UNIVERSIDAD DE GRANADA

DOCTORAL THESIS

Calculation of the electronic structure and transport properties of semiconductor nanowires

Author:
Celso Jesús Martínez Blanque

Supervisors:
Dr. Andrés Godoy Medina
Dr. Francisco Javier García Ruiz

A thesis submitted in fulfillment of the requirements
to obtain the International Doctor degree as part of the
Programa de Doctorado en Física y Ciencias del Espacio
in the

Nanoelectronics Research Group

Departamento de Electrónica y Tecnología de los Computadores

Granada, 5$^{\text{th}}$ February, 2015

Declaration of authorship

Mr. Celso Jesús Martínez Blanque, as Ph.D Candidate, and Dr. Francisco Javier García Ruiz and Dr. Andrés Godoy Medina, as Ph.D Supervisors and Professors of Electronics at the Departamento de Electrónica y Tecnología de los Computadores of the Universidad de Granada in Spain,

Guarantee by signing this thesis:

that the research work contained in the present report, entitled ***Calculation of the electronic structure and transport properties of semiconductor nanowires***, has been performed under the full guidance of the Ph.D Supervisors and, as far as our knowledge reaches, during the work, it has been respected the right of others authors to be cited, when their publications or their results have been used.

Granada, 5th February, 2015.

Celso Jesús Martínez Blanque
Ph.D Candidate

Dr. Andrés Godoy Medina
Full Professor of Electronics

Dr. Francisco Javier García Ruiz
Tenured Professor of Electronics

To my Parents

Physical constants

ϵ_0	vacuum permittivity	$8.85418782 \cdot 10^{-12}$ F/m
m_0	electron rest mass	$9.10938291 \cdot 10^{-31}$ Kg
h	Planck's constant	$6.62606957 \cdot 10^{-34}$ J·s
$\hbar$	reduced Planck's constant	$1.05457172 \cdot 10^{-34}$ J·s
k_{B}	Boltzmann's constant	$1.38064881 \cdot 10^{-23}$ J/K
q	elementary charge	$1.60217656 \cdot 10^{-19}$ C

Acronyms

AC Acoustic phonons.

AD Alloy disorder.

BTE Boltzmann transport equation.

BZ Brillouin zone.

CB Conduction band.

CRS Crystal reference system.

DFT Density functional theory.

DOS Density of states.

DRS Device reference system.

EOT Equivalent oxide thickness.

FCC face centered cubic.

FEM Finite element method.

FinFET Fin field effect transistor.

GAA Gate all around.

GCA Gradual channel approximation.

HEMT High electron mobility transistor.

HH Heavy hole.

IC Integrated circuit.

LA Longitudinal acoustic phonon mode.

LH Light hole.

LO Longitudinal optical phonon mode.

MOSFET Metal oxide semiconductor field effect transistor.

MRT Momentum relaxation time.

MuG Multi gate.

NMOS N-type MOSFET.

NW Nanowire.

OP Optical phonons.

PMOS P-type MOSFET.

SCE Short channel effects.

SO Spin orbit.

SOI Silicon on insulator.

SS Subthreshold swing.

TA Transversal acoustic phonon mode.

TB Tight binding.

TO Transversal optical phonon mode.

VB Valence band.

List of Symbols

a_0	Lattice constant.
$\widehat{abc}$	Triangle formed by the nodes a,b, and c.
a_{v}, b, d	Alternative strain parameters for the six band $\boldsymbol{k}{\cdot}\boldsymbol{p}$ method.
$\tilde{A}_{\mathrm{c}}$	Conduction band parameter for eight band $\boldsymbol{k}{\cdot}\boldsymbol{p}$ method.
c	Vacuum light velocity.
C_{ij}	Stiffness constants.
C_{G}	Total gate capacitance.
C_{ins}	Insulator capacitance.
C_{inv}	Inversion capacitance.
C_{s}	Semiconductor capacitance.
D	Domain of f.
$D_{ij}^{\alpha,\beta}$	Deformation potentials for the strain.
D_{ac}	Acoustic phonon constant.
$D_{\mathrm{a}}, D_{\mathrm{b}}, D_{\mathrm{l}}, D_{\mathrm{m}}, D_{\mathrm{n}}$	Deformation potentials for the eight band $\boldsymbol{k}{\cdot}\boldsymbol{p}$ method.
D_{op}	Optical phonon constant.
$D_{\mathrm{u}}, D_{\mathrm{sh}}$	Uniaxial and shear deformation potentials for the two band $\boldsymbol{k}{\cdot}\boldsymbol{p}$ method for Si.
E_{PH}	Energy of the phonon.
E_i	Energy of the state i.
E_{c}	Conduction band edge.
E_{F}	Fermi energy level.
E_{g}	Energy gap (measurable).
E_{g}'	Energy gap without SO coupling ($E_{\mathrm{c}} - E_{\mathrm{v}}'$).

E_P	Optical matrix parameter.
E_v	Valence band edge.
E_v'	Valence band edge without SO coupling.
f_0	Fermi distribution function.
f	Generic function for FEM.
$\tilde{f}$	Interpolated f function.
f_n	Occupation probability function.
$\boldsymbol{F}$	External electric field.
F_c	Conduction band Kane parameter.
$\mathbb{F}$	External electric force.
F_{ij}	Form factor.
g	Density of states.
g_s	Spin degeneracy.
$\hat{\boldsymbol{H}}$	Hamiltonian operator.
$\boldsymbol{H_{\mathbf{kp},2}}$	2 band $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ matrix.
$\boldsymbol{H_{\mathbf{kp},3}}$	3 band $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ matrix.
$\boldsymbol{H_{\mathbf{kp},6}}$	6 band $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ matrix.
$\boldsymbol{H_{\mathbf{kp},8}}$	8 band $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ matrix.
$\boldsymbol{H}$	Hamiltonian matrix.
$\mathbf{H}_{ij}$	Subblock ij of the matrix $\boldsymbol{H}$.
$\boldsymbol{H}_\mathrm{SO}$	Spin Orbit matrix.
H_st	Strain component of the hamiltonian.
i	Imaginary unit.
$i = |n, k_\mathrm{z}\rangle$	i state corresponding to the n-th subband and wave vector k_z.
$j = |n', k_\mathrm{z}'\rangle$	j state corresponding to the n'-th subband and wave vector k_z'.
J_n	Current density of the n-th subband.
$\boldsymbol{k}$	Wave vector.
$\hat{\boldsymbol{k}}^\mathrm{C}$	Momentum operator $\hat{\boldsymbol{k}}$ in the CRS.
$\hat{\boldsymbol{k}}^\mathrm{D}$	Momentum operator $\hat{\boldsymbol{k}}$ in the DRS.
k_z	z component of the wave vector.
$\mathcal{L}$	Device length .
L, M, N	Parameters for six band $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ method.

$\tilde{L}, \tilde{M}, \tilde{N}$	Reduced parameters for eight the band $\boldsymbol{k} \cdot \boldsymbol{p}$ method.
m	Generic particle mass.
m_{HH}^*	Heavy hole effective mass.
m_{l}	Electron longitudinal effective mass.
m_{LH}^*	Light hole effective mass.
m_{t}	Electron transversal effective mass.
M_{ij}	Scattering matrix element.
$\boldsymbol{M}$	$\boldsymbol{M}$ matrix for FEM.
$\mathbf{M}_{ij}$	Subblock ij of the matrix $\boldsymbol{M}$.
n	Electron concentration.
n_{PH}	Number of phonons.
N_1, N_2, N_3	Nodal functions for the first, second and third node of the triangle.
N_{A}	Acceptor concentration.
N_{c}	Number of coefficients of the $\boldsymbol{k} \cdot \boldsymbol{p}$ method.
N_{D}	Donor concentration.
N_{k}	Number of discrete points in k_{z}.
N_{n}	Number of nodes of the mesh in the FEM.
N_{s}	Number of subbands.
N_{t}	Number of triangles of the mesh in the FEM.
p	Hole concentration.
$\mathbf{p}_{mn}$	Momentum matrix element.
$\hat{\mathbf{p}}$	Momentum operator $\hat{\mathbf{p}} = -\imath \hbar \nabla$.
$\hat{\mathbf{p}}_{\mathbf{r}}$	Momentum operator in the confined coordinate system.
P	Momentum interband matrix element for eight band $\boldsymbol{k} \cdot \boldsymbol{p}$ method.
P_{I}	Hole inversion density.
Q_{G}	Total charge in the gate.
Q_{inv}	Inversion charge.
Q_{s}	Semiconductor charge.
$\boldsymbol{r}$	Position vector in spatial coordinates $\boldsymbol{r} = (x, y, z)$.

$\mathbf{r}$	Spatial coordinates in the confined directions. For 2D confined systems $\mathbf{r} = (x, y)$.
$\boldsymbol{z}$	Spatial coordinates in the non confined directions. For 2D confined systems $\boldsymbol{z} = z$.
$\boldsymbol{R}$	Lattice periodic vector.
$\boldsymbol{R}^{\mathrm{CD}}$	Rotation matrix from DRS to CRS.
R_{s}	Radius of the channel in cylindrical NWs.
$\boldsymbol{R}_{\mathrm{SO}}$	Spin rotation matrix.
$\mathcal{S}$	Area of the device cross section.
S	Area of the triangular element in FEM.
S_{ij}	Scattering rate from state i to state j.
S_{in}	Scattering-in flux.
S_{out}	Scattering-out flux.
t	Time.
T	Temperature.
T_{ins}	Insulator thickness.
$u_{n\boldsymbol{k}}$	Lattice periodic function for the Bloch waves.
v_i	Velocity of the carrier in state i.
v_{s}	sound velocity.
V	Potential energy.
V_{cr}	Potential energy of the lattice.
V_{G}	Gate voltage.
V_{T}	Threshold voltage.
W_{nk}	Eigenvalue of the perturbation theory for an state $|n, \boldsymbol{k}\rangle$.
z_{I}	Inversion charge centroid.

Greek Symbols

$\tilde{\rho}_{\mathrm{q}}$	Prediction of the charge density.
δ	Dirac delta function.
δf	Deviation of the occupation probability function.
δ_{ij}	Kronecker delta function.
ΔU	AD scattering parameter.
Δ_{SO}	Spin orbit splitting.
Δ, Λ, Σ	High symmetry directions of the lattice.
ε	Dielectric constant.
$\boldsymbol{\epsilon}$	Strain matrix.
ϵ_{q}	Maximum relative error for the charge allowed.
$\gamma_1, \gamma_2, \gamma_3$	Luttinger parameters for $\boldsymbol{k}{\cdot}\boldsymbol{p}$ method.
$\Gamma, \mathrm{X}, \mathrm{L}, \mathrm{K}, \mathrm{U}$	High symmetry points of the lattice.
Λ_a	Basis function of the node a.
μ	Mobility.
∇	Partial derivative operator defined for the spatial coordinates $\boldsymbol{r}$.
$\nabla_{\boldsymbol{k}}$	Partial derivative operator defined in the wave number space $\boldsymbol{k}$.
$\nabla_{\mathbf{r}}$	Partial derivative operator defined for the confined spatial coordinates $\mathbf{r}$.
ω	Angular frequency.
Ω_{B}	Brillouin zone volume.
Ω_{C}	Unit cell volume.

ϕ	Spatial dependence of the wave functions defined in the confined spatial coordinates $\mathbf{r}$.
Φ_{ms}	Work function difference.
ψ_{s}	Surface potential in the interface.
Ψ	Quantum state.
Ψ_{t}	Full spatial-temporal description of the wave function.
ρ_{q}	Charge density.
ρ	Material density.
$\boldsymbol{\sigma}_{\mathrm{P}}$	Pauli matrices.
$\boldsymbol{\sigma}$	Stress matrix.
τ	Momentum relaxation time.
$\hat{\theta}$	Time reversal operator.
$\xi_{n\boldsymbol{k}}$	Envelope function of the wave functions in the confined coordinates $\mathbf{r}$.
ζ	Wave function.

Abstract

Nanoelectronics Research Group
Departamento de Electrónica y Tecnología de los Computadores

Calculation of the electronic structure and transport properties of semiconductor nanowires

by Celso Jesús Martínez Blanque

The main objective of this PhD Thesis is the study of the performance of nanowire transistors, as they are postulated as an alternative for future technological nodes. To do so, this work presents the physical background and the numerical tools employed to achieve an accurate description of the electrostatic and transport properties of such devices, accounting for the dominant quantum effects which they undergo.

We begin with the self-consistent solution of the Poisson and Schrödinger equations in the 2D cross-section of an arbitrary nanowire, to obtain the potential and the charge distribution. The simulator is especifically designed to achieve a satisfactory description of the holes behavior in devices fabricated with Si, Ge, $Si_{1-x}Ge_x$ and III-V materials. For this purpose the Schrödinger equation is solved by means of the $\boldsymbol{k}\cdot\boldsymbol{p}$ method, extended to 2D confined systems using the envelope function approximation. An important part of this manuscript is devoted to the implementation from scratch of this $\boldsymbol{k}\cdot\boldsymbol{p}$ method, including the physical treatment of the problem, the mathematical and numerical implementation of the method, and the validation of the results.

Then, the numerical simulator is used to carry out several electrostatic and transport studies of Si, Ge and III-V NWs, for different orientations and geometries. The influence of the NW size on the bandstructure and the charge distribution as a function of the applied gate bias is assessed. As the diameter is reduced, the LH subbands split from the HH ones. For the smallest sizes the LH subbands predominate with respect to the HH ones, yielding a reduction of the mean effective mass. This effect is stressed

in Si devices compared to Ge ones, and similar mean effective masses can be achieved for the smallest devices, specially along the [111] orientation.

Furthermore, in order to determine the transport properties, the linearization of the 1D Boltzmann Transport Equation using the momentum relaxation time approximation is implemented, accounting for phonons and alloy disorder as scattering mechanisms. Finally, the mobility is calculated by means of the Kubo-Greenwood formula.

In this regard, we have focused on the hole mobility in Si, Ge and $Si_{1-x}Ge_x$ nanowires. We demonstrate that for low inversion charge, Ge NWs clearly outperform their Si counterpart. However, in the high inversion regime the hole mobility shows a significant degradation in Ge channels, with a reduced improvement compared to Si NWs. For $Si_{1-x}Ge_x$ devices, the influence of the AD on the total mobility was evaluated, demonstrating its dominance at low inversion, whereas for high inversion the optical phonon contribution degrades considerably the hole mobility for any Ge molar fraction, reducing the beneficial effect of the alloy. Furthermore, the impact of using two common approximations for the overlap integrals is assessed. These approaches neglect the dependence of the wave function with the wave vector k_z. Both approximations lead to a high dispersion of the hole mobility values calculated compared with our full band procedure.

Resumen

Nanoelectronics Research Group

Departamento de Electrónica y Tecnología de los Computadores

Calculation of the electronic structure and transport properties of semiconductor nanowires

by Celso Jesús Martínez Blanque

Esta tesis tiene por objeto el estudio de las propiedades electrónicas y de transporte de los nanohilos semiconductores. La importancia tecnológica de estos dispositivos se basa en su potencial como alternativa de futuro a los dispositivos tradicionales dada la mejora en las prestaciones que presentan. Precisamente para poder predecir de manera precisa las prestaciones de estos dispositivos hemos comenzado con el estudio de la estructura de bandas en sistemas semiconductores confinados en dos dimensiones. Para ello hemos hecho uso del método $\boldsymbol{k}\cdot\boldsymbol{p}$ que permite una descripción correcta de la relación $E - k$ en estos sistemas con un coste computacional limitado. Hemos comprobado la calidad de los resultados mediante comparaciones con el método Tight-Binding, que corresponde a una descripción atomística. Hemos observado ligeras desviaciones, dentro de lo previsto en aproximaciones basadas en parámetros semiempíricos.

A continuación hemos resuelto de manera autoconsistente las ecuaciones de Poisson y Schrödinger en la sección transversal de un nanohilo. Como resultado obtenemos la distribución de potencial y de carga. Nos hemos centrado en el estudio de huecos como portadores de carga, puesto que se trata de un tema poco analizado en la literatura si lo comparamos con la cantidad de trabajos dedicados a los electrones. El simulador numérico nos ha proporcionado una gran cantidad de resultados relativos a propiedades electroestáticas y de transporte en nanohilos compuestos de Si, Ge, $Si_{1-x}Ge_x$ y compuestos III-V. Hemos analizado la influencia del tamaño y la tensión de puerta aplicada sobre la estructura de bandas. Se ha observado que al reducir el tamaño dominan las

bandas de huecos ligeros sobre los pesados, produciendo una reducción de la masa efectiva promedio, sobre todo a lo largo de la orientación [111].

Para el análisis de las propiedades de transporte se ha procedido a linealizar la ecuación de Boltzmann mediante la aproximación MRT (Momentum Relaxation Time). Se ha considerado la dispersión de los portadores producida por fonones acústicos y ópticos así como por el desorden en las aleaciones (alloy disorder AD). Finalmente la movilidad de los huecos se ha estimado mediante el fórmula de Kubo-Greenwood.

Con este simulador nos hemos centrado en la movilidad de huecos en nanohilos fabricados con Si, Ge y $Si_{1-x}Ge_x$. Se ha comprobado que en baja inversión la movilidad de huecos en Ge es muy superior a la que se obtiene con Si. No obstante, en fuerte inversión, esta movilidad se degrada considerablemente debido al efecto de los fonones ópticos. Se ha realizado un estudio pormenorizado de los nanohilos compuestos de $Si_{1-x}Ge_x$, en concreto del efecto del AD sobre la movilidad de los huecos. Se ha comparado el efecto de distintos parámetros que modelan la importancia relativa de este mecanismo de dispersión. Además, hemos analizado distintas aproximaciones que se realizan con frecuencia al calcular la integral de solapamiento entre estados iniciales y finales en los elementos de matriz y se han comparado con la situación más precisa obtenida al considerar la evolución de las funciones de onda con k.

Contents

Part I

Introduction

Chapter 1

Introduction

Electronics have powered the advances in technology and science for the last decades, producing a huge social impact on the industry and society that has induced changes in the world beyond anything that could have been imagined before. This amazing progress has been possible due to the astonishing performance increase of the Metal-Oxide-Semiconductor Field-Effect-Transistor (MOSFET) on account of its continuous size reduction [30]. The cost of a single transistor on an integrated circuit (IC) has been reduced in more than a millionfold, whereas the number of devices on a single chip has increased in a similar rate during the last decades. This exponential size reduction follows the so-called Moore's law [95]. The most direct consequence of this progression is the increase in the integration capacity of transistors in an IC, and hence, in the computational power.

Integrated circuits use the MOSFET as a basic switching element for digital logic applications and as an amplifier for analog ones, and its supremacy is expected to continue in the future. Nevertheless, it is not clear that the semiconductor industry can keep on with the scaling process in the future, due to the numerous technological challenges arising from both device physics and manufacturing capabilities. However, the size reduction trend still remains, supported by the society demand on new products, and new technological nodes are reached every few years [3].

Each processor generation involves a leap forward in the fabrication technology, and it materializes in a noticeable enhancement of the device and system performance. Nowadays, new challenges have to be addressed in order to keep up this trend and to respond to the voracity of the technology market for new groundbreaking products.

Intel claims that between the 45 nm of 2007, and the 14 nm of 2014, they have achieved to double the performance of their CPUs.

The performance enhancement comes along with an essential challenge: the reduction of power consumption. This is a very relevant issue since the portable devices are of major importance in the society and there is a strong demand for, not only computational power, but also energy saving and autonomy. Moreover, there is not a halt expected on the incessant progression of the semiconductor industry: the next generation is expected for 2016, the 10 nm node [3].

1.1 Problematic and limits of downscaling

Since the birth of electronics, the MOSFET downscaling has led to a general improvement of the performance in several aspects: the smaller the device size, the better the performance in terms of density, speed, cost, functionality and power figures. These improvements are the result of a better control of the channel by the gate terminal, which involves less energy and time to gather and remove the charge. Therefore, the boost of the performance was subjected to the reduction of the device size, which is purely a technological challenge but not a design problem. Back then, the end of the road was still supposed to be caused by technological issues: in the seventies, the lithography resolution seemed to be the main concern [57, 144]; while in the early eighties, tunneling through the insulator, and the resulting inadmissible current leakage and device breakdown, was thought to be the ultimate restriction [59, 127].

Nonetheless, the benefits obtained by shrinking the devices as they approximate to the end of the roadmap [51] have some limitations due to the arising of undesirable effects such as:

1. The short channel effects (SCEs) which result in the lost of control of the gate terminal over the channel charge. They arise when the drain-source field controls a large fraction of the charge located below the gate. This effects can be exacerbated as the channel length is in the nanometer range.

2. The increase in the circuit power density, resulting from the impossibility to further scaling the supply bias and threshold voltage while keeping a high $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio.

3. Unacceptable high device variability, as a consequence of technological fabrication

issues.

Moreover, these undesirable effects are strongly interrelated, complicating the problem: an improvement of one of them may lead to an aggravation of the other.

The main concern to achieve a good operation of the device is to keep a reasonable $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio, which allows to differentiate both states. A reduction of the device size leads to an increase in the integration density and consequently in the power density for the same current levels. Therefore, a scaling of I_{ON} is required to keep the power density within safe levels. The reduction of I_{ON} could be achieved by lowering the supply voltage V_{DD}, resulting in a lower power consumption. However, this trend cannot be hold indefinitely since there are two constraints: the limitation of the V_{DD} reduction and the enhancement of the I_{OFF}.

To compensate the I_{ON} drop and keep the ON-OFF ratio constant, the I_{OFF} current should be reduced in the same amount. However, when the V_{DD} is lowered, the threshold voltage V_{T} has to be reduced to keep the same gate overdrive voltage. As a result, there is an exponential increase in I_{OFF} due to the dependence of the subthreshold current with the gate voltage. This increase is determined by the subthreshold swing, SS, which for MOSFETs at room temperature under ideal conditions faces to a theoretical barrier of $60\,\mathrm{mV/dec}$. In the practice, the best MOSFET implementations cannot bring SS$<$70–80mV/dec, which leads to an even worse situation [28, 73]. New designs such as tunnel FETs should be considered to achieve lower SS values.

As for the SCEs, the channel length shortening may also contribute making possible the direct tunneling between the source and the drain, and the drain induced barrier lowering (DIBL) [124] which strongly affects the ON-OFF ratio by yielding a punch-through in the channel. A traditional solution for the SCEs is to reduce the oxide thickness to increase the gate control over the channel by augmenting the transverse electric field. The reduction in the SiO_2 thickness to avoid short channel effects halts due to direct tunneling. The limit for the scaling of the SiO_2 thickness is around to 3 atomic layers, which corresponds to $1\,\mathrm{nm}$. The technology required to manipulate the material structure in such a precise manner has some limitations, and the variability becomes a critical feature which affects the device performance [5, 133].

Finding solutions to these issues has slowed down the shrinking of the device size, leading to the use of new materials and device designs. The electronics industry has been able to reinvent itself keeping a continuous improvement of their product's performance. There is still a long way to go before the MOSFET technology is exhausted,

as it is detailed in the next section.

1.2   Beyond the limits

We have already enumerated some of the critical issues which affect the traditional miniaturization process. However, the scaling of the electronic devices has continued using new strategies that have overcome the adversities. Among them, we focus on the use of strain, high-κ insulators (e.g., $Zr0_2$, HfO_2, La_2O_3), and metal gate materials, since they have driven the last generations of planar technology and will be also determinant for the future generations.

Strain was adopted as a carrier mobility booster for the first time in the 90 nm technology node, and it has been commonly used up to now for advanced planar and non-planar devices. It is based on the modification of the lattice structure of the material produced by applying external forces, resulting in a compressed or stretched lattice. This change in the arrangement of the atoms cause modifications on the band structure of the material, leading to an enhancement of the carrier mobility [4]. The chosen material, the strain direction, its intensity, and the device orientation give rise to a vast number of possibilities in the tailoring of the transport properties that can be denominated as strain engineering. However, there are also some restrictions, as not all the possibilities lead to performance improvements and, unfortunately, the best option for one type of carrier is not necessarily convenient for the other type [135, 136].

Beginning at the 45 nm node, the use of high-κ insulators has allowed further improvements concerning the gate leakage current of ultrascaled planar Si CMOS devices. The ultimate goal is to improve the gate control over the channel keeping a reasonable insulator thickness to avoid gate leakage currents [45, 147]. The main drawback of this technology is the low quality of the interface between the insulator and the semiconductor, although stacking the high-κ material on a thin SiO_2 film can overcome this problem [94]. Nevertheless, active research is being conducted to achieve high-quality insulators directly on the semiconductor, reducing the equivalent oxide thickness (EOT). Some promising results are being obtained for La_2O_3 to reduce the EOT below 1 nm [70]. Along with the new insulators, the use of metallic gates has mitigated the harmful polysilicon depletion contribution to the EOT. The combination of these two techniques has successfully extended the lifetime of planar silicon CMOS to the 32 nm technology node.

From this point on, further device improvements require more revolutionary modifications in the design of the Field Effect Transistors (FETs). Due to the higher influence of SCEs, bulk planar technology can no longer satisfy the miniaturization requirements. In this new scenario, different proposals affecting the traditional structure of the device are being considered to enhance the gate control over the channel.

One possible solution is the use of silicon-on-insulator (SOI) technology, which consists on the introduction of a buried insulator layer underneath the active semiconductor, which isolates it from the bulk substrate. The insulator reduces the active volume of the transistor and, consequently, enhances the gate control mitigating the SCEs. The compatibility and versatility of this technology has resulted in an industrial consolidation and a promising future [122]. In addition, it also presents other benefits, such as the reduction of the negative effects produced by ionizing radiation, reduction of the parasitic capacitance [25], or the possibility of integration of different components such as Micro-Electro-Mechanical Systems (MEMS) or waveguides.

MultiGate (MuG) architectures were proposed to increase the control of the gate over the semiconductor channel. The idea is to add gates around the channel in order to enclose it as much as possible. The first of these devices, the Double Gate (DG) MOSFET, was postulated by *Sekigawa et al.* in the early eighties [121]. They demonstrated that these devices controlled more efficiently the energy barrier between the source and the drain, alleviating the SCEs. Since then, different MuG devices have been postulated to exploit their properties. A variation of the DG MOSFET is the FinFET, where the two gates are placed in a vertical direction, and they are altogether covered by the gate terminal [137]. The main advantage of the FinFET device is that it is CMOS compatible, while maintaining the main advantages of DG devices. The Tri-Gate device appeared as an evolution of the FinFET, where the insulator thickness of the top region is reduced, creating a top channel [34]. Finally, in the Gate All Around (GAA) or Nanowire (NW), the channel is completely surrounded by the gate [152]. GAA structures are considered as the ultimate architecture to achieve the best gate control over the channel. However, there is still a lot of work to be done before NW devices can be commercially feasible, especially from a technological point of view.

Another alternative to improve the device performance is the use of alternative semiconductor materials. Silicon has some limitations and, in terms of mobility, there are different materials with better properties. Consequently, along the history of electronics, many candidates have emerged.
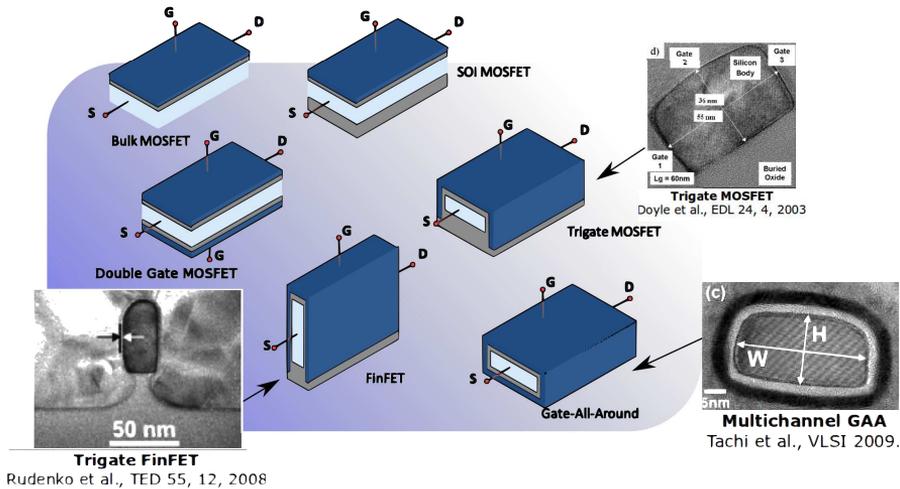
Figure 1.1: Evolution of the design in the MOSFET technology.

Among the present day alternatives, one of the strongest candidates are the III-V compound materials. These materials are well-known in the electronic industry since they have been extensively used in specific applications, such as high electron mobility transistor (HEMTs), and optoelectronics applications (light emitting diode, laser,..), and they have been postulated for a long time as the replacement of Si in logic applications. The main interest of these materials is their exceptional electron mobility, which is due to the low Γ valley effective mass, and their high injection velocity. However, the small effective mass also means a small density of states and, therefore, a low carrier concentration which can result in a lower conductivity. Therefore, a trade-off must be adopted between the carrier concentration and the carrier velocity in order to enhance I_{ON}. Moreover, there are different compounds, with their particular characteristics, which can be combined to form ternary alloys. Strain can also be applied to modify the material features. So that, researchers have at their disposal a lot of designing possibilities.

While III-V compound materials are a good alternative to Silicon for NMOS fabrication, they do not show promising features for hole transport (although antimonide compounds may be an exception [150]). Thus, for PMOS, Germanium (Ge) and SiGe compound remain as the most promising materials. Hybrid technology, including III-V materials for NMOS and Ge for PMOS, has already been experimentally demonstrated [131].

The implantation of these materials replacing silicon constitutes a challenge since

there are some issues that must be overcome. The combination of materials and strain are being studied in order to get an optimum CMOS design. Moreover, some technical problems, such as the choice of an appropriate gate insulator, or the incompatibility with the Si fabrication process, have to be addressed. In any case, III-V compounds and Ge are considered as a possible alternative to boost the performance of future electronic devices [3].

1.3   Computational Modeling

The fabrication of new semiconductor nanostructures involves costly cutting-edge technological processes which are only affordable by a few companies and research labs. Consequently, the experiments are prohibitive and must be performed only when strictly necessary. In this context, it is of prime importance to be able to predict the performance of these devices in advance. This is not a straightforward task, since the consideration of different physical phenomena is mandatory to estimate the device electrical properties. Thus, appropriate physical models must be implemented and validated to achieve reliable results.

In this context, computational modeling becomes very useful since it provides a feasible opportunity to implement the physical models, and therefore, to reveal the details of the device behavior, describing different features from the macroscopic description to the atomic scale. Once the models are implemented and validated, they provide a valuable insight at a reduced economic cost and in a short time interval.

As the technological progress allows the fabrication of smaller devices, quantum phenomena become dominant and the traditional classical models are no longer acceptable. This fact obliges to adopt new strategies and to develop new models accounting for these new issues. The quantum phenomena involved are well known by physicists and there are different theories that can accurately describe the behavior of the carriers in such small devices at different levels of approximation. While some complex methods, such as the Density Functional Theory (DFT) or Tight-Binding (TB) can model the atomic interaction with high accuracy at a high computational cost, other macroscopic models can provide information at a different level of description, creating a hierarchy of physical models.

In any case, the adoption of new models to describe the features of a device is not an easy task. Many factors influence the behavior of the carriers in a device, and the

assumptions must be validated by testing the results with experimental data.

1.4   Objectives

This PhD Thesis is devoted to the study of one of the most promising alternative device to continue the MOSFET downscaling process beyond the 14nm technology node: the Gate All Around transistor. Also, we have placed an special emphasis in SiGe and III-V based devices as an alternative to Si ones.

Our study is particularly aimed to the description of the valence band for bidimensional confined structures. To achieve this goal, the $\boldsymbol{k}\cdot\boldsymbol{p}$ method has been implemented, and the hole mobility has been modeled using the Momentum Relaxation Time (MRT) approximation to solve the Boltzmann Transport Equation (BTE).

The main goals of this work are:

1. Implementation of the simulation tools required for the solution of the equations involved.

2. Study of the electrostatic behavior of GAA devices using the $\boldsymbol{k}\cdot\boldsymbol{p}$ method.

3. Analysis of the hole mobility in SiGe NWs including the most relevant scattering mechanisms that influence the hole transport properties.

1.5   Methodology

The aim of this manuscript is to present the algorithms that have been developed for the SP2D simulation tool. Some preliminary results are also reported in this document; however, the full potential of the developed simulator has to be still exploited in future research works.

The manuscript is structured as follows:

Chapter 2 In this chapter, the theoretical background needed for the numerical and analytical solution of the 2D Schrödinger and Poisson equations is summarized.

Chapter 3 The $\boldsymbol{k}\cdot\boldsymbol{p}$ method as well as the different implementations that have been developed in this thesis for different materials and carrier types are described: two band $\boldsymbol{k}\cdot\boldsymbol{p}$ for electrons in Si, six band $\boldsymbol{k}\cdot\boldsymbol{p}$ for holes in diamond and zinc

blende semiconductors, and eight band $\boldsymbol{k}\cdot\boldsymbol{p}$ for electrons and holes in III-V semiconductors and alloys.

Chapter 4 The implementation of the 2D electrostatic simulator is described, with specific emphasis in its main features: arbitrary geometries, orientation effects, efficiency boosters, etc. A special attention is paid to the description and implementation of the numerical method to solve the $\boldsymbol{k}\cdot\boldsymbol{p}$ hamiltonian, the Finite Element Method (FEM).

Chapter 5 The implemented simulator is validated, comparing its results with those obtained using Tight-Binding (TB). In particular, the band structure of III-V NWs with different dimensions are compared.

Chapter 6 This chapter is devoted to the electrostatic analysis of NWs using the $\boldsymbol{k}\cdot\boldsymbol{p}$ method. Si, Ge and III-V devices are simulated, and parameters such as the mean effective mass, the hole linear density and the gate capacitance are studied.

Chapter 7 The hole mobility of SiGe NWs is studied, by means of the Momentum Relaxation Time (MRT) approach resulting from the linearization of the Boltzmann Transport Equation (BTE). Specifically, the mobility dependence on the NW diameter and the Ge molar fraction is assessed, and the role of AD is determined.

Chapter 8 The main conclusions of this Thesis, along with some future work, are summarized.

Appendixes Some additional explanations useful for a better understanding of the results are provided.

$\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ implementation

Part II

$\boldsymbol{k \cdot p}$ implementation

Chapter 2

Electrostatics of nanowires: background

In this Chapter we briefly introduce some physical concepts which will be useful for the analysis and simulation of the nanoelectronic devices analyzed in this manuscript. These concepts are thoroughly explained in general quantum mechanic and solid state textbooks [6, 117], nano and microelectronics textbooks [29, 128, 129, 148], and course lectures [1].

In this thesis we are interested in the study of semiconductor nanowires as a potential option for the development of the future technology nodes. In general terms, the MOSFET behavior is regulated by the combination of two electric fields: the transversal field produced by the gate contact, whose function is to set the charge in the channel, and the longitudinal field induced by the drain-source voltage, which drifts the carriers in the channel. Assuming a long channel we can apply the gradual channel approximation (GCA), that establishes the dominance of the transversal field over the longitudinal one. Under this approximation the charge distribution in a cross-section of the device can be described through the self-consistent solution of the Poisson and Schrödinger equations.

So that, both equations are required to describe the electrostatic properties of the device. First, the Poisson equation is solved to provide the relationship between the potential and the charge distribution. Second, the Schrödinger equation yields the energy levels and their corresponding wave functions for the potential well previously obtained. Therefore, these two equations are closely related and must be self-consistently solved.

This Chapter is organized as follows. In Section 2.1 we present the general Schrödinger equation, which is solved for crystalline lattices in Section 2.2, introducing the concept of band structure. Section 2.3 deals with the methods to calculate the band structure, particularly, the $\boldsymbol{k}{\cdot}\boldsymbol{p}$ method, which will be extended in the next Chapter. Section 2.4 outlines the envelope function approximation for confined devices which will require the introduction of operator ordering in Section 2.5. Finally, Section 2.6 is devoted to briefly explain the Poisson equation to calculate the electrostatic potential, particularized in Section 2.7 for 2D confined devices, providing the expression for the carrier concentration.

2.1   Schrödinger Equation

The Schrödinger Equation plays for the quantum mechanics the same role as the Newton Laws for the classical mechanics. In particular, it is of major relevance for the description of the carrier behavior in semiconductor nanodevices, where the quantum mechanic laws gain relevance as dimensions are reduced to the nanometer range. The dynamic Schrödinger equation for a particle of mass m can be written as:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r},t)\right)\Psi(\boldsymbol{r},t) = \mathrm{i}\hbar\frac{\partial\Psi(\boldsymbol{r},t)}{\partial t}\,, \tag{2.1}$$

where i is the imaginary unit, $\hbar = h/2\pi$ is the reduced Planck constant, $\boldsymbol{r}$ is the position vector, t is the time, and $V(\boldsymbol{r},t)$ is the potential energy. This equation describes the time evolution of the quantum state of a physical system. The Hamiltonian $\hat{\mathrm{H}}$ (the term into parenthesis in the left hand side of Eq. (2.1)) yields the contribution of the kinetic and the potential energy terms. $\Psi(\boldsymbol{r},t)$ corresponds to the wave function of the system that provides a probabilistic description of its properties.

The complexity of Eq. (2.1) prevents from achieving an analytical solution. However, some approximations can be carried out to obtain a solution in some scenarios. In some cases it is useful to consider static potentials to look for solutions with separated variables, as the product of a position dependent function and a time dependent one: $\Psi(\boldsymbol{r},t) = \phi(\boldsymbol{r})\cdot\zeta(t)$. So that, the new equation reads as

$$\frac{1}{\phi(\boldsymbol{r})}\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r})\right)\phi(\boldsymbol{r}) = \frac{1}{\zeta(t)}i\hbar\frac{\partial\zeta(t)}{\partial t}\,. \tag{2.2}$$

This equation can be split in two eigenvalue problems involving the same constant E as:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r})\right)\phi(\boldsymbol{r}) = E\phi(\boldsymbol{r})\,, \tag{2.3}$$

$$i\hbar\frac{\partial\zeta(t)}{\partial t} = E\zeta(t)\,. \tag{2.4}$$

This Equation is the so-called static Schrödinger equation and the term in parenthesis is the Hamiltonian operator[1] $\hat{\mathrm{H}}$ which provides the description of the energy of the system. This equation is an eigenvalue problem for the Hamiltonian, which provides the states of the system. For each state, the eigenvalue of the Hamiltonian corresponds to its energy E and the eigenfunction to its wave function $\phi(\boldsymbol{r})$. The Hamiltonian encompasses the kinetic and potential energy contributions.

A system may have different solutions, that can be differentiated by a subbindex n, or principal quantum number, in their corresponding energy and wave function. Depending on the type of problem it can be necessary to use more quantum numbers to unequivocally define a particular solution of the problem.

The temporal evolution of the state is described by Eq. (2.4), the solution of which is harmonic and depends on the energy of the state

$$\zeta(t) = K\mathrm{e}^{-\mathrm{i}Et/\hbar}\,, \tag{2.5}$$

where K is an arbitrary constant.

In a system composed by several particles, the Hamiltonian includes sums over all of them and their possible cross interactions, resulting in an extremely complex problem. This is the case for semiconductors that present a crystalline structure where the atoms are periodically arranged over all the space. Therefore, more approximations [148] should be done to simplify the scenario. In this context, the electrons are separated into two groups: 1) Core electrons which are close and tightly bound to the core of the atom, and usually considered as lumped with it, creating the ion cores. 2) The valence electrons which are responsible for the electronic properties and chemical reactivity. They are usually located in the incomplete outermost shell of the electronic structure.

The adiabatic approximation assumes that the ions are almost static for the elec-

---

[1]We use the notation ˆ to define operators.

trons, which therefore feel a stationary potential due to the core. The vibration of the atomic core is treated in a different way, considering it as quasi-particles, the so-called phonons. Finally, another important approximation to be made is the mean-field approximation, that assumes that all the electrons experience the same average potential $V(\boldsymbol{r})$.

Consequently, the potential energy included in the Schrödinger Eq. (2.3), can be split into two components: the one resulting from the crystalline lattice V_{cr}, and the contribution of other electric fields, such as those coming from external fields and lumped in a term $V(\boldsymbol{r})$.

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V_{\text{cr}}(\boldsymbol{r}) + V(\boldsymbol{r})\right)\phi(\boldsymbol{r}) = E\phi(\boldsymbol{r})\,. \tag{2.6}$$

The lattice potential is not known a priori. However, due to the periodic nature of the crystal, symmetry can be used to unveil some of its properties. Here "symmetry" means geometrical transformations that leave the crystal unchanged, and can be used to simplify the calculations. The study of systems with these symmetries is addressed using group theory, the explanation of which is out of scope of this manuscript. However, some basic concepts will be used. More details can be found elsewhere [148].

2.2 Translational Symmetry and Brillouin zone

The crystalline lattices are highly symmetric and the most characteristic symmetry that they exhibit is the Translational Symmetry, that states the invariance of the crystal under certain spatial translation operations when no external fields are present. We only consider the periodic lattice potential $V_{\text{cr}}(\boldsymbol{r})$ of period $\boldsymbol{R}$ in Eq. (2.6). The crystal lattice is composed of unit cells, where the potential has an specific form that is repeated periodically in all the space, ideally to infinity. A direct consequence of this invariance is the Bloch Theorem, which imposes the periodicity of the carrier wave function corresponding to a particular state n:

$$\phi_n(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega_{\text{C}}}}u_{n\boldsymbol{k}}(\boldsymbol{r})\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\,, \tag{2.7}$$

where $\phi_n(\boldsymbol{r})$ is the so-called Bloch function which is normalized to Ω_{C}, the volume of the unit cell, and $u_{n\boldsymbol{k}}(\boldsymbol{r})$ is a lattice-periodic function thereby having the same periodicity
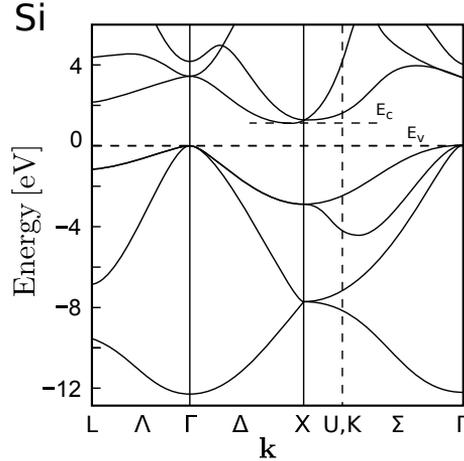
Figure 2.1: Bandstructure for Si from [26].

of the potential:

$$u_{n\boldsymbol{k}}(\boldsymbol{r}) = u_{n\boldsymbol{k}}(\boldsymbol{r} + \boldsymbol{R}). \tag{2.8}$$

The Bloch function of Eq. (2.7) depends on the wave vector $\boldsymbol{k}$, which determines the periodicity of the phase of the wave functions. Therefore, the solutions of the system are periodic functions enveloped by a planar wave with wave vector $\boldsymbol{k}$, and whose momentum, the so-called crystal momentum, is $\hbar\boldsymbol{k}$.

Since there is a solution for all the possible values of $\boldsymbol{k}$, we can consider the $\boldsymbol{k}$space as a reciprocal space where the states are defined. Therefore, the different energies calculated for each specific value of $\boldsymbol{k}$form the band structure $E_n(\boldsymbol{k})$. Figure 2.1 shows the band structure calculated for Si. The solutions of the Hamiltonian can be indexed as $|n, \boldsymbol{k}\rangle$. The bands defined in the reciprocal space are also periodic, and therefore their study can be restricted to the primitive cell of the reciprocal space, the Brillouin zone (BZ) [148].

Crystalline lattices also present other symmetries that keep the problem invariant, as stated by group theory. If the crystal remains invariant after a rotation it means that this rotation applied to a solution $\phi_n(\boldsymbol{k})$ is also a solution with the same energy. This property leads to the definition of equivalent points of high symmetry in the BZ, and, therefore, to a symmetrically repeated bandstructure inside the Brillouin zone according to the symmetries of the crystal. The wave functions associated to those states are the result of applying the symmetry operations to $\phi_n(\boldsymbol{k})$. As a result,
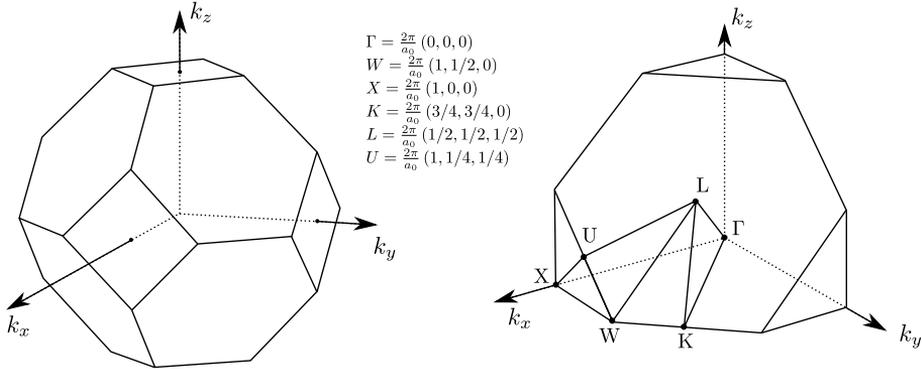
Figure 2.2: Brillouin zone of face centered cubic lattices. The high symmetry points are detailed. Extracted from [135].

the complete characterization of the band structure is finally restricted to a wedge of the BZ limited by the main symmetry points. Figure 2.2 illustrates the BZ and its high symmetry points for the particular case of face centered cubic (fcc) lattices, very common in semiconductors: in particular, Si and Ge are of this type.

2.3 Band structure determination

There are different methods to achieve solutions to equation (2.6), depending on the considered assumptions [129]. As a result, the complexity of the problem is variable, from atomistic methods, where the position of each atom is taken into account, leading to quite complex problems, to the most simple Effective Mass Approach (EMA). Depending on the material, device, dimensions, purpose, and required accuracy, one can choose among the different options to obtain the best solution with a reasonable computational cost.

In this manuscript, we mainly address the calculation of the valence band of diamond and zinc blende semiconductors. In those cases the valence band structure is not accurately described by the EMA, specially when the states are mixed due to the effect of broken symmetries. Therefore, a more accurate approach is mandatory. The chosen option in this work is the $\boldsymbol{k}\cdot\boldsymbol{p}$ method. This method was initially introduced by *Bardeen* [9] and *Seitz* [120] and it has been widely used by many researchers [35, 65, 81]. The full theory is extensively covered by *Bir and Pikus* [17]. This method derives the analytical expressions for the band structure, using symmetry arguments and experimental measurements, and it is particularly useful to describe with high precision the

band structure around the Γ point of the Brillouin zone for this type of semiconductors.

The $\boldsymbol{k} \cdot \boldsymbol{p}$ theory begins with the substitution of the Bloch functions in the Schrödinger equation, where we use the definition of the momentum operator $\hat{\mathbf{p}} = -\mathrm{i}\hbar\nabla$:

$$\left[\frac{\hat{\mathbf{p}}^2}{2m_0} + V(\boldsymbol{r})\right]\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}u_{n\boldsymbol{k}}(\boldsymbol{r}) = E_n(\boldsymbol{k})\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}u_{n\boldsymbol{k}}(\boldsymbol{r})\,. \tag{2.9}$$

After some algebra, this equation reads:

$$\left[\frac{\hbar^2|\boldsymbol{k}|^2}{2m_0} + \frac{\hbar}{m_0}\boldsymbol{k}\cdot\hat{\mathbf{p}} + \frac{\hat{\mathbf{p}}^2}{2m_0} + V(\boldsymbol{r})\right]u_{n\boldsymbol{k}}(\boldsymbol{r}) = E_n(\boldsymbol{k})u_{n\boldsymbol{k}}(\boldsymbol{r})\,. \tag{2.10}$$

In this new Hamiltonian the contribution of the planar wave component $\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}$, and of the lattice-periodic functions $u_{n\boldsymbol{k}}$, are separated. The result is a Hamiltonian acting on the Bloch functions, where the planar wave influence is considered by means of its wave vector $\boldsymbol{k}$ and how it couples with the momentum of the Bloch functions, achieved by applying the momentum operator $\hat{\mathbf{p}}$.

The resulting Hamiltonian is split into two terms, H_0 and H_1, where H_1 is considered as a perturbation of the unperturbed term H_0:

$$\left(\hat{\mathrm{H}}_0 + \hat{\mathrm{H}}_1\right)u_{n\boldsymbol{k}}(\boldsymbol{r}) = W_{n\boldsymbol{k}}u_{n\boldsymbol{k}}(\boldsymbol{r})\,, \tag{2.11}$$

$$\hat{\mathrm{H}}_0 = \frac{\hat{\mathbf{p}}^2}{2m_0} + V(\boldsymbol{r})\,, \tag{2.12}$$

$$\hat{\mathrm{H}}_1 = \frac{\hbar}{m_0}\boldsymbol{k}\cdot\hat{\mathbf{p}}\,, \tag{2.13}$$

$$W_{n\boldsymbol{k}} = E_n(\boldsymbol{k}) - \frac{\hbar^2|\boldsymbol{k}|^2}{2m_0}\,. \tag{2.14}$$

This equation system can be solved using perturbation theory. Thus, considering the known-solutions for an initial $\boldsymbol{k}_0$ as a set of basis functions, it is possible to calculate the solution at a value $\boldsymbol{k}$ close to the original $\boldsymbol{k}_0$. In the next Chapter, a more detailed discussion on this method will be presented, including the solution of the previous equation system.
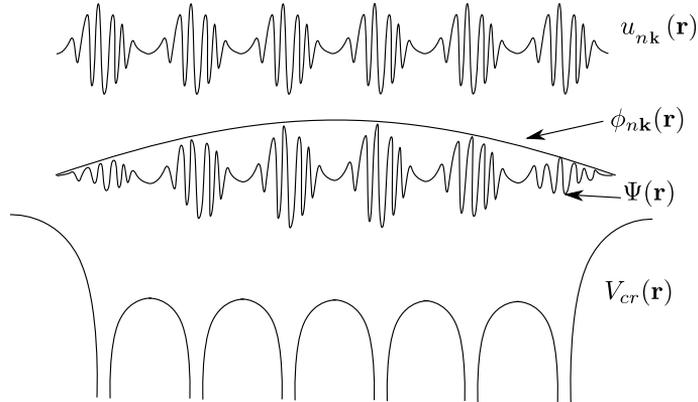
Figure 2.3: Sketch of a Bloch function for the envelope function approximation, represented along the confined direction. The size of the unit cell has been exaggerated for clarity. For the sake of simplicity, a single lattice-periodic function, and therefore, only the corresponding envelope function, has been represented.

2.4 Envelope function

In the previous Section, the Translational Symmetry of the crystal was assumed. Thus, the crystal was presumed to be homogeneous and infinitely extended. When the Translational Symmetry is not fulfilled in any of the directions, this assumption is no longer valid. Consequently, the carriers are confined in the direction where the symmetry is broken, but can still propagate in the other directions[2]. In the case of a quantum well, the symmetry is broken in one direction, for a quantum wire in two directions, and for a quantum dot in the three directions. As a result, the number of directions where the carriers can freely propagate, is also reduced. Hence, the spatial coordinates can be separated as $r = (\mathbf{r}, z)$, where the z accounts for the directions where carriers can freely move as planar waves and $\mathbf{r}$ are the coordinates in which the carrier are confined.

The wave functions have to be rewritten as a combination of lattice-periodic functions $u_m(r)$ [13], which form an orthonormal basis function, weighted by the envelope functions $\boldsymbol{\xi}_{nk}(\mathbf{r})$ in the confined directions. Thus, a wave function for an state $|n, \boldsymbol{k}\rangle$ can be expressed as:

$$\phi_{n\boldsymbol{k}}(\mathbf{r}, \boldsymbol{z}) = \sum_m u_m(\boldsymbol{r}) \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{z}} \xi_{n\boldsymbol{k},m}(\mathbf{r}) \,. \tag{2.15}$$

---

[2]As a consequence, the wave vector $\boldsymbol{k}$ reduces its dimensionality.

In Eq. (2.15) the formerly planar waves enveloping the lattice-periodic functions are separated into planar waves for the non confined direction $\boldsymbol{z}$ and a slowing varying envelope function $\boldsymbol{\xi}_{n\boldsymbol{k}}(\mathbf{r})$ for the confined directions, where planar waves cannot propagate. The plane wave expansion is restricted to the first BZ, which fulfills the boundary conditions. As a consequence, the crystal momentum in the confined directions cannot take continuous values and the states are quantized, resulting in bands splitting into the so-called subbands.

In this case, the crystal momentum in the confined direction is no longer defined as a constant $\hbar\boldsymbol{k}$, but as the result of applying the momentum operator $-\mathrm{i}\hbar\nabla_{\mathbf{r}}$, with $\nabla_{\mathbf{r}}$ defined in the confined coordinates $\mathbf{r}$. Traditionally, the wave vector in the confined directions is substituted as [139]:

$$\boldsymbol{k} \to \hat{\boldsymbol{k}} = -\mathrm{i}\nabla\,, \tag{2.16}$$

$$k_\alpha \to \hat{\mathrm{k}}_\alpha = -\mathrm{i}\frac{\partial}{\partial\alpha}\,. \tag{2.17}$$

With this replacement, the solutions for the resulting Hamiltonian $\hat{\boldsymbol{H}}(-\mathrm{i}\nabla_{\mathbf{r}})$ are the envelope functions of the problem, providing the eigenenergies of the system:

$$\hat{\boldsymbol{H}}\boldsymbol{\xi}(\mathbf{r}) = E\,\boldsymbol{\xi}(\mathbf{r})\,. \tag{2.18}$$

Assuming that the lattice constant is much smaller than the confinement size, the slowly varying envelope functions provide a description of the probability distribution along the confinement directions. The spatial probability density for the state $|n,\boldsymbol{k}\rangle$, neglecting the rapid oscillation near the core, can be calculated as $\boldsymbol{\xi}_{n\boldsymbol{k}}(\mathbf{r})^* \cdot \boldsymbol{\xi}_{n\boldsymbol{k}}(\mathbf{r}) = |\boldsymbol{\xi}_{n\boldsymbol{k}}(\mathbf{r})|^2$.

Hereinafter, when referring to the wave functions, we will be actually working with their envelope functions.

2.5   Operator Ordering

The replacement proposed in Eq. (2.16) is a widely accepted and usual approximation, but it gives rise to non trivial difficulties in the treatment of the interfaces of the device. As explained in [139], the effect is estimated to be small in abrupt interfaces and for

slowly-varying envelopes. So that, its influence on the energies calculation is believed to be rather small, although it may be relevant to study the effect of interface mixing. For further explanations, see [139] and [23]. Some ad-hoc fixes can be carried out to achieve accurate solutions.

One usual fix, justified to assure the continuity of the probability flux of the wave function in both sides of the interface, which is defined for the second order operator as:

$$-\frac{\hbar^2 \nabla^2}{2m(z)} \rightarrow -\nabla \frac{\hbar^2}{2m(z)} \nabla \, . \tag{2.19}$$

This operator ordering [14] also assures the hermiticity of the operator for spacial dependent parameters as for example the effective mass. However, it produces problems when dealing with crossed differential terms of the form $\hat{k}_i \, C(\mathbf{r}) \, \hat{k}_j$ with $i \neq j$. In this case, the operation is not reversible, and the chosen ordering for $\hat{k}_i$ and $\hat{k}_j$ when constructing the system is essential. The usual fix is to choose an average of the two terms:

$$\hat{k}_i \, C(\mathbf{r}) \, \hat{k}_j \rightarrow -\frac{\partial}{\partial i} \frac{C(\mathbf{r})}{2} \frac{\partial}{\partial j} - \frac{\partial}{\partial j} \frac{C(\mathbf{r})}{2} \frac{\partial}{\partial i} \, . \tag{2.20}$$

This ordering is referred to as *symmetric operator ordering* and might provide wrong results in some cases. In the next Chapter we detail the operator ordering chosen for the $\boldsymbol{k} \cdot \boldsymbol{p}$ calculations.

2.6   Poisson equation

In order to solve Eq. (2.6) it is necessary to find the potential energy, $V(\boldsymbol{r})$, due to the external forces and to the rest of electrons in the system. This relationship is determined by the Poisson equation, which relates the electrostatic potential, and the charge density in the device, ρ_q as:

$$\nabla[\varepsilon(\boldsymbol{r})\nabla V(\boldsymbol{r})] = -\rho_\mathrm{q}(\boldsymbol{r}) \, , \tag{2.21}$$

where $\varepsilon(\boldsymbol{r})$ is the dielectric constant. The charge density, $\rho_\mathrm{q}(\boldsymbol{r})$, is given by:

$$\rho_{\mathrm{q}}(\boldsymbol{r}) = q\left[p(\boldsymbol{r}) - n(\boldsymbol{r}) + N_{\mathrm{D}}^{+} - N_{\mathrm{A}}^{-}\right], \tag{2.22}$$

where $p(\boldsymbol{r})$ and $n(\boldsymbol{r})$ are the electron and hole concentrations respectively, q is the elementary charge, and N_{D}^{+} and N_{A}^{-} are the ionized donor and acceptor impurity concentrations respectively.

The Poisson equation together with the Schrödinger equation presented in Sec. 2.1 complete the physical background necessary for the determination of the electrostatic behavior of the devices under study in this manuscript.

2.7 Quantum carrier concentration

Carries in confined systems are distributed according to the weighted distribution functions determined by the squared norm of the wave functions, which are solution of the Schrödinger equation. To calculate the charge density it is necessary to know the number of available states and their occupation probability. The first magnitude is determined by a function $g(\boldsymbol{k})$ known as the density of states (DOS).

The periodicity of the crystal lattice implies a discretization of the wave vector space. Each $\boldsymbol{k}$ state occupies a volume in the wave vector space given by Ω_{B}/N where Ω_{B} is the volume of the Brillouin zone and N is the number of unit cells in the real space for the given volume, $\mathcal{V}$, of the semiconductor: $N = \mathcal{V}/\Omega_{\mathrm{C}}$, being Ω_{C} the unit cell volume. Then, the density of states is given by:

$$g(\boldsymbol{k}) = g_{\mathrm{s}}\frac{N/\Omega_{\mathrm{B}}}{\mathcal{V}} = g_{\mathrm{s}}\frac{(\mathcal{V}/\Omega_{\mathrm{C}})/\Omega_{\mathrm{B}}}{\mathcal{V}} = \frac{g_{\mathrm{s}}}{(2\pi)^3}, \tag{2.23}$$

where $\Omega_{\mathrm{B}}\Omega_{\mathrm{C}} = (2\pi)^3$ and g_{s} accounts for the spin degeneracy. As can be seen, the states are uniformly distributed over all the wave vector space.

