Investigation of dimensionality effects on capacitorless memory and trench power MOSFET

Silvio Pierro

November 2011

The "Collegio dei docenti" of the Ph.D. in
Fisica e Tecnologie Quantistiche

Coordinator:
Giovanni Falcone

Tutor:
Antonello Sindona
Calogero Pace

acknowledgement
My first thank is to my tutor Prof. Calogero Pace. He bring me advanced knowledge about power electronic devices, has always given me any equipment and instrument which I need to, and gave me the possibility to face this Ph.D.
Another thank to Prof. Felice Crupi and Ing. Gino Giusi that helped me by gaving some research ideas and suggestions.
Another thank goes to STMicroelectronics, in particular to Ing. Giuseppe Consentino and Vincenzo Cilia that followed me in my internship experience to learn power electronic devices.
Last but not the least, I wish to thank my parents, my family and my dolly Manuela. They were a constant support during these three years of PhD.

to my family

Contents

Preface ix

List of Pubblications x

Introduction xi

1 TCAD Simulation setup 1
1.1 Drift-diffusion model 2
1.1.1 Poisson equation 2
1.1.2 Continuity equations 2
1.1.3 Density Of State 4
1.1.4 Mobility model 6
1.1.5 Impact ionization 9
1.1.6 Band To Band Tunneling 9
1.2 Self-consistent Poisson-Shroedinger model 10
1.2.1 Shroedinger equation 10
1.2.2 Statistic 11
1.3 Numerical technique 12
1.3.1 Mesh definition 12
1.3.2 Discretization method 13

I Power MOSFET device analysis 17

2 Power mosfet devices 19
2.1 Power diode 20
2.1.1 Basic p-n physics 21
2.1.2 junction breakdown 24
2.2 Power MOSFET 29
2.2.1 I-V characteristic 30
2.3 Trench Power MOSFET 32
2.3.1 Trench power MOSFET breakdown 32

3 Trench power MOSFET simulations 35
3.1 Simulation setup 36
3.1.1 pn reference structure setup 38
3.1.2 measurement setup 39
3.2 pn structure simulations 42
3.2.1 Parameter analysis 43
3.3 Trench structure simulations 45
3.3.1 Pitch exclusion 46
3.3.2 Drift region analysis 48
3.3.3 Trench's shape analysis 49

4 Novel Breakdown model 55
4.1 Pn breakdown model 56
4.1.1 Model analysis 60
4.2 Trench breakdown model 62
4.2.1 Electrical consideration 64
4.2.2 novel breakdown model 65
4.2.3 model synthesis 68

II memory device simulations 71

5 capacitorless device description 73
5.1 Memory device concepts 74
5.1.1 Dynamic RAM 75
5.2 SOI MOSFET devices 77
5.3 FinFET devices 79
5.4 Bipolar mode description 80
5.4.1 Write operation 82
5.4.2 Hold operation 83
5.4.3 Read operation 84
5.5 transient analysis 86

6 Capacitorless simulations 89
6.1 simulation setup 90
6.1.1 Initial simulation 91
6.1.2 first result 95
6.2 Scaling study of ZRAMcell 96
6.2.1 Longitudinal scaling 97
6.2.2 Transversal scaling 98
6.2.3 scaling considerations 100

A FinFET simulations 103

B List of Symbols **107**

Conclusions **109**

Bibliography **116**

Preface

Investigation of geometrical characteristics of novel device structures is a crucial activity in modern electronics. Due to its higher cost and to higher time to develop, it's not possible to make real device structures for large scale measurements. In order to obtain good results in lower time and with reasonable resources it's better to analyze the device characteristic's variations with a simulation setup and to extract model and trends with this simulation setup, and to use measurements for models check only.

With the simulation approach, since we are able to analyze the internal device electrical characteristics, it's easier to analyze device's electrical phenomena and so extract the model.

In my research activities this methodology has been robustly applied to electronics devices from memory low power device to high power MOSFET devices.

