POLITECNICO DI MILANO
Scuola di Ingegneria Industriale e dell'Informazione
Laurea magistrale in Ingegneria Matematica

3D Finite Element Drift-Diffusion Simulation of Semiconductor Devices

Relatore: Prof. Riccardo SACCO
Correlatore: Dott. Aurelio MAURI

Tesi di Laurea di:
Andrea BORTOLOSSI
Matr. n. 783023

Anno Accademico 2013–2014

alla mia famiglia
che mi ha sostenuto,
e a Francesca
che mi ha sopportato

Indice

Elenco delle figure

Elenco delle tabelle

Sommario

Questo elaborato di tesi si inserisce nel progetto FEMOS (*Finite Element Method Oriented Solver*) che costituisce all'interno dell'azienda Micron Technology una piattaforma per la simulazione 3D multifisica (termo-elettro-chimica-meccanica) delle memorie elettroniche. In particolare questo lavoro di tesi si è occupato della trattazione dell'approccio Drift-Diffusion [Jac84] per i semiconduttori la cui risoluzione si è basata sull'algoritmo della mappa di Gummel [Gum64]. La discretizzazione delle equazioni è stata realizzata secondo il metodo di Galerkin agli elementi finiti (FEM) ed in particolare per il problema di Poisson si è scelta una formulazione agli spostamenti, mentre per l'equazione di continuità ci si è affidati allo schema numerico EAFE presentato in [XZ99]. Il problema non lineare di Poisson è stato affrontato con il metodo di Newton.

La parte più originale del lavoro è costituita dalle tecniche sviluppate al fine di calcolare la corrente ai contatti e all'interno dei dispositivi. Nel primo caso abbiamo esteso al caso tridimensionale il metodo dei residui presentato in [GS06]. Per il secondo abbiamo proposto due schemi innovativi volti all'estensione tridimensionale della formula di Scharfetter-Gummel [GS69].

Sono stati condotti numerosi test di simulazione su vari dispositivi a semiconduttore (diodo, n-MOSFET/p-MOSFET) ed i risultati ottenuti sono stati confrontati con quelli forniti da un solutore commerciale di uso comune nel presente contesto dimostrando un ottimo accordo.

Abstract

This thesis is part of the FEMOS (*Finite Element Method Oriented Solver*) project which is a modular numerical code designed for the treatment of multiphysical effects (thermal-electrical-chemical-mechanical) applied to the most modern memory devices. More precisely in this work the Gummel map algorithm [Gum64] is employed to solve the Drift-Diffusion model [Jac84] for semiconductors. The Non Linear Poisson equation has been discretized using the Galerkin finite element method [QV08] following a displacement formulation and the Continuity equations have been treated using the EAFE scheme [XZ99].

The original part of this work is the calculation of the current both at contacts and inside the device. In the first case we extended to the 3D framework the *residual method* [GS06], while in the second one we proposed two novel schemes in order to extend the Scharfetter-Gummel formula [GS69] to the 3D case.

The code has been thoroughly tested on different semiconductor devices (p-n junction, p-n junction in oxide and n-channel/p-channel MOSFET), comparing the results with a commercial tool as reference reaching a very good agreement.

Estratto della tesi

Lo sviluppo tecnologico nell'industria dei semiconduttori è in rapidissima crescita in questi ultimi anni e in particolare non solo secondo il tradizionale *shrinking tecnologico*, ma grazie anche all'utilizzo di nuovi principi fisici complementari o alternativi alla fisica dei semiconduttori. In particolare, l'utilizzo di nuovi materiali richiede capacità di sviluppo e comprensione decisamente superiori rispetto al passato. Proprio per rispondere a queste esigenze in Micron Technology (una delle compagnie leader fra le industrie di semiconduttori) è nato il progetto FEMOS (*Finite Element Method Oriented Solver*). FEMOS è un codice numerico modulare in grado di affrontare problematiche di diversa natura fisico-chimica, di trasporto termico ed elettronico in mezzi anisotropi e di meccanica dei continui, applicate ai più moderni dispositivi di memoria (a cambiamento di fase, a movimento ionico). A completamento di questo progetto si è resa necessaria la possibilità di simulare i materiali a semiconduttore: questo lavoro di tesi si inserisce proprio in questo contesto. In particolare ci siamo occupati della trattazione dell'approccio Drift-Diffusion [Jac84] per i semiconduttori la cui risoluzione si è basata sull'algoritmo della mappa di Gummel [Gum64]. La discretizzazione delle equazioni è stata realizzata secondo il metodo di Galerkin agli elementi finiti (FEM) ed in particolare per il problema di Poisson si è scelta una formulazione agli spostamenti, mentre per l'equazione di continuità ci si è affidati allo schema numerico EAFE presentato in [XZ99]. Il problema non lineare di Poisson è stato affrontato con il metodo di Newton.

La parte più originale del lavoro è costituita dalle tecniche sviluppate al fine di calcolare la corrente ai contatti e all'interno dei dispositivi. Nel primo caso abbiamo esteso il metodo dei residui presentato in [GS06] al caso tridimensionale. Per il secondo abbiamo proposto due schemi innovativi volti all'estensione al caso 3D della formula di Scharfetter-Gummel [GS69].

Sono stati condotti test di simulazione su vari dispositivi a semiconduttore (diodo, n-MOSFET/p-MOSFET) ed i risultati sono stati confrontati con un solutore commerciale di uso comune nel contesto in esame, dimostrandosi in ottimo accordo con essi.

L'elaborato è organizzato come segue:

Capitolo 1 richiamiamo brevemente le proprietà fisiche dei semiconduttori ed enunciamo le principali relazioni che intercorrono fra le grandezze fondamentali (potenziale elettrostatico, campo elettrico, densità di portatori e di corrente). Presentiamo il modello Drift-Diffusion e alcuni dei principali modelli di mobilità dei portatori e dei fenomeni di R/G.

Capitolo 2 questo capitolo è diviso in due sezioni. Nella prima ci occupiamo di introdurre le geometrie considerate durante le simulazioni e le relative notazioni. Nella seconda parte illustriamo gli algoritmi usati al fine di trattare il modello esposto nel primo capitolo (mappa di Gummel).

Capitolo 3 la buona posizione delle equazioni trattate e i metodi utilizzati per discretizzarle sono approfonditi in questo capitolo. Particolare attenzione viene posta su alcuni aspetti che ne rendono difficile la risoluzione numerica.

Capitolo 4 questo capitolo contiene i risultati ottenuti. La validazione dei tests è stata condotta su dispositivi a semiconduttore tipici delle applicazioni microelettroniche (diodo, n-MOSFET/p-MOSFET) confrontando le soluzioni con un software commerciale. La parte finale del capitolo riguarda l'estensione del *metodo dei residui* al caso 3D per il calcolo della corrente ai contatti [GS06].

Capitolo 5 in questo capitolo vengono trattate alcune tecniche che permettono la ricostruzione delle densità di corrente all'interno dei dispositivi. Proponiamo due schemi innovativi al fine di estendere la formula di Scharfetter-Gummel [GS69] al caso 3D. Infine confrontiamo i risulati con il simulatore commerciale.

Capitolo 6 contiene una sintesi delle tematiche e dei metodi proposti ed impiegati nella tesi, e affronta alcuni punti aperti che saranno oggetto della futura ricerca in questo ambito.

Sono infine incluse due Appendici dedicate a illustrare brevemente la struttura del codice FEMOS.

Introduction

In the last years there has been a continously technological development in the semiconductor industry according to the traditional technological shrinking but also for the introduction of new complementary or alternative physical principles into the modern devices. This has led to the use of new materials which need higher development and comprehension skills than in the past. In order to satisfy these requirements in Micron Technology (one of the leader companies among semiconductor industries) the FEMOS project (*Finite Element Method Oriented Solver*) was born. FEMOS is a modular numerical code designed for the treatment of different physical and chemical phenomena, thermal and electrical transport inside anisotropic materials and mechanical effects applied to the most modern memory devices (i.e. phase change memories, ionic transport). In order to complete the FEMOS project the urgent need of simulating semiconductor materials has recently manifested: this is the object of this work of thesis. More precisely the Gummel map algorithm [Gum64] is employed to solve the Drift-Diffusion model [Jac84]. The Non Linear Poisson has been discretized using the Galerkin finite element method [QV08] following a displacement formulation and the Continuity equations have been treated using the EAFE scheme [XZ99]. The Non Linear Poisson problem is solved with the Newton method.

The original part of this work is the calculation of the current both at contacts and inside the device. In the first case we extended to the 3D framework the *residual method* [GS06], while in the second one we proposed two novel schemes in order to extend the Scharfetter-Gummel fromula to the 3D case [GS69].

The code has been thoroughly tested on different semiconductor devices (p-n junction, p-n junction in oxide and n-channel/p-channel MOSFET), comparing the results with a commercial tool as reference reaching a very good agreement.

The thesis is organized as follows:

Chapter 1 we briefly recall the semiconductor material properties, physical behaviour and relations between the fundamental quantities (electrostatic potential, electric field, carrier densities and current densities). The classical Drift-Diffusion model is discussed in detail; with the needed models for carrier mobilities and generation/recombination phenomena.

Chapter 2 this chapter consists of two main sections. The first one presents the geometry framework and introduces some useful notation. The second one illustrates the algorithms used in order to treat the equations of the first chapter (decoupled Gummel map approach).

Chapter 3 the well-posedness analysis and the numerical approximation of the equations is discussed in detail. Special emphasis is also devoted to illustrating several issues that make the numerical approximation a difficult task.

Chapter 4 in this chapter we present the numerical results. The validation tests are performed on typical semiconductor devices (p-n junction, n-channel/p-channel MOSFET) comparing the results with a commercial software. At the end the calculation of the current at contacts is performed, extending the *residual method* [GS06] to the 3D case.

Chapter 5 we investigate some techniques that allow to reconstruct the current density inside the device. We propose two novel schemes in order to extend the 1D Scharfetter-Gummel formula [GS69] to the 3D case. Finally, we compare the results with the commercial software.

Chapter 6 contains summarising conclusions of the thesis work and addresses future research perspectives.

Two concluding Appendices are devoted to briefly illustrating the FEMOS code.

Chapter 1

Physical models for charge transport in semiconductor materials

In this chapter we present the basic physical properties of semiconductor materials according to the quantum mechanics theory [YT09] and the Drift-Diffusion model [Jac84].

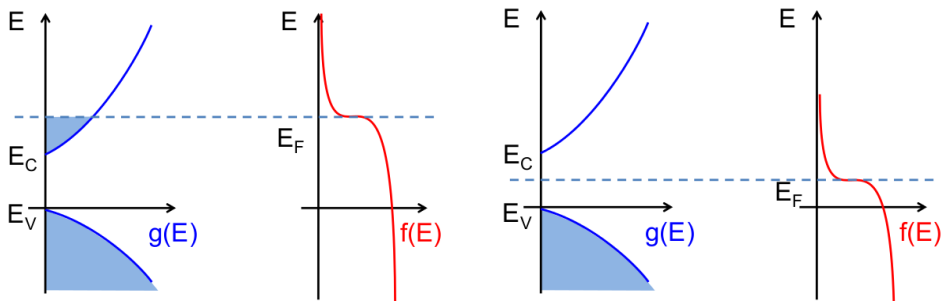
1.1 Basic Device Physics

Since the most used material in the fabrication of VLSI devices technology is silicon, the following description is based on this material choice.

1.1.1 Intrinsic semiconductor

In a silicon crystal each atom has four valence electrons to share with its four neighboring atoms. The valence electrons are shared in a paired configuration called covalent bond. In a solid semiconductor energy levels of electrons are grouped into bands separated by regions of not allowed energy, the so-called forbidden gaps. The highest energy band completely filled by electrons at 0[K] is called *valence band* (E_V), the next band is called *conduction band* (E_C).

As in silicon the band gap is 1.11 [eV] [BGB00], at room temperature a small fraction of the electrons are excited into the conduction band, leaving behind vacancies (called *holes*) in the valence band. In contrast, an insulator has a much larger forbidden gap making room-temperature conduction virtually impossible, while metals have partially filled conduction bands even



(a) Metal **(b)** Insulator

Figure 1.1: Two typical examples of state density occupation (g(E)) and probability distribution (f(E)).

at absolute zero temperature, making them excellent conductors at any temperature.

A suitable formulation of the electron concentration in the conduction band is given by the following integral

$$n = \int_{E_c}^{\infty} g(E) f(E)\, dE \tag{1.1}$$

where $g(E)dE$ represents the number of electronic states per unit volume with an energy between E and $E + dE$ in the conduction band and $f(E)$ is the *Fermi-Dirac distribution function*, which gives the probability that an electronic state at energy E is occupied by an electron

$$f_D(E) = \frac{1}{1 + \exp\left(\dfrac{E - E_f}{k_B T}\right)}. \tag{1.2}$$

In (1.2) $k_B = 1.38 \times 10^{-23}[J/K]$ is Boltzmann's constant, T is the absolute temperature and E_f is the *Fermi level*. Fig.1.1 shows the state density occupation $g(E)$ and the probability distribution $f(E)$ for a metal and a semiconductor. In order to obtain an analytic formula for the state density occupation $g(E)$, we can consider the well known parabolic approximation of the conduction band [PN]

$$E = E_c + \frac{\hbar^2}{2m_e^*}\mathrm{k}^2 \tag{1.3}$$

where $\hbar = h/2\pi$ and $h = 6.63 \times 10^{-34}[J\,s]$ is Planck's constant, m_e^* the electron effective mass, E_c the minimum value of the conduction band and

k the wavenumber. Using this approximation, conduction band density of states can be calculated as

$$g(E) = \frac{m_e^* \sqrt{2m_e^*(E - E_c)}}{\pi^2 \hbar^3}. \tag{1.4}$$

We point that (1.1) is a Fermi integral of order 1/2 and must be evaluated numerically.

Definition 1.1. The Fermi level (E_f) is the energy at which the probability of occupation of an energy state by an electron is equal to 1/2.

In most cases, when the energy is at least several $k_B T$ above or below the Fermi level (non degenerate semiconductor), equation (1.2) can be well approximated by the Maxwell-Boltzmann statistics, which reads as follows:

$$f_D(E) \simeq f_{MB}(E) = \begin{cases} \exp\left(-\dfrac{E - E_f}{k_B T}\right) & E \gg E_f \\ 1 - \exp\left(-\dfrac{E_f - E}{k_B T}\right) & E \ll E_f. \end{cases} \tag{1.5}$$

The Fermi level plays an essential role for the equilibrium of a system, it is important to keep in mind the following observation.

Observation 1.1. When two systems in contact are in thermal equilibrium with no current flow between them, their Fermi levels must be equal: in other words for a continuous region (of metals or semiconductors in contact), the Fermi level at thermal equilibrium is spatially constant.

Replacing (1.5) and (1.4) into (1.1) we obtain

$$n = N_c \exp\left(-\frac{E_c - E_f}{k_B T}\right). \tag{1.6}$$

Using the similar approach for the holes, the concentration p in the valence band is

$$p = N_v \exp\left(-\frac{E_f - E_v}{k_B T}\right). \tag{1.7}$$

N_c and N_v are the *effective density of states* while E_v is the maximum value of the valence band. In an intrinsic semiconductor $n = p$ and the *intrinsic Fermi level* E_i can be calculated by (1.6) and (1.7) as:

$$E_i = E_f = \frac{E_c + E_v}{2} - \frac{k_B T}{2} ln\left(\frac{N_c}{N_v}\right). \tag{1.8}$$

By replacing (1.8) in (1.6) we obtain the intrinsic carrier concentration $n_i = n = p$

$$n_i = \sqrt{N_c N_v} exp\left(-\frac{E_g}{2k_B T}\right) \tag{1.9}$$

where $E_g = E_c - E_v$ is the semiconductor energy gap.

Observation 1.2. Since the thermal energy, $k_B T$ is much smaller than the usual semiconductor bandgap E_g, the intrinsic Fermi level is very close to the midgap.

Equations (1.6) and (1.7) can be written in terms of the intrinsic carrier density (n_i) and energy (E_i) as:

$$n = n_i \exp\left(\frac{E_f - E_i}{k_B T}\right) \tag{1.10}$$

$$p = n_i \exp\left(\frac{E_i - E_f}{k_B T}\right). \tag{1.11}$$

Finally we remark that at thermal equilibrium the *mass action law* holds

$$np = n_i^2. \tag{1.12}$$

The analysis of the work principles of devices can be effectively done by the band diagram (Fig.1.2), which summarizes the information presented above.

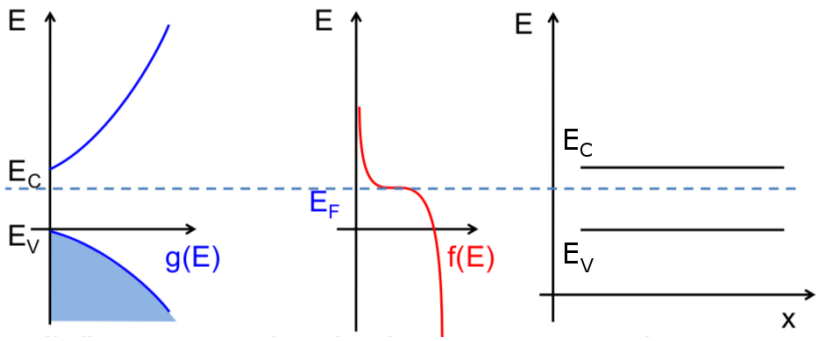


Figure 1.2: Construction of the band diagram.

1.1.2 Extrinsic semiconductor

At room temperature an intrinsic semiconductor has an extremely low free-carrier concentration, therefore, its resistivity is very high. In order to improve the conductivity of the semiconductor, impurities atoms are added in the material. This introduces additional energy levels in the forbidden gap: the impurities are easily ionized adding either electrons to the conduction band or holes to the valence band, in such a way that the electrical conductivity is dominated by the type and concentration of the impurity atoms.

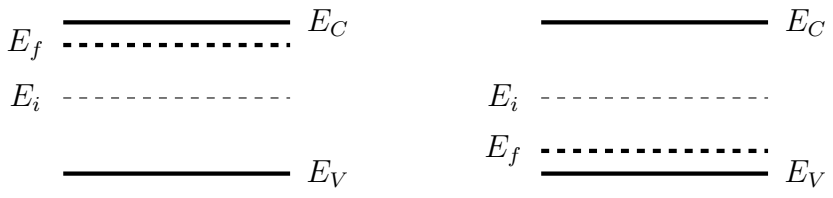
Two are the types of impurities which are electrically active: those from column V of the Periodic Table such as arsenic or phosphorus, and those from column III such as boron or indium.

The thermal energy at room temperature is sufficient to ionize the impurities and free the extra electron to the conduction band (column V) or accept an electron from valence band (column III). Column V impurities are called *donors*; they become positively charged when ionized. Silicon material doped with column-V impurities or donors is called *n-type* silicon.

Column III impurities are called *acceptors*: they become negatively charged when ionized. Silicon material doped with column-III impurities or acceptors is called *p-type* silicon.

A p-type or an n-type is named as *extrinsic* silicon. In terms of the energy-band diagram, donors add allowed electron states in the bandgap close to the conduction-band edge, while acceptors add allowed states just above the valence-band edge.

The Fermi level in n-type silicon moves up towards the conduction band while in p-type silicon it moves down towards the valence band. This behaviour is presented in the band diagrams of Fig.1.3.



(a) n-type **(b)** p-type

Figure 1.3: Band diagrams of extrinsic silicon for (1.18) and (1.19).

The position of the Fermi level depends on both the ionization energy and the concentration of dopants. For the sake of simplicity we consider that at room temperature all impurties are ionized ($N_d = N_d^+$ and $N_a = N_a^-$). For an n-type material with a donor impurity concentration, N_d, the charge

neutrality condition requires that

$$n = N_d^+ + p \tag{1.13}$$

where N_d^+ is the density of ionized donors. Similarly for a p-type material with acceptor impurity concentration N_a we have

$$p = N_a^- + n. \tag{1.14}$$

Since the magnitude of impurities is in the range of $10^{16} \div 10^{20}[cm^{-3}]$, and intrinsic carrier concentration in the order of $10^{10}[cm^{-3}]$, we can approximate the carrier concentrations as:

$$\begin{aligned} n &\simeq N_d^+, \qquad p \simeq \frac{n_i^2}{N_d^+} \quad (n-type) \\ p &\simeq N_a^-, \qquad n \simeq \frac{n_i^2}{N_a^-} \quad (p-type). \end{aligned} \tag{1.15}$$

Replacing (1.15) in (1.6) and (1.7) in (1.13) and (1.14) and solving the corresponding algebraic equation, we have:

$$E_c - E_f = k_B T \ln\left(\frac{N_c}{N_d^+}\right) \tag{1.16}$$

$$E_f - E_v = k_B T \ln\left(\frac{N_v}{N_a^-}\right). \tag{1.17}$$

Equations (1.16) and (1.17) can be written in a more useful form using (1.9) and (1.8) (for n_i and E_i):

$$E_f - E_i = k_B T \ln\left(\frac{N_d^+}{n_i}\right) \tag{1.18}$$

$$E_i - E_f = k_B T \ln\left(\frac{N_a^-}{n_i}\right). \tag{1.19}$$

Observation 1.3. The distance between the Fermi level and the intrinsic Fermi level is a logarithmic function of doping concentration.

1.1.3 Carrier densities at nonequilibrium condition

In VLSI devices a nonequilibrium condition is often possible: the densities of one or both types of carriers depart from their equilibrium as given by

(1.10) and (1.11). In particular, the minority carrier concentration can be easily overwhelmed by the injection from neighboring regions. Under these circumstances, while electrons and holes are in local equilibrium with themselves, they are not in equilibrium with each other. In order to extend the relationship between Fermi level and densities discussed above, we can introduce different Fermi levels for electrons and holes. They are called *quasi Fermi levels* defined as:

$$E_{fn} = E_i + k_B T \ln\left(\frac{n}{n_i}\right) \tag{1.20}$$

$$E_{fp} = E_i - k_B T \ln\left(\frac{p}{n_i}\right). \tag{1.21}$$

Considering the well known relation between electrostatic potential and energy $\varphi = -E/q$, (1.20) and (1.21) can be written as:

$$n = n_i \exp\left(\frac{\varphi_i - \varphi_n}{k_B T/q}\right) \tag{1.22}$$

$$p = n_i \exp\left(\frac{\varphi_p - \varphi_i}{k_B T/q}\right) \tag{1.23}$$

where φ_n and φ_p are the quasi Fermi potential levels and φ_i is the midgap potential level, while $q = 1.602e^{-19}[C]$ is the elementary charge.

Observation 1.4. In non equilibrium conditions, the quasi Fermi levels have the same physical meaning in terms of the state occupancy as the Fermi level, therefore the electron (hole) density in the conduction band can be calculated using E_{fn} (E_{fp}).

1.1.4 Carrier transport in a semiconductor

Carrier transport or current flow is driven by two different mechanisms:

- the **drift**, which is caused by the presence of an electric field;
- the **diffusion**, which is caused by a spatial gradient of electron or hole concentration.

Drift current - Ohm's law

When an electric field is applied to a device, the free carriers are accelerated and acquire a drift velocity superimposed upon their random thermal motion.

Observation 1.5. The drift velocity of holes (h) is in the direction of the applied field, and the drift velocity of electrons (e) is opposite to the field.

The velocity of the carriers does not increase indefinitely under field acceleration, since they are scattered frequently and lose their acquired momentum after each collision. During their motion throughout the lattice structure, carriers travel at an average speed defined by:

$$\mathbf{v}_d^e = -\frac{q\mathbf{E}\tau_e}{m_e^*}\,, \qquad \mathbf{v}_d^h = +\frac{q\mathbf{E}\tau_h}{m_h^*} \tag{1.24}$$

where $\mathbf{E}$ is the electric field, τ_e, τ_h the average times between two consecutive scattering events and m_e^*, m_h^* the effective masses of electron and hole respectively. The coefficient $q\tau_e/m_e$ ($q\tau_p/m_p$) characterizes how quickly a carrier move through the lattice and is known as carrier mobility $[cm^2V^{-1}s^{-1}]$. In general, to include different scattering mechanisms *Mathiessen's rule* is used to calculate the resulting mobility

$$\frac{1}{\mu} = \frac{1}{\mu_L} + \frac{1}{\mu_I} + \cdots \tag{1.25}$$

where μ_L and μ_I correspond to the lattice and impurity scattering (for a more detailed description of mobility models see [YT09]).

Therefore the drift electron (hole) current density reads as follows:

$$\mathbf{J}_n = -\,qn\mathbf{v}_d^n = qn\mu_n\mathbf{E} = \sigma_n\mathbf{E} \tag{1.26}$$

$$\mathbf{J}_p = +\,qp\mathbf{v}_d^p = qp\mu_p\mathbf{E} = \sigma_p\mathbf{E}. \tag{1.27}$$

The scalar coefficient $qn\mu_n(qp\mu_p)$ is called electron (hole) conductivity $\sigma_n(\sigma_p)$.

Relations (1.26) and (1.27) express the well known *Ohm' law* stating that the current density is directly proportional to the applied electric field.

Diffusion current - Fick's law

In semiconductor devices it is very common to have different profiles of dopant in order to allow specific electrical behaviors. This implies a non uniform concentration of carriers which also diffuse as a result of a gradient concentration. This leads to an additional current contribution according to the classical *Fick's law*:

$$\mathbf{J}_n = -D_n(-q\nabla n) \tag{1.28}$$

$$\mathbf{J}_p = -D_p(+q\nabla p). \tag{1.29}$$

The constants D_n and D_p are called electron and hole diffusion coefficients and have units of $[cm^2 s^{-1}]$. Drift and diffusion are closely associated with the random thermal motion of carriers and their collisions with the silicon lattice in thermal equilibrium. The *Einstein relation* (1.30) expresses the relation between diffusivity and mobility

$$D_n = \frac{k_B T}{q}\mu_n, \qquad D_p = \frac{k_B T}{q}\mu_p. \tag{1.30}$$

Drift-Diffusion transport equations

By considering (1.26), (1.27), (1.28) and (1.29), the electron and hole current densities become [Sie84]:

$$\mathbf{J}_n = qn\mu_n \mathbf{E} + qD_n \nabla n \tag{1.31}$$
$$\mathbf{J}_p = qp\mu_p \mathbf{E} - qD_p \nabla p. \tag{1.32}$$

The total conduction current density is

$$\mathbf{J} = \mathbf{J}_n + \mathbf{J}_p.$$

Equations (1.31) and (1.32) are called constitutive laws and can be written in two other ways highlighting different physical explanations of the same phenomenon. These interpreations give also different starting points for the mathematical formulation and solution of the Drift-Diffusion model.

Considering that the electric field is related to the scalar potential through the quasi-static approximation

$$\mathbf{E} = -\nabla\varphi, \tag{1.33}$$

using (1.30) the current densities can be written as:

$$\mathbf{J}_n = -qn\mu_n \left(\nabla\varphi - \frac{k_B T}{qn}\nabla n\right) \tag{1.34}$$
$$\mathbf{J}_p = -qp\mu_p \left(\nabla\varphi + \frac{k_B T}{qp}\nabla p\right). \tag{1.35}$$

Considering equations (1.22) and (1.23) the current densities are equal to:

$$\mathbf{J}_n = -qn\mu_n \nabla\varphi_n \tag{1.36}$$
$$\mathbf{J}_p = -qp\mu_p \nabla\varphi_p. \tag{1.37}$$

With these equations we underly an important aspect which occurs in a semiconductor:

Observation 1.6. The current density is proportional to the gradient of the quasi Fermi potential.

The third way to represent the current density is based on *Slotboom variables* which are particularly suited for the mathematical analysis of the semiconductor equations:

$$u_n = n_i \exp\left(-\frac{\varphi_n}{V_{th}}\right) \tag{1.38}$$

$$u_p = n_i \exp\left(\frac{\varphi_p}{V_{th}}\right) \tag{1.39}$$

where $V_{th} = k_B T/q$. Replacing these equations into (1.31) and (1.32) we obtain:

$$\mathbf{J}_n = qD_n \exp\left(\frac{\varphi}{V_{th}}\right) \nabla u_n \tag{1.40}$$

$$\mathbf{J}_p = -qD_p \exp\left(-\frac{\varphi}{V_{th}}\right) \nabla u_p. \tag{1.41}$$

Observation 1.7. The drift-diffusion current density in a semiconductor is a purely diffusive flux of a new kind of carrier with a properly modified diffusion coefficient.

1.2 Drift Diffusion Model for semiconductors

In the study of integrated devices the *Drift Diffusion model* (DD) is the most widely used mathematical approach, in particular in industrial simulation. In this section we show how the DD model can be obtained.

1.2.1 Drift Diffusion formulation

The system of Maxwell equations describes the propagation of electromagnetic signal in a medium [Jac84]:

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{1.42}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{1.43}$$

$$\nabla \cdot \mathbf{D} = \rho \tag{1.44}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.45}$$

with the following set of constitutive laws that characterize the electromagnetic properties of the medium:

$$\begin{aligned} \mathbf{D} &= \epsilon \mathbf{E} \\ \mathbf{B} &= \mu_m \mathbf{H} \end{aligned} \tag{1.46}$$

where ϵ is the material dielectric permittivity $[Fcm^{-1}]$ and μ_m is the magnetic permeability $[Hcm^{-1}]$. Since $\nabla \cdot (\nabla \times \mathbf{A}) = 0$ for any vector $\mathbf{A}$, (1.45) is satisfied by introducing a vector potential $\mathbf{A}$ such that $\mathbf{B} = \nabla \cdot \mathbf{A}$. We replace it in (1.43) to obtain

$$\nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = 0. \tag{1.47}$$

From this we can state that there exists a scalar potential φ such that

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla \varphi. \tag{1.48}$$

Applying the divergence operator and using (1.33), (1.46) and (1.44), relation (1.48) becomes

$$\rho + \frac{\partial \mathbf{A}}{\partial t} = -\nabla \cdot (\epsilon \nabla \varphi). \tag{1.49}$$

We now assume that $\dfrac{\partial \mathbf{A}}{\partial t} = 0$ (quasi static approximation) and we have the *Poisson Equation*

$$\nabla \cdot (\epsilon \nabla \varphi) = \rho. \tag{1.50}$$

Applying the divergence operator to (1.42) and considering (1.50) we get the *Continuity Equation*

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0. \tag{1.51}$$

To close system given by (1.50) and (1.51), we need specify the mathematical form of the electric charge density (ρ) and the electric conduction current density ($\mathbf{J}$). Considering (1.15), ρ can be expressed as

$$\rho = \underbrace{q(p-n)}_{\rho_{free}} + \underbrace{q(N_D - N_A)}_{\rho_{fixed}} \tag{1.52}$$

in which we can distinguish the two following contributions:

- free charge (ρ_{free}) (free electron and holes carriers),
- fixed charge (ρ_{fixed}) (ionoized dopant impurities).

Notice that we assume N_D^+ and N_A^- to be time invariant ($\partial N_D^+/\partial t = \partial N_A^-/\partial t = 0$).

Splitting the continuity equation (1.51) into two distinct equations (for holes and electrons), Drift Diffusion (DD) model formulation reads as follows:

$$\begin{cases} \nabla \cdot (-\epsilon \nabla \varphi) &= q(p - n + N_D^+ - N_A^-) \\ -q\dfrac{\partial n}{\partial t} + \nabla \cdot (-q\mu_n n \nabla \varphi + qD_n \nabla n) &= qR \\ q\dfrac{\partial p}{\partial t} + \nabla \cdot (-q\mu_p p \nabla \varphi - qD_p \nabla p) &= -qR \end{cases} \tag{1.53}$$

where the arbitrarily introduced function $R = R(\mathbf{x}, t)$ can be considered as the net rate of generation and recombination. The system is an incompletely parabolic initial value/boundary problem in three scalar unknown dependent variables $\varphi(\mathbf{x}, t)$, $n(\mathbf{x}, t)$ and $p(\mathbf{x}, t)$: the presence of the drift terms ($n\nabla\varphi$ and $p\nabla\varphi$) makes (1.53) a nonlinear coupled system of PDE's.

