Simulation of three-dimensional nanoelectronic devices

Gianluca Fiori

Tutore: Prof. Giuseppe Iannaccone
Tutore: Prof. Bruno Pellegrini

Pisa, Febbraio 2005
## Contents

1 Introduction .......................... 1

2 Quantum Simulations .................... 3
   2.1 Density Functional Theory .......... 5
   2.2 The Bloch Theorem .................. 8
   2.3 Effective mass equation ............ 9
   2.4 Quantum confined electrons ........ 15
      2.4.1 Integration over the \( k \) space . 15
      2.4.2 One-dimensional quantum confinement . 16
      2.4.3 Two-dimensional quantum confinement ... 20
      2.4.4 Three-dimensional quantum confinement ... 21
   2.5 3D Poisson/Schrödinger equation .... 22

3 Quantum point contact and single electron devices 25
   3.1 The Landauer Formula .............. 26
   3.2 Simulation of Quantum Point Contacts .... 29
      3.2.1 Experiment .................... 31
      3.2.2 Surface states .................. 33
      3.2.3 Simulation results .............. 35
   3.3 Single Electron Box ................. 37
   3.4 SET Capacitance extraction .......... 40
   3.5 Single Electron Transistor .......... 44
      3.5.1 Device structure ............... 45
      3.5.2 Results ....................... 47
   3.6 Concluding remarks ................. 50
# 4 Ultra-short MOSFET Devices

4.1 Ballistic Transport in nanoscaled MOSFETs
   4.1.1 Ballistic electron and current densities in 2D subbands
   4.1.2 25 nm “Well-Tempered” MOSFET

4.2 Random distribution of dopants in nanoscaled MOSFETs
   4.2.1 Random distribution effects in “Well-Tempered” MOSFETs
   4.2.2 Alternative method for the statistical study of random dopant effects

4.3 Silicon Nanowire Transistor
   4.3.1 Physical models
   4.3.2 Results and Discussions

4.4 Concluding remarks

# 5 Simulation of nanocrystal memories

5.1 Device fabrication and simulated structure
5.2 Simulations
5.3 Concluding remarks

# 6 Bohm Quantum Potential

6.1 Density gradient approach
6.2 Bohm quantum potential approach
   6.2.1 Bohm’s potential
   6.2.2 Effective Quantum Potential

6.3 Discretization
   6.3.1 Maxwell-Boltzmann statistics
   6.3.2 Fermi-Dirac statistics
   6.3.3 Boundary conditions

6.4 Numerical Algorithm
6.5 Fitting parameters
6.6 Simulation examples
   6.6.1 MOS capacitor
   6.6.2 FinFET or triple gate FET

6.7 Concluding remarks

# A ViDES User Manual
Acknowledgments

First of all, I would like to thank my advisor Prof. Giuseppe Iannaccone for having taught me most of what I know. Prof. Iannaccone is a remarkable person and I have benefited the experience of working with him throughout the course of this work.

I am also indebted with Prof. Bruno Pellegrini for his advice in improving this work.

I wish to express my gratitude for the benefits that I have gained from conversations with the other members of the Nanoelectronic group, in particular Prof. Massimo Macucci, Dr. Marco Pala, Gilberto Curatola, Pietro Coli, Dr. Luca Bonci, Dr. Mario Gattobigio, Giorgio Mugnaini, Luca Perniola and Andrea Campera.

I gratefully acknowledge Prof. Mark Lundstrom for having given me the opportunity to spend a very fruitful period in Purdue, and Prof. Gerhard Klimeck, who has taught me a lot on “the art” of programming device simulators. Thanks also to the folks in the INAC/NCN laboratory, especially to Dr. Diego Kienle and Dr. Avik Ghosh.

Big thanks to Andrea Cappa, who has helped me lots of times during the internship in SILVACO.

Finally, I would like to thank my parents, Angelo and Laura, my grandmother Liana, my brother Federico, and my girlfriend Lucia, for their indispensable support during these years.
Chapter 1

Introduction

Nanoelectronics has become a reality which we experience everyday. MOSFET devices with effective channel length equal to 90 nm are presently in production, and the 65 nm technology node is close to first shipment [1].

This hectic rush toward reduction of geometries, has its main motivation in the continuous demand for lower power consuming systems in the field of portable electronics (laptop and cellular phones), as well as faster circuits for computational applications.

However, as geometries are shrunk, problems not encountered in the microscopic regime arise. We refer, for example, to the quantum confinement of carriers, which eventually degrades device performances or to the discrete distribution of dopants in the depletion region, which yields statistical fluctuations of device parameters. Indeed, the reduced dimensions diminishes the number of electrons and impurities to hundreds in the inversion layer, therefore fluctuations in their number and position can affect significantly device performance, seriously compromising the correct operation of the device itself.

It is evident that the conventional CMOS process technology can not be scaled indefinitely, and the “red brick wall” has been foreseen by the ITRS’03 in the 2010 in correspondence of the 45 nm node, since, at the state of the art, there are not known solution in most fabrication technical areas.

As a consequence, modifications to convectional process are needed, using strained layers which enhance carrier mobility, or, most of all, exploring alternative non-conventional CMOS architectures, as well as
“exotic” devices such as single electron transistor or molecular devices, which have attracted the attention of the research community in the last few years.

It is evident that Field Effect transistors will remain, in the mean time, the main devices, but, it is mandatory to start thinking right now about the performance and the feasibility of new promising architectures, in order to propose a solution to overcome the limits predicted by the ITRS. To this purpose, device simulations can represent an indispensable tool, since they can give design guidelines of conceptually new devices, and explanations to effects observed experimentally.

In this thesis, we present a simulation framework capable of managing a wide variety of nanoscale devices, ranging from conventional planar MOSFET devices, to nanowire transistors, nanocrystal memories, as well as single electron transistors or quantum point contacts.

In particular, in Chapter 2 we will introduce the theoretical background of the thesis, focusing our attention on the Density Functional Theory and deriving the effective mass approximation, which will be used extensively in all the performed simulations. In the last sections, instead we will derive the expressions for the electron concentrations in the case of different quantum confinements, which will be used in the self-consistent Poisson/Schrödinger iterative scheme.

In the first part of Chapter 3, we will validate a simple model for surface states, comparing simulations and experiments performed on Quantum Point Contacts, realized by the University of Würzburg, and in the last part, a simulation study of Single Electron Transistor defined by split gates on AlGaAs/GaAs heterostructure, will be shown.

In the fourth chapter, we will study quantum confinement and discrete dopant distribution effects in the so called “Well Tempered” MOSFETs and we will also perform an out of the equilibrium analysis, considering ballistic transport in the channel. In the last section, we will study the performance of Silicon Nanowire Transistors, considering the limiting cases of both drift-diffusion and ballistic transport in one-dimensional subbands.

In Chapter 5, Silicon on insulator nanocrystal memories will be studied, and an alternative explanation, as compared to those found in literature, to the observed decrease of the threshold voltage shift as a function of increasing channel width will be given. In Chapter 6 instead, the Bohm Quantum Potential approach as well as its application on semiclassical model will be introduced in order take into account quantum effects.
Quantum mechanics has become an indispensable tool for the analysis of modern electronic devices.

In the many-body system with \( N \) electrons of coordinates \( \mathbf{r}_i \), supposing atoms to be fixed as in the Born-Oppenheimer approximation, the Schrödinger equation in the bulk reads,

\[
\mathcal{H}\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \tag{2.1}
\]

where \( E \) is the total energy, \( \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \) is the many-body wave function, \( m_0 \) is the free electron mass. The Hamiltonian operator \( \mathcal{H} \) can be expressed as

\[
\mathcal{H} = \sum_i -\frac{\hbar^2}{2m_0} \nabla_i^2 + \sum_{i\neq j} U_I(\mathbf{r}_i, \mathbf{r}_j) + \sum_i U_L(\mathbf{r}_i) \tag{2.2}
\]

where \( \nabla_i \) is the differential operation with respect to \( \mathbf{r}_i \), and the potential energy is expressed as the sum of the electron-electron coulomb interactions \( [U_I(\mathbf{r}_i)] \), and the periodic potential due to atoms in the crystal lattice \( [U_L(\mathbf{r}_i)] \) [2].

The direct solution of (2.1) would require many-body methods, and consequently an unbearable computational burden if the target of the simulation is the analysis of complex device out of the equilibrium.

An interesting approach, which has become popular in the last few years, is the Density Functional Theory (DFT), in which, the many-body system under the effect of the potential energy \( U \), is reduced to a one-body problem in a single-particle effective potential energy \( U_{eff} \).
In particular, $U_{eff}$ reads

$$U_{eff}(\mathbf{r}) = U_L(\mathbf{r}) + V_H(\mathbf{r}) + V_{exc}(\mathbf{r}),$$

(2.3)

where $V_H(\mathbf{r})$ is the Hartree potential, which is related to the electrostatic potential $\phi$ by the formula

$$V_H(\mathbf{r}) = -q\phi(\mathbf{r}),$$

(2.4)

and $V_{exc}(\mathbf{r})$ is the exchange-correlation term, which takes into account the coulomb and quantum interaction between electrons.

In particular, equation (2.1) reduces to the Kohn-Sham equation

$$\left[-\frac{\hbar^2}{2m_0} \nabla^2 + U_{eff}(\mathbf{r})\right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}),$$

(2.5)

where $\psi_i$'s, ($i = 1..N$) are the orthonormal orbitals of the system of non-interacting particles, whose density $n_0$ is equal to the electron density $n$ of the interacting particles system, and $\epsilon_i$ are the associated energy levels.

Equation (2.5) is unfortunately still not manageable, because of the presence of the lattice potential.

In order to simplify (2.5), the most widely used approach is based on the effective mass approximation, which avoids treating explicitly the lattice potential, considering indirectly the periodicity of the potential in the effective mass $m^*$,

$$\left[-\frac{\hbar^2}{2m^*} \nabla^2 + E_C(\mathbf{r}) + V_H(\mathbf{r}) + V_{exc}(\mathbf{r})\right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r})$$

(2.6)

and in the potential energy $E_C(\mathbf{r})$, which correspond to the bottom of the flat conduction band.

If we relax the hypothesis of a bulk material and we consider the case of an electronic device, the Hartree potential is not derivable directly by the expression based on the point charges, but boundary conditions have to be taken into account. As a consequence, a proper treatment requires coupling the Schrödinger equation with the Poisson equation

$$\nabla \cdot \varepsilon \nabla \phi = -\rho$$

(2.7)

where $\rho$ is the charge density.

In this chapter, we will put the basis for the physical background of the results that will be presented in the following chapters. In particular,
2.1 Density Functional Theory

Density Functional Theory (DFT) has become in the last few years a very popular approach in the semiconductor and molecular device simulation community. Its popularity is mainly due to the fact that, as compared to the other variational methods as the Hartree-Fock method, DFT deals with the one-body electron density instead that with many-electron wavefunctions.

Its development is due to the work of Hohenberg, Kohn and Sham [3],[4]. In particular the Hohenberg and Kohn theorem states that taken a system of $N$ particles, there is a one-to-one correspondence between the ground-state density $n(r)$ and the external potential acting on it (that we assume, from here on, to be given by electron-nuclei interactions). This means not only that given the external potential and hence the Hamiltonian, the electron density is uniquely determined, but that also the inverse is true, i.e. given the electron density we can uniquely determine the external potential, or in other words, we can find a functional $F$ that

$$U_L = F[n(r)].$$

(2.8)

As an important consequence, the ground state energy of the system $E$ is a function of $n$ as well as the kinetic energy $T$ and the potential in the Hamiltonian operator.

Indeed

$$E[n] = T[n] + U_I[n] + \int U_L(r)n(r)dr.$$  

(2.9)

By means of a variational principle with respect to the charge density, the total energy of the system can then be computed.

We note that in (2.9), the term $T[n] + U_I[n]$ does not depend on the external potential acting on the system ($U_L$), so it is referred as a universal functional.

The Hohenberg-Kohn theorem, however, provides the basis for the calculation of $E$, but tells nothing about how to construct the functional. To this purpose, the Kohn-Sham equations provide useful guidelines [4].
In particular, a simple representation of the charge density is provided by assuming a system of \( N \) non-interacting particles with density \( n_0 \),

\[
n_0(r) = \sum_i N_i \langle \psi_i(r) | \psi_i(r) \rangle \tag{2.10}
\]

where \( \psi_i \) are orthonormal orbitals and \( N_i \) is the occupation factor.

If we now assume that \( n_0(r) = n(r) \), the Hohenberg-Kohn theorem states that the external potential \( U_{L0}(r) \) that produces the ground state density \( n_0(r) \), is equal to \( U_L(r) \).

Rewriting (2.9) as

\[
E[n] = T_0[n] + V_H[n] + \int U_L(r)n(r)dr + E_{exc}[n], \tag{2.11}
\]

where \( T_0[n] \) is the kinetic energy of the non-interacting electrons system, \( V_H[n] \) is the Hartree potential

\[
V_H[n] = \frac{1}{8\pi\varepsilon} \int n(r) \frac{q^2}{|r - r'|} n(r')drdr', \tag{2.12}
\]

\( \varepsilon \) is the dielectric constant and \( E_{exc}[n] \) is the exchange-correlation functional

\[
E_{exc}[n] = T[n] - T_0[n] + U_I[n] - V_H[n], \tag{2.13}
\]

performing a variational calculation, we obtain the Kohn-Sham equation

\[
\left[-\frac{\hbar^2}{2m_0} \nabla^2 + U_L(r) + V_H(r) + V_{exc}(r)\right] \psi_i(r) = \epsilon_i \psi_i(r), \tag{2.14}
\]

where

\[
V_{exc} = \frac{\partial E_{exc}[n]}{\partial n(r)} \tag{2.15}
\]

As can be noted, the system of interacting particles has been reduced to a system of non-interacting particles in an effective single-particle potential. The total energy reads,

\[
E = \sum_i N_i \epsilon_i - V_H[n] + E_{exc}[n] - \int V_{exc}(r)n(r)dr. \tag{2.16}
\]

We have to point out, that the \( \epsilon_i \)’s, strictly speaking, have not a physical meaning since they are the Lagrange multipliers used to solve the differential equations. Nevertheless, with sufficient accuracy, they can
be considered as single-particle energies associated to the wavefunctions \( \psi_i(r) \).

Within the DFT framework, it is easy to include the exchange-correlation term in the Schrödinger equation, as in (2.6). Obviously, for different systems, different exchange-correlation terms have to be considered.

As an example, we can refer to the exchange potential derived by Slater [5], for a homogeneous electron gas at 0 K, embedded in a uniform neutralizing background of positive charges. As demonstrated in [6] the exchange potential reads

\[
V_{\text{exc}} = -\frac{q^2 k_F}{2\pi^2 \varepsilon} G\left(\frac{k}{k_F}\right),
\]

where \( k_F \) is the wave vector of the Fermi sphere in the \( k \)-space that separates occupied by unoccupied states and \( G \) reads

\[
G(x) = \frac{1}{2} + \frac{1-x^2}{4x} \ln \left| \frac{1+x}{1-x} \right|.
\]

The average of \( G(x) \) within the Fermi sphere reads,

\[
\bar{G} = \frac{\int_0^1 x^2 G(x)dx}{\int_0^1 x^2dx} = \frac{3}{4}.
\]

Replacing then \( G(x) \) with its average value, we find that the average exchange potential is

\[
V_{\text{exc}} = -\frac{3q^2}{8\pi^2 \varepsilon} k_F.
\]

Let us now express (2.20) as a function of the electron density \( n \), which can be expressed as

\[
n = \frac{N}{V} \quad (2.21)
\]

where \( N \) is the number of electron in the volume \( V \). In the \( k \)-space, the volume occupied by one state is \( \frac{(2\pi)^3}{V} \), so the number of electron within the Fermi sphere is

\[
N = 2 \frac{4\pi k_F^3}{(2\pi)^3},
\]

where we have considered spin degeneracy. From (2.21) and (2.22) we obtain

\[
k_F^3 = 3n\pi^2.
\]
Substituting (2.23) in (2.20) we finally obtain
\[ V_{exc} = \frac{3q^2}{8\pi^2\varepsilon}(3\pi^2 n)^{\frac{3}{2}} \] (2.24)

### 2.2 The Bloch Theorem

Let us consider a one-dimensional lattice at \( T = 0 \) K without any impurities, with lattice constant \( a \), and length \( L = Na \) [7]. Supposing \( L \) very long, the 1D system can be considered infinite. Due to the translational symmetry, the potential \( V(x) \) felt by electrons is periodic
\[ V(x, t) = V(x + na, t), \] (2.25)
and, as a consequence, we can express \( V \) by the Fourier expansion
\[ V(x, t) = \sum_n V_n(t)e^{jnax}, \] (2.26)
where \( h_n = \frac{2\pi n}{a} \).

In the particular case of constant \( V \) (and hence periodic), the eigenfunctions have the known form:
\[ W_k = \frac{1}{\sqrt{L}}e^{jKx}, \] (2.27)
where \( k \) is
\[ k = \frac{2\pi m}{L}, \text{ } m = 1, 2, .. \] (2.28)

Taken \( W_k \), and applying the \( \mathcal{H} \) operator, the span of \( \mathcal{H} \) will have the following basis set
\[ S_k = \{ W_k(x), W_{k+h_1}(x), W_{k-h_1}(x), W_{k+h_2}(x), W_{k-h_2}(x), \ldots \} \] (2.29)

The eigenfunctions \( \Psi_k \) of the time dependent Schrödinger equation
\[ j\hbar \frac{\partial \Psi(r, t)}{\partial t} = \mathcal{H}(r, t) \] (2.30)
in which the potential term is the periodic potential \( V(r) \), can be expressed as a linear combination of vectors belonging to \( S_k \).
\[ \Psi_k(x, t) = \sum_n c_n(k, t)e^{j(k+h_n)x} \] (2.31)
We note that if we take another \( k' \neq k + h_n \), we obtain another basis set \( S_{k'} \) that is not a linear combination of \( S_k \): we can then limit the study to an interval \(-\frac{\pi}{a} \leq k \leq \frac{\pi}{a}\) called first Brillouin zone, since the eigenfunctions associated to a \( k \) belonging outside the interval, can be expressed as a linear combination of the eigenfunction belonging the first Brillouin zone.

Equation (2.31) can be further expressed as

\[
\Psi_k(x, t) = e^{j k x} \sum_n c_n(k, t) e^{j h_n x} = e^{j k x} u(k, x) \quad (2.32)
\]

where the dependence on time is implicit and \( u(k, x) = u(k, x + na) \) is a periodic function since,

\[
u(k, x) = \sum_n c_n(k, t) e^{j 2\pi n x} = \sum_n c_n(k, t) e^{j 2\pi n (x + na)} = u(k, x + na). \quad (2.33)
\]

As a first conclusion, we can state that, in general, an electron propagating in a periodic potential has an eigenfunction that is equal to the eigenfunction of an electron propagating in the vacuum, multiplied by a periodic function \( u(k, x) \).

Equivalently, we can write (2.32)

\[
\Psi_k(x + na, t) = e^{j k (x + na)} u(k, x + na) = e^{j k x} e^{j k n a} u(k, x) = e^{j k n a} \Psi_k(x, t). \quad (2.34)
\]

We can generalize then to the three-dimensional lattice case, and find

\[
\Psi_k(r + R, t) = e^{j k R} \Psi_k(r, t), \quad (2.35)
\]

where \( r \) is the position vector in the three-dimensional domain, while \( R \) is the translational vector [8].

We have then demonstrated the **Bloch Theorem**: given a periodic potential \( V(r) = V(r + R) \), \( \forall \Psi, \forall R, \exists k : \Psi_k(r + R) = e^{j k R} \Psi_k(r) \).

### 2.3 Effective mass equation

Given a periodic potential, we can demonstrate [9] that periodic energy bands rise in the first Brillouin zone and : if \( E_\nu(k) \) is the \( \nu \)-th band the associated eigenfunction can be defined as

\[
\Psi_\nu k(r, t) \overset{\text{def}}{=} |\nu k\rangle, \quad (2.36)
\]
which is the exact solution of the equation
\[ \left[ -\frac{\hbar^2}{2m_0} \nabla^2 + U_L(r) \right] |\nu k\rangle = \mathcal{E}_\nu(\mathbf{k})|\nu k\rangle, \quad (2.37) \]
or equivalently
\[ \mathcal{H}_0 |\nu k\rangle = \mathcal{E}_\nu(\mathbf{k})|\nu k\rangle. \quad (2.38) \]

Let us now consider a potential \( V_{\text{ext}} \) slightly varying in time and in space, so that an electron in the \( \nu \)-th band does not transfer to another band.

We want to find the eigenfunction \( \Psi_\nu(r, t) \) that are the exact solution of
\[ (\mathcal{H}_0 + V_{\text{ext}}) \Psi_\nu(r, t) = j\hbar \frac{\partial}{\partial t} \Psi_\nu(r, t). \quad (2.39) \]

If the energy bands are not degenerate and do not cross each other, we can express \( \Psi_\nu(r, t) \) as a linear combination of \( |\nu k\rangle \):
\[ \Psi_\nu(r, t) = \sum_k f_\nu(k, t)|\nu k\rangle. \quad (2.40) \]

Substituting (2.40) into (2.39), and making use of (2.38), we obtain
\[ \sum_k f_\nu(k, t)\mathcal{E}_\nu(\mathbf{k})|\nu k\rangle + V_{\text{ext}} \Psi_\nu(r, t) = j\hbar \frac{\partial}{\partial t} \Psi_\nu(r, t). \quad (2.41) \]

Noting that \( \mathcal{E}_\nu(\mathbf{k}) \) is a periodic function in the \( k \) space, we can express it by its Fourier transform
\[ \mathcal{E}_\nu(\mathbf{k}) \overset{\text{def}}{=} \sum_R \mathcal{E}_\nu(\mathbf{R})e^{j\mathbf{R} \cdot \mathbf{k}}. \quad (2.42) \]

Substituting formally in eq. (2.42) the argument of the energy \( \mathbf{k} \) with the operator \( -j\nabla \) [10], we obtain
\[ \mathcal{E}_\nu(-j\nabla) = \sum_R \mathcal{E}_\nu(\mathbf{R})e^{\mathbf{R} \cdot \nabla}, \quad (2.43) \]
where the operator \( e^{\mathbf{R} \cdot \nabla} \) is called translational operator. Indeed for a
2.3. Effective mass equation

generic function of \( r \ f(r) \),
\[
\mathcal{E}_\nu(-j\nabla)f(r) = \sum_R \mathcal{E}_\nu(R)e^{R \cdot \nabla}f(r) \\
= \sum_R \mathcal{E}_\nu(R) \left[ 1 + R \cdot \nabla + \frac{1}{2}(R \cdot \nabla)^2 + \ldots \right] f(r) \\
= \sum_R \mathcal{E}_\nu(R) \left[ f(r) + R \cdot \nabla f(r) \right. \\
\left. + \frac{1}{2} \left\{ R^2 \frac{\partial^2 f(r)}{\partial x^2} + \ldots \right\} \right] \\
= \sum_R \mathcal{E}_\nu(R)f(r + R). \tag{2.44}
\]

Instead of a generic function \( f(r) \), we can consider in the above relation
the eigenfunction \( |\nu k\rangle \), which, for the Bloch theorem, can be expressed
as
\[
|\nu k\rangle = e^{j\mathbf{k} \cdot \mathbf{r}}u_\nu(k, r). \tag{2.45}
\]

We obtain
\[
\mathcal{E}_\nu(-j\nabla)|\nu k\rangle = \sum_R \mathcal{E}_\nu(R)e^{R \cdot \nabla}e^{j\mathbf{k} \cdot \mathbf{r}}u_\nu(k, r) \\
= \sum_R \mathcal{E}_\nu(R)e^{j\mathbf{k} \cdot (\mathbf{r} + \mathbf{R})}u_\nu(k, \mathbf{r} + \mathbf{R}) \\
= \sum_R \mathcal{E}_\nu(R)e^{j\mathbf{k} \cdot \mathbf{R}}e^{j\mathbf{k} \cdot \mathbf{r}}u_\nu(k, \mathbf{r}) \\
= \sum_R \mathcal{E}_\nu(R)e^{j\mathbf{k} \cdot \mathbf{R}}|\nu k\rangle = \mathcal{E}_\nu(k)|\nu k\rangle, \tag{2.46}
\]

where the periodicity of the function \( u_\nu(k, \mathbf{r}) \) has been exploited. Applying this relation to the first addendum of the first term of (2.41), we obtain
\[
\sum_k f_\nu(k, t)|\nu k\rangle = \sum_k f_\nu(k, t)\mathcal{E}_\nu(-j\nabla)|\nu k\rangle. \tag{2.47}
\]

The term \( \mathcal{E}_\nu(-j\nabla) \) can be put out of the summation, so (2.47) becomes
\[
\mathcal{E}_\nu(-j\nabla)\sum_k f_\nu(k, t)|\nu k\rangle = \mathcal{E}_\nu(-j\nabla)\Psi_\nu(r, t) \tag{2.48}
\]
(2.39) reduces then to the so-called *Single Band Equation*

\[ [E_\nu(-j\nabla) + V_{\text{ext}}] \Psi_\nu(r, t) = j\hbar \frac{\partial \Psi_\nu(r, t)}{\partial t}. \]  

(2.49)

If the field is weak, we can limit the study in the \( k \)-space around the minimum of the potential energy (\( k \approx k_0 \)). If we also suppose that \( u_\nu(k, r) \) is slowly varying in \( k \), we can express the eigenfunction as

\[
\Psi_\nu(r, t) = \sum_k f_\nu(k, t) e^{jk \cdot r} u_\nu(k_0, r) \approx \sum_k f_\nu(k, t) e^{jk \cdot r} u_\nu(k_0, r) = \sum_k f_\nu(k, t) e^{j(k-k_0) \cdot r} u_\nu(k_0, r) = \sum_k f_\nu(k, t) e^{j(k-k_0) \cdot r} |\nu k_0 \rangle = F_\nu(r, t) |\nu k_0 \rangle,
\]

(2.50)

where \( F_\nu(r, t) \) is the *Envelope Function*. From the single band equation, we then obtain the *Envelope Equation*

\[ [E_\nu(-j\nabla) + V_{\text{ext}}] F_\nu(r, t) = j\hbar \frac{\partial F_\nu(r, t)}{\partial t}. \]  

(2.51)

In the parabolic band approximation, the energy in correspondence of the minimum for \( k = k_0 \) can be expressed as

\[ E_\nu(k) = E_C + \frac{\hbar^2(k-k_0)^2}{2m^*} + o((k-k_0)^2), \]  

(2.52)

where \( m^* \) is the effective mass, defined as

\[ m^* = \hbar^2 \left[ \frac{\partial^2 E_\nu(k)}{\partial k^2} \right]^{-1} |_{k_0}, \]  

(2.53)

and

\[ E_C = E_\nu(k)|_{k_0}. \]  

(2.54)

We have to point out that \( m^* \) in (2.53) is a scalar only if \( E_\nu(k) \) is isotropic, i.e. \( E_\nu(k) = E_\nu(|k|) \), otherwise it is a tensor.