- My PhD study started with simulation study of novel semiconductor devices for memory application, that involves the use of self consistent Poisson-Schrodinger model in order to simulate very thin devices.
- The second year of study focused on power applications, with particular attention to breakdown voltage analysis on MOSFET devices and its literature model.
- The last PhD year started with a collaboration with ST Microelectronics and a small stage period of two months, which focused attention to a new power MOSFET architecture called Trench power MOSFET on which has been made a breakdown characterization as a function of device's characteristics.

.

List of Pubblications

[1] Magnone P., Subramanian V., Pace C., Crupi F., and Pierro S. *"Gate voltage and geometry dependence of the series resistance and of the carrier mobility inn FinFET devices"*, Microelectronic engineering, Vol.**85**, issue 8, Elsevier, 2008.

[2] Giusi G., Alam M., Crupi F., and Pierro S. *"Bipolar Mode Operation and scalaility of Double Gate Capacitorless 1T DRAM Cells"*, IEEE Transactions on electron devices, Vol.**57**, issue 8, pp.1743-1750, 2010.

Conference paper

[3] Pace C., Piacente A., Vescio F., Pierro S., and Sigh B.G. *"An Ultra-Low-Noise Source-Measuring Unit for Semiconductor Device Noise Characterization"*, I2MTC 2010, Austin TX, USA, 3-6 May 2010.

Submitted paper

[4] Pierro S., Cilia V., Cosentino G., and Pace C. *"Novel Avalanche Breakdown Calculation Model for Trench Power MOSFETs"*, Submitted on *Microelectronics Engineering* 2011.

Introduction

The characterization of electronic devices is a crucial step for understanding device properties and their use in applied electronics; in particular it helps to know how to change the structure to obtain better performance, or how to reduce production costs and produce devices more competitive in the market.

The main device characterization technique involves measuring the electrical characteristics of the device according to the different geometrical parameters to extract the dependencies of electrical parameters over geometric variations of device. The production of standard MOS devices allow us to realize devices with a large number of different geometric characteristics on a single wafer with the same process step, for example, you can create arrays with wide range of devices with different gate length, called Larray or variable width, called Warray.

For standard MOS devices it's possible to perform this analysis, but for innovative devices is not always possible to change geometrical parameters in order to realize device arrays because theese parameters are set by the technological processes, or set by wafer characteristics. An example of a limitation due to the characteristics of the wafer is shown in Figure 1a, where we present a SOI MOSFET, which is a MOSFET made on a semiconductor layer bottom isolated by a field oxide, the semiconductor depth is determined during the SOI wafer production, and therefore cannot be changed.

Another example is trench power MOSFET device, the process flow cannot allow to change the depth of drift region, realized by a deposition process, and therefore must be constant throughout the wafer, it is not possible to realize trench lengths arbitrary, since it would require separate etching processes for each device.

For these kind of devices, the realization of device arrays to be studied is virtually impossible, so the previous method of device characterization is inapplicable.

Fortunately, the modern computing power and deep knowledge of semiconductor physics allow you to make any kind of simulation of electronic

devices even in stationary state than in time domain; The software that allow us to perform these simulations are so called Technology Computer Aided device simulators (TCAD), and allow us to simulate any device with doping and size arbitrary chosen and fit for any value of voltage applied to his contacts.

This technique allows us to draw many different technological devices with lengths and to obtain realistic simulations of these devices in a similar way to the measurement setup made on real devices, with the advantage of obtain also the device's electrical characteristics, such as electrostatic potential and carrier concentration, this feature is very useful to explain the macroscopic phenomena of the device.

The use of a TCAD simulator involves the transition from a real device to simulation of the same device, then we need to make a series of checks to ensure the accuracy of the results, in particular, check that the device has been made correctly, that the number of device's simulation points is a good trade-off between good results and simulation times, and that includes all physical phenomena that occur in the device.

This characterization method was applied to evaluate the electrical characteristics of a Trench power MOSFET device, in particular to extract a model taking into account the device breakdown to obtain the breakdown voltage. In addition it was used to characterize a double-gate FinFET memory device in order to check the validity of capacitorless memory and assess what might be future developments in terms of technology scaling.