From Maxwell equations we are able to guarantee only that $\mathbf{J}$ is a solenoidal field. The stationary form can be easily deduced from (1.53) by neglecting the temporal derivatives.

1.2.2 Generation and Recombination phenomenon

The modelling of $R(\mathbf{x}, t)$ is fundamental for device simulation due to its role in determining the current-voltage characteristic.

It is important to keep in mind that electrons and holes are in continuos fluctuation due to their thermal energy, the macroscopic result is that the net recombination rate at equilibrium is identically zero. Our interest is to analyze the deviations from this condition.

While generation events are usually due to thermal agitation or an external input source, the recombination events happen in order to neutralize an excess of charge.

The phenomenological model for the net recombination rate R is often given in the following

$$R(n,p) = (pn - n_i^2)F(n,p) \tag{1.54}$$

where F is a function accounting for specific recombination/generation (R/G) events. In the following we present the classical theory that includes three specific kind of contributions.

Shockley-Read-Hall recombination (SRH)

Electron and hole generation and recombination can take place directly between the valence band and the conduction band, or mediated via trap centers in the energy gap. Shockley-Read-Hall phenomena is a two-particle process which matematically expresses the probability that:

- an electron in the conduction band neutralizes a hole at the valence band through the mediation of an unoccupied trapping level located at the energy gap (R_{SRH}),
- an electron is emitted from the valence band to the conduction band, through the mediation of an unoccupied trapping level located at the energy gap (G_{SRH}).

The modeling function F is given by

$$F_{SRH}(n,p) = \frac{1}{\tau_n \left(p + n_i \cosh\left(\frac{E_T}{k_B T} \right) \right) + \tau_p \left(n + n_i \cosh\left(\frac{E_T}{k_B T} \right) \right)} \tag{1.55}$$

where E_T is the energy level of the traps, τ_n and τ_p are called *carrier lifetimes* and are physically defined as the reciprocals of the capture rates. The typical order of magnitude of the lifetimes lies in the range of $10^{-3}\mu s \div 1\mu s$ (see [VOT83] and [GH92]).

Auger recombination (AU)

Auger R/G is a three-particle process and takes place directly between the valence band and the conduction band. We distinguish four cases which depend on the type of carriers involved:

Parameter	Unit	Electrons	Holes
τ	s	1.0×10^{-5}	3.0×10^{-6}
E_T	eV	0.0	0.0

Table 1.1: Parameters in the Shockley-Read-Hall generation/recombination model.

$R_{AU}^{2n,1p}$ a high-energy electron in the conduction band moves to the valence band where it neutralizes a hole, transmitting the excess energy to another electron in the conduction band;

$G_{AU}^{2n,1p}$ an electron in the valence band moves to the conduction band by taking the energy from a high energy electron in the conduction band and leaves a hole in the valence band;

$R_{AU}^{2p,1n}$ an electron in the conduction band moves to the valence band where it neutralizes a hole, transmitting the excess energy to another hole in the valence band;

$G_{AU}^{2p,1n}$ an electron in the valence band moves to the conduction band by taking the energy from a high energy hole in the valence band and leaves a hole in the valence band.

The modeling function F is

$$F_{AU}(n,p) = C_n n + C_p p \tag{1.56}$$

where the quantities C_n and C_p are the so called Auger capture coefficients tipically in the order of magnitude of $10^{-25}[cm^6 s^{-1}]$ [LH80]. Note that Auger R/G is relevant only when both carrier densities are high.

Parameter	Unit	Magnitude
C_n	$cm^6 s^{-1}$	2.9×10^{-31}
C_p	$cm^6 s^{-1}$	1.028×10^{-31}

Table 1.2: Parameters in Auger generation/recombination model.

Impact ionization (II)

The impact ionization mechanism is a three-particle phenomena where carrier generation is triggered by the presence of a high electric field: due to this

field an electron could gain enough energy to excite an electron-hole pair out of a silicon lattice bond. Then the process can be repeated until an avalanche of generated carriers is produced within the region: this process can not be described by a relation of the form (1.54).

Among the different formulations for the impact ionization generation we choose the van Overstraeten - de Man model [vOdM70], based on the Chynoweth law [Chy58]:

$$G_{II}(n,p) = \alpha_n n|\mathbf{v}_n| + \alpha_p p|\mathbf{v}_p| \tag{1.57}$$

with:

$$\alpha(E_{ava}) = \gamma a \exp\left(-\frac{\gamma b}{E_{ava}}\right) \tag{1.58}$$

$$\gamma = \frac{\tanh\left(\dfrac{\hbar\omega_{op}}{2k_B T_0}\right)}{\tanh\left(\dfrac{\hbar\omega_{op}}{2k_B T}\right)} \tag{1.59}$$

where $\hbar\omega_{op}$ is the phonon energy, b the critical electric field and γ the temperature dependence of the phonon gas against which carriers are accelerated. Two sets of coefficients a and b are used for high and low ranges of electric field. The values a_{low}, b_{low} are used in the low field range up to E_0 and the values a_{high}, b_{high} apply in the high field region above E_0. E_{ava} is the driving force that can be computed as:

- the component of the electrostatic field in the direction of current flow

$$E_{ava}^{n,p} = \frac{\mathbf{E}\cdot\mathbf{J}_{n,p}}{||\mathbf{J}_{n,p}||} \tag{1.60}$$

- the module of the quasi fermi gradient

$$E_{ava}^{n,p} = |\nabla\varphi_{n,p}|. \tag{1.61}$$

1.2.3 Mobility models

In this section we illustrate the most common phenomenological models for carrier mobilities. The main physical phenomena underlying a mobility reduction with respect to its bulk value are:

- interaction with the silicon atoms (due to thermal vibrations);
- interaction with ionized dopant impurities in the crystal.

Parameter	Unit	Electrons	Holes	Valid range of electric field
E_0	$V\,cm^{-1}$	4.0×10^5	4.0×10^5	
a_{high}	cm^{-1}	7.03×10^5	6.71×10^5	E_0 to 6.0×10^5
a_{low}	cm^{-1}	7.03×10^5	1.582×10^6	1.75×10^5 to E_0
b_{high}	$V\,cm^{-1}$	1.231×10^6	1.693×10^6	E_0 to 6.0×10^5
b_{low}	$V\,cm^{-1}$	1.231×10^6	2.036×10^6	1.75×10^5 to E_0
$\hbar\omega_{op}$	eV	0.063	0.063	

Table 1.3: Parameters in van Overstraeten-de Man model.

Scattering with lattice

Carrier mobility is a decreasing function of temperature, as we expect collisions become more and more frequent as T gets higher (see [Lom88]). This can be represented by

$$\mu_\nu^L = \mu_\nu^0 \left(\frac{T}{T_0}\right)^{-\beta_\nu} \qquad \nu = n, p \tag{1.62}$$

where μ_ν^0 is the low-field mobility, β_ν are positive numbers and T_0 is a reference temperature, typically $T_0 = 300[K]$.

Parameter	Unit	Electrons	Holes
μ^0	$cm^2V^{-1}s^{-1}$	1417.0	470.5
β	1	2.5	2.2

Table 1.4: Parameters for mobility models including scattering from lattice thermal vibrations.

Scattering from ionized impurities

Dopant ionized impurities induce local perturbations of the periodic silicon lattice, they strongly influence the carrier motion through electrostatic interaction, reducing the mobility. To take into account this physical effect the following model has been proposed in [MS83]

$$\mu = \mu_{min1} \exp\left(-\frac{P_c}{N_{tot}}\right) + \frac{\mu^L - \mu_{min2}}{1 + \left(\frac{N_{tot}}{C_r}\right)^\alpha} - \frac{\mu_1}{1 + \left(\frac{C_s}{N_{tot}}\right)^\beta} \tag{1.63}$$

where $N_{tot} = N_D^+ + N_A^-$, μ_ν^L is given by (1.62), μ_{min1} and μ_{min2} are the minimum values of μ; P_c, C_r and C_s are reference doping values.

Parameter	Unit	Electrons	Holes
μ_{min1}	$cm^2V^{-1}s^{-1}$	52.2	44.9
μ_{min2}	$cm^2V^{-1}s^{-1}$	52.2	0
μ_1	$cm^2V^{-1}s^{-1}$	43.4	29.0
P_c	cm^{-3}	0	9.23×10^{16}
C_r	cm^{-3}	9.68×10^{16}	2.23×10^{17}
C_s	cm^{-3}	3.43×10^{20}	6.10×10^{20}
α	1	0.680	0.719
β	1	2.0	2.0

Table 1.5: Parameters for models including scattering from ionized dopant impurities.

Veclocity saturation at high electric field

Under the assumption of low electric field, mobilities are reasonably constant and the carrier drift velocity is proportional to the electric field. As the applied field strength increases, the above assumption predicts an unbounded carrier velocity as $|\mathbf{E}| \to \infty$. This outcome is physically incorrect, indeed at high fields, carriers lose their energy by optical-phonon emission [YT09]. We can take this into account by considering that

$$\lim_{|\vec{E}|\to\infty} \mu|\mathbf{E}| = v_{sat}. \tag{1.64}$$

A common adopted formula is the *Canali model* [Can75]

$$\mu = \frac{\mu_L}{\left[1 + \left(\frac{\mu_L|\mathbf{E}|}{v_{sat}}\right)^\beta\right]^{1/\beta}} \tag{1.65}$$

where μ_L is (1.62) while v_{sat} and β are given by

$$v_{sat} = v_0 exp\left(\frac{300}{T}\right)^{v_{exp}} \qquad \beta = \beta_0\left(\frac{T}{300}\right)^{\beta_{exp}}. \tag{1.66}$$

where v_0 and β_{exp} are fitting parameters.

Parameter	Unit	Electrons	Holes
v_0	$cm\,s^{-1}$	1.07×10^7	8.37×10^6
v_{exp}	1	0.87	0.52
β_0	1	1.109	1.213
β_{exp}	1	0.66	0.17

Table 1.6: Parameters for mobility models including scattering from velocity saturation.

Chapter 2

Solution of the Drift-Diffusion system

In this chapter we introduce geometry and boundary conditions for the stationary form of system (1.53) and we discuss the functional iteration algorithms used to decouple the problem.

2.1 Geometry and boundary conditions

In order to close the *Poisson equation* and the *Drift Diffusion equation* for electrons and holes in the stationary form of problem (1.53), suitable boundary conditions must be considered.

Let us consider the device domain as the union of two open disjoint subsets, Ω_{Si} (doped silicon part), and Ω_{ox} (oxide part), such that their intersection $\partial\Omega_{Si} \cap \partial\Omega_{ox} = \Gamma_{int}$ is the interface. The oxide region Ω_{ox} is assumed to be a perfect insulator so that:

$$\begin{aligned} n = p = 0 \\ \mathbf{J}_n = \mathbf{J}_p = \mathbf{0}. \end{aligned} \tag{2.1}$$

The device boundary $\partial\Omega$ is divided into two disjoint subsets: $\partial\Omega_c$ and $\partial\Omega_a$. The subset $\partial\Omega_c$ includes the so called *ohmic contacts* (with ohmic contacts we define every electrical terminal of the device on which the external input voltages are applied). Ohmic contacts are assumed to be *ideal*, they are equipotential surfaces and no voltage drop occurs at the interface between the contact and the neighbouring domain. This is well represented by suitable Dirichlet boundary conditions, therefore in the following we set $\partial\Omega_c = \Gamma_D$ and enforce:

$$\begin{aligned} \varphi &= \varphi_D \\ n &= n_D \qquad on\, \Gamma_D. \\ p &= p_D \end{aligned} \tag{2.2}$$

We point that in the case of a perfect insulator domain, (2.2) reduces to the only condition on the electrostatical potential.

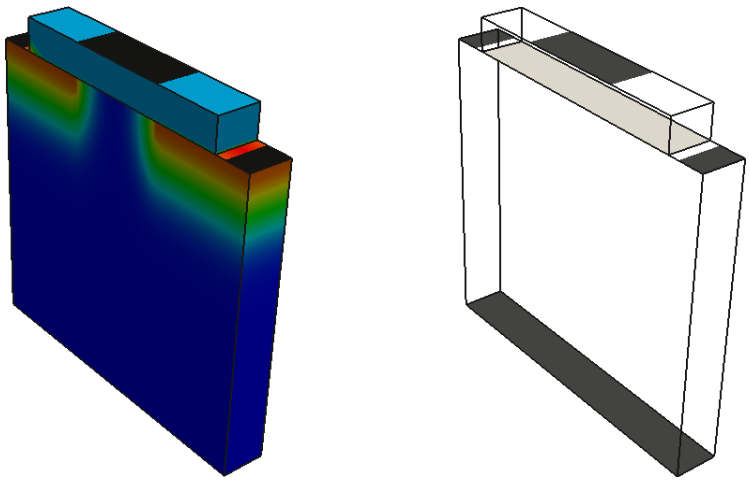
Artificial boundaries ($\partial\Omega_a$) are needed in order to obtain a self-contained simulation domain. On these boundaries no electric and current flux is exchanged with the surrounding environment, this fact being well represented by homogeneous Neumann boundary conditions ($\partial\Omega_a = \Gamma_N$):

$$\begin{aligned} \mathbf{D} \cdot \mathbf{n} &= 0 \\ \mathbf{J}_n \cdot \mathbf{n} &= 0 \qquad on\, \Gamma_N \\ \mathbf{J}_p \cdot \mathbf{n} &= 0 \end{aligned} \tag{2.3}$$

where $\mathbf{n}$ is the outward unit normal vector defined over $\partial\Omega$. As we noted before on $\partial\Omega_{ox} \cap \Gamma_N$ condition (2.3) is reduced to the first equation.

When oxide is present, the silicon boundaries for continuity equations become

$$\begin{aligned} \Gamma_{D,Si} &= \Gamma_D \cap \partial\Omega_{Si} \\ \Gamma_{N,Si} &= \Gamma_N \cap \partial\Omega_{Si} \cup \Gamma_{int}. \end{aligned} \tag{2.4}$$



(a) **(b)**

Figure 2.1: (a) MOS device with net dopant concentration distributed according to a gaussian profile and Γ_D colored in black. The oxide layer is colored in light blue. (b) Outline of the MOS device with Γ_{int} in light gray.

Fig.2.1 shows an example of boundary setting for a MOS device: in Fig.2.1a contacts are colored in black and in Fig.2.1b with light gray we indicate the interface between oxide and silicon.

Thermodynamical equilibrium and charge neutrality are the physical characteristic of an ideal contact. These conditions correspond to the following algebraic system for n_D and p_D:

$$\begin{cases} p_D n_D & = & n_i^2 \\ p_D - n_D + N_D^+ - N_A^- & = & 0 \end{cases} . \tag{2.5}$$

Solving (2.5) on $\Gamma_{D,Si}$ we have:

$$n_D = \frac{D + \sqrt{D^2 + 4n_i^2}}{2} \tag{2.6}$$

$$p_D = \frac{-D + \sqrt{D^2 + 4n_i^2}}{2} \tag{2.7}$$

where $D := N_D^+ - N_A^-$ is the net doping concentration. Furthermore at each contact, the quasi Fermi potential levels of silicon are aligned with the external applyed voltage V_{ext}

$$\varphi_n = \varphi_p = \varphi_f = V_{ext}. \tag{2.8}$$

where $\varphi_f = -E_f/q$ is the unique quasi Fermi potential level defined at the contacts. As a consequence, we can easily determine potential condition on $\Gamma_{D,Si}$ using (1.10) and (1.11)

$$\varphi_D = \varphi_f + V_{th} \ln\left(\frac{n_D}{n_i}\right) = \varphi_f - V_{th} \ln\left(\frac{p_D}{n_i}\right). \tag{2.9}$$

When $\Omega_{ox} \neq \varnothing$ we set φ_D equal to the external applied voltage on $\Gamma_D/\Gamma_{D,Si}$.

The stationary form of (1.53) can be now written in closed form as:

$$
\begin{aligned}
\nabla \cdot (-\epsilon \nabla \varphi) - q(p - n) &= qD \quad && in\, \Omega = \Omega_{ox} \cup \Omega_{Si} \\
\varphi &= \varphi_D \quad && on\, \Gamma_D \\
\nabla \varphi \cdot \mathbf{n} &= 0 \quad && on\, \Gamma_N \\
\\
\nabla \cdot (q\mu_n n \nabla \varphi - qD_n \nabla n) &= -qR \quad && in\, \Omega_{Si} \\
n &= n_D \quad && on\, \Gamma_{D,Si} \\
\nabla n \cdot \mathbf{n} &= 0 \quad && on\, \Gamma_{N,Si} \\
\\
\nabla \cdot (-q\mu_p p \nabla \varphi - qD_p \nabla p) &= -qR \quad && in\, \Omega_{Si} \\
p &= p_D \quad && on\, \Gamma_{D,Si} \\
\nabla p \cdot \mathbf{n} &= 0 \quad && on\, \Gamma_{N,Si}.
\end{aligned}
\tag{2.10}
$$

The highly nonlinear coupled nature of system (2.10) makes an analytical treatment very difficult, if even not impossible. For this reason, numerical schemes must be used to compute an approximate solution.

2.2 Iteration algorithms

The most used algorithms for the iterative treatment of (2.10) are *the fully coupled Newton's method* and *the decoupled Gummel map*. System (2.10) can be written in compact form as

$$
\mathbf{F}(\mathbf{U}) = \mathbf{0} \tag{2.11}
$$

where

$$
\mathbf{U} := [\varphi, n, p]^T, \qquad \mathbf{F}(\mathbf{U}) := \begin{bmatrix} F_1(\mathbf{U}) \\ F_2(\mathbf{U}) \\ F_3(\mathbf{U}) \end{bmatrix} \tag{2.12}
$$

having set:

$$
\begin{aligned}
F_1(\mathbf{U}) &= \nabla \cdot (-\epsilon \nabla \varphi) - q(p - n + D) \\
F_2(\mathbf{U}) &= \nabla \cdot (q\mu_n n \nabla \varphi - qD_n \nabla n) + qR \\
F_3(\mathbf{U}) &= \nabla \cdot (-q\mu_p p \nabla \varphi - qD_p \nabla p) + qR.
\end{aligned}
$$

Problem (2.11) is the generalization of the search of a zero for a real function $f : \mathbb{R} \to \mathbb{R}$. Since the vector function $\mathbf{F}$ is a nonlinear differential

operator, the associated problem which we intend to solve is: given a functional space V and the operator $\mathbf{F} : V \to V$, find $\mathbf{U} \in V$ such that (2.11) is satisfied.

In our application, the function space V is typically a subset of the Sobolev space $[H^1(\Omega)]^d$ (where d is the number of component of $\mathbf{F}$). The general form of a Sobolev space for an integer $m \geq 0$ is

$$H^m(\Omega) := \left\{ v : D^\alpha v \in L^2(\Omega), \forall |\alpha| \leq m \right\}. \tag{2.13}$$

where $L^2(\Omega)$ is the space of square integrable functions on Ω defined as

$$L^2(\Omega) := \left\{ v : \int_\Omega |v|^2 d\,\Omega = ||v||^2_{L^2(\Omega)} < +\infty \right\}. \tag{2.14}$$

On these spaces, we use the semi-norm

$$|v|^2_{m,\Omega} = \sum_{|\alpha|=m} ||D^\alpha v||^2_{L^2(\Omega)} \tag{2.15}$$

and the norm

$$||v||^2_{m,\Omega} = \sum_{k \leq m} |D^\alpha v|^2_{k,\Omega}. \tag{2.16}$$

We also need consider functions that vanish on either the entire or a part of the boundary:

$$H^1_0 := \left\{ v : v \in H^1(\Omega), v|_{\partial\Omega} = 0 \right\} \tag{2.17}$$

$$H^1_{0,\Gamma_D} := \left\{ v : v \in H^1(\Omega), v|_{\Gamma_D} = 0 \right\} \tag{2.18}$$

For $v \in H^1_0(\Omega), H^1_{0,\Gamma_D}(\Omega)$ we have the *Poincaré - Friedrich's inequality* [Sal10]

$$|v|_{0,\Omega} \leq C(\Omega)|v|_{1,\Omega} \tag{2.19}$$

from which it follows that the seminorm $|\cdot|_\Omega$ is actually a norm in $H^1(\Omega)$, equivalent to $||\cdot||_{1,\Omega}$.

The above function spaces (for a more detailed description see [AF03]) are used widely in this work, especially during the well-posedness analysis as reported in Chapter 3.

2.2.1 Newton's method

Definition 2.1 (Frechèt differentiability)**.** Let be X and Y two vector spaces. Given $f, g \in X$ and a functional $F : X \to Y$, the functional F is Frechét differentiable if there exists a linear bounded operator $A_f : X \to Y$ such that

$$\lim_{||g||\to 0} \frac{||F(f+g) - F(f) - A_f(g)||_Y}{||g||_X} = 0, \tag{2.20}$$

where $||\cdot||_X$ and $||\cdot||_Y$ are the norms on X and Y respectively. If the above limit exists, we write $DF(f) = A_f$ and call it the Frechét derivative of F at f.

Considering the functional operator (2.12) we can easily compute the associated *Jacobian matrix* $\mathbf{F}'$, whose $(i,j) - th$ entry represents the Frechét derivative of the $i - th$ row of the non linear operator with respect to the $j - th$ variable, defined as

$$\mathbf{F}'_{ij}(\mathbf{U})[\mathbf{V}]_j := \lim_{\eta\to 0} \frac{F_i(\mathbf{U} + \eta[\mathbf{V}]_j) - F_i(\mathbf{U})}{\eta} \qquad \mathbf{V} \in V \tag{2.21}$$

where $[\mathbf{V}]_j \in V$ is the projection of $\mathbf{V}$ in the $j - th$ direction.

$\mathbf{F}'_{ij}(\cdot)$ is a linear operator from V into the space $L(V, V)$ of linear continuos functionals from V into V, while $\mathbf{F}'_{ij}(\mathbf{U})$ is the Frechét derivative of the functional F_i with respect to the variable $[\mathbf{U}]_j$.

According to the above definitions the Newton method reads as follows:

Newton's method

Let X, Y be two vector spaces and $\mathbf{F} : X \to Y$ a function operator Frechèt differentiable, given an initial datum $\mathbf{U}^0 \in X$ and $toll > 0$, for all $k \geq 0$ solve the following linear problem:

$$\begin{aligned} \mathbf{F}'(\mathbf{U}^k)\delta\mathbf{U}^k &= -\mathbf{F}(\mathbf{U}^k) \\ \mathbf{U}^{k+1} &= \mathbf{U}^k + \delta\mathbf{U}^k \end{aligned} \tag{2.22}$$

until $||\mathbf{F}(\mathbf{U}^{k+1})||_Y < toll$.

The application of Newton's method has transformed the original problem (2.11) into the *fixed-point problem* of finding $\mathbf{U} \in V$ such that

$$\mathbf{U} = T_{\mathbf{F}}(\mathbf{U}) \tag{2.23}$$

where

$$T_{\mathbf{F}}(\mathbf{U}) = \mathbf{F}'(\mathbf{U})^{-1}(\mathbf{F}'(\mathbf{U})\mathbf{U} - F(\mathbf{U})) \tag{2.24}$$

is the *iteration function* associated with the Newton method. The main result about the convergence of this method is the following [OR70].

Theorem 2.1. *Let $\mathbf{U} \in V$ be a solution of problem (2.11). Assume that $\mathbf{F}'$ is Lipschitz continuous in the ball $\mathcal{B}(\mathbf{U}, \delta)$, i.e., that there exists $K > 0$ such that*

$$||\mathbf{F}'(\mathbf{v}) - \mathbf{F}'(\mathbf{z})||_{L(V,V)} \leq K||\mathbf{v} - \mathbf{z}||_V \qquad \forall \mathbf{v}, \mathbf{z} \in \mathcal{B}(\mathbf{U}, \delta),\ \mathbf{v} \neq \mathbf{z}. \tag{2.25}$$

Then there exists in correspondence $\delta' > 0$, with $\delta' \leq \delta$, such that for all $\mathbf{U}^0 \in \mathcal{B}(\mathbf{U}, \delta')$ the sequence $\left\{\mathbf{U}^k\right\}$ generated by (2.22) converges quadratically to $\mathbf{U}$, i.e., there exists $C > 0$ such that, for a suitable $k_0 \geq 0$ we have

$$||\mathbf{U} - \mathbf{U}^{k+1}||_V \leq C||\mathbf{U} - \mathbf{U}^k||_V^2 \qquad \forall k \geq k_0. \tag{2.26}$$

2.2.2 Fully coupled Newton's method

If we consider the linearization of system (2.10) the Jacobian matrix of the Newton method is a 3x3 matrix and the associated problem is

$$\begin{bmatrix} F_{1,\varphi} & F_{1,n} & F_{1,p} \\ F_{2,\varphi} & F_{2,n} & F_{2,p} \\ F_{3,\varphi} & F_{3,n} & F_{3,p} \end{bmatrix} \begin{bmatrix} \delta\varphi \\ \delta n \\ \delta p \end{bmatrix} = \begin{bmatrix} -F_1(\varphi, n, p) \\ -F_2(\varphi, n, p) \\ -F_3(\varphi, n, p) \end{bmatrix}. \tag{2.27}$$

Each row of the above matrix is a PDE that can be discretized using the FEM. Denoting by N_{dof} the number of degrees of freedom (dofs) to represent $\delta\varphi$, δn and δp we see that the structure of the discrete problem associated with (2.27) is the following linear algebraic system

$$\begin{bmatrix} \mathbf{K}_{1,\varphi} & \mathbf{K}_{1,n} & \mathbf{K}_{1,p} \\ \mathbf{K}_{2,\varphi} & \mathbf{K}_{2,n} & \mathbf{K}_{2,p} \\ \mathbf{K}_{3,\varphi} & \mathbf{K}_{3,n} & \mathbf{K}_{3,p} \end{bmatrix} \begin{bmatrix} \delta\varphi \\ \delta n \\ \delta p \end{bmatrix} = \begin{bmatrix} -\mathbf{F}_1(\varphi, n, p) \\ -\mathbf{F}_2(\varphi, n, p) \\ -\mathbf{F}_3(\varphi, n, p) \end{bmatrix} \tag{2.28}$$

where each matrix $\mathbf{K}$ is a block of size $N_{dof} \times N_{dof}$. This implies that at every iteration step we have to solve a linear problem of $3 \times N_{dof}$ variables. Moreover, to ensure convergence of the Newton iterative process, it is important to provide a very good initial guess vector $[\varphi^0, n^0, p^0]$. Because the problem variables have different orders of magnitude and the Jacobian matrix is often quite ill-conditioned, appropriate scaling and balancing techniques are needed in order to avoid problems associated with round-off error.

2.2.3 Gummel map algorithm

In 1964 H. K. Gummel proposed an original and alternative to (2.27) approach in order to solve system (2.10) in a semiconductor device in one spatial dimension [Gum64]. The main idea of the algorithm is to move the nonlinearity to the Poisson equation only, and once obtained the electric potential profile, both continuity equations are solved in linear form. This is possible if we consider the Maxwell-Boltzmann approximation for electrons (1.10) and holes (1.11) obtaining

$$F_1(\varphi) = \nabla \cdot (-\epsilon \nabla \varphi) - q(n_i(e^{(\varphi_p - \varphi)/V_{th}} - e^{(\varphi - \varphi_n)/V_{th}}) + D). \tag{2.29}$$

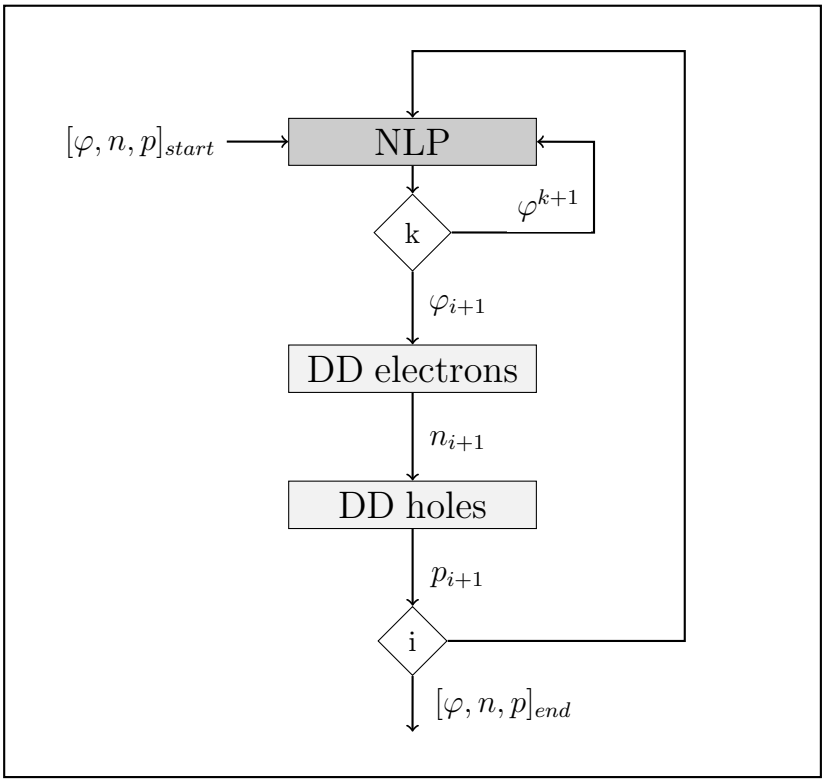


Figure 2.2: Flow chart of Gummel algorithm.

The Gummel algorithm is represented by the following iteration.

Decoupled Gummel map.

0. Give a suitable initial condition for φ^0 and set a positive parameter $toll_{GM} > 0$ (Gummel Map tollerance)

1. Fix a positive parameter $toll_{NLP} > 0$ (Non Linear Poisson tolerance), solve the linearized Non Linear Poisson equation (NLP) in Ω using the Newton method until $||F_1(\varphi^{k+1})|| > toll_{NLP}$:

$$\begin{cases} \nabla \cdot (-\epsilon_{Si} \nabla \delta\varphi^k) + \dfrac{1}{V_{th}} \sigma^k_{Si} \delta\varphi^k = f^k_{Si} & in\ \ \Omega_{Si} \\ \nabla \cdot (-\epsilon_{ox} \nabla \delta\varphi^k) = f^k_{ox} & in\ \ \Omega_{ox} \\ \delta\varphi^k = 0 & on\ \ \Gamma_D \\ \nabla \delta\varphi^k \cdot \mathbf{n} = 0 & on\ \ \Gamma_N \\ \varphi^{k+1} = \varphi^k + \delta\varphi^k \end{cases} \tag{2.30}$$

having set:

$$\sigma^k_{Si}(\varphi^k) = q n_i \left[e^{(\varphi_p - \varphi^k)/V_{th}} - e^{(\varphi^k - \varphi_n)/V_{th}} \right]$$
$$f^k_{Si}(\varphi^k) = \nabla \cdot (-\epsilon \nabla \varphi^k) + q n_i \left[e^{(\varphi_p - \varphi^k)/V_{th}} - e^{(\varphi^k - \varphi_n)/V_{th}} + D \right]$$
$$f^k_{ox}(\varphi^k) = \nabla \cdot (-\epsilon \nabla \varphi^k).$$

2. Solve the Linear Electron Continuity Equation (LEC):

$$\begin{cases} \nabla \cdot (q\mu_n n \nabla \varphi^i - q D_n \nabla n) = -qR(n^{i-1}, p^{i-1}) & in\ \ \Omega_{Si} \\ n = n_D & on\ \ \Gamma_{D,Si} \\ \nabla n \cdot \mathbf{n} = 0 & on\ \ \Gamma_{N,Si} \end{cases} \tag{2.31}$$

3. Solve the Linear Hole Continuity Equation (LHC):

$$\begin{cases} \nabla \cdot (-q\mu_p p \nabla \varphi^i - q D_p \nabla p) = -qR(n^{i-1}, p^{i-1}) & in\ \ \Omega_{Si} \\ p = p_D & on\ \ \Gamma_{D,Si} \\ \nabla p \cdot \mathbf{n} = 0 & on\ \ \Gamma_{N,Si} \end{cases} \tag{2.32}$$

4. If $\max\{||\varphi^i - \varphi^{i-1}||_{L^\infty}, ||p^i - p^{i-1}||_{L^\infty}, ||n^i - n^{i-1}||_{L^\infty}\} > toll_{GM}$ restart from step (1).