Formally, we can express \( E_\nu(-j\nabla) \) as

\[ E_\nu(-j\nabla) = E_C - \frac{\hbar^2 \nabla^2}{2m^*} + o(\nabla^2). \]  

(2.55)
Substituting (2.55) in the envelope equation, and not considering the $o(\nabla^2)$, we obtain the Effective Mass Equation

$$\left[ -\frac{\hbar^2 \nabla^2}{2m^*} + E_C + V_{ext} \right] F_\nu(r,t) = j \hbar \frac{\partial F_\nu(r,t)}{\partial t}.$$ (2.56)

As can be noticed, we have reduced the problem of an electron propagating in a periodic potential, to a problem of a particle with mass equal to $m^*$ which undergoes a potential equal to $E_C + V_{ext}$. Such a problem, as introduced at the beginning of this chapter, is simpler to be solved as compared to the original problem described by (2.1).

At this point, after introducing the concept of effective mass, it can be interesting to introduce, in the semiclassical picture, also the concept of velocity.

For simplicity, but without losing generality, let us consider the one dimensional case, defining the momentum operator $p = -j\hbar \frac{\partial}{\partial x}$

The Hamiltonian $\mathcal{H}$ can be expressed as

$$\mathcal{H} = \frac{p^2}{2m^*} + V,$$ (2.57)

while the derivative of $\mathcal{H}$ with respect to the momentum reads

$$\frac{\partial \mathcal{H}}{\partial p} = \frac{p}{m^*}.$$ (2.58)

Let us consider now the expectation value of the momentum operator in the $|\nu k\rangle$ state divided by the effective mass, which can be related to the semiclassical concept of velocity

$$v(\nu, k) = \langle \nu k | \frac{p}{m^*} | \nu k \rangle.$$ (2.59)

As a consequence

$$v(\nu, k) = \langle \nu k | \frac{\partial \mathcal{H}}{\partial p} | \nu k \rangle = \frac{\partial}{\partial p} \langle \nu k | \mathcal{H} | \nu k \rangle = \frac{\partial E_\nu}{\partial p}.$$ (2.60)

Defining $\hbar k$ as the quasi-momentum or crystal momentum, from (2.60), we obtain

$$v(\nu, k) = \frac{1}{\hbar} \frac{\partial}{\partial k} E_\nu,$$ (2.61)

which is the group velocity of the electron wave packet. In three dimensions (2.61) becomes

$$v(\nu, k) = \frac{1}{\hbar} \nabla k E_\nu,$$ (2.62)
where \( \nabla_k \) is the gradient operator in the \( k \) space.

The introduced quantity \( \hbar k \) is referred as quasi-momentum, because the Bloch function \( |\nu k\rangle \) is not an eigenfunction of the momentum operator. Indeed

\[
p|\nu k\rangle = -j\hbar \frac{\partial}{\partial x} \left[ e^{jkz} u_\nu(k, x) \right] = \hbar k|\nu k\rangle - je^{jkz} \hbar \frac{\partial}{\partial x} u_\nu(k, x).
\]  

(2.63)

Let us consider now the effect of an external potential on the electrons, due to the presence of a uniform force \( F \) along the direction \( r \)

\[
V_{\text{ext}} = qF \cdot r.
\]  

(2.64)

The Hamiltonian \( \mathcal{H}' \) can be expressed as

\[
\mathcal{H}' = \mathcal{H} + qF \cdot r.
\]  

(2.65)

Supposing that at the time \( t = 0 \) the electron occupies the state \( k_0 \), the generic eigenfunction can be expressed as,

\[
\Psi(r, k, t) = \exp \left( -\frac{j}{\hbar} \mathcal{H} + \frac{qF \cdot r}{\hbar} t \right) \Psi(r, k_0, 0).
\]  

(2.66)

Considering now the same wavefunction translated by \( R \), the translational vector, and making use of the Bloch theorem, we obtain

\[
\Psi(r + R, k, t) = e^{-\frac{j}{\hbar}(\mathcal{H} + qF \cdot R)t} \Psi(r + R, k_0, 0)
\]

\[
= e^{-\frac{j}{\hbar}(\mathcal{H} + qF \cdot R)t} e^{jk_0 \cdot R} e^{-j\frac{qF \cdot R}{\hbar} t} e^{jk_0 \cdot (r + R) u_\nu(r, k)}
\]

\[
= e^{jk_0 \cdot R} e^{-j\frac{qF \cdot R}{\hbar} t} \Psi(r, k, t).
\]  

(2.67)

Comparing (2.67) with (2.35), we note that they can be reduced to the same form, if we define

\[
k = -\frac{q}{\hbar} Ft + k_0.
\]  

(2.68)

If we now derive \( k \) with respect to time, we obtain

\[
\frac{\partial}{\partial t}(\hbar k) = -qF.
\]  

(2.69)

As can be noted, the quasi-momentum obeys the relation in the semiclassical picture between the semiclassical momentum and the force applied to the particle.
2.4 Quantum confined electrons

We will now focus our attention on the electron charge densities in case of one, two and three-dimensional confinement. All the relations will be then used in the self-consistent solution of the Poisson/Schrödinger equation.

2.4.1 Integration over the \( k \) space

Consider a function \( f \) defined in the \( k \)-space. The sum of \( f \) over \( k \) is

\[
I = \sum_k f(k)
\]

(2.70)

(2.70) can be written as

\[
\sum_k f(k) = \frac{L^3}{(2\pi)^3} \sum_k f(k) \frac{(2\pi)^3}{L^3},
\]

(2.71)

where \( L \) is the sample size along each direction, and \( \frac{(2\pi)^3}{L^3} \) is the volume occupied by one state in the \( k \)-space. If the available states are continuous rather than discrete, the quantity \( \frac{(2\pi)^3}{L^3} \) can be seen as an infinitesimal volume \( (\Delta k)^3 \). As a consequence, the sum becomes an integral, and (2.70) can be expressed as

\[
\sum_k f(k) = \frac{L^3}{(2\pi)^3} \sum_k f(k) (\Delta k)^3 \approx \frac{L^3}{(2\pi)^3} \int f(k)dk.
\]

(2.72)

In general, we can generalize to a generic dimensionality \( d \) and make the following substitution

\[
\sum_k f(k) \rightarrow 2 \times \frac{L^d}{(2\pi)^d} \int f(k)(dk)^d
\]

(2.73)

where we have also considered the spin degeneracy. This relation will result to be very useful in the derivation of the electron density whose general expression at the equilibrium is [11]

\[
n(r) = \sum_i f_0(E_i)|\Psi_i(r)|^2
\]

(2.74)

where the sum is performed over the single energy levels \( E_i \), and \( f_0 \) is the Fermi-Dirac distribution function.
2.4.2 One-dimensional quantum confinement

In devices such as MOSFETs or AlGaAs/GaAs heterostructure transistors, quantum confinement is strong only along one direction (the direction perpendicular to the Si/SiO$_2$ or the AlGaAs/GaAs interface).

For this reason, we can decouple the Schrödinger equation into a 1D equation in the confining direction (or “vertical” direction) and a 2D equation in the $y - z$ (or “horizontal” direction) plane: the density of states in the horizontal plane is well approximated by the semiclassical expression, since there is no in-plane confinement, while discrete states are evident in the vertical direction.

The expression of the single particle Schrödinger equation in 3D, in case of anisotropic effective mass reads

\[
-\frac{\hbar^2}{2} \left[ \frac{\partial}{\partial x} \frac{1}{m_x} \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \frac{1}{m_y} \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \frac{1}{m_z} \frac{\partial}{\partial z} \right] \Psi + V \Psi = E \Psi \tag{2.75}
\]

We can arbitrarily write the wave function $\Psi(x, y, z)$ as

\[
\Psi(x, y, z) = \psi(x, y, z) \chi(y, z), \tag{2.76}
\]

substituting (2.76) in (2.75), we obtain the following expression

\[
-\frac{\hbar^2}{2} \chi \left( \frac{\partial}{\partial x} \frac{1}{m_x} \frac{\partial}{\partial x} \psi \right) - \left[ \frac{\hbar^2}{2} \frac{\partial}{\partial y} \frac{1}{m_y} \frac{\partial}{\partial y} + \frac{\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m_z} \frac{\partial}{\partial z} \right] \psi \chi + V \psi \chi = E \psi \chi, \tag{2.77}
\]

where the dependence on $x$, $y$ and $z$ is implicit. We take $\psi$ as the solution of the Schrödinger equation along the $x$ direction

\[
-\frac{\hbar^2}{2} \frac{\partial}{\partial x} \frac{1}{m_x} \frac{\partial}{\partial x} \psi + V \psi = E_1(y, z) \psi. \tag{2.78}
\]

Equation (2.77) can be written as

\[
- \left[ \frac{\hbar^2}{2} \frac{\partial}{\partial y} \frac{1}{m_y} \frac{\partial}{\partial y} + \frac{\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m_z} \frac{\partial}{\partial z} \right] \psi \chi + \left[ \left( -\frac{\hbar^2}{2} \frac{\partial}{\partial x} \frac{1}{m_x} \frac{\partial}{\partial x} + V \right) \psi \right] \chi = E \psi \chi; \tag{2.79}
\]
by substituting (2.78) in (2.79) we obtain
\[
- \left[ \frac{\hbar^2}{2} \frac{\partial}{\partial y} \frac{1}{m_y} \frac{\partial}{\partial y} + \frac{\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m_z} \frac{\partial}{\partial z} \right] \psi \chi + E_1(y, z) \psi \chi = E \psi \chi.
\] (2.80)

Finally, if we assume that \( \psi(x, y, z) \) is weakly dependent on \( y \) and \( z \), equation (2.80) reduces to
\[
- \left[ \frac{\hbar^2}{2} \frac{\partial}{\partial y} \frac{1}{m_y} \frac{\partial}{\partial y} \right] \chi + E_1(y, z) \chi = E \chi.
\] (2.81)

Since \( E_1(y, z) \) is rather smooth in the \( y-z \) plane, we will assume that eigenvalues of Eq. (2.81) essentially obey the 2D semiclassical density of states.

Let us now show [12] in practical cases of interest, that by decoupling the 3D Schrödinger equation we introduce only a negligible error. Let us define the operator
\[
\hat{T}_{yz} = - \frac{\hbar^2}{2} \frac{\partial}{\partial y} \frac{1}{m_y} \frac{\partial}{\partial y} - \frac{\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m_z} \frac{\partial}{\partial z},
\] (2.82)
and let us call \( a(x, y, z) \) the term that we have neglected in passing from (2.80) to (2.81):

\[
a(x, y, z) \equiv \hat{T}_{yz} \psi \chi - \psi \hat{T}_{yz} \chi; \quad (2.83)
\]

if the approximation is valid \( a(x, y, z) \) must be much smaller than \( E - E_{1i}(y, z) \) in any point of the domain, which means that the parameter \( \delta \), which we define as

\[
\delta \equiv \max_{x, y, z} \left| \frac{a(x, y, z)}{[E - E_{1i}(y, z)]} \right| \psi(x, y, z) \chi(y, z)
\]

must be much smaller than 1.

Since \( \chi \) obeys the 2D density of states, it can be written as \( \chi = A e^{jk_y y + k_z z} \), where

\[
\frac{\hbar^2 k_y^2}{2m_y} + \frac{\hbar^2 k_z^2}{2m_z} = E - E_{1i} \quad (2.85)
\]

As a numerical example, we have considered the case of a silicon MOSFET with channel length equal to 50 nm and oxide thickness equal to 2.5 nm investigated in [12] considering the low lying subband.

In Fig. 2.1 we plot \( \delta \) as a function of the gate voltage for a “continuous” doping profile (solid line) and for an “atomic” doping profile (dashed line) : in both cases \( \delta \) is smaller than \( 10^{-3} \).

The eigenfunction \( \Psi \) can be then expressed as

\[
\Psi_{i,k_y,k_z}(x, y, z) = \frac{e^{jk_y y + k_z z}}{\sqrt{L_y L_z}} \psi_i(x, y, z) \quad (2.86)
\]

where \( L_y \) and \( L_z \) are the system size in the \( y \) and \( z \) direction, respectively, while \( \psi_i(x, y, z) \) is the \( i \)-th eigenfunction. Using (2.74), we obtain

\[
n(x, y, z) = \sum_i \sigma_i |\psi_i(x, y, z)|^2 \quad (2.87)
\]

where

\[
\sigma_i \overset{\text{def}}{=} \sum_{k_y, k_z} \frac{f_0(E_{i,k_y,k_z})}{L_y L_z} \quad (2.88)
\]

\[
E_{i,k_y,k_z} = \epsilon_i + \frac{\hbar^2 k_y^2}{2m_y} + \frac{\hbar^2 k_z^2}{2m_z} \quad (2.89)
\]
and $\epsilon_i$ is the $i$-th energy level in the $x$-direction. Making the following variable transformation

$$k_y' = \frac{k_y}{\sqrt{2m_y}}; \quad k_z' = \frac{k_z}{\sqrt{2m_z}}$$

(2.90)

we obtain

$$dk^2 = dk_ydk_z = 2\sqrt{m_ym_z}dk_y'dk_z'$$

(2.91)

while (2.89) becomes

$$E_{i,k_y',k_z'} = \epsilon_i + \hbar^2 k_y'^2 + \hbar^2 k_z'^2 = \epsilon_i + \hbar^2 k'^2$$

(2.92)

By means of eqs. (2.73), and (2.91), and using cylindrical coordinates [13]

$$\sum_{k_y,k_z} \rightarrow 2 \times \frac{L_yL_z}{4\pi^2} 2\sqrt{m_ym_z} \int_0^{2\pi} d\theta \int_0^\infty k'dk'$$

(2.93)

we can evaluate $\sigma_i$

$$\sigma_i = \int_0^{2\pi} d\theta \int_0^\infty \frac{\sqrt{m_ym_z}}{\pi^2} k'dk' f_0(E_{i,k_y',k_z'})$$

(2.94)

From (2.92), we can replace $dk'$ in terms of $dE_{i,k_y',k_z'}$

$$dE_{i,k_y',k_z'} = 2\hbar^2 k'dk'$$

(2.95)

and obtain

$$\sigma_i = \int_{\epsilon_i}^\infty \frac{\sqrt{m_ym_z}}{\pi\hbar^2} f_0(E_{i,k_y',k_z'}) dE_{i,k_y',k_z'}$$

(2.96)

and finally

$$\sigma_i = \frac{\sqrt{m_ym_z}K_BT}{\pi\hbar^2} \ln \left[ 1 + \exp \left( \frac{\epsilon_i - E_F}{K_BT} \right) \right]$$

(2.97)

where $K_B$ is the Boltzmann constant, $T$ is the temperature and $E_F$ is the Fermi level.

The electron concentration can be then expressed as

$$n(x,y,z) = \sum_i |\psi_i(x,y,z)|^2 \frac{\sqrt{m_ym_z}K_BT}{\pi\hbar^2} \ln \left[ 1 + \exp \left( \frac{\epsilon_i(x,y,z) - E_F}{K_BT} \right) \right]$$

(2.98)
2.4.3 Two-dimensional quantum confinement

The considerations in the previous section, similarly apply when quantum confinement is predominant in the $y-z$ plane. The two-dimensional Schrödinger equation reads,

$$
-\left[ \frac{\hbar^2}{2} \frac{\partial}{\partial y} \frac{1}{m_y} \frac{\partial}{\partial y} + \frac{\hbar^2}{2} \frac{\partial}{\partial z} \frac{1}{m_z} \frac{\partial}{\partial z} \right] \chi_i
= E_{i,k_x} \chi_i.
$$

(2.99)

while the three-dimensional eigenfunctions can be expressed as

$$
\Psi_{k_x,i} = \frac{e^{jk_x x}}{\sqrt{L_x}} \chi_i(y,z)
$$

(2.100)

where $\chi_i(y,z)$ are the eigenfunctions associated to the energy level $\epsilon_i(x)$ and the total energy reads

$$
E_{i,k_x} = \epsilon_i(x) + \frac{\hbar^2 k_x^2}{2m_x}
$$

(2.101)

The electron concentration is

$$
n(x, y, z) = \sum_i |\chi_i(x, y, z)|^2 \alpha_i
$$

(2.102)

where

$$
\alpha_i = \sum_{k_x} \frac{f_0(E_{i,k_x})}{L_x}
$$

(2.103)

Using (2.73),

$$
\alpha_i = \frac{2L_x}{2\pi} \int_{-\infty}^{+\infty} \frac{f_0(E_{i,k_x})}{L_x} dk_x = \frac{2}{\pi} \int_{0}^{+\infty} \frac{f_0(E_{i,k_x})}{L_x} dk_x
$$

(2.104)

Expressing $k_x$ as

$$
k_x = \frac{\sqrt{2m_x(E - \epsilon_i)}}{\hbar}
$$

(2.105)

and deriving with respect to $E$ we obtain

$$
dk_x = \frac{\sqrt{2m_x}}{2\hbar\sqrt{E - \epsilon_i}} dE
$$

(2.106)
Substituting (2.106) in (2.104), we obtain

$$
\alpha_i = \frac{2m_x}{\pi \hbar} \int_{\epsilon_i}^{\infty} \frac{1}{\sqrt{(E - \epsilon_i)}} \frac{1}{1 + e^{(E - \epsilon_i)/k_B T}} dE
$$

(2.107)

and multiplying and dividing by $\sqrt{K_B T}$, we obtain the final expression for the electron density

$$
n(y, z) = \frac{1}{\pi} \left( \frac{2m_x K_B T}{\hbar^2} \right)^{\frac{1}{4}} \sum_i |\chi_i(x, y, z)|^2 \mathcal{F}_{-\frac{1}{2}} \left( \frac{E_F - \epsilon_i(x)}{K_B T} \right)
$$

(2.108)

where $\mathcal{F}_{-\frac{1}{2}}$ is the Fermi integral of order $-\frac{1}{2}$:

$$
\mathcal{F}_{-\frac{1}{2}}(y) = \int_0^\infty \frac{x^{-\frac{1}{2}}}{1 + e^{y-x}} dx
$$

(2.109)

### 2.4.4 Three-dimensional quantum confinement

For what concerns three-dimensional quantum confinement, (2.74) reduces to

$$
n(r) = \sum_i |\Psi_i(r)|^2 \frac{1}{1 + e^{(E_i - E_F)/k_B T}}
$$

(2.110)

where $\Psi_i$ is the orbital associated to the $i$-th eigenvalue, with energy $E_i$. However for a system with very few electrons, we cannot define a Fermi level, thence it is improper to fix the Fermi energy.

A more proper approach is then represented by fixing the number of electrons in the system and using in (2.74) the Gibbs statistic. However the Gibbs statistics in the grand canonical ensemble is rather complicated, so it is worth following a simpler approach represented by considering the ground-state of the confined region and filling the energy levels starting from the lowest.

The electron density, can then be expressed as

$$
n(r) = 2 \sum_{i=1}^m |\Psi_i(r)|^2 + (N - 2m)|\Psi_{m+1}(r)|^2,
$$

(2.111)

where $m = \lfloor \frac{N}{2} \rfloor$ is the number of fully occupied single electron levels.
2.5 3D Poisson/Schrödinger equation

As introduced at the beginning of this chapter, as dimensions becomes smaller and smaller, quantum confinement effects becomes relevant. A proper simulation approach requires thence solving the Schrödinger equation together with the Poisson equation, in a self-consistent iterative scheme. However, at nanoscale dimensions, not only quantum confinement affects device behavior, but also other effects whose nature is purely electrostatic, limit the performance of the device. We refer for example to the charge sharing effects that lower the threshold voltage in MOSFET devices, or to the random distribution of dopants that can affect the conductivity in the channel.

In order to take into account this wide variety of effects, a three-dimensional simulation approach has to be followed.

In particular the Poisson equation in the three-dimensional simulation domain reads

\[
\nabla \left[ \varepsilon(r) \nabla \phi(r) \right] = -q \left[ p(r) - n(r) + N_D^+(r) - N_A^-(r) + \rho_{\text{fix}}(r) \right], \tag{2.112}
\]

where \( \phi \) is the electrostatic potential, \( \varepsilon \) is the dielectric constant, \( p \) and \( n \) are the hole and electron densities, respectively, \( N_D^+ \) is the concentration of ionized donors, \( N_A^- \) is the concentration of ionized acceptors, and \( \rho_{\text{fix}} \) is the fixed charge concentration.

In all the simulations that we have performed (which will be shown in the following chapters), we have considered quantum confinement of the electrons, while a semiclassical expression has been assumed for the other densities.

For what concern the exchange term \( V_{\text{exc}} \), we have considered the expression in [4]

\[
V_{\text{exc}} = -\frac{q^2}{4\pi^2\varepsilon_0 \varepsilon_r} \left[ 3\pi^2 n(r) \right]^{\frac{4}{3}}, \tag{2.113}
\]

which differ by the exchange potential found by Slater by a factor of \( \frac{2}{3} \).

We have found that the above expression, valid for a 3D system, has an impact on the results only inside the confined regions, while it turns out that exchange corrections to the potential are negligible in the other regions of the analyzed structures.

From a numerical point of view, the coupled Poisson and Schrödinger equations are solved by means of a Newton-Raphson method with a predictor-corrector scheme [14], that we have verified to be faster and
more stable than a code previously developed by our group, based on the multigrid algorithm [15].

Numerical details follows in Appendix A, where ViDES (Versatile DEvice Simulator), which implements all the previously explained models, is presented.
The development of precision growth and lithography techniques has led to the realization of nanoscale structures, in which the dimensions are smaller than the characteristic phase coherence length of the electrons.

By means of these new techniques, a wide variety of structures can be envisioned and studied, not only to reproduce experimentally effects demonstrated theoretically, but also to find alternative device architectures. In such a picture, simulations play an important role, since they can represent an indispensable tool for both the design, and analysis of the experimental data.

In this chapter, we will focus on devices realized on AlGaAs/GaAs heterostructures defined by split-gates. In particular, we will first analyze Quantum Point Contacts (QPCs), showing the quantization of the conductance and validating a simple model for surface states and, in the last section, we will study a Single Electron Transistor (SET),
Chapter 3

Quantum point contact and single electron devices

considering the transfer of electrons from the reservoirs to the inner three-dimensional quantum confined region.

