strain fabrication techniques

In the past few years strain has represented one of the most important means to enhance the performance of MOS transistors [1]. Strain has been introduced for the first time in an industrial 90nm CMOS technology by Intel [2] and, starting from that moment, a lot of effort was devoted to optimize the strained devices on the technological side and to understand the effects of strain on the modeling side.

The strain affects the device behavior in many ways. First of all, it has a direct impact on the carrier transport in the device channel. Besides, other effects such as threshold voltage shift, leakage of source-drain junctions, gate leakage current and reliability are affected by the strain.

Mechanical strain translates in a small change of the atomic lattice constant. Since all electrical and optical properties depend on the inter-atomic distance, strain has a great impact on device behavior. Nowadays engineers are able to produce strained devices in a controlled and stable way and over large manufacturing scale, so that a large bulk of experimental results is also available.

This chapter is devoted to the strain technology. The first part briefly covers the existing fabrication techniques to induce global and local strain. Section 3.2 introduces the theoretical framework used to model the strain within the elastic approximation. Finally, Section 3.3 shows how we can include the strain effects in the **k·p** formalism and gives a few examples of strained-band structure results.

3.1.1 Global strain techniques: the biaxial strain

The first experimental demonstrations where strain was exploited to enhance the carrier mobility date back to the early 90s. The first papers regarding strain in MOS devices involved biaxially strained silicon layers epitaxially grown on relaxed SiGe virtual substrates [3, 4]. This is a global strain technique because the deformation is induced over a large surface, which often corresponds to the entire wafer. We say that the silicon layer is biaxially strained because the same strain is imposed along the two axes perpendicular to the growth direction.

As explained in Fig.3.1, the biaxial strain is induced by the lattice mismatch between the silicon layer and the SiGe virtual substrate. The latter is much thicker and features a larger lattice constant. When the silicon is grown on top of the SiGe, the silicon atoms are displaced and the layer is under mechanical strain. By changing the mole fraction of Ge the magnitude of the strain can be varied. The dual condition is also possible: a strained SiGe epitaxially grown on top of a silicon virtual substrate. In this case the sign of the strain is opposite, namely the strain is compressive.

The main disadvantage of the biaxial strain is that it induces the same strain configuration for all devices. This represents a serious limitation because n- and p-MOS transistors are affected in different ways by the strain. For example, a compressive biaxial strain increases the hole mobility but, at the same time, decreases electron mobility.

3.1.2 Local strain techniques: the uniaxial strain

In the past few years, a lot of effort has been devoted to develop manufacture process steps able to induce a local strain inside the channel of a single transistor. The two main techniques we mention here are the Contact-Etch-Stop-Liners (CESL) [5, 6] and the SiGe or SiC source-drain stressors [7, 8, 9]. Both techniques can be used to induce tensile or compressive strain inside the channel. As we will see in Chapter 5, tensile and compressive strain are favorable for n- and p-MOS transistors, respectively. The local strain techniques induce almost uniaxial strain configurations, i.e. there is a dominant strain component along a particular direction.

Fig.3.2 shows a sketch of a silicon transistor with SiGe embedded in the source-drain regions. This configuration induces a compressive uniaxial strain inside the channel and thus is favorable

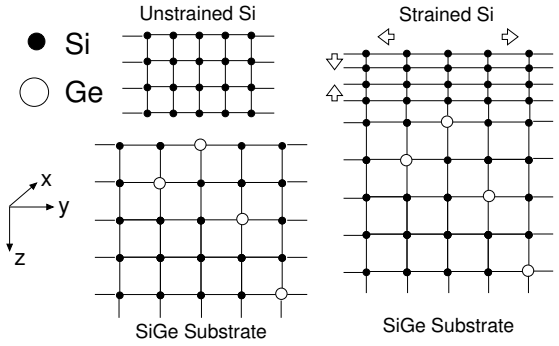


Figure 3.1: Left: relaxed Si and SiGe lattices. Right: biaxially strained silicon grown on a SiGe virtual substrate. The arrows show the strain directions.

for p-type transistors.

The main drawback of local strain techniques is that the strain levels are usually lower with respect to global techniques, so the performance enhancements are reduced.

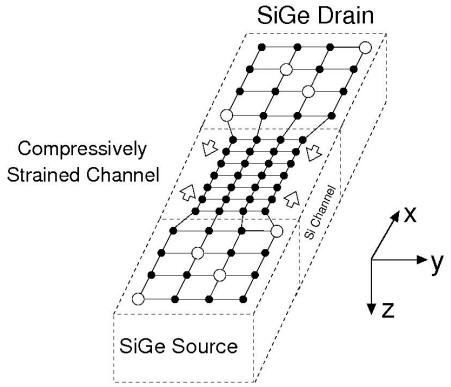


Figure 3.2: Sketch of the strain produced SiGe source-drain stressors in the channel of a MOS transistor.

3.2 Elastic deformation of a cubic crystal

This section presents the mathematical framework for the description of the effect of strain in solids, with a particular emphasis on crystalline semiconductors. We treat the material as a continuous medium rather than a lattice of atoms. By doing so and limiting our analysis to small deformations, we can use elastic theory to couple stress and strain. The notations used here for stress and strain are consistent with [10].

3.2.1 Stress and strain: definitions and relations

Stress

According to the theory of elasticity, the deformation is produced by internal forces within a body. To describe these forces we use the stress tensor $\mathbf{T}$ as illustrated in Fig.3.3, where the components of the full stress tensor have been sketched as forces per unit area (pressure) acting on an infinitesimal cube. For example, T_{xx} is the pressure applied normally on the x-face, T_{xy} is the pressure on the x-face applied along the y-direction and so on. The sign of the axial

stress is positive if the force is oriented outward from the inside of the cube and vice-versa. By doing so, tensile and compressive stress have positive and negative sign, respectively.

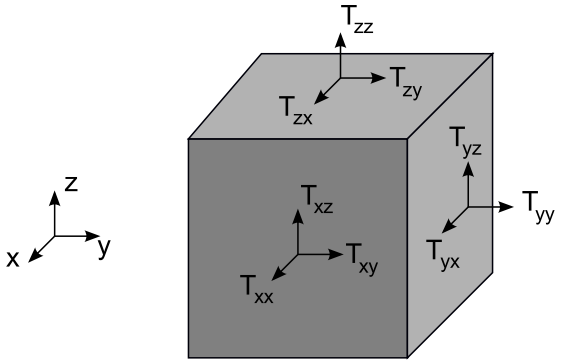


Figure 3.3: Components of the stress tensor.

The stress tensor has in general 9 components which can be organized in a matrix:

$$\mathbf{T} = \begin{pmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{pmatrix} \tag{3.1}$$

where T_{ii} are the axial components and T_{ij} are the shear components (with i=x,y,z and $j \neq i$). The equilibrium condition requires that

$$T_{xy} = T_{yx}; \quad T_{yz} = T_{zy}; \quad T_{xz} = T_{zx} \tag{3.2}$$

Hence, the tensor in Eq.3.1 is symmetric and can be written as a column vector with 6 independent components

$$\mathbf{T} = \begin{pmatrix} T_{xx} \\ T_{yy} \\ T_{zz} \\ T_{yz} \\ T_{xz} \\ T_{xy} \end{pmatrix} = \begin{pmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \end{pmatrix} \tag{3.3}$$

Strain

The strain in solids causes the change of the position of each unit cell. We start by defining a coordinate system where the origin is the position of a single atom of the unstrained material and the sample point is denoted with $\mathbf{r}$=(x,y,z). When strain is applied, both the origin and the sample point are displaced, thus the new position for the sample point $\mathbf{r}'$ can be written as

$$\mathbf{r}' = \mathbf{r} + \mathbf{u}(\mathbf{r}) \tag{3.4}$$

where we have introduced the vector $\mathbf{u}(\mathbf{r})$, which represents the strain induced relative displacement for the point $\mathbf{r}$ under weak strain. We assume that $\mathbf{u}(\mathbf{0})$=$\mathbf{0}$. For small strain values, the

displacement $\mathbf{u}(\mathbf{r})$ can be approximated in terms of Taylor expansion:

$$\begin{aligned}
\begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} &\simeq \begin{pmatrix} \frac{\partial u_x}{\partial x} & \frac{\partial u_x}{\partial y} & \frac{\partial u_x}{\partial z} \\ \frac{\partial u_y}{\partial x} & \frac{\partial u_y}{\partial y} & \frac{\partial u_y}{\partial z} \\ \frac{\partial u_z}{\partial x} & \frac{\partial u_z}{\partial y} & \frac{\partial u_z}{\partial z} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \\
&= \begin{bmatrix} \frac{\partial u_x}{\partial x} & \frac{1}{2}\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right) & \frac{\partial u_y}{\partial y} & \frac{1}{2}\left(\frac{\partial u_y}{\partial z} + \frac{\partial u_z}{\partial y}\right) \\ \frac{1}{2}\left(\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial u_y}{\partial z} + \frac{\partial u_z}{\partial y}\right) & \frac{\partial u_z}{\partial z} \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \\
&+ \begin{bmatrix} 0 & \frac{1}{2}\left(\frac{\partial u_x}{\partial y} - \frac{\partial u_y}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial u_x}{\partial z} - \frac{\partial u_z}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial u_x}{\partial y} - \frac{\partial u_y}{\partial x}\right) & 0 & \frac{1}{2}\left(\frac{\partial u_y}{\partial z} - \frac{\partial u_z}{\partial y}\right) \\ \frac{1}{2}\left(\frac{\partial u_x}{\partial z} - \frac{\partial u_z}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial u_y}{\partial z} - \frac{\partial u_z}{\partial y}\right) & 0 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}
\end{aligned} \tag{3.5}$$

If there is no rotational motion, the second term in Eq.3.5 must be zero. Hence, we can write the displacement as

$$\begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{xz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{3.6}$$

where we have defined the strain tensor

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{xz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{pmatrix} \tag{3.7}$$

As it can be inferred from the definitions, the strain represents the relative lattice deformation of a solid under a weak mechanical stress and is a unitless quantity. As the stress tensor $\mathbf{T}$, the strain $\boldsymbol{\varepsilon}$ is symmetric, i.e.

$$\varepsilon_{xy} = \varepsilon_{yx}; \quad \varepsilon_{yz} = \varepsilon_{zy}; \quad \varepsilon_{xz} = \varepsilon_{zx} \ , \tag{3.8}$$

and can be written as a 6-components column vector

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \varepsilon_{yz} \\ \varepsilon_{xz} \\ \varepsilon_{xy} \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix} \tag{3.9}$$

The vector notation is useful when we write the equations relating stress and strain.

3.2.2 Relations between stress and strain

A solid crystal is generally anisotropic, thus the relation between stress and strain is a tensorial equation

$$\varepsilon_{ij} = \sum_{k,l}^{3} S_{ijkl}\, T_{kl} \tag{3.10}$$

where $\boldsymbol{\varepsilon}$ and $\mathbf{T}$ are 3×3 matrices and $\mathbf{S}$ is the fourth-order compliance tensor. We have already seen that at the equilibrium we can write the stress and the strain as 6-components column vectors, hence Eq.3.10 is re-written as:

$$\varepsilon_i = \sum_j^3 S_{ij}\, T_j \tag{3.11}$$

The symmetries of cubic semiconductors (like silicon, germanium and zinc-blende III-V materials) allow to express the compliance matrix $\mathbf{S}$ in terms of only three independent compliance constants, thus the matrix relation between stress and strain reads

$$\begin{pmatrix} \varepsilon_{c,xx} \\ \varepsilon_{c,yy} \\ \varepsilon_{c,zz} \\ \varepsilon_{c,yz} \\ \varepsilon_{c,xz} \\ \varepsilon_{c,xy} \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} & S_{12} & 0 & 0 & 0 \\ S_{12} & S_{11} & S_{12} & 0 & 0 & 0 \\ S_{12} & S_{12} & S_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{44}/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & S_{44}/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & S_{44}/2 \end{pmatrix} \begin{pmatrix} T_{c,xx} \\ T_{c,yy} \\ T_{c,zz} \\ T_{c,yz} \\ T_{c,xz} \\ T_{c,xy} \end{pmatrix} \tag{3.12}$$

Eq.3.12 holds if the stress and strain are written in the Crystal Coordinate System (CCS), and for this reason we have introduced the symbols $T_{c,ij}$ and $\varepsilon_{c,ij}$.

The inverse relation of Eq.3.12 reads

$$\begin{pmatrix} T_{c,xx} \\ T_{c,yy} \\ T_{c,zz} \\ T_{c,yz} \\ T_{c,xz} \\ T_{c,xy} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & 2C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & 2C_{44} \end{pmatrix} \begin{pmatrix} \varepsilon_{c,xx} \\ \varepsilon_{c,yy} \\ \varepsilon_{c,zz} \\ \varepsilon_{c,yz} \\ \varepsilon_{c,xz} \\ \varepsilon_{c,xy} \end{pmatrix} \tag{3.13}$$

where C_{ij} are the elements of the elastic stiffness matrix $\mathbf{C}$. The compliance and stiffness parameters are related by

$$\begin{aligned} S_{11} &= \frac{C_{11}+C_{12}}{(C_{11}-C_{12})(C_{11}+2C_{12})} \\ -\frac{S_{12}}{S_{11}} &= \frac{C_{12}}{C_{11}+C_{12}} \\ S_{44} &= \frac{1}{C_{44}} \end{aligned} \tag{3.14}$$

The numerical values for S_{ij} and C_{ij} can be found in Tab.3.1. We emphasize that Eqs.3.12 and 3.13 are valid if the stress and the strain are written in the CCS. In the following section we will see what happens if this condition is not true, i.e. how we can generalize the elastic law for general stress and strain conditions.

3.2.3 Coordinate transformation for tensors of the second rank

If we have a general stress configuration and we want to calculate the corresponding strain condition, a more general expression than Eq.3.12 is needed. We need to know how a tensor of the second rank (matrices $\mathbf{S}$ or $\mathbf{C}$ in our case) transforms under a rotation of the coordinate axes. In order to do so, we need to define two coordinate systems: the Crystal Coordinate System (CCS) and the Device Coordinate System (DCS) (see Fig.3.4), which differ by a rotation.

A general vector $\mathbf{r}$, which is represented by the triplet (x_c,y_c,z_c) in the CCS, can be expressed in the DCS by using the rotation matrix $\mathbf{R}_{C\to D}$ from the CCS to the DCS defined as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{R}_{C\to D} \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} \tag{3.15}$$

	Si	Ge	GaAs	InP	InAs
C_{11} [10^{10}Pa]	16.6	12.6	11.88	10.22	8.33
C_{12} [10^{10}Pa]	6.4	4.4	5.38	5.73	4.53
C_{44} [10^{10}Pa]	7.96	6.77	5.94	4.42	3.96
S_{11} [10^{-12}Pa^{-1}]	7.67	9.69	11.73	16.39	19.45
S_{12} [10^{-12}Pa^{-1}]	-2.13	-2.51	-3.66	-5.89	-6.85
S_{44} [10^{-11}Pa^{-1}]	1.26	1.48	1.69	2.26	2.53

Table 3.1: Elastic compliance and stiffness parameters for Si, Ge (from [10]) and some III-V compounds (from [11]).

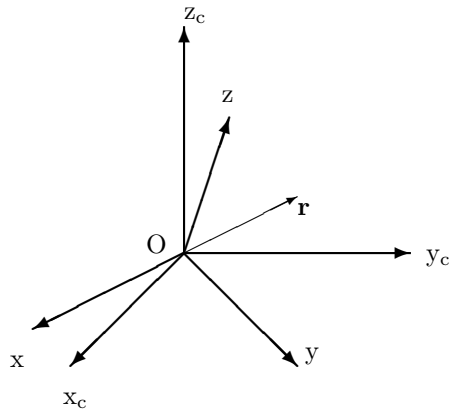


Figure 3.4: Sketch of the Crystal Coordinate System (CCS) (x_c,y_c,z_c) and Device Coordinate System (DCS) (x,y,z).

where

$$\mathbf{R}_{C\to D} = \begin{pmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \end{pmatrix} \tag{3.16}$$

The columns of $\mathbf{R}_{C\to D}$ are the components, expressed in the DCS, of the three unit vectors in the CCS. The inverse relation reads

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \mathbf{R}_{D\to C} \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} \tag{3.17}$$

where

$$\mathbf{R}_{D\to C} = \mathbf{R}^{-1}_{C\to D} \tag{3.18}$$

Since $\mathbf{R}_{C\to D}$ is a unitary matrix, then

$$\mathbf{R}_{D\to C} = \mathbf{R}^{t}_{C\to D} \tag{3.19}$$

where the superscript t denotes the transpose operation.

It can be verified that any matrix $\mathbf{A}_c$ representing a linear relation in the CCS can be represented in the DCS (where we call it $\mathbf{A}$) via the transformation

$$\mathbf{A} = \mathbf{R}_{C\to D}\, \mathbf{A}_c\, \mathbf{R}^{t}_{C\to D} \tag{3.20}$$

The strain $\boldsymbol{\varepsilon}_c$ and the stress $\mathbf{T}_c$ in the CCS follow the same rule:

$$\begin{aligned}\boldsymbol{\varepsilon} &= \mathbf{R}_{C\to D}\,\boldsymbol{\varepsilon}_c\,\mathbf{R}^t_{C\to D}\\ \mathbf{T} &= \mathbf{R}_{C\to D}\,\mathbf{T}_c\,\mathbf{R}^t_{C\to D}\end{aligned} \tag{3.21}$$

If we combine Eqs.3.21 and 3.16 and we use the vector notation for stress and strain (see Eq.3.9) we can write

$$\begin{pmatrix}\varepsilon_{xx}\\ \varepsilon_{yy}\\ \varepsilon_{zz}\\ \varepsilon_{yz}\\ \varepsilon_{xz}\\ \varepsilon_{xy}\end{pmatrix} = \begin{pmatrix} l_1^2 & m_1^2 & n_1^2 & 2m_1n_1 & 2n_1l_1 & 2l_1m_1\\ l_2^2 & m_2^2 & n_2^2 & 2m_2n_2 & 2n_2l_2 & 2l_2m_2\\ l_3^2 & m_3^2 & n_3^2 & 2m_3n_3 & 2n_3l_3 & 2l_3m_3\\ l_2l_3 & m_2m_3 & n_2n_3 & m_2n_3+m_3n_2 & n_2l_3+n_3l_2 & m_2l_3+m_3l_2\\ l_3l_1 & m_3m_1 & n_3n_1 & m_3n_1+m_1n_3 & n_3l_1+n_1l_3 & m_3l_1+m_1l_3\\ l_1l_2 & m_1m_2 & n_1n_2 & m_1n_2+m_2n_1 & n_1l_2+n_2l_1 & m_1l_2+m_2l_1\end{pmatrix}\begin{pmatrix}\varepsilon_{c,xx}\\ \varepsilon_{c,yy}\\ \varepsilon_{c,zz}\\ \varepsilon_{c,yz}\\ \varepsilon_{c,xz}\\ \varepsilon_{c,xy}\end{pmatrix} \tag{3.22}$$

which can be written in a compact way as

$$\boldsymbol{\varepsilon} = \mathbf{R}_{6,C\to D}\,\boldsymbol{\varepsilon}_c \tag{3.23}$$

where

$$\mathbf{R}_{6,C\to D} = \begin{pmatrix} l_1^2 & m_1^2 & n_1^2 & 2m_1n_1 & 2n_1l_1 & 2l_1m_1\\ l_2^2 & m_2^2 & n_2^2 & 2m_2n_2 & 2n_2l_2 & 2l_2m_2\\ l_3^2 & m_3^2 & n_3^2 & 2m_3n_3 & 2n_3l_3 & 2l_3m_3\\ l_2l_3 & m_2m_3 & n_2n_3 & m_2n_3+m_3n_2 & n_2l_3+n_3l_2 & m_2l_3+m_3l_2\\ l_3l_1 & m_3m_1 & n_3n_1 & m_3n_1+m_1n_3 & n_3l_1+n_1l_3 & m_3l_1+m_1l_3\\ l_1l_2 & m_1m_2 & n_1n_2 & m_1n_2+m_2n_1 & n_1l_2+n_2l_1 & m_1l_2+m_2l_1\end{pmatrix} \tag{3.24}$$

The inverse relation reads

$$\boldsymbol{\varepsilon}_c = \mathbf{R}^{-1}_{6,C\to D}\,\boldsymbol{\varepsilon} \tag{3.25}$$

Similar expressions exist for the stress $\mathbf{T}$, that is we have

$$\mathbf{T} = \mathbf{R}_{6,C\to D}\,\mathbf{T}_c \tag{3.26}$$

and

$$\mathbf{T}_c = \mathbf{R}^{-1}_{6,C\to D}\,\mathbf{T} \tag{3.27}$$

When we simulate a device in presence of strain, we often know the stress $\mathbf{T}$ in the DCS. In order to calculate the strained band-structure, however, we need the strain in the CCS $\boldsymbol{\varepsilon}_c$ (see Section 3.3). To this purpose we can combine Eqs.3.12 and 3.27 to obtain

$$\boldsymbol{\varepsilon}_c = \mathbf{S}\,\mathbf{R}^{-1}_{6,C\to D}\,\mathbf{T} \tag{3.28}$$

As an example, we consider the presence of a uniaxial stress equal to $T_{xx} = \sigma$ along the [110]/(001) direction. The stress in the DCS is

$$\mathbf{T} = \begin{pmatrix}T_{xx}\\ T_{yy}\\ T_{zz}\\ T_{yz}\\ T_{xz}\\ T_{xy}\end{pmatrix} = \begin{pmatrix}\sigma\\ 0\\ 0\\ 0\\ 0\\ 0\end{pmatrix} \tag{3.29}$$

and the rotation matrix $\mathbf{R}_{C\to D}$ from the CCS to the DCS reads

$$\mathbf{R}_{C\to D} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0\\ -1/\sqrt{2} & 1/\sqrt{2} & 0\\ 0 & 0 & 1\end{pmatrix} \tag{3.30}$$

We can readily calculate $\mathbf{R}^{-1}_{6,C\to D}$:

$$\mathbf{R}^{-1}_{6,C\to D} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & -1 \\ 1/2 & 1/2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/2 & -1/2 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.31}$$

According to Eq.3.27 the stress in the CCS is equal to

$$\mathbf{T}_c = \mathbf{R}^{-1}_{6,C\to D}\,\mathbf{T} = \begin{pmatrix} \sigma/2 \\ \sigma/2 \\ 0 \\ 0 \\ 0 \\ \sigma/2 \end{pmatrix} \tag{3.32}$$

Now we can use Eq.3.12 to calculate the strain in the CCS:

$$\begin{aligned} \boldsymbol{\varepsilon}_c = \mathbf{S}\;\mathbf{T}_c \quad &= \quad \begin{pmatrix} S_{11} & S_{12} & S_{12} & 0 & 0 & 0 \\ S_{12} & S_{11} & S_{12} & 0 & 0 & 0 \\ S_{12} & S_{12} & S_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & S_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & S_{44} \end{pmatrix} \begin{pmatrix} \sigma/2 \\ \sigma/2 \\ 0 \\ 0 \\ 0 \\ \sigma/2 \end{pmatrix} \\ &= \quad \begin{pmatrix} (S_{11}+S_{12})\,\sigma/2 \\ (S_{11}+S_{12})\,\sigma/2 \\ S_{12}\,\sigma \\ 0 \\ 0 \\ S_{44}/2\,\sigma \end{pmatrix} = \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \\ \epsilon_{yz} \\ \epsilon_{xz} \\ \epsilon_{xy} \end{pmatrix} \end{aligned} \tag{3.33}$$

Further examples can be found in [10].

3.3 Inclusion of strain in the k·p method

3.3.1 Strained 6-bands k·p Hamiltonian

The effect of strain can be naturally included in the 6-bands **k·p** model for the calculation of the band-structure [12, 13]. In particular, strain can be included in the total Hamiltonian by modifying Eq.2.18 as

$$\left[\mathbf{H}\left(\mathbf{k}, -\mathrm{i}\frac{\partial}{\partial z}\right) + \mathbf{H}_\varepsilon + \mathbf{I}\,U(z)\right]\psi_{\mathbf{k}}(z) = E(\mathbf{k})\,\psi_{\mathbf{k}}(z) \tag{3.34}$$

where $\mathbf{H}_\varepsilon$ is the strain operator, namely a **k**-independent 6x6 matrix. This operator is given by [14]

$$\mathbf{H}_\varepsilon = \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \tag{3.35}$$

where $\mathbf{D}$ is the 3×3 matrix

$$\mathbf{D} = \begin{pmatrix} l\varepsilon_{c,xx} + m(\varepsilon_{c,yy} + \varepsilon_{c,zz}) & n\varepsilon_{c,xy} & n\varepsilon_{c,xz} \\ n\varepsilon_{c,xy} & l\varepsilon_{c,yy} + m(\varepsilon_{c,xx} + \varepsilon_{c,zz}) & n\varepsilon_{c,yz} \\ n\varepsilon_{c,xz} & n\varepsilon_{c,yz} & l\varepsilon_{c,zz} + m(\varepsilon_{c,xx} + \varepsilon_{c,yy}) \end{pmatrix} \tag{3.36}$$

and $\varepsilon_{c,ij}$ are the components of the strain tensor in the CCS. The deformation potentials l, m, n have been calibrated by comparison to the results of pseudo-potential calculations and the values for silicon and germanium are reported in Tab.2.1. It has been demonstrated that the **k·p** model is quite accurate for the calculation of band-structure in either strained materials or quantized systems by comparing the calculated band-structure with the results of a full-band model [15] (namely the Linear Combination of Bulk Bands).

Examples of strained band-structure

The first step in order to assess the potentials of strain in semiconductor devices is the calculation of the strained energy dispersion. In the following we present a few examples of strained valence band-structures calculated for different strain conditions and surface orientations.

Fig.3.5 shows the impact of biaxial strain on the lowest subband in Si for different values of the Ge mole fraction. This type of strain does not change the in-plane symmetry, but produces a remarkable deformation of the band-structure. In particular, if we increase the biaxial tensile strain by augmenting the Ge content of the virtual substrate, the curvature and consequently the effective mass are reduced.

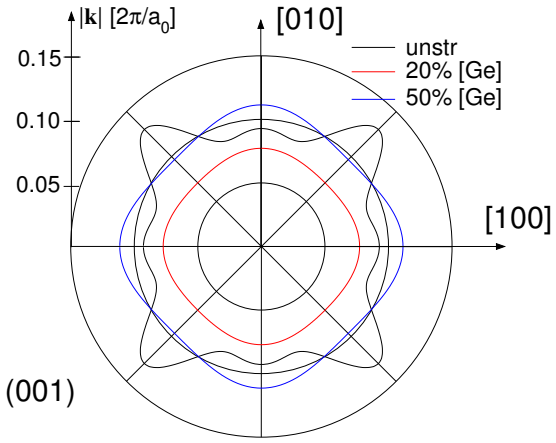


Figure 3.5: (001) hole inversion layer. Equi-energy curves for the lowest subband at 25meV from the subband minimum. Solid line: unstrained. Dashed line: 20% Ge. Dotted-dashed: 50% Ge. Triangular well with a confining field F_c =0.7 MV/cm.

Fig.3.6 compares the impact of [110]/(001) uniaxial strain on the electrons and hole equi-energy lines. The stress is tensile and compressive for n- and p-MOSFETs, respectively. The conduction band curves have been calculated according to the analytical expressions of [16]. As it can be seen, the effect of band warping is much more pronounced in the case of holes, for the same stress magnitude of 1GPa. Loosely speaking, we may say that the compressive [110] stress reduces significantly the hole effective transport mass, that is related to the magnitude k of the wavevector in the [110] direction of the plot. The reduction of the electron transport mass obtained with the tensile [110] stress is only modest compared to what observed for holes. It should be noted that, contrary to the biaxial stress configuration of Fig.3.5, the uniaxial strain changes the in-plane symmetry.

Fig.3.7 reports a self-consistent band structure calculated for an unstrained and uniaxially strained p-type FinFET (see Section 5.2), where the quantization direction is perpendicular to the (110) plane. Also in this case the uniaxial strain induces a strong deformation of the hole band-structure for both compressive and tensile stress. The **k·p** model developed in this work is able to take into account for arbitrary crystal orientations and strain configurations by calculating the band-structure symmetries and exploiting them by limiting the computation of

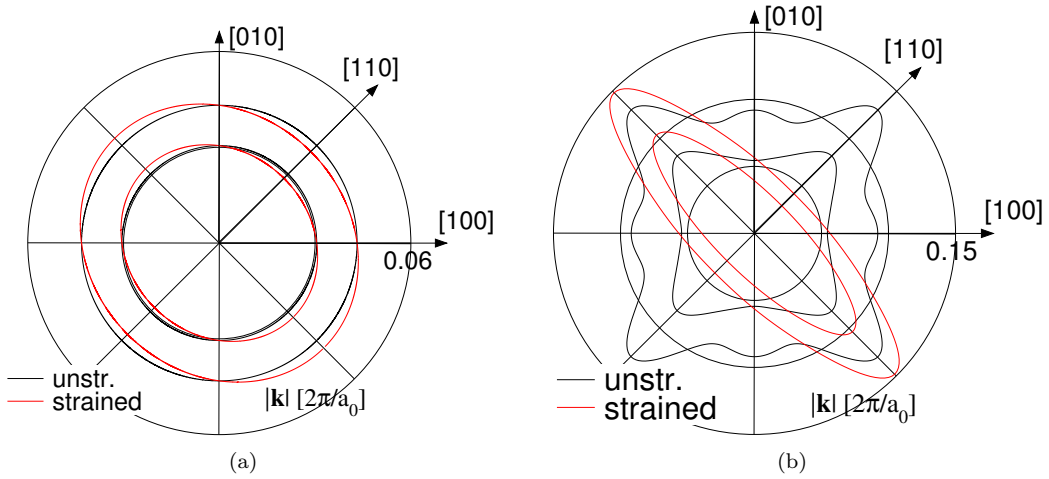


Figure 3.6: (a) (001) electron inversion layer and (b) (001) hole inversion layer. Equi-energy curves for the lowest subband at an energy 25 meV and 50 meV above the subband minimum. Solid line: unstrained. Dashed line: 1GPa tensile/compressive stress along the [110] direction for (a)/(b). The stress yields a strong deformation of the band-structure. A triangular well with a confining field F_c =0.7 MV/cm is imposed in both (a) and (b) cases.

the bands to the smallest possible portion of the 2D **k**-space plane from which the complete energy dispersion can be obtained by simple symmetry transformations (see Fig.2.12).

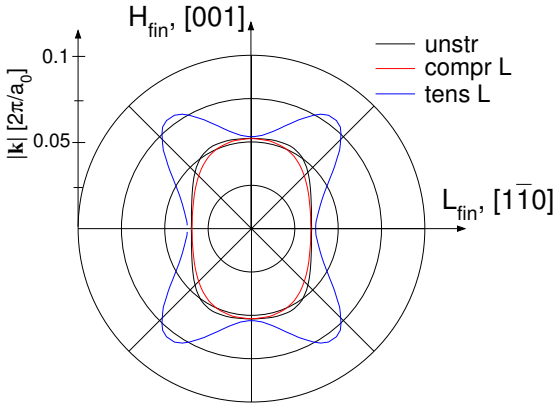


Figure 3.7: Self-consistent simulation of a p-type FinFET. Hole equi-energy contour plots for the lowest subband at 100meV from the subband minimum for unstrained or strained (110) inversion layer. Tensile or compressive uniaxial stress in the [1$\bar{1}$0] direction ±1.5GPa. The inversion hole density is P_{inv}=8·10^{12} cm^{-2}.

As a final remark, it should be noted that is not so trivial to examine these results in terms of insightful quantities as effective masses or energy shifts (as already done for the conduction band in [16]). This is caused by the strong band warping that makes it problematic to write simple analytical expression for the band edges and effective masses. What can be done is to calculate a sort of *mobility effective mass* for each subband, which is weighted over the in-plane angle and depends on the energy [17]. This information together with the subband occupation helps understand the effects of strain and has been applied in the case of uniaxial strain in (001) hole inversion layers in [18].

3.3.2 Strained 8-bands k·p Hamiltonian

When we include the strain in the 8-bands **k·p** theory, Eq.2.49 must be modified as

$$(\mathbf{H}_8 + \mathbf{H}_{8,\varepsilon})\,\boldsymbol{\psi}_n = E_n \boldsymbol{\psi}_n \tag{3.37}$$

Following [19], the Hermitian 8×8 strain interaction matrix $\mathbf{H}_{8,\varepsilon}$ entering Eq.3.37 reads

$$\mathbf{H}_{8,\varepsilon} = \begin{pmatrix} a_c\,e & 0 & -v^* & 0 & -\sqrt{3}\,v & \sqrt{2}\,u & u & -\sqrt{2}\,v^* \\ & a_c\,e & \sqrt{2}\,u & -\sqrt{3}\,v^* & 0 & v & -\sqrt{2}\,v & -u \\ & & -p+q & -s^* & r & 0 & \sqrt{3/2}\,s & -\sqrt{2}\,q \\ & & & -p-q & 0 & r & -\sqrt{2}\,r & s/\sqrt{2} \\ & & & & -p-q & s^* & s^*/\sqrt{2} & \sqrt{2}\,r^* \\ & & & & & -p+q & \sqrt{2}\,q & \sqrt{3/2}\,s \\ & & & & & & -a_v\,e & 0 \\ & & & & & & & -a_v\,e \end{pmatrix} \tag{3.38}$$

where

$$\begin{aligned} e &= \varepsilon_{c,xx} + \varepsilon_{c,yy} + \varepsilon_{c,zz} p = a_v\,e \\ q &= b\,[\varepsilon_{c,zz} - (\varepsilon_{c,xx} + \varepsilon_{c,yy})/2] \\ u &= \tfrac{P_0}{\sqrt{3}}(\varepsilon_{c,xz}\,k_x + \varepsilon_{c,yz}\,k_y + \varepsilon_{c,zz}\,k_z) \\ v &= \tfrac{P_0}{\sqrt{6}}\,[(\varepsilon_{c,xx} - \mathrm{i}\,\varepsilon_{c,xy})\,k_x + (\varepsilon_{c,xy} - \mathrm{i}\,\varepsilon_{c,yy})\,k_y + (\varepsilon_{c,xz} - \mathrm{i}\,\varepsilon_{c,yz})\,k_z] \\ r &= \tfrac{\sqrt{3}}{2}\,b\,(\varepsilon_{c,xx} - \varepsilon_{c,yy}) - \mathrm{i}\,d\,\varepsilon_{c,xy} \\ s &= -d\,(\varepsilon_{c,xz} - \mathrm{i}\,\varepsilon_{c,yz}) \end{aligned} \tag{3.39}$$

and $\varepsilon_{c,ij}$ is the general component of the strain tensor in the CCS. The strain deformation potentials for some III-V compound semiconductors are listed in Tab. 3.2. It should be noted that, contrary to the 6-bands **k·p** case, the strain interaction matrix $\mathbf{H}_{8,\varepsilon}$ depends on the wave-vector components k_i. However, the inclusion of strain in the **k·p** method does not increase the computational burden because the rank of the matrices does not change.

	a_c	a_v	b	d
GaAs	-7.17	-1.16	-2.0	-4.8
InP	-6.0	-0.6	-2.0	-5.0
InAs	-5.08	-1.00	-1.8	-3.6

Table 3.2: Strain deformation potentials entering Eqs.3.38 and 3.39 (in [eV]) taken from [20].

The examples of strained band-structure calculated with the 8-band **k·p** model will be presented in Chapter 7, where we will see how the strain affects the performances of InAs nanowire Tunnel-FETs designed for low power consumption applications.

Bibliography

[1] S. Thompson, S. Suthram, Y. Sun, S. Parthasarathy, M. Chu, and T. Nishida, "Future of Strained Si/Semiconductors in Nanoscale MOSFETs," in *IEEE IEDM Technical Digest*, 2006.

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

[3] J. Welser, J. Hoyt, and J. Gibbons, "NMOS and PMOS transistors fabricated in strained silicon/relaxed silicon-germanium structures," in *IEEE IEDM Technical Digest*, dec 1992, pp. 1000 –1002.

[4] J. Welser, J. Hoyt, S. Takagi, and J. Gibbons, "Strain dependence of the performance enhancement in strained-Si n-MOSFETs," in *IEEE IEDM Technical Digest*, dec 1994, pp. 373 –376.

[5] S. Ito, H. Namba, K. Yamaguchi, T. Hirata, K. Ando, S. Koyama, S. Kuroki, N. Ikezawa, T. Suzuki, T. Saitoh, and T. Horiuchi, "Mechanical stress effect of etch-stop nitride and its impact on deep submicron transistor design," in *IEEE IEDM Technical Digest*, 2000, pp. 247–251.

[6] H.S. Yang, R. Malik, S. Narasimha, Y. Li, R. Divakaruni, P. Agnello, S. Allen, A. Antreasyan, J.C. Arnold, K. Bandy *et al.*, "Dual stress liner for high performance sub-45nm gate length SOI CMOS manufacturing," in *IEEE IEDM Technical Digest*, 2004, p. 1075 V1077.

[7] S. Eneman, P. Verheyen, R. Rooyackers, F. Nouri, L. Washington, R. Degraeve, B. Kaczer, V. Moroz, A. De Keersgieter, R. Schreutelkamp, M. Kawaguchi, Y. Kim, A. Samoilov, L. Smith, P. P. Absil, K. De Meyer, M. Jurczak, S. Biesemans, "Layout impact on the performance of a locally strained p-MOSFET," in *IEEE Symposium on VLSI Technology - Technical Digest*, 2005, pp. 22–23.

[8] M. Horstmann, A. Wei, T. Kammler, J. Höntschel, H. Bierstedt, T. Feudel, K. Frohberg, M. Gerhardt, A. Hellmich, K. Hempel, J. Hohage, P. Javorka, J. Klais, G. Koerner, M. Lenski, A. Neu, R. Otterbach, P. Press, C. Reichel, M.Trentsch, B. Trui, H. Salz, M. Schaller, H.-J. Engelmann, O. Herzog, H. Ruelke, P. Hübler, R. Stephan, D. Greenlaw, M. Raab, N. Kepler, "Integration and optimization of embedded-SiGe, compressive and tensile stressed liner films, and stress memorization in advanced SOI CMOS technologies," in *IEEE IEDM Technical Digest*, 2005, pp. 233–236.

[9] K.-W. Ang, K.J. Chui, V. Bliznetsov, A. Du, N. Balasubramanian, M.F. Li, G. Samudra, Y.-C. Yeo, "Enhanced performance in 50 nm n-MOSFETs with silicon-carbon source/drain regions," in *IEEE IEDM Technical Digest*, 2004, pp. 1069–1071.

[10] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS transistors: Semi-classical modeling and applications*. United Kingdom: Cambridge University Press, 2010.

[11] S. Adachi, *Properties of Semiconductor Alloys*. Wiley, 2009.

[12] T. Manku and A. Nathan, "Valence energy-band structure for strained group-IV semiconductors," *Journal of Applied Physics*, vol. 73, no. 3, pp. 1205–1213, 1993.

[13] M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, "Six-band kp calculation of the hole mobility in silicon inversion layers: Dependence on surface orientation, strain, and silicon thickness," *Journal of Applied Physics*, vol. 94, no. 2, pp. 1079–1095, 2003.

[14] J. M. Hinckley and J. Singh, "Monte Carlo studies of ohmic hole mobility in silicon and germanium: Examination of the optical phonon deformation potential," *Journal of Applied Physics*, vol. 76, no. 7, pp. 4192–4200, 1994.

[15] V. Sverdlov, D. Esseni, O. Baumgartner, F. Schanovsky, H. Kosina, and S. Selberherr, "The Linear Combination of Bulk Bands-Method for Electron and Hole Subband Calculations in Strained Silicon Films and Surface Layers," in *International Workshop on Computational Electronics*, 2009, pp. 49–52.

[16] E. Ungersboeck, S. Dhar, G. Karlowatz, V. Sverdlov, H. Kosina, and S. Selberherr, "The effect of general strain on the band structure and electron mobility of silicon," *IEEE Transactions on Electron Devices*, vol. 54, no. 9, pp. 2183–2190, 2007.

[17] F.Conzatti, M. Michielis, D.Esseni, and P.Palestri, "Drain Current Improvements in Uniaxially Strained p-MOSFETs: A Multi-Subband Monte Carlo Study," in *Proc. European Solid State Device Res. Conf.*, 2008, p. 250.

[18] F. Conzatti, M. De Michielis, D. Esseni, and P. Palestri, "Drain Current Improvements in Uniaxially Strained p-MOSFETs: A Multi-Subband Monte Carlo Study," *Solid State Electronics*, vol. 53, pp. 706–711, 2009.

[19] T. B. Bahder, "Eight-band k·p model of strained zinc-blende crystals," *Phys. Rev. B*, vol. 41, no. 17, pp. 11 992–12 001, Jun 1990.

[20] 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.

Chapter 4

Semi-classical transport approach

This chapter is devoted to the description of the semi-classical approach with a special emphasis on the momentum relaxation time approximation for holes.

4.1 Introduction to the semi-classical approach

THE high electric fields present in modern nanoelectronic devices and the reduced dimensions push the carriers far from equilibrium, hence we need a model capable to handle any type of transport condition, from equilibrium to far-from-equilibrium. Many approaches have been developed in the past years to simulate carrier transport in nanodevices. Among them, we mention here the full-quantum approaches, capable to take into account the wave-nature of charged particles. One popular full-quantum method is the so called Non-Equilibrium Green's Function formalism [1], which will be treated in detail in Chapter 6. A full-quantum approach is often very expensive from the computational point of view, hence semi-classical models have been successfully employed over the years.

In the semi-classical approach [2, 3, 4] each carrier is treated like an electrically charged classical particle which obeys the Newton and Maxwell laws of classical physics. It is called semi-classical because quantum mechanics enters the calculation of the subbands and the expressions for scattering rates, i.e. when we model the interaction between the carrier and phonons, impurities or spatial roughness of the interfaces.

A necessary condition for the semi-classical approach to be valid is that the potentials induced from the outside be slowly-varying compared to the crystal potential and to the thermal wavelength of the carriers $\lambda = \hbar/\sqrt{2\pi m^* k_B T}$. Moreover, the mean-free-path between consecutive collisions must be much longer than the De Broglie wavelength λ_{DB}, because the carriers must have a well-defined energy. Finally, since scattering events are described as instantaneous events and are localized in space, there is a lower limit to the time scale and to the device dimensions we can simulate [5].

Section 4.2 introduces the basic equation of the semi-classical picture, namely the Boltzmann Transport Equation (BTE). In the same section of this chapter we treat the theory of the Momentum Relaxation Time (MRT), which is an approximation of the BTE and the method we use to compute low-field mobility. Section 4.3 presents the problem of the self-consistency together with the Poisson equation, which controls the electrostatics inside the device. Section 4.4 covers the theory of scattering, with a special focus on the expressions we derived for hole inversion layers in the MRT framework. Finally, Section 4.5 presents the calibration results for both unstrained and strained bulk p-MOSFETs.

4.2 The Boltzmann Transport Equation (BTE)

In order to describe the device behavior within the semiclassical framework, the most important quantity we need to know is the *distribution function* $f(\mathbf{r}, \mathbf{k}, t)$, which represents the generalization of the Fermi-Dirac distribution for an out-of-equilibrium carrier gas. $f(\mathbf{r}, \mathbf{k}, t)$ expresses the probability of occupation of a state in the $(\mathbf{r}, \mathbf{k})$-space (namely the phase space), that is the probability that the state in $\mathbf{r}$ with wave-vector $\mathbf{k}$ is filled at the time instant t. Once we know the distribution function we can calculate the charge, the average carrier velocities, the current density and the average kinetic energy of the carriers.

The distribution function $f(\mathbf{r}, \mathbf{k}, t)$ satisfies the Boltzmann Transport Equation (BTE) which, for a 3D carrier gas, reads

$$\frac{\partial f(\mathbf{r}, \mathbf{k}, t)}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{k}, t) + e\mathbf{F}(\mathbf{r}) \cdot \frac{1}{\hbar} \nabla_{\mathbf{k}} f(\mathbf{r}, \mathbf{k}, t) = \left. \frac{\partial f(\mathbf{r}, \mathbf{k}, t)}{\partial t} \right|_{scatt} \tag{4.1}$$

where $\mathbf{v}$ is the carrier velocity, e is the electron charge, $\mathbf{F}$ is the electric field and $\partial f(\mathbf{r}, \mathbf{k}, t)/\partial t|_{scatt}$ is the rate of variation of the distribution function related to the scattering mechanisms. The

BTE can be viewed as a balance of particle fluxes in the phase space where the distribution is different from the equilibrium-one because of a non-uniform distribution function in real-space $\nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{k}, t)$, or due to an applied external field $\mathbf{F}$ or any scattering event.

One of the most popular methods to solve exactly the BTE is the Monte-Carlo method (see for instance [6, 7, 8]), which is a widely used technique to solve integro-differential equations. When applied to the BTE, the basic algorithm consists in the simulation of a wide number of classical particles' trajectories, which are stopped by random scattering events. The detailed treatment of the BTE and of the Monte Carlo method goes beyond the scope of this thesis; interested readers can find further details in [2, 3, 9].

4.2.1 The BTE in inversion layers

When we model the carrier transport in inversion layers, where we cannot neglect the presence of subbands, Eq.4.1 cannot be used in a straightforward manner. The dimensionality of the problem changes, thus the 3D vectors $\mathbf{r}$ and $\mathbf{k}$ of Eq.4.1 become 2D vectors, and, moreover, the state of a carrier is described by the subband index i. When dealing with multiple valley systems, we have to consider also the valley index, which complicates somewhat the picture and the notation. In the case of hole inversion layers modeled with the $\mathbf{k}\cdot\mathbf{p}$ approximation all relevant subbands are at the Γ point, so, in what follows, the valley index will not be included in the notation.

A basic assumption of the semi-classical approach is that a particle can change subband only due to scattering events [10]. This is not the case for full-quantum approaches, as we will see in Chapter 6.

The BTE for an inversion layer reads

$$\frac{\partial f_i(\mathbf{r}, \mathbf{k}, t)}{\partial t} + \mathbf{v}_g \cdot \nabla_{\mathbf{r}} f_i(\mathbf{r}, \mathbf{k}, t) - \nabla_{\mathbf{k}} f_i(\mathbf{r}, \mathbf{k}, t) \cdot \frac{1}{\hbar} \nabla_{\mathbf{r}} \varepsilon_i = \left. \frac{\partial f_i(\mathbf{r}, \mathbf{k}, t)}{\partial t} \right|_{scatt} \tag{4.2}$$

where $f_i(\mathbf{r}, \mathbf{k}, t)$ is the probability to find a particle in the position $\mathbf{r}$, with 2D wave-vector $\mathbf{k}$ in the subband i and at time t. If we can separate the $\mathbf{r}$ and $\mathbf{k}$ dependencies of the energy of the energy eigenvalue $E_i(\mathbf{r}, \mathbf{k})$ and write it as the sum of a parallel (in-plane) energy $E_p(\mathbf{k})$ and a subband energy $\varepsilon_i(\mathbf{r})$ [3], i.e.

$$E_i(\mathbf{r}, \mathbf{k}) = E_p(\mathbf{k}) + \varepsilon_i(\mathbf{r}) \quad , \tag{4.3}$$

the group velocity $\mathbf{v}_{\mathrm{g}}$ can be written as as

$$\mathbf{v}_{\mathrm{g}} = \frac{1}{\hbar} \nabla_{\mathbf{k}} E_i(\mathbf{k}) \tag{4.4}$$

The gradient of ε_i in Eq.4.2 is the driving force:

$$\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = -\frac{1}{\hbar} \nabla_{\mathbf{r}} \varepsilon_i(\mathbf{r}) \tag{4.5}$$

For the case of non-parabolic band structures is not possible to separate the $\mathbf{r}$ and $\mathbf{k}$ dependencies of the energy and thus writing Eq.4.3, and the definition of $E_i(\mathbf{r}, \mathbf{k})$ is in general not additive.