When the 1D electron gas approximation is considered, Ω_{B} and Ω_{C} represent lengths instead of volumes, hence $\Omega_{\mathrm{B}}\Omega_{\mathrm{C}} = 2\pi$. The wave vector density of states is thus:

$$g(k_{\mathrm{z}}) = g_{\mathrm{s}}\frac{N/\Omega_{\mathrm{B}}}{\mathcal{L}} = g_{\mathrm{s}}\frac{(\mathcal{L}/\Omega_{\mathrm{C}})\Omega_{\mathrm{B}}}{\mathcal{L}} = \frac{g_{\mathrm{s}}}{2\pi}. \tag{2.24}$$

where $\mathcal{L}$ is the length of the device.

As for the occupation of each state, it depends on its energy E by means of the Fermi-Dirac function:

$$f_0(E) = \frac{1}{1 + e^{\frac{E - E_F}{k_B T}}} ,\tag{2.25}$$

being k_B the Boltzmann constant and E_F the Fermi level. This is a statistical expression that determines the probability of occupation of a certain state with an energy E.

In statistical terms, the contribution of an electron in the state $|n, \boldsymbol{k}\rangle$ to the total electron density $n(\boldsymbol{r})$ is the result of weighting the probability distribution of such state with the Fermi-Dirac function, $|\phi_{n\boldsymbol{k}}(\boldsymbol{r})|^2 f_0(E_n(\boldsymbol{k}))$. Since a hole is the absence of an electron, the hole density p is calculated as the probability of an state not being occupied, and can be calculated by substituting $f_0(E)$ by $(1 - f_0(E))$ in the previous expression. Then, the total electron-hole density for a 1D gas can be calculated as:

$$n(\boldsymbol{r}) = \frac{g_s}{2\pi} \sum_n \int |\phi_{nk_z}(\boldsymbol{r})|^2 f_0(E_n(k_z)) \, \mathrm{d}k_z ,\tag{2.26}$$

$$p(\boldsymbol{r}) = \frac{g_s}{2\pi} \sum_n \int |\phi_{nk_z}(\boldsymbol{r})|^2 (1 - f_0(E_n(k_z))) \, \mathrm{d}k_z ,\tag{2.27}$$

where the summation runs over all the subbands of the system, and the wave functions are normalized to the area of the cross-section $\mathcal{S}$ of the device. These equations involve the wave functions for all the states of the system. Due to complexity, it is usual to simplify by means of approximations, such as employ only the wave function at $k_z = 0$ [40]. Nevertheless, we will deal with the full band calculation of Eqs. (2.26) and (2.27). The accuracy of these two approaches is assessed in Chapter 6.

As for ionized impurities due to dopants, classical expressions are considered.

$$N_A^- = N_A \, f_0(E_A) = \frac{N_A}{1 + \frac{1}{g_A} e^{\frac{E_A - E_F}{k_B T}}} ,\tag{2.28}$$

$$N_D^+ = N_D[1 - f_0(E_D)] = \frac{N_D}{1 + g_D e^{-\frac{E_D - E_F}{k_B T}}} ,\tag{2.29}$$

where N_A (N_D), g_A (g_D) and E_A (E_D) are the acceptor (donor) concentration, level degeneracy and energy, respectively.

Chapter 3

The $k \cdot p$ method

In Chapter 2, we mentioned that the electronic structure of semiconductor nanowires can be studied using different methods. Among them, the $k \cdot p$ method has been chosen for this purpose in this thesis. This method is widely used to calculate the band structure around high symmetry points of the reciprocal lattice, where the edges of the electronic bands are located. Besides, it has raised a high interest as a numerical tool to calculate the bandstructure in quantum devices, with many works employing the $k \cdot p$ approach to analyze different phenomena [18, 46, 111, 139]. So that, in this chapter we provide a more extended insight into the $k \cdot p$ method, which is necessary for its later implementation into the simulation tool.

Among the strengths of this method that make it so popular, we can enumerate a few:

1. It is a continuum approach, which simplifies the lay out of the problem and allows to keep the semi-classical point of view used in simpler methods such as the EMA.

2. A reduced set of parameters, which can be empirically adjusted, is enough to determine the band structure near the band edges.

3. It is able to analyze the influence of the lattice symmetry breaking due to different phenomena such as strain, which can be included as additional perturbations.

4. It also captures the spin orbit interaction of the different states.

5. It suits nicely with the envelope function approximation, thereby allowing the description of nanostructures.

Nevertheless, there are also some weaknesses that must be taken into account:

1. It only reproduces the behavior of the band structure around a point of interest, and therefore the representation is less accurate for more distant $\boldsymbol{k}$ values. Nevertheless, it works well enough when the conduction carriers are located around a band maximum or a minimum.

2. It is not an atomistic method, thereby it is not able to capture all the physical processes involved.

3. For confined devices, the envelope function approximation is considered, making the problem much more complex since the $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ Hamiltonian is transformed into a partial differential equation system.

The $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ method is based on the perturbation theory, making use of it to study the band structure evolution from a point of interest in the reciprocal space. In [36], the importance of the $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ approach as a basis for the semi empirical determination of the band structure was established. By using symmetry arguments, it was shown that the evolution of the band structure in a small region near a point of interest of the $\boldsymbol{k}$ space depends only on a few parameters. Extensive derivations and calculations using the $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ method, and many review papers by E.O. Kane et al. [64–67], have transformed this perturbation approach into a very extended method in Solid State Physics.

A thorough study of the $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ theory involves the knowledge of the lattice symmetry as well as the different symmetry groups that the basis functions belongs to. Therefore, an initial knowledge in group theory is recommended, e.g. in [148]. This will be useful since the matrix elements are calculated mixing the basis states by means of the momentum operator. Those matrix elements determine the evolution of the band structure in the reciprocal space. In this chapter we provide a general insight into these $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ features and the resulting mathematical problem to be solved. A more detailed view of the symmetry operations which lead to the final layout of the model is provided in [139].

Some discrepancies are found in the literature when defining the different sets of parameters to be used in $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ calculations, which might cause confusion. Therefore, our intention here is to provide a clear definition of them, remarking the possible misunderstandings.

3.1 The $\boldsymbol{k}\cdot\boldsymbol{p}$ formalism

As introduced in Sec. 2.3, the $\boldsymbol{k}\cdot\boldsymbol{p}$ formalism is based on the consideration of the Bloch functions as solutions of the single electron Schrödinger equation (Eq. (2.6)) for a crystalline lattice. Let us rewrite the resulting system for clarity:

$$\left(\hat{H}_0 + \hat{H}_1\right) u_{n\boldsymbol{k}}(\boldsymbol{r}) = W_{n\boldsymbol{k}} u_{n\boldsymbol{k}}(\boldsymbol{r}) \,, \tag{3.1}$$

$$\hat{H}_0 = \frac{\hat{\mathbf{p}}^2}{2m_0} + V_{\mathrm{cr}}(\boldsymbol{r}) \,, \tag{3.2}$$

$$\hat{H}_1 = \frac{\hbar}{m_0} \boldsymbol{k} \cdot \hat{\mathbf{p}} \,, \tag{3.3}$$

$$W_{n\boldsymbol{k}} = E_n(\boldsymbol{k}) - \frac{\hbar^2 k^2}{2m_0} \,. \tag{3.4}$$

Equation (3.1) can be solved using perturbation theory, as outlined in Appendix A, assuming that the states for the unperturbed Hamiltonian $\hat{H}_0$, i.e., the lattice-periodic functions for $k = 0$, are known[1]. The perturbation term H_1, after which the method is named, contains the product of the wave vector $\boldsymbol{k}$ and the momentum operator $\hat{\mathbf{p}}$. Therefore, the evolution of the electronic states with $\boldsymbol{k}$ can be estimated from the mixing of the unperturbed Hamiltonian states.

Using perturbation theory for non degenerate states, as described in Appendix A, the term W_{nk} can be approximated as:

$$W_{nk} \approx W_{nk}^{(0)} + W_{nk}^{(1)} + W_{nk}^{(2)} = E_{n0} + \sum_{m \neq n} \frac{|\boldsymbol{k} \cdot \mathbf{p}_{mn}|^2}{E_{m0} - E_{n0}} \,, \tag{3.5}$$

where the term $W_{nk}^{(1)}$ vanishes since $\langle u_{n0}|\hat{\mathbf{p}}|u_{n0}\rangle = 0$, and the momentum matrix elements $\mathbf{p}_{mn}$ are given by

$$\mathbf{p}_{mn} = \langle u_{m0}|\hat{\mathbf{p}}|u_{n0}\rangle \,. \tag{3.6}$$

Then, the energy $E_n(\boldsymbol{k})$ can be calculated by substituting Eq. (3.5) into (3.4):

$$E_n(\boldsymbol{k}) = E_{n0} + \frac{\hbar^2 k^2}{2m_0} + \sum_{m \neq n} \frac{|\boldsymbol{k} \cdot \mathbf{p}_{mn}|^2}{E_{m0} - E_{n0}} \,, \tag{3.7}$$

---

[1]Actually, the method can be solved using the functions at any $\boldsymbol{k}_0$, but usually, for simplicity, $\boldsymbol{k}_0 = 0$ is considered.

As for the wave functions, they can be calculated in a first order approximation as:

$$u_{n\boldsymbol{k}} \approx u_{n0} + \sum_{m \neq n} \frac{\boldsymbol{k} \cdot \mathbf{p}_{mn}}{E_{m0} - E_{n0}} u_{m0} \tag{3.8}$$

Equation (3.7) demonstrates that the E-$\boldsymbol{k}$ relation can be approximated by a parabolic expression. After some algebra, and considering the symmetry of the cubic semiconductors, this expression can be reduced to the effective mass approximation [129]:

$$E_n(\boldsymbol{k}) = E_{n0} + \frac{\hbar^2 k^2}{2m^*}, \tag{3.9}$$

where m^* is the effective mass, given by

$$\frac{1}{m^*} = \frac{1}{m_0} \left[1 + 2 \sum_{m \neq n} \frac{|\mathrm{p}_{\mathrm{x},mn}|^2}{E_{m0} - E_{n0}} \right], \tag{3.10}$$

and $\mathrm{p}_{\mathrm{x},mn}$ stands for the x component of the momentum matrix element $\mathbf{p}_{mn}$, since its three components are equal for the particular case of cubic semiconductors. Observe that the effective mass contains a term depending on the bare electron mass and a term accounting for the coupling with other states.

3.2 The $\boldsymbol{k}\cdot\boldsymbol{p}$ method for holes: Luttinger-Kohn Hamiltonian

The study of holes in diamond type semiconductors, such as Si and Ge, can be restricted to the edge of the valence band (VB) around the Γ point. In that case, the VB edge can be described by means of degenerate perturbation theory, which is briefly presented in Appendix A.1. According to this theory, there are three degenerate states, denoted as $|X\rangle$, $|Y\rangle$ and $|Z\rangle$, while the rest of the bands are accounted for as remote states. The three degenerate states are related to the bonding states of the orbitals $|p_{\mathrm{x}}\rangle$, $|p_{\mathrm{y}}\rangle$, and $|p_{\mathrm{z}}\rangle$, thereby keeping their symmetry [129] and odd parity.

Different approaches can be considered to solve the resulting Hamiltonian. In this document, we start outlining the three band $\boldsymbol{k}\cdot\boldsymbol{p}$, even though it does not yield an accurate approximation since it does not appropriately consider the effect of the spin

coupling. However, this method is of interest since it allows to present the origin of the different valence bands, from the mixing of the $|X\rangle$, $|Y\rangle$ and $|Z\rangle$ states. Later, the six band $\boldsymbol{k \cdot p}$ method, which includes the spin orbit coupling, is analyzed.

3.2.1 Three band $\boldsymbol{k \cdot p}$ theory

As a first approximation, the spin of the states forming the VB at the Γ point ($|X\rangle$, $|Y\rangle$ and $|Z\rangle$) is not considered. So that, the problem is reduced to three degenerate bands, resulting in:

$$H_{nm} = \left[E'_{\text{v}} + \frac{\hbar^2 k^2}{2m_0} \right] \delta_{nm} + \frac{\hbar^2}{m_0^2} \sum_{\beta} \frac{\boldsymbol{k} \cdot \mathbf{p}_{n\beta} \boldsymbol{k} \cdot \mathbf{p}_{\beta m}}{E'_{\text{v}} - E_{\beta}} , \tag{3.11}$$

where δ_{nm} is the Kronecker delta function, E'_{v} is the valence band edge[2], and the sum runs over all the remote states β.

The matrix elements can be simplified making use of the symmetries of the basis states, and how they couple with the momentum operator [139]. For example, the element H_{11}, resulting from the self-coupling of the $|X\rangle$ state reads:

$$H_{11} = E'_{\text{v}} + \sum_{j=\text{x,y,z}} \left[\frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0^2} \sum_{\beta} \frac{|\langle X | \hat{\text{p}}_j | \beta \rangle|^2}{E'_{\text{v}} - E_{\beta}} \right] k_j^2 \tag{3.12}$$

It can be proven that, as a consequence of the orbital symmetry, it is fulfilled:

$$|\langle X | \hat{\text{p}}_{\text{y}} | \beta \rangle|^2 = |\langle X | \hat{\text{p}}_{\text{z}} | \beta \rangle|^2 , \tag{3.13}$$

and therefore, Eq. (3.12) can be rewritten as:

$$H_{11} = E_1 + L k_{\text{x}}^2 + M(k_{\text{y}}^2 + k_{\text{z}}^2), \tag{3.14}$$

with

$$L = \frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0} \sum_{\beta} \frac{|\langle X | \hat{\text{p}}_{\text{x}} | \beta \rangle|^2}{E'_{\text{v}} - E_{\beta}} \tag{3.15}$$

---

[2] Actually, E'_{v} is the valence band edge before SO coupling, or the energy position of the states $|X\rangle$, $|Y\rangle$ and $|Z\rangle$. After SO coupling, the valence band edge varies its position, as it will be shown later.

and

$$M = \frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0} \sum_{\beta} \frac{|\langle X|\hat{p}_y|\beta\rangle|^2}{E'_v - E_\beta} \,. \tag{3.16}$$

Similarly,

$$H_{12} = Nk_x k_y \,, \tag{3.17}$$

with

$$N = \frac{\hbar^2}{m_0^2} \sum_{\beta} \frac{\langle X|\hat{p}_x|\beta\rangle \langle \beta|\hat{p}_y|Y\rangle + \langle X|\hat{p}_y|\beta\rangle \langle \beta|\hat{p}_x|Y\rangle}{E'_v - E_\beta} \,. \tag{3.18}$$

Therefore, the three band **k·p** Hamiltonian can be written as the matrix:

$$\boldsymbol{H_{kp,3}} = \tag{3.19}$$

$$\begin{pmatrix}
& |X\rangle & |Y\rangle & |Z\rangle \\
\hline
|X\rangle & E'_v + Lk_x^2 + M(k_y^2 + k_z^2) & Nk_x k_y & Nk_x k_z \\
|Y\rangle & Nk_y k_x & E'_v + Lk_y^2 + M(k_z^2 + k_x^2) & Nk_y k_z \\
|Z\rangle & Nk_z k_x & Nk_z k_y & E'_v + Lk_z^2 + M(k_x^2 + k_y^2)
\end{pmatrix}$$

As can be seen, the evolution of the valence bands for a bulk semiconductor can be calculated using only three different parameters: L, M and N. These parameters depend on the coupling of the non perturbed states $|X\rangle$, $|Y\rangle$ and $|Z\rangle$ with the rest of the states, as a consequence of the **k·p** perturbation. The resulting valence band is threefold degenerate in the Γ point and can evolve into non degenerate states for different reciprocal space orientations. This is a direct consequence of the symmetry of the lattice, and therefore, of the states composing the valence bands.

The L and M parameters defined in Eqs. (3.15) and (3.16), respectively, are composed by a term related to the free electron dispersion and a summation depending on the coupling with the remote states. An alternative, traditional definition of the L and M parameters takes into account only the coupling part, and therefore the free electron dispersion must be accounted for externally to the parameter in the diagonal of the matrix, e.g. as in [7, 36]. For practical reasons, we have chosen to include the free electron dispersion term into L and M to simplify the notation, as in [129, 139]. This issue has to be taken into account when seeking parameter values in the literature for different materials.

3.2.2 The six band $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ method: spin orbit coupling

The three band $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ method yields an intuition on how the valence bands behave in relation with the symmetry of the semiconductor: there are three degenerate bands in the center zone that evolve anisotropically into non degenerate ones. However, this method does not provide an accurate description of these bands since it ignores the effect of the spin orbit (SO) coupling. Up to now, the spin of the states was ignored, that is, the independence of the spin states was wrongly assumed.

The spin orbit coupling is produced by the coupling of the spin with the orbital angular momentum, which contributes to the total energy of the system, and can be accounted for introducing the following term into the Hamiltonian operator:

$$\hat{H}_{SO} = \frac{\hbar}{4m_0^2 c^2}(\nabla V_{cr} \times \hat{\mathbf{p}}) \cdot \boldsymbol{\sigma}_P \,. \tag{3.20}$$

where c is the vacuum light velocity, ∇V_{cr} is the crystal potential divergence, and $\boldsymbol{\sigma}_P = (\boldsymbol{\sigma}_x, \boldsymbol{\sigma}_y, \boldsymbol{\sigma}_z)$ are the Pauli matrices. In the $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ formalism, $\hat{H}_{SO}$ splits into two terms [7], which are also treated as perturbations. The first one accounts for the momentum

$$\hat{H}_{SO,\mathbf{p}} = \frac{\hbar}{4m_0^2 c^2}(\nabla V_{cr} \times \hat{\mathbf{p}}) \cdot \boldsymbol{\sigma}_P \,, \tag{3.21}$$

and the second one for the wave vector

$$\hat{H}_{SO,\mathbf{k}} = \frac{\hbar}{4m_0^2 c^2}(\nabla V_{cr} \times \boldsymbol{k}) \cdot \boldsymbol{\sigma}_P \,. \tag{3.22}$$

Then, after applying the perturbation theory, total Hamiltonian matrix reads:

$$\boldsymbol{H} = \boldsymbol{H_{kp}} + \boldsymbol{H}_{SO,\mathbf{p}} + \boldsymbol{H}_{SO,\mathbf{k}} \,, \tag{3.23}$$

where the last term due to the interaction between p-type states vanishes [7], and thus, the total SO matrix can be approximated as $H_{SO} \approx H_{SO,\mathbf{p}}$.

The basis states, accounting for spin, are $|X\uparrow\rangle$, $|Y\uparrow\rangle$, $|Z\uparrow\rangle$, $|X\downarrow\rangle$, $|Y\downarrow\rangle$, and $|Z\downarrow\rangle$. Each of these basis functions does not couple with opposite spin basis functions by means of the $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ operator, and the spin contribution depends strictly on the contribution of the SO operator. Thus, the total six band $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ matrix accounting for SO

coupling can be written as

$$H_{\mathbf{kp},6} = \begin{pmatrix} H_{\mathbf{kp},3} & 0 \\ 0 & H_{\mathbf{kp},3} \end{pmatrix} + H_{\mathrm{SO}}, \tag{3.24}$$

where $H_{\mathbf{kp},3}$ is the $\boldsymbol{k}\cdot\boldsymbol{p}$ matrix for the three band $\boldsymbol{k}\cdot\boldsymbol{p}$ method (Eq. (3.19)). However, the spin orbit perturbation couples up and down spin states, thereby yielding the SO matrix:

$$H_{\mathrm{SO}} = \frac{\Delta_{\mathrm{SO}}}{3} \begin{pmatrix} 0 & -i & 0 & 0 & 0 & 1 \\ i & 0 & 0 & 0 & 0 & -i \\ 0 & 0 & 0 & -1 & i & 0 \\ 0 & 0 & -1 & 0 & i & 0 \\ 0 & 0 & -i & -i & 0 & 0 \\ 1 & i & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{3.25}$$

where Δ_{SO} is the spin orbit splitting, an intrinsic parameter of the material under study that can be experimentally measured.

In order to show the effect of the SO coupling, the total $\boldsymbol{H}$ matrix can be expressed in a different basis, for which H_{SO} is diagonal. For this basis there are two eigenvalues: a fourfold degenerate $\Delta_{\mathrm{SO}}/3$ and a twofold degenerate $-2\Delta_{\mathrm{SO}}/3$. Therefore, the valence band edge is shifted from E_{v}', the energy of the p-type states that coincides with the valence band edge in the three band $\boldsymbol{k}\cdot\boldsymbol{p}$ model, to the actual valence band edge E_{v} after considering SO coupling:

$$E_{\mathrm{v}} = E_{\mathrm{v}}' + \frac{\Delta_{\mathrm{SO}}}{3}. \tag{3.26}$$

This relationship is of major importance, since the actual valence band edge E_{v}, which can be determined empirically, does not lie on the position introduced by the three band Hamiltonian E_{v}', but it is lifted $\Delta_{\mathrm{SO}}/3$.

The valence bands now can be classified in three types with a twofold spin degeneracy each one: heavy holes (HH) and light holes (LH), which are degenerate in the Γ point, and spin orbit band (SO), which lies Δ_{SO} below HH and LH. This scenario is schematically depicted in Fig. 3.1.

The eigenvectors of H_{SO} have the physical meaning that they form the basis for the total angular momentum [129]. The new basis as a function of the p-states is:
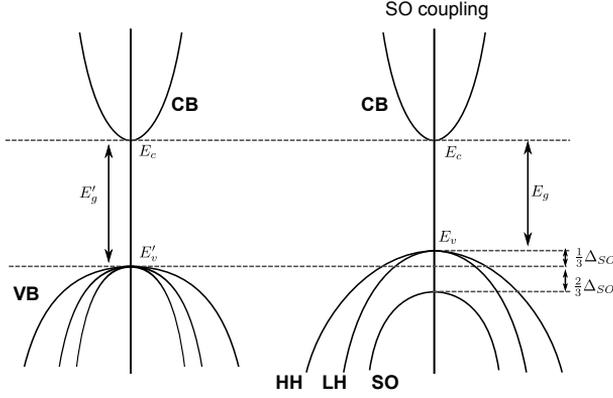
Figure 3.1: Band representation obtained with the $\boldsymbol{k}\cdot\boldsymbol{p}$ method for an arbitrary orientation in the reciprocal space: before considering SO coupling (left) and after considering it (right). The degeneration in the Γ point is partially removed, and the valence bands are split into: degenerate HH, LH; and SO, which lies a Δ_{SO} energy below the previous ones.

$$|HH\uparrow\rangle = -\frac{1}{\sqrt{2}}(|X\uparrow\rangle + \mathrm{i}\,|Y\uparrow\rangle),$$

$$|HH\downarrow\rangle = \frac{1}{\sqrt{2}}(|X\downarrow\rangle - \mathrm{i}\,|Y\downarrow\rangle),$$

$$|LH\uparrow\rangle = -\frac{1}{\sqrt{6}}(|X\downarrow\rangle + \mathrm{i}\,|Y\downarrow\rangle - 2\,|Z\uparrow\rangle),$$

$$|LH\downarrow\rangle = \frac{1}{\sqrt{6}}(|X\uparrow\rangle - \mathrm{i}\,|Y\uparrow\rangle + 2\,|Z\downarrow\rangle),$$

$$|SO\uparrow\rangle = -\frac{1}{\sqrt{3}}(|X\downarrow\rangle + \mathrm{i}\,|Y\downarrow\rangle + |Z\uparrow\rangle),$$

$$|SO\downarrow\rangle = \frac{1}{\sqrt{3}}(|X\uparrow\rangle - \mathrm{i}\,|Y\uparrow\rangle - |Z\downarrow\rangle).$$

(3.27)

From Eq. (3.27), the rotation matrix $\boldsymbol{R}_{\text{SO}}$ between these two bases can be readily obtained. Therefore, a Hamiltonian with a different set of parameters, usually labeled as P, Q, R, and S [129], can be constructed. This set of parameters is widely used in the literature, since it allows to express the wave functions as a linear combination of physically meaningful states, the spin states for the valence bands, i.e, the HH, LH and SO states.

There is another useful set of parameters, the Luttinger parameters, that are related to L, M and N, as:

$$-\frac{\hbar}{2m_0}\gamma_1 = \frac{1}{3}(L + 2M)\,, \tag{3.28}$$

$$-\frac{\hbar}{2m_0}\gamma_2 = \frac{1}{6}(L - M)\,, \tag{3.29}$$

$$-\frac{\hbar}{2m_0}\gamma_3 = \frac{N}{6}\,. \tag{3.30}$$

The Eqs. (3.28-3.30) are valid for using the L, M, and N parameters as defined in this manuscript[3] in Eqs. (3.15), (3.16) and (3.18).

This alternative parameter set is usually provided in the literature, and hence, it is important to know its relationship with the rest of parameters presented in this manuscript. These parameters are related to the carrier effective masses along different orientations and, therefore, can be extracted from experimental measurements of the effective masses. To do it, the following relationships between HH effective mass for two different orientations and the Luttinger parameters can be used [106, 129]:

$$\left(\frac{m_0}{m^*_{\text{HH}}}\right)^{[001]} = \gamma_1 - 2\gamma_2\,, \tag{3.31}$$

$$\left(\frac{m_0}{m^*_{\text{HH}}}\right)^{[011]} = \gamma_1 - \sqrt{\gamma_2^2 + 3\gamma_3^2}\,. \tag{3.32}$$

Whereas, the LH effective masses are related to the Luttinger parameters using:

$$\left(\frac{m_0}{m^*_{\text{LH}}}\right)^{[001]} = \gamma_1 + 2\gamma_2\,, \tag{3.33}$$

$$\left(\frac{m_0}{m^*_{\text{LH}}}\right)^{[011]} = \gamma_1 + \sqrt{\gamma_2^2 + 3\gamma_3^2}\,. \tag{3.34}$$

3.3 The $\boldsymbol{k}\cdot\boldsymbol{p}$ method for direct semiconductors: Kane's model

In the previous Section we have considered that the valence band can be calculated in a basis of p type states, considering the rest of states as remote interactions. This model is appropriate for large band gaps, since the contributions of remote states to the perturbation is usually small. However, for small direct gap semiconductors, as

---

[3]Let us recall that a different definition of L and M can be found which does not account for the free electron dispersion implicitly.

it is the case of narrow gap III-V semiconductors, the coupling between the valence and the conduction bands is strong. Among these semiconductors, we are focusing on those presenting a zinc blende structure, which present the minimum of the conduction band at the Γ symmetry point. These materials have been widely studied due to their interest for optical and high frequency applications, and the $\boldsymbol{k \cdot p}$ method has been usually employed to calculate their main features.

The inclusion of the conduction band, composed by s-type states $|S\rangle$, is recommended to achieve an accurate representation of the conduction and valence bands around the Γ point. Hence, this approach is known as the eight band $\boldsymbol{k \cdot p}$ method.

Now, according to the Löwdin perturbation theory, which is described in Appendix B, the $|S\uparrow\rangle$ and $|S\downarrow\rangle$ states are accounted for directly as A class states, introducing their self-interaction and the interaction with the three main p-states directly in the Hamiltonian. If we ignore the SO coupling at this point, a 4×4 submatrix $\boldsymbol{H_{kp,4}}$ can be constructed, in a similar way to the three band $\boldsymbol{k \cdot p}$ method, by adding a new row and column for the $|S\rangle$ state [7]:

$$\boldsymbol{H_{kp,4}} = \tag{3.35}$$

$$
\begin{pmatrix}
 & |S\rangle & |X\rangle & |Y\rangle & |Z\rangle \\
\hline
|S\rangle & E_c + \tilde{A}_c k^2 & iPk_x + Bk_y k_z & iPk_y + Bk_x k_z & iPk_z + Bk_x k_y \\
|X\rangle & -iPk_x + Bk_y k_z & E_v' + \tilde{L}k_x^2 + \tilde{M}(k_y^2 + k_z^2) & \tilde{N}k_x k_y & \tilde{N}k_x k_z \\
|Y\rangle & -iPk_y + Bk_x k_z & \tilde{N}k_y k_x & E_v' + \tilde{L}k_y^2 + \tilde{M}(k_z^2 + k_x^2) & \tilde{N}k_y k_z \\
|Z\rangle & -iPk_z + Bk_x k_y & \tilde{N}k_z k_x & \tilde{N}k_z k_y & E_v' + \tilde{L}k_z^2 + \tilde{M}(k_x^2 + k_y^2)
\end{pmatrix}
$$

Here, we would like to highlight some aspects of the previous Hamiltonian. First, the tilde in the parameters is used to remark they are different from the analogous parameters in the six band $\boldsymbol{k \cdot p}$ model, and to keep coherence with the notation in [139].

The terms coupling the conduction and the valence band in Eq. (3.35) encompass a linear term with $\boldsymbol{k}$ and a quadratic one. The linear term is the first order term in the Löwdin perturbation theory, which is only non zero for the momentum applied in the direction of the basis state. It is characterized by the momentum interband matrix element P,

$$P = -i\frac{\hbar}{m_0} \langle S | \hat{p}_x | X \rangle , \tag{3.36}$$

which is usually given in terms of the optical energy parameter E_P

$$E_\mathrm{P} = \frac{2m_0}{\hbar^2} P^2 \,. \tag{3.37}$$

Regarding the second order term of the conduction and valence bands coupling, governed by the parameter B, it vanishes for diamond structures. However, this is not the case for zinc blende structures as they do not show inversion symmetry [7]. Nevertheless, it is usually neglected [139] and, furthermore, its value is not easily found in the literature. Hence, we also consider it negligible.

As for the term in position (1,1), it corresponds to the self-coupling of the $|S\rangle$ state, and involves the conduction band parameter $\tilde{A}_\mathrm{c}$. In a single band perturbation theory, with the rest of remote subbands being neglected, this term would be directly related to the electron effective mass, as in Eq. (3.10). However, in this model, the p-type states are accounted for explicitly, thus their effect must be subtracted from the parameter.

The conduction band effective mass is related to $\tilde{A}_\mathrm{c}$ as:

$$\left(\frac{m_0}{m_\mathrm{el}^*}\right) = A_\mathrm{c} = \tilde{A}_\mathrm{c} + \frac{P^2}{E_\mathrm{g}'} \tag{3.38}$$

It is usual to find an alternative parameter F_c in the literature [143] instead of $\tilde{A}_\mathrm{c}$, which stands for the coupling between $\tilde{A}_\mathrm{c}$ and the remote, class B states:

$$F_\mathrm{c} = \frac{1}{m_0} \sum_\beta^B \frac{|\langle S \,|\, \hat{\mathrm{p}}_\mathrm{x} \,|\, u_r \rangle|^2}{E_\mathrm{c} - E_\beta} \,, \tag{3.39}$$

with

$$\tilde{A}_\mathrm{c} = \frac{\hbar^2}{2m_0} (1 + 2F_\mathrm{c}) \,. \tag{3.40}$$

The rest of the parameters, corresponding to the coupling between the p-type states, are similar to those found in the six band model, although in this case the coupling with the $|S\rangle$ must not be included. Therefore, the set of parameters corresponding to the six band approach have to be renormalized to be used in the eight band Hamiltonian [139]:

$$L = \tilde{L} + \frac{P^2}{E_\mathrm{g}'} \,, \tag{3.41}$$

$$M = \tilde{M} \,, \tag{3.42}$$

$$N = \tilde{N} + \frac{P^2}{E_\mathrm{g}'} \,. \tag{3.43}$$

Observe that $E_\mathrm{g}' = E_\mathrm{c} - E_\mathrm{v}'$ is not the real band gap, but the difference with the energy corresponding to the p-type states. Thus, E_g' is given by

$$E_\mathrm{g}' = E_\mathrm{g} + \frac{\Delta_\mathrm{SO}}{3} \,. \tag{3.44}$$

Regarding the SO coupling, it is not necessary to add new terms as the s-type states do not couple with either themselves or the p-type ones. Therefore, the matrix $\boldsymbol{H}_\mathrm{SO}$ only has to be resized[4], adding zeros to the rows and columns corresponding to the s-states.

Then, the final form of the whole Hamiltonian, accounting for the SO coupling is

$$\boldsymbol{H}_{\mathbf{kp},8} = \begin{pmatrix} \boldsymbol{H}_{\mathbf{kp},4} & 0 \\ 0 & \boldsymbol{H}_{\mathbf{kp},4} \end{pmatrix} + \boldsymbol{H}_\mathrm{SO} \,, \tag{3.45}$$

3.4 Strain for the valence band

As mentioned in Chapter 1, strain engineering has had a major relevance in the latest CMOS technology nodes, since it has been able to improve the device performance. Furthermore, when dealing with heterostructures, strain might appear spontaneously as a consequence of the lattice mismatch between materials [135]. Thus, strain plays a very important role in the device design, and it is of major importance to account for it in the models.

One of the strengths of the $\boldsymbol{k \cdot p}$ method is that strain can be easily accounted for as an additional perturbation. In this Section we briefly explain the main concepts related to strain and how to account for it using the $\boldsymbol{k \cdot p}$ method. The topic is further addressed in many publications, such as [17, 129, 130, 135].

Strain is produced when the lattice constant of a crystalline semiconductor is modified by a technological process. A common example is achieved in an epitaxial process,

---

[4]The term $\boldsymbol{H}_{\mathrm{SO},\mathbf{k}}$ is also neglected for the zinc blende semiconductors

when the overlayer and the substrate have different lattice constants. Assuming small deformations, the previously developed $\boldsymbol{k}\cdot\boldsymbol{p}$ models are still valid and the effect of the strain can be included making use of the perturbation theory. In this respect, the $\boldsymbol{k}\cdot\boldsymbol{p}$ approach is able to translate the strain effect into the band structrure.

The deformation of the solid leads to the creation of a force per unit area or stress. The Hooke's law states that the relationship between the stress σ and the strain ϵ is linear, and determined by the tensor formed by the elastic stiffness constants C. For cubic semiconductors, such as diamond and zinc blende ones, many of the terms of C vanish, and the relationship between the stress and the strain reduces to:

$$
\begin{pmatrix} \sigma_{\mathrm{xx}} \\ \sigma_{\mathrm{yy}} \\ \sigma_{\mathrm{zz}} \\ \sigma_{\mathrm{yz}} \\ \sigma_{\mathrm{xz}} \\ \sigma_{\mathrm{xy}} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{pmatrix} \cdot \begin{pmatrix} \epsilon_{\mathrm{xx}} \\ \epsilon_{\mathrm{yy}} \\ \epsilon_{\mathrm{zz}} \\ 2\epsilon_{\mathrm{yz}} \\ 2\epsilon_{\mathrm{xz}} \\ 2\epsilon_{\mathrm{xy}} \end{pmatrix} . \tag{3.46}
$$

According to the linear deformation potential theory, originally developed by *Bardeen* [10] and generalized by *Herring* [55], the strained Hamiltonian can be written by summing an additional term dependent on ϵ, the so-called strain interaction matrix:

$$
[H_{\mathrm{st}}(\epsilon)]_{ij} = \sum_{\alpha,\beta}^{\mathrm{x,y,z}} D_{ij}^{\alpha\beta} \epsilon_{\alpha\beta} \tag{3.47}
$$

being $D_{ij}^{\alpha\beta}$ the deformation potential constants. Therefore, the strain tensor ϵ must be calculated from the stress σ in the lattice, resulting from the external forces (see Appendix C for details). Then, the deformation potential constants are used to define a strain interaction matrix that is added to the Hamiltonian.

Particularly, for the three band $\boldsymbol{k}\cdot\boldsymbol{p}$ theory without spin orbit coupling, the strain interaction matrix can be expressed as:

$$
\boldsymbol{H}_{\mathrm{st}} = \begin{pmatrix} D_{\mathrm{l}}\,\epsilon_{\mathrm{xx}} + D_{\mathrm{m}}(\,\epsilon_{\mathrm{yy}} + \epsilon_{\mathrm{zz}}) & D_{\mathrm{n}}\,\epsilon_{\mathrm{xy}} & D_{\mathrm{n}}\,\epsilon_{\mathrm{xz}} \\ D_{\mathrm{n}}\,\epsilon_{\mathrm{yx}} & D_{\mathrm{l}}\,\epsilon_{\mathrm{yy}} + D_{\mathrm{m}}(\,\epsilon_{\mathrm{xx}} + \epsilon_{\mathrm{zz}}) & D_{\mathrm{n}}\,\epsilon_{\mathrm{yz}} \\ D_{\mathrm{n}}\,\epsilon_{\mathrm{zx}} & D_{\mathrm{n}}\,\epsilon_{\mathrm{zy}} & D_{\mathrm{l}}\,\epsilon_{\mathrm{zz}} + D_{\mathrm{m}}(\,\epsilon_{\mathrm{xx}} + \epsilon_{\mathrm{yy}}) \end{pmatrix},
$$

$$\tag{3.48}$$

where only the reduced set of deformation potentials[5] D_l, D_m and D_n is required because of the crystal symmetry [7]. However, it is easier to find the alternative parameters a_v, b and d in the literature, which depend on the latter as:

$$a_v = \frac{1}{3}(D_l + 2D_m)\,, \qquad b = \frac{1}{3}(D_l - D_m)\,, \qquad d = \frac{1}{\sqrt{3}}D_n\,. \tag{3.49}$$

As can be observed, due to the crystal symmetry, the strain interaction matrix exhibits a similar form to the $\boldsymbol{k\cdot p}$ Hamiltonian given by Eq. (3.19).

If the conduction band is accounted for, the strain interaction matrix must be expanded using the additional deformation potentials[6] D_a and D_b as [7]:

$$\boldsymbol{H}_{st} = \tag{3.50}$$

$$\begin{pmatrix} D_a(\epsilon_{xx}+\epsilon_{yy}+\epsilon_{zz}) & D_b\,\epsilon_{yz} - iP\sum_j \epsilon_{xj}k_j & D_b\,\epsilon_{xz} - iP\sum_j \epsilon_{yj}k_j & D_b\,\epsilon_{xy} - iP\sum_j \epsilon_{zj}k_j \\ D_b\,\epsilon_{yz} + iP\sum_j \epsilon_{xj}k_j & D_l\,\epsilon_{xx} + D_m\,(\epsilon_{yy}+\epsilon_{zz}) & D_n\,\epsilon_{xy} & D_n\,\epsilon_{xz} \\ D_b\,\epsilon_{xz} + iP\sum_j \epsilon_{yj}k_j & D_n\,\epsilon_{yx} & D_l\,\epsilon_{yy} + D_m\,(\epsilon_{xx}+\epsilon_{zz}) & D_n\,\epsilon_{yz} \\ D_b\,\epsilon_{xy} + iP\sum_j \epsilon_{zj}k_j & D_n\,\epsilon_{zx} & D_n\,\epsilon_{zy} & D_l\,\epsilon_{zz} + D_m\,(\epsilon_{xx}+\epsilon_{yy}) \end{pmatrix}$$

In this case, the CB and the VB are coupled with two terms. The first one depends on the shear deformation potential constant D_b, which is zero for diamond structures, and is usually ignored for zinc blende ones. For this reason it is not usually tabulated in the parameter collections, e.g., in Ref. [143]. The second term depends linearly on the wave vector, and it is also usually ignored in the literature [129, 135, 139], assuming that the wave vector is small enough. Thus, the effect of the strain in the coupling of the CB and the VB can be neglected.

Finally, to account for the spin orbit interaction, the up and down spin states must be considered, thus leading to an strain matrix interaction term for the spin orbit coupling, as shown in [7]. However, the usual model also neglects this interaction and then, the total spin orbit interaction matrix is formed as follows, for both six and eight band $\boldsymbol{k\cdot p}$ methods:

$$\boldsymbol{H}_{st,so} = \begin{pmatrix} \boldsymbol{H}_{st} & 0 \\ 0 & \boldsymbol{H}_{st} \end{pmatrix} \tag{3.51}$$

[5]These parameters are usually denoted as l, m, and n because of the analogy with the $\boldsymbol{k\cdot p}$ parameters. However we use D to keep the coherence with the denomination of the deformation parameters.
[6]As previously, we use denote them as D to keep coherence with our notation. These potentials are usually denoted as a_c and b' [7, 143].

3.5 Two band $\boldsymbol{k}\cdot\boldsymbol{p}$ method for electrons in Si

A different $\boldsymbol{k}\cdot\boldsymbol{p}$ model is used to calculate the conduction band in Si. In this material, the edge of the conduction band is not in the Γ point but in the Δ direction, near the X point at a distance $\boldsymbol{k}_0 = 0.15(2\pi/a_0)$, denoted as Δ_1 valley. Due to the periodicity of the reciprocal space, the valley is replicated symmetrically with respect to the X point, where it is degenerate, thereby defining a second valley, placed in $+\boldsymbol{k}_0$ and denoted Δ_2'. These valleys can interact and mix when external perturbations, such as strain, are applied. Therefore, using these two bands it is possible to apply the $\boldsymbol{k}\cdot\boldsymbol{p}$ method to get the following expression for the two band $\boldsymbol{k}\cdot\boldsymbol{p}$ matrix [54]

$$\boldsymbol{H}_{\mathbf{kp},2} = \begin{pmatrix} H_+ & H_{\mathrm{bc}} \\ H_{\mathrm{bc}} & H_- \end{pmatrix}, \tag{3.52}$$

where the matrix elements of the diagonal $H_\pm$ and the coupling terms H_{bc} are

$$H_\pm = E_c + \frac{\hbar^2 k_z^2}{2m_{\mathrm{l}}} + \frac{\hbar^2 \left(k_x^2 + k_y^2\right)}{2m_{\mathrm{t}}} \pm \frac{\hbar^2 k_0 k_z}{2m_{\mathrm{l}}}, \tag{3.53}$$

$$H_{\mathrm{bc}} = \frac{\hbar^2 k_x k_y}{M}. \tag{3.54}$$

with the + and - sign denoting H_+ and H_-, respectively. These elements describe the two parabolic bands, with their corresponding longitudinal m_{l} and transversal m_{t} effective masses and the parameter M is given by

$$\frac{1}{M} = \frac{2}{m_0^2} \left| \sum_{l\neq1,2} \frac{\langle \Delta_1 | \hat{\mathrm{p}}_x | l \rangle \langle l | \hat{\mathrm{p}}_x | \Delta_{2'} \rangle}{E_{\Delta_{2'}} - E_{\Delta_1}} \right|, \tag{3.55}$$

and is usually approximated as:

$$\frac{1}{M} \simeq \frac{1}{m_{\mathrm{t}}} - \frac{1}{m_0}. \tag{3.56}$$

The resulting Hamiltonian can be completed by taking into account the effect of strain. This only requires adding the shifting of the conduction band $D_{\mathrm{u}}\epsilon_{zz}$ to the diagonal elements, and the shear strain effect $2D_{\mathrm{sh}}\epsilon_{xy}$ to the non diagonal elements, where D_{u} and D_{sh} are the uniaxial and shear deformation potentials for the conduction[7] band of Si and the parameters $\epsilon_{\alpha\beta}$ are the crystal deformations due to the applied strain.

---

[7]They can usually be found in the literature as Ξ_u and Ξ_u' [126, 135].

3.6 $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ envelope function approximation

Burt [21–23] carried out an extensive work to demonstrate that the exact envelope equations can be derived from the Schrödinger equation. However, Burt's theory is not widely adopted as the envelope function approximation is preferred by most of the authors. An interesting analysis of the derivation of the exact envelope function and the assumptions made to reach the approximated theory is developed in [139]. The main concerns on this approximation come from: 1) those terms related to the interfaces between materials, where the periodicity of the potential is broken, and 2) the operator ordering. By neglecting the interface terms, which are rarely considered in literature [44], the envelope function approximation is generally valid. However, the operator ordering still remains as a major concern for the numerical stability of the envelope equations [140, 142]. Foreman [43] firstly described the correct operator ordering for zinc blende crystals, which is commonly referred to as Burt-Foreman operator ordering.

Then, the envelope function approximation, introduced in Sec. 2.4, can be used in the $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ approach. As a result of replacing the wave vector $\boldsymbol{k}$ by the *momentum-like* operator[8] $\hat{\boldsymbol{k}}$ (Eq. (2.16)) applied to the envelope functions, the $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ Hamiltonian becomes a matrix of operators. This matrix does not exhibit an unique form since the operators may be ordered in different ways. However, it is essential to assure the hermiticity in order not to obtain imaginary energies as solutions. The general form of the Hamiltonian is thus of the form

$$
\begin{aligned}
\hat{\mathbf{H}}_{\mathbf{kp}} =& \boldsymbol{H}^{(0)}(\mathbf{r};\boldsymbol{k}) \\
&+ \sum_i \left(\boldsymbol{H}^{(1)}_{i;L}(\mathbf{r};\boldsymbol{k})\frac{\partial}{\partial i} + \frac{\partial}{\partial i}\boldsymbol{H}^{(1)}_{i;R}(\mathbf{r};\boldsymbol{k}) \right) \\
&- \sum_{i,j} \frac{\partial}{\partial i}\boldsymbol{H}^{(2)}_{ij}(\mathbf{r};\boldsymbol{k})\frac{\partial}{\partial j}\,,
\end{aligned}
\tag{3.57}
$$

where i and j are the spatial coordinates along the confined directions. The total $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ matrix is then constructed using matrices $\boldsymbol{H}^{(0)}$, $\boldsymbol{H}^{(1)}_{i;L}$, $\boldsymbol{H}^{(1)}_{i;R}$, and $\boldsymbol{H}^{(2)}_{ij}$, corresponding to the zero-th, first and second order momentum operators. These matrices account for the $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ parameters, which may depend on the position, and the wave vectors along

---

[8]The so-called *momentum-like operator* differs from the momentum operator $\hat{\mathbf{p}}$ only in a $\hbar$ constant. Hereinafter, when we refer to the momentum operator in the context of the EFA for the $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ method, we actually mean $\hat{\mathbf{k}} = (\hat{\mathbf{k}}_x, \hat{\mathbf{k}}_y, \hat{\mathbf{k}}_z)$, acting on the envelope function. The missed $\hbar$ is intrinsically assimilated in the parameters.

the non confined directions $\boldsymbol{k}$. Using this general expression, the hermiticity of the operator is assured if the following conditions are fulfilled:

$$\boldsymbol{H}^{(0)} = \left(\boldsymbol{H}^{(0)}\right)^{\dagger},\tag{3.58}$$

$$\boldsymbol{H}^{(1)}_{i;L} = \left(\boldsymbol{H}^{(1)}_{i;R}\right)^{\dagger},\tag{3.59}$$

$$\boldsymbol{H}^{(2)}_{ij} = \left(\boldsymbol{H}^{(2)}_{ji}\right)^{\dagger}.\tag{3.60}$$

Any ordering fulfilling these conditions is physically meaningful since it provides a set of real energies. However, as aforementioned, the ordering can be crucial for the numerical stability of the problem, mainly in the eight-band $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ model. So that, we are focusing in describing the operator ordering for such model. The six band approach can be readily extracted from the eight band approach by neglecting the dependence on the conduction band.

The derivation of the eight band approach stems from the four band bulk Hamiltonian in Eq. (3.35). Some modifications must be performed to assure the hermiticity of the matrix operator adapting it to the form shown in Eq. (3.57). The resulting matrix operator accounting for the correct operator ordering is shown in Eq. (3.63).

In this representation, the arbitrary symmetrization of the operators shown in Eq. (2.20) is not used. Instead, the Burt-Foreman symmetrization has been used in the coupling terms of the p-type states, i.e., the terms related to the N parameter in the bulk approximation. Actually, this parameter naturally lies in the contribution of different sets of remote functions [82], giving rise to two different parameters N_+ and N_- related to two different operator orderings, e.g., $\hat{k}_x N_+ \hat{k}_y + \hat{k}_y N_- \hat{k}_x$ [44]. Since these parameters are not directly accessible from the Luttinger ones, Eqs. (3.28-3.30), they can be derived from the rest as:

$$N_- = M - \frac{\hbar^2}{2m_0},\tag{3.61}$$

$$N_+ = N - N_-.\tag{3.62}$$

For those elements coupling the CB and the VB, the ordering of the first order terms, e.g. $iP\hat{k}_x$ and $-i\hat{k}_x P$, is an ad-hoc choice to assure the hermiticity of the matrix, in particular when P varies in an interface. The symmetric ordering of the second order term, $\hat{k}_y B\hat{k}_z + \hat{k}_z B\hat{k}_y$, is derived directly from the symmetry of the problem [139]. However, as aforementioned, B is usually negligible and dropped from the Hamiltonian.

$$\hat{H}_{\mathbf{kp},4} = \tag{3.63}$$

$$
\begin{pmatrix}
E_{\mathrm{c}} + \boldsymbol{k}\tilde{A}_{\mathrm{c}}\boldsymbol{k} & iP\hat{k}_{\mathrm{x}} + \hat{k}_{\mathrm{y}}B\hat{k}_{\mathrm{z}} + k_{\mathrm{z}}B\hat{k}_{\mathrm{y}} & iP\hat{k}_{\mathrm{y}} + \hat{k}_{\mathrm{x}}B\hat{k}_{\mathrm{z}} + \hat{k}_{\mathrm{z}}B\hat{k}_{\mathrm{x}} & iP\hat{k}_{\mathrm{z}} + \hat{k}_{\mathrm{x}}B\hat{k}_{\mathrm{y}} + \hat{k}_{\mathrm{y}}B\hat{k}_{\mathrm{x}} \\
-i\hat{k}_{\mathrm{x}}P + \hat{k}_{\mathrm{y}}B\hat{k}_{\mathrm{z}} + \hat{k}_{\mathrm{z}}B\hat{k}_{\mathrm{y}} & E_{\mathrm{v}}' + \hat{k}_{\mathrm{x}}\tilde{L}\hat{k}_{\mathrm{x}} + \hat{k}_{\mathrm{y}}\tilde{M}\hat{k}_{\mathrm{y}} + \hat{k}_{\mathrm{z}}\tilde{M}\hat{k}_{\mathrm{z}} & \hat{k}_{\mathrm{x}}\tilde{N}_{+}\hat{k}_{\mathrm{y}} + \hat{k}_{\mathrm{y}}\tilde{N}_{-}\hat{k}_{\mathrm{x}} & \hat{k}_{\mathrm{x}}\tilde{N}_{+}\hat{k}_{\mathrm{z}} + \hat{k}_{\mathrm{z}}\tilde{N}_{-}\hat{k}_{\mathrm{x}} \\
-i\hat{k}_{\mathrm{y}}P + B\hat{k}_{\mathrm{x}}B\hat{k}_{\mathrm{z}} + \hat{k}_{\mathrm{z}}B\hat{k}_{\mathrm{x}} & \hat{k}_{\mathrm{y}}\tilde{N}_{+}\hat{k}_{\mathrm{x}} + \hat{k}_{\mathrm{x}}\tilde{N}_{-}\hat{k}_{\mathrm{y}} & E_{\mathrm{v}}' + \hat{k}_{\mathrm{y}}\tilde{L}\hat{k}_{\mathrm{y}} + \hat{k}_{\mathrm{x}}\tilde{M}\hat{k}_{\mathrm{x}} + \hat{k}_{\mathrm{z}}\tilde{M}\hat{k}_{\mathrm{z}} & \hat{k}_{\mathrm{y}}\tilde{N}_{+}\hat{k}_{\mathrm{z}} + \hat{k}_{\mathrm{z}}\tilde{N}_{-}\hat{k}_{\mathrm{y}} \\
-i\hat{k}_{\mathrm{z}}P + B\hat{k}_{\mathrm{x}}B\hat{k}_{\mathrm{y}} + \hat{k}_{\mathrm{y}}B\hat{k}_{\mathrm{x}} & \hat{k}_{\mathrm{z}}\tilde{N}_{+}\hat{k}_{\mathrm{x}} + \hat{k}_{\mathrm{x}}\tilde{N}_{-}\hat{k}_{\mathrm{z}} & \hat{k}_{\mathrm{z}}\tilde{N}_{+}\hat{k}_{\mathrm{y}} + \hat{k}_{\mathrm{y}}\tilde{N}_{-}\hat{k}_{\mathrm{z}} & E_{\mathrm{v}}' + \hat{k}_{\mathrm{z}}\tilde{L}\hat{k}_{\mathrm{z}} + \hat{k}_{\mathrm{x}}\tilde{M}\hat{k}_{\mathrm{x}} + \hat{k}_{\mathrm{y}}\tilde{M}\hat{k}_{\mathrm{y}}
\end{pmatrix}
$$

$$
\boldsymbol{H}_{\mathrm{SO},\hat{\mathbf{p}}} = \frac{\Delta_{\mathrm{SO}}}{3}
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -i & 0 & 0 & 0 & 0 & 1 \\
0 & i & 0 & 0 & 0 & 0 & 0 & -i \\
0 & 0 & 0 & 0 & 0 & -1 & i & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & i & 0 \\
0 & 0 & 0 & -i & 0 & -i & 0 & 0 \\
0 & 1 & i & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\tag{3.64}
$$

$$
\hat{\mathbf{H}} = \begin{pmatrix} \hat{\mathbf{H}}_{\mathbf{kp},4} & 0 \\ 0 & \hat{\mathbf{H}}_{\mathbf{kp},4} \end{pmatrix} + \boldsymbol{H}_{\mathrm{SO}} \tag{3.65}
$$

3.7 Spurious solutions

In the previous Section we have shown how to modify the intrinsically bulk $\boldsymbol{k}\cdot\boldsymbol{p}$ method to a confined system. Thus, this method provides a procedure to calculate the band structure of devices composed by different semiconductor materials, i.e., heterostructures. However, the numerical implementation of this method is sometimes affected by a serious problem, the appearance of spurious solutions, which was firstly reported for the eight band $\boldsymbol{k}\cdot\boldsymbol{p}$ method in [119, 145]. This issue consists on the appearance of solutions that do not have a physical interpretation. Some examples are subbands warping in the wrong direction (CB bending downwards or VB bending upwards), subbands located inside the semiconductor gap, highly oscillatory wave functions that violate the assumption of a slowly varying envelope function, or wave functions that lead to a charge distribution located at forbidden regions. Some mathematical artifices have been proposed to overcome this issue and to obtain a set spurious-free solutions. We can mention a few of them: adding parameters to the Hamiltonian [58], imposing special boundary conditions [113], filtering out any out-of-zone components [49] or simply discarding the suspicious solutions.

A thorough study based on the mathematical nature of the problem is performed by Veprek et al. in Refs. [139–141], who attribute the spurious solutions to a wrong selection of the parameters, and not to the numerical method employed. The goal is to assure the ellipticity of the $\boldsymbol{k}\cdot\boldsymbol{p}$ method through a careful selection of the $\boldsymbol{k}\cdot\boldsymbol{p}$ parameters, thereby achieving the expected band behaviour.

The Burt-Foreman operator ordering intrinsically provides values to the parameters close to fulfill the ellipticity of the operator, thus reducing the incidence of the problem. However, it does not eliminate it and further efforts must be done to assure the criterion of ellipticity. For the six band $\boldsymbol{k}\cdot\boldsymbol{p}$ method, the ellipticity of the operator is assured if the parameters fulfill [140] the following relations:

$$A_{\mathrm{c}} > 0\,, \quad M - N_- < 0\,, \quad M + N_- < 0\,, \tag{3.66}$$
$$L - N_- < 0\,, \quad L + 2N_+ < 0\,.$$

For the eight band model, in principle it is more difficult to assure the ellipticity since the valence and the conduction bands are considered at the same time. However, as the second order terms coupling the CB and the VB are negligible compared to the rest, the ellipticity can be assured by achieving the ellipticity for the CB and the VB

independently. Therefore, the conditions imposed in Eq. (3.66) must be fulfilled for the analogous renormalized parameters $\tilde{A}_c$, $\tilde{L}$, $\tilde{M}$, $\tilde{N}_+$, and $\tilde{N}_-$.

The reason for the greater incidence of spurious solutions in the eight band model is related to the renormalization of the parameters which, when not being carefully chosen, is more likely to violate the ellipticity criteria.

Therefore, when spurious solutions appear, a reduction of the E_P parameter is suggested. This procedure is not physically justified. However, the E_P value is normally used to fit experimental data and thus it can be tuned [143]. When the rest of the parameters are correctly renormalized, the effective masses resulting from these parameters do not change, and therefore the results are still reliable. Obviously, the band structure away from the Γ point depends on the parameters chosen. Further studies are required for a better adjustment of the parameters through the comparison with either experimental data or other methods. However, we can assume that the main electrostatic features will not change drastically (subband energies, effective masses, density of states, and envelope functions), leading to an accurate description of the device behavior.

3.8 Parameter calculation in alloys

An alloy is a mixture of two or more semiconductors yielding a new material with modified properties, such as the band-gap energy and the carrier effective mass, useful to optimize and improve the device performance [15, 60, 96, 97, 149]. These tailored materials also provide an extra degree of freedom in the design of heterostructure devices, since the potential barriers and energy gaps can be accurately designed.

The alloys of interest in this manuscript are $Si_{1-x}Ge_x$, composed by Si and Ge in different molar fraction x, and ternary III-V alloys. These ternary alloys can be viewed as the mixture of two binary III-V compounds which share one of their constituent elements. For example, $Al_xGa_{1-x}As$ results as a composition of AlAs with GaAs, with x being the AlAs molar fraction.

Alloys have a well-defined crystal structure, but the randomness of atoms on the lattice sites breaks down the periodicity in the background crystalline potential. A rigorous treatment of this problem obliges to use the virtual crystal approximation (VCA) [13], where the random potential is replaced by an average periodic potential and the random variation is treated as a perturbation. Under some assumptions, the

$\boldsymbol{k}\cdot\boldsymbol{p}$ method can be employed to calculate the band structure of different compounds. In particular, SiGe is a diamond like semiconductor and the six band $\boldsymbol{k}\cdot\boldsymbol{p}$ method is appropriate to calculate its band structure, whereas the ternary alloys are zinc blende semiconductors that require the implementation of the eight band $\boldsymbol{k}\cdot\boldsymbol{p}$ method when the gap is small.

Since the constituent materials are of similar nature, the properties of the alloys are expected to be a hybridization of them. Therefore, the parameters can be calculated as a function of the parameters of the original materials.

3.8.1 Ternary III-V alloys

The most important parameter of the III-V ternary alloys analyzed in this manuscript is the band gap. The dependence of the energy gap on the alloy composition is generally assumed to fit a simple equation [138]:

$$E_{\mathrm{g}}(\mathrm{A_{1\text{-}x}B_x}) = (1-x)E_{\mathrm{g}}(\mathrm{A}) + xE_{\mathrm{g}}(\mathrm{B}) - x(1-x)b\,, \tag{3.67}$$

where the band gap is interpolated between the extremal values for the two pure binary compounds A and B, using a bowing parameter b to add a quadratic correction. Depending on the binary compound, the conduction band minimum can be set by a different valley; therefore, a gap must be defined for each specific conduction band valley, and the interpolation must be performed accounting for it in both materials.