3.1 The Landauer Formula

Let us consider the following system, composed by a one-dimensional channel (a quantum wire) with length $L$, and two metallic reservoirs with electrochemical potential $\mu_L$ and $\mu_R$ ($\mu_L > \mu_R$) (Fig. 3.1). We suppose that only the first one-dimensional subband is occupied, that the electrons in the channel does not suffer any scattering mechanism, i.e. transport in the channel is ballistic, and that the electrons entering the reservoirs contacts are instantaneously in equilibrium with them. For the moment, let the temperature be equal to zero ($T=0$ K) and the contacts reflectionless, that means that the transmission probability from the contact to contact is unitary. Being a two-dimensional confined system, the dispersion relation is equal to (2.101), where $\epsilon_i$ is also referred as the cut-off energy of $i$-th 1D subband or transversal mode. The positive current, carried by $k > 0$ states in the $i$-subband, reads [16]

$$I_+^i = \frac{q}{L} \sum_{k>0} v_i(k) f(E - \mu_L) \quad (3.1)$$

where $v_i$ is the velocity and $f$ is the Fermi-Dirac distribution function. From (2.61) we obtain

$$I_+^i = \frac{q}{\pi} \int_{\infty}^{0} v_i(k) f(E - \mu_L) dk \quad (3.2)$$

and from (2.73), we can express (3.1) as

$$I_+^i = \frac{2q}{\hbar} \int_{\epsilon_i}^{\infty} f(E - \mu_L) dE \quad (3.3)$$
where \( h \) is the Planck constant. Equivalently, the negative current reads
\[
I_i^< = \frac{2q}{h} \int_{\epsilon_i}^{\infty} f(E - \mu_R) dE \tag{3.4}
\]
If we now define the function \( M(E) \) as
\[
M(E) = \sum_i u(E - \epsilon_i), \tag{3.5}
\]
where \( u(E) \) is the Heavyside function, we can express the total positive current as
\[
I^+ = \sum_i I_i^+ = \sum_i \frac{2q}{h} \int_{\epsilon_i}^{\infty} f(E - \mu_L) dE
\]
\[
= \frac{2q}{h} \int_{-\infty}^{+\infty} f(E - \mu_L) M(E) dE \tag{3.6}
\]

Working at \( T = 0 \) K, the Fermi-Dirac function is a step function, and considering \( M(E) \) constant and equal to an integer \( \mathcal{M} \) between the energy range \( [\mu_R, \mu_L] \), the total current reads
\[
I = I^+ - I^< = \frac{2q^2}{h} \mathcal{M} \frac{\mu_L - \mu_R}{q} \tag{3.7}
\]

We note that (3.7) can be reduced to the form
\[
I = GV \tag{3.8}
\]
where \( V \) is the applied potential \( (\mu_L - \mu_R)/q \) and \( G \) is the conductance, that in this case is
\[
G = \frac{2q^2}{h} \mathcal{M} \tag{3.9}
\]

Since the channel is purely ballistic, the resistance of the channel is equal to zero. This means that all the dissipation of the energy happens in the reservoirs, since a charge propagating from the left transverse the channel conserving the energy, and when it reaches the right reservoir, it loses energy thermalizing in the contact. That’s why the inverse of \( G \) is often referred as the contact resistance.

If we now relax the hypothesis of reflectionless contacts, we have to consider in (3.6) \( T(E) \), the probability that an electron originated by one electrode reaches the other reservoir propagating through the channel.
Chapter 3. Quantum point contact and single electron devices

Figure 3.2: Electrons propagating from the left contact undergo elastic scattering and $T$ is the probability of reaching the other reservoir.

If $I_L^>$ is the influx of electrons from the left reservoir, $I_L^<$ is the flux of back-scattered electrons in the left reservoir, and $I_R^>$ is the flux of electrons that has reached the right reservoir (Fig. 3.2), if we consider $T(E)$ constant and equal to $\tau$ between the energy range $[\mu_R, \mu_L]$, we deal with the following relations

$$I_L^> = \frac{2e}{h} M(\mu_L - \mu_R) \quad (3.10)$$

$$I_L^< = \frac{2e}{h} M(1 - \tau)(\mu_L - \mu_R) \quad (3.11)$$

$$I_R^> = \frac{2e}{h} M\tau(\mu_L - \mu_R) \quad (3.12)$$

The total current $I$ is then equal to

$$I = I_L^> - I_L^< = I_R^> = \frac{2e^2}{h} M\tau \frac{\mu_L - \mu_R}{q} \quad (3.13)$$

and the conductance in case of reflecting contacts can be expressed as

$$G = \frac{2e^2}{h} M\tau \quad (3.14)$$

that is the Landauer Formula.

We can now study the case in which the temperature is not equal to zero. The total current now reads

$$I = \frac{2e}{h} \int_{-\infty}^{+\infty} [f(E - \mu_L) - f(E - \mu_R)] M(E)T(E)dE \quad (3.15)$$
3.2 Simulation of Quantum Point Contacts

If \( \mu_R = \mu \) and \( \mu_L = \mu + \delta \mu \), with \( \delta \mu \ll \mu \), (3.15) becomes

\[
I = \frac{2q}{h} \int_{-\infty}^{+\infty} \left[ f(E - (\mu + \delta \mu)) - f(E - \mu) \right] M(E)T(E)dE \quad (3.16)
\]

and by means of the Taylor expansion

\[
f(E - (\mu + \delta \mu)) - f(E - \mu) = -\delta \mu \frac{\partial f(E, \mu)}{\partial E} \quad (3.17)
\]

we can express (3.16) as

\[
I = \frac{2q^2}{h} \left[ \int_{-\infty}^{+\infty} -\frac{\partial f(E, \mu)}{\partial E} M(E)T(E)dE \right] \frac{\delta \mu}{q} \quad (3.18)
\]

that can be reduced to the form expressed in (3.8), if we define

\[
G(\mu) = \frac{2q^2}{h} \int_{-\infty}^{+\infty} -\frac{\partial f(E, \mu)}{\partial E} M(E)T(E)dE \quad (3.19)
\]

where \(-\frac{\partial f(E, \mu)}{\partial E}\) is often referred as the broadening function.

As a numerical example, we can consider the case of a QPC with reflectionless contacts in which the transversal potential is a harmonic potential [17]. The transversal modes are equal to

\[
E_{tn} = (n - \frac{1}{2}) \hbar \omega_0. \quad (3.20)
\]

In Fig. 3.3 we show the conductance at different temperatures, as a function of the electrochemical potential of the reservoir (\( \mu \)), supposing a small voltage \( \delta \mu \) is applied between the two reservoir, and \( \hbar \omega_0 = 0.1 \text{ eV} \). As can be noted, as the temperature is increased, the step-like function of the conductance at \( T = 0 \text{ K} \) is smoothed. Such an effect is due to the fact that in (3.19), the broadening function “broadens” over the energy as the temperatures increases (Fig. 3.4) : Equation (3.19) can be then seen as an average over a window that increases as the temperature is increased.

### 3.2 Simulation of Quantum Point Contacts

We will now investigate the effect of surface states on transport properties of quantum point contacts, through experiments and simulations.
Chapter 3. Quantum point contact and single electron devices

Figure 3.3: Conductance of a QPC with reflectionless contact at different temperatures, in which the transversal potential has been approximated with a harmonic potential ($\hbar \omega_0 = 0.1 \text{ eV}$).

Figure 3.4: Broadening function computed at different temperatures. At $T=0$ K the broadening function is a delta-Dirac function with area equal to one.
3.2. Simulation of Quantum Point Contacts

on the simplest devices exhibiting mesoscopic transport, i.e., quantum point contacts (QPCs) defined by split gates on an AlGaAs/GaAs heterostructure.

Indeed, in order to obtain strong electron confinement and, consequently, to observe the transport properties typical of the mesoscopic regime at higher temperatures, the two dimensional electron gas must be very close to the surface.

In the case of devices defined electrostatically by metal gates evaporated on the semiconductor surface, stronger confinement is obtained if the heterointerface where the two-dimensional electron gas is located is at a small distance from the surface. On the other hand, this implies that surface properties strongly affect device behavior.

From the point of view of fabrication, controlling the surface properties is crucial for the reproducibility of device characteristics. On the other hand, as far as device modeling is concerned, proper treatment of the surface is necessary to perform quantitatively accurate simulations.

Surface states are usually treated with drastic approximations, such as Fermi level pinning or the assumption of a frozen charge at the air-semiconductor interface [18, 19, 20]. However, these approximations are unable to reproduce the experimental results [21, 22]. In the following sections, a slightly more complex model, typically used for metal-semiconductor contacts, is shown to be adequate for reproducing the experimental pinch-off voltages.

3.2.1 Experiment

Quantum point contacts defined by split gates on AlGaAs/GaAs modulation doped heterostructures have been fabricated and characterized by means of transport measurements at 0.3 K at the University of Würzburg. Samples were grown by Molecular Beam Epitaxy, with the layer structure shown in Fig. 3.5, consisting of an undoped GaAs substrate, an undoped 12 nm spacer layer of Al$_{0.2}$Ga$_{0.8}$As, a 31 nm layer of doped GaAs (donor concentration $N_D \approx 10^{18}$ cm$^{-3}$), and an undoped 9 nm GaAs cap layer. The Hall mobility of the structure is $1 \times 10^6$ cm$^2$/Vs with a carrier concentration of $3.7 \times 10^{11}$ cm$^{-2}$ at 4.2 K in the dark.

UV lithography was used to expose the Hall bar and the contact pads on top of the high mobility heterostructure. Wet etching was used to remove 200 nm and to separate electrically 12 single devices for each sample. In a second step the contacts were realized by a contact mask and evaporation of Au/Ge/Ni layers followed by alloying. High resolu-
Chapter 3. Quantum point contact and single electron devices

Figure 3.5: Layer structure of the AlGaAs/GaAs heterostructure on which the quantum point contacts are defined

<table>
<thead>
<tr>
<th>Cap layer</th>
<th>9nm GaAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doped layer (Si)</td>
<td>31 nm</td>
</tr>
<tr>
<td>( \text{Al}<em>{0.2} \text{Ga}</em>{0.8} \text{As} )</td>
<td></td>
</tr>
<tr>
<td>Spacer</td>
<td>12 nm</td>
</tr>
<tr>
<td>( \text{Al}<em>{0.2} \text{Ga}</em>{0.8} \text{As} )</td>
<td></td>
</tr>
<tr>
<td>500 nm GaAs</td>
<td></td>
</tr>
</tbody>
</table>

Electron beam lithography at 100 kV was used to define masks for the split gates.

Three different gate layouts have been defined, corresponding to a nominal lithographic gap of 50 nm, 100 nm, and 150 nm, respectively. For each gate layout, several samples have been realized and characterized by standard lock-in techniques, in order to evaluate the dispersion of the transport characteristics of the QPCs. A SEM picture of the gate layout of a QPC with a nominal gap of 100 nm is shown in Fig. 3.6; the actual lithographic gap, measured from the SEM image, is 112 nm. For the simulations described in the following sections, we consider an “actual” gate layout of the split gate structure, extracted from the SEM image, for each of the three nominal layouts considered.

Various structures exhibit conductance quantization at integer multiples of \( 2q^2/h \), as indication of ballistic transport of electrons through a quantum point contact defined by depletion of electrons beneath the metallic gates. For negative gate voltages exceeding the pinch-off voltage, the conductance increases until reaching a plateau at \( 2q^2/h \), which corresponds to the population of the first subband in the constriction [23, 24]. With increasing gate voltage further subbands are occupied, which re-
3.2 Simulation of Quantum Point Contacts

Figure 3.6: SEM picture of the gate layout of one of the considered QPCs with actual lithographic gap of 112 nm

results in a step like function of the conductance. The average pinch-off voltage and its standard deviation for the three gate layouts considered are reported in Table 3.1.

3.2.2 Surface states

Surface states can strongly influence the behavior of devices defined on a 2DEG in proximity of the surface. We assume a model for surface states typically applied to metal-semiconductor interfaces [18] and based on two parameters: an effective work function $\Phi^*$ and an uniform density of surface states per unit area per unit energy $D_s$, that are properties of the semiconductor surface and depend on the fabrication process. The effective work function, as shown in Fig. 3.7, represents the difference between the Fermi level at the surface and the vacuum energy $E_0$ when the surface charge density is zero. We make the additional assumption that all surface states below $\Phi^*$ behave as donors and all surface states above $\Phi^*$ behave as acceptors. The surface charge density due to surface states can be therefore expressed as

$$\rho_{ss} = q D_s [E_0 - \Phi^* - E_F]$$  (3.21)

If we assume that surface states are very effective in shielding the electric field, so that the electric field in the air is zero, we can limit our
Chapter 3. Quantum point contact and single electron devices

<table>
<thead>
<tr>
<th>structure type</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominal gap (nm)</td>
<td>50</td>
<td>100</td>
<td>150</td>
</tr>
<tr>
<td>actual gap (nm)</td>
<td>57</td>
<td>112</td>
<td>140</td>
</tr>
<tr>
<td>number of samples</td>
<td>14</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>average $V_P$</td>
<td>-456</td>
<td>-665</td>
<td>-1219</td>
</tr>
<tr>
<td>(Experiments (mV))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard deviation</td>
<td>81</td>
<td>52</td>
<td>61</td>
</tr>
<tr>
<td>(Experiments (mV))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V_P$ (simulations (mV))</td>
<td>-465</td>
<td>-656</td>
<td>-1198</td>
</tr>
</tbody>
</table>

Table 3.1: Experimental and theoretical values of the pinch-off voltage for several quantum point contacts with three different nominal lithographic gaps [25]

<table>
<thead>
<tr>
<th>Temperature (K)</th>
<th>Conductance [Units of $2^2/h$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>$2.36 \times 10^{-10}$</td>
</tr>
<tr>
<td>77</td>
<td>$4.29 \times 10^{-7}$</td>
</tr>
<tr>
<td>50</td>
<td>$5.6 \times 10^{-4}$</td>
</tr>
<tr>
<td>35</td>
<td>0.0132</td>
</tr>
<tr>
<td>25</td>
<td>0.065</td>
</tr>
<tr>
<td>20</td>
<td>0.118</td>
</tr>
<tr>
<td>15</td>
<td>0.186</td>
</tr>
<tr>
<td>10</td>
<td>0.254</td>
</tr>
<tr>
<td>5</td>
<td>0.299</td>
</tr>
<tr>
<td>4.2</td>
<td>0.299</td>
</tr>
</tbody>
</table>

Table 3.2: Conductance simulated at different temperatures for a single quantum point contact. $V_G=-0.45 \text{ V}$.
simulation domain to the semiconductor region and apply the following boundary condition at the air-semiconductor interface [25], : 
\[ \vec{E} \cdot \vec{n} = \frac{q}{\varepsilon} D_s |E_0 - \Phi^* - E_F|, \]  
(3.22)
where \( \vec{E} \) is the electric field in the semiconductor, and \( \vec{n} \) is the unit vector normal to the semiconductor surface.

If \( D_s \to \infty \), the Fermi level at the surface is pinned at the level \( E_0 - \Phi^* \) and the surface practically behaves as a metallic layer with work function \( \Phi^* \).

The parameters of the surface state model and the concentration of donors in the doped layer have been extracted from measurements on purposely fabricated test structures [22]: \( \Phi^* = 4.85 \) eV, \( D_S = 5 \times 10^{12} \) cm\(^{-2} \) eV\(^{-1} \).

3.2.3 Simulation results

For our simulation, we have used the actual gate layout extracted from an SEM picture (for example the one shown in Fig. 3.6). Once the 3D Poisson-Schrödinger equation is solved, we extract the first subband profile \( \tilde{E}_1(y, z) \), which is the only populated subband, and compute the conductance of the quantum point contact with a method based on recursive Green’s functions [26].
Convergence problems arise as the temperature is close to zero, because of the discontinuity of the Fermi-Dirac distribution. For this reason we performed our simulations at the temperature of 4.2 K (measurements have been performed at 0.3 K). To achieve convergence at 4.2 K, we have to perform a “cooling” procedure: we start with a simulation at 100 K, and progressively decrease the temperature down to 4.2 K. Actually, the conductance almost saturates when the temperature of 4.2 K is reached, as can also be seen in Table 3.2, therefore we reasonably compare the pinch-off voltages computed at 4.2 K with the experimental values at 0.3 K.

We have noticed that the pinch-off voltage is strongly dependent on the donor concentration in the doped layer $N_D$, which is known from experiments with insufficient accuracy. For this reason, we have considered $N_D$ as a fitting parameter in order to reproduce the experimental pinch-off voltages of QPCs with different lithographic gaps. The best fit is provided by $N_D = 0.8 \times 10^{18}$ cm$^{-3}$. The electron concentration in the 2DEG is $4 \times 10^{11}$ cm$^{-2}$, close to the experimental value.

In Table 3.1 theoretical threshold voltages are compared with the experimental values, exhibiting a very good agreement. Theoretical $G-V$ curves of QPCs with actual lithographic gap of 57, 112 and 140 nm, corresponding to a nominal gap of 50, 100, and 150 nm, respectively, are shown in Fig. 3.8.
3.3 Single Electron Box

In order to investigate the dependence of conductance on the parameters of the model for surface states, we have computed the $G$-$V$ curve for $\Phi^*$ ranging from 4.8 to 4.9 eV (Fig. 3.9) and $D_S$ ranging from $2.5 \times 10^{16}$ to $10^{17}$ m$^{-2}$eV$^{-1}$ (Fig. 3.10). The pinch-off voltage depends significantly on both parameters: an accurate evaluation of their values is therefore mandatory to perform quantitative simulations of split gate structures.

### 3.3 Single Electron Box

We now focus on the derivation of the charging energy of the Single Electron Box device, which will be very useful for the analysis of Single Electron Transistors.

The Single Electron Box device is composed by an isolated metallic island (BOX) which is coupled via a tunnel junction with capacitance $C_2$ to an electrode and via a capacitance $C_1$ to a voltage source $V_G$ (Fig. 3.11).

The charging energy (or free energy) $\mathcal{F}$ is defined as the energy necessary to add $N$ electrons in the BOX, and it is equal to

$$\mathcal{F} = E_S - \mathcal{L},$$

where $E_S$ is the system energy and $\mathcal{L}$ is the work done by external forces.

Figure 3.9: Simulated conductance as a function of gate voltage with varying $\Phi^*$ for the QPC with actual lithographic gap of 112 nm
Figure 3.10: Simulated conductance as a function of gate voltage for $D_s$ equal to $2.5 \times 10^{16}$, $5 \times 10^{16}$ and $1 \times 10^{17} \text{ m}^{-2} \text{eV}^{-1}$. The QPC considered has an actual lithographic gap of $a = 112$ nm.

to the system.

If $n$ electrons are present in the BOX, the following equations are verified

$$\begin{align*}
q_2 - q_1 &= -nq \\
V_G &= \frac{q_1}{C_1} + \frac{q_2}{C_2}
\end{align*}$$

(3.24)

where $q_1$ and $q_2$ are the charges on the $C_1$ and $C_2$ positive capacitor plates, respectively, and $q$ is the elementary unit charge. If we now define

$$C \overset{\text{def}}{=} C_1 + C_2,$$

(3.25)

we obtain

$$\begin{align*}
q_2 &= q_1 - nq \\
C_1C_2V_G &= C_2q_1 + C_1q_2
\end{align*}$$

(3.26)

and further

$$\begin{align*}
q_1 &= \frac{C_2C_2V_G + nqC_1}{C} \\
q_2 &= \frac{C_1C_2V_G - nqC_2}{C}
\end{align*}$$

(3.27)
Figure 3.11: Equivalent circuit of the Single Electron Box: $C_2$ is the tunnel capacitance.

The system energy $E_S$ can be then computed as follows

$$E_S = \frac{q_1^2}{2C_1} + \frac{q_2^2}{2C_2} = \frac{(C_1C_2V_G)^2 + C_1^2(nq)^2 + 2C_1C_2V_Gnq}{2C_1C^2} + \frac{(C_1C_2V_G)^2 + C_2^2(nq)^2 - 2C_2C_1C_2V_Gnq}{2C_2C^2} \quad (3.28)$$

$$= \frac{C_2(C_1C_2V_G)^2 + C_2C_1^2(nq)^2 + 2C_1C_2C_1C_2V_Gnq}{2C_1C_2C^2} + \frac{C_1(C_1C_2V_G)^2 + C_1C_2^2(nq)^2 - 2C_1C_2C_1C_2V_Gnq}{2C_1C_2C^2} \quad (3.29)$$

$$= \frac{(C_1C_2V_G)^2(C_1 + C_2) + C_1C_2(nq)^2(C_1 + C_2)}{2C_1C_2C^2} = \quad (3.30)$$

$$= \frac{C_1C_2V_G^2 + (nq)^2}{2C}. \quad (3.31)$$

while the external work $\mathcal{L}$,

$$\mathcal{L} = q_1V_G. \quad (3.32)$$

As a consequence the free energy reads

$$\mathcal{F}(n,V_G) = \frac{C_1C_2V_G^2 + (nq)^2}{2C} - \frac{C_1C_2V_G + nqC_1}{C}V_G \quad (3.33)$$

$$= -\frac{C_1C_2V_G^2 - 2C_1nqV_G + (nq)^2}{2C}. \quad (3.34)$$
The number of electrons in the BOX at the equilibrium, can be found imposing
\[ \frac{\partial F}{\partial n} = 0. \] (3.35)
Substituting (3.34) in (3.35), we obtain
\[ n = \frac{C_1 V_G}{q} \] (3.36)
from which we can compute the value of \( F \) at the equilibrium
\[ F = -\frac{C_1 C_2 V_G^2 - C_1^2 V_G^3}{2 C}. \] (3.37)
Since \( F \) is a potential, we can refer it with respect to an arbitrary potential, so that \( F \) is equal to zero at the equilibrium. In the end we obtain,
\[ F(n, V_G) = \frac{(C_1 V_G - nq)^2}{2 C}. \] (3.38)

### 3.4 SET Capacitance extraction

The Single Electron Transistor device is composed by a three-dimensional quantum confined region (dot), which is coupled by tunneling junctions to the reservoirs, and by capacitances to the other gates, which determine the number of electrons in the dot.

Using the above results, we can extract the capacitances in the equivalent circuit of the single electron transistor (SET) depicted in Fig. 3.12: \( V_S \), \( V_D \) and \( V_G \) are the source, drain and gate voltage source, respectively. \( C_S \) and \( C_D \) are the tunnel capacitances, while \( C_G \) and \( C_m \) are the capacitance between the dot, the gate and the back-gate, respectively.

As shown in the previous section, the free energy can be expressed as
\[ F = \frac{(-nq + Q_0)^2}{2C_{TOT}}, \] (3.39)
where
\[ Q_0 \overset{\text{def}}{=} C_G V_G + C_S V_S + C_D V_D \] (3.40)
\[ C_{TOT} \overset{\text{def}}{=} C_S + C_D + C_G + C_m \] (3.41)
(3.42)
and \( n = N - N_0 \), \( N \) is the number of the electrons in the dot and \( N_0 \) is the number of the electrons in the dot when \( V_G = 0 \) V. For simplicity we assume \( N_0 = 0 \).

In the total free energy expression we have also to consider the quantum confinement of the electrons in the dot [27], so

\[
\mathcal{F} = \sum_{i=1}^{N} E_i + \frac{(-qN + Q_0)^2}{2C_{TOT}},
\]

(3.43)

where \( E_i \) is the energy of \( i \)-th electron in the dot, relative to the bottom of the conduction band. The chemical potential \( \mu(N) \), defined as the energy needed to add one electron to the dot, can be expressed as

\[
\mu(N) = \mathcal{F}(N) - \mathcal{F}(N - 1),
\]

(3.44)
Substituting (3.43) in (3.44), we obtain

\[
\mu(N) = \sum_{i=1}^{N} E_i + \frac{[-qN + Q_0]^2}{2C_{TOT}} - \sum_{i=1}^{N-1} E_i - \frac{[-q(N-1) + Q_0]^2}{2C_{TOT}} = E_N + \frac{Q_0^2 - 2qNQ_0 + (Nq)^2}{2C_{TOT}} - \frac{[Q_0^2 + q^2(N^2 - 2N + 1) - 2Q_0q(N - 1)]}{2C_{TOT}} = E_N - \frac{Q_0q}{C_{TOT}} + \frac{(N - \frac{1}{2})q^2}{C_{TOT}},
\]

(3.45)

From (3.45), once computed the chemical potential as a function of the electrons in the dot and of the applied voltages, it is possible to derive all the capacitances in the circuit of Fig. 3.12 [28].

For example, let us consider the chemical potential as a function of the gate voltage \( V_G \) and of the number of electron \( N \), fixed \( V_S = V_D = 0 \) V (Fig. 3.13). Fixing the gate voltage \( V_G \), we can compute the quantity

\[
\Delta \mu = \mu(N) - \mu(N-1)|_{V_G} = E_N - E_{N-1} + \frac{q^2}{C_{TOT}}|_{V_G}.
\]

(3.46)

Considering a SET defined on a AlGaAs/GaAs heterostructure, each eigenstate has a degeneracy equal to two, so, for each couple of electrochemical potential, \( E_N = E_{N-1} \) and (3.46) can be reduced as

\[
\Delta \mu = \frac{q^2}{C_{TOT}}|_{V_G},
\]

(3.47)

from which we can compute the total capacitance

\[
C_{TOT} = \frac{q}{\Delta \mu |_{V_G}},
\]

(3.48)

if we express \( \Delta \mu \) in eV.