The most complicated part of the BTE is represented by the scattering mechanisms, which enter the right-hand side of Eq.4.2. We can re-write the scattering term as the difference between the incoming flux S_{in} and the out-coming one S_{out}:

$$\left. \frac{\partial f_i(\mathbf{r}, \mathbf{k}, t)}{\partial t} \right|_{scatt} = S_{in} - S_{out} \tag{4.6}$$

The out-coming flux reads:

$$S_{out} = f_i(\mathbf{r}, \mathbf{k}, t) \sum_{\mathbf{k}', i'} S_{i,i'}(\mathbf{k}, \mathbf{k}', \mathbf{r})[1 - f_{i'}(\mathbf{r}, \mathbf{k}', t)] \tag{4.7}$$

where $S_{i,i'}(\mathbf{k},\mathbf{k}',\mathbf{r})$ is the scattering rate from state $\mathbf{k}$ in subband i to state $\mathbf{k}'$ in subband i' at the position $\mathbf{r}$ [3]. The in-coming flux reads:

$$S_{in} = \sum_{\mathbf{k}',i'} f_{i'}(\mathbf{r},\mathbf{k}',t) S_{i',i}(\mathbf{k}',\mathbf{k},\mathbf{r})[1 - f_i(\mathbf{r},\mathbf{k},t)] \tag{4.8}$$

Starting from this derivations, the next section presents the method of the Momentum Relaxation Time approximation to solve the BTE for small displacements from the equilibrium and for a uniform transport condition.

4.2.2 Momentum Relaxation Time (MRT) approximation

The MRT method is a widely used approximation of the BTE to calculate carrier mobility in low-dimensional as well as bulk systems. In this work we will discuss only the MRT for hole inversion layers. The MRT is valid only for small displacements from the equilibrium and when for a uniform transport condition, hence when the distribution function f_i, the scattering rates $S_{i,i'}$ and all macroscopic quantities do not depend on the position $\mathbf{r}$.

Let us start by writing the distribution function as the equilibrium one plus a small displacement, i.e.

$$f_i(\mathbf{k}) = f_0(E_i(\mathbf{k})) + f_i^1(\mathbf{k}) \tag{4.9}$$

where f_0 is the equilibrium Fermi-Dirac distribution function, which reads

$$f_0(E) = \frac{1}{1 + \exp\left(\frac{E-E_F}{k_B T}\right)} \tag{4.10}$$

and $f_i^1(\mathbf{k})$ represents the difference between the actual distribution function and the equilibrium one. $E_i(\mathbf{k})$ represents the total energy corresponding to the wave-vector $\mathbf{k}$, E_F is the Fermi energy, k_B is the Boltzmann constant and T is the temperature.

If we now consider the uniform transport condition, i.e. $\nabla_{\mathbf{r}} f$=0, and we suppose that the gas is subject to a small electric field $\mathbf{F}$=$F_d\,\hat{\mathbf{e}}_d$ (where $\hat{\mathbf{e}}_d$ is the unit vector along the direction d=x or y), we can re-write $f_i^1(\mathbf{k})$ as [3]

$$\begin{aligned} f_i^1(\mathbf{k}) &= e\,\tau_{i,d}(\mathbf{k})\,\mathbf{F}\cdot\mathbf{v}_i(\mathbf{k})\,\frac{\partial f_0(E_i(\mathbf{k}))}{\partial E} \\ &= e\,\tau_{i,d}(\mathbf{k})\,F_d\,v_{i,d}(\mathbf{k})\,\frac{\partial f_0(E_i(\mathbf{k}))}{\partial E}, \quad d \in \{x,y\} \end{aligned} \tag{4.11}$$

where $v_{i,d}$ is the group velocity component along the direction d in the subband i and F_d is the electric field component along d. Eq.4.11 defines the quantity $\tau_{i,d}(\mathbf{k})$, which is the relaxation time along the direction d and in the subband i.

Starting from Eq.4.11 and after some manipulations [3], it can be demonstrated that the right-hand side of Eq.4.6 can be expressed in terms of the relaxation time $\tau_{i,d}(\mathbf{k})$ as

$$S_{in} - S_{out} = -\frac{f_i(\mathbf{k}) - f_0(E_i(\mathbf{k}))}{\tau_{i,d}(\mathbf{k})} = -\frac{f_i^1(\mathbf{k})}{\tau_{i,d}(\mathbf{k})} \tag{4.12}$$

If the electric field F is zero, we can combine Eqs.4.2 and 4.12 to write:

$$\frac{\partial f_i(\mathbf{k},t)}{\partial t} = -\frac{f_i(\mathbf{k}) - f_0(E_i(\mathbf{k}))}{\tau_{i,d}(\mathbf{k})} \tag{4.13}$$

which tells us that the relaxation time $\tau_{i,d}(\mathbf{k})$ is the time constant for the system to decay to its equilibrium condition thanks to scattering events. The larger $\tau_{i,d}(\mathbf{k})$ the more time the system needs to recover equilibrium condition, thus the weaker the scattering and vice-versa.

From Eq.4.11 we infer that the problem of the calculation of low-field mobility can be divided in two fundamental steps: the first step consists in calculating the relaxation time $\tau_{i,d}(\mathbf{k})$ and then in computing the mobility. The remaining part of this section gives the mathematical formulations used to calculate both quantities, starting from the relaxation time.

Relaxation time calculation

It can be demonstrated that, starting from Eqs.4.6–4.7 and Eq.4.11, with some manipulations and in the spirit of small deviations from equilibrium, the total relaxation time $\tau_{i,d}$ along d and in subband i reads [3, 11]

$$\frac{1}{\tau_{i,d}(\mathbf{k})} = \sum_{j,\mathbf{k}'} S_{i,j}(\mathbf{k},\mathbf{k}') \, \frac{1-f_0(E_j(\mathbf{k}'))}{1-f_0(E_i(\mathbf{k}))} \, \left[1 - \frac{\tau_{j,d}(\mathbf{k}')v_{j,d}(\mathbf{k}')}{\tau_{i,d}(\mathbf{k})v_{i,d}(\mathbf{k})}\right] \tag{4.14}$$

where, according to Fermi's golden rule, the scattering rate is written as

$$S_{i,j}(\mathbf{k},\mathbf{k}') = \frac{2\pi}{\hbar}|M_{i,j}(\mathbf{q})|^2 \, \delta[E_j(\mathbf{k}') - E_i(\mathbf{k}) - \Delta E(\mathbf{q})] \tag{4.15}$$

where $M_{i,j}$ is the matrix element between the subbands i and j for an exchanged wavevector $\mathbf{q} = (\mathbf{k}' - \mathbf{k})$ and ΔE is the energy exchanged during the scattering event, which can be either positive (i.e. absorption of inelastic phonons) or negative (emission of inelastic phonons).

Eq.4.14 does not represent an explicit expression for $\tau_{i,d}$ because the momentum relaxation time appears also in the r.h.s. This complicates the calculation of the relaxation time and forces us to solve a problem where the relaxation times in all subbands are coupled by Eq.4.14.

For isotropic scattering mechanisms such as electron-phonon scattering, however, it can be demonstrated [3] that the term in square brackets in Eq.4.14 simplifies to 1. Hence, by combining Eqs.4.14 and 4.15 the relaxation time for a generic isotropic scattering mechanism can be written as

$$\frac{1}{\tau_{i,d}(\mathbf{k})} = \frac{2\pi}{\hbar} \sum_{j,\mathbf{k}'} |M_{i,j}(\mathbf{q})|^2 \, \frac{1-f_0(E_j(\mathbf{k}'))}{1-f_0(E_i(\mathbf{k}))} \, \delta[E_j(\mathbf{k}') - E_i(\mathbf{k}) - \Delta E_0] \tag{4.16}$$

where ΔE_0 is the phonon energy, and the term $v_{j,d}(\mathbf{k}')/v_{i,d}(\mathbf{k})$ reduces to zero when we perform the sum over $\mathbf{k}'$.

For anisotropic scattering mechanisms, the matrix element and the exchanged energy depend on the the exchanged wavevector $\mathbf{q}$, thus we cannot use Eq.4.16. For scattering with surface roughness and ionized impurities it is widely accepted to treat as elastic ($E(\mathbf{k}') = E(\mathbf{k})$) and include only intra-subband scattering events (i=j). The reader can find in Section 4.4.2 the final formulation for surface roughness relaxation time.

Section 4.4 presents the expression we used for the calculation of the relaxation time for the case of hole inversion layers with the $\mathbf{k}\cdot\mathbf{p}$ model and for both acoustic and optical phonons and surface roughness scattering mechanisms.

Low field mobility

In general, the distribution function $f_i^1(\mathbf{k})$ can produce non-zero current density components both along x and y directions. For a 2-D carrier gas we say that either the current density $J_{i,d}$ or the relaxation time $\tau_{i,d}$ can have non-zero x and y components, hence the low-field mobility, which relates the current density and the electric field F_d, is a 2×2 matrix:

$$\begin{aligned} J_{i,x} &= e\,P_i\,(\mu_{i,xx}F_x + \mu_{i,xy}F_y) \\ J_{i,y} &= e\,P_i\,(\mu_{i,yx}F_x + \mu_{i,yy}F_y) \end{aligned} \tag{4.17}$$

where P_i is the inversion hole density in the subband i. We re-write the first equation of 4.17 as

$$J_{i,x} = \frac{e}{A} \sum_{\mathbf{k}} v_{i,x}(\mathbf{k}) \, f_i^1(\mathbf{k}) \tag{4.18}$$

By substituting Eq.4.11 into Eq.4.18 we can readily calculate the xx and xy hole mobility components in the subband i as

$$\begin{aligned}\mu_{i,xx} &= \tfrac{e}{P_i}\tfrac{1}{A}\textstyle\sum_{\mathbf{k}} v_{i,x}^2(\mathbf{k})\,\tau_{i,x}(\mathbf{k})\left|\tfrac{\partial f_0(E_i(\mathbf{k}))}{\partial E}\right| \\ \mu_{i,xy} &= \tfrac{e}{P_i}\tfrac{1}{A}\textstyle\sum_{\mathbf{k}} v_{i,x}(\mathbf{k})\,v_{i,y}(\mathbf{k})\,\tau_{i,y}(\mathbf{k})\left|\tfrac{\partial f_0(E_i(\mathbf{k}))}{\partial E}\right|\end{aligned} \tag{4.19}$$

Eq.4.19 is also known as the Kubo-Greenwood formula.

The sums over $\mathbf{k}$ in Eq.4.19 are converted to integrals as explained in Section 2.4.1. By doing so and using Eqs.2.38 and 2.39 to re-write Eq.4.19 we readily obtain

$$\mu_{i,xy} = \frac{e}{k_B T\, P_i}\,\frac{1}{(2\pi)^2}\sum_S \int_S \mathrm{d}^2k\; v_{i,x}(S)\,v_{i,y}(S)\,\tau_{i,x}(\mathbf{k})\left|\frac{\partial f_0(E_i(\mathbf{k}))}{\partial E}\right| \tag{4.20}$$

where we have assumed constant velocities inside each simplex (see Eq.2.37). Another useful approximation consists in supposing a constant relaxation time for each simplex, whose value corresponds to the average between the relaxation time values at the vertexes of each triangle. We verified that, for a sufficiently dense triangular mesh, this approximation does not change the results in terms of mobility and hole concentration.

Now we can proceed as already done for the hole concentration and substitute k_y as in Eq.2.46 to obtain

$$\begin{aligned}\mu_{i,xy} \;=\; & \frac{e}{k_B T\, P_i}\,\frac{1}{(2\pi)^2}\sum_S v_{i,x}(S)\,v_{i,y}(S)\,\tau_{i,x}(S) \\ & \times\frac{1}{a_{i,y}^S}\int_{k_{xm}}^{k_{xM}}\mathrm{d}k_x\left[\frac{1}{1+\exp\left(\dfrac{E_{i,m}(k_x)-E_f}{k_BT}\right)}-\frac{1}{1+\exp\left(\dfrac{E_{i,M}(k_x)-E_f}{k_BT}\right)}\right]\end{aligned} \tag{4.21}$$

which can be finally simplified as

$$\begin{aligned}\mu_{i,xy} \;=\; & \frac{e}{k_B T\, P_i}\,\frac{1}{(2\pi)^2}\sum_S v_{i,x}(S)\,v_{i,y}(S)\,\tau_{i,x}(S) \\ & \times\frac{k_BT}{a_{i,y}^S}\log\left[\frac{1+\exp\left(\dfrac{E_f-(E_{i,0}^S+a_{i,x}^S\,k_{xm}+a_{i,y}^S\,k_{ym})}{k_BT}\right)}{1+\exp\left(\dfrac{E_f-(E_{i,0}^S+a_{i,x}^S\,k_{xM}+a_{i,y}^S\,k_{yM})}{k_BT}\right)}\right]\end{aligned} \tag{4.22}$$

All above considerations are valid for the calculation of the mobility in a single subband i. At room temperature many subbands contribute to the current density, thus the total current density is given by the sum of the current carried by all subbands. Defining P_{inv} as the total hole inversion charge

$$P_{\mathrm{inv}} = \sum_i P_i \quad , \tag{4.23}$$

the total mobility is equal to

$$\mu_{xy} = \sum_i \frac{P_i\,\mu_{i,xy}}{P_{\mathrm{inv}}} \tag{4.24}$$

By doing so, the most occupied subbands will dominate the final value of the low-field mobility.

What we have derived in Eqs.4.20-4.24 is valid for all μ_{ij} components. In particular, when we consider low-field mobility in long channel devices, the only component we are interested in is μ_{xx}, which is usually referred to as the effective channel mobility.

Multiple scattering mechanisms

In case of multiple scattering mechanisms the MRT further allows us to investigate all scattering mechanisms one by one. On the other hand, when we want to consider the overall effect on the low-field mobility, the relaxation time in the subband i (τ_i) entering Eqs.4.19 and 4.20 should be the total relaxation time in the subband i. This quantity can be evaluated approximately as:

$$\frac{1}{\tau_i(\mathbf{k})} = \sum_m \frac{1}{\tau_{i,m}(\mathbf{k})} \tag{4.25}$$

where m denotes the scattering mechanism. Section 4.4 describes how we computed the relaxation time for different sources of scattering, namely acoustic and optical phonons and surface roughness scattering.

4.3 Self-consistent solution

When we calculate the band-structure with the **k·p** model discussed in Section 2.2, the electrostatic potential is an input for the calculation of the subbands. Obviously, the first guess we use for the potential is usually far from the *real* profile, thus we need to couple the Schrödinger equation (**k·p** in our case) and the Poisson equation in a self-consistent loop [3]. During the first part of the PhD activity, we developed a complete simulator which couples the **k·p** model for hole inversion layers and the Poisson equation, and then calculates the low-field mobility in inversion layers as explained in Section 4.2.2.

This section presents the simulator structure together with the Poisson equation.

4.3.1 Simulator structure

In Fig.4.1 the simulator flowchart has been reported. With this approach we are able to simulate any device with one-dimensional confinement such as bulk MOSFETs, single gate (SG) and double gate (DG) ultra-thin-body UTB-SOI FETs. Once we have defined the device structure and other simulation parameters such as lattice temperature, semiconductor type and crystal orientation, the first guess is calculated with a Poisson solver based on classical charge profile. The electrostatic potential is the input for the **k·p** solver, whose outputs are the eigenvalues and the eigenfunctions. Thus we can calculate the hole concentration that enters the 1-D Poisson equation solver. Once the charge and the electrostatic potential has converged to stable values, the self-consistent band structure can be used to calculate hole mobility. This has been done according to the MRT approximation and the *Kubo-Greenwood* formula (see Section 4.2.2).

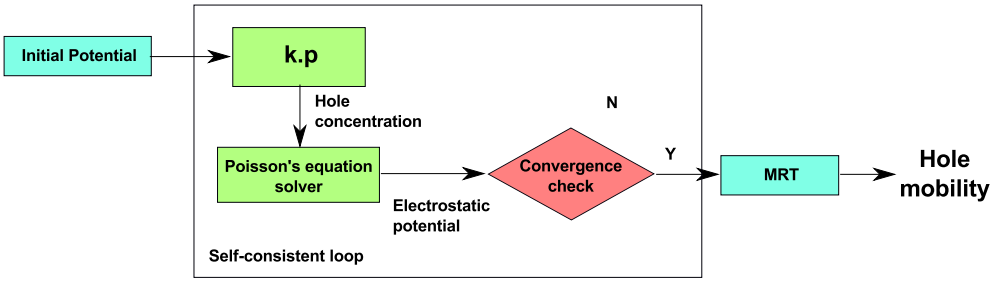


Figure 4.1: Flowchart of the simulator structure. The self-consistent loop consists of the **k·p** block and the 1-D Poisson equation solver. They are coupled with the electrostatic potential and the hole concentration.

4.3.2 Poisson equation

The general formulation of the Poisson equation discretized with the FDM reads

$$\nabla \cdot (\epsilon \nabla \phi(z)) = -e\,(p(z) - n(z) + N_{\mathrm{D}} - N_{\mathrm{A}}) \tag{4.26}$$

where ϵ is the permittivity of the material and ϕ is the electrostatic potential. p and n are the hole and electron concentrations, respectively.

We employed a non-uniform one-dimensional grid in real space, which is sketched in Fig.4.2. With this discretization we can handle multiple discontinuities of the material. Δ_i is the grid step while σ_i is the segment associated with the charge. The convergence of the self-

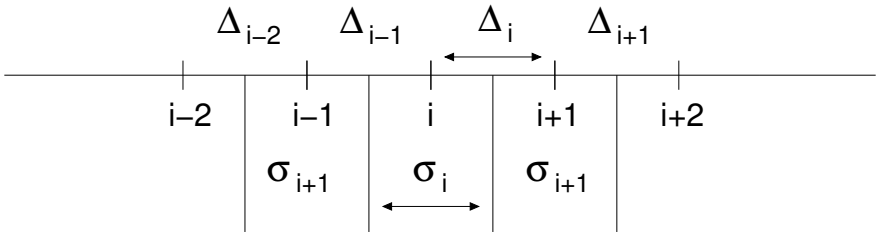


Figure 4.2: Poisson discretization grid. Δ_i is the grid step while σ_i is the segment associated with the charge.

consistent loop is achieved using a non-linear formulation of the Poisson equation that includes the electrostatic potential of the actual (n) and the preceding $(n-1)$ iterations in the right-hand-side [12, 13]:

$$\nabla \cdot (\epsilon \nabla \phi^{(n)}) = -e\left(p^{(n)} \exp\left(e\frac{\phi^{(n)} - \phi^{(n-1)}}{k_B T}\right) - n_b \exp\left(e\frac{\phi^{(n)} - \phi_b}{k_B T}\right) + N_{\mathrm{D}} - N_{\mathrm{A}}\right) \tag{4.27}$$

where p is the hole charge calculated with the 6-band $\mathbf{k}\cdot\mathbf{p}$ solver described in Chapter 2 and calculated according to Eq.2.47. n_b and ϕ_b are the concentration and the potential experienced by the electrons deep in the bulk substrate.

4.4 Scattering Mechanisms

4.4.1 Phonon scattering

At finite temperature atoms in the crystal lattice oscillate with respect to their rest positions and this represents an important source of scattering for both electrons and holes. These vibrations produce a perturbation potential that can be modeled using a simple mechanical model [3]. The phonon concept originates from the quantum mechanical nature of the lattice vibrations. Here we will present only the final expressions for the acoustic and optical phonon relaxation times, while interested readers may find further details in [2, 3].

Acoustic electron-phonon interaction is treated as an elastic and isotropic scattering mechanism. The acoustic relaxation time in the subband i and at the energy $E(\mathbf{k})$ reads

$$\frac{1}{\tau^i_{ac}(E(\mathbf{k}))} = \frac{2\pi k_B T D^2_{acc}}{\rho v_l^2 \hbar} \sum_j I_{i,j}\, D_j(E(\mathbf{k})) \tag{4.28}$$

where k_B is the Boltzmann constant, T is the lattice temperature, D_{acc} is the acoustic deformation potential, ρ is the mass density and v_l is the longitudinal sound velocity. $D_j(E(\mathbf{k}))$ is the

density of states at the energy $E(\mathbf{k})$ defined in Eq.2.43, while $I_{i,j}$ is the form factor involving the subbands i and j which, in principle, depends on the 2D wavevectors $\mathbf{k}$ and $\mathbf{k}'$:

$$I_{i,j,\mathbf{k}} = \int dz \, |\boldsymbol{\psi}^{\dagger}_{i,\mathbf{k}}(z)\boldsymbol{\psi}_{j,\mathbf{k}'}(z)|^2 \tag{4.29}$$

By following what already done in [11], we neglected the $\mathbf{k}$-dependence of $I_{i,j}$ and calculated:

$$I_{i,j} = I_{i,j,0} = \int dz \, |\boldsymbol{\psi}^{\dagger}_{i,0}(z)\boldsymbol{\psi}_{j,0}(z)|^2 \tag{4.30}$$

that is by using the wave functions at $\mathbf{k}$=0.

Optical phonon scattering is instead treated as an inelastic mechanism. The relaxation time inside the subband i is written as

$$\begin{aligned}\frac{1}{\tau^i_{opt}(E(\mathbf{k}))} = \frac{\pi D^2_{opt}}{\rho\,\omega_{opt}} \sum_j \left[N_{opt}(\hbar\omega_{opt}) + \frac{1}{2} \pm \frac{1}{2} \right] \\ \times \frac{1 - f_0\left(E(\mathbf{k}) \mp \hbar\omega_{opt}\right)}{1 - f_0(E(\mathbf{k}))} I_{i,j}\, D_j(E(\mathbf{k}) \mp \hbar\omega_{opt})\end{aligned} \tag{4.31}$$

where D_{opt} is the optical deformation potential, ω_{opt} is the phonon frequency, N_{opt} and f_0 are the Bose-Einstein and Fermi-Dirac distributions, respectively. The upper and lower signs correspond to emission and absorption processes, respectively. Tab.4.1 summarizes the phonon scattering parameters used for the simulation of hole inversion layers.

	D_{acc} [eV]	D_{opt} [10^8eV/cm]	ω_{opt} [meV]
Si	10.2	15	62
Ge	11	6	38

Table 4.1: Calibrated phonon scattering parameters for Si and Ge.

4.4.2 Surface roughness scattering

Surface roughness is the dominant scattering mechanism at high inversion densities for both electron and hole mobilities [14]. The model implemented here has been derived starting from the general expression for surface roughness relaxation time in the i-th subband and along the direction x which reads [11]

$$\begin{aligned}\frac{1}{\tau^x_{i,SR}(\mathbf{k})} &= \frac{2\pi}{\hbar} \sum_j \frac{1}{(2\pi)^2} \int \mathrm{d}^2k' \, |M_{i,j}(\mathbf{k},\mathbf{k}')|^2 \, \delta(E_i(\mathbf{k}) - E_j(\mathbf{k}')) \\ &\times \left(1 - \frac{v^j_x(\mathbf{k}')\,\tau^x_j(\mathbf{k}')\, f_0(E_j(\mathbf{k}'))}{v^i_x(\mathbf{k})\,\tau^x_i(\mathbf{k}) f_0(E_i(\mathbf{k}))} \right)\end{aligned} \tag{4.32}$$

where $M_{i,j}$ is the surface roughness matrix element. In App. A the reader can find the formulation used for $M_{i,j}$ for either bulk or SOI structures. As already done in [11], we simplify the problem by considering only intra-subband scattering (i.e. i=j) and by considering sufficiently slow variations of $\tau_i(\mathbf{k})$ with $\mathbf{k}$, thus $\tau_i(\mathbf{k}) \simeq \tau_i(\mathbf{k}')$. It is clear from Eq.4.32 that we are treating as elastic all the transitions assisted by surface roughness scattering because the initial and final energies ($E_i(\mathbf{k})$ and $E_j(\mathbf{k}')$) must be the same in order to give a non-zero rate. Another useful

approximation concerns the ratio of the velocities in the integral of Eq.4.32. If we consider only intra-subband scattering events we can approximately write

$$\frac{v_x^i(\mathbf{k}')}{v_x^i(\mathbf{k})} \simeq \cos\theta_{\mathbf{q}} \tag{4.33}$$

where $\theta_{\mathbf{q}}$ is the angle of the exchanged momentum defined as $\mathbf{q} = \mathbf{k}' - \mathbf{k}$.

Using Eqs.2.39 and 2.43 we can simplify Eq.4.32 as

$$\frac{1}{\tau_{SR}^i(\mathbf{k})} \simeq \frac{1}{2\pi\hbar} \sum_{j,S} \frac{1}{a_y^S} \int_{kxm}^{kxM} \mathrm{d}k_x' \, |M_{i,j}(\mathbf{k},\mathbf{k}')|^2 (1 - \cos\theta_{\mathbf{q}}) \tag{4.34}$$

4.4.3 Static dielectric screening

For a 2D carrier gas the dielectric screening problem is related with the solution of a self-consistent electrostatic problem. In presence of free charge a perturbing potential (e.g. the potential produced by fluctuations of the Si-SiO_2 interface) induces a change in the spatial distribution of the charge which in turn creates an induced potential which sums algebraically to the source perturbing potential. This is clearly a self-consistent problem which can be solved writing the relation between the induced matrix element $M_{m,m'}^{(ind)}(\mathbf{q})$ and the perturbed one $M_{n,n'}(\mathbf{q})$ as [15, 3]:

$$M_{m,m'}^{(ind)}(\mathbf{q}) = \frac{e^2}{q(\epsilon_{si} + \epsilon_{ox})} \sum_{n,n'} \Pi_{n,n'}(\mathbf{q}) \, F_{m,m'}^{n,n'}(q) \, M_{n,n'}(\mathbf{q}) \tag{4.35}$$

where e is the electron charge and q is the modulus of the exchanged momentum $\mathbf{q}$. If we rely on Linhard screening [16, 17], $\Pi_{n,n'}(\mathbf{q})$ is the static polarization factor which takes the form

$$\Pi_{n,n'}(\mathbf{q}) = \frac{1}{A} \sum_{\mathbf{k}} \frac{f_{n'}(\mathbf{k}+\mathbf{q}) - f_n(\mathbf{k})}{E_{n'}(\mathbf{k}+\mathbf{q}) - E_n(\mathbf{k})} \tag{4.36}$$

and $F_{m,m'}^{n,n'}(q)$ is the unitless screening form factor defined as

$$F_{m,m'}^{n,n'}(q) = \int dz \, \psi_m^\dagger(z) \, \psi_{m'}(z) \int dz_0 \psi_n^\dagger(z_0) \, \psi_{n'}(z_0) \, \phi_{pcN}(q,z,z_0) \tag{4.37}$$

In this last expression the term $\phi_{pcN}(q,z,z_0)$ is the normalized point-charge potential

$$\phi_{pcN}(q,z,z_0) = \frac{q(\epsilon_{si} + \epsilon_{ox})}{e} \, \phi_{pc}(q,z,z_0) \tag{4.38}$$

and $\phi_{pc}(q,z,z_0)$ is the electrostatic potential calculated as the solution of a 1-D Poisson equation along the quantization direction z and for a point-charge located at z_0 ($\mathbf{r}_0$ coincides with the origin, i.e. $\mathbf{r}_0$=0).

Depending on the device structure, different expressions are obtained for $\phi_{pc}(q,z,z_0)$. For a bulk MOS transistor, where the interface between the dielectric and the channel is taken at z=0, $\phi_{pc}(q,z,z_0)$ reads:

$$\phi_{pc}(q,z,z_0) = \frac{e}{2q\epsilon_{si}} \exp(-q|z-z_0|) + \left(\frac{\epsilon_{si} - \epsilon_{ox}}{\epsilon_{ox} + \epsilon_{si}}\right) \frac{e}{2q\epsilon_{si}} \exp(-q(z+|z_0|)), \quad \text{for } z>0 \tag{4.39}$$

$$\phi_{pc}(q,z,z_0) = \frac{e}{2q\epsilon_{ox}} \exp(-q|z-z_0|) + \left(\frac{\epsilon_{ox} - \epsilon_{si}}{\epsilon_{ox} + \epsilon_{si}}\right) \frac{e}{2q\epsilon_{ox}} \exp(q(z-|z_0|)), \quad \text{for } z<0 \tag{4.40}$$

where we have assumed that the silicon and dielectric regions correspond to $z>0$ and $z<0$, respectively. we areFor a SOI MOS structure, the scattering potential in the silicon region (i.e. $0<z<T_{SI}$) can be written as

$$\phi_{pc}(q,z,z_0) = \frac{e}{2q\epsilon_{si}} \left[\exp(-q|z-z_0|) + C_1\exp(-qz) + C_2\exp(-qz)\right] \tag{4.41}$$

where

$$\begin{aligned} C_1 &= \frac{(\epsilon_{si}-\epsilon_{ox})^2\exp(-q|z_0|) + (\epsilon_{si}^2-\epsilon_{ox}^2)\exp[-q(|T_{SI}-z_0|-T_{SI})]}{(\epsilon_{si}+\epsilon_{ox})^2\exp(2qT_{SI}) - (\epsilon_{si}-\epsilon_{ox})^2} \\ C_2 &= \frac{(\epsilon_{si}-\epsilon_{ox})(C_1+\exp(-q|z_0|))}{(\epsilon_{si}+\epsilon_{ox})} \end{aligned} \tag{4.42}$$

Eq.4.35 provides a relation between the unscreened matrix element $M_{n,n'}(\mathbf{q})$ and the induced matrix element $M^{(ind)}_{m,m'}(\mathbf{q})$. Obviously, this latter induced potential causes itself another perturbation in the charge distribution, hence the need for a self consistent solution, which can be obtained by substituting the unscreened matrix element with the total screened one $M^{(scr)}_{m,m'}(\mathbf{q})$ and the left hand side of Eq.4.35 with the difference between the screened and the unscreened matrix elements which corresponds by definition to the perturbation due to the screened matrix element [16]:

$$M^{(scr)}_{m,m'}(\mathbf{q}) - M_{m,m'}(\mathbf{q}) = \frac{e^2}{q(\epsilon_{si}+\epsilon_{ox})} \sum_{n,n'} \Pi_{n,n'}(\mathbf{q})\, F^{n,n'}_{m,m'}(q)\, M^{scr}_{n,n'}(\mathbf{q}) \tag{4.43}$$

Eq.4.43 represents a linear system of equations which links the unknown screened matrix elements with the screened ones and can be rewritten as

$$M_{m,m'}(\mathbf{q}) = \sum_{n,n'} \varepsilon^{n,n'}_{m,m'}(\mathbf{q})\, M^{(scr)}_{n,n'}(\mathbf{q}) \tag{4.44}$$

where $\varepsilon^{n,n'}_{m,m'}(\mathbf{q})$ is the dielectric function tensor

$$\varepsilon^{n,n'}_{m,m'}(\mathbf{q}) = \delta_{n,m}\,\delta_{n',m'} - \frac{e^2}{q(\epsilon_{si}+\epsilon_{ox})}\,\Pi_{n,n'}(\mathbf{q})\,F^{n,n'}_{m,m'}(q) \tag{4.45}$$

Despite the fact that in the $\mathbf{k}\cdot\mathbf{p}$ framework the wave function $\psi_{n,\mathbf{k}}(z)$ depends also on the wavevector $\mathbf{k}$, in what follows we will neglect this dependence, which would otherwise increase the computational burden because of the huge number of integrals implied by Eqs.4.36 and 4.37. This choice is consistent with the approximation used for the calculation of the relaxation times due to surface roughness and ionized impurities which employs the wave functions at $\mathbf{k}$=0.

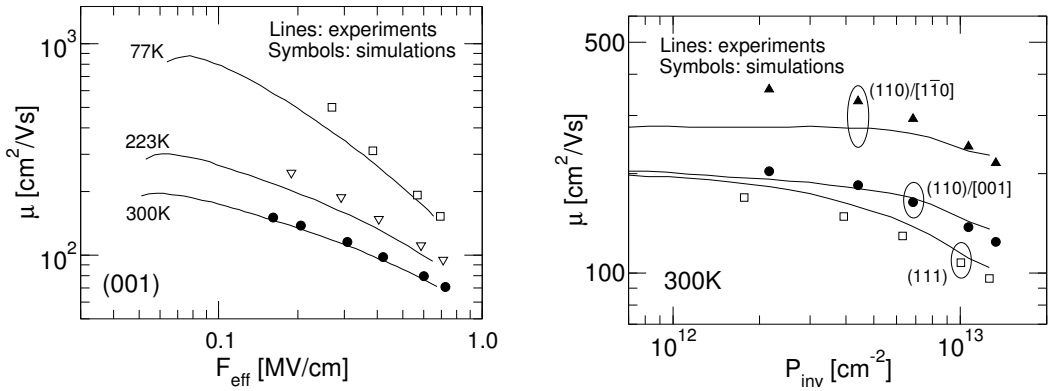
The anisotropic matrix element already calculated in Secs.4.4.2 is valid only for intra-subband transitions due to the approximations needed to find a closed expression for the relaxation time. In this spirit the dielectric function and its components are calculated only for intra-subband transitions.

4.5 Model calibration

The consistency of the theoretical models with the experimental results is of utmost importance in order to be able to have good qualitative and quantitative prediction capabilities. This section presents the comparison of mobility simulations with the experimental results in literature. The good agreement we found after the calibration step for either unstrained or strained conditions legitimates us to extend the use of our theoretical models to study novel device architectures, novel materials and non-conventional strain conditions, as we will see in Chapter 5.

4.5.1 Unstrained bulk devices with arbitrary orientation

Fig.4.3 illustrates hole mobility simulations compared to the experimental data for (001), (110) and (111) silicon inversion layers. The results were obtained using the self-consistent **k·p**-Poisson solver and then calculating the mobility with the Kubo-Greenwood formalism (see Section 4.2.2), and accounting for acoustic and optical phonons and surface roughness scattering.



(a) (b)

Figure 4.3: (a) Silicon (001). Hole mobility versus effective field at different temperatures. The transport is aligned with the [100] direction. Lines: experiments from [14]. Symbols: simulations. (b) Silicon (110) and (111). Hole mobility versus inversion charge density at 300K.Lines: experiments from [18]. Symbols: simulations.

Surface Roughness: (Exp. spectrum)	Holes	Electrons
Δ[nm]	0.24	0.9
Λ [nm]	4	1.2
Phonons:	Holes	Electrons
Acoustic D_{acc} [eV]	10.2	13
Optical D_{opt} [10^8eV/cm]	15	from [6]
Optical Ph. Energy [meV]	62	from [6]

Table 4.2: Scattering parameters used in all simulations. Electron optical phonons model from [6]. Screening is not accounted for in the hole mobility calculations, hence the Λ is relatively large to achieve a good agreement with the experimental mobility data at large inversion densities [19].

The parameters for the phonon and the surface roughness scattering are reported in Tab.4.2 for electrons and holes. Since the screening is not accounted for in the **k·p** hole mobility simulations, then it was necessary to use a relatively large value of the surface roughness correlation length Λ in order to reproduce the measured mobility versus F_{eff} slope at large F_{eff}.

As it can be seen, the differences between the hole mobilities in (001), (110) and (111) inversion layers are nicely reproduced by the simulations with a fixed set of scattering parameters, thus showing that the band-structure effects dominate the mobility dependence on the crystal orientation. In particular, Fig.4.3(b) shows that, for (110) hole inversion layers, the **k·p** simulations agree well with the experimentally observed mobility dependence on the transport direction.

These results demonstrate good predictive capabilities of the semi-classical transport modelling as far as the mobility dependence on the crystal orientation is concerned.

4.5.2 Strained bulk devices

A vast literature exists for both the experimental analysis and the modelling of mobility in strained devices. We hereafter present simulation results for a few relevant cases and discuss the comparison with the experiments.

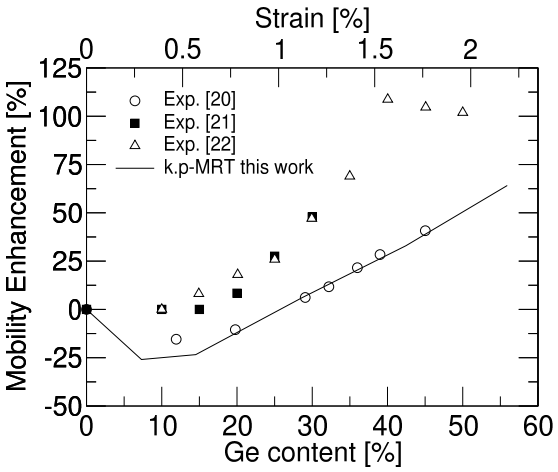


Figure 4.4: (001) silicon. Hole mobility enhancement as a function of the germanium content for biaxially strained p-MOSFETs. Effective field F_{eff} = 0.7 MV/cm. Symbols: experiments [20, 21, 22]. Line: **k·p** simulations.

Fig. 4.4 shows simulations and experiments for hole mobility enhancements in biaxially strained p-MOSFETs. As it can be seen, the simulations obtained with the MRT approach and the self-consistent **k·p**-Poisson solver can reproduce reasonably well the experimental behavior without any change in the surface roughness parameters with the strain level. Both in the experiments and in the simulations we observe a non monotonic dependence of the mobility on the biaxial strain. The non monotonic behavior of mobility enhancement is due to two counteracting effects: on one hand the strain-induced band structure warping leads to mobility improvement, whereas on the other hand the scattering mechanisms, at low strain levels, tend to decrease hole mobility.

Another widely explored strain configuration is the uniaxial strain in the (110) quantization plane. This type of strain can be present in the lateral sidewalls of FinFET devices and can be exploited in order to enhance the performances either in terms of mobility or drain current [23]. Fig.4.5 shows the experimental and simulated low-field mobility in uniaxially strained (110)/[1$\bar{1}$0] planar MOSFETs. Good agreement between simulations and experiments has been obtained for both tensile and compressive strain. As it can be seen, compressive uniaxial strain increases the hole mobility, while tensile strain causes a reduction of the hole mobility.

4.5.3 Calibration in presence of dielectric screening

The scattering parameters have been calibrated also in presence of dielectric screening (see Section 4.4.3). Fig.4.6 shows the calibrated hole mobility for a bulk p-MOSFET and for different temperature values. As it can be seen, by changing the scattering parameters, we reproduce

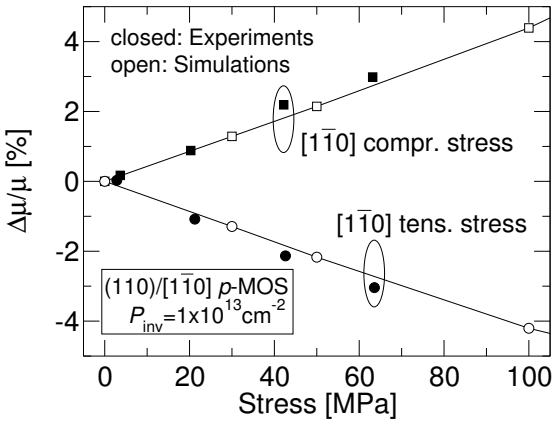


Figure 4.5: Simulations and measurements [18] of hole mobility in uniaxially strained (110)/⟨110⟩ bulk planar p-MOSFETs.

the measured mobilities versus the effective field in a wide temperature range. Fig.4.6 reports also the calibrated parameters for phonon and surface roughness scattering.

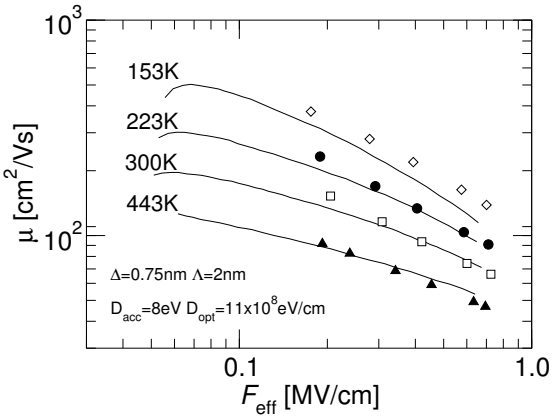


Figure 4.6: Silicon (001). Hole mobility versus effective field at different temperatures. Lines: experiments from [14]. Symbols: simulations accounting for the effect of screening. The figure illustrates also the calibrated scattering parameters, which are somewhat different with respect to Tab.4.2 because the dielectric screening changes both the mobility value and the slope versus the effective field.

Bibliography

[1] S. Datta, *Quantum Transport - Atom to Transistor*. United Kingdom: Cambridge University Press, 2005.

[2] M.S.Lundstrom, *Fundamentals of Carrier Transport*. New York: Addison Wesley, 1990.

[3] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS transistors: Semi-classical modeling and applications*. United Kingdom: Cambridge University Press, 2010.

[4] A. Schenk, *Advanced Physical Models for Silicon Device Simulation*. Wien New York: Springer-Verlag, 1998.

[5] E. Sangiorgi, P. Palestri, D. Esseni, C. Fiegna, A. Abramo, and L. Selmi, "Device Simulation for Decananometer MOSFETs ," *Material Science in Semiconductor Processing*, vol. 6, no. 1–3, pp. 93–105, 2003.

[6] 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, 1983.

[7] M. V. Fischetti and S. E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects," *Phys. Rev. B*, vol. 38, pp. 9721–9745, 1988.

[8] L. Lucci, P. Palestri, D. Esseni, L. Bergagnini, and L. Selmi, "Multisubband Monte Carlo Study of Transport, Quantization, and Electron-Gas Degeneration in Ultrathin SOI n-MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 5, pp. 1156–1164, 2007.

[9] F. M. Bufler, A. Schenk, and W. Fichtner, "Efficient Monte Carlo Device Modeling," *IEEE Trans. Electron Devices*, vol. 47, pp. 1891–1897, 2000.

[10] D. Esseni and P. Palestri, "Theory of the motion at the band crossing points in bulk semiconductor crystals and in inversion layers," *Journal of Applied Physics*, vol. 105, no. 5, p. 053702, 2009.

[11] M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, "Six-band kp calculation of the hole mobility in silicon inversion layers: Dependence on surface orientation, strain, and silicon thickness," *Journal of Applied Physics*, vol. 94, no. 2, pp. 1079–1095, 2003.

[12] A. Pacelli, "Self-consistent solution of the Schrodinger equation in semiconductor devices by implicit iteration," *IEEE Trans. Electron Devices*, vol. 44, no. 7, pp. 1169 –1171, jul 1997.

[13] P. Palestri, N. Barin, D. Esseni, and C. Fiegna, "Revised Stability Analysis of the Non-Linear Poisson Scheme in Self-Consistent Monte-Carlo Device Simulations," *IEEE Trans. Electron Devices*, vol. 53, no. 6, pp. 1443–1451, 2006.

[14] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the Universality of Inversion-layer Mobilty in Si MOSFETs. Part I- Effect of Substrate Impurity Concentration," *IEEE Trans. Electron Devices*, vol. 41, no. 12, pp. 2357–62, 1994.

[15] 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.

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

[17] B. K. Ridley, *Quantum Processes in Semiconductors.* Oxford Science Publication, 1993.

[18] H. Irie, K. Kita, K. Kyuno, and A. Toriwni, "In-Plane Mobility Anisotropy and Universality Under Uni-axial Strains in n- and p-MOS Inversion Layers on (100), (110), and (111) Si," in *IEEE IEDM Technical Digest*, 2004, pp. 225–228.

[19] N. Serra, F. Conzatti, D. Esseni, M. De Michielis, P. Palestri, L. Selmi, S. Thomas, T. Whall, E. Parker, D. Leadley, L. Witters, A. Hikavyy, M. Hytch, F. Houdellier, E. Snoeck, T. Wang, W. Lee, G. Vellianitis, M. van Dal, B. Duriez, G. Doornbos, and R. Lander, "Experimental and physics-based modeling assessment of strain induced mobility enhancement in FinFETs," in *IEEE IEDM Technical Digest*, dec. 2009, pp. 71 –74.

[20] K. Rim, K. Chan, L. Shi, D. Boyd, J. Ott, N. Klymko, F. Cardone, L. Tai, S. Koester, M. Cobb, D. Canaperi, B. To, E. Duch, I. Babich, R. Carruthers, P. Saunders, G. Walker, Y. Zhang, M. Steen, and M. Ieong, "Fabrication and mobility characteristics of ultra-thin strained Si directly on insulator (SSDOI) MOSFETs," in *IEEE IEDM Technical Digest*, dec. 2003, pp. 3.1.1 – 3.1.4.

[21] M. Currie, C. Leitz, T. Langdo, G. Taraschi, and E. Fitzgerald, "Carrier mobilities and process stability of strained Si n- and p-MOSFETs on SiGe virtual substrates," *J. Vac. Sci. Technol. B*, vol. 19, pp. 2268–2279, 2001.

[22] C. Leitz, M. Currie, M. Lee, Z. Cheng, D. Antoniadis, and E. A. Fitzgerald, "Hole mobility enhancements and alloy scattering-limited mobility in tensile strained Si/SiGe surface channel metal-oxide-semiconductor field-effect transistors," *Journal of Applied Physics*, vol. 92, pp. 3745–3751, 2002.

[23] F. Conzatti, N. Serra, D. Esseni, M. D. Michielis, A. Paussa, P. Palestri, L. Selmi, S. Thomas, T. Whall, D. Leadley, E. Parker, L. Witters, M. Hytch, E. Snoeck, T. Wang, W. Lee, G. Doornbos, G. Vellianitis, M. van Dal, and R. Lander, "Investigation of strain engineering in FinFETs comprising experimental analysis and numerical simulations," *IEEE Trans. Electron Devices*, vol. 58, no. 6, pp. 1583–1593, June 2011.

Chapter 5

Strain and material engineering in advanced CMOS transistors

5.1 Introduction

THIS chapter is devoted to the application of the semiclassical model presented in Chapter 4 to study the effect of strain and novel channel materials on the performances of advanced nanoscale MOSFET devices.