The general assumption is to consider that a quadratic correction also leads to a correct representation of other modeling parameters [143]. In that case, an expression analogous to Eq. (3.67) can be used. In this manuscript we use the parameters tabulated in Ref. [143]. For those parameters that are not tabulated there, a linear interpolation is employed. Usually, the parameters that can be interpolated, with or without bowing, are those assigned to direct magnitudes, such as the lattice constant, energy gaps, effective masses, or the optical matrix element E_{P}. However, it is not recommended to interpolate indirect parameters such as the Luttinger ones, since they depend inversely on the effective masses.

3.8.2 SiGe

As a rule of thumb, the $Si_{1-x}Ge_x$ parameters are calculated in a similar way to III-V alloys. Therefore, a quadratic interpolation can be carried out. However, further studies of the parameters performed by *Rieger and Vogl* [112] found that the L, M and N parameters could be derived from the pure Si and Ge ones using the following expressions:

$$C(x) = C(0) + \alpha\ln\left(1 - Sx^{\beta}\right), \qquad (3.68)$$

$$S = 1 - \exp(C(1) - C(0)/\alpha), \qquad (3.69)$$

where $C(x)$ stands for any of the three parameters as a function of the Ge molar fraction x. The parameters for pure Si and Ge, and the adjusting parameters α and β are the following:

$$\alpha = 6.7064, \qquad (3.70)$$

$$\beta = 1.35, \qquad (3.71)$$

$$L(0) = -6.69, L(1) = -21.65, \qquad (3.72)$$

$$M(0) = -4.62, M(1) = -5.02, \qquad (3.73)$$

$$L(0) = -8.56, L(1) = -23.48. \qquad (3.74)$$

Let us remark two issues regarding these parameters. First, they are in $\hbar^2/2m_0$ units, so that they must be multiplied by this quantity before using them in the construction of the Hamiltonian. Second, these parameters are defined without considering the free electron dispersion relation. Thus, to be consistent with our notation, it must be added to them, as mentioned in Sec. 3.2.1. Therefore, it is necessary to add 1 (the free electron dispersion is the unity in this normalization) to the L and M parameter after the interpolation carried out using Eq. (3.68).

Chapter 4

SP2D Simulation Tool

In the previous Chapters we have outlined general concepts and equations related to the physics of semiconductor devices. Now we are ready to address the most specific problem that concerns us in this manuscript: the simulation of 2D confined nanodevices. So that, this chapter is devoted to the computational treatment of the equations that have to be solved to achieve an accurate physical description of semiconductor nanowires.

In order to achieve an accurate description of the behavior of the carriers inside the semiconductor nanowires, it will be necessary to deal with several differential and integral equations and systems. In some scenarios, approximations and simplifications of the problems are enough to provide either with a description or an intuition about the device behavior. Unfortunately, as the devices are shrunk, the quantum effects arise, requiring the use of more sophisticated models able to describe the carrier behavior.

The study of realistic devices usually complicates the resolution of the equations as several issues have to be taken into account. Some of them are: 1) different geometries that may not be easily represented in a rectangular coordinate system; 2) different materials, introducing discontinuities in the most important parameters of the problem; 3) boundary conditions that may not be described by analytical expressions. The simplicity required for the analytical treatment is not fulfilled because it is confronted to the complexity needed to properly describe the device underlying physics. Consequently, in most of the cases, the problem can not be analytically solved.

For these reasons, the assistance of computers in the simulation process has played an important role among the nanoelectronic community for many years. The devel-

opment of TCAD (Technology Computer Aided Design) tools has allowed a better understanding of the semiconductor devices and has helped to establish procedures to improve their performance. In this context, different simulation tools have been developed by various companies, such as Sentaurus^{TM} from Synopsis®, Atlas^{TM} from Silvaco®, VSP from Global TCAD Solutions, or Garand from Gold Standard Simulations.

One of the main goals of this thesis deals with the development of a simulation tool fully implemented within our group, SP2D, to analyze 2-D confined structures. Despite the variety of simulators in the market, the decision of implementing a simulation tool is mainly motivated by:

- Absolute control of the code: commercial simulation tools usually are conceived as black boxes with insufficient information about what is happening inside.

- Versatility, adaptability and extensibility: modifications, improvements and extensions can be added to the simulator in order to satisfy the necessities at the time they arise.

- Complete knowledge of the limitations of the solutions: as we choice the models used to solve the problem, we can evaluate each of them in different scenarios. This is of major importance in research, since the limitations of the models and their range of application is sometimes not well defined.

- Economic saving: the licenses of commercial simulators are usually quite expensive.

There are nonetheless trade-offs that have to be taken into account:

- Time consuming: the development process requires a lot of time and effort, which translates into an economic cost. Moreover, the results are not achieved until the whole program is finished.

- Deep knowledge on mathematics and physics: an extensive background in these fields is needed to put down the whole problem and to face all the issues that can emerge.

- General programming skills: the problem is not only related to mathematics or physics, but also requires a deep knowledge about programming.

- No technical support: there is not a supporting staff that can help you to solve the problems found during the programming process.

4.1 TCAD Simulator

The first decision to be made to develop a simulation tool is the choice of the programming tool or language. Different options can be considered: a general purpose language (C++, Python) or a higher level programming tool (MATLAB). We have chosen MATLAB as a high-performance language for technical computing, which is integrated in a whole development environment which easily defines, debugs, solves, and represents the mathematical problems and solutions. The main motivation was the huge library of functions and tools, including algorithms and methods suitable for addressing most of the scientific and engineering problems. Nevertheless, there are many other advantages that make MATLAB a very interesting option for technical computing:

- Extensive and thorough documentation of all the built-in functions and algorithms.

- Clear and straight syntax specifically created for mathematics, and specially for matrix processing. Thus, the learning curve is more steep than for other general purpose languages.

- Easy debugging by using breakpoints. The development environment provides a debugging tool which allows checking line by line for errors.

- The graphical output is optimized for interaction. Data can be easily plotted, and interactive tools are provided for changing colors, sizes, scales, etc, in order to get high quality graphics.

- Technical support from a professional organization.

- Large user community providing help and free code implementations.

For these reasons, MATLAB is a very interesting option during the process of programming the simulation tools. Nevertheless, there are also some disadvantages that were also evaluated:

- It is a proprietary language, subject to licenses. The sharing of the code is restricted to those having a license.

- The functionality of the simulator is subject to the renewal of the license. The simulator will be no longer useful if the license is not renewed.

- The code is not easily portable to other languages. Sometimes it is not even compatible between different MATLAB versions due to the use of specific toolboxes.

- There are restricted functions that work as black boxes.

- It is a platform where the code does not run directly on the operating system but in a virtual machine. This additional layer reduces the efficiency of the codes even though the algorithms are optimized.

In particular, MATLAB includes a very useful toolbox to solve a large variety of numerical problems involving partial differential equations: the PDE Toolbox$^{\text{TM}}$ [2]. It is a built-in set of functions providing the functions needed to define and solve problems that can be described by a system of 2D partial differential equations. The key feature of PDE Toolbox$^{\text{TM}}$ is the discretization method. Instead of using the Finite Differences Method (FDM) for a regular discretization of the equations, PDE Toolbox$^{\text{TM}}$ is based on the Finite Element Method (FEM). The main advantages of FEM over FDM, can be summarized as:

- More flexibility in terms of dealing with complex geometries and thin sections.

- Better treatment of inhomogeneous media.

- Reduced requirements on the regularity or smoothness of the solutions.

- Flexible discretization of the geometry under study.

The PDE toolbox implementation of FEM discretize the equations using a non regular grid composed of triangles which allows a better adaptation to curved complex geometries.

Other remarkable characteristics of PDE Toolbox$^{\text{TM}}$ are:

- Versatile boundary condition specification.

- Flexible non-linear solvers.

- Visualization tools for solutions, mesh and geometries.

The toolbox offers a very complete set of functions able to solve the different parts of the FEM problem at different levels (mesh, algorithms, eigenvalues, visualization, integration, etc.), which are extremely useful during the implementation of the problem. However, the numerical solution of the $\boldsymbol{k \cdot p}$ matrix will produce a coupled system of partial differential equations that precludes the direct use of these functions making necessary a deeper understanding of the problem in order to implement the necessary functions.

4.2   Outline of the SP2D Simulator

In this section we provide a brief description of the SP2D simulator, whose main features will be detailed in subsequent sections. As aforementioned, the simulator deals with semiconductor nanowires designed to work as transistors. These devices are based on the CMOS technology, thus, they present at least three different materials and regions: a metallic gate, a semiconductor channel where the carriers flow and an insulator preventing the carriers to flow from the gate to the channel. As the traditional planar MOSFETs, the nanowire FETs are ended by a source and a drain regions, and the carriers flow from one to the other, controlled by the potential applied to the gate contact. The particularity of the devices under consideration, the nanowire (NW) or Gate All Around (GAA), is that the gate completely surrounds the channel, as shown in Fig. 4.1. In our study we assume that the length of the channel (transport direction) is much larger than the length of the confined directions. So that, the transverse electric field generated by the gate on the channel is much higher than the electric field along the channel, and the gradual channel approximation can be applied. At this point, one can simplify the problem to a cross-section of the device, extracting valuable information of the electrostatics and, later, of the transport properties of the device.

The electrostatic description of the device involves the calculation of the potential and the charge distribution on the cross-section of the nanowire. The charge distribution is determined by: 1) the wave functions that yield the charge distribution, 2) the energy of the states and the relative position of the Fermi level, which determines the probability of finding a carrier in that state. The wave functions and energies are provided by the Schrödinger Equation, which depends on the potential of the structure. Whereas, the potential is obtained from the charge distribution using the Poisson Equation. Therefore, these two equations are coupled, and have to be self-consistently
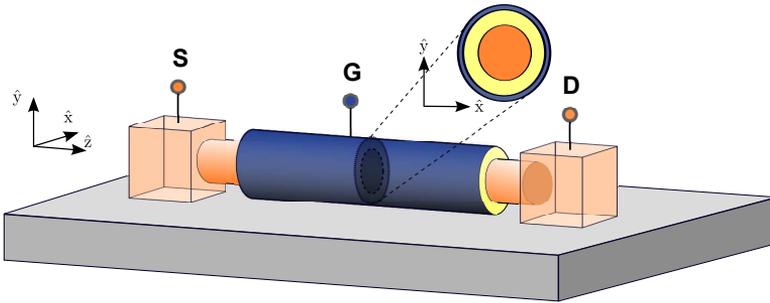
Figure 4.1: 3D schematic view of a SGT device. The insulator is colored in yellow, the semiconductor channel in orange while the gate contact, in blue. The three terminals are pinned as G (gate), D (drain) and S (source).

solved. A typical approach is to consider that these two equations are decoupled and the potential is a flat well. This approximation is very rough since in nanostructures, with a considerable charge density, the potential varies strongly along the structure, affecting to the solutions of the Schrödinger Equation. For this reason, we have decided to take into account the coupling of these two equations by self consistently solving the Schrödinger and Poisson's equations, thereby obtaining more realistic results.

In this manuscript the Schrödinger Equation is solved using the $\bm{k}\cdot\bm{p}$ method, which leads to a more accurate description of the band structure than the EMA, specially for the valence band in diamond and zinc blende semiconductors.

4.3 Confinement of the device

Let us now take an insight into the geometry of the devices. Figure 4.1 depicts a nanowire transistor. As we already know, assuming a weak longitudinal field, the study can be reduced to the section of the device. Therefore, we have defined the following coordinate system, using the device as a reference:

- The x and y coordinates, $\mathbf{r} = (x, y)$, are the coordinates of the cross-sectional plane of the device, where the carriers are confined.

- The z coordinate denotes the transport direction, along which planar waves can travel freely.

In order to solve the device band structure, the $\boldsymbol{k}\cdot\boldsymbol{p}$ method, detailed in Chapter 3, is used. Since the device is confined, the envelope function approximation is used. As seen in Sec. 2.4, the wave function $\phi_{n\boldsymbol{k}}(\boldsymbol{r})$ is fully determined, if neglecting the rapid oscillation of the core, by the envelope function $\boldsymbol{\xi}_{nk_z}(\mathbf{r})$, which only depends on the confined coordinates $\mathbf{r} = (x, y)$. Then, the calculation of the envelope functions corresponding to the $\boldsymbol{k}\cdot\boldsymbol{p}$ method using the EFA, detailed in 3.6, involves the substitution of the wave vectors (k_x, k_y) by the the momentum-like operators [1] $(\hat{k}_x, \hat{k}_y)$, acting on the envelope functions. As a result, the simple $\boldsymbol{k}\cdot\boldsymbol{p}$ matrix turns into a matrix of operators $\hat{\mathbf{H}}(\nabla_{\mathbf{r}}, k_z)$, as explained in Sec. 3.6. The system to be solved is then transformed into a non-linear system of partial differential equations, the solutions of which depend on the confined coordinates x and y:

$$\hat{\boldsymbol{H}}(\mathbf{r}, k_z)\boldsymbol{\xi}_{nk_z}(\mathbf{r}) = E_{nk_z}\boldsymbol{\xi}_{nk_z}(\mathbf{r}). \tag{4.1}$$

The consequence of the confinement is that the carriers are quantized in the $\mathbf{r}$ directions, yielding a subband pattern. As for the resulting envelope function, it is a function of k_z and the position $\mathbf{r}$. This may be interpreted as follows: the orbital solutions used in the $\boldsymbol{k}\cdot\boldsymbol{p}$ method as initial solutions mix differently in each position $\mathbf{r}$, according to the weights of the different components of the envelope function. As the periodicity is not altered along the transport direction, and the envelope function is not a function of the corresponding coordinates, the Blöch Theorem still holds and the dependence on these coordinates keeps the periodicity of the crystal, enveloped by the factor $\exp(ik_z z)$. Nevertheless, considering the envelope function approximation, the oscillation of the crystal solutions are negligible in the calculation of the charge distribution, so that, the envelope function will be enough to determine the charge distribution and the information related to the coupling between states.

4.4 Arbitrary Orientation

Up to here, we have defined coordinates in the Device Reference System (DRS). However, the $\boldsymbol{k}\cdot\boldsymbol{p}$ matrix is referred to the Crystal Reference System (CRS) and it does not have to match necessarily with the principal axes of the device. This issue is treated in this Section. The device orientation determines the device quantization, and the momenta along the different orientations. In contrast, the $\boldsymbol{k}\cdot\boldsymbol{p}$ method is defined in

---

[1]Hereinafter, simply momentum operator.

a different reference system where the lattice exhibits some symmetries, including the anisotropical properties and behavior of the carriers as a function of the wave vectors defined in the CRS. Since the confinement forces us to consider the momenta instead of the wave vector, it is necessary to calculate them in the CRS so as to evaluate the $\boldsymbol{k}\cdot\boldsymbol{p}$ matrix. The direct consequence is that the alignment of the crystal structure with the device is decisive to the carriers behavior, because it determines the periodicity of the lattice along the device transport direction.

To address this problem we must choose the reference system to lay out the equations. The momentum operator in the DRS, the one used so far, can be readily defined as $\hat{\boldsymbol{k}}^{\mathrm{D}} = \hbar(\hat{\mathrm{k}}_{\mathrm{x}}, \hat{\mathrm{k}}_{\mathrm{y}}, k_{\mathrm{z}})^{\mathrm{D}}$. A rotation of this operator must be performed so as to get the momentum in the CRS, which can be substituted in the $\boldsymbol{k}\cdot\boldsymbol{p}$ matrix. Since the momentum operator is linear, it can be rotated using the matrix rotation between the two coordinate systems:

$$\hat{\boldsymbol{k}}^{\mathrm{C}} = \boldsymbol{R}^{\mathrm{CD}}\hat{\boldsymbol{k}}^{\mathrm{D}} \,, \tag{4.2}$$

where $\boldsymbol{R}^{\mathrm{CD}}$ is the rotation matrix from the DRS to the CRS. After performing this rotation, each component of the momentum in the CRS may be described as a mixing of partial derivatives and the wave vector k_{z}.

$$\hat{\mathrm{k}}_{\mathrm{x}}^{\mathrm{C}} = R_{11}^{\mathrm{CD}}\hat{\mathrm{k}}_{\mathrm{x}}^{\mathrm{D}} + R_{12}^{\mathrm{CD}}\hat{\mathrm{k}}_{\mathrm{y}}^{\mathrm{D}} + R_{13}^{\mathrm{CD}}k_{\mathrm{z}}^{\mathrm{D}} \,, \tag{4.3}$$

$$\hat{\mathrm{k}}_{\mathrm{y}}^{\mathrm{C}} = R_{21}^{\mathrm{CD}}\hat{\mathrm{k}}_{\mathrm{x}}^{\mathrm{D}} + R_{22}^{\mathrm{CD}}\hat{\mathrm{k}}_{\mathrm{y}}^{\mathrm{D}} + R_{23}^{\mathrm{CD}}k_{\mathrm{z}}^{\mathrm{D}} \,, \tag{4.4}$$

$$\hat{\mathrm{k}}_{\mathrm{z}}^{\mathrm{C}} = R_{31}^{\mathrm{CD}}\hat{\mathrm{k}}_{\mathrm{x}}^{\mathrm{D}} + R_{32}^{\mathrm{CD}}\hat{\mathrm{k}}_{\mathrm{y}}^{\mathrm{D}} + R_{33}^{\mathrm{CD}}k_{\mathrm{z}}^{\mathrm{D}} \,. \tag{4.5}$$

Substituting this expressions into the $\boldsymbol{k}\cdot\boldsymbol{p}$ matrix we obtain the final partial differential system, which, applying the boundary conditions of the device, must be solved to calculate the electronics states.

4.5 Finite Element Method

At this point we have achieved a system of partial differential equations and now we will look for its numerical solution using the Finite Element Method. This Section is devoted to the mathematical implementation of the FEM for two dimensional function,

required in the case of two dimensional confinement . The description of this method can easily be found in many handbooks, but particularly, we have followed Ref. [110], to which we refer for further details.

FEM allows to get an approximated discrete solution over a domain where the differential equations are defined in weak form. The weak formulation consists in a transformation of the partial derivatives and integrals into equations that can be operated using linear algebra over a vectorial space of finite dimension, or over a Hilbert space of functions. In algebraic terms it means to define a basis of functions where any solution of the problem can be represented as a vector with a finite number of dimensions.

In the FEM, the solution is sought over a finite number of points in the spatial domain forming a triangular mesh, being each triangle called *element*. The denser the mesh is, the more accurate the solution gets. The main advantage of this type of numerical method compared by its more popular alternative, the Finite Difference Method, is that the nodes can be placed anywhere to adapt any complex geometry.

If N_n is the number of nodes, which are placed in the positions $\mathbf{r}_i$, with $i \in 1, \ldots N_\mathrm{n}$, we define the set of positions as the discrete domain D. In this scenario, N_n coefficients will be necessary to achieve the definition of a given function f at the N_n nodes: $f(\mathbf{r}_i)$, with $\mathbf{r} \in D$. The function in the rest of the spatial domain can be evaluated as a linear interpolation from $f(\mathbf{r}_i)$.

In this section, we present the main features of the FEM and its implementation on the SP2D simulator.

4.5.1 Definition of the discretized function

As aforementioned, this method focuses on the discretization of the continuous domain and tries to give an approximated solution for those points belonging to the discretized domain. Our problem involves the discretization of the cross-section of the device, the plane labeled[2] as XY, thus, it is a bidimensional problem. The element resulting from connecting all the adjacent nodes of D is triangular. Each triangular element enclosed by the nodes placed in $\mathbf{r}_a$, $\mathbf{r}_b$ and $\mathbf{r}_c$ is labeled as $\widehat{abc}$. It is a continuous subdomain where the functions defined on it can be forced to take any value on the vertexes and to be linearly interpolated in the rest of the triangle. Therefore, we can imagine the

---

[2]In the following, the default coordinate system is the DRS.
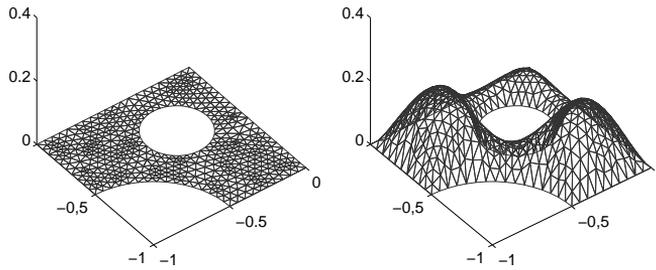
Figure 4.2: Example of a mesh corresponding to the domain of the function f and approximated function $\tilde{f}$ using the discretization patterns of the FEM.

function $\tilde{f}$ in an element as a triangular plane whose vertexes match the values of $f(\mathbf{r}_i)$, as seen in Fig. 4.2.

We want to reformulate the partial differential equations as an algebraic problem. To do so, the function f is expressed as a vector in a basis of functions, obtaining the approximated function $\tilde{f}$. The basis of functions is composed by N_n functions $\Lambda_i(\mathbf{r})$, associated to every node i, defined as:

- The value at node $\mathbf{r}_i$ is 1.

- The value at any node $j \neq i$ is 0.

- The function away of the nodes is calculated as the interpolation of the adjacent nodes.

Thus, a basis function evaluated only at the nodes can be expressed as:

$$\Lambda_i(\mathbf{r}_j) = \delta_{ij} \,. \tag{4.6}$$

From this definition, only the basis function Λ_i contributes to the value in node i. The approximation $\tilde{f}$ of the function f can thus be defined as:

$$f(\mathbf{r}) \simeq \tilde{f}(\mathbf{r}) = \sum_{i=1}^{N_\mathrm{n}} f(\mathbf{r}_i)\Lambda_i(\mathbf{r}) \,. \tag{4.7}$$

As a result, the approximated function $\tilde{f}$ can be unequivocally defined by a set of coefficients, which match the value of function f in the nodes of the domain D. The
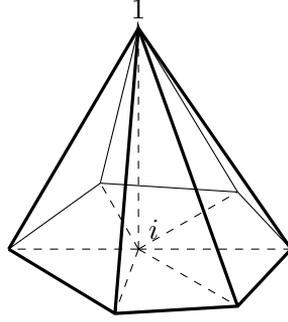
Figure 4.3: Representation of the basis function Λ_i. Its shape is a pyramid whose base (the domain D_i) is composed by the triangular elements sharing the node i.

conversion of the functions into vectors of coefficients is the first step for the weak formulation, where the derivatives and integrals must only be calculated on the basis functions.

We can regard the basis function Λ_i as a pyramid of irregular basis and height 1 (Fig. 4.3). The base of the pyramid is formed by all the triangular elements of the mesh sharing the node $\mathbf{r}_i$ as a vertex. These are the only elements we must consider when operating with the basis function Λ_i, since the basis function in the rest of the space is 0, and thus we refer to it as the domain D_i of the wave function. This visualization will be very useful to understand the operations involving the basis functions.

The result of summing all the basis functions associated to the vertexes of a triangular element is the interpolation of the values of $f(\mathbf{r}_a)$, $f(\mathbf{r}_b)$ and $f(\mathbf{r}_c)$ in the subdomain of the triangular element, as can be seen in Fig. 4.4. These are the only three basis functions to be accounted for when operating the function in that subdomain.

To evaluate the basis function in a triangular element, the result is an interpolation plane whose value is 1 in the i vertex and 0 in the other two. We denote these new functions as nodal functions $N(x, y)$, and their value is supposed to be zero out of the subdomain. Thus, a basis function Λ_i is composed by the summation of all the nodal functions whose value is 1 in the vertex i.

The nodal functions are associated to a triangular element, and each triangle has three nodal functions: one per vertex. To accurately define a nodal function, we must provide the triangle $\widehat{abc}$ it belongs to and enumerate the node whose value is 1. As a result, the three nodal functions associated to a node can be labeled as $N_1(x, y)$, $N_2(x, y)$ and $N_3(x, y)$ (Fig. 4.5), as the 1 valued node is a, b or c, respectively. Nev-
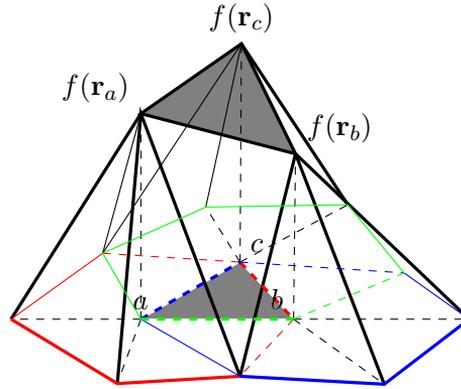
Figure 4.4: Representation of $\tilde{f}$ in the domain of the triangle $\widehat{abc}$. The contribution of the weighted basis functions involved in the interpolation Λ_a, Λ_b and Λ_c (red, blue and green respectively) is shown.
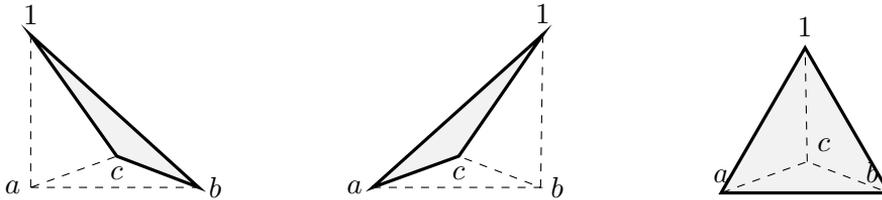
Figure 4.5: Nodal functions N_1, N_2 and N_3 for the triangular element $\widehat{abc}$.

ertheless, according to this nomenclature, there are different ways to define the same nodal function, for instance, N_1 in $\widehat{abc}$ is the same as N_2 in $\widehat{bac}$.

Thus, when dealing with the basis functions, we can subdivide the problem into the different nodal functions, operate on them and sum all the contributions. For example, to calculate the value of $\tilde{f}$ in a triangular element $\widehat{abc}$, the only contributing basis functions are Λ_a, Λ_b and Λ_c, as shown in Fig. 4.6. Particularly, the part of the function Λ_a circumscribed to $\widehat{abc}$ has the form N_1 for Λ_a, N_2 for Λ_b and N_3 for Λ_c. As a consequence, the interpolation on the triangle $\widehat{abc}$ can be written in terms of the nodal functions as:

$$\tilde{f}(x,y) = f(\mathbf{r}_a)N_1^{\widehat{abc}}(x,y) + f(\mathbf{r}_b)N_2^{\widehat{abc}}(x,y) + f(\mathbf{r}_c)N_3^{\widehat{abc}}(x,y) \quad ,(x,y) \in \widehat{abc}. \quad (4.8)$$

To sum up, we have shown in this section how to get an approximate function $\tilde{f}$
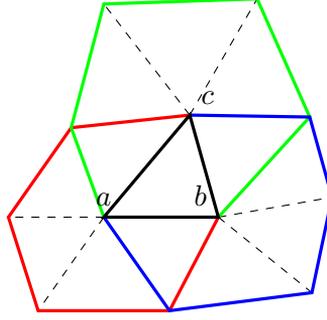
$\boldsymbol{k \cdot p}$ implementation

Figure 4.6: Representation of the domains D_a, D_b and D_c of the three basis functions Λ_a, Λ_b and Λ_c (red, blue and green respectively) sharing a triangular element $\widehat{abc}$.

that can be expressed as a vector on a basis of functions, defined by the functions $\Lambda_i(\mathbf{r})$. Since the basis functions are chosen to be 1 in a node and 0 in the rest, the value of the function $\tilde{f}(\mathbf{r}_i)$ depends only on one basis function Λ_i. Therefore, the coefficients of the resulting vector are the values of f discretized in D and the position dependency of the function has been removed. Therefore, we can treat it as a vector using conventional linear algebra.

4.5.2 Discretization of the system

Once we now how the functions are discretized in the 2D FEM, we are going to analyze how to implement the $\boldsymbol{k}\cdot\boldsymbol{p}$ equation system. The system is formed by N_c differential equations, which correspond to the number of bands of the $\boldsymbol{k}\cdot\boldsymbol{p}$ method, and therefore, the number of components of the wave functions. For the sake of simplicity, we use a system of only two equations as an example, $N_c = 2$, which can be written in a generic form as:

$$
\begin{aligned}
\hat{H}_{11}\xi_1(x,y) + \hat{H}_{12}\xi_2(x,y) &= E\xi_1(x,y) \\
\hat{H}_{21}\xi_1(x,y) + \hat{H}_{22}\xi_2(x,y) &= E\xi_2(x,y)\,.
\end{aligned}
\tag{4.9}
$$

Where $\hat{H}_{ij}$ is a generic operator, ξ_1 and ξ_2 are the coefficients of the envelope wave functions for the confined coordinates (hereinafter we call them simply 'wave functions') which are still position dependent. We rewrite the wave functions in their discretized form as:

$$\xi_1(x,y) \simeq \sum_{i=1}^{N_n} \xi_{1i}\, \Lambda_i(x,y)\,,$$

$$\xi_2(x,y) \simeq \sum_{i=1}^{N_n} \xi_{2i}\, \Lambda_i(x,y)\,. \tag{4.10}$$

Substituting Eq. (4.10) into Eq. (4.9), we obtain a system of $N_c \times N_n$ unknowns:

$$\hat{H}_{11} \sum_{i=1}^{N_n} \xi_{1i}\, \Lambda_i(x,y) + \hat{H}_{12} \sum_{i=1}^{N_n} \xi_{2i}\, \Lambda_i(x,y) = E \sum_{i=1}^{N_n} \xi_{1i}\, \Lambda_i(x,y)$$

$$\hat{H}_{21} \sum_{i=1}^{N_n} \xi_{1i}\, \Lambda_i(x,y) + \hat{H}_{22} \sum_{i=1}^{N_n} \xi_{2i}\, \Lambda_i(x,y) = E \sum_{i=1}^{N_n} \xi_{2i}\, \Lambda_i(x,y)\,. \tag{4.11}$$

To completely eliminate the dependence of the system on the position (x,y), we project the equations on a basis function Λ_j, i.e , we multiply by[3] $\Lambda_j^*(x,y)$ and integrate in the whole domain. We use the Dirac notation, which defines a generic inner product between the states represented as Λ_i and Λ_j as:

$$\langle \Lambda_j \,|\, \Lambda_i \rangle = \int \Lambda_j^*\, \Lambda_i \,\mathrm{d}\mathbf{r}\,, \tag{4.12}$$

$$\left\langle \Lambda_j \,\middle|\, \hat{H}_{lm} \,\middle|\, \Lambda_i \right\rangle = \int \Lambda_j^*\, \hat{H}_{lm}\, \Lambda_i \,\mathrm{d}\mathbf{r}\,. \tag{4.13}$$

Then, the whole system reads:

$$\sum_{i=1}^{N_n} \left(\xi_{1i} \left\langle \Lambda_j \,\middle|\, \hat{H}_{11} \,\middle|\, \Lambda_i \right\rangle + \xi_{2i} \left\langle \Lambda_j \,\middle|\, \hat{H}_{12} \,\middle|\, \Lambda_i \right\rangle \right) = E \sum_{i=1}^{N_n} \xi_{1i} \langle \Lambda_j \,|\, \Lambda_i \rangle$$

$$\sum_{i=1}^{N_n} \left(\xi_{1i} \left\langle \Lambda_j \,\middle|\, \hat{H}_{21} \,\middle|\, \Lambda_i \right\rangle + \xi_{2i} \left\langle \Lambda_j \,\middle|\, \hat{H}_{22} \,\middle|\, \Lambda_i \right\rangle \right) = E \sum_{i=1}^{N_n} \xi_{2i} \langle \Lambda_j \,|\, \Lambda_i \rangle\,. \tag{4.14}$$

Since the bra-kets affect to the known basis functions, these terms can be easily

---

[3]Although Λ is a real function, we use the conjugation so as to obtain a most general expression.

calculated. Thus, the system of N_c partial differential equations has turned into a linear equation system determined by two matrices $\boldsymbol{H}$ and $\boldsymbol{M}$ [105], whose eigenvalues and eigenvectors must be found,

$$\boldsymbol{H}\boldsymbol{\xi} = E M \boldsymbol{\xi} \tag{4.15}$$

where $\boldsymbol{\xi}$ is a vector formed by the coefficients of ξ_{1i} and ξ_{2i}, $\boldsymbol{H}$ and $\boldsymbol{M}$ are matrices of dimension $N_c N_n \times N_c N_n$, that can be subdivided into $N_n \times N_n$ sublocks of size $N_c \times N_c$. The subblock ij corresponds to the inner product of the two basis functions Λ_i and Λ_j, either the direct product in the case of M_{ij}, or applying the matrix operator for H_{ij} as shown in Eq. (4.16).

$$\mathsf{H}_{ij} = \begin{pmatrix} \left\langle \Lambda_i \left| \hat{\mathrm{H}}_{11} \right| \Lambda_j \right\rangle & \left\langle \Lambda_i \left| \hat{\mathrm{H}}_{12} \right| \Lambda_j \right\rangle \\ \left\langle \Lambda_i \left| \hat{\mathrm{H}}_{21} \right| \Lambda_j \right\rangle & \left\langle \Lambda_i \left| \hat{\mathrm{H}}_{22} \right| \Lambda_j \right\rangle \end{pmatrix}, \tag{4.16}$$

$$\mathsf{M}_{ij} = \begin{pmatrix} \left\langle \Lambda_i \left| \Lambda_j \right\rangle & 0 \\ 0 & \left\langle \Lambda_i \left| \Lambda_j \right\rangle \end{pmatrix}. \tag{4.17}$$

$$\tag{4.18}$$

And the matrices $\boldsymbol{H}$ and $\boldsymbol{M}$ read as

$$\boldsymbol{H} = \begin{pmatrix} \mathsf{H}_{11} & \mathsf{H}_{12} & \dots & \mathsf{H}_{1\,N_n} \\ \mathsf{H}_{21} & \mathsf{H}_{22} & \dots & \mathsf{H}_{2\,N_n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{H}_{N_n 1} & \mathsf{H}_{N_n 2} & \dots & \mathsf{H}_{N_n N_n} \end{pmatrix}, \tag{4.19}$$

$$\boldsymbol{M} = \begin{pmatrix} \mathsf{M}_{11} & \mathsf{M}_{12} & \dots & \mathsf{M}_{1\,N_n} \\ \mathsf{M}_{21} & \mathsf{M}_{22} & \dots & \mathsf{M}_{2\,N_n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{M}_{N_n 1} & \mathsf{M}_{N_n 2} & \dots & \mathsf{M}_{N_n N_n} \end{pmatrix}. \tag{4.20}$$

The inner products are non zero only where the basis functions overlap $D_i \cap D_j$, namely, they share any triangular element in their domains. These products are calculated as the sum of all the contributions of each triangular element shared by both

basis functions.

Let us focus on the calculation of the inner product in the domain of a triangular element shared by two basis functions. For each of the involved basis functions, its representation in the subdomain of the triangular element can be characterized as a nodal function. We will operate with the corresponding nodal functions summing the contribution of all the triangular elements involved.

So that, there are two types of inner products between basis functions. First, a function Λ_i multiplied by itself, which corresponds to the diagonal subblocks H_{ii} and M_{ii} of the matrices $\boldsymbol{H}$ and $\boldsymbol{M}$, which share the same triangular elements. Second, those corresponding to two different basis functions Λ_i and Λ_j, which corresponds to non diagonal subblocks H_{ij} and M_{ij}, where only two triangular elements are shared, as can be seen in Fig. 4.6.

4.5.3   Evaluation of the inner products

In order to calculate each element of $\boldsymbol{H}$ and $\boldsymbol{M}$ it is necessary to study the inner products of the form in Eqs. (4.12) and (4.13) involved in the construction of the $\boldsymbol{k}\cdot\boldsymbol{p}$ matrices. We use the generic operator $\hat{\mathcal{O}}$ so as to refer the operator in the inner products of the type of Eq. (4.13). This operator can involve derivatives of the second order, which are not well defined in the basis functions. For this reason, it is desirable to split these operators into two, each one involving derivatives of the first order, $\hat{\mathcal{O}} = \hat{\mathcal{O}}_1 \hat{\mathcal{O}}_2$, giving expressions of the form:

$$\left\langle \Lambda_i \left| \hat{\mathcal{O}} \right| \Lambda_j \right\rangle = \left\langle \Lambda_i \left| \hat{\mathcal{O}}_1 \hat{\mathcal{O}}_2 \right| \Lambda_j \right\rangle , \tag{4.21}$$

which can be rewritten using the adjoint of the operator $\hat{\mathcal{O}}_1$ as:

$$\left\langle \Lambda_i \left| \hat{\mathcal{O}} \right| \Lambda_j \right\rangle = \left\langle \hat{\mathcal{O}}_1^{\dagger} \Lambda_i \left| \hat{\mathcal{O}}_2 \Lambda_j \right\rangle , \tag{4.22}$$

Let us now analyze the products involved in the calculation of each subblock H_{ij}, corresponding to the inner products of two basis functions Λ_i and Λ_j. According to the operator ordering for the $\boldsymbol{k}\cdot\boldsymbol{p}$ operator introduced in Sec. 3.6, the products needed to construct the $\boldsymbol{k}\cdot\boldsymbol{p}$ matrices, in the integral form, are:

$$\int \Lambda_i(x,y)\, C\, \Lambda_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{4.23a}$$

$$\int \Lambda_i(x,y)\, \hat{\mathrm{k}}_\alpha\, C\, \Lambda_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{4.23b}$$

$$\int \Lambda_i(x,y)\, C\, \hat{\mathrm{k}}_\alpha\, \Lambda_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{4.23c}$$

$$\int \Lambda_i(x,y)\, \hat{\mathrm{k}}_\alpha\, C\, \hat{\mathrm{k}}_\beta\, \Lambda_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{4.23d}$$

$$\int \Lambda_i(x,y)\, \hat{\mathrm{k}}_\alpha\, C\, k_{\mathrm{z}}\, \Lambda_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{4.23e}$$

$$\int \Lambda_i(x,y)\, k_{\mathrm{z}}\, C\, \hat{\mathrm{k}}_\alpha\, \Lambda_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{4.23f}$$

$$\int \Lambda_i(x,y)\, V(x,y)\, \Lambda_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,. \tag{4.23g}$$

where α and β refer to either of the confined directions x or y; k_{z} is constant; and C refers to a generic parameter that depends on the material, i.e., it is considered constant inside a triangular element, but can change abruptly from one element to another if their materials are different. Finally, $V(x,y)$ refers to the potential depending on the position, which is interpolated using the nodal functions[4]:

$$V(x,y) = \sum_{i=1}^{3} V_i\, N_i\,. \tag{4.24}$$

As a result of the application of the property in Eq. (4.22), the inner products involving the momentum operator can be rewritten as:

$$\int \Lambda_i^{*}\hat{\mathrm{k}}_\alpha\, C\, k_{\mathrm{z}}\Lambda_j\, \mathrm{d}x\, \mathrm{d}y = \int \left(\hat{\mathrm{k}}_\alpha \Lambda_i\right)^{*} C\, k_{\mathrm{z}}\Lambda_j\, \mathrm{d}x\, \mathrm{d}y\,, \tag{4.25}$$

$$\int \Lambda_i^{*}\hat{\mathrm{k}}_\alpha\, C\, \hat{\mathrm{k}}_\beta\Lambda_j\, \mathrm{d}x\, \mathrm{d}y = \int \left(\hat{\mathrm{k}}_\alpha \Lambda_i\right)^{*} C\, \left(\hat{\mathrm{k}}_\beta\Lambda_j\right)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{4.26}$$

The integrals are carried out over all the XY plane. However, the basis functions are non zero in a small subdomain D_i. These inner products between basis functions can be calculated by summing the contribution of the integrals performed in the subdomain

---

[4]For a simpler notation, the dependence of the nodal and basis functions with (x,y) is no longer included.

of each triangular element. Only the elements in the intersection of the subdomains $D_i \cap D_j$ of both basis functions contribute to the total integral. Then, the integrals restricted to the subdomain of a triangular element can be expressed in terms of the nodal functions with their corresponding shape N_1, N_2 or N_3.

$$\int N_l^* \hat{O} N_m \, \mathrm{d}x \, \mathrm{d}y = \int \left(\hat{O}_1^\dagger N_l \right)^* \hat{O}_2 N_m \, \mathrm{d}x \, \mathrm{d}y \,, \tag{4.27}$$

where $l, m = 1, 2, 3$ denotes the specific nodal function.

In the case that concerns us, the operator $\hat{O}_{1,2}$ stands for any of the momentum operators $\hat{k}_x, \hat{k}_y, k_z$ or for the unit operator. The parameter C depends also on the position, since it can change abruptly between materials. Thus, it must be considered as a part of $\hat{O}_2$ for consistence.

$$\hat{O}_1 \to \hat{k}_x, \hat{k}_y, k_z \,, \tag{4.28a}$$

$$\hat{O}_2 \to C\hat{k}_x, C\hat{k}_y, Ck_z \,. \tag{4.28b}$$

The momentum operators are self adjoint, but the operators involving the multiplication by a function P depending on the position are not. Then, the adjoint operators of the previous ones are, respectively:

$$\hat{O}_1^\dagger \to \hat{k}_x, \hat{k}_y, k_z, C \,, \tag{4.29a}$$

$$\hat{O}_2^\dagger \to \hat{k}_x C, \hat{k}_y C, k_z C \,. \tag{4.29b}$$

Thus, the adjoint of an operator with a generic form $\hat{k}_\alpha C \hat{k}_\beta$ is given by:

$$\hat{k}_\alpha C \hat{k}_\beta \xrightarrow{\dagger} \hat{k}_\beta C \hat{k}_\alpha \,. \tag{4.30}$$

However, since the parameter C depending on the material is constant in the domain of a triangular element, it can be pulled outside of the integral of the nodal functions. Therefore, it is not considered in the next subsections when operating with the nodal functions.

These operations can be numerically calculated by means of a change of variable as a function of the coordinates (x_i, y_i) of the vertexes of the triangle and its surface S, which is also calculated from these coordinates. The change of variable is described in Appendix D.
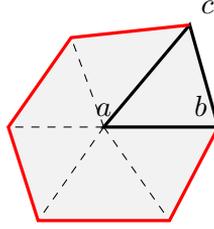
Figure 4.7: Triangular elements involved in the calculation of the diagonal subblocks. All the triangles composing the basis function Λ_a take part in the product between Λ_a and itself.

4.5.4 Construction of H and M

Once we know how to deal with the differential operators in the inner products, we can proceed to calculate the matrices $\boldsymbol{H}$ and $\boldsymbol{M}$. We already know that these matrices are composed by submatrices or subblocks, corresponding to the inner products between the basis functions. Therefore, since there are N_n basis functions, there are also $N_\mathrm{n} \times N_\mathrm{n}$ inner products, with only some of them non-zero because most of the basis functions do not overlap.

As it was already commented, the product of two basis functions can be decomposed into the summation of the products of the nodal functions in each triangle where the two basis functions overlap. In the domain of a triangular element, each basis function takes the shape of a nodal function, and therefore, the product can be expressed as a sum of terms with the form of Eq. (4.27). Therefore, for each block it is necessary to identify the triangular elements that overlap, and their corresponding nodal functions for each one.

The submatrices can be classified into two groups: those placed on the diagonal and labeled as H_{ii} and M_{ii}, which involve only one basis function; and those involving two different basis functions, placed out of the diagonal and labeled as H_{ij} and M_{ij}.

Diagonal Submatrices

Let us start with the inner products of a basis function Λ_a (centered at node a) with itself. These terms correspond to the submatrices H_{aa} and M_{aa}, located at the diagonal. The equations must be solved for $\Lambda_i = \Lambda_j = \Lambda_a$. We denote this products generically as $\left\langle \Lambda_a \left| \hat{\mathcal{O}} \right| \Lambda_a \right\rangle$, where $\hat{\mathcal{O}}$ stands for any of the operators in Eq. (4.16).

We divide the integral in a sum of the integrals over the triangular elements over-

lapped in the product between the two basis functions. In this case, since the function to be multiplied is the same, all the triangles belonging to the domain of the basis function contribute to the total integral (Fig. 4.7).

Considering a triangular element $\widehat{abc}$, the nodal function associated to Λ_a is N_1, i.e., the nodal function which is one in the node a. We remark that there are different ways of defining the same nodal function for different ordering of the vertexes of the triangles. Nevertheless, we always consider the ordering of the triangle with the a vertex in the first position to use the nodal function N_1. The calculation of the inner products of this type is detailed in Apendix E. The results are collected in Table 4.1, where x_i and y_i correspond to the coordinates of the triangle vertexes, and S is the area of the triangle (defined in Appendix D).

The expressions in the tables show $\pm$ and $\mp$ signs that are related to the sign of S. According to our convention of positive surfaces and counterclockwise ordering of vertexes, we take the upper sign.

Non Diagonal submatrices

As for calculating the non diagonal submatrices $\mathbf{H}_{ab}$ and $\mathbf{M}_{ab}$, we must consider the inner products between two different basis functions Λ_a and Λ_b. Since the basis functions are non zero only in a small subdomain, there are only a few non zero products, corresponding to overlapping basis function, i.e., those basis functions centered in the adjacent nodes. This means that for a basis function with n adjacent nodes, there are only n non zero out of diagonal submatrices. For example, the node a in Fig. 4.8 is the adjacent node of six basis functions centered in the six external nodes of Λ_a.

As in the previous case, the inner products are calculated for each triangle. For this case, there are only two contributing elements shared by the two basis functions (Fig. 4.8). The functions Λ_a and Λ_b show different nodal functions in the triangle $\widehat{abc}$. The same product can be labeled in different ways, for instance, N_3 times N_1 in the triangle $\widehat{bca}$ (nodes are labeled counterclockwise). Since any product can be expressed in terms of the type (N_1, N_2) or (N_2, N_1) according to a determined ordering of the vertexes, we define these products in Tab. 4.2. The calculation of these products is detailed in Apendix E.

Summing the contribution of the inner products calculated for the two triangles, we obtain the complete inner product of Λ_a with Λ_b. The opposite ordering of the nodal functions yields the inner product $\left\langle \Lambda_b \left| \hat{\mathcal{O}} \right| \Lambda_a \right\rangle$.

$\hat{\mathcal{O}}$	$\int N_1 \, \hat{\mathcal{O}} \, N_1 \, \mathrm{d}x \, \mathrm{d}y$
$\hat{\mathrm{k}}_\mathrm{x}\hat{\mathrm{k}}_\mathrm{x}$	$\dfrac{(y_2 - y_3)^2}{4|S|}$
$\hat{\mathrm{k}}_\mathrm{x}\hat{\mathrm{k}}_\mathrm{y}$	$\dfrac{(y_2 - y_3)(x_3 - x_2)}{4|S|}$
$\hat{\mathrm{k}}_\mathrm{x}k_\mathrm{z}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{y_2 - y_3}{6}$
$\hat{\mathrm{k}}_\mathrm{y}\hat{\mathrm{k}}_\mathrm{x}$	$\dfrac{(y_2 - y_3)(x_3 - x_2)}{4|S|}$
$\hat{\mathrm{k}}_\mathrm{y}\hat{\mathrm{k}}_\mathrm{y}$	$\dfrac{(x_3 - x_2)^2}{4|S|}$
$\hat{\mathrm{k}}_\mathrm{y}k_\mathrm{z}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{x_3 - x_2}{6}$
$k_\mathrm{z}\hat{\mathrm{k}}_\mathrm{x}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{y_3 - y_2}{6}$
$k_\mathrm{z}\hat{\mathrm{k}}_\mathrm{y}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{x_2 - x_3}{6}$
$k_\mathrm{z}k_\mathrm{z}$	$\dfrac{k_\mathrm{z}^2|S|}{6}$
C	$C\dfrac{|S|}{6}$
V	$|S|\dfrac{3V_1 + V_2 + V_3}{30}$

Table 4.1: Constituent terms of the elements in the diagonal blocks.

Once we know which nodal functions are involved in each subblock, let us emphasize on the importance of the correct labeling of the vertexes of the triangles. The choice of the products (N_1, N_1), (N_1, N_2) and (N_2, N_1) is not made randomly. Each triangular element contributes to nine different interactions between the three basis functions that share it. Assuming that the vertexes are always ordered in the same orientation[5], a triangle can be ordered in three different ways, for which the latter products represent all the nine overlaps between functions, as shown in Table 4.3.

Consequently, it is possible to follow a simple algorithm so as to avoid either the skipping or the repetition of contributions. The algorithm consists in a loop for the triangular elements, where the nine products are calculated and their contribution summed to the corresponding subblock. Thus, when the loop is finished, all the subblocks will have been completed with all their contributions.

---

[5] We use counterclockwise ordering.
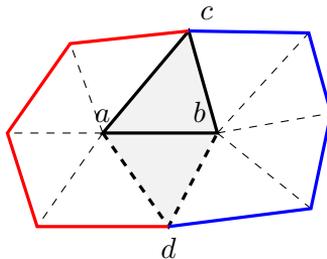
Figure 4.8: Triangular elements involved in the calculation of the non diagonal subblocks. Only the two triangles shared by the basis function Λ_a and Λ_b take part in the product between them.

Rotation of the Triangular Elements

The terms calculated so far are referred to the DRS. However, as seen in Sec. 4.4, the $\boldsymbol{k} \cdot \boldsymbol{p}$ hamiltonian is referred to the CRS, and consequently, it is necessary a rotation between both reference systems.

The rotation operation is linear, and therefore, the inner products involving the momentum in the CRS can be calculated as

$$\left\langle \Lambda_a \left| \hat{\boldsymbol{k}}^{\mathrm{C}} \right| \Lambda_b \right\rangle = \left\langle \Lambda_a \left| \boldsymbol{R}^{\mathrm{CD}} \hat{\boldsymbol{k}}^{\mathrm{D}} \right| \Lambda_b \right\rangle = \boldsymbol{R}^{\mathrm{CD}} \left\langle \Lambda_a \left| \hat{\boldsymbol{k}}^{\mathrm{D}} \right| \Lambda_b \right\rangle , \qquad (4.31)$$

where $\left\langle \Lambda_a \left| \hat{\boldsymbol{k}}^{\mathrm{D}} \right| \Lambda_b \right\rangle$ is a vector whose components are the inner products of $\hat{\mathrm{k}}_{\mathrm{x}}$, $\hat{\mathrm{k}}_{\mathrm{y}}$ and k_{z} calculated for Λ_a and Λ_b.

The same change of reference system must be done for the second order momentum operator, though it is represented by a tensor which in the DRS is:

$$\hat{\mathbf{K}}^{\mathrm{D}} = \begin{pmatrix} -\partial^2/\partial x^2 & -\partial^2/\partial xy & -\mathrm{i}\partial/\partial x\, k_{\mathrm{z}} \\ -\partial^2/\partial yx & -\partial^2/\partial y^2 & -\mathrm{i}\partial/\partial y\, k_{\mathrm{z}} \\ -\mathrm{i}\, k_{\mathrm{z}}\partial/\partial x & -\mathrm{i}\, k_{\mathrm{z}}\partial/\partial y & k_{\mathrm{z}}^2 \end{pmatrix} . \qquad (4.32)$$

The rotation of the coordinate system in this case is

$$\left\langle \Lambda_a \left| \hat{\mathbf{K}}^{\mathrm{C}} \right| \Lambda_b \right\rangle = \boldsymbol{R}^{\mathrm{CD}} \left\langle \Lambda_a \left| \hat{\mathbf{K}}^{\mathrm{D}} \right| \Lambda_b \right\rangle \left(\boldsymbol{R}^{\mathrm{CD}} \right)^{\mathrm{T}} , \qquad (4.33)$$

where $\left\langle \Lambda_a \left| \hat{\mathbf{K}}^{\mathrm{C}} \right| \Lambda_b \right\rangle$ corresponds to a matrix formed by the arrange of the inner products in Tabs. 4.1 or 4.2, depending on whether or not $\Lambda_a = \Lambda_b$ respectively. Thus, the inner products in the CRS can be calculated by simply rotating the inner products in

$\hat{\mathcal{O}}$	$\int N_1 \, \hat{\mathcal{O}} \, N_2 \, \mathrm{d}x \, \mathrm{d}y$	$\int N_2 \, \hat{\mathcal{O}} \, N_1 \, \mathrm{d}x \, \mathrm{d}y$
$\hat{\mathrm{k}}_\mathrm{x}\hat{\mathrm{k}}_\mathrm{x}$	$\dfrac{(y_2 - y_3)(y_3 - y_1)}{4|S|}$	$\dfrac{(y_2 - y_3)(y_3 - y_1)}{4|S|}$
$\hat{\mathrm{k}}_\mathrm{x}\hat{\mathrm{k}}_\mathrm{y}$	$\dfrac{(y_2 - y_3)(x_1 - x_3)}{4|S|}$	$\dfrac{(y_3 - y_1)(x_3 - x_2)}{4|S|}$
$\hat{\mathrm{k}}_\mathrm{x}k_\mathrm{z}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{y_2 - y_3}{6}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{y_3 - y_1}{6}$
$\hat{\mathrm{k}}_\mathrm{y}\hat{\mathrm{k}}_\mathrm{x}$	$\dfrac{(x_3 - x_2)(y_3 - y_1)}{4|S|}$	$\dfrac{(x_1 - x_3)(y_2 - y_3)}{4|S|}$
$\hat{\mathrm{k}}_\mathrm{y}\hat{\mathrm{k}}_\mathrm{y}$	$\dfrac{(x_3 - x_2)(x_1 - x_3)}{4|S|}$	$\dfrac{(x_1 - x_3)(x_3 - x_2)}{4|S|}$
$\hat{\mathrm{k}}_\mathrm{y}k_\mathrm{z}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{x_3 - x_2}{6}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{x_1 - x_3}{6}$
$k_\mathrm{z}\hat{\mathrm{k}}_\mathrm{x}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{y_1 - y_3}{6}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{y_3 - y_2}{6}$
$k_\mathrm{z}\hat{\mathrm{k}}_\mathrm{y}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{x_3 - x_1}{6}$	$\pm \mathrm{i}k_\mathrm{z}\dfrac{x_2 - x_3}{6}$
$k_\mathrm{z}k_\mathrm{z}$	$\dfrac{k_\mathrm{z}^2|S|}{12}$	$\dfrac{k_\mathrm{z}^2|S|}{12}$
C	$C\dfrac{|S|}{12}$	$C\dfrac{|S|}{12}$
V	$|S|\dfrac{2V_1 + 2V_2 + V_3}{60}$	$|S|\dfrac{2V_1 + 2V_2 + V_3}{60}$

Table 4.2: Constituent terms of the elements in the non diagonal blocks

the DRS. As a consequence, all the momenta expressed in the CRS, which are used in the $\boldsymbol{k\cdot p}$ matrix, may depend on the three orientation DRS momenta[6] $\hat{\mathrm{k}}_\mathrm{x}$, $\hat{\mathrm{k}}_\mathrm{y}$ and k_z.

4.5.5 Hermiticity of the operators

In order to provide physically meaningful energies, the operator Hamiltonian $\hat{\mathbf{H}}$ has to be Hermitian:

$$\hat{\mathbf{H}} = \hat{\mathbf{H}}^\dagger, \tag{4.34}$$

that is to say, the element in position ij is the adjoint operator of that in position ji

---

[6]From now on, the superscript D is dropped since the DRS is assumed to be the default reference system.
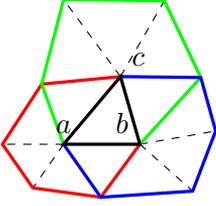
	$\widehat{abc}$	$\widehat{bca}$	$\widehat{cab}$
N_1, N_1	(Λ_a, Λ_a)	(Λ_b, Λ_b)	(Λ_c, Λ_c)
N_1, N_2	(Λ_a, Λ_b)	(Λ_b, Λ_c)	(Λ_c, Λ_a)
N_2, N_1	(Λ_b, Λ_a)	(Λ_c, Λ_b)	(Λ_a, Λ_c)

Table 4.3: Correspondence between the products of the nodal functions evaluated in the triangular element $\widehat{abc}$ (and its rotations) and the product of the basis functions they contribute to.

$$\hat{\mathrm{H}}_{ji} = \hat{\mathrm{H}}_{ij}^{\dagger}. \tag{4.35}$$

Each element in the matrix may consist of the sum of different operators like those used in Eq. (4.23), and the form of the operator matrix so as to fulfill the hermiticity condition was introduced in Section 3.6.

Regarding the inner products of the nodal functions in Tables 4.1 and 4.2, it is shown that the numerical calculation fulfills the hermiticity condition of the operators for each triangular element, since:

$$\int N_l \hat{\mathrm{k}}_\alpha P \hat{\mathrm{k}}_\beta N_m \, \mathrm{d}x \, \mathrm{d}y = \left(\int N_m \hat{\mathrm{k}}_\beta P \hat{\mathrm{k}}_\alpha N_l \, \mathrm{d}x \, \mathrm{d}y \right)^*, \tag{4.36}$$

$$\int N_l \hat{\mathrm{k}}_\alpha P N_m \, \mathrm{d}x \, \mathrm{d}y = \left(\int N_m \hat{\mathrm{k}}_\beta P \hat{\mathrm{k}}_\alpha N_l \, \mathrm{d}x \, \mathrm{d}y \right)^*. \tag{4.37}$$

As a result, the inner products of the basis functions can be calculated also to be Hermitian summing the contribution of each triangular element:

$$\int \Lambda_a \hat{\mathrm{k}}_\alpha C \hat{\mathrm{k}}_\beta \Lambda_b \, \mathrm{d}x \, \mathrm{d}y = \left(\int \Lambda_b \hat{\mathrm{k}}_\beta C \hat{\mathrm{k}}_\alpha \Lambda_a \, \mathrm{d}x \, \mathrm{d}y \right)^*, \tag{4.38}$$

$$\int \Lambda_a \hat{\mathrm{k}}_\alpha C \Lambda_b \, \mathrm{d}x \, \mathrm{d}y = \left(\int \Lambda_b \hat{\mathrm{k}}_\beta C \hat{\mathrm{k}}_\alpha \Lambda_a \, \mathrm{d}x \, \mathrm{d}y \right)^*. \tag{4.39}$$

This condition, if the operator matrix is constructed according to Eq. (4.34), is enough to assure that the total $\boldsymbol{k \cdot p}$ matrix in its weak form $\boldsymbol{H}$ is Hermitian, and therefore, to assure that it provides real energies, since all the subblocks fulfill:

$$\mathbf{H}_{ab} = \mathbf{H}_{ba}^{\dagger}. \tag{4.40}$$

Let us now take a deeper insight into the form of the subblocks $\mathbf{H}_{ab}$, in two different scenarios: for a uniform material and for an interface between two materials.