If instead of fixing \( V_G \), we fix \( \mu \), we can extract \( C_G \). Indeed,

\[
\mu(N, V_G) = E_N + \frac{(N - \frac{1}{2})q^2}{C_{TOT}} - q \frac{C_G}{C_{TOT}} V_G,
\]

(3.49)
3.4. SET Capacitance extraction

\begin{align*}
\mu(N+1, V_G + \Delta V_G) &= E_{N+1} + \frac{(N + \frac{1}{2}) q^2}{C_{TOT}} \\
&\quad - q \frac{C_G}{C_{TOT}} V_G - q \frac{C_G}{C_{TOT}} \Delta V_G. 
\end{align*}

(3.50)

Imposing $\mu(N, V_G) = \mu(N + 1, V_G + \Delta V_G)$ we obtain

$$
\Delta V_G = \left. \frac{q}{C_G} \right|_{\mu},
$$

(3.51)

and finally

$$
C_G = \left. \frac{q}{\Delta V_G} \right|_{\mu}.
$$

(3.52)

$C_S (C_D)$, can be computed once extracted $C_{TOT}$ and $C_G$, if we consider the elettrochemical potential as a function of the electrons in the dot and of the source (drain) voltage, once fixed the other voltages. Deriving (3.45) with respect to $V_S$, we obtain

$$
\frac{\partial \mu(N)}{\partial V_S} = \frac{\partial E_N}{\partial V_S} \bigg|_{a} + \frac{\partial}{\partial V_S} \left[ \frac{(N + \frac{1}{2}) q^2}{C_{TOT}} \right] \bigg|_{b} - q C_S \frac{C_G}{C_{TOT}}. 
$$

(3.53)
The term $b$ in (3.53) is zero since we have fixed the number of electrons in the dot. If the capacitative coupling between the reservoir and the dot is small, the term $a$ can be reasonably considered equal to zero, so (3.53) reduces to

$$\frac{\partial \mu(N)}{\partial V_S} = -q \frac{C_S}{C_{TOT}},$$

(3.54)

and the dot-to-source capacitance reads

$$C_S = -\frac{C_{TOT}}{q} \frac{\partial \mu(N)}{\partial V_S}.$$  

(3.55)

The electrochemical potential can be determined in a simple way by means of the Slater formula [5], imposing the number of the electrons $N$ in the dot. Indeed, the electrochemical potential $\mu(N)$, can be expressed as

$$\mu(N) = E(N) - E(N - 1) = \varepsilon \left( N - \frac{1}{2} \right),$$

(3.56)

where $E$ is the total energy of the dot, while $\varepsilon$ is the energy of the half-occupied highest Kohn-Sham orbital of a system with $N - \frac{1}{2}$ electrons.

### 3.5 Single Electron Transistor

The improvement of fabrication techniques in the last couple of decades has made possible the realization and characterization of devices capable of exploiting the discreteness of the electron charge. Single Electron Transistors (SET) represent the prototype of such type of devices, allowing to control a current consisting of electrons traversing the device one by one, by varying the charge on the control gate by a fraction of an electron [29]. While SETs are extremely and excessively sensible to the presence of single charged defects, impressive reproducibility have been recently obtained through refined fabrication techniques [30, 31].

In simulating this kind of devices, it is necessary to take into account different kind of quantum confinement, typically three dimensional in the dot region, two dimensional (2D) in the wire connecting the dot to reservoirs, and one dimensional (1D) in the two dimensional electron gas (2DEG) reservoirs.

In this section we focus our attention on a single electron transistor defined by split gates on an AlGaAs/GaAs modulation doped heterostructure.
We consider such device as a typical test bench structure, due to the popularity of structures defined by gate voltages. Further, it presents at the same time two kinds of different quantum confinements since there is 3D confinement inside the dot and 1D confinement the 2D electron gas at the interface. From a physical point of view, such SET is interesting since it is characterized by a variable size of the dot depending on the number of electrons in the central island and by shell filling effects.

### 3.5.1 Device structure

The layer structure is shown in Fig. 3.14a and consists of a $p$-doped GaAs bulk layer with unintentional acceptor concentration of $10^{15} \text{cm}^{-3}$, on which a 20 nm intrinsic AlGaAs spacer is grown, then a 50 nm AlGaAs $n$-doped layer with donor concentration of about $10^{18} \text{cm}^{-3}$, and a ten nanometer GaAs cap layer.

In Fig. 3.14b the gate layout considered in the simulation is shown.

Surface states have been considered on the exposed surface at the top of the structure. Indeed the region in which electron transport occurs is very close to the surface, so device characteristics are strongly affected by surface properties. We have assumed the model for surface states explained in Sec. 3.2.2 and we have used parameters extracted from measurements on similar structures [22]: $\Phi^* = 4.85 \text{ eV}$ and $D_s = 5 \times 10^{12} \text{cm}^{-2} \text{eV}^{-1}$. Donor concentration has been taken as $N_D = 3.5 \times 10^{17} \text{cm}^{-3}$.

In the two-dimensional electron gas, two kinds of confinement have been considered. As shown in Fig. 3.14b, in the center of the 2DEG, the three dimensional Schrödinger equation has been solved, since quantum confinement is strong in all the three directions, while, in the rest of the 2DEG, the Schrödinger equation has been solved only in the vertical direction, since confinement along the direction perpendicular to the AlGaAs/GaAs interface is predominant. The solutions of the 3D and of the 1D Schrödinger equation are matched at the interface between the two regions by imposing Dirichlet boundary conditions to the wave functions. As long as tunnel barriers are opaque (tunneling probability much smaller than one) and the boundaries of the confined region are taken in order to allow penetration of the wave-function in the barriers, such approximation has a negligible effect on the energy eigenvalues and on the shape of the eigenfunctions.
Chapter 3. Quantum point contact and single electron devices

Figure 3.14: a) Layer structure and b) gate layout of the SET defined by split gate on an AlGaAs-GaAs heterostructure. In the picture the regions are shown in which 1D and 3D quantum electron density has been computed. Gates 1, 3, 4 and 6 are the external gates that define the quantum dot, while 2 and 3 are the inner gate pair that determine the number of electrons in the dot.
3.5.2 Results

In Fig. 3.15 is depicted the first subband at the AlGaAs/GaAs interface. Fig. 3.16 shows the computed electrochemical potential as a function of the inner gate pair voltage ($V_{G1}$) for a fixed external gate voltage ($V_{G1} = -1.1 \text{ V}$). From these curves it is possible to extract the capacitances of the equivalent circuit of the set (Fig.3.12). In this case $C_{TOT}$ is defined as

$$C_{TOT} = 2C_j + C_g + C_m,$$  \hspace{1cm} (3.57)

where $C_j = C_S = C_D$ is the capacitance between the dot and each reservoir and $C_m$ is the capacitance between the dot and the other gates (gates 1,3,4,6).

Curves are almost linear and parallel, but are not equally spaced. This means that while $C_{TOT}$ and $C_G$ vary as the number of electrons in the dot increase, the ratio $C_{TOT}/C_G$ remains constant. The values of the capacitance extracted from Fig. 3.16 are shown in Tab. 3.3. $C_j$ is obtained as described in the previous section and the remaining $C_m$ capacitance can be obtained from (3.57). $C_j$’s are quite small for the considered structure, meaning that the dot is weakly coupled to the reservoirs.

From Fig. 3.16 we can also extract the gate voltages for which the dot electrochemical potential is equal to the Fermi level of the reservoirs, at which conductance exhibits a peak (Table 3.4). In this case, an odd-even
Figure 3.16: Electrochemical potential as a function of the inner gate voltage for a number of electrons ranging from 1 to 4.

Table 3.3: Total dot capacitance and gate-dot capacitance derived from simulations as a function of the number of electrons in the dot for the AlGaAs-GaAs SET.

<table>
<thead>
<tr>
<th>transition (# electrons)</th>
<th>1 $\rightarrow$ 2</th>
<th>2 $\rightarrow$ 3</th>
<th>3 $\rightarrow$ 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_g$</td>
<td>3.84 aF</td>
<td>1.68 aF</td>
<td>4.67 aF</td>
</tr>
<tr>
<td>$C_T$</td>
<td>105.21 aF</td>
<td>46.93 aF</td>
<td>131.95 aF</td>
</tr>
<tr>
<td>$C_j$</td>
<td>0.713 aF</td>
<td>0.297 aF</td>
<td>0.805 aF</td>
</tr>
<tr>
<td>$C_m$</td>
<td>99.94 aF</td>
<td>44.66 aF</td>
<td>125.67 aF</td>
</tr>
</tbody>
</table>
Table 3.4: Gate Voltages corresponding to conductance peaks of the AlGaAs-GaAs SET.

<table>
<thead>
<tr>
<th>Peak number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage (V)</td>
<td>-0.420</td>
<td>-0.379</td>
<td>-0.283</td>
<td>-0.249</td>
</tr>
</tbody>
</table>

Figure 3.17: Transversal section of the conduction band at the Al-GaAs/GaAs interface for $V_g=-0.283$ V and in the case of half of an electron in the dot.

pairing is evident in the position of the conductance peaks, and is due to the fact that each Kohn-Sham orbital is occupied by two electrons, and the conduction band in GaAs has a single minimum at the $\Gamma$ point.

For these values, the conduction band in a region under the outer QPC has been extracted, as shown in Fig. 3.17. Indeed, in order to compute the conductance, it is necessary to find the transmission coefficient through the tunnel potential barrier created by the outer QPCs at the AlGaAs/GaAs interface. This goal has been accomplished by means of a scattering matrix code described in [32].

In Fig. 3.18 the isosurfaces are shown in correspondence of peak voltage, computed for an electron density of $1.3 \times 10^{16}$ cm$^{-3}$ as a function of half electrons imposed in the dot, in agreement with the Slater formula.

The computed conductance is negligibly small, since for a small number of electrons in the dot, the dot is small and weakly coupled to the
reservoirs. In reality, that would correspond to a dot that becomes practically insulated from the reservoirs even when it still contains several electrons.

### 3.6 Concluding remarks

In this chapter, we have focused our attention on device structures defined by split gates on AlGaAs/GaAs heterostructures. In particular, in the first part of the chapter we have developed a simple model for surface states based on two parameters: an “effective” work function of the surface states and the density of surface states per unit area per unit energy.

In order to validate the model, several quantum point contacts with different lithographic gaps have been realised and characterised from the electrical point of view in the regime of conductance quantization.

Pinch-off voltages obtained from experiments and simulations exhibit a very good agreement with just one fitting parameter, the donor concentration in the doped layer, which is usually known with insufficient precision to perform accurate simulations. Also parameters of the density of surface states must be known with reasonable accuracy to perform simulations, and in our case they were independently obtained.

In the second part instead, we have studied a single electron transistor in which subregions with different types of confinement are present. Indeed, SETs are a very useful testbench for assessing the capabilities of
codes for the simulation of nanoelectronic solid states devices.

We have shown that the followed approach allows to take into account the detailed geometry of the structure, and to compute parameters useful for a higher level description of device behavior, in terms of normal and tunneling capacitances, to be exploited by circuit simulators.
Chapter 4

Ultra-short MOSFET Devices

CMOS feature sizes have by now reached the sub 100 nm regime, and in the 2007 25 nm channel length transistors are going to be produced, as foreseen by the ITRS’01 [33]. In this perspective, from a design point of view, modern device simulation tools need to address typical aspects of nanoscale transport: “atomistic” distribution of dopants, quantum confinement of carriers and far from equilibrium transport in the channel.

In MOSFET devices of the latest technology node, the number of impurities in the depletion region is of the order of hundreds, therefore intrinsic fluctuations of the number and of the position of impurities strongly influence the value of the threshold voltage.

Moreover, already in 0.13 μm MOSFETs, a significant fraction of electrons traverse the channel without energy loss and with conservation of transversal momentum [34]. In this perspective, accurate simulations need to consider the ballistic fraction of electron contributing to the transport.

A modern device simulator must also provide a guidance for the development of alternative device architectures to conventional planar bulk MOSFETs, in order to reach the end of the roadmap expected in 2014.

In this chapter, we will investigate ballistic transport and the effect of random distribution of dopants in MOSFETs devices with channel length shorter than 50 nm, with particular attention to the so-called “Well-tempered” MOSFET, proposed by the group of Prof. Antoniadis
at MIT [35].

In the last section, we will focus on an unconventional MOSFET structure, the Silicon Nanowire Transistor (SNWT), which is considered by the nanoelectronic community a valid alternative to the Bulk planar MOSFET architecture.

4.1 Ballistic Transport in nanoscaled MOSFETs

In Metal-Oxide-Semiconductor Field Effect Transistors (MOSFETs) with channel length of a few tens of nanometers expected in the next few years [33], a significant fraction of electrons contributing to the drain current traverse the channel without undergoing inelastic scattering. The “ballistic” component is expected to increase with further scaling, and to predominate over inelastically scattered electrons in devices with channel length shorter than 30 nm [36, 37].

An analytical model for ballistic MOSFETs has been initially proposed by Natori [38], and recent simulations based on semiclassical Monte Carlo codes [39] and on a scattering theory of MOSFETs [34] exhibit significant differences with respect to simulations based on drift-diffusion or energy balance models.

Nanoscale MOSFETs also exhibit a significant degree of quantum confinement in the channel due to the high electric field in the direction perpendicular to the Si/SiO$_2$ interface. A quantum simulation is consequently required to take into account the 2D subband splitting and the lifting of the six-fold degeneracy of silicon conduction band. This is especially required to reproduce the experimental MOSFET threshold voltage $V_T$, since semiclassical simulations may underestimate $V_T$ by more than 100 mV [12].

We will now focus our attention on ballistic transport in nanoscaled MOSFETs, in which quantum confinement in the channel is properly taken into account, and electrons are assumed to conserve energy and transversal momentum until they reach the drain contact.

4.1.1 Ballistic electron and current densities in 2D subbands

Let us consider a 2D subband $E_i$, derived from 1D confinement in the $x$-direction, translationally invariant along the $z$-direction, and varying
along the $y$ direction as in the sketch of Fig. 4.1, where $E_{i_{\text{max}}}$ is the maximum of the subband and $y_{\text{max}}$ is the corresponding coordinate on the $y$ axis. We assume that the contacts are reflectionless, that electrons with energy lower than $E_{i_{\text{max}}}$ are in equilibrium with the contacts, while those with higher energy manage to traverse the channel conserving the energy. In particular, states propagating from the left to the right are in equilibrium with the left contact (whose electrochemical potential is $\mu_L$), while modes propagating from the right to the left are in equilibrium with the right contact (whose electrochemical potential is $\mu_R$).

The electron density can be expressed in a form similar to that in (2.87), where $\sigma_i$ this time reads

$$\sigma_i = \sigma_{i1} + \sigma_{i2} + \sigma_{i3}$$ (4.1)
where

\[ \sigma_{i1} = \sum_{k_z} \frac{f_0(E_i, k_y, k_z, \mu_{LR})}{L_y L_z} \]

\[ \sigma_{i2} = \sum_{k_z} \frac{f_0(E_i, k_y, k_z, \mu_L)}{L_y L_z} \]

\[ \sigma_{i3} = \sum_{k_z} \frac{f_0(E_i, k_y, k_z, \mu_R)}{L_y L_z} \]

(4.2)

and

\[ \mu_{LR} = \begin{cases} 
\mu_L & \text{if } y \leq y_{\text{max}} \\
\mu_R & \text{if } y > y_{\text{max}} 
\end{cases} \]

(4.3)

Since in the parabolic band approximation \( f_0 \) is an even function with respect to \( k_y \) and \( k_z \), \( \sigma_{i1} \) reads

\[ \sum_{k_z} f_0(E_i, k_y, k_z, \mu_{LR}) = 2 \sum_{k_z} f_0(E_i, k_y, k_z, \mu_{LR}) \]

\[ = \frac{2L_y}{\pi} \sum_{k_z \leq k_y_{\text{max}}} \int_{-\infty}^{+\infty} f_0(E_i, k_y, k_z, \mu_{LR}) \frac{dE_z}{E_z} \]

(4.5)

where we have used the relation expressed in (2.73). If we now define

\[ \Gamma \text{ def} = \frac{\hbar^2 k_y^2}{2m_y K_B T} \]

(4.6)

the integrals in eq. (4.5) becomes

\[ \int_{-\infty}^{+\infty} f_0(E_i, k_y, k_z, \mu_{LR}) \frac{dE_z}{L_y L_z} = \]

\[ \frac{\sqrt{2m_y K_B T}}{\hbar L_y L_z} \int_{0}^{+\infty} \frac{1}{\sqrt{\Gamma}} \frac{d\Gamma}{1 + \exp \left( \Gamma + \frac{E_i + \frac{\hbar k_y^2}{2m_y} - \mu_{LR}}{K_B T} \right)} \]

(4.7)
and by means of the Fermi integral with order $-\frac{1}{2}$, $\sigma_{i1}$ reads

$$\sigma_{i1} = \frac{2\sqrt{2m_y}K_BT}{\pi \hbar L_y} \sum_{k_y < k_{y_{max}}} \mathcal{F} \left[ -\frac{E_i + \frac{\hbar^2 k_y^2}{2m_y} - \mu_{LR}}{K_BT} \right]$$

$$= \frac{\sqrt{2m_y}K_BT}{\pi^2 \hbar} \int_0^{k_{y_{max}}} \mathcal{F} \left[ -\frac{E_i + \frac{\hbar^2 k_y^2}{2m_y} - \mu_{LR}}{K_BT} \right] dk_y \quad (4.8)$$

Defining

$$\Gamma' = \frac{\hbar^2 k_y^2}{2m_y K_BT}, \quad (4.9)$$

eq. (4.8) can finally be expressed as

$$\sigma_{i1} = \frac{K_BT \sqrt{m_y m_z}}{\pi^2 \hbar^2} \int_{\frac{E_{y_{max}} - E_i}{\hbar k_y}}^{+\infty} \mathcal{F} \left[ -\Gamma' - \frac{E_i - \mu_{LR}}{K_BT} \right] d\Gamma' \quad (4.10)$$

The terms $\sigma_{i2}$ and $\sigma_{i3}$ can be derived in the same way. In particular we have

$$\sigma_{i2} = \frac{K_BT \sqrt{m_y m_z}}{\pi^2 \hbar^2} \int_{\frac{E_{y_{max}} - E_i}{\hbar k_y}}^{+\infty} \mathcal{F} \left[ -\Gamma' - \frac{E_i - \mu_{R}}{K_BT} \right] d\Gamma', \quad (4.11)$$

$$\sigma_{i3} = \frac{K_BT \sqrt{m_y m_z}}{\pi^2 \hbar^2} \int_{\frac{E_{y_{max}} - E_i}{\hbar k_y}}^{+\infty} \mathcal{F} \left[ -\Gamma' - \frac{E_i - \mu_{R}}{K_BT} \right] d\Gamma' \quad (4.12)$$

and the electron density finally reads

$$n = \sum_i |\phi_i|^2 (\sigma_{i1} + \sigma_{i2} + \sigma_{i3}). \quad (4.13)$$

Following the same considerations, we can derive the current density in the $i$-th subband $J_i$

$$J_i = J_i^> - J_i^<, \quad (4.14)$$

where $J_i^>$ and $J_i^<$ are the current density due to electrons propagating from the left and from the right, respectively

$$J_i^> = \sum_{k_y > k_{y_{max}}} \frac{\hbar k_y f_0(E_i, k_y, k_z, \mu_{L})}{m_y} \frac{1}{L_y L_z}, \quad (4.15)$$
\[
J_i^< = \sum_{k_y < -k_{y_{\max}}} \frac{q \hbar k_y f_0(E_i, k_y, k_z, \mu_R)}{m_y L_y L_z}, \quad (4.16)
\]

and \(\frac{\hbar k_y}{m_y}\) is the group velocity in the \(y\)-direction.

Let us derive, for example, \(J_i^>\).

\[
J_i^> = \sum_{k_y > k_{y_{\max}}} 2L_z \frac{2}{\pi} \int_{-\infty}^{+\infty} \frac{q \hbar k_y f_0(E_i, k_y, k_z, \mu_R)}{m_y L_y L_z} dk_z
\]

\[
= \sum_{k_y > k_{y_{\max}}} \frac{2q}{\pi L_y} \sqrt{\frac{2m_z K_B T}{\hbar^2}} \frac{\hbar k_y}{m_y} f_{-\frac{1}{2}} \left[ \frac{E_i + \frac{\hbar^2 k_y^2}{2m_y} - \mu_L}{K_B T} \right]
\]

\[
= \frac{q \sqrt{2m_z K_B T}}{2\pi^2} \int_{k_{y_{\max}}}^{\infty} \frac{E_i + \frac{\hbar^2 k_y^2}{2m_y} - \mu_L}{K_B T} dk_y
\]

\[
= \frac{q \sqrt{2m_z K_B T}}{2\pi^2 \hbar^2} \int_{E_{i_{\max}} - E_i}^{+\infty} \left\{ f_{-\frac{1}{2}} \frac{E_i + E - \mu_L}{K_B T} - f_{-\frac{1}{2}} \left[ \frac{E_i + E - \mu_R}{K_B T} \right] \right\} dE \quad (4.17)
\]

The total current \(J_i\) can then be expressed as

\[
J_i(m_z, E_i, E_{i_{\max}}) = \frac{q \sqrt{2m_z K_B T}}{2\pi^2 \hbar^2} \int_{E_{i_{\max}} - E_i}^{+\infty} \left\{ f_{-\frac{1}{2}} \frac{E_i + E - \mu_L}{K_B T} - f_{-\frac{1}{2}} \left[ \frac{E_i + E - \mu_R}{K_B T} \right] \right\} dE. \quad (4.18)
\]

As can be noticed, both the electron and the current densities have to be solved via numerical integration, since they cannot be expressed in a closed form, as it happens for the quasi-equilibrium electron density [eq. (2.98)]. As a consequence, the computational burden greatly increases compared to simulations performed at quasi equilibrium; in particular, we have experienced a simulation time at least three times larger.
4.1.2 25 nm “Well-Tempered” MOSFET

The described model [40] has been applied to the simulation of the so-called “Well Tempered” MOSFET with channel length of 25 nm proposed by D. Antoniadis et al. [35] as a reasonable candidate for nanoscale MOSFETs and as a benchmark for the validation of simulation codes. The doping profile is shown in Fig. 4.2. Source and drain profiles are gaussian, while super-halo doping is implanted in the channel in order to reduce charge sharing effects. The oxide thickness is 1.5 nm and the channel length is defined by the points where the source and drain doping concentration falls below $2 \times 10^{19} \text{ cm}^{-3}$. The polysilicon gate has donor concentration $5 \times 10^{20} \text{ cm}^{-3}$.

Quantum confinement is predominant along the $x$-direction, so the 1D Schrödinger equation has been solved in the direction perpendicular to the Si/SiO$_2$ interface, while a semiclassical expression has been assumed for in-plane states.

Considering all the 2D subbands resulting from mass anisotropy and the six conduction band minima, the total current density per unit length can be written as

$$J = \sum_{i=1}^{\infty} (J_{il} + J_{it}),$$  \hspace{1cm} (4.19)$$

where $J_{il}$ and $J_{it}$ are the current densities computed for the two degen-
Figure 4.3: First subband profile computed for $V_{GS} = 1.0$ V and several values of $V_{DS}$.

erate minima in the $k_x$ direction and the four degenerate minima in the $k_y$ and $k_z$ directions, respectively:

$$J_{\text{il}} = 2J(m_t, E_{\text{il}}, E_{\text{il}}^{\max})$$
$$J_{\text{it}} = 2J(m_t, E_{\text{it}}, E_{\text{it}}^{\max}) + 2J(m_t, E_{\text{it}}, E_{\text{it}}^{\max});$$

(4.20)

where $E_{\text{il}}^{\max}$ and $E_{\text{it}}^{\max}$ are the maximum values of $E_{\text{il}}(y)$ and $E_{\text{it}}(y)$, respectively.

The computed profiles of the first subband for gate voltage $V_{GS} = 1$ V and different values of drain-to-source voltage ($V_{DS}$) are plotted in Fig. 4.3. Subband profiles are strongly affected by the overlap of the polysilicon gate with the drain and source regions. The super-halo $p$ doping allows short channel effects to be kept under control: as can be seen, the channel barrier is only slightly reduced when $V_{DS}$ is significantly increased.

Subband splitting is much larger than the thermal energy, so that practically only the first subband is occupied. Indeed, simulations show that in strong inversion more than 95% of electrons in the channel are in the first subband. The common source output characteristics are shown in Fig. 4.4. The threshold voltage $V_T$ is close to 0.8 V, depending on the definition used, and is apparently too high for CMOS circuits with a supply voltage of about 1.5 V. It should be considered that a semiclassical simulation would underestimate the threshold voltage by 170 mV [12]. The output characteristics saturate for small values of $V_{DS}$, between 0.1
4.1. Ballistic Transport in nanoscaled MOSFETs

and 0.2 V, since ideal contacts have been considered with zero series resistance.

In Fig. 4.5 the drain-to-source current ($I_{DS}$) and the transconductance ($g_m$) are plotted as a function of $V_{GS}$. The device exhibits a subthreshold slope close to 100 mV/decade, which is still adequate for obtaining a good $I_{on}/I_{off}$ ratio. From the $I_{DS}(V_{GS})$ curves for different values of $V_{DS}$, we can extract a value of the DIBL of 100 mV/V.

In the same figure we also plot, for comparison, the results obtained with the Medici simulator in Ref. [35] with a quantum-corrected drift diffusion model. As can be seen, the subthreshold slope and the DIBL are similar, while an appreciable difference can be observed in the saturation current and in the transconductance. Indeed, in the ballistic regime, a higher current drive is obtained compared to that predicted by a drift-diffusion model.

Fig. 4.6 shows the electron charge density in the channel at a distance from the Si/SiO$_2$ interface corresponding to the peak of the electron density (1 nm), the maximum velocity in the channel for electrons with energy equal to the Fermi level of the source, and the average velocity of electrons in the channel. In particular, we can notice that the injection velocity of electrons in the channel is close to the thermal velocity in the longitudinal direction ($1.55 \times 10^7$ cm/s), in agreement with the fact that
Figure 4.5: Current and transconductance versus gate voltage for $V_{DS} = 0.5$ V and 1 V obtained with the ballistic model and from MEDICI simulations reported in Ref. [35].