Fig.2.2 shows a flow chart of the Gummel algorithm where k is the iteration step of the inner loop, while i is the iteration step of the Gummel Map.

For the Gummel map, a result analogous to Thm.2.1 can be found in [Jer96], but the convergence rate is linear, although heuristic experience often shows superlinear convergence behaviour.

There are several advantages which make Gummel map algorithm more attractive than the Fully Coupled Newton's Method: first of all simulation experience shows that the Gummel process is much more insensitive to the choice of the initial guess than Newton's method. This is particularly important in multidimensional problems where it is far from trivial to design a good starting point for initializing the iterative procedure. Another important feature is the reduced computational effort and memory cost: at each iteration step the Gummel algorithm requires the successive solution of three problems, each one of size equal to $N_{dof} \times N_{dof}$.

Let us discuss again steps 2-3 of *Decoupled Gummel map*. According with (1.54) the general R/G phenomenon can be separated in a reaction term and a force term (except for the II which is only a force term contribution). Letting:

$$\begin{array}{ll} R_n^{i-1}(n) & = \sigma_n^{i-1} n - f^{i-1} \\ R_p^{i-1}(p) & = \sigma_p^{i-1} p - f^{i-1} \end{array} \tag{2.33}$$

where

$$\sigma_n = \frac{p^{i-1}}{F(p^{i-1}, n^{i-1})} \qquad \sigma_p = \frac{n^{i-1}}{F(p^{i-1}, n^{i-1})} \qquad f = \frac{n_i^2}{F(p^{i-1}, n^{i-1})}, \tag{2.34}$$

we can write systems (2.31) and (2.32) as:

$$\begin{cases} \nabla \cdot (q\mu_n n \nabla \varphi^i - qD_n \nabla n) + q\sigma_n^{i-1} n = q f^{i-1} & in \ \ \Omega_{Si} \\ n = n_D & on \ \ \Gamma_{D,Si} \\ \nabla n \cdot \mathbf{n} = 0 & on \ \ \Gamma_{N,Si} \end{cases} \tag{2.35}$$

$$\begin{cases} \nabla \cdot (-q\mu_p p \nabla \varphi^i - qD_p \nabla p) + q\sigma_p^{i-1} p = q f^{i-1} & in \ \ \Omega_{Si} \\ p = p_D & on \ \ \Gamma_{D,Si} \\ \nabla p \cdot \mathbf{n} = 0 & on \ \ \Gamma_{N,Si}\,. \end{cases} \tag{2.36}$$

The above splitting of R/G term is called *lagging approach* [Jer96] and corresponds to extending to the non-linear case the classical *Jacobi* method for the iterative solution of linear algebraic systems. Since equations are sequentially solved, it is possible to take advantage of the solution at the present step. Indeed an alternative approach would be use the solution of the first solved equation to compute the R/G contribution in the second equation. In such a case, the lagging method corresponds to extending to the nonlinear case the classical *Gauss-Seidel* method for the iterative solution in linear algebraic systems.

Chapter 3

Finite element discretization

In this chapter we present the weak formulation of problems (2.30), (2.35) and (2.36). For each weak problem we discuss the well-posedness analysis and describe the finite element discretization.

3.1 Non Linear Poisson Equation: weak form

Let us write problem (2.30) in compact form:

$$\begin{cases} \nabla \cdot (-\epsilon \nabla \delta\varphi^k) + \sigma^k \delta\varphi^k &= f^k & in\ \Omega \\ \delta\varphi^k &= 0 & on\ \Gamma_D \\ \nabla \delta\varphi^k \cdot \mathbf{n} &= 0 & on\ \Gamma_N \\ \varphi^{k+1} &= \varphi^k + \delta\varphi^k & \end{cases} \tag{3.1}$$

having set:

$$\begin{aligned} \epsilon &= \epsilon_s \mathcal{I}_{\Omega_{Si}} + \epsilon_{ox} \mathcal{I}_{\Omega_{ox}} \\ f &= f_s \mathcal{I}_{\Omega_{Si}} + f_{ox} \mathcal{I}_{\Omega_{ox}} \\ \sigma &= \sigma_s \mathcal{I}_{\Omega_{Si}} \end{aligned}$$

where $\mathcal{I}_A(\mathbf{x})$ is equal to 1 if $\mathbf{x} \in A$ and 0 otherwise. System (3.1) is a Diffusion-Reaction (DR) problem in Ω, with respect to the dependent variable $\delta\varphi^k$. Now we multiply the first equation in (3.1) by a test function $v \in H^1_{\Gamma_D}(\Omega)$ and integrating over all the domain we obtain

$$-\int_\Omega \nabla \cdot (-\epsilon \nabla \delta\varphi^k) v \, d\Omega + \int_\Omega \sigma^k \delta\varphi^k v \, d\Omega = \int_\Omega f^k v \, d\Omega \qquad \forall v \in H^1_{\Gamma_D}(\Omega). \tag{3.2}$$

Applying Green's formula and including the boundary conditions, we obtain the weak formulation of (3.1) which reads: find $\delta\varphi^k \in H^1_{\Gamma_D}(\Omega)$ such that

$$\int_\Omega \epsilon\nabla\delta\varphi^k\nabla v\,d\Omega + \int_\Omega \sigma^k\delta\varphi^k v\,d\Omega = \int_\Omega f^k v\,d\Omega \qquad \forall v \in H^1_{\Gamma_D}(\Omega). \tag{3.3}$$

We are able to define the following bilinear form

$$a : H^1_{\Gamma_D}(\Omega)\times H^1_{\Gamma_D}(\Omega)\to\mathbb{R}, \qquad a(u,v) = \int_\Omega \epsilon\nabla u\nabla v\,d\Omega + \int_\Omega \sigma^k uv\,d\Omega. \tag{3.4}$$

and the linear and bounded functional

$$F : H^1_{\Gamma_D}(\Omega)\to\mathbb{R}, \qquad F(v) = \int_\Omega f^k v\,d\Omega. \tag{3.5}$$

In order to prove the existence and uniqueness of the solution of (3.3), we apply the *Lax-Milgram theorem* [Sal10]. Well-posedness is ensured by several physical hypotheses:

- $\epsilon \in L^\infty(\Omega)$ and $\epsilon(\mathbf{x}) > 0$ a.e. in Ω;
- $\forall k \geq 0\ \sigma^k \in L^\infty(\Omega)$ and $\sigma^k(\mathbf{x}) > 0$ a.e. in Ω_{Si}.

We define some useful quantities:

$$\begin{array}{ll} \epsilon_M = \max_\Omega \epsilon & \epsilon_m = \min_\Omega \epsilon \\ \sigma_M = \max_\Omega \sigma & \sigma_m = \min_\Omega \sigma = 0 \end{array}$$

Taking into account the above hypotheses it is possible to prove the following properties:

- **Continuity of the bilinear form:**

  $\forall u, v \in H^1_{\Gamma_D}$

  $$\begin{array}{ll} |\int_\Omega \epsilon\nabla u\nabla v + \int_\Omega \sigma^k uv| & \leq \epsilon_M||\nabla u||_{L^2}||\nabla v||_{L^2} + \sigma_M||u||_{L^2}||v||_{L^2} \\ & \leq \max\{\epsilon_M,\sigma_M\}\,(||\nabla u||_{L^2}||\nabla v||_{L^2} + ||u||_{L^2}||v||_{L^2}) \\ & \leq \max\{\epsilon_M,\sigma_M\}||u||_{H^1_{\Gamma_D}}||v||_{H^1_{\Gamma_D}}. \end{array}$$

- **Coercivity of the bilinear form:**

$$\forall u \in H^1_{\Gamma_D}$$

$$\begin{aligned} |\int_\Omega \epsilon \nabla u \nabla u + \int_\Omega \sigma^k u^2| &\geq \epsilon_m ||\nabla u||^2_{L^2} + \sigma_m ||u||^2_{L^2} \\ &= \epsilon_m ||\nabla u||^2_{L^2} \\ &= \epsilon_m |\nabla u|^2_{H^1_{\Gamma_D}} \equiv \epsilon_m ||u||^2_{H^1_{\Gamma_D}} . \end{aligned}$$

- **Continuity of the functional:**

$$|\int_\Omega f^k v| \leq ||f^k||_{L^2} ||v||_{H^1_{\Gamma_D}} \qquad \forall v \in H^1_{\Gamma_D}.$$

Then we can state that $\forall k \geq 0$ there exists a unique weak solution of the linearized Non Linear Poisson equation.

3.2 Continuity Equations: weak form

Without loss of generality we consider only the electron continuity equation. System (2.35) is a diffusion-advection-reaction (DAR) problem in conservative form. With a suitable change of variables we are able to treat these PDE likewise the linearized Non Linear Poisson equation in the previous section. In fact, using the Slotboom variable (1.38), we can write system (2.35) as:

$$\begin{cases} \nabla \cdot \left(-qD_n e^{\varphi^i / V_{th}} \nabla u_n \right) + \sigma_n^{i-1} e^{\varphi^i / V_{th}} u_n &= f^{i-1} & in\ \Omega_{Si} \\ u_n &= n_D e^{-\varphi^i / V_{th}} & on\ \Gamma_{D,Si} \\ \nabla u_n \cdot \mathbf{n} &= 0 & on\ \Gamma_{N,Si} . \end{cases} \tag{3.6}$$

Proceeding as in Section 3.1, the weak formulation of the Electron Continuity equation is:

find $u_n \in H^1_{\Gamma_{D,Si}}(\Omega)$ such that

$$\int_{\Omega_{Si}} qD_n e^{\varphi^i / V_{th}} \nabla u_n \nabla v \, d\Omega + \int_{\Omega_{Si}} \sigma_n^{i-1} e^{\varphi^i / V_{th}} u_n v \, d\Omega = \int_{\Omega_{Si}} f^{i-1} v \, d\Omega \qquad \forall v \in H^1_{\Gamma_{D,Si}}. \tag{3.7}$$

Existence and uniqueness of the unknown variable u_n ensures the same properties on n, thanks to the relation (1.38) between u_n and n. Further hypotheses on the coefficients $\forall i \geq 0$ are:

- $qD_n e^{\varphi^i / V_{th}} \in L^\infty(\Omega_{Si})$ and $D_n(\mathbf{x}) > 0$ a.e. in Ω_{Si};
- $\sigma_n^{i-1} e^{\varphi^i / V_{th}} \in L^\infty(\Omega_{Si})$ and $\sigma_n^{i-1}(\mathbf{x}) > 0$ a.e. in Ω_{Si}.

We define the bilinear form

$$a(u,v) = \int_{\Omega_{Si}} qD_n e^{\varphi^i/V_{th}} \nabla u_n \nabla v \, d\Omega + \int_{\Omega_{Si}} \sigma_n^{i-1} e^{\varphi^i/V_{th}} u_n v \, d\Omega \tag{3.8}$$

and the linear and bounded functional

$$F(v) = \int_{\Omega_{Si}} f^{i-1} v \, d\Omega \tag{3.9}$$

Well-posedness of problem (3.7) is verified using the same arguments as in Section 3.1.

3.3 Numerical approximation

In this section we introduce the Galerkin method to approximate the weak formulations (3.3) and (3.7) (see [QV08]). Each of them can be represented in compact form as:

find $u \in V$ such that

$$a(u,v) = F(v) \qquad \forall v \in V \tag{3.10}$$

where V is the space of admissible functions, i.e., $H^1_{\Gamma_D}(\Omega)$ or $H^1_{\Gamma_{D,Si}}(\Omega_{Si})$. Let us introduce V_h which is a family of finite-dimensional subspaces of V, depending on a positive parameter h, such that

$$V_h \subset V, \quad \dim V_h < \infty \qquad \forall h > 0 \tag{3.11}$$

The *Galerkin problem* associated with (3.10) reads:

find $u_h \in V_h$ such that

$$a(u_h, v_h) = F(v_h) \qquad \forall v_h \in V_h. \tag{3.12}$$

Unique solvability of (3.12) is an immediate consequence of the analysis carried out in Sections 3.1 and 3.2.

Let $\mathcal{T}_h$ be a partition of Ω, and K a generic element of $\mathcal{T}_h$ such that $\bar{\Omega} = \bigcup \bar{K}$. In this case the parameter h represents the characteristic dimension of each element K. Let us introduce the general finite element spaces of the polynomial elementwise functions

$$X_h^r(\Omega) := \{v_h \in C^0(\bar{\Omega}) : v_h|_K \in \mathbb{P}_r(K), \forall K \in \mathcal{T}_h\} \tag{3.13}$$

and the associated space where functions vanish on boundaries

$$X^r_{h,\Gamma_D}(\Omega) := \{v_h \in X^r_h : v_h|_{\Gamma_D} = 0\}. \tag{3.14}$$

If $\Omega \subset \mathbb{R}^3$ we have

$$\dim\mathbb{P}_r(K) := \binom{3+r}{r}. \tag{3.15}$$

We approximate $H^1_{\Gamma_D}(\Omega)$ with $X^1_{h,\Gamma_D}(\Omega)$ and $H^1_{\Gamma_{D,Si}}(\Omega_{Si})$ with $X^1_{h,\Gamma_{D,Si}}(\Omega_{Si})$. Therefore according to (3.15) we have:

$$\begin{aligned} \dim\mathbb{P}_1(K) &= 4 \\ \dim X^1_h &= N_h \\ \dim X^1_{h,\Gamma_D} &= N_h - N_g \end{aligned}$$

where N_h is the number of vertices of the partition $\mathcal{T}_h$ and N_g is the number of vertices that belong to the Dirichlet boundary.

We denote by $\{\psi_j\}_{j=1}^{N_h}$ the Lagrangian basis of the space X^1_h in such a way that

$$u_h(\mathbf{x}) = \sum_{j=1}^{N_h} u_j \psi_j(\mathbf{x}). \tag{3.16}$$

Since each function of V_h is a linear combination of ψ_i for $i = 1, \ldots, N_h$, the Galerkin problem (3.12) becomes:

find $[u_1, u_2, ..., u_{N_h}]^T \in \mathbb{R}^{N_h}$ such that

$$\sum_{j=1}^{N_h} u_j a(\psi_j, \psi_i) = F(\psi_i) \qquad \forall i = 1, \ldots, N_h. \tag{3.17}$$

In order to implement problem (3.17) it is convenient to express the bilinear form $a(\cdot,\cdot)$ and the linear functional $F(\cdot)$ with respect to each element of the partition $\mathcal{T}_h$ as

$$\sum_{j=1}^{N_h} u_j \sum_{K\in\mathcal{T}_h} a_K(\psi_j, \psi_i) = \sum_{K\in\mathcal{T}_h} F_K(\psi_i) \qquad \forall i = 1, \ldots, N_h. \tag{3.18}$$

3.3.1 Geometrical discretization

Each element $K \in \mathcal{T}_h$ is a tetrahedron of volume $|K|$. From now on, we assume that there exists a constant $\delta > 0$ such that

$$\frac{h_K}{\rho_K} \leq \delta \qquad \forall K \in \mathcal{T}_h \tag{3.19}$$

where $h_k = \text{diam}(K) = \max_{x,y \in K} |x - y|$ and ρ_K is the diameter of the sphere inscribed in the tetrahedron K. Condition (3.19) is the so called *mesh regularity condition* [Qua08] [QV08]. We denote with $\mathcal{E}_h$, $\mathcal{V}_h$ and $\mathcal{F}_h$ the set of all the edges, vertices and faces of $\mathcal{T}_h$ respectively, and for each $K \in \mathcal{T}_h$ we denote by ∂K and $\mathbf{n}_{\partial K}$ the boundary of the element and its outward unit normal.

We notice that $\mathcal{T}_h$ is built in such a way that every K belongs to a single region, while it is possible that its vertices belong to different regions.

3.3.2 Linearized Non Linear Poisson equation

Concerning with the linearized NLP equation we have

$$a(\psi_j, \psi_i) = \int_\Omega \epsilon \nabla \psi_j \nabla \psi_i \, d\Omega + \int_\Omega \sigma^k \psi_j \psi_i \, d\Omega \tag{3.20}$$

and the restriction on each element K is

$$a_K(\psi_j, \psi_i) = \int_K \epsilon \nabla \psi_j \nabla \psi_i \, dK + \int_K \sigma^k \psi_j \psi_i \, dK. \tag{3.21}$$

Equation (3.21) contains two distinct contributions: the first one identifies the diffusive contribution and generates the so-called *stiffness matrix*, while the second refers to the reaction term and generates the *mass matrix*.

The coefficient ϵ is a piecewise constant function, which changes on different material regions. Therefore ϵ is constant over each element and the first integral in (3.21) becomes easier to compute.

As a consequence of choosing the discrete space X_h^1, we can not expect a better convergence rate than the first order in $|| \cdot ||_{1,\Omega}$ with respect to h [QV08]. This implies that the trapezoidal rule is enough accurate and an high-order quadrature rule is not needed. The main consequence of the using trapezoidal quadrature rule is that the mass-matrix becomes diagonal. This technique is well known as *lumping procedure* applied on the mass-matrix.

The entries of the local system matrix A_K^k are

$$[A_K^k]_{ij} = \epsilon_K L_{ij} + \frac{|K|}{4} \sigma_i^k \tag{3.22}$$

having set:

$$\begin{aligned} L_{ij} &= \int_K \nabla\psi_i \nabla\psi_j \, d\Omega \\ \sigma_i^k &= \sigma^k(\mathbf{x}_i). \end{aligned} \tag{3.23}$$

The construction of the right hand side of (3.18) using the trapezoidal rule yields

$$[F_K]_i^k = f_i^k \frac{|K|}{4} \simeq \int_\Omega f^k \psi_i \, d\Omega. \tag{3.24}$$

The local contributions of each element K are assembled in the global matrix A as follows. Let I be the global index of a generic vertex belonging to the partition $\mathcal{T}_h$. We denote by $\mathcal{J}_K : \mathcal{V}_{\mathcal{T}_h} \to \mathcal{V}_K$ the map which connects I to its corresponding local index $i = 1, \ldots, 4$ in the element K. Then we have

$$A_{IJ}^k = \sum_{\substack{\forall K \in \mathcal{T}_h s.t. \\ \mathcal{J}_K(I), \mathcal{J}_K(J) \subset \mathcal{V}_K}} [A_K]_{ij}^k. \tag{3.25}$$

Analogously the force term $\mathbf{b}^k$ is

$$b_I^k = \sum_{\substack{\forall K \in \mathcal{T}_h s.t. \\ \mathcal{J}_K(I) \subset \mathcal{V}_K}} [F_K]_i^k. \tag{3.26}$$

Once we have built the global matrix A^k and the global vector $\mathbf{b}^k$ we need to take into account the essential boundary conditions. In fact the displacement formulation is a primal method which enforces Dirichlet boundary condition in a strong manner. Therefore we have to modify the algebraic system. We choose the *diagonalization* technique which does not alter the matrix pattern nor introduce ill-conditioning in the linear system. Let i_D be the generic index of a Dirichlet node, we denote by $[\delta\varphi_D]_i$ (which in this case is equal to zero) the known value of the solution $\delta\varphi$ at the node. We consider the Dirichlet condition as an equation of the form $\beta[\delta\varphi]_i = \beta[\delta\varphi_D]_i$, where $\beta \neq 0$ is a suitable coefficient. In order to avoid degradation of the global matrix condition number, we take β equal to the diagonal element of the matrix at row i_D.

Finally, the discretization of step 1 in the Gummel algorithm, reads:

$$\begin{cases} A^k \delta\varphi^k &= \mathbf{b}^k \\ \varphi^{k+1} &= \varphi^k + \delta\varphi^k. \end{cases} \tag{3.27}$$

As every iteration procedure, problem (3.27) needs a suitable convergence break criterion. A good approach is based on checking the satisfaction of the

fixed point equation (2.11) by the k-th solution. In this case the inner loop of the Gummel Map reads as: given a tolerance $toll > 0$ solve problem (3.27) until

$$||\mathbf{b}(\varphi^{k+1})||_2 > toll \tag{3.28}$$

where $||\cdot||_2$ is the usual Euclideian norm for a vector.

Damping

Despite the validity of Thm.2.1, the use of the Newton method may be affected by numerical implications. The main problem is the fact that the method overestimate the length of the correction step. This phenomenon is frequently indicated as *overshoot*. In the case of the semiconductor equations this overshoot problem can be treated by simply limiting the size of the correction vector ($\delta\varphi$) determined by Newton's method. The usual established modifications to avoid overshoot are given by the following formulation

$$\tilde{A}(\varphi_k) = \frac{1}{t_k} A(\varphi_k) \tag{3.29}$$

where t_k is a positive parameter to be properly chosen. With $t_k = 1$ the modified Newton method reduces to the classical Newton method. For the case (3.29) a simple criterion suggested by Deuflhard [Deu74], prescribes t_k to be taken in $(0, 1]$ in such a way that for any norm, we have

$$||A(\varphi_k)^{-1}\mathbf{b}(\varphi_k - t_k A(\varphi_k)^{-1}\mathbf{b}(\varphi_k))|| < ||A(\varphi_k)^{-1}\mathbf{b}(\varphi_k)||. \tag{3.30}$$

Condition (3.30) guarantees that the norm of the residual is decreasing with k. This condition is hardly to be evaluated because of the presence of the inverse of A. If the solution is accomplished using a direct method like the LU factorization, the evaluation of the argument of the norm on the left hand side of (3.30) is reduced to a forward and backward substitution and the evaluation of $\mathbf{b}(\varphi)$. However we use an iterative method (BCG solver based on [PTVF07]) and this implies serious difficulties to the application of the criterion (3.30). In order to overcome this problem we replace A in (3.30) with the main diagonal $D(\varphi_k)$:

$$||D(\varphi_k)^{-1}\mathbf{b}(\varphi_k - t_k D(\varphi_k)^{-1}\mathbf{b}(\varphi_k))|| < ||D(\varphi_k)^{-1}\mathbf{b}(\varphi_k)||. \tag{3.31}$$

This criterion has been adopted in our code. However the value to use for t_k is a question of trial and error. Frequently the following sequence is used:

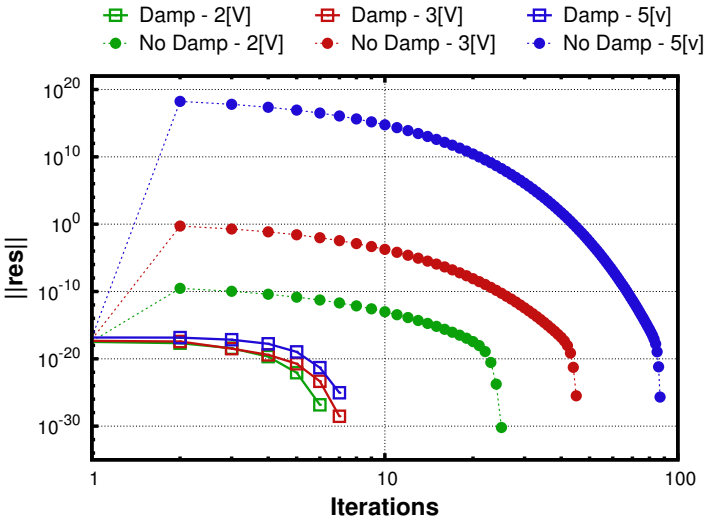
$$t_k = \frac{1}{2^i} \tag{3.32}$$

$$t_k = \frac{1}{2^{\frac{i(i+1)}{2}}} \tag{3.33}$$

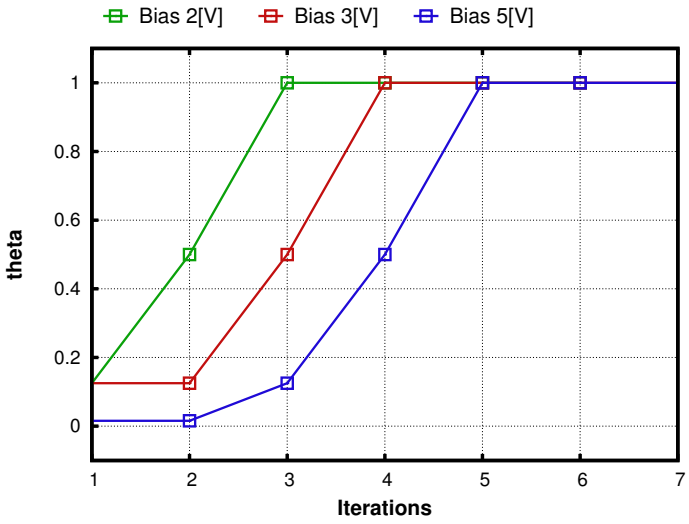
where i is the number of subiterations of the damping procedure needed to satisfy (3.31). Close to the solution, (3.30) (and so (3.31)) will be satisfied with $t_k = 1$ so that the quadratic convergence properties of the classical Newton method Thm.2.1 are recovered.

The benefits due to the damping techinque are visible in Fig.3.1a, where for different voltages applied to a p-n junction, the evolution of the residual, for the first Gummel map iteration, is shown. When damping is switched off the first iterations are critical, because the solution found is very far from the real one. When damping is switched on this problem is removed and the scheme converges with fewer iterations. In some heuristic sense this procedure guarantees a progressive approaching of the solution to the ball mentioned in Thm.2.1, where the convergence rate is quadratic. In Fig.3.1b the evolution of the coefficient t_k as a function of the iteration counter k is shown. The curves are monotonic from 0 to 1 below and this means that damping procedure is more relevant in the first iterations than in the last, where the standard Newton method is recovered.

Finally, we notice that for high voltages the scheme needs more iterations to converge. This phenomenon is strictly related to the shape of the initial guess as we shall see in detail in Section 4.1.1.



(a) *Non Linear Poisson residual: damping benefit prodedure.*



(b) t_k *parameter.*

Figure 3.1: (a) Number of iteration against residual for different voltages in a diode test case. (b) Magnitude of the damping parameter t_k.

3.3.3 Continuity equations

Concerning with equation (3.6) we can write the bilinear form as

$$a(u,v) = \int_{\Omega_{Si}} qD_n e^{\varphi^i/V_{th}} \nabla\psi_j \nabla\psi_i \, d\Omega + \int_{\Omega_{Si}} \sigma_n^{i-1} e^{\varphi^i/V_{th}} \psi_j \psi_i \, d\Omega. \tag{3.34}$$

Even if this form allows an immediate analysis of well-posedness, the choice of using Slotboom variables u_n and u_p causes the onset of overflow problems due to the evaluation of $\exp(\varphi/V_{th})$, which can be a rapidly varying function according to the behaviour of the potential φ.

Therefore special care has to be taken in the treatment of the diffusion coefficient. In view of further discussion we introduce some useful notation. For each set $S \subset \Omega$ having measure $|S|$, we introduce the following averages of a given function g that is integrable on S:

$$\mathcal{M}_S(g) = \frac{\int_S g \, dS}{|S|}, \qquad \mathcal{H}_S = (\mathcal{M}_S(g^{-1}))^{-1}. \tag{3.35}$$

Notice that $\mathcal{M}_S$ is the usual integral average, while $\mathcal{H}_S$ is the *harmonic average*. It is well-known that the use of the harmonic average provides a superior approximation performance in one spatial dimension [IE83].

The weak form (3.34) is the result of a standard displacement approach, although different variational formulations and therefore different finite element approximations may be used, like a primal mixed approach (PM). First of all it is convenient to formulate problem (2.35) by using relations (1.38) and (1.40) in a more generic form as:

$$\begin{cases} \nabla \cdot \mathbf{J}_n(n) + \sigma n = f & in \ \ \Omega_{Si} \\ \mathbf{J}_n(n) = qD_n e^{\varphi/V_{th}} \nabla(e^{-\varphi/V_{th}} n) & in \ \ \Omega_{Si} \\ n = n_D & on \ \ \Gamma_{D,Si} \\ \mathbf{J}_n(n) \cdot \mathbf{n} = 0 & on \ \ \Gamma_{N,Si}. \end{cases} \tag{3.36}$$

Problem (3.36) can be discretized using the *Edge Averaged Finite Elements* (EAFE). The complete derivation of this scheme can be found in [XZ99] [ZL12].

The EAFE scheme is particularly suited for problems with a highly variable diffusion coefficient. Furthermore this approach has several good properties, i.e., in a 2D framework if $\mathcal{T}_h$ is a Delaunay partition the system matrix is an M-matrix [BCC98a]. The main consequence of this statement is that the solution satisfies the *Discrete Maximum Principle*. This is a notable property which implies that no negative concentrations are admitted. Unfortunately

this property is not anymore valid in a 3D framework, because the Delaunay condition on the mesh is not enough to guarantee that the system matrix is an M-matrix. A more general condition is presented in [XZ99].

Theorem 3.1 (Zikatanov condition)**.** *The system matrix of the EAFE scheme is an M-matrix if and only if for any fixed edge E of the partition $\mathcal{T}_h$ the following inequality holds*

$$\omega_E = \frac{1}{d(d-1)} \sum_{K \supset E} |k_E^K| cot\theta_E^K \geq 0, \tag{3.37}$$

where $\sum_{K\supset E}$ means summation over all simplexes K containing E, θ_E^K is tha angle between the faces f_i, $f_j \in \mathcal{T}_h$ such that $f_i \bigcap f_j = E$ and k_E^K is the edge in K which does not share any verticies with E.

Observation 3.1. For $d = 2$, condition (3.37) means that the sum of the angles opposite to any edge is less than or equal to π, which implies that the partition is a Delaunay triangulation.

Observation 3.2. Condition (3.37) highlights that in order to satisfy the discrete maximum principle, a partition without obtuse angles is preferable.

We remark that presently meshing algorithms are oriented to care about the minimum angle of the elements, rather than the maximum, this implying that to obtain a mesh which satisfies condition (3.37) is a really difficult task.

Fig.3.2 shows a simple partition of a cube performed with the Synopsis tool SNMESH. For every element we evaluated how many edges do not satisfy condition (3.37). It is clear that there are a lot of edges which do not fulfil the condition and a precise pattern cannot be singled out. When several bad edges belong to a single element we can identify the presence of many obtuse angles.

In order to avoid this problem some alternative solutions are proposed in the literature. In particular conditions the system matrix can be changed to a M-matrix using the *Orthogonal Subdomain Collocation method* [PC98], but also this approach is not a definite solution.

Therefore, in 3D numerical simulation, in presence of a negative concentration the most used technique to recover carrier density positivity is to use local mesh refinement in the regions where trouble occurs, which often are the ones where the carrier density decreases.

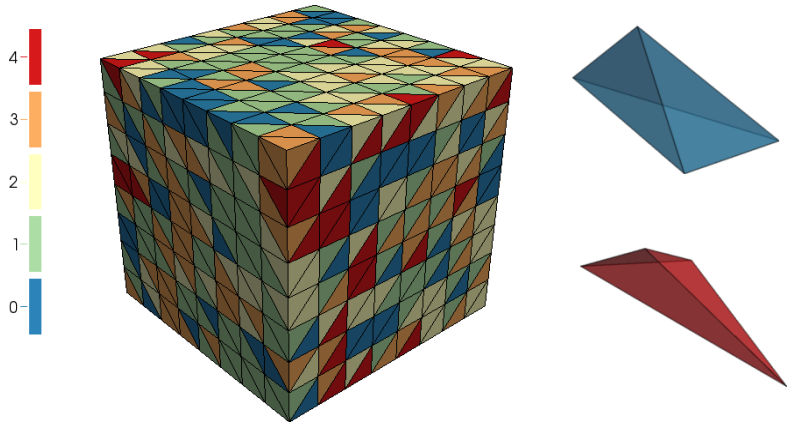


Figure 3.2: Verification of Thm.3.1 over a simple partition. Red elements do not satisfy condition (3.37) over four edges while blue elements fully satisfy the criterion.