This thesis make simulations of two different devices, so it was divided into two parts. The first part deals with the simulation of Trench Power MOSFET devices in stationary conditions, and the second part make time-domain simulations of capacitorless memory devices.

- Chapter 1 talk about the main physics phenomena that describes devices under study; In particular the drift-diffusion model and self consistent Schrodinger-Poisson models with its statistical models, mobility and impact generation models. Finally the numerical method used to solve differential equation is described.

- Chapter 2 starts discussion of trench power MOSFET devices by describing the main power MOSFET and power diode features with a particular attention on breakdown characteristics and a comparison with the equivalent low power device.

- Chapter 3 discuss massive simulation on trench power MOSFET and diode equivalent structure to describe the variation on breakdown voltage due to geometrical and doping variation and identify main trends on breakdown voltage versus main geometrical characteristics.

- Chapter 4 presents the physical model used to calculate the breakdown voltage in a power diode and compares the results with the simulations described in chapter 3 for the reference junction structure. Finally it uses the same method to derive the appropriate correction of equipment breakdown voltage in the trench structure using appropriate correction parameters.
- Chapter 5 starts the study of storage devices, with an introduction to memory device, a particular attention to SOI MOSFET devices and its application to use as a volatile Dynamic RAM memory with the explanation of its bipolar mode operation, so called Type II.
- Chapter 6 briefly describes the simulation setup used on the memory storage structure from which begin the discussion on the technology scaling, with particular attention to the length scaling, the resulting problems and solutions to apply.

Chapter 1

TCAD Simulation setup

Physical phenomena in semiconductor devices are very complicated and, depending on applications, are described by partial differential equations of different level of complexity [1]. Coefficients and boundary conditions of equations (such as mobility, generation–recombination rate, material–dependent parameters, interface and contact boundary conditions) can be very complicated and can depend on microscopic physics, the structure of the device, and the applied bias; This knowledge of semiconductor physics and numeracy solution method allows us the use of a device simulation setup to obtain good estimates on the behavior of described devices.
This chapter describes the main characteristics of solid-state physics, with particular attention to the drift-diffusion model is under quasi stationary in the time domain, in addition to the model of self-consistent Poisson-shroedinger. Finally, were considered out important techniques used to obtain numerical solutions of the previous physical models with their main characteristics.

1.1 Drift-diffusion model

Derivations based upon the Boltzmann transport theory have shown that the current densities in the continuity equations may be approximated by a drift-diffusion model [2].
This kind of model is a good approximation to device physics for a very large devices, especially for mosfet with low power dissipation condition and with relatively large area ($> 100nm$). The model includes the Poisson equation and two transport equations for electrons and holes respectively. The model require also a mobility model, a generation model, and a boundary condition setup [3].

1.1.1 Poisson equation

Poissons Equation relates the electrostatic potential to the space charge density:

$$\nabla(\epsilon\nabla\Psi) = \rho \tag{1.1}$$

where Ψ is the electrostatic potential, ϵ is the local permittivity, and ρ is the local space charge density. The reference potential can be defined in various ways. The most used way to define it is the intrinsic Fermi potential Ψ_i which will be defined later. The local space charge density is the sum of contributions from all mobile and fixed charges, including electrons, holes, and ionized impurities. By replacing the local space charge density the previous equation becomes

$$\nabla(\epsilon\nabla\Psi) = q(p - n + N_D - N_A + G - R) - \rho_{trap} \tag{1.2}$$

where q is the elementary electronic charge, n and p are the electron and hole densities, N_D is the concentration of ionized donors, N_A is the concentration of ionized acceptors, G and R are respectively generation and recombination rates, and ρ_{trap} is the charge density contributed by traps and fixed charges.
The electric field is obtained from the gradient of the potential

$$\vec{E} = -\nabla\Psi \tag{1.3}$$

1.1.2 Continuity equations

The continuity equations for electrons and holes are defined by equations:

$$\frac{\partial n}{\partial t} = \frac{1}{q} div\vec{J_n} + G_n - R_n \tag{1.4}$$

$$\frac{\partial p}{\partial t} = \frac{1}{q} div \vec{J_p} + G_p - R_p \tag{1.5}$$

where n and p are the electron and hole concentration, $\vec{J_n}$ and $\vec{J_p}$ are the electron and hole current densities, G_n and G_p are the generation rates for electrons and holes, and R_n and R_p are the recombination rates for electrons and holes. Normally, a drift-diffusion simulation includes both Equations (1.4) and (1.5). In some circumstances, like a unilateral diode, it is sufficient to solve only one carrier continuity equation.