In particular, Section 5.2 combines direct measurements of strain, electrical mobility measurements and a rigorous modeling approach to provide insight about strain induced mobility enhancement in FinFETs and guidelines for device optimization. Good agreement between simulated and measured mobility is obtained using strain components measured directly at device level by a novel holographic technique. A large vertical compressive strain is observed in metal gate FinFETs and the simulations show that this helps recover the electron mobility disadvantage of the (110) FinFETs lateral interfaces with respect to (100) interfaces, with no degradation of the hole mobility. The model is then used to explore systematically the impact of stress components in the fin-width, fin-height and fin-length direction on the mobility of both n-type and p-type FinFETs and to identify optimal stress configurations. Finally, self-consistent Monte Carlo simulations are used to investigate how the most favorable stress configurations can improve the on-current of nanoscale MOSFETs.

Section 5.3 presents a simulation study of the electron and hole mobility enhancements in biaxially strained MOS transistors. In particular, this section presents a critical examination of a recently proposed interpretation of the experimental data, according to which the strain significantly modifies not only the r.m.s. value but also the correlation length of the surface roughness spectrum. We present a systematic comparison between comprehensive numerical simulations and experiments which supports such an interpretation.

Finally, Section 5.4 analyzes the $I_{\rm on}$ in Si, sSi, Ge and sGe n- and p-MOSFETs by accounting for all the relevant scattering mechanisms (including the remote surface-optical phonons (SOph) and remote Coulomb scattering (RCS) related to high-κ dielectrics), in which strain is implicitly introduced by a modification of the band structure. Our models are first validated against experiments for both mobility and $I_{\rm DS}$ in nanoscale transistors. Then the $I_{\rm on}$ in Ge and Si MOSFETs is compared for different crystal orientations and strain conditions.

5.2 Investigation of strain engineering in FinFET devices

5.2.1 Context of this study

FinFETs are considered promising candidates for nanoscale CMOS ICs [1, 2, 3] and the strain engineering is of the utmost importance for the optimization of the transistor performances [4, 3, 5]. Strain induced mobility enhancement in planar MOSFETs has been studied extensively [6, 7, 8, 9], but the complex stress configurations in FinFETs still demand further experimental analysis and physics-based models to support the device design. In this respect, Fig.5.1 shows that in a FinFET structure there are three main stress components, namely the component $T_{\rm fL}$ along the fin length direction and the $T_{\rm fH}$ and $T_{\rm fW}$ respectively along the fin height and fin width direction.

The stress distribution in FinFETs have been so far typically inferred from process simulations and the analysis of mobility data has been based on the piezo-resistive coefficients [3, 4, 10], that do not offer physical insight and may be inaccurate at high stress levels and for complex stress configurations as in Fig.5.1 [11]. Furthermore, for narrow fins, mobility data are limited and strain measurement is challenging for established techniques [12, 13].

Despite the recent theoretical studies devoted to the physical mechanisms behind the stress induced mobility enhancements in FinFETs [14], a joint experimental and theoretical analysis of this topic is necessary and preliminary results have been presented in [15].

In this chapter we present assessment of the strain induced mobility enhancement based on direct strain measurements, electrical mobility measurements and numerical physics-based

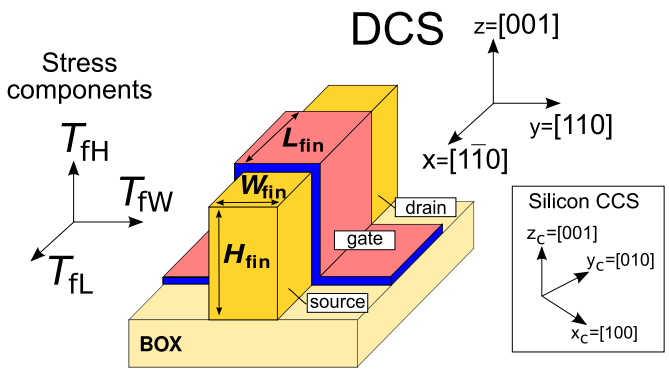


Figure 5.1: Schematic of a (110)/⟨110⟩ FinFET obtained with a standard Manhattan layout in a (100) wafer. The figure illustrates the stress components in the fin height T_{fH}, width T_{fW} and length T_{fL} direction in the Device Coordinate System (DCS). Also shown is the silicon Crystal Coordinate System (CCS).

modelling. Mobilities in narrow n- and p-FinFETs, measured from 4.2K to 300K, are compared with state-of-the-art transport models based on the Boltzmann Transport Equation (BTE) in inversion layers. These models account for all the relevant scattering mechanisms and strain is introduced by modification of the silicon band structure. Simulations are used to clarify the physical mechanisms for mobility enhancement and to provide guidelines for the device optimization. Finally, we investigated how the most favorable stress configurations can translate into on-current improvements by using self-consistent Monte Carlo simulations of nanoscale MOSFETs accounting for realistic values of series resistances.

The results of this section are the outcome of a collaboration between the University of Udine (Italy), NXP-TSMC in Leuven (Belgium), the University of Warwick (UK) and the CEMES-CNRS, University of Toulouse (France).

5.2.2 Device fabrication and characterization

Device fabrication

22nm node FinFETs were fabricated by NXP-TSMC in Leuven using a VLSI-compatible process [1, 2], with a starting material 300 mm (100) Silicon-on-Insulator (SOI) wafers featuring a 145nm buried SiO_2 and 65nm crystalline Si. The SOI layers were lowly p-type (B) doped with a dopant concentration of approximately $10^{15}\mathrm{cm}^{-3}$. After optimizing the 193nm immersion lithography process, reactive ion etch (RIE) and wet etching conditions, fins with (110) sidewalls and widths down to 16nm were fabricated. To repair for RIE damage on the fin outer surface, a sacrificial oxide was grown and subsequently removed by wet etching. Gate stacks were formed on 1nm chemical oxide with deposition of 2.3nm $Hf_{0.4}Si_{0.6}O$ by atomic layer deposition. This was followed by physical vapor deposition of 5nm TiN and the stack was capped with 100 nm plasma-enhanced chemical vapor deposited amorphous-Si at 600°C. Gate patterning was performed, leaving an oxide hard mask of about 30nm. The oxide hard mask is preserved in order to avoid undesirable growth of Si on top of the gate during the selective epitaxial growth (SEG) on S/D areas. After gate patterning, extensions were implanted (As for n-MOS and BF_2 for p-MOS) at oblique angles (45°) to implant both top and side wall of the fin. Afterwards spacers were formed. In our study, 30nm Si SEG was used for S/D areas. Finally, the oxide hard mask was removed by wet etching and highly doped drain implantation (As plus P for n-MOS; B for p-MOS) and NiSi (10nm Ni) were used respectively for gate and S/D contacts; the doping concentration is estimated to be about $10^{20}\mathrm{cm}^{-3}$ in the LDD areas. A H_2 anneal was used to round the corners of the fins and smoothen the sidewalls.

Holographic strain measurements

Strain components ε_{yy}, ε_{zz} (y, z being in the fin-width and fin-height directions respectively, see DCS in Fig.5.1) of fully-processed devices were quantified directly by dark-field electron holography (DFEH) carried out in a transmission electron microscope (TEM) at the CEMES-CNRS, University of Toulouse (France) [16, 17]. This novel technique is capable of measuring strain with nanometer spatial resolution and for relatively thick samples, thus limiting the influence of sample relaxation compared to high-resolution TEM strain measurements [18]. Cross-sectional TEM samples were prepared to a uniform thickness of about 150nm at a $[1\bar{1}0]$ zone-axis by focused-ion beam (FIB). Electron holography was performed on the SACTEM-Toulouse, a Tecnai F20 TEM (FEI) operating at 200kV, equipped with field-emission gun, electrostatic biprism and imaging aberration corrector (CEOS). Pairs of dark-field holograms were recorded using the $(11\bar{1})$ and $(\bar{1}\,\bar{1}\,\bar{1})$ diffracted beams. Biprism voltages were typically 130V, fringe spacings 1.5nm and overlap regions 450nm. Strain maps were extracted from these holograms using the software package HoloDark 1.0 (HREM Research), a plug-in for the DigitalMicrograph image processing package (Gatan).

The strain in the fins is determined by measuring the distortion of the crystalline lattice in the fins compared with the silicon substrate, used as the reference.

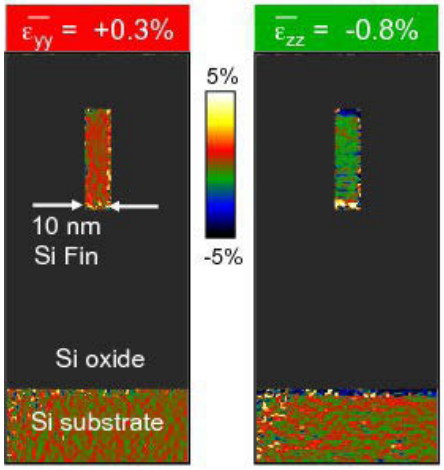


Figure 5.2: Measured ε_{yy} (left) and ε_{zz} (right) strain maps for 10nm wide fins (see Fig.5.1 for the definition of the y and z directions). The deformations along the lateral (W_{fin}, left) and vertical (H_{fin}, right) directions have been calculated from (111) and $(\bar{1}\,\bar{1}1)$ holographic fringes, respectively. The region of the silicon dioxide has been blacked out as the measurements are not meaningful in the amorphous material.

Fig.5.2 shows the ε_{yy} and ε_{zz} strain maps. The fluctuations in the substrate are due to noise and can be used to estimate the precision of the technique, which in these measurements is about 0.05% at a spatial resolution of 3nm. It can be seen that the average strain values within the fin are ε_{yy}=0.30±0.05% and ε_{zz}=−0.80±0.05%. From the strain maps in Fig.5.2 it can be seen that, within the experimental precision of the technique, the two strain components are uniform across the body of the fin.

Electrical characterization

Mobility was extracted in 10μm long, 65nm tall SOI FinFETs consisting of 10 fins in parallel, and of total gate area equal to $10(2H_{\text{fin}}+W_{\text{fin}})$L. The measurements were performed at the University of Warwick (UK). Gate-channel capacitance, C_{gc}, was measured using a frequency of 100kHz from which the inversion density N_{inv} was calculated by integration under

the curve, taking care to subtract parasitic offsets. Gate leakage current I_G limited accurate calculation of $N_{\rm inv}$ up to values of approximately $2{\cdot}10^{13}{\rm cm}^{-2}$. The drain current $I_{\rm DS}$ characteristics versus $V_{\rm GS}$ were measured at $V_{\rm DS}$=10mV and corrected for gate leakage current using $I_{\rm DS}$=$I_{\rm D}^{meas}$+$I_G/2$ [19], where $I_{\rm D}^{meas}$ is the current measured at the drain side. Effective mobility was calculated using [20, 21]

$$\mu_{eff} = \frac{I_{\rm DS}(V_{\rm GS})\, L^2}{V_{\rm DS} \int C_{\rm GC}(V_{\rm GS})\, {\rm d}V_{\rm GS}} \tag{5.1}$$

Preliminary tests showed that a drain voltage $V_{\rm DS}$ of 10mV was low enough to lead to negligible error in the extraction of mobility [22].

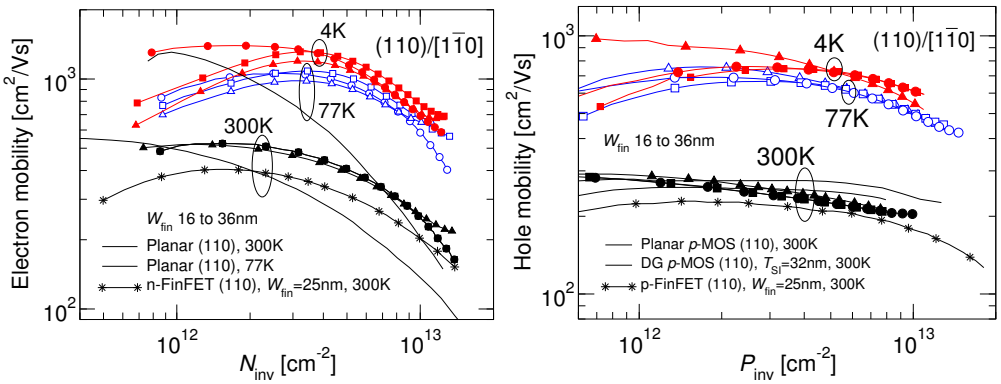


Figure 5.3: (a) Measured electron mobility at T=300K, 77K and 4K and for three fin widths W_{fin}=16nm (circles), 26nm (squares) and 36nm (triangles). A comparison to mobility data for (110) planar MOSFETs [23] and for (110) n-FinFETs [24] is also shown. (b) Measured hole mobility at T=300K,77K and 4K and for two fin widths W_{fin}=16nm (circles) and 36nm (squares). A comparison to mobility data for (110) planar bulk or SOI MOSFETs [25, 26] and (110) p-FinFETs [24] is also shown.

The measured electron and hole mobilities are compared in Fig.5.3 to literature data corresponding to (110) planar MOSFETs with SiO_2 gate dielectric [23, 25, 26] and to (110) n- and p-FinFETs with high-κ and metal gate [24]. For both electrons and holes the mobility is fairly independent of the fin width W_{fin} in the range considered in the figures. Fig.5.3(a) shows that the electron mobility in the high-κ, metal gate FinFETs of this work is significantly larger than for planar MOSFETs in the $N_{\rm inv}$ range of practical interest (i.e. for $N_{\rm inv}$ larger than about $2{\cdot}10^{12}{\rm cm}^{-2}$). At low $N_{\rm inv}$ values the measured mobility shows a clear roll-off, stronger than in the reference data, most likely due to scattering by Coulomb centers in the high-κ stack [27]. Fig.5.3(a) shows that the electron mobility in our samples is also larger than the mobility extracted in the n-type FinFETs of [24]; quite interestingly the electron mobility in either FinFETs is larger than the one in planar SiO_2 MOSFETs [23, 25, 26]. This comparison to literature data suggests that the FinFETs of this work have high quality interfaces and, furthermore, that both in our samples and in the devices of [24] a physical mechanism must be responsible for a significant electron mobility enhancement with respect to planar (110) MOSFETs. Fig.5.3(b) finally shows that the hole mobility is comparable to available literature data for both (110) planar p-MOSFETs [25, 26] and (110) p-type FinFETs [24].

5.2.3 Mobility modelling and comparison with the experiments

The transport analysis is focused on the lateral (110)/⟨110⟩ interfaces of the FinFETs simulated as double-gate devices. For a fin height of 65nm this assumption is appropriate for fin widths below approximately 20nm. In this study we did not include the remote phonon and remote Coulomb scattering mechanisms in our calculations, which may be relevant for mobility in high-κ transistors [27, 28]. In the devices of this work, due to the essentially undoped channel, such scattering mechanisms are expected to affect the mobility mainly at small inversion densities. Furthermore, the impact of remote Coulomb scattering increases at low temperatures because of the relative phonon scattering suppression [29]. Because of this simplification in the model, the comparison between the mobility simulations and the experiments is not quantitatively meaningful at small inversion densities, especially at low temperature.

In the following the models for the electron and hole mobility calculation will be summarized and the low field transport results will be presented.

Electron inversion layers

The n-type devices have been simulated with the Multi Subband Monte Carlo (MSMC) approach that accounts for the subband quantization and for the most important scattering mechanisms [30]. The electron band structure is described according to the effective mass approximation and non-parabolic corrections are used for the energy dependence on the wave-vector $\mathbf{k}$ in the transport plane. The effect of arbitrary crystal and transport directions is accounted for by an appropriate calculation of the relevant quantization and transport effective masses [30].

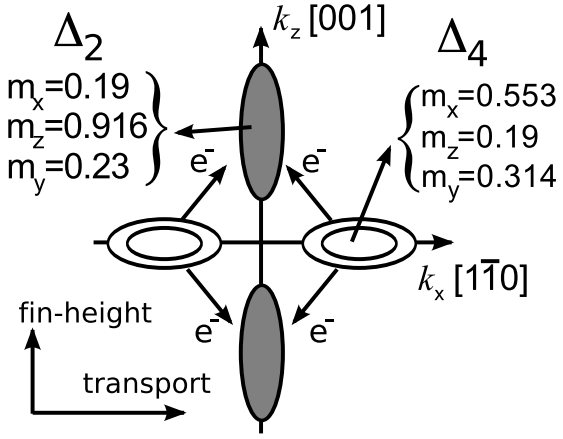


Figure 5.4: Sketch illustrating the Δ_4 and Δ_2 electron valleys in the (110)/[1$\bar{1}$0] silicon inversion layer corresponding to the sidewall interface of the FinFETs as defined in Fig.5.1. Consistently with Fig.5.1, m_{x}, m_{y} and m_{z} are the effective masses respectively in the fin length, fin width and fin height direction. In an unstrained (110) inversion layer the Δ_4 are the lowest valleys due to the largest effective mass m_{y} in the fin width direction.

Fig.5.4 shows a sketch of the Δ_4 and the Δ_2 valleys in a (110)/[1$\bar{1}$0] electron inversion layer. In unstrained silicon the Δ_4 valleys are the lowest due the largest quantization mass m_{y}=0.315m$_0$. It has been argued that, due to the non-parabolicity of the Δ bulk silicon valleys in the [110] direction, the parabolic effective mass approximation is inaccurate for the Δ_2 valleys in (110) inversion layers [31]. To overcome this limitation we used a quantization model based on the Linear Combination of Bulk Bands (LCBB [32]) to extract an empirically modified quantization mass $m_{y,\mathrm{eff}}$=0.23m$_0$ for the Δ_2 valleys. The effective mass approximation can achieve a good agreement with the LCBB results by using such an adjusted value of the quantization mass $m_{y,\mathrm{eff}}$ [33].

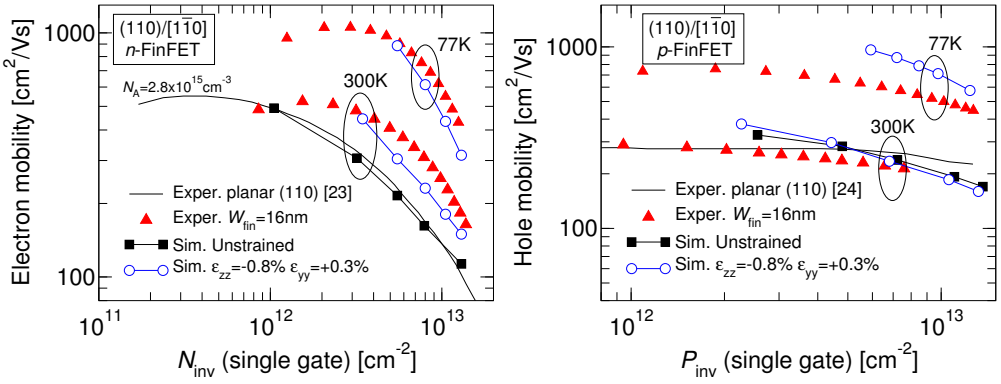
Hole inversion layers

Hole mobility was calculated with the self-consistent **k·p** solver described in Chapter 2. We accounted for acoustic phonon, optical phonon and surface roughness scattering and the scattering parameters for electrons and holes has already been reported in Tab.4.2, which result in a good agreement with electron and hole mobility in bulk MOSFETs [15]. These parameters were used for all the simulations throughout this chapter and not changed with the strain conditions.

Strain effects on electron and hole mobility

The inclusion of strain in our simulations starts by converting the T_{fH}, T_{fW}, T_{fL} stress components in the DCS to the corresponding strain matrix $\boldsymbol{\varepsilon_c}$ in the CCS (see Fig.5.1), by using appropriate rotation matrices and the silicon elastic compliance constants [14, 34, 35]. Then for electrons we calculated, for each strain configuration $\boldsymbol{\varepsilon_c}$, the Δ_4 and the Δ_2 valley splitting and the modifications of the effective masses by using the analytical expressions reported in [36], which were derived from a two band **k·p** approach for the conduction band and validated against pseudo-potential calculations [37]. These values were directly used in the MSMC simulations [14].

For holes, instead, the $\boldsymbol{\varepsilon_c}$ enters directly the matrix of the **k·p** calculations [6], and consequently the band-structure and the mobility determination. As already illustrated in [15], experimental electron and hole mobility enhancements in uniaxially strained (110)/⟨110⟩ planar devices are reproduced well by our models with the same model parameters used for unstrained (110) planar inversion layers.



(a) (b)

Figure 5.5: Measured and simulated electron mobility (a) and hole mobility (b) in n-FinFETs at 300K and 77K. Strain values ε_{zz}=−0.8% and ε_{yy}=+0.3% were taken from the measurements reported in Fig.5.2.

To quantitatively investigate the effect of the stress along the fin height direction T_{fH}, Fig.5.5 shows that the electron and hole mobility simulations obtained by using the measured strain values of Fig.5.2 are in fairly good agreement with measured data; in this respect, the DCS strain configuration of Fig.5.2 (ε_{zz}=−0.8% and ε_{yy}=+0.3%) corresponds to a dominant T_{fH} stress component of approximately −1.1GPa (compressive). The mobility calculations have not been extended to small inversion densities N_{inv} and P_{inv} because, as already mentioned, we did not account for the remote Coulomb scattering due to possible fixed charges in the high-κ stack, which is expected to be relevant at low N_{inv} and P_{inv} values.

The physical interpretation of the stress induced electron mobility enhancements is illustrated in Fig.5.6 reporting the relative occupation of the Δ_2 and Δ_4 valleys (see Fig.5.4) versus

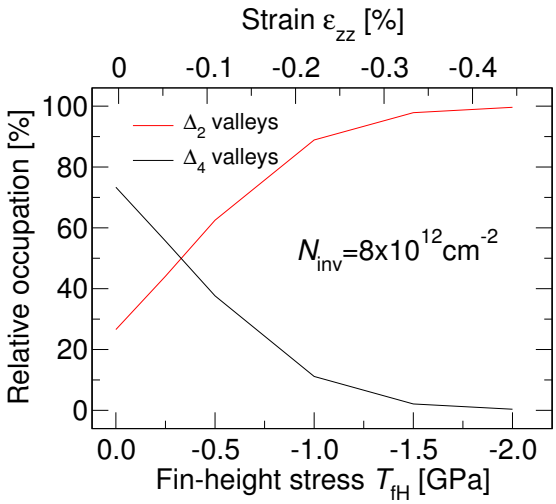


Figure 5.6: (110)/[1$\bar{1}$0] n-FinFET. $W_{\rm fin}$=25nm. Simulated re-population of the Δ_2 valleys versus compressive $T_{\rm fH}$ (bottom x-axis). The top x-axis shows the corresponding axial strain component ε_{zz} in the fin height direction (see Fig.5.1). These results are approximately independent of $W_{\rm fin}$ for fin widths down to approximately 10nm.

the compressive stress $T_{\rm fH}$. As it can be seen, a compressive $T_{\rm fH}$ yields a remarkable re-population of the Δ_2 valleys due to a lowering of the corresponding minima in the strained silicon conduction band minima [14]. Since the Δ_2 have a smaller transport mass m_x than the Δ_4 valleys, as sketched Fig.5.4, the re-population of the Δ_2 valleys results in a significant electron mobility enhancement.

The compressive stress in the fin height direction is consistent with the strain induced by differences in the thermal expansion coefficient following plastic relaxation at high temperature [15].

Since the T_{fH} is mainly ascribed to the metal gate processing, it is relevant to understand if stress components additional to the T_{fH} can provide further improvements in the electron and hole mobility and to identify the most effective stress configurations. To this purpose we performed a large number of mobility simulations where T_{fL} and T_{fW} are varied for a fixed T_{fH}=-1GPa. The results are illustrated in Figs.5.7 in the form of contour plots for the electron (a) and hole (b) mobility enhancements in the $T_{\rm fL}$, $T_{\rm fW}$ plane. Since T_{fH} is -1GPa the mobility enhancements are non null for T_{fL}=T_{fW}=0. It can be seen that the tensile $T_{\rm fL}$ further improves electron mobility with respect to the values obtained with only a compressive $T_{\rm fH}$; this effect is mainly related to the reduction of the transport effective mass m_x of the Δ_2 valleys with the strain [14]. Furthermore the compressive $T_{\rm fL}$ increases the hole mobility. It is worth noting that the combination of tensile $T_{\rm fL}$ and compressive $T_{\rm fW}$ stress is more effective in enhancing electron mobility than compressive $T_{\rm fL}$ and tensile $T_{\rm fW}$ are for the hole mobility. In fact, for T_{fH}=-1GPa, the former configuration can reach electron mobility enhancements of about 120%, while the latter leads to about a 50% hole mobility increase for $T_{\rm fL}\simeq-1.5$GPa.

We can summarize the results of this section by noting that the optimal stress configuration requires compressive $T_{\rm fH}$ and tensile $T_{\rm fL}$ for n-FinFETs, whereas compressive $T_{\rm fL}$ enhances hole mobility. At the same time, a compressive and tensile $T_{\rm fW}$ can further improve carrier mobility in n-FinFETs and p-FinFETs, respectively.

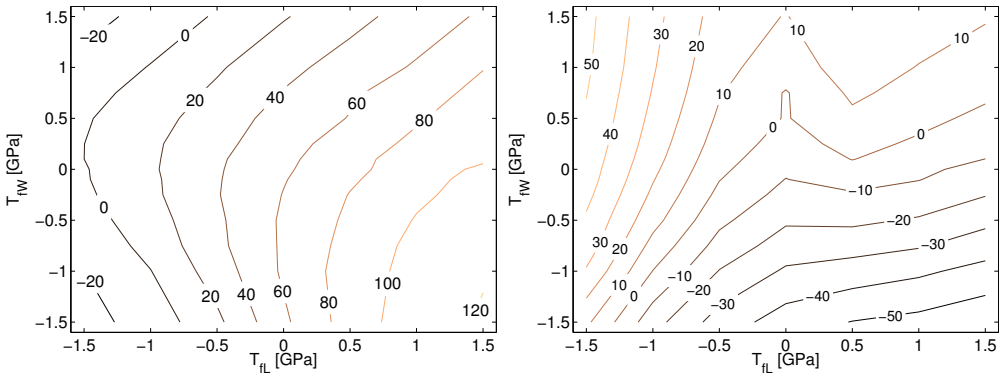


Figure 5.7: (a) (110)/[1$\bar{1}$0] n-FinFET. Simulated stress induced electron mobility variations with respect to the unstrained case versus T_{fL} and T_{fW} for T_{fH}=-1GPa. (b) Same for p-FinFET. The inversion density is N_{inv}=8×10^{12} cm^{-2} for both devices.

5.2.4 Drain current in strained n-FinFETs

The same transport model used so far for the calculation of the electron mobility has been employed to extend our analysis to the simulation of nanoscale n-FinFETs and to study the possible drain current improvements in the presence of multiple stress components. As in the case of mobility analysis, we simulated DG-SOI transistors corresponding to a section of the FinFET structure normal to the fin height, hence the I_{DS} values will be expressed in Ampere per unit width. The transistors feature a channel length L_{G}=25nm, a fin width W$_{fin}$=10nm and and equivalent oxide thickness of about 1.1nm, which result in an almost ideal sub-threshold slope SS=67mV/decade and in a drain induced barrier lowering of about 50mV/V. The series resistances were accounted for as external lumped elements and the V_{GS} and V_{DS} values reported hereafter are always the extrinsic voltages. Unless otherwise stated, we used R_{SD}=140$\Omega\mu$m by following the projections of the ITRS for the 22nm technology node [38]. It should be noticed that a possible effect of strain on the value of series resistance was not taken into account.

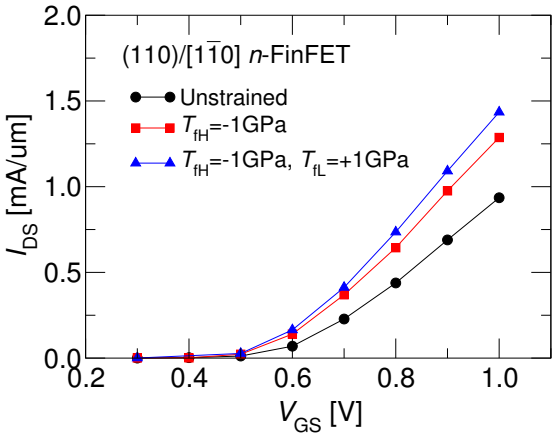


Figure 5.8: (110)/[1$\bar{1}$0] n-FinFET. L_{G}=25nm. Simulated drain current versus gate voltage in the unstrained (circles) case and for T_{fH}=−1GPa (squares) and T_{fH}=−1GPa, T_{fL}=+1GPa (triangles). The series resistance is R_{SD}=140$\Omega\mu$m according to the ITRS [38].

Fig.5.8 shows the simulated $I_{\rm DS}$ versus $V_{\rm GS}$ characteristics for both unstrained and strained (110)/[1$\bar{1}$0] n-FinFETs. We assumed that the threshold voltage $V_{\rm T}$ can be set to the same value for all the devices and stress configurations thanks to an ideal work function flexibility; in particular, the saturation $V_{\rm T}$, defined as the $V_{\rm GS}$ yielding $I_{\rm DS}$=1μA/μm for $V_{\rm DS}$=$V_{\rm DD}$, is about 450mV for all the devices. We analyzed in detail two strain configurations: the first one consisting of only one stress component $T_{\rm fH}$=−1GPa and the second one with $T_{\rm fH}$=−1GPa and $T_{\rm fL}$=+1GPa. According to Fig.5.7 the second stress configuration allows for a significant electron mobility improvement with respect to the case where only $T_{\rm fH}$ is used. Fig.5.8 shows that the on-current I_{ON} (here defined as the $I_{\rm DS}$ at $V_{\rm DS}$=$V_{\rm GS}$=1V) is appreciably enhanced by the stress, but the relative I_{ON} improvements are smaller than the mobility enhancements reported in Fig. 5.7. The additive effect of different strain components on the I_{ON} has been qualitatively confirmed by experiments reporting stress induced $I_{\rm DS}$ improvements in nanoscale FinFETs [15].

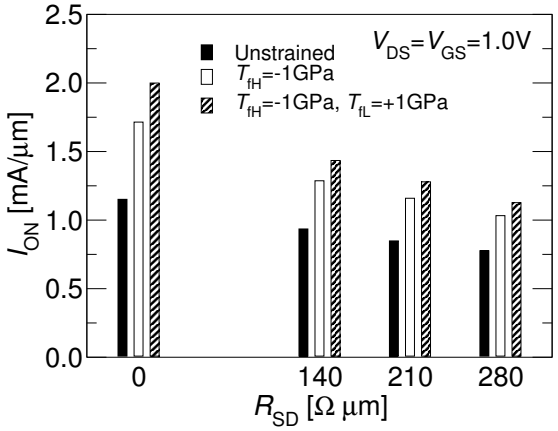


Figure 5.9: (110)/[1$\bar{1}$0] n-FinFET. $L_{\rm G}$=25nm. On-current as a function of series resistance for $V_{\rm DS}$=$V_{\rm GS}$=1.0V in the unstrained and strained cases. Same device and stress levels as in Fig.5.8.

Fig.5.9 shows the simulated I_{ON} for different strain conditions and for several values of the series resistance R_{SD}. As it can be seen, the strain induced I_{ON} enhancement is substantially reduced with the increase of the R_{SD} value. These results emphasize that the optimization of the series resistances is an essential step in order to fully exploit the technology boosters employed at the intrinsic device level, such as the stress engineering.

In order to interpret the strain induced I_{ON} improvements in Fig.5.9 we examined the basic constituents of the I_{ON} according to the quasi-ballistic transport model for MOS transistors [39]

$$I_{ON} \simeq Q_{inv}\, v^{+} \left[\frac{1-r}{1+r}\right] \tag{5.2}$$

where Q_{inv} is the density of inversion charge at the virtual source, v^{+} is the velocity at the virtual source for the electrons moving from source to drain and r is the channel reflection coefficient. In particular, Tab.5.1 reports, for the same device and stress conditions as in Fig.5.8, the velocity v^{+}, the reflection coefficient r as well as the long channel mobility μ_{LC}, to which r is known to be related in a quasi-ballistic transistor [39]. The μ_{LC} has been extracted in a long channel device for the same inversion density calculated at the virtual source of the 25nm n-FinFET. It is interesting to note that the stress results in a substantial enhancement of the injection velocity due to the favorable subband re-population discussed in Section 5.2.3, whereas the improvements of the backscattering coefficient are only modest.

Stress		v^+ [10^7cm/s]	r	μ_{LC} [cm^2/Vs]
Unstrained	R_{SD}=0 $\Omega\mu$m	0.96	0.19	144
	R_{SD}=140 $\Omega\mu$m	0.93	0.19	
T_{fH}=−1GPa	R_{SD}=0 $\Omega\mu$m	1.46	0.17	192
	R_{SD}=140 $\Omega\mu$m	1.29	0.18	
T_{fH}=−1GPa, T_{fL}=+1GPa	R_{SD}=0 $\Omega\mu$m	1.56	0.17	333
	R_{SD}=140 $\Omega\mu$m	1.43	0.16	

Table 5.1: Comparison of injection velocity v^+ at the virtual source, backscattering coefficient r and low field, long channel mobility μ_{LC} for the same device and stress conditions as in Fig.5.8 for V_{DS}=V_{GS}=1.0V. Results are shown both for realistic series resistance and for a null series resistance.

5.3 Surface roughness scattering in biaxially strained MOSFETs

5.3.1 Context of this study

Strain is recognized as one of the most effective boosters for the silicon CMOS technologies [40] and the device optimization requires the understanding of the physical mechanisms underlying the strain-induced mobility enhancements. At this regard, however, even for the supposedly simple biaxial strain configuration the behavior of the electron and hole mobility is still poorly understood at large inversion densities, where the mobility is mainly limited by surface roughness (SR) scattering.

In fact it is known that a strain induced reduction of the SR spectrum is needed to obtain a good agreement between the simulated mobility results and the experimental data for n-MOSFETs [41], and such a decrease of the interface roughness has been also reported in experiments [42, 8] and theoretically predicted by *ab initio* calculations [43]. However several experimental sets have shown that, for a biaxial strain able to induce a significant electron mobility enhancement, the corresponding hole mobility change is much smaller than for the electrons [44, 45, 46, 47, 48]. To explain such a behavior it was recently proposed that the strain reduces not only the r.m.s. value Δ but also the correlation length Λ of the surface roughness spectrum [49]. Such an interpretation stemmed from the analysis of experiments at 4.2K and by using a single subband mobility model that assumes parabolic circular bands for both electrons and holes, includes only surface roughness scattering and neglects the effect of the screening [49]. The simplifications in the modeling approach proposed in [49] call for a detailed analysis of the available experiments with a more complete model.

This section presents an extended investigation and a critical re-examination of the interpretation proposed in [49], and shows a systematic comparison between several sets of mobility experiments at room temperature and the results of physically based mobility simulations. Our models go well beyond the simplified approach of [49] and account for a realistic energy dispersion for both electrons and holes, for all the most relevant scattering mechanisms and for the large effect produced by the carrier screening at high inversion densities. Our results show that the overall agreement between simulations and experiments is improved by assuming that the biaxial strain reduces the SR correlation length Λ and therefore substantiate the assumption made in [49] with the support of detailed and accurate models.

5.3.2 Numerical mobility models

Two different approaches were used to study the low field mobility in electron and hole inversion layers. The electron mobility was obtained through the use of a Multi Subband Monte Carlo

(MSMC) simulator that accounts for the subband quantization and for the most important scattering mechanisms: acoustic and optical phonons, SR and ionized impurities [30]. The strain induced splitting of the Delta silicon valleys was accounted for by following [36].

For hole inversion layers the mobility was calculated starting from the band-structure obtained with a self-consistent **k·p** solver coupled to the Poisson equation, and then by using the Momentum Relaxation Time (MRT) approximation and accounting for phonons and surface roughness scattering according to the model presented in Chapter 4.

For both electrons and holes we used an exponential SR power spectrum $S(q)$ (see App.A)

$$S(q) = \frac{\pi\Delta^2\Lambda^2}{(1+q^2\Lambda^2/2)^{3/2}} \tag{5.3}$$

and we accounted for the screening produced by the carriers in the inversion layer by using the dielectric function approach [50, 51].

5.3.3 Simulation results

Here we proceeded in a different way with respect to what done in Sec.5.2 for FinFET devices. In particular, we calibrated our scattering models in order to reproduce electron and hole mobility experiments with the same parameter set. Fig.5.10 compares the experimental data [52] and the simulated mobility results as a function of the effective field F_{eff} for lightly doped unstrained bulk n- and p-MOS transistors. The same set of surface roughness parameters, that is Δ=0.65nm and Λ=1.2nm, was used for both electrons and holes.

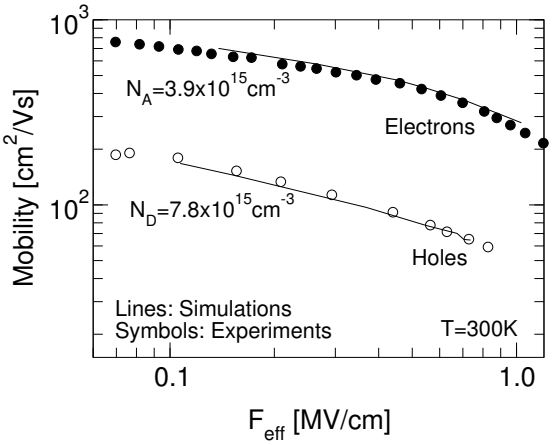


Figure 5.10: Unstrained Si (001)/[110] electron and hole mobility versus effective field at room temperature. Lines: simulated electron and hole mobility. Symbols: experimental data [52]. Δ=0.65nm and Λ=1.2nm for both electron and hole simulations.

The experimental data are taken from [44, 45, 46, 47, 48] simulation results were obtained considering three different sets of SR parameters. In the set A the Δ and Λ values are the same as in Fig.5.10; in the set B the Δ is as in Fig.5.10 and the Λ is reduced as the Ge content increases; in the set C the Δ is decreased whereas the Λ is kept as in Fig.5.10. The Λ and Δ value respectively in set B and set C have been chosen to give the same electron mobility enhancements with the two sets (see Fig.5.11(a)). As for this latter choice, we here notice that the key question addressed in this work is whether a single set of Δ and Λ values (set A) and the two extreme scenarios of a single modulation of these parameters with the strain (set B and C) is able to explain both the electron and the hole mobility improvements. In this respect, we believe that forcing set B and C to give the same electron mobility enhancements reduces the arbitrariness by which one may change Λ and Δ with the strain. The three sets of SR

parameters are summarized in Tab.5.2. Fig.5.11 shows the simulated mobility enhancements

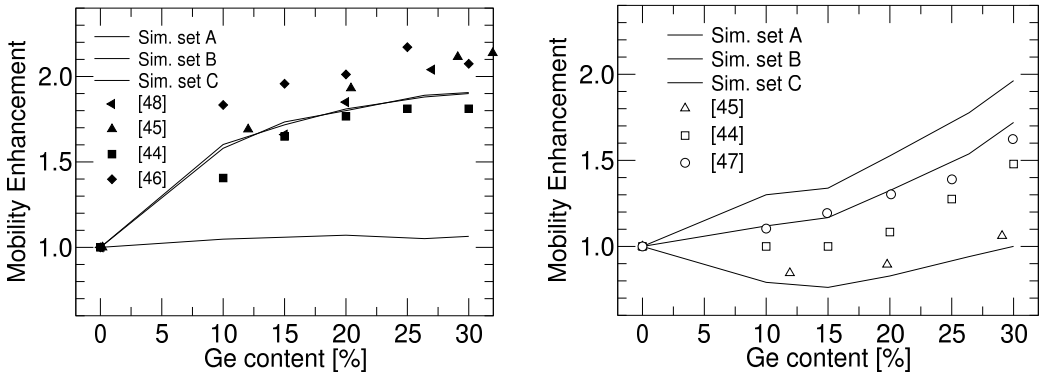


Figure 5.11: Electron (a) and hole (b) mobility enhancements as a function of the Ge content for biaxially strained MOSFETs. Effective field F_{eff}=0.7 MV/cm. Symbols denote the experimental values [44, 45, 46, 47, 48], while lines are the simulations for the Δ and Λ values corresponding to the sets A, B and C of Tab.5.2.

and the experimental data as a function of the Ge content in the SiGe substrate for both n-MOS and p-MOSFETs at a high effective field F_{eff}=0.7MV/cm.

	A		B		C	
Ge[%]	Δ[nm]	Λ[nm]	Δ[nm]	Λ[nm]	Δ[nm]	Λ[nm]
0	0.65	1.2	0.65	1.2	0.65	1.2
10	0.65	1.2	0.65	0.55	0.38	1.2
15	0.65	1.2	0.65	0.45	0.34	1.2
20	0.65	1.2	0.65	0.4	0.3	1.2
26.4	0.65	1.2	0.65	0.38	0.28	1.2
30	0.65	1.2	0.65	0.35	0.25	1.2

Table 5.2: The three sets A, B, and C of SR parameters Δ, Λ.

As expected, with the set A the simulated electron mobility enhancements (dashed lines in Fig.5.11(a)) cannot reproduce the experiments: in fact the simulated electron mobility enhancement is essentially negligible [41]. Also the corresponding simulated hole mobility enhancements underestimate the experimental values. On the other hand, the simulated hole mobility enhancements are in quite good agreement to the experiments when set B is used, whereas set C yields enhancements somewhat larger than the experimental ones (see Fig.5.11(b)).

In Fig.5.12 we extended the analysis to different F_{eff} values using the parameters for the surface roughness scattering corresponding to set B. Fig.5.12(a) shows that the insensitivity of the electron mobility enhancements to F_{eff} obtained with the simulations is in good agreement with the experiments in [48]. In Fig.5.12(b) we compare simulated and measured hole mobility enhancements for F_{eff}=0.7MV/cm and 0.9MV/cm; the data at F_{eff}=0.9MV/cm were obtained by extrapolating the curves reported in Fig.8 of [44]. As it can be seen the sensitivity of the hole mobility enhancement to F_{eff} is somewhat overestimated in the simulations, however the overall agreement between simulations and measurements remain fairly satisfactory in the explored F_{eff} range. The results of Fig.5.12 show that the F_{eff} taken for the analysis is not so critical as long as F_{eff} is large enough for the surface roughness to be the dominant scattering mechanism.

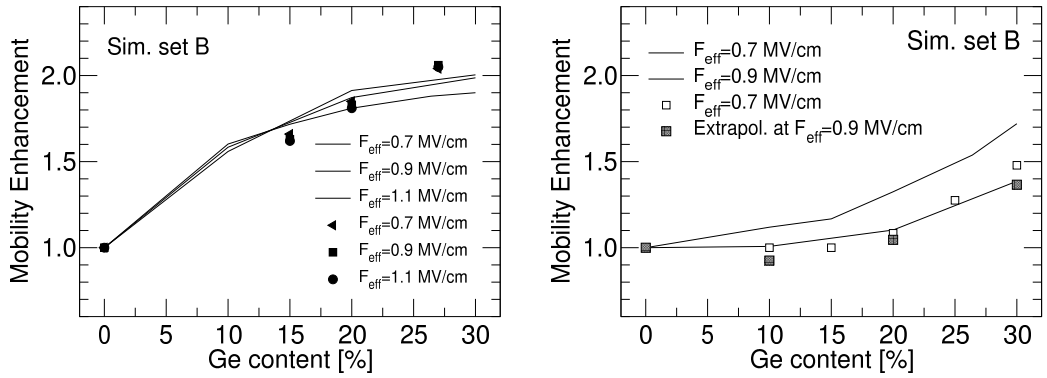


Figure 5.12: Electron (a) and hole (b) mobility enhancements as a function of the Ge content for different F_{eff}. Symbols denote the experimental values, while lines are the simulations obtained by using the surface roughness parameters corresponding to set B of Tab.5.2. The experimental values were taken from [44] and [48] for (a) and (b), respectively.

We also studied the strain induced mobility enhancements in the temperature range from 230K to 300K. In our simulations both electron and hole mobility enhancements are fairly insensitive to the temperature down to 230K (not shown), and this behavior is in reasonably good agreement with the available experimental data.

5.3.4 Discussion

In this section we interpret the results of Fig.5.11 by analyzing the role of the SR spectrum on the n- and p-MOSFET mobility. To this purpose we first define $q_{aver}^{(e)}$ and $q_{aver}^{(h)}$ as the average of the magnitude of the wavevector $\mathbf{q}=(\mathbf{k}'-\mathbf{k})$ exchanged during a SR scattering event for electrons and holes, respectively. The value for $q_{aver}^{(e)}$ is determined by definition as the centroid of the q values recorded during the MSMC simulations, whereas $q_{aver}^{(h)}$ has been calculated as

$$q_{aver}^{(h)} = \frac{\sum_{\mathbf{k},n} f_0 \sum_{\mathbf{k}',n'} |\mathbf{k}'-\mathbf{k}| S_{n,n'}(\mathbf{k},\mathbf{k}')(1-f_0')}{\sum_{\mathbf{k},n} f_0 \sum_{\mathbf{k}',n'} S_{n,n'}(\mathbf{k},\mathbf{k}')(1-f_0')} \tag{5.4}$$

where $\mathbf{k}$, $\mathbf{k}'$ are the wavevectors before and after the scattering event, n and n' are the initial and final subbands and f_0 and f_0' are the Fermi-Dirac occupation function $f_0(E_n(\mathbf{k}))$ and $f_0(E_{n'}(\mathbf{k}'))$, respectively.

Fig.5.13 reports the SR spectrum as a function of q for the sets A, B and C and the average exchanged wavevectors (symbols) for electrons and holes in both the unstrained and strained case.

Note that neither $q_{aver}^{(e)}$ nor $q_{aver}^{(h)}$ have a pronounced sensitivity to the Δ and Λ parameters of the roughness spectrum. As it can be seen for all the (Δ, Λ) sets the average exchanged wavevector $q_{aver}^{(e)}$ for electrons is smaller than $q_{aver}^{(h)}$. It is worth pointing out that the set B features a SR spectrum reduction (with respect to the unstrained spectrum, set A) almost identical to set C for q close to $q_{aver}^{(e)}$. For q values close to $q_{aver}^{(h)}$, instead, the set C results in a larger reduction of the SR spectrum (w.r.t. set A). As a result, set C provides a hole mobility enhancement higher than set B (see Fig.5.11(b)).

We have also highlighted the average exchanged wavevectors for both nMOS and pMOSFETs: $q_{aver}^{(e)}$ is smaller and less sensitive to the Ge content than $q_{aver}^{(h)}$. In fact $q_{aver}^{(h)}$ tends to increases with the Ge mole fraction. Note that for small Ge content a fast increase of the

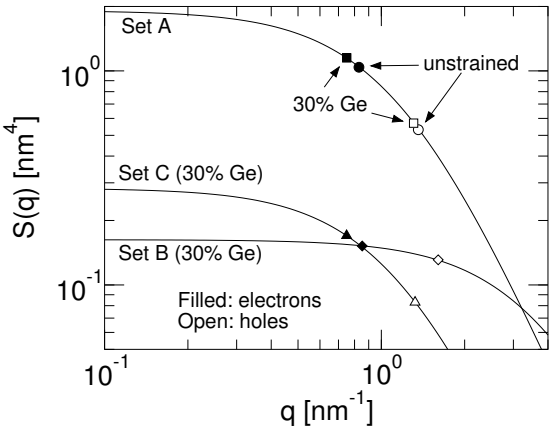


Figure 5.13: Lines: SR spectrum $S(q)$ as a function of the exchanged wavevector q obtained with the three different sets of SR parameters A, B and C. The unstrained and the strained case (30% of Ge content) are reported. Symbols: average exchanged wavevector of electrons (filled symbols) and holes (open symbols) corresponding to each set of parameters.

electron mobility is found (see Fig.5.11) because $q_{aver}^{(e)}$ is centered in a wavevector range where the reduction of the spectrum is higher (see Fig.5.13). A similar mobility increase is not found in the case of holes because $q_{aver}^{(h)}$ is in the range of higher q where the reduction of the spectrum is lower.

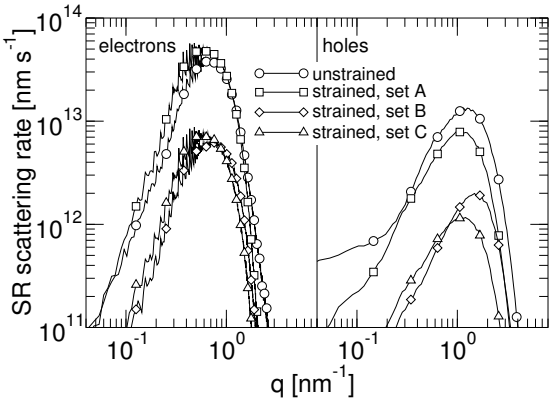


Figure 5.14: SR scattering rates for electrons (on the left) and holes (on the right) as a function of q=$|\mathbf{k}'-\mathbf{k}|$ in the unstrained and in the strained case (30% of Ge content) using sets A, B and C.

In order to gain further insight about the SR scattering in unstrained and strained transistors, Fig.5.14 reports the average SR scattering rate versus q=$|\mathbf{k}'-\mathbf{k}|$ for electrons and holes and for the three (Δ, Λ) sets of Tab.5.2. More precisely, the plots illustrate the average number of SR scattering events having $|\mathbf{k}'-\mathbf{k}|$=q per unit time and unit q, so that the integral over q of the curves in Fig.5.14 provides the average SR scattering rate in the inversion layer. The results in Fig.5.14 have been normalized to the unstrained case for both electrons and holes. As it can be seen, the maxima of the electron statistics are centred at lower q with respect to holes, which is consistent to the $q_{aver}^{(e)}$ and $q_{aver}^{(h)}$ values shown in Fig.5.13.

The set A provides quite the same scattering rates as the unstrained case for both electrons and holes. The slightly larger variation in the hole case can be attributed to the strain-induced

deformation of the bands. As for the electrons, set B and C yield approximately the same reduction of the SR scattering rate w.r.t the unstrained case (Fig.5.14), in fact they were calibrated in order to produce the same mobility enhancements (Fig.5.11(a)). For holes, instead, the set C provides a lower scattering rate and a consistently larger mobility than set B (see Fig.5.11(b)).

5.4 Simulation study of Ge nano-MOSFETs

5.4.1 Context of this study

The competitive edge of Ge compared to Si MOSFETs is a hot topic already investigated for n-MOSFETs by using quantum ballistic simulations [53, 54] and for p-MOSFETs with a semi-classical approach [55]. Recently, promising experimental data for electron and hole mobility in Ge MOSFETs was reported [56, 57, 58], as well as the first data for the electron mobility enhancement in strained Ge (sGe) transistors [59]. Thus a more complete and quantitative study of the on-current $I_{\rm on}$ in nanoscale Ge MOSFETs is now possible.

In this section, we first validate our models comparing our simulations to recent mobility experiments and then study the $I_{\rm on}$ in both n- and p-MOSFETs devices designed for a HP application. We demonstrate that Ge does not outperform strained Si (sSi), but the introduction of uniaxial strain can greatly enhance the Ge performance thanks to band-structure modifications which increase the carrier velocity. The large series resistance of Ge devices, if not solved, can substantially degrade the $I_{\rm on}$.

5.4.2 Transport model and validation

The mobility and the $I_{\rm on}$ were simulated with the Multi-Subband Monte Carlo (MSMC) approach [34] that accounts for the quantization in the inversion layer, for a wide set of scattering mechanisms and for the non-local transport in nanoscale MOSFETs. The MSMC simulators for n-MOS and p-MOS transistors are described in [30, 60] and reproduce well the effective mobility for unstrained n- and p-MOSFETs [30, 60]. The models for surface optical phonons (SOph) and remote Coulomb scattering (remQ) were presented in [27] and have been extended to p-MOSFETs consistently with the hole band structure of [61].

The model has been validated against measured long-channel electron and hole mobilities in [62] for n- and p-MOS transistors of a high-κ metal gate SOI technology. Our model reproduces well both electron and hole mobilities (μ_n and μ_p) by using the same density of fixed charge at the SiO_2-HfO_2 interface. Moreover, our simulations are also in agreement with the measured $I_{\rm DS}(V_{\rm GS})$ characteristics for the corresponding nano-scale 32nm n- and p-MOSFETs [62].

The simulations of Ge n-MOS transistors account for the Λ, Δ and Γ valleys within the effective mass approximation (EMA) for the different crystal orientations. Fig.5.15 shows the conduction band equi-energy curves for germanium. The lowest valleys for germanium are the Λ ones [34], which are strongly anisotropic and non-parabolic. We accounted for the non-parabolicity in the transport plane [63]. We also accounted for the effect of the Δ and Γ valleys, because they are quite close in energy to the Λ valleys, so their contribution to the transport might be non-negligible.

In order to improve the accuracy in the calculation of the different valleys subband minima, whose relative position critically affects the subband population and hence the transport, we used the Linear Combination of Bulk Bands (LCBB) quantization model [32] to extract modified quantization masses m_q to be used in the simplified EMA model. The m_q values used in this work are reported in Tab.5.3 and, to improve the EMA accuracy, we used the Linear Combination of bulk Bands (LCBB) quantization model [32] to extract the modified quantization masses m_q in Tab.5.3. (with respect to the values inferred by transverse and longitudinal

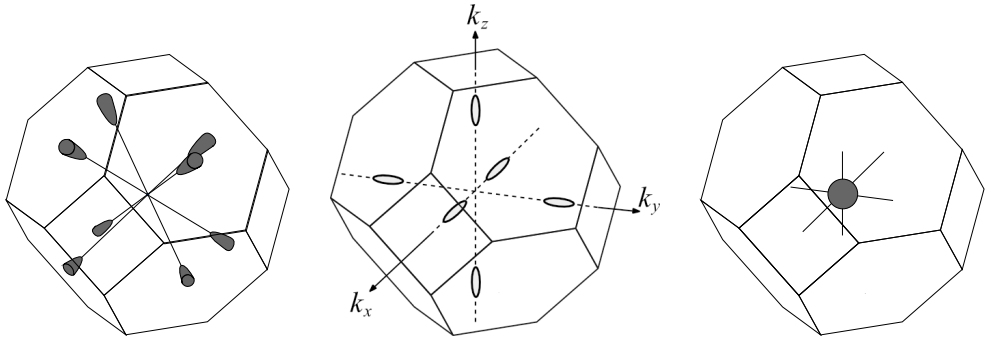


Figure 5.15: Sketch of the main conduction band equi-energy curves in the first Brillouin zone for germanium. (a) Four-fold Λ valleys. The Λ minima lay along the ⟨111⟩ equivalent directions. (b) Six-fold Δ valleys. The Δ minima lay along the ⟨100⟩ equivalent directions. The CCS crystal axis are reported. (c) Spherical Γ valley.

	Ge (100)		**Ge (111)**		**Ge (110)**	
	ν	m_q	ν	m_q	ν	m_q
Λ	4	0.1539 (0.1185)	1	2.0756 (1.588)	2	0.2726 (0.2205)
			3	0.1187 (0.0905)	2	0.1039 (0.081)
Δ	2	0.8355 (1.353)	6	0.3094 (0.3929)	2	0.2341 (0.29)
	4	0.1943 (0.29)			4	0.3136 (0.4776)

Table 5.3: Quantization mass m_q extracted by fitting with an EMA model the subband minima calculated with the LCBB method for triangular wells with different confining electric fields. The values in parenthesis are those inferred from transverse and the longitudinal masses of bulk Ge [63].

masses of bulk Ge [63]). The quantization masses m_q are reported in Tab.5.3 and Fig.5.16 shows the good agreement between the subband minima obtained with either the EMA or the LCBB model (using the same confining potential).

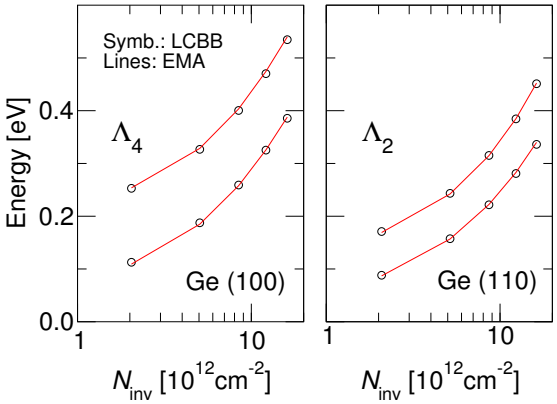


Figure 5.16: Subband minima calculated with the LCBB (symbols) and the EMA model (lines, m_q from Tab.5.3) for the same confining potential obtained with an EMA based self-consistent SchrödingerPoisson solver. The two lowest subbands for the (100) (left) and the (110) (right) crystal orientations are shown.

Parameter name	**Holes**	**Electrons**		
		(100)	(110)	(111)
Surface Roughness:				
Δ[nm]	0.368	1.2	1.2	1.0
Λ [nm]	3	1.3	1.3	1.5
Phonons:				
Acoustic D_{acc} [eV]	11	from [64]		
Optical D_{opt} [10^8eV/cm]	6	from [64]		
Optical Ph. Energy [meV]	38	from [64]		

Table 5.4: Scattering parameters used in all Ge simulations. Intra-valley and inter-valley electron phonons are from [64] but the deformation potential of intra-valley acoustic phonons has been empirically increased by (13/9) (with respect to [64]), namely by the same factor used also in Si to reproduce phonon limited mobility in inversion layers [66, 67]. Hole scattering parameters from [69].

With the phonon set in [64] our simulations reproduce the velocity versus field curves in bulk Ge [65]. For Ge MOSFET simulations, instead, the deformation potential for acoustic phonons is empirically increased with respect to [64] by the same enhancement factor used also in Si inversion layers [66, 67].

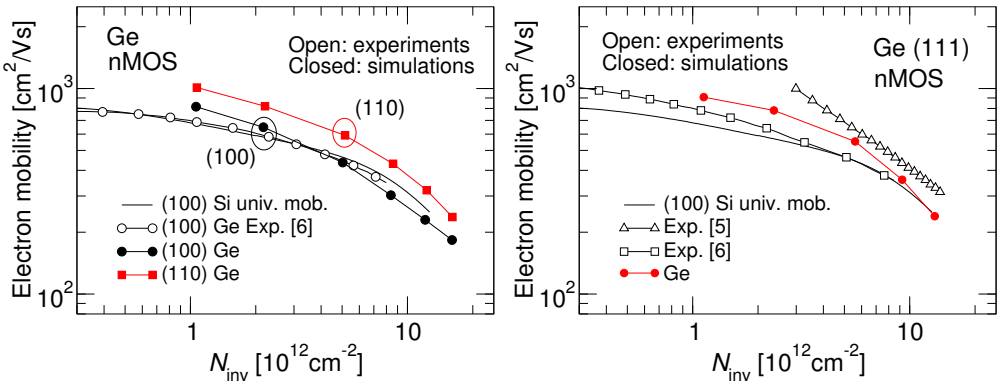


Figure 5.17: (a) Experimental [58] and simulated (100) and (110) Ge electron mobility versus electron inversion density for different crystal orientations. A GeO_2-Al_2O_3 gate stack was assumed with a density N_Q=$6{\cdot}10^{12}cm^{-2}$ of charges at the GeO_2-Al_2O_3 interface (GeO_2 thickness 0.5nm) and a density D_{it}=$4{\cdot}10^{11}cm^{-2}$ of interface states at the Ge-GeO_2 consistent with [57]. (b) Experimental [57, 58] and simulated (111) Ge electron mobility versus electron inversion density for different crystal orientations. Same gate stack and trap densities as in (a). Scattering parameters from Tab.5.4 for both plots. The (100) Si universal mobility curve has been reported for comparison.

Fig.5.17 show the good agreement between experiments [57, 58] and simulations for the μ_n in Ge (100) and (111) MOSFETs, obtained using the experimental value of the fixed charge density at the GeO_2-Al_2O_3 reported in [57]. For the same roughness parameters the effective mobility is somewhat better for the (110) compared to the (100) orientation (see Fig.5.17). The SOph for the GeO_2 dielectric were included according to the GeO_2 parameters in [68]; their impact on mobility is modest (about 3% at high effective field F_{eff}).

The MSMC simulations of Ge *p*-MOSFETs were obtained by calibrating the anisotropic and non-parabolic hole band-structure model of [61] using an extensive comparison to **k·p**

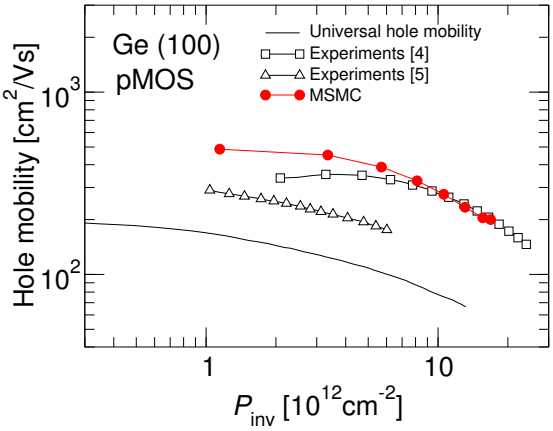


Figure 5.18: Experimental [56, 57] and simulated hole mobility versus hole inversion density. The universal hole mobility has been reported for comparison. Scattering parameters from Tab.5.4.

calculations. Fig.5.18 shows the good agreement with the largest reported experimental μ_p values with the same scattering parameters as in [69]. Tab.5.4 summarizes the scattering parameters for electrons and holes in Ge that were used throughout the paper.

The uniaxial stress in Si n-MOSFETs was described by introducing a splitting between the Δ_4 and Δ_2 subbands and by changing the transport masses of the Δ_2 according to [36]. In order to obtain a good agreement with experimental mobility enhancement we also assumed that the RMS value Δ_{sr} of the surface roughness is reduced by the strain, similarly to the case of the biaxial strain [70]. The strain for Si p-MOSFETs was implicitly accounted for by calibrating the hole band-structure model for each stress condition [71]; no strain dependent Δ_{sr} values were used for p-MOS devices.

The modeling of the uniaxial stress for p-type Si MOSFETs has been described in detail in Chapter 3.

The uniaxial strain in (100) Ge n-MOSFETs was modeled introducing in the simulations the splitting between the $\Lambda_{||}$ and the $\Lambda_{\perp}$ valleys reported in [59] and obtained with DFT calculations; the corresponding strain induced modulation of the transport masses was considered to be negligible [59]. Fig.5.19 illustrates the good agreement between the simulated and experimental μ_n enhancements employing the same scattering parameters as in Fig.5.17 (see Tab.5.4). Fig.5.20(a) shows that the physical mechanism responsible for the μ_n improvement is the stress induced repopulation of the $\Lambda_{\perp}$ valleys that have a smaller transport mass than the $\Lambda_{||}$ valleys (as sketched in Fig.5.20(b)). The population of the Δ_2, Δ_4 and Γ valleys is always negligible.

5.4.3 I_{on} in Si and Ge MOSFETs

The transport model validated in Fig.5.19 was used to simulate high performance SOI double-gate (DG) devices with L_G=25nm and 16nm with either Si or Ge channel. The devices were designed as high performance MOSFETs and ideal work function flexibility is assumed to achieve the same V_T (defined as V_{GS} at I_{DS}=1$\mu A/\mu$m) of about 200mV for all the transistors. Series resistances R_{SD} were accounted for as external lumped elements; unless otherwise specified we used R_{SD}=140$\Omega\mu$m and R_{SD}=160$\Omega\mu$m for L_G=25nm and L_G=16nm, respectively, consistently with the ITRS roadmap. I_{on} is defined as I_{DS} for extrinsic voltages V_{GS}=V_{DS}=1V. The I_{DS} is always reported per unit width, per gate. For Si and sSi devices we assumed an SiO_2-HfO_2 gate stack and for Ge devices a GeO_2 gate dielectric which give EOT=1nm and 0.85nm for L_G=25nm and L_G=16nm, respectively [58]. The appropriate SOph and RCS scattering mechanisms were accounted for, but they have a modest impact on the I_{on}.

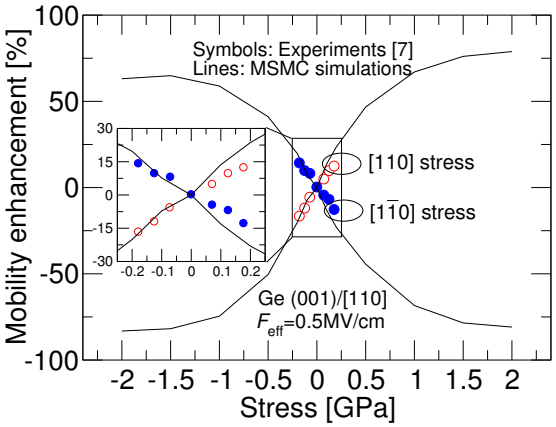


Figure 5.19: (100)/[110] Ge. Measured (symbols, [59]) and simulated (lines) electron mobility versus uniaxial tensile stress along the [110] and [1$\bar{1}$0] directions and at F_{eff}=0.5MV/cm. The inset shows a zoom for low stress values. The simulated mobility enhancement saturates at about 1.5GPa.

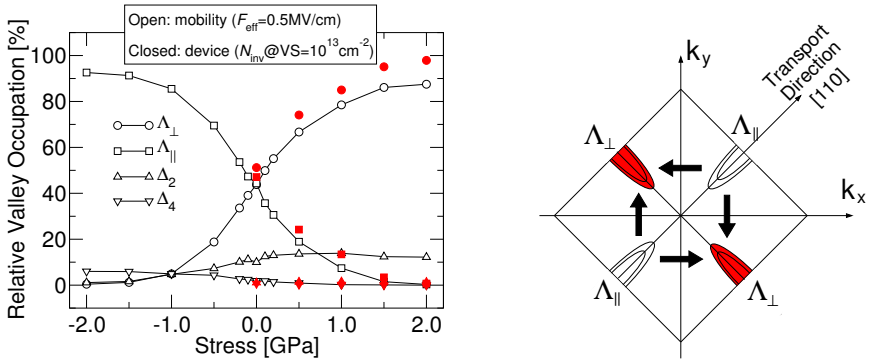


Figure 5.20: (a) (100) Ge. Simulated occupation of the $\Lambda_{||}$, $\Lambda_\perp$ (see (b)), and Δ_2, Δ_4 valleys versus stress along the [110] direction. Open symbols are results obtained with an equilibrium self-consistent Schrödinger-Poisson solver used for mobility calculation and for F_{eff}=$0.5MV/cm$ (as in Fig.5.19). Tensile stress results in a strong repopulation of the $\Lambda_\perp$ valley. The population of the Δ_2, Δ_4 valleys is always negligible. Closed symbols indicate the relative population at the virtual source of a 25nm Ge n-MOS obtained with self-consistent MSMC simulations and for an inversion density at the VS N_{inv}=10^{13}cm^{-2}. (b) Sketch of the Λ valleys for a (100) Ge n-MOSFET with [110] transport direction. The $\Lambda_{||}$ valleys are aligned with transport and have transport mass 1.086m$_0$; the $\Lambda_\perp$ valleys have a remarkably lower transport mass 0.082m$_0$ [63]. The arrows indicate the stress induced repopulation of the $\Lambda_\perp$ valleys (see (a)).

Fig.5.21 shows that unstrained Ge n- and p-MOSFETs are competitive with but do not outperform sSi devices (for a 1.5GPa uniaxial stress). To investigate this point, Fig.5.22 reports the hole equi-energy contour lines for Si, sSi and Ge (same stress as in Fig.5.21) showing that the band curvature for a fixed energy in the [110] direction is similar for sSi and Ge.

At the same time, Fig.5.24(a) demonstrates that the injection velocities at the virtual source (VS) are comparable for sSi and Ge n-MOSFETs. Fig.5.21 also shows that the tensile stress can improve remarkably the I_{on} of Ge n-MOSFETs. Fig.5.23(a) reports the I_{on} versus the stress for both Si and Ge n-MOSFETs; different crystal orientations for unstrained Ge are also

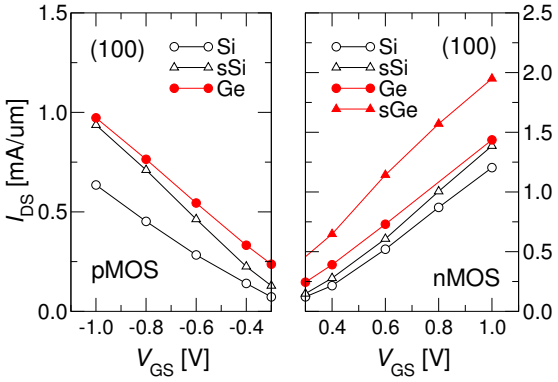


Figure 5.21: Simulated drain current versus gate voltage for the 25nm DG device and for unstrained and strained Si and Ge (EOT=1nm, T_{SI}=11nm). Left: p-MOS. Right: n-MOS. The 1.5GPa stress is along the [110] transport direction and it is tensile and compressive for n-MOS and p-MOS, respectively. The DIBL is about 100mV/V for all the devices.

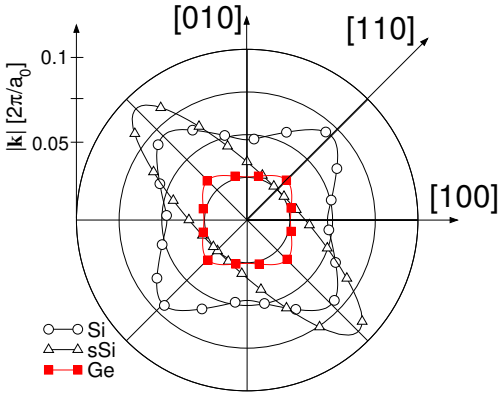


Figure 5.22: (100) Hole equi-energy contour lines at 20meV above the minimum of the lowest subband and for an effective field of F_{eff}=0.7MV/cm. Results from quantized $\mathbf{k}\cdot\mathbf{p}$ calculations.

shown. The I_{on} for (110) and (111) Ge is larger than for sSi, essentially because of the larger injection velocity shown in Fig.5.24(a). Furthermore, the strained (100) Ge can outperform sSi significantly for a given R_{SD} value. Fig.5.23(b) confirms that the I_{on} is similar in Ge and sSi p-MOSFETs up to largest stress values. Fig.5.20 suggests that the I_{on} enhancement is due to the repopulation of the $\Lambda_{\perp}$ valleys at the VS [72], which is the same physical mechanism responsible for the mobility enhancement. Fig.5.24 consistently shows the increase of the injection velocity at the VS of n-MOSFETs with the tensile stress; the reflection coefficient r, instead, is not much improved by the stress. The advantages of the sGe are confirmed also for the MOSFETs scaled to 16nm (see Fig.5.23(c)). Fig.5.23(a) finally shows that, if the R_{SD} of the Ge n-MOSFETs is increased by 50% with respect to the Si devices, then the I_{on} improvement of the sGe is completely lost.

5.5 Resume

In this chapter we have presented the main results obtained by using the semiclassical models developed during the work. We mainly focused our attention to study, understand and optimize

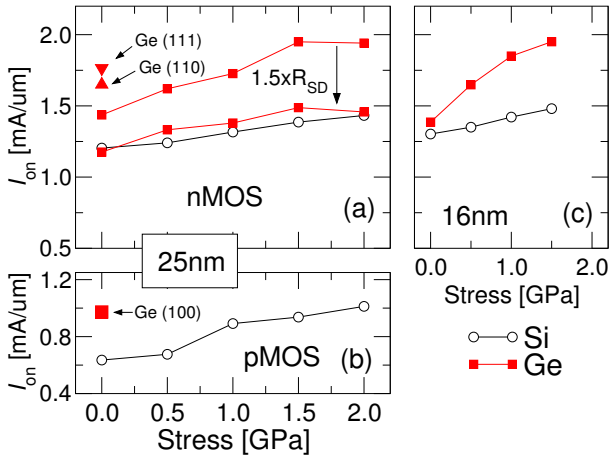


Figure 5.23: (a) $I_{\rm on}$ vs. stress for Si and Ge n-MOSFETs; different crystal orientations for unstrained Ge are also shown. (b) $I_{\rm on}$ versus stress for Si p-MOSFETs; the unstrained Ge is also shown. (c) Same as (a) for a 16nm n-MOSFET (EOT=0.85nm, T_{SI}=7nm).

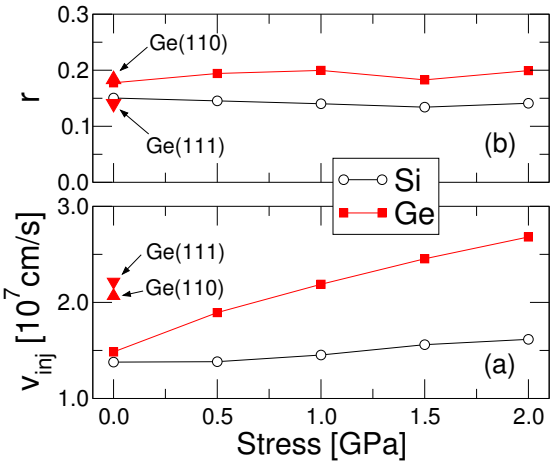


Figure 5.24: Injection velocity at the VS (a) and reflection coefficients (b) for some of the Si and Ge n-MOSFETs of Fig.5.23(a).

the strain in advanced nanoscale MOSFET devices.

Section 5.2 addressed the strain engineering in n- and p-type FinFETs by using direct strain measurements, mobility characterization and physics-based modelling. We identified the presence of a large vertical compressive strain in the transistor fin, which improves the electron mobility substantially without reducing the hole mobility. We noticed that both the electron and the hole measured mobility was found to be fairly insensitive to the fin width in the explored W_{fin} range. Considering the marked electron mobility sensitivity to the strain, one may interpret the mobility independence of W_{fin} as an evidence that the strain configuration is fairly independent of the fin width. The numerical simulations show that the electron mobility can be further improved by using a tensile strain in fin length direction, whereas hole mobility can be increased by using a compressive strain in the same direction. The effect of multiple stressors was investigated also for nanoscale n-type FinFETs by using Multi Subband Monte Carlo simulations. We observed that the stress configurations identified by the mobility analysis yield substantial improvements in the $\rm I_{ON}$ of nanoscale transistors and that a careful

optimization of the series resistance is a crucial step to exploit at best the potentials of the strain engineering.

Section 5.3 presented a simulation study using comprehensive models for electron and hole mobilities to critically examine a recent interpretation for the mobility enhancements in biaxially strained MOSFETs [49], according to which the strain affects the correlation length Λ of the surface roughness spectrum. Our simulations, carefully calibrated against universal mobility curves, have been systematically compared to a large bulk of experiments at room temperature, which allowed us to study the impact on the calculated mobility of different possible assumptions for surface roughness parameters in strained devices. Our results confirmed that a strain-induced reduction of the correlation length Λ of the surface roughness spectrum is a plausible explanation for the mobility enhancement observed in the presence of biaxial strain.

Finally, in Section 5.4 our simulations extensively verified with experiments demonstrated that Ge MOSFETs are competitive with but do not outperform sSi devices in terms of on-current. The sGe has great potentials for n-MOSFETs, however the engineering of the R_{SD} is a crucial issue to exploit the potential advantages of Ge transistors.

Bibliography

[1] M. van Dal, N. Collaert, G. Doornbos, G. Vellianitis, G. Curatola, B. Pawlak, R. Duffy, C. Jonville, B. Degroote, E. Altamirano, E. Kunnen, M. Demand, S. Beckx, T. Vandeweyer, C. Delvaux, F. Leys, A. Hikavyy, R. Rooyackers, M. Kaiser, R. Weemaes, S. Biesemans, M. Jurczak, K. Anil, L. Witters, and R. Lander, "Highly manufacturable FinFETs with sub-10nm fin width and high aspect ratio fabricated with immersion lithography," in *VLSI Technology, IEEE Symposium on*, 2007, pp. 110–111.

[2] G. Vellianitis, M.J.H. van Dal, L. Witters, G. Curatola, G. Doornbos, N. Collaert, C. Jonville, C. Torregiani, L.-S. Lai, J. Petty, B.J. Pawlak, R. Duffy, M. Demand, S. Beckx, S. Mertens, A. Delabie, T. Vandeweyer, C. Delvaux, F. Leys, A. Hikavyy, R. Rooyackers, M. Kaiser, R.G. Weemaes, F. Voogt, H. Roberts, D. Donnet, S. Biesemans, M. Jurczak, and R.J.R. Lander, "Gatestacks for scalable high-performance FinFETs," in *IEEE IEDM Technical Digest*, 2007, pp. 681–684.

[3] S. Suthram, M.M. Hussain, H.R. Harris, C. Smith, H.-H. Cheng, R. Jammy, and S.E. Thompson, "Comparison of Uniaxial Wafer Bending and Contact-Etch-Stop-Liner Stress Induced Performance Enhancement on Double-Gate FinFETs," *IEEE Electron Device Lett.*, vol. 29, no. 5, pp. 480–482, 2008.

[4] K.-M. Tan, T.-Y. Liow, R.T.P. Lee, C.-H. Tung, G.S. Samudra, W.-J. Yoo, and Y.-C. Yeo, "Drive-Current Enhancement in FinFETs Using Gate-Induced Stress," *IEEE Electron Device Lett.*, vol. 27, no. 9, pp. 769–771, 2006.

[5] T.-Y. Liow, K.-M. Tan, D. Weeks, L. Rinus, M. Zhu, K.-M. Hoe, C.-H. Tung, M. Bauer, J. Spear, S.G. Thomas, G.S. Samudra, N. Balasubramanian, and Y.-C. Yeo, "Strained n-Channel FinFETs Featuring In Situ Doped Silicon Carbon ($Si_{1-y}C_y$) Source and Drain Stressors With High Carbon Content," *IEEE Trans. Electron Devices*, vol. 55, no. 9, pp. 2475–2483, 2008.

[6] M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, "Six-band kp calculation of the hole mobility in silicon inversion layers: Dependence on surface orientation, strain, and silicon thickness," *Journal of Applied Physics*, vol. 94, no. 2, pp. 1079–1095, 2003.

[7] T. Krishnamohan, K. Donghyun, 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*, 2008, pp. 899–902.

[8] Y. Zhao, M. Takenaka, and S. Takagi, "Comprehensive understanding of surface roughness and Coulomb scattering mobility in biaxially-strained Si MOSFETs," in *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, Dec. 2008, pp. 1–4.

[9] M. Saitoh, N. Yasutake, Y. Nakabayashi, T. Numata, and K. Uchida, "Comprehensive performance assessment of scaled (110) CMOSFETs based on understanding of STI stress effects and velocity saturation," in *IEEE IEDM Technical Digest*, Dec. 2008, pp. 573–576.

[10] M. Saitoh, A. Kaneko, K. Okano, T. Kinoshita, S. Inaba, Y. Toyoshima, and K. Uchida, "Three-dimensional stress engineering in FinFETs for mobility/on-current enhancement and gate current reduction," in *Symposium on VLSI Technology*, 2008, pp. 18–19.

[11] A. Pham, C. Jungemann, and B. Meinerzhagen, "Modeling of piezoresistive coefficients in Si hole inversion layers," in *IEEE International ULIS conference*, March 2009, pp. 121–124.

[12] P. Zhang, A. Istratov, E. Weber, C. Kisielowski, H. He, C. Nelson, and J. Spence, "Direct strain measurement in a 65 nm node strained silicon transistor by convergent-beam electron diffraction," *Applied Physics Letters*, vol. 89, no. 161907, 2006.

[13] J. Li, D. Anjum, R. Hull, G. Xia, and J. Hoyt, "Nanoscale stress analysis of strained-Si metal-oxide-semiconductor field-effect transistors by quantitative electron diffraction contrast imaging," *App. Phys. Lett.*, vol. 87, no. 22, pp. 222 111.1–222 111.3, Nov. 2005.

[14] N. Serra and D. Esseni, "Mobility enhancement in strained n–FinFETs: basic insight and stress engineering," *IEEE Trans. Electron Devices*, vol. 57, no. 2, pp. 482–490, 2010.

[15] N. Serra, F. Conzatti, D. Esseni, M. De Michielis, P. Palestri, L. Selmi, S. Thomas, T. Whall, E. Parker, D. Leadley, L. Witters, A. Hikavyy, M. Hytch, F. Houdellier, E. Snoeck, T. Wang, W. Lee, G. Vellianitis, M. van Dal, B. Duriez, G. Doornbos, and R. Lander, "Experimental and physics-based modeling assessment of strain induced mobility enhancement in FinFETs," in *IEEE IEDM Technical Digest*, dec. 2009, pp. 71 –74.

[16] M. Hÿtch, F. Houdellier, F. Hüe, and E. Snoeck, "Nanoscale holographic interferometry for strain measurements in electronic devices," *Nature*, vol. 453, pp. 1086–1089, June 2008.

[17] ——, "Patent Application PCT N PCT/FR2008/001302," 2008.

[18] Hüe, Florian and Hÿtch, Martin and Bender, Hugo and Houdellier, Florent and Claverie, Alain , "Direct mapping of strain in a strained silicon transistor by high-resolution electron microscopy," *Phys. Rev. Lett.*, vol. 100, no. 15, p. 156602, Apr 2008.

[19] P. Zeitzoff, C. Young, G. Brown, and Y. Kim, "Correcting Effective Mobility Measurements for the Presence of Significant Gate Leakage Current," *IEEE Electron Device Lett.*, vol. 24, no. 4, pp. 275 – 277, apr. 2003.

[20] D. Esseni, M.Mastrapasqua, G.K.Celler, C.Fiegna, L.Selmi, and E.Sangiorgi, "Low Field Electron and Hole Mobility of SOI Transistors Fabricated on Ultra-Thin Silicon Films for Deep Sub-Micron Technology Application," *IEEE Trans. Electron Devices*, vol. 48, no. 12, pp. 2842–2850, 2001.

[21] T. Rudenko, N. Collaert, S. D. Gendt, V. Kilchytska, M. Jurczak, and D. Flandre, "Effective mobility in FinFET structures with HfO_2 and SiON gate dielectrics and TaN gate electrode," *Microelectronic Engineering*, vol. 80, pp. 386 – 389, 2005, 14th biennial Conference on Insulating Films on Semiconductors.

[22] S. Thomas, T. Whall, E. Parker, D. Leadley, R. Lander, G. Vellianitis, and J. Watling, "Improved effective mobility extraction in MOSFETs," *Solid-State Electronics*, vol. 53, no. 12, pp. 1252 – 1256, 2009.

[23] S.Takagi, A.Toriumi, M.Iwase, and H.Tango, "On the Universality of Inversion-layer Mobilty in Si MOSFETs. Part I- Effect of Surface Orientations," *IEEE Trans. Electron Devices*, vol. 41, no. 12, pp. 2357–62, 1994.

[24] T. Rudenko, V. Kilchytska, N. Collaert, M. Jurczak, A. Nazarov, and D. Flandre, "Carrier Mobility in Undoped Triple-Gate FinFET Structures and Limitations of Its Description in Terms of Top and Sidewall Channel Mobilities," *IEEE Trans. Electron Devices*, vol. 55, no. 12, pp. 3532 –3541, Dec 2008.

[25] H. Irie, K. Kita, K. Kyuno, and A. Toriumi, "In-plane mobility anisotropy and universality under uni-axial strains in n- and p-MOS inversion layers on (100), [110], and (111) Si," in *IEEE International Electron Devices Meeting*, 2004, pp. 225–228.

[26] G. Tsutsui and T. Hiramoto, "Mobility and threshold-voltage comparison between [110]- and (100)-oriented ultrathin-body silicon MOSFETs," *IEEE Transactions on Electron Devices*, vol. 53, no. 10, pp. 2582–2588, Oct. 2006.

[27] P. Toniutti, P. Palestri, D. Esseni, and L. Selmi, "Revised analysis of the mobility and ION degradation in high-k gate stacks: Surface optical phonons vs. remote Coulomb scattering," in *ESSDERC*, Sept. 2008, pp. 246–249.

[28] M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, "Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-K insulator: The role of remote phonon scattering," *Journal of Applied Physics*, vol. 90, no. 9, pp. 4587–4608, 2001.

[29] P.Toniutti, M. Michielis, P.Palestri, F.Driussi, D.Esseni, and L.Selmi, "Understanding the mobility reduction in MOSFETs featuring high-κ dielectrics," in *Proceeding of the International Conference on Ultimate Integration on Silicon (ULIS)*, 2010, pp. 65–68.

[30] L. Lucci, P. Palestri, D. Esseni, L. Bergagnini, and L. Selmi, "Multisubband Monte Carlo Study of Transport, Quantization, and Electron-Gas Degeneration in Ultrathin SOI n-MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 5, pp. 1156–1164, 2007.

[31] K. Uchida, A. Kinoshita, and M. Saitoh, "Carrier transport in (110) nMOSFETs: subband structures, non-parabolicity, mobility characteristics, and uniaxial stress engineering," in *IEEE International Electron Devices Meeting*, 2006, pp. 1019–1021.

[32] D. Esseni and P. Palestri, "Linear combination of bulk bands method for investigating the low-dimensional electron gas in nanostructured devices," *Phys. Rev. B*, vol. 72, no. 16, pp. 165 342.1–165 342.14, Oct 2005.

[33] N. Serra and D. Esseni, "Basic insight about the strain engineering of n-type FinFETs," in *IEEE ULIS Conference*, Aachen, March 2009, pp. 113–116.

[34] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS transistors: Semi-classical modeling and applications*. United Kingdom: Cambridge University Press, 2010.

[35] D. Esseni, F. Conzatti, M. De Michielis, N. Serra, P. Palestri, L. Selmi, "Semi-classical transport modelling of CMOS transistors with arbitrary crystal orientations and strain engineering," *Journal of Computational Electronics*, vol. 8, pp. 209–224, 2009.

[36] E. Ungersboeck, S. Dhar, G. Karlowatz, V. Sverdlov, H. Kosina, and S. Selberherr, "The effect of general strain on the band structure and electron mobility of silicon," *IEEE Transactions on Electron Devices*, vol. 54, no. 9, pp. 2183–2190, 2007.

[37] S. E. Ungersboeck, "Advanced modeling of strained CMOS technology," Ph.D. dissertation, Technischen Universität Wien, Wien, Austria, 2007.

[38] *International Technology Roadmap for Semiconductors: 2009 Update*, 2009.

[39] M. Lundstrom and Z. Ren, "Essential Physics of carrier transport in nanoscale MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no. 1, pp. 133–141, 2002.

[40] S. Thompson, G. Sun, Y. S. Choi, and T. Nishida, "Uniaxial-process-induced strained-Si: extending the CMOS roadmap," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 1010–1020, 2006.

[41] M.V.Fischetti, F.Gamiz, and W.Hansch, "On the enhanced electron mobility in strained-silicon inversion layers," *Journal of Applied Physics*, vol. 92, p. 7320, 2002.

[42] 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, p. 063715, 2008.

[43] M.H.Evans, M.Caussanel, R.D.Schrimpf, and S.Pantelides, "First-Principle Modeling of Double-Gate UTSOI MOSFETs," in *IEEE IEDM Technical Digest*, 2005, p. 600.

[44] M. Currie, C. Leitz, T. Langdo, G. Taraschi, and E. Fitzgerald, "Carrier mobilities and process stability of strained Si n- and p-MOSFETs on SiGe virtual substrates," *J. Vac. Sci. Technol. B*, vol. 19, pp. 2268–2279, 2001.

[45] K. Rim, K. Chan, L. Shi, D. Boyd, J. Ott, N. Klymko, F. Cardone, L. Tai, S. Koester, M. Cobb, D. Canaperi, B. To, E. Duch, I. Babich, R. Carruthers, P. Saunders, G. Walker, Y. Zhang, M. Steen, and M. Ieong, "Fabrication and mobility characteristics of ultra-thin strained Si directly on insulator (SSDOI) MOSFETs," in *IEEE IEDM Technical Digest*, dec. 2003, pp. 3.1.1 – 3.1.4.

[46] S. Olsen, A. O'Neill, P. Dobrosz, S. J. Bull, L. Driscoll, S. Chattopadhyay, and K. Kwa, "Study of strain relaxation in Si/SiGe metal-oxide-semiconductor field-effect transistors," *Journal of Applied Physics*, vol. 97, p. 114504, 2005.

[47] C. Leitz, M. Currie, M. Lee, Z. Cheng, D. Antoniadis, and E. A. Fitzgerald, "Hole mobility enhancements and alloy scattering-limited mobility in tensile strained Si/SiGe surface channel metal-oxide-semiconductor field-effect transistors," *Journal of Applied Physics*, vol. 92, pp. 3745–3751, 2002.

[48] F.Driussi, D.Esseni, L.Selmi, P.-E. Hellstrom, G.Malm, J.Hallstedt, M.Ostling, T.J.Grasby, D.R.Leadley, and X.Mescot, "On the Electron Mobility Enhancement in biaxially strained Si MOSFETs," *Solid State Electronics*, vol. 52, no. 4, pp. 498–505, 2008.

[49] Y.Zhao, M.Takenaka, and S.Takagi, "On surface roughness scattering-limited mobilities of electrons and holes in biaxially tensile-strained Si MOSFETs," *IEEE Electron Device Lett.*, vol. 9, pp. 987–989, 2009.

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

[51] D.Esseni and A.Abramo, "Modeling of Electron Mobility Degradation by Remote Coulomb Scattering in Ultra-Thin Oxide MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 7, 2003.

[52] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the Universality of Inversion-layer Mobilty in Si MOSFETs. Part I- Effect of Substrate Impurity Concentration," *IEEE Trans. Electron Devices*, vol. 41, no. 12, pp. 2357–62, 1994.

[53] T.Low, Y.T. Hou, M.F.Li, C.Zhu, A.Chin, G.Samudra, L.Chan, and D.L.Kwong, "Investigation of Performance Limits of Germanium Double-Gated MOSFETs," in *IEEE IEDM Technical Digest*, 2003, p. 691.

[54] A. Rahman, G. Klimeck, and M. Lundstrom, "Novel channel materials for ballistic nanoscale MOSFETs: bandstructure effects," in *IEEE IEDM Technical Digest*, 2005, pp. 615–618.

[55] T. Krishnamohan, D. Kim, S. Raghunathan, and K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With record high drive currents and < 60mV/dec subthreshold slope," in *IEEE IEDM Technical Digest*, dec. 2008, pp. 947–951.

[56] P. Zimmerman, G. Nicholas, B. De Jaeger, B. Kaczer, A. Stesmans, L.-A. Ragnarsson, D.P. Brunco, F.E. Leys, M.Caymax, G. Winderickx, K. Opsomer, M. Meuris, and M. M. Heyns, "High performance Ge pMOS devices using a Si-compatible process flow," in *IEEE IEDM Technical Digest*, 2006, pp. 655–658.

[57] D. Kuzum, T. Krishnamohan, A. Nainani, Y. Sun, P. Pianetta, H.-P. Wong, and K. Saraswat, "Experimental demonstration of high mobility Ge NMOS," in *IEEE IEDM Technical Digest*, 2009, pp. 1–4.

[58] C. Lee, T. Nishimura, N. Saido, K. Nagashio, K. Kita, and A. Toriumi, "Record-high electron mobility in Ge n-MOSFETs exceeding Si universality," in *IEEE IEDM Technical Digest*, 2009, pp. 1 –4.

[59] M. Kobayashi, T. Irisawa, B. Magyari-Kope, K. Saraswat, H.-S. Wong, and Y. Nishi, "Uniaxial Stress Engineering for High-Performance Ge NMOSFETs," *IEEE Trans. Electron Devices*, vol. 57, no. 5, pp. 1037–1046, may 2010.

[60] M. D. Michielis, D. Esseni, P. Palestri, and L. Selmi, "Semiclassical Modeling of Quasi-Ballistic Hole Transport in Nanoscale pMOSFETs Based on a Multi-Subband Monte Carlo Approach," *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 2081–2091, 2009.

[61] M. De Michielis, D.Esseni, and F.Driussi, "Analytical models for the insight into the use of alternative channel materials in ballistic nano-MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 1, pp. 115–123, 2007.

[62] F. Conzatti, P. Toniutti, D. Esseni, P. Palestri, and L. Selmi, "Simulation study of the on-current improvements in Ge and sGe versus Si and sSi nano MOSFETs," in *IEEE IEDM Technical Digest*, 2010, pp. 15.2.1–15.2.4.

[63] F. Stern and W. E. Howard, "Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit ," *Physical Review*, vol. 163, no. 3, pp. 816–835, 1967.

[64] C. Jacoboni, F. Nava, C. Canali, and G. Ottaviani, "Electron drift velocity and diffusivity in Germanium," *Phys. Rev. B*, vol. 24, p. 1014, 1981.

[65] Q.Rafhay, R.Clerc, G.Ghibaudo, and G.Pananakakis, "Impact of source to drain tunneling on the Ion/Ioff trade-off of alternative channel material MOSFETs," in *International Semiconductor Device Research Symposium*, 2007, pp. 1–2.

[66] 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.

[67] D.Esseni, A.Abramo, L.Selmi, and E.Sangiorgi, "Physically Based Modeling of Low Field Electron Mobility in Ultra-Thin Single and Double-Gate SOI n-MOSFETs." *IEEE Trans. Electron Devices*, vol. 50, no. 12, pp. 2445–2455, 2003.

[68] D. Roessler and W. A. Jr., "Infrared reflectance of single crystal tetragonal GeO2," *Journal of Physics and Chemistry of Solids*, vol. 33, no. 2, pp. 293 – 296, 1972.

[69] A.T.Pham and C. Jungemann and B. Meinerzhagen, "Physics-Based Modeling of Hole Inversion-Layer Mobility in Strained-SiGe-on-Insulator," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2174–2182, 2007.

[70] D.Ponton, L.Lucci, P.Palestri, D.Esseni, and L.Selmi, "Assessment of the Impact of Biaxial Strain on the Drain Current of Decanometric n-MOSFET," in *Proc. European Solid State Device Res. Conf.*, 2006, pp. 166–169.

[71] F. Conzatti, M. De Michielis, D. Esseni, and P. Palestri, "Drain Current Improvements in Uniaxially Strained p-MOSFETs: A Multi-Subband Monte Carlo Study," *Solid State Electronics*, vol. 53, pp. 706–711, 2009.

[72] M. Lundstrom, "Elementary Scattering Theory of the Si MOSFET ," *IEEE Electron Device Lett.*, vol. 18, no. 7, pp. 361–363, 1997.

Chapter 6

Quantum transport approach

6.1 Introduction

The continued scaling of semiconductor device dimensions accelerates the development of new simulation models and software tools which can correctly model quantum effects in extremely small devices. Chapter 4 has presented the semiclassical approach in the framework of the **k·p** theory, where the quantum mechanical effects have been introduced only in the calculation of the subbands and the scattering rates, so it does not represent a full-quantum approach. As already said, the semi-classical approach poses a lower limit to the device dimensions we can simulate.

The full-quantum simulation approaches, on the other side, can overcome this limit, but are usually far more demanding from the computational standpoint with respect to the semiclassical ones. The full-quantum models can naturally include quantum effects along the transport direction like band-to-band-tunneling (BTBT) and interference effects, which are problematic to include in a semi-classical model. In the past few years, the increased performances of modern computers allowed to implement full-quantum approaches and to simulate realistic device structures. One possible way to tackle the full quantum transport problem in small devices is to directly solve the Schrödinger equation by using the finite difference or finite elements techniques and imposing open boundary conditions [1]. Another possibility to include interference and tunneling effects along the transport direction would be to use the Wigner function theory [2].

In this chapter, we rather focus our attention on the Green's function (GF) formalism [3, 4], where we solve the equations for the single-particle GF instead of considering the Schrödinger equation discretized on a real-space mesh. The main advantage of the GF formalism is not related with the computational cost, but instead the fact that such formalism is very general and can be used to include also the electron-phonon interaction (as we will see in Sec.6.4). In this respect, the GF method can be used with any kind of Hamiltonian, from atomistic to effective mass ones, and many techniques have been developed in order to reduce the CPU time.

In the course of the PhD program, we developed a 3-D Poisson-NEGF solver based on a 8-band **k·p** Hamiltonian (see Section 2.5) to simulate transport in direct-bandgap semiconductor materials. In particular, the next chapter will present the results of a simulation study which employs this model in order to simulate InAs tunneling-based nanowire transistors (Tunnel-FETs) and to systematically benchmark them against strained-silicon nanowire MOSFETs for low-power and ultra-low-power applications.

This chapter is devoted to the introduction of the formalism we used in order to simulate realistic devices in the full quantum approach. In particular, Sec.6.2 briefly reviews the Non-Equilibrium Green's function (NEGF) approach together with the recursive algorithms we implemented in order to calculate the GF for a coherent simulation. Subsequently, Sec.6.3 introduces the coupled mode-space approach, which is a widely used technique to reduce the computational burden. The electron-phonon interactions within the NEGF model are treated in Sec.6.4, and Sec.6.5, finally, presents the validation of our model, obtained both by inspecting internal quantities and by comparing the results of our model with the simulations found in literature for the same structures.

6.2 The Non Equilibrium Green's Function formalism

6.2.1 Basic concepts

Let us write a time-independent, closed boundary Schrödinger equation in the operator notation:

$$[E - \hat{H}]\,\psi = 0 \quad , \tag{6.1}$$

where $\hat{H}$ is a general single-particle Hamiltonian and E can take only the values E_n corresponding to the eigen-energies of the system. The retarded GF operator $\hat{G}$ associated to Eq.6.1 is

defined as

$$[E - \hat{H}]\,\hat{G}(E) = 1 \tag{6.2}$$

and $\hat{G}(E)$ depends on the energy E, that is now a parameter in Eq.6.2 and can take any possible value. The general expression for $\hat{G}(E)$ is given by

$$\hat{G}(E) = (E - \hat{H})^{-1} \tag{6.3}$$

which is defined for any energy E, except for the singularities corresponding to the eigenenergies E_n of $\hat{H}$. In order to avoid such singularities, we can re-write Eq.6.3 as

$$\hat{G}(E) = (E + \mathrm{i}0^+ - \hat{H})^{-1} \tag{6.4}$$

where we have introduced the small imaginary quantity $\mathrm{i}0^+$. As already said, the retarded GF $\hat{G}$ in Eq.6.4 depends on the energy E and this represents a big conceptual step with respect to the standard Schrödinger equation, where the eigenenergies are the solutions of an eigenvalue problem (see, for instance, Eq.2.1). In fact, the energy E entering Eq.6.4 is an independent variable and represents the energy of excitation from an external source. Hence, in order to evaluate the quantities we are interested in, we have to calculate the GF for a dense energy grid, which typically results in a large computational burden.

Physically speaking, the inverse Fourier transform of the retarded GF represents the causal impulse response of the Schrödinger equation [3], and expresses how the system reacts to an impulse given at a certain time instant. In a similar way we can define the advanced GF operator G^a, which is a valid mathematical solution of Eq.6.3, but does not have a direct physical interpretation, because it is related to the anti-causal response of the system:

$$\hat{G}^a(E) = (E - \mathrm{i}0^+ - \hat{H})^{-1} \tag{6.5}$$

It should be noted that Eqs.6.4 and 6.5 hold for any Hamiltonian (from effective mass to atomistic ones), thus the GF formalism is indeed very general. In what follows we will limit our considerations to the matrix formulation of the problem, that is to the formulation obtained by using a discretization scheme based on a grid in real space of the operator $\hat{H}$, which turns $\hat{H}$ into the matrix $\mathbf{H}$. An example of this discretization procedure is described in Section 2.2.1.

To be more explicit in this respect, we denote with $\mathbf{r}_d$=($\mathbf{r}_1$,$\mathbf{r}_2$,...,$\mathbf{r}_N$) the vector of N discretization points in the 3-D real-space mesh (see Fig.6.1). If we adopt the EMA, $\hat{H}$ is a scalar operator, hence the corresponding $\mathbf{H}$ is a $N \times N$ matrix.

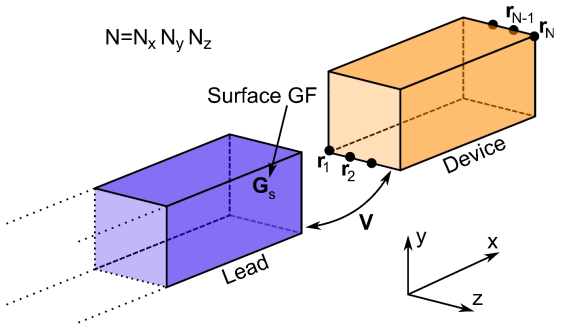


Figure 6.1: Sketch of the simulated device attached to a semi-infinite lead. The circles highlight some grid points labeled as $\mathbf{r}_i$. The total number of points N corresponds to N=$N_x N_y N_z$. We also reported the surface GF $\mathbf{G}_\mathrm{s}$ of the lead and the hopping matrix $\mathbf{V}$ between the lead and the device.

Besides, also the retarded (advanced) GFs $\mathbf{G}$ ($\mathbf{G}^a$) becomes a matrix with the same rank as $\mathbf{H}$, i.e. $N \times N$. Eq.6.4 can be thus re-written as:

$$\mathbf{G}(E) = [(E + \mathrm{i}0^+)\mathbf{I} - \mathbf{H}]^{-1} \tag{6.6}$$

where $\mathbf{I}$ is an $N \times N$ identity matrix. Moreover, it can be demonstrated that the advanced GF is the complex conjugate of the retarded GF, that is

$$\mathbf{G}^a = \mathbf{G}^\dagger \tag{6.7}$$

Eqs.6.6 and 6.7 are valid only if we consider an isolated device described by the Hamiltonian $\mathbf{H}$ and with closed boundary conditions. One of the most important parts of the GF theory concerns the modeling of the open boundary conditions: this condition is described by assuming semi-infinite leads attached to the device and is modeled as a self-energy matrix $\mathbf{\Sigma}_{\text{Lead}}$ [3] which adds to the Hamiltonian matrix $\mathbf{H}$. Thus, Eq.6.6 is re-written as

$$\mathbf{G}(E) = [(E + \mathrm{i}0^+)\mathbf{I} - \mathbf{H} - \mathbf{\Sigma}_{\text{Lead}}]^{-1} \tag{6.8}$$

where $\mathbf{\Sigma}_{\text{Lead}}$ is a $N \times N$ matrix which contains the interaction with the contacts. If we consider for instance, a channel connected with two leads labeled as S and D, the matrix $\mathbf{\Sigma}_{\text{Lead}}$ can be written as

$$\mathbf{\Sigma}_{\text{Lead}} = \begin{bmatrix} \mathbf{\Sigma}_{\text{S}} & 0 & \cdots & \\ 0 & 0 & \cdots & \\ \cdots & \cdots & \cdots & \\ & \cdots & 0 & 0 \\ & \cdots & 0 & \mathbf{\Sigma}_{\text{D}} \end{bmatrix} \tag{6.9}$$

It can be demonstrated that the self-energy of a single contact can be calculated starting from the surface GF $\mathbf{G_s}$ [3]:

$$\mathbf{\Sigma}_{\text{C}} = \mathbf{V}\,\mathbf{G_s}\,\mathbf{V}^\dagger \tag{6.10}$$

where $\mathbf{V}$ is the coupling Hamiltonian between the channel and the contact, as depicted in Fig.6.1. In the case of a MOS device, the subscript C refers to the source S or drain D contacts. In Section 6.2.6 we will present the recursive algorithm required to compute $\mathbf{G_s}$.

In a coherent simulation, the knowledge of the retarded GF allows us to calculate the most important physical quantities, namely the charge and the current. The density matrix $\boldsymbol{\rho}$ reads [3]:

$$\boldsymbol{\rho}(\mathbf{r}_d, \mathbf{r}'_d) = \frac{1}{\Omega}\int (\mathrm{d}E/2\pi)[\mathbf{G}\mathbf{\Gamma}_{\text{S}}\mathbf{G}^\dagger f_0(E - E_{f,S}) + \mathbf{G}\mathbf{\Gamma}_{\text{D}}\mathbf{G}^\dagger f_0(E - E_{f,D})] \tag{6.11}$$

where Ω is the volume of an element in the 3-D grid, f_0 is the Fermi-Dirac distribution function, $E_{f,S}$ and $E_{f,D}$ are the Fermi-energies at the source and drain contacts, respectively. The term $\mathbf{\Gamma}_C$ is the broadening matrix and is equal to the anti-hermitian part of $\mathbf{\Sigma}_{\text{C}}$:

$$\mathbf{\Gamma}_C = \mathrm{i}[\mathbf{\Sigma}_{\text{C}} - \mathbf{\Sigma}_{\text{C}}^\dagger], \qquad \text{C} = \text{S}, \text{D} \tag{6.12}$$

The diagonal elements of $\boldsymbol{\rho}$ represent the electron concentration

$$n(\mathbf{r}_d) = \boldsymbol{\rho}(\mathbf{r}_d, \mathbf{r}_d) \tag{6.13}$$

By using the Landauer theory, the current can be calculated as [3]

$$I = \frac{e}{h}\int \mathrm{d}E\, T(E)\, [f_0(E - E_{f,S}) - f_0(E - E_{f,D})] \tag{6.14}$$

It is clear from Eq.6.14 that the current is non-zero only if the the source and drain Fermi levels are different, that is if a non-zero voltage difference exists between the drain and source contacts. The real and scalar quantity $T(E)$ is the transmission coefficient, which is always less than one and can be expressed as a function of the retarded GF and the broadening matrices at the two contacts as [3]

$$T = \text{Trace}\{\mathbf{\Gamma}_{\text{S}}\mathbf{G}\mathbf{\Gamma}_{\text{D}}\mathbf{G}^\dagger\} \tag{6.15}$$

In practice, the matrix inversion of Eq.6.8 can be performed directly only if the system is small, i.e. the rank of $\mathbf{H}$ is also small. The aim of our work, however, is to simulate a full 3-D nano-scale transistor with realistic dimensions, hence we deal with very large Hamiltonians and the term in curly brackets in the right-hand side of Eq.6.8 cannot be directly inverted. We will see in Section 6.2.2 that, in order to evaluate the current density and the carrier concentration, we do not need the full retarded GF matrix, but only small parts of it. Hence, if the Hamiltonian is written as a tri-diagonal block matrix, it can be demonstrated that the full matrix inversion of Eq.6.8 can be converted into a block-inversion, thus reducing dramatically the computational cost. Section 6.2.2 introduces the Dyson equation and derives the iterative procedure to calculate the retarded GF in a ballistic case.

6.2.2 The Dyson equation approach

This section derives the Dyson equation and illustrates simple 1-D and 2-D examples based on EMA Hamiltonians. Indeed, we developed a NEGF solver based on the 8-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian introduced in Section 2.5, but, once we write the Hamiltonian as a tri-diagonal block matrix, all the transformations and the recursive algorithms are very general and remain the same also for different Hamiltonians.

Consider an unperturbed system whose Hamiltonian matrix is $\mathbf{H}_0$. The corresponding (unperturbed) retarded GF $\mathbf{g}$ is defined as

$$\mathbf{g} = (E\mathbf{I} - \mathbf{H}_0)^{-1} \tag{6.16}$$

In Eq.6.16 we have implicitly assumed that the small imaginary term $\mathrm{i}0^+$ introduced in Eq.6.4 has been included in the energy term.

Assume that we perturb the system with a small Hamiltonian $\mathbf{V}$. We can write the new retarded GF $\mathbf{G}$ as

$$\mathbf{G} = (E\mathbf{I} - \mathbf{H})^{-1} \tag{6.17}$$

where H is the perturbed Hamiltonian defined as

$$\mathbf{H} = \mathbf{H}_0 + \mathbf{V} \quad . \tag{6.18}$$

Using Eqs.6.16 and 6.18 we can rewrite Eq.6.17 as

$$\mathbf{G} = (E\mathbf{I} - \mathbf{H}_0 - \mathbf{V})^{-1} = \left(\mathbf{g}^{-1} - \mathbf{V}\right)^{-1} \tag{6.19}$$

which, with straight-forward matrix algebra, gives the *Dyson equation* [4, 5]

$$\mathbf{G} = \mathbf{g} + \mathbf{g}\,\mathbf{V}\,\mathbf{G} = \mathbf{g} + \mathbf{G}\,\mathbf{V}\,\mathbf{g} \tag{6.20}$$

Eq.6.20 can be written explicitly as

$$G(m,n) = g(m,n) + \sum_{p,q} g(m,p)\,V(p,q)\,G(q,n) = g(m,n) + \sum_{p,q} G(m,p)\,V(p,q)\,g(q,n) \tag{6.21}$$

which relates the perturbed GF elements $G(m,n)$ to the unperturbed ones $g(m,n)$ via the elements $V(p,q)$ of the perturbation Hamiltonian. The indexes m, n, p and q run over the N discretization points.

6.2.3 A simple 1-D example

Consider now a semi-infinite chain of sites as in Fig.6.2(a) characterized by an unperturbed GF $\mathbf{g}$ and by a single energy level ε_0 for each site. The binding energy (hopping) between two adjacent sites in the EMA is equal to $t = \hbar/2m^*a^2$. Assume to perturb the system by adding a

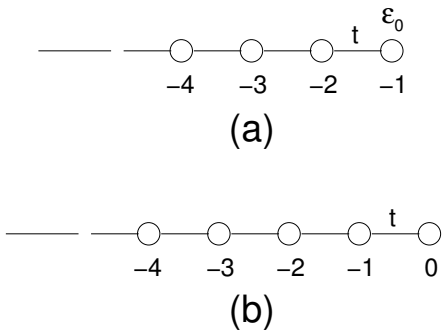


Figure 6.2: (a) Unperturbed semi-infinite chain of sites, each characterized by a single energy level ε_0. The binding energy (hopping) between two adjacent sites in the EMA is equal to $t = \hbar/2m_0a^2$. (b) Adding a single site perturbs the system.

further site (Fig.6.2(b)). This simplified example mimics the coupling of a semi-infinite contact with a real device [3]. Using Eq.6.21 one can write the following system of equations

$$\begin{cases} G_{0,0} &= g_{0,0} + g_{0,0}\, V_{0,-1}\, G_{-1,0} \\ G_{-1,0} &= g_{-1,-1}\, V_{-1,0}\, G_{0,0} \end{cases} \tag{6.22}$$

which relates the unperturbed ($g_{0,0}$ and $g_{-1,-1}$) to the perturbed ($G_{0,0}$ and $G_{-1,0}$) GF components. All the other terms in the summations of Eq.6.22 vanish (e.g. $g_{-1,0}$ is zero because in the unperturbed case we do not have the site labeled as 0). Now we can conjecture that the final system is identical to the initial one, because in both cases we are dealing with a lead taking the form of a semi-infinite set of sites. Hence, we may identify the diagonal GF elements of the last sites in both systems and write

$$g_{-1,-1} = G_{0,0} \tag{6.23}$$

Considering a constant hopping term t between sites and writing the unperturbed GF as $g_{0,0} = (E-\varepsilon_0)^{-1}$ we can combine Eqs.6.22 and 6.23 in order to obtain the following closed relationship for the unknown element $G_{0,0}$ of the perturbed Hamiltonian

$$t^2\, G_{0,0}^2 - (E - \varepsilon_0)\, G_{0,0} + 1 = 0 \tag{6.24}$$

The term $G_{0,0}$ is the surface GF introduced in Eq.6.10. Solving Eq.6.24 and taking only the negative solution one can obtain

$$G_{0,0} = \frac{1}{t}\left(X - i\sqrt{1 - X^2}\right) \tag{6.25}$$

where

$$X = \frac{E - \varepsilon_0}{2t} \tag{6.26}$$

6.2.4 A 2-D example

General Formulation

Once we have calculated the GF for the simplified one dimensional case, it is quite easy to extend the results to a two dimensional case (see Fig.6.3) if we make use of the EMA approach. In this case we have a vector of eigenenergies for each device section (equal to the number of

points in the vertical direction if we work with EMA), hence, consistently with Eq.6.25, the contact GF $\mathbf{G}_\mathrm{s}$ in the eigenvector representation is a diagonal matrix whose elements are

$$\mathbf{G}_\mathrm{s}(n,n) = \frac{1}{t_x}\left(X_n - i\sqrt{1-X_n^2}\right) \tag{6.27}$$

where n is the subband index and

$$X_n = \frac{E + 2t_x - \varepsilon_n}{2t_x} \tag{6.28}$$

and t_x is defined in Eq.6.38. The matrix $\mathbf{G}_\mathrm{s}$ has dimension $N_y \times N_y$, N_y being the number of discretization points in the vertical direction. In this case ε_n is the n-th eigenvalue from a 1D Schrödinger solution.

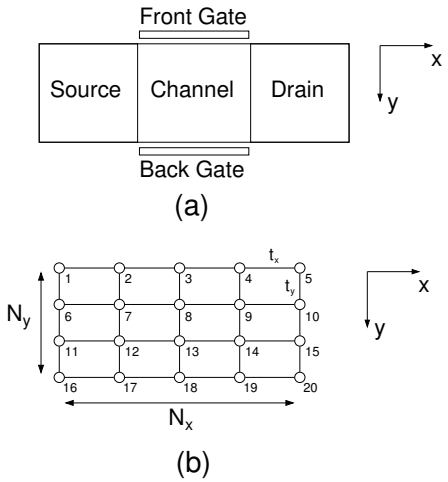


Figure 6.3: (a) 2D sketch of a MOS device with the coordinate system. (b) Corresponding discretization grid with the natural site ordering.

The real space matrix representation $\mathbf{G}_\mathrm{RS}$ of the contact GF $\mathbf{G}_\mathrm{s}$ can be calculated with the unitary transformation

$$\mathbf{G}_\mathrm{RS} = \mathbf{U}\,\mathbf{G}_\mathrm{s}\,\mathbf{U}^\dagger \tag{6.29}$$

where $\mathbf{U}$ is the $(N_y \times N_y)$ eigenvector matrix , thus the GF has dimensions equal to $(N_y \times N_y)$. It can be demonstrated that the contact self-energy $\mathbf{\Sigma}_\mathrm{C}$ is given by [3]

$$\mathbf{\Sigma}_\mathrm{C} = \mathbf{V}_\mathrm{DC}^\dagger\,\mathbf{G}_\mathrm{RS}\,\mathbf{V}_\mathrm{DC} \tag{6.30}$$

where $\mathbf{V}_\mathrm{DC}$ is the hopping matrix between the contact and the device, which can be written as

$$\mathbf{V}_\mathrm{DC} = -t_x\mathbf{I} \tag{6.31}$$

where t_x is defined in Eq.6.38. This result is very general, as it does not depend on the Hamiltonian nor on the discretization scheme.

Once we have calculated the self-energies for each contact we can calculate the complete retarded GF $\mathbf{G}$ starting from Eq.6.8 as

$$\mathbf{G} = (E\mathbf{I} - \mathbf{H}_{2D} - \mathbf{\Sigma}_\mathrm{Lead})^{-1} \tag{6.32}$$

where $\mathbf{H}_{2D}$ is the 2D Hamiltonian of the system and can be written as

$$\mathbf{H}_{2D} = \begin{bmatrix} \mathbf{H}_{1,1} & -t_x\mathbf{I} & 0 & \cdots & & \\ -t_x\mathbf{I} & \mathbf{H}_{2,2} & -t_x\mathbf{I} & 0 & \cdots & \\ 0 & -t_x\mathbf{I} & \mathbf{H}_{3,3} & -t_x\mathbf{I} & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & \cdots & 0 & -t_x\mathbf{I} & \mathbf{H}_{N_x-1,N_x-1} & -t_x\mathbf{I} \\ & & \cdots & 0 & -t_x\mathbf{I} & \mathbf{H}_{N_x,N_x} \end{bmatrix} \tag{6.33}$$

The elements $\mathbf{H}_{i,i}$ constitute a $N_y \times N_y$ tri-diagonal matrix, which can be written as

$$\mathbf{H}_{i,i} = \begin{bmatrix} h_{1,1} & -t_y & 0 & \cdots & \\ -t_y & h_{2,2} & -t_y & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & -t_y & h_{N_y-1,N_y-1} & -t_y \\ & \cdots & 0 & -t_y & h_{N_y,N_y} \end{bmatrix} \tag{6.34}$$

and $h_{i,i} = 2t_x + 2t_y - qV_i(x)$. The total self-energy $\mathbf{\Sigma}_{\text{Lead}}$ entering Eq.6.32 is a $N_xN_y \times N_xN_y$ matrix which contains the self-energies of the source and drain contacts and has only two $N_y \times N_y$ non-zero sub-matrices. The $\mathbf{\Sigma}_{\text{Lead}}$ reads:

$$\mathbf{\Sigma}_{\text{Lead}} = \begin{bmatrix} \mathbf{\Sigma}_{\text{S}} & 0 & \cdots & \\ 0 & 0 & \cdots & \\ \cdots & \cdots & \cdots & \\ & \cdots & 0 & 0 \\ & \cdots & 0 & \mathbf{\Sigma}_{\text{D}} \end{bmatrix} \tag{6.35}$$

If we adopt the EMA and the finite differences discretization scheme, then the inverse of the GF is a block tri-diagonal matrix consisting of $N_x \times N_x$ matrices each with dimensions $N_y \times N_y$ and reads

$$\mathbf{G} = \begin{bmatrix} \mathbf{D}_{1,1} - \mathbf{\Sigma}_{\text{S}} & t_x\mathbf{I} & 0 & \cdots & & \\ t_x\mathbf{I} & \mathbf{D}_{2,2} & t_x\mathbf{I} & 0 & \cdots & \\ 0 & t_x\mathbf{I} & \mathbf{D}_{3,3} & t_x\mathbf{I} & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & \cdots & 0 & t_x\mathbf{I} & \mathbf{D}_{N_x-1,N_x-1} & t_x\mathbf{I} \\ & & \cdots & 0 & t_x\mathbf{I} & \mathbf{D}_{N_x,N_x} - \mathbf{\Sigma}_{\text{D}} \end{bmatrix}^{-1} \tag{6.36}$$

where $\mathbf{D}_{i,i}$ is a $N_y \times N_y$ matrix and has a tridiagonal form e.g.

$$\mathbf{D}_{i,i} = \begin{bmatrix} d_{1,1} & t_y & 0 & \cdots & \\ t_y & d_{2,2} & t_y & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & t_y & d_{N_y-1,N_y-1} & t_y \\ & \cdots & 0 & t_y & d_{N_y,N_y} \end{bmatrix} \tag{6.37}$$

and $d_{i,i} = E - 2t_x - 2t_y + qV_i(x)$. In this latter expression $V_i(x)$ is the electrostatic potential for the i-th slice. The hopping terms in Eqs.6.36 and 6.37 represent the coupling between adjacent sites in x and y directions. These terms can be written as

$$t_x = \frac{\hbar^2}{2m_x^*\Delta_x^2} \quad \text{and} \quad t_y = \frac{\hbar^2}{2m_y^*\Delta_y^2} \tag{6.38}$$

Eq.6.36 can be formally re-written as

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{G}_{1,2} & \cdots & \mathbf{G}_{1,N_x} \\ \mathbf{G}_{2,1} & \mathbf{G}_{2,2} & \cdots & \mathbf{G}_{2,N_x} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{N_x,1} & \mathbf{G}_{N_x,2} & \cdots & \mathbf{G}_{N_x,N_x} \end{bmatrix} \tag{6.39}$$

where each block $\mathbf{G}_{i,i}$ has dimension $N_y \times N_y$. From Eqs.6.36 and 6.39 we see that the inverse of $\mathbf{G}^r$ is block tri-diagonal, but $\mathbf{G}$ itself is dense.

The local density of states $LDOS$ at a generic discretization point $\mathbf{r}_{i,j}$=(x_i,y_j) in real space is given by

$$LDOS(\mathbf{r}_{i,j}, E) = -\frac{1}{2\pi}\mathrm{Im}[G(\mathbf{r}_{i,j}, \mathbf{r}_{i,j}, E) - (G(\mathbf{r}_{i,j}, \mathbf{r}_{i,j}, E))^\dagger]_{j,j} \tag{6.40}$$

The total density of states $DOS(E)$ is given by the sum of the diagonal elements

$$DOS(E) = -\frac{1}{2\pi}\mathrm{Im}\{\mathrm{Trace}[\mathbf{G}(E) - (\mathbf{G}(E))^\dagger]\} \tag{6.41}$$

By starting from Eqs.6.11, 6.13 for the expression of $n(x_i, y_j)$ and then Eqs.6.12, 6.35 for the broadening matrix, it can be demonstrated that the electronic concentration for each grid point (x_i, y_j) is given by

$$n(x_i, y_j) = g\, A_{k_z} \int \frac{dE}{2\pi}\Big[\mathbf{G}_{i,1}\, \mathbf{\Gamma}_\mathrm{S}\, (\mathbf{G}^\dagger)_{1,i}\, F_{-1/2}((E_{f,S} - E)/KT) + \mathbf{G}_{i,N_x}\, \mathbf{\Gamma}_\mathrm{D}\, (\mathbf{G}^\dagger)_{N_x,i}\, F_{-1/2}((E_{f,D} - E)/KT)\Big]_{j,j} \tag{6.42}$$

where g is the degeneration factor, the broadening of the density of states reads

$$\mathbf{\Gamma}_{S,D} = \mathrm{i}[\mathbf{\Gamma}_{S,D} - \mathbf{\Gamma}^\dagger_{S,D}] \tag{6.43}$$

and the constant term A_{k_z} is equal to

$$A_{k_z} = \sqrt{\frac{2m_z KT}{\pi\hbar^2}} \tag{6.44}$$

The symbol $F_{-1/2}(x)$ indicates the Fermi integral of order $-1/2$ and $\mu_{S,D}$ is the Fermi level at the S/D contact. From Eq.6.42 it is clear that if we want to determine the electronic concentration we need only the first and last columns of the full GF of Eq.6.39 i.e.

$$\mathbf{G} = \begin{bmatrix} \boxed{\begin{matrix}\mathbf{G}_{1,1}\\ \mathbf{G}_{2,1}\\ \vdots\\ \mathbf{G}_{N_x,1}\end{matrix}} & \begin{matrix}\mathbf{G}_{1,2}\\ \mathbf{G}_{2,2}\\ \vdots\\ \mathbf{G}_{N_x,2}\end{matrix} & \begin{matrix}\cdots\\ \cdots\\ \ddots\\ \cdots\end{matrix} & \boxed{\begin{matrix}\mathbf{G}_{1,N_x}\\ \mathbf{G}_{2,N_x}\\ \vdots\\ \mathbf{G}_{N_x,N_x}\end{matrix}} \end{bmatrix} \tag{6.45}$$

In the case of coherent transport, starting from Eqs.6.14 and 6.15 it can be demonstrated that the current reads [3]

$$I = g\, A_{k_z} \frac{e}{h} \int dE\, \mathrm{Trace}\left\{\mathbf{\Gamma}_\mathrm{S}\, \mathbf{G}_{1,N_x}\, \mathbf{\Gamma}_\mathrm{D}\, (\mathbf{G}^\dagger)_{N_x,1}\right\} \times \left[F_{-1/2}((E_{f,S} - E)/KT) - F_{-1/2}((E_{f,D} - E)/KT)\right] \tag{6.46}$$

Hence for the current we only need the small matrix $\mathbf{G}_{1,N_x}$, which is the upper-right block of the full GF:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{G}_{1,2} & \cdots & \boxed{\mathbf{G}_{1,N_x}} \\ \mathbf{G}_{2,1} & \mathbf{G}_{2,2} & \cdots & \mathbf{G}_{2,N_x} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{N_x,1} & \mathbf{G}_{N_x,2} & \cdots & \mathbf{G}_{N_x,N_x} \end{bmatrix} \tag{6.47}$$

Similar expression can be derived for a 3-D EMA case, except for the fact that the matrices involved are much larger.

6.2.5 Recursive Method

To invert the r.h.s of Eq.6.36 and determine $\mathbf{G}$ can be a daunting task because this matrix has $(N_x \times N_y)^2$ elements and, moreover, it has to be done for a large number of energies E defined on a dense energy grid. There is an alternative way to determine exactly $\mathbf{G}$ or sub-matrices of $\mathbf{G}$ using the *Dyson equation*. It is important to remember that, in order to calculate carrier concentration and current, we need only a small part of the GF (see Eqs.6.45 and 6.47), and so it is not necessary to know all the elements and to invert the entire matrix in the r.h.s of Eq.6.32.

Here we should point out that the presented recursive algorithms can be used only if the matrix $[\mathbf{H} + \mathbf{\Sigma}]$ is a tri-diagonal block matrix. If this is not the case, the recursive algorithm cannot be used and Eq.6.36 has to be inverted directly. Fig.6.4 sketches the device sections

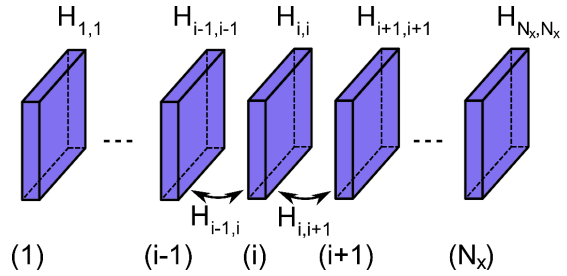


Figure 6.4: Sketch of the device sections along the transport direction x. The Hamiltonian of a general section is $\mathbf{H}_{i,i}$, while $\mathbf{H}_{i,j}$, $j{=}i \pm 1$ is the hopping term.

along the transport direction x. The perturbation of the system consists in adding one slice at a time on one side of the device, starting from one contact. In particular, using Eq.6.19 we can write the relation between the unperturbed GF of the first slice $g_{1,1}$ and the perturbed $g^{\text{Left}}_{1,1}$, which is not yet the final GF, because it is perturbed only on the left side, since we add the first slice to the source contact (see Fig.6.4). In particular we can write:

$$\mathbf{g}^{\text{Left}}_{1,1} = (\mathbf{g}^{-1}_{1,1} - \mathbf{\Sigma}_{\text{S}})^{-1} \tag{6.48}$$

where the superscript *Left* means *coming from the left side*, because we're entering the device and we're going to calculate the equilibrium GF $\mathbf{g}$ for each slice.

For a generic slice (i.e. $i = 2, \ldots, N_x - 1$) Eq.6.48 becomes

$$\mathbf{g}^{\text{Left}}_{i,i} = (\mathbf{g}^{-1}_{i,i} - \mathbf{V}_{i,i-1}\, \mathbf{g}^{\text{Left}}_{i-1,i-1}\, \mathbf{V}^{\dagger}_{i,i-1})^{-1} \tag{6.49}$$

In this case we have used the result of self-energies calculation (see Eq.6.30) which is completely general. This means that it can be used both when connecting the contact and the device and inside the device when we go from one section to the adjacent one (see Fig.6.4). Once we arrive at the right-end side of the device we know the self-energy of the drain $\mathbf{\Sigma}_{\text{D}}$, hence we can calculate the *real* perturbed GF $\mathbf{G}$. Hence, the last equation of the chain will be

$$\mathbf{G}_{N_x,N_x} = (\mathbf{g}^{-1}_{N_x,N_x} - \mathbf{V}_{N_x,N_x-1}\, \mathbf{g}^{\text{Left}}_{N_x-1,N_x-1}\, \mathbf{V}^{\dagger}_{N_x,N_x-1} - \mathbf{\Sigma}_{\text{D}})^{-1} \tag{6.50}$$

Eq.6.50 gives the final out-of-equilibrium GF for the last slice. In Eq.6.50 we can identify the contributions of the fluxes coming from the left and from the right sides. Once $\mathbf{G}_{N_x,N_x}$ is known, one can proceed in the calculation of the parts of the total retarded GF needed to calculate current and charge concentration (see Eqs.6.45 and 6.47). In particular, using the *Dyson Equation* we can calculate the last column of block matrices G_{i,N_x} using Eq.6.21 to write

$$\mathbf{G}_{i-1,N_x} = \mathbf{g}^{\text{Left}}_{i-1,i-1}\, \mathbf{V}_{i-1,i}\, \mathbf{G}_{i,N_x} \tag{6.51}$$

which allows one to calculate $\mathbf{G}_{i,N_x}$ for any $i \leq (N_x - 1)$. In Eq.6.51 we have used the fact that $\mathbf{g}^{\text{Left}}_{i-1,i}$ is null in the unperturbed problem because the device sections are not connected.

The procedure from Eq.6.48 to 6.50 can be repeated starting from the right side to calculate the $\mathbf{G}_{i,1}$ terms [6].

Finally, starting from the $\mathbf{G}_{N_x,N_x}$ calculated in Eq.6.50, one can calculate the blocks on the diagonal by using:

$$\mathbf{G}_{i-1,i-1} = \mathbf{g}^{\text{Left}}_{i-1,i-1} + \mathbf{g}^{\text{Left}}_{i-1,i-1}\,\mathbf{V}_{i-1,i}\,\mathbf{G}_{i,i}\,\mathbf{V}^{\dagger}_{i-1,i}\,\mathbf{g}^{\text{Left}}_{i-1,i-1} \tag{6.52}$$

We stress once again that the recursive algorithm to calculate the retarded GF is very general and has the same expressions for any form of Hamiltonian. In particular, we employed such algorithm to calculate the 3-D GF for the 8-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian introduced in Section 2.5.

A direct application of this procedure with the Hamiltonians discretized directly in real space is still too demanding from the computational point of view, thus we employed the mode-space transformation (see Section 6.3), in order to reduce the dimension of the matrices involved in the calculation.

6.2.6 Sancho-Rubio iteration algorithm

For a generic Hamiltonian $\mathbf{H}$ the problem of writing the surface Green's function and thus the self energy of the reservoir is not trivial and, in general, does not provide a closed form expression as in Eq.6.27. The iterative algorithm to determine the surface GF for semi-infinite leads first proposed in [7] starts by writing the GF for the isolated slice i and for a given energy E as

$$\mathbf{G}_{i,i} = (E\mathbf{I} - \mathbf{H}_{i,i})^{-1} \tag{6.53}$$

If we assume of a homogeneous reservoir, i.e. $\mathbf{H}_{0,0} = \mathbf{H}_{1,1} = \mathbf{H}_{2,2} = \ldots$ and $\mathbf{H}_{0,1} = \mathbf{H}_{1,2} = \mathbf{H}_{2,3} = \ldots$, we can re-write Eq.6.21 as

$$\begin{aligned}
&\mathbf{G}_{0,0} = \mathbf{g}_{0,0} + \mathbf{g}_{0,0}\,\mathbf{H}_{0,1}\,\mathbf{G}_{1,0} \\
&\mathbf{G}_{1,0} = \mathbf{g}_{0,0}\,\mathbf{H}^{\dagger}_{0,1}\,\mathbf{G}_{0,0} + \mathbf{g}_{0,0}\,\mathbf{H}_{0,1}\,\mathbf{G}_{2,0} \\
&\ldots \\
&\mathbf{G}_{n,0} = \mathbf{g}_{0,0}\,\mathbf{H}^{\dagger}_{0,1}\,\mathbf{G}_{n-1,0} + \mathbf{g}_{0,0}\,\mathbf{H}_{0,1}\,\mathbf{G}_{n+1,0}
\end{aligned} \tag{6.54}$$

where positive indexes denote sites inside the contact. If we now multiply on the left all the expressions in Eq.6.54 by $\mathbf{g}^{-1}_{0,0}$, we can re-write the chain of matrix equations inside the semi-infinite contact as:

$$\begin{aligned}
&(E\mathbf{I} - \mathbf{H}_{0,0})\,\mathbf{G}_{0,0} = \mathbf{I} + \mathbf{H}_{0,1}\,\mathbf{G}_{1,0} \\
&(E\mathbf{I} - \mathbf{H}_{0,0})\,\mathbf{G}_{1,0} = \mathbf{H}^{\dagger}_{0,1}\,\mathbf{G}_{0,0} + \mathbf{H}_{01}\,\mathbf{G}_{2,0} \\
&\ldots \\
&(E\mathbf{I} - \mathbf{H}_{0,0})\,\mathbf{G}_{n,0} = \mathbf{H}^{\dagger}_{0,1}\,\mathbf{G}_{n-1,0} + \mathbf{H}_{01}\,\mathbf{G}_{n+1,0}
\end{aligned} \tag{6.55}$$

where I is the identity matrix, $\mathbf{H}_{00}$ and $\mathbf{H}_{01}$ are the diagonal and off-diagonal block matrices of the total Hamiltonian $\mathbf{H}$, respectively. It is worth noting that such formulation is independent on the representation of the Hamiltonian as well as on the dimensionality of our problem.

The general term for $n \geq 1$ of Eq.6.55 can be written as:

$$\mathbf{G}_{n,0} = \mathbf{t}_0\,\mathbf{G}_{n-1,0} + \tilde{\mathbf{t}}_0\,\mathbf{G}_{n+1,0} \tag{6.56}$$

where

$$\begin{aligned}
\mathbf{t}_0 &= (E\mathbf{I} - \mathbf{H}_{0,0})^{-1}\,\mathbf{H}^{\dagger}_{0,1} \\
\tilde{\mathbf{t}}_0 &= (E\mathbf{I} - \mathbf{H}_{0,0})^{-1}\,\mathbf{H}_{0,1}
\end{aligned} \tag{6.57}$$

If we re-write Eq.6.56 for the $n-1$ and $n+1$ GF we readily obtain:

$$\begin{aligned} \mathbf{G}_{n-1,0} &= \mathbf{t}_0\,\mathbf{G}_{n-2,0} + \tilde{\mathbf{t}}_0\,\mathbf{G}_{n,0} \\ \mathbf{G}_{n+1,0} &= \mathbf{t}_0\,\mathbf{G}_{n,0} + \tilde{\mathbf{t}}_0\,\mathbf{G}_{n+2,0} \end{aligned} \tag{6.58}$$

Substituting Eq.6.58 into Eq.6.56 one obtains

$$\mathbf{G}_{n,0} = \mathbf{t}_1\,\mathbf{G}_{n-2,0} + \tilde{\mathbf{t}}_1\,\mathbf{G}_{n+2,0} \tag{6.59}$$

where $n \geq 2$ and

$$\begin{aligned} \mathbf{t}_1 &= (\mathbf{I} - \mathbf{t}_0\,\tilde{\mathbf{t}}_0 - \tilde{\mathbf{t}}_0\,\mathbf{t}_0)^{-1}\,t_0\,t_0 \\ \tilde{\mathbf{t}}_1 &= (\mathbf{I} - \mathbf{t}_0\,\tilde{\mathbf{t}}_0 - \tilde{\mathbf{t}}_0\,\mathbf{t}_0)^{-1}\,\tilde{\mathbf{t}}_0\,\tilde{\mathbf{t}}_0 \end{aligned} \tag{6.60}$$

This procedure can be iterated to obtain the general expression

$$\mathbf{G}_{n,0} = \mathbf{t}_i\,\mathbf{G}_{n-2^i,0} + \tilde{\mathbf{t}}_i\,\mathbf{G}_{n+2^i,0} \tag{6.61}$$

and

$$\begin{aligned} \mathbf{t}_i &= (\mathbf{I} - \mathbf{t}_{i-1}\,\tilde{\mathbf{t}}_{1-i} - \tilde{\mathbf{t}}_{1-i}\,\mathbf{t}_{1-i})^{-1}\,\mathbf{t}_{1-i}\,\mathbf{t}_{1-i} \\ \tilde{\mathbf{t}}_i &= (\mathbf{I} - \mathbf{t}_{1-i}\,\tilde{\mathbf{t}}_{1-i} - \tilde{\mathbf{t}}_{1-i}\,\mathbf{t}_{1-i})^{-1}\,\tilde{\mathbf{t}}_{1-i}\,\tilde{\mathbf{t}}_{1-i} \end{aligned} \tag{6.62}$$

Eq.6.61 holds for $n \geq 2^i$. If we set $n = 2^i$ in Eq.6.61 we can write another chain of equations:

$$\begin{aligned} &\mathbf{G}_{1,0} = \mathbf{t}_0\,\mathbf{G}_{0,0} + \tilde{\mathbf{t}}_0\,\mathbf{G}_{2,0} \\ &\mathbf{G}_{2,0} = \mathbf{t}_1\,\mathbf{G}_{0,0} + \tilde{\mathbf{t}}_1\,\mathbf{G}_{4,0} \\ &\mathbf{G}_{4,0} = \mathbf{t}_2\,\mathbf{G}_{0,0} + \tilde{\mathbf{t}}_2\,\mathbf{G}_{4,0} \\ &\ldots \\ &\mathbf{G}_{2^n,0} = \mathbf{t}_n\,\mathbf{G}_{0,0} + \tilde{\mathbf{t}}_n\,\mathbf{G}_{2^{n+1},0} \end{aligned} \tag{6.63}$$

Substituting the expressions for $\mathbf{G}_{2^n,0}$ for $n \geq 1$ back in the first of the Eqs.6.63, we can express $\mathbf{G}_{1,0}$ as

$$\mathbf{G}_{1,0} = (\mathbf{t}_0 + \tilde{\mathbf{t}}_0\,\mathbf{t}_1 + \tilde{\mathbf{t}}_0\,\tilde{\mathbf{t}}_1\,\mathbf{t}_2 + \ldots + \tilde{\mathbf{t}}_0\,\tilde{\mathbf{t}}_1 \cdots \tilde{\mathbf{t}}_{n-1}\,\mathbf{t}_n)\,\mathbf{G}_{0,0} + \tilde{\mathbf{t}}_n\,\mathbf{G}_{2^{n+1},0} \tag{6.64}$$

In Eq.6.64 the effect of the n-th term should vanish rapidly because it represents the coupling between the 0-th and the n-th layer, so we can truncate the sum when convergence is reached. Thus, Eq.6.64 becomes

$$\mathbf{G}_{1,0} \simeq \mathbf{T}\,\mathbf{G}_{0,0} \tag{6.65}$$

where we have defined the transfer matrix $\mathbf{T}$ as

$$\mathbf{T} = \mathbf{t}_0 + \sum_{n=1}^{N} \left(\prod_{j=0}^{n-1} \tilde{\mathbf{t}}_j \right) \mathbf{t}_n \tag{6.66}$$

and substituting $\mathbf{G}_{1,0}$ in the first of Eq.6.55, we can finally obtain an expression for the surface GF $\mathbf{G}_{0,0}$ which is:

$$\mathbf{G}_{0,0} = (E\mathbf{I} - \mathbf{H}_{0,0} - \mathbf{H}_{0,1}\,\mathbf{T})^{-1} \tag{6.67}$$

6.3 The mode-space approach

If we discretize the 3-D 8-bands $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian of Eq.2.50 directly in real-space by using a finite difference method for each point $\mathbf{r}=(x_i,\, y_j,\, z_k)$, we obtain a block tri-diagonal matrix with sub-matrices $\mathbf{H}(x_i, x_j)$, where i identifies the i-th section. Here x is the transport direction and carriers are confined in the y-z plane (see Fig.6.1). We have already seen in Section 6.2.2

that the tri-diagonal form for the Hamiltonian is necessary in order to use the Dyson equation to calculate the retarded GF. The complete real-space Hamiltonian $\mathbf{H}_R$ reads

$$\mathbf{H}_R = \begin{pmatrix} \mathbf{H}(x_1,x_1) & \mathbf{H}(x_1,x_2) & & & \\ \mathbf{H}(x_2,x_1) & \mathbf{H}(x_2,x_2) & \mathbf{H}(x_2,x_3) & & \\ & \mathbf{H}(x_3,x_2) & \mathbf{H}(x_3,x_3) & & \\ & & & \ddots & \\ & & & \mathbf{H}(x_{N_x},x_{N_x-1}) & \mathbf{H}(x_{N_x},x_{N_x}) \end{pmatrix} \tag{6.68}$$

Each $\mathbf{H}(x_i,x_j)$ has rank $N_b\,N_y\,N_z$, where N_b=8 is the number of bands of the $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian and N_y, N_z are the number of discretization points along the y and z directions, respectively. If we denote with N_x the number of sections in the transport direction, then the total Hamiltonian has rank $N_bN_xN_yN_z$. It can be demonstrated that the non-diagonal matrices $\mathbf{W}_R$ are the same for all sections and the diagonal matrices $\mathbf{H}_R(i)$=$\mathbf{H}(x_i,x_i)$ differ only by the electrostatic potential. Eq.6.68 can thus be rewritten as

$$\mathbf{H}_R = \begin{pmatrix} \mathbf{H}_R(1) & \mathbf{W}_R & & & \\ \mathbf{W}_R^\dagger & \mathbf{H}_R(2) & \mathbf{W}_R & & \\ & \mathbf{W}_R^\dagger & \mathbf{H}_R(3) & & \\ & & & \ddots & \\ & & & \mathbf{W}_R^\dagger & \mathbf{H}_R(N_x) \end{pmatrix} \tag{6.69}$$

where

$$\mathbf{H}_R(i) = \mathbf{H}_{R,yz} + \mathbf{V}_R(i) \tag{6.70}$$

and $\mathbf{H}_{R,yz}$ is the discretized kinetic Hamiltonian of the i-th device section, whereas $\mathbf{V}_R(i)$ is the diagonal matrix containing the energy potential in the device section. The expression for $\mathbf{H}_{R,yz}$ can be found in App.B.

The manipulation of matrices of such dimensions is an almost intractable problem from a computational standpoint. We thus resorted to the theoretically equivalent Coupled Mode-Space formulation [8, 9, 10]. With this approach, the rank of the diagonal and off-diagonal blocks of the Hamiltonian can be reduced by means of the unitary transformation

$$\mathbf{H}_{\mathrm{MS}}(x_i,x_j) = \mathbf{U}(x_i)^\dagger\,\mathbf{H}_R(x_i,x_j)\,\mathbf{U}(x_j) \quad j=\{i,i+1,i-1\} \tag{6.71}$$

where

$$\mathbf{U}(x_i) = [\boldsymbol{\chi}_1(x_i)\cdots\boldsymbol{\chi}_M(x_i)] \tag{6.72}$$

is the rectangular unitary matrix composed by the M eigenvectors of the 2-D Schrödinger problem at the i-th device section expressed by the equation [10]

$$[\mathbf{H}_R(i)+\mathbf{W}_R+\mathbf{W}_R^\dagger]\,\boldsymbol{\chi}_m(x_i) = E_m(x_i)\,\boldsymbol{\chi}_m(x_i) \tag{6.73}$$

Eq.6.71 stems from the transformation of the complete 3-D Hamiltonian of Eq.6.69. Given that $\mathbf{H}_R$ is a tri-diagonal block-matrix, also the mode-space Hamiltonian $\mathbf{H}_{\mathrm{MS}}$ preserves the tri-diagonal form and we can express the blocks as in Eq.6.71.

The eigenvalues $E_m(x_i)$ and wave-functions $\boldsymbol{\chi}_m(x_i)$ are referred to as the modes of the section x_i and the number M of the modes retained in each section sets the rank of the sub-matrices $\mathbf{H}_{\mathrm{MS}}(x_i,x_j)$ of the mode space Hamiltonian. For the problem of the determination of the modes we refer to Section 2.5.1. In the mode-space, Eq.6.8 is re-written as

$$\mathbf{G}_{\mathrm{MS}}(E) = [(E+\mathrm{i}0^+)\mathbf{I} - \mathbf{H}_{\mathrm{MS}} - \boldsymbol{\Sigma}_{\mathrm{MS},C}]^{-1} \tag{6.74}$$

where $\boldsymbol{\Sigma}_{MS,C}$ is the mode-space self-energy for the contact C, which has been transformed as in Eq.6.71.

We extensively validated the mode-space approach in Section 6.5 by calculating some important quantities such as density of states or transmission as a function of the number of modes.

6.4 The electron-phonon interaction in the NEGF model

So far we have only considered coherent transport, thus no phase-breaking scattering effects have been taken into account when we solved the transport equations. We have assumed that all phase-breaking processes occur into the contacts, thus keeping them at the equilibrium. Indeed, it has been demonstrated that scattering events play an important role even in nano-scale devices, and this effect becomes more and more important when the channel length increases [11]. The phonon scattering is important for conventional MOS transistors, where thermo-ionic transport dominates the current and can be important in tunneling-based devices as well, where the transport is dominated by the tunneling components.

Electron-phonon interaction in the NEGF formalism can be viewed as an additional contact (see Fig.6.5) which spills-out and re-injects electrons according to the phonon self-energy $\mathbf{\Sigma}_{\text{ph}}$ [3].

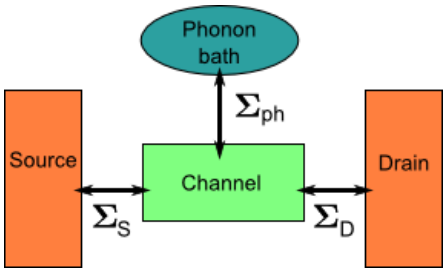


Figure 6.5: Idealized simulated device within the NEGF model. The source and the drain contacts are modeled as semi-infinite reservoirs and are described by their self-energy matrices, $\mathbf{\Sigma}_{\text{S}}$ and $\mathbf{\Sigma}_{\text{D}}$, respectively. The electron-phonon interaction is modeled as an additional contact, with its own self-energy $\mathbf{\Sigma}_{\text{ph}}$.

We will limit our analysis to the so-called self-consistent Born approximation [3], where the interaction is modeled as the coupling between neighboring configurations of the overall electron-phonon system with one more/less phonons. By doing so, Eq.6.8 for the retarded GF modifies as

$$\mathbf{G}(E) = [(E + \mathrm{i}0^+)\mathbf{I} - \mathbf{H} - \mathbf{\Sigma}_{\text{Lead}} - \mathbf{\Sigma}_{\text{ph}}]^{-1} \tag{6.75}$$

where $\mathbf{\Sigma}_{\text{ph}}$ is the retarded self-energy of the phonon bath. Moreover, in order to be able to use the recursive algorithms, we will consider only local interactions, i.e. the phonon self-energies are diagonal matrices in real space.

The NEGF formalism requires to introduce two additional GFs, namely the lesser-than GF $\mathbf{G}^<$ and the greater-than GF $\mathbf{G}^>$, which are related to the electron and hole concentrations, respectively. By writing a generalized Dyson Equation, it can be demonstrated that the lesser-than, greater-than, and retarded GFs are non-linearly coupled and need to be self-consistently solved [12]. The matrix system to be solved reads

$$[(E + \mathrm{i}0^+)\mathbf{I} - \mathbf{H} - \mathbf{\Sigma}_{\text{Lead}} - \mathbf{\Sigma}_{\text{ph}}]\,\mathbf{G} = \mathbf{I} \tag{6.76a}$$

$$\mathbf{G}^< = \mathbf{G}\,\mathbf{\Sigma}^<\,\mathbf{G}^\dagger \tag{6.76b}$$

$$\mathbf{G}^> - \mathbf{G}^< = \mathbf{G} - \mathbf{G}^\dagger \tag{6.76c}$$

$$\mathbf{\Sigma}^> - \mathbf{\Sigma}^< = \mathbf{\Sigma} - \mathbf{\Sigma}^\dagger \tag{6.76d}$$

and the calculation of $\mathbf{G}$, $\mathbf{G}^<$ and $\mathbf{G}^>$ is coupled because the $\mathbf{\Sigma}_{\text{ph}}$ entering Eq.6.76a is expressed in terms of $\mathbf{G}^<$ and $\mathbf{G}^>$. Hence, we implemented an additional self-consistent loop inside the

NEGF solver in order to calculate the different GFs. Here, the lesser-than self-energies for phonon scattering are expressed within the self-consistent Born approximation:

$$\mathbf{\Sigma}^{<(>)} = \mathbf{D}_0^{<(>)} \mathbf{G}^{<(>)} \tag{6.77}$$

where $\mathbf{D}_0^{<(>)}$ is proportional to the Green's function of the unperturbed phonon bath [13]. More precisely, the mode-space lesser-than self-energy for acoustic phonons and for the $n-$th mode reads

$$\mathbf{\Sigma}^{<\,n,n}_{\mathrm{MS,ac}}(x_i, x_i; E) = \frac{D_{\mathrm{ac}}^2 k_B T}{\rho v_s^2} \sum_m I^{m,n}(x_i, x_i)\, \mathbf{G}^{<\,m,m}_{\mathrm{MS}}(x_i, x_i; E)\,, \tag{6.78}$$

where

$$I^{m,n}(x_i, x_j) = \int \mathrm{d}y \int \mathrm{d}z\, |\chi_m(x_i, y, z)|^2\, |\chi_n(x_j, y, z)|^2 \tag{6.79}$$

is the usual form factor, ρ is the material density, v_s is the sound velocity, T is the temperature and D_{ac} is the acoustic deformation potential (see Tab.6.1). The lesser-than self-energy for dispersionless optical phonons can be expressed as

$$\mathbf{\Sigma}^{<\,n,n}_{\mathrm{MS,opt}}(x_i, x_i; E) = \frac{\hbar D_{\mathrm{opt}}^2}{2\rho\omega_j} \sum_m I^{m,n}(x_i, x_i)\, \mathbf{G}^{<\,m,m}_{\mathrm{MS}}(x_i, x_i; E \pm \hbar\omega_j) \left[N_j + \frac{1}{2} \pm \frac{1}{2} \right] \tag{6.80}$$

where ω_j is the frequency of the $j-$th optical branch, D_{opt} is the optical deformation potential and N_j is the equilibrium phonon density given by the Bose statistics. Acoustic and optical phonon parameters for sSi MOSFETs were taken from [14].

	D_{ac}[eV]	D_{pop}[10^8 eV/cm]	$\hbar\omega_{\mathrm{pop}}$[meV]
Electrons	5.8	20	30
Holes	1.0	20	30

Table 6.1: Phonon scattering parameters for InAs. Acoustic deformation potentials D_{ac} and phonon energy $\hbar\omega_{\mathrm{pop}}$ for polar phonons are from [15]. Mass density ρ=5.67 g/cm^3 and sound velocity v_s=3·10^5 cm/s are from [16]. The value of D_{pop} is discussed in the text.

Polar optical phonon (POP) scattering was also included in InAs device simulations due to its relevance in III-V compounds. However, although this is a non-local mechanism, it was accounted for by means of a local self-energy similar to Eq. 6.80 with effective deformation potential D_{pop} and frequency ω_{pop} reported in Tab.6.1. This approximation was adopted to circumvent the numerical difficulties of implementing non local-self-energies, which would prevent the use of the recursive algorithms used to handle 3D quantum transport problems [17].

In order to solve Eq.6.76, we used a more general recursive scheme for the calculation of the lesser-than GF [18] starting from Eq.6.76.

When phase-breaking transport in included in the simulation, the electron and hole concentrations can be computed as [3]

$$n(\mathbf{r}_d) = \frac{1}{2\pi} \mathrm{Im} \left[\int \mathrm{d}E\, \mathbf{G}^{<}(\mathbf{r}_d, \mathbf{r}_d, E) \right] \tag{6.81}$$

$$p(\mathbf{r}_d) = -\frac{1}{2\pi} \mathrm{Im} \left[\int \mathrm{d}E\, \mathbf{G}^{>}(\mathbf{r}_d, \mathbf{r}_d, E) \right] \tag{6.82}$$

The drain current, instead, reads [3]

$$I = -\frac{e}{\hbar} \int \frac{\mathrm{d}E}{2\pi} \left[\mathrm{Trace}[\mathbf{\Gamma}_{\mathrm{D}} \mathbf{G} \mathbf{\Gamma}_{\mathrm{D}} \mathbf{G}^{\dagger}] f_0(E - E_{f,D}) - \mathrm{Trace}[\mathbf{\Gamma}_{\mathrm{D}}\, \mathbf{G}^{<}] \right] \tag{6.83}$$

6.5 Validation of the mode-space approach

As a necessary step, we extensively verified the solidity of our models by calculating some important quantities as a function of the number of modes. In this section we report a few examples of prominent relevance, where we validated the results of our mode-space approach and compared the calculated drain-current for a Tunnel-FET against atomistic tight-binding simulations taken from the literature.

First of all, we calculated the total density of states *DOS* for an infinite InAs nanowire, starting from the bulk GF calculated with the Sancho-Rubio algorithm (see Section 6.2.6) and using the mode-space procedure (see Section 6.3). Fig.6.6 shows the *DOS* as a function of the total energy and calculated for different number of modes. The first thing we notice is

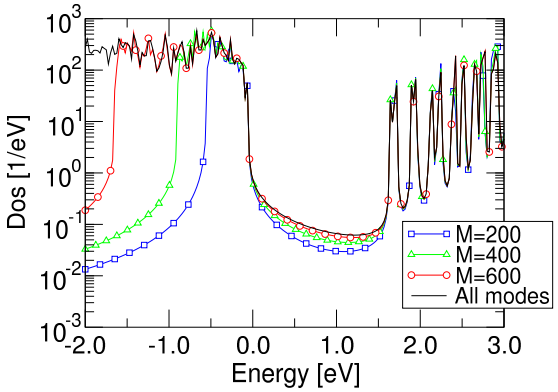


Figure 6.6: Total density of states *DOS* for an infinite bulk InAs nanowire with a 2nm square section and for a flat potential. The energy reference is taken at 0eV for the top of the valence band. By increasing the number of modes included in the calculation, we recover the results obtained with the full-mode model.

that by increasing the number of modes, we reproduce better the *DOS* for energies farther from the energy gap. Moreover, we used a particular mode-selection algorithm, where we selected the modes for the valence (conduction) band starting from the top (bottom) of the valence (conduction) band and going towards lower (higher) energies. By doing so, we consider only the most occupied subbands, which are the ones giving the largest contribution to the transmission and hence to the drain current.

Here we anticipate some concepts and results that will be treated in detail in Chapter 7. In particular, we refer to a simulation study of InAs Tunnel-FET devices, which represents the main application of the full-quantum model we investigated in this thesis. Fig.6.7(left) reports the subband profile for an InAs nanowire Tunnel-FET (see Section 7.2.2), while Fig.6.7(right) shows the corresponding transmission for a ballistic case as a function of the energy. It can be seen that the simulation results are essentially independent of the number of modes M for M larger than about 40. In our simulations we typically used 50 modes and selected the eigenvalues so that approximately $M/6$ modes are taken from the conduction band, while the remaining part from the valence band. This is justified by the smaller effective mass in conduction band for III-V materials like InAs or GaAs, and consequently by a stronger band splitting effect with quantization.

A possible limitation of our band-structure model for InAs Tunnel-FETs based on the 8-band $\mathbf{k}\cdot\mathbf{p}$ approach is the fact that we do not account for the Λ and Δ minima of the conduction band, which may lead to an underestimation of the tunneling current in the n-type Tunnel-FET (see Section 7.2). Quantitatively speaking, however, we think that this is not a serious limitation for an InAs device, because the Λ and Δ minima are respectively about 0.72eV and 1.0eV above the Γ minimum in the bulk material [15], hence they are not expected to provide

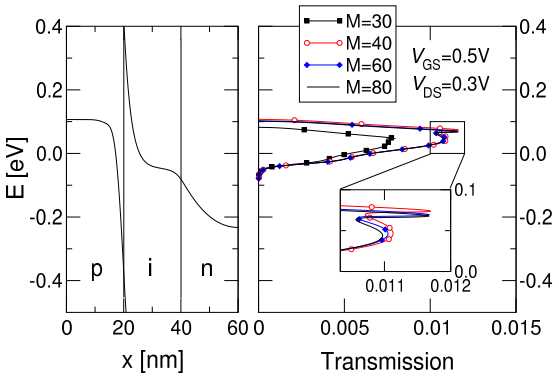


Figure 6.7: Left: subbands profile; the source Fermi level E_{Fs} is taken as zero. Right: transmission as a function of energy for different number M of modes. V_{GS}=0.5V, V_{DS}=0.3V. The simulation results are essentially independent of M for M larger than about 40. The inset shows a zoom of the transmission peak, where we see that a stable value is obtained with only 40 modes.

a large contribution to the current unless we go to very narrow nanowires, which in turn is not advisable for the design of the device.

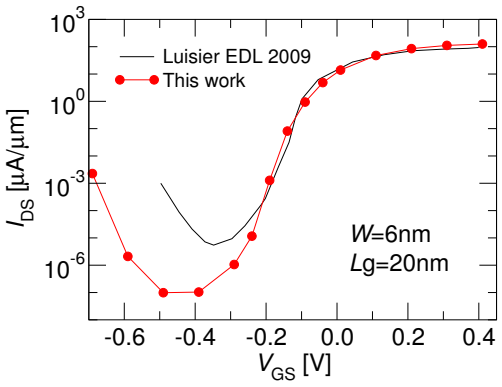


Figure 6.8: Drain current versus gate voltage characteristics for an InAs Tunnel-FET with L_{G}=20nm. Solid Line: simulation results from [19] for a circular nanowire with a 6nm diameter. Symbols: simulation results of this work for a squared nanowire with W=H=6nm. The difference between the two curves at low V_{GS} originates mainly from the different geometry of the nanowires' cross sections.

To test the validity of our simulation approach we compared our results with the simulations obtained with an atomistic tight-binding Hamiltonian [19]. Fig.6.8 illustrates the I_{DS} versus V_{GS} characteristics obtained with our model for a squared nanowire with W=H=6nm, channel length L_{G}=20nm and a symmetric source-drain doping (N_{S}=N_{D}=$5{\cdot}10^{19}\mathrm{cm}^{-3}$), together with the corresponding results reported in [19] for a cylindrical nanowire with a 6nm diameter; the currents are normalized to W=H or to the diameter for the squared and the circular nanowire, respectively. As it can be seen the agreement between the two models is very good in the turn-on region and also in deep subthreshold for I_{DS} down to about 10pA/μm, that is essentially in the entire range of practical interest. The minimum I_{DS} value and the onset of the ambipolar behavior is instead quantitatively different in our results with respect to [19]. We ascribe this discrepancy mainly to a different geometry of the device cross sections.

Bibliography

[1] S. Laux, A. Kumar, and M. Fischetti, "Ballistic FET modeling using QDAME: quantum device analysis by modal evaluation," *IEEE Transactions on Nanotechnology*, vol. 1, no. 4, pp. 255 – 259, dec 2002.

[2] M. D. Croitoru, V. N. Gladilin, V. M. Fomin, J. T. Devreese, W. Magnus, W. Schoenmaker, and B. Sorée, "Quantum transport in a nanosize double-gate metal-oxide-semiconductor field-effect transistor," *Journal of Applied Physics*, vol. 96, no. 4, pp. 2305–2310, 2004.

[3] S. Datta, *Quantum Transport - Atom to Transistor*. United Kingdom: Cambridge University Press, 2005.

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

[5] S. Davison and K. Sulston, *Green-Function Theory of Chemisorption*. Springer, 2006.

[6] G. Klimeck, "Numerical Aspects of NEGF: The Recursive Green Function Algorithm," Aug 2005. [Online]. Available: http://nanohub.org/resources/165

[7] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio, "Quick iterative scheme for the calculation of transfer matrices: application to Mo (100)," *Journal of Physics F: Metal Physics*, vol. 14, no. 5, p. 1205, 1984.

[8] R. Venugopal, Z. Ren, S. Datta, M. S. Lundstrom, and D. Jovanovic, "Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches," *Journal of Applied Physics*, vol. 92, p. 3730, 2002.

[9] M. Luisier, A. Schenk, and W. Fichtner, "Quantum transport in two- and three-dimensional nanoscale transistors: Coupled mode effects in the Nonequilibrium Green's Function formalism," *Journal of Applied Physics*, vol. 100, p. 043713, 2006.

[10] M. Shin, "Full-quantum simulation of hole transport and band-to-band tunneling in nanowires using the k·p method," *Journal of Applied Physics*, vol. 106, no. 5, p. 054505, 2009.

[11] P. Palestri, D. Esseni, S. Eminente, C. Fiegna, E. Sangiorgi, and L. Selmi, "Understanding Quasi-Ballistic Transport in nano-MOSFETs. Part I: Scattering in the Channel and in the Drain," *IEEE Trans. Electron Devices*, vol. 52, no. 12, pp. 2727–2735, 2005.

[12] G. Mahan, *Many-particle physics*. New York: Kluwer, 2000.

[13] S. Poli and M. Pala, "Channel-Length Dependence of Low-Field Mobility in Silicon-Nanowire FETs," *IEEE Electron Device Lett.*, vol. 30, no. 11, pp. 1212 –1214, nov. 2009.

[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. Electron Devices*, vol. 55, no. 8, pp. 2086 – 2096, nov. 2008.

[15] 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.

[16] M. V. Fischetti, "Monte Carlo simulation of transport in technologically significant semiconductors of the Diamond and Zinc-Blende structures - Part I: homogeneous transport," *IEEE Trans. Electron Devices*, vol. ED-38, p. 634, 1991.

[17] M. Luisier and G. Klimeck, "Atomistic full-band simulations of silicon nanowire transistors: Effects of electron-phonon scattering," *Phys. Rev. B*, vol. 80, p. 155430, Oct 2009.

[18] R. Lake, G. klimeck, R. Bowen, and D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices," *Journal of Applied Physics*, vol. 81, p. 7845, 1997.

[19] M. Luisier and G. Klimeck, "Atomistic Full-Band Design Study of InAs Band-to-Band Tunneling Field-Effect Transistors," *IEEE Electron Device Lett.*, vol. 30, no. 6, pp. 602 –604, june 2009.

Chapter 7

Tunnel-FETs and strain optimization

7.1 Introduction

THIS chapter presents the simulation results obtained with the full-quantum model developed during this PhD work. In particular, we focused our attention on the simulation of InAs nanowire Tunnel-FETs, which are very promising devices for ultra low-voltage and low-energy applications. The main advantage of tunneling-based devices resides in the competitive values of the inverse sub-threshold slope (SS) that can be obtained both in silicon and in III-V transistors, much below the theoretical limit of 60mV/decade of conventional MOSFETs. This turns into very low off-current values and high switching frequencies. However, since drain current is dominated by the tunneling components, the on-current values that can be reached are far below with respect to conventional MOSFET technologies and ITRS requirements [1]. In this respect, we performed a simulation study of InAs Tunnel-FETs with a systematic benchmark against conventional silicon MOSFETs in terms of performances and surface roughness induced variability.

In particular, Section 7.2 investigates the strain engineering in n-type InAs nanowire Tunnel-FETs. To this purpose we used the NEGF formalism based on a 8-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian already presented in Chapter 6. The model accounts for arbitrary crystal orientations and describes the strain implicitly by a modification of the band-structure. Acoustic and optical phonon scattering is also accounted for in the local approximation. Our results show that appropriate strain conditions in InAs Tunnel-FETs enable: (a) a remarkable enhancement of the I_{on} with no significant degradation of the sub-threshold slope; (b) large improvements in the I_{off} versus I_{on} tradeoff for low I_{off} and V_{DD} values; (c) significant widening of the range of I_{off} and V_{DD} values where Tunnel-FETs can compete with strained silicon MOSFETs.

Subsequently, Section 7.3 presents a comparative study of the surface roughness induced variability at low supply voltage V_{DD}=0.3V in nanowire InAs Tunnel-FETs and Si MOSFETs. By exploiting a 3-D full-quantum approach based on the NEGF formalism, we show that, contrary to Si MOSFETs, InAs Tunnel-FETs present an I_{on} variability much smaller than the I_{off} variability. We explain this result by noting that in the source depletion region, where the tunneling mainly occurs for V_{GS}=V_{DD}, microscopic subband fluctuations induced by surface roughness are small compared to the macroscopic band bending due to the built-in potential and to the gate bias.

7.2 Effect of strain on InAs Tunnel-FET devices

7.2.1 Context of the study

The design of most integrated circuits and systems is nowadays constrained by the budget for the available power or energy and ultra low power applications have become pervasive [2]. In particular, energy autonomous systems for ambient intelligence should be designed for minimum energy operation and are expected to work at supply voltages V_{DD} below 0.5V [3, 4]. The inverse sub-threshold slope (SS) of conventional MOSFETs (limited to 60mV/dec at room temperature) is a fundamental limit to the V_{DD} scaling and the ITRS 2009 has recognized the steep slope transistors as emerging devices and singled out tunnel FETs (Tunnel-FETs) as the most promising steep slope transistors [5]. However, silicon Tunnel-FETs exhibit poor on currents (I_{on}) [6], and improving the I_{on} while preserving small off currents (I_{off}) and SS values is probably the greatest challenge for Tunnel-FETs [1]. To this purpose the use of high-κ dielectrics, spacers and underlaps [7, 8], of strain in Si, Ge and SiGe Tunnel-FETs [9, 10], of broken and staggered bandgaps [11, 12, 13] and of graphene based Tunnel-FETs [14, 15, 16] has been investigated. Today, however, it is still unclear if there exist I_{off} and V_{DD} domains where Tunnel-FETs may outperform MOSFETs.

This section presents a simulation study exploring the improvements in the performance of homo-junction InAs nanowire Tunnel-FETs obtained by using appropriate stress conditions [17,

18]. To this purpose, we developed a three-dimensional quantum transport simulator in the framework of the non equilibrium Green's function (NEGF) formalism and based on the eight-band **k·p** Hamiltonian discussed in [19]. The model can deal with arbitrary crystal orientations and we investigated both coherent and dissipative transport to test the influence of phonon scattering on the device performance.

Our simulation results identify stress conditions that reduce the gap as well as the imaginary wavevector in the energy gap of InAs, which are thus very favorable for Tunnel-FETs applications. Such stress configurations result in remarkable drain current improvements with a modest impact on the sub-threshold slope. We also develop a systematic comparison between InAs Tunnel-FETs and silicon nanowire MOS transistors and show that the stress engineering in InAs Tunnel-FETs can help widen the window of I_{off} and V_{DD} values where Tunnel-FETs may outperform silicon MOSFETs.

We performed both coherent and dissipative quantum transport simulations based on the self-consistent solution of the 3-D Poisson and Schrödinger equations within the NEGF formalism. We simulated strained silicon (sSi) MOSFETs according to the effective mass approximation [20, 21], whereas for InAs Tunnel-FETs we developed a new simulator based on the 8-band **k·p** Hamiltonian already described in Chapter 6. Sec.7.2.2 deals with the structure of InAs nanowires and Sec.7.2.3 illustrates the effect of some promising stress conditions. Finally, in Sec.7.2.4 we compare InAs Tunnel-FETs and Si nanowires in terms of simple delay and energy metrics.

7.2.2 Design of the Tunnel-FETs

The gate-all-around (GAA), rectangular nanowire transistors studied in this work are sketched in Fig.7.1, where x is the transport direction and carriers are confined in the y-z plane; the coordinates (x,y,z) form the device coordinate system (DCS) and the figure shows also the crystal coordinate system (CCS).

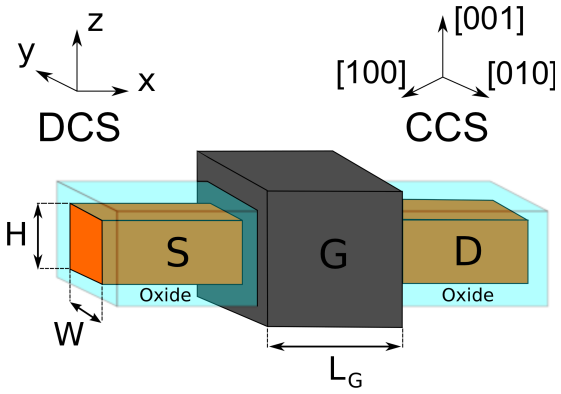


Figure 7.1: Sketch of the simulated nanowire transistors. The Device Coordinate System (DCS) and Crystal Coordinate System (CCS) are also shown.

The simulated nanowire transistors have the parameters of Tab.7.1, if not otherwise stated. The sSi MOSFETs have a 2% tensile strain in the [100] direction, that corresponds to the transport direction. Both devices have the same structure but different doping species and concentrations.

Fig.7.2(a) reports the calculated energy dispersion in unstrained InAs nanowires for the three main transport directions, namely [100], [110] and [111]. It can be seen that the energy relations for the real branches as well as for the fundamental imaginary branch are very similar for the different orientations. Fig.7.2(b) shows the corresponding IV curves for the unstrained

L_G [nm]	$W=H$ [nm]	L_{SD} [nm]	T_{ox} [nm]	N_S [cm^{-3}]	N_D [cm^{-3}]	Transport direction x
20	5	20	1	$5{\cdot}10^{19}$	10^{18}	[100]

Table 7.1: Device parameters of the InAs tunnel FETs used in the simulations (if not stated otherwise). The gate dielectric is SiO_2 with κ=3.9. L_{SD} is the length of source and drain regions, N_S and N_D are the doping concentrations in the source and drain (respectively acceptor and donor type). The sSi MOSFETs have the same device structure as the Tunnel-FETs but both source and drain have a donor type doping concentration of 10^{20}cm^{-3}.

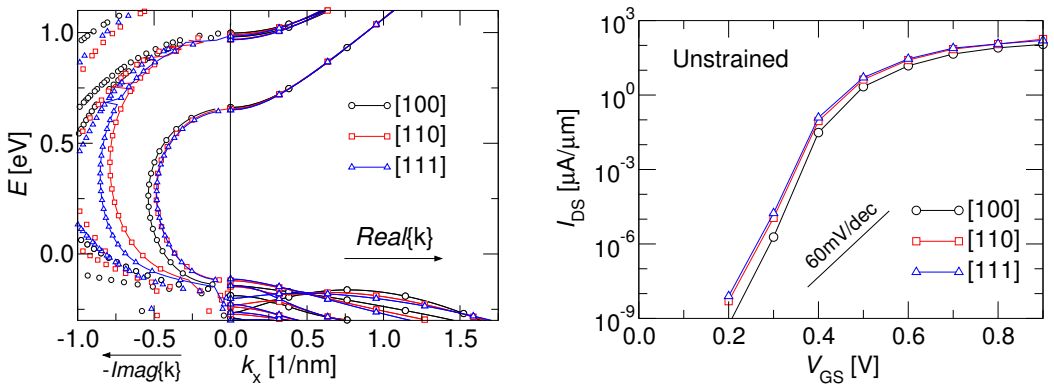


Figure 7.2: (a) Energy dispersion for a 1D carrier gas in an InAs nanowire (W=H=5nm) and for different transport directions x (see Fig.7.1). The potential energy inside the wire is taken as zero. The energy relation in the gap is very similar for the different x. (b) Drain current versus gate voltage characteristics of the InAs Tunnel-FET nanowire for different transport directions. The biaxial strain is always perpendicular with respect to the transport direction. Electron-phonon interaction has not been included. Gate work-function Φ_M=4.7eV.

and biaxially strained case. The three channel orientations exhibit very close drain currents and a slightly degraded SS for the non-[100] orientations, hence in the rest of the paper we will focus on [100] nanowires.

Fig.7.3 (left) compares the I_{DS} versus V_{GS} curves (IV) for an L_G=20nm unstrained InAs Tunnel-FET and a sSi MOSFET; the Tunnel-FET has a far better SS ($\simeq$22mV/dec), especially at T=400K. The SS value has been averaged for I_{DS} between 10pA/μm and 10nA/μm at T=300K and 400K. The drain current I_{DS} is always normalized to W=H. As discussed in [17], a reduction of the gate length L_G to 15nm slightly degrades the SS value and does not improve the I_{DS}, while a larger drain doping N_D deteriorates SS because of the tunneling at the drain (not shown). A cross section W=H=5nm is also a good choice because smaller W, H values improve only slightly the SS but reduce substantially the I_{DS} per unit width [17].

The possible impact of gradual and thus more realistic doping profiles is illustrated in Fig.7.3 (right), which compares the IV characteristics for InAs Tunnel-FETs for both an exponential decay of the source doping, identified by the inverse logarithmic slope (1 decade every 2nm), and a Gaussian decay (σ=2nm and σ=3nm). The Gaussian doping profile causes a non-negligible reduction of the on-current for σ greater than approximately 2nm, while the exponential profile does not affect appreciably either the sub-threshold slope or the on-current. The simulations reported in the remaining part of the paper are obtained for abrupt doping profiles in the source-drain extensions.

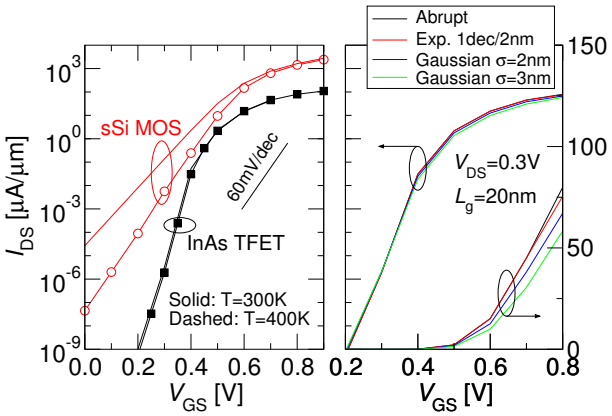
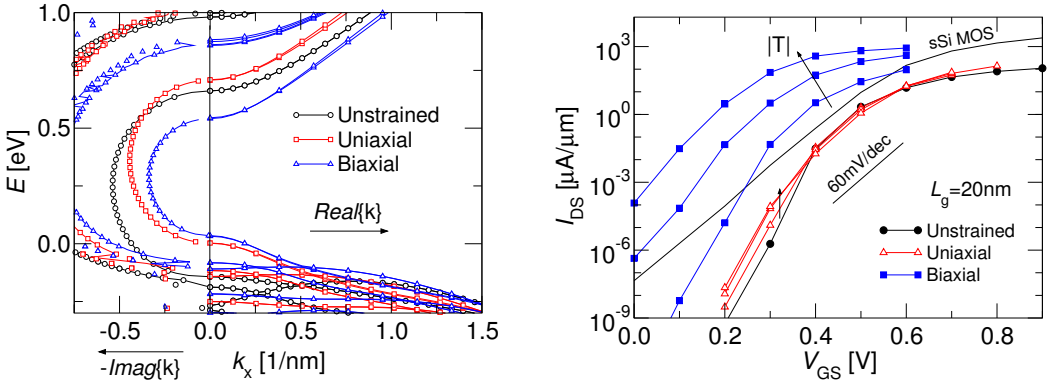


Figure 7.3: Left: drain current versus gate voltage characteristics for an InAs Tunnel-FET and a sSi MOSFET at T=300K and T=400K. The drain current is always normalized to W=H=5nm. The MOSFET has an SS (averaged for I_{DS} between $10pA/\mu$m and $10nA/\mu$m) of about 62 and 81mV/dec respectively at T=300K and 400K, while the Tunnel-FET has an SS≈22mV/dec independently of T. Right: drain current versus gate voltage of the InAs Tunnel-FET in logarithmic and linear scale for an exponential and a Gaussian doping profile and compared to the results for abrupt doping profiles. Electron-phonon interaction has not been included. Gate work-function Φ_M=4.7eV.

7.2.3 Strain engineering

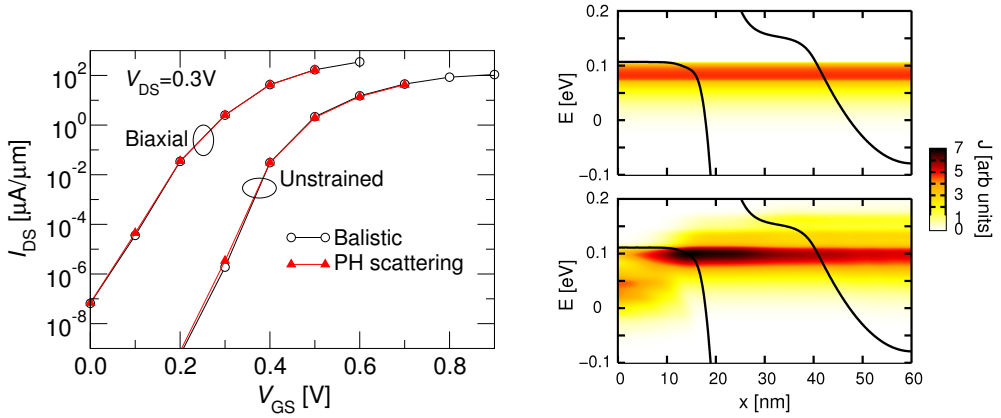


(a) (b)

Figure 7.4: (a) Energy dispersion in the gap for an InAs nanowire (W=H=5nm) and for k_x along the [100] direction. Comparison between unstrained, compressive uniaxial stress (T_{xx}=-2GPa) and tensile biaxial (T_{yy}=T_{zz}=2GPa) stress condition. The potential energy inside the wire is taken as zero. (b) Drain current versus gate voltage characteristics for different strain conditions. Compressive uniaxial stress is T_{xx}=-1, -2 and -3GPa and tensile biaxial stress is T_{yy}=T_{zz}=1, 2 and 3GPa. The drain current for the sSi MOSFET has been reported for comparison. Electron-phonon interaction has not been included. Device parameters from Tab.7.1. V_{DS}=0.3V. Gate work-function Φ_M=4.7eV.

Fig.7.4(a) illustrates the energy dispersion for unstrained and strained InAs nanowires. As it can be seen, both uniaxial compressive and biaxial tensile stress shift up the valence band, but the biaxial stress also lowers the conduction band and thus obtains the largest reduction

of the energy gap and of the imaginary wavevector in the gap. Fig.7.4(b) compares the IV curves of Tunnel-FETs for different stress conditions. The biaxial stress reduces the V_T of the transistors and increases remarkably the I_{DS}, while the uniaxial stress has a smaller impact on the IV curves.



(a) (b)

Figure 7.5: (a) Drain current versus gate voltage characteristics of the InAs Tunnel-FET for coherent (open symbols) and dissipative simulations (filled symbols). (b) Current spectrum for the coherent (top) and dissipative (bottom) simulations. Scattering parameters from Tab.6.1. Device parameters from Tab.7.1. V_{GS}=0.3V, V_{DS}=0.3V.

Fig.7.5(left) shows that the phonon scattering has a very modest effect on the IV of these short-channel InAs Tunnel-FETs. We verified that it has an impact of less than 10% in the ON condition, consistently with [22]),even if we chose a deliberately large D_{pop} value (ten times larger than for InAs inter-valley optical phonons [23]), in order to test the impact of inelastic scattering for an overestimated scattering rate (see Tab.6.1). In fact, the tunneling path assisted by phonon absorption that is observed in Fig.7.5(right) (consistently with [13]), results in only a tiny I_{DS} increase that hardly affects the IV curves. Moreover, the transistor is short enough that essentially coherent transport conditions hold in the conduction band. The very modest impact of the scattering illustrated in Fig.7.5 is fully consistent with the results in [22].

The potentials of the strain are further clarified by studying the I_{on} at fixed I_{off}, that is the I_{DS} at V_{DS}=V_{DD} and V_{GS}=($V_{GS,off}$+V_{DD}), where $V_{GS,off}$ is the V_{GS} giving I_{off} [13]. This analysis is illustrated in Fig.7.6 for V_{DD}=0.3V and the I_{off} targets projected by the ITRS roadmap are also indicated: I_{off}=5pA/μm for the low stand-by power (LSTP), 5nA/μm for the low-operating power (LOP) and 100nA/μm for high-performance (HP) applications [5]. The biaxial strain improves remarkably the I_{on} to I_{off} tradeoff of Tunnel-FETs. Furthermore, due to the larger SS values, the I_{on} deteriorates more in sSi MOSFETs than in Tunnel-FETs when I_{off} is reduced below about 10nA/μm, so that for low target I_{off} the Tunnel-FETs exhibit an I_{on} advantage that increases at T=400K. The comparison in Fig.7.6 presumes a large flexibility in the gate work-function Φ_M, which is implicitly changed to meet the desired I_{off} values for the different devices. For most of the Tunnel-FETs in Fig.7.4(b), Φ_M should be increased thus moving towards midgap metal gates. In this respect, Fig.7.7 reports the Φ_M values necessary to meet the I_{off} requirements for different uniaxial or biaxial stress conditions; as it can be seen realistic Φ_M values are obtained ranging roughly from midgap down to 4.3V.

The physical mechanism for the I_{on} enhancements at fixed I_{off} is investigated in Fig.7.8, comparing the subband profiles and the current spectra. The current spectra are peaked at the source Fermi level E_{fS}, as expected, and the two strain conditions result in a very similar tunneling distance at E=E_{fS}, only slightly smaller than in the unstrained device. Hence, the

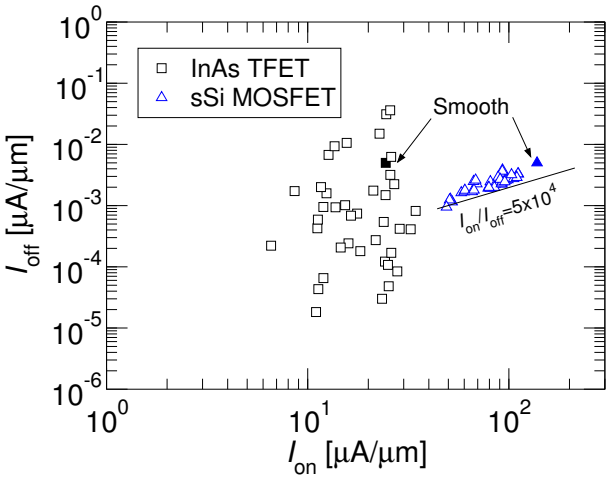


Figure 7.6: Off current versus on current for V_{DD}=0.3V (obtained as explained in the text), for sSi MOSFET (T=300K and T=400K) and for InAs Tunnel-FETs (T=300K). Results for InAs Tunnel-FETs at T=400K (not shown) are almost identical to those for T=300K (see Fig.7.3(left)). The arrow indicates the stress induced I_{on} improvement. The I_{off} targets projected by the ITRS roadmap for the low stand-by power (LSTP), low-operating power (LOP) and high-performance (HP) applications are indicated [5] Device parameters from Tab.7.1. Uniaxial stress: T_{xx}=-2GPa; biaxial stress: T_{yy}=T_{zz}=2GPa.

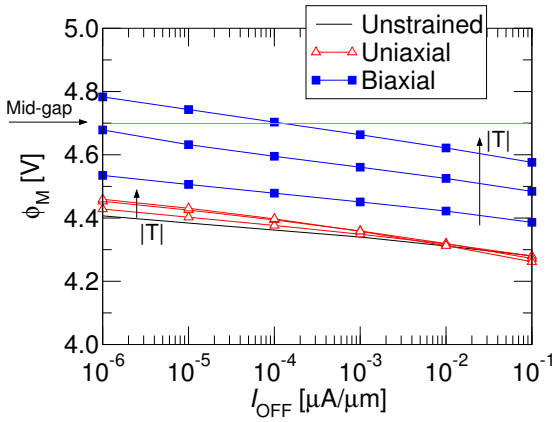


Figure 7.7: Gate work-function Φ_M implicitly assumed in InAs Tunnel-FETs to achieve different target values of I_{off} (on the x-axis) in the unstrained case and for different uniaxial or biaxial stress conditions. The arrows indicate increasing magnitudes of the compressive uniaxial stress (T_{xx}=−1, −2 and −3GPa) and of the tensile biaxial stress T_{yy}=T_{zz}=1, 2 and 3GPa. All the I_{off} values and stress conditions considered in this work lead to realistic Φ_M values.

larger current density for the biaxial stress mainly stems from the smaller imaginary wavevector in the energy gap (see Fig.7.4(a)).

A possible concern about Tunnel-FETs for ultra low voltage applications is that the output characteristics typically lack a good linear region. In fact a threshold-like behavior with respect to V_{DS} is observed with a drastic I_{DS} degradation at small V_{DS} [24, 25]. Fig.7.9 investigates this point showing the I_{DS} versus V_{DS} curves for the InAs Tunnel-FETs studied in this work and for V_{DD}=0.3V. As it can be seen, the Tunnel-FETs show a very good saturation region and, furthermore, exhibit only a modest degradation of the output characteristics at small V_{DS}

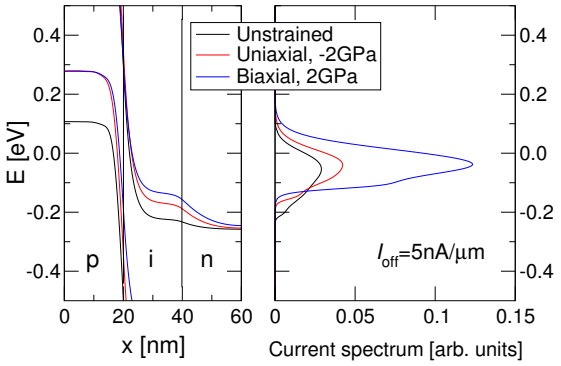


Figure 7.8: Left: subbands profile; the source Fermi level is taken as zero. Right: current spectrum as a function of energy for the unstrained (solid), uniaxially strained (dashed) and biaxially strained (dash-dotted) Tunnel-FETs. $I_{\rm off}$=5nA/μm and ($V_{\rm GS}$-$V_{\rm GS,off}$)=$V_{\rm DS}$=0.3V.

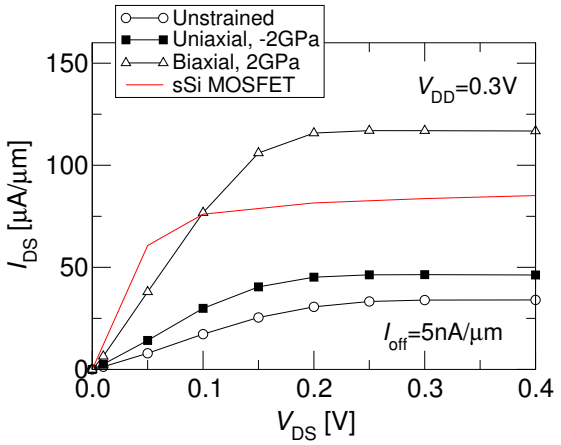


Figure 7.9: Drain current versus drain voltage characteristics for an InAs Tunnel-FET and a sSi MOSFET. The $V_{\rm GS}$ is equal to ($V_{\rm GS,off}$+$V_{\rm DD}$), where $V_{\rm GS,off}$ is the $V_{\rm GS}$ giving $I_{\rm off}$=5nA/μm. $V_{\rm DD}$=0.3V.

compared to the sSi MOSFET. The biaxial strain improves the $I_{\rm DS}$ in the entire $V_{\rm DS}$ range.

7.2.4 Delay and energy metrics

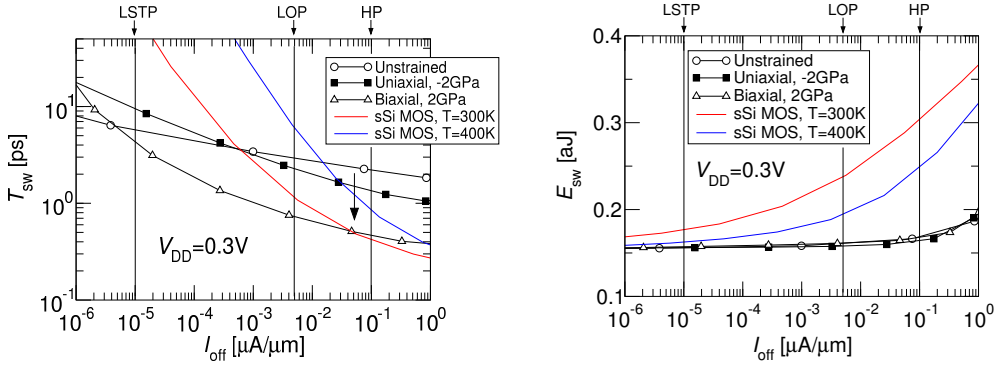
In order to move one step further in the assessment of the possible advantages related to stress engineering, we evaluated some simple metrics for the delay and the energy corresponding to the switching of a digital inverter. In particular, we define a first order estimate of the switching time $T_{\rm sw}$ as

$$T_{\rm sw} = \left[\frac{(Q_{\rm on} - Q_{\rm off}) + C_p V_{\rm DD}}{I_{\rm on}} \right], \tag{7.1}$$

where $Q_{\rm on}$ and $Q_{\rm off}$ are the channel charges in the on and off state of the transistor, and they were obtained by integrating numerically the charge density in the devices (see Eq.6.13). The gate-drain parasitic capacitance C_p=1.7aF is inferred from the fringing capacitance 170aF/μm projected by the ITRS roadmap at $L_{\rm G}$ around 20nm [5]. More refined estimates of the switching time can be obtained by replacing the $I_{\rm on}$ in Eq.7.1 with effective drive currents [26], however Eq.7.1 is adequate for the purposes of our analysis. A similarly simple metric for the switching

energy E_{sw} can be expressed as

$$E_{sw} = V_{DD}(Q_{on} - Q_{off}) + C_p V_{DD}^2. \quad (7.2)$$



(a) (b)

Figure 7.10: (a) T_{sw} (defined in Eq.7.1) versus I_{off} for a parasitic capacitance C_p=1.7aF. V_{DD}=0.3V. The arrow indicates the stress induced T_{sw} improvement. (b) E_{sw} (defined in Eq.7.2) versus I_{off} for a parasitic capacitance C_p=1.7aF. V_{DD}=0.3V. In both plots the I_{off} targets projected by the ITRS roadmap for LSTP, LOP and HP applications are indicated [5]. Device parameters from Tab.7.1.

Fig.7.10(a) shows that for V_{DD}=0.3V the biaxially stressed Tunnel-FETs can achieve T_{sw} values smaller than MOSFETs in a wide I_{off} range (that increases with the temperature). This is partly due to the much smaller charge Q_{on} in the on-state of Tunnel-FETs, that never reach strong inversion for V_{DS}=V_{DD}, as confirmed by the E_{sw} in Fig.7.10(b), showing that the charge in Tunnel-FETs is smaller than in MOSFETs so that E_{sw} is dominated by the contribution $C_p V_{DD}^2$ ≃0.15aJ. The smaller Q_{on} value is a distinct advantage of Tunnel-FETs for ultra-low energy applications and suggests that if C_p could be proportionally reduced, even better performance with respect to CMOS could be achieved.

The analysis in Figs.7.6 and 7.10 was extended for V_{DD} ranging from 0.2V to 0.5V. Fig.7.11 reports the I_{on} and T_{sw} versus V_{DD} at fixed I_{off} and reveals that the strain engineering enlarges significantly the range of I_{off} and V_{DD} values where Tunnel-FETs may outperform MOSFETs.

7.3 Surface roughness study in InAs Tunnel-FETs

7.3.1 Context

Tunnel FETs are probably the most promising electron devices that, by reducing the inverse subthreshold slope (SS) below 60mV/dec at room temperature, may lead to digital integrated circuits operating at a supply voltage, V_{DD}, substantially lower than 0.5V [1]. Such small V_{DD} values are strongly demanded by minimum energy design methodologies [3], and make the noise margins and the related variability issues extremely critical [3].

For double-gate Tunnel FETs the sensitivity to the device parameter fluctuations has been recently investigated by using a TCAD based approach [27], however, despite the fact that gate-all-around (GAA) nanowire (NW) structures are ideally suited to achieve small SS values in Tunnel-FETs [28], the variability of NW Tunnel-FETs is still largely unexplored. While surface roughness (SR) variability is known to affect quite critically NW MOS transistors [29, 30, 31, 32], the differences in the working principle may lead to a different sensitivity to SR for NW Tunnel-FETs with respect to NW MOSFETs.

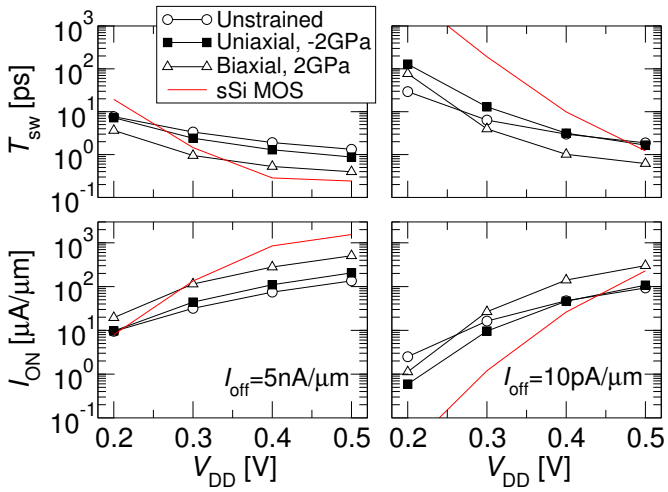


Figure 7.11: T_{sw} (top) (defined in Tab.3) and I_{on} (bottom) versus V_{DD} and for the LOP target I_{off}=5nA/μm (left) and the LSTP target I_{off}=10pA/μm (right). Device parameters from Tab.7.1.

In this section we employed the same 3-D full-quantum approach used in the previous section to develop a comparative study concerning the surface roughness variability in InAs NW Tunnel-FETs and in strained silicon (sSi) NW MOSFETs with identical, nominal geometrical features. Our results show that the Tunnel-FETs have a somewhat larger I_{on} variability and a much worse I_{off} variability compared to MOSFETs. Furthermore, the I_{on} variability is substantially smaller than the I_{off} variability in Tunnel FETs, whereas in MOSFETs the I_{on} and I_{off} variability is similar for a V_{DD} as low as 0.3V. Our simulations provide a physical insight to explain the much smaller sensitivity to SR of the I_{on} compared to the I_{off} observed in Tunnel-FETs.

The phonon scattering has a very modest impact on the IV characteristics of nanoscale InAs Tunnel-FETs (see Section 7.2.3), hence in this study we used coherent simulations for these devices. Phonon scattering is instead appreciable in Si NW MOSFETs [21] and it was thus included with the acoustic and optical phonon parameters taken from [33]. The sSi MOSFETs have a 2% tensile strain in the [100] direction, introduced in the simulations as explained in [18].

Surface roughness was introduced in our simulations via a geometrical description of interface fluctuations according to an exponential autocovariance function $C(\mathbf{r}) = \Delta_m^2 \exp\left(-\sqrt{2}r/L_m\right)$ [21]. Throughout this work we used a r.m.s Δ_m=0.4nm and a correlation length L_m=1.0nm. The Δ_m has been deliberately set to a value somewhat large than previously used in the literature [21, 30, 31], in order to emphasize the possible difference between Tunnel-FETs and MOSFETs. However, we expect to find the same trends also for different Δ_m values.

7.3.2 Results and discussion

Figs.7.12(a),(b) illustrate the simulated I_{DS} versus V_{GS} characteristics at V_{DS}=0.3V for Tunnel-FETs and MOSFETs and for about 40 realizations of both devices. Here, we should underline that the choice to consider realistic devices, with a relatively large cross section (W=H=5nm) and a channel length L_G=20nm, resulted in a huge computational effort to develop a statistical study based on 3-D full quantum simulations.

As it can be seen in Fig.7.12(a),(b), the work-function for the devices (i.e. Φ_M=4.15V and 4.49V respectively for Tunnel-FETs and MOSFETs) was chosen in order to have the same I_{off}=5nA/μm for both the nominal transistors, namely for the devices with no roughness. The nominal Tunnel-FET and MOSFET have an SS value of respectively 23mV/dec and 60mV/dec (averaged in the I_{off} range between 10pA/μm and 10nA/μm). Furthermore,

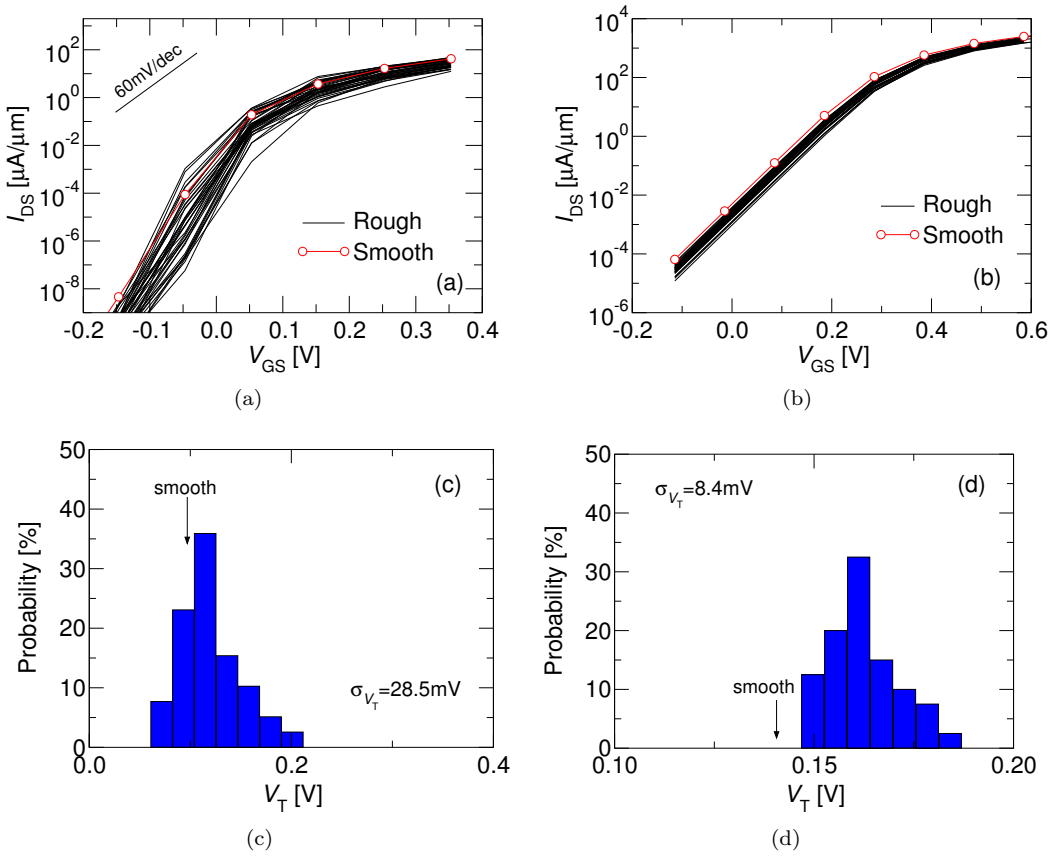


Figure 7.12: Top: I_{DS} versus V_{GS} (IV) curves for the InAs Tunnel-FET (a) and for the sSi MOSFET (b) for about 40 realizations of SR. The IV curves for the nominal devices have been reported also for comparison (smooth, open circles). SR parameters: Δ_m=0.4nm, L_m=1.0nm. Bottom: Distribution of the threshold voltage V_T for InAs Tunnel-FET (c) and sSi MOSFET (d); V_T is defined as the V_{GS} yielding I_{DS}=1μA/μm for V_{DS}=V_{DD}=0.3V.

for a small V_{DD}=0.3V the sSi MOSFETs are still in sub-threshold or weak inversion for the V_{GS}=V_{DS}=V_{DD} bias conditions that identify the I_{on}. As a result, the impact of the surface roughness on the I_{DS} is similar at all V_{GS} between 0V and 0.3V and, in particular, it is similar for I_{off} and I_{on}.

Differently from the MOSFETs, for V_{GS}=V_{DS}=V_{DD}=0.3V the InAs Tunnel-FETs do not operate in the steep subthreshold region, so that the I_{on} appears to be less sensitive to surface roughness than the I_{off}. Furthermore, Fig. 7.12(c) shows that the V_T dispersion in InAs Tunnel FETs has a standard deviation of σ_{V_T}=28.5mV, which is considerably larger than the σ_{V_T}=8.2mV extracted from Fig. 7.12(d) for MOSFETs. The σ_{V_T} for MOSFETs is reasonably consistent with the V_T spreads reported in the literature for similar Si NW MOSFETs [30, 31, 32, 34].

The different current variability for Tunnel-FETs and MOSFETs is best illustrated by Fig.7.13, reporting the I_{off} versus I_{on} for the different device realizations. As it can be seen, for the MOSFETs the spread for I_{off} and I_{on} is comparable (see the straight line corresponding to I_{on}/I_{off}=5×10^4), because, as already said, the devices operate in sub-threshold or weak inversion for V_{GS}=V_{DS}=V_{DD}=0.3V. For the Tunnel-FETs, instead, the variability is very modest for I_{on} compared to I_{off}, and, furthermore, it is much smaller than the I_{on} variability of the

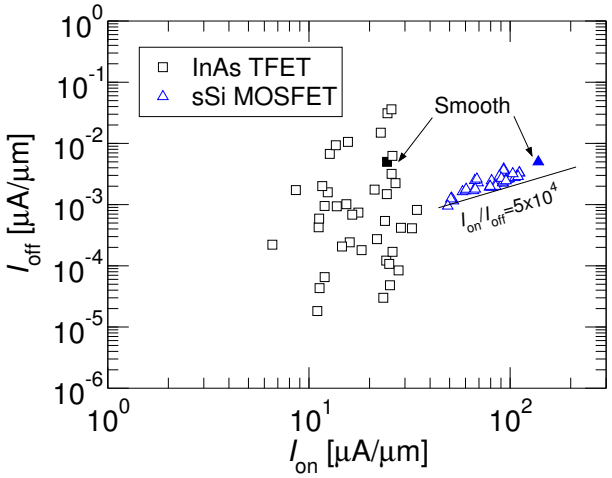
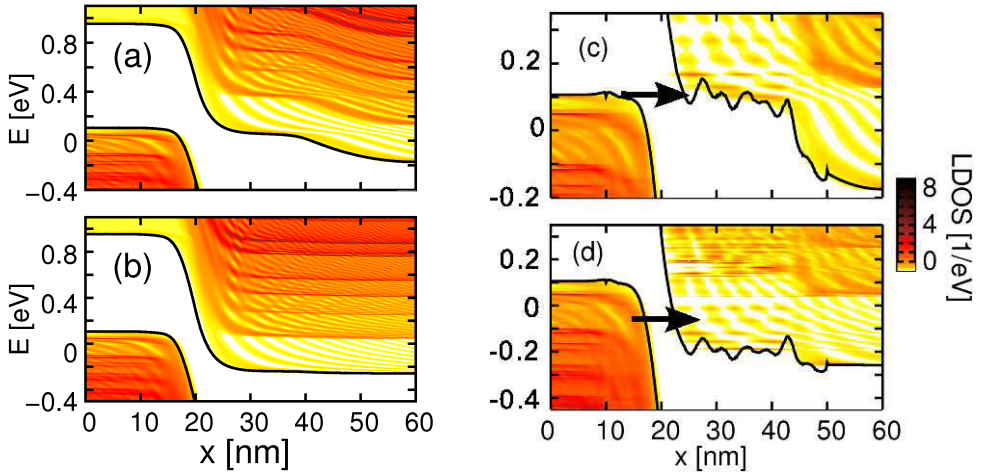


Figure 7.13: I_{off} versus I_{on} extracted from Fig.7.12 for the InAs Tunnel-FET (squares) and for the sSi MOSFET (triangles). The smooth devices values (filled symbols) have been reported for comparison.

MOSFETs.



(a) (b)

Figure 7.14: Profile for the conduction and valence subband and local density of states for the InAs Tunnel-FETs. (a),(c) correspond to the I_{off} bias condition (i.e. V_{GS}=0V and V_{DS}=V_{DD}=0.3V), while (b),(d) are for the I_{on} bias condition (i.e. V_{GS}=V_{DS}=V_{DD}=0.3V). The figure compares the results for the nominal device in (a),(b) (having I_{off}=5nA/μm and I_{on}=24.43μA/μm) with those for a device with SR in (c),(d) (having I_{off}=30pA/μm and I_{on}=23.44μA/μm). The horizontal arrows in (c),(d) indicate the tunneling path corresponding to the energy for maximum transmission.

In order to gain a physical insight about the smaller I_{on} compared to I_{off} variability in InAs Tunnel-FETs, we analyze in Fig.7.14 the conduction and valence subbands and the local density of states (LDOS) for the nominal device and for a transistor with non null surface roughness. Fig.7.14(a) and (c) refer to the I_{off} bias conditions (i.e. V_{GS}=0V and V_{DS}=V_{DD}=0.3V), whereas Fig.7.14(b) and (d) are for the I_{on} bias conditions (i.e. V_{GS}=V_{DS}=V_{DD}=0.3V). The gate elec-

trode extends from 20nm to 40nm and the interface roughness is non null from 10nm to 50nm, hence it enters for a 10nm length the source and drain regions. The inspection of the current spectrum (not shown) reveals that the current density is maximum for energies about 70meV above $E_{\rm fs}$ for $V_{\rm GS}$=0V (i.e. in Figs. 7.14(a),(c)) and for energies very close to $E_{\rm fs}$ for $V_{\rm GS}$=0.3V (i.e. in Figs. 7.14(b),(d)). This information allowed us to mark with a thick arrow the dominant tunneling path in Figs. 7.14(c) and (d).

If we now focus on Fig.7.14(a) and (c) we can observe that for $V_{\rm GS}$=0V the tunneling distance is relatively long; the dominant contribution to the current stems from energies about 70meV above the source Fermi level $E_{\rm fs}$=0 (not shown). In particular, the tunneling occurs mainly from the source to the central part of the channel, where the SR influences significantly the conduction subband and the LDOS. Consequently, the $I_{\rm off}$ for the device of Fig.7.14(c) is much smaller than the nominal value 5nA/μm.

Fig.7.14(c) and (d) show that for $V_{\rm GS}$=$V_{\rm DS}$=$V_{\rm DD}$=0.3V. The tunneling path is much shorter, as expected, and that the tunneling occurs mostly in the source depletion region and extends into the channel for only very few nanometers. Furthermore, Fig.7.14(d) indicates that surface roughness has a relatively weak influence on the subbands at the source side end of the channel and in the source depletion region. The results of Fig.7.14 explain why in Tunnel-FETs the SR induced variability is much smaller for $I_{\rm on}$ than it is for $I_{\rm on}$.

Then one may ask why the conduction subbands appear relatively smoother in the source depletion region (i.e. roughly from 15 to 22nm), than they are in the channel region. It should be noticed that here *smooth* means close to the profile of the nominal device, because the roughness induced variations with respect to nominal device are what really matters for the I_{DS} variability. At this regard, two quantities must be compared: (a) the subband fluctuations produced by the *microscopic roughness*; (b) the *macroscopic band bending* produced by built-in potentials or external biases.

The SR induced fluctuations can be inferred from the subbands profile between about 25nm and 40nm in Figs. 7.14(c),(d), that reveals an amplitude for the fluctuations of at most 100meV and over a spatial range of about 3-4nm. The macroscopic band bending in the source depletion region, instead, is illustrated by Figs. 7.14(a),(b), and it is quantified in almost 1eV for x between about 15nm and 22nm. Hence, in the source depletion region the conduction subband fluctuations induced by the microscopic roughness are much smaller than the macroscopic band bending, so that the roughness has a relatively weak influence on the subband profile.

7.4 Resume

This chapter has presented the results of a simulation study of InAs Tunnel-FETs with a constant benchmark against conventional silicon MOSFETs in terms of either performances or variability.

In particular, in Section 7.2 we have investigated the effect of strain on the performance of homo-junction n-type InAs Tunnel-FETs using a 3-D full-quantum transport simulator based on an 8-band **k·p** Hamiltonian. We have presented a systematic comparison with sSi MOSFETs and evaluated some relevant figures of merit for digital applications. Our results suggest that biaxial stress perpendicular to the transport direction can improve remarkably the $I_{\rm off}$ versus $I_{\rm on}$ tradeoff in InAs Tunnel-FETs and help open an $I_{\rm off}$ and $V_{\rm DD}$ window where Tunnel-FETs may outperform sSi MOSFETs. The identified stress condition is favorable both in terms of switching time and switching energy for LSTP and LOP digital applications.

Subsequently, Section 7.3 has presented a comparative study of surface roughness induced variability in InAs NW Tunnel-FETs and in sSi NW MOSFETs obtained by using a 3-D full-quantum simulation approach. Our results show that, for a low $V_{\rm DD}$=0.3V, the $I_{\rm on}$ and $I_{\rm off}$ variability is similar in sSi MOSFETs, whereas in Tunnel-FETs the variability is much smaller for $I_{\rm on}$ than for $I_{\rm off}$ and, furthermore, the $I_{\rm on}$ variability is somewhat larger than in

sSi MOSFETs. These results are interesting for the ultra low V_{DD} applications that may be enabled by steep-slope transistors.

Bibliography

[1] A. Seabaugh and Q. Zhang, "Low-Voltage Tunnel Transistors for Beyond CMOS Logic," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2095 –2110, dec. 2010.

[2] J. Rabaey, J. Ammer, T. Karalar, B. O. S. Li, M. Sheets, and T. Tuan, "Comprehensive understanding of surface roughness and Coulomb scattering mobility in biaxially-strained Si MOSFETs," in *IEEE International Solid-State Circuits Conference (ISSCC),Digest Technical Papers*, 2002, pp. 200–201.

[3] A. Wang and A. Chandrakasan, "A 180-mV subthreshold FFT processor using a minimum energy design methodology," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 1, pp. 310 – 319, jan. 2005.

[4] S. Hanson, M. Seok, D. Sylvester, and D. Blaauw, "Nanometer Device Scaling in Subthreshold Logic and SRAM," *IEEE Trans. Electron Devices*, vol. 55, no. 1, pp. 175 –185, jan. 2008.

[5] *International Technology Roadmap for Semiconductors: 2009 Update*, 2009.

[6] W. Y. Choi, B.-G. Park, J. D. Lee, and T.-J. K. Liu, "Tunneling Field-Effect Transistors (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec," *IEEE Electron Device Lett.*, vol. 28, no. 8, pp. 743 –745, aug. 2007.

[7] K. Boucart and A. Ionescu, "Double-Gate Tunnel FET With High-k Gate Dielectric," *IEEE Trans. Electron Devices*, vol. 54, no. 7, pp. 1725 –1733, july 2007.

[8] A. S. Verhulst, W. G. Vandenberghe, K. Maex, and G. Groeseneken, "Tunnel field-effect transistor without gate-drain overlap," *Applied Physics Letters*, vol. 91, no. 5, p. 053102, 2007.

[9] F. Mayer, C. Le Royer, J.-F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, "Impact of SOI, Si1-xGexOI and GeOI substrates on CMOS compatible Tunnel FET performance," in *IEEE IEDM Technical Digest*, dec. 2008, pp. 1 –5.

[10] T. Krishnamohan, D. Kim, S. Raghunathan, and K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With record high drive currents and < 60mV/dec subthreshold slope," in *IEEE IEDM Technical Digest*, dec. 2008, pp. 947–951.

[11] M. Luisier and G. Klimeck, "Atomistic Full-Band Design Study of InAs Band-to-Band Tunneling Field-Effect Transistors," *IEEE Electron Device Lett.*, vol. 30, no. 6, pp. 602 –604, june 2009.

[12] J. Knoch and J. Appenzeller, "Modeling of High-Performance p-Type III-V Heterojunction Tunnel FETs," *IEEE Electron Device Lett.*, vol. 31, no. 4, pp. 305 –307, april 2010.

[13] S. Koswatta, S. Koester, and W. Haensch, "On the Possibility of Obtaining MOSFET-Like Performance and Sub-60-mV/dec Swing in 1-D Broken-Gap Tunnel Transistors," *IEEE Trans. Electron Devices*, vol. 57, no. 12, pp. 3222 –3230, dec. 2010.

[14] J. Appenzeller, Y.-M. Lin, J. Knoch, and P. Avouris, "Band-to-Band Tunneling in Carbon Nanotube Field-Effect Transistors," *Phys. Rev. Lett.*, vol. 93, no. 19, p. 196805, Nov 2004.

[15] S. Poli, S. Reggiani, A. Gnudi, E. Gnani, and G. Baccarani, "Computational Study of the Ultimate Scaling Limits of CNT Tunneling Devices," *IEEE Trans. Electron Devices*, vol. 55, no. 1, pp. 313 –321, jan. 2008.

[16] S. Koswatta, M. Lundstrom, and D. Nikonov, "Performance Comparison Between p-i-n Tunneling Transistors and Conventional MOSFETs," *IEEE Trans. Electron Devices*, vol. 56, no. 3, pp. 456 –465, march 2009.

[17] F. Conzatti, M. Pala, D. Esseni, E. Bano, and L. Selmi, "A simulation study of strain induced performance enhancements in InAs nanowire Tunnel-FETs," in *IEEE IEDM Technical Digest*, accepted for publication, 2011.

[18] ——, "Strain Induced Performance Improvements in InAs Nanowire Tunnel-FETs," *submitted to IEEE Trans. Electron Devices*, 2011.

[19] T. B. Bahder, "Eight-band k·p model of strained zinc-blende crystals," *Phys. Rev. B*, vol. 41, no. 17, pp. 11 992–12 001, Jun 1990.

[20] S. Poli, M. Pala, T. Poiroux, S. Deleonibus, and G. Baccarani, "Size Dependence of Surface-Roughness-Limited Mobility in Silicon-Nanowire FETs," *IEEE Trans. Electron Devices*, vol. 55, no. 11, pp. 2968 –2976, nov. 2008.

[21] S. Poli and M. Pala, "Channel-Length Dependence of Low-Field Mobility in Silicon-Nanowire FETs," *IEEE Electron Device Lett.*, vol. 30, no. 11, pp. 1212 –1214, nov. 2009.

[22] M. Luisier and G. Klimeck, "Simulation of nanowire tunneling transistors: From the Wentzel-Kramers-Brillouin approximation to full-band phonon-assisted tunneling," *Journal of Applied Physics*, vol. 107, no. 8, p. 084507, 2010.

[23] M. V. Fischetti, "Monte Carlo simulation of transport in technologically significant semiconductors of the Diamond and Zinc-Blende structures - Part I: homogeneous transport," *IEEE Trans. Electron Devices*, vol. ED-38, p. 634, 1991.

[24] S. Koswatta, S. Koester, and W. Haensch, "1D Broken-Gap Tunnel Transistor with MOSFET-like On-currents and Sub-60mV/dec Subthreshold Swing," in *IEEE IEDM Technical Digest*, dec. 2009, pp. 909 –912.

[25] Y. Khatami and K. Banerjee, "Steep Subthreshold Slope n- and p-Type Tunnel-FET Devices for Low-Power and Energy-Efficient Digital Circuits," *IEEE Trans. Electron Devices*, vol. 56, no. 11, pp. 2752 –2761, nov. 2009.

[26] S. Mookerjea, R. Krishnan, S. Datta, and V. Narayanan, "Effective Capacitance and Drive Current for Tunnel FET (TFET) CV/I Estimation," *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 2092 –2098, sept. 2009.

[27] K. Boucart, A. Ionescu, and W. Riess, "A simulation-based study of sensitivity to parameter fluctuations of silicon Tunnel FETs," in *Proc. European Solid State Device Res. Conf.*, sept. 2010, pp. 345 –348.

[28] J. Knoch, S. Mantl, and J. Appenzeller, "Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices," *Solid State Electronics*, vol. 51, pp. 572–578, April 2007.

[29] M. Luisier, A. Schenk, and W. Fichtner, "Atomistic treatment of interface roughness in Si nanowire transistors with different channel orientations," *Applied Physics Letters*, vol. 90, no. 10, p. 102103, 2007.

[30] C. Buran, M. Pala, M. Bescond, M. Dubois, and M. Mouis, "Three-Dimensional Real-Space Simulation of Surface Roughness in Silicon Nanowire FETs," *IEEE Trans. Electron Devices*, vol. 56, no. 10, pp. 2186 –2192, oct. 2009.

[31] A. Martinez, N. Seoane, A. Brown, J. Barker, and A. Asenov, "Variability in Si Nanowire MOSFETs Due to the Combined Effect of Interface Roughness and Random Dopants: A Fully Three-Dimensional NEGF Simulation Study," *IEEE Trans. Electron Devices*, vol. 57, no. 7, pp. 1626 –1635, july 2010.

[32] A. Cresti, M. Pala, S. Poli, M. Mouis, and G. Ghibaudo, "A Comparative Study of Surface-Roughness-Induced Variability in Silicon Nanowire and Double-Gate FETs," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2274 –2281, aug. 2011.

[33] 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. Electron Devices*, vol. 55, no. 8, pp. 2086 – 2096, nov. 2008.

[34] 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. Electron Devices*, vol. 58, no. 5, pp. 1371 –1380, may 2011.

Chapter 8

Concluding remarks and future developments

DURING the course of the PhD program we have explored the potentials of the main technology boosters for advanced CMOS transistors and also for Tunnel-FETs. To do so, we developed two TCAD simulation tools based on advanced techniques to describe the materials band-structure and the transport in the devices. The former is a self-consistent **k·p**-Poisson solver for the calculation of the hole mobility in inversion layers, computed within the momentum-relaxation-time approximation. With this simulator we investigated the effect of strain and channel orientation on the hole mobility in both bulk and DG device structures. In this respect, Chapter 5 provides an insightful view on the strain induced mobility enhancement in FinFETs and guidelines for device optimization. In particular, we used our predictive models to systematically explore the impact of 3-D stress components on the mobility of both n-type and p-type FinFETs and to identify optimal stress configurations. Then, we presented a simulation study of the electron and hole mobility enhancements in biaxially strained MOS transistors and, in particular, a critical examination of a recently proposed interpretation of the experimental data, according to which the strain significantly modifies not only the r.m.s. value but also the correlation length of the surface roughness spectrum. Finally, we have explored the potentials in terms of on-current (I_{on}) for Ge and sGe n- and p-MOSFETs versus Si and sSi MOSFETs and shown that, according to our simulation model, sGe MOSFETs can outperform sSi MOSFETs in terms of I_{on}, but the engineering of the series-resistance is a crucial issue to exploit the potential advantages of Ge transistors.

We have demonstrated that the modeling of the numerical band-structure is of utmost importance in order to correctly predict the device performances. Moreover, scattering still plays an important role in nano-scale transistors, and all relevant scattering mechanisms should be taken into account in order to correctly describe the charge transport.

Recently, it has been proposed that, for future CMOS devices, the integration of III-V materials in n-type devices and SiGe alloys for p-type ones could help improve their performances due to a higher carrier mobility. Hence, a possible extension of this work would be to study the effect of strain in both III-V and SiGe alloys by means of semi-classical simulations.

During the second part of the PhD activity we developed a 3-D NEGF-Poisson solver based on a 8-band **k·p** Hamiltonian. With this model we investigated in depth the effect of strain and the variability issues in InAs Tunnel-FETs, which are very promising devices for future low-power applications.

In particular, in Chapter 7 we presented a systematic comparison of strained InAs Tunnel-FETs with sSi MOSFETs and evaluated some relevant figures of merit for digital applications. Our results suggest that biaxial stress perpendicular to the transport direction can improve remarkably the I_{off} versus I_{on} tradeoff in InAs Tunnel-FETs and help open an I_{off} and V_{DD} window where Tunnel-FETs may outperform sSi MOSFETs. In the same chapter, we studied the surface roughness induced variability effects in InAs nanowire Tunnel-FETs and in sSi nanowire MOSFETs. Our results show that in Tunnel-FETs the variability is much smaller for I_{on} than for I_{off} and, furthermore, the I_{on} variability is small compared to sSi MOSFETs.

We have shown that, with this type of quantum transport model, we can provide insight in the behavior of direct-bandgap Tunnel-FET and, moreover, give some important guidelines to the device engineers in order to identify the strain configurations which are most favorable in order to enhance the device performances.

As of today, however, experiments show poor performances for Tunnel-FETs compared to the simulations results. In particular, the inverse sub-threshold slope values obtained in the experiments are very high. This can be ascribed to different effects. The existence of traps which induce a trap-assisted-tunneling (TAT) could be a possible source of performance degradation. With the 3-D NEGF code developed during this PhD program, a simulation study devoted to the effect of the traps on the IV characteristics of Tunnel-FETs can be very interesting and it is considered a possible future development of this work.

Appendix A

Surface Roughness Matrix element for Hole Inversion Layers

The **k·p** method requires different calculations to obtain the matrix element either in the case of Bulk or SOI structures. In both cases we assume that the Hamiltonian coefficients do not depend on the quantization direction z (i.e. do not account for material changes).

We can rewrite the Schrödinger equation as:

$$\mathbf{H}\,\boldsymbol{\psi}_i = E_i \boldsymbol{\psi}_i \tag{A.1}$$

where the unperturbed Hamiltonian operator for $\mathbf{k} = 0$ is defined as:

$$\mathbf{H}(\mathbf{r}, z) = \mathbf{A}\frac{d^2}{dz^2} + \mathbf{B} + \mathbf{I}[U(z) + \phi_B\,\theta(-z)] \tag{A.2}$$

Here all matrices are 6x6 and, moreover, only **A** depends on crystal orientation and **B** includes spin-orbit and strain matrices. Furthermore, **I** is the identity matrix, $U(z)$ is the potential energy, ϕ_B is the barrier height at the interface and θ is the step function.

According to perturbation theory, the matrix element for the unscreened case is given by:

$$M_{ij}(\mathbf{k}, \mathbf{k}') = \int\int (\boldsymbol{\psi}_i^{unp})^\dagger \left[\mathbf{H}^{pert}(\mathbf{r}, z) - \mathbf{H}^{unp}(z)\right] \boldsymbol{\psi}_j^{unp} dz\, \frac{e^{i\mathbf{q}\cdot\mathbf{r}}}{A}\, d\mathbf{r} \tag{A.3}$$

where $\mathbf{q} = \mathbf{k} - \mathbf{k}'$ is the exchanged momentum and A is the normalization area. The superscripts *unp* and *pert* indicates the unperturbed and perturbed cases respectively. In the following, the superscript *unp* will be dropped only for the wave functions.

Bulk Structure

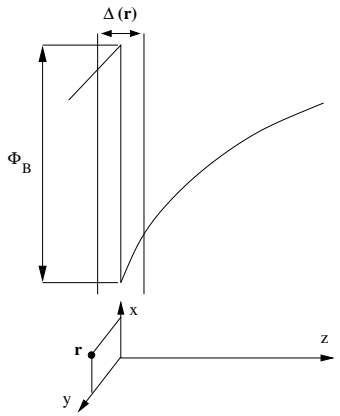


Figure A.1: Sketch of the potential energy well close to the Si-oxide interface of a MOSFET. $\Delta(\mathbf{r})$ is the possible fluctuation of the rough interface position. Φ_{ox} is the energy barrier height.

We write the perturbed Hamiltonian $\mathbf{H}^{pert}$ for a displacement $\Delta(\mathbf{r})$ (see Fig.A.1) as:

$$\mathbf{H}^{pert}(\mathbf{r}, z) = \mathbf{A}\frac{d^2}{dz^2} + \mathbf{B} + \mathbf{I}[U(z) + \phi_B\theta(-z + \Delta(\mathbf{r}))] \tag{A.4}$$

If we denote $z' = z - \Delta(\mathbf{r})$ and we expand $U(z) \simeq U(z') + \Delta(\mathbf{r})dU/dz'$, we can rewrite the perturbed Hamiltonian as a function of z':

$$\begin{aligned} \mathbf{H}^{pert}(\mathbf{r}, z) &= \mathbf{A}\frac{d^2}{dz'^2} + \mathbf{B} + \mathbf{I}\left[U(z') + \Delta(\mathbf{r})\frac{dU}{dz'} + \phi_B\theta(-z')\right] \\ &= \mathbf{H}^{unp}(z') + \mathbf{I}\,\Delta(\mathbf{r})\frac{dU}{dz} \end{aligned} \tag{A.5}$$

Before performing the entire integral, it is worth expanding the wave function as $\boldsymbol{\psi}_i(z) \simeq \boldsymbol{\psi}_i(z') + \Delta(\mathbf{r}) d\boldsymbol{\psi}_i/dz$. Projecting the difference of the perturbed and unperturbed Hamiltonians as in Eq.A.3 and neglecting higher order terms (i.e. Δ^2) we readily obtain:

$$\begin{aligned}
\langle i\,|\mathbf{H}^{pert} - \mathbf{H}^{unp}|\,j\rangle &= \int \left[\boldsymbol{\psi}_i(z') + \Delta(\mathbf{r})\frac{d\boldsymbol{\psi}_i}{dz'}\right]^{\dagger} \left[\mathbf{H}^{unp}(z') + \mathbf{I}\Delta(\mathbf{r})\frac{dU}{dz'} - \mathbf{H}^{unp}(z)\right] \\
&\quad \times \left[\boldsymbol{\psi}_j(z') + \Delta(\mathbf{r})\frac{d\boldsymbol{\psi}_j}{dz'}\right] dz' \\
&= E_j\delta_{ij} + \Delta(\mathbf{r})E_i \int \boldsymbol{\psi}_i^{\dagger}\frac{d\boldsymbol{\psi}_i}{dz'}dz' + \Delta(\mathbf{r})E_j \int \frac{d\boldsymbol{\psi}_i^{\dagger}}{dz'}\boldsymbol{\psi}_j(z')dz' \\
&\quad + \Delta(\mathbf{r}) \int \boldsymbol{\psi}_i^{\dagger}(z') \left[\mathbf{I}\frac{dU}{dz'}\right] \boldsymbol{\psi}_j(z')dz' - E_j\delta_{ij}
\end{aligned} \tag{A.6}$$

Recalling Eq.A.3 and substituting z' with z we can write the surface roughness matrix element as:

$$M_{ij}(\mathbf{q}) = \Delta(\mathbf{q}) \left[(E_i - E_j) \int \boldsymbol{\psi}_i^{\dagger}\frac{d\boldsymbol{\psi}_j}{dz}dz + \int \boldsymbol{\psi}_i^{\dagger}\left[\mathbf{I}\frac{dU}{dz}\right]\boldsymbol{\psi}_j\,dz\right] \tag{A.7}$$

which is the integral formulation similar to what derived in [1] for electrons. The term $\Delta(\mathbf{q})$ corresponds to:

$$\Delta(\mathbf{q}) = \frac{1}{A}\int \Delta(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}d\mathbf{r} \tag{A.8}$$

The squared matrix element is given by:

$$|M_{ij}|^2 = \frac{S(\mathbf{q})}{A}\left|(E_i - E_j) \int \boldsymbol{\psi}_i^{\dagger}\frac{d\boldsymbol{\psi}_j}{dz}dz + \int \boldsymbol{\psi}_i^{\dagger}\left[\mathbf{I}\frac{dU}{dz}\right]\boldsymbol{\psi}_j\,dz\right|^2 \tag{A.9}$$

where

$$S(\mathbf{q}) = \frac{1}{A}\left|\int \Delta(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}d\mathbf{r}\right|^2 \tag{A.10}$$

is the spectrum of the surface roughness at the interface.

The most common expression used in literature for the surface roughness spectrum are the Gaussian and exponential ones [2]. The Gaussian spectrum is given by:

$$S(\mathbf{q}) = \pi\Delta^2\Lambda^2 \exp\left(-\frac{\Delta^2\Lambda^2}{4}\right) \tag{A.11}$$

For all simulations of this thesis, we considered an exponential expression for the surface roughness spectrum, where $S(\mathbf{q})$ reads

$$S(\mathbf{q}) = \frac{\pi\Delta^2\Lambda^2}{\left[1 + \frac{q^2\Lambda^2}{2}\right]^{3/2}} \tag{A.12}$$

By changing the parameters Δ and Λ we reproduced well the experimental mobility measured in bulk devices and the result of the calibration procedure is summarized in Tab.4.2.

SOI Structure

As sketched in Fig.A.2, a SOI device has two interfaces and, denoting with $\Delta_F(\mathbf{r})$ and $\Delta_B(\mathbf{r})$ the possible fluctuations at the front and the back interface respectively, we can write the unperturbed and the perturbed Hamiltonians with the $\mathbf{k}\cdot\mathbf{p}$ formalism as:

$$\mathbf{H}^{unp}(z) = \mathbf{A}\frac{d^2}{dz^2} + \mathbf{B} + \mathbf{I}\left[U(z) + \Phi_B\theta(-z) + \Phi_B\theta(z - T_{SI})\right]$$

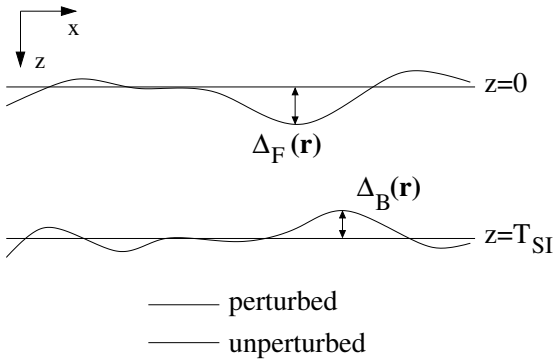


Figure A.2: Sketch of a rough interface for a SOI device. $\Delta_F(\mathbf{r})$ and $\Delta_B(\mathbf{r})$ denote, respectively, the possible fluctuations at the front and the back interface.

$$\mathbf{H}^{pert}(z) = \mathbf{A}\frac{d^2}{dz^2} + \mathbf{B} + \mathbf{I}[U(z) + \Phi_B\theta(-z + \Delta_F(\mathbf{r})) + \Phi_B\theta(z - T_{SI} - \Delta_B(\mathbf{r}))] \tag{A.13}$$

where T_{SI} is the silicon film thickness. Since we embrace the single material approximation, we cannot account for material changes (i.e. different masses in the dielectric). Defining:

$$z' = \frac{T_{SI}}{T_{SI} + \Delta_B - \Delta_F}(z - \Delta_F) \tag{A.14}$$

and proceeding similarly to what done in the bulk case, we can express the perturbed Hamiltonian as a function of the unperturbed one:

$$\mathbf{H}^{pert}(z) = \underbrace{\mathbf{A}\frac{d^2}{dz'^2} + \mathbf{B} + \mathbf{I}\left[U(z') + \Phi_B\theta(-z') + \Phi_B\theta(z' - T_{SI})\right]}_{H^{unp}(z')} - 2\frac{\Delta_B - \Delta_F}{T_{SI}}\mathbf{A}\frac{d^2}{dz'^2} + \mathbf{I}\frac{dU}{dz'}\left(\Delta_F + \frac{\Delta_B - \Delta_F}{T_{SI}}z'\right) \tag{A.15}$$

where we have dropped the dependence on the position $\mathbf{r}$. The unperturbed wave function will be expanded as:

$$\boldsymbol{\psi}_i(z) \simeq \boldsymbol{\psi}_i(z') + \frac{d\boldsymbol{\psi}_i}{dz'}\left(\Delta_F + \frac{\Delta_B - \Delta_F}{T_{SI}}z'\right) \tag{A.16}$$

Now we can use Eq.A.3 and project the difference of the Hamiltonians obtaining:

$$\begin{aligned}\langle i\,|\mathbf{H}^{pert} - \mathbf{H}^{unp}|\,j\rangle = \int &\left[\boldsymbol{\psi}_i + \frac{d\boldsymbol{\psi}_i}{dz'}\left(\Delta_F + \frac{\Delta_B - \Delta_F}{T_{SI}}z'\right)\right]^\dagger \\ \times &\left[\mathbf{H}^{unp}(z') - 2\frac{\Delta_B - \Delta_F}{T_{SI}}\mathbf{A}\frac{d^2}{dz'^2} + \mathbf{I}\frac{dU}{dz'}\left(\Delta_F + \frac{\Delta_B - \Delta_F}{T_{SI}}z'\right)\right] \\ \times &\left[\boldsymbol{\psi}_j + \frac{d\boldsymbol{\psi}_j}{dz'}\left(\Delta_F + \frac{\Delta_B - \Delta_F}{T_{SI}}z'\right)\right]\left(1 + \frac{\Delta_B - \Delta_F}{T_{SI}}\right)dz' - E_j\delta_{ij}\end{aligned} \tag{A.17}$$

Neglecting higher order terms we can finally write the matrix element as:

$$\begin{aligned} M_{ij}(\mathbf{q}) &= \Delta_F(\mathbf{q})\left[(E_i - E_j)\int \boldsymbol{\psi}_i^\dagger \frac{d\boldsymbol{\psi}_j}{dz}dz + \int \boldsymbol{\psi}_i^\dagger \left(\mathbf{I}\frac{dU}{dz}\right)\boldsymbol{\psi}_j dz\right] \\ &+\frac{\Delta_B(\mathbf{q})-\Delta_F(\mathbf{q})}{T_{SI}}(E_i - E_j)\int \boldsymbol{\psi}_i^\dagger z\frac{d\boldsymbol{\psi}_j}{dz}dz \\ &+\frac{\Delta_B(\mathbf{q})-\Delta_F(\mathbf{q})}{T_{SI}}\left[\int \boldsymbol{\psi}_i^\dagger \left(\mathbf{I}\frac{dU}{dz}\right) z\boldsymbol{\psi}_j dz - 2\int \boldsymbol{\psi}_i\dagger\, \mathbf{A}\frac{d^2}{dz^2}\boldsymbol{\psi}_j dz\right] \end{aligned} \tag{A.18}$$

If the spectra of the two interfaces are uncorrelated, the squared matrix element is given by:

$$\begin{aligned} |M_{ij}|^2 &= \frac{S_F(\mathbf{q})}{A}\left|(E_i - E_j)\int \boldsymbol{\psi}_i^\dagger \frac{d\boldsymbol{\psi}_j}{dz}dz + \int \boldsymbol{\psi}_i^\dagger \left(\mathbf{I}\frac{dU}{dz}\right)\boldsymbol{\psi}_j dz\right. \\ &- \frac{1}{T_{SI}}\left[(E_i - E_j)\int \boldsymbol{\psi}_i^\dagger z\frac{d\boldsymbol{\psi}_j}{dz}dz\right. \\ &+ \left.\left.\int \boldsymbol{\psi}_i^\dagger \left(\mathbf{I}\frac{dU}{dz}\right) z\boldsymbol{\psi}_j dz - 2\int \boldsymbol{\psi}_i^\dagger\, \mathbf{A}\frac{d^2}{dz^2}\boldsymbol{\psi}_j dz\right]\right|^2 \\ &+ \frac{S_B(\mathbf{q})}{A\cdot T_{SI}^2}\left|(E_i - E_j)\int \boldsymbol{\psi}_i^\dagger z\frac{d\boldsymbol{\psi}_j}{dz}dz + \int \boldsymbol{\psi}_i^\dagger \left(\mathbf{I}\frac{dU}{dz}\right) z\boldsymbol{\psi}_j dz\right. \\ &- \left.2\int \boldsymbol{\psi}_i^\dagger\, \mathbf{A}\frac{d^2}{dz^2}\boldsymbol{\psi}_j dz\right|^2 \end{aligned} \tag{A.19}$$

where

$$S_F(\mathbf{q}) = \frac{1}{A}\left|\int \Delta_F(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}d\mathbf{r}\right|^2 \tag{A.20}$$

$$S_B(\mathbf{q}) = \frac{1}{A}\left|\int \Delta_B(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}}d\mathbf{r}\right|^2 \tag{A.21}$$

are the spectra of the surface roughness at the two interfaces.