Chapter 4

Simulation results

In this chapter we present the work done in order to validate the numerical implementation of the discretization method illustrated in Chapter 3 in the FEMOS 3D code compared with a reference simulation tool (SDEVICE, commercialized by Synopsis [Sde13]). In particular we illustrate (and compare) the algorithm used to calculate the current at the ohmic contacts.

4.1 Test cases

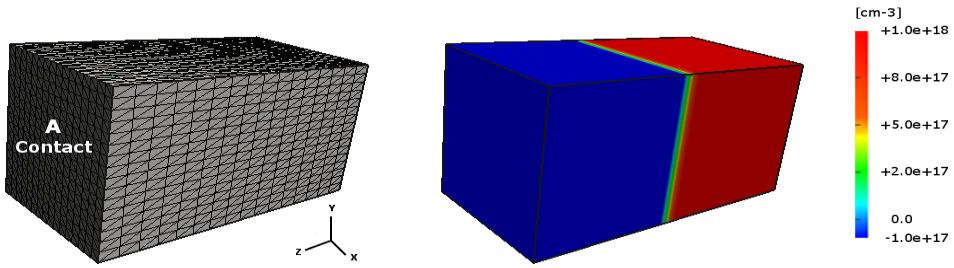
We consider three kinds of semiconductor devices:

- **p-n junction**
- **p-n junction in oxide**
- **n-channel / p-channel MOSFET**

4.1.1 p-n junction

In this example we consider a simple p-n junction. Fig.4.1 presents the partition and the doping profile for this test. The section of the parallelepiped is a $0.05 \times 0.05[\mu m^2]$ square while the device is $0.1[\mu m]$ long. The number of vertices are 4933, while the elements are 24576. The doping concentration is obtained setting a constant profile of acceptors over all the domain ($N_A^- = 1.0 \times 10^{17}[cm^{-3}]$) overwhelmed by a doping profile of donors ($N_D^+ = 1.0 \times 10^{18}[cm^{-3}]$) bounded on one side of the device and resulting in an almost abrupt junction. Two contacts are defined: (A) contact is placed at $Z = 0.1[\mu m]$ and (B) contact is placed at $Z = 0.0[\mu m]$.

In order to analyze the operating function of the diode, two cases of direct bias are performed: $0.3[V]$-$1.0[V]$. The setting values and the parameters



(a) *Mesh.* **(b)** *Doping concentration.*

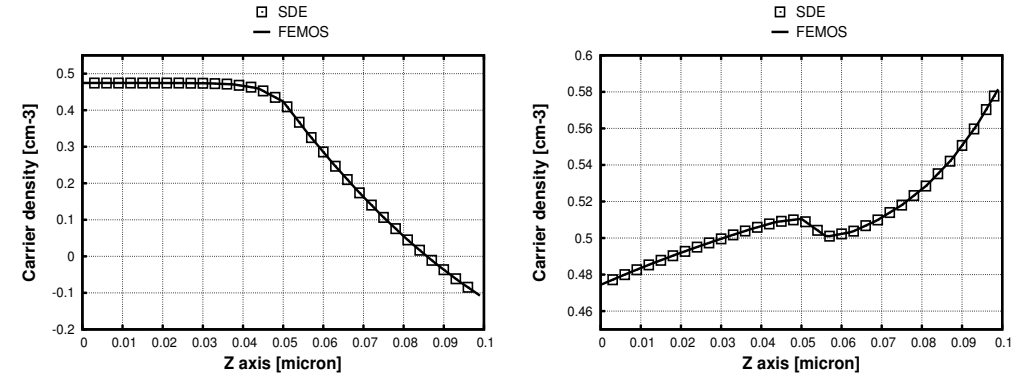
Figure 4.1: p-n junction.

are summarized in Tab.4.1. Fig.4.2 reports the solutions for $V_A = 0.3[V]$ and $V_A = 1.0[V]$, along a line parallel to the Z-axis and placed at the center of the device. Because the built-in voltage is around $0.7 \div 0.8[V]$ the behaviour of the device is different when the applied bias is below or above this threshold. At a bias voltage of $0.3[V]$, potential drop is almost bounded around the junction, and due to the asymmetric doping, is mostly extended in the p-side. Carriers cannot cross the potential barrier and this causes low current flux inside the device. At $1.0[V]$ the minority carrier density becomes almost ten order bigger, resulting in a large amount of current toward the contacts: the device turns from exponential to linear resistive. This is clear in Fig.4.2b where the potential shape becomes similar to a that of a resistance voltage profile (linear potential profile). Comparing the quasi Fermi potentials of Fig.4.2e and Fig.4.2f the boundary layers at contacts increase with the applied bias. This effect is related to the ohmic contact hypothesis, and can be changed by adopting different boundary condition: this occurs also for the carrier concentration because at the contacts, charge neutrality and thermodynamic equilibrium are imposed.

Figs.:4.3-4.5 show the comparison between SDEVICE and FEMOS in 3D plots for electrostatic potential, electron and hole densities at $0.3[V]$, while Figs.:4.6-4.8 show the same comparison at $1.0[V]$. In both conditions the agreement is very good.

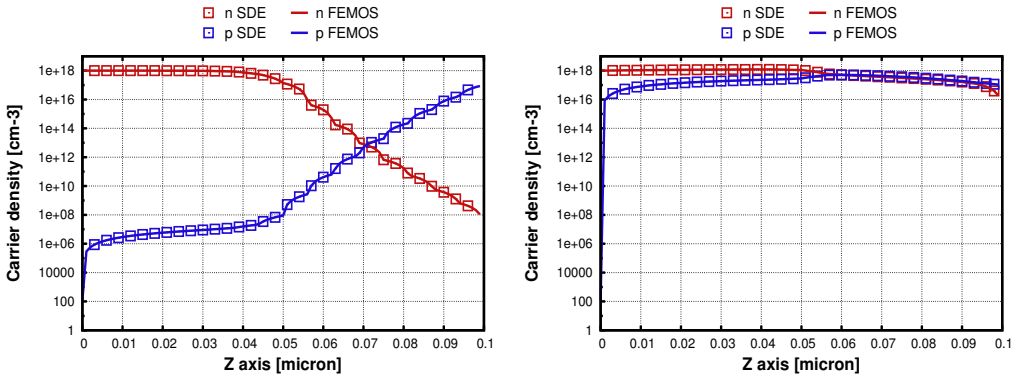
Test case $[V]$	Mobility model $[cm^2V^{-1}s^{-1}]$	R/G model	ϵ_{Si}
$V_A = 0.3$	$\mu_n = 1417$, $\mu_p = 470.5$	SRH, Auger	11.6
$V_A = 1.0$	$\mu_n = 1417$, $\mu_p = 470.5$	SRH, Auger	11.6

Table 4.1: p-n junction - list of settings, parameters and models.



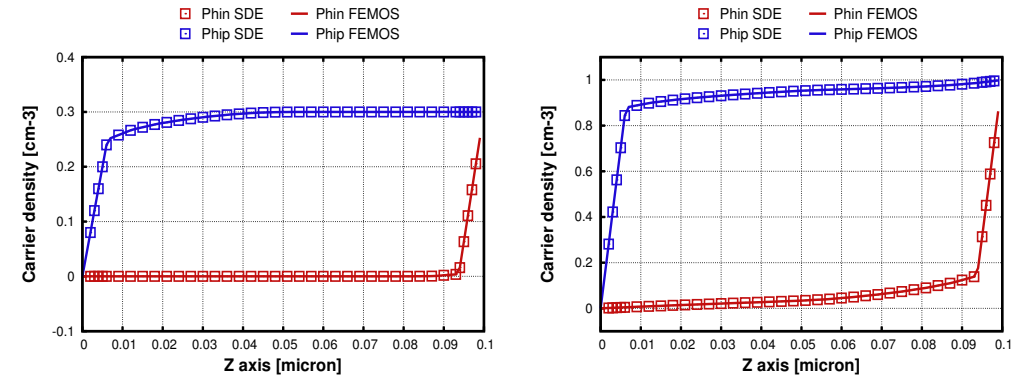
(a) *Potential*

(b) *Potential*



(c) *Carriers*

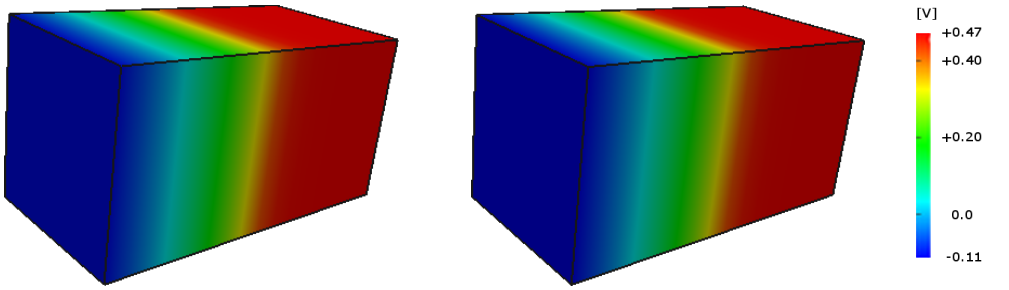
(d) *Carriers*



(e) *Quasi Fermi potentials*

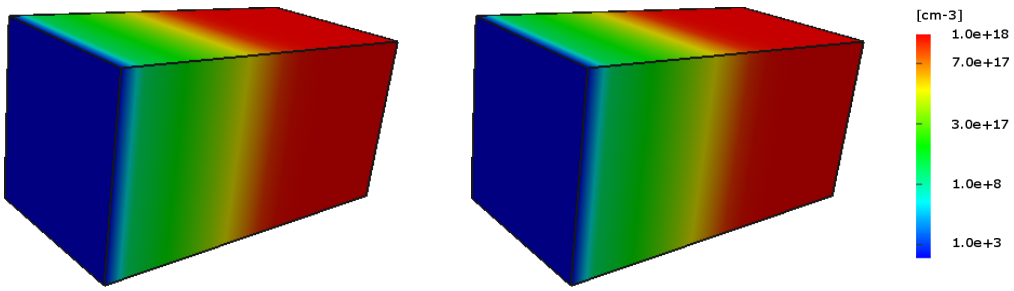
(f) *Quasi Fermi potentials*

Figure 4.2: 1D plots of the solutions and the quasi fermi potential levels along the line parallel to the Z-axis and placed at the center of the device of Fig.4.1. On the left is presented the test case at $V_A = 0.3[V]$ while on the right at $V_A = 1.0[V]$.



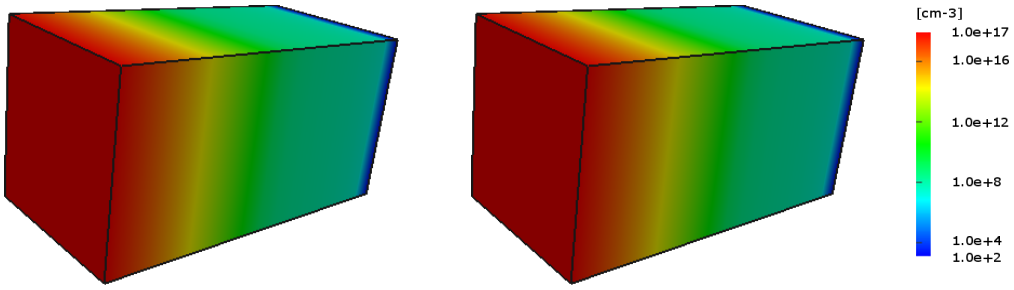
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.3: p-n junction 0.3[V] - Electrostatic Potential.



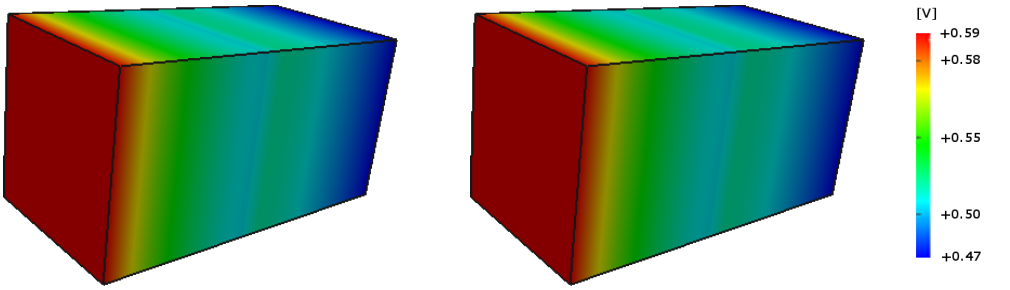
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.4: p-n junction 0.3[V] - Electron density.



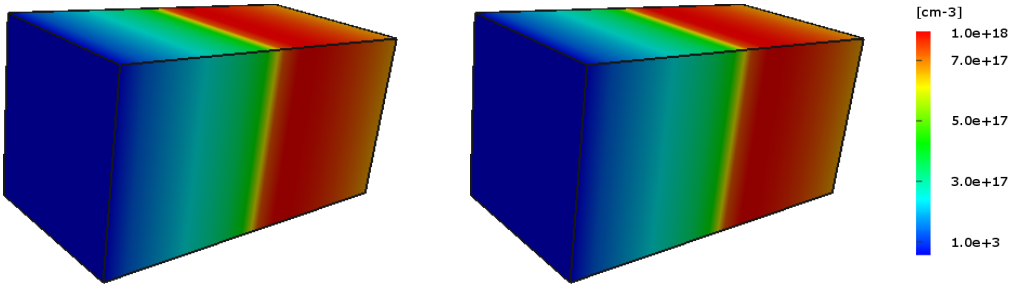
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.5: p-n junction 0.3[V] - Hole density.



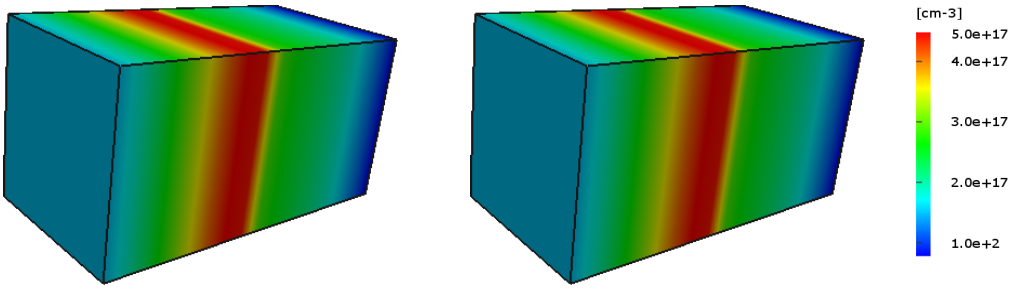
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.6: p-n junction 1.0[V] - Electrostatic Potential.



(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.7: p-n junction 1.0[V] - Electron density.



(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.8: p-n junction 1.0[V] - Hole density.

Computational cost and initial condition

Computational experience demonstrates that the convergence of the Gummel algorithm is strictly related to the kind of chosen initial condition: the closest to solution, the better the convergence. However, to predict in every situation the possible shape of the solutions is hard (if not even impossible). For this reason we have adopted a common and general approach splitting the domain in several regions according to their doping concentration: each of the semiconductor regions are treated as they are in equilibrium with the nearest contact, then the initial guess for φ is obtained using the relations (1.22) or (1.23). This choice corresponds to a case close to equilibrium and guarantees good performance of the algorithm.

In order to analyze the response of the system at different bias an additional test is realized: in the range between $0.0[V]$ and $3.0[V]$ several voltages are applied on the previous device and for each bias point the initial guess is computed as described.

Fig.4.9 shows how the computational cost increases as the applied bias is increased. Moreover as expected if the mesh is finer, the time needed to find the solution increases, resulting in a rigid upper shift of the curve.

Let us consider the case with a coarse mesh. In Fig.4.10 it is clear how the average time spent to solve the NLP and the DD equations remains almost unchanged. On the contrary the number of GM iterations needed by the system to reach the solution, increases for voltages above $\approx 1.5[V]$.

A possible explanation of this trend can be found comparing solution

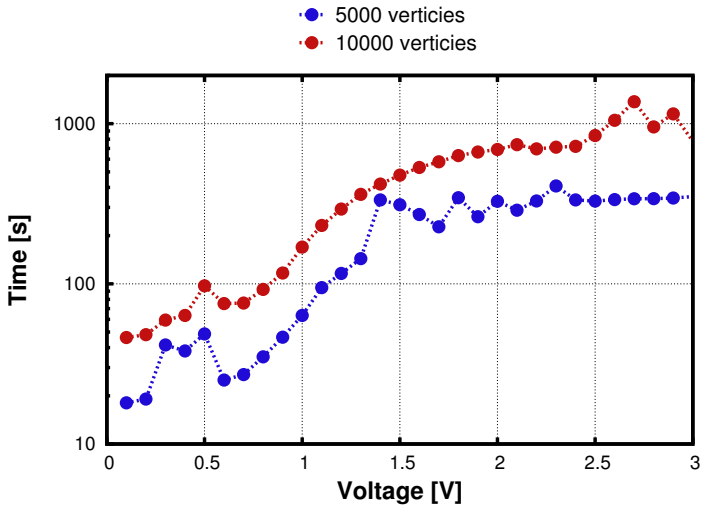


Figure 4.9: Total time Gummel Map.

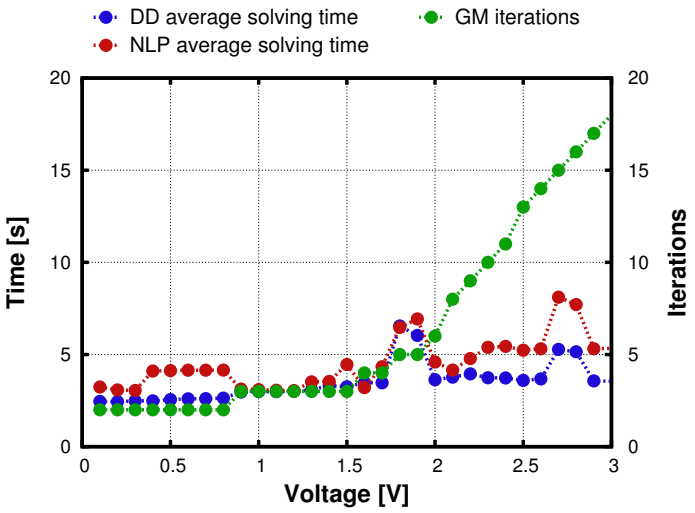


Figure 4.10: Time to solve the NLP and DD equations, and number of iterations of the Gummel map.

and initial guess for a bias below and above $1.5[V]$ (similar considerations can be done for carrier densities). When voltage is low, like in Fig.4.11a ($V_A = 0.1[V]$), the potential shape is well predicted by the initial guess, resulting in a better convergence for the Gummel map algorithm. On the contrary in Fig.4.11b ($V_A = 1.6[V]$) the device operates as a resistance and the potential profile is close to a linear function: this implies that the solution is far from the initial guess of equilibrium condition and the algorithm needs more steps.

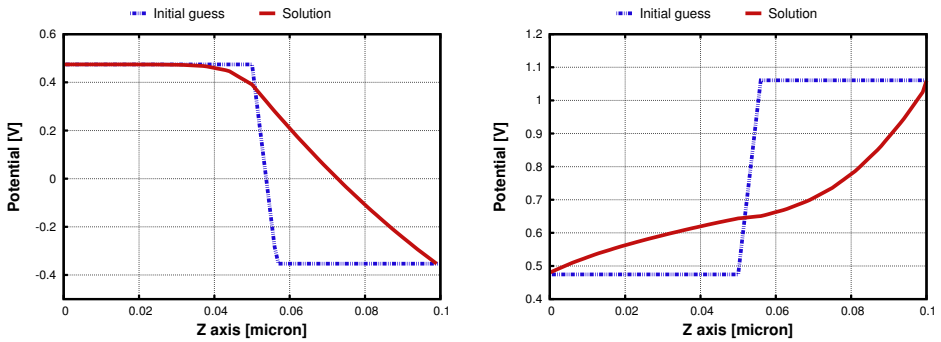
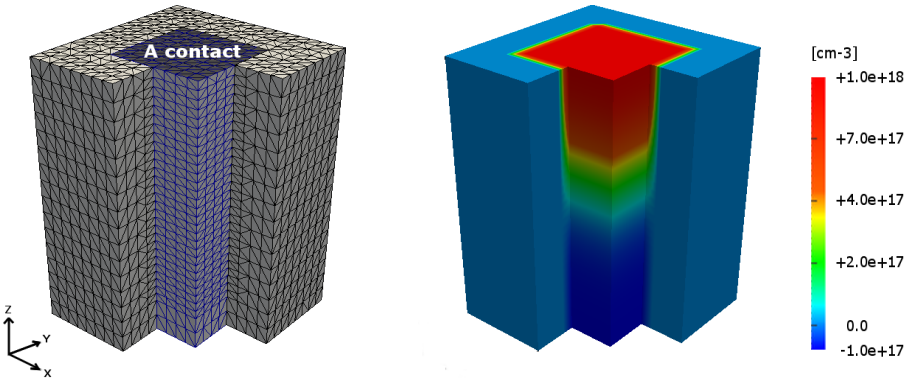


Figure 4.11: Initial guess for different bias compared with the final solution of the device of Fig.4.1.

4.1.2 p-n junction in oxide

In this test case a silicon p-n junction $0.3[\mu m]$ long is surrounded by an oxide layer $0.025[\mu m]$ thick. The section of the silicon part is a $0.1 \times 0.1[\mu m^2]$ square. We employ 6334 vertices and 33121 elements overall the domain. The structure and the doping are shown in Fig.4.12. The setting of the electrodes is similar to the previous test case and contacts are defined only on silicon surface. Tab.4.2 reports settings, models and parameters used in the simulations.

Fig.4.13 shows the solutions and the quasi Fermi potential levels along a line parallel to the Z-axis and placed at the center of the device. The main features are similar to the previous test case, also for the boundary layers at contact for carriers and quasi Fermi potentials. Figs.:4.14-4.16 show the 3D solutions for the test at $0.3[V]$, while Fig.4.17-4.19 refer to the case at $1.0[V]$. Both the 1D cuts and 3D plots agreement with the commercial software are very good.

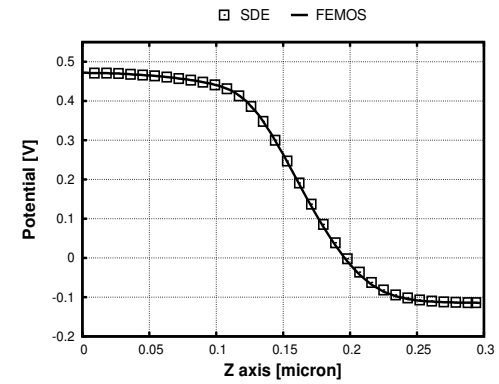


(a) *Mesh* **(b)** *Doping concentration*

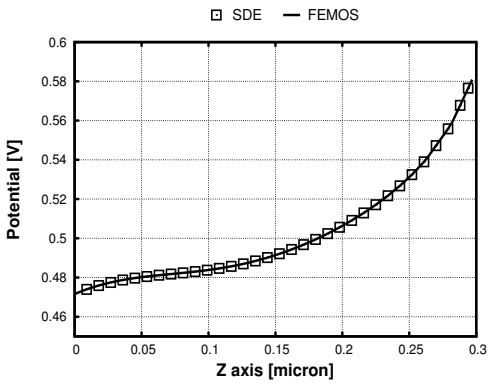
Figure 4.12: Test case p-n junction in oxide.

Test case $[V]$	Mobility model $[cm^2V^{-1}s^{-1}]$	R/G model	ϵ_{Si}	ϵ_{0x}
$V_A = 0.3$	$\mu_n = 1417$, $\mu_p = 470.5$	SRH, Auger	11.6	3.9
$V_A = 1.0$	$\mu_n = 1417$, $\mu_p = 470.5$	SRH, Auger	11.6	3.9

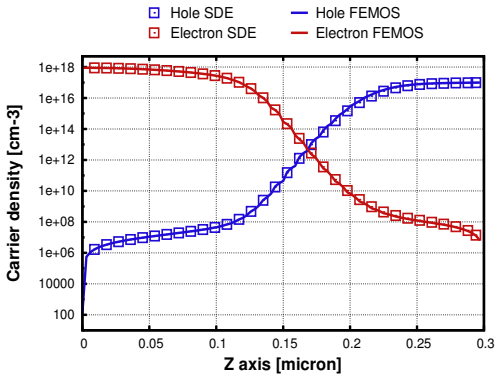
Table 4.2: p-n junction in oxide - list of settings, parameters and models.



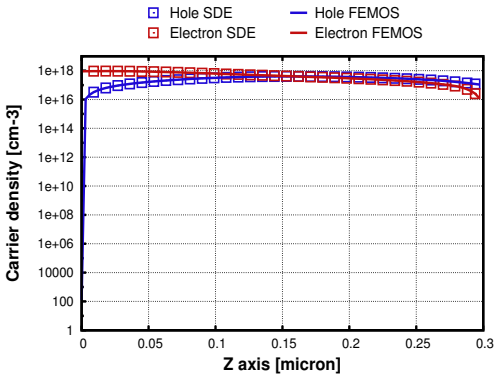
(a) *Electrostatic potential.*



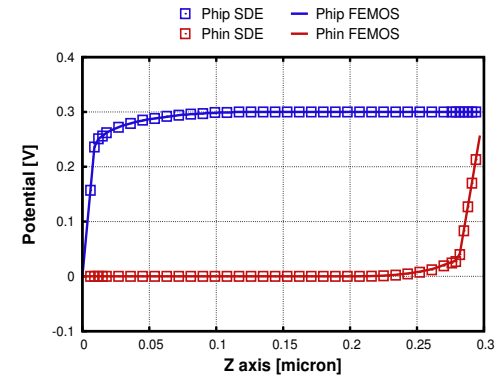
(b) *Electrostatic potential.*



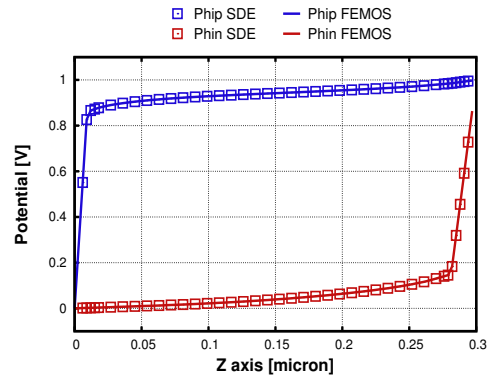
(c) *Hole and electron densities.*



(d) *Hole and electron densities.*

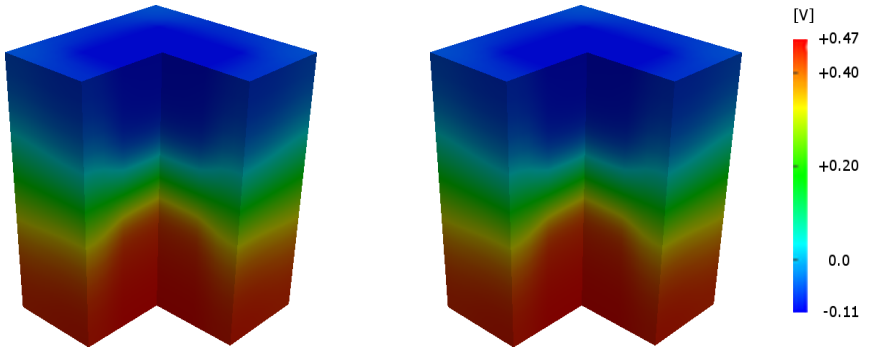


(e) *Quasi Fermi potential levels.*



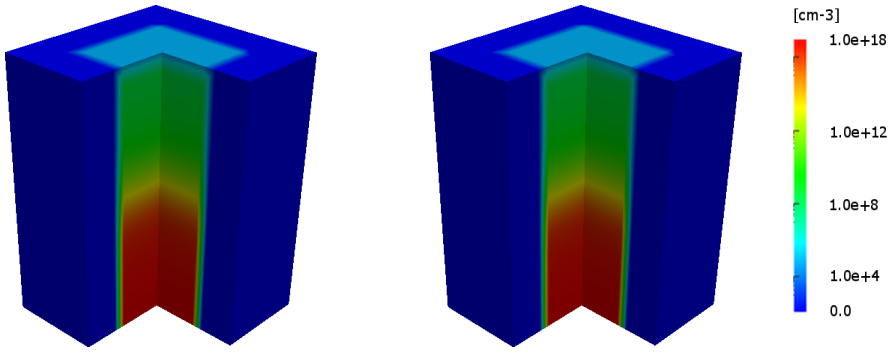
(f) *Quasi Fermi potential levels.*

Figure 4.13: 1D plots of the solutions and the quasi Fermi potential levels along the line parallel to the Z-axis and placed at the center of the device of Fig.4.12. On the left test case at $V_A = 0.3[V]$ is reported while on the right at $V_A = 1.0[V]$.



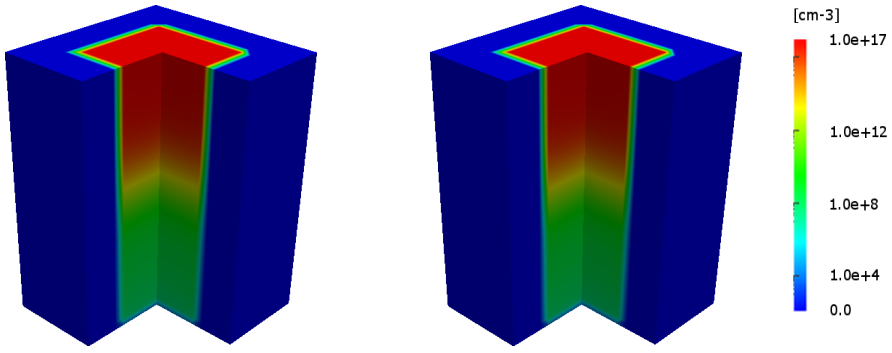
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.14: p-n junction in oxide 0.3[V] - ElectrostaticPotential.



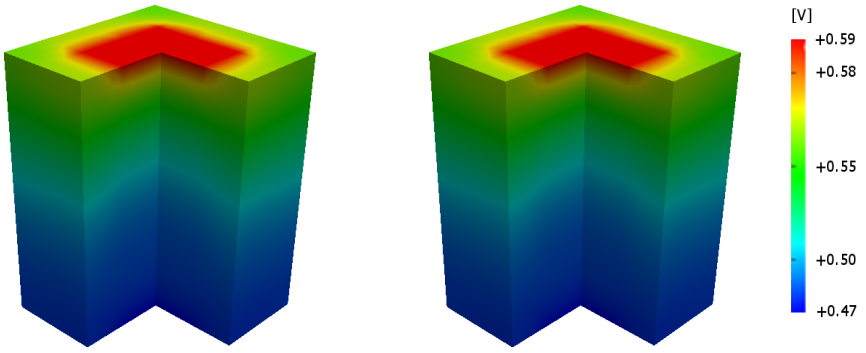
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.15: p-n junction in oxide 0.3[V] - Electron density.