Uniform material

As seen before, the hermiticity is a sufficient condition to obtain real energies. However, as the momentum operators are linear, if the parameter C is constant in the domain of a inner product, the operator ordering should be indifferent, i.e., $\hat{\mathcal{O}}_1 \hat{\mathcal{O}}_2 = \hat{\mathcal{O}}_2 \hat{\mathcal{O}}_1$:

$$\left\langle \Lambda_a \left| \hat{k}_\alpha \hat{k}_\beta \right| \Lambda_b \right\rangle = \left\langle \Lambda_a \left| \hat{k}_\beta \hat{k}_\alpha \right| \Lambda_b \right\rangle . \tag{4.41}$$

Clearly, regarding the Tables 4.1 and 4.2, this condition does not seem to be fulfilled. Let us now look into this issue. Firstly, we consider the $k_z \hat{k}_\alpha$ term (with $\alpha = x, y$). According to the previous calculations, we can obtain the integrals corresponding to the diagonal terms $\langle \Lambda_{v1} | \hat{\mathcal{O}}_1 \hat{\mathcal{O}}_2 | \Lambda_{v1} \rangle$:

$$\int N_1 k_z \hat{k}_x N_1 \mathrm{d}x \mathrm{d}y = \pm \mathrm{i}k_z \frac{y_3 - y_2}{6} , \qquad \int N_1 \hat{k}_x k_z N_1 \mathrm{d}x \mathrm{d}y = \pm \mathrm{i}k_z \frac{y_2 - y_3}{6} ,$$

$$\int N_1 k_z \hat{k}_y N_1 \mathrm{d}x \mathrm{d}y = \pm \mathrm{i}k_z \frac{x_2 - x_3}{6} , \qquad \int N_1 \hat{k}_y k_z N_1 \mathrm{d}x \mathrm{d}y = \pm \mathrm{i}k_z \frac{x_3 - x_2}{6} . \tag{4.42}$$

These terms do not fulfill the commutative property of the operators. However, this operation is only performed in the domain of a single triangular element, and the full product is obtained by summing the contribution of all triangles that share the same vertex, multiplied each one by their respective C parameter, which is constant for all of them. It can be shown that this summation vanishes. If we label vertexes such that v_1 is the common vertex and $v_1, v_{2,t}$ and $v_{3,t}$ are in counterclockwise order (see Fig. 4.9), we obtain for the $k_z \hat{k}_x$ term:

$$\sum_{t=1}^{n} \mathrm{i}k_z \frac{y_{3,t} - y_{2,t}}{6} = \mathrm{i}k_z \frac{y_{3,n} - y_{2,1}}{6} = 0 , \tag{4.43}$$

since $y_{3,n} \equiv y_{2,1}$ are the same node with different labeling. The same result is obtained
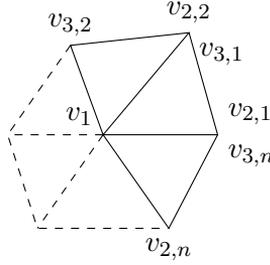
Figure 4.9: Labeling of vertexes for the sum in a diagonal subblock.

for the term $\hat{k}_y k_z$. Therefore, the hermiticity is fulfilled, and these terms are expected to vanish when solving analytically the product.

As for the integrals containing products of elements centered on different vertexes we obtain:

$$\int N_1 k_z \hat{k}_x N_2 \mathrm{d}x \mathrm{d}y = \pm i k_z \frac{y_1 - y_3}{6}\,, \qquad \int N_1 \hat{k}_x k_z N_2 \mathrm{d}x \mathrm{d}y = \pm i k_z \frac{y_2 - y_3}{6}\,,$$
$$\int N_1 k_z \hat{k}_y N_2 \mathrm{d}x \mathrm{d}y = \pm i k_z \frac{x_3 - x_1}{6}\,, \qquad \int N_1 \hat{k}_y k_z N_2 \mathrm{d}x \mathrm{d}y = \pm i k_z \frac{x_3 - x_2}{6}\,. \qquad (4.44)$$

Summing the contribution of the two triangles in the product $\langle \Lambda_{v1}|\hat{\mathcal{O}}_1 \hat{\mathcal{O}}_2|\Lambda_{v2}\rangle$ (Fig. 4.10), it yields:

$$i k_z \frac{y_1 - y_{3,1}}{6} + i k_z \frac{y_{3,2} - y_1}{6} = i k_z \frac{y_{3,2} - y_{3,1}}{6} \quad \text{(for } k_z \hat{k}_x \text{ term),}$$
$$i k_z \frac{y_2 - y_{3,1}}{6} + i k_z \frac{y_{3,2} - y_2}{6} = i k_z \frac{y_{3,2} - y_{3,1}}{6} \quad \text{(for } \hat{k}_x k_z \text{ term).} \qquad (4.45)$$

which are the same, thereby fulfilling the commutative condition. An analogous calculation can be done for the term containing $\hat{k}_y$.

Let us calculate now the product $\langle \Lambda_{v2}|\hat{\mathcal{O}}_1 \hat{\mathcal{O}}_2|\Lambda_{v1}\rangle$:

$$i k_z \frac{y_{3,1} - y_2}{6} + i k_z \frac{y_2 - y_{3,2}}{6} = i k_z \frac{y_{3,1} - y_{3,2}}{6} \quad \text{(for } k_z \hat{k}_x \text{ term),}$$
$$i k_z \frac{y_{3,1} - y_1}{6} + i k_z \frac{y_1 - y_{3,2}}{6} = i k_z \frac{y_{3,1} - y_{3,2}}{6} \quad \text{(for } \hat{k}_x k_z \text{ term).} \qquad (4.46)$$

We can observe that these terms are equal, thereby fulfilling the symmetry of the operator.
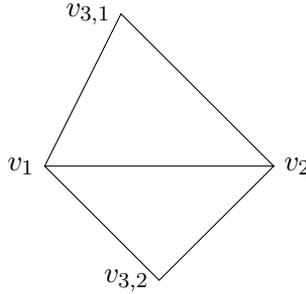
Figure 4.10: Labeling of vertexes for the sum in a non diagonal subblock.

Another pair of terms that are not symmetric at first sight is $\hat{k}_x\hat{k}_y$ and its reverse. The diagonal subblocks are symmetric, as shown in Tab. 4.1. As for non diagonal subblocks, the product is not trivially symmetric so we have to consider the sum over the two triangles in Fig. 4.10. Using Eq. (D.5) we can write:

$$\frac{(y_2 - y_3)(x_1 - x_3)}{4|S|} = \frac{2S + (x_3 - x_2)(y_3 - y_1)}{4|S|} = \pm\frac{1}{2} + \frac{(x_3 - x_2)(y_3 - y_1)}{4|S|}\,, \quad (4.47)$$

where the $+(-)$ sign must be used for counterclockwise (clockwise) order. When summing over the two triangles as in Fig. 4.10, the two $\pm 1/2$ terms cancel out because of the differently ordered labelings, so that the matrix element of the two terms $\hat{k}_x\hat{k}_y$ and $\hat{k}_y\hat{k}_x$ coincides.

As a consequence of the commutativity of these operators for constant C value, the $\mathbf{H}_{ab}$ subblocks are symmetric matrices (not Hermitian then). Nevertheless, this issue does not conflict with the hermiticity of the total matrix of the system $\boldsymbol{H}$.

Heterostructures

In an interface between different materials, the material parameter C becomes a function of the position, and therefore

$$\hat{\mathcal{O}}_1\hat{\mathcal{O}}_2 \neq \hat{\mathcal{O}}_2\hat{\mathcal{O}}_1\,, \quad (4.48)$$

where $\hat{\mathcal{O}}_1$ and $\hat{\mathcal{O}}_2$ are those defined in Eq. (4.28). Thus, we get:

$$\left\langle \Lambda_a \left| \hat{\mathrm{k}}_\alpha C \hat{\mathrm{k}}_\beta \right| \Lambda_b \right\rangle \neq \left\langle \Lambda_a \left| \hat{\mathrm{k}}_\beta C \hat{\mathrm{k}}_\alpha \right| \Lambda_b \right\rangle . \tag{4.49}$$

As a result, the subblocks $\mathbf{H}_{ab}$ are not symmetric. In this case, any linear combination of the operator orderings can be chosen as long as the conditions in Sec. 3.6 are fulfilled. It is an usual approach to force the symmetry of these subblock matrices corresponding to the interface of the device. However, as it was introduced in Sec. 3.7, the symmetrization has been reported to provide spurious solutions in some cases [27]. Thus, the Burt-Foreman operator ordering is chosen so as to mitigate this issue. Let us then remark that, for this operator ordering, the subblock matrices corresponding to the interface are not symmetric, but it does not conflict with the hermiticity of the total matrix $\boldsymbol{H}$.

4.6 Self-Consistent Schrödinger-Poisson Solver

In order to evaluate the electrostatic behavior of long-channel nanowires, the potential as a function of the charge distribution has to be evaluated using the Poisson equation. However, the charge distribution also depends on the potential through the solution of the Schrödinger equation (in our case, using the $\boldsymbol{k}\cdot\boldsymbol{p}$ evaluation), since an especific potential yields to a bandstructure and the corresponding set of wave function. The occupation of these states, according to the Fermi distribution, determines the total charge distribution in the cross-section of the device. As a result, we get a coupled system whose solution has to be self-consistent. To achieve this self-consistency, an iterative method must be adopted, where the final solution is sought by using the result of one of the equations as the input of the other. To start the iterative scheme, an initial assumption is chosen for any of the two equations. Depending on this choice the algorithm can quickly converge or even not converge at all.

To sum up, the description of a self-consistent algorithm is:

1. The potential $V(\mathbf{r})$ for the iteration $j+1$ is calculated using the Poisson equation, and it depends on the charge distribution $\rho_{\mathrm{q}}(\mathbf{r})$ calculated in the previous iteration j, or the initial solution assumed if it is the first iteration.

$$\nabla \varepsilon \nabla V^{(j+1)}(\mathbf{r}) = -\rho_{\mathrm{q}}^{(j)}(\mathbf{r}) . \tag{4.50}$$

2. Once the potential is known, the envelope wave functions and the corresponding energies are calculated by solving the partial differential equation system that yields from the discretization of the $\boldsymbol{k}{\cdot}\boldsymbol{p}$ equation.

$$\hat{\mathbf{H}}\left(V^{(j+1)(\mathbf{r})}\right)\boldsymbol{\xi}_{nk_z}^{(j+1)}(\mathbf{r}) = E_{nk_z}^{(j+1)}\boldsymbol{\xi}_{nk_z}^{(j+1)}(\mathbf{r}). \tag{4.51}$$

3. The charge is calculated using the envelope wave functions, to determine the probability distribution of the charge density, and their energies, to estimate the probability of occupation.

$$\rho_q^{(j+1)}(\mathbf{r}) = \rho_q\left(E_{nk_z}^{(j+1)}, \boldsymbol{\xi}_{nk_z}^{(j+1)}(\mathbf{r})\right). \tag{4.52}$$

4. The convergence is evaluated for the charge density in the last two iterations. The relative error must be below a limit ϵ_q.

$$\frac{\left|\rho_q^{(j+1)}(\mathbf{r}) - \rho_q^{(j)}(\mathbf{r})\right|^2}{\left|\rho_q^{(j+1)}(\mathbf{r})\right|^2} \leq \epsilon_q. \tag{4.53}$$

This process may be problematic since convergence is not assured. The convergence may depend on the validity of the initial solution, and even if it converges, it usually occurs in too many iterations. An usual approach to assure the convergence is the underrelaxation of the solution by using a weighted solution between iterations [134], i.e., a solution of one iteration depends also on the previous as

$$\rho_q^{j+1} = \omega\rho_{q,\text{raw}}^{(j+1)} + (1 - \omega)\rho_q^j, \tag{4.54}$$

where $\rho_{q,\text{raw}}$ stands for the charge obtained using the regular expression in step 3, and ω is the weighting factor, which may also depend on the iteration if the algorithm is designed to be adaptive. The weighting factor is determinant to assure the convergence since it prevents the oscillation of the solution. However, if it is too small, many iterations would be required. Due to the computational burden of each iteration, it is very important to minimize its number. Therefore, we must adopt a better algorithm able to assure and accelerate the convergence.

4.6.1 Predictor-Corrector Algorithm

To improve the performance of the iterative method and assure the convergence, we implemented a predictor-corrector scheme similar to the one described in Ref. [134]. This algorithm is based in a prediction of the charge density $\tilde{\rho}_q(\mathbf{r})$, instead of the value $\rho_q(\mathbf{r})$, based in an estimation that explicitly accounts for the potential. This prediction of the charge is then corrected by solving the Schrödinger equation with the new potential.

In the first step of this algorithm, the potential $V^{(j+1)}(\mathbf{r})$ of the new iteration is calculated from the charge density given by the previous iteration, modified to take into account the variation of the potential, yielding a more accurate description of the charge. Consequently, this step becomes more complex because of the transformation into a non linear differential equation. However, this non linear equation can be solved using the Newton's Method and leads to an acceleration of the whole algorithm by reducing the number of iterations needed.

Including the modifications, the algorithm (see Fig. 4.11) can be described as:

1. The potential $V^{(j+1)}(\mathbf{r})$ in the cross-section of the device is calculated taking into account the prediction of the charge distribution $\tilde{\rho}_q(\mathbf{r})$, which depends on the new potential, as well as on the values of energy and wave functions of the previous iteration. The system is solved considering the boundary conditions set by the gate potential for the external borders.

$$\nabla\varepsilon\nabla V^{(j+1)}(\mathbf{r}) = -\tilde{\rho}_q\left(V^{(j+1)}(\mathbf{r}); V^{(j)}(\mathbf{r}), \boldsymbol{\xi}_{nk_z}^{(j)}(\mathbf{r}), E_{nk_z}^{(j)}\right). \qquad (4.55)$$

2. The $\boldsymbol{k}\cdot\boldsymbol{p}$ method is solved to get the envelope wave functions and the energies as a function of k_z.

$$\hat{\mathbf{H}}\left(V^{(j+1)}(\mathbf{r})\right)\boldsymbol{\xi}_{nk_z}^{(j+1)}(\mathbf{r}) = E_{nk_z}^{(j+1)}\boldsymbol{\xi}_{nk_z}^{(j+1)}(\mathbf{r}). \qquad (4.56)$$

3. The value of the charge for the $(j+1)-th$ iteration is calculated with the new wave functions and energies, obtained from the new potential.

$$\rho_q^{(j+1)} = \rho_q\left(E_{nk_z}^{(j+1)}, \boldsymbol{\xi}_{nk_z}^{(j+1)}(\mathbf{r})\right). \qquad (4.57)$$

4. The convergence is checked for the charge density in the last two iterations. The
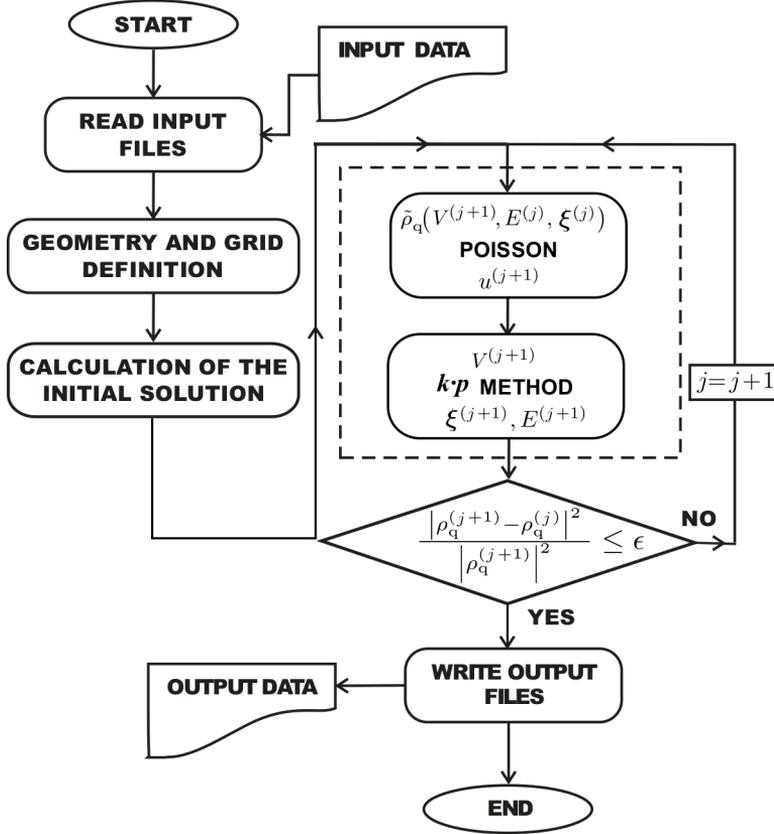
Figure 4.11: Flow diagram of the simulator SP2D. The iterative part is remarked in a dashed rectangle.

relative error must be below a limit ϵ_{q}.

$$\frac{\left|\rho_{\mathrm{q}}^{(j+1)}(\mathbf{r}) - \rho_{\mathrm{q}}^{(j)}(\mathbf{r})\right|^2}{\left|\rho_{\mathrm{q}}^{(j+1)}(\mathbf{r})\right|^2} \leq \epsilon_{\mathrm{q}}\,. \tag{4.58}$$

The predicted value of charge, $\tilde{\rho}_{\mathrm{q}}(\mathbf{r})$, is calculated as in Ref. [134], although taking into account that for $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ the band profile is not parabolic and has to be integrated numerically from the expression:

$$\tilde{\rho}_{\mathrm{q}}[V(\mathbf{r})] = \frac{1}{2\pi}\sum_n \int \left|\boldsymbol{\xi}_{nk_{\mathrm{z}}}^{(j)}(\mathbf{r})\right|^2 f_0\!\left(\frac{E_F - E_{nk_{\mathrm{z}}}^{(j)} + q\big(V(\mathbf{r}) - V^{(j)}(\mathbf{r})\big)}{k_B T}\right) \mathrm{d}k_{\mathrm{z}}\,, \tag{4.59}$$

where the probability distribution of an state $|n, k_z\rangle$ is calculated as the square modulus of the vectorial envelope function $\left|\boldsymbol{\xi}_{nk_z}^{(j)}(\mathbf{r})\right|^2$, and f_0 is the Fermi distribution, which depends in the position. Therefore, the prediction is based in the modification of the probability distribution of every state by using the Fermi distribution as a position dependent weighting function. As the difference between the potential of two consecutive iterations is reduced, Eq. (4.59) tends to fulfill the convergence criteria.

The application of this algorithm achieves a remarkable acceleration of the convergence, as it was already shown in Ref. [134].

4.7 Performance Improvements

The self-consistent solution of the $\boldsymbol{k}\cdot\boldsymbol{p}$ method and the Poisson equation demands a large amount of computational resources, particularly, when implementing the $\boldsymbol{k}\cdot\boldsymbol{p}$ calculation. This method has to be solved for each point of the discretized k_z grid, and each one involves: 1) the creation of all the matrices of the system, with size is $N_c N_p \times N_p N_c$, and 2) the calculation of the eigenvalues of the system, and the storage of all the resulting wave functions. The calculation of the eigenvalues is usually the most expensive task due to the size of the equations system. Moreover, we have to recall that the $\boldsymbol{k}\cdot\boldsymbol{p}$ method is usually solved several times until convergence is achieved. Hence, any simplification in the algorithm can result in a considerably reduction in the computational cost and time. Next, we enumerate the main optimizations carried out to the program.

4.7.1 Optimization of the calculation of H

Regarding the first point, the matrix $\boldsymbol{H}$ must be created for each k_z value. Although it is not the most time consuming part of the algorithm, it has to be solved several times for each iteration. However, once we have calculated the matrix for an specific k_z, this calculation can be skipped for the rest of k_z values. Let us detail the procedure.

As it is well known, the $\boldsymbol{k}\cdot\boldsymbol{p}$ matrix in confined systems turns into an operator matrix $\hat{\mathbf{H}}$ which depends on the momentum operators of first and second order referred to the crystal reference system, $\hat{\mathbf{H}}(\nabla^2, \nabla)$. The operators in this reference system can be expressed as a linear combination of the operators in the device reference system, as seen in Sec. 4.5.4. After the discretization of the problem, the operators are evaluated

as a linear combination of the terms gathered in Tables 4.1 and 4.2, which either depend on k_z or k_z^2, or are constants. Therefore, the operators corresponding to the momenta of first and second order for the construction of the $\boldsymbol{k} \cdot \boldsymbol{p}$ matrix yield a polynomial function of first and second order of k_z when the FEM is applied:

$$
\begin{aligned}
\nabla &\rightarrow \mathbf{p}_1(k_z) \,, \\
\nabla^2 &\rightarrow \mathbf{p}_2(k_z) \,.
\end{aligned}
\tag{4.60}
$$

Consequently, the whole $\boldsymbol{k} \cdot \boldsymbol{p}$ matrix is also a linear combination of powers of k_z, and can be written as:

$$
\boldsymbol{H} = \boldsymbol{H}_0 + \boldsymbol{H}_1 k_z + \boldsymbol{H}_2 k_z^2 \,.
\tag{4.61}
$$

We refer to the matrices $\boldsymbol{H}_0$, $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ as matrices of zeroth, first and second order, respectively. They are calculated accounting for the constant, first order and second order parts of the operators. The advantage of calculating these matrices, is that they depend on the geometry of the problem and the potential V of the iteration, but are constant for any value of k_z. Consequently, by calculating them once for an iteration, the matrix system for the rest of values of k_z is simply calculated using Eq. (4.61), avoiding the costly computational process of recalculating the system matrix from scratch.

4.7.2 Calculation for positive k_z

The positive k_z states and the negative k_z states are related to each other. So that, it is not necessary to solve both systems if the relationship is known, saving as much as half of the computational time and memory. This relationship stems from the time reversal property of a physical system. Since the Hamiltonian is invariant under time reversal, the following pair of eigenstates linked by the time reversal operator will be degenerate:

$$
|\boldsymbol{k}, \uparrow\rangle \xrightarrow{\Theta} |-\boldsymbol{k}, \downarrow\rangle \,,
\tag{4.62}
$$

where Θ is the time reversal operator. Eq. (4.62) means that for an state $\boldsymbol{k}$, or in our case, k_z, the opposite state $-k_z$ is the result of time reversal symmetry, and it has the same energy and opposite spin. Therefore, the energy levels are known and there exists

a link between the wave functions. Thus,If this relationship is known, the problem does not need to be solved for negative values of k_z.

Let us start with the six band $\boldsymbol{k}\cdot\boldsymbol{p}$ as an example, and then, extend the reasoning to the eight band $\boldsymbol{k}\cdot\boldsymbol{p}$. Equation (4.62) suggests that the relationship between the wave functions in $\pm k_z$ lies on the spin eigenstates. The six band $\boldsymbol{k}\cdot\boldsymbol{p}$ hamiltonian uses the orbital states $|X\rangle$, $|Y\rangle$ and $|Z\rangle$ with their corresponding up and down spin as a basis. Not applying by now the spin orbit coupling, i.e., taking into account only the $\boldsymbol{k}\cdot\boldsymbol{p}$ operator matrix $\hat{\mathbf{H}}(k_z)$, the application of the time reversal operator results into its Hermitian:

$$\hat{\mathbf{H}}(-k_z) = \Theta\hat{\mathbf{H}}(k_z) = \hat{\mathbf{H}}^\dagger(k_z)\,. \tag{4.63}$$

In this case, the eigenfunctions for $-k_z$ are the complex conjugate of the eigenfunctions for k_z. However, we have seen that these states mix due to the spin orbit coupling term. Therefore, in this basis, we cannot relate the wave functions with their opposite. To do that, the $\boldsymbol{k}\cdot\boldsymbol{p}$ Hamiltonian must be represented in a different basis consisting of the eigenfunctions of the spin orbit interaction. As already seen in Sec. 3.2.2, there is a rotation matrix $\boldsymbol{R}_{\mathrm{SO}}$ which diagonalizes the spin orbit matrix $\mathbf{M}_{\mathrm{SO}}$. This rotation mixes the original states of the Hamiltonian to obtain the spin eigenfunctions corresponding to the states $|HH\rangle$, $|LH\rangle$ and $|SO\rangle$ and their corresponding up and down spins:

$$\mathbf{M}_{\mathrm{SO}} = \boldsymbol{R}_{\mathrm{SO}}\,\boldsymbol{D}_{\mathrm{SO}}\,\boldsymbol{R}_{\mathrm{SO}}^\dagger\,. \tag{4.64}$$

Then, applying this rotation to the full $\boldsymbol{k}\cdot\boldsymbol{p}$ Halmitonian

$$\hat{\mathbf{H}}\boldsymbol{\xi} = E\,\boldsymbol{\xi}\,, \tag{4.65}$$

$$\boldsymbol{R}_{\mathrm{SO}}(\hat{\mathbf{H}}_{\mathrm{D}} + \boldsymbol{D}_{\mathrm{SO}})\boldsymbol{R}_{\mathrm{SO}}^\dagger\,\boldsymbol{\xi} = E\,\boldsymbol{\xi}\,, \tag{4.66}$$

$$(\hat{\mathbf{H}}_{\mathrm{D}} + \boldsymbol{D}_{\mathrm{SO}})\boldsymbol{R}_{\mathrm{SO}}^\dagger\,\boldsymbol{\xi} = E\,\boldsymbol{R}_{\mathrm{SO}}^\dagger\boldsymbol{\xi}\,, \tag{4.67}$$

where $\hat{\mathbf{H}}_{\mathrm{D}}$ is the representation of the $\hat{\mathbf{H}}$ matrix in the basis of the eigenfunctions of the spin states. The rotation matrix $\boldsymbol{R}_{\mathrm{SO}}$ is an antiunitary operator which conserves the norm of the eigenfunctions and does not affect to the eigenvalues either. The result

is a new Hamiltonian in the spin basis for which the eigenfunctions are:

$$\boldsymbol{\xi}_{\mathrm{D}} = \boldsymbol{R}_{\mathrm{SO}}^{\dagger}\boldsymbol{\xi}\,. \tag{4.68}$$

In this new basis, the Hamiltonian for the opposite wave vector $\hat{\mathbf{H}}_{\mathrm{D}}(-k_{\mathrm{z}})$ can be calculated from the direct wave vector Hamiltonian $\hat{\mathbf{H}}_{\mathrm{D}}(k_{\mathrm{z}})$ as a combination of complex conjugate and a shuffling of the up and down spin states:

$$\hat{\mathbf{H}}_{\mathrm{D}}(-k_{\mathrm{z}}) = \varTheta\hat{\mathbf{H}}_{\mathrm{D}}(k_{\mathrm{z}}) = (-1)\Big(\boldsymbol{S}\,\hat{\mathbf{H}}_{\mathrm{D}}(k_{\mathrm{z}})\,\boldsymbol{S}\Big)^{*}. \tag{4.69}$$

where S depends on the $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ method employed. For six band $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ method:

$$\boldsymbol{S}_{6} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}, \tag{4.70}$$

is the shuffling matrix which interchanges the corresponding spin up and down states.

According to this transformation, now the wave functions of the state k_{z} and his opposite can be readily calculated as:

$$\boldsymbol{\xi}_{\mathrm{D},-k_{\mathrm{z}}} = S\,\boldsymbol{\xi}_{\mathrm{D},k_{\mathrm{z}}}^{*}\,. \tag{4.71}$$

This transformation conserves the norm, and therefore, the spacial probability distribution of the eigenstate.

For the eight band $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$, the reasoning is similar, since the spin orbit coupling is the same for the valence band states and they do not couple with the conduction band ones. Therefore, the reasoning can be easily expanded using the following shuffling matrix :

$$S_8 = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}. \tag{4.72}$$

Finally, the S matrix for the two band $\boldsymbol{k} \cdot \boldsymbol{p}$ method reads:

$$S_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{4.73}$$

and the $\boldsymbol{R}_{\mathrm{SO}}$ is the identity since in the two conduction bands in this method does not couples by means of the spin.

4.7.3 Optimization of the eigenvalue interval

The slowest part of the program corresponds to the search of the eigenvalues. The Arnoldi algorithm is used to solve the problem, which slows down when the number of eigenvalues in the interval of search is large. The optimization of the algorithm, which is provided by the *pdetool* package, is not an easy task, since it already provides an efficient computational implementation. However, we can reduce the calculation time by carefully choosing the interval where the eigenvalues are sought.

One of the limits of the interval is clear: the edge of the conduction or valence band. The other is linked, but not solely, to the probability of occupation of the state, i.e., the Fermi distribution function, which determines the weight of the state. Therefore, the width of the interval will depend on the precision required for the algorithm, and which probabilities are considered negligible. The main problem is that, a priori, these limits are not well known, and the interval is usually chosen arbitrarily large to ensure that all the relevant states are included.

In order to determine the band edges, a simple approximation can be done, taking into account that no energy levels can lie in the gap. A better approximation can be done if the conduction and valence band shifting due to the quantization can be
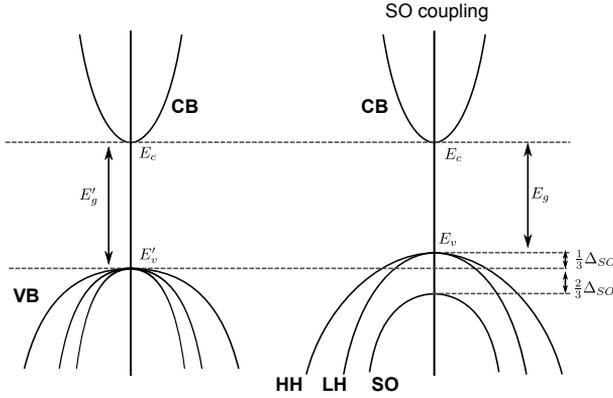
Figure 4.12: Representation of the edges of the bulk band structure in a potential well: the flat well is represented in black and the self-consistent potential, in red. The position of the resulting bandstructure edge is marked with a dashed lines.

evaluated. As the potential yielding the lowest quantization is the flat one (Fig. 4.12), we use it as a minimum energy solution. The quantization shifting due to a flat well can be calculated for a wire, and used to correct the limits estimated in the bulk case.

Once the bandstructure is known, it is easy to readjust the interval in order to accelerate subsequent iterations. In this case, the definition of the total charge distribution as a function of the energy is useful to accurately determine which energies contribute noticeably to the total charge, and which ones have a negligible contribution. Therefore, we can again use information coming from previous iterations to get the final estimation of the optimal limits of the interval and, thus, reduce the computation time.

4.7.4 Memory saving

The memory usage is another aspect to care about when programming, since the management of big amounts of data and the overflow issue affects severely the performance. The main memory consumer structure is that corresponding to the wave function at each state. We spotlight some important considerations to reduce its size.

The calculation of the band structure only for positive k_z values implies a reduction of the memory storage to the half. However, the amount of consumed memory is still huge due to the amount of wave functions to be stored. In principle, each wave function needs to store $N_c \times N_n$ imaginary numbers. Furthermore, the wave functions are calculated at each value of the k_z grid, with size N_k. Let us assume that we want to

calculate a maximum of N_s subbands. If all the data are saved in a single array, a very large four dimensional array is required, with dimensions $N_c \times N_n \times N_s \times N_k$. A better solution is to use a matrix of $N_s \times N_k$ pointers to save the defined wave functions, as for many k_z values most of the wave function are not evaluated.

Another reduction of the size can be achieved by only keeping in the program memory the squared wave function, corresponding to the probability distribution of the state, which is needed to evaluate the charge distribution of each state. By doing this, each wave function is reduced to N_n real numbers, instead of $N_c \times N_n$ imaginary numbers. As the complete wave functions are required for the mobility calculations, they cannot be discarded, but can be stored in the disk instead of keeping them in the program memory.

4.7.5   Estimation of the bandstructure

Some additional improvements in terms of computational time efficiency can be achieved by carefully choosing the initial solution of the problem before starting the iteration. We can assume, once the convergence is assured by the algorithm described above, that the closer the initial solution is to the sought one, the faster the algorithm will converge. Some rough choices are the zero potential solution, or the solution achieved after solving the classical problem. However, in this type of problem, a battery of simulations is performed varying the gate potential. As a consequence, usually some results have been previously calculated for close V_G values which can yield some valuable information to accelerate the convergence of our problem.

For a new value of V_G, the external boundary condition is changed. The immediate effect is a shifting on the Fermi level, and therefore, the occupation probability of the states in the bandstructure and the charge distribution. This charge affects the potential, which, in turn, changes the bandstructure.

Using the solutions for a previous gate potential, the Hamiltonian of which as a function of k_z is $\hat{H}_0(k_z)$, and its eigenfunctions, ξ_{0,nk_z}, the new Hamiltonian can be expressed as a perturbed one:

$$\left(\hat{H}_0(k_z) + \Delta V(\mathbf{r})\right)\boldsymbol{\xi}_{nk_z} = W_{nk_z}\boldsymbol{\xi}_{nk_z} \tag{4.74}$$

where the perturbation is the variation of the potential in the structure $\Delta V(\mathbf{r})$ between the two Hamiltonians, with

$$W_{nk_{\mathrm{z}}} \approx W_{nk_{\mathrm{z}}}^{(0)} + W_{nk_{\mathrm{z}}}^{(1)} = E_{0,nk_{\mathrm{z}}}(\boldsymbol{k}) + \langle \boldsymbol{\xi}_{0,nk_{\mathrm{z}}} | \Delta V(\mathbf{r}) | \boldsymbol{\xi}_{0,nk_{\mathrm{z}}} \rangle \tag{4.75}$$

and we can calculate the resulting variation of the new energies with respect to the previous ones, $\Delta E_{nk_{\mathrm{z}}}$, for each state $|n, k_{\mathrm{z}}\rangle$, as:

$$\Delta E_{nk_{\mathrm{z}}} = \langle \boldsymbol{\xi}_{0,nk_{\mathrm{z}}} | \Delta V(\mathbf{r}) | \boldsymbol{\xi}_{0,nk_{\mathrm{z}}} \rangle \, . \tag{4.76}$$

Only the first order term of the perturbation is considered, since the second order term vanishes for nearly constant values of ΔV. Weighting each state with its probability function and summing over all the states we have

$$
\begin{aligned}
\sum_n \int f_0(E_{nk_{\mathrm{z}}}) \Delta E_{nk_{\mathrm{z}}} \, \mathrm{d}k_{\mathrm{z}} &= \sum_n \int f_0(E_{nk_{\mathrm{z}}}) \langle \boldsymbol{\xi}_{nk_{\mathrm{z}}} | \Delta V(\mathbf{r}) | \boldsymbol{\xi}_{nk_{\mathrm{z}}} \rangle \, \mathrm{d}k_{\mathrm{z}} \\
&= \sum_n \int f_0(E_{nk_{\mathrm{z}}}) \int |\boldsymbol{\xi}_{nk_{\mathrm{z}}}|^2 \Delta V(\mathbf{r}) \, \mathrm{d}\mathbf{r} \, \mathrm{d}k_{\mathrm{z}} \\
&= \int \Delta V(\mathbf{r}) \sum_n \int f_0(E_{nk_{\mathrm{z}}}) |\boldsymbol{\xi}_{nk_{\mathrm{z}}}|^2 \, \mathrm{d}k_{\mathrm{z}} \, \mathrm{d}\mathbf{r} \\
&= \int \Delta V(\mathbf{r}) \rho_{\mathrm{q}}(\mathbf{r}) \, \mathrm{d}\mathbf{r} \, .
\end{aligned}
\tag{4.77}
$$

For small or nearly constant $\Delta V(\mathbf{r})$ values, we can expect a small variation of the bandstructure, and the shift of the gate potential can be accounted for as a shifting of the Fermi level. This effect is shown in Fig. 4.13, where, for different values of V_{G}, we observe that the first subband of a 5nm width [001]-oriented Si NW, is mainly shifted, but does not changes its shape dramatically. As a result, the ΔE is constant for all the states, and therefore

$$\sum_n \int f_0(E_{nk_{\mathrm{z}}}) \Delta E \, \mathrm{d}k_{\mathrm{z}} = \Delta E \int \rho_{\mathrm{q}}(\mathbf{r}) \, \mathrm{d}\mathbf{r} \, . \tag{4.78}$$

From Eqs. (4.77) and (4.78), a prediction of the new bandstructure is the old one shifted:

$$\Delta E = \frac{\int \Delta V(\mathbf{r}) \rho_{\mathrm{q}}(\mathbf{r}) \, \mathrm{d}\mathbf{r}}{\int \rho_{\mathrm{q}}(\mathbf{r}) \, \mathrm{d}\mathbf{r}} \, . \tag{4.79}$$

This expression corresponds to the variation of the potential energy averaged with
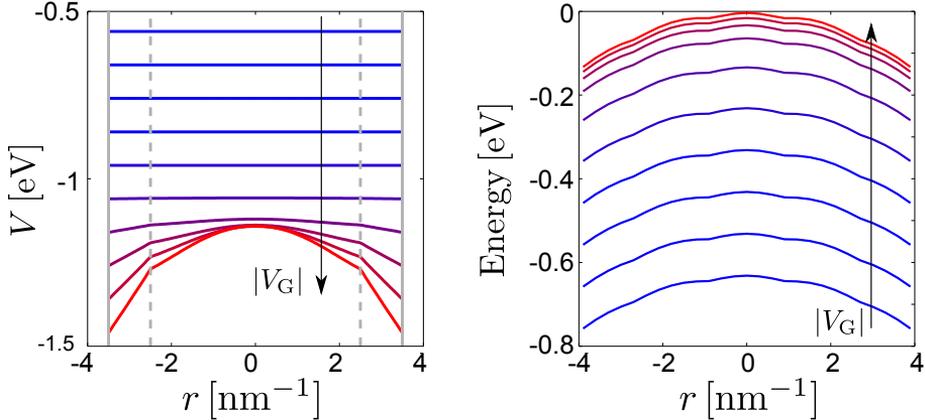
Figure 4.13: Behavior of the self consistent potential for different gate voltages along the radial direction r of a 5nm width [001]-oriented Si NW. The arrow aims toward higher values of $|V_G|$. On the left, the cross-section of the potential is plotted. On the right, the first subband is plotted, using as a reference the Fermi level. The gray dashed lines denote the oxide-semiconductor interface position.

the charge distribution ρ_q. According to that, the variation of the Fermi level is expected to be equivalent to the variation of the potential energy.

Therefore, based on a previous result, we can perform a self consistent loop where the $\mathbf{k} \cdot \mathbf{p}$ calculation is substituted by the calculation of the new Fermi level for the previous bandstructure. The solution of this loop will be a reasonable approximation to start the final self-consistent loop, where the $\mathbf{k} \cdot \mathbf{p}$ is used. Let us recall that if this approximation fails to predict a good solution, it will be corrected by the posterior self-consistent loop.

Fig. 4.14 illustrates the potential and the charge distribution for each iteration before the convergence as dashed lines. The final solution is depicted as a thicker solid line. As can be seen, the first iteration is a reasonable prediction of the final value.

This algorithm is specially useful for the low inversion calculations, where the potential is flat. In this case, the self-consistent solutions are a mere shifting of the Fermi level, which is calculated by this algorithm. Therefore, if there is a previous solution in the low inversion regime, no iterations of the $\mathbf{k} \cdot \mathbf{p}$ algorithm are required.
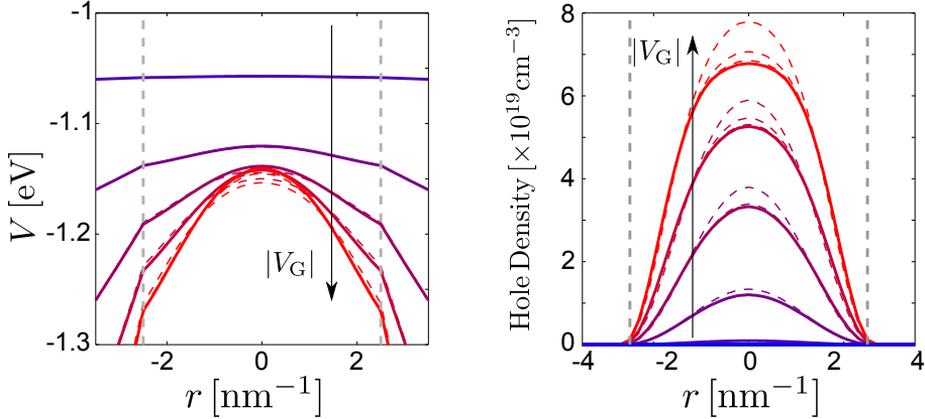
Figure 4.14: Representation of the potential and the charge density for each iteration at different gate voltages along the radial direction r of a 5nm width [001]-oriented Si NW. The arrow aims toward higher values of $|V_\mathrm{G}|$. The dashed lines stand for the iterations before the convergence, the solid thicker line is the final result. On the left, a zoomed version of the potential shown in Fig. 4.13 for the higher gate voltages. On the right, the hole density for each iteration along the NW diameter. The gray dashed lines denote the oxide-semiconductor interface position.

4.8   Summary

A self-consistent Schrödinger-Poisson solver was developed to simulate nanowires FETs, which is able to deal with different materials, geometries and orientations. The $\boldsymbol{k}\!\cdot\!\boldsymbol{p}$ method is used as an approximation of the Schrödinger Equation to calculate the band-structure and the wave functions, leading to the electrostatic description of the device. In order to deal with the numerical problem, a Finite Element Method has been implemented. A comprehensive description of the discretization problem has been provided, and the process followed to lay out a partial differential equations system. Finally, some particularities of the code such as the optimization strategies are exposed.

Chapter 5

SP2D simulator evaluation and validation

Once the theoretical bases and the implementation of the SP2D simulator have been introduced, the next step is the validation and analysis of its results. The main topic of this thesis is the implementation and use of the $\boldsymbol{k}\cdot\boldsymbol{p}$ method to calculate the band structure of different materials and device structures. So that, the aim of this chapter is twofold. First, we will check that the simulation results are coherent and there are no errors on its implementation; second, we will check that the results provided using the $\boldsymbol{k}\cdot\boldsymbol{p}$ method are accurate in the description of the band structure for multigate devices.

As already mentioned in Chapter 3, the $\boldsymbol{k}\cdot\boldsymbol{p}$ method is appropriate to estimate the band edges in bulk materials [17, 129], providing a more accurate description than the effective mass approximation, which is no longer valid to describe the valence band in diamond and zinc blende semiconductors. However, the $\boldsymbol{k}\cdot\boldsymbol{p}$ method still holds the continuum approach, which means that it avoids the atomistic description of the material. This characteristic is desirable in terms of simplicity and reduction of the computational burden. Moreover, this method is also able to capture the physics related to the symmetry of the underlying crystalline structure and the electronic states mixing.

Most of the electrostatic and transport properties of a semiconductor device are determined by the carries located around the band edges. So that, as long as the $\boldsymbol{k}\cdot\boldsymbol{p}$ method provides an accurate description of the band structure around the Fermi level, it can be considered as an appropriate choice for our calculations.

Our SP2D code is intended to provide a good physical description for devices with

2D confinement. In particular, we are interested in the performance of NWs, as they have demonstrated a good control of the SCEs. In some indirect materials, such as Si and Ge, the band structure can be accurately reproduced by considering that there is no coupling between the conduction and the valence band. Consequently, the valence band is depicted considering a six band $\boldsymbol{k \cdot p}$ method, described in Sec. 3.2.2, and the conduction band is calculated making use of the EMA accounting for non parabolicity, or a two band $\boldsymbol{k \cdot p}$ method for Si. However, for III-V materials with zinc blende structure, the energy gap is much smaller. Thus, the CB and the VB are close enough so as to couple to each other. In this case, an eight band $\boldsymbol{k \cdot p}$ method that couples the VB with the CB becomes mandatory (Sec. 3.3).

The validation of the results obtained with our simulator is carried out through the comparison with a different approach: the Tight Binding method (TB). This method considers the position of each atom in the structure explicitly (see Fig. 5.1). Hence, it deals with the confinement of the devices in a transparent manner. On the other hand, the $\boldsymbol{k \cdot p}$ method makes an infinite bulk assumption to calculate the electrostatic properties of an infinite lattice. The envelope function method, outlined in Section 3.6, must be used to apply the $\boldsymbol{k \cdot p}$ method to confined systems in order to adapt a continuum approach to the confined device. Regarding the theoretical bases of both approaches, the Tight Binding method provides a full band description, not limited to the band edges. Notwithstanding, the $\boldsymbol{k \cdot p}$ method has demonstrated to be effective in the description of the band structure near the band edges, which are the energy intervals of interest for electronic applications [143]. For example, it is widely used for the calculation of the bulk band structure around the Γ symmetry point ($\boldsymbol{k} = 0$), where the top of the Valence Band (VB) and the bottom of the Conduction Band (CB) are placed for most of the III-V compound semiconductors. Since the $\boldsymbol{k \cdot p}$ is a computationally efficient algorithm, it becomes a very interesting alternative to the more complex atomistic approaches. Furthermore, it can be described with a reduced set of parameters commonly used in the literature. Therefore, the aim is to unveil the extent to which this method can accurately reproduce the band structure of the NWs.

The TB simulation data used in our comparisons have been gently provided by Dr. Triozon and his research group in CEA-LETI, Grenoble (France). The TB simulations account for the sp^{3}d^{5}s* orbitals [61].
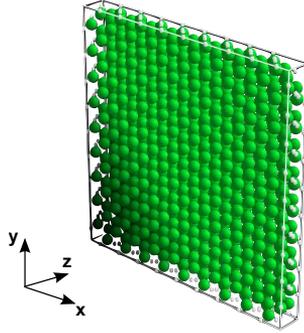
Figure 5.1: Unit cell employed with the Tight Binding Model for a squared $5\,\text{nm} \times 5\text{nm}$ standalone Si nanowire. The nanowire is free standing and passivated with hydrogen.

5.1 Validation of the two-band $\boldsymbol{k{\cdot}p}$ simulator

As seen in Chapter 3, the simplest implementation of the $\boldsymbol{k{\cdot}p}$ method is the two-band model of the conduction band in Si. This approximation is very similar to the EMA but implicitly accounts for the non parabolicity, and strain effect can be easily added [125]. Although this is not the main focus of this manuscript, we consider it as a good starting point to validate our simulator. The $\boldsymbol{k{\cdot}p}$ matrix is simple enough to test and validate the FEM implementation, debug the program, and analyze the first results of the $\boldsymbol{k{\cdot}p}$ method, which can be compared with those shown by other authors in the literature as well as from our collaboration with LETI.

The two-band $\boldsymbol{k{\cdot}p}$ method calculates the interaction between two adjacent Δ valleys found at $\pm 0.85 \frac{2\pi}{a_0}[1,0,0]$, close to the X point (being a_0 the lattice constant). Because of the periodicity of the band structure, another Δ valley lies just in the opposite place of the X-point. As a consequence, these two valleys are close enough to interact, specially when strong confinement is applied. Let us recall that there are three equivalent Δ directions in the reciprocal space, and therefore six equivalent conduction valleys on Si. When 2D confinement is considered, these valleys project on the transport direction. This projection depends on the orientation of the nanowire with respect to the crystal reference system [125]. As an example, for the [001]-oriented NWs, four of the valleys project on the Γ point, whereas the other two remain shifted from that point. The former are called the unprimed subbands, and the latter, the primed ones. This terminology is used hereinafter to denote the valleys of the conduction band in any
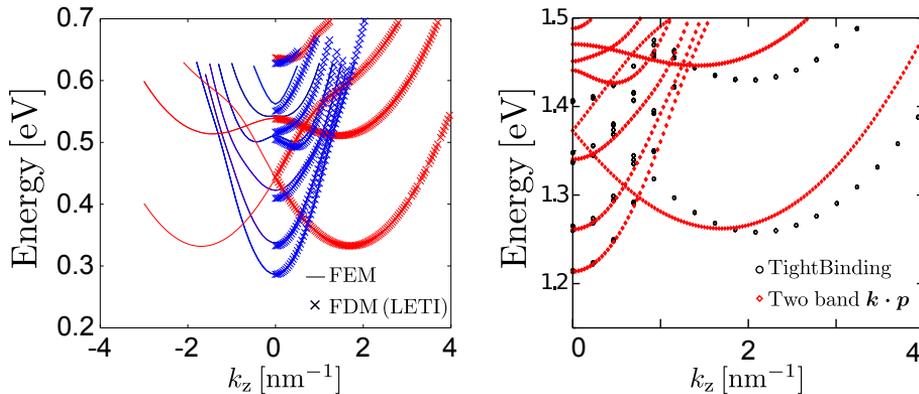
Figure 5.2: Comparisons of the different implementations of the 2 band $\boldsymbol{k}\cdot\boldsymbol{p}$ method for the conduction band of a 5 nm × 5 nm squared Si-NWs. Left figure: the FEM method (UGR-solid line) is compared to the FDM (LETI-crosses). Right figure: comparison between the $\boldsymbol{k}\cdot\boldsymbol{p}$ method and the TB method.

orientation.

The first comparison to validate our simulator has been performed against a two-band $\boldsymbol{k}\cdot\boldsymbol{p}$ simulator developed at LETI. The main difference is that this simulator is implemented using the Finite Difference Method. In order to perform a fair comparison, the size of the elements and differences has to be very small in the two implementations in order to minimize the numerical error: 10000 triangles for FEM and 0.5Å for FDM. The simulations have been performed for squared Si nanowires, since the FDM is conceived for rectangular structures. The size of the device is 5 nm × 5 nm, it is oriented in the [001] direction, and an infinite potential well is considered.

The left plot in Fig. 5.2 illustrates the results of these two simulations, using solid lines for FEM and crosses for FDM. The primed subbands are depicted in red, and the unprimed ones, in blue. As can be observed, the results match almost perfectly, except for some small shift in some of the high energy subbands. We can conclude that the implementations of the FEM and FDM tend to the same solution when their meshes are sufficiently refined.

Then, the $\boldsymbol{k}\cdot\boldsymbol{p}$ method is compared to the Tight Binding method. The results are represented in the right plot of Fig. 5.2. Both simulations match in terms of energies and curvature. Notwithstanding, the valleys along the transport direction are slightly shifted in k_z. This shift is reported in the literature to depend on the set of parameters used to implement the TB [89] and can be easily reproduced by the $\boldsymbol{k}\cdot\boldsymbol{p}$
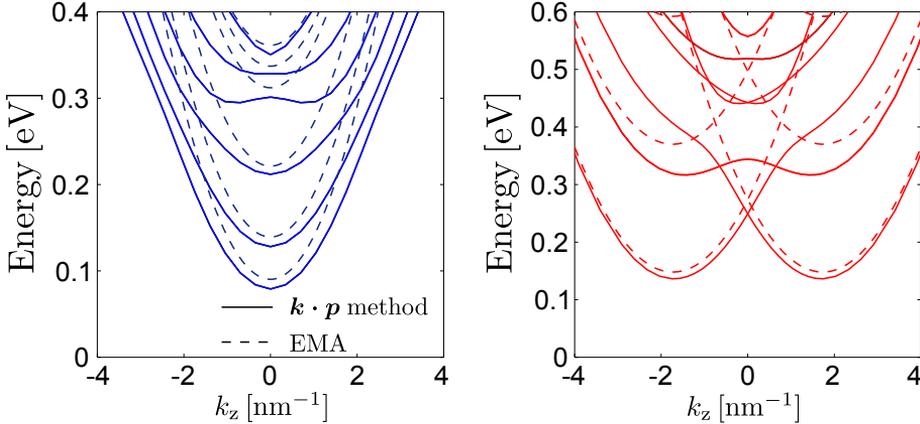
Figure 5.3: Comparison of the two band $\mathbf{k} \cdot \mathbf{p}$ method and the EMA for Si $5 \, \text{nm} \times 5 \, \text{nm}$ [001] squared NWs. The unprimed valleys are plotted on the left and the primed ones on the right.

method changing the position k_0 of the minimum of the valleys. Nevertheless, this fact does not affect the electrostatic behavior of the device, as the density of states and the position of the subbands are correctly described. Therefore, we conclude that the two-band $\mathbf{k} \cdot \mathbf{p}$ method is able to yield an accurate description of the band structure for the conduction band in Si nanowires.

We can also observe that both approaches provide nearly parabolic bands as a result. Because of that, we wonder if the EMA could also reproduce these results. The comparison is shown in Fig. 5.3 (for the same nanowire as in Fig. 5.2). As can be seen, the EMA represents moderately well the shape of the valleys. In the case of the unprimed subbands, it nearly reproduces exactly the quantization of the levels and the effective mass. However, the EMA is not able to capture the non parabolicity by itself, and this has to be done by include a non parabolic effective mass approach. As for the primed subbands, the effect of the coupling between the two adjacent valleys is evident: the lowest subbands match perfectly in the two approaches, whereas, for higher subbands, the EMA cannot reproduce the coupling between the two valleys. This coupling becomes more noticeable for smaller sizes and under strain conditions, as demonstrated in [125], justifying the use of the $\mathbf{k} \cdot \mathbf{p}$ approach for small size NWs.

5.2   Band structure for the valence band

Once we have tested the first results of the SP2D simulator and we have proved it works properly, we continue with the validation of the $\boldsymbol{k}\cdot\boldsymbol{p}$ method for the valence band. As explained in Chapter 3, the $\boldsymbol{k}\cdot\boldsymbol{p}$ method has been widely used for the description of the VB of diamond and zinc blende semiconductors. In particular, for diamond bulk semiconductors the six band $\boldsymbol{k}\cdot\boldsymbol{p}$ method is proved to yield good results. However, for zinc blende semiconductor compounds, the reduced and direct gap suggests that there exists interaction between the CB and the VB that must be accounted for. An extension of the six band $\boldsymbol{k}\cdot\boldsymbol{p}$ method, accounting for the $|S\rangle$ states of the CB, is expected to achieve a good description, resulting in the eight band $\boldsymbol{k}\cdot\boldsymbol{p}$ method. Therefore, we assume that the validation of the results achieved with the eight band $\boldsymbol{k}\cdot\boldsymbol{p}$ for NWs must automatically validate the six band approach.

The aim of this Section is the assessment of the results obtained for bulk materials by the eight band $\boldsymbol{k}\cdot\boldsymbol{p}$ method and the TB, so that we can compare and adjust the $\boldsymbol{k}\cdot\boldsymbol{p}$ parameters. To this end, we will analyze two III-V alloys: GaAs and InAs. Both of them are direct gap materials. They have been chosen because the GaAs has a relatively high energy gap, (1.41 eV), compared with the rest of III-V materials, whereas InAs has a small one (0.37 eV). This difference will help us later to evaluate the interaction of the VB and the CB in NWs.

Both methods are defined by a set of semi-empirical parameters. As a starting point, we must define the parameters which provide the best fitting between both methods. First, we use parameters reported in the literature by *Veprek et al.* [140], and *Vurgaftman et al.* [143] for the $\boldsymbol{k}\cdot\boldsymbol{p}$ simulations; and those by *Boykin et al.* [19], and *Jancu et al.* [61] for the TB simulations provided by CEA-LETI.

The results depicted in Fig. 5.4 demonstrate that the $\boldsymbol{k}\cdot\boldsymbol{p}$ method provides the band structure description around the Γ point, which is the region of interest for direct gap semiconductors. The $\boldsymbol{k}\cdot\boldsymbol{p}$ results given by the two set of data are quite similar, and they are in good agreement with the TB simulations around the Γ point for both CB and VB. Nevertheless, satellite valleys of the CB placed in the L and X point respectively, are not modeled by $\boldsymbol{k}\cdot\boldsymbol{p}$. Normally, these valleys are weakly populated for a bulk material at low inversion charges and can be neglected. However, we must be aware of their existence and of their effect when confinement is present. Nevertheless, as the employed parameters have been obtained from different authors, some mismatches
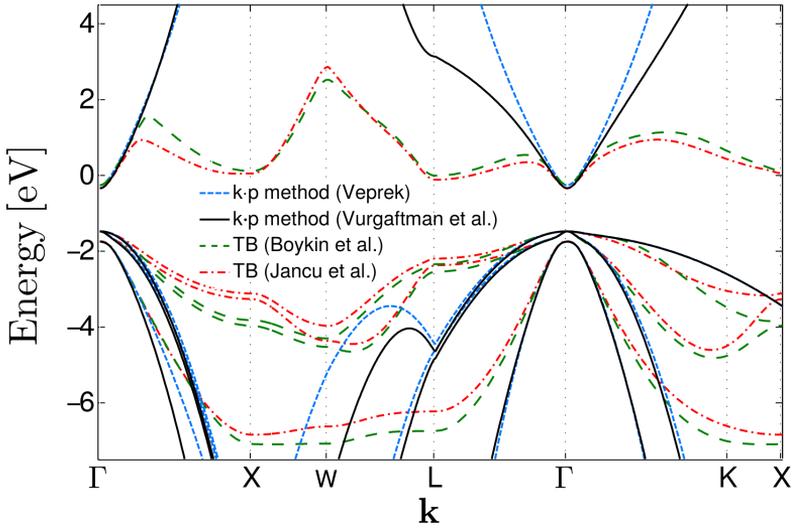
Figure 5.4: Band structure for the bulk GaAs calculated using two different methods: eight band $\boldsymbol{k}\cdot\boldsymbol{p}$ and TB. For each method two different set of parameters have been employed: those provided by *Veprek et al.* (Ref. [140]), and *Vurgaftman et al.* (Ref [143]) for the $\boldsymbol{k}\cdot\boldsymbol{p}$ simulations; and those by *Boykin et al.* (Ref. [19]), and *Jancu et al.* (Ref. [61]), for the TB simulations. For the TB simulations, only the bands with correspondence with the $\boldsymbol{k}\cdot\boldsymbol{p}$ simulations have been plotted.

in the gap and curvature are found. A better agreement can be achieved by tuning these bulk parameters so as to make a fairer comparison. Later, we will achieve a better fitting of the $\boldsymbol{k}\cdot\boldsymbol{p}$ simulation to those of the TB obtained using the *Boykin et al.* [19] parameters as a reference.

However, the small mismatches are not so relevant compared with another effect also observed in the figure: the CB warps down near the L point for the $\boldsymbol{k}\cdot\boldsymbol{p}$ parameters from Ref. [143]. If we took a look at the evolution of this band outside the first BZ, i. e. , beyond the L point, we would observe that its warping leads to a conduction band tending to minus infinity, and therefore, to unphysical states. This effect is not relevant in bulk analysis, when restricting the calculations to the first BZ. However, it is critical in the introduction of spurious solutions when confinement is added, since these unphysical conduction band states outside the BZ interact and are projected into the reduced BZ. Actually, the set of parameters provided in Ref. [139] are free of spurious solutions, as discussed in Sec. 3.7. Consequently, tunning the parameters is essential

to avoid spurious solutions.