In drift-diffusion model, the current densities are expressed in terms of the quasi-Fermi levels ϕ_n and ϕ_p as:

$$\vec{J_n} = -q\mu_n n \div \phi_n \tag{1.6}$$

$$\vec{J_p} = -q\mu_p p \div \phi_p \tag{1.7}$$

where μ_n and mu_p are the electron and hole mobility. The quasi-Fermi levels are then linked to the carrier concentrations and the potential through the two Boltzmann approximations [4]:

$$n = n_i exp\left[\frac{q\left(\psi - \phi_n\right)}{kT_L}\right] \tag{1.8}$$

$$p = n_i exp\left[\frac{-q\left(\psi - \phi_p\right)}{kT_L}\right] \tag{1.9}$$

where n_i is the effective intrinsic concentration, T_L is the lattice temperature, and k is the Boltzmann constant. The first two quantities has been assumed to be constant for all simulations; and its value are respectively $T_L = 300K$ and $n_i = 9.38e^{19}cm^{-3}$ for Silicon.

These two equations may then be re-written to define the quasi-Fermi potentials:

$$\phi_n = \Psi - \frac{kT_L}{q} ln \frac{n}{n_i} \tag{1.10}$$

$$\phi_p = \Psi + \frac{kT_L}{q} ln \frac{p}{n_i} \tag{1.11}$$

By substituting these equations into the current density expressions (equations (1.4) and (1.5)), the following adapted current relationships are obtained:

$$\vec{J_n} = qD_n\nabla n - qn\mu_n\nabla\Psi - \mu_n n(kT_L\nabla(lnn_i) \tag{1.12}$$

$$\vec{J_p} = qD_p\nabla p - qp\mu_p\nabla\Psi + \mu_p p(kT_L\nabla(lnn_i) \tag{1.13}$$

The final term accounts for the gradient in the effective intrinsic carrier concentration, which takes account of bandgap narrowing effects. Effective electric fields change from equation (1.3) to

$$\vec{E_n} = -\nabla \left(\Psi + \frac{kT_L}{q} lnn_i \right) \tag{1.14}$$

$$\vec{E_p} = -\nabla \left(\Psi - \frac{kT_L}{q} lnn_i \right) \tag{1.15}$$

which then allows the more conventional formulation of drift-diffusion equations to be written as

$$\vec{J_n} = qn\mu_n \vec{E_n} + qD_n \nabla n \tag{1.16}$$

$$\vec{J_n} = qp\mu_p \vec{E_p} - qD_p \nabla p \tag{1.17}$$

It should be noted that this derivation of the drift-diffusion model has tacitly assumed that the Einstein relationship holds. In the case of Boltzmann statistics [5] this corresponds to:

$$D_n = \frac{kT_L}{q} \mu_n \tag{1.18}$$

$$D_p = \frac{kT_L}{q} \mu_p \tag{1.19}$$

If Fermi-Dirac statistics are assumed for electrons, Diffusivity for electron becomes

$$D_n = \frac{\left(\frac{kT_L}{q} \mu_n \right) F_{1/2} \left\{ \frac{1}{kT_L} \left[\epsilon_{Fn} - \epsilon_C \right] \right\}}{F_{-1/2} \left\{ \frac{1}{kT_L} \left[\epsilon_{Fn} - \epsilon_C \right] \right\}} \tag{1.20}$$

where F_α is the Fermi-Dirac integral of order α and $\epsilon_{Fn} = -q\phi_n$. An analogous expression is used for holes with Fermi-Dirac statistics; The Fermi-Dirac statistic is used instead of Boltzmann statistic when high doping concentrations are used in semiconductor ($10^{19} cm^{-3}$).