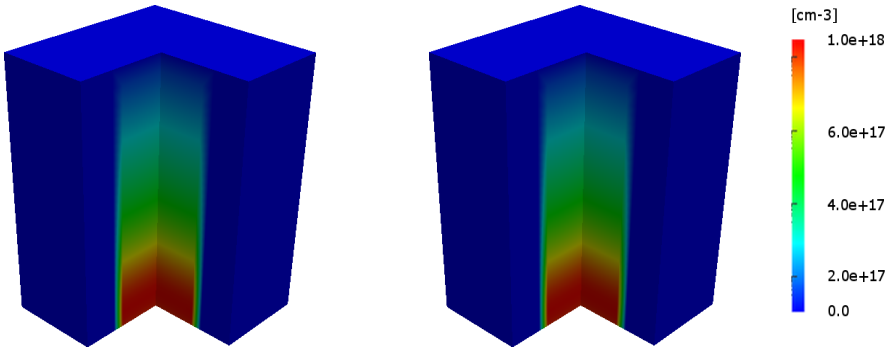
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.16: p-n junction in oxide 0.3[V] - Hole density.



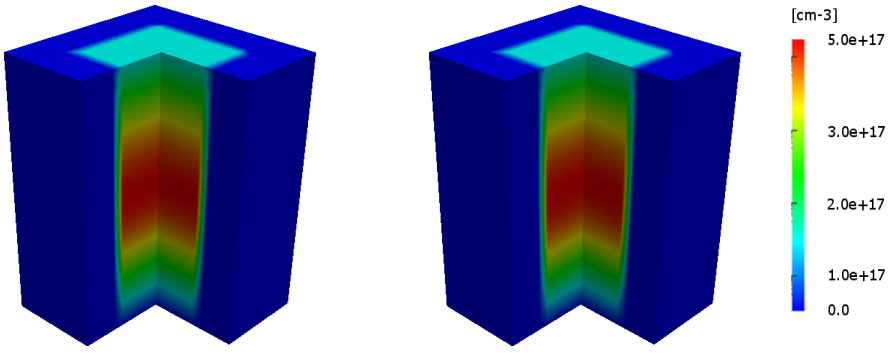
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.17: p-n junction in oxide 1.0[V] - Electrostatic Potential.



(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.18: p-n junction in oxide 1.0[V] - Electron density.



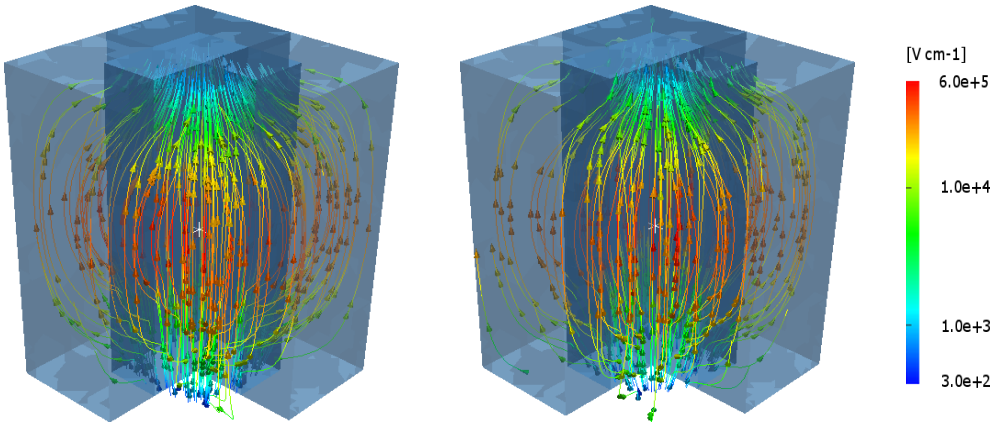
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.19: p-n junction in oxide 1.0[V] - Hole density.

3D effect on the electric field

Fig.4.14 shows how, at low bias, the electrostatic potential behaves in a different manner in the oxide with the respect to the silicon region. Since we impose $\nabla\varphi \cdot \mathbf{n} = 0$ on the oxide boundary no field lines of the electric field can cross that boundary and, as a consequence, field lines can start and end only at contact A or B. The electric field inside the device is due only to the displacement effect in the junction of the silicon which imposes the electric response also in the oxide material where the solution is no more linear. Fig.4.20 reports the electric field lines in the case of $V_A = 0.3[V]$ compared with the commercial software: agreement is very good. The lower magnitude of the electric field in the oxide, results in a more diffused potential.

At high bias (Fig.4.17) the influence of the contacts (A,B) becomes higher and the electrostatic potential is much more similiar in the two subdomains.



(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.20: Test case dide p-n in oxide 0.3[V] - Electric field.

It is important to notice that the displacement formulation approach does not satisfy in a strong manner the action reaction principle at inter-element boundaries. This is equivalent to saying that given two elements $K_i, K_j \in \mathcal{T}_h$ such that $K_i \bigcap K_j = f_i$ where $f_i \in \mathcal{F}_h$ and denoting by $\mathbf{n}_i$ the outward normal vector of ∂K_i and $\mathbf{n}_j$ the outward normal vector of ∂K_j, the following equations are satisfied only in a weak sense:

$$\mathbf{D}|_{K_i} \cdot \mathbf{n}_i = \mathbf{D}|_{K_j} \cdot \mathbf{n}_j \tag{4.1}$$
$$\mathbf{J}_n|_{K_i} \cdot \mathbf{n}_i = \mathbf{J}_n|_{K_j} \cdot \mathbf{n}_j \tag{4.2}$$
$$\mathbf{J}_p|_{K_i} \cdot \mathbf{n}_i = \mathbf{J}_p|_{K_j} \cdot \mathbf{n}_j. \tag{4.3}$$

In fact, taking in consideration equation (4.1), we can state that along the interface silicon-oxide the following equality holds

$$\epsilon_{ox}\mathbf{E}|_{K_i} \cdot \mathbf{n}_i = \epsilon_{Si}\mathbf{E}|_{K_j} \cdot \mathbf{n}_j \quad \Rightarrow \quad [\mathbf{E}|_{K_i}]_y = \frac{\epsilon_{Si}}{\epsilon_{ox}}[\mathbf{E}|_{K_j}]_y. \tag{4.4}$$

According with the parameters used in the simulations the jump of the electric field component from oxide to semiconductor is around 3.00. Fig.4.21 shows the plots of E_y along a line parallel to the Y-axis crossing both oxide and silicon: the ratio expressed by (4.4) is almost 3.6 at $y = 0.05[\mu m]$ and $y = 0.15[\mu m]$ where the interfaces are located.

Each tetrahedral interface of the partition is affected by the same problem, which means that the normal component of the electric field from one element of the grid to the neighbouring is not conserved even if the material is homogeneous and there is no charge at the interface. Despite this drawback, the solutions are acceptable. But if we would like to satisfy equation (4.1) in a strong manner, the use of a mixed-hybrid formulation is needed to ensure the conservation of the flux also under possible strong discontinuities of material properties.

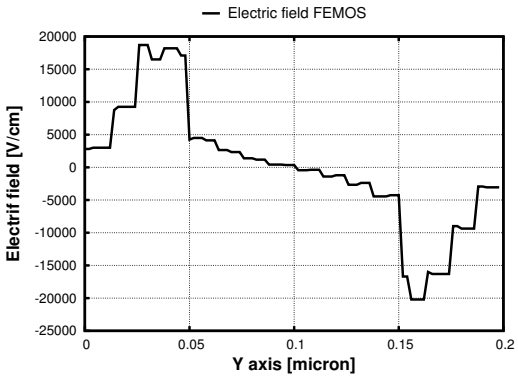


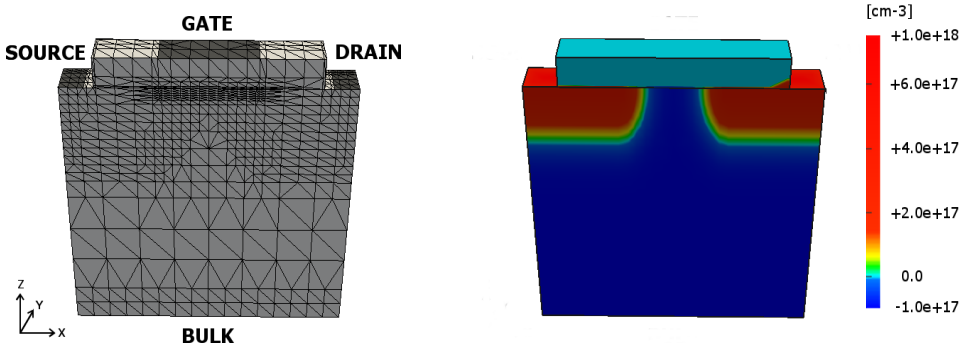
Figure 4.21: E_y along a line parallel to Y-axis, $z = 0.22[\mu m]$ and $x = 0.1[\mu m]$.

4.1.3 n-channel MOSFET

The description of the working principle of a MOSFET device can be found in [YT09]: it is a four-terminal device with the electrodes designated as gate (G), source (S), drain (D) and substrate or bulk (B). The gate electrode is usually made of metal or heavily doped polysilicon and is separated from the substrate by a thin silicon dioxide. The surface region under the gate oxide between source and drain is called *channel* region. Because the current in a MOSFET is due to carriers of one polarity, the MOSFET is usually referred as a unipolar or majority-carrier device: n-channel (n-MOSFET) and p-channel (p-MOSFET) are considered in the test. A *n-MOSFET* (p-MOSFET) consists of a p-type (n-type) silicon substrate into which two n-regions (p-regions) are designed as source and drain. The n-regions (p-regions) are doped according to a Gaussian profile as in a real implantation process. Fig.4.22a and Fig.4.22b show the geometry and the doping concentration for the n-MOSFET with a coarse mesh (2739 vertices and 12338 elements). We note that the mesh has been refined where the more interesting phenomena occur, i.e., along channel and at drain/source contacts.

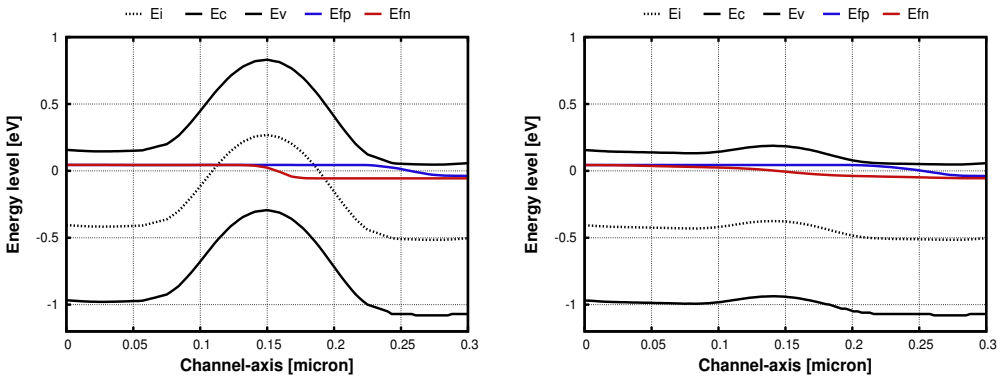
If a sufficiently large positive voltage is applied to the gate, the silicon surface is inverted to n-type (p-type), which forms a conducting channel between the source and drain: applying a small positive voltage to the drain (source) the electrons (holes) start to flow from source (drain) to drain (source) and therefore a current is generated.

The visualization of the MOSFET working principle is better clarified in Fig.4.23 where the band profile along the channel axis is reported for two different gate bias. The voltage applied to the gate tends to decrease the energy barrier in the channel region: a little drain voltage causes the flow of the electrons. Fig.4.24a shows the profile of carrier concentrations in the



(a) *Mesh* **(b)** *Doping concentration*

Figure 4.22: Geometry of a n-channel MOSFET.



(a) $V_G = 0.0[V]$, $V_D = 0.1[V]$ **(b)** $V_G = 2.0[V]$, $V_D = 0.1[V]$

Figure 4.23: Energy band levels for a n-MOSFET along the channel.

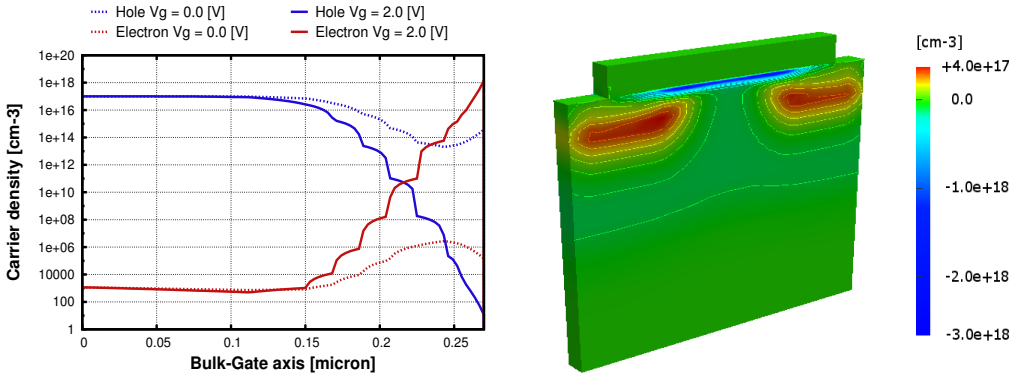
middle of the channel with a cut perpendicular to the gate, for off-state and on-state: the inversion occurs when the gate voltage is higher than the well known threshold voltage. Fig.4.24b shows the 3D view of the n-channel space charge after inversion.

Finally Fig.4.25 reports the streamline plot of the electric field inside the device for FEMOS and the SDEVICE in the case of on-state MOSFET.

Settings, parameters and models used for these simulations are summarized in Tab.4.3. Figs.:4.26-4.28 show the 3D view of the electrostatic potential, electron and hole densities obtained by FEMOS and by the commercial code in the off-state ($V_G = 0.0[V]$), while Figs.:4.29-4.31 refer to the on-state ($V_G = 2.0[V]$): the agreement is very good.

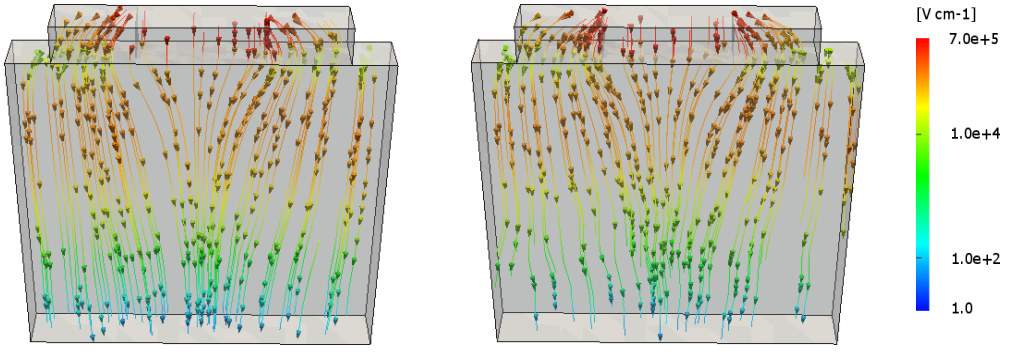
Test case $[V]$	Mobility model $[cm^2V{-}1s^{-1}]$	R/G model	ϵ_{Si}	ϵ_{0x}
$V_G = 0.0[V]$, $V_D = 0.0[V]$, $V_S = V_B = 0.0[V]$	$\mu_n = 1417$ $\mu_p = 470.5$	SRH, Auger, II	11.6	3.9
$V_G = 2.0[V]$, $V_D = 0.1[V]$, $V_S = V_B = 0.0[V]$	$\mu_n = 1417$ $\mu_p = 470.5$	SRH, Auger	11.6	3.9

Table 4.3: n-MOSFET - list of settings, parameters and models.



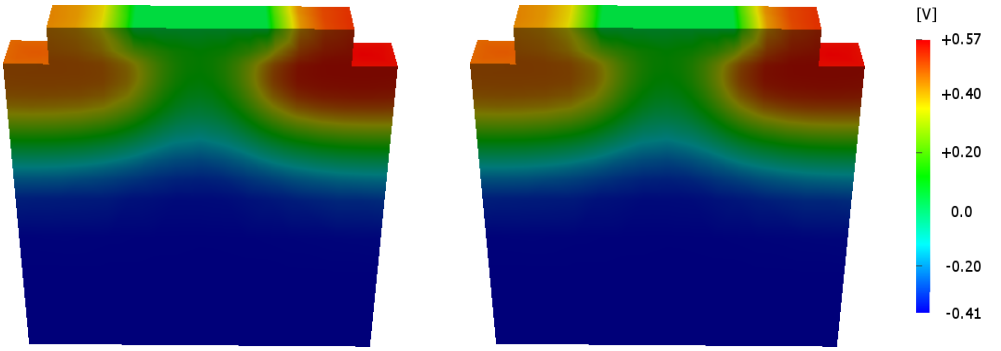
(a) 1D cut perpendicular to the channel. **(b)** Space charge.

Figure 4.24: Channel of the n-MOSFET.



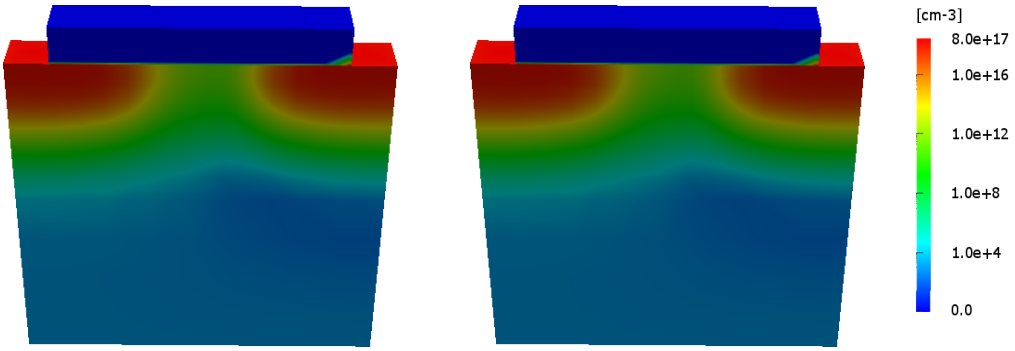
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.25: n-MOSFET $V_G = 2.0[V]$ - Electric field.



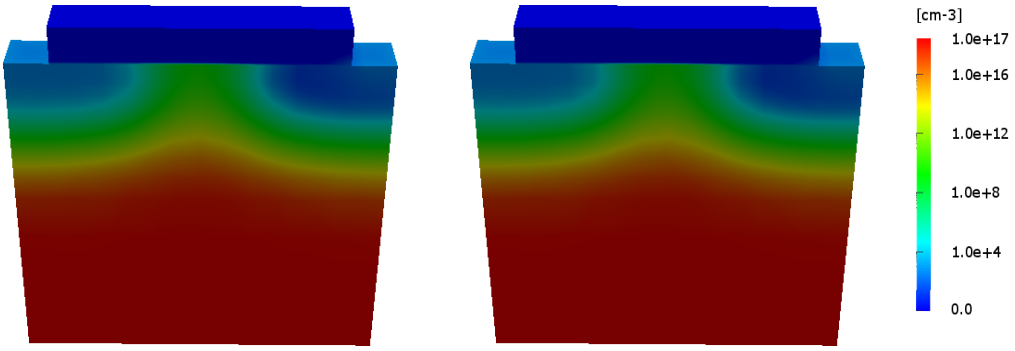
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.26: n-MOSFET $V_G = 0.0[V]$ - Electrostatic potential.



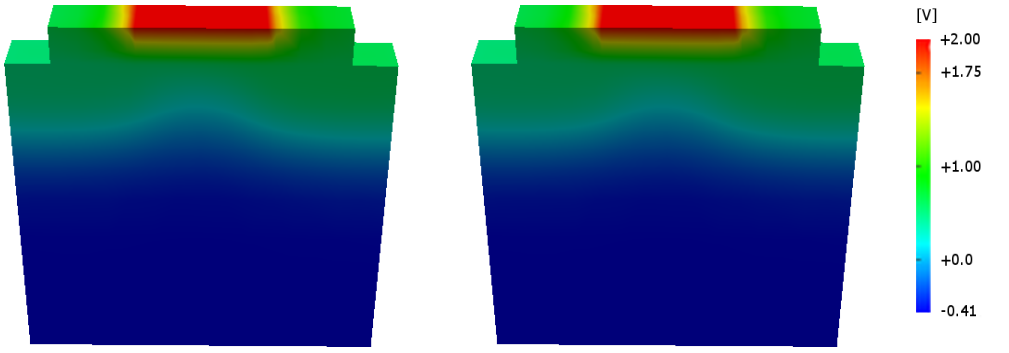
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.27: n-MOSFET $V_G = 0.0[V]$ - Electron density.



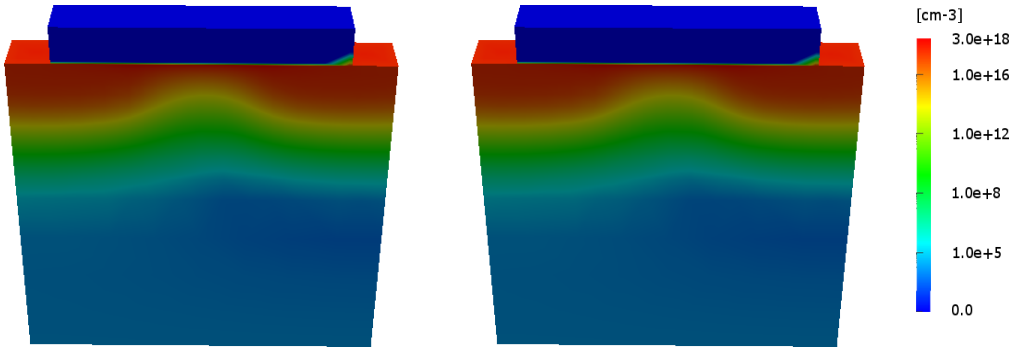
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.28: n-MOSFET $V_G = 0.0[V]$ - Hole density.



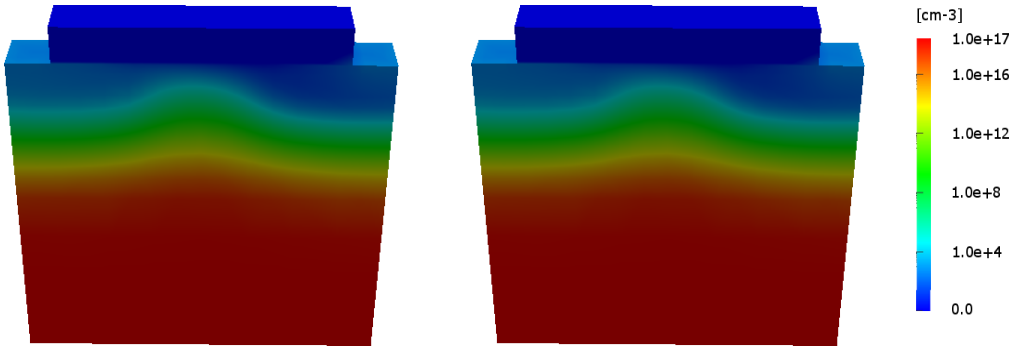
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.29: n-MOSFET $V_G = 2.0[V]$ - Electrostatic potential.



(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.30: n-MOSFET $V_G = 2.0[V]$ - Electron density.



(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.31: n-MOSFET $V_G = 2.0[V]$ - Hole density.

Reverse bias

In Section 3.3.3 we pointed out that the discretization scheme (EAFE) cannot satisfy the discrete maximum principle in 3D simulations unless we satisfy condition (3.37). Therefore it is possible to encounter situations where negative solutions are obtained and this usually happens when the concentration of electrons and holes become low.

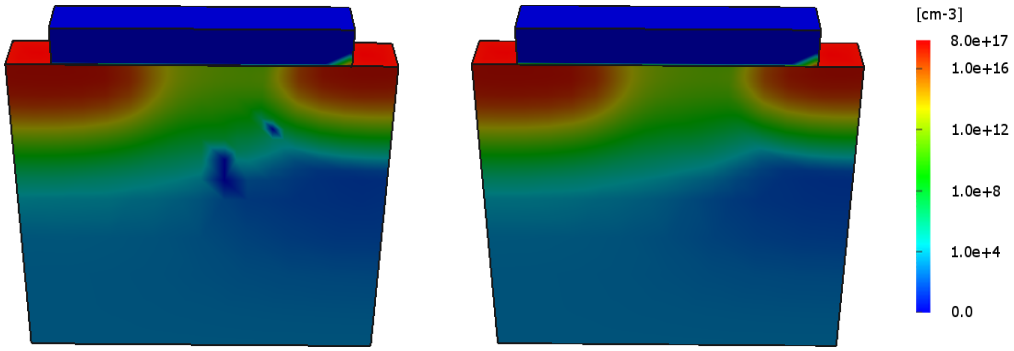
In order to highlight this possible critical situation, a n-channel MOSFET is simulated in reverse bias regime by grounding all the contacts except the drain which is ramped to $0.5[V]$: Tab.4.4 reports the needed indications for the simulation.

Fig.4.32 reports the electron density computed with FEMOS and SDEVICE using the mesh presented in Fig.4.22b: the results are comparable, but near the drain-bulk junction FEMOS presents some points with negative concentrations. Increasing drain bias the phenomenon tends to spread over a larger area, until it affects irremediably the simulation of the device. As we anticipated in Chapter 3, the most practice technique to limit this problem is local mesh refinement. Fig.4.33a represents a finer mesh with 13000 points and 67388 elements. Using this mesh the correctness of the solution is recovered Fig.4.33b. Fig.4.34 shows how the satisfaction of (3.37) changes between the different meshes: increasing the number of vertices over the critical region a better fulfillment of (3.37) is guaranteed. Results suggest that in order to treat this situation it may be useful to implement a suitable a-posteriori error estimation and adaptive mesh refinement techniques.

Finally Fig.4.35 and Fig.4.36 report FEMOS electrostatic potential and hole density for the finer mesh compared with the commercial tool: the agreement is very good.

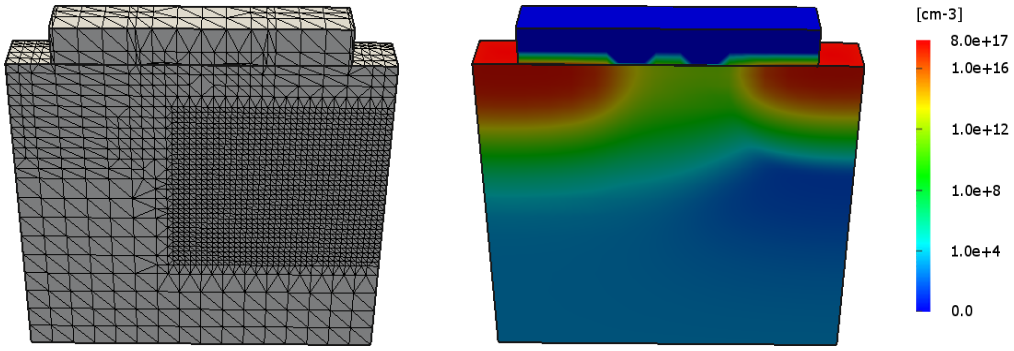
Test case $[V]$	Mobility model $[cm^2V{-}1s^{-1}]$	R/G model	ϵ_{Si}	ϵ_{0x}
$V_G = 0.0$, $V_D = 0.5$, $V_S = V_B = 0.0$	$\mu_n = 1417$ $\mu_p = 470.5$	SRH, Auger, II	11.6	3.9

Table 4.4: n-MOSFET (reverse bias) - list of settings, parameters and models.



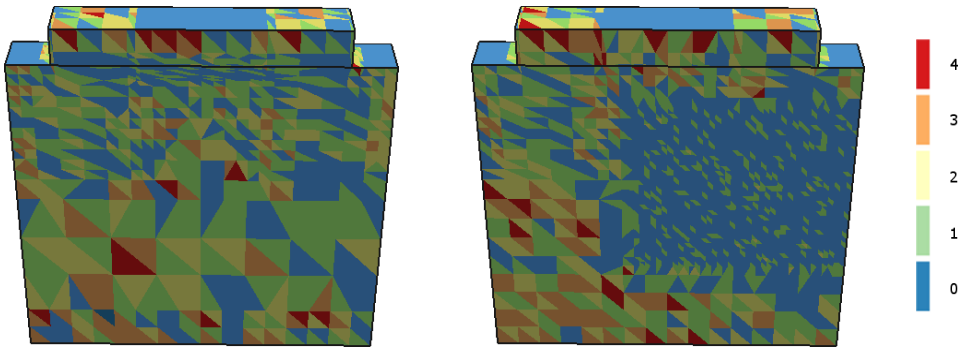
(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.32: n-MOSFET reverse bias: negative carriers spots for the electron density solution.



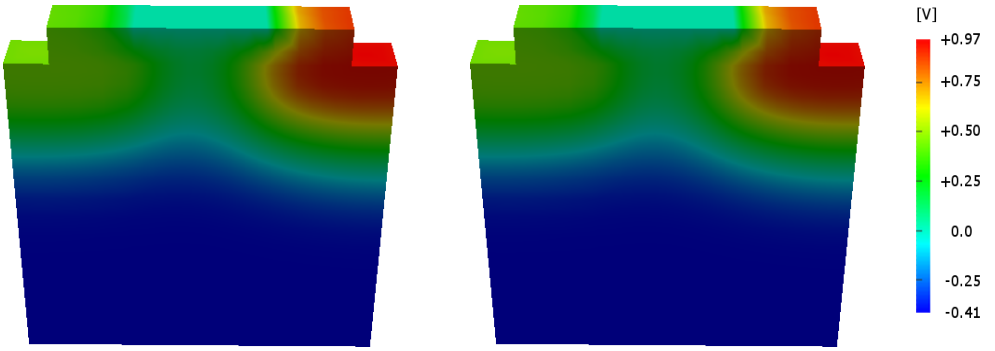
(a) *Mesh* **(b)** *FEMOS*

Figure 4.33: n-MOSFET reverse bias: electron density with finer mesh.



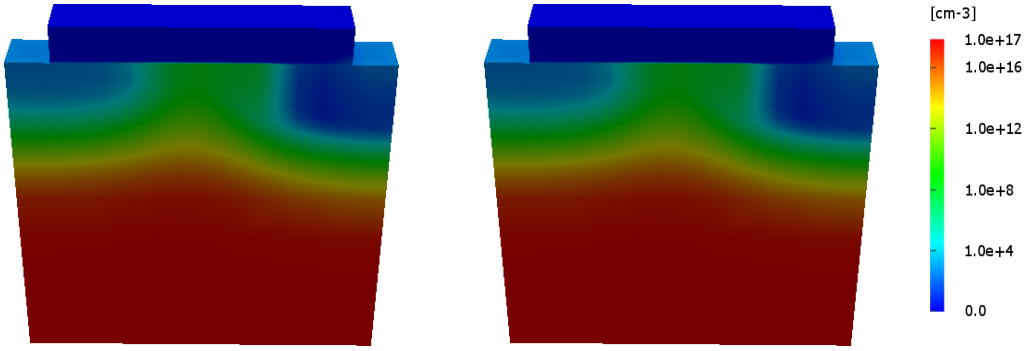
(a) *Coarse mesh.* **(b)** *Fine mesh.*

Figure 4.34: n-MOSFET: verification of (3.37) condition.



(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.35: n-MOSFET reverse bias - Electrostatic potential.



(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.36: n-MOSFET reverse bias - Hole density.

4.2 Current evaluation at the Ohmic contacts

During the analysis of an electric device, one of the most important information is the electrical response at terminals. In order to accomplish this target we have to compute the integral of the electron and hole current density over a generic 2D electrode. We refer here to the procedure found in [GS06] (*residual method*) for the 2D case: the analysis is easily extendable to the 3D case if we consider also [HEML00]. Moreover we remark that the method can be successfully applied to a wide spread of applications, including contact charges, carrier quantum probability fluxes and heat fluxes.