Before going on with the comparisons, let us summarize how to obtain the Luttinger parameters from the experimental measurements. As seen in Chapter 3, for the characterization of the valence bands, some parameters are necessary: L, M, N, and Δ_{SO}. However, the most common parameters reported in the literature for the VB are the Luttinger parameters γ_1, γ_2, and γ_3, whose relationship with the L, M and N parameters is shown in Eqs. (3.28 -3.30). These parameters can be in turn calculated from the effective masses as [38]:

$$
\begin{aligned}
\gamma_1 &= \frac{1}{2}\left(\frac{m_0}{m_{LH}^{001}} + \frac{m_0}{m_{HH}^{001}}\right), \\
\gamma_2 &= \frac{1}{4}\left(\frac{m_0}{m_{LH}^{001}} - \frac{m_0}{m_{HH}^{001}}\right), \\
\gamma_3 &= \frac{1}{4}\left(\frac{m_0}{m_{LH}^{111}} - \frac{m_0}{m_{HH}^{111}}\right).
\end{aligned}
\tag{5.1}
$$

Therefore, from the TB simulations, a set of parameters useful to carry out the $\boldsymbol{k}\cdot\boldsymbol{p}$ simulations can be calculated whether the hole effective masses are known. The effective masses we have obtained from the TB simulations of GaAs and InAs are presented in Table 5.1.

The parameters calculated so far correspond to the six band model, and they have to be renormalized by means of the E_P matrix element, according to Eq. (3.41), so as to use them in the eight band model ($\tilde{L}, \tilde{M}, \text{and } \tilde{N}$). As discussed in Sec. 3.7, when the coupling is considered, the Luttinger parameters must be modified and, therefore, may no longer fulfill the ellipticity of the Hamiltonian [140]. When this happens, some modifications must be done to guarantee the ellipticity, and assure a spurious-free solution.

The proposal of *Veprek et al.* [140, 142] is to reduce the parameter E_P, in order to decrease the interaction, and assure that the modified parameters still fulfill the elliptic condition. The smaller the E_P parameter, the more parabolic is the conduction band. This non-parabolicity of the band is the cause of its warping. By reducing it, a spurious-free solution can be achieved, but at the expense of a less accurate description of the non-parabolic conduction band. For this reason, we have considered the E_P used in [140] to guarantee a spurious free solution.

GaAs	[001]	[011]	[111]
Electron mass (m_c)	0. 06574	–	–
Heavy hole mass (m_{HH})	0. 37686	0. 65307	83905
Light hole mass (m_{LH})	0. 08249	0. 07550	0. 07362
Split-Off hole mass (m_{SO})	0. 16243	–	–
Gap (E_g)	1. 41585	–	–
Spin-Orbit Splitting (Δ_{SO})	0. 32648	–	–
InAs	[001]	[011]	[111]
Electron mass (m_c)	0. 02353	–	–
Heavy hole mass (m_{HH})	0. 35158	0. 56111	0. 69813
Light hole mass (m_{LH})	0. 02811	0. 02729	0. 02703
Split-off hole mass (m_{SO})	0. 09959	–	–
Gap (E_g)	0. 36991	–	–
Spin-Orbit Splitting (Δ_{SO})	0. 39317	–	–

Table 5.1: Parameters for GaAs and InAs, obtained from TB simulations along different directions: [001], [011] and [111]. For clarity, when the parameter is isotropic for the three orientations, the second and third cells are filled with '−'. The effective masses are provided relative to the bare electron mass m_0 and the gap in eV.

Taking into account all the previous considerations, we reach to the parameters for the $\boldsymbol{k \cdot p}$ simulations shown in Table 5.2, which will be used hereinafter

The comparison between the bulk band structure for the TB and $\boldsymbol{k \cdot p}$ simulations after the adjustment of the parameters is shown in Figures 5.5 and 5.6 for GaAs and InAs, respectively, showing a very good agreement between the $\boldsymbol{k \cdot p}$ and the TB for the region around the Γ point. As mentioned before, GaAs is a III-V direct gap material with a larger energy gap, but also has a L valley close in energy (0.5 eV) which may become important in some cases. On the other hand, InAs has a much smaller gap resulting on a higher coupling of the CB and the VB. For this material, the rest of the valleys lie on higher energies, and as a consequence, they are scarcely populated. The main divergence between the two approaches lies in the SO band in InAs.

We can also conclude that the evolution of the band structure for high values of k_z cannot be depicted by the eight band $\boldsymbol{k \cdot p}$ approach, and this fact is specially remarkable for the CB of the GaAs. These divergences occur far from the band edges, and therefore, these subbands should not be heavily populated, minimizing their impact on the behavior of the device. Nevertheless, it may become significant in those cases where high fields are applied and the Fermi level penetrates deeply in either the conduction
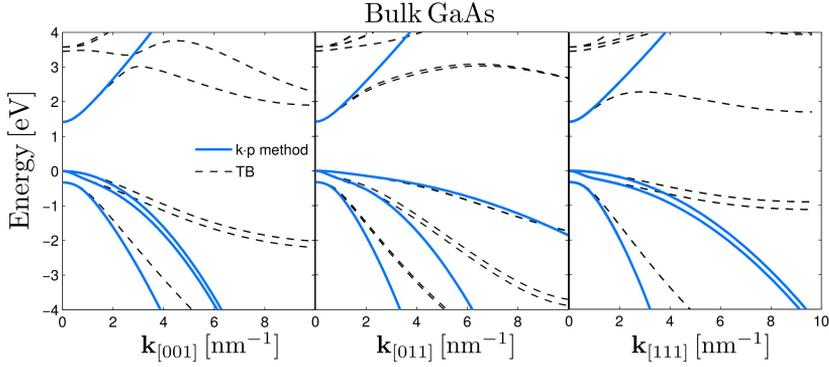
Figure 5.5: Comparison of the bulk GaAs band structure for TB and $\boldsymbol{k}\cdot\boldsymbol{p}$ method along different orientations: [001], [011] and [111].

	GaAs	**InAs**
E_{g}	1. 41585	0. 36991
Δ_{SO}	0. 32648	0. 39316
E_{P}	27. 96 ($\rightarrow$20)	19. 7 ($\rightarrow$18)
m_{c}	0. 06574	0. 02353
γ_1	7. 39	19. 2
γ_2	2. 37	8. 18
γ_3	3. 10	8. 89

Table 5.2: $\boldsymbol{k}\cdot\boldsymbol{p}$ parameters for GaAs and InAs. In parenthesis, the modified E_{P} value to avoid the spurious states.

or the valence band.

Another minor effect which cannot be reproduced by the eight band $\boldsymbol{k}\cdot\boldsymbol{p}$ method is the splitting of the conduction and LH bands into non-degenerate for the [110] orientation in both materials (see the center plot in Figs. 5.5 and 5.6). To reproduce this behavior, a ten-band $\boldsymbol{k}\cdot\boldsymbol{p}$ model would be needed, which is able to account for states close in energy to the conduction band [108]. Nevertheless, this band splitting is so weak that the effect on the resulting band structure is not relevant for electrostatic purposes, and it does not justify the extra computational cost.

$\boldsymbol{k}\cdot\boldsymbol{p}$ implementation
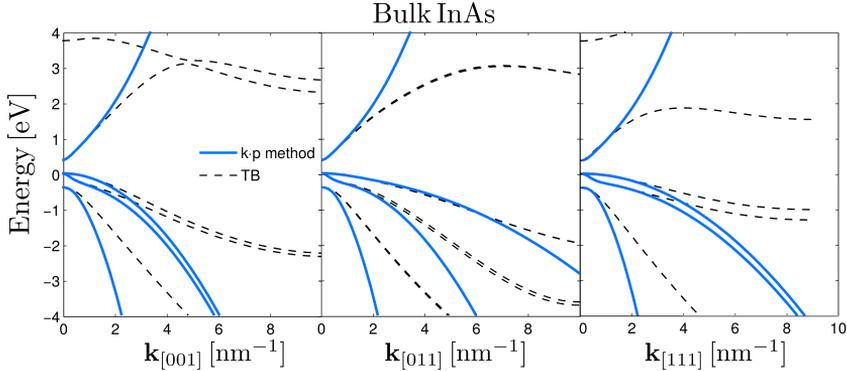
Figure 5.6: Comparison of the bulk InAs band structure for TB and $\boldsymbol{k}\cdot\boldsymbol{p}$ method with respect to the wave vector along different orientations:[001], [011] and [111].

5.3   III-V Nanowires

In this Section we will study the effect of the confinement on the band structure of NWs. As aforementioned, the $\boldsymbol{k}\cdot\boldsymbol{p}$ method intrinsically considers infinite bulk crystal and it is necessary to make use of the envelope function approach to account for the confinement. Therefore, it is important to assess the validity of the $\boldsymbol{k}\cdot\boldsymbol{p}$ method in highly confined systems.

Simulations are performed for three different squared NWs with lateral sizes of $5\,\mathrm{nm}$, $7\,\mathrm{nm}$ and $10\,\mathrm{nm}$ respectively. GaAs and InAs are considered as the channel materials, and different orientations are taken into account. In order to carry out TB simulations, standalone NWs are considered since this method does not account for the insulator atoms, and hydrogen passivation of the channel is considered instead. So that, to compare with the $\boldsymbol{k}\cdot\boldsymbol{p}$ model, an infinite flat potential well is used. Therefore, no self-consistent potential is accounted for.

The $\boldsymbol{k}\cdot\boldsymbol{p}$ parameters used for this simulation are those obtained in the previous Section and presented in Table 5.2. In the comparisons, we will focus on those features that determine the electrostatic and transport behavior of the device, i. e. , the energy gap, the quantization of the levels, the effective masses and the density of states.
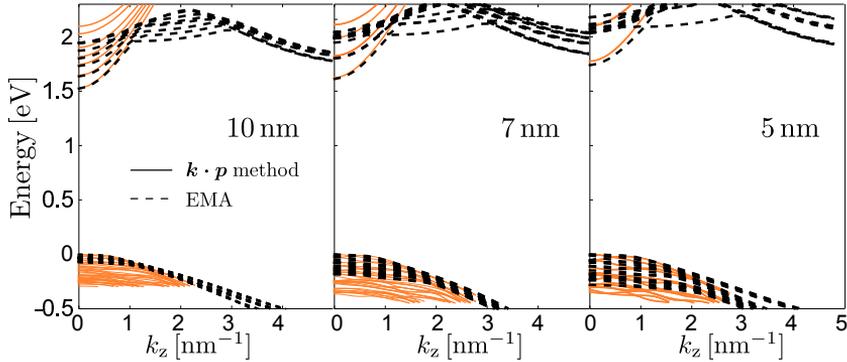
Figure 5.7: Comparison of the band structure obtained from TB and $\boldsymbol{k}\cdot\boldsymbol{p}$ simulations for [001]-oriented GaAs square NWs with lateral sizes of 10nm (left), 7nm (center), and 5 nm (right).

5.3.1 Gallium Arsenide

In this wide gap material, the effect of the coupling between the conduction and the valence bands is small, and therefore, the CB is expected to be more parabolic. However, since the satellite valleys (such as the L and X) are close in energy to the minimum of the CB (located at the Γ valley), their effect can be relevant in the overall electrostatic behavior of the device [84].

Figure 5.7 depicts a global overview of the band structure for [001]-oriented nanowires of different sizes. In general, the $\boldsymbol{k}\cdot\boldsymbol{p}$ method reproduces properly the energy gap, the curvature of the bands and the energy of the subbands for the VB and for the CB at the Γ valley. However, as we already mentioned, the $\boldsymbol{k}\cdot\boldsymbol{p}$ method used here does not take into account satellite L and X valleys. In this orientation, the L valleys, which are the closest in energy, project along the transport direction, and lie closer in energy to the CB edge (Γ valley) than for the bulk case. Taking into account the four-fold degeneration of the L valley in this orientation (eight-fold degeneration accounting for the spin), and its high conduction effective mass, we can conclude that it would contribute strongly to the total population if the Fermi level penetrated deeply into the conduction band. As for the X valley, it is projected to $k_z = 0$, and its energy also gets closer to the band edge as quantization is stronger. As it exhibits a higher density of states, it reduces the relevance of the Γ valley projection in the overall electrostatic behavior. The effect of these valleys in the bans structure makes difficult to represent the CB structure accurately using exclusively the $\boldsymbol{k}\cdot\boldsymbol{p}$ method.

In these cases, it is necessary to consider these valleys somehow. The easiest way would be to use of the EMA to describe these valleys, since they will have a small coupling with the rest of valleys. The non-parabolic EMA has indeed been proven to be a good approximation to model III-V NWs [83]. Anyway, to determine the actual position of the Fermi level into the conduction band, and therefore, the relative contribution of the L valleys, a self consistent simulation is required. If the charge associated to the Γ valley is high enough, the gate potential will be screened and therefore the L valleys will not be very relevant. Therefore, a thorough study must be performed to determine in which situations is necessary to consider these valleys.

A more detailed view of the band structure for the different sizes is shown in Figs. 5.8 and 5.9, for [001] and [011] oriented devices respectively. These figures focus on the region of interest for the Γ valleys, around $k_z = 0$, where both methods can be compared. Fig. 5.8 shows that the $\boldsymbol{k}\cdot\boldsymbol{p}$ method correctly depicts the band gap and curvature for both, CB and VB. The quantization of the states, i. e. , the energetic position of the subbands, is also generally well determined, although some mismatches appear as the device is shrunk. For the VB, a slightly higher separation of the subbands can be observed, whereas the CB calculated with TB exhibits a zone of energies with a denser number of subbands above 2eV. This additional subbands are due to the projection at $k_z = 0$ of the X valley which cannot be captured using the 8 band $\boldsymbol{k}\cdot\boldsymbol{p}$, as aforementioned. Notwithstanding, the lowest conduction bands, which corresponds to the Γ valley, are in good agreement with the results of TB.

As for [011]-oriented NWs (see Fig. 5.9), a good agreement between the results provided by both methods is again achieved, except for the projections of the satellite valleys, as they appear at slightly lower energies. Therefore, their influence on the electrostatic behavior is expected to be more relevant. Regarding the valence band, it is well represented by the $\boldsymbol{k}\cdot\boldsymbol{p}$ calculation although there is a higher separation among the subbands compared with the TB results. A better agreement could be achieved by tuning the $\boldsymbol{k}\cdot\boldsymbol{p}$ parameters as a function of the device size [123].
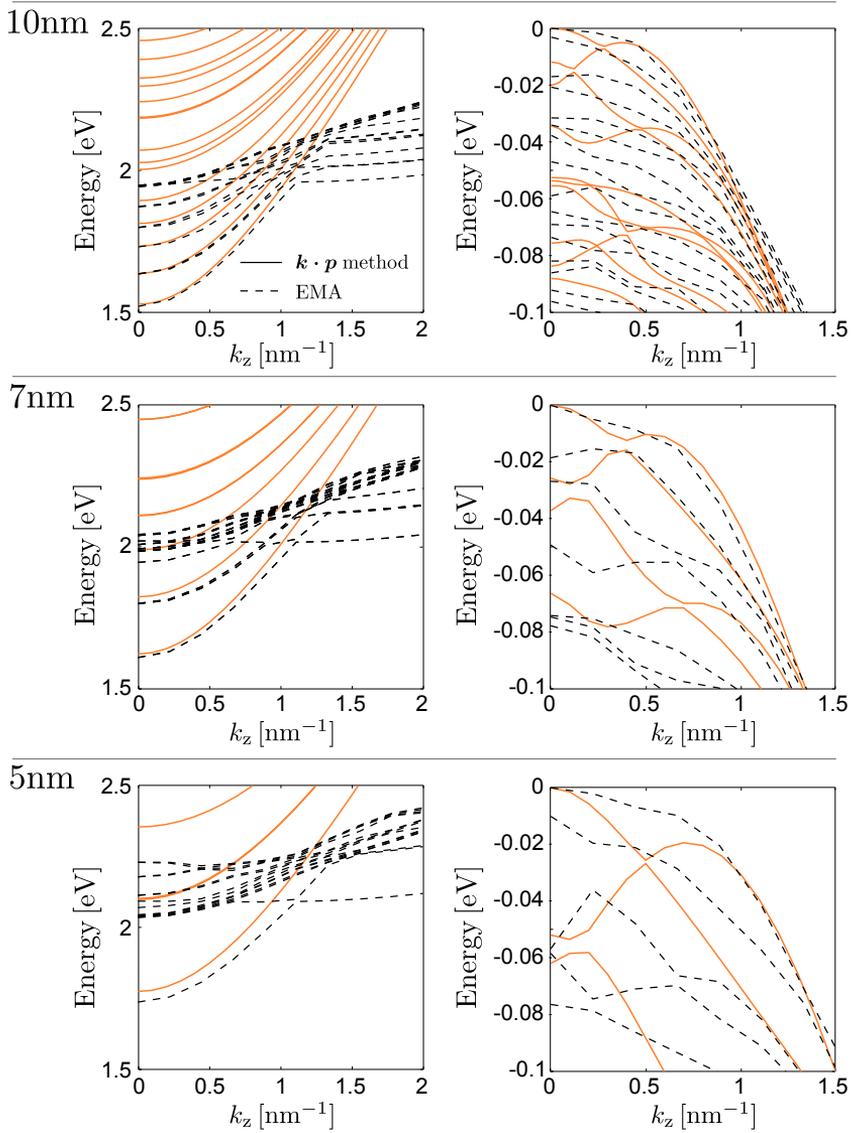
Figure 5.8: Detail of the conduction (left) and the valence (right) bands in [011]-oriented GaAs square NWs with width of 10nm (top), 7nm (center) and 5 nm(bottom).
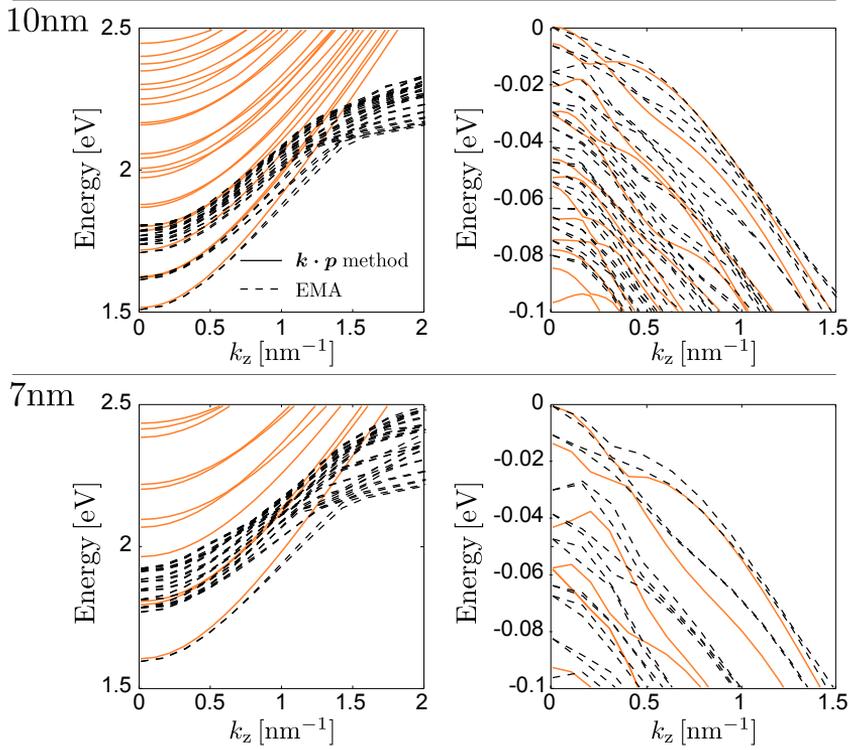
Figure 5.9: Detail of the conduction (left) and the valence (right) bands in [011]-oriented square GaAs NWs with width of 10nm (top) and 7 nm (bottom).
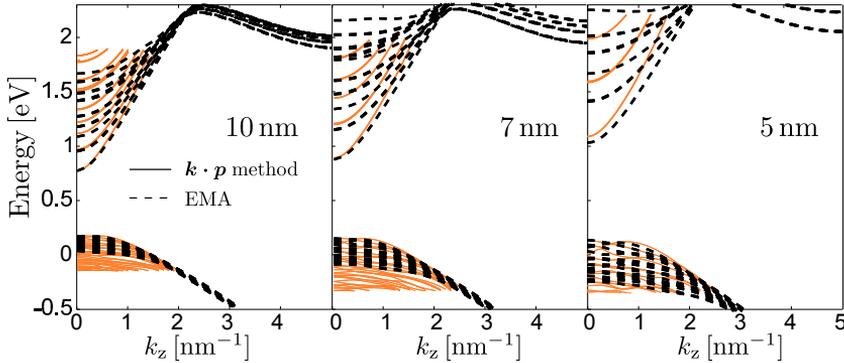
Figure 5.10: Detail of the conduction (left) and the valence (right) bands in [001]-oriented square InAs NWs and width of 10nm (top), 7nm (center), and 5 nm (bottom).

5.3.2 Indium Arsenide

This material has a narrow bandgap, and therefore, the effect of the coupling between conduction and valence bands is noticeable. Besides, satellite valleys lie at a much higher energy, and their effect on the conduction properties is negligible.

Figure 5.10 shows a general overview of the band structure of InAs nanowires of different sizes. A more detailed view of the band structure for the different sizes is shown on Figs. 5.12 and 5.11, for [001] and [011]-oriented devices, respectively. By inspection of these figures, we can assert that the $\boldsymbol{k} \cdot \boldsymbol{p}$ method provides fine results for the valence band along the different orientations considered. As expected, the projection of the Γ valley (in $k_z = 0$) is the main conduction band, with the L valley falling more than 1eV above the CB edge. The CB exhibits high non parabolicity because of the strong coupling with the valence band. However, the non parabolic behavior is not fully captured by the $\boldsymbol{k} \cdot \boldsymbol{p}$ method. For the larger device, the representation of the CB seems to be quite accurate. However, for the smaller sizes, the non parabolicity effect is more evident, affecting the separation between the subbands. The reason of this mismatch on the parabolicity is the artificial reduction of the coupling parameter E_P, needed to cancel the spurious solutions, as discussed in Sec. 3.7. A smaller value of that parameter reduces the coupling between bands, thereby resulting in a more parabolic representation of the CB. Despite these differences, the effective masses near the band edges are still correct.

In general terms, it is safe to conclude that the $\boldsymbol{k} \cdot \boldsymbol{p}$ method provides a correct

Figure 5.11: Detail of the conduction (left) and the valence (right) bands in [011]-oriented square InAs NWs with width of 7nm (top), and 5 nm (bottom).

representation of the energy gap and the curvature of the bands. There are some improvements which may be achieved in the calculation of the separation between levels and the non parabolic behavior for high k_z values. A careful selection of the $\boldsymbol{k \cdot p}$ parameters could lead to those improvements.
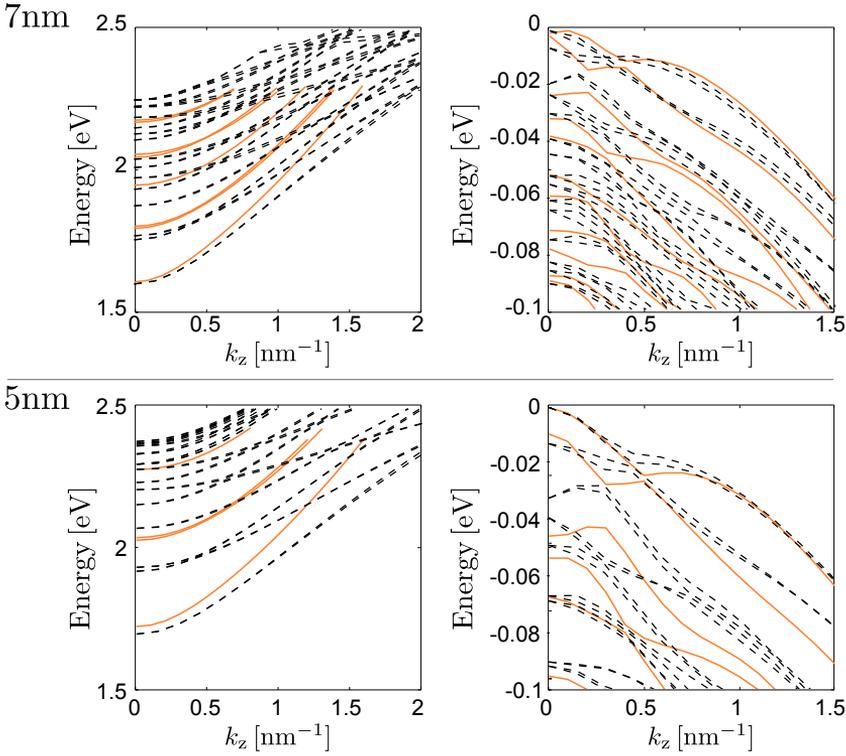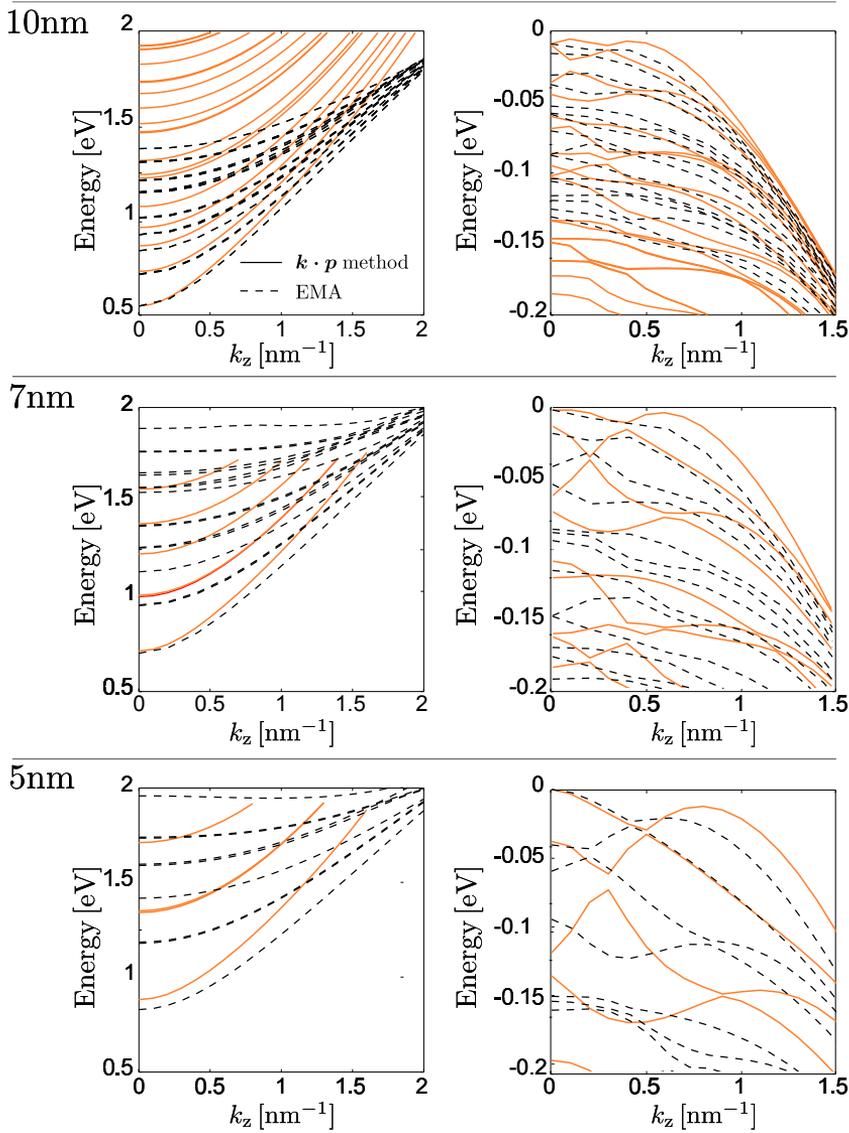
Figure 5.12: Detail of the conduction (left) and the valence (right) bands in [001]-oriented square InAs NWs with width of 10nm (top), 7nm (center), and 5 nm (bottom).

5.3.3 Linear Carrier Density in III-V NWs

It is difficult to evaluate the performance of the $\boldsymbol{k}\cdot\boldsymbol{p}$ method only by comparing the calculated band structure with that obtained by a different method like TB. A more practical approach is the comparison between the carrier densities yielded by each one. In this Subsection, we calculate the carrier density per unit length confined in the cross-section of the NW. To do that, we perform the integration over every state weighted by its probability of occupation, which depends on the relative position of the Fermi level.

Figure 5.13 illustrates the linear carrier density calculated with both methods in a GaAs and InAs square NW, [001] oriented, for the three consid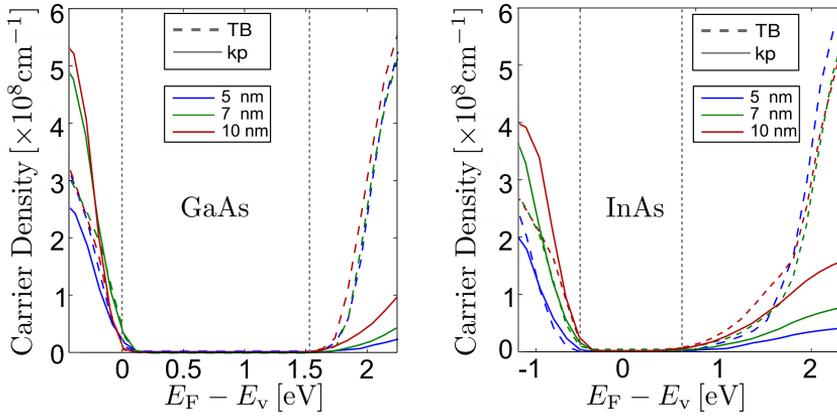ered sizes: 5nm (blue), 7nm (green), and 10 nm (red). The reference is the valence band edge, namely, the zero energy corresponds to the top of the VB. In general, we can observe that the representation of the carrier densities is very similar for both methods for low penetration of the Fermi level into the valence band. Thus, as we demonstrated in the previous Subsections, the $\boldsymbol{k}\cdot\boldsymbol{p}$ method yields fine results. Nevertheless, a better matching could be achieved by tuning the parameters. For lower positions of the Fermi level, we observe a decrease in the slope for the TB. This issue results from the fact that the data provided from the TB simulations only consisted of a limited number of subbands, those around the band edges. Therefore, the carrier density is poorly estimated for very high inversion.

For the conduction bands, higher differences can be appreciated. The electron density for the GaAs NW is much lower in $\boldsymbol{k}\cdot\boldsymbol{p}$ simulation than in TB one, even when a good match in the conduction band is achieved. The reason is that for small energies (0. 1 or 0. 2 eV above the CB), the L and X valleys projections start to be strongly populated because of their high density of states. The energy at which satellite valleys get populated is remarkably evident in the case of InAs devices (right figures in Fig. 5.13), where a change in the slope is observed. The effect of these satellite valleys masks the good matching existing between the population of the Γ valley achieved with both approaches. To prove this fact, a more detailed representation of the population of each valley is shown in Fig. 5.14 for a 10 nm wide [001] oriented NW. We observe a good match of the electron density in the Γ valley calculated using the TB and the $\boldsymbol{k}\cdot\boldsymbol{p}$ methods. For the InAs, the results seem to be better because of the higher energetic distance between the Γ and the L valley. In this case, the electron density calculated using the TB method shows a steep slope when the L valley starts to be populated.

Figure 5.13: Carrier density per unit length in GaAs (left) and InAs (right) [001]-oriented square NWs, for three sizes: 5nm (blue), 7nm (green), and 10 nm (red). Dashed lines correspond to TB and solid lines to $\boldsymbol{k}\cdot\boldsymbol{p}$. The carried density is calculated by varying the position of the Fermi level with respect to the VB edge. Dotted vertical lines remark the positions of the CB and VB edges.
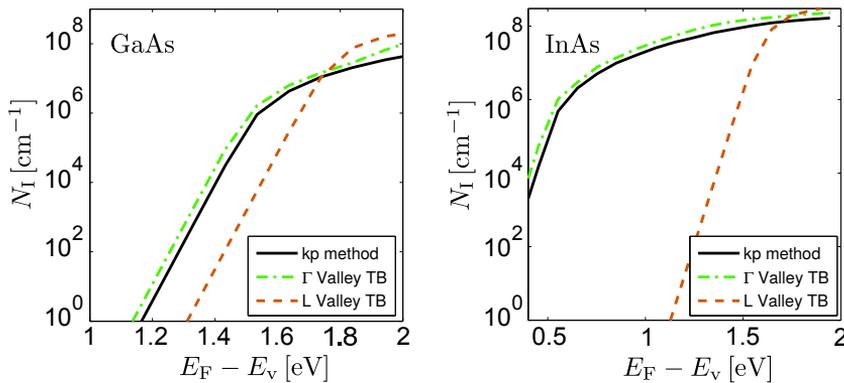
Figure 5.14: Linear electron density corresponding to the Γ and the L valleys in GaAs (left) and InAs (right) NWs (10x10nm, [001] orientation). Solid line corresponds to the $\boldsymbol{k}\cdot\boldsymbol{p}$ simulation and dashed line to the TB one.

5.3.4   Summary

In this chapter we have calculated and compared the electrostatic characteristics of different NWs making use of the TB and $\boldsymbol{k}\cdot\boldsymbol{p}$ methods. The results show a good agreement for different magnitudes such as the energy gap, curvature and separation of the Γ valley subbands, which is very important in III-V materials. Nevertheless, for very high electron densities, the L valley should also be considered [83, 85–87]. This valley is naturally included by the TB calculation. Unfortunately this not the case when the 8 bands $\boldsymbol{k}\cdot\boldsymbol{p}$ approach is employed, and alternative methods should be used to account for it.

As for the valence band, it can be concluded that results yielded by both methods are in good agreement, although slight differences are shown. This fact was expected, as there is a large number of parameters involved which are semiempirically estimated. Besides, the TB approach has its own limitations since it is applied in an standalone wire with surface states passivated with hydrogen. This means that an infinite potential barrier is assumed, precluding the wave function penetration inside the insulator gate. As a consequence, differences are justified and can be reduced by an appropriate tunning of the parameters.

To sum up, we can conclude that our simulation tool has been tested and it provides good enough results to characterize the electrostatic behavior of NWs.

Part III

Results

Chapter 6

$k \cdot p$ simulation of NW electrostatics

This chapter is devoted to the study of electrostatic properties of NWs with different shapes, materials, and geometries. We will focus on the effects of the 2D confinement on the carriers, which leads to: (a) a shift of the charge centroid, pushed away from the semiconductor-insulator interface [115] and (b) a reduction of the density of states due to the separation of the energy subbands.

We will analyse the results provided by the SP2D simulation tool described in previous chapters. Different magnitudes are employed to assess the electrostatic performance of the device, such as the gate capacitance, effective mass, charge distribution, etc.

In this chapter we will analyse the results provided by the SP2D simulation tool described in the previous chapters. The direct outcomes of the $k \cdot p$ simulator are: 1) the band structure of the device in a discrete set of points of the wave vector space k_z, and 2), their associated eigenfunctions $\boldsymbol{\xi}_{nk_z}(x, y)$ in the real space discretized by means of the FEM method. However, to efficiently assess the performance of different devices and compare them, several parameters must be estimated: gate capacitance, charge centroid, effective mass, and group velocity. Along this chapter, these parameters and their estimation will be briefly introduced.

The Chapter is organized as follows. First, in Sec. 6.1, we analyze the band structure of NWs with different size and materials, introducing some important parameters such as the group velocity and the effective mass, and dealing with the subbands crossing issue. Then, in Sec. 6.2 the behavior of the charge density is analyzed, and the

evolution of the charge centroid as a function of the device bias is studied. Next, the electrostatic behavior of SiGe NWs is studied in Sec. 6.3, in terms of the inversion charge and the gate capacitance. Finally, the main conclusions are drawn.

6.1 Band structure analysis

In this Section, we deal with the band structure of different semiconductor NWs. One relevant consideration regarding the band structure calculation is the complexity of the problem in terms of the large amount of data to be managed. So as to achieve an efficient solution of the problem, a tradeoff between accuracy, computational cost, and memory expense must be made. Sometimes, some assumptions are made in order to simplify such a complex problem. We have shown that the band structure in NWs is composed by different subbands which evolve as a function of k_z. Under some approaches, such as the EMA, they are characterized as a whole by some parameters, e.g., the wave function in the edge, the effective mass, or a linear group velocity depending on k_z. Then, it is quite common to use the same definitions and approaches to refer to the subbands in more complex methods, such as the $\boldsymbol{k \cdot p}$ method. Unfortunately, these definitions must be taken with a grain of salt since these interpretations may yield to confusion and wrong results.

On the one hand, since the band structure resulting from our simulations is discretized in k_z, it is not easy to determine which values correspond to each subband, since either a crossing or an anti-crossing can occur between these subbands [39, 92] as we will treat in the following Subsection. On the other hand, a subband no longer evolves independently to the rest. Each subband is formed by the contribution of different basis states at each k_z point and it varies continuously. Consequently, the wave function can drastically change its shape for different k_z values.

6.1.1 Subband crossing and anti-crossing

In Fig. 6.1 the band structure for two different NWs is shown (cylindrical GaAs NW with $2R_s = 5\text{nm}$ for [001] (a) and [011] (b) orientations respectively). The approaching bands exhibit two possibilities: a) a crossing between them, and b) the so-called anti-crossing. For the representation of that figure, the subbands are ordered by their energy: first subband corresponds to the higher energy for each k_z and so on. As can be seen
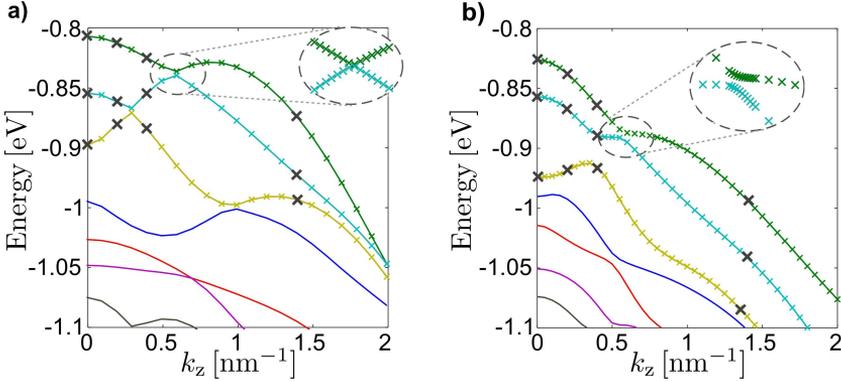
Figure 6.1: Representation of the crossing (a) and anti-crossing (b) of subbands in two GaAs NWs with $2R_s = 5$nm for [001] (a) and [011] (b) orientations, respectively. The crosses mark the discrete point were the energies are calculated in the first subbands. The area of interest is zoomed to show the actual behavior of the subbands. The dark larger crosses remark the states corresponding to the wave functions represented in Fig. 6.2.

in Fig. 6.1(a), with this ordering the subbands change their slope abruptly, which can be viewed as an abrupt change of the carrier velocity for a low variation in k_z.

This issue is usually eluded in the literature when the **k·p** method is implemented. However, in some cases it may be of major relevance if some cautions are not taken. Firstly, the wrong ordering of the subbands induces additional numerical error in the charge calculation, since the evolution of the subband in the crossing is misrepresented. This error can be kept low with a fine discretization of k_z. Furthermore, an evident consequence is the difficulty to determine the magnitudes involving the derivation of the energy with respect to k_z, which would not be well defined in the crossing points. This problem is aggravated by the numerical calculation of the derivatives as we will see in the next Subsection.

Notwithstanding, the most serious concern arises when doing some approximations without taking the necessary cautions. Due to the large amount of data required to define the wave functions in all the calculated states, it is not unusual to find some approximations involving the reduction of the total number of wave functions used to determine either the charge or other magnitudes depending on them (see Ref. [40]). One of these approximations is to use the same wave function of the edge of the subband for the whole subband, thus neglecting the dependence of the wave functions on k_z. In this case, the charge distribution can be severely affected depending on the chosen
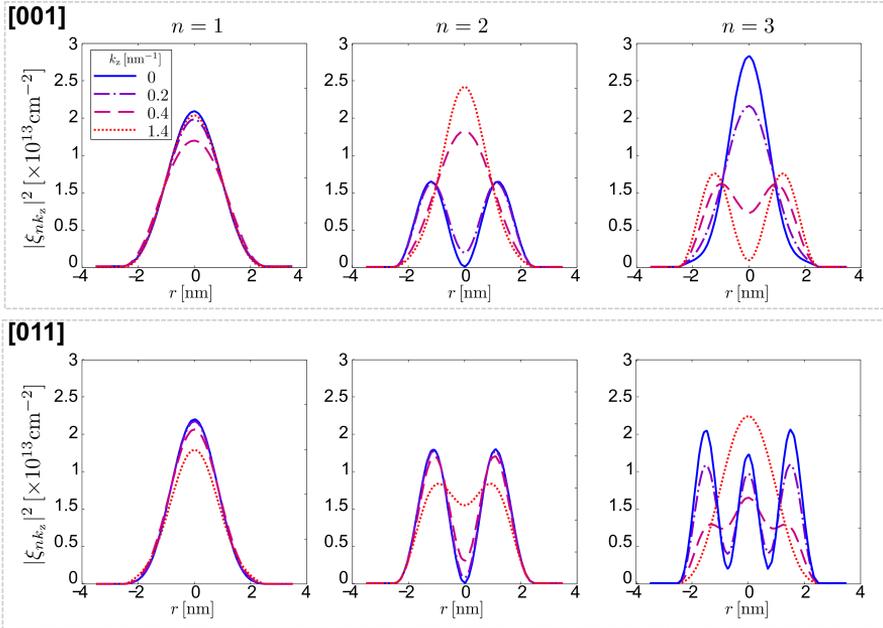
Figure 6.2: Probability density distribution $|\boldsymbol{\xi}_{nk_z}|^2$, for a state $|n, k_z\rangle$, along the section of the NW corresponding to the states remarked in Fig. 6.1. These states are those corresponding to the three first subbands (not considering two-fold degeneracy) and $k_z = (0, 0.2, 0.4, 1.4)\mathrm{nm}^{-1}$. The upper row corresponds to the [001]-oriented GaAs NW (crossing case), whereas the lower row corresponds to the [011]-oriented GaAs NW (anti-crossing case).

ordering.

Let us now take an insight on this approximation. Fig. 6.2 shows a slice of $|\boldsymbol{\xi}_{nk_z}|^2$, for the first three subbands, along the x direction ($y = 0$) and at the k_z values remarked in Fig. 6.1. Since all of them are two-fold spin degenerate, the contribution of both states is summed to get the actual charge distribution associated to each state. As illustrated in Fig. 6.2(a), it is clear that the second and the third subbands exchange their wave function after an undetected crossing. This issue may be solved by elaborating a method to detect the subband crossing. However, as shown in Fig. 6.2(b), even when the subbands present anti-crossing, the $|\boldsymbol{\xi}_{nk_z}|^2$ distribution still varies appreciably with k_z even for the same subband.

We can assert that the approximation of a single wave function for the whole subband is highly inaccurate in the definition of the probability distribution of the states, and therefore, the charge distribution provided is not correct, as Fig. 6.3 illustrates.
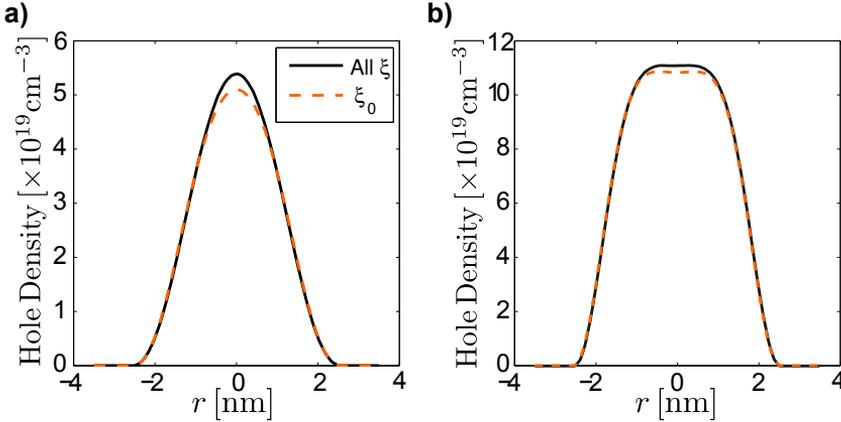
Figure 6.3: Hole density along the section of a NW (the same as in Fig. 6.1(a)). Solid line represent the calculation of the charge using the complete set of wave functions, whereas dashed lines represent the calculation considering only the wave function at $k_z = 0$ for the whole subband. For (a) the charge has been computed directly from the band structure, and in (b) a self-consistent iteration has been employed.

Here, the charge has been calculated in the high inversion regime using the band structure represented in Fig. 6.1(a), and setting the Fermi level at 0eV, for both approaches: solid lines for the complete set of wave functions, dashed lines for the wave function calculated at $k_z = 0$. As can be seen, a relevant error ($\sim 1.5\%$) in the charge distribution can be appreciated for both cases.

6.1.2 Carrier velocity and effective mass

The carrier velocity and the effective mass of a subband are calculated through the first and second derivatives of its energy with respect to k_z, respectively:

$$v_n(k_z) = \frac{1}{\hbar} \frac{\partial E_n}{\partial k_z}, \tag{6.1}$$

$$\frac{1}{m_{nk_z}} = \frac{1}{\hbar^2} \frac{\partial^2 E_n}{\partial k_z^2}. \tag{6.2}$$

When a parabolic energy band is assumed, the effective mass has a well defined value for each subband. However, for non parabolic bands as those shown in Fig. 6.1, the effective mass can be different for each state $|n, k_z\rangle$. It would be quite useful to be

able to estimate a global effective mass that provides information regarding the overall behavior of the carriers in the device. Thus, to take into account the contribution of all the occupied states to the mean effective mass, we define a mean effective mass for the transport $\langle m_{\text{eff}} \rangle$ [33], where each state is weighted by its occupation probability function:

$$\frac{1}{\langle m_{\text{eff}} \rangle} = \left\langle \frac{1}{m_{nk_z}} \right\rangle = \sum_n \int \frac{1}{m_{nk_z}} f_0(E_n(k_z)) \, \mathrm{d}k_z \Bigg/ \sum_n \int f_0(E_n(k_z)) \, \mathrm{d}k_z \qquad (6.3)$$

In theory, the velocities and effective masses could be calculated by numerical differentiation. However, this procedure is not advisable due to the difficulty of tracking the evolution of a subband in a discrete k_z space. Thus, a fine grid for k_z would be required to obtain accurate results, which involves the resolution of the $\boldsymbol{k} \cdot \boldsymbol{p}$ algorithm in a high number of points for each iteration. Furthermore, a subband ordering accounting for the crossing of the subbands must be defined. Nevertheless, peaks might appear when performing the numerical differentiation, where the subband crossing calculation has failed and an abrupt change of the trend of the band is located (see upper plots in Fig. 6.4). The calculation of the effective mass would be even more troublesome since a second derivative is required.

As we will see in Chapter 7, it is of major importance to calculate a reliable value of the carrier velocity to reduce the numerical error in the mobility calculation. For this reasons, we have chosen the procedure proposed by *Stanojevic et al.* [126] to calculate the first derivative. This approach stems directly from the Hamiltonian and uses only the local information of the point where the derivative is calculated, avoiding information coming from adjacent points, yielding a reduction of the number of points in k_z. This method is based in the non-degenerate perturbation theory described in Appendix A, which states that in the presence of a small perturbation $\delta \hat{\boldsymbol{H}}(k_z)$ the energy of a state will change according to

$$\delta E_n(k_z) \approx \left\langle \boldsymbol{\xi}_{nk_z} \left| \delta \hat{\boldsymbol{H}}(k_z) \right| \boldsymbol{\xi}_{nk_z} \right\rangle + \sum_{j \neq i} \frac{\left| \left\langle \boldsymbol{\xi}_{nk_z} \left| \delta \hat{\boldsymbol{H}}(k_z) \right| \boldsymbol{\xi}_{n'k_z} \right\rangle \right|^2}{E_n(k_z) - E_{n'}(k_z)} . \qquad (6.4)$$

We assume $\delta \hat{\boldsymbol{H}}(k_z)$ as the difference in the Hamiltonian between points k_z and
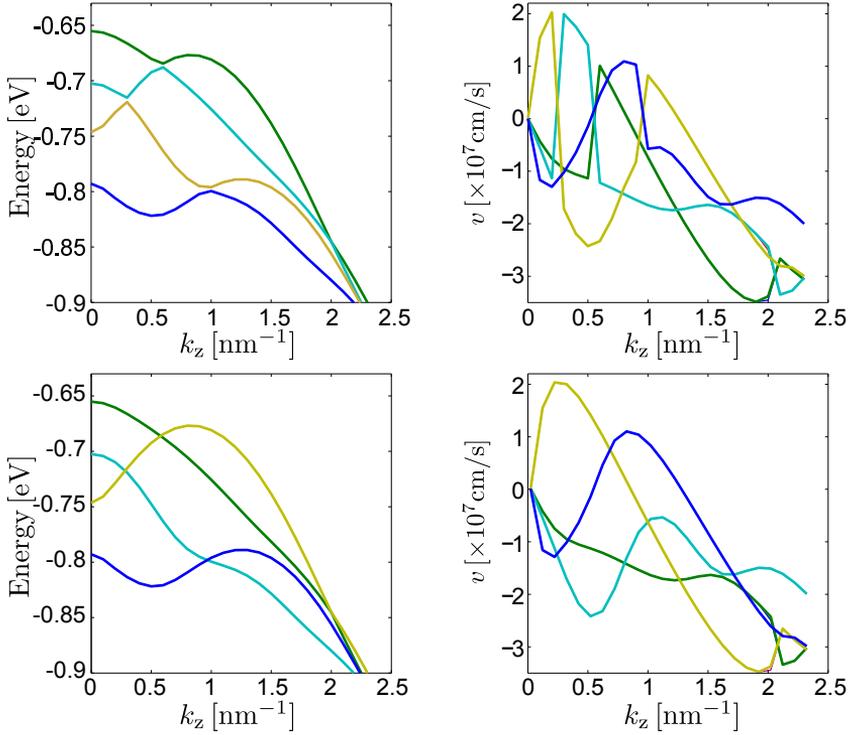
Figure 6.4: Comparison of the four first subbands (left plots) and their corresponding carrier velocity (right plots) for the disordered (upper plots) and ordered (lower plots) cases.

$k_z + \delta k_z$ and approximate it using the derivative of the Hamiltonian to get

$$\delta E_n(k_z) \approx \left\langle \boldsymbol{\xi}_{nk_z} \left| \frac{\partial \hat{\boldsymbol{H}}}{\partial k_z} \right| \boldsymbol{\xi}_{nk_z} \right\rangle \delta k_z \, , \qquad (6.5)$$

and obtain the derivative of the energy, which is related to the carrier velocity of the n-th subband as

$$\frac{\partial E_n}{\partial k_z} = \hbar v_n(k_z) \approx \left\langle \boldsymbol{\xi}_{nk_z} \left| \frac{\partial \hat{\boldsymbol{H}}}{\partial k_z} \right| \boldsymbol{\xi}_{nk_z} \right\rangle \, . \qquad (6.6)$$

The calculation of the group velocity for a state $|n, k_z\rangle$ involves the inner product of the derivative of the Hamiltonian. This derivative is easily calculated since the dependence of the Hamiltonian with respect to k_z is a polynomial of second order. As

it was explained in Section 4.5, the $\boldsymbol{k}\cdot\boldsymbol{p}$ matrix can be constructed as:

$$\boldsymbol{H}(k_{\mathrm{z}}) = \boldsymbol{H}_0 + \boldsymbol{H}_1 k_{\mathrm{z}} + \boldsymbol{H}_2 k_{\mathrm{z}}^2 \,. \tag{6.7}$$

Thus, its derivative is given by:

$$\frac{\partial \boldsymbol{H}}{\partial k_{\mathrm{z}}} = \boldsymbol{H}_1 + 2\boldsymbol{H}_2 k_{\mathrm{z}} \,. \tag{6.8}$$

So that, the derivative of the energy can be approximated numerically by a simple matrix product

$$\frac{\partial E_n}{\partial k_{\mathrm{z}}} = \boldsymbol{\xi}_{nk_{\mathrm{z}}}^{\dagger} \left(\frac{\partial \boldsymbol{H}}{\partial k_{\mathrm{z}}} \right) \boldsymbol{\xi}_{nk_{\mathrm{z}}} \,, \tag{6.9}$$

where now $\boldsymbol{\xi}_{nk_{\mathrm{z}}}$ stands for the discrete form of the envelope vector function (Eq. (4.10)).

This procedure was described for non degenerate states. However, the resulting band structure is two-fold degenerate due to the presence of the spin. Notwithstanding, this theory is still applicable since the spin degeneracy is not lifted along k_{z}, and therefore $\left\langle \boldsymbol{\xi}_{nk_{\mathrm{z}}} \left| \delta \hat{\boldsymbol{H}}(k_{\mathrm{z}}) \right| \boldsymbol{\xi}_{n'k_{\mathrm{z}}} \right\rangle$ vanishes for two degenerate subbands n and n'. In the case that the degeneration of the subbands is local, as in the crossings, the degenerate perturbation theory should be used.

From Eq. (6.4) we can also obtain the second derivative of the energy by including the second order term. This is not practical though, since the calculation of the second order term would requires to calculate all the interactions between all the wave functions, or a sufficiently large number of them to keep a low error. On the contrary, the first derivative only requires the diagonal matrix elements, i.e., the interaction of a wave function with itself. Therefore, so as to calculate the second derivative, it is usually a satisfactory approach to use the numerical derivative of the first derivative in Eq.(6.6), in regions where there are not subband crossing. However, we are interested in the mean effective mass, and therefore, the local values of m_{eff} are weighted by their corresponding population. Consequently, certain amount of error is assumed in this magnitude, which is not of major relevance since this magnitude is used as an estimation. We have implemented an ordering algorithm for the subbands, which detects as many band crossing as possible, being a good way to reduce this error as much as possible.

The ordering algorithm is based in the fact that the energy of a subband and its derivative (the velocity) are already known. Figure 6.4 shows the disordered band

structure and their corresponding velocity, which abruptly changes where a crossing is present. It is evident that an abrupt slope change in a subband produces an unphysical step in v_n. Then, the criterion to order subbands must be to achieve a smooth evolution of the velocity of every subband. The detection of the steps in v_n leads to identify the crossings in the band structure. The lower plots in Fig. 6.4 depict the band structure and the group velocities after applying this algorithm. It is shown that it correctly captures the main crossing points and the subbands no longer exhibit abrupt changes. A numerical derivative can be performed to obtain a reasonable second derivative, used to calculate the mean effective mass.

6.1.3 Valence Band structure of Si and Ge NWs

The bandstructure of bulk diamond and zinc blende semiconductors was described in Chapter 3. The most relevant bands in the VB were denominated LH, HH, and SO. These bands evolve independently and anisotropically in the wave vector space $\boldsymbol{k}$. However, as we saw, when confinement is added, the envelope function approximation in Sec. 3.6 must be applied and these energy bands are quantized in subbands, which no longer correspond to the bulk LH, HH, and SO, but to a mixture of them. Here, we will show how the quantization affects the valence band structure in cylindrical NWs made of Si and Ge, for the usual transport orientations ([001], [011], and [111]), and different diameters from 3nm to 20nm. The insulator thickness is kept constant to $T_{\text{ins}} = 1\text{nm}$ for all the devices and the metal has been chosen to be midgap for Si (the same metal has been used for Ge). The results of this section were calculated under the flat potential well assumption.

Figure 6.5 illustrates the band structure for [111] oriented Si NWs (left column) and Ge NWs (right column) for two different diameters, 3nm (upper row) and 10nm (lower row). The energy reference in these figures is the maximum of the VB. We can corroborate the lighter overall effective mass for Ge than for Si, as in the bulk scenario. The consequence is that Ge is also more affected by quantization of the levels, i.e, the shifting is larger and the density of subbands lower. As a result, the total density of states is also reduced.

As can be seen, in all cases the subbands do not show a parabolic behavior as they evolve in k_z, and so, an effective mass can not be formally attributable to them. However, in a general overview, we can identify two marked trends of the overall ensemble of subbands: one corresponding to a large curvature (lighter effective mass) and other
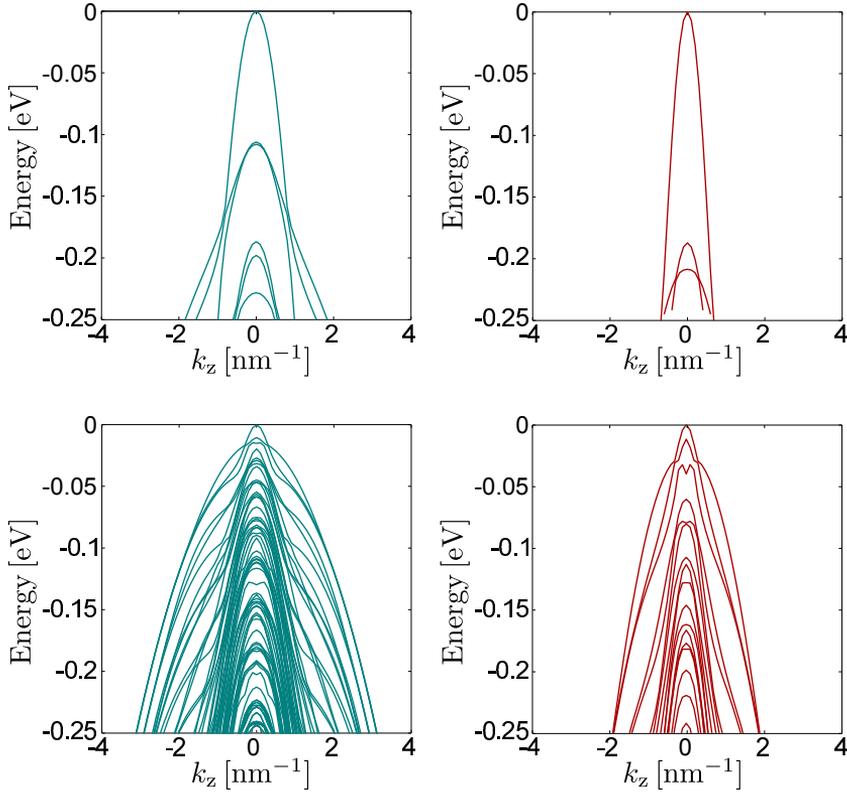
Figure 6.5: Subband structure for cylindrical Si (left column) and Ge (right column) [111]-oriented NWs. Two diameters $2R_s$ are represented: 3nm (upper row) and 10nm (lower row).

corresponding to a smaller curvature (heavier effective mass). Then, we can keep in mind the idea of the LH and HH, which are degenerated in bulk semiconductors, splitting differently for different confinements. Indeed, there are studies which use the $\boldsymbol{k}\cdot\boldsymbol{p}$ results for adjusting an effective mass approach, as in [123].