Figure 4.6: Electron density, source Fermi level and average velocity in the channel for $V_{GS} = 1.2$ V and $V_{DS} = 0.5$ V.
4.2 Random distribution of dopants in nanoscaled MOSFETs

In state of the art transistors, the number of ionized impurities is of the order of hundreds, and stochastic variations of the number and of the position of the dopants occur, due to the random nature of ion implantation and diffusion. These microscopic variations greatly affect macroscopic quantities such as the off-state current and the threshold voltage, as demonstrated experimentally in [41].

In order to understand, predict and control the effects of discrete dopants, analytical modeling and numerical simulations have been performed in the last decade. A simple one-dimensional analytical model that takes into account the fluctuations of the total number of dopants and of their position, has been proposed in [46], [45, 47] and considerably improved in [48]. Three-dimensional simulations capable of taking into account the random position of ionized impurities in all directions, that is important in very short devices, have been performed in [42, 43, 49, 50]. It is worth noticing that such works are based on semiclassical simulations, in which quantum effects are not considered at all.

However, devices with channel lengths of the order of tens of nanometers and oxide thickness in the range of 1-3 nm, are significantly affected by quantum confinement [46, 50]. The high electric field in the direction perpendicular to the Si/SiO$_2$ interface, due to the reduced gate oxide thickness and the increased bulk doping, strongly confines electrons in the channel and induces well-separated 2D subbands. Such effect must be considered to make reliable prediction of device properties.

The influence of quantum effect on the dispersion of the threshold voltage has been addressed by Asenov et al. [44] by means of the density gradient formalism.

Such approach, however, is a rude approximation of the self-consistent solution of the Poisson/Schrödinger equation, requires a parameter calibration for the particular structure, and the error is not easily controllable in the case of rapidly varying potentials. Therefore, we propose to perform a full quantum simulation, based on the self-consistent solution of the Poisson/Schrödinger equation with density functional theory.

In particular we will present results for the “Well tempered” bulk-Si
n-MOSFETs with channel length and width of 25 and 50 nm, and, in the last section, we will introduce an alternative method for the statistical study of random dopant effects, making comparisons with other works present in literature.

### 4.2.1 Random distribution effects in “Well-Tempered” MOSFETs

The structure considered in the simulations is depicted in Fig. 4.7.

In order to perform a statistical simulation, we need to choose a convenient way to extract the threshold voltage ($V_T$).

One method is to compute the conductance $g_0$, which has the following approximate expression:

$$g_0 \equiv \frac{\partial I_D}{\partial V_{DS}} \bigg|_{V_{DS}=0} \approx \mu_n \frac{W}{L} C_{ox} (V_{GS} - V_T),$$

(4.21)

where $\mu_n$ is the electron mobility in the channel, $C_{ox}$ is the oxide capacitance per unit area, $W$ and $L$ are the channel width and length,
4.2. Random distribution of dopants in nanoscaled MOSFETs

Figure 4.8: Region considered for the calculation of conductance and associated boundary conditions

respectively, while \( V_{GS} \) and \( V_{DS} \) are the gate and the drain-to-source voltage.

\( V_T \) can be then obtained as the intercept of the curve \( g_0-V_{GS} \) in the strong inversion region with the \( V_{GS} \) axis. The conductance is computed by solving a simplified continuity equation [51]. In particular the current density can be expressed as

\[
\vec{J}_n = -qn\mu_n \nabla \phi_n, \tag{4.22}
\]

where \( \mu_n \) is the electron mobility and \( \phi_n \) is the quasi fermi level for the electrons. If we now suppose to work at quasi equilibrium, (4.22) becomes,

\[
\vec{J}_n = -qn_0\mu_n \nabla \phi_n, \tag{4.23}
\]

where \( n_0 \) is the charge density computed for zero drain-to-source voltage. Considering null generation-recombination term, the continuity equation for the electrons reads,

\[
\nabla \cdot J_n = 0. \tag{4.24}
\]

Substituting (4.23) in (4.24) and considering a constant electron mobility, we obtain

\[
\nabla \cdot (n_0 \nabla \phi) = 0. \tag{4.25}
\]

The boundary condition for (4.25) are shown in Fig. 4.8 for a region of the MOSFET containing the channel. A small voltage \( \Delta \phi \) is applied between the source and the drain, while a zero current density is fixed through the lateral faces of the region, imposing a zero component of
the gradient of the quasi-Fermi level in the direction perpendicular to the surface (n).

Another method is to derive the threshold voltage directly from the transfer characteristics, extracting the intersection of the line approximating the drain-to-source current curve as a function of the gate voltage at fixed \( V_{DS} \). To this purpose, we have used the ballistic model explained in the previous section. Compared to the other, this approach is the only able to take into account drain-induced barrier lowering, but, on the other hand, requires a much larger computational burden.

In the following, we will then extract the \( V_T \) by means of the simplified drift-diffusion method.

The “atomistic” distribution of impurities has been obtained assuming that the implanted ions in the channel obey the Poisson distribution. In particular, for each grid point we have considered the associated volume element and multiplied its volume \( \Delta V \) by the nominal doping concentration. Then, a random number \( N' \) has been extracted with Poisson distribution and divided by \( \Delta V \) in order to have the “actual” doping concentration in the volume element. The standard deviation of \( V_T \) has been then obtained by simulating a large number of devices with the same nominal doping, but with different actual dopant distribution. Fig. 4.9 show the distribution of threshold voltage computed on an ensemble of 100 nominally identical devices, where \( V_{T,\text{nom}} \) is the quantum computed threshold voltage in case of uniform doping distribution, while in Tab. 4.1 the standard deviations of the threshold voltage are shown.

The considered number of simulated devices is a good compromise between precision and computational burden. Indeed, the standard deviation of the error on \( \sigma_{V_T} \) (standard deviation of the threshold voltage) is about 7%, which is still accurate. It comes out from the performed simulations that \( 3\sigma_{V_T} \) is comparable with the threshold voltage shift due to the quantum effects: for an accurate simulation, random dopant dispersion has to be therefore considered as well as quantum confinement.

In Tab. 4.1, the first column represents the threshold voltage shift in the case of uniform doping due to quantum effects: \( V_{T,sc} \) and \( V_{T,Q} \) are the threshold voltage computed with semiclassical and quantum models, respectively. \( \sigma_{sc} \) and \( \sigma_Q \) are the semiclassical and the quantum standard deviations respectively, while \( \sigma_s \) and \( \sigma_i \) are the standard deviations computed with the analytical model described in [45] using, respectively, the one-dimensional doping profile extracted in the middle of the channel and the one-dimensional profile derived from the average integration of the three-dimensional profile over the \( y-z \) plane.
Figure 4.9: Threshold voltage distribution in the two simulated devices: in each sub-plot is indicated the channel length of the associated device is indicated.
Table 4.1: $V_{T_Q}$ and $V_{T_{sc}}$ are the quantum and semiclassical threshold voltage. $\sigma_{sc}$ and $\sigma_Q$ are the standard deviation computed semiclassically and with the quantum model, respectively. $\sigma_s$ and $\sigma_i$ are the standard deviation computed as in [48] using the 1D doping profile in the middle of the channel and the average 1D profile in the $y-z$ plane.

As can be noticed, the standard deviation computed by means of quantum model is larger than that computed semiclassically since the threshold voltage standard deviation is a rising monotonic function of the oxide thickness, and quantum confinement effects produce an increase of the effective oxide thickness itself. Moreover, the semiclassical standard deviation obtained from the simulations differs from that computed analytically even in the case of the mean integrated doping profile. Indeed, contrary to the performed simulations, the analytical model is derived for constant doping in the bulk and does not consider the charge sharing effects that are relevant for these geometries.

In Tab. 4.1 the ratios $\frac{\sigma_Q}{V_{T_{Qnom}}}$ are also shown: even if the ratios are small with respect to unity, the obtained threshold voltage fluctuations can be significant in minimum-geometry devices.

For the two structures, we have also computed the threshold voltage standard deviation as a function of the doping concentration (Fig. 4.10). In particular we have multiplied the doping profile for a factor and then we have computed the standard deviation as explained above.

The slopes of the least mean square fitting lines, which in the case of constant doping concentration is approximately the $N_A$ exponent, differ from that computed using the mono-dimensional analytical model described in [46], [45] and [48]. Extracting the 1D doping profile in the middle of the channel, the analytical simulations have yielded slopes equal to 0.262 and 0.247 for the devices with channel length equal to...
25 and 50 nm, respectively. Indeed, in our three-dimensional simulator, charge sharing effects are taken into account as well as quantum confinement.

In Fig. 4.11 the mean value of the threshold voltage variation ($\Delta V_T = V_{TQ} - V_{TQ,nom}$) is plotted as a function of the doping factor, which is defined as the factor by which the doping profile is multiplied: as expected theoretically [46] it decreases for increasing doping factor.

To investigate the influence of dopant positions on the threshold voltage, we have computed for each randomly generated dopant profile and for the 25 nm case, the term $M$, defined as

$$M = \int_0^{W_d} dx \int_0^W dy \int_{-W_d}^{W_d} dz \left( 1 - \frac{x}{W_d} \right) N_A(x, y, z), \quad (4.26)$$

where $W_d$ is the depletion width.

As shown in Fig. 4.12, we have then plot the scatter plot $M$ vs $V_{TQ}$ for each of the random samples. As it can be seen, the correlation is about 0.601, that is quite far from 0.964 found in [52] in case of constant bulk doping.

Indeed (4.26) not only confers a larger weight to shallower impurities, but also tacitly assumes that the $x$-position of the impurities is the dominant factor of threshold voltage dispersion. The scattering plot then
Figure 4.11: Mean value of the threshold voltage variation ($\Delta V_T$) as a function of the multiplying factor of the doping profile, which is defined as the factor by which the doping profile is multiplied.

![Figure 4.11](image)

Figure 4.12: Scatter plot of $M$ versus $V_T$ for the $L=25$ nm device for each of the 100 devices, and least mean square fitting line. The correlation coefficient is 0.601.

![Figure 4.12](image)
4.2. Random distribution of dopants in nanoscaled MOSFETs

Figure 4.13: a) Threshold voltage computed by means of the ballistic and the simplified drift-diffusion approach (straight lines) and threshold voltage extracted from the charge plot computed in a region of dimension $50 \times 50 \times \delta \text{ nm}^{-3}$ b) Transfer characteristics extracted for $\delta = 20 \text{ nm}$ and $\delta = 13 \text{ nm}$, which fit well the simplified drift-diffusion and the ballistic transfer characteristics, respectively.

states that in the three-dimensional doping profile case, the position of the impurities on the $y-z$ plane is relevant.

4.2.2 Alternative method for the statistical study of random dopant effects

As shown in the previous section, the threshold voltage definition is based on the extraction of the intercept with the voltage axis of the line that provides the best fit of the transfer characteristics at low $V_{DS}$ in the strong inversion region.

Such characteristics can be computed in the limit of fully ballistic transport, and in the case of drift-diffusion transport, from the conductance obtained by solving a simplified continuity equation.

Both approaches, however, are too computationally demanding to be used in statistical simulations, therefore we propose a simplified method [53], consisting in extracting $V_T$ from the intercept of the fitting line of the curve representing the charge integrated in a region in the channel as a function of $V_G$.

Fig. 4.13 show the threshold voltage computed by means of the bal-
Figure 4.14: Semiclassical and quantum standard deviation computed with the simplified drift/diffusion model as a function of the doping concentration. Standard deviation computed in [43], [44], [45] and in [48] are also shown.

listic and the simplified drift/diffusion model (straight lines) and the threshold voltage obtained computing the charge in the channel (square points) as a function of the width of the region of integration in the channel direction. As can be noticed, the length of the intervals centered in the middle of the channel equal to 26 nm and 40 nm fit well the ballistic and the quasi-equilibrium threshold voltage, respectively. Indeed, in case of the ballistic model, only the carriers very close to the maximum of the subband contribute roughly to the current, while in the drift/diffusion model, all regions of the channel contribute equally to the channel conductance.

The above described threshold voltage extraction has been used to compute the threshold voltage statistics, performing simulations over 100 devices with the same nominal doping profile, but with different atomistic distribution.

Fig. 4.14 shows the quantum and semiclassical threshold voltage standard deviation as a function of the doping concentration in the bulk, in the quasi-equilibrium case. As expected, the standard deviation computed semiclassically ($\sigma_{sc}$) is smaller than that computed by means of quantum models ($\sigma_Q$), due to the increase of the effective oxide thickness caused by quantum confinement, as explained in the previous section.
For the same reason, we can notice that the two curves are almost shifted. Indeed, as the carrier centroid along the vertical direction varies in the range of $1 \div 1.3$ nm as the doping factor is varied, the percentage increase of the oxide thickness is about 13.5%, that in the semilog plot corresponds to a shift of approximately $55 \times 10^{-2}$.

In Fig. 4.14 semiclassical and quantum results obtained in [43, 44] are also shown, and those derived by the analytical model described in [45, 48]. As can be noticed, our results differ from those in Refs. [43, 44], and this is probably due to the different definition of threshold volt-

<table>
<thead>
<tr>
<th>Doping factor</th>
<th>ballistic $\sigma_Q/\sigma_{sc}$</th>
<th>quasi equilibrium $\sigma_Q/\sigma_{sc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1.253</td>
<td>1.196</td>
</tr>
<tr>
<td>0.8</td>
<td>1.18</td>
<td>1.163</td>
</tr>
<tr>
<td>0.6</td>
<td>1.159</td>
<td>1.164</td>
</tr>
<tr>
<td>0.4</td>
<td>1.158</td>
<td>1.162</td>
</tr>
<tr>
<td>0.2</td>
<td>1.16</td>
<td>1.172</td>
</tr>
</tbody>
</table>

Table 4.2: Ratio of quantum to semiclassical standard deviation of the threshold voltage ($\sigma_Q$ and $\sigma_{sc}$, respectively) derived from simulations and from linear fitting of the scatter plot.
Figure 4.16: Scatter plot of the threshold voltage computed by means of quantum model versus semiclassical threshold voltage.

age. Indeed, while in our case the threshold voltage is defined in deep inversion, in the case of Ref. [43] the threshold voltage is defined as the gate voltage at which the source-to-drain current is equal to $10^{-8}A$, i.e., in weak inversion or in subthreshold conditions, where random percolating paths may have a role in increasing the dispersion of threshold voltage.

Semiclassical simulations are instead in very good agreement with analytical expression based on a simple 1D model [48], since the dispersion of $V_T$ depends in practice only on the vertical distribution of dopants and also because in the considered devices charge sharing effects are negligible: indeed, simulations performed on a MOS structure with the same doping yield a threshold voltage that differs from that of the MOSFET only by 30 mV.

In Fig. 4.15 the standard deviation is plotted as a function of doping, in case of ballistic and quasi-equilibrium simulations of threshold voltage. In this way it is possible to define an upper and a lower limit for the standard deviation of $V_T$, depending on the degree of ballistic transport. In particular, considering quantum effects, the device should belong to the highlighted region of Fig. 4.15.

From simulations it has also been possible to obtain an interesting relation between the random variables $V_{TQ}$ and $V_{Tsc}$, i.e., the threshold voltages computed with quantum and semiclassical models, respectively. In particular, the two random variables are strictly correlated (the cor-
relation factor is always larger than 0.997 for all simulated devices), and a linear relation can be found for them: \( V_{TQ} = mV_{Tsc} + q \) (Fig. 4.16).

In Tab. 4.2 the best fitting values of \( m \) are compared with the ratio \( \sigma_Q/\sigma_{sc} \) for both the ballistic and the quasi equilibrium case, and are practically identical, suggesting that charge centroid and quantum capacitance are responsible for the difference between \( \sigma_Q \) and \( \sigma_{sc} \).

Since results derived from the analytical formula and semiclassical simulations are in good agreement, and since quantum and semiclassical threshold voltages are linearly dependent, in case of reduced charge sharing effects, the above shown results suggest a simple way to correct the semiclassical expression for the threshold voltage dispersion, in order to include quantum effects.

### 4.3 Silicon Nanowire Transistor

Silicon Nano Wire Transistors (SNWT) represent a promising alternative architecture to the conventional planar technology for devices at the end of the ITRS roadmap [1], because of the improved electrostatic control of the channel via the gate voltage and the consequent suppression of short channel effects [54, 55, 56, 57]. Since electrostatic play an important role in these devices, simulations can represent a valid tool in order to understand device behavior and to give design guidelines. Three-dimensional drift-diffusion simulations have been performed in Refs. [54, 55] to study the optimum configuration that reduces short channel effects, and in [58, 59, 60], where quantum corrections to the electron density have been considered. In [61] instead, a simulation study of ballistic SNWTs with different cross sections, has shown advantages with respect to Double Gate MOSFETs, as far as downsizing is concerned.

Here we focus on some important aspects of the electrical properties of SNWTs that are still open. Indeed, as demonstrated in [62], silicon MOSFETs with effective channel length smaller than 50 nm will not be fully ballistic, so it is reasonable to assume that even for SNWTs scattering events can occur and a fully ballistic assumption can be considered as a limiting case. Transport in SNWT is then likely to be in an intermediate regime between ballistic and drift-diffusion, and its simulation would require the detailed knowledge of the scattering rates of electrons in the 1D subbands. Relevant information can be obtained by considering the two limiting cases (ballistic and drift-diffusion), assuming that a
partially ballistic transistor would have an intermediate behavior. Another important aspect of interest is source-to-drain tunneling, that is expected to have a significant impact on the shortest devices.

In this section, we investigate the electrical properties of SNWTs in order to understand what are the scaling prospective of such devices, the achievable performance, and the relevance of source-to-drain tunneling.

In particular, we have considered a device structure with rectangular cross section (5 nm × 5 nm), in which fully ballistic and drift-diffusion transport in 1D subbands has been considered. As a consequence we have been able to define an upper and a lower limit for the SNWT performance, deriving significant quantities such as the DIBL and the subthreshold swing as a function of the channel length.

4.3.1 Physical models

Since in the considered devices the confinement is strong in the plane perpendicular to the current direction, we have decoupled the Schrödinger equation in a two dimensional equation in the plane of confinement, while continuous states have been considered in the direction of propagation [64].

Energy levels splits in 1D subbands, which we have assumed to be uncoupled.

In particular, in order to define an upper limit for the device performance, let us derive the electron concentration and the current in case of ballistic transport in 1D subbands.

By means of the same consideration of section 4.1 (the sketch of Fig. 4.1 is still valid), the electron concentration has the same form of eq. (2.102),

\[
\alpha_i = \alpha_{i1} + \alpha_{i2} + \alpha_{i3} \quad (4.27)
\]

where

\[
\alpha_{i1} = \sum_{|k_y|<k_{ymax}} \frac{f_0(E_i, k_y, \mu_{LR})}{L_y},
\]

\[
\alpha_{i2} = \sum_{k_y>k_{ymax}} \frac{f_0(E_i, k_y, \mu_L)}{L_y},
\]

\[
\alpha_{i3} = \sum_{k_y<-k_{ymax}} \frac{f_0(E_i, k_y, \mu_R)}{L_y}. \quad (4.28)
\]
4.3. Silicon Nanowire Transistor

and

\[ \mu_{LR} = \begin{cases} \mu_L & \text{if } y \leq y_{\text{max}} \\ \mu_R & \text{if } y > y_{\text{max}} \end{cases} \]  

(4.29)

In particular, \( \alpha_{i1} \) can be expressed as

\[
\alpha_{i1} = \sum_{|k_y|<k_{ymax}} \frac{f_0(E_i, k_y, \mu_{LR})}{L_y} = 2 \sum_{k_y<k_{ymax}} \frac{f_0(E_i, k_y, \mu_{LR})}{L_y} 
= \frac{2}{\pi} \int_{k_{ymax}}^{k_y} f_0(E_i, k_y, \mu_{LR}) dk_y 
= \sqrt{2m_yK_BT} \int_{\epsilon_{iymax}-E_i}^{+\infty} \frac{dE}{\hbar \pi} \frac{1}{\sqrt{E}} \exp \left( \frac{E_i - \epsilon_{iymax} - \mu_L}{K_BT} \right) 

(4.30)
\]

that can be computed using the incomplete Fermi integrals of order \(-\frac{1}{2}\) [63].

\( \alpha_{i2} \) and \( \alpha_{i3} \) can be derived in a similar way:

\[
\alpha_{i2} = \sqrt{2m_yK_BT} \int_{\epsilon_{iymax}-E_i}^{+\infty} \frac{dE}{\hbar \pi} \frac{1}{\sqrt{E}} \exp \left( \frac{E_i - \epsilon_{iymax} - \mu_R}{K_BT} \right) 

(4.31)
\]

\[
\alpha_{i3} = \sqrt{2m_yK_BT} \int_{\epsilon_{iymax}-E_i}^{+\infty} \frac{dE}{\hbar \pi} \frac{1}{\sqrt{E}} \exp \left( \frac{E_i - \epsilon_{iymax} - \mu_R}{K_BT} \right) 

(4.32)
\]

For what concerns the current \( I_i \), it can be expressed as

\[ I_i = I_i^\geq - I_i^< \]  

(4.33)

where \( I_i^\geq \) and \( I_i^< \) are the current due to electrons propagating from the left and from the right, respectively.
In particular

\[ I_i^> = \sum_{k_y > k_{y_{max}}} q \hbar k_y \frac{f_0(E_i, k_y, \mu_L)}{m_y L_y} \]

\[ = \frac{q}{\pi} \int_{k_{y_{max}} \pi}^{+\infty} \frac{\hbar k_y}{m_y} dk_y \frac{dE}{1 + \exp \left( \frac{E_i + \frac{\hbar^2 k_y^2}{2m_y} - \mu_L}{K_B T} \right)} \]

\[ = \frac{qK_B T}{\pi \hbar} \int_{E_{i_{max}} - E_i}^{+\infty} \frac{\exp \left( -E - \frac{E_i - \mu_L}{K_B T} \right)}{1 + \exp \left( -E - \frac{E_i - \mu_L}{K_B T} \right)} dE \]

\[ = -\frac{qK_B T}{\pi \hbar} \ln \left[ 1 + \exp \left( -E - \frac{E_i - \mu_L}{K_B T} \right) \right]_{E_{i_{max}} - E_i}^{+\infty} \]

\[ = \frac{qK_B T}{\pi \hbar} \mathcal{F}_0\left( \frac{\mu_L - E_{i_{max}}}{K_B T} \right) \]

(4.34)

where \( \mathcal{F}_0 \) is the Fermi-Dirac integral of zero order.

Equivalently

\[ I_i^< = \frac{qK_B T}{\pi \hbar} \mathcal{F}_0\left( \frac{\mu_R - E_{i_{max}}}{K_B T} \right) \]

(4.35)

and finally,

\[ I_i = \frac{qK_B T}{\pi \hbar} \left[ \mathcal{F}_0\left( \frac{\mu_L - E_{i_{max}}}{K_B T} \right) - \mathcal{F}_0\left( \frac{\mu_R - E_{i_{max}}}{K_B T} \right) \right] \]

\[ = \frac{qK_B T}{\pi \hbar} \ln \left[ \frac{1 + \exp \left( \frac{\mu_L - E_{i_{max}}}{K_B T} \right)}{1 + \exp \left( \frac{\mu_R - E_{i_{max}}}{K_B T} \right)} \right] \]

(4.36)

The general expression for the source-to-drain current \( (I_{DS_i}) \) in the generic \( i \)-th subband \( E_i \), considering quantum tunneling, reads

\[ I_{DS_i} = \frac{qK_B T}{\pi \hbar} \int_0^\infty dE \tau_i(E) \left[ f_0(E - \mu_L) - f_0(E - \mu_R) \right], \]

(4.37)
where $\tau_i(E)$ is the transmission coefficient, that, in the semiclassical case (eq. (4.36)), is zero for energy below the maximum of $E_i$ and equal to one otherwise [40], while, in the quantum case, it can be computed by means of the transmission matrix formalism. We have to point out that, in the semiclassical approximation for the transmission coefficient, (4.37) reduces to (4.36).

Since scattering events may occur even at these geometries, it is worth solving the continuity equation also in the drift-diffusion approximation, with the Scharfetter and Gummel scheme [65].

In particular, the density current for electrons in the $i – th$ subband ($J_{n_i}$) reads

$$J_{n_i} = q\mu_n n_{1D_i} \frac{\partial E_i}{\partial x} + qD_n \frac{\partial n_{1D_i}}{\partial x} \quad (4.38)$$

where $n_{1D_i}$ is the linear electron density, and $\mu_n$ and $D_n$ are the mobility and the diffusion coefficients for electrons, respectively.

For what concerns the mobility, velocity saturation has been taken into account by means of the Caughey-Thomas model [66],

$$\mu_n = \frac{\mu_0}{1 + \left(\frac{\mu_0 \mathcal{E}_l}{v_{sat}}\right)^\gamma} \quad (4.39)$$

where $\mathcal{E}_l$ is the longitudinal field, $v_{sat}$ is the velocity saturation ($v_{sat} = 1.1 \times 10^7$ cm/s), $\gamma$ is a constant fitting parameter ($\gamma = 2$ is a common value for electrons at room temperature), while $\mu_0$ has been computed by means of the Mathiessen’s rule combining lattice, and impurities scattering described by the unified mobility model proposed by Klassen [67].