A contact is defined by a surface: we can consider $\Gamma_{D,Si} = \bigcup_{c=1}^{d} \Gamma_c$ where d is the number of terminals on the device and $\forall c = 1, ..., d$, Γ_c is the c-th contact. For each contact we need to compute the total current $\mathcal{I}_c$ as

$$\mathcal{I}_c = \mathcal{I}_c^n + \mathcal{I}_c^p \tag{4.5}$$

where $\mathcal{I}_c^n$ and $\mathcal{I}_c^p$ are the contribution of the electron and hole current. For a given contact Γ_c, the flux of the current density is defined as

$$\mathcal{I}_c^\nu = \int_{\Gamma_c} \mathbf{J}_\nu(\nu) \cdot \mathbf{n}\, d\Gamma \qquad \nu = \{n, p\} \tag{4.6}$$

where $\mathbf{n}$ is the unit outward normal of the domain boundary. It is well known that the evaluation of boundary integrals is a difficult task. Most problems in (4.6) arise from singularities in spatial derivatives of the approximate solution n_h or p_h near the contact edges, due to a change in the boundary condition type from Dirichlet to Neumann.

Let η be the set of all vertices of the partition $\mathcal{T}_h$ for the discretized electron continuity problem (3.34). We can split the set of total nodes in contact nodes $\eta_g \in \Gamma_{D,Si}$ and the complementary part $\eta_n \in \Gamma_{N,Si}$. We define an *auxiliary flux* H_h on $\Gamma_{D,Si}$ as

$$H_h = \sum_{i \in \eta_g} H_{h,i} \psi_i. \tag{4.7}$$

Now, given the spaces:

$$\begin{aligned} \mathcal{V}_h &= \text{span}\{\psi_i\}_{i \in \eta_n} \\ \mathbf{V}_h &= \text{span}\{\psi_i\}_{i \in \eta_g} \\ \mathcal{S}_h &= \{u \in \mathcal{V}_h \oplus \mathbf{V}_h : u|_{\Gamma_{D,Si}} = n_D\} \end{aligned}$$

it is possible to write a modified form of Galerkin method which reads as:

find $n_h \in \mathcal{S}_h$ and $H_h \in \mathbf{V}_h$ such that

$$(W_h, H_h)_{\Gamma_{D,Si}} = a(W_h, n_h) - F(W_h) \qquad \forall W_h \in \mathcal{V}_h \oplus \mathbf{V}_h \tag{4.8}$$

where $a(\cdot,\cdot)$ is the bilinear form (3.34) and $F(\cdot)$ the associated linear functional. Equation (4.8) splits into two subproblems:

$$0 = a(w_h, n_h) - F(w_h) \qquad \forall w_h \in \mathcal{V}_h \tag{4.9}$$

$$(W_h, H_h)_{\Gamma_{D,Si}} = a(W_h, n_h) - F(W_h) \qquad \forall W_h \in \mathbf{V}_h. \tag{4.10}$$

Problem (4.9) is identical to the unmodified case and can be treated as before or using a different discretization scheme. Once obtained the solution n_h, problem (4.10) is fully decoupled from (4.9) and we can determine H_h by solving the following equation

$$(H_h, \psi_i)_{\Gamma_{D,Si}} = a(\psi_i, n_h) - F(\psi_i) \qquad \forall i \in \eta_g. \tag{4.11}$$

In [HEML00], it is shown that H_h defines the conserved total flux along $\Gamma_{D,Si}$ and the following equality is obtained starting from (4.11)

$$\int_{\Gamma_{D,Si}} H_h \, d\Gamma = -\int_{\Omega_{Si}} qR \, d\Omega. \tag{4.12}$$

On the other hand if we apply the divergence theorem to (2.31) we get

$$\int_{\Gamma_{D,Si}} \mathbf{J}_n \cdot \mathbf{n} \, d\Gamma = \int_{\Omega} -qR \, d\Omega. \tag{4.13}$$

Equations (4.12) and (4.13) lead us to conclude that for all contacts it holds

$$\mathcal{I}_c^n = \int_{\Gamma_c} H_h \, d\Gamma. \tag{4.14}$$

In order to compute (4.14), let η_c be the set of nodes of the contact Γ_c. Then, the following equalities hold

$$\sum_{l \in \eta_c} \int_{\Gamma_c} H_h \psi_l \, d\Gamma = \int_{\Gamma_c} H_h \sum_{l \in \eta_c} \psi_l \, d\Gamma = \int_{\Gamma_c} H_h \, d\Gamma. \tag{4.15}$$

According to (4.15) we can interpret $(H_h, \psi_i)_{\Gamma_{D,Si}}$ as the contribution to the flux at node i and therefore the current at contact c is given by summing this quantity over the vertices η_c.

The residual method is thus defined as: given the system matrix A of the Drift-Diffusion equation, the solution n_h and the right hand side $\mathbf{b}$, the contribution to the total contact current $\forall c = 1, ..., d$ is

$$\mathcal{I}_c^n = (An_h - \mathbf{b}) \cdot \mathbb{I}_c \tag{4.16}$$

where

$$[\mathbb{I}_c]_i := \begin{cases} 0 & i \notin \eta_c \\ 1 & i \in \eta_c. \end{cases} \tag{4.17}$$

The above results hold also for the hole continuity equation.

4.2.1 Simulation results

The residual method is applied to the already analysed devices and compared with the SDEVICE results. In this section, different mobility and recombination/generation models are tested and verified along with the solution of the commercial code.

p-n junction

Considering the p-n junction of Section 4.1.1 we ground the B contact and then A contact is ramped from $-7.5[V]$ to $2.0[V]$ in order to obtain the well known diode characteristic. Tab.4.5 reports the parameters used in the simulation. In Fig.4.37 we plot the electron and hole current at contact A. Diode breakdown voltage is appearing around $-7.0[V]$ and it is quite well aligned with SDEVICE.

Test case $[V]$	Mobility model $[cm^2V^{-1}s^{-1}]$	R/G model	ϵ_{Si}
$V_A = -7.5 \div 1.5$, $V_B = 0.0$	$\mu_n = 1417$ $\mu_p = 470.5$	SRH II, Auger	11.6

Table 4.5: p-n junction (characteristic) - list of settings, parameters and models.

Fig.4.38 shows the different behaviour of the SRH and Auger R/G contribution at two voltages (plots are made along a line parallel to the Z-axis and placed at the center of the device). Under the built-in condition, both SRH and Auger mechanisms are not significant. Auger contribution decreases in

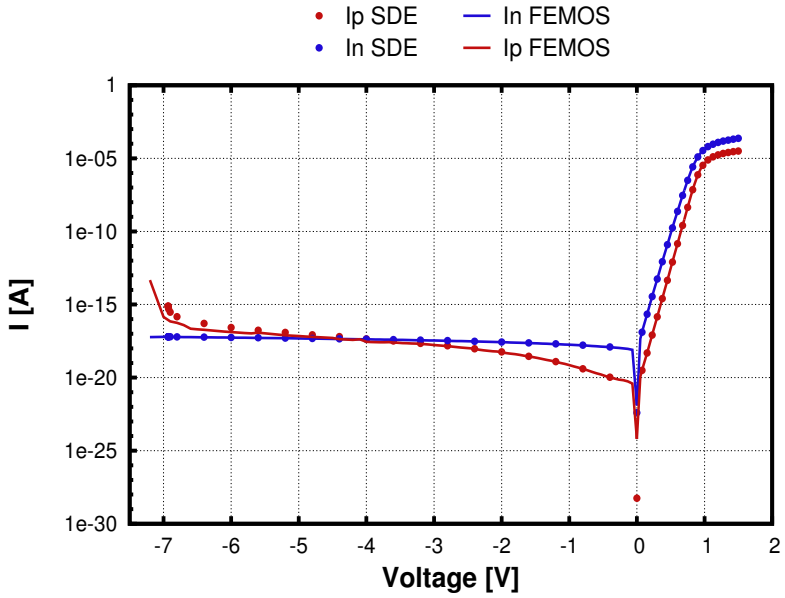
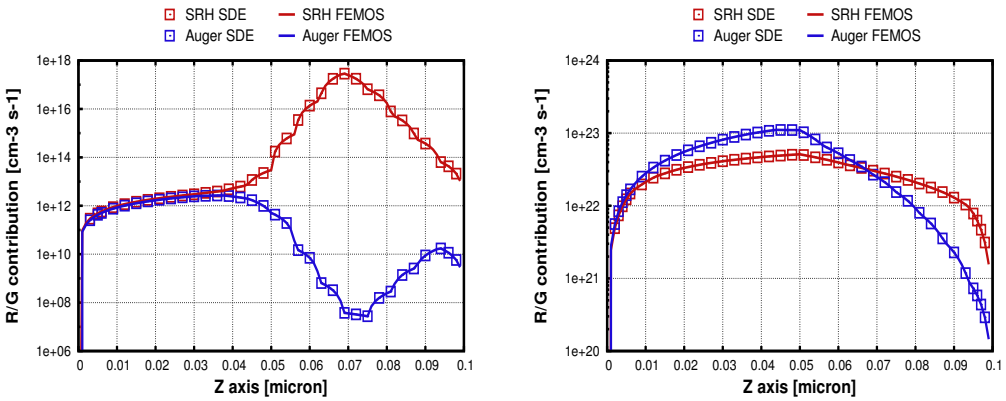


Figure 4.37: p-n junction characteristic.

the depletion region due to its strict dependence on the carrier concentrations. At the built-in condition the R/G phenomenon quite increase and are distributed over the entire device causing its saturation. The agreement with the commercial tool is very good.



(a) $V_A = 0.3[V]$. **(b)** $V_A = 1.0[V]$.

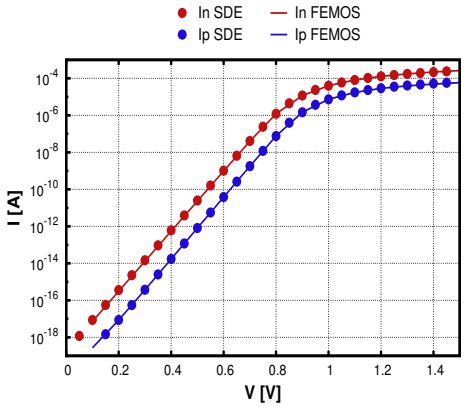
Figure 4.38: p-n junction SRH and Auger RG contribution.

p-n junction in oxide

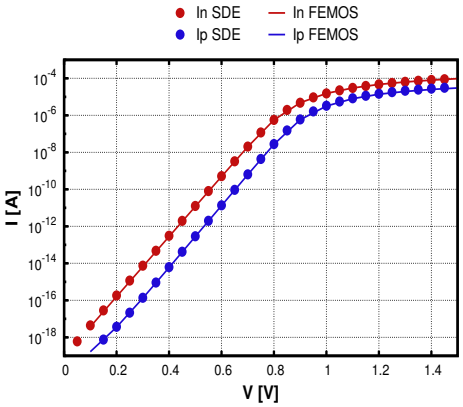
For diode in oxide we analyze the influence of the Masetti mobility model introduced in Section 1.2.3. Fig.4.39a reports electron and hole currents at contact A using the constant mobility model, while Fig.4.39b refers to the Masetti mobility model. This model predicts a decreased value of current flow due to the scattering with impurity dopants, as shown in the 1D plot proposed by Fig.4.39d. Agreement with SDEVICE is very good.

Test case $[V]$	Mobility model $[cm^2V{-}1s^{-1}]$	R/G model	ϵ_{Si}	ϵ_{0x}
$V_A = 0.0 \div 1.5$, $V_B = 0.0$	$\mu_n = 1417$ $\mu_p = 470.5$	SRH II, Auger	11.6	3.9
$V_A = 0.0 \div 1.5$, $V_B = 0.0$	Masetti	SRH II, Auger	11.6	3.9

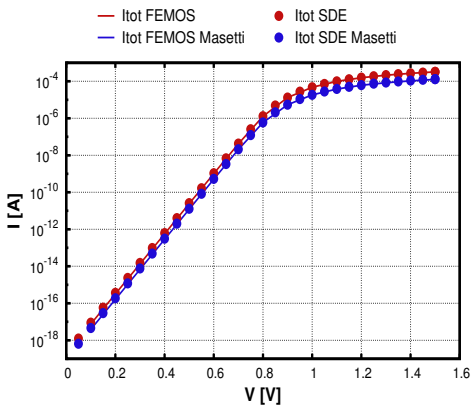
Table 4.6: p-n junction in oxide - list of settings, parameters and models.



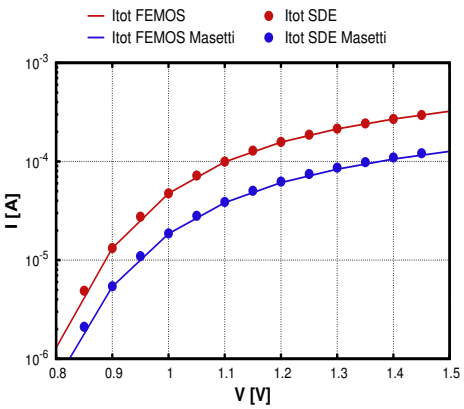
(a) Constant mobility.



(b) Masetti mobility.



(c) Total current.



(d) Total current (Zoom).

Figure 4.39: p-n junction in oxide current at contact - Forward bias.

n-channel/p-channel MOSFET

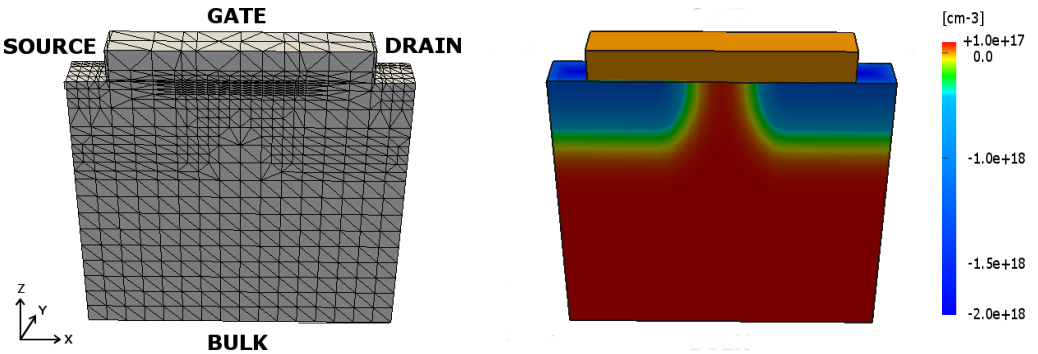
Results for the n-MOSFET and the p-MOSFET are shown in this section. For the n-channel device we refer to Fig.4.22a while for the p-channel device Fig.4.40 shows mesh (2618 vertices and 11514 elements) and doping profile (a Gaussian profile in source and drain regions). In order to validate the code we test the following device conditions:

1. $I_D - V_G$ **characteristic at low drain bias with several mobility models**;
2. $I_D - V_G$ **characteristic for different drain bias**;
3. $I_D - V_D$ **characteristic in off-state (reverse bias)**.

Tab.4.7 and Tab.4.8 show simulation settings of the first test case for the n-MOSFET and the p-MOSFET respectively. Fig.4.41 reports the results for the n-MOSFET: Figs.:4.41a-4.41d show how the current densities change as a function of the use of different mobility models, the agreement with SDEVICE is very good.

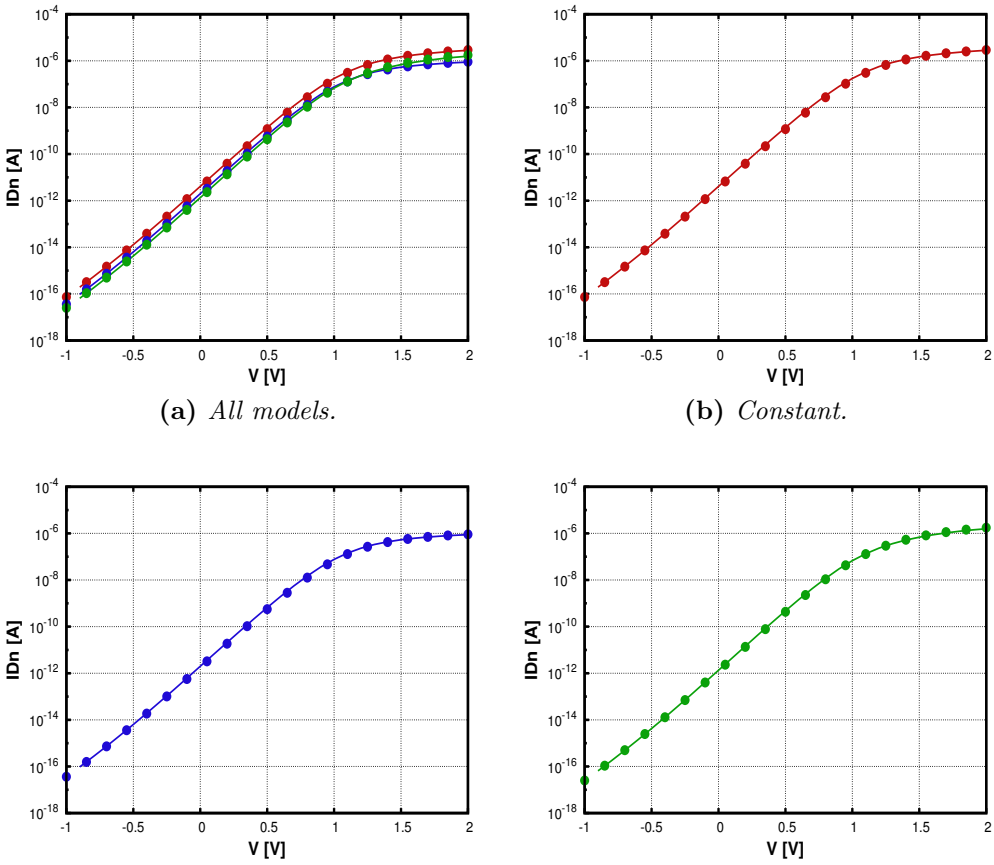
Similar test is performed for the p-MOSFET, but considering that a p-channel usually operats for negative values of the gate bias, we have to apply a positive bias at the source terminal. The results are shown in Fig.4.42: the agreement with SDEVICE is again very good.

Fig.4.43 presents the characteristic of the n-MOSFET at different values of the drain voltage. The analogous test for the p-MOSFET is presented in Fig.4.44. Both are very well in agreement with the commercial code results.



(a) *Mesh.* **(b)** *Doping profile.*

Figure 4.40: p-MOSFET.



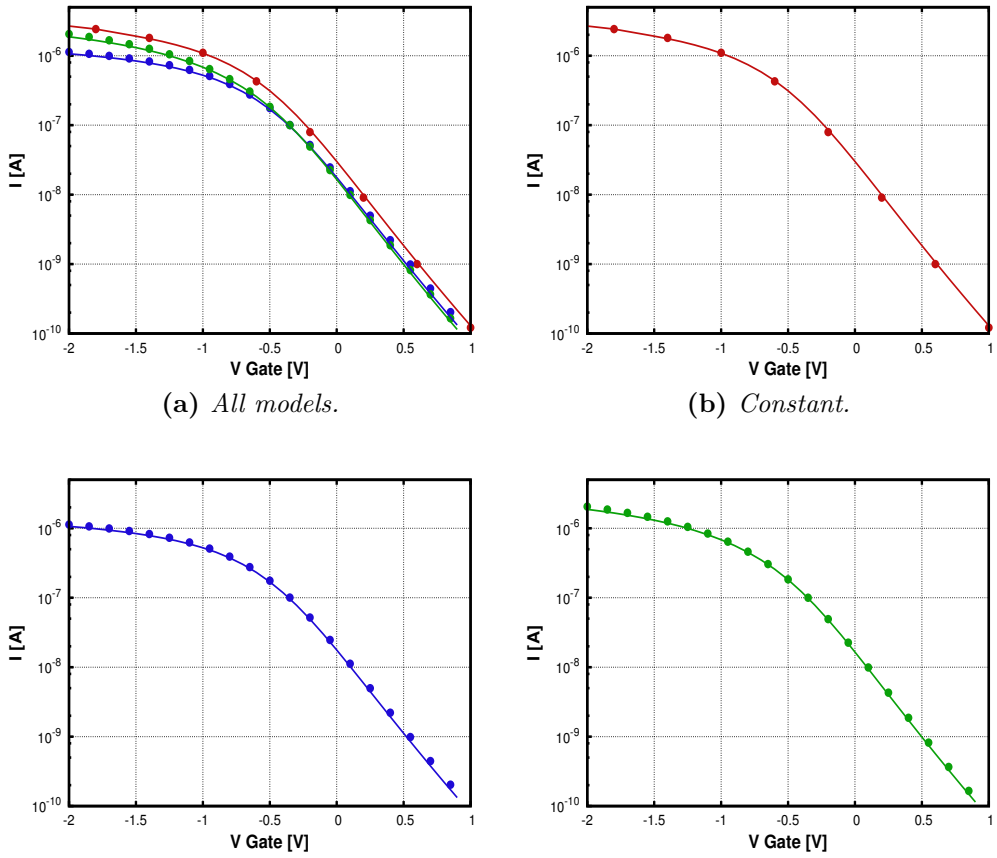
(a) *All models.* **(b)** *Constant.*

(c) *Masetti.* **(d)** *Canali.*

Figure 4.41: $I_D - V_G$ n-MOSFET characteristic - mobility models.

Test case $[V]$	Mobility model $[cm^2V^{-1}s^{-1}]$	R/G model	ϵ_{Si}	ϵ_{0x}
$V_G = -0.5 \div 2.0$, $V_D = 0.1$,$V_S = V_B = 0.0$	$\mu_n = 1417$ $\mu_p = 470.5$	SRH, Auger	11.6	3.9
$V_G = -0.5 \div 2.0$, $V_D = 0.1$, $V_S = V_B = 0.0$	Masetti	SRH, Auger	11.6	3.9
$V_G = -0.5 \div 2.0$, $V_D = 0.1$, $V_S = V_B = 0.0$	Canali	SRH, Auger	11.6	3.9

Table 4.7: $I_D - V_G$ n-MOSFET characteristic - list of settings, parameters and models.



(a) *All models.* **(b)** *Constant.*

(c) *Masetti.* **(d)** *Canali.*

Figure 4.42: $I_S - V_G$ p-MOSFET characteristic - mobility models.

Test case $[V]$	Mobility model $[cm^2V^{-1}s^{-1}]$	R/G model	ϵ_{Si}	ϵ_{0x}
$V_G = -1.5 \div 0.5$, $V_S = 0.1$,$V_D = V_B = 0.0$	$\mu_n = 1417$ $\mu_p = 470.5$	SRH, Auger	11.6	3.9
$V_G = -1.5 \div 0.5$, $V_S = 0.1$, $V_D = V_B = 0.0$	Masetti	SRH, Auger	11.6	3.9
$V_G = -1.5 \div 0.5$, $V_S = 0.1$, $V_D = V_B = 0.0$	Canali	SRH, Auger	11.6	3.9

Table 4.8: $I_S - V_G$ p-MOSFET characteristic - list of settings, parameters and models.

All values and models used for these simulations are summarized in Tab.4.9 for the n-MOSFET and Tab.4.10 for the p-MOSFET.

Test case $[V]$	Mobility model $[cm^2V^{-1}s^{-1}]$	R/G model	ϵ_{Si}	ϵ_{0x}
$V_G = -0.5 \div 2.0$, $V_D = 0.1$, $V_S = V_B = 0.0$	Canali	SRH, II	11.6	3.9
$V_G = -0.5 \div 2.0$, $V_D = 0.2$, $V_S = V_B = 0.0$	Canali	SRH, II	11.6	3.9
$V_G = -0.5 \div 2.0$, $V_D = 0.5$, $V_S = V_B = 0.0$	Canali	SRH, II	11.6	3.9
$V_G = -0.5 \div 2.0$, $V_D = 1.0$, $V_S = V_B = 0.0$	Canali	SRH, II	11.6	3.9
$V_G = -0.5 \div 2.0$, $V_D = 2.0$, $V_S = V_B = 0.0$	Canali	SRH, II	11.6	3.9

Table 4.9: $I_D - V_G$ n-MOSFET for different drain voltages - list of settings, parameters and models.

Test case $[V]$	Mobility model $[cm^2V{-}1s^{-1}]$	R/G model	ϵ_{Si}	ϵ_{0x}
$V_G = -1.5 \div 0.5$, $V_S = 0.05$, $V_D = V_B = 0.0$	Canali	SRH, II	11.6	3.9
$V_G = -1.5 \div 0.5$, $V_S = 0.1$, $V_D = V_B = 0.0$	Canali	SRH, II	11.6	3.9
$V_G = -1.5 \div 0.5$, $V_D = 0.2$, $V_D = V_B = 0.0$	Canali	SRH, II	11.6	3.9
$V_G = -1.5 \div 0.5$, $V_S = 0.5$, $V_D = V_B = 0.0$	Canali	SRH, II	11.6	3.9

Table 4.10: $I_S - V_G$ p-MOSFET for different source voltages - list of settings, parameters and models.

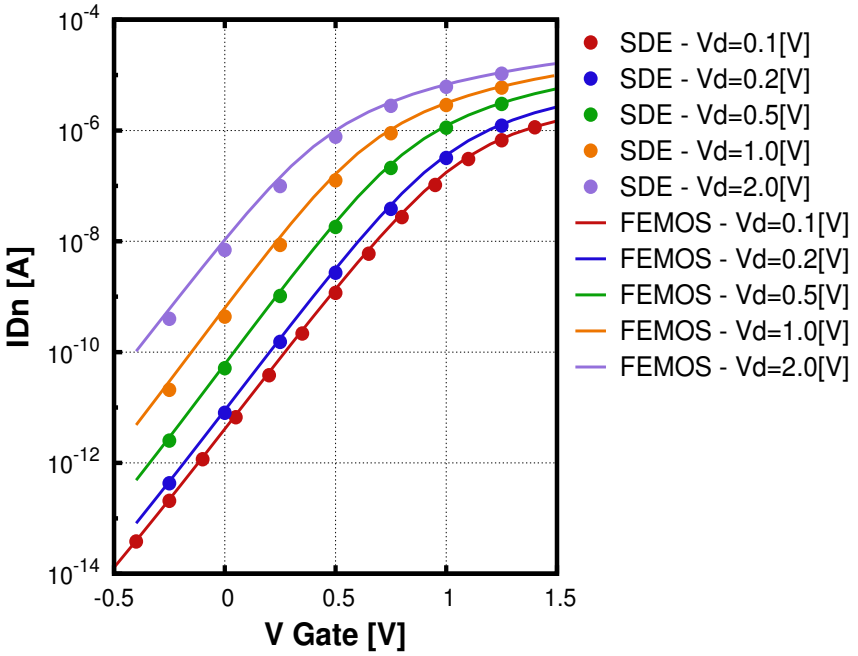


Figure 4.43: $I_D - V_G$ n-MOSFET for different drain voltages.

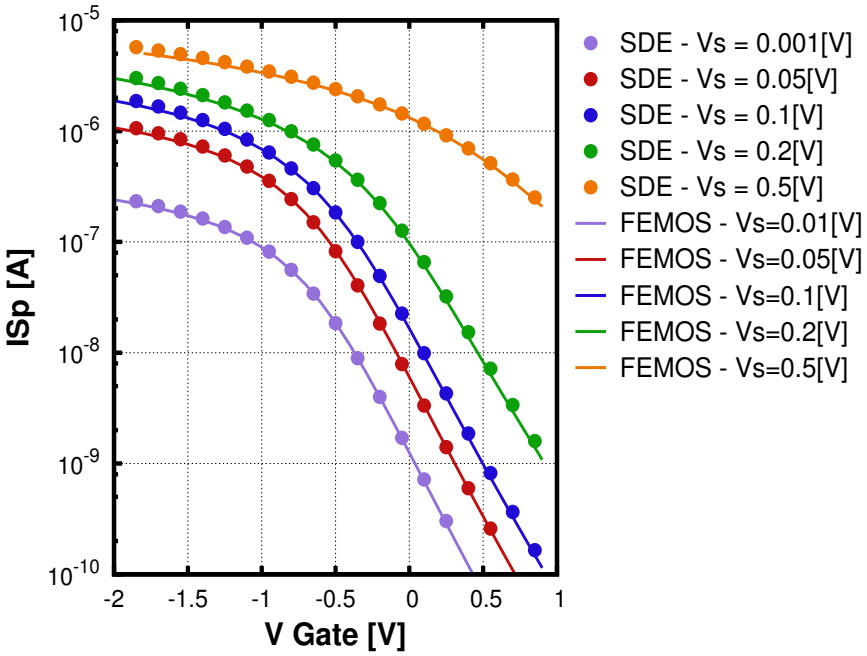


Figure 4.44: $I_S - V_G$ p-MOSFET for different source voltages.

Impact-ionization model

To visualize the effects of impact ionization (II, Van Overstraeten - de Man model see Section 1.2.2) we investigate the devices in off-state with increasing drain (n-MOSFET) or source voltage (p-MOSFET).

The parameters are summarized in Tab.4.11 for the n-MOSFET and in Tab.4.12 for the p-MOSFET.

In order to avoid possibly negative concentration also for the p-channel MOSFET, we add degrees of freedom at the source-bulk junction region as shown in Fig.4.45 (15504 vertices and 81587 elements).

For the n-MOSFET, increasing drain voltage a large amount of generation is produced around the drain-bulk junction as shown in Fig.4.46, where we compare the contribution due to the II model computed at $V_D = 0.5[V]$ between FEMOS and SDEVICE. Some differences arise but the discrepancy does not affect the computation of the current at contacts. Figs.:4.49 and 4.50 confirm the very good agreement with the commercial tool also for the bulk current. In off-state condition no channel is formed beneath the oxide layer, therefore no preferential path is allowed for electron (or hole) conduction: carriers can move toward source (drain) and bulk contact.

Due to the difference in the doping level between n-MOSFET and p-MOSFET the generation of carriers for II is different as shown in Fig.4.46 (n-MOSFET) compared with Fig.4.47 (p-MOSFET) for $V_S = V_D = 0.5[V]$ and Fig.4.48 also for $V_S = 1.2[V]$.

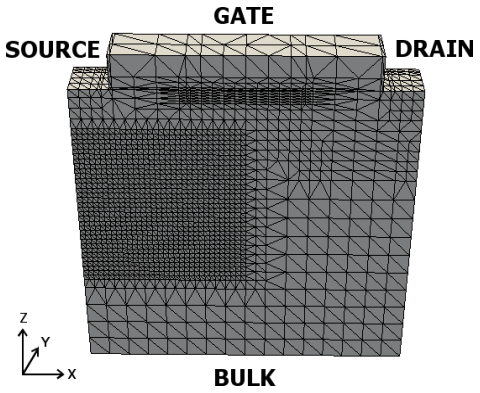


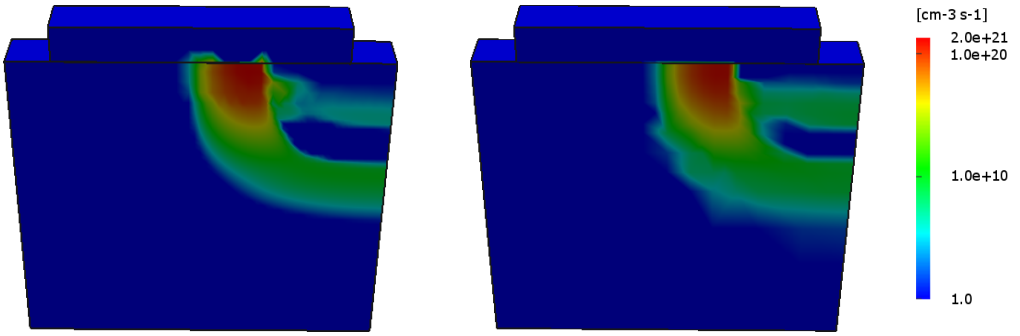
Figure 4.45: pMOSFET mesh.