The Drift-diffusion model is so realized by Poisson equation (eq. (1.3)) and the two equations for electron and hole concentrations (eq. (1.12) and (1.13)).

1.1.3 Density Of State

In order to evaluate Density of State for electrons and holes in semiconductor devices, a distribution probability as energy function must be used; Identical discussion must been made for electrons and holes, for simplicity we report discussion for electrons only.

Electrons in thermal equilibrium at temperature T_L with a semiconductor lattice obey Fermi-Dirac statistics. That is the probability $f(\epsilon)$ that an available electron state with energy ϵ is occupied by an electron is:

$$f(\epsilon) = \frac{1}{1 + exp\left[\frac{\epsilon - E_F}{kT_L}\right]} \tag{1.21}$$

where E_F is a spatially independent reference energy known as the Fermi level and k is Boltzmanns constant. In the limit that, $\epsilon - E_F \gg kT_L$, Equation (1.21) can be approximated as:

$$f(\epsilon) = exp\left[\frac{E_F - \epsilon}{kT_L}\right] \tag{1.22}$$

Statistics based on the use of Equation (1.22) are referred to as Boltzmann statistics. The use of Boltzmann statistics instead of Fermi-Dirac statistics makes subsequent calculations much simpler. The use of Boltzmann statistics is normally justified in semiconductor device theory, but Fermi-Dirac statistics are necessary to account for certain properties of very highly doped materials.

Integrating the Fermi-Dirac statistics over a parabolic density of states in the conduction and valence bands, whose energy minimum is located at energies E_C and E_V respectively, yields the following expressions for the electron and hole concentrations:

$$n = N_C f_{1/2}\left(\frac{E_F - E_C}{kT_L}\right) \tag{1.23}$$

$$p = N_V f_{1/2}\left(\frac{E_V - E_F}{kT_L}\right) \tag{1.24}$$

where $F(1/2)(\eta)$ is referred to as the Fermi-Dirac integral of order 1/2. If Equation (1.22) is a good approximation, then Equations (1.23) and (1.24) can be simplified to

$$n = N_C exp\left(\frac{E_F - E_C}{kT_L}\right) \tag{1.25}$$

$$p = N_V exp\left(\frac{E_V - E_F}{kT_L}\right) \tag{1.26}$$

which are referred to as the Boltzmann approximation. N_C and N_V are referred to as the effective density of states for electrons and holes and are given by:

$$N_C = 2\left(\frac{2\pi m_n kT_L{}^{3/2}}{h^2}\right) \tag{1.27}$$

$$N_V = 2\left(\frac{2\pi m_p k T_L}{h^2}^{3/2}\right) \tag{1.28}$$

where h is the Planck constant, and m_n and m_p are effective mass for electron and holes respectively.

1.1.4 Mobility model

Free electrons present in conduction band are substantially free particles that tend to move in a Brownian motion with a medium speed v_{th}. The kinetic energy of the electrons can be expressed by the relation:

$$\frac{1}{2}m_n v_{th}^2 = \frac{3}{2}kT_L \tag{1.29}$$

At room temperature, the mean speed for Silicon is approximately $10^7 cm/s$ [6].

Since the effects of various scattering phenomena, like lattice vibrations (phonons), impurity ions, other carriers, surfaces, and other material imperfections, the macroscopic mobility will be introduced by the transport equations. This mobility is therefore function of the local electric field, lattice temperature, doping concentration, and so on [7].

Mobility modeling is normally divided into:

(i) low field behavior;

(ii) high field behavior;

(iii) bulk semiconductor regions;

(iv) inversion layers.

The low electric field behavior has carriers almost in equilibrium with the lattice and the mobility has a characteristic low-field value that is commonly denoted by the symbol μ_{n0},μ_{p0}. The value of this mobility is dependent upon phonon and impurity scattering. Both of which act to decrease the low field mobility.