Once computed $n_{1D_i}$, the three-dimensional electron density ($n$) reads

$$n = \sum_i |\psi_i|^2 n_{1D_i} \quad (4.40)$$

where $|\psi_i|$ are the eigenfunctions associated with the $i – th$ subband.

### 4.3.2 Results and Discussions

In Fig. 4.17 the simulated SNWT structure is shown. The oxide thickness is 1.5 nm, and the channel length $L$ ranges from 7 to 25 nm. Degenerate statistics is considered in the wire. For simplicity, the gate is metallic.

In Fig. 4.18 the isosurface of the electron concentration for a device with channel length equal to 15 nm nanometer is shown: the gate and
the source-to-drain voltages are equal to 0.5 V, i.e. the device is in the saturation regime, as confirmed by the constriction in correspondence of the drain.

In Fig. 4.19 the transfer characteristics are plotted for $L=25$ nm and $L=7$ nm, respectively, for a drain-to-source voltage $V_{DS}=0.5$ V. It is evident that for both devices the current in the ballistic case is much higher than in the drift-diffusion case. For a channel length of 7 nm source-to-drain tunneling is significant both in sub-threshold and in strong inversion conditions, while for $L=25$ nm tunneling is quantitatively relevant only in strong inversion. Such behavior is more clear if we consider $I_{on}$ and $I_{off}$ as a function of $L$, shown in Figs. 4.20 and 4.21, respectively. We have defined $I_{off}$ as the current obtained for $V_{DS}=0.5$ V and $V_{GS}=-0.5$ V, and $I_{on}$ as the current obtained for $V_{DS}=0.5$ V and $V_{GS}=0.5$ V. The dependence of $I_{on}$ on $L$ is rather small, in all transport regimes. Quantum tunneling gives a contribution only slightly dependent on $L$, and significant already for $L=25$ nm.

The reason can be understood considering that, for example, for $L=25$ nm the product of the transmission probability times the occupation factor has the energy dependence shown in Fig. 4.22, for semiclassical tunneling (thin line) and quantum tunneling (thick line).
Figure 4.18: Electron density isosurface \((n=1.4 \times 10^{19} \text{ cm}^{-3})\) computed for the SNWT with channel length equal to 15 nm, for a gate voltage \(V_{GS} = 0.5\) V and a source-to-drain voltage \(V_{DS} = 0.5\) V.

Figure 4.19: Transfer characteristics computed for \(V_{DS} = 0.5\) V, for the \(L = 25\) nm and \(L = 7\) nm devices.
Figure 4.20: $I_{On}$ current as a function of the channel length

Figure 4.21: $I_{Off}$ current as a function of the channel length
Figure 4.22: Argument of eq.(4.37). The upper curve is the argument in the semiclassical case, while the lower curve the argument in the quantum case. Tunneling current is larger than the semiclassical current, since the area $a$ is larger than the area $b$.

Figure 4.23: Sub-threshold slope as a function of the channel length
The subband current is proportional to the area below the curve. Quantum tunneling adds a contribution that is proportional to the difference between the area $a$ and the area $b$. The shape of the subband peak, shown in the inset, determines $a$ and $b$, and depends essentially on the vertical electrostatics, and only marginally on $L$. Even for $L=25$ nm the barrier depth is only about 3 nm. On the other hand, $I_{off}$ is affected by quantum tunneling only for $L=10$ nm, and is extremely sensitive to the channel length. Such dependence is mainly due to the degradation of the subthreshold slope $S$ with decreasing length (shown in Fig. 4.23). Down to 15 nm, $S$ is very good ($<70$ mV/dec).

For smaller $L$, $S$ is degraded by charge sharing, but is still acceptable. Conversely, Drain Induced Barrier Lowering (DIBL) is much higher for ballistic than for drift-diffusion transport, as shown in Fig. 4.24. In the former case, current is essentially controlled by the subband peak, while in the latter, it is roughly dependent on the whole region between the source and the subband peak, that is farther from the drain.
4.4 Concluding remarks

In this chapter, we have focused our attention on ultra-short planar MOSFETs, considering the effects of the random dopant distribution, quantum confinement of electrons, and ballistic transport in the channel.

The threshold voltage standard deviation computed by means of quantum models is larger than that computed semiclassically, due to the increase of the effective oxide thickness.

Moreover, the performed simulations have shown that even the ballistic transport in the channel increase the dispersion of the threshold voltage. In particular, fluctuations of $V_T$ in the case of ballistic transport depend on fluctuations of the subband maximum, that are mainly due to fluctuations of the impurity number in a small volume in the central region of the device (at small drain voltage).

Quantum confinement has on the dispersion of $V_T$ an effect that is shown to be strongly correlated to that on $V_T$, and is basically due to the inversion layer capacitance.

In the last section instead, we have investigated the electrical properties of SNWTs, considering both drift-diffusion and ballistic transport, in order to evaluate the limiting cases for the performance of partially ballistic devices. SNWTs offer promising scaling perspectives down to 7 nm channel lengths, predicted at the end of the ITRS Roadmap. In addition, we have shown that the impact of quantum tunneling is significant also in strong inversion, and depends essentially on the vertical electrostatics rather than on $L$. 
Chapter 5

Simulation of nanocrystal memories

Nanocrystal memories [68] are promising candidates for future non-volatile memories for two main reasons: on the one hand, their fabrication requires only few additional process steps with respect to conventional CMOS technology; on the other hand, they offer the perspective of better scalability than conventional double-poly Flash EEPROMs, that would translate into reduced program/erase times, lower power consumption, and higher integration density [69, 70, 71, 72, 73].

Recently, for SOI nanocrystal floating gate memories with ultra-narrow channel [74, 75], an increasing threshold voltage shift has been observed with decreasing channel width, at constant program time and write voltage. Such behavior might suggest that wider programming window, which is defined as the difference between the threshold voltages measured when dots are charged (after programming) and when dots are discharged (after erasing), is achievable with the use of ultra-narrow channels, and therefore deserves detailed investigation.

A possible mechanism responsible for such effect has been proposed by [74, 75], and it is based on the role played by percolating paths in the channel due to the random distribution of charged dots above the silicon wire. However, we will show experiments in which the behavior under investigation is also observed when the threshold voltage shift is measured in strong inversion, in a situation in which percolating paths cannot possibly be present.

In this chapter, we present an alternative physical explanation based
on experiments and on detailed simulations of very narrow SOI nanocrystal memories [76]. In particular, we will show that the above mentioned effect can be consistently explained by the preferential injection of electrons near the edges of the channel, that makes the dots in the oxide region surrounding the edges much more charged than those over the center of the channel.

5.1 Device fabrication and simulated structure

In Figure 5.1 we show the longitudinal and transversal sections of the fabricated devices. (Figs. 5.1a-b, respectively). The 15 nm-thick (100)-oriented channel with varying width $W$ and length $L$ of 50 nm is defined by electron beam lithography (EBL) and reactive ion etching (RIE). The SiO$_2$ tunneling layer has a thickness of 2.5 nm and is grown by rapid thermal oxidation.

The layer of silicon dots, with a density of about $5 \times 10^{11} \text{ cm}^{-2}$ and average dot diameter of 6 nm, is obtained by the phase separation of a silicon rich oxide (SiO$_{0.5}$) layer through an annealing step at 1050 C for about 3 minutes under N$_2$ [77]. The dot layer is then covered with a deposited 20 nm control oxide and, finally, polysilicon gates are defined.
5.2 Simulations

Figure 5.2: Measured threshold voltage shift as function of the channel width for a device, in the strong inversion and in the sub-threshold regime, as obtained with a writing voltage of 10 V and a programming time of 10 s.

by EBL and RIE.

In Figure 5.2 we show the threshold voltage shifts as a function of the channel width, obtained in the strong inversion and in the sub-threshold regime, for a writing voltage of 10 V and a programming time of 10 s.

As can be seen, the effect is also present in strong inversion, when the Debye length [46] can be estimated to about 3 nm, for an electron concentration \( n = 2 \times 10^{18} \text{cm}^{-3} \); percolating paths in this condition, cannot be a relevant cause of the observed effect, since for such electron density charged dots are not able to deplete the channel underneath.

5.2 Simulations

In order to investigate the decrease of the threshold voltage shift as a function of the channel width, we have performed detailed simulations of six different devices with same \( L \) (50 nm), but with \( W \) varying from 30 to 80 nm. The longitudinal and the transversal sections of the simplified structure used in the simulation of the nanocrystal memory are shown in Figures 5.1c and 5.1d: the channel section is rectangular rather than trapezoidal, the gates are metallic, and the charged dots have been mod-
Figure 5.3: Simulated threshold voltage shift as function of the channel width for a continuous fixed charge in the dot layer, computed in the sub-threshold and in the strong inversion regime.

eled by a layer of fixed charge with a thickness of 4 nm. The threshold voltage has been computed with the simplified drift/diffusion transport model described in section 4.2.1. In particular, in the sub-threshold regime, the threshold voltage has been defined as the gate voltage at which the drain-to-source current is equal to 1 nA; in the strong inversion regime it has been defined as the intercept with the gate voltage axis of the line that fits the transfer characteristic in strong inversion.

As a first attempt, in order to reproduce by simulations the effect observed in Figure 5.2, we have considered a uniform and continuous distribution of fixed charge in the dot layer, corresponding to two electrons per dot and to an experimental dot density of $5 \times 10^{11}$ cm$^{-2}$.

In Figure 5.3 the threshold voltage shift computed in the sub-threshold and in the strong inversion regime is plotted as a function of $W$. As can be noticed, simulation results are in evident contrast with experiments, since the shift slightly increases with decreasing $W$.

The discrete nature of dot and charge distribution is an aspect that might have a role, and that can be considered by assuming that charged dots obey a Poisson distribution in space.

In Fig. 5.4 we show the isosurface of the charge density ($n = 2 \times$
Figure 5.4: Isosurface of the charge density \((n = 2 \times 10^{18} \text{ cm}^{-3})\) computed for a \(V_{GS} = 1.6 \text{ V}\) and of the discrete fixed charge density in the storage layer (average charge in the dot layer \(2 \times 10^{18} \text{ cm}^{-3}\)).

\(10^{18} \text{ cm}^{-3}\) computed for \(V_{GS} = 1.6 \text{ V}\) : the spots around the channel are the isosurfaces of the discrete fixed charge density \((\rho_{fix} = 2 \times 10^{18} \text{ cm}^{-3})\), and represent each individual "charged dot". The shape of each spot is instead not particularly meaningful, as it depends both on the number of electrons per dot and on the grid.

Despite the new assumption, the same contrast is observed if a discrete distribution of fixed charge in the dot layer is considered. We have computed the threshold voltage shift over a sample of twelve devices (as in the experiments) with the same nominal dot density, but with a different discrete distribution of charged dots. Results in the sub-threshold regime are shown in Figure 5.5.

We have therefore removed the hypothesis of uniformly charged dots, and we have considered the possibility of non-homogeneous charging of the dots.

Indeed, in the devices considered, for the applied program voltage (10 V), electrons traverse the oxide via direct tunneling. An analytical
formula for the direct tunneling current reads [78]

\[ J_{DT}(\psi_b, E_{ox}) = \frac{C_1}{\left[1 - \left(\frac{\psi_b + q E_{ox} t_{ox}}{q\psi_s}\right)^{0.5}\right]^2} \times \exp \left[-\frac{C_2}{E_{ox}} \left[1 - \left(1 - \frac{q t_{ox} E_{ox}}{\psi_b E_{ox}}\right)^{1.5}\right]\right] \]  

(5.1)

with

\[ C_1 = \frac{q^4}{16\pi^2\hbar \psi_s}, \]

(5.2)

and

\[ C_2 = \frac{4}{3} \left(\frac{2m_{ox}}{\hbar q}\right)^{0.5} \psi_s^{1.5}, \]

(5.3)

where \( \psi_b \) is difference between the top of the conduction band and the Fermi level in the bulk, \( \hbar \) is the reduced Plank's constant, \( t_{ox} \) is the oxide thickness, \( m_{ox} \) is the electron effective mass in the oxide, and \( E_{ox} \) is the electric field in the oxide.
By extracting for each point at the Si/SiO$_2$ interface both $\psi_b$ and the local $E_{ox}$ in the tunnel oxide, we have been able to compute the local tunneling current.

Indeed, the electrostatics of the structure is such that close to the channel edges the electric field in the oxide is higher and the tunnel barrier lower, as extracted from simulations and shown in Figures 5.6a and 5.6b. Therefore the direct tunneling current through the oxide is larger at the edges than at the center of wire (Fig. 5.7).

Since dots are changed by tunneling current, we have then made the assumption that the charge locally stored in the dot layer is proportional to the local direct tunneling current, as in eq. (5.1).

The narrower the channel, the more effective the charge stored near the edges on the device behavior. In the limit of infinitely large channel, electrons stored in dots near the edges have no effect, and the threshold voltage shift is only due to the smaller amount of charge injected in the middle of the channel. Results are shown in Figure 5.8. As can be noticed, simulations are now able to reproduce the experimental behavior.

In Fig. 5.9 we show the charge density computed for a $V_{GS} = -0.4$ V and $n = 7.5 \times 10^{17}$ cm$^{-3}$ for devices with channel width equal to 30 and 80 nm: semitransparent surfaces correspond to simulation results when dots are discharged, while opaque surfaces are those when dots are charged. As it appears, charged dots at the corner are more effective in depleting the channel in the case the width is smaller.

We want to point out that this effect is also present if the structure does not have sharp edges, but only round edges, and that the presence of a curvature is sufficient to ensure preferential injection. For example, we consider the limit in which the sharp square edges are replaced by the arc with the minimum curvature compatible with the device structure, shown in Figure 5.10a. To this purpose, we have performed a two-dimensional simulation of the cross section, which corresponds to a device with an infinite channel length. Results in the strong inversion and in the sub-threshold regime are shown in Figure 5.10b: also this structure is able to reproduce the observed dependence of the programming window on the channel width. Indeed, also in this case, tunneling is strongly enhanced through the thin oxide in the curved regions, and therefore the dots in the regions of higher curvature are the most charged. For completeness, the same kind of plot of Figure 5.10b is shown in Figure 5.10c for a rectangular cross section.
Figure 5.6: a) Electric field in the direction perpendicular to the Si/SiO$_2$ interface and b) local barrier height seen by the electrons in the channel as a function of transversal position in the channel (shown in Fig. 5.1d) for different values of $W$. 
Figure 5.7: Local direct tunneling current density in the correspondence of the dot layer as a function of the z-coordinate.

Figure 5.8: Threshold voltage shift as a function of the channel width for a charge density in the dot layer proportional to the local tunnel current, computed in the sub-threshold and in the strong inversion regime.
Figure 5.9: Isosurface of the electron density \((n = 7.5 \times 10^{17} \text{ cm}^{-3})\) computed for a \(V_{GS} = -0.4 \text{ V}\) in case of fixed charge proportional to the tunneling current, for devices with channel width equal to 30 and 80 nm.
Figure 5.10: a) Structure of the “round” edge device; threshold voltage shift as a function of the channel width for a fixed charge in the dot proportional to the direct tunnel current for the “round” structure b) and for a rectangular cross section. c) Channel lengths tends to infinity in both cases.
5.3 Concluding remarks

In this chapter, we have studied the dependence of the programming window of SOI nanocrystal memories on the channel width. As confirmed by the experiments, the observed effect can be explained by preferential injection of charge in correspondence of channel edges, due to the reduced barrier height and increased electric field in the tunnel oxide near the channel edges.

In addition, in our case the threshold voltage shift is always proportional to the stored charge, for all the channel widths. Therefore also the threshold voltage shift is expected to decay exponentially with time as observed in experiments in Ref. [74]. In the subthreshold regime, percolating paths proposed in [74] can represent an additional mechanism, that, together with the main mechanism explained in this chapter, can reinforce such effect.
Chapter 6

Bohm Quantum Potential

The theoretical basis for the definition of an effective quantum potential is Bohm’s interpretation of quantum mechanics, which dates back to 1952 [79], and is at the origin of other more recent derivations based on a first order expansion of the Wigner equation [80], or on the so-called density gradient approach (DG) [81].

The great advantage of such a method consists in the possibility to correctly model quantum effects without the explicit solution of the Schrödinger equation and, more importantly, to be easy integrable in widespread simulation approaches based on drift-diffusion or hydrodynamic equations. Therefore, the capability of commercial device simulators, generally based on the approximate solution of the semiclassical Boltzmann equation, can be conveniently extended also to the quantum regime which is relevant in recent nanoscale semiconductor technology.

In this chapter, after the theoretical derivation of the DG, we will present an alternative approach to the DG formalism, based on Bohm quantum theory, which for this reason, we have referred as the Bohm Quantum Potential (BQP) approach [82], and we will make comparisons with the self-consistent solution of the Poisson/Schrödinger equations, applying both models on various nanoscale structures, characterized by different kind of quantum confinement.
6.1 Density gradient approach

Since the direct solution of the Boltzmann transport equation (BTE) is too computationally demanding, the balance equation approach [83] is generally adopted in order to achieve faster but still accurate results. However, due to the semiclassical treatment of electrons dynamic, these equations do not take into account quantum confinement: the Wigner equation has then to be solved, in order to consider quantum effects.

In the following, starting from the Wigner transport equation (WTE) [84], and following the same guidelines adopted to derive the drift-diffusion and energy balance equations from the BTE [85], we will compute the first two momenta of the WTE, in order to find the quantum corrections to the drift-diffusion and to the energy balance equations.

Quantum Drift-Diffusion equation

Considering time dependence implicit, the one particle Wigner equation [86] reads

\[
\frac{\partial f_W(r, p)}{\partial t} + v \nabla f_W(r, p) = \frac{2}{\hbar} \sin \left( \frac{\hbar \mathbf{p} \cdot \nabla r}{2} \right) V(r) f_W(r, p),
\]

(6.1)

where \(v\) is the velocity, \(V(r)\) is the potential and \(\nabla r\) is the gradient with respect to \(r\), applied only to the potential. If the argument of the \(\sin\) operator is very small, from the Taylor expansion \(\sin(\alpha) \approx \alpha - \frac{\alpha^3}{6}\), so the \(\sin\) operator can be expressed as

\[
\frac{2}{\hbar} \sin \left( \frac{\hbar \mathbf{p} \cdot \nabla r}{2} \right) \approx \frac{2}{\hbar} \left[ \frac{\hbar \mathbf{p} \cdot \nabla r}{2} - \frac{1}{6} \left( \frac{\hbar \mathbf{p} \cdot \nabla r}{2} \right)^3 \right].
\]

(6.2)

Considering in the expansion only the terms up to the second order of \(\hbar\), the Wigner equation is reduced to the Boltzmann equation,

\[
\frac{\partial f_W(r, p)}{\partial t} + v \nabla f_W(r, p) = -qE\nabla_p f_W(r, p) + \partial_t f_W(r, p) |_C
\]

(6.3)

where the collision term has been introduced and \(E\) is the electric field. For simplicity suppose now to work in one dimension both in the direct and in the indirect space.

In the stationary case and using the RTA approximation, eq. (6.3) becomes

\[
-\frac{p}{m} \partial_r f_W(r, p) - (-q)E \partial_p f_W(r, p) = \frac{f_W(r, p) - f_{W_0}(r, p)}{\tau_c}
\]

(6.4)
where $\tau_c$ is the average relaxation time and $f_{W_0}(r,p)$ is the Wigner function at the equilibrium.

Let us now calculate the first momentum of the Wigner equation, multiplying both members by $\frac{p}{m}$, and integrating over $p$.

$$-\frac{p^2}{m^2} \partial_r f_W(r,p) - \frac{p}{m} (-q) E \partial_p f_W(r,p) = \frac{p}{m} \left( \frac{f_W(r,p)}{\tau_c} - \frac{f_{W_0}(r,p)}{\tau_c} \right).$$  \hspace{1cm} (6.5)

From the second term on the left hand side, we obtain

$$\int_{-\infty}^{+\infty} \frac{p}{m} \frac{f_W(r,p)}{\tau_c} dp = \frac{v_d n}{\tau_c}, \hspace{1cm} (6.6)$$

where $v_d$ is the drift velocity and $n$ is the electron concentration.

From the second term on the right hand side, we obtain

$$\int_{-\infty}^{+\infty} \frac{p}{m} \frac{f_{W_0}(r,p)}{\tau_c} dp = 0 \hspace{1cm} (6.7)$$

because $f_{W_0}(r,p)$ is symmetric in the $p$-space, while $p$ is an odd function, so the integration over $p$ is null.

From the first term on the left hand side, we obtain

$$\int_{-\infty}^{+\infty} \frac{p^2}{m^2} \partial_r f_{W_0}(r,p) dp = \frac{\partial_r}{m^2} \int_{-\infty}^{+\infty} p^2 f_{W_0}(r,p) dp \hspace{1cm} (6.8)$$

Let us solve the integral $\int p^2 f_{W_0}(r,p) dp$.

From [87], the Wigner function can be expressed as

$$f_W(r,p) = \frac{e^{-\beta \left( \frac{p^2}{2m} + V \right)}}{h} \left[ 1 - \frac{\hbar^2 \beta^2}{8m} \left[ \partial_r^2 V - \frac{\beta}{3} \left( (\partial_r V)^2 + \frac{p^2}{m} \partial_r^2 V \right) \right] + O(\hbar^2) \right] \hspace{1cm} (6.9)$$

where $\beta = \frac{1}{K_B T}$.

The following formulas will result to be very useful

$$\int_{-\infty}^{+\infty} e^{-\alpha p^2} dp = \sqrt{\frac{\pi}{\alpha}} \hspace{1cm} (6.10)$$

$$\int_{-\infty}^{+\infty} p^2 e^{-\alpha p^2} dp = \frac{1}{2\alpha} \sqrt{\frac{\pi}{\alpha}} \hspace{1cm} (6.11)$$
where $\alpha = \frac{\beta}{2m}$.

First of all, it is necessary to compute the electron density

$$n(r) = \int_{-\infty}^{+\infty} f_W(r,p) dp$$

(6.12)

Following the integrals above and from eq.(6.9),

$$n = f_W(r) = \int_{-\infty}^{+\infty} e^{-\frac{\beta}{\hbar} \left( \frac{\mathbf{p}^2}{2m} + V \right)} \left[ 1 - \frac{\hbar^2 \beta^2}{8m} \left[ \partial_r^2 V + \frac{\beta}{m} \partial_r^2 V \right] \right] dp$$

$$= e^{-\beta V} \frac{\sqrt{\pi} \alpha}{\hbar} \left[ 1 - \frac{\hbar^2 \beta^2}{8m} \left[ \partial_r^2 V + \frac{2m}{m\beta^2} \partial_r^2 V \right] \right]$$

$$= e^{-\beta V} \frac{\sqrt{2m\pi}}{\beta} \left[ 1 - \frac{\hbar^2 \beta^2}{8m} \left[ \frac{2}{3} \partial_r^2 V - \frac{\beta}{3} \left( \partial_r V \right)^2 \right] \right]$$

$$= e^{-\beta V} \frac{\sqrt{2m\pi}}{\beta} \left[ 1 - \frac{\hbar^2 \beta^2}{12m} \left[ \frac{2}{3} \partial_r^2 V - \frac{\beta}{2} \left( \partial_r V \right)^2 \right] \right].$$

(6.13)

If we had performed an integration over the three momenta $p_x$, $p_y$ and $p_z$, we would have found the following form (as in [87])

$$n(r) = \left[ \frac{m^*}{2\pi\hbar^2} \right]^\frac{3}{2} e^{-\beta V} \left[ 1 - \frac{\hbar^2 \beta^2}{12m} \left( \nabla^2 V - \frac{\beta}{2} \nabla V \cdot \nabla V \right) + O(\hbar^2) \right]$$

(6.14)

Using the above results and the following integral,

$$\int_{-\infty}^{+\infty} p^4 e^{-\alpha p^2} dp = \frac{3}{4\alpha^2} \sqrt{\frac{\pi}{\alpha}}$$

(6.15)
6.1. Density gradient approach

the integral \( \int p^2 f_W(r,p) dp \) becomes

\[
\int_{-\infty}^{+\infty} p^2 f_W(r,p) dp = e^{-\beta V} \frac{m}{2\alpha h} \sqrt{\frac{\pi}{\alpha}} \left[ 1 - \frac{\hbar^2 \beta}{8m} [\partial_r^2 V - \frac{\beta}{3} \left( (\partial_r V)^2 + \frac{3}{2m\alpha} \partial_r^2 V \right)] \right]
\]

\[
= e^{-\beta V} \frac{m}{2\alpha h} \sqrt{\frac{\pi}{\alpha}} \left[ 1 - \frac{\hbar^2 \beta}{8m} [\partial_r^2 V - \frac{\beta}{3} \left( (\partial_r V)^2 + \frac{3}{2m\alpha} \partial_r^2 V \right)] \right]
\]

\[
= e^{-\beta V} \frac{m}{2\alpha h} \sqrt{\frac{\pi}{\alpha}} \left[ 1 + \frac{\hbar^2 \beta^3}{24m} (\partial_r V)^2 \right].
\]

(6.16)

From eq.(6.13), the electronic charge density can be expressed as

\[ n(r) \approx \frac{e^{-\beta V}}{h} \sqrt{\frac{2m\pi}{\beta}}, \]

(6.17)

and consequently,

\[ V = -\frac{1}{\beta} \ln \left[ \frac{n}{\frac{1}{h} \sqrt{\frac{2m\pi}{\beta}}} \right]. \]

(6.18)

In the end, the integral can be expressed as

\[
\int_{-\infty}^{+\infty} p^2 f_W(r,p) dp = \frac{m}{\beta} n + \frac{m}{\beta} \frac{\hbar^2 \beta^3}{24m} \left[ \partial_r \left( -\frac{1}{\beta} \ln n \right) \right]^2
\]

\[
= \frac{m}{\beta} n + \frac{\hbar^2}{24} [\partial_r (\ln n)]^2
\]

(6.19)

Let us now consider the first term of the right hand side of (6.5)

\[
\int_{-\infty}^{+\infty} \frac{p}{m} (-q) \mathcal{E} \partial_p f_W(r,p) dp,
\]

(6.20)

Noting that

\[ p \partial_p f_W(r,p) = \partial_p (pf_W(r,p)) - f_W(r,p) \partial_p p \]

(6.21)

and that
\[ \partial_p (pf_W(r,p)) \approx 0 \quad (6.22) \]

because \( f_W(r,p) \) approaches zero rapidly for large \( p \), eq.(6.20) becomes

\[
\int_{-\infty}^{+\infty} \frac{p}{m} (-q) \xi \partial_p f_W(r,p) dp = - \int_{-\infty}^{+\infty} \frac{(-q)}{m} \xi f_W(r,p) dp = \frac{(-q) \xi n}{m} . \quad (6.23)
\]

The momentum of order one of the Wigner equation has therefore the following form :

\[
\frac{v_d n}{\tau_c} = - \frac{\partial_r m}{m^2} \left[ \frac{m}{\beta} n + \frac{\hbar^2}{24} n \left( \partial_r \ln(n) \right)^2 \right] + \frac{(-q) \xi n}{m} , \quad (6.24)
\]

We can express (6.24) as a function of the current density, noting that \( J_n = (-q)n v_d \). Multiplying both members of (6.24) by \(-q\), we obtain

\[
\frac{J_n}{\tau_c} = q \frac{\partial_r m}{m^2} \left[ \frac{m}{\beta} n + \frac{\hbar^2}{24} n \left( \partial_r \ln(n) \right)^2 \right] + \frac{q^2 \xi n}{m} . \quad (6.25)
\]

and defining

\[
\mu_n \triangleq \frac{q \tau_c}{m} \quad (6.26)
\]

we finally obtain

\[
J_n = \mu_n K_B T \partial_r n + q \mu_n n \xi + \frac{q \tau_c}{m^2} \partial_r \left[ \frac{\hbar^2}{24} n \left( \partial_r \ln(n) \right)^2 \right] . \quad (6.27)
\]

We note that eq. (6.27), as compared to the drift-diffusion equation, has an additional term which directly derives from the approximate solution of the WTE, and that, for this reason, can be considered as a quantum correction term.