Test case $[V]$	Mobility model $[cm^2V^{-1}s^{-1}]$	R/G model	ϵ_{Si}	ϵ_{0x}
$V_D = 0.0 \div 1.0$, $V_G = V_S = V_B = 0.0$	$\mu_n = 1417$ $\mu_p = 470.5$	SRH Auger, II	11.6	3.9

Table 4.11: n-MOSFET $I_D - V_D$ off-state characteristic - list of settings, parameters and models.

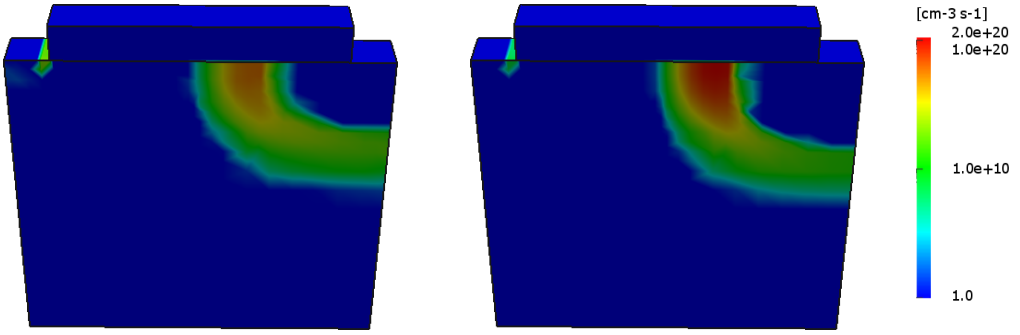
Test case $[V]$	Mobility model $[cm^2V^{-1}s^{-1}]$	R/G model	ϵ_{Si}	ϵ_{0x}
$V_S = 0.0 \div 1.5$, $V_G = V_D = V_B = 0.0$	$\mu_n = 1417$ $\mu_p = 470.5$	SRH Auger, II	11.6	3.9

Table 4.12: p-MOSFET $I_D - V_S$ off-state characteristic - list of settings, parameters and models.



(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.46: n-MOSFET $V_D = 0.5[V]$: contribution of impact ionization with the Van Overstraeten - de Man model.



(a) *FEMOS* **(b)** *SDEVICE*

Figure 4.47: p-MOSFET $V_S = 0.5[V]$: contribution of impact ionization with the Van Overstraeten - de Man model.

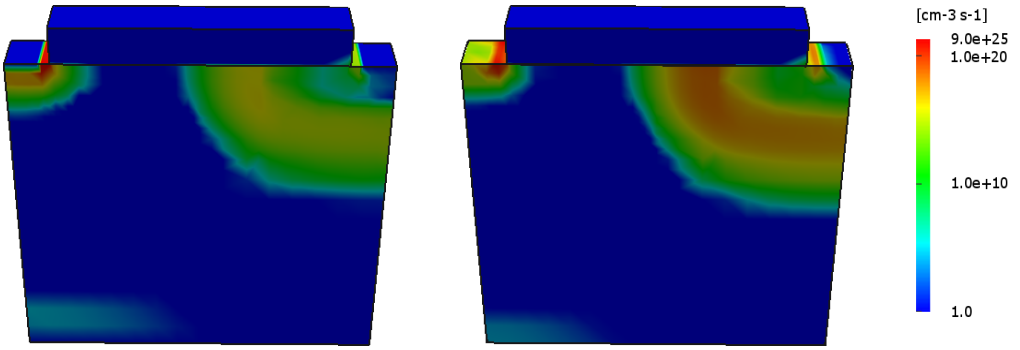


Figure 4.48: p-MOSFET $V_S = 1.2[V]$: contribution of impact ionization with the Van Overstraeten - de Man model.

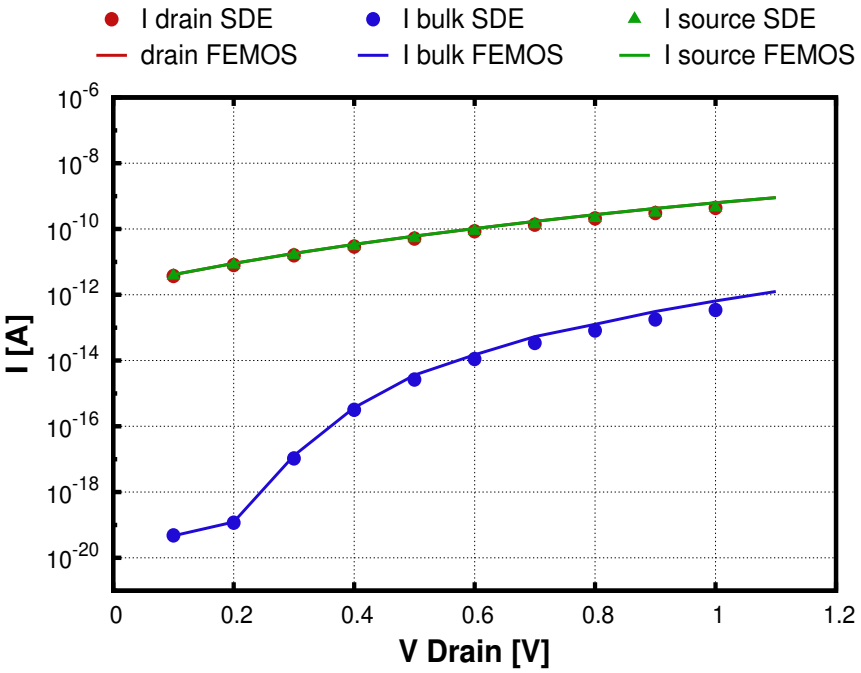


Figure 4.49: n-MOSFET off-state characteristic.

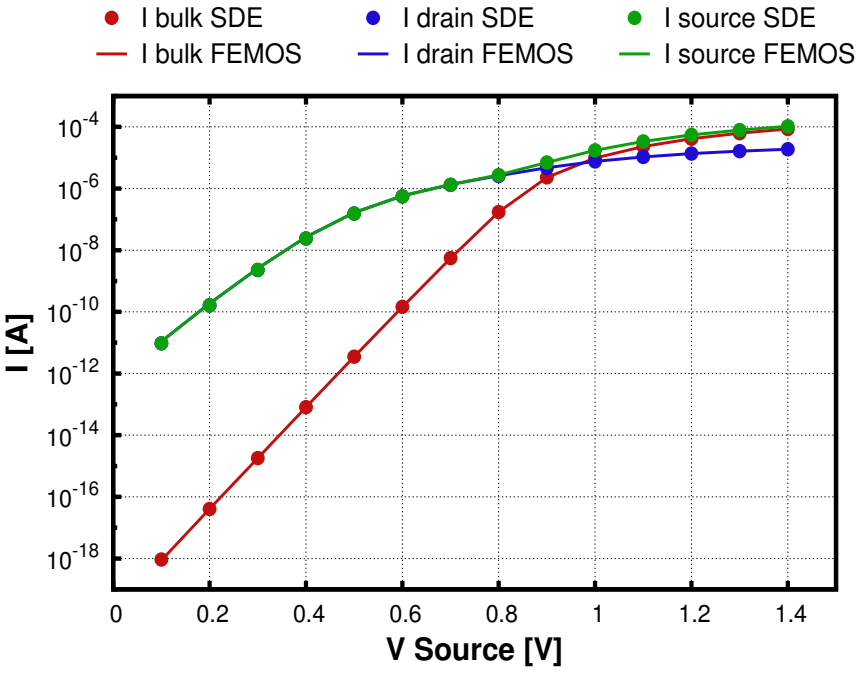


Figure 4.50: p-MOSFET off-state characteristic.

Chapter 5

Post-Processing techniques for current density calculation

In many physical and engineering problems the real interesting variable of the conservation law is the flux inside the domain. The study of micro and nano electronics devices does not except this observation, so that, an accurate description of the current density is a basic requirement. However, we recall that with a displacement-based finite element approach the current density is not a dependent variable of the system but rather a post-calculated quantity.

In this chapter we present the standard Drift-Diffusion formula and we propose an extension of the Scharfetter-Gummel scheme [GS69] to the 3D case through two novel schemes based on:

- *edge average approximation*;
- *alternative upwinding technique.*

The results obtained are compared with a well tested field simulator as a reference (SDEVICE).

5.1 Drift-Diffusion formula

In Section 1.1.4 we have presented three different but mathematically equivalent ways to represent the current density, but not all of them are appropriate for numerical implementation. In particular we exluded from our analysis the *Slotboom equations* (1.40)-(1.41) because the exponential dependency on the factor φ/V_{th} brings unavoidable numerical instability. The classical *Drift-Diffusion formula* (1.31)-(1.32) presents also some difficulties: the drift and diffusion contributions are respectively well defined but their combination may give rise to unphysical oscillations due to numerical cancellation.

Let us introduce some useful notation: with the subscript K we refer to a quantity defined on elements, while the subscript h refers to a quantity defined on vertices. The solutions φ_h, n_h and p_h obtained with the discretization scheme presented in Section 3.3 are piecewise linear continuous functions over $\mathcal{T}_h$. According to (1.31) and (1.32), in order to compute $\mathbf{J}_n$ and $\mathbf{J}_p$ numerical differentiation of the solutions must be carried out. Notice that $\mathbf{J}_n, \mathbf{J}_p \in [X_h^0]^3$. If we want to combine solutions and their derivatives, we have to compute appropriate projection of n_h and p_h:

$$n|_K := < n_h >$$
$$p|_K := < p_h > .$$

where with the symbol $< \cdot >$ we refer to a suitable average on the element, as presented in Section 3.3.3. If the diffusion and the mobility coefficients are variable functions of the space and defined on vertices they also have to be projected on the space X_h^0.

We implemented a numerical differentiation based on Lagrange polynomial interpolation:

$$\nabla n \simeq \nabla(\Pi_h^1 n) = \sum_{i=1}^{N_h} n_i \nabla\psi_i = \nabla n_h$$
$$\nabla p \simeq \nabla(\Pi_h^1 p) = \sum_{i=1}^{N_h} p_i \nabla\psi_i = \nabla p_h.$$

Notice that ∇n_h, $\nabla p_h \in X_h^0$. The discretized form of equations (1.34), (1.35) reads as:

$$\mathbf{J}_n|_K = -qn|_K \mu_n \nabla\varphi_h + qD_n n|_K \nabla n_h \tag{5.1}$$
$$\mathbf{J}_p|_K = -qp|_K \mu_p \nabla\varphi_h - qD_p p|_K \nabla p_h \tag{5.2}$$

where for sake of simplicity we assume constant diffusion and mobility coefficients in K. Equations (5.1) and (5.2) can be easily computed over each element of $\mathcal{T}_h$.

5.2 Edge averaging techniques

It is well known that the classical Scharfetter-Gummel (SG) scheme for discretizing drift-diffusion models has proven to be the workhorse for semiconductor device modeling codes [GS69]. As a matter of fact the EAFE

scheme proposed in Section 3.3.3 is strictly related to the FVSG (Finite Volume Scharfetter-Gummel) method presented by Bank, Fichtner and Rose [BRF83].

In this section we recall the Scharfetter-Gummel formula in a 1D spatial domain and we report the extension of the method to the 2D case proposed in [BCC98b]. Finally we present a novel method in order to extend the Scharfetter-Gummel approach to the 3D framework.

5.2.1 The 1D Scharfetter-Gummel scheme

Consider the solution of the electron continuity equation along a one-dimensional domain. For sake of simplicity, we assume a uniform partition. Moreover at every node is defined φ_h, and in every element the associated electrostatic field $\mathbf{E}_K$.

In 1969 D. Scharfetter and H.K. Gummel (two scientists of Bell Labs), introduced a formula to compute the current densities given φ_h and n_h, p_h at each node of the discretization grid.

The constitutive law for the current density is composed by a drift component, which depends on the electric field, and a diffusion component, which depends on the variation of the carrier density. Considering the geometry shown in Fig.5.1, given a generic element K and defining the voltage drop $\Delta\varphi|_K = \varphi_{i+1} - \varphi_i$ we can distinguish three limit situations:

- $\Delta\varphi|_K \gg 0$, mainly drift component from right to left
- $\Delta\varphi|_K \ll 0$, mainly drift component from left to right
- $\Delta\varphi|_K \simeq 0$, mainly diffusion component

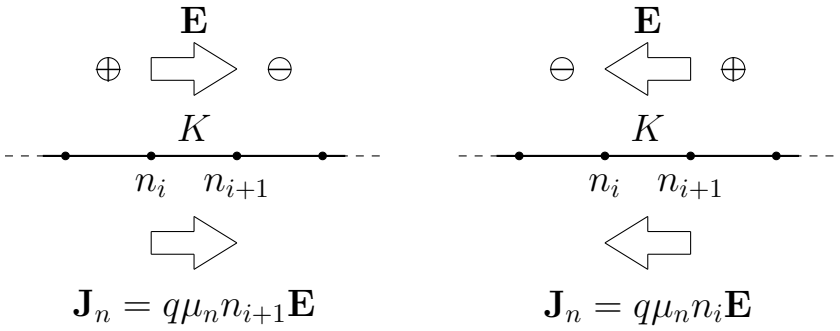


Figure 5.1: Effect of a high electric field over the current density of electron.

All of these situations can be accounted for by the following unified formula

$$J_n|_K = q\frac{D_n}{h}\left[n_{i+1}\mathcal{B}\left(\frac{\Delta\varphi|_K}{V_{th}}\right) - n_i\mathcal{B}\left(-\frac{\Delta\varphi|_K}{V_{th}}\right)\right]. \tag{5.3}$$

In the case $\Delta\varphi|_K = 0$, the SG formula becomes simply

$$J_n|_K = qD_n\frac{n_{i+1} - n_i}{h} \tag{5.4}$$

which is the correct approximation of the current density using a $\mathbb{P}_1$ basis for n_h. When $\Delta\varphi|_K \gg 0$, the SG formula becomes

$$J_n|_K = q\mu_n n_i\frac{\Delta\varphi|_K}{h} \tag{5.5}$$

while for $\Delta\varphi|_K \ll 0$ we have

$$J_n|_K = q\mu_n n_{i+1}\frac{\Delta\varphi|_K}{h}. \tag{5.6}$$

The current density on the element K becomes similar to the Ohm's law where the carrier transported is n_{i+1} when $\Delta\varphi|_K \ll 0$ or n_i when $\Delta\varphi|_K \gg 0$. These situations are well illustrated in Fig.5.1. Analogous considerations holds for holes, and the associated formula for the current density is

$$J_p|_K = q\frac{D_p}{h}\left[p_{i+1}\mathcal{B}\left(-\frac{\Delta\varphi|_K}{V_{th}}\right) - p_i\mathcal{B}\left(\frac{\Delta\varphi|_K}{V_{th}}\right)\right]. \tag{5.7}$$

5.2.2 The 2D Scharfetter-Gummel scheme

One of the main results of [BCC98b] is the equivalence between the finite volume approach and the finite element Galerkin discretizations of the continuity equation. In order to facilitate the connection between these two different discretization approaches the authors introduce for each $K \in \mathcal{T}_h$ a linear map $\mathcal{J}_K : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$\mathcal{J}_K(\{\gamma_i\}_{i=1}^3) = \frac{1}{|K|}\sum_{i=1}^{3}\gamma_i|e_i|s_i\mathbf{t}_i \tag{5.8}$$

where s_i is the measure of the segment from the midpoint of e_i to the intersection of the perpendicular edge bisectors and $\mathbf{t}_i$ denotes the unit tangent vector of the edge e_i (see Fig.5.2). $\mathcal{J}_K$ has the following properties:

$$\mathcal{J}_K(\{\mathbf{J}\cdot\mathbf{t}_i\}_{i=1}^3) = \mathbf{J} \tag{5.9}$$

$$\mathcal{J}_K(\{s_i^{-1}\}_{i=1}^3) = 0 \tag{5.10}$$

$$\int_K \mathcal{J}_K(\{\gamma_i\}_{i=1}^3)\cdot\nabla\psi\, dK = \gamma_{i+1}s_{i+1} - \gamma_{i-1}s_{i-1}. \tag{5.11}$$

Equation (5.9) states that if we are able to compute the tangential component of the current density over all edges, we can combine these values according to (5.8) and obtain the current density $\forall K \in \mathcal{T}_h$.

Using the EAFE scheme, in the case of electrons we have

$$\mathbf{J}_n\cdot\mathbf{t}_i = qD_n\frac{\mathcal{B}(\delta_i(\varphi_h/V_{th}))n_{h,k} - \mathcal{B}(-\delta_i(\varphi_h/V_{th}))n_{h,j}}{|e_i|} \tag{5.12}$$

where

$$\delta_i(\varphi_h/V_{th}) = \frac{\varphi_{h,k}-\varphi_{h,j}}{V_{th}}. \tag{5.13}$$

The extension of this procedure to the 3D case is non trivial, because the characterization of the cross-section s_i becomes much more complex.

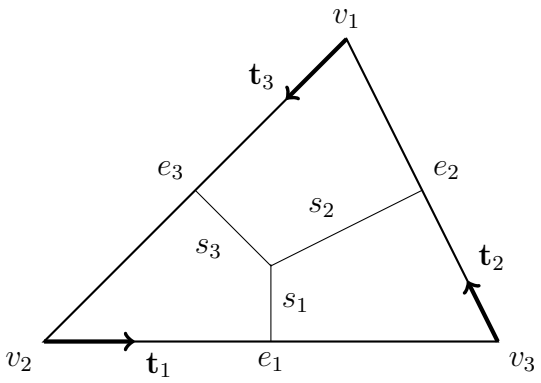


Figure 5.2: Parameters associated with element K for the current density calculation.

5.2.3 The 3D Scharfetter-Gummel scheme

In this section we present a novel method for the calculation of the electron current density over each element of the grid that is based on the so-called Primal-Mixed formulation [RT91].

We start by recalling the formula for the electron current density expressed as a function of the quasi Fermi potential

$$\mathbf{J}_n = -q\mu_n n \nabla \varphi_n. \tag{5.14}$$

Relation (5.14) can be written considering equation (1.22) as

$$\mathbf{J}_n \frac{\exp\left(\dfrac{\varphi_n - \varphi}{V_{th}}\right)}{q\mu_n n_i} + \nabla \varphi_n = 0. \tag{5.15}$$

Let $\mathbf{J}_n \in [L^2(\Omega)]^3$ and $\varphi_n, \varphi \in H^1(\Omega)$. We multiply (5.15) with a generic function $\mathbf{q} \in [L^2(\Omega)]^3$ and then intagrate over the domain Ω

$$\int_\Omega \frac{\exp\left(\dfrac{\varphi_n - \varphi}{V_{th}}\right)}{q\mu_n n_i} \mathbf{J}_n \cdot \mathbf{q}\, d\Omega + \int_\Omega \nabla \varphi_n \cdot \mathbf{q}\, d\Omega = 0. \tag{5.16}$$

We proceed using the discrete space of the piecewise constant functions over $\mathcal{T}_h$

$$V_h = \left\{ w \in L^2(\Omega) : w|_K \in \mathbb{P}_0 \forall K \in \tau_h \right\}. \tag{5.17}$$

Now the discrete quantities are $\mathbf{J}_n^h \in [V_h]^3$ and $\nabla \varphi_n^h \in V_h$. We consider the following choice of the test function $\mathbf{q}_h \in [V_h]^3$

$$\mathbf{q}_{1,2,3}^h = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}. \tag{5.18}$$

From (5.16) we obtain a system of equations defined for each $K \in \mathcal{T}_h$:

$$\int_K \frac{\exp\left(\dfrac{\varphi_n - \varphi}{V_{th}}\right)}{q\mu_n n_i} \mathbf{J}_n^h \cdot \mathbf{q}_i^h\, dK + \int_K \nabla \varphi_n^h \cdot \mathbf{q}_i^h\, dK = 0 \qquad i = 1, 2, 3. \tag{5.19}$$

After integration we have for the generic component of the current density the following equation

$$[\mathbf{J_n}]_i = -\mathcal{H}_K \left(q\mu_n n_i \exp\left(\frac{\varphi - \varphi_n}{V_{th}}\right)\right) \frac{\partial \varphi_n^h}{\partial x_i} \quad i = 1, 2, 3, \quad \forall K \in \mathcal{T}_h. \tag{5.20}$$

We do not evaluate the harmonic average with an exact 3D integration because it may be computationally expensive. Therefore, we approximate $\mathcal{H}_K(\cdot)$ by the following quadrature

$$\left(\frac{\int_K f^{-1}\,dK}{|K|}\right)^{-1} \simeq \left(\frac{\int_{e*} f^{-1}\,de}{|e^*|}\right)^{-1} \tag{5.21}$$

where

$$f = q\mu_n n_i \exp((\varphi - \varphi_n)/V_{th})\,,$$

and e^* is the edge of ∂K where the maximum drop of f occurs. The choice of the approximate formula (5.21) is motivated by the fact that the diffusion coefficient inside the element K may be well represented by the edge where the diffusion phenomenon is more important rather than considering the entire element: the problem now is the indentification of the correct edge. Let us consider a quantity defined at the vertices

$$\Phi := \frac{\varphi - \varphi_n}{V_{th}} \tag{5.22}$$

which is the difference between the electrostatic potential and the quasi Fermi potential level. Now for every element consider two vertices: $\mathbf{x}_m$ s.t. $\Phi(\mathbf{x}_m) = \Phi_m := \min_K(\Phi)$ and $\mathbf{x}_M$ s.t. $\Phi(\mathbf{x}_M) = \Phi_M := \max_K(\Phi)$. Obviously there exists only one edge which connects these two points: on this edge (e^*) we perform the 1D integration (5.21).

Along the edge e^* we have

$$f(s) = q\mu_n n_i \exp\left(\Phi_m + (\Phi_M - \Phi_m)\frac{s - s_m}{|e^*|}\right) \tag{5.23}$$

where $s \in [s_m, s_M]$ is the parameter refered to the edge e^* s.t. $f(s_m) = f(\mathbf{x}_m)$ and $f(s_M) = f(\mathbf{x}_M)$. We can solve (5.21) with the following change of variables $\eta := (s - s_m)/|e^*|$ and proceed with trivial integration steps, to obtain

$$\begin{aligned}\int_{e^*} f^{-1}\,de &= |e^*| \int_0^1 \frac{\exp\left(-\Phi_m - (\Phi_M - \Phi_m)\eta\right)}{q\mu_n n_i}\,d\eta \\ &= |e^*| \frac{\exp(-\Phi_m)}{q\mu_n n_i} \frac{\exp(\Phi_m - \Phi_M) - 1}{\Phi_m - \Phi_M} \\ &= |e^*| \frac{\exp(-\Phi_m)}{q\mu_n n_i} \frac{1}{\mathcal{B}(\Phi_m - \Phi_M)}\end{aligned}$$

from which we finally get

$$\int_K f^{-1}\,dK \simeq q\mu_n n_i \exp(\Phi_m)\mathcal{B}(\Phi_m - \Phi_M). \tag{5.24}$$

Similar results may be obtained repeating the integration and considering s_M as starting point

$$\int_K f^{-1}\, dK \simeq q\mu_n n_i \exp(\Phi_M)\mathcal{B}(\Phi_M - \Phi_m)\,. \tag{5.25}$$

Equation (5.24) and (5.25) can be combined to find

$$\mathbf{J}_n|_K = -q\mu_n \left[\frac{n_{\min}\mathcal{B}(-\Delta\Phi_{max}) + n_{\max}\mathcal{B}(\Delta\Phi_{max})}{2}\right] \nabla\varphi_n^h \tag{5.26}$$

where $n_{\min} = n_i e^{\Phi_m}$ and $n_{\max} = n_i e^{\Phi_M}$ while $\Delta\Phi_{\max} := \Phi_M - \Phi_m$. If we consider equation (5.26) over a one-dimensional domain we can recover equation (5.3), which shows that the above described approach is the natural extension of the Scharfetter-Gummel formula to the 3D case. Following the same procedure we obtain for the hole current density

$$\mathbf{J}_p|_K = -q\mu_p \left[\frac{p_{\min}\mathcal{B}(\Delta\Phi_{max}) + p_{\max}\mathcal{B}(-\Delta\Phi_{max})}{2}\right] \nabla\varphi_p^h\,. \tag{5.27}$$

5.3 Upwinding techniques

It is well known that the classical finite element method applied to a convection-diffusion problem is unstable when the solution presents boundary layers. This has led to the introduction of upwinding techniques which in 1D case consists of adding an artificial diffusion term to the original problem.

Using the finite element space X_h^1, in the case of the 1D electron continuity equation we have the following perturbed problem

$$-\partial_x(qD_n(1 + \Phi(\mathbb{P}e|_K))\partial_x n - q\mu_n n\partial_x\varphi) = -qR. \tag{5.28}$$

where Φ is the stabilization function and $\mathbb{P}e|_K$ is the local Pèclet number defined as

$$\mathbb{P}e|_K = \frac{h\partial_x\varphi_h}{2V_{th}} = \frac{\Delta\varphi}{2V_{th}}.$$

The weak form associated with problem (5.28) is

$$a_h(n, v) = a(n, v) + \sum_{K\in\mathcal{T}_h} \int_K \Phi(\mathbb{P}e|_K)\partial_x n \cdot \partial_x v\, dK \tag{5.29}$$

and relation (5.1) becomes

$$J_n|_K = -qn|_K \mu_n \partial_x \varphi_h + qD_n(1 + \Phi(\mathbb{P}e|_K))n|_K \partial_x n_h. \tag{5.30}$$

where, considering local indices for the vertices of K, we have:

$$\begin{aligned} n|_K &= \frac{\int_K n_h \, dx}{|K|} = \frac{n_1 + n_2}{2} \\ \partial_x \varphi_h &= \frac{\varphi_2 - \varphi_1}{h} = \frac{\Delta \varphi}{h} \\ \partial_x n_h &= \frac{n_2 - n_1}{h}. \end{aligned}$$

In order to guarantee the consistency of (5.29) with respect to the standard Galerkin weak form the stabilization function must satisfy

$$\lim_{\mathbb{P}e|_K \to 0} \Phi(\mathbb{P}e|_K) = 0 \qquad \forall K \in \mathcal{T}_h. \tag{5.31}$$

The efficiency of an upwinding scheme is related to the choice of $\Phi(\mathbb{P}e|_K)$ and as we are perturbing the problem we would like to satisfy some interesting limiting cases for the current density in the the original problem:

1 **Constant carrier concentrations** (only drift contribution),

$$\mathbf{J}_n = q\mu_n n \mathbf{E}\,, \qquad \mathbf{J}_p = q\mu_p p \mathbf{E}\,.$$

2 **Constant potential** ($\mathbf{E} = 0$, only diffusive contribution),

$$\mathbf{J}_n = qD_n \nabla n\,, \qquad \mathbf{J}_p = -qD_p \nabla p\,.$$

3 **Constant quasi Fermi potential**, which implies that $n = C_1 e^{\varphi/V_{th}}$ and $p = C_2 e^{-\varphi/V_{th}}$ where C_1 and C_2 are two arbitrary constants such that

$$C_1 = \exp(-\bar{\varphi}_n/V_{th}) \qquad C_2 = \exp(\bar{\varphi}_p/V_{th})$$

where $\bar{\varphi}_n$ and $\bar{\varphi}_p$ are given contact values. Under this assumption from equations (1.34) and (1.35) we have:

$$\begin{aligned} \mathbf{J}_n &= -q\mu_n(n\nabla\varphi - V_{th}(\frac{C_1}{V_{th}}\nabla\varphi e^{\varphi/V_{th}}) = 0 \\ \mathbf{J}_p &= -q\mu_p(p\nabla\varphi + V_{th}(-\frac{C_1}{V_{th}}\nabla\varphi e^{-\varphi/V_{th}}) = 0\,. \end{aligned}$$

Constant quasi Fermi potentials correspond to thermodynamical equilibrium condition for the carrier densities and imply no current flow in the device.

Thanks to assumption (5.31) the stabilized current (5.30) satisfies case 1 and case 2. Case 3 is recovered imposing

$$J_n|_K(\Pi_1^k(Ce^{\varphi/V_{th}})) = 0. \tag{5.32}$$

Using (5.32) in (5.30) we have:

$$\begin{aligned} q\mu_n < n_h > \partial_x\varphi_h &= qD_n(1+\Phi(\mathbb{P}e|_K)\partial_x n_h \\ < n_h > \partial_x\varphi_h &= V_{th}(1+\Phi(\mathbb{P}e|_K))\partial_x n_h \end{aligned}$$

and finally we get the following relation for the stabilization function

$$\Phi(\mathbb{P}e|_K) = \sigma\mathbb{P}e|_K\frac{n_1+n_2}{n_2-n_1} - 1 \tag{5.33}$$

where

$$\sigma = sign(\Delta\varphi).$$

Now we impose the constant quasi Fermi potential hypothesis ($n_i = \exp(\varphi_i/V_{th})$)

$$\begin{aligned} \Phi(\mathbb{P}e|_K) &= \sigma\mathbb{P}e|_K\frac{e^{\varphi_1/V_{th}}+e^{\varphi_2/V_{th}}}{e^{\varphi_2/V_{th}}-e^{\varphi_1/V_{th}}} - 1 \\ &= \sigma\mathbb{P}e|_K\frac{e^{\Delta\varphi/V_{th}}+1}{e^{\Delta\varphi/V_{th}}-1} - 1 \\ &= \sigma\mathbb{P}e|_K\frac{e^{2\sigma\mathbb{P}e|_K}+1}{e^{2\sigma\mathbb{P}e|_K}-1} - 1. \end{aligned}$$

Setting $X := 2\sigma\mathbb{P}e|_K$ we have

$$\begin{aligned} \Phi(X) &= \frac{X}{2}\left(\frac{e^X}{e^X-1}+\frac{1}{e^X-1}\right) - 1 \\ &= \frac{1}{2}\left(\mathcal{B}(-X)+\mathcal{B}(X)\right) - 1 \\ &= \frac{1}{2}\left(X+\mathcal{B}(X)+\mathcal{B}(X)\right) - 1 \\ &= \mathcal{B}(X) + \frac{X}{2} - 1. \end{aligned}$$

Replacing the definition of X we obtain for both $\Delta\varphi > 0$ and $\Delta\varphi < 0$

$$\Phi(\mathbb{P}e|_K) = \mathcal{B}(2\mathbb{P}e|_K) + \mathbb{P}e|_K - 1. \tag{5.34}$$

Replacing (5.34) in (5.30) we have the well known 1D Scharfetter-Gummel discretization scheme proposed by Allen and Southwell in [AS55]. In a 3D framework it is not intuitive how to evaluate equation (5.34), but considering (5.33) we can say that a straightforward extension to the 3D case of the 1D Scharfetter-Gummel stabilization can be found considering a 3×3 diagonal tensor $\underline{\underline{\mathbf{\Phi}}}^K$ defined on each element as follows

$$\underline{\underline{\mathbf{\Phi}}}^K_{ii} = -\frac{< \Pi_1^k(e^{\varphi/V_{th}}) > \partial_{x_i}\varphi}{\partial_{x_i}\Pi_1^k(e^{\varphi/V_{th}})V_{th}} - 1 \qquad i = 1, 2, 3, \tag{5.35}$$

In (5.36) the argument of the exponential can be a highly varying function over K, therefore it is better to consider a reference value for the electrostatic potential. Observing that $\varphi|_K \in [\varphi_{min}, \varphi_{max}]$ we can use one of these values as reference and having

$$\underline{\underline{\mathbf{\Phi}}}^K_{ii} = -\frac{< \Pi_1^k(e^{(\varphi-\varphi_{\min})/V_{th}}) > \partial_{x_i}\varphi}{\partial_{x_i}\Pi_1^k(e^{(\varphi-\varphi_{\min})/V_{th}})V_{th}} - 1 \qquad i = 1, 2, 3. \tag{5.36}$$

Finally, the 3D electron current density is

$$\mathbf{J}_n|_K = -qn|_K\mu_n\nabla\varphi_h + qD_n n|_K(\underline{\underline{\mathcal{I}}} + \underline{\underline{\mathbf{\Phi}}}_K)\nabla n_h \tag{5.37}$$

where $\underline{\underline{\mathcal{I}}}$ is the 3×3 identity tensor and $n|_K = \mathcal{M}(n_h)$, where $\mathcal{M}(\cdot)$ is the integral average defined in (3.35).