The high electric field behavior shows that the carrier mobility declines with electric field because the carriers that gain energy can take part in a wider range of scattering processes. The mean drift velocity no longer increases linearly with increasing electric field, but rises more slowly. Eventually, the velocity doesnt increase any more with increasing field but saturates at a constant velocity. This constant velocity is commonly denoted by the symbol v_{sat}. Impurity scattering is relatively insignificant for energetic carriers, and so v_{sat} is primarily a function of the lattice temperature.

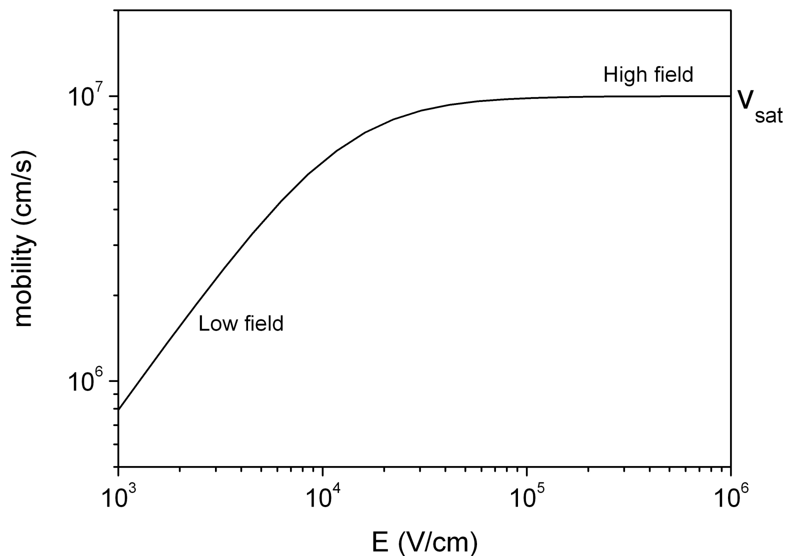


Figure 1.1: Electron mobility as function of electric field at room temperature.

Modeling mobility in bulk material involves: (i) characterizing μ_{n0} and μ_{p0} as a function of doping and lattice temperature, (ii) characterizing v_{sat} as a function of lattice temperature, and (iii) describing the transition between the low field mobility and saturated velocity regions.

Modeling carrier mobility in inversion layers introduces additional complications. Carriers in inversion layers are subject to surface scattering, extreme carrier-carrier scattering, and quantum mechanical size quantization effects. These effects must be accounted for in order to perform accurate simulation of MOS devices. The transverse electric field is often used as a parameter that indicates the strength of inversion layer phenomena.

To use all the mobility model, the different mobility contributions are combined according to the following scheme: different mobility contributions are combined following Mathiessens rule [8]:

$$\frac{1}{\mu} = \frac{1}{\mu_1} + \frac{1}{\mu_2} + \frac{1}{\mu_3} \cdots + \frac{1}{\mu_N} \tag{1.30}$$

Low field mobility model

The low field mobility parameter taking into account the effects of lattice scattering, impurity scattering (with screening from charged carriers), carrier-carrier scattering, and impurity clustering effects at high concentration.

The total mobility described by Matthiessens rule add two mobility components [9]

$$\frac{1}{\mu_n} = \frac{1}{\mu_{nL}} + \frac{1}{\mu_{nDAP}} \tag{1.31}$$

μ_n is the total low field electron and hole mobility, μ_{nL} is the electron mobility due to lattice scattering, μ_{nDAP} the electron and hole mobility due to donor (D), acceptor (A), screening (P) and carrier-carrier scattering. A similar mobility equation could be evaluated for electrons.

The lattice scattering component, μ_{nL} is given as:

$$\mu_{nL} = v_{nmax}\left(\frac{300}{TL}\right)^{\gamma} \tag{1.32}$$

where v_{nmax} is the maximum electron velocity that for silicon can be evaluated as $1417(cm^2/V \cdot s)$, and γ is a correction factor of 2.2 ; the equivalent value for holes are $v_{hmax} = 470.5(cm^2/V \cdot s)$ and $\gamma = 2.24$.