The effect of confinement in the bandstructure is very remarkable, and so that, we detail it in Figure 6.6, where the evolution of the first two subbands is represented for different diameters in Ge NWs oriented along the [001] and [111] directions.

As the NW diameter is reduced, the lighter subbands are predominant in the energy region close to the VB edge. These subbands are the most relevant in the electrostatic behavior of the smaller device and also in the transport, and so that, the performance in smaller devices is driven by these light holes subbands. Lighter holes have two implications, one negative and other positive for electronic purposes: 1) a reduced density of states which degrades the gate voltage control over the channel charge, and
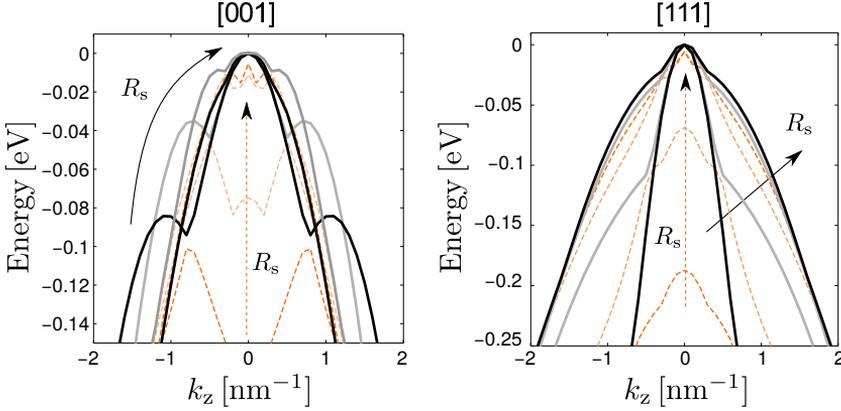
Figure 6.6: The subband structure of the first two subbands (solid line for the first and dashed line for the second) are depicted for Ge NWs oriented along the [001] (left figure) and [111] (right figure) directions for different diameters: $2R_s = 3, 5, 10, 20$nm. The plots for the extremal diameters are highlighted in darker tonalities. The arrows aim to larger devices.

2) a higher injection velocity and a higher mobility which are desirable for transport. Regarding both issues, and although a comprehensive study must be performed, Ge NWs are expected to exhibit better transport performance than Si, but driving less charge.

Another manifestation of the subband quantization is the shifting of the ground state, which drops for smaller sizes, as can be seen in Fig. 6.7. This dropping of the VB edge means that the threshold voltage will be higher (in absolute value) and a larger negative gate voltage is thus required to yield a certain inversion charge. As can be seen in Fig. 6.7, large devices present a nearly constant E_1 value which drops for sizes smaller than 15nm in Ge and 10nm in Si. The variation in Ge is specially relevant, since devices with a diameter of 3nm will undergo a decay of ~ 0.35eV with respect to the valence band edge of large devices. This effect is consistent and attributable to a smaller effective mass in Ge which magnifies the quantum effects.

To finalize this brief analysis of the effect of geometric confinement on the band structure, we will study the mean effective mass for these devices, as defined in Eq. (6.3), which provides an overall insight on the weight of each subband.
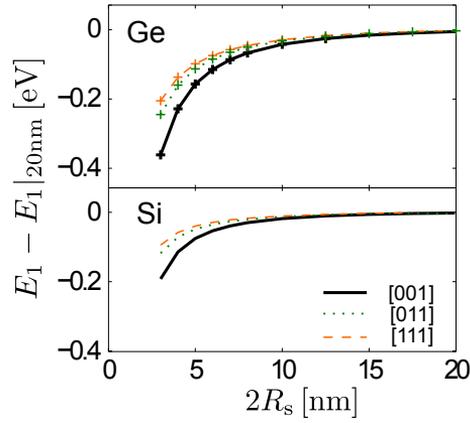
Figure 6.7: Representation of the ground state (VB), E_1, referenced to the E_1 value achieved for the largest device considered, as a function of the diameter of the channel. Si and Ge NWs (crosses) are oriented along the [001] (solid line), [011] (dotted line) and [111] (dashed lines).
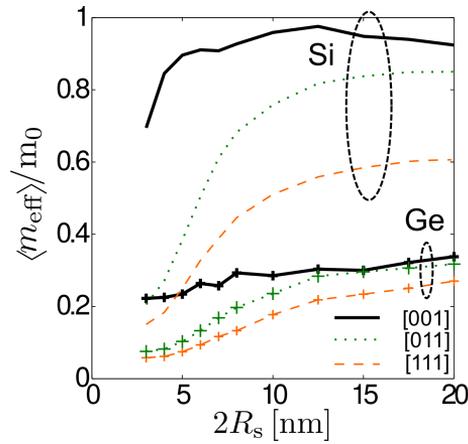
Figure 6.8: Mean effective mass of the valence band for Si and Ge NWs (marked with crosses) oriented in the directions [001] (solid line), [011] (dotted line) and [111] (dashed lines) as a function of $2R_s$.

Figure 6.8 illustrates the mean effective mass for Si and Ge NWs for the different device orientations as a function of the device size. As can be seen, the mean effective mass is smaller in all the cases as the devices are shrunk. This is a consequence of the aforementioned splitting between the lighter and heavier subbands. In this case, Si NWs experiment a higher drop in the effective mass as they are shrunk, and for the smallest sizes its value gets closer to that calculated for Ge NWs. We can also corroborate that the devices oriented in the [111] direction, exhibit the lowest effective mass and therefore they concentrate the potential to improve the transport properties [98]. These results are in agreement with those presented in the literature [52, 98, 103, 123].

6.1.4 Self Consistency

Previously we have studied the effect of confinement in the band structure under the flat band assumption. However, we are also interested in the effect of the applied gate voltage V_G on the band structure and therefore on the electrostatic behavior of the device. To do that, the self-consistent iteration procedure described in Sec. 4.6 is used to determine the potential and charge in the device, for an arbitrary V_G value. In that Section, the necessity of adopting such procedure to assure accurate results for the charge and the potential distributions was justified. The use of self-consistency for the resolution of the $\boldsymbol{k} \cdot \boldsymbol{p}$ method has been recurrent in the last decades [8, 31, 98, 104, 151]. In this Section, we assess how the self-consistent solution modifies the band structure and compare its results with those achieved with a non self-consistent approximation.

First we observe the effect of the self-consistent solution on the band structure. In Fig. 6.9, the highest energy subband for a [001]-oriented GaAs NW with $2R_s = 5$nm [001] oriented NW is shown for different gate voltages, from -0.8V to -1.6V. These voltages are chosen so as to depict the region where the device is driven to high inversion. As V_G increases, the inversion charge becomes more relevant and sets up a position dependent potential in the semiconductor channel, which simultaneously affect to the solutions of the $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian. It can be seen the noticeable change of the first subband as V_G grows, modifying its curvature. As a consequence, an appreciable increase of the mean effective mass is observed (Fig. 6.9.b) as the device is driven deeply to inversion. This effect is not captured by the flat band approximation.
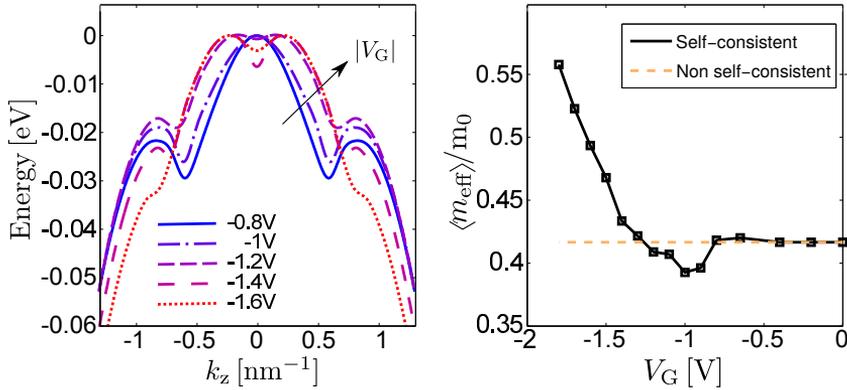
Figure 6.9: First subbands of a $R_s = 5$nm [001]-oriented GaAs NW for different values of $V_G = -0.8, -1, -1.2, -1.4, -1.6$V. The arrow aims to higher $|V_G|$. The subbands are referenced to the VB edge at $k_z = 0$.
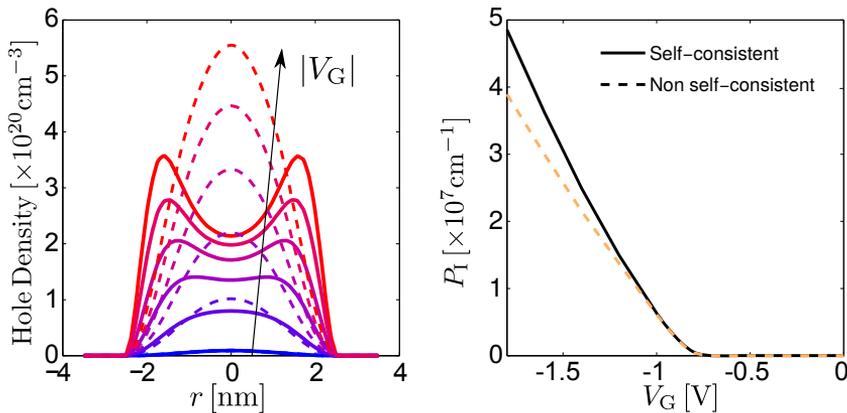
Figure 6.10: Comparison of the hole density calculated in GaAs $2R_s = 5$nm [001]-oriented NWs using the self-consistent (solid lines) algorithm and the non self-consistent (dashed lines) solution. On the left, hole density along the section of the device is represented for different values of $V_G = -0.8, -1, -1.2, -1.4, -1.6$V (from blue to red tonalities respectively) with the arrow aiming to higher values of $|V_G|$. On the right, the linear hole density as a function of V_G is depicted.

Figure 6.10 illustrates the comparison of the charge density for different values of V_G. On the left, the surface hole density is shown for both approaches. In the self-consistent case (solid lines), the charge gets closer to the interface as V_G increases. The flat well approximation is not able to reproduce this effect, and the charge is positioned around the center for all the V_G values. This is a manifestation of the fact that not only the shape of the band structure is modified by the potential but also the wave functions associated to each state. Fig. 6.10 (right) also depicts the total integrated charge for this device, showing that not only the charge position but also the total integrated charge is affected by the self-consistent solution, which therefore becomes mandatory to find accurate results.

6.2   Hole density and centroid calculation

In this Section, we will have a deeper insight in the charge distribution along the NW, as a function of the applied gate bias and the device orientation.

Figure 6.11 illustrates the hole density in the cross section of cylindrical Ge NWs, with 4nm of diameter, and for different orientations. A gate overdrive voltage of $V_G - V_T = -1V$ has been considered. The threshold voltage was obtained from the maximum of the second derivative of the charge with respect to the gate voltage [85]. As it can be observed, the charge is mainly placed close to the interface due to the high applied V_G. A very remarkable issue is the anisotropy of the charge distribution in a highly symmetric structure as a cylindrical NW, specially for the [011] orientation, which stems from the anisotropy of the valence band.
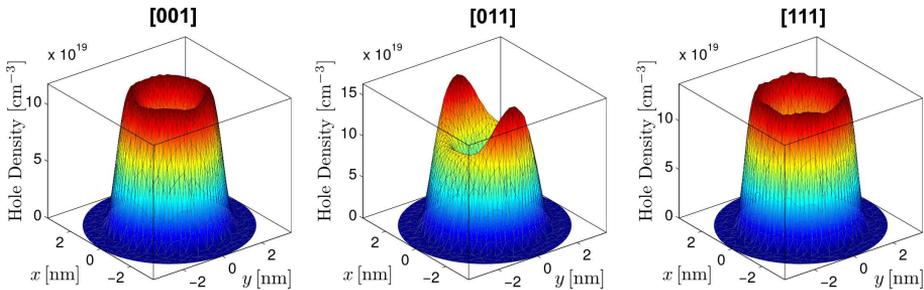
Figure 6.11:   Hole density calculated for cylindrical Ge NWs with 4nm of diameter and different orientations: [001] left, [011] center, and [111] right. A gate overdrive voltage $V_G - V_T = -1$ has been considered.
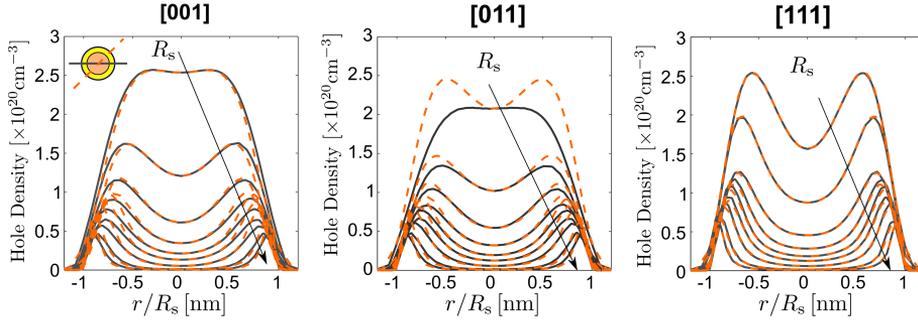
Figure 6.12: Section of the hole density calculated for cylindrical Ge NWs with three different orientations at a gate overdrive voltage of $V_G - V_T = -1$, and for different diameters $2R_s = 3, 4, 5, 8, 10, 15, 20$ nm. Two slices are considered to study the anisotropy (shown in the inset): an horizontal slice (solid lines), and a diagonal one (dashed lines). Arrows aims to larger devices.

A more intuitive vision of the charge distribution is shown in Fig. 6.12, where the charge density along an horizontal direction (solid lines) and a diagonal direction (dashed lines) are represented for Ge NWs of different diameters, ranging from 3nm to 20nm, and three orientations ([001], [011] and [111]). In order to compare the shape of the charge distributions, the position along the channel has been normalized by each device radius. We can observe that the smaller the device, the more confined the charge is in the center of the device, pushing the carriers away from the interface. Furthermore, the peak of the charge density is larger for the smaller devices. It is also shown that [111]-oriented NWs exhibit an almost isotropic behavior, whereas the [001] and [011] (in particular the latter) oriented NWs are anisotropic for all the sizes.

6.2.1 Inversion charge centroid

The inversion charge centroid is an useful magnitude, which assesses the average distance of the inversion charge from the semiconductor-insulator interface. This magnitude is related to the gate capacitance, and is also related to the interaction between the charge and the surface-related scattering mechanisms, such as surface roughness or Coulomb scattering.

In a classical description of the device behavior, the charge is located right at the insulator-semiconductor interface. However, one of the more relevant consequences of the quantum effects in a NW regards with the distribution of the charge inside the
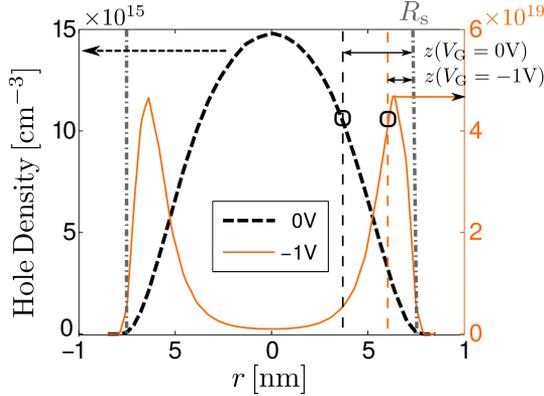
Figure 6.13: Representation of the inversion charge centroid for a $2R_\mathrm{s} = 15\,\mathrm{nm}$ Ge NW for two different V_G values: 0V (dashed), -1V (solid). The hole density for both polarization have been represented in two different scales: the reference for $V_\mathrm{G} = 0\mathrm{V}$ is the right axis, whereas for $V_\mathrm{G} = -1\mathrm{V}$ is the right axis.

device, which is shifted away from the interface.

To measure this effect, the inversion charge centroid z_I was defined [77, 114, 115]. In a cylindrical NW, it can be calculated as:

$$z_\mathrm{I} = R_s - \Delta\,, \tag{6.10}$$

with

$$\Delta = \frac{\int_0^{R_s} r^2 \rho_\mathrm{q}(r)\,\mathrm{d}r}{\int_0^{R_s} r \rho_\mathrm{q}(r)\,\mathrm{d}r}\,, \tag{6.11}$$

where R_s is the radius of the cylindrical semiconductor channel and ρ_q is the charge density. For a better intuition on the centroid meaning, Fig. 6.13 shows an example of their calculation for a $2R_\mathrm{s} = 15\,\mathrm{nm}$ cylindrical Ge NW at low and high inversion.

Figure 6.14 illustrates the centroid calculated as a function of the total hole density for Si (left) and Ge (right) [001]-oriented NWs. We can observe that the charge is mainly placed in the center of the device for low inversion and, as the device is driven to high inversion, the carriers approach to the interface (centroid is reduced). This effect is negligible for very small devices, which means that the charge remains in the center of the device.
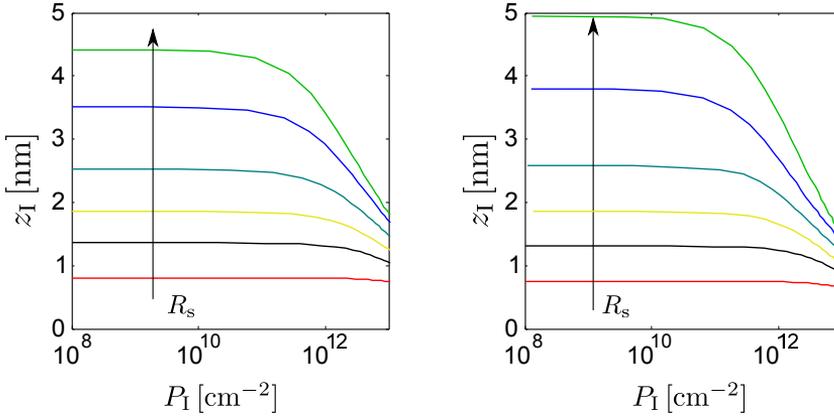
Figure 6.14: Charge centroid calculated for Si (left figure) and Ge (right figure) NWs oriented along the [001] direction as a function of the inversion hole density for different diameters $2R_s = 3, 5, 7, 10, 15, 20$nm. The arrow aims to larger diameters.

6.3 Charge and capacitance analysis of SiGe NWs

In this section, we will analyze the behavior of $Si_{1-x}Ge_x$ NWs, in terms of their charge and capacitance behavior, as a function of the Ge molar fraction x. To do it, let us study first the behavior of Si and Ge NWs, which constitute the limit cases for $x = 0$ and $x = 1$, respectively.

Figure 6.15 illustrates the linear hole density as a function of the gate overdrive voltage $(V_G - V_T)$ for NWs of Si and Ge of different diameters ($2R_s = 3, 5, 7, 10, 15, 20$nm) along two orientations ([001] and [111]). As can be observed, the differences for the total charge are not very relevant between both orientations. However, there are important differences between the Si and Ge in terms of the threshold voltage variation with the device size (which is depicted in the insets). This stronger influence of the confinement on V_T for Ge is due to its lower mean effective mass.

Another very relevant parameter in the performance of semiconductor devices is the gate capacitance C_G, defined as

$$C_G = \frac{\partial Q_G}{\partial V_G} = -\frac{\partial}{\partial V_G}(Q_{ins} + Q_s) \approx -\frac{\partial Q_s}{\partial V_G}, \tag{6.12}$$

where Q_G, Q_{ins}, and Q_s stand for the gate charge, the insulator charge and the semiconductor charge, respectively. In the previous equation we assumed that the insulator charge is negligible compared to Q_s. In order to express the gate capacitance in terms
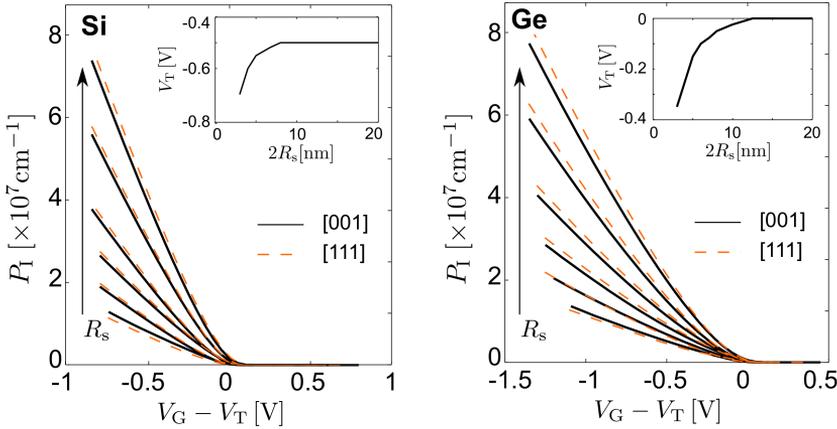
Figure 6.15: Hole linear density vs gate overdrive voltage $(V_G - V_T)$ for Si (left figure) and Ge (right figure) for different diameters $2R_s = 3, 5, 7, 10, 15, 20$nm and for two orientations: solid lines for [001] and dashed lines for [111]. The arrow aims to larger diameters.

of its different contributions we write

$$\frac{1}{C_G} = -\frac{\partial V_G}{\partial Q_s} = -\frac{\partial(V_G - \psi_s)}{\partial Q_s} - \frac{\partial \psi_s}{\partial Q_s},\tag{6.13}$$

where ψ_s is the semiconductor surface potential that can be related to V_G by means of the Gauss's law:

$$V_G - \Phi_{ms} = \psi_s - \frac{Q_s}{C_{ins}}\tag{6.14}$$

with C_{ins} the insulator capacitance per unit length and Φ_{ms} is the difference between the metal and semiconductor work functions. Neglecting any contribution coming from the depletion charge we can write:

$$\frac{1}{C_G} = \frac{1}{C_{ins}} + \frac{1}{C_{inv}}\tag{6.15}$$

where $C_{inv} = -\frac{\partial Q_{inv}}{\partial \psi_s}$ is the inversion capacitance.

The inversion capacitance measures the gate control over the inversion charge in the channel, and it is desirable to keep it as large as possible. One important challenge faced when device dimensions are shrunk is to achieve high values for that capacitance. C_{ins} is determined by the insulator thickness, the geometry of the structure and by the insulator permittivity.
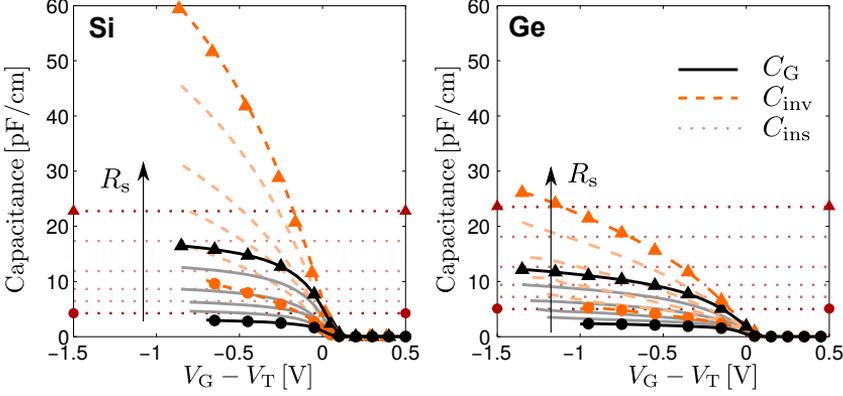
Figure 6.16: Capacitances of Si (left figure) and Ge (right figure) [001] oriented NWs for different diameters $2R_s = 3, 5, 7, 10, 15, 20$nm ($T_{ins} = 1$nm for all cases). The plots for the extremal diameters are highlighted: triangles for 20nm, and circles for 3nm, whereas the intermediate values are depicted in a lighter tonality for clarity. The total C_G (solid lines) and their contributions C_{ins} (dotted lines) and C_{inv} (dashed lines) are represented.

For cylindrical NWs, the insulator capacitance per unit length can be calculated analytically as [116]

$$C_{ins} = \frac{2\pi\epsilon_s}{\ln\left(1 + \frac{T_{ins}}{R}\right)}. \tag{6.16}$$

From Eq. (6.16), it can be inferred that the value of the insulator capacitance could be increased by reducing the insulator thickness or by choosing an insulator with high permittivity, i.e., high-κ insulators. Both approaches face numerous technological challenges. Meanwhile, the value of C_{inv} depends on the density of states, and thereby, on the subband distribution, but also on the charge distribution along the channel [116]. This effect is also related to the charge centroid presented in the previous Section: the higher the charge centroid, the lower C_{inv} and, as a consequence, C_G.

Let us then compare Si and Ge NWs in terms of their C_G behavior. The gate capacitance and its components can be calculated from the simulated Q-V curves. Figure 6.16 shows the different capacitances calculated from the curves depicted in Fig. 6.15. As can be seen, Si NWs exhibit a better performance in terms of the total gate capacitance C_G (solid lines) than Ge NWs. These differences are produced by C_{inv}, which depends on the charge centroid and the density of states. In this case, Si
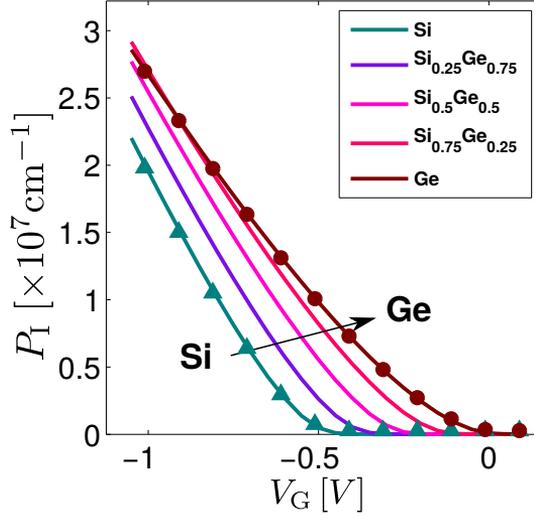
Figure 6.17: Hole linear density for $Si_{1-x}Ge_x$ 10nm NWs [001] oriented of different Ge molar fractions $x = 0, 0.25, 0.5, 0.75, 1$.

outperforms Ge in terms of the electrostatic performance.

We have finally simulated $Si_{1-x}Ge_x$ NWs with different Ge molar fraction. Fig. 6.17 illustrates the Q-V characteristic for different x values for a [001]-oriented $Si_{1-x}Ge_x$ NW with 10nm diameter. It can be observed that the intermediate values are a progressive mixture of the characteristics of the pure Si and Ge, as expected. It should be highlighted that the threshold voltage is tuned by the x value, due to the different work functions of Si and Ge. In Fig. 6.18, the resulting capacitances are represented and it again confirms the progressive transformation of the capacitances from those corresponding to Si to the Ge ones. As can be seen, high Ge molar fractions degrade the total gate capacitance, due to the reduction of the inversion capacitance.
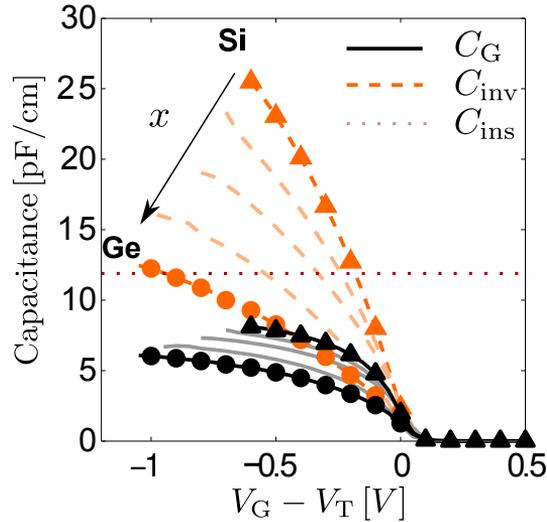
Figure 6.18: Capacitances of $Si_{1-x}Ge_x$ NWs with 10nm of diameter along the [001] direction for different Ge molar fractions $x = 0, 0.25, 0.5, 0.75, 1$. The plots for Si are marked with triangles, and Ge with circles, whereas the intermediate values are depicted in a lighter tonality for clarity. The total C_G (solid lines) and their contributions C_{ins} (dotted lines) and C_{inv} (dashed lines) are represented.

6.4 Conclusions

In this Chapter, different electrostatic magnitudes have been analyzed using the results provided by the SP2D simulator. The self consistent solution and the dependence of the band structure on the gate voltage have been assessed. It was demonstrated that the self-consistent solution is needed to get accurate results.

We have studied the electrostatic results provided by the $\boldsymbol{k\cdot p}$ method for NWs made of Si, Ge, and III-V semiconductors from 3nm to 20nm. Different orientations, geometries and materials were considered. The evolution of the band structure as a function of the device size in Si and Ge NWs was shown. As the diameter is reduced, the LH subbands split from the HH ones. For the smallest sizes the LH subbands predominate with respect the HH, and, as a consequence, a reduction of the mean effective mass is expected. This effect is more accentuated in Si devices than in Ge ones, and similar mean effective masses can be achieved for the smallest devices, specially along the [111] orientation.

It was demonstrated that the mean effective mass increases at high inversion since

the HH states get more populated. Regarding the hole density, a strong anisotropy has been observed for [011] oriented NWs, whereas, for the [001] and [011] orientations, the anisotropy is much smaller.

We have analyzed the hole linear density and the gate capacitance for Si and Ge NWs. A higher capacitance for the Si NWs has been reported, due to its higher density of states. A study on $Si_{1-x}Ge_x$ NWs has depicted the progressive variation of the electrostatic properties as a function of the Ge molar fraction.

Chapter 7

Carrier transport in semiconductor NWs

This chapter is focused on the study of the carrier mobility in semiconductor NWs. To perform this study, a semi-classical approach, that implies the resolution of the Boltzmann Transport Equation (BTE) is employed [40, 79]. This approach proposes a classical model where many particles can fly freely in a volume until a scattering event occurs and its momentum and/or energy change. Moreover, quantum effects are included to achieve an accurate description of the problem.

The lay out of the problem is very difficult due to the huge complexity of modeling statistically all the possible events that might occur. As it happened with the Schrödinger and Poisson equations in the electrostatic description, it is not straightforward to achieve an analytical solution of the BTE in most practical cases.

For that reason, a numerical solution of the BTE is commonly used to analyze the transport behavior in different types of electronic devices. Among others, we can highlight the deterministic description [74], the Monte Carlo approach, which solves the BTE collecting statistics [47, 118], or the Momentum Relaxation Time (MRT) approximation, which linearizes the BTE [12, 37, 63, 83, 100, 107]. In particular, in this work we will make use of the latter one, the MRT, to solve the BTE. In this regard, an in-house simulation tool has been developed to implement this method, that will allow to estimate the carrier mobility in semiconductor NWs. More details on this theory can be found elsewhere [83].

The main goals of this Chapter are: 1) to present a brief summary on the theoretical

background of the BTE, in particular for a 1D carrier gas, 2) to solve the BTE by means of the MRT discussing the main approaches employed, highlighting those related to the $\boldsymbol{k} \cdot \boldsymbol{p}$ method used to calculate the devices band structure and wave functions, and 3) to employ this approach to analyze a high-interest practical case, the mobility of SiGe NWs.

The Chapter is outlined as follows. In Section 7.1 we introduce the BTE to describe the carrier transport in a 1D gas. In Section 7.2, the linearization of the BTE and the definition of the MRT are presented. Section 7.3 shows how to calculate the MRT, explaining the explicit approximation used in this work. Then, in Section 7.4 the different types of scattering events are enumerated and characterized. Section 7.5 is devoted to the Kubo-Greenwood formula for the calculation of the mobility from the solutions of the linearized BTE. Section 7.6 uses the developed method to determine the hole mobility in $\mathrm{Si}_{1-\mathrm{x}}\mathrm{Ge}_{\mathrm{x}}$ NWs as a function of the Ge molar fraction, analyzes the influence of the form factor approach used to evaluate the scattering mechanisms MRTs, and studies the influence of the alloy disorder on the behavior of the devices. Finally, Section 7.7 recapitulates the main conclusions obtained throughout the Chapter.

7.1 Boltzmann Transport Equation

This Section is devoted to the BTE and its particularization to the carrier transport in a 1D gas scenario. The BTE is based on a statistical description of the classical particle behaviour, providing with the probability distribution of the particle evolution in real space, momentum space and time. To do this, we define $f = f(\boldsymbol{r}, \boldsymbol{k}, t)$ as the probability of finding an occupied state determined by a certain momentum $\boldsymbol{k}$, in the position $\boldsymbol{r}$, at an instant t. Its derivation can be found in several textbooks [29, 40, 41, 53, 79]. The most general formulation establishes that the change of the distribution function f with respect to the time can be written as:

$$\frac{\partial f(\boldsymbol{r}, \boldsymbol{k}, t)}{\partial t} \pm \frac{1}{\hbar} \nabla_{\boldsymbol{k}} E(\boldsymbol{r}, \boldsymbol{k}) \cdot \nabla f(\boldsymbol{r}, \boldsymbol{k}, t) \mp \frac{q\boldsymbol{F}}{\hbar} \cdot \nabla_{\boldsymbol{k}} f(\boldsymbol{r}, \boldsymbol{k}, t) = S_{\mathrm{in}}(\boldsymbol{r}, \boldsymbol{k}, t) - S_{\mathrm{out}}(\boldsymbol{r}, \boldsymbol{k}, t),$$

(7.1)

where the upper sign is used for electrons and the lower sign for holes (this notation will be kept hereinafter). The second term on the lhs of Eq. (7.1) takes into account the change in f due to the diffusion of particles, accounting for changes in the position, and

can be expressed in terms of the group velocity of the particle, $v_n(\boldsymbol{k}) = \pm \nabla_{\boldsymbol{k}} E_n(\boldsymbol{k})/\hbar$. The third term on the lhs of Eq. (7.1) stands for the external forces producing a change in the carrier momentum. Finally, the rhs of the equation is due to the presence of random scattering events which change the carrier trajectory in the classical sense, i.e., making the particle to change from one state to another. So that, S_{in} and S_{out} stand for the scattering rate of particles jumping from any state to $|\boldsymbol{k}, \boldsymbol{r}\rangle$ after a collision and the scattering rate of particles changing from state $|\boldsymbol{k}, \boldsymbol{r}\rangle$ to another one, respectively.

Equation (7.1) determines the rate of change in the occupation probability assigned to each state in a 3D carrier gas under a semi-classical approach. However, in this manuscript the object of study are NWs, and thus the carriers are confined in 2D. Therefore, the BTE defined for classical system, must be reassembled to take into account the quantum effects determined by the solution of the Schrödinger equation by means of the $\boldsymbol{k} \cdot \boldsymbol{p}$ method.

To do so, we make use of the Gradual Channel Approximation (GCA) which assumes that the variation of the electric field along the transport direction, z, is much lower than the corresponding variation in the confinement plane. Then, the Hamiltonian is assumed to be constant along z and the inclusion of $V(z)$ is modeled as a variation of the Fermi level along the NW length that has no effect on the quantum description of the device [63]. In this scenario, the determination of the quantum behavior of the device is restricted to the solution of the $\boldsymbol{k} \cdot \boldsymbol{p}$ equation in the cross-section of the device, defined by the coordinates $\mathbf{r} = (x, y)$, as described in Chapter 4. This slowly varying potential along the z axis causes the drift of the carriers.

From the previous reasoning we infer that the three dimensional problem involving a six dimensional space formed by $(\boldsymbol{r}, \boldsymbol{k})$ is reduced to (z, k_{z}) for 2D confined systems[1]. In this scenario, equation (7.1) is particularized for the 1D carrier gas as:

$$\frac{\partial f_n(z, k_{\text{z}}, t)}{\partial t} \pm \frac{1}{\hbar} \frac{\partial E_n(k_{\text{z}})}{\partial k_{\text{z}}} \frac{\partial f_n(z, k_{\text{z}}, t)}{\partial z} \mp \frac{qF_{\text{z}}}{\hbar} \frac{\partial f_n(z, k_{\text{z}}, t)}{\partial k_{\text{z}}} =$$
$$= S_{\text{in},n}(z, k_{\text{z}}, t) - S_{\text{out},n}(z, k_{\text{z}}, t), \quad (7.2)$$

where S_{in}, S_{out} and f are particularized for each subband n and $F_{\text{z}} = -\partial V(z)/\partial z$ is the longitudinal component of the electric field.

---

[1] Let us recall that the momentum of a carrier in 2D systems is described by two quantum numbers: the wave vector k_{z} and the subband index n.

Once the probability distribution f_n is known, several macroscopic magnitudes such as the carrier concentration (n for electrons or p for holes), and the current I_z can be determined.

The carrier concentration is determined from the probability distribution of each state, which in NWs reads

$$n(\boldsymbol{r}, t) = \frac{1}{\mathcal{L}} \sum_n f_n(z, k_z, t) |\boldsymbol{\xi}_{nk_z}(\mathbf{r})|^2 \,, \tag{7.3}$$

$$p(\boldsymbol{r}, t) = \frac{1}{\mathcal{L}} \sum_n f_n(z, k_z, t) |\boldsymbol{\xi}_{nk_z}(\mathbf{r})|^2 \,, \tag{7.4}$$

for electrons and holes, respectively. The current can be calculated as:

$$I_z = \pm \frac{q\mathcal{S}}{\mathcal{L}} \sum_n \sum_{k_z} v_n(k_z) f_n(z, k_z, t) \,, \tag{7.5}$$

where the plus sign is for electrons and the minus for holes.

The $S_{\text{in},n}$ and $S_{\text{out},n}$ can be attributed to random scattering events that modify the carrier free flights, changing their ballistic trajectory, and therefore, their momentum. There are three assumptions to be considered in the modeling of the scattering rates [79]: 1) the collisions occur in a very short time scale almost instantaneous; 2) the collisions produce a change in the particles trajectory modifying their wavenumber but they do not change their position; and 3) the scattering interaction is weak. Under these three assumptions, $S_{\text{out},n}$ can be written as:

$$S_{\text{out},n}(z, k_z, t) = f_n(z, k_z, t) \sum_{n', k_z'} S_{nn'}(k_z, k_z') \big[1 - f_{n'}(z, k_z', t) \big] \,, \tag{7.6}$$

where $S_{nn'}(k_z, k_z')$ is the so-called scattering rate and denotes the probability per unit time that a particle suffers a transition from an initial state $|n, k_z\rangle$ to the final state $|n', k_z'\rangle$, and the sum runs over all the final states. The interpretation of Eq. (7.6) is that the rate of electrons leaving the state i depends on the probability of occupation of that state multiplied by the sum of all the probabilities that a transition to an empty state j happens. A similar reasoning can be inferred to obtain the scattering-in flux,

$S_{\mathrm{in},n}$:

$$S_{\mathrm{in},n}(z, k_z, t) = [1 - f_n(z, k_z, t)] \sum_{n', k_z'} S_{nn'}(k_z, k_z') f_{n'}(z, k_z', t). \qquad (7.7)$$

Substituting Eqs. (7.6) and (7.7) into Eq. (7.2), the resulting 1D BTE for the n-th subband in terms of the occupation probability of all the states involved is:

$$\frac{\mathrm{d}f_n(z, k_z, t)}{\mathrm{d}t} - \frac{1}{\hbar} \frac{\partial E_n(k_z)}{\partial k_z} \frac{\partial f_n(z, k_z, t)}{\partial z} + \frac{qF_z}{\hbar} \frac{\partial f_n(z, k_z, t)}{\partial k_z} =$$
$$[1 - f_n(z, k_z, t)] \sum_{n' k_z'} S_{nn'}(k_z, k_z') f_{n'}(z, k_z', t) - f_n(z, k_z, t) \sum_{n' k_z'} S_{n'n}(k_z', k_z) \big[1 - f_{n'}(z, k_z', t) \big]$$

$$(7.8)$$

where the scattering rate $S_{nn'}(k_z, k_z')$ is defined independently for the different physical processes. The scattering events are consequence of a small perturbation of the Hamiltonian describing the system. Under this premise it is possible to use the time dependent perturbation theory which lead us to the so-called Fermi Golden Rule:

$$S_{nn'}(k_z, k_z') = \frac{2\pi}{\hbar} \big| M_{nn'}(k_z, k_z') \big|^2 \delta(E_n(k_z) - E_{n'}(k_z')). \qquad (7.9)$$

This expression will be employed to determine the scattering rate associated to each mechanism. A comprehensive derivation for confined system can be found in different textbooks [29, 40, 79].

7.2 Momentum Relaxation Time

So far, an expression to calculate the evolution of the distribution function $f_n(z, k_z, t)$ has been provided using a description of the semi-classical BTE. However, this expression is a complex integro-differential non linear equation and further approximations must be carried out to address it. In this Section we present the Momentum Relaxation Time approximation (MRT), which can be used to solve the BTE and then to calculate the mobility, as shown in Section 7.5.

The MRT assumes small displacements of the distribution function $f \approx f_0 + \delta f$ from the equilibrium due to an external stimulus such as the electric field in the transport direction F_z. It also assumes uniform transport conditions, i.e., the deviation from

equilibrium is constant along the transport direction:

$$f_n(z, k_z, t) = f_n(k_z, t) \,, \tag{7.10}$$

$$f_n(z, k_z, t) = f_0(E_n(k_z)) + \delta f_n(k_z, t) \,, \tag{7.11}$$

where the equilibrium occupation probability $f_0(E)$ is the Fermi-Dirac function, and $\delta f_n(k_z)$ is the deviation of the distribution function $f_n(k_z, t)$ from the equilibrium due to an external stimulus, the electric field F_z .

The calculation of the mobility depends on the initial and the final states, which are determined by their subband index and wave vector. In the following, we simplify the notation by only using the subindex i or j so as to denote the initial and final states respectively:

$$i \rightarrow |n, k_z\rangle \,, \tag{7.12}$$

$$j \rightarrow |n', k_z'\rangle \,. \tag{7.13}$$

The MRT assumes that the rhs term in Eq. (7.8), corresponding to the input and output scattering fluxes, can be expressed as:

$$S_{\mathrm{in},i} - S_{\mathrm{out},i} = -\frac{\delta f_i}{\tau_i} \,, \tag{7.14}$$

where τ_i is the momentum relaxation time of the state i because it is related to the time that δf_i takes to vanish or relax from the instant that the stimulus F_z ceases. From Eq. (7.8), assuming that $F_z = 0$, we can write:

$$\frac{\mathrm{d}f_i(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\delta f_i(t) = -\frac{\delta f_i(t)}{\tau_i} \rightarrow \delta f_i(t) = \delta f_{i0}\, \mathrm{e}^{-t/\tau_i} \,. \tag{7.15}$$

So that, τ_i measures the mean time that the distribution function f_i needs to be reduced in a factor e after the origin of the perturbation is switched off.

However, the MRT approach also assumes stationary behavior, i.e., the system has evolved to stationary equilibrium and $\partial f_i / \partial t = 0$. Therefore, under a small electric field F_z, Eq. (7.2) can then be written as:

$$-\frac{qF_z}{\hbar} \frac{\partial f_i}{\partial k_z} = -\frac{\delta f_i}{\tau_i} \,. \tag{7.16}$$

The partial derivative of f_i on the left, using the definition in Eq. (7.11), can be expressed as:

$$\frac{\partial f_i}{\partial k_z} = \frac{\partial \delta f_i}{\partial k_z} + \frac{\partial f_0(E_i)}{\partial k_z} = \frac{\partial \delta f_i}{\partial k_z} + \hbar \frac{\partial f_0(E_i)}{\partial E} v_i \,, \tag{7.17}$$

where v_i is the carrier velocity in the transport direction, $v_i = 1/\hbar \, \partial E_i / \partial k_z$. The first term on the rhs, corresponding to the small deviation, can be neglected when it is multiplyed by the small F_z. Then, Eq. (7.16) leads to:

$$\delta f_i = -q F_z v_i \tau_i \frac{\partial f_0(E_i)}{\partial E} \,. \tag{7.18}$$

Equation (7.18) shows that the deviation δf_i, and thus, the probability function f_i can be uniquely determined by the momentum relaxation time τ_i. Therefore, once the momentum relaxation time is calculated, an estimation of macroscopic quantities which are determined from f_i can be evaluated. The next step is to calculate this relaxation time.

7.3 Calculation of the momentum relaxation time

In this Section, we calculate the momentum relaxation time from the scattering-in and scattering-out fluxes. We firstly relate δf_i to the scattering rates in Eqs. (7.6) and (7.7), to achieve an expression for τ_i as a function of the scattering rate S_{ij} and δf_n. The difference of the input and output fluxes can be written as [40]:

$$S_{\text{in},i} - S_{\text{out},i} = -\delta f_i \sum_j S_{ij} \left[\frac{1 - f_0(E_j)}{1 - f_0(E_i)} - \frac{\delta f_j}{\delta f_i} \frac{f_0(E_i)}{f_0(E_j)} \right] \,, \tag{7.19}$$

where we have substituted $f_i = \delta f_i + f_0(E_i)$, discarded the second order terms involving the product $\delta f_i \delta f_j$, and used the expression of the flux balance at equilibrium:

$$f_0(E_j) S_{ij}(1 - f_0(E_i)) = f_0(E_i) S_{ij}(1 - f_0(E_j)) \,. \tag{7.20}$$

The ratio $\delta f_j / \delta f_i$ can be calculated using Eq. (7.18) as:

$$\frac{\delta f_j}{\delta f_i} = \frac{q \tau_j F_z v_j \frac{\partial f_0(E_j)}{\partial E}}{q \tau_i F_z v_i \frac{\partial f_0(E_i)}{\partial E}} = \frac{\tau_j v_j f_0(E_j)[1 - f_0(E_j)]}{\tau_i v_i f_0(E_i)[1 - f_0(E_i)]} \,, \tag{7.21}$$

where the derivative of the Fermi-Dirac function $\partial f_0(E_i)/\partial E = (1/k_{\mathrm{B}}T)\, fo(E_i)[1 - f_0(E_i)]$ is employed. Then, substituting Eq. (7.21) in the rhs term Eq. (7.19), and (7.14) in the lhs, the following expression for the momentum relaxation time τ_i is obtained:

$$\frac{1}{\tau_i} = \sum_j S_{ij} \left[\frac{1 - f_0(E_j)}{1 - f_0(E_i)}\right]\left[1 - \frac{\tau_j v_j}{\tau_i v_i}\right].$$

(7.22)

When different scattering mechanisms are accounted for, the scattering rate S_{ij} includes the contribution of each of them (S_{ij}^m) as:

$$S_{ij} = \sum_m S_{ij}^m,$$

(7.23)

where m runs over all the possible scattering mechanisms.

The MRT associated to the m-th scattering mechanism can be calculated as in Eq.(7.22):

$$\frac{1}{\tau_i^m} = \sum_j S_{ij}^m \left[\frac{1 - f_0(E_j)}{1 - f_0(E_i)}\right]\left[1 - \frac{\tau_j v_j}{\tau_i v_i}\right].$$

(7.24)

So that, the total MRT τ_i can be calculated once the MRT corresponding to each scattering mechanism τ_i^m is known as:

$$\frac{1}{\tau_i} = \sum_m \frac{1}{\tau_i^m}.$$

(7.25)

This expression is the general formulation for the calculation of the momentum relaxation times for all the scattering mechanisms. Notwithstanding, τ_i^m can not be calculated directly since it depends on the total τ_i (Eq. (7.24)), and thus, on the MRT of the rest of states τ_j. Hence, an implicit calculation would be needed to solve Eqs. (7.24) and (7.25). This problem is still computationally affordable and an example about how to do deal with it can be found in [83], particularized for the EMA. However, a drastic simplification of the problem can be achieved under the so-called explicit approximation.

To avoid the complexity inherent to the implicit model, a drastic simplification is carried out using the explicit approximation for the calculation of the MRT, which has been extensively used in the literature [12, 37, 63, 69, 99, 107]. So that, we have decided to make use of this approach to simplify the calculations.

The usual assumption of the explicit calculation considers that the momentum relaxation times for the initial τ_i and final τ_j states exhibit small differences between them. Then, it can be considered that $\tau_j/\tau_i \simeq 1$, and the general form for the MRT in Eq. (7.22) can be simplified to:

$$\frac{1}{\tau_i^m} = \sum_m S_{ij}^m \left[\frac{1 - f_0(E_j)}{1 - f_0(E_i)}\right]\left[1 - \frac{v_j}{v_i}\right]. \tag{7.26}$$

Therefore, the explicit approach allows to calculate the MRT for each state independently, avoiding the dependence on the total τ_i. Once the MRTs for the different scattering mechanisms are calculated, Eq. (7.25) is employed to calculate the total MRT. The explicit assumption has been proven to provide excellent results in different cases [42, 72].

The evaluation of Eq. (7.26) can be simplified for different scattering mechanisms. For example, if an elastic scattering mechanism is considered, no change in energy between the initial and the final state is produced, and therefore the term $(1 - f_0(E_j))/(1 - f_0(E_i)) = 1$, reducing the equation to:

$$\frac{1}{\tau_i^m} = \sum_j S_{ij}^m \left[1 - \frac{v_j}{v_i}\right]. \tag{7.27}$$

Another useful case in which Eq. (7.26) can be simplified is in isotropic mechanisms, where S_{ij} can be written as $S_{nn'}$, i.e., there is no dependence with k_z and k_z'. In that case, the second term into brackets disappears, reducing the complexity of the problem. Nevertheless, it should be highlighted that, under the $\boldsymbol{k}\cdot\boldsymbol{p}$ method, this situation is not usual, as the initial and final wave functions depend on k_z and k_z', respectively, and therefore so it does S_{ij}.

Figure 7.1 presents the available transitions from an initial state i to the final states j for elastic transitions (left) and inelastic transitions (right), where $\Delta E = \pm\hbar\omega$ is the energy emission and absorption of the inelastic transition, respectively. As can be seen, several states can be reached in a transition. When the final subband coincides with the initial one, it is denominated intrasubband transition. Otherwise, it is a intersubband transition. It should be noticed that in $\boldsymbol{k}\cdot\boldsymbol{p}$ band structures, for a given energy, there can be several states corresponding to the same subband. This fact will be important when dealing with the mobility calculation in Sec. 7.5.
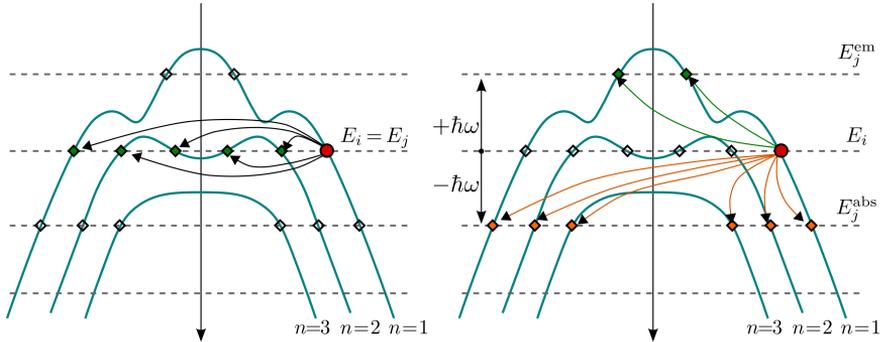
Figure 7.1: Available transitions from an initial state i (marked with a circle) to other states (marked with diamonds). The elastic transitions are shown on the left while the inelastic transitions are plotted on the right figure, where $\Delta E = \pm\hbar\omega$ is the energy for the emission and absorption respectively.

7.4 Implementation of scattering mechanisms

Carrier transport in NWs is determined by different scattering mechanisms: surface roughness, coulomb dispersion, bulk phonons (optical and acoustic), polar optical phonons and alloy disorder [83]. In this chapter we will focus on acoustic and optical phonons and alloy disorder. Our aim is to implement these scattering mechanism taking into account the $\boldsymbol{k}\cdot\boldsymbol{p}$ method particularities.

7.4.1 Phonons

Acoustic (AC) and optical phonons (OP) are usually the main cause of scattering in semiconductors. They are related to the vibration of the lattice atoms around their equilibrium positions. This vibration produces a perturbation in the potential that affects the motion of the carriers by changing their momentum and/or energy. The mathematical description of the carrier-phonon interaction can be found in several textbooks [40, 148]. In this manuscript, we sketch the main ideas necessary to estimate the matrix elements that describe the interaction.

The usual scheme is to model the atoms of the lattice as harmonic oscillators, with wave vector $\boldsymbol{q}$, and energy $E_{\mathrm{ph}} = \hbar\omega$, being ω their oscillation frequency. There are two modes of oscillations according to how adjacent atoms behave: a) when the two atoms oscillate with the same phase, they are named as acoustic (AC) phonons, b) when the
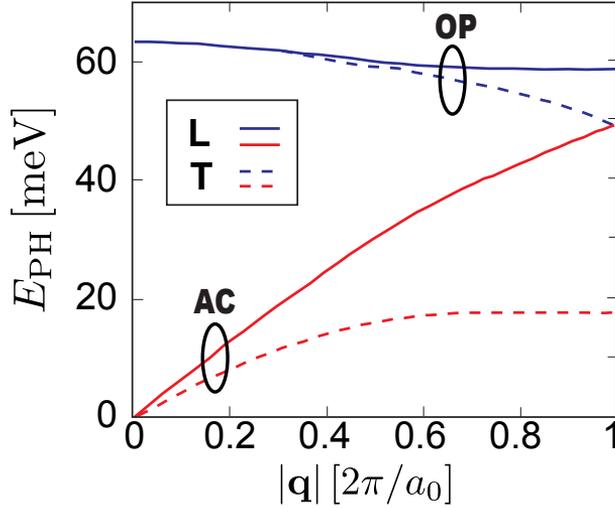
Figure 7.2: Phonon energy E_{Ph} of the different modes for bulk Si as a function of the phonon wave vector magnitude $|\boldsymbol{q}|$ along the [001] direction. Transversal and longitudinal modes are plotted as dashed and solid lines respectively. Acoustic and optic modes are marked as AC and OP, respectively. After Refs. [20, 76].

two atoms oscillate with opposite phases, they are named as optical (OP) phonons. Regarding the direction of vibration of the modes, they can be transverse (T) when the atoms vibrate perpendicularly to the phonon wave vector $\boldsymbol{q}$ direction (there are two transverse modes), and longitudinal (L) when the atoms vibrate aligned with $\boldsymbol{q}$ (there is one longitudinal mode).

The phonon oscillation modes are shown in Fig. 7.2, where the phonon energy is plotted as a function of the phonon wave vector magnitude $|\boldsymbol{q}|$. The transversal modes (TA, TO) are degenerated. Under the assumption of low $|\boldsymbol{q}|$ values, acoustic phonon modes exhibit a linear dependence of the energy with respect to the phonon wave vector $E_{PH} = \hbar v_s |\boldsymbol{q}|$, where v_s is the slope of the $\hbar\omega$-$|\boldsymbol{q}|$ curve and it is the so called sound velocity in that direction. For optical phonons, LO and TO, a constant energy with value proportional to the angular frequency of the phonon $E_{PH}(|\boldsymbol{q}|) = \hbar\omega_{PH}$, can be assumed.

The potential induced by the vibrations of the crystal is given by [40]:

$$U_{PH}(\boldsymbol{r}, t) = D(\boldsymbol{q}) \sqrt{\frac{\hbar}{2\rho\Omega\omega_{PH}(\boldsymbol{q})}} a_\nu(\boldsymbol{q}) \mathrm{e}^{\pm\mathrm{i}(\boldsymbol{q}\boldsymbol{r} - \omega_{PH}(\boldsymbol{q})t)} , \qquad (7.28)$$

where the $\pm$ sign corresponds to phonon emission and absorption processes, respectively, $D(\boldsymbol{q})$ is the phonon deformation potential, ρ is the semiconductor density, $\omega_{\mathrm{PH}} = E_{\mathrm{PH}}/\hbar$ is the phonon frequency and $a_\nu(\boldsymbol{q})$ is the wave magnitude of the phonon branch ν, which is related to the number of phonons, $n_{\mathrm{PH}}(\boldsymbol{q})$, as $a_\nu(\boldsymbol{q}) = \sqrt{n_{\mathrm{PH}}(\boldsymbol{q}) + \frac{1}{2} \pm \frac{1}{2}}$, being $n_{\mathrm{PH}}(\boldsymbol{q})$ given by the Bose-Einstein distribution:

$$n_{\mathrm{PH}}(\boldsymbol{q}) = \frac{1}{\mathrm{e}^{\frac{E_{\mathrm{PH}}}{k_{\mathrm{B}} T}} - 1} . \tag{7.29}$$

7.4.2 Acoustic phonons

In the case of acoustic phonons, small values of $|\boldsymbol{q}|$ lead to negligible energies for the phonons compared to those of electrons, and then, the scattering mechanism can be modeled as an elastic process. So that, three simplifications can be made [79]: a) the time dependence in Eq. (7.28) can be dropped, b) Eq. (7.29) is approximated to $n_{\mathrm{PH}}(|\boldsymbol{q}|) \approx k_{\mathrm{B}} T / E_{\mathrm{PH}} = k_{\mathrm{B}} T / \hbar v_{\mathrm{s}} |\boldsymbol{q}|$, and c) isotropy of the phonons [40, 42]. These assumptions lead to the following definition for the matrix elements corresponding to transitions from the initial state $i = |n, k_{\mathrm{z}}\rangle$ to the final one $j = |n', k_{\mathrm{z}}'\rangle$:

$$|M_{ij}|^2 = \frac{1}{\mathcal{L}} \frac{D_{\mathrm{ac}}^2 k_{\mathrm{B}} T}{\rho v_{\mathrm{s}}^2} F_{ij} , \tag{7.30}$$

where D_{ac} is the acoustic deformation potential, $\mathcal{L}$ the wire length, ρ is the density of the material, and v_{s} is the sound velocity. F_{ij} is the so-called form factor, an overlap integral coupling the initial and final envelope functions calculated on the cross-section S of the NW as

$$F_{ij} = \int_S \left| \boldsymbol{\xi}_j^\dagger(\mathbf{r}) \cdot \boldsymbol{\xi}_i(\mathbf{r}) \right|^2 \mathrm{d}\mathbf{r} . \tag{7.31}$$

Eq. (7.30) includes both, emission and absorption processes. According to (7.30) the transition from one state i to other j is determined by the coupling of the wave functions expressed in Eq. (7.31), and does not depend on the phonon wave vector q_{z}.

Then, our major concern will be the calculation of the form factors. This is not a trivial problem since it involves a huge amount of data to be stored and processed. Some simplifications can be applied to these form factors, such as consider just one wave function for the complete subband, assuming that it does not changes too much (as proposed in [40]), or even approximate F_{ij} as a constant depending on whether the

transition is intrasubband or intersubband [80]. The limitations and consequences of applying these approaches are analyzed in Section 7.6.1.