**Quantum energy balance equation**

We are now focusing our attention to the modification of the energy balance equation, adding a quantum correction term \([88]\). The second momentum of the Boltzmann equation reads

\[
\partial_t w = - \nabla \cdot F_w + J_n \cdot \xi - \frac{w - w_0}{\tau_c} \quad (6.28)
\]
where \( w \) is the kinetic energy density, \( F_w \) is its related flux and \( w_0 \) is the kinetic energy density at the equilibrium. Let us consider the kinetic energy density \( w \) within the Wigner function formalism.

\[
\int_{-\infty}^{+\infty} \frac{p^2}{2m} f_W (r, p) dp = \frac{K_B T n}{2} + \frac{\hbar^2}{48} n (\partial_r \ln n)^2
\]  

(6.29)

Considering the equipartition energy theorem, in 3D we have

\[
\int_{-\infty}^{+\infty} \frac{p^2}{2m} f_W (r, p) dp = \frac{3}{2} K_B T n + \frac{\hbar^2}{48} n |\nabla \ln n|^2.
\]  

(6.30)

Till this moment we have neglected the kinetic term, or better, we have neglected the kinetic term due to the drift component of the velocity as compared to the thermal energy. If we consider also this term, the kinetic energy density becomes

\[
w = \frac{3}{2} K_B T n + \frac{1}{2} m |v_d|^2 + \frac{\hbar^2}{48} n |\nabla \ln n|^2,
\]  

(6.31)

that is to say

\[
w = w' + \frac{\hbar^2}{48} n |\nabla \ln n|^2,
\]  

(6.32)

where \( w' \) is the kinetic energy density generally used in the energy balance equation. We can then consider as well, the \( \frac{\hbar^2}{48} n |\nabla \ln n|^2 \) term as a quantum correction term to the kinetic energy.

### 6.2 Bohm quantum potential approach

We will now focus our attention on the derivation of the Bohm quantum potential approach. With respect to the previously described density gradient approach, the presented approach has few advantages:

- It is independent of the transport model (drift-diffusion or hydrodynamic).
- It can straightforwardly include Fermi-Dirac statistics
- It provides an additional degree of freedom for calibration, which is particularly useful for considering non planar device structures.
- It can be discretized in a way to exhibit very stable convergence properties.
6.2.1 Bohm’s potential

Let us suppose that we want to solve the many body Schrödinger equation with density functional theory.

According to Bohm’s interpretation of quantum mechanics, if we write the wave function $\Psi$ in a polar representation,

$$\Psi = R \exp \left( \frac{iS}{\hbar} \right), \quad (6.33)$$

where $R$ is the modulus of $\Psi$ and $S$ has the dimension of an “action” (energy $\times$ time) and is the phase of $\Psi$ multiplied by $\hbar$, the Schrödinger equation reads

$$-\frac{\hbar^2}{2} \nabla \left\{ M^{-1} \nabla \left[ R \exp \left( \frac{iS}{\hbar} \right) \right] \right\} + VR \exp \left( \frac{iS}{\hbar} \right) = ER \exp \left( \frac{iS}{\hbar} \right) \quad (6.34)$$

All terms in (6.34) are real quantities, except $i$, therefore it is possible to split the above equation in a real part, which reads

$$\nabla (M^{-1}R^2\nabla S) = 0, \quad (6.35)$$

and in an imaginary part, which is

$$-\frac{\hbar^2}{2} \nabla \left( M^{-1} \nabla R \right) + \frac{1}{2} \nabla S \cdot M^{-1} \cdot \nabla S + V = E \quad (6.36)$$

Let us just remind that $R^2$ a probability density per unit volume and $M^{-1}\nabla S$ is the local velocity of the probability density flux (or, according to Bohm’s interpretation of quantum mechanics, the local velocity of the particle associated to the wave function).

Supported by these considerations, equation (6.35) may be interpreted as a continuity equation of the probability density, since it says that the divergence of the probability density flux $R^2 \times M^{-1}\nabla S$ is zero.

Equation (6.36), on the other hand, ensures energy conservation: the total energy $E$ is equal to the sum of the potential energy $V$, of the kinetic energy $\frac{1}{2} \nabla S \cdot M^{-1} \cdot \nabla S$, and of the term

$$Q = -\frac{\hbar^2}{2} \frac{\nabla (M^{-1} \nabla R)}{R}, \quad (6.37)$$

which is interpreted as the “quantum potential” (the so called “Bohm’s potential”).
Indeed, it can be easily observed that if we remove Q, or force it to zero, equations (6.35) and (6.36) describe a completely classical system. In other terms, this means that the Schrödinger equation can be recast in two coupled continuity and energy conservation equations, in which quantum mechanics manifests itself only in the quantum potential term.

The possibility of taking into account all quantum mechanical aspects of the system by only adding the quantum potential to the mechanical energy of the system is very intriguing from the point of view of the numerical simulation of a quantum system, if one is able to compute Q efficiently.

Let us consider more closely the properties of Q. From (6.37) it can be seen that Q depends on the amplitude of the wave function, that can only be known exactly if one solves the Schrödinger equation or the system given by equations (6.35-6.36). Of course, from the point of view of simulations, one would prefer to compute Q in an easier, even approximate, way.

In addition, every eigenfunction of the Schrödinger equation is associated with its own quantum potential. For example, the probability density of the i-th eigenfunction Ψ_i is subjected the quantum potential

\[ Q_i = -\frac{\hbar^2}{2} \nabla \left( \frac{M^{-1} \nabla R_i}{R_i} \right) \]

where \( R_i = |\Psi_i| \). In other words, to exactly simulate a many-particle quantum system with density functional theory, we should consider a different quantum potential for each single particle (Kohn-Sham) eigenfunction.

### 6.2.2 Effective Quantum Potential

To make the problem manageable from the numerical point of view, we would like to have a quantum potential correction term that i) may be computed without solving the Schrödinger equation, and ii) is the same for all identical particles in the system.

To this aim, we can use a sort of “mean field approximation” such as that used in the case of electrostatic interaction in a many body system.

We define an “Effective quantum potential” \( Q_{\text{eff}} \) defined as a weighted average of the quantum potentials of all eigenstates, i.e.:

\[
Q_{\text{eff}}(r) = \frac{\sum_i f_i R_i^2(r) Q_i(r)}{\sum_i f_i R_i^2(r)} = -\frac{\hbar^2}{2} \frac{\sum_i f_i R_i(r) \nabla \left[ M^{-1} \nabla R_i(r) \right]}{\sum_i f_i R_i^2(r)}
\] (6.39)
The weights are represented by the probability density of all states in the considered position ($f_i$ is the occupation factor of state $i$).

Now, we would like to express $Q_{\text{eff}}$ in terms of $R_i^2$, instead of $R_i$, therefore we rewrite (6.39) as

$$Q_{\text{eff}} = -\frac{\hbar^2}{2} \sum_i f_i \sqrt{R_i^2} \nabla \left( M^{-1} \nabla \sqrt{R_i^2} \right) \sum_i f_i R_i^2,$$

(6.40)

which, after upwinding the derivatives, becomes

$$Q_{\text{eff}} = -\frac{\hbar^2}{2} \left\{ \sum_i \frac{1}{2} f_i \nabla \left[ M^{-1} \nabla (R_i^2) \right] - \sum_i \frac{1}{4} f_i \nabla (R_i^2) \frac{M^{-1} \nabla (R_i^2)}{R_i^2} \right\} \sum_i f_i R_i^2,$$

(6.41)

If, without loss of generality, we assume to consider a quantum system of electrons, the electron density $n$ is represented by the sum term at the denominator of (6.41), i.e.

$$n = \sum_i f_i R_i^2.$$  

(6.42)

We want to point out again that the occupation factor $f_i$ enters $Q_{\text{eff}}$ only as a weighting term, and therefore is not differentiated. Indeed, the quantum potential for each state $i$ does not depend on state occupancy, but only on the shape of the eigenfunction, i.e., of the density of states. However, at equilibrium $f_i$ does not depend on position, therefore can be moved to the right of the differentiation operators. In out of equilibrium conditions, this choice will have to be reconsidered.

In addition, if $M$ does not depend on the state considered, it can be taken out of the sum operator. Often, this is not the case: in silicon, for example, in the conduction band where we have three possible values of $M$, one for each pair of minima; in the valence band, we must have two different effective mass tensors for light and heavy holes. Therefore, it is useful to define the parameter $\gamma$ as follows:

$$\gamma \equiv \frac{\sum_i f_i \nabla M^{-1} \cdot \nabla (R_i^2) - \sum_i f_i M^{-1} \nabla^2 (R_i^2)}{
\nabla M^{-1} \cdot \sum_i f_i \nabla (R_i^2) - \sum_i f_i \nabla^2 (R_i^2)}$$

$$= \frac{\sum_i \nabla \left[ M^{-1} \nabla (f_i R_i^2) \right]}{\n\nabla \left[ M^{-1} \nabla (\sum_i f_i R_i^2) \right]}.$$  

(6.43)
so that we can write the first sum term of eq. (6.41) as

\[
\sum_i \nabla [M^{-1} \nabla (f_i R_i^2)] = \gamma \nabla [M^{-1} \nabla \left( \sum_i f_i R_i^2 \right)]
\]

\[= \gamma \nabla (M^{-1} \nabla n) \quad (6.44)\]

It is evident that if the effective mass tensor is constant for all states, as for example in GaAs conduction band in the parabolic band approximation, we have \(\gamma = 1\). Error is minimized if the effective mass tensor which is taken out of the sum operation is the most probable value.

To write the second sum of (6.41) is a simple way we have to define a second parameter \(\eta\) as

\[
\eta \equiv \frac{\sum_i f_i R_i^2}{\sum_i f_i \nabla R_i^2 \cdot \sum_i f_i \nabla [M^{-1} (R_i^2)]} \sum_i \frac{f_i \nabla (R_i^2) M^{-1} \nabla (R_i^2)}{R_i^2} \quad (6.45)
\]

Here, it is worth noticing that in the case of only one occupied state, we have \(\eta = 1\).

Using results of equations (6.42), (6.43), and (6.45) the second sum between parenthesis of (6.41) becomes

\[
\sum_i \frac{1}{4} \frac{f_i \nabla (R_i^2) M^{-1} \nabla (R_i^2)}{R_i^2} = \gamma \eta \frac{1}{4} \frac{\nabla n \cdot M^{-1} \cdot \nabla n}{n} \quad (6.46)
\]

Finally, we can write (6.41) as

\[
Q_{\text{eff}} = -\frac{\hbar^2}{2} \gamma \left[ \frac{\nabla (M^{-1} \nabla n)}{2n} - \frac{\eta \nabla n \cdot M^{-1} \cdot \nabla n}{4 n^2} \right]
\]

\[= -\frac{\hbar^2}{2} \gamma \frac{\nabla \left[ M^{-1} \nabla (n^\alpha) \right]}{n^\alpha} \quad (6.47)\]

The last equality is obtained in a straightforward way by defining \(\alpha \equiv 1 - \eta/2\).

First, we notice that \(\gamma\) and \(\eta\) (or, equivalently, \(\alpha\)) are not known a priori, and depend in general on position, on the device structure, and on the applied bias. Instead, to make the expression (6.47) tractable, we have at least to consider them constant parameters. We will treat them as fitting parameters, i.e., will choose for \(\gamma\) and \(\eta\) (or \(\alpha\)) the values that provide the best fit between the full Poisson-Schrödinger solution and the semiclassical solution corrected with the effective quantum potential.
To improve the fitting, retain some flexibility without losing in simplicity, we shall try to extract fitting parameters for a few different classes of device structures.

**Out of equilibrium**

Out of equilibrium, the occupation factor $f_i$ of state $i$ may depend on position, and the above equations are not directly applicable. Therefore, we cannot assume, as we do in equilibrium, that $\nabla n = \sum_i f_i \nabla (R_i^2)$. In that case we have to simply define a special partial differentiation operator $\tilde{\nabla}$ as

$$\tilde{\nabla} n = \sum_i f_i \nabla (R_i^2)$$

(6.48)

Of course the operator $\tilde{\nabla}$ reduces to $\nabla$ at equilibrium. Instead, out of equilibrium we have to use $\tilde{\nabla}$. If, for example, a transport model such as the hydrodynamic model is used, the occupation factor of an electron state depends on the quasi-Fermi level $E_{fn}$ and the carrier temperature $T_e$, which are both a function of position, i.e., $f_i = f_i[E_{fn}(\mathbf{r}), T_e(\mathbf{r})]$, then the relationship between $\tilde{\nabla}$ and $\nabla$ is simply

$$\nabla n = \tilde{\nabla} n + \left( \frac{\partial n}{\partial E_{fn}} \nabla E_{fn} + \frac{\partial n}{\partial T_e} \nabla T_e \right)$$

(6.49)

In plain language, $\tilde{\nabla} n$ is a partial gradient of $n$, that does not takes into account temperature and quasi-fermi level gradients. Out of equilibrium, independently of the transport model used, equation (6.47) becomes:

$$Q_{\text{eff}} = -\frac{\hbar^2}{2} \gamma M^{-1} \tilde{\nabla} \left[ \frac{M^{-1} \tilde{\nabla} (n^\alpha)}{n^\alpha} \right]$$

(6.50)

### 6.3 Discretization

Suppose we want to discretize (6.50) on a Delauney mesh using box integration. Some care is required in handling the operator $\tilde{\nabla}$. Let us consider Fig. 6.1, which shows the Voronoi cell around $\mathbf{r}_l$, the position of the $l$-th point of the mesh. We want to use box integration to discretize the equation around point $l$. First, let us define a function $\tilde{n}(\mathbf{r}, l)$ as follows

$$\tilde{n}(\mathbf{r}, l) = \sum_i f_i(\mathbf{r}_l)R_i^2(\mathbf{r}).$$

(6.51)
6.3. Discretization

Figure 6.1: Voronoi cell associated to the $l$-th point of the mesh, at position $r_l$.

As can be seen, for $r = r_l$, $\hat{n}$ and $n$ are coincident. More importantly, we have

$$\nabla n(r) = \nabla \hat{n}(r, l). \quad (6.52)$$

In order to perform box integration we multiply all terms of (6.50) by $n^\alpha$ and integrate over the volume $\Omega_l$ of the Voronoi cell

$$\int_{\Omega_l} Q_{\text{eff}} n^\alpha dr = \int_{\Omega_l} -\frac{\hbar^2}{2} \frac{\gamma}{\beta} \left[ M^{-1} \hat{n}^\alpha \right] dr$$

$$= \int_{\Omega_l} -\frac{\hbar^2}{2} \nabla \left[ M^{-1} \nabla (\hat{n}^\alpha) \right] dr. \quad (6.53)$$

In the left-hand side we approximate $n$ with as a constant term $n = n(r_l)$ in the Voronoi cell. In the right-hand side we can use, as usual, Green’s theorem, since now we are dealing with a “normal” divergence operator. With box integration (6.53) becomes

$$Q_{\text{eff}}(r_l) n^\alpha(r_l) \Omega_l = -\frac{\hbar^2}{2} \gamma \sum_j m_{ij} \left( \hat{n}^\alpha(r_j, l) - n^\alpha(r_l) \right) d_{ij} \sigma_{ij}, \quad (6.54)$$

where $d_{ij}$ is the distance between $r_j$ and $r_l$, $\sigma_{ij}$ in 3D is the area of the face of the Voronoi cell separating point $l$ and $j$, as can be seen in
Fig. 6.1. If we call \( \chi_{lj} = \{\chi_{ljx}, \chi_{ljy}, \chi_{ljz}\} \) the unitary vector representing the direction from \( r_l \) to \( r_j \), we have

\[
m_{lj}^{-1} \equiv m_x^{-1} \chi_{ljx} + m_y^{-1} \chi_{ljy} + m_z^{-1} \chi_{ljz}.
\]

(6.55)

Such definition of \( m_{lj} \) is rigorously correct only if the direction from \( r_l \) to \( r_j \) is a principal direction of the effective mass tensor. Otherwise, it is typically a good approximation. From (6.54), dividing both terms by \( n^\alpha \), we have

\[
Q_{\text{eff}}(r_l)\Omega_l = \frac{\hbar^2}{2} \sum_j \sigma_{lj} m_{lj}^{-1} \left\{ 1 - \left[ \frac{\tilde{n}(r_l, l)}{n(r_l)} \right]^\alpha \right\}
\]

(6.56)

One possibility could be to directly use this discretization. However, we have found that it is often unstable or slow to converge. We have obtained better results by making some modifications to equation (6.56), that depend on the statistics of charge carriers.

### 6.3.1 Maxwell-Boltzmann statistics

In the case of Maxwell-Boltzmann (MB) statistics the electron density is given by the following expression

\[
n(r) = N_C(r) \exp \left( -\frac{E_C(r) + Q_{\text{eff}}(r) - E_{fn}(r)}{k_B T_e(r)} \right),
\]

(6.57)

where the expression already includes the effective quantum potential, as a term simply added to the conduction band \( E_C \). Here \( N_C \) is the effective density of states in the conduction band. In general, out of equilibrium, \( E_{fn} \) and \( T_e \) depend on position. Since \( E_{fn} \) and \( T_e \) do not depend on the density of states, but only on the transport properties, their position dependence is removed in the definition of \( \tilde{n} \), which is given by

\[
\tilde{n}(r, l) = N_C(r) \exp \left( -\frac{E_C(r) + Q_{\text{eff}}(r) - E_{fn}(r_l)}{k_B T_e(r_l)} \right)
\]

(6.58)

Equation (6.56) becomes:

\[
Q_{\text{eff}}(r_l)\Omega_l = \frac{\hbar^2}{2} \sum_j \sigma_{lj} m_{lj}^{-1} \times \left\{ 1 - \left[ \frac{N_C(r_j)}{N_C(r_l)} \right]^\alpha \right. \exp \left( \frac{E_C(r_l) + Q_{\text{eff}}(r_l) - E_C(r_j) - Q_{\text{eff}}(r_j)}{k_B T_e(r_l)} \right) \left\} \right\}
\]

(6.59)
6.3. Discretization

Care has to be taken with insulating materials in this case, because putting \( N_C = 0 \) would cause a division by zero error. The problem is avoided by considering the insulating layer as a semiconductor with a finite value of \( N_C \) but different bandgap and electron affinity. For example, in MOS structures we assign to all dielectric materials the same \( N_C \) as silicon. This allows to remove the ratio of \( N_C \)'s in the preceding equation, and to speed up the calculations.

All considerations equally apply to holes.

It is worth noticing that the nonlinear algebraic system of equations described by (6.59) does not depend on the transport regime. From our simulations, it is the most stable discretization scheme.

6.3.2 Fermi-Dirac statistics

In the case of Fermi-Dirac (FD) statistics the electron density is given by

\[
n(r) = N_C(r) F_{1/2} \left( -\frac{E_C(r) + Q_{\text{eff}}(r) - E_{fn}(r)}{k_B T_e(r)} \right), \tag{6.60}
\]

where \( F_{1/2} \) is the Fermi-Dirac function of order one half. For a generic mesh point \( r_l \) we define

\[
F_{1/2}^l \equiv F_{1/2} \left( -\frac{E_C(r_l) + Q_{\text{eff}}(r_l) - E_{fn}(r_l)}{k_B T_e(r_l)} \right). \tag{6.61}
\]

We also define \( \tilde{F}_{1/2}^{j,l} \) as

\[
\tilde{F}_{1/2}^{j,l} \equiv F_{1/2} \left( -\frac{E_C(r_j) + Q_{\text{eff}}(r_j) - E_{fn}(r_l)}{k_B T_e(r_l)} \right) \tag{6.62}
\]

\( \tilde{n} \) is given by

\[
\tilde{n}(r, l) = N_C(r) F_{1/2} \left( -\frac{E_C(r) + Q_{\text{eff}}(r) - E_{fn}(r_l)}{k_B T_e(r_l)} \right). \tag{6.63}
\]

Simple substitution in (6.56) yields:

\[
Q_{\text{eff}}(r_l) \Omega_l = \frac{\hbar^2}{2} \sum_j \frac{\sigma_{lj}}{d_{lj}} m_{ij}^{-1} \left\{ 1 - \left[ \frac{N_C(r_j) \tilde{F}_{1/2}^{j,l}}{N_C(r_l) F_{1/2}^l} \right]^\alpha \right\}. \tag{6.64}
\]
One could use this discretization, but a little simplification allows us to have a more manageable equation, without loss of generality. To first order,

$$\tilde{F}_{1/2}^{j,l} = F_{1/2} + \left( \frac{E_C(r_l) + Q_{\text{eff}}(r_l) - E_C(r_j) - Q_{\text{eff}}(r_j)}{k_B T_e(r_l)} \right) F_{-1/2}^l,$$

(6.65)

where $F_{-1/2}$ is the Fermi integral of order $-1/2$ computed in $r_l$:

$$F_{-1/2}^l \equiv F_{-1/2}^l \left( \frac{E_C(r_l) + Q_{\text{eff}}(r_l) - E_{fn}(r_l)}{k_B T_e(r_l)} \right).$$

(6.66)

By substituting (6.65) in (6.64) we obtain

$$Q_{\text{eff}}(r_l) \Omega_l = \hbar \sum_j \sigma_{lj} m_j^{-1} \times \left\{ 1 - \left( \frac{N_C(r_l)}{N_C(r)} \right)^{\alpha} \left( 1 + \frac{F_{1/2}^l}{F_{-1/2}^l} \left( \frac{E_C(r_l) + Q_{\text{eff}}(r_l) - E_{fn}(r_l)}{k_B T_e(r_l)} \right) \right) \right\}$$

(6.67)

which can be conveniently written as

$$Q_{\text{eff}}(r_l) \Omega_l = \hbar \sum_j \sigma_{lj} m_j^{-1} \times \left\{ 1 - \left( \frac{N_C(r_l)}{N_C(r)} \right)^{\alpha} \left[ \exp \left( \nu_l \alpha \frac{E_C(r_l) + Q_{\text{eff}}(r_l) - E_{fn}(r_l)}{k_B T_e(r_l)} \right) \right] \right\},$$

(6.68)

where

$$\nu_l = \frac{F_{-1/2}^l}{F_{1/2}^l} \equiv \nu \left( \frac{E_C(r_l) + Q_{\text{eff}}(r_l) - E_{fn}(r_l)}{k_B T_e(r_l)} \right).$$

(6.69)

It is worth noticing that when the Fermi-Dirac statistics tends to the Maxwell-Boltzmann statistics, $F_{1/2}^l$ tends to $F_{1/2}^l$ and therefore $\nu_l$ tends to 1 and (6.68) reduces to (6.59). In typical operating conditions, the term $\nu_l$ is rather close to one.

### 6.3.3 Boundary conditions

The following boundary conditions are applied to the equation for the solution of the algebraic system of equations (6.59) and (6.68), depending on the nature of the boundary point:
• If the \( m \)-th grid point belongs to a contact, or to a metal gate, we force Dirichlet boundary conditions, i.e. \( Q_{\text{eff}}(r_m) = 0 \).

• If the \( m \)-th grid point belongs to a point to which zero-Field Neumann boundary conditions for the electrostatic potential are enforced, we force Neumann boundary conditions on the term \( (E_C + Q_{\text{eff}}) \). This means putting to zero the argument of the exponential in the term of the sums of equations (6.59) and (6.68) corresponding to the boundary.

## 6.4 Numerical Algorithm

In this section we describe how the correction due to the effective quantum potential may be added to an existing device simulation code, in which both drift-diffusion and hydrodynamic transport models are implemented.

As already said, the great advantage of the present method is that the correction does not depend on the transport model adopted.

In our opinion, the addition of an external cycle provides a non-invasive solution, and ensures that convergence properties of the pre-existing algorithm are not undermined.

Our proposed algorithm is the following

1. Start from an initial guess of the potential and carrier temperature (for MB statistics) and of the potential, carrier temperature, and quasi-Fermi level (or electron density) for FD statistics.

2. Solve with Newton-Raphson the system of algebraic equations (6.59) for MB statistics and (6.68) for FD statistics. In the latter case, \( \nu_l \) can be discarded in the Jacobian.

3. Now \( Q_{\text{eff}} \) should be considered fixed, simply added to the conduction band profile, and the pre-existing system of Poisson, continuity, and possibly, energy balance equation should be solved with the usual means (Newton-Raphson, Gummel, and the like).

4. Repeat step 2, with the updated values of potential, temperature, and electron density

5. Repeat step 3 and 4 until convergence on the external cycle.
In typical conditions, we have found that only 3-5 external iteration are needed, leading to an increase of computing time by a factor 2.

Quantum correction can be extended to all the device simulation domain.

Systems defined by (6.59) and (6.68) converge without particular problems with Newton-Raphson.

6.5 Fitting parameters

We have already said that $\gamma$ and $\alpha$ are approximated with constant parameters providing the best fit with results from the Poisson-Schrödinger equation.

With respect to recently proposed density gradient corrections, the method proposed has the significant advantage of providing two fitting parameters, instead of one, and therefore an additional degree of freedom for calibration.

While it is clear that a different pair of parameters $\gamma$ and $\alpha$ is required for each material, we also believe that different classes of structures should require different pairs of parameters. For example, bulk devices, with a triangular confining potential, ultra-thin Soi MOSFETs, or FinFETs, with a rectangular confining potential, and triple gate FETs, with a two-dimensional quantum confinement, are likely to require different sets of effective potential parameters.

In the following, we shall present a comparison between results from a self consistent Poisson-Schroedinger solver and a semiclassical solver with MB distribution corrected with an effective quantum potential. As shall be seen, parameters $\gamma$ and $\alpha$ widely vary in the examples. When the calibration will be completed, the software user will have to choose among few pairs of $\gamma$ and $\alpha$ depending on the device structure at hand.

6.6 Simulation examples

Here we present simulations performed on simple structures of general interest in order to understand and verify the accuracy of the proposed model. In particular, we focus on devices in which the quantum confinement of electrons is predominant along one direction (MOS capacitors) and those in which electrons are confined in a quasi-one dimensional wire (FinFETs).
To this purpose, we have compared results obtained with the self-consistent solution of the Poisson/Schrödinger equation with those derived by with a semiclassical code corrected with the effective Bohm quantum potential using, for each structure, $\alpha$ and $\gamma$ as fitting parameters.

### 6.6.1 MOS capacitor

The first considered structure is a MOS capacitor with oxide thickness of 2 nm and acceptor concentration in the bulk $N_A = 1 \times 10^{18}$ cm$^{-3}$. In our simulations we have included a metal gate, in order to study only the confinement of electrons in the channel, and to neglect other effects such as quantization and depletion of the polysilicon.

We have assumed a longitudinal mass for the electrons along the direction perpendicular to the Si/SiO$_2$ interface, since the longitudinal states have the highest occupancy.
In Fig. 6.2 we show a comparison of the electron density profiles computed coupling the Poisson equation both with the one-dimensional Schrödinger equation and with the effective Bohm quantum potential equation, for different applied gate voltages and for fixed values of the $\gamma$ and $\alpha$ parameters. In particular, the parameters $\gamma = 1.089$ and $\alpha = 0.5$, fit well as the gate voltage is varied, except for small voltages, since quantum confinement is weaker. Such choice of parameters, for a 1D silicon structure at equilibrium, make the proposed effective quantum potential identical to the density gradient correction used by Wettstein et al. [89].

Using the above mentioned values, we have computed the capacitance as a function of the gate voltage (Fig. 6.3). Since there are no confined electrons in the accumulation regime, for low voltages, the quantum corrected curve is in good agreement with that derived by the solution of the Poisson/Schrödinger equation, while a small discrepancy between the two curves can be noted in strong inversion.

Using the two degrees of freedom, it is possible to improve the fitting: by performing a systematic exploration of the $\alpha$ and $\gamma$ parameters, we have found that the values for $\alpha$ and $\gamma$ that best fit the capacitance curve for this structure are 0.4 and 1.7, respectively (Fig. 6.4).
### 6.6. Simulation examples

#### 6.6.2 FinFET or triple gate FET

The device has the structure depicted in Fig. 6.5 in which is also shown the simulated two-dimensional cross section. The oxide thickness is 2 nm and the acceptor concentration is $N_A = 10^{18} \text{ cm}^{-3}$. Quantum confinement is strong both in $x$ and $y$ direction, so a complete solution of the Schrödinger equation in two dimensions is required. In particular we have considered two devices with different cross sections and strong confinement: one with $W = 10 \text{ nm}$ and $L = 20 \text{ nm}$, and one with $W = 5 \text{ nm}$ and $L = 10 \text{ nm}$.

In Fig. 6.6 is shown the derivative of the charge in the FinFET as a function of the gate voltage computed solving the 2D Poisson/Schrödinger equation and using the effective quantum potential for the structure with $W = 10 \text{ nm}$ and $L = 20 \text{ nm}$. For this kind of device, we have considered the effective mass components along the $x$ and $y$ directions equal to silicon longitudinal and transversal effective mass, respectively. As it is apparent from Fig. 6.6, the for the $\alpha = 0.5$ and $\gamma = 0.685$ a curve in very good agreement with the curve computed by the 2D Poisson/Schrödinger solver is obtained.

In Fig. 6.7, the electron densities computed using the two models are shown for a fixed gate voltage: even if they do not match in detail, the integrated electron densities are almost identical, as can be seen in Fig. 6.6.

---

![Figure 6.4: Capacitance computed solving the Poisson/Schrödinger equation and using the Bohm quantum potential as a function of the gate voltage for $\gamma = 1.7$ and $\alpha = 0.4$.](image)
Figure 6.5: FinFET device and transversal cross section of the silicon wire.

Figure 6.6: Derivative of the charge in the wire as a function of the gate voltage for the wire with $W = 10$ nm and $L = 20$ nm.
The considered structure is far from the condition of only one subband occupied, which is the usual situation in conventional MOSFETs. When, instead, more than one subband is populated, a proper choice of the electron effective masses along the \( x \) and \( y \) directions is required. The values used in the present simulations are \( m_x = m_y = 0.7m_0 \).

### 6.7 Concluding remarks

In summary, we have proposed an expression for an effective quantum potential, an appropriate algorithm and a discretization scheme. With respect to previously published density gradient methods [89], the main advantages of our proposal are represented by the addition of one degree of freedom, which may provide better fit and improved calibration, by the possibility of including both Maxwell Boltzmann and Fermi Dirac Statistics, and by its independence of the transport model adopted.
Appendix A

ViDES User Manual

The implemented simulation code consists of two different subprograms: grid3d and ViDES. These programs run separately but must be executed in the above order, since the output files of one are the input files for the other.

grid3d: Discretization of the domain

The first step is the definition of the three dimensional rectangular grid necessary to discretize the spatial domain. The input file of grid3d is grid.in where the position of each grid point in the x, y and z directions are defined, respectively. An example is presented below:

```
2   Number of X-regions
1    section one
0    X Initial coordinate [nm]
10   X Final coordinate [nm]
 1    Initial x-grid spacing [nm]
 0.1  Final x-grid spacing [nm]
2    section two
10   X Initial coordinate [nm]
100  X Final coordinate [nm]
 0.1  Initial x-grid spacing [nm]
 10   Final x-grid spacing [nm]
1    Number of Y-regions
1    section one
```
The first line represents the number of sections, along the considered direction, into which the solution domain is divided. The characteristics of each section are specified by means of five terms:

- The section number
- The initial coordinate (in nm)
- The final coordinate (in nm)
- The initial grid spacing
- The final grid spacing

For each layer, the program calculates the ratio of two successive grid spacings so that the value of the last grid point is equal to the coordinate specified in the input data file.

The output files are: `points.out` in which the number of points along each direction are indicated, and `gridx.out`, `gridy.out` and `gridz.out` which represent the x, y and z coordinates, respectively.

**ViDES User's Manual**

ViDES (Versatile DEvice Simulator) is a self-consistent Poisson/Schrödinger solver in 3D, which is based, from a numerical point of view, on the Newton-Raphson (NR) method with a predictor/corrector scheme.

In Fig. A.1 we sketched a flow-chart of the whole code. In particular, the Schrödinger equation is solved at the beginning of each NR cycle, starting from an initial potential $\tilde{\Phi}$. Then, the eigenfunctions are kept constant until the NR cycle converges (i.e. the correction on the potential is smaller than `poissnorm`), while the eigenvalues are adjusted...
according to the predictor-corrector scheme [14], by a quantity equal to
$q(\Phi - \Phi)$, where $\Phi$ is the electrostatic potential at the each step inside the
NR cycle. In this way, the Jacobian results to be simplified, since a direct
relation between the electronic density and the electrostatic potential to
be computed is established.

Here, as an example, we write the modified expression for the electron
density in case of one-dimensional and two-dimensional confinement
(isotropic masses).

• **1D**: quantum confinement is along the $x$ direction. $\psi_i(x, y, z)$ and
$\epsilon_i(y, z)$ are the computed eigenfunctions and eigenvalues, respectively.

\[
n(x, y, z) = \sum_i |\psi_i(x, y, z)|^2 \frac{\sqrt{m_y m_z K_B T}}{\pi \hbar^2} \times \\
\ln \left[ 1 + \exp \left( \frac{\epsilon_i(y, z) + q(\Phi - \Phi) - E_F}{K_B T} \right) \right]
\]  
(A.1)

• **2D**: quantum confinement is in the $y-z$ plane. $\chi_i(x, y, z)$ and $\epsilon_i(x)$
are the the computed eigenfunctions and eigenvalues, respectively.

\[
n(y, z) = \frac{1}{\pi} \left( \frac{2m_x K_B T}{\hbar^2} \right)^{\frac{1}{2}} \\
\sum_i |\chi_i(x, y, z)|^2 \mathcal{F}_{-\frac{1}{2}} \left( \frac{E_F - \epsilon_i(x) + q(\Phi - \Phi)}{K_B T} \right)
\]  
(A.2)

\(\mathcal{F}_{-\frac{1}{2}}\) is the Fermi integral of order \(-\frac{1}{2}\).

An under relaxation can also be performed inside each NR cycle
(underrel \(\neq 0\)). At the end of the NR cycle, the Schrödinger equation is
solved again and a new NR cycle is performed. The algorithm is ended
when the norm-two of the difference between the electrostatic potentials
obtained at the end of two successive NR cycles ($\Phi_f$ is smaller than a
given predetermined value (normad).

The structure of the device to be simulated is described in the input file
**str3d.in**, which consist in a sequence of regions which differs
from each other by physical parameters such as material, doping con-
centration, mole fraction, etc. The regions can be spatially contiguous
or overlapped. In case of overlapping, the grid points are assigned the
Figure A.1: Flow-chart of the self-consistent 3D Poisson-Schrödinger solver.
physical properties of the last defined region. At first the program assigns to each point of the three dimensional grid obtained after launching `grid3d`, the physical parameters, depending on the properties of the corresponding layer, for example the Fermi level, the electron affinity, the dielectric constant, and so on.

Inside ViDES is implemented a parser (case insensitive), which has the syntax and the grammar that will be discussed in the following. In particular, we will now focus our attention on the commands that has to be declared in the input file, in order to define the structure to be simulated and the kind of quantum analysis that has to be performed.

We have to note, all the command defined as default in the following list of commands can be undeclared in the input file.

**List of commands**

- `#` : comment command (it has to be followed by a space).
- `prev` : the non linear solver starts from the solution provided in the file `Phi.out`.
- `flat` : the non linear solver starts from the flat band solution (default).
- `poiss` : the Poisson equation is solved (default).
- `ps` : the self-consistent Poisson/Schrödinger equation is solved.
- `underel` : under relaxation coefficient (default : underel=1).
- `tolldomn` : inner tolerance of the linear solver (default : tolldomn=0.1).
- `poissnorm` : tollerance inside the NR cycle (default : poissnorm=10^{-3})
- `normad` : tollerance of the outer cycle (default : normad=10^{-3})
- `temp` : temperature (K) (default : temp=300)
- `complete` : complete ionization of donors and acceptors.
- `incompl` : incomplete ionization of donors and acceptors (default).
- **field**: the electric field on the six lateral surface of the simulation domain (V/m). If declared it has to be followed by six number corresponding to the electric field at each surface. By default the electric field is imposed to zero on each surface (Null Neumann Boundary condition).

- **maxwell**: Maxwell-Boltzmann statistic for the semiclassical electron density (**default**).

- **fermi**: Fermi-Dirac statistic for the semiclassical electron density.

- **expsurf**: impose surface states boundary condition [25] on the boundary surface of the three-dimensional domain. It has to be followed by a number $N$ varying between 1 and 6 (**unset by default**).
  - If $N$ is equal to 1 the exposed surface is the $x = x_{\text{min}}$ surface.
  - If $N$ is equal to 2 the exposed surface is the $x = x_{\text{max}}$ surface.
  - If $N$ is equal to 3 the exposed surface is the $y = y_{\text{min}}$ surface.
  - If $N$ is equal to 4 the exposed surface is the $y = y_{\text{max}}$ surface.
  - If $N$ is equal to 5 the exposed surface is the $z = z_{\text{min}}$ surface.
  - If $N$ is equal to 6 the exposed surface is the $z = z_{\text{max}}$ surface.

- **enned**: surface states density ($m^{-2}eV^{-1}$) (**default** : $\text{enned}=0$).

- **phitilde**: surface states effective work function (eV) (**default** : $\text{phitilde}=0$).

- **solid**: the solid command defines, in the three-dimensional domain, the regions of different materials that belong to the device to be simulated. To this purpose, the following ensemble of commands has to be declared and included in brackets { } after the **solid** statement:
  - **rect**: impose the shape of the region to be rectangular (**default**).
  - **xmin**: define the minimum coordinate (nm) in the $x$-direction of the rectangular region. It has to be followed by a number. (**default** : $\text{xmin}=\text{minimum } x \text{ coordinate of the grid}$).
- \texttt{\textit{xmax}}: define the maximum coordinate (nm) in the \textit{x}-direction of the rectangular region. It has to be followed by a number. 
  \textbf{(default : xmax=maximum \textit{x} coordinate of the grid)}.

- \texttt{\textit{ymin}}: define the minimum coordinate (nm) in the \textit{y}-direction of the rectangular region. It has to be followed by a number. 
  \textbf{(default : ymin=minimum \textit{y} coordinate of the grid)}.

- \texttt{\textit{ymax}}: define the maximum coordinate (nm) in the \textit{y}-direction of the rectangular region. It has to be followed by a number. 
  \textbf{(default : ymax=maximum \textit{y} coordinate of the grid)}.

- \texttt{\textit{zmin}}: define the minimum coordinate (nm) in the \textit{z}-direction of the rectangular region. It has to be followed by a number. 
  \textbf{(default : zmin=minimum \textit{z} coordinate of the grid)}.

- \texttt{\textit{zmax}}: define the maximum coordinate (nm) in the \textit{z}-direction of the rectangular region. It has to be followed by a number. 
  \textbf{(default : zmax=maximum \textit{z} coordinate of the grid)}.

- \texttt{\textit{sphere}}: impose the shape of the region to be a spherical shell. It has to be followed by 5 numbers (nm): the \textit{x}, \textit{y}, \textit{z} coordinates of the center (nm), and the inner and outer radius of the sphere.

- \texttt{\textit{polys}}: define a general geometry for the solid region as shown in Fig. A.2. To this purpose the 8 points P1..P8 has to be declared. For example, the point P1 is declared by the command \texttt{P1} followed by three numbers which represent the \textit{x}, \textit{y} and \textit{z} coordinates of \textit{P1} respectively in the 3D domain.

- \texttt{\textit{cylinderx}}: imposes the shape of the region to be a cylindrical shell. The axis of the cylinder is parallel to the \textit{x}-axis. The command has to be follows by four number: the \textit{y} and \textit{z} coordinates of the \textit{x}-axis, the inner and outer radius of the cylinder.

- \texttt{\textit{cylindery}}: imposes the shape of the region to be a cylindrical shell. The axis of the cylinder is parallel to the \textit{y}-axis. The command has to be follows by four number: the \textit{x} and \textit{z} coordinates of the \textit{y}-axis, the inner and outer radius of the cylinder.

- \texttt{\textit{silicon}}: declares that the material of the solid is Silicon.
Figure A.2: General structure of the region defined by the *polys* command.

- *sio2*: declares that the material of the solid is $SiO_2$.
- *air*: declares that the material of the solid is Air.
- *si3n4*: declares that the material of the solid is $Si_3N_4$.
- *AlGaAs*: declares that the material of the solid is AlGaAs.
- *Al*: defines the molar fraction of Aluminium in case the AlGaAs material has been defined.
- *nulln*: imposes the semiclassical electron concentration to zero.
- *nullp*: imposes the semiclassical hole concentration to zero.
- *Na*: defines the nominal acceptors concentration ($m^{-3}$). It has to be followed by a number (default: $Na=0$).
- *Nd*: defines the nominal donors concentration ($m^{-3}$). It has to be followed by a number (default: $Nd=0$).
- *rho*: fixed charge concentration ($m^{-3}$). It has to be followed by a number (default: $rho=0$).
- *Ef*: defines the Fermi level of the region (eV). It has to be followed by a number (default: $Ef=0$).
• *gate*: the gate command defines the gate region in the three-dimensional domain. The first defined gate is the referring gate, whose electrostatic potential is set to zero. The following ensemble of commands has to be declared and included in brackets { } after the gate statement:

  – *xmin*: define the minimum coordinate (nm) in the x-direction of the rectangular region. It has to be followed by a number.
    (default : *xmin*=minimum x coordinate of the grid).

  – *xmax*: define the maximum coordinate (nm) in the x-direction of the rectangular region. It has to be followed by a number.
    (default : *xmax*=maximum x coordinate of the grid).

  – *ymin*: define the minimum coordinate (nm) in the y-direction of the rectangular region. It has to be followed by a number.
    (default : *ymin*=minimum y coordinate of the grid).

  – *ymax*: define the maximum coordinate (nm) in the y-direction of the rectangular region. It has to be followed by a number.
    (default : *ymax*=maximum y coordinate of the grid).

  – *zmin*: define the minimum coordinate (nm) in the z-direction of the rectangular region. It has to be followed by a number.
    (default : *zmin*=minimum z coordinate of the grid).

  – *zmax*: define the maximum coordinate (nm) in the z-direction of the rectangular region. It has to be followed by a number.
    (default : *zmax*=maximum z coordinate of the grid).

  – *Ef*: defines the Fermi energy of the gate (eV) (default : *Ef*=0).

  – *workf*: defines the work function of the gate (eV) (default : *workf*=4.1).

• *region*: the region command defines where and which kind of quantum analysis has to be performed. The following ensemble of commands has to be declared and included in brackets { } after the region statement:

  – *xmin*: define the minimum coordinate (nm) in the x-direction of the rectangular region. It has to be followed by a number.
    (default : *xmin*=minimum x coordinate of the grid).
- `xmax`: define the maximum coordinate (nm) in the `x`-direction of the rectangular region. It has to be followed by a number. *(default: `xmax=maximum x coordinate of the grid`)*.
- `ymin`: define the minimum coordinate (nm) in the `y`-direction of the rectangular region. It has to be followed by a number. *(default: `ymin=minimum y coordinate of the grid`)*.
- `ymax`: define the maximum coordinate (nm) in the `y`-direction of the rectangular region. It has to be followed by a number. *(default: `ymax=maximum y coordinate of the grid`)*.
- `zmin`: define the minimum coordinate (nm) in the `z`-direction of the rectangular region. It has to be followed by a number. *(default: `zmin=minimum z coordinate of the grid`)*.
- `zmax`: define the maximum coordinate (nm) in the `z`-direction of the rectangular region. It has to be followed by a number. *(default: `zmax=maximum z coordinate of the grid`)*.
- `1d`: the 1D Schrödinger equation is solved in the `x`-direction.
- `2d`: the 2D Schrödinger equation is solved in the `x-z`-plane.
- `3d`: the 3D Schrödinger equation is solved.
- `nauto`: specifies the maximum number of eigenvalues computed by the Schrödinger equation in the 1D and 2D cases, while it specifies the number of half-electrons in the 3D region.
- `ball1d`: the 2D Schrödinger equation is solved in the `x-z`-plane and the ballistic transport is computed in 1D subbands along the `y` direction.
- `drdf1d`: the 2D Schrödinger equation is solved in the `x-z`-plane and the drift-diffusion transport equation is computed in 1D subbands along the `y` direction. The computed current is saved in the "jaydf.out" file.
- `efs`: specifies the chemical potential (eV) of the left contact in case the `ball1d` or `drdf1d` command is specified *(default: `efs=0`)*.
- `efd`: specifies the chemical potential (eV) of the right contact in case the `ball1d` or `drdf1d` command is specified *(default: `efd=0`)*.
– outfile : specifies the output file for the current in case the ball1d command is specied. In particular, The following ensemble of commands has to be declared and included in brackets {
} after the region statement :

* 1dballsemi : the current is computed assuming semiclassical transmission coefficient along the 1d barrier. The computed current is saved in “jaysemi.out” file.

* 1dballtunn : the current is computed assuming quantum transmission coefficient along the 1d barrier. The computed current is saved in “jaytunn.out” file.

Output files

• “Phi.out” : Electrostatic potential.

• “Ec.out” : Conduction Band.

• “ncar.out” : electron density.

• “pcar.out” : hole density.

• “Nam.out” : Ionized acceptors concentration.

• “Ndp.out” : Ionized donors concentration.

• “autoval3d” : computed eigenvalues in case the 3D Schrödinger equation is solved.

section

The programs sectionx, sectiony, sectionz extract a slice along the x, y and z direction, respectively of the three-dimensional output file. After running the program, the user is asked to specified the name of the file and the coordinate of the section.

Depending on the called program, the plottable slice is saved in sectionx.out, sectiony.out or sectionz.out, and it can be visualized by the gnuplot command splot.

Example : splot “sectionx.out” w l
Example: Nanocrystal memory

In this section, we will give an example of the “str3d.in” file, for a nanocrystal memory, whose structure is depicted in Fig. A.3. Semiclassical hole and electron densities are computed by means of the Fermi-Dirac statistic, and complete ionization is assumed for donors and acceptors. Both gates are aluminum gates, so their work function is equal to 4.1 eV. Moreover we applied a positive voltage to the top gate equal to 0.5 V, (i.e. $E_F=-0.5$ eV). A 3D quantum analysis is performed in correspondence of the dot, imposing 1.5 electrons in the confined region, while the 1D Schrödinger equation is solved at the Si/SiO$_2$ interface, with a maximum number of eigenvalues equal to 20.
```plaintext
str3d.in

ps prev normad 0.5e-4 poissnorm 1.5e-3
# this line is commented
complete fermi
solid from solid to { everything is considered as a comment
  { SiO2 nulln nullp
  }
solid dot
  { xmin 2 xmax 6
    ymin 2 ymax 6
    zmin 2 zmax 6
    Nd 1e26
  }
solid bulk
  { xmin 8
    Na 1e24
  }
gate from gate to { everything is considered as a comment
  { xmin 50
  }
gate top
  { xmax 0
    Ef -0.5
  }
region from region to { everything is considered as a comment
  {
    xmin 1 xmax 7
    ymin 1 ymax 7
    zmin 1 zmax 7
    3d
```
nauto 3
}
region 1d quantum analysis
{
   xmin 7.5
   1d
   nauto 20
}
Bibliography