The impurity-carrier scattering component of the total electron mobility is given by

$$\mu_{nDAP} = \mu_{N,n}\frac{N_{nsc}}{N_{nsc,eff}} + \mu_{nc}\left(\frac{n+p}{N_{nsc,eff}}\right) \tag{1.33}$$

where $\mu_{N,n}$ is the impurity scatter component, $N_{nsc} = N_D + N_A + p$, $N_{nsc,eff} = N_D + \alpha N_A + \beta p$ and μ_{nc} is the carrier-carrier scattering component.

Inversion layer mobility model

The inversion layer mobility model must take into account for mobility reduction due to MOS inversion layer, the transverse field, doping dependent and temperature dependent parts of the mobility are given by three components that are combined using Matthiessens rule [10]. These components are μ_{AC}, and μ_{sr}. The first component, is the surface mobility limited by scattering with acoustic phonons:

$$\mu_{nAC} = \frac{B_n}{E_{\perp}^{\alpha AC}} + \frac{N^{\alpha AC}}{T_L E_{\perp}^{\beta sr}} \tag{1.34}$$

the second component is the surface roughness factor, and is given by:

$$\mu_{nsr} = \frac{N^{\alpha sr}}{T_L E_{\perp}^{\beta sr}} \tag{1.35}$$

where T_L is the lattice temperature, $E_{\perp}$ is the perpendicular electric field, and N is the total doping concentration.

A_n	$7.03e^5\ cm^{-1}$
A_p	$6.73e^5\ cm^{-1}$
B_n	$1.23e^6\ V/cm$
B_p	$1.69e^6\ V/cm$
β_n	1
β_p	1

Table 1.1: Numerical parameters of impact ionization formula

1.1.5 Impact ionization

Impact ionization phenomena modeling is essential in the simulation object of this thesis, because this phenomena is responsible of avalanche breakdown in power semiconductors, and is is responsible of carrier increase in memory devices; Therefore we must choose the model to use with extreme accuracy. The general impact ionization process is described by [11]

$$G = \alpha_n \left|\vec{J_n}\right| + \alpha_p \left|\vec{J_p}\right| \tag{1.36}$$

where, G is the local generation rate of electron-hole pairs, α_n and α_p are the ionization coefficient for electrons and holes, and J_n and J_p are their current densities. The ionization coefficient represents the number of electron-hole pairs generated by a carrier per unit distance travelled.

The generation rate can be calculated by using the following formula

$$\alpha_n = A_n exp\left[-\left(\frac{B_n}{E}\right)^{\beta n}\right] \tag{1.37}$$

$$\alpha_p = A_p exp\left[-\left(\frac{B_p}{E}\right)^{\beta p}\right] \tag{1.38}$$

where, E is the electric field in the direction of current flow at their structure's position and A_n, A_p, B_n, B_p, β_n, and β_p are parameters defined by the model with its valued shown on table 1.1.

1.1.6 Band To Band Tunneling

If a sufficiently high electric field exists within a device local band bending may be sufficient to allow electrons to tunnel, by internal field emission, from the valence band into the conduction band. An additional electron is therefore generated in the conduction band and a hole in the valence band.

This generation mechanism is implemented into the right-hand side of the continuity equations. The tunneling generation rate is [12] as:

$$G_{BTBT} = BB_A E^{BB\gamma} exp\left(-\frac{BB_B}{E}\right) \quad (1.39)$$

where E is the magnitude of local electric field and $BB_A = 9.6615e^{18}$, $BB_B = 3e^7 V/cm$, and $BB\gamma = 2$ are constant parameters. This band-to-band tunneling model uses the electric field value at each node to give a Generation rate at that point due to the tunneling. In reality, the tunneling process is non-local and is necessary to take into account the spatial profile of the energy bands. It is also necessary to take into account the spatial separation of the electrons generated in the conduction band from the holes generated in the valence band. A model for this process assumes that the tunneling can be modeled as being one-dimensional in nature so that it must be calculated using a special rectangular mesh superimposed over and coupled to the normal mesh. This mesh needs to include the junction region of interest and the direction of the Band-to-Band tunneling must be specified. We must adjust the extent and resolution of the mesh to obtain the best results.

1.2 Self-consistent Poisson-