7.4.3 Optical phonons

In the case of optical phonons, small q does not imply small energy. From Figure 7.2, it can be observed that the energy of the optical phonons is almost constant for any q value. The number of phonons (n_{PH}) and the deformation potential (D_{op}) can also be assumed to be constant. Therefore, the energy of these phonons can be approximated by a constant $E = \hbar\omega_{\mathrm{PH}}$. Hence, the carrier scattering by optical phonons is an inelastic process.

The square matrix element for NWs can then be written as [62, 68]:

$$|M_{ij}|^2 = \frac{D_{\mathrm{op}}^2}{\mathcal{L}} \frac{\hbar}{2\rho\omega_{\mathrm{op}}} \left(n_{\mathrm{PH}} + \frac{1}{2} \pm \frac{1}{2} \right) F_{ij}. \qquad (7.32)$$

According to this expression, like for acoustic phonons, the phonon matrix element does not dependen on the phonon wave vector q_{z}, being therefore modeled as an inelastic isotropic mechanism. Again, our main concern is the calculation of the form factors for the available transitions, in this case, the levels with energies $E_j = E_i \pm \hbar\omega_{\mathrm{PH}}$ (the plus and minus signs are for emission and absorption respectively). Then, the complete scattering rate for emission and absorption of optical phonons can be written as:

$$S_{ij}^{\mathrm{op,em}} = \frac{2\pi}{\hbar} \frac{D_{\mathrm{op}}^2}{\mathcal{L}} \frac{\hbar}{2\rho\omega_{\mathrm{op}}} \left(n_{\mathrm{PH}} + 1 \right) F_{ij} \, \delta(E_i + \hbar\omega - E_j), \qquad (7.33)$$

$$S_{ij}^{\mathrm{op,abs}} = \frac{2\pi}{\hbar} \frac{D_{\mathrm{op}}^2}{\mathcal{L}} \frac{\hbar}{2\rho\omega_{\mathrm{op}}} n_{\mathrm{PH}} F_{ij} \, \delta(E_i - \hbar\omega - E_j). \qquad (7.34)$$

7.4.4 Alloy Disorder

In alloy semiconductors, such as ternary III-V and SiGe, there is some randomness in the distribution of atoms in the lattice due to the substitution of some species of atoms by others. For $\mathrm{Si_{1-x}Ge_x}$, this randomness consists in the substitution of the atoms of a pure Si lattice by atoms of Ge to a certain molar fraction x. In the case of ternary III-V alloys with zinc blende structure, the scenario is different since the pure lattice

already has two different types of atoms. The zinc blende lattice can be imagined as two interpenetrating fcc lattices, corresponding to the elements of the compound. The ternary alloy only substitutes the atoms of one of these fcc lattices by a different one. For example, $In_{1-x}Ga_xAs$ is composed by InAs and GaAs that share the As. The ternary compound alloy consists in a fcc of pure As interpenetrated by a fcc composed by a mixture of In and Ga atoms in a ratio determined by the molar fraction x.

The random distribution produces a perturbation potential that scatters the carriers in their movement, the so-called Alloy Disorder (AD) scattering. To address this problem, the most common approach is the virtual crystal approximation, which assumes a periodic potential produced by a virtual lattice of average atoms, i.e., the potential is the mean of the potential produced by both types of atoms. An statistical study can be found in Refs. [13, 91]. The resulting scattering mechanism is considered to be elastic and the matrix element is written as:

$$|M_{ij}|^2 = \frac{2\pi}{\hbar} x(1-x) \Delta U^2 \Omega F_{ij} \,, \tag{7.35}$$

where Ω is the volume of the unit cell.

Regarding to the AD deformation potential ΔU, there is a certain controversy on its definition. Some authors consider it as the difference between the band edges of the different components of the alloy [13], while others try to physically explain its origin using atomistic models [91]. In other cases, it is simply considered as a fitting parameter, adjusted to reproduce experimental data [109].

7.5   Mobility calculation

As described in Section 7.2, under the assumption of small displacements from equilibrium, the MRT approximation can be employed and the distribution function $f_n(\mathbf{r}, k_z, t)$ used to calculate different macroscopic quantities. The mobility μ can be estimated using these quantities by means of the Kubo-Greenwood formula [50, 71]. In this section, we gather the necessary expressions for the mobility calculation in confined systems.

Assuming a linear relation between the current and the electric field (which is at the basis of the MRT approach), the current density of the n-th subband can be written as

$$J_n = q P_{I,n} \mu_n F_z \,, \tag{7.36}$$

where $P_{\mathrm{I},n}$ is the hole inversion density of the n-th band.

We can define the current density per unit length for the n-th subband as the contribution of all the initial states $i = |n, k_{\mathrm{z}}\rangle$ in that subband [40] as

$$J_n = \mp \frac{q}{L} \sum_{k_{\mathrm{z}}} v_n(k_{\mathrm{z}}) f_n(k_{\mathrm{z}}) = \mp \frac{q}{L} \sum_{k_{\mathrm{z}}} v_n(k_{\mathrm{z}}) \delta f_n(k_{\mathrm{z}}) \,, \tag{7.37}$$

where $f_n(k_{\mathrm{z}}) = f_0(E_n(k_{\mathrm{z}})) + \delta f_n(k_{\mathrm{z}})$ and the product $v_n(k_{\mathrm{z}}) f_0(E_n(k_{\mathrm{z}}))$ vanishes due to its odd symmetry with k_{z}, which is coherent with the fact that there is no current flow at equilibrium, and it depends exclusively on the variation of the probability distribution induced by the external field. Due to the continuity of the subband n in k_{z}, the sum must be converted to an appropriate integral by using the general prescription for 1D $\sum_{k_{\mathrm{z}}} = \mathcal{L}/2\pi \int \mathrm{d}k_{\mathrm{z}}$ [40] , resulting into:

$$J_n = \mp \frac{q}{2\pi} \int_{-\infty}^{\infty} v_n(k_{\mathrm{z}}) \delta f_n(k_{\mathrm{z}}) \, \mathrm{d}k_{\mathrm{z}} \,. \tag{7.38}$$

Substituting Eq. (7.38) in Eq. (7.18) we obtain

$$J_n = \frac{q^2}{2\pi} F_{\mathrm{z}} \int_{-\infty}^{\infty} v_n^2(k_{\mathrm{z}}) \tau_n(k_{\mathrm{z}}) \frac{\partial f_0(E_n(k_{\mathrm{z}}))}{\partial E} \, \mathrm{d}k_{\mathrm{z}} \,. \tag{7.39}$$

Finally, we get the mobility μ_n for the n-th subband from Eqs. (7.39) and (7.36):

$$\mu_n = \frac{q}{P_{\mathrm{I},n} 2\pi} \int_{-\infty}^{\infty} v_n^2(k_{\mathrm{z}}) \tau_n(k_{\mathrm{z}}) \frac{\partial f_0(E_n(k_{\mathrm{z}}))}{\partial E} \, \mathrm{d}k_{\mathrm{z}} \,, \tag{7.40}$$

which is the Kubo-Greenwood formula for the mobility of a 1D carrier gas [50, 71], or after operating on the derivative:

$$\mu_n = \frac{q}{P_{\mathrm{I},n} 2\pi k_{\mathrm{B}} T} \int_{-\infty}^{\infty} v_n^2(k_{\mathrm{z}}) \tau_n(k_{\mathrm{z}}) f_0(E_n(k_{\mathrm{z}}))[1 - f_0(E_n(k_{\mathrm{z}}))] \, \mathrm{d}k_{\mathrm{z}} \,, \tag{7.41}$$

It is interesting to express Eq. (7.41) in terms of the energy. To do this, it is necessary to perform a change of variable using the relation $\mathrm{d}k_{\mathrm{z}} = \mathrm{d}E/\hbar v_n(k_{\mathrm{z}})$

$$\mu_n = \frac{q}{P_{\mathrm{I},n} \hbar \pi k_{\mathrm{B}} T} \int_{-\infty}^{E_n} v_n(E) \tau_n(E) f_0(E)[1 - f_0(E)] \, \mathrm{d}E \,. \tag{7.42}$$

This change of variable is only valid when integrating monotonic functions. Then,
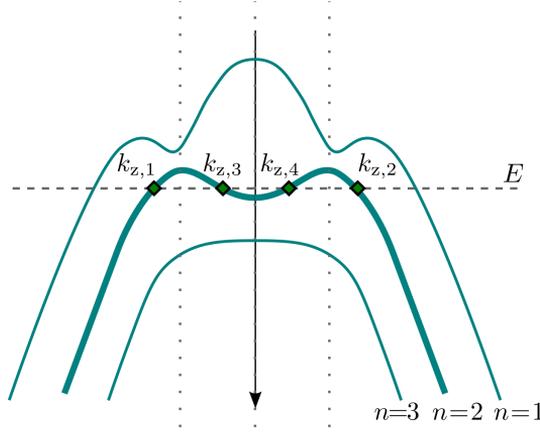
Figure 7.3: Representation of the intersecting points of subband E_2 with the energy E for non monotonic subbands. Vertical dotted lines represent the intervals for which E_2 is monotonic.

a monotonic dispersion relations $E_n(k_z)$ is required. This is the case for example in EMA for only the positive (or negative) values of k_z. So that, the integral in Eq. (7.42) is performed only in the positive k_z using $\int_{-\infty}^{\infty} \mathrm{d}k_z = 2 \int_{0}^{\infty} \mathrm{d}k_z$.

Notwithstanding, the band structure in $\boldsymbol{k \cdot p}$ is non monotonic since for a certain value of energy, E_n may be defined for different values of k_z, as shown in Fig. 7.3. Then, Eq. (7.41) must be split in subintervals K_a of k_z where the dispersion is monotonic

$$\mu_n = \frac{q}{P_{\mathrm{I},n}\hbar\pi k_{\mathrm{B}}T} \sum^{K_a} \int v_n(E)\tau_n(E)f_0(E)[1 - f_0(E)]\,\mathrm{d}E \,. \tag{7.43}$$

Since (7.43) is going to be solved numerically for discrete values of E, we can express it in a more intuitive manner as

$$\mu_n = \frac{q}{P_{\mathrm{I},n}\hbar\pi k_{\mathrm{B}}T} \int_{-\infty}^{E_n} \left[\sum_{k_z}^{E_n(k_z)=E} |v_n(k_z)\tau_n(k_z)| \right] f_0(E)[1 - f_0(E)]\,\mathrm{d}E \,, \tag{7.44}$$

which means that the summation $|v_n(k_z)\tau_n(k_z)|$, for a certain energy E must be performed on the k_z values such that the $E_n(k_z) = E$. The absolute value is needed to take into account the change on the sign of the integral limits in each K_a subinterval.

7.6 Results

This section is focused on the assessment of the hole mobility in SiGe NWs. Its interest stems from the fact that SiGe is postulated as a potential alternative to pure silicon technology for high performance pMOS devices [93]. Moreover it will be useful to test and show the performance of the mobility simulator implemented in this thesis. The main results of this work were published in [88].

It has been experimentally demonstrated that SiGe produces a boost in the hole mobility compared with pure Si, while keeping the compatibility with the Si technology. However, the behavior of SiGe NWs still has to be studied. As mentioned in Sec. 7.4.4, for this compound material there is an additional source of scattering with respect to the pure materials, the alloy disorder, which can counteract the hypothetical benefits that alloying would provide on the mobility due to their lower effective mass. Hence, our aim in this section is to assess the influence of the AD scattering on the hole mobility of $Si_{1-x}Ge_x$ NWs as a function of the Ge molar fraction x.

To calculate the band-structure and the envelope eigenfunctions, we have used the self-consistent solution of the Poisson equation and the six-band $\boldsymbol{k}\cdot\boldsymbol{p}$ method in the cross-section of a cylindrical NW with a diameter of 10 nm, and three different transport orientations: [001], [011] and [111]. The band structures for pure Si and Ge are shown in Fig. 7.4, where low ($P_I = 10^{10} cm^{-2}$) and high inversion ($P_I = 5 \times 10^{12} cm^{-2}$) are represented in solid and dashed lines respectively [2].

The mobility is calculated as described in this chapter. The scattering mechanisms accounted for are phonons (acoustic and optical) and AD. To evaluate the MRT for acoustic phonons, we have considered the bulk phonon model, which has been proven to deviate only slightly from the results provided by the confined phonons model [32].

In SiGe, there are three different optical modes to be considered, which stem from the vibration of the different pair of atom species Si-Si, Ge-Ge and Si-Ge. These modes are weighted by the probability of occurrence of each pair on the lattice [16]. Therefore, the MRT must be calculated for these three types of phonons, averaging them according to their probability of appearance [42].

AD is also considered as an elastic mechanism and the relaxation time is calculated in a similar way to acoustic phonons, as described in Sec. 7.4.4. The AD scattering

---

[2]The charge per unit of surface is calculated as the linear charge density in a cross-section divided by its perimeter.
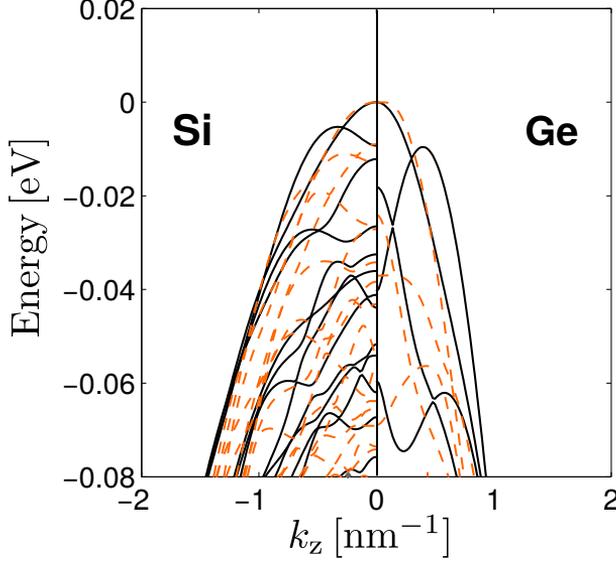
Figure 7.4: Bandstructure for Si (left side) and Ge (right side) NWs with a diameter of 10nm along the [001] orientation. Low inversion is represented in solid lines whereas high inversion is represented in dashed lines ($P_I = 5 \times 10^{12}$cm^{-2}). The energy reference is the top of the valence band edge in each case.

rate has been modeled using different parameters reported in the literature. As it was previously mentioned, there is a strong controversy on the definition of the ΔU parameter and different values can be found in the literature [42]. Due to the lack of comprehensive study of SiGe NWs, we have considered interesting to carry out simulations using different values of ΔU to analyze its influence on the total hole mobility as a function of the Ge molar fraction x.

---

[1]Ref. [42]
[2]Ref. [16]

Parameter	Literature	Our value
D_{ac} (eV)	$7.12 + 0.98\,x$ [1]	$7.12 + 2.8\,x$
D_{op} (10^8eV cm^{-1})	$13.24 - 1.07\,x$ [1]	–
$\hbar\omega_{\text{Si-Si}}$ (eV)	61.2[2]	–
$\hbar\omega_{\text{Si-Ge}}$ (eV)	45.0[2]	–
$\hbar\omega_{\text{Ge-Ge}}$ (eV)	37.0[2]	–

Table 7.1: Parameters of the Si$_{1-x}$Ge$_x$ alloy used for the calculation of the relaxation times.

First of all, the simulator has been benchmarked against results found in the literature for similar devices made on pure Si and Ge. To do this we have used the parameters presented in Table 7.1, where we have just modified the acoustic deformation potential to adjust our simulations with those presented in Ref. [101] for Ge. The results, shown in Table 7.2, are in good agreement with those provided by *Niquet et al.* in Ref. [102] for pure Si, and in Ref. [101] for pure Ge. Those results were calculated using the Tight Binding method for obtaining the band structure of the device. The comparison with these results validates the use of the $\boldsymbol{k}\cdot\boldsymbol{p}$ method to describe the electrostatic properties of Si and Ge NWs.

For the chosen parameters, however, there is a noticeable difference between these results and those found in [99], which were also calculated using the TB method. A better fit with them can be obtained by choosing a larger acoustic deformation constant for Si, e.g., $D_{ac} = 9.2\,\text{eV}$ [109], as it will be shown later.

We have depicted in Fig. 7.5 the hole mobility as a function of the Si NW diameter. As can be seen, the performance of the [111]-oriented NWs is much higher than for the other orientations. This is mainly due to their lower effective mass, as it was already demonstrated by other authors [99, 102], and also corresponds to our observations in Sec. 2.3. Moreover, the increase on the mobility for the [011] and [111]-oriented devices with the reduction of the diameter is also related to the reduction of their corresponding effective mass, which was depicted in Fig. 6.8.

7.6.1 Form Factors. Overlap integral calculation

The accurate calculation of the overlap integrals $F_{i,j}$ in Eq. (7.31) may be troublesome because of the high computational cost of operating with all the envelope functions. Therefore, some approximations are usually made. The simplest one is to consider

	Literature		Our results	
Orientation	Si (Refs. [102],[99])	Ge (Ref. [101])	Si	Ge
[001]	245, 75	1175	300	1297
[011]	435, 145	2030	373	2059
[111]	655, 275	2790	604	2901

Table 7.2: Comparison of the mobilities (in $\text{cm}^2\,\text{V}^{-1}\,\text{s}^{-1}$) for pure Si and Ge with the results of other authors. The simulations are performed for cylindrical, $10\,\text{nm}$ diameter, NWs.
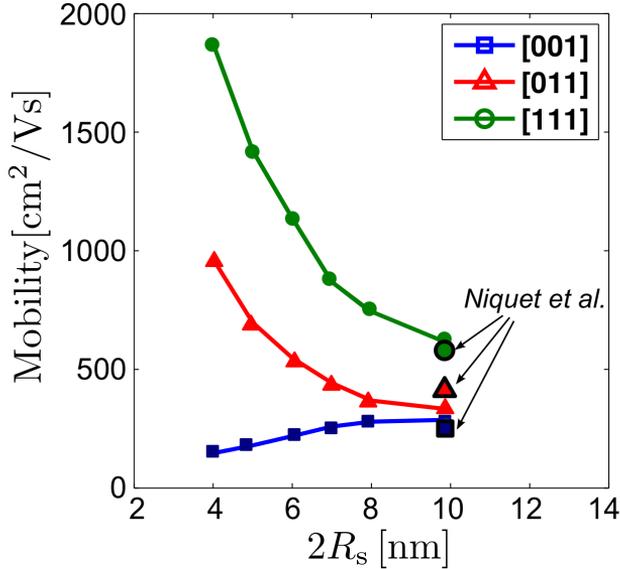
Figure 7.5: Mobility for Si NWs of different diameters calculated with our simulator (solid lines with solid symbols). The results provided for $R_s = 10$ nm in Ref. [101] are also included as larger filled symbols. The three standard orientations [001],[011], and [111] are represented as squares, triangles, and circles.

this overlap integral as a constant, depending on whether the transition between states is considered intersubband ($F_{i,j} = 1/\mathcal{S}$) or intrasubband ($F_{i,j} = 9/4\mathcal{S}$) (e.g. in Ref. [90]). Another approach is to use the envelope function calculated at $k_z = 0$ for the whole subband, i.e, $F_{i,j} = F_{n,n'}(0,0)$ [40]. A priori, there are no physical arguments supporting the validity of any of these approximations.

To assess the impact of these approaches we have performed simulations considering the coupling between the initial and final states as they are provided by the band structure calculation, and compared these results with those obtained from the two aforementioned approaches.

Figure 7.6 shows the overlap integrals calculated for a $Si_{0.5}Ge_{0.5}$ NW of 10 nm diameter at a low hole density. Two orientations are considered, [001] on the left column and [111] on the right column. The first (second) row shows the overlap integral of the first (second) subband with the first four subbands as a function of k'_z. The intrasubband overlap integrals are represented with a line thicker than those corresponding to intersubband ones. The two approximations are also depicted: dashed lines for the constant approximation, and circles highlighting the value $F_{n,n'}(0,0)$ used in the second
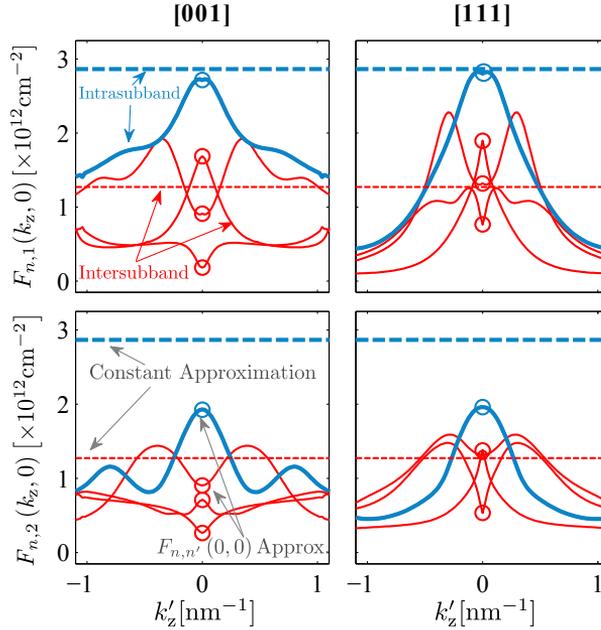
Figure 7.6: Form factors calculated for a $Si_{0.5}Ge_{0.5}$ NW of 10 nm diameter at low inversion, for [001] (left) and [111] (right) device orientation. The first (second) row shows the overlap integral of the first (second) subband with the first four subbands as a function of k'_z. The dashed lines represent the constant approximation, while the circles highlight the value $F_{n,n'}(0,0)$.

approximation.

As can be seen in that Figure, the evolution of the overlap integral with k_z can not be considered as a constant. The weight of the intrasubband transition decays quickly for values away from $k_z = 0$. This effect is neglected in both approximations due to the lack of knowledge about the evolution of $F_{i,j}$ with k_z. In general, for the cases represented, the form factor in intrasubbands transitions is overestimated, whereas the intersubband transitions are in some cases overestimated and underestimated in others. Anyhow, resulting from this misrepresentation of the overlap integrals as a function of k_z, there is an unpredictable weighting of the transitions. Obviously, the so-called constant approximation is more incorrect since its estimation of the intrasubband and intersubband overlap integrals does not match the actual value at any point.

Let us now remark that for the representation in Fig. 7.6, the subbands evolution in k_z has been carefully chosen so as to reduce errors induced by a misrepresentation of the subband structure, i.e., the subband crossing points have been calculated using
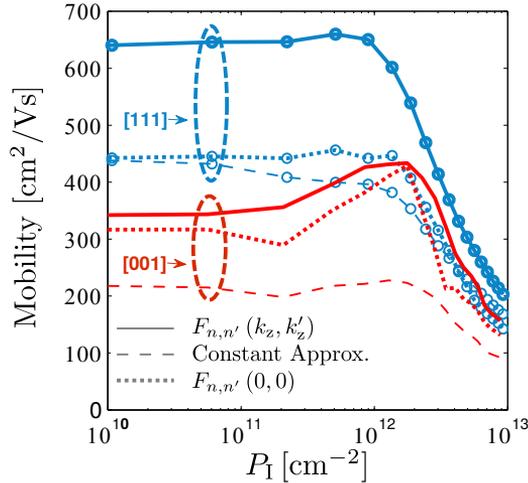
Figure 7.7: Mobility calculated for a $Si_{0.5}Ge_{0.5}$ NW using three models: considering all the form factors (solid line), approximating by constants the intersubband and intrasubband terms (dashed line) and considering that the wave function remains constant (dotted line) to the value obtained for $k_z = 0$. Circles are used to distinguish the [111] orientation from the [001] one.

the procedure in Sec. 6.1. If these cautions are not taken, the uncertainty is even greater, since the choice of the intrasubband or intersubband overlap integral can be erroneous if a crossing point is not detected properly. Moreover, the treatment of the intrinsic two-fold degeneracy of the subbands due to the spin, makes the problem also cumbersome. It is evident that the transition from a subband to its degenerate is an intersubband transition with small probability, and thus the overlap integral should be negligible. The problem is that the wave functions are not unequivocally determined for an state and its degenerate, since all their linear combinations are also solutions of the problem. Therefore, when one tries to track the evolution of a subband, the wave function provided by the $\boldsymbol{k}\cdot\boldsymbol{p}$ method for a k_z value may differ drastically to the wave function of the same subband at $k_z + \Delta k_z$, as it is the result of a different mixing of the degenerate wave functions. These issues do not have any influence in the case where all the overlap integrals are calculated, since each transition is evaluated individually. However, in both of the two approximations presented before, we had to account for them carefully, in order to reduce the uncertainty in the mobility results.

Then, in order to address the effect of the different approximations on the total mobility, we proceed to calculate the hole mobility in a $Si_{0.5}Ge_{0.5}$ NW with 10 nm di-

ameter making use of the three aforementioned models. Figure 7.7 shows the total mobilities calculated for the different approaches, and we can assert that none of the approximations fit the values of the mobility obtained considering the detailed calculation of the form factors. Whereas the $F_{n,n'}(0,0)$ approximation seems to follow the curve for [001] NWs, it clearly fails in the [111] orientation. The constant approach (dashed line) severely underestimates the hole mobility for both orientations. This uncertainty of the mobility results stem from the overestimation of the matrix element of the intrasubband and the intersubband transitions, which depends on the device orientation and the wave vector. Hence, these crude approximations limit the capability of the simulator to predict the performance of the device under study.

We can assert that, although these approximations provide a much more simple and efficient calculation of the mobility, they are not based on physical arguments and therefore preclude the use of these simulators to predict the device performance. The uncertainty in the results might even compromise the benefits achieved by using accurate models for the band structure calculation. Therefore, these approximations must be avoided as far as possible to achieve reliable results for the mobility.

7.6.2   Phonon-limited mobility

Once we have analyzed the non negligible effects of the different approaches used to calculate the form factors, let us now take a deeper insight on the phonon-limited mobility of Si and Ge NWs, where we have used the accurate procedure to guarantee the best results.

As shown in Fig. 7.7, the mobility is roughly constant at low inversion charges and quickly decays for high densities. However, according to Fig. 7.4, the mean effective mass for high inversion seems to be dominated by the light holes, and a higher mobility could be expected. Contrarily, the behavior of the mobility is just the opposite. To explain this situation, a thorough study on the different components of the mobility should be carried out.

Figure 7.8 illustrates the total mobility and the contribution of the mobilities limited by the acoustic and optical phonons in Si and Ge NWs for the [111] orientation. It can be observed that for high inversion both types of phonon-limited mobilities drop quickly. However, the rate of decay is mainly determined by the optical phonons.As a consequence, the Ge NWs undergo a high degradation of the hole mobility for high inversion charges, which might counteract the benefits of a lower effective mass.
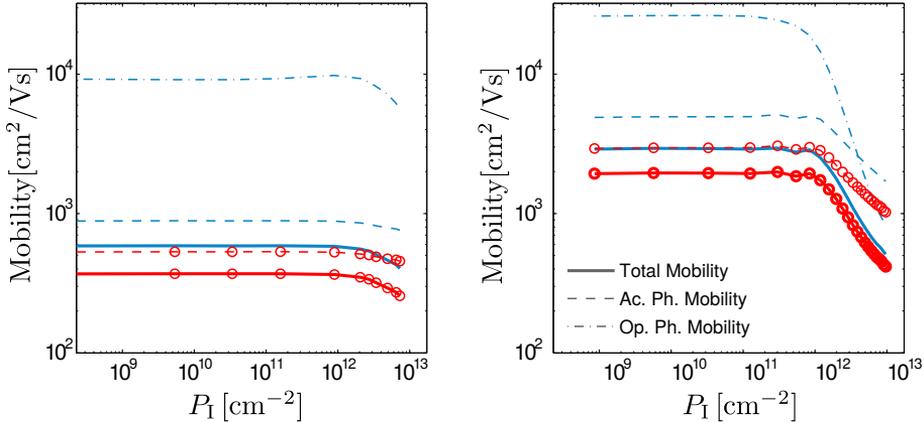
Figure 7.8: Mobility broken down in the contribution of acoustic phonons (dashed lines) and optical phonons (dash dot) for Si (left plot) and Ge (right plot) 10nm NWs along the [111] orientation. Lines marked with circles stand for the acoustic phonon limited mobilities and the corresponding total mobilities calculated for a higher D_{ac} (9.2eV for Si and 11eV for Ge as in Ref. [109]).

Nonetheless, usually a larger value of the acoustic deformation potential is needed to fit the experimental mobility data in highly confined systems, and then the previous conclusions should be revised. This fact could reduce the relative importance of optical phonon scattering with respect to other mechanisms. To check it, we have represented in the same Figure (marked with circles) the mobility calculated using the Si and Ge acoustic deformation potentials employed in Ref. [109]. As can be seen, the mobility is reduced as expected due to the higher acoustic phonon scattering rate. However, the influence of the optical phonons is still quite relevant, and the total mobility is degraded a general degradation but still exhibit the same behavior. Consequently, a noticeable degradation of the mobility for high inversion densities can be expected for Ge, such that its value drops more than in Si.

7.6.3 Alloy disorder (AD) influence

Let us focus now on the influence of the AD on the hole mobility of $Si_{1-x}Ge_x$ NWs. As already mentioned, the alloying of Si with Ge is expected to improve the performance with respect to pure Si NWs, since Ge exhibits a lower effective mass and thus a larger mobility. However, when alloying, there is an additional scattering mechanism which is not present in the pure materials, the AD. In this section we study the influence of
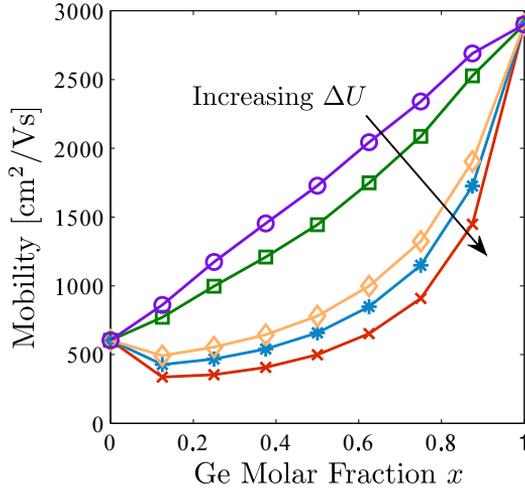
Figure 7.9: Total hole mobility in 10 nm cylindrical SiGe NWs at low hole inversion, [111] orientation, as a function of the Ge molar fraction. Five values of ΔU are used: 0 eV ($\circ$), 0.2 eV [109] ($\square$), 0.513 eV [132]($\diamond$), 0.6 eV [42]($*$), and 0.75 eV [91] ($\times$).

this mechanism on the total mobility.

Figure 7.9 depicts the dependence of the hole mobility on the Ge molar fraction, for [111]-oriented SiGe NWs and low inversion charge. The following values of ΔU were considered: 0 eV (circles), 0.2 eV [109] (squares), 0.513 eV [132] (diamonds), 0.6 eV [42] (asterisks), and 0.75 eV [91] (crosses). In this figure we can see that ΔU has a strong influence on the total mobility behavior. For the two smallest values, the alloy disorder is lightweight compared to the phonon scattering, and, therefore, the mobility shows a monotonic (quasi linear) increase. This scenario does not seem very realistic since it has been widely studied and proven that the alloy scattering affects strongly the mobility in SiGe [24, 42, 48]. Nevertheless, for the rest of ΔU values, we observe a behavior more similar to that measured in bulk SiGe, where the alloy scattering dominates in the balanced alloys, counteracting the hypothetical improvement of the mobility produced by the reduction of the hole effective mass. In fact, for low Ge concentration alloys there is a decay on the mobility compared to the pure Si as it happens in bulk and inversion layers. Thus, high Ge molar fractions, at least larger than 0.6, are needed to achieve an actual improvement on the mobility.

Next, we will use $\Delta U = 0.6$eV and study the behavior at low and high inversion charge for the three channel orientations (see Fig. 7.10). As can be seen, at high inversion charge the mobility undergoes a strong degradation compared to the results
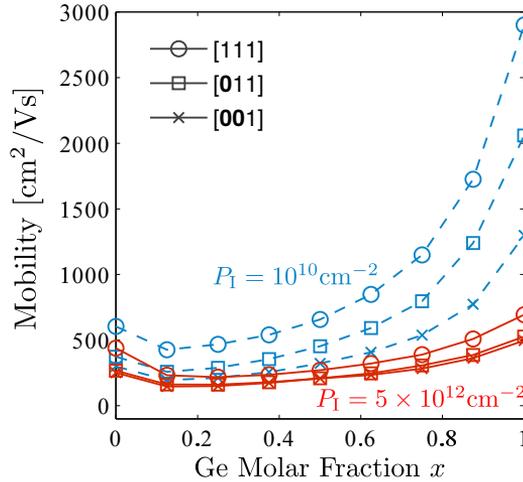
Figure 7.10: Total mobility in $10\,\text{nm}$ cylindrical SiGe NWs of different orientations for a fixed $\Delta U = 0.6\text{eV}$ in high inversion $p = 5 \times 10^{12}\text{cm}^{-2}$ (solid line) and low inversion $p = 10^{9}\text{cm}^{-2}$ (dashed line).

achieved at low inversion (dashed lines), particularly for rich Ge alloys. The origin of this degradation has to be traced on the higher influence of optical phonon scattering, as discussed in the previous Section.

This reduction of the hole mobility for high inversion reduces severely the advantage of using SiGe alloys or even pure Ge NWs. In this scenario, the actual potential of mixing both elements lies on the possibility of adding strain, reducing the hole effective mass and the scattering due to intrasubband transitions. This technique has been proven experimentally to enhance the NW performance [146]. An in-depth study of this topic is planned as a continuation of this work.

7.7 Conclusions

In this Chapter, we have introduced the Boltzmann Transport Equation as the mathematical tool necessary to describe the transport of carriers in NWs. So as to solve the BTE, the momentum relaxation time approximation is employed. This approximation simplifies the problem to the calculation of the scattering rates, and is intrinsically related to the distribution of probabilities for each state in the system. Then, the Kubo-Greenwood is used to calculate the mobility in the device from the MRT.

Finally, a specific case has been considered: the study of the hole mobility in SiGe NWs limited by phonons and alloy disorder. It was demonstrated that the influence of the AD is dominant at low inversion charge, whereas for high inversion the optical phonon contribution degrades considerably the hole mobility for any Ge molar fraction, reducing the beneficial effect of the alloy.

We have also given an insight in the calculation of the overlap integrals, needed to evaluate the scattering rates. Two approaches, that neglect the dependency of the wave function with k_z to reduce the computational cost, have been compared with our full band procedure, depicting a high dispersion of the calculated hole mobility values.

Therefore, we conclude that an accurate estimation of the hole mobility in SiGe NWs can be affected by different sources of uncertainty that can be reduced carrying out accurate calculations as those shown in this manuscript.

Part IV

Conclusions and future work

Chapter 8

Conclusion and future work

8.1 Conclusion

The main objective of this PhD Thesis has been the study of the electrostatic and transport properties of MOSFET devices based in nanowires. The research has been focused on holes as the charge carriers, as they have been usually circumvented due to their inherent difficulties.

In this context, the main contributions of this work are listed next:

1. Implementation of the $\boldsymbol{k} \cdot \boldsymbol{p}$ method for different scenarios: two band, six band and eight band. They have been applied to estimate the conduction and the valence bands in diamond semiconductors (Si, Ge, and SiGe) and both of them simultaneously for zinc blende semiconductors (III-V compounds). A spurious free solution for the eight band $\boldsymbol{k} \cdot \boldsymbol{p}$ method has been achieved by modifying the parameters to keep the ellipticity of the system.

2. Development of an electrostatic simulator which implements the self-consistent solution of the Poisson and Schrödinger equations in the cross-section of the NWs. The Schrödinger equation is solved by means of the $\boldsymbol{k} \cdot \boldsymbol{p}$ method restricted to two dimensional confined systems using the envelope function approximation. The simulator is preferentially designed to achieve a satisfactory description of the holes.

3. Comparison of the $\boldsymbol{k} \cdot \boldsymbol{p}$ results obtained for NWs with Tight-Binding simulations. The $\boldsymbol{k} \cdot \boldsymbol{p}$ method suffers the limitation of the lack of an accurate description of

satellite valleys. This drawback can be overcome by considering the effective mass approach for these satellite valleys. In general, a good agreement has been achieved between both descriptions although slight differences are found. This fact was expected as there is a large number of parameters which are empirically estimated. As a consequence, differences are justified and can be reduced by an appropriate tunning of the parameters.

4. Study of the electrostatic results provided for NWs made of Si, Ge, and III-V semiconductors from 3nm to 20nm. Different orientations, geometries and materials were considered. The evolution of the band structure as a function of the device size in Si and Ge NWs was shown. As the diameter is reduced, the LH subbands split from the HH ones. For the smallest sizes the LH subbands predominate with respect the HH, and, as a consequence, a reduction of the mean effective mass is expected. This effect is more accentuated in Si devices than in Ge ones, and similar mean effective masses can be achieved for the smallest devices, specially along the [111] orientation.

5. Assessment of the self consistent solution and also of the dependence of the band structure on the gate voltage. It was demonstrated that the mean effective mass increases at high inversion since the HH states gets more populated. Regarding the hole density, a strong anisotropy has been observed for [011] oriented NWs, whereas, for the [001] and [011] orientations, the anisotropy is much smaller.

6. Analysis of the hole linear density and the gate capacitance for Si and Ge NWs. A higher capacitance for the Si NWs has been reported. A study on $Si_{1-x}Ge_x$ NWs has depicted the progressive variation of the electrostatic properties as a function of the Ge molar fraction.

7. Solution for the subband crossing issue based on the group velocity, which has been calculated from the Hamiltonian rather than from the numerical derivative.

8. Calculation of the carrier mobility for a 1D gas based on the linearization of the Boltzmann Transport Equation (BTE) using the momentum relaxation time approximation (MRT). Then, the mobility is calculated using the Kubo-Greenwood formula. Bulk phonons (acoustic and optical) and alloy disorder scattering mechanisms have been implemented.

9. Calculation of the hole mobility in Si, Ge and $Si_{1-x}Ge_x$ nanowires. Ge NWs

exhibit better mobility than Si in the low inversion regime. However, for high inversion this difference is severely mitigated. In $Si_{1-x}Ge_x$ devices, it was demonstrated that the influence of the AD is dominant at low inversion, whereas for high inversion the optical phonon contribution degrades considerably the hole mobility for any Ge molar fraction, reducing the beneficial effect of the alloy.

10. Assessment of the impact of using two common approaches for the calculation of the overlap integrals that neglect the dependency of the wave function with k. The two approximations depict a high dispersion of the hole mobility values compared with our full band procedure.

8.2 Future Work

The simulation tools implemented in this Thesis exhibit a great potential for future works and our aim is to exploit it with further studies on the matter. There is still a lot of work to do concerning NWs and we think these simulators are letting us continue with the research in that field.

The future work planning is described on the following lines:

1. The study of the carrier mobility will be completed with the inclusion of other scattering mechanisms such as: surface roughness, coulomb dispersion, and polar optical phonons. Their influence on the carrier mobility as a function of the geometry, size and applied bias is of major relevance in order to determine the feasibility of these devices for future technological nodes.

2. We will perform further investigations concerning the use of a full band description, where all the wave functions corresponding to different subbands and wave vectors are employed to achieve an accurate calculation of the electrostatic and the transport problem. Since the storage of the wave functions and the operation with them are the more limiting issues in the simulator, it is quite important to assess the limits of different approaches commonly used in the literature to reduce the computational burden.

3. A thorough study of the effect of strain on the device performance will be addressed. This task should be linked with a realistic distribution of strain along

the channel of the device as it has been demonstrated that strain depicts a non-uniform distribution as the dimensions are shrunk to the nanometer scale.

4. Core-shell structures will be studied using the $\boldsymbol{k}\cdot\boldsymbol{p}$ simulator. These devices are NWs composed by different layers of semiconductor materials and are promising to provide higher currents as they reduce the surface roughness and produce strained channels [11].

5. We will develop physically based compact models which fit the simulation data. These simulators would be useful to be included in circuit simulators.

Part V

Appendixes

Appendix A

Perturbation theory

The perturbation theory separates the total Hamiltonian $\hat{H}$ into an unperturbed one $\hat{H}_0$, the solutions of which are known, and a perturbed one $\hat{H}_1$. The total solutions for $\hat{H}$, labeled as $u_{mk}(\boldsymbol{r})$, are sought in terms of the solutions of $\hat{H}_0$, $u_{m0}(\boldsymbol{r})$, which naturally form an orthonormal basis of functions, if their eigenvalues are not degenerate[1]:

$$u_{n\boldsymbol{k}}(\boldsymbol{r}) = \sum_m c_{m,n}(\boldsymbol{k}) u_{m0}(\boldsymbol{r}) \tag{A.1}$$

So as to seek the solutions of the system, a parameter λ is included to decouple the solutions in different orders[2]:

$$\left(\hat{H}_0 + \lambda\hat{H}_1\right) u_{n\boldsymbol{k}} = W_{n\boldsymbol{k}} u_{n\boldsymbol{k}} \tag{A.2}$$

For this system, the solutions in a polynomial form are sought:

$$u_{n\boldsymbol{k}} = u_{n\boldsymbol{k}}^{(0)} + \lambda u_{n\boldsymbol{k}}^{(1)} + \lambda^2 u_{n\boldsymbol{k}}^{(2)} + \lambda^3 u_{n\boldsymbol{k}}^{(3)} + \dots \tag{A.3}$$

$$W_{n\boldsymbol{k}} = W_{n\boldsymbol{k}}^{(0)} + \lambda W_{n\boldsymbol{k}}^{(1)} + \lambda^2 W_{n\boldsymbol{k}}^{(2)} + \lambda^3 W_{n\boldsymbol{k}}^{(3)} + \dots \tag{A.4}$$

Introducing Eqs. (A.3) and (A.4) into (A.2) and rearranging terms, it reads:

---

[1] In the case that there are degenerate eigenvalues, still a set of orthonormal function can be found among the solutions

[2] Hereinafter, we drop the explicit dependency in $\boldsymbol{r}$ for the sake of a simpler notation.

$$
\begin{aligned}
&\left(\hat{\mathrm{H}}_0 - W_{n\boldsymbol{k}}^{(0)}\right)u_{n\boldsymbol{k}}^{(0)} && 0 \\
&+\lambda\left(\hat{\mathrm{H}}_0 - W_{n\boldsymbol{k}}^{(0)}\right)u_{n\boldsymbol{k}}^{(1)} && +\lambda\left(W_{n\boldsymbol{k}}^{(1)} - \hat{\mathrm{H}}_1\right)u_{n\boldsymbol{k}}^{(0)} \\
&+\lambda^2\left(\hat{\mathrm{H}}_0 - W_{n\boldsymbol{k}}^{(0)}\right)u_{n\boldsymbol{k}}^{(2)} \;=\; && +\lambda^2\left[\left(W_{n\boldsymbol{k}}^{(1)} - \hat{\mathrm{H}}_1\right)u_{n\boldsymbol{k}}^{(1)} + W_{n\boldsymbol{k}}^{(2)}u_{n\boldsymbol{k}}^{(0)}\right] \\
&+\lambda^3\left(\hat{\mathrm{H}}_0 - W_{n\boldsymbol{k}}^{(0)}\right)u_{n\boldsymbol{k}}^{(3)} && +\lambda^3\left[\left(W_{n\boldsymbol{k}}^{(1)} - \hat{\mathrm{H}}_1\right)u_{n\boldsymbol{k}}^{(2)} + W_{n\boldsymbol{k}}^{(2)}u_{n\boldsymbol{k}}^{(1)} + W_{n\boldsymbol{k}}^{(3)}u_{n\boldsymbol{k}}^{(0)}\right] \\
&\qquad + \ldots && \qquad + \ldots
\end{aligned}
\tag{A.5}
$$

The different values of $u_{n\boldsymbol{k}}^{(x)}$ and $W_{n\boldsymbol{k}}^{(x)}$ can be obtained by equaling the same order terms. The total solution is finally obtained for $\lambda = 1$. The solution to the perturbed system is usually achieved by only considering the first order solutions and neglecting the superior orders. Next we present the first solutions up to second order.

Zeroth order solution It corresponds to the terms with $\lambda = 0$, and the solutions are straightforwardly calculated as:

$$
\left(\hat{\mathrm{H}}_0 - W_{n\boldsymbol{k}}^{(0)}\right)u_{n\boldsymbol{k}}^{(0)} = 0\,,
\tag{A.6}
$$

$$
W_{n\boldsymbol{k}}^{(0)} = E_{n0}\,,
\tag{A.7}
$$

$$
u_{n\boldsymbol{k}}^{(0)} = u_{n0}\,.
\tag{A.8}
$$

These solutions are those of the unperturbed system and do not provide new information.

First order solution The equation for the terms of first order is:

$$
\left(\hat{\mathrm{H}}_0 - W_{n\boldsymbol{k}}^{(0)}\right)u_{n\boldsymbol{k}}^{(1)} = \left(W_{n\boldsymbol{k}}^{(1)} - \hat{\mathrm{H}}_1\right)u_{n\boldsymbol{k}}^{(0)}
\tag{A.9}
$$

As aforementioned, the solutions, for any i-th order solution, are sought as coordinates in a basis formed by the unperturbed solutions:

$$
u_{n\boldsymbol{k}}^{(x)} = \sum_m c_{m,n}^{(x)} u_{m0}\,,
\tag{A.10}
$$

Thus, the coefficients $c_{m,n}^{(x)}$, which are the projection of the solution $u_{n\boldsymbol{k}}^{(x)}$ in the basis function u_{m0}, are needed:

$$c_{m,n}^{(x)} = \langle u_{m0}|u_{n\boldsymbol{k}}^{(x)}\rangle \ . \tag{A.11}$$

These coefficients can be calculated by projecting Eq.(A.9) in each basis function:

$$\langle u_{n0}|\hat{\mathrm{H}}_0|u_{n\boldsymbol{k}}^{(1)}\rangle - E_{n0}\,\langle u_{n0}|u_{n\boldsymbol{k}}^{(1)}\rangle = W_{n\boldsymbol{k}}^{(1)} - \langle u_{n0}|H_1|u_{n0}\rangle \ , \tag{A.12}$$

where $u_{n\boldsymbol{k}}^{(1)}$ can be expanded as in Eq. (A.10), and we can use the relationship $\langle u_{m0}|\hat{\mathrm{H}}_0|u_{n0}\rangle = E_{n0}\delta_{mn}$.

Then, for $m = n$, the first order energy term can be obtained as:

$$W_{n\boldsymbol{k}}^{(1)} = \langle u_{n0}|\hat{\mathrm{H}}_1|u_{n0}\rangle \ . \tag{A.13}$$

This term might vanish for some perturbations, such as for the $\boldsymbol{k}\cdot\boldsymbol{p}$ one, since $\hat{\mathrm{H}}_1$ is related to the momentum, which applied in Eq. (A.13) yields 0.

The coefficients $c_{m,n}^{(1)}$, for $m \neq n$, are obtained as:

$$c_{m,n}^{(1)} = \frac{\langle u_{m0}|\hat{\mathrm{H}}_1|u_{n0}\rangle}{E_{m0} - E_{n0}} \ . \tag{A.14}$$

Consequently, the wave function approximation of first order reads:

$$u_{n\boldsymbol{k}}^{(1)} = \sum_{m\neq n} \frac{\langle u_{m0}|\hat{\mathrm{H}}_1|u_{n0}\rangle}{E_{m0} - E_{n0}} u_{m0} \ . \tag{A.15}$$

Second order solution When the first order solution for the energy either vanishes or is not accurate enough to describe the total solution, the second order solution is needed:

$$\left(\hat{\mathrm{H}}_0 - W_{n\boldsymbol{k}}^{(0)}\right)u_{n\boldsymbol{k}}^{(2)} = \left(W_{n\boldsymbol{k}}^{(1)} - \hat{\mathrm{H}}_1\right)u_{n\boldsymbol{k}}^{(1)} + W_{n\boldsymbol{k}}^{(2)}u_{n\boldsymbol{k}}^{(0)} \ . \tag{A.16}$$

Working as in the first order solution, Eq. (A.16) is projected on the basis functions. Now, $u_{n\boldsymbol{k}}^{(2)}$ is the function expanded as in Eq. (A.10), and $u_{n\boldsymbol{k}}^{(1)}$ is substituted using Eq. (A.15).

Then, for $m = n$, the second order energy term reads:

$$W_{nk}^{(2)} = \sum_{m \neq n} \frac{\left| \langle u_{n0} | \hat{H}_1 | u_{m0} \rangle \right|^2}{E_{m0} - E_{n0}} \tag{A.17}$$

Although the second order functions can also be obtained for $m \neq n$, their complexity makes them not useful in most of the cases.

Therefore, the energies of the perturbed system can be approximated as:

$$W_{nk} \approx W_{nk}^{(0)} + W_{nk}^{(1)} + W_{nk}^{(2)} = E_{n0} + \langle u_{n0} | \hat{H}_1 | u_{n0} \rangle + \sum_{m \neq n} \frac{\left| \langle u_{n0} | \hat{H}_1 | u_{m0} \rangle \right|^2}{E_{m0} - E_{n0}}. \tag{A.18}$$

As can be seen, the evolution of the non perturbed energy level $n0$ into nk depends on the coupling of the initial state $n0$ with the rest of the states $m0$ by means of the perturbation hamiltonian $\hat{H}_1$. However, not all the states contribute equally to the final solution, being those states with closer energy more relevant due to the term $E_{m0} - E_{n0}$ in the denominator of the last term of Eq. (A.18).

A.1   Perturbation theory for degenerate states

If there are states with the same energy but different wave function forms (degenerate states), their contribution to Eq. (A.17) would be undetermined. A priori, the eigenfunctions associated to two degenerate states might not be orthonormal, since a subspace of solutions is formed with all their linear combinations. Therefore, the actual unperturbed function associated to a degenerate state which might evolve to non degenerate states, is unknown. However, a new set of orthonormal eigenfunctions $|n\rangle'$ can be found which fulfills this condition. These eigenfunctions are a linear combination of the N eigenfunctions $|1\rangle$, $|2\rangle$,..$|N\rangle$, which share the same eigenstate $E_1 = E_2 = \cdots = E_N$, as in Eq. (A.19).

Consequently, the basis functions can be divided in two classes: one class consists of the N degenerate states, and the second class, consists of the rest of the remote states.

Zeroth order solution   The degenerate states belonging to the new set of eigenfunctions, can be expressed as a sum of the old ones:

$$|n\rangle' = \sum_{m=1}^{N} |m\rangle \langle m|n\rangle' \tag{A.19}$$

The remote states remain unchanged and also do the energies.

First order solution   Using a similar reasoning as for the non degenerate perturbation theory, considering the evolution of a degenerate state, the first order equation reads:

$$\left(\hat{H}_0 - W_{n\boldsymbol{k}}^{(0)}\right) \sum_{m=1}^{N} |m\rangle \langle m|n\rangle' = \left(W_{n\boldsymbol{k}}^{(1)} - \hat{H}_1\right) \sum_{m=1}^{N} |m\rangle \langle m|n\rangle' \tag{A.20}$$

Projecting that equation in a degenerate function $|j\rangle$:

$$W_{n\boldsymbol{k}}^{(1)} \langle j|n\rangle' - \sum_{m=1}^{N} \langle j|\hat{H}_1|m\rangle \langle m|n\rangle' = 0 \qquad \text{for } |j\rangle \text{ in } N \tag{A.21}$$

$$\langle j|n\rangle' = \sum_{m=1}^{N} \frac{\langle j|\hat{H}_1|m\rangle}{E_1 - E_j} \langle m|n\rangle' \qquad \text{for } j \notin N \tag{A.22}$$

Now, the first order energy $W_{n\boldsymbol{k}}^{(1)}$ is computed as a N-coupled equation system. The calculation for the rest of non degenerate states is not affected, and is performed as in the previous section.

For the $\boldsymbol{k}\cdot\boldsymbol{p}$ perturbation, the first order energy also vanishes, and therefore the second order perturbation must be calculated.

Second order solution   When projecting a $|j\rangle$ function on the second order equation, Eq. (A.16), the second order energy and the coefficients can be calculated as an N-coupled equation system, which must be solved for the different eigenvalues $W_{n\boldsymbol{k}}^{(2)}$ and for the coefficients of $|n\rangle'$, which are the eigenvectors of the system:

$$\sum_{l=1}^{N} A_{ln'} \langle l|n\rangle' - W_{n\bm{k}}^{(2)} \langle j|n\rangle' = 0 \qquad \text{para} \quad j \in N \,, \tag{A.23}$$

where the matrix elements $A_{ln'}$ are calculated as the coupling of all the remote states by means of the perturbation:

$$A_{ln'} = \sum_{\alpha \notin N} \frac{\langle j|\hat{H}_1|\alpha\rangle \, \langle \alpha|\hat{H}_1|l\rangle}{E_1 - E_\alpha} \tag{A.24}$$

Appendix B

Löwdin perturbation theory

The Löwdin perturbation theory was postulated in [78]. This theory aims at expressing the eigenfunctions of a Hamiltonian as the coefficients in a basis functions of a reduced subspace A, and another subspace B. The main idea is that the states of the class A are more relevant than those in class B, which can be treated as remote states that contribute to the total solution as corrections.

Lets consider a generic Hamiltonian:

$$\hat{H}\phi = E\phi \,, \tag{B.1}$$

whose eigenfunctions ϕ can be expressed in a basis ϕ_{m0} which forms a space of functions S. This space can be split into the subspaces A, the reduced subspace, and B, the remote states subspace. Then the solutions are:

$$\phi = \sum_{m}^{A,B} c_m \phi_{m0} \,. \tag{B.2}$$

Substituting Eq. (B.2) into Eq. (B.1) and projecting in a state ϕ_{n0}:

$$\sum_{n}^{A,B} \hat{H}_{mn} c_n = E\, c_m \,, \tag{B.3}$$

$$\sum_{n\neq m}^{A,B} \hat{H}_{mn} c_n = (E - H_{mm})\, c_m \,, \tag{B.4}$$

where the Hamiltonian matrix elements are

$$H_{mn} = \langle \phi_{m0} | \hat{H} | \phi_{n0} \rangle \ . \tag{B.5}$$

Therefore, according to Eq. (B.4), a coefficient c_m depends on the rest as:

$$c_m = \sum_{\alpha, m \neq \alpha}^{A} \frac{H_{m\alpha}}{E - H_{mm}} c_\alpha + \sum_{\beta, m \neq \beta}^{B} \frac{H_{m\beta}}{E - H_{mm}} c_\beta \tag{B.6}$$

By substituting all the coefficients $\beta \in B$ in the Eq. (B.6), one can express the coefficients c_m with $m \in A$ in terms of only coefficients belonging to A. Thus, the system for the α coefficients is reduced to:

$$\sum_{\alpha}^{A} U_{m\alpha} c_\alpha = E \, c_m \,, \qquad m \in A \tag{B.7}$$

with

$$U_{m\alpha} \equiv H_{m\alpha} + \sum_{\beta, \beta \neq m}^{B} \frac{H_{m\beta} H_{\beta\alpha}}{E - H_{\beta\beta}} + \sum_{\beta, \beta \neq m}^{B} \sum_{\gamma, \gamma \neq m}^{B} \frac{H_{m\beta} H_{\beta\gamma} H_{\gamma\alpha}}{(E - H_{\beta\beta})(E - H_{\gamma\gamma})} + \cdots \qquad m \in A \tag{B.8}$$

As can be seen, the hamiltonian has a term accounting for the direct coupling of the class A states $H_{m\alpha}$, whereas the class B states are considered as remote ones in a similar manner than in the regular perturbation theory.

The rest of coefficients, those corresponding to class B states, can be calculated in terms of the coefficients of class A :

$$c_m = \sum_{\alpha}^{A} \frac{U_{m\alpha}}{E - H_{mm}} c_\alpha \,. \qquad m \in B \tag{B.9}$$

These coefficients, if corresponding to remote states, are expected to be negligible, since they are divided by the difference of energies. Therefore, an approximated solution can be obtained by only considering the coefficients of class A states. They are obtained from the solution of the more reduced system in Eq. (B.7), which is a correction of the original hamiltonian matrix elements.

Unfortunately, the system in Eq. (B.7) cannot be directly solved, since the coeffi-

cients depend in turn on E. Therefore, some strategies must developed to eliminate this dependence [75] to express the coefficients in terms of the known unperturbed energies.

Appendix C

Strain matrix calculation

In this Appendix, the calculation of the strain matrix $\boldsymbol{\epsilon}$ from the stress σ of the lattice [135] is addressed. The stress is the result of applying external forces to the structure which deform the lattice. The two principal types of deformations are uniaxial and biaxial strain.

C.1 Uniaxial strain

The uniaxial stress along an arbitrary direction consists in the application of a pressure of magnitude P along that direction.

Firstly, we adopt a coordinate system (x', y', z') referred to the strain, in which the x' axis is parallel to the stress direction. This coordinate system is related to the crystal reference system (x, y, z) by a rotation U

$$U(\phi, \theta) = \begin{pmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \\ \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \end{pmatrix}, \tag{C.1}$$

where θ and ϕ are the polar and azimuthal angles of the stress direction relative to the crystal reference system. In the primed coordinate system, the stress tensor has only one non-zero component, $\sigma'_{xx} = P$.

When uniaxial stress is applied along one of the directions of main interest [001], [011], and [111], stress tensors can be calculated using the rotation matrix and their

expressions are given by:

$$\boldsymbol{\sigma}_{[001]} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & P \end{pmatrix}, \qquad \boldsymbol{\sigma}_{[011]} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & P/2 & P/2 \\ 0 & P/2 & P/2 \end{pmatrix}, \qquad \text{(C.2)}$$

$$\boldsymbol{\sigma}_{[011]} = \begin{pmatrix} P/3 & P/3 & P/3 \\ P/3 & P/3 & P/3 \\ P/3 & P/3 & P/3 \end{pmatrix}. \qquad \text{(C.3)}$$

C.2 Biaxial strain

The biaxial strain usually appears in electronics as the result of the epitaxial growth of a material with relaxed lattice constant a_0 over a substrate with lattice constant a_0. The strain in the plane of the interface can be determined from the lattice mismatch

$$\epsilon_{||} = \frac{a_s - a_0}{a_0} \,. \qquad \text{(C.4)}$$

The strain tensor for arbitrary substrate orientations can be directly calculated according to [56] and it is given here for the most frequently used substrate orientations:

$$\boldsymbol{\epsilon}_{(001)} = \epsilon_{||} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\dfrac{2C_{12}}{C_{11}} \end{pmatrix}, \qquad \text{(C.5)}$$

$$\boldsymbol{\epsilon}_{(011)} = \epsilon_{||} \begin{pmatrix} \dfrac{2C_{44} - C_{12}}{C_{11} + C_{12} + 2C_{44}} & -\dfrac{C_{11} + 2C_{12}}{C_{11} + C_{12} + 2C_{44}} & 0 \\ -\dfrac{C_{11} + 2C_{12}}{C_{11} + C_{12} + 2C_{44}} & \dfrac{2C_{44} - C_{12}}{C_{11} + C_{12} + 2C_{44}} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \text{(C.6)}$$

$$\boldsymbol{\epsilon}_{(111)} = \epsilon_{||} \begin{pmatrix} \dfrac{4C_{44}}{C_{11} + 2C_{12} + 4C_{44}} & -\dfrac{C_{11} + 2C_{12}}{C_{11} + 2C_{12} + 4C_{44}} & -\dfrac{C_{11} + 2C_{12}}{C_{11} + 2C_{12} + 4C_{44}} \\ -\dfrac{C_{11} + 2C_{12}}{C_{11} + 2C_{12} + 4C_{44}} & \dfrac{4C_{44}}{C_{11} + 2C_{12} + 4C_{44}} & -\dfrac{C_{11} + 2C_{12}}{C_{11} + 2C_{12} + 4C_{44}} \\ -\dfrac{C_{11} + 2C_{12}}{C_{11} + 2C_{12} + 4C_{44}} & -\dfrac{C_{11} + 2C_{12}}{C_{11} + 2C_{12} + 4C_{44}} & \dfrac{4C_{44}}{C_{11} + 2C_{12} + 4C_{44}} \end{pmatrix},$$
$$\text{(C.7)}$$

where the interface plane is denoted in the Miller notation and the C parameters are the elastic stiffness constants.