Following the same procedure we obtain the hole current density relation

$$\mathbf{J}_p|_K = -qp|_K\mu_p\nabla\varphi_h - qD_p p|_K(\underline{\underline{\mathcal{I}}} + \underline{\underline{\mathbf{\Phi}}}_K)\nabla p_h \tag{5.38}$$

where

$$\underline{\underline{\mathbf{\Phi}}}^K_{ii} = -\frac{< \Pi_1^k(e^{(\varphi_{\min}-\varphi)/V_{th}}) > \partial_{x_i}\varphi}{\partial_{x_i}\Pi_1^k(e^{(\varphi_{\min}-\varphi)/V_{th}})V_{th}} - 1 \qquad i = 1, 2, 3. \tag{5.39}$$

5.3.1 Results

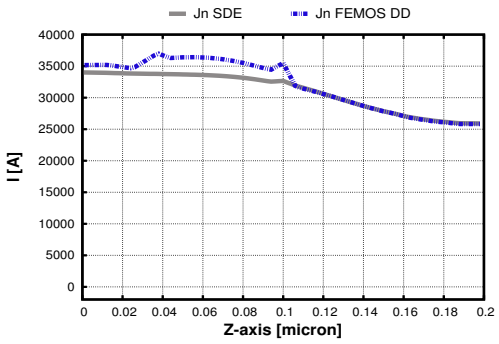
In this section we compare the performance of the different current computational methods . The tests are the p-n junction (Tab.4.1 $V_A = 1.0[V]$) and the n-MOSFET (Tab.4.3 on-state condition). We recall the procedures we used for the calculation:

- **Drift-Diffusion** defined by (5.1) and (5.2);
- **3D edge average Scharfetter-Gummel** defined by (5.26) and (5.27);
- **3D upwinding Scharfetter-Gummel** defined by (5.37) and (5.38).

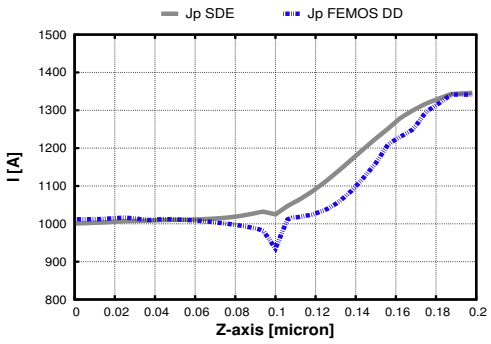
The electron and hole current densities for the forward biased p-n junction are depicted in Fig.5.3 using 1D plots along a line parallel to the Z-axis and placed at the center of the device. We note that the best agreement is obtained by the 3D upwinding SG method. The expected critical behaviour of Figs.:5.3a and 5.3b, due to the wrong balance between drift and diffusion contributions, is fixed by the upwinding technique as shown in Figs.:5.3e and 5.3f.

Considering Fig.5.3c, the 3D edge average SG scheme is well aligned with the commercial tool inside the device but we note important differences at the contact where electrons are minoritary carriers. (Fig.5.3d shows the dual effect for the hole current). Physically speaking the current densities computed by the SG 3D formula are not wrong: in fact assuming ideal contacts we are enforcing that all the recombination of the excess carriers occurs at the contact surface. This phenomenon is restricted to the elements close to the Ohmic contacts and therefore it is strictly related to mesh refinement as shown in Fig.5.4. Fig.5.4a well depictes the behaviour of the boundary layers of the hole quasi Fermi potential, while Fig.5.4b shows that as we refine the grid the effect is more confined at the contact leading to a better prediction of the current density.

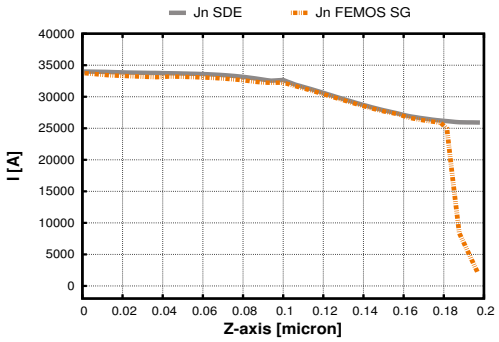
The results for the n-MOSFET are shown in Fig.5.5 compared with SDEVICE (Fig.5.5a). In order to avoid the above problem we refined each Ohmic contacts and as a consequence the 3D edge average SG scheme performs very well (Fig.5.5b). Fig.5.5c shows that when the current density presents more than one main component or equation (5.36) is next to critical computations (zero division or overflow) the upwinding scheme behaves worse. In fact, the principal problem of the 3D upwinding SG scheme is the numerical evaluation of (5.36). However, the numerical issues can be partially solved using a finer mesh as shown in Fig.5.5d, where we use a mesh with 35342 vertices (about ten times the number of degrees of freedom of the standard mesh). Fig.5.5e shows the current density computed with the standard Drift-Diffusion formula using the same mesh. Comparing Fig.5.5c, Fig.5.5e and Fig.5.5d we can say that even if the 3D upwinding SG scheme is unstable, the addition to the standard Drift-Diffusion formula of an artificial diffusion modulates by the coefficient (5.36), gives better results also for a coarse mesh.



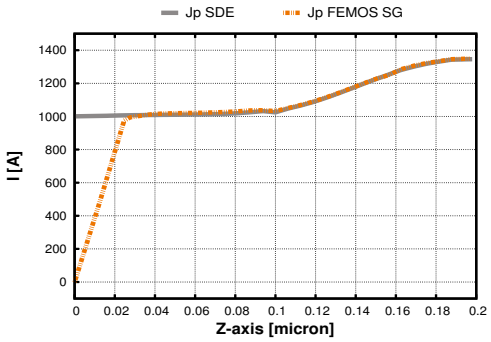
(a) *Standard DD - Jn*



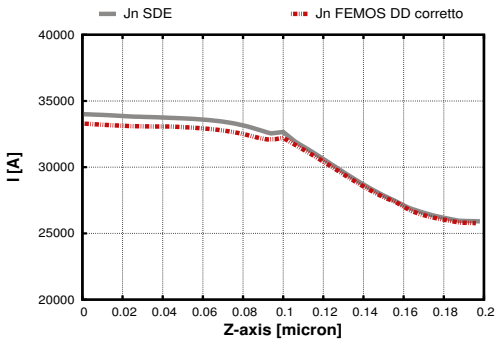
(b) *Standard DD - Jp*



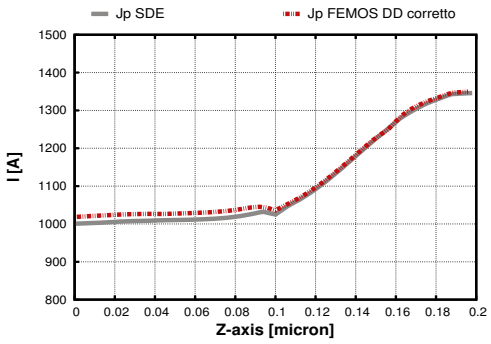
(c) *3D edge average SG - Jn*



(d) *3D edge average SG - Jp*

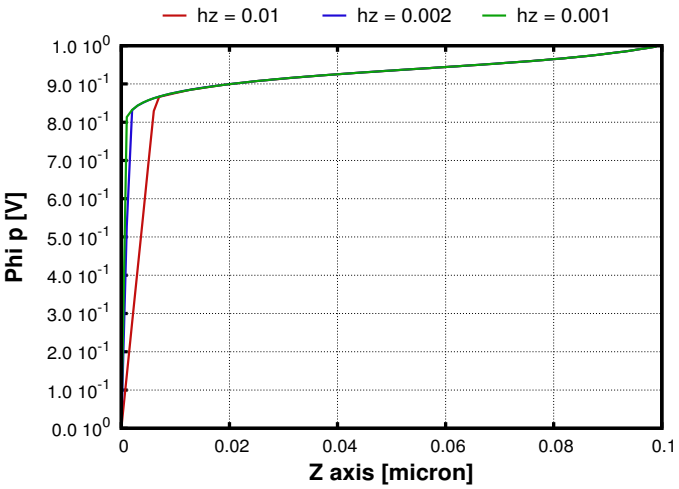


(e) *3D upwinding SG - Jn*

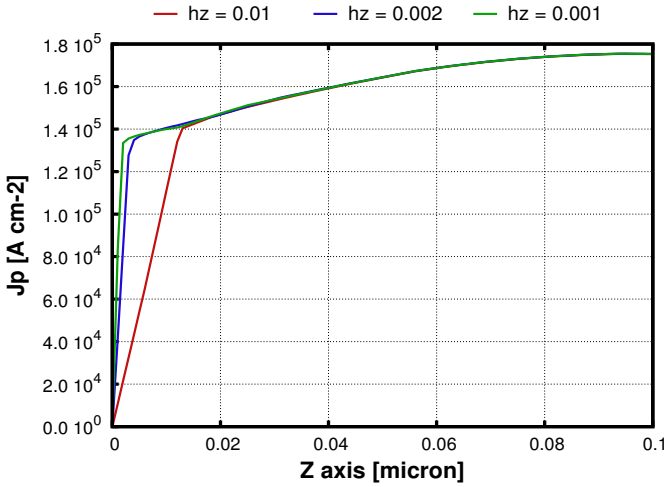


(f) *3D upwinding SG - Jp*

Figure 5.3: 1D plot p-n junction, hole and electron current densities - $V_A = 1.0[V]$

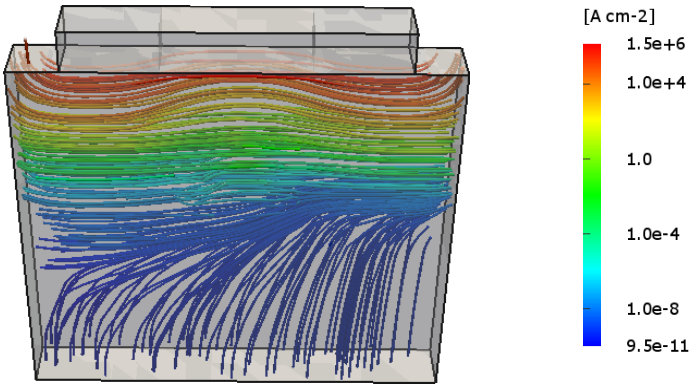


(a) *Hole quasi Fermi potential.*

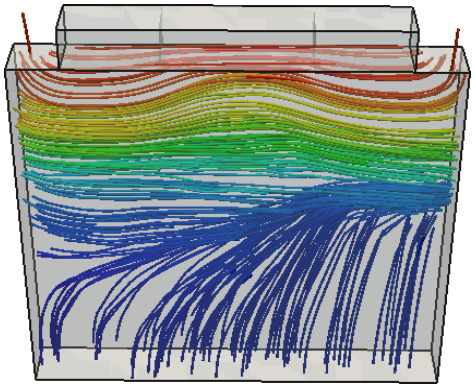


(b) *Jp.*

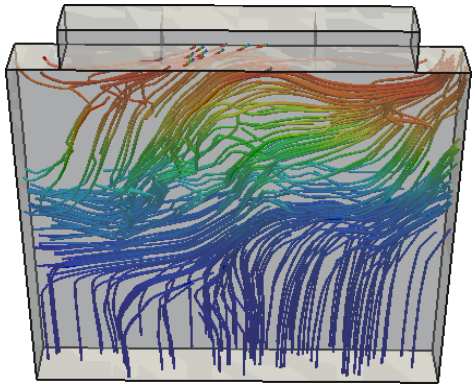
Figure 5.4: p-n junction forward biased, hole quasi Fermi potential and hole current density for different meshes refined at contacts (h_Z is the grid step used along Z-axis near the contact)



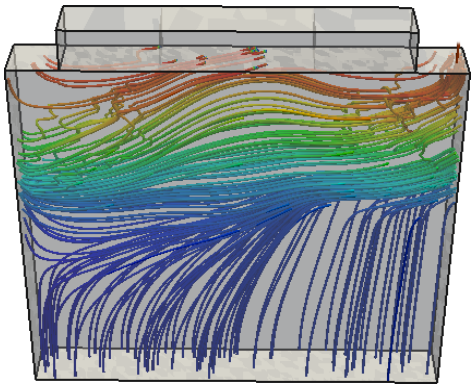
(a) *SDEVICE*



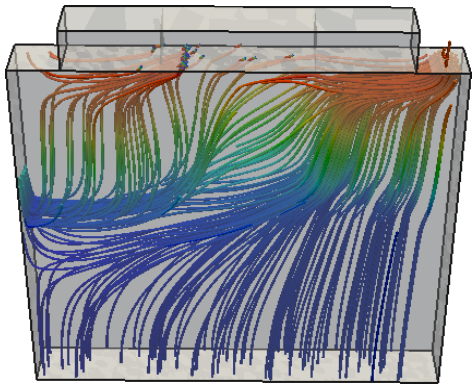
(b) *3D edge average SG*



(c) *3D upwinding SG*



(d) *3D upwinding SG (finer mesh)*



(e) *Standard DD (finer mesh)*

Figure 5.5: n-MOSFET on-state, calculation of J_n with different methods compared with the commercial code.

Chapter 6

Conclusions and future work

In this MD Thesis, we have addressed the simulation of semiconductor devices in 3D framework. The Gummel map algorithm [Gum64] is employed to solve the Drift-Diffusion model [Jac84]. The Non Linear Poisson has been discretized using the Galerkin finite element method [QV08] following a displacement-based formulation and improving the convergence algorithm with the damping technique [Deu74]. The Continuity equations have been treated using the EAFE scheme [XZ99] reserving a particular attention to the Zikatanov condition (presented in [XZ99]) in order to discuss the discrete maximum principle.

This mathematical framework has been implemented through shared libraries using an object-oriented programming language (C++).

The code has been thoroughly tested on different semiconductor devices (p-n junction, p-n junction in oxide and n-channel/p-channel MOSFET), comparing the results with a commercial tool as reference.

A great effort has been spent in order to compute the current both at contacts and inside the device. In the first case we extended the *residual method* [GS06] to the 3D framework with excellent results.

Concerning the evaluation of the current density we proposed two novel schemes to post-process the solutions extending the Scharfetter-Gummel scheme to the 3D case:

- *edge average technique*, based on a primal mixed formulation of the continuity equation, followed by the approximation of the harmonic average of the diffusion coefficient;

- *alternative upwinding technique*, based on a generalization of the 1D

stabilization function associated with the Scharfetter-Gummel technique.

To conclude, we wish to mention some possible directions for future activities and investigations:

- the novel schemes presented in Chapter 5 could be also used to discretize the continuity equation. Moreover, as regards the edge average technique, some care must be taken due to the highly non linear dependence of the diffusion coefficient on the quasi Fermi potential levels;
- in order to complete the integration with the other modules of FEMOS the coupling with the multiphysical environment must be developed;
- the available mobility and R/G models must be extended to include other physical effects;
- the treatment of floating semiconductor regions;
- the development of the mixed hybrid strategy in order to guarantee the conservation of the flux across each internal face of the triangulation;
- implementation of the fully coupled Newton method.

Appendices

Appendix A

FEMOS inputfile

In this appendix we address the information in order to personalize the input file (`input.txt`) of the FEMOS executable. The parsing of the input file is managed by the library `reading_files.so`. The file is organized in sections identified by a name and a scope; each voice of a section could be a vector of string (:), a vector of double (=) or a vector of pair of string and double (:=).

Now we describe each section. The section `File` sets the input and output path.

```
File {

  Input       : ./input/MosMesh.tdr ;
  # sets mesh file path

  Output      : ./output/MosSimulation ;
  # sets output file path and name

  Output_type : xmf ;
  # sets output file (.vtk/.xmf)

  Par         : ./input/MosParameter.par
  # sets parameter file path

}
```

In the `Mesh` section it is possible to activate some checks on the mesh and the related prints (if one activated check fails simulation is aborted).

```
Mesh {
```

```
  RegularityClassic : no ;
  # activate the regularity check on the mesh and the associated
    print

  RCtoll = 1000 ;
  # sets the tolerance for the regularity check

  RegularityDelauny : no
  # activates the Zikatanov condition check on the mesh and the
    associated print

}
```

The `Equation` section sets which equations are solved in the general Gummel Map. At the stage of code development, semiconductor physics is working by activating carriers type when silicon material is present in the mesh.

```
Equation {

  Poisson : no ;
  # activates the Linear Poisson equation (not used for
    semiconductor material)

  Carriers_type : Electron , Hole ;
  # defines carrier types

  Carriers :=  Electron -1 , Hole 1 ;
  # defines carrier charges

  Thermal : no ;
  # activates the Thermal equation

  Chemicals_type : no ;
  # defines chemical species

  Chemicals := no 0;
  # defines effective charge of chemical species

  Mechanical : no
  # activates the Mechanical equation
}
```

Contacts are defined in the section `Electrode`. The ramping procedure is managed by the variables listed in the section `Stationary`.

```
Electrode {

  # definition of contacts and their initial voltage [V]
  Contact       :=  Source 0.0 , Drain 0.1 ,
                    Bulk 0.0 , Gate 0.0 ;

  # definition of the workfunction of each electrode [eV]
  Workfunction  :=  Source 4.65 , Drain 4.65 ,
                    Bulk 4.65 , Gate 4.65
}



Stationary {
  # definition of the final voltage for each contact [V]
  GoalContact   := Source 0.0 , Drain 0.1 ,
                    Bulk 0.0 , Gate 2.0 ;

  # defines the number of steps between intial e final voltage
  Steps         = 20 ;

  # defines the minimum bias step (only for semiconductor) if
    the ramping procedure is active
  MinStep       = 0.01 ;

  # actives the ramping procedure
  Checkramp     : yes
}
```

Each physical module of FEMOS has a default group of printable variables defined inside the library `libUtility.so`. The output can be personalized by changing the `Plot` section of the input file.

```
Plot {

  # defines which variable print in the output file (variables
    for semiconductor)
  eDensity : yes ;                  # [cm-3]
  hDensity : yes ;                  # [cm-3]
  DopantConc : yes ;                # [cm-3]
  DonorConc : yes ;                 # [cm-3]
  AcceptorConc : yes ;              # [cm-3]
  ElectrostaticPotential : yes ;    # [V]
  ElectricField : yes ;             # [V cm-1]
  eQuasiFermiPotential : yes ;      # [V]
  hQuasiFermiPotential : yes ;      # [V]
```

```
  MiddleBandLevel : yes ;          # [eV]
  CondBand : yes ;                 # [eV]
  ValBand : yes ;                  # [eV]
  eQuasiFermiBand : yes ;          # [eV]
  hQuasiFermiBand : yes ;          # [eV]
  SRH : yes ;                      # [cm-3 s-1]
  Auger : yes ;                    # [cm-3 s-1]
  ImpactIonization : yes ;         # [cm-3 s-1]
  Jn_vertex : yes ;                # [A cm-2]
  Jp_vertex : yes ;                # [A cm-2]
  Jn_element : yes ;               # [A cm-2]
  Jp_element : yes ;               # [A cm-2]

}
```

In the `Math` section it is possible to choose the kind of solver and the related tolerance for the different equations. In the case of semiconductor devices other information are added in the `Semiconductor` section for the manage of the Drift-Diffusion Gummel map. Moreover in the same section it is possible to set mobility and R/G models, current density calculation methodologies and ASCII output files for convergence and contact current plots.

```
Math {

  Solver : BCG_NR;
  # defines the kind of solver use for the linear problems. The
    following solver are available:
  # BCG_NR,          bigradient coniugate implemented with
                     Numerical Recipes library
  # BCG_E,           bigradient coniugate implemented with Eigen
                     library
  # GMRES_E,         generalized minimal residual method
                     implemented with Eigen
  # LU_NR,           LU factorization implemented with Numerical
                     Recipes library
  # PartialLU_E,     partial LU factorization implemented with
                     Eigen
  # FullLU_E,        full LU factorization implemented with Eigen
  # QR_E,            QR factorization implemented with Eigen

  Poisson_err = 1e-40 ;
  # sets the tolerance for the iteration solver for Poisson and
    Non Linear Poisson equation

  GM_err = 5e-3 ;
```

```
  # sets the tolerance for the general Gummel map

  GM_max_iter = 5 ;
  # sets the maximum number of iteration for the general Gummel
    map

  Carrier_err = 1e-40 ;
  # sets the tolerance for the Carrier equation

  Temperature_err = 1e-7 ;
  # sets the tolerance for the Thermal equation

  Chemical_err = 1e-6 ;
  # sets the tolerance for the Chemical equations

  GM_chem_err = 5e-3 ;
  # sets the tolerance for the inner Gummel map of the Chemical
    equations

  GM_chem_max_iter = 4
 # sets the maximum number of iteration for the inner Gummel map
     of the Chamical equations

  Mechanical_err = 1e-10
  # sets the tolerance for the Mechanical equation

}


Semiconductor (type : model) {
  model_lib : /home/LIBRERIE/libmodels_lib.so ;
  #  sets the path where find the model library

  J_bulk_method : DDcorretto ;
  # sets the method for current density calculation:
  # SG,            3D edge average SG  ;
  # DD,            standard DD ;
  # DDmodified,    3D upwinding SG;
  # QF,            quasi fermi.

  initguess = 2 ;
  # sets the kind of initial guess performed for the Non Linear
    Poisson problem:
  # 1,  guess computed take into account the ramping procedure,
  # 2,  guess based on the thermal equilibrium hypothesis

  damp : yes ;
  # activates the damping procedure
```

```
  tollNLP = 1.0e-26 ;
  # sets the tolerance for the Non Linear Poisson Newton
    algorithm

  maxitNLP = 100 ;
  # sets the maximum number of iteration for the Non Linear
    Poisson Newton algorithm

  tollGM = 1.0e-24 ;
  # sets the tolerance for the Gummel map

  maxitGM = 40 ;
  # sets the maximum number of iteration for the Gummel map

  showRAMP : yes ;
  # activates the semiconductor print section

  TXTvi : yes ;
  # activates the print of .txt file for the contact current
    plots

  TXTnlp : yes ;
    # activates the print of .txt file for Non Linear Poisson
    convergence plots

  TXTgm : no ;
    # activates the print of .txt file for Gummel map
    convergence plots

   Mobility: Constant , Masetti ;
   # sets the mobility models:
   # Constant,
   # Masetti,
   # Canali

  RG: SRH , II
  # sets the R/G models:
  # SRH
  # Auger
  # II

}
```

Appendix B

FEMOS organization notes

FEMOS code is composed by 28 internal libraries plus the external HDF5-1.8.10 library, with a total number of 50000 rows. The compile session is managed by `makefile`. The output files are visualized with Paraview.

The semiconductor part is integrated in the FEMOS project through five libraries with a total number of 5000 rows. In the following we give a brief description of the contents of each library:

Semiconductor the main object of this library is the `Semiconductor` class, where the fundamental relations between the characterizing quantity of a semiconductor material are implemented.

NLPoisson this library gathered the utility functions to manage the Newton method applied to the Poisson equation.

DD_semiconductor this library includes both the functions assigned to the Continuity equation solver and the post-processing utilities for the evaluation of the current at contacts and inside the device.

ModelManage the main object of this library is the `ModelsManage` class which puts in contact the internal variables of the code with the models (mobility or R/G) loaded by the user.

Models this library stores all the mobility and R/G models. Moreover new models can be added easily without changing other parts of the code.

Bibliografia

[AF03] R. A. Adams and J. J. F. Fournier. *Sobolev Spaces*. Academic Press, 2003.

[AS55] D. N. G. Allen and R. V. Southwell. Relaxation methods applied to determine the motion, in two dimensions, of a viscous fluid past a fixed cylinder. *Quart. J. Mech. Appl. Math.*, 8:129–145, 1955.

[BCC98a] R. E. Bank, W. Coughran, and L. C. Cowsar. The finite volume sharfetter-gummel method for steady convection diffusion equations. *Computing and Visualization in Science*, 1:123–136, 1998.

[BCC98b] R. E. Bank, W. M. Coughran, and L. C. Cowsar. The finite volume Scharfetter-Gummel method for steady convection diffusion equations. *Computing and Visualization in Science*, 1:123–136, 1998.

[BGB00] Streetman Ben G. and Sanjay Banerjee. *Solid State electronic Devices*. New Jersey: Prentice Hall, 2000.

[BRF83] R. E. Bank, D. J. Rose, and W Fichtner. Numerical methods for semiconductor device simulation. *SIAM J. Sci. Stat.*, 4:416–514, 1983.

[Can75] C. Canali. Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature. *IEEE Transactions on Electron Devices*, 22:1045–1047, 1975.

[Chy58] A. G. Chynoweth. Ionization rates for electrons and holes in silicon. *Physical Review*, 109:1537–1540, 1958.

[Deu74] P. Deuflhard. A modified newton method for the solution of ill-conditioned system of nonlinear equations with application to multiple shooting. *Numer. Math*, 22:289–315, 1974.

[GH92] H Goebel and K. Hoffmann. Full dynamic power diode model including temperature behavior for use in circuit simulations. *Proceedings of the 4th International Symposium on Power Semiconductor Devices & ICs*, pages 130–135, 1992.

[GS69] H. K. Gummel and D. Scharfetter. Large-signal analysis of a silicon read diode oscillator. *IEEE Trans. Electron Devices*, pages 64–77, 1969.

[GS06] R. Gusmeroli and A. S. Spinelli. Accurate boundary integrals calculation in semiconductor device simulation. *IEEE transactions on electron devices*, 53:1730–1733, 2006.

[Gum64] H. K. Gummel. A self-consistent iterative scheme for one-dimensional steady state transistor calculations. *Electron Devices*, pages 455–465, 1964.

[HEML00] Thomas J.R. Hughes, Gerald Engel, Luca Mazzei, and Mats G. Larson. The continuous Galerkin method is locally conservative. *Journal of Computational Physics*, 163:467–488, 2000.

[IE83] Babuška I. and Osborn J. E. Generlized finite element methods: Their performance and their relation to mixed methods. *Numer. Anal.*, 20:510–536, 1983.

[Jac84] J. D. Jackson. *Elettrodinamica Classica*. Zanichelli, 1984.

[Jer96] J. W. Jerome. *Analysis of Charge Transport*. Springer, 1996.

[LH80] W. Lochmann and A. Haug. Phonon-assisted auger recombination in si with direct calclation of the overlap integrals. *Solid State Communications*, 35:553–556, 1980.

[Lom88] C. Lombardi. A physically based mobility model for numerical simulation of nonplanar devices. *IEEE Transactions on Computer-Aided Design*, 7:1164–1171, 1988.

[MS83] G. Masetti and M. Severi. Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon. *IEEE Transactions on Electron Devices*, 7:764–769, 1983.

[OR70] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, 1970.

[PC98] M. Putti and C. Cordes. Finite element approximation of the diffusion operator on tetrahedra. *Society for Industrial and Applied Mathematics*, 19:1154–1168, 1998.

[PN] R. F. Pierret and G. W. Neudeck. *Advanced Semiconductor Fundamentals*, volume 6. Prentice Hall.

[PTVF07] William H. Press, Saul A. Teukoisky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes, The Art of Scientific Computing*. Cambridge University Press, New York, 2007.

[Qua08] Alfio Quarteroni. *Modellistica Numerica per Problemi Differenziali*. Springer Italia, Milan, 2008.

[QV08] Alfio Quarteroni and Alberto Valli. *Numerical Approximation of Partial Differential Equations*. Springer, 2008.

[RT91] J. Roberts and J. M. Thomas. Mixed and hybrid methods. *Handbook of Numerical Analysis*, 2:523–633, 1991.

[Sal10] Sandro Salsa. *Equazioni a Derivate Parziali, metodi, modelli e applicazioni*. Springer Italia, Milan, 2010.

[Sde13] *Sentaurus Device User Guide*. Synopsis Inc., 2013.

[Sie84] Selberherr Siegfried. *Analysis and Simulation of Semiconductor Devices*. Springer-Verlag, Wien, 1984.

[vOdM70] R. van Overstraeten and H. de Man. Measurement of the ionization rates in diffused silicon p-n junctions. *Solid-State Electronics*, 13:583–608, 1970.

[VOT83] R. Van Overstraeten and M. S. Tyagi. Minority carrier recombination in heavily-doped silicon. *Solid-State Electronics*, 26:577–597, 1983.

[XZ99] Jinchao Xu and Ludmil Zikatanov. A monotone finite element scheme for convection-diffusion equations. *Mathematics of Computation*, 68:1429–1446, 1999.

[YT09] Taur Yuan and H. Ning Tak. *Fundamentals of Modern VLSI Devices*. Cambridge University Press, 2009.

[ZL12] L. T. Zikatanov and R. D. Lazarov. An exponential fitting scheme for general convection-diffusion equations on tetrahedral meshes. *Computational and Applied Mathematics*, 1:60–69, 2012.

Ringraziamenti

Quando porti a termine un importante progetto vi è un momento in cui ti ritrovi a tirare le somme e ti domandi cosa accadrà dopo. Sei privato dell'unico pensiero che ti ha assillato negli ultimi mesi e per un istante ti senti perso. Il futuro non è nitido e non ti rimane altro che rifugiarti nel passato. All'improvviso ti accorgi che questo momento è il frutto di un'incredibile combinazione di incontri, coincidenze ed eventi che mai avresti potuto pianificare. Un turbinio di emozioni ti assale rimanendo aggrovigliato in gola e lasciandoti un'impalpabile velo di maliconia e al contempo di serentità.

In quel preciso istante ho realizzato che mi stavo laureando, ma per sciogliere quel nodo mancava ancora un passo fondamentale: ringraziare tutte le persone che hanno preso parte al raggiungimento di questo obiettivo. Siete tantissimi e non riuscirei mai a citarvi tutti, ma prometto che farò del mio meglio.

Vorrei cominciare col ringraziare il mio tutor Aurelio che ha saputo guidarmi in ogni momento di questo progetto. Ringrazio Riccardo Sacco i cui saggi consigli sono stati il faro guida di questa tesi. Vorrei poi ringraziare i compagni di lavoro Giada, Paolo, Davide, Riccardo e tutti quelli del gruppo TCAD con i quali ho passato un anno di lavoro sereno e pieno di bei ricordi.

Un ringraziamento speciale va a Ribo, Steve, Cruz e tutti i componenti dei *bulli* con i quali ho condiviso gli ultimi sei anni della mia vita. Ringrazio inoltre tutti i miei compagni di corso, che mi sono stati vicini e in particolare vorrei citare Matteo, col quale ho affrontato gli esami più belli della specialistica. Un grazie di cuore a Monica e a Fede che sono stati in grado di sopportarmi fra le mura strette di CastellAgrumi.

Voglio ringraziare con tutto me stesso mia mamma, sempre pronta a tirarmi su di morale con un sorriso, mio padre, i cui consigli sono stati fonte di ispirazione fin dalla tenera età, e tutta la mia famiglia che mi ha sempre supportato nelle mie scelte.

Infine, se ancora oggi conservo un briciolo di sanità mentale lo devo a Francesca che ha saputo starmi vicino e sostenermi nei momenti più duri.

Grazie a tutti quelli che non ho nominato ma sono stati parte della mia

vita e che volenti o nolenti hanno contribuito a rendermi la persona che sono oggi, grazie grazie e grazie ancora, mi sembra di impazzire ma non riesco proprio a farvi capire quanto vi sono grato.

Tu prova ad avere un mondo nel cuore,
e non riesci ad esprimerlo con le parole ...

(Un Matto, Fabrizio De André)