Differential formulation

Bulk Case

In order to obtain a differential formulation for the bulk case, we can start from the unperturbed Schrödinger equation:

$$\mathbf{A}\frac{d^2\boldsymbol{\psi}_i}{dz^2} + \left[\mathbf{B} + \mathbf{I}U + \Phi_B\theta(-z)\right]\boldsymbol{\psi}_i = E_i\boldsymbol{\psi}_i \tag{A.22}$$

If Eq.A.22 holds thus:

$$\frac{d^2\boldsymbol{\psi}_j^\dagger}{dz^2}\mathbf{A} + \boldsymbol{\psi}_j^\dagger\left[\mathbf{B} + \mathbf{I}U + \Phi_B\theta(-z)\right] = E_j\boldsymbol{\psi}_j^\dagger \tag{A.23}$$

where $\mathbf{A}$ is real and $\mathbf{B}$ is hermitian. Multiplying Eq.A.22 on the left by $d\boldsymbol{\psi}_j^\dagger/dz$ and Eq.A.23 on the right by $d\boldsymbol{\psi}_i/dz$ we obtain:

$$\begin{cases} \frac{d\boldsymbol{\psi}_j^\dagger}{dz}\left(\mathbf{A}\frac{d^2\boldsymbol{\psi}_i}{dz^2}\right) + \frac{d\boldsymbol{\psi}_j^\dagger}{dz}\left[\mathbf{B} + \mathbf{I}U + \Phi_B\theta(-z)\right]\boldsymbol{\psi}_i = E_i\frac{d\boldsymbol{\psi}_j^\dagger}{dz}\boldsymbol{\psi}_i \\ \left(\frac{d^2\boldsymbol{\psi}_j^\dagger}{dz^2}\mathbf{A}\right)\frac{d\boldsymbol{\psi}_i}{dz} + \boldsymbol{\psi}_j^\dagger\left[\mathbf{B} + \mathbf{I}U + \Phi_B\theta(-z)\right]\frac{d\boldsymbol{\psi}_i}{dz} = E_j\boldsymbol{\psi}_j^\dagger\frac{d\boldsymbol{\psi}_i}{dz} \end{cases} \tag{A.24}$$

Summing on both sides and intergrating from 0 to $+\infty$ after some manipulations it follows that:

$$\left.\frac{d\boldsymbol{\psi}_j^\dagger}{dz}\right|_0 \mathbf{A} \left.\frac{d\boldsymbol{\psi}_i}{dz}\right|_0 = (E_i - E_j)\int_0^{+\infty} \frac{d\boldsymbol{\psi}_j^\dagger}{dz}\boldsymbol{\psi}_i dz + \int_0^{+\infty} \boldsymbol{\psi}_j^\dagger \mathbf{I}\frac{dU}{dz}\boldsymbol{\psi}_i dz \tag{A.25}$$

where we have imposed closed boundaries (i.e. $\boldsymbol{\psi}(z=0)=0$) and supposed that $d\boldsymbol{\psi}/dz = 0$ for $z \to +\infty$. The RHS of Eq.A.25 is equal to the matrix element of Eq.A.7 hence, for bulk case:

$$|M_{ij}|^2 = \frac{S(\mathbf{q})}{A}\left|\frac{d\boldsymbol{\psi}_j^\dagger}{dz}(0)\mathbf{A}\frac{d\boldsymbol{\psi}_i}{dz}(0)\right|^2 \tag{A.26}$$

which is consistent with [3].

SOI Case

Considering the Schrödinger equation for a hole inversion layer (with the same Hamiltonian as in Eq.A.2) and its conjugate transpose written as:

$$\mathbf{A}\frac{d^2\boldsymbol{\psi}_i}{dz^2} + (\mathbf{B} + \mathbf{I}U)\boldsymbol{\psi}_i = E_i\boldsymbol{\psi}_i \tag{A.27}$$

$$\frac{d^2\boldsymbol{\psi}_j\dagger}{dz^2}\mathbf{A} + \boldsymbol{\psi}_j^\dagger(\mathbf{B} + \mathbf{I}U) = E_j\boldsymbol{\psi}_j^\dagger \tag{A.28}$$

Multiplying by $z d\boldsymbol{\psi}_j^\dagger/dz$ and by $z d\boldsymbol{\psi}_i/dz$ the first and the second equation, respectively, integrating from 0 to T_{SI} both and summing we obtain:

$$\begin{aligned} &T_{SI}\left.\frac{d\boldsymbol{\psi}_j^\dagger}{dz}\right|_{T_{SI}} \mathbf{A}\left.\frac{d\boldsymbol{\psi}_i}{dz}\right|_{T_{SI}} - \int_0^{T_{SI}} \frac{d\boldsymbol{\psi}_j^\dagger}{dz}\mathbf{A}\frac{d\boldsymbol{\psi}_i}{dz}dz - \int_0^{T_{SI}} \boldsymbol{\psi}_j^\dagger(\mathbf{B}+\mathbf{I}U)\boldsymbol{\psi}_i dz \\ &- \int_0^{T_{SI}} \boldsymbol{\psi}_j^\dagger z\mathbf{I}\frac{dU}{dz}\boldsymbol{\psi}_i dz = (E_i - E_j)\int_0^{T_{SI}} \frac{d\boldsymbol{\psi}_j^\dagger}{dz} z\boldsymbol{\psi}_i dz - E_j\int_0^{T_{SI}} \boldsymbol{\psi}_j^\dagger\boldsymbol{\psi}_i dz \end{aligned} \tag{A.29}$$

After some manipulations we obtain:

$$\begin{aligned} T_{SI}\left.\frac{d\boldsymbol{\psi}_j^\dagger}{dz}\right|_{T_{SI}} \mathbf{A}\left.\frac{d\boldsymbol{\psi}_i}{dz}\right|_{T_{SI}} &= \int_0^{T_{SI}} \frac{d\boldsymbol{\psi}_j^\dagger}{dz}\mathbf{A}\frac{d\boldsymbol{\psi}_i}{dz}dz + (E_i - E_j)\int_0^{T_{SI}} \frac{d\boldsymbol{\psi}_j^\dagger}{dz} z\boldsymbol{\psi}_i dz \\ &- 2\int_0^{T_{SI}} \frac{d^2\boldsymbol{\psi}_j^\dagger}{dz^2}\mathbf{A}\boldsymbol{\psi}_i dz \end{aligned} \tag{A.30}$$

Now we can recall Eq.A.25 (remembering that the derivative of the wave functions does not vanish for $z = T_{SI}$) together with Eq.A.30 in order to re-write the RHS of Eq.A.18: by doing so we find the differential formulation for the SOI case:

$$|M_{ij}|^2 = \frac{S_F(\mathbf{q})}{A}\left|\frac{d\boldsymbol{\psi}_j^\dagger}{dz}(0)\mathbf{A}\frac{d\boldsymbol{\psi}_i}{dz}(0)\right|^2 + \frac{S_R(\mathbf{q})}{A}\left|\frac{d\boldsymbol{\psi}_j^\dagger}{dz}(T_{SI})\mathbf{A}\frac{d\boldsymbol{\psi}_i}{dz}(T_{SI})\right|^2 \tag{A.31}$$

where $S_F(\mathbf{q})$ and $S_R(\mathbf{q})$ are the spectra of the surface roughness at the front and at the back interface, respectively.

Bibliography

[1] D. Esseni, "On the Modeling of Surface Roughness Limited Mobility in SOI MOSFETs and its Correlation to the Transistor Effective Field ," *IEEE Trans. Electron Devices*, vol. 3, no. 3, pp. 394–401, 2004.

[2] S.M.Goodnick, D.K.Ferry, C.W.Wilmsen, Z.Liliental, D.Fathy, and O.L.Krivanek, "Surface roughness at the Si(100)-SiO_2 interface," *Phys. Rev. B*, vol. 32, pp. 8171–8186, 1985.

[3] M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, "Six-band kp calculation of the hole mobility in silicon inversion layers: Dependence on surface orientation, strain, and silicon thickness," *Journal of Applied Physics*, vol. 94, no. 2, pp. 1079–1095, 2003.

Appendix B

Discretized 8-bands k·p Hamiltonians

B.1 Real Space

We can write $\mathbf{H}_{R,yz}$ of Eq.6.70 as a block matrix composed of 64 terms, each with a rank of $N_y \times N_z$. The non-zero entries are:

$$
\begin{aligned}
&\mathbf{H}_{11} = E_g\,\mathbf{I}_{yz} + \tfrac{\hbar^2}{2m_0}\,\gamma_c\left(\tfrac{2}{\Delta_x^2}\,\mathbf{I}_{yz} - \mathbf{D}_{yy} - \mathbf{D}_{zz}\right) \quad ; \quad \mathbf{H}_{22} = \mathbf{H}_{11}\\
&\mathbf{H}_{13} = P_0\,\mathbf{D}_y/\sqrt{6} \quad ; \quad \mathbf{H}_{31} = -P_0\,\mathbf{D}_y/\sqrt{6}\\
&\mathbf{H}_{15} = -P_0\,\mathbf{D}_y/\sqrt{2} \quad ; \quad \mathbf{H}_{51} = P_0\,\mathbf{D}_y/\sqrt{2}\\
&\mathbf{H}_{16} = i\sqrt{\tfrac{2}{3}}\,P_0\,\mathbf{D}_z \quad ; \quad \mathbf{H}_{61} = \mathbf{H}_{16}\\
&\mathbf{H}_{17} = iP_0\,\mathbf{D}_z/\sqrt{3} \quad ; \quad \mathbf{H}_{71} = \mathbf{H}_{17}\\
&\mathbf{H}_{18} = P_0\,\mathbf{D}_y/\sqrt{3} \quad ; \quad \mathbf{H}_{81} = -P_0\,\mathbf{D}_y/\sqrt{3}\\
&\mathbf{H}_{23} = \mathbf{H}_{16} \quad ; \quad \mathbf{H}_{32} = \mathbf{H}_{61}\\
&\mathbf{H}_{24} = -P_0\,\mathbf{D}_y/\sqrt{2} \quad ; \quad \mathbf{H}_{42} = P_0\,\mathbf{D}_y/\sqrt{2}\\
&\mathbf{H}_{26} = P_0\,\mathbf{D}_y/\sqrt{6} \quad ; \quad \mathbf{H}_{62} = -P_0\,\mathbf{D}_y/\sqrt{6}\\
&\mathbf{H}_{27} = -P_0\,\mathbf{D}_y/\sqrt{3} \quad ; \quad \mathbf{H}_{27} = P_0\,\mathbf{D}_y/\sqrt{3}\\
&\mathbf{H}_{28} = -\mathbf{H}_{17} \quad ; \quad \mathbf{H}_{82} = -\mathbf{H}_{71}\\
&\mathbf{H}_{33} = \tfrac{\hbar^2}{2m_0}\left[(\gamma_2-\gamma_1)\tfrac{2}{\Delta_x^2}\,\mathbf{I}_{yz} - (\gamma_2-\gamma_1)\,\mathbf{D}_{yy} + (\gamma_1+2\gamma_2)\,\mathbf{D}_{zz}\right] \quad ; \quad \mathbf{H}_{66} = \mathbf{H}_{33}\\
&\mathbf{H}_{34} = i\tfrac{\hbar^2}{2m_0}\,\gamma_3\,2\sqrt{3}\,\mathbf{D}_{yz} \quad ; \quad \mathbf{H}_{43} = -i\tfrac{\hbar^2}{2m_0}\,\gamma_3\,2\sqrt{3}\,\mathbf{D}_{yz}\\
&\mathbf{H}_{35} = -\tfrac{\hbar^2}{2m_0}\,\sqrt{3}\left(\gamma_2\tfrac{2}{\Delta_x^2}\,\mathbf{I}_{yz} + \gamma_2\,\mathbf{D}_{yy}\right) \quad ; \quad \mathbf{H}_{53} = \mathbf{H}_{35}\\
&\mathbf{H}_{37} = i\tfrac{\hbar^2}{2m_0}\,\gamma_3\,3\sqrt{2}\,\mathbf{D}_{yz} \quad ; \quad \mathbf{H}_{73} = -i\tfrac{\hbar^2}{2m_0}\,\gamma_3\,3\sqrt{2}\,\mathbf{D}_{yz}\\
&\mathbf{H}_{38} = -\tfrac{\hbar^2}{2m_0}\,\gamma_2\,\sqrt{2}\left(\tfrac{2}{\Delta_x^2}\,\mathbf{I}_{yz} - \mathbf{D}_{yy} + 2\mathbf{D}_{zz}\right) \quad ; \quad \mathbf{H}_{83} = \mathbf{H}_{38}\\
&\mathbf{H}_{44} = -\tfrac{\hbar^2}{2m_0}\left[(\gamma_1+\gamma_2)\tfrac{2}{\Delta_x^2}\,\mathbf{I}_{yz} - (\gamma_1+\gamma_2)\mathbf{D}_{yy} - (\gamma_1-2\gamma_2)\mathbf{D}_{zz}\right] \quad ; \quad \mathbf{H}_{55} = \mathbf{H}_{44}\\
&\mathbf{H}_{46} = \mathbf{H}_{35} \quad ; \quad \mathbf{H}_{64} = \mathbf{H}_{53}\\
&\mathbf{H}_{47} = -\sqrt{2}\mathbf{H}_{46} \quad ; \quad \mathbf{H}_{74} = -\sqrt{2}\mathbf{H}_{64}\\
&\mathbf{H}_{48} = \mathbf{H}_{37}/\sqrt{3} \quad ; \quad \mathbf{H}_{84} = \mathbf{H}_{73}/\sqrt{3}\\
&\mathbf{H}_{56} = -\mathbf{H}_{34} \quad ; \quad \mathbf{H}_{65} = -\mathbf{H}_{43}\\
&\mathbf{H}_{57} = \mathbf{H}_{56}/\sqrt{2} \quad ; \quad \mathbf{H}_{75} = \mathbf{H}_{65}/\sqrt{2}\\
&\mathbf{H}_{58} = -\mathbf{H}_{74} \quad ; \quad \mathbf{H}_{85} = -\mathbf{H}_{47}\\
&\mathbf{H}_{67} = -\mathbf{H}_{38} \quad ; \quad \mathbf{H}_{76} = -\mathbf{H}_{83}\\
&\mathbf{H}_{68} = \sqrt{3}\mathbf{H}_{57} \quad ; \quad \mathbf{H}_{86} = \sqrt{3}\mathbf{H}_{75}\\
&\mathbf{H}_{77} = -\tfrac{\hbar^2}{2m_0}\,\gamma_1\left(\tfrac{2}{\Delta_x^2}\,\mathbf{I}_{yz} - \mathbf{D}_{yy} - \mathbf{D}_{zz}\right) - \Delta_{so}\,\mathbf{I}_{yz} \quad ; \quad \mathbf{H}_{88} = \mathbf{H}_{77}
\end{aligned}
\tag{B.1}
$$

where $\mathbf{I}_{yz}$ is the identity matrix of rank N_yN_z. By numbering the grid point with the lexicographic ordering, the D_{ij} are the 2-D differentiation matrices calculated as

$$
\begin{aligned}
\mathbf{D}_y &= \mathbf{D}_y^{1D} \otimes \mathbf{I}_z\\
\mathbf{D}_{yy} &= \mathbf{D}_{yy}^{1D} \otimes \mathbf{I}_z\\
\mathbf{D}_z &= \mathbf{I}_y \otimes \mathbf{D}_z^{1D}\\
\mathbf{D}_{zz} &= \mathbf{I}_y \otimes \mathbf{D}_{zz}^{1D}\\
\mathbf{D}_{yz} &= \mathbf{D}_y^{1D} \otimes \mathbf{D}_z^{1D}
\end{aligned}
\tag{B.2}
$$

where $\mathbf{D}_i^{1D}$ (i=y,z) is the differentiation matrix in 1-D and $\otimes$ denotes the Kronecker tensor product [1]. If we use a the Pseudo-spectral methods discretization scheme (see Sect.2.3), $\mathbf{D_i}^{1D}$

is a $(N_i+1)\times(N_i+1)$ matrix and reads [1]

$$\begin{array}{ll} D_{i\,00}^{1D} = \frac{2N_i^2+1}{6}, & D_{i\,N_iN_i}^{1D} = -\frac{2N_i^2+1}{6}, \\ D_{i\,jj}^{1D} = \frac{-x_j}{2(1-x_j^2)}, & j=1,\ldots,N_i-1 \\ D_{i\,ij}^{1D} = \frac{c_i}{c_j}\frac{(-1)^{i+j}}{(x_i-x_j)}, & i\neq j \qquad i,j=0,\ldots N_i, \end{array} \tag{B.3}$$

where x_j are the Chebyshev extremal nodes defined in Eq.2.28 and

$$c_i = \begin{cases} 2, & i=0 \;\; or \;\; N_i, \\ 1, & otherwise \end{cases} \tag{B.4}$$

The matrix $\mathbf{D_i}^{1D}$ in Eq.B.3 is indexed from 0 to N_i.

The non-zero entries of the $\mathbf{W}_R$ matrix read:

$$\begin{array}{l}
\mathbf{W}_{11} = \frac{\hbar^2}{2m_0}\gamma_c\left(-\frac{1}{\Delta_x^2}\mathbf{I}_{yz}\right) \quad ; \quad \mathbf{W}_{22}=\mathbf{W}_{11} \\
\mathbf{W}_{13} = -i\,P_0\,\frac{1}{2\Delta_x}\mathbf{I}_{yz}/\sqrt{6} \quad ; \quad \mathbf{W}_{31}=\mathbf{W}_{13} \\
\mathbf{W}_{15} = -i\,P_0\,\frac{1}{2\Delta_x}\mathbf{I}_{yz}/\sqrt{2} \quad ; \quad \mathbf{W}_{51}=\mathbf{W}_{15} \\
\mathbf{W}_{18} = -i\,P_0\,\frac{1}{2\Delta_x}\mathbf{I}_{yz}/\sqrt{3} \quad ; \quad \mathbf{W}_{81}=\mathbf{W}_{18} \\
\mathbf{W}_{24} = i\,P_0\,\frac{1}{2\Delta_x}\mathbf{I}_{yz}/\sqrt{2} \quad ; \quad \mathbf{W}_{42}=\mathbf{W}_{24} \\
\mathbf{W}_{26} = i\,P_0\,\frac{1}{2\Delta_x}\mathbf{I}_{yz}/\sqrt{6} \quad ; \quad \mathbf{W}_{62}=\mathbf{W}_{26} \\
\mathbf{W}_{27} = -i\,P_0\,\frac{1}{2\Delta_x}\mathbf{I}_{yz}/\sqrt{3} \quad ; \quad \mathbf{W}_{72}=\mathbf{W}_{27} \\
\mathbf{W}_{33} = \frac{\hbar^2}{2m_0}\left[-(\gamma_2-\gamma_1)\frac{1}{\Delta_x^2}\mathbf{I}_{yz}\right] \quad ; \quad \mathbf{W}_{66}=\mathbf{W}_{33} \\
\mathbf{W}_{34} = i\frac{\hbar^2}{2m_0}\gamma_3\,2\sqrt{3}\,(-i\,\frac{1}{2\Delta_x}\mathbf{D}_z) \quad ; \quad \mathbf{W}_{43}=\mathbf{W}_{34} \\
\mathbf{W}_{35} = -\frac{\hbar^2}{2m_0}\sqrt{3}\left(-\gamma_2\frac{1}{\Delta_x^2}\mathbf{I}_{yz}+2i\gamma_3\,\frac{1}{2\Delta_x}\mathbf{D}_y\right) \\
\mathbf{W}_{53} = -\frac{\hbar^2}{2m_0}\sqrt{3}\left(-\gamma_2\frac{1}{\Delta_x^2}\mathbf{I}_{yz}-2i\gamma_3\,\frac{1}{2\Delta_x}\mathbf{D}_y\right) \\
\mathbf{W}_{37} = -\frac{\hbar^2}{2m_0}\gamma_3\,3\sqrt{2}\,\frac{1}{2\Delta_x}\mathbf{D}_z \quad ; \quad \mathbf{W}_{73}=\mathbf{W}_{37} \\
\mathbf{W}_{38} = \frac{\hbar^2}{2m_0}\gamma_2\sqrt{2}\left(\frac{1}{\Delta_x^2}\mathbf{I}_{yz}\right) \quad ; \quad \mathbf{W}_{83}=\mathbf{W}_{38} \\
\mathbf{W}_{44} = -\frac{\hbar^2}{2m_0}\left[-(\gamma_1+\gamma_2)\frac{1}{\Delta_x^2}\mathbf{I}_{yz}\right] \quad ; \quad \mathbf{W}_{55}=\mathbf{W}_{44} \\
\mathbf{W}_{46} = \mathbf{W}_{35} \quad ; \quad \mathbf{W}_{64}=\mathbf{W}_{53} \\
\mathbf{W}_{47} = -\sqrt{2}\mathbf{W}_{46} \quad ; \quad \mathbf{W}_{74}=-\sqrt{2}\mathbf{W}_{64} \\
\mathbf{W}_{48} = \mathbf{W}_{37}/\sqrt{3} \quad ; \quad \mathbf{W}_{84}=\mathbf{W}_{73}/\sqrt{3} \\
\mathbf{W}_{56} = -\mathbf{W}_{34} \quad ; \quad \mathbf{W}_{65}=-\mathbf{W}_{43} \\
\mathbf{W}_{57} = \mathbf{W}_{56}/\sqrt{2} \quad ; \quad \mathbf{W}_{75}=\mathbf{W}_{65}/\sqrt{2} \\
\mathbf{W}_{58} = -\mathbf{W}_{74} \quad ; \quad \mathbf{W}_{85}=-\mathbf{W}_{47} \\
\mathbf{W}_{67} = -\mathbf{W}_{38} \quad ; \quad \mathbf{W}_{76}=-\mathbf{W}_{83} \\
\mathbf{W}_{68} = \sqrt{3}\mathbf{W}_{57} \quad ; \quad \mathbf{W}_{86}=\sqrt{3}\mathbf{W}_{75} \\
\mathbf{W}_{77} = -\frac{\hbar^2}{2m_0}\gamma_1\left(-\frac{1}{\Delta_x^2}\mathbf{I}_{yz}\right) \quad ; \quad \mathbf{W}_{88}=\mathbf{W}_{77}
\end{array} \tag{B.5}$$

B.2 K-space

The 8-bands $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian can be expanded as:

$$\mathbf{H} = \mathbf{H}_1\,\hat{k}_y^2 + \mathbf{H}_2\,\hat{k}_z^2 + \mathbf{H}_3\,\hat{k}_y + \mathbf{H}_4\,\hat{k}_z + \mathbf{H}_5\,\hat{k}_y\hat{k}_z + \mathbf{H}_6 \tag{B.6}$$

where each $\mathbf{H}_i$ is a 8×8 constant matrix:

$$\mathbf{H}_1 = \begin{pmatrix} C_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & C_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & -P_1+Q_1 & 0 & R_1 & 0 & 0 & -\sqrt{2}Q_1 \\ & & & -P_1-Q_1 & 0 & R_1 & -\sqrt{2}R_1 & 0 \\ & & & & -P_1-Q_1 & 0 & 0 & \sqrt{2}R_1^* \\ & & & & & -P_1+Q_1 & \sqrt{2}Q_1 & 0 \\ & & & & & & -P_1 & 0 \\ & & & & & & & -P_1 \end{pmatrix} \tag{B.7}$$

$$\mathbf{H}_2 = \begin{pmatrix} C_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & C_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & -P_2+Q_2 & 0 & 0 & 0 & 0 & -\sqrt{2}Q_2 \\ & & & -P_2-Q_2 & 0 & 0 & 0 & 0 \\ & & & & -P_2-Q_2 & 0 & 0 & 0 \\ & & & & & -P_2+Q_2 & \sqrt{2}Q_2 & 0 \\ & & & & & & -P_2 & 0 \\ & & & & & & & -P_2 \end{pmatrix} \tag{B.8}$$

$$\mathbf{H}_3 = \begin{pmatrix} 0 & 0 & P_3^+/\sqrt{6} & 0 & P_3^-/\sqrt{2} & 0 & 0 & P_3^+/\sqrt{3} \\ & 0 & 0 & -P_3^+/\sqrt{2} & 0 & -P_3^-/\sqrt{6} & P_3^-/\sqrt{3} & 0 \\ & & 0 & 0 & R_3 & 0 & 0 & 0 \\ & & & 0 & 0 & R_3 & -\sqrt{2}R_3 & 0 \\ & & & & 0 & 0 & 0 & \sqrt{2}R_3^* \\ & & & & & 0 & 0 & 0 \\ & & & & & & 0 & 0 \\ & & & & & & & 0 \end{pmatrix} \tag{B.9}$$

$$\mathbf{H}_4 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -\sqrt{\frac{2}{3}}P_0 & -P_0/\sqrt{3} & 0 \\ & 0 & -\sqrt{\frac{2}{3}}P_0 & 0 & 0 & 0 & 0 & P_0/\sqrt{3} \\ & & 0 & -S_4^* & 0 & 0 & \sqrt{\frac{3}{2}}S_4 & 0 \\ & & & 0 & 0 & 0 & 0 & S_4/\sqrt{2} \\ & & & & 0 & S_4^* & S_4^*/\sqrt{2} & 0 \\ & & & & & 0 & 0 & \sqrt{\frac{3}{2}}S_4^* \\ & & & & & & 0 & 0 \\ & & & & & & & 0 \end{pmatrix} \tag{B.10}$$

$$\mathbf{H}_5 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 0 & -S_5^* & 0 & 0 & \sqrt{\frac{3}{2}}S_5 & 0 \\ & & & 0 & 0 & 0 & 0 & S_5/\sqrt{2} \\ & & & & 0 & S_5^* & S_5^*/\sqrt{2} & 0 \\ & & & & & 0 & 0 & \sqrt{\frac{3}{2}}S_5^* \\ & & & & & & 0 & 0 \\ & & & & & & & 0 \end{pmatrix} \tag{B.11}$$

$$\mathbf{H}_6 = \begin{pmatrix} C_6 & 0 & P_6^+/\sqrt{6} & 0 & P_6^-/\sqrt{2} & 0 & 0 & P_6^+/\sqrt{3} \\ & C_6 & 0 & -P_6^+/\sqrt{2} & 0 & -P_6^-/\sqrt{6} & P_6^-/\sqrt{3} & 0 \\ & & -P_6+Q_6 & 0 & R_6 & 0 & 0 & -\sqrt{2}Q_6 \\ & & & -P_6-Q_6 & 0 & R_6 & -\sqrt{2}R_6 & 0 \\ & & & & -P_6-Q_6 & 0 & 0 & \sqrt{2}R_6^* \\ & & & & & -P_6+Q_6 & \sqrt{2}Q_6 & 0 \\ & & & & & & -P_6-\Delta_{so} & 0 \\ & & & & & & & -P_6-\Delta_{so} \end{pmatrix} \tag{B.12}$$

where

$$\begin{aligned}
&C_1 = C_2 = \tfrac{\hbar^2}{2m_0}\,\gamma_c \\
&C_6 = E_g + \tfrac{\hbar^2}{2m_0}\,\gamma_c\,(\tfrac{2}{\Delta_x^2}) \\
&P_1 = P_2 = \tfrac{\hbar^2}{2m_0}\,\gamma_1 \\
&P_6 = \tfrac{\hbar^2}{2m_0}\,\gamma_1\,(\tfrac{2}{\Delta_x^2}) \\
&Q_1 = \tfrac{\hbar^2}{2m_0}\,\gamma_2 \\
&Q_2 = \tfrac{\hbar^2}{2m_0}\,\gamma_2\,(-2) \\
&Q_6 = \tfrac{\hbar^2}{2m_0}\,\gamma_2\,(\tfrac{2}{\Delta_x^2}) \\
&R_1 = -\tfrac{\hbar^2}{2m_0}\sqrt{3}\,(-\gamma_2) \\
&R_3 = 0 \\
&R_6 = -\tfrac{\hbar^2}{2m_0}\sqrt{3}\,(\gamma_2\,(\tfrac{2}{\Delta_x^2})) \\
&S_4 = 0 \\
&S_5 = \tfrac{\hbar^2}{2m_0}\,\gamma_3\,2\sqrt{3}\,(-i) \\
&P_3^{\pm} = \pm i\,P_0 \\
&P_6^{+} = P_6^{-} = 0
\end{aligned} \tag{B.13}$$

The differential parts of the Hamiltonian are obtained according to the standard prescriptions of quantum mechanics:

$$\hat{k}_y \to -\imath\frac{\partial}{\partial y} \qquad \hat{k}_z \to -\imath\frac{\partial}{\partial z} \tag{B.14}$$

If we write the generic 2-D base function u_{km} as [2]:

$$u_{km}(y,z) = \frac{2}{\sqrt{W_y\,W_z}}\,\sin\!\left(\frac{\pi k}{W_y}y\right)\,\sin\!\left(\frac{\pi m}{W_z}z\right) \tag{B.15}$$

we can calculate the integrals corresponding to the projection of each differential operator into this basis:

$$\begin{aligned}
\left\langle l\left|\tfrac{\partial^2}{\partial y^2}\right|k\right\rangle &= \tfrac{4}{W_yW_z}\int_0^{W_y}\sin\left(\tfrac{\pi\,l}{W_y}y\right)\left[\tfrac{\partial^2}{\partial y^2}\sin\left(\tfrac{\pi\,k}{W_y}y\right)\right]\mathrm{d}y = -\left(\tfrac{\pi\,k}{W_y}\right)^2\,\delta_{l,k} \\
\left\langle n\left|\tfrac{\partial^2}{\partial z^2}\right|m\right\rangle &= \tfrac{4}{W_yW_z}\int_0^{W_z}\sin\left(\tfrac{\pi\,l}{W_z}z\right)\left[\tfrac{\partial^2}{\partial z^2}\sin\left(\tfrac{\pi\,z}{W_z}z\right)\right]\mathrm{d}z = -\left(\tfrac{\pi\,n}{W_z}\right)^2\,\delta_{n,m} \\
\left\langle l\left|\tfrac{\partial}{\partial y}\right|k\right\rangle &= \tfrac{4}{W_yW_z}\int_0^{W_y}\sin\left(\tfrac{\pi\,l}{W_y}y\right)\left[\tfrac{\partial}{\partial y}\sin\left(\tfrac{\pi\,y}{W_y}y\right)\right]\mathrm{d}y = \\
&\quad -\tfrac{k}{W_y}\left[\tfrac{\cos[\pi(l+k)]-1}{l+k}+\tfrac{\cos[\pi(l-k)]-1}{l-k}\right] \quad l\neq\pm k \\
\langle n\left|\tfrac{\partial}{\partial z}\right|m\rangle &= \tfrac{4}{W_yW_z}\int_0^{W_z}\sin\left(\tfrac{\pi\,l}{W_z}z\right)\left[\tfrac{\partial}{\partial z}\sin\left(\tfrac{\pi\,z}{W_z}z\right)\right]\mathrm{d}z = \\
&\quad -\tfrac{m}{W_z}\left[\tfrac{\cos[\pi(n+m)]-1}{n+m}+\tfrac{\cos[\pi(n-m)]-1}{n-m}\right] \quad n\neq\pm m
\end{aligned} \tag{B.16}$$

Thus, the total Hamiltonian matrix in k-space H_K becomes:

$$\begin{aligned}
H_K(r,c) &= H_1(i,j)\left(\frac{\pi\,k}{W_y}\right)^2\,\delta_{k,l}\,\delta_{n,m} + H_2(i,j)\left(\frac{\pi\,n}{W_z}\right)^2\,\delta_{k,l}\,\delta_{n,m} - \\
&\quad - iH_3(i,j)\,\alpha_{k,l}\,\delta_{n,m} - iH_4(i,j)\,\alpha_{n,m}\,\delta_{k,l} \\
&\quad - H_5(i,j)\,\alpha_{k,l}\,\alpha_{n,m} + H_6(i,j)\,\delta_{k,l}\,\delta_{n,m}
\end{aligned} \tag{B.17}$$

where

$$\begin{aligned}
r &= m + N_z^{KS}(k-1) + N_y^{KS}N_z^{KS}(i-1) \\
c &= n + N_z^{KS}(l-1) + N_y^{KS}N_z^{KS}(j-1)
\end{aligned} \tag{B.18}$$

with N_y^{KS} and N_z^{KS} denoting the number of sine test functions along y and z, respectively, and $i, j = 1, \ldots, 8$. In Eq.B.17 we have also implicitly defined the symbols

$$\begin{aligned} \alpha_{l,k} &= \left\langle l \left| \tfrac{\partial}{\partial y} \right| k \right\rangle \\ \alpha_{n,m} &= \left\langle n \left| \tfrac{\partial}{\partial z} \right| m \right\rangle \end{aligned} \tag{B.19}$$

Once we have solved the k-space problem, the calculated eigenfunctions U_K are the coefficients of the sine basis of Eq.B.15. The real space eigenfunctions U_R can be calculated as

$$\mathbf{U}_R = \mathbf{U}_{RK}\, \mathbf{U}_K \tag{B.20}$$

where the matrix $\mathbf{U}_{RK}$ is written as

$$\mathbf{U}_{RK} = \begin{pmatrix} \mathbf{U}_s & & & & \\ & \mathbf{U}_s & & & \\ & & \mathbf{U}_s & & \\ & & & \ddots & \\ & & & & \mathbf{U}_s \end{pmatrix} \tag{B.21}$$

and

$$U_s(r,c) = \frac{2}{\sqrt{(N_y+1)(N_z+1)}} \sin\left(\frac{\pi\, k\, y_a}{N_y+1}\right) \sin\left(\frac{\pi\, m\, z_b}{N_z+1}\right) \tag{B.22}$$

where

$$\begin{aligned} r &= z_b + N_z\,(y_a - 1) \\ c &= m + N_z^{KS}\,(k-1) \end{aligned} \tag{B.23}$$

and

$$\begin{aligned} y_a &= 1, \ldots, N_y \\ z_b &= 1, \ldots, N_z \\ k &= 1, \ldots, N_y^{KS} \\ m &= 1, \ldots, N_z^{KS} \end{aligned} \tag{B.24}$$

A special care must be used when we select the number of points in k-space (namely N_y^{KS} and N_z^{KS}). In order to prevent the aliasing effect, the number of points in real space must be grater than the number of points in k-space, i.e.

$$\begin{aligned} N_z &\geq N_z^{KS} \\ N_y &\geq N_z^{KS} \end{aligned} \tag{B.25}$$

Bibliography

[1] L. N. Trefethen, *Spectral Methods in MATLAB*. Philadelphia: PA:SIAM, 2000.

[2] M. Shin, "Full-quantum simulation of hole transport and band-to-band tunneling in nanowires using the k·p method," *Journal of Applied Physics*, vol. 106, no. 5, p. 054505, 2009.

Appendix C

Arbitrary crystal orientations with the k·p method

This appendix describes the method we used in order to simulate arbitrary crystal orientations with the 8-band **k·p** method. Given the bulk Hamiltonian $\mathbf{H}_8$ of Eq.2.50, which is a function of the 3-D wave vector $\mathbf{k}$ written in the CCS, it can always be decomposed as:

$$\begin{aligned}\mathbf{H}_8(k_x,k_y,k_z) &= \mathbf{H}_{xx}k_x^2+\mathbf{H}_{yy}k_y^2+\mathbf{H}_{zz}k_z^2+\mathbf{H}_{xy}k_xk_y+\mathbf{H}_{yz}k_yk_z+\mathbf{H}_{zx}k_zk_x+\\ &\quad +\mathbf{H}_xk_x+\mathbf{H}_yk_y+\mathbf{H}_zk_z+\mathbf{H}_0\end{aligned} \tag{C.1}$$

where all $\mathbf{H}_{ij}$ are 8×8 matrices. These matrices can be calculated in an automatic way by computing the Hamiltonian of Eq.2.50 for various values of $\mathbf{k}$:

$$\begin{aligned}\mathbf{H}_{xx} &= [\mathbf{H}_8(\mathbf{e}_x)+\mathbf{H}_8(-\mathbf{e}_x)]/2\\ \mathbf{H}_x &= [\mathbf{H}_8(\mathbf{e}_x)-\mathbf{H}_8(-\mathbf{e}_x)]/2\\ \mathbf{H}_{yy} &= [\mathbf{H}_8(\mathbf{e}_y)+\mathbf{H}_8(-\mathbf{e}_y)]/2\\ \mathbf{H}_y &= [\mathbf{H}_8(\mathbf{e}_y)-\mathbf{H}_8(-\mathbf{e}_y)]/2\\ \mathbf{H}_{zz} &= [\mathbf{H}_8(\mathbf{e}_z)+\mathbf{H}_8(-\mathbf{e}_z)]/2\\ \mathbf{H}_z &= [\mathbf{H}_8(\mathbf{e}_z)-\mathbf{H}_8(-\mathbf{e}_z)]/2\\ \mathbf{H}_{xy} &= \mathbf{H}_8(\mathbf{e}_x+\mathbf{e}_y)-\mathbf{H}_{xx}-\mathbf{H}_{yy}-\mathbf{H}_x-\mathbf{H}_y\\ \mathbf{H}_{yz} &= \mathbf{H}_8(\mathbf{e}_y+\mathbf{e}_z)-\mathbf{H}_{yy}-\mathbf{H}_{zz}-\mathbf{H}_y-\mathbf{H}_z\\ \mathbf{H}_{zx} &= \mathbf{H}_8(\mathbf{e}_z+\mathbf{e}_x)-\mathbf{H}_{zz}-\mathbf{H}_{xx}-\mathbf{H}_z-\mathbf{H}_x\\ \mathbf{H}_0 &= \mathbf{H}_8(0,0,0)\end{aligned} \tag{C.2}$$

where $\mathbf{e}_i$ is the unit versor along the direction i in the CCS.

If we now define the wave vector in the Device Coordinate System (DCS) $\mathbf{k}_d$=($k_{x,c}$,$k_{y,c}$,$k_{z,c}$) as

$$\mathbf{k}_d = \mathbf{R}_{C\to D}\mathbf{k} \tag{C.3}$$

where the 3×3 unitary rotation matrix has been introduced in Section 3.2.3, we can re-write Eq.C.1 in the DCS:

$$\begin{aligned}\mathbf{H}_8(k_{x,d},k_{y,d},k_{z,d}) &= \mathbf{H}_{xx,d}k_{x,d}^2+\mathbf{H}_{yy,d}k_{y,d}^2+\mathbf{H}_{zz,d}k_{z,d}^2+\\ &\quad +\mathbf{H}_{xy,d}k_{x,d}k_{y,d}+\mathbf{H}_{yz,d}k_{y,d}k_{z,d}+\mathbf{H}_{zx,d}k_{z,d}k_{x,d}+\\ &\quad +\mathbf{H}_{x,d}k_{x,d}+\mathbf{H}_{y,d}k_{y,d}+\mathbf{H}_zk_{z,d}+\mathbf{H}_{0,d}\end{aligned} \tag{C.4}$$

where

$$\begin{aligned}\mathbf{H}_{xx,d} &= [\mathbf{H}_8(\mathbf{e}_{x,d})+\mathbf{H}_8(-\mathbf{e}_{x,d})]/2\\ \mathbf{H}_{x,d} &= [\mathbf{H}_8(\mathbf{e}_{x,d})-\mathbf{H}_8(-\mathbf{e}_{x,d})]/2\\ \mathbf{H}_{yy,d} &= [\mathbf{H}_8(\mathbf{e}_{y,d})+\mathbf{H}_8(-\mathbf{e}_{y,d})]/2\\ \mathbf{H}_{y,d} &= [\mathbf{H}_8(\mathbf{e}_{y,d})-\mathbf{H}_8(-\mathbf{e}_{y,d})]/2\\ \mathbf{H}_{zz,d} &= [\mathbf{H}_8(\mathbf{e}_{z,d})+\mathbf{H}_8(-\mathbf{e}_{z,d})]/2\\ \mathbf{H}_{z,d} &= [\mathbf{H}_8(\mathbf{e}_{z,d})-\mathbf{H}_8(-\mathbf{e}_{z,d})]/2\\ \mathbf{H}_{xy,d} &= \mathbf{H}_8(\mathbf{e}_{x,d}+\mathbf{e}_{y,d})-\mathbf{H}_{xx,d}-\mathbf{H}_{yy,d}-\mathbf{H}_{x,d}-\mathbf{H}_{y,d}\\ \mathbf{H}_{yz,d} &= \mathbf{H}_8(\mathbf{e}_{y,d}+\mathbf{e}_{z,d})-\mathbf{H}_{yy,d}-\mathbf{H}_{zz,d}-\mathbf{H}_{y,d}-\mathbf{H}_{z,d}\\ \mathbf{H}_{zx,d} &= \mathbf{H}_8(\mathbf{e}_{z,d}+\mathbf{e}_{x,d})-\mathbf{H}_{zz,d}-\mathbf{H}_{xx,d}-\mathbf{H}_{z,d}-\mathbf{H}_{x,d}\\ \mathbf{H}_{0,d} &= \mathbf{H}_0\end{aligned} \tag{C.5}$$

All the $\mathbf{H}_{ij}$ terms are then easily mapped to the matrices we have written in App. B, either in real or in k-space. For example, in the real space case the terms $\mathbf{H}_{yy,d}$, $\mathbf{H}_{zz,d}$, $\mathbf{H}_{y,d}$, $\mathbf{H}_{z,d}$ and $\mathbf{H}_{0,d}$ enter the on-diagonal matrix $\mathbf{H}_{R,yz}$ of Eq.B.1, while the matrices $\mathbf{H}_{x,d}$, $\mathbf{H}_{xy,d}$, $\mathbf{H}_{xz,d}$ enter the off-diagonal matrix $\mathbf{W}_R$ of Eq.B.5. The matrix $\mathbf{H}_{xx,d}$ give a contribution to both on- and off-diagonal operators.

All the 8×8 matrices we have written so far should be converted to the $8N_yN_z \times 8N_yN_z$ matrices of Eqs.B.1 and B.5 by combining them with the right discretization matrices. To this purpose, we have used the Kronecker products (denoted by the symbol ⊗) to write:

$$\begin{aligned}\mathbf{H}_{R,yz} &= \frac{2}{\Delta^2 x}\mathbf{H}_{xx,d}\otimes\mathbf{I}_{yz}-\mathbf{H}_{yy,d}\otimes\mathbf{D}_{yy}-\mathbf{H}_{zz,d}\otimes\mathbf{D}_{zz}\\ &\quad -i\,\mathbf{H}_{y,d}\otimes\mathbf{D}_y-i\,\mathbf{H}_{z,d}\otimes\mathbf{D}_z-\mathbf{H}_{yz,d}\otimes\mathbf{D}_{yz}+\mathbf{H}_{0,d}\otimes\mathbf{I}_{yz}\end{aligned} \tag{C.6}$$

$$\begin{aligned}\mathbf{W}_{R,yz} &= -\frac{1}{\Delta^2 x}\mathbf{H}_{xx,d}\otimes\mathbf{I}_{yz} - \frac{i}{2\Delta x}\mathbf{H}_{x,d}\otimes\mathbf{I}_{yz} - \frac{1}{2\Delta x}\mathbf{H}_{xy,d}\otimes\mathbf{D}_y \\ &\quad -\frac{1}{2\Delta x}\mathbf{H}_{xz,d}\otimes\mathbf{D}_z \end{aligned} \tag{C.7}$$

where the differentiation matrices $\mathbf{D}_{ij}$ have been defined in Eq.B.2.

Publications of the Author

Articles in international journals

F. Conzatti, M. De Michielis, D. Esseni and P. Palestri, "Drain Current Improvements in Uniaxially Strained p-MOSFETs: A Multi-Subband Monte Carlo Study", *Solid State Electronics*, Vol. 53, n. 7, July 2009, pp. 706-711.

D. Esseni, F. Conzatti, M. De Michielis, N. Serra, P. Palestri and L. Selmi, "Semi-classical Transport Modeling of CMOS Transistors with Arbitrary Crystal Orientations and Strain Engineering", *Journal of Computational Electronics*, Vol. 8, October 2009, pp. 209-224, Review invited paper.

A. Paussa, F. Conzatti, D. Breda, R. Vermiglio, D. Esseni and P. Palestri, "Pseudospectral Methods for the Efficient Simulation of Quantization Effects in Nanoscale MOS Transistors", *IEEE Transactions on Electron Devices*, Vol. 57, n. 12, December 2010, pp. 3239-3249.

F. Conzatti, N. Serra, D. Esseni, M. De Michielis, A. Paussa, P. Palestri, L. Selmi, S. Thomas, T.E. Whall, D.R. Leadley, E.H.C. Parker, L. Witters, M.J. Hytch, E. Snoeck, T.J. Wang, W.C. Lee, G. Doornbos, G. Vellianitis, M.J.H. van Dal and R.J.P. Lander, "Investigation of strain engineering in FinFETs comprising experimental analysis and numerical simulations", *IEEE Transactions on Electron Devices*, 2011, pp. 1583-1593.

M. De Michielis, F. Conzatti, D. Esseni and L. Selmi, "On the surface-roughness scattering in biaxially-strained n- and p-MOS transistors", *IEEE Transactions on Electron Devices*, Vol. 58, n. 9, September 2011, pp. 3219-3223

S.M. Thomas, M.J. Prest, T.E. Whall, D.R. Leadley, P. Toniutti, F. Conzatti, D. Esseni, L. Donetti, F. Gamiz, R.J.P. Lander, G. Vellianitis, P.-E. Hellstrom, and M. Ostling, "On the role of Coulomb scattering in hafnium-silicate gated silicon n and p-channel metal-oxide-semiconductor-field-effect-transistors", in *Journal of Applied Physics*, 2011, accepted for publication.

F. Conzatti, M.G. Pala, D. Esseni, E. Bano and L. Selmi, "Strain Induced Performance Improvements in InAs Nanowire Tunnel-FETs", submitted to *IEEE Transactions on Electron Devices*, 2011.

F. Conzatti, M.G. Pala and D. Esseni, "Surface Roughness Induced Variability in Nanowire InAs Tunnel-FETs", submitted to *IEEE Electron Device Letters*, 2011.

Presentations in international conferences

F. Conzatti, M. De Michielis, D. Esseni and P. Palestri, "Drain Current Improvements in Uniaxially Strained p-MOSFETs: A Multi-Subband Monte Carlo Study", in *Proc. ESSDERC*, 2008, pp. 250-253.

N. Serra, F. Conzatti, D. Esseni, M. De Michielis, P. Palestri, L. Selmi, S. Thomas, T.E. Whall, E.H.C. Parker, D.R. Leadley, L. Witters, A. Hikavyy, M.J. Hytch, F. Houdellier, E. Snoeck, T.J. Wang, W.C. Lee, G. Vellianitis, M.J.H. van Dal, B. Duriez, G. Doornbos and R.J.P. Lander, "Experimental and Physics-based Modeling Assessment of Strain Induced Mobility Enhancement in FinFETs", in *IEEE IEDM Technical Digest*, 2009, pp. 71-74.

A. Paussa, F. Conzatti, D. Breda, R. Vermiglio and D. Esseni, "Pseudo-spectral method for the modelling of quantization effects in nanoscale MOS transistors", in *Proc. SISPAD*, 2010, pp. 299-302.

F. Conzatti, P. Toniutti, D. Esseni, P. Palestri and L. Selmi, "Simulation Study of the On-Current Improvements in Ge and sGe versus Si and sSi Nano-MOSFETs", in *IEEE IEDM Technical Digest*, 2010, pp. 15.2.1-15.2.4.

F. Conzatti, M.G. Pala, D. Esseni, E. Bano and L. Selmi, "A simulation study of strain of strain induced performance enhancements in InAs nanowire Tunnel-FETs", in *IEEE IEDM Technical Digest*, 2011, pp. 5.2.1-5.2.4.

F. Conzatti, M.G. Pala, D. Esseni, E. Bano and L. Selmi, "Investigation of localized versus uniform strain as a performance booster in InAs Tunnel-FETs", in *IEEE ULIS*, 2012, accepted.