Appendix D

Change of coordinates for the FEM

Due to the fact that the mesh in the FEM is irregular, the solution of the integrals and derivatives is more complex than in FDM. To deal with this issue, a change of coordinates (x, y) to (ξ, η) simplifies the task. The particularity of the new set of coordinates is that the triangle is standardized, as Fig. D.1 illustrates. The advantage is that the partial derivatives and integrals of the nodal functions defined in that triangle are readily calculated since the nodal functions can be easily estimated in the new coordinates as

$$
\begin{aligned}
N_1(\xi, \eta) &= 1 - \xi - \eta \\
N_2(\xi, \eta) &= \xi \\
N_3(\xi, \eta) &= \eta
\end{aligned}
\tag{D.1}
$$

Using the new nodal functions, we can relate the interpolation of a function $f(x, y)$ with the new coordinate system

$$
f(x, y) \simeq \tilde{f}(x, y) = \sum_{i=1}^{3} f_i N_i(\xi, \eta).
\tag{D.2}
$$

This is the general case for any f function, then we can particularize for $f(x, y) = x$ and $f(x, y) = y$ to achieve the direct relation between the two coordinate systems only
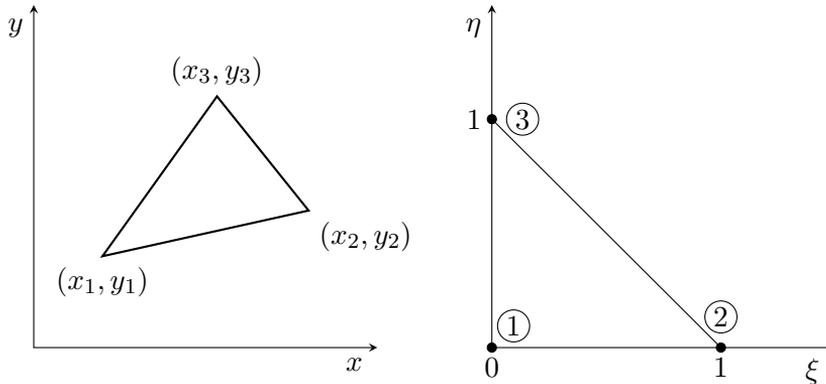
Figure D.1: Mapping of the irregular triangular element to the standard rectangular triangle.

by knowing the coordinates of the triangle vertexes

$$x = x_1 N_1(\xi, \eta) + x_2 N_2(\xi, \eta) + x_3 N_3(\xi, \eta)$$
$$y = y_1 N_1(\xi, \eta) + y_2 N_2(\xi, \eta) + y_3 N_3(\xi, \eta)$$

(D.3)

After substituting and rearranging terms, the change of variables reads:

$$\begin{pmatrix} x - x_1 \\ y - y_1 \end{pmatrix} = \begin{pmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \boldsymbol{J}_\mathrm{F} \begin{pmatrix} \xi \\ \eta \end{pmatrix}$$

(D.4)

Where the right-hand matrix corresponds to the Jacobian matrix $\boldsymbol{J}_\mathrm{F}$, and its determinant (Jacobian) is the transformation rate of the differential element $\mathrm{d}x\,\mathrm{d}y$ and $\mathrm{d}\xi\,\mathrm{d}\eta$, and its value is twice the area of the triangle. We are aware that the Jacobian might be positive or negative, thus the so-defined area could also be positive or negative, depending on the ordering of the vertexes. For the sake of simplicity we convey ordering the vertexes counterclockwise in order to get a positive S.

$$\det(\boldsymbol{J}_\mathrm{F}) = (\boldsymbol{r}_2 - \boldsymbol{r}_1) \times (\boldsymbol{r}_3 - \boldsymbol{r}_1)$$
$$= (x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)$$
$$= x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2) = 2S$$

(D.5)

The relation between the surface differentials is

$$\mathrm{d}x\,\mathrm{d}y = 2|S|\,\mathrm{d}\xi\,\mathrm{d}\eta. \tag{D.6}$$

After some algebra, the partial derivatives of $\tilde{f}$ are[110]:

$$
\begin{aligned}
\frac{\partial}{\partial x} f(x,y) &= \frac{\partial}{\partial x} \sum_{i=1}^{3} f_i N_i(\xi,\eta) = \sum_{i=1}^{3} f_i \frac{\partial}{\partial x} N_i \\
&= \frac{1}{2S} \sum_{i,j} f_i y_j \left(\frac{\partial N_i}{\partial \xi}\frac{\partial N_j}{\partial \eta} - \frac{\partial N_i}{\partial \eta}\frac{\partial N_j}{\partial \xi} \right) \\
&= \frac{1}{2S} \left[(f_2 - f_1)(y_3 - y_1) - (f_3 - f_1)(y_2 - y_1) \right] \\
\frac{\partial}{\partial y} f(x,y) &= \frac{1}{2S} \sum_{i,j} f_i x_j \left(-\frac{\partial N_i}{\partial \xi}\frac{\partial N_j}{\partial \eta} + \frac{\partial N_i}{\partial \eta}\frac{\partial N_j}{\partial \xi} \right) \\
&= \frac{1}{2S} \left[(f_3 - f_1)(x_2 - x_1) - (f_2 - f_1)(x_3 - x_1) \right]
\end{aligned}
\tag{D.7}
$$

From this formulation, we conclude that the partial derivatives can be calculated as a function of the coordinates x and y of the vertexes as:

$$
\begin{array}{lll}
\dfrac{\partial N_1}{\partial x} = \dfrac{y_2 - y_3}{2S} & \dfrac{\partial N_2}{\partial x} = \dfrac{y_3 - y_1}{2S} & \dfrac{\partial N_3}{\partial x} = \dfrac{y_1 - y_2}{2S} \\[2mm]
\dfrac{\partial N_1}{\partial y} = \dfrac{x_3 - x_2}{2S} & \dfrac{\partial N_2}{\partial y} = \dfrac{x_1 - x_3}{2S} & \dfrac{\partial N_3}{\partial y} = \dfrac{x_2 - x_1}{2S}
\end{array}
$$

Furthermore, to solve the integrals of the nodal function products it is necessary to know some integrals defined in the standard triangle:

$$I = \int_0^1 \int_0^{1-\xi} f(\xi,\eta)\,\mathrm{d}\eta\,\mathrm{d}\xi$$

where $f(\xi,\eta)$ is a nodal function or a product of up to three of them, shown in Table D.1. According to [110, §9.3, pg. 227] we can write:

$$I = \int_0^1 \int_0^{1-\xi} \xi^\alpha \eta^\beta (1-\xi-\eta)^\gamma\,\mathrm{d}\eta\,\mathrm{d}\xi = \frac{\alpha!\beta!\gamma!}{(\alpha + \beta + \gamma + 2)!}$$

$f(\xi,\eta)$	I
1	1/2
N_1,N_2,N_3	1/6
N_1^2,N_2^2,N_3^2	1/12
N_1N_2,N_1N_3,N_2N_3	1/24
N_1^3,N_2^3,N_3^3	1/20
$N_1^2N_2,N_1N_2^2,N_1^2N_3,\cdots$	1/60
$N_1N_2N_3$	1/120

Table D.1: Integrals in the standard triangle.

Appendix E

Inner products for the nodal functions

In this Appendix, we calculate the inner products of the nodal functions N_i affected by the Hamiltonian operators, which are needed to calculate the FEM equation system:

$$\int N_i(x,y)\, C\, N_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{E.1a}$$

$$\int N_i(x,y)\, k_{\mathrm{x,y}}\, C\, k_{\mathrm{x,y}}\, N_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{E.1b}$$

$$\int N_i(x,y)\, k_{\mathrm{x,y}}\, C\, k_{\mathrm{z,0}}\, N_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{E.1c}$$

$$\int N_i(x,y)\, k_{\mathrm{z}}\, k_0\, N_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{E.1d}$$

$$\int N_i(x,y)\, V(x,y)\, N_j(x,y)\, \mathrm{d}x\, \mathrm{d}y\,, \tag{E.1e}$$

where $i,j = 1,2$. The expressions show $\pm$ and $\mp$ signs which are related to the sign of S. According to our convention of *positive surfaces* and counterclockwise ordering of vertexes, we take the upper sign.

Firstly, we calculate the products involving the same nodal function $\langle N_1|\hat{\mathcal{O}}|N_1\rangle$, which are used for the construction of the diagonal subblocks[1]:

---

[1]The parameter C, constant in the triangular element, have been dropped for simplicity in the notation.

$$\int N_1 \hat{k}_x \hat{k}_x N_1 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial x} N_1\right)^2 \mathrm{d}x\mathrm{d}y = \left(\frac{y_2 - y_3}{2S}\right)^2 2|S| \int \mathrm{d}\xi\mathrm{d}\eta = \frac{(y_2 - y_3)^2}{4|S|}$$

$$\int N_1 k_y k_y N_1 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial y} N_1\right)^2 \mathrm{d}x\mathrm{d}y = \left(\frac{x_3 - x_2}{2S}\right)^2 2|S| \int \mathrm{d}\xi\mathrm{d}\eta = \frac{(x_3 - x_2)^2}{4|S|}$$

$$\int N_1 k_z k_z N_1 \mathrm{d}x\mathrm{d}y = k_z^2 \int N_1^2 \mathrm{d}x\mathrm{d}y = k_z^2 2|S| \int N_1^2 \mathrm{d}\xi\mathrm{d}\eta = \frac{k_z^2 |S|}{6}$$

$$\int N_1 \hat{k}_x k_y N_1 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial x} N_1\right)\left(\frac{\partial}{\partial y} N_1\right) \mathrm{d}x\mathrm{d}y = \frac{y_2 - y_3}{2S} \frac{x_3 - x_2}{2S} 2|S| \int \mathrm{d}\xi\mathrm{d}\eta$$
$$= \frac{(y_2 - y_3)(x_3 - x_2)}{4|S|}$$

$$\int N_1 \hat{k}_x k_z N_1 \mathrm{d}x\mathrm{d}y = \mathrm{i}k_z \int \left(\frac{\partial}{\partial x} N_1\right) N_1 \mathrm{d}x\mathrm{d}y = \mathrm{i}k_z \frac{y_2 - y_3}{2S} 2|S| \int N_1 \mathrm{d}\xi\mathrm{d}\eta$$
$$= \pm \mathrm{i}k_z \frac{y_2 - y_3}{6}$$

$$\int N_1 k_y \hat{k}_x N_1 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial y} N_1\right)\left(\frac{\partial}{\partial x} N_1\right) \mathrm{d}x\mathrm{d}y = \frac{x_3 - x_2}{2S} \frac{y_2 - y_3}{2S} 2|S| \int \mathrm{d}\xi\mathrm{d}\eta$$
$$= \frac{(y_2 - y_3)(x_3 - x_2)}{4|S|}$$

$$\int N_1 k_y k_z N_1 \mathrm{d}x\mathrm{d}y = \mathrm{i}k_z \int \left(\frac{\partial}{\partial y} N_1\right) N_1 \mathrm{d}x\mathrm{d}y = \mathrm{i}k_z \frac{x_3 - x_2}{2S} 2|S| \int N_1 \mathrm{d}\xi\mathrm{d}\eta$$
$$= \pm \mathrm{i}k_z \frac{x_3 - x_2}{6}$$

$$\int N_1 k_z \hat{k}_x N_1 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \int N_1 \left(\frac{\partial}{\partial x} N_1\right) \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \frac{y_2 - y_3}{2S} 2|S| \int N_1 \mathrm{d}\xi\mathrm{d}\eta$$
$$= \pm \mathrm{i}k_z \frac{y_3 - y_2}{6}$$

$$\int N_1 k_z k_y N_1 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \int N_1 \left(\frac{\partial}{\partial y} N_1\right) \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \frac{x_3 - x_2}{2S} 2|S| \int N_1 \mathrm{d}\xi\mathrm{d}\eta$$
$$= \pm \mathrm{i}k_z \frac{x_2 - x_3}{6}$$

$$\int N_1 k_0 \hat{k}_x N_1 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \int N_1 \left(\frac{\partial}{\partial x} N_1\right) \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \frac{y_2 - y_3}{2S} 2|S| \int N_1 \mathrm{d}\xi\mathrm{d}\eta$$
$$= \pm \mathrm{i}k_0 \frac{y_3 - y_2}{6}$$

$$\int N_1 k_0 k_y N_1 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \int N_1 \left(\frac{\partial}{\partial y} N_1\right) \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \frac{x_3 - x_2}{2S} 2|S| \int N_1 \mathrm{d}\xi\mathrm{d}\eta$$
$$= \pm \mathrm{i}k_0 \frac{x_2 - x_3}{6}$$

$$\int N_1 k_0 k_\mathrm{z} N_1 \mathrm{d}x\mathrm{d}y = k_\mathrm{z}\,k_0 \int N_1^2 \mathrm{d}x\mathrm{d}y = k_0\,k_\mathrm{z}2|S|\int N_1^2 \mathrm{d}\xi\mathrm{d}\eta = \frac{k_0\,k_\mathrm{z}|S|}{6}$$

$$\int N_1 N_1 \mathrm{d}x\mathrm{d}y = 2|S|\int N_1^2 \mathrm{d}\xi\mathrm{d}\eta = \frac{|S|}{6}$$

$$\int N_1 N_1 V \mathrm{d}x\mathrm{d}y = 2|S|\int N_1^2 (V_1 N_1 + V_2 N_2 + V_3 N_3)\mathrm{d}\xi\mathrm{d}\eta$$

$$= 2|S|\left(\frac{V_1}{20} + \frac{V_2 + V_3}{60}\right) = |S|\frac{3V_1 + V_2 + V_3}{30}$$

Finally, we calculate the products involving different nodal functions, which, as seen in Subsection 4.5.4, can be always be expressed as a product of the nodal functions N_1 and N_2 for a determined ordering of the triangle vertexes. These products, which have the form either $\langle N_1|\hat{\mathcal{O}}|N_2\rangle$ or $\langle N_2|\hat{\mathcal{O}}|N_1\rangle$, are used for the construction of the non diagonal subblocks:

$$\int N_1 \hat{\mathrm{k}}_\mathrm{x}\hat{\mathrm{k}}_\mathrm{x} N_2 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial x}N_1\right)\left(\frac{\partial}{\partial x}N_2\right)\mathrm{d}x\mathrm{d}y = \frac{y_2 - y_3}{2S}\frac{y_3 - y_1}{2S}2|S|\int \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{(y_2 - y_3)(y_3 - y_1)}{4|S|}$$

$$\int N_2 \hat{\mathrm{k}}_\mathrm{x}\hat{\mathrm{k}}_\mathrm{x} N_1 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial x}N_2\right)\left(\frac{\partial}{\partial x}N_1\right)\mathrm{d}x\mathrm{d}y = \frac{y_2 - y_3}{2S}\frac{y_3 - y_1}{2S}2|S|\int \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{(y_2 - y_3)(y_3 - y_1)}{4|S|}$$

$$\int N_1 \hat{\mathrm{k}}_\mathrm{x} k_y N_2 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial x}N_1\right)\left(\frac{\partial}{\partial y}N_2\right)\mathrm{d}x\mathrm{d}y = \frac{y_2 - y_3}{2S}\frac{x_1 - x_3}{2S}2|S|\int \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{(y_2 - y_3)(x_1 - x_3)}{4|S|}$$

$$\int N_2 \hat{\mathrm{k}}_\mathrm{x} k_y N_1 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial x}N_2\right)\left(\frac{\partial}{\partial y}N_1\right)\mathrm{d}x\mathrm{d}y = \frac{y_3 - y_1}{2S}\frac{x_3 - x_2}{2S}2|S|\int \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{(y_3 - y_1)(x_3 - x_2)}{4|S|}$$

$$\int N_1 \hat{\mathrm{k}}_\mathrm{x} k_\mathrm{z} N_2 \mathrm{d}x\mathrm{d}y = \mathrm{i}k_\mathrm{z}\int \left(\frac{\partial}{\partial x}N_1\right)N_2 \mathrm{d}x\mathrm{d}y = \mathrm{i}k_\mathrm{z}\frac{y_2 - y_3}{2S}2|S|\int N_2 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm \mathrm{i}k_\mathrm{z}\frac{y_2 - y_3}{6}$$

$$\int N_2 \hat{k}_x k_z N_1 \mathrm{d}x\mathrm{d}y = \mathrm{i}k_z \int \left(\frac{\partial}{\partial x}N_2\right)N_1\mathrm{d}x\mathrm{d}y = \mathrm{i}k_z \frac{y_3 - y_1}{2S}2|S|\int N_1 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm\mathrm{i}k_z\frac{y_3 - y_1}{6}$$

$$\int N_1 k_y \hat{k}_x N_2 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial y}N_1\right)\left(\frac{\partial}{\partial x}N_2\right)\mathrm{d}x\mathrm{d}y = \frac{x_3 - x_2}{2S}\frac{y_3 - y_1}{2S}2|S|\int \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{(x_3 - x_2)(y_3 - y_1)}{4|S|}$$

$$\int N_2 k_y \hat{k}_x N_1 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial y}N_2\right)\left(\frac{\partial}{\partial x}N_1\right)\mathrm{d}x\mathrm{d}y = \frac{x_1 - x_3}{2S}\frac{y_2 - y_3}{2S}2|S|\int \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{(x_1 - x_3)(y_2 - y_3)}{4|S|}$$

$$\int N_1 k_y k_y N_2 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial y}N_1\right)\left(\frac{\partial}{\partial y}N_2\right)\mathrm{d}x\mathrm{d}y = \frac{x_3 - x_2}{2S}\frac{x_1 - x_3}{2S}2|S|\int \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{(x_3 - x_2)(x_1 - x_3)}{4|S|}$$

$$\int N_2 k_y k_y N_1 \mathrm{d}x\mathrm{d}y = \int \left(\frac{\partial}{\partial y}N_2\right)\left(\frac{\partial}{\partial y}N_1\right)\mathrm{d}x\mathrm{d}y = \frac{x_1 - x_3}{2S}\frac{x_3 - x_2}{2S}2|S|\int \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{(x_1 - x_3)(x_3 - x_2)}{4|S|}$$

$$\int N_1 k_y k_z N_2 \mathrm{d}x\mathrm{d}y = \mathrm{i}k_z \int \left(\frac{\partial}{\partial y}N_1\right)N_2\mathrm{d}x\mathrm{d}y = \mathrm{i}k_z \frac{x_3 - x_2}{2S}2|S|\int N_2 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm\mathrm{i}k_z\frac{x_3 - x_2}{6}$$

$$\int N_2 k_y k_z N_1 \mathrm{d}x\mathrm{d}y = \mathrm{i}k_z \int \left(\frac{\partial}{\partial y}N_2\right)N_1\mathrm{d}x\mathrm{d}y = \mathrm{i}k_z \frac{x_1 - x_3}{2S}2|S|\int N_1 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm\mathrm{i}k_z\frac{x_1 - x_3}{6}$$

$$\int N_1 k_z \hat{k}_x N_2 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \int N_1\left(\frac{\partial}{\partial x}N_2\right)\mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \frac{y_3 - y_1}{2S}2|S|\int N_1 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm\mathrm{i}k_z\frac{y_1 - y_3}{6}$$

$$\int N_2 k_z \hat{k}_x N_1 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \int N_2\left(\frac{\partial}{\partial x}N_1\right)\mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \frac{y_2 - y_3}{2S}2|S|\int N_2 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm\mathrm{i}k_z\frac{y_3 - y_2}{6}$$

$$\int N_1 k_z k_y N_2 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \int N_1\left(\frac{\partial}{\partial y}N_2\right)\mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \frac{x_1 - x_3}{2S}2|S|\int N_1 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm\mathrm{i}k_z\frac{x_3 - x_1}{6}$$

$$\int N_2 k_z k_y N_1 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \int N_2 \left(\frac{\partial}{\partial y}N_1\right)\mathrm{d}x\mathrm{d}y = -\mathrm{i}k_z \frac{x_3 - x_2}{2S}2|S|\int N_2 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm\mathrm{i}k_z \frac{x_2 - x_3}{6}$$

$$\int N_1 k_z k_z N_2 \mathrm{d}x\mathrm{d}y = k_z^2 \int N_1 N_2 \mathrm{d}x\mathrm{d}y = k_z^2 2|S|\int N_1 N_2 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{k_z^2|S|}{12}$$

$$\int N_2 k_z k_z N_1 \mathrm{d}x\mathrm{d}y = k_z^2 \int N_2 N_1 \mathrm{d}x\mathrm{d}y = k_z^2 2|S|\int N_1 N_2 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{k_z^2|S|}{12}$$

$$\int N_1 k_0 \hat{\mathrm{k}}_x N_2 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \int N_1 \left(\frac{\partial}{\partial x}N_2\right)\mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \frac{y_3 - y_1}{2S}2|S|\int N_1 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm\mathrm{i}k_0 \frac{y_1 - y_3}{6}$$

$$\int N_2 k_0 \hat{\mathrm{k}}_x N_1 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \int N_2 \left(\frac{\partial}{\partial x}N_1\right)\mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \frac{y_2 - y_3}{2S}2|S|\int N_2 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm\mathrm{i}k_0 \frac{y_3 - y_2}{6}$$

$$\int N_1 k_0 k_y N_2 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \int N_1 \left(\frac{\partial}{\partial y}N_2\right)\mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \frac{x_1 - x_3}{2S}2|S|\int N_1 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm\mathrm{i}k_0 \frac{x_3 - x_1}{6}$$

$$\int N_2 k_0 k_y N_1 \mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \int N_2 \left(\frac{\partial}{\partial y}N_1\right)\mathrm{d}x\mathrm{d}y = -\mathrm{i}k_0 \frac{x_3 - x_2}{2S}2|S|\int N_2 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \pm\mathrm{i}k_0 \frac{x_2 - x_3}{6}$$

$$\int N_1 k_0 k_z N_2 \mathrm{d}x\mathrm{d}y = k_0\ k_z \int N_1 N_2 \mathrm{d}x\mathrm{d}y = k_0\ k_z 2|S|\int N_1 N_2 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{k_0\ k_z|S|}{12}$$

$$\int N_2 k_0 k_z N_1 \mathrm{d}x\mathrm{d}y = k_0\ k_z \int N_2 N_1 \mathrm{d}x\mathrm{d}y = k_0\ k_z 2|S|\int N_1 N_2 \mathrm{d}\xi\mathrm{d}\eta$$

$$= \frac{k_0\ k_z|S|}{12}$$

$$\int N_1 N_2 \mathrm{d}x\mathrm{d}y = 2|S|\int N_1 N_2 \mathrm{d}\xi\mathrm{d}\eta = \frac{|S|}{12}$$

$$\int N_2 N_1 \mathrm{d}x\mathrm{d}y = 2|S|\int N_1 N_2 \mathrm{d}\xi\mathrm{d}\eta = \frac{|S|}{12}$$

$$\int N_1 N_2 V \,\mathrm{d}x\mathrm{d}y = 2|S| \int N_1 N_2 (V_1 N_1 + V_2 N_2 + V_3 N_3) \mathrm{d}\xi\mathrm{d}\eta$$

$$= 2|S| \left(\frac{V_1 + V_2}{60} + \frac{V_3}{120} \right) = |S| \frac{2V_1 + 2V_2 + V_3}{60}$$

$$\int N_2 N_1 V \,\mathrm{d}x\mathrm{d}y = 2|S| \int N_1 N_2 (V_1 N_1 + V_2 N_2 + V_3 N_3 + E_v) \mathrm{d}\xi\mathrm{d}\eta$$

$$= 2|S| \left(\frac{V_1 + V_2}{60} + \frac{V_3}{120} + \frac{E_v}{24} \right) = |S| \frac{2V_1 + 2V_2 + V_3 + 5E_v}{60}$$

Part VI

References

List of publications

Journal papers

- **Celso Martinez-Blanque**, Francisco J. García Ruiz, Andres Godoy, Enrique G. Marin, Luca Donetti, and Francisco Gámiz. Influence of alloy disorder scattering on the hole mobility of SiGe nanowires. *Journal of Applied Physics*, 116(24):244502, December 2014.

- Enrique G. Marin, Francisco J. García Ruiz, Andres Godoy, Isabel M. Tienda-Luna,**Celso Martinez-Blanque**, and Francisco Gámiz. Theoretical interpretation of the electron mobility behavior in InAs nanowires. *Journal of Applied Physics*, 116(17):174505, November 2014.

- Enrique G. Marin, Francisco J. García Ruiz, Andres Godoy, **Celso Martinez-Blanque**, Luca Donetti, and Francisco Gámiz. Impact of the Back-Gate Biasing on Trigate MOSFET Electron Mobility. *IEEE Transactions on Electron Devices*, 62(1):224, January 2015.

- Isabel M. Tienda-Luna, Juan B. Roldán,Francisco J. García Ruiz, **Celso Martinez-Blanque**, and Francisco Gámiz. An analytical mobility model for square Gate-All-Around MOSFETs. *Solid-State Electronics*, 90:18, January 2013.

- Isabel M. Tienda-Luna, Francisco J. García Ruiz, Andres Godoy, Luca Donetti, **Celso Martinez-Blanque**, and Francisco Gámiz. Effect of confined acoustic phonons on the electron mobility of rectangular nanowires. *Applied Physics Letters*, 103(16):163107, October 2013.

Conference contributions

- **Celso Martinez-Blanque**, Francisco J. García Ruiz, Andres Godoy, Luca Donetti, Isabel M. Tienda-Luna, and Francisco Gámiz. *EUROSOI,2012.*

- Isabel M. Tienda-Luna, Juan B. Roldán, Francisco J. García Ruiz,**Celso Martinez-Blanque** , and Francisco Gámiz. An analytical mobility model for square Gate-All-Around MOSFETs *EUROSOI,2012.*

- Francisco G. Ruiz, Enrique G. Marin, Isabel M. Tienda-Luna, Andres Godoy, **Celso Martinez-Blanque** and Francisco Gámiz. Influence of the back-gate bias on the electron mobility of Trigate MOSFETs. *Simulation of Semiconductor Processes and Devices (SISPAD), 2013 International Conference on.*

- Francisco G. Ruiz, Enrique G. Marin, I. M. Tienda-Luna, A. Godoy, **C.Martinez-Blanque** and F. Gámiz. Back-Gate Biasing Influence on the Electron Mobility and the Threshold Voltage of Ultra Thin Box Multigate MOSFET. *Silicon Nanoelectronics Workshop (SiNW), 2013 IEEE.*

Bibliography

[1] nanohub.org.

[2] The Mathworks, Inc. *Partial Differential Equation Toolbox*$^{\text{TM}}$. *User's guide* The Mathworks,Inc.

[3] *International Technology Roadmap for Semiconductors. Available online http://www.itrs.net.* 2014.

[4] D. A. Antoniadis, I. Aberg, C. N. Chléirigh, O. Nayfeh, A. Khakifirooz, and J. L. Hoyt. Continuous MOSFET performance increase with device scaling: The role of strain and channel material innovations. *IBM Journal of Research and Development*, 50(4):363–376, 2006.

[5] A. Asenov, S. Roy, R. A. Brown, G. Roy, C. Alexander, C. Riddet, C. Millar, B. Cheng, A. Martinez, N. Seoane, D. Reid, M. F. Bukhori, X. Wang, and U. Kovac. Advanced simulation of statistical variability and reliability in nano CMOS transistors. In *2008 IEEE International Electron Devices Meeting*, pages 1–1. IEEE, Dec. 2008.

[6] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Saunders College, Philadelphia, 1976.

[7] T. B. Bahder. Eight-band $\boldsymbol{k}\cdot\boldsymbol{p}$ model of strained zinc-blende crystals. *Physical Review B*, 41(17):11992–12001, 1990.

[8] E. Bangert and G. Landwehr. Self-consistent calculations of electric subbands in p-type silicon inversion layers. *Surface Science*, 58(1):138–140, Aug. 1976.

[9] J. Bardeen. An Improved Calculation of the Energies of Metallic Li and Na. *The Journal of Chemical Physics*, 6:367, 1938.

[10] J. Bardeen and W. Shockley. Deformation Potentials and Mobilities in Non-Polar Crystals. *Physical Review*, 80(1):72–80, Oct. 1950.

[11] S. Barraud, J.-M. Hartmann, V. Maffini-Alvaro, L. Tosti, V. Delaye, and D. Lafond. Top-Down Fabrication of Epitaxial SiGe/Si Multi-(Core/Shell) p-FET Nanowire Transistors. *IEEE Transactions on Electron Devices*, 61(4):953–956, Apr. 2014.

[12] S. Barraud, E. Sarrazin, and a. Bournel. Temperature and size dependences of electrostatics and mobility in gate-all-around MOSFET devices. *Semiconductor Science and Technology*, 26(2):025001, Feb. 2011.

[13] G. Bastard. *Wave mechanics applied to semiconductor heterostructures*. Les éditions de Physique, 1992.

[14] D. J. BenDaniel and C. B. Duke. Space-Charge Effects on Electron Tunneling. *Physical Review*, 152(2):683, 1966.

[15] B. R. Bennett, M. G. Ancona, J. B. Boos, and B. V. Shanabrook. Mobility enhancement in strained p-InGaSb quantum wells. *Applied Physics Letters*, 91(4):042104, July 2007.

[16] K. Bhaumik, B. K. Ridley, and Y. Shacham-Diamand. Hole-phonon scattering in strained SiGe quantum wells. *Journal of Applied Physics*, 74(9):5546, 1993.

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

[18] S. Birner. *Modeling of semiconductor nanostructures and semiconductor-electrolyte interfaces*. PhD thesis, 2011.

[19] T. B. Boykin, G. Klimeck, R. Bowen, and F. Oyafuso. Diagonal parameter shifts due to nearest-neighbor displacements in empirical tight-binding theory. *Physical Review B*, 66(12):125207, Sept. 2002.

[20] B. N. Brockhouse. Lattice vibrations in silicon and germanium. *Physical Review Letters*, 2:256–258, 1959.

[21] M. G. Burt. A new effective-mass equation for microstructures. *Semiconductor Science and Technology*, 3:1224–1226, 1988.

[22] M. G. Burt. The justification for applying the effective-mass approximation to microstructures. *Journal of Physics: Condensed Matter*, 6651(4):6651–6690, 1992.

[23] M. G. Burt. Fundamentals of envelope function theory for electronic states and photonic modes in nanostructures. *Journal of Physics: Condensed Matter*, 11(53), 1999.

[24] G. Busch and O. Vogt. Elektrische Leitfähigkeit und Halleffekt von Ge-Si-Legierungen. *Helv. Phys. Acta*, 33:437, 1960.

[25] G. K. Celler and S. Cristoloveanu. Frontiers of silicon-on-insulator. *Journal of Applied Physics*, 93(9):4955, 2003.

[26] J. Chelikowsky and M. Cohen. Electronic structure of silicon. *Physical Review B*, 10(12):5095–5107, Dec. 1974.

[27] S. Chuang and C. Chang. $\boldsymbol{k} \cdot \boldsymbol{p}$ method for strained wurtzite semiconductors. *Physical Review B*, 54:2491–2504, 1996.

[28] J.-P. Colinge, J. Alderman, and C. Cleavelin. Quantum-mechanical effects in trigate SOI MOSFETs. *IEEE Transactions on Electron Devices*, 53(5):1131–1136, May 2006.

[29] S. Datta. *Quantum transport: atom to transistor*. Cambridge University Press, 2005.

[30] R. Dennard, F. Gaensslen, V. Rideout, E. Bassous, and A. LeBlanc. Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE Journal of Solid-State Circuits*, 9(5):256–268, Oct. 1974.

[31] L. Donetti, F. Gamiz, A. Godoy, and N. Rodriguez. Fully self-consistent $\boldsymbol{k} \cdot \boldsymbol{p}$ solver and Monte Carlo simulator for hole inversion layers. In *ESSDERC 2008 - 38th European Solid-State Device Research Conference*, pages 254–257. IEEE, 2008.

[32] L. Donetti, F. Gamiz, N. Rodriguez, and A. Godoy. Hole Mobility in Ultrathin Double-Gate SOI Devices: The Effect of Acoustic Phonon Confinement. *IEEE Electron Device Letters*, 30(12):1338–1340, Dec. 2009.

[33] L. Donetti, F. Gamiz, S. Thomas, T. E. Whall, D. R. Leadley, P.-E. Hellström, G. Malm, and M. Östling. Hole effective mass in silicon inversion layers with different substrate orientations and channel directions. *Journal of Applied Physics*, 110(6):063711, 2011.

[34] B. Doyle, B. Boyanov, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, R. Rios, and R. Chau. Tri-Gate fully-depleted CMOS transistors: fabrication, design and layout. In *2003 Symposium on VLSI Technology. Digest of Technical Papers (IEEE Cat. No.03CH37407)*, pages 133–134. Japan Soc. Applied Phys, 2003.

[35] G. Dresselhaus. Spin-Orbit Coupling Effects in Zinc Blende Structures. *Physical Review*, 100(2):580–586, Oct. 1955.

[36] G. Dresselhaus, A. Kip, and C. Kittel. Cyclotron Resonance of Electrons and Holes in Silicon and Germanium Crystals. *Physical Review*, 98(2):368–384, Apr. 1955.

[37] J. Dura, F. Triozon, S. Barraud, D. Munteanu, S. Martinie, and J. L. Autran. Kubo-Greenwood approach for the calculation of mobility in gate-all-around nanowire metal-oxide-semiconductor field-effect transistors including screened remote Coulomb scattering-Comparison with experiment. *Journal of Applied Physics*, 111(10):103710, 2012.

[38] R. Enderlein, G. Sipahi, L. Scolfaro, and J. Leite. Parameters of the Kane Model from Effective Masses: Ambiguities and Instabilities. *physica status solidi (b)*, 206(2):623–633, Apr. 1998.

[39] 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*, 105(5):053702, 2009.

[40] D. Esseni, P. Palestri, and L. Selmi. *Nanoscale MOS Transistors: Semi-Classical Tranasport and Applications*. Cambridge University Press, New York, 2011.

[41] D. K. Ferry and C. Jacoboni, editors. *Quantum transport in semiconductors*. Plenum Press, New York, 1992.

[42] 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*, 80(4):2234, 1996.

[43] B. A. Foreman. Effective-mass Hamiltonian and boundary conditions for the valence bands of semiconductor microstructures. *Physical Review B*, 48(7):4964, 1993.

[44] B. A. Foreman. Analytical Envelope-Function Theory of Interface Band Mixing. *Physical Review Letters*, 81(2):425–428, July 1998.

[45] M. M. Frank. High-k / metal gate innovations enabling continued CMOS scaling. In *2011 Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, pages 25–33. IEEE, Sept. 2011.

[46] C. Galeriu. $\boldsymbol{k}\cdot\boldsymbol{p}$ Theory of Semiconductor Nanostructures. 2005.

[47] F. Gamiz, J. Lopez-Villanueva, J. B. Roldán, J. Carceller, and P. Cartujo. Monte Carlo simulation of electron transport properties in extremely thin SOI MOSFET's. *IEEE Transactions on Electron Devices*, 45(5):1122–1126, May 1998.

[48] M. Glicksman. Mobility of Electrons in Germanium-Silicon Alloys. *Physical Review*, 111(1):125–128, July 1958.

[49] M. Godfrey and A. Malik. Boundary conditions and spurious solutions in envelope-function theory. *Physical Review B*, 53(24):16504–16509, June 1996.

[50] D. A. Greenwood. The Boltzmann Equation in the Theory of Electrical Conduction in Metals. *Proceedings of the Physical Society*, 71(4):585–596, 1958.

[51] W. Haensch, E. J. Nowak, R. H. Dennard, P. M. Solomon, A. Bryant, O. H. Dokumaci, A. Kumar, X. Wang, J. B. Johnson, and M. V. Fischetti. Silicon CMOS Devices Beyond Scaling. *IBM Journal of Research and Development*, 50(4.5):339–361, July 2006.

[52] C. Harris and E. O'Reilly. Nature of the band gap of silicon and germanium nanowires. *Physica E: Low-dimensional Systems and Nanostructures*, 32(1-2):341–345, May 2006.

[53] S. Harris. *An Introduction to the Theory of the Boltzmann Equation*. Dover Publications, New York, 2004.

[54] J. Hensel, H. Hasegawa, and M. Nakayama. Cyclotron Resonance in Uni-axially Stressed Silicon. II. Nature of the Covalent Bond. *Physical Review*, 138(1A):A225–A238, Apr. 1965.

[55] C. Herring and E. Vogt. Transport and deformation-potential theory for many-valley semiconductors with anisotropic scattering. *Physical Review*, 101(3):994, 1956.

[56] J. Hinckley and J. Singh. Influence of substrate composition and crystallographic orientation on the band structure of pseudomorphic Si-Ge alloy films. *Physical Review B*, 42(6):3546–3566, Aug. 1990.

[57] B. Hoeneisen and C. Mead. Fundamental limitations in microelectronics-I. MOS technology. *Solid-State Electronics*, 15(7):819–829, July 1972.

[58] M. Holm, M.-E. Pistol, and C. Pryor. Calculations of the electronic structure of strained InAs quantum dots in InP. *Journal of Applied Physics*, 92(2):932, 2002.

[59] C. Hu. Gate oxide scaling limits and projection. In *International Electron Devices Meeting. Technical Digest*, pages 319–322. IEEE, 1996.

[60] M. S. Islam and M. Alam. InxGa1-xSb n-channel MOSFET: Effect of interface states on CV characteristics. In *INEC*, pages 197–200, 2013.

[61] J.-M. Jancu, R. Scholz, F. Beltram, and F. Bassani. Empirical spds* tight-binding calculation for cubic semiconductors: General method and material parameters. *Physical Review B*, 57(11):6493–6507, Mar. 1998.

[62] S. Jin, M. V. Fischetti, and T.-W. Tang. Modeling of electron mobility in gated silicon nanowires at room temperature: Surface roughness scattering, dielectric screening, and band nonparabolicity. *Journal of Applied Physics*, 102(8):083715, 2007.

[63] S. Jin, T.-W. Tang, and M. V. Fischetti. Simulation of Silicon Nanowire Transistors Using Boltzmann Transport Equation Under Relaxation Time Approximation. *IEEE Transactions on Electron Devices*, 55(3):727–736, 2008.

[64] E. O. Kane. Energy band structure in p-type germanium and silicon. *Journal of Physics and Chemistry of Solids*, 1(1-2):82–99, Sept. 1956.

[65] E. O. Kane. Band structure of indium antimonide. *Journal of Physics and Chemistry of Solids*, 1(4):249–261, Jan. 1957.

[66] E. O. Kane. The semi-empirical approach to band structure. *Journal of Physics and Chemistry of Solids*, 8:38–44, Jan. 1959.

[67] E. O. Kane. *Narrow Gap Semiconductors Physics and Applications*, volume 133 of *Lecture Notes in Physics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1980.

[68] R. Kotlyar, T. D. Linton, R. Rios, M. D. Giles, S. M. Cea, K. J. Kuhn, M. Povolotskyi, T. Kubis, and G. Klimeck. Does the low hole transport mass in ⟨110⟩ and ⟨111⟩ Si nanowires lead to mobility enhancements at high field and stress: A self-consistent tight-binding study. *Journal of Applied Physics*, 111(12):123718, 2012.

[69] R. Kotlyar, C. Weber, L. Shifren, S. Cea, M. D. Giles, and M. Stettler. Effect of band warping and wafer orientation on NMOS mobility under arbitrary applied stress. *Journal of Computational Electronics*, 7(3):95–98, Dec. 2007.

[70] T. Koyanagi, K. Tachi, K. Okamoto, K. Kakushima, P. Ahmet, K. Tsutsui, N. Sugii, T. Hattori, and H. Iwai. Electrical Characterization of La_2O_3 - Gated Metal Oxide Semiconductor Field Effect Transistor with Mg Incorporation. *Japanese Journal of Applied Physics*, 48(5):05DC02, May 2009.

[71] R. Kubo. Statistical-mechanical theory of irreversible processes. I. general theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12:570–586, 1957.

[72] P. Lawaetz. Low-Field Mobility and Galvanomagnetic Properties of Holes in Germanium with Phonon Scattering. *Physical Review*, 174(3):867–880, Oct. 1968.

[73] C.-W. Lee, A. N. Nazarov, I. Ferain, N. D. Akhavan, R. Yan, P. Razavi, R. Yu, R. T. Doria, and J.-P. Colinge. Low subthreshold slope in junctionless multigate transistors. *Applied Physics Letters*, 96(10):102106, 2010.

[74] 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 Transactions on Electron Devices*, 55:2086–2096, 2008.

[75] J. P. Loehr. *Physics of Strained Quantum Well Lasers*. Springer US, Boston, MA, 1998.

[76] D. Long. Scattering of conduction electrons by lattice vibrations in silicon. *Physical Review*, 120(8):2024–2032, Aug. 1960.

[77] J. Lopez-Villanueva, P. Cartujo-Casinello, J. Banqueri, F. Gamiz, and S. Rodriguez. Effects of the inversion layer centroid on MOSFET behavior. *IEEE Transactions on Electron Devices*, 44(11):1915–1922, 1997.

[78] P. Löwdin. A Note on the Quantum-Mechanical Perturbation Theory. *The Journal of Chemical Physics*, 19(11):1396–1401, 1951.

[79] M. Lundstrom. Fundamentals of Carrier Transport,. *Measurement Science and Technology*, 13(2):230–230, 2002.

[80] M. Lundstrom and Z. Ren. Essential physics of carrier transport in nanoscale MOSFETs. *IEEE Transactions on Electron Devices*, 49(1):133–141, 2002.

[81] J. M. Luttinger. Quantum Theory of Cyclotron Resonance in Semiconductors: General Theory. *Physical Review*, 102(4):1030–1041, May 1956.

[82] J. M. Luttinger and W. Kohn. Motion of Electons and Holes in Perturbed Periodic Fields. *Physical Review*, 97(4):869–883, 1955.

[83] E. G. Marin. *Modeling and Simulation of Semiconductor Nanowires for Future Technology Nodes*. PhD thesis, University of Granada, 2014.

[84] E. G. Marin, F. G. Ruiz, a. Godoy, and F. Gámiz. The unexpected beneficial effect of the L-valley population on the electron mobility of GaAs nanowires The unexpected beneficial effect of the L -valley population on the electron mobility of GaAs nanowires. *Applied Physics Express*, 022113:1–5, 2015.

[85] E. G. Marin, F. J. G. Ruiz, I. Tienda-Luna, A. Godoy, and F. Gámiz. Analytical model for the threshold voltage of III-V nanowire transistors including quantum effects. *Solid-State Electronics*, 92:28–34, Feb. 2014.

[86] E. G. Marin, F. J. G. Ruiz, I. M. Tienda-Luna, A. Godoy, and F. Gamiz. Analytical Gate Capacitance Modeling of III-V Nanowire Transistors. *IEEE Transactions on Electron Devices*, 60(5):1590–1599, May 2013.

[87] E. G. Marin, F. J. G. Ruiz, I. M. Tienda-Luna, A. Godoy, P. Sánchez-Moreno, and F. Gamiz. Analytic potential and charge model for III-V surrounding gate metal-oxide-semiconductor field-effect transistors. *Journal of Applied Physics*, 112(8):084512, 2012.

[88] C. Martinez-Blanque, F. J. G. Ruiz, A. Godoy, E. G. Marin, L. Donetti, and F. Gámiz. Influence of alloy disorder scattering on the hole mobility of SiGe nanowires. *Journal of Applied Physics*, 116(24):244502, Dec. 2014.

[89] A. Martins, T. B. Boykin, G. Klimeck, and B. Koiller. Conduction-band tight-binding description for Si applied to P donors. *Physical Review B*, 72(19):193204, Nov. 2005.

[90] S. Mehrotra, P. Long, M. Povolotskyi, and G. Klimeck. Atomistic simulation of phonon and alloy limited hole mobility in Si 1- x Ge x nanowires. *Physica Status Solidi (RRL) - Rapid Research Letters*, 7(10):903–906, Oct. 2013.

[91] S. Mehrotra, A. Paul, and G. Klimeck. Atomistic approach to alloy scattering in $Si_{1-x}Ge_x$. *Applied Physics Letters*, 98(17):173503, 2011.

[92] T. Meng, J. Klinovaja, and D. Loss. Renormalization of anticrossings in interacting quantum wires with Rashba and Dresselhaus spin-orbit couplings. *Physical Review B*, 89(20):205133, May 2014.

[93] D. B. Migas and V. E. Borisenko. Structural, electronic, and optical properties of ⟨001⟩-oriented SiGe nanowires. *Physical Review B*, 76(3):035440, July 2007.

[94] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C.-H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. McIntyre, P. Moon, J. Neirynck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki. A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging. In *2007 IEEE International Electron Devices Meeting*, pages 247–250. IEEE, 2007.

[95] G. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8), 1965.

[96] A. Nainani, B. R. Bennett, J. Brad Boos, M. G. Ancona, and K. C. Saraswat. Enhancing hole mobility in III-V semiconductors. *Journal of Applied Physics*, 111(10):103706, 2012.

[97] A. Nainani, Z. Yuan, T. Krishnamohan, B. R. Bennett, J. B. Boos, M. Reason, M. G. Ancona, Y. Nishi, and K. C. Saraswat. $In_xGa_{1-x}Sb$ channel p-Metal-Oxide-Semiconductor field effect transistors: Effect of strain and heterostructure design. *Journal of Applied Physics*, 110(1):014503, 2011.

[98] N. Neophytou, O. Baumgartner, Z. Stanojević, and H. Kosina. Bandstructure and mobility variations in p-type silicon nanowires under electrostatic gate field. *Solid-State Electronics*, 90:44–50, Dec. 2013.

[99] N. Neophytou and H. Kosina. Atomistic simulations of low-field mobility in Si nanowires: Influence of confinement and orientation. *Physical Review B*, 84(8):085313, Aug. 2011.

[100] N. Neophytou and H. Kosina. Strong anisotropy and diameter effects on the low-field mobility of silicon nanowires. *2011 International Conference on Simulation of Semiconductor Processes and Devices*, pages 31–34, Sept. 2011.

[101] Y.-M. Niquet and C. Delerue. Carrier mobility in strained Ge nanowires. *Journal of Applied Physics*, 112(8):084301, 2012.

[102] Y.-M. Niquet, C. Delerue, D. Rideau, and B. Videau. Fully atomistic simulations of phonon-limited mobility of electrons and Holes in $\langle 001 \rangle$, $\langle 110 \rangle$, and $\langle 111 \rangle$ Oriented Si Nanowires. *IEEE Transactions on Electron Devices*, 59(5):1480–1487, May 2012.

[103] Y.-M. Niquet, A. Lherbier, N. Quang, M. Fernández-Serra, X. Blase, and C. Delerue. Electronic structure of semiconductor nanowires. *Physical Review B*, 73(16):165319, Apr. 2006.

[104] R. Oberhuber, G. Zandler, and P. Vogl. Subband structure and mobility of two-dimensional holes in strained Si/SiGe MOSFET's. *Physical Review B*, 58(15):9941–9948, Oct. 1998.

[105] M. Ogawa, T. Kunimasa, T. Ito, and T. Miyoshi. Finite-element analysis of quantum wires with arbitrary cross sections. *Journal of Applied Physics*, 84(6):3242, 1998.

[106] S. Oktyabrsky and P. Ye. *Fundamentals of III-V Semiconductor MOSFETs*. Springer US, Boston, MA, 2010.

[107] T. P. O'Regan, M. V. Fischetti, B. Sorée, S. Jin, W. Magnus, and M. Meuris. Calculation of the electron mobility in III-V inversion layers with high-κ dielectrics. *Journal of Applied Physics*, 108(10):103705, 2010.

[108] P. Pfeffer and W. Zawadzki. Five-level $\boldsymbol{k}\cdot\boldsymbol{p}$ model for the conduction and valence bands of GaAs and InP. *Physical Review B*, 53(19):12813–12828, May 1996.

[109] A.-T. Pham, C. Jungemann, and B. Meinerzhagen. Physics-Based Modeling of Hole Inversion-Layer Mobility in Strained-SiGe-on-Insulator. *IEEE Transactions on Electron Devices*, 54(9):2174–2182, Sept. 2007.

[110] L. R. Ram-Mohan. *Finite element and boundary element applications in quantum mechanics*. Oxford texts in applied and engineering mathematics. Oxford Univ. Press, Oxford, 2002.

[111] V. V. Ravi Kishore. Electronic structure of core-shell nanowires. 2013.

[112] M. M. Rieger and P. Vogl. Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on Si1-yGey substrates. *Physical Review B*, 48(19):14276–14287, Nov. 1993.

[113] A. Rodina, A. Alekseev, A. Efros, M. Rosen, and B. Meyer. General boundary conditions for the envelope function in the multiband $\boldsymbol{k}\cdot\boldsymbol{p}$ model. *Physical Review B*, 65(12):125302, Feb. 2002.

[114] J. B. Roldán, F. Gamiz, J. Lopez-Villanueva, P. Cartujo, and a. Godoy. Strained-Si on $Si_{1-x}Ge_x$/ MOSFET inversion layer centroid modeling. *IEEE Transactions on Electron Devices*, 48(10):2447–2449, 2001.

[115] J. B. Roldán, A. Godoy, F. Gámiz, and M. Balaguer. Modeling the centroid and the inversion charge in cylindrical surrounding gate MOSFETs, including quantum effects. *IEEE Transactions on Electron Devices*, 55(1):411–416, 2008.

[116] F. J. G. Ruiz, I. M. Tienda-Luna, A. Godoy, L. Donetti, and F. Gamiz. A Model of the Gate Capacitance of Surrounding Gate Transistors: Comparison With Double-Gate MOSFETs. *IEEE Transactions on Electron Devices*, 57(10):2477–2483, Oct. 2010.

[117] J. J. Sakurai. *Modern quantum mechanics*. Adison-Wesley, 1995.

[118] C. Sampedro, F. Gámiz, A. Godoy, R. Valín, a. García-Loureiro, N. Rodríguez, I. Tienda-Luna, F. Martinez-Carricondo, and B. Biel. Multi-Subband Ensemble Monte Carlo simulation of bulk MOSFETs for the 32nm-node and beyond. *Solid-State Electronics*, 65-66:88–93, Nov. 2011.

[119] M. F. H. Schuurmans and G. W. Hooft. Simple calculations of confinement states in a quantum well. *Physical Review B*, 31(12):8041–8048, June 1985.

[120] F. Seitz. *Modern theory of solids*. Mc Graw Hill, New York, 1940.

[121] T. Sekigawa and Y. Hayashi. Calculated threshold-voltage characteristics of an XMOS transistor having an additional bottom gate. *Solid-State Electronics*, 27(8):827–828, Aug. 1984.

[122] G. G. Shahidi. SOI technology for the GHz era. *IBM Journal of Research and Development*, 46(2.3):121–131, Mar. 2002.

[123] M. Shin, S. Lee, and G. Klimeck. Computational Study on the Performance of Si Nanowire pMOSFETs Based on the $\boldsymbol{k} \cdot \boldsymbol{p}$ Method. *IEEE Transactions on Electron Devices*, 57(9):2274–2283, Sept. 2010.

[124] T. Skotnicki, J. Hutchby, T. King, H.-S. Wong, and F. Boeuf. The end of CMOS scaling: toward the introduction of new materials and structural changes to improve MOSFET performance. *IEEE Circuits and Devices Magazine*, 21(1):16–26, Jan. 2005.

[125] Z. Stanojević, O. Baumgartner, V. Sverdlov, and H. Kosina. Electronic band structure modeling in strained Si-nanowires: Two band $\boldsymbol{k} \cdot \boldsymbol{p}$ versus tight binding. In *2010 14th International Workshop on Computational Electronics*, pages 1–4. IEEE, Oct. 2010.

[126] Z. Stanojević, V. Sverdlov, O. Baumgartner, and H. Kosina. Subband engineering in n-type silicon nanowires using strain and confinement. *Solid-state electronics*, 70:73–80, Apr. 2012.

[127] J. H. Stathis and D. J. DiMaria. Reliability projection for ultra-thin oxides at low voltage. In *Electron Devices Meeting, 1998. IEDM '98. Technical Digest., International*, 1998.

[128] B. Streetman and S. Banerjee. *Solid state electronic devices*. Prentice Hall, 1995.

[129] Y. Sun, S. E. Thompson, and T. Nishida. *Strain Effect in Semiconductors*. Springer US, Boston, MA, 2010.

[130] V. Sverdlov. *Strain-Induced Effects in Advanced MOSFETs*. Computational Microelectronics. Springer Vienna, Vienna, 2011.

[131] S. Takagi, S.-H. Kim, M. Yokoyama, R. Zhang, N. Taoka, Y. Urabe, T. Yasuda, H. Yamada, O. Ichikawa, N. Fukuhara, M. Hata, and M. Takenaka. High mobility cmos technologies using iii-v/ge channels on si platform. *Solid-State Electronics*, 88:2–8, 2013. cited By 5.

[132] K. Takeda, K. Sakui, A. Taguchi, and M. Sakata. Effect of intervalence-band interaction on relaxation time and transport coefficients for holes in non-polar semiconductors. *Journal of Physics C: Solid State Physics*, 16(4):729–745, Feb. 1983.

[133] A. V. Y. Thean, Z. H. Shi, L. Mathew, T. Stephens, H. Desjardin, C. Parker, T. White, M. Stoker, L. Prabhu, R. Garcia, B. Y. Nguyen, S. Murphy, R. Rai, J. Conner, B. E. White, and S. Venkatesan. Performance and variability comparisons between multi-gate FETs and planar SOI transistors. In *Technical Digest - International Electron Devices Meeting, IEDM*, pages 1–4. IEEE, 2006.

[134] A. Trellakis, A. T. Galick, A. Pacelli, and U. Ravaioli. Iteration scheme for the solution of the two-dimensional Scrödinger-Poisson equations in quantum structures Schro. *Journal of Applied Physics*, 81(12):7880–7884, 1997.

[135] S.-E. Ungersbock. *Advanced Modeling of Strained CMOS Technology*. PhD thesis, Technischen Unversität Wien, 2007.

[136] 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*, 54(9):2183–2190, Sept. 2007.

[137] 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 *2007 IEEE Symposium on VLSI Technology*, pages 110–111. IEEE, June 2007.

[138] J. V. Vechten and T. Bergstresser. Electronic structures of semiconductor alloys. *Physical Review B*, 641(8), 1970.

[139] R. G. Veprek. *Computational Modeling of Semiconductor Nanostructures for Optoelectronics*. PhD thesis, Swiss Federal Institut of Technology Zürich, 2009.

[140] R. G. Veprek, S. Steiger, and B. Witzigmann. Ellipticity and the spurious solution problem of $\boldsymbol{k}\cdot\boldsymbol{p}$ envelope equations. *Physical Review B*, 76(16):165320, Oct. 2007.

[141] R. G. Veprek, S. Steiger, and B. Witzigmann. Reliable $\boldsymbol{k}\cdot\boldsymbol{p}$ band structure calculation for nanostructures using finite elements. *Journal of Computational Electronics*, 7(4):521–529, Oct. 2008.

[142] R. G. Veprek, S. Steiger, and B. Witzigmann. Operator ordering, ellipticity and spurious solutions in $\boldsymbol{k}\cdot\boldsymbol{p}$ calculations of III-nitride nanostructures. *Optical and Quantum Electronics*, 40(14-15):1169–1174, Jan. 2009.

[143] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan. Band parameters for III-V compound semiconductors and their alloys. *Journal of Applied Physics*, 89(11):5815, 2001.

[144] J. T. Wallmark. Fundamental physical limitations in integrated electronic circuits. In *Institute of Physics Conference*, page 133, 1975.

[145] S. R. White and L. J. Sham. Electronic Properties of Flat-Band Semiconductor Heterostructures. *Phys. Rev. Lett.*, 47(12):879–882, Sept. 1981.

[146] J. Woo Lee, D. Jang, M. Mouis, K. Tachi, G. Tae Kim, T. Ernst, and G. Ghibaudo. Short channel mobility analysis of SiGe nanowire p-type field effect transistors: Origins of the strain induced performance improvement. *Applied Physics Letters*, 101(14):143502, 2012.

[147] G.-F. Yeap, S. Krishnan, and M.-R. Lin. Fringing-induced barrier lowering (FIBL) in sub-100 nm MOSFETs with high-K gate dielectrics. *Electronics Letters*, 34(11):1150, 1998.

[148] P. Y. Yu and M. Cardona. *Fundamentals of semiconductors*. Graduate Texts in Physics. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[149] Z. Yuan, A. Nainani, and X. Guan. Tight-binding study of Γ-L bandstructure engineering for ballistic III-V nMOSFETs. *Simulation of Semiconductor Processes and Devices (SISPAD)*, pages 71–74, 2011.

[150] Z. Yuan, A. Nainani, and A. Kumar. InGaSb: Single channel solution for realizing III-V CMOS. *VLSI Technology (VLSIT), 2012 Symposium on*, 96(c):185–186, 2012.

[151] Y. Zhang, J. Kim, and M. V. Fischetti. Self-consistent calculation for valence subband structure and hole mobility in p-channel inversion layers. *Journal of Computational Electronics*, 7(3):176–180, Dec. 2007.

[152] Y. Zheng, C. Rivas, R. Lake, K. Alam, T. B. Boykin, and G. Klimeck. Electronic Properties of Silicon Nanowires. *IEEE Transactions on Electron Devices*, 52(6):1097–1103, June 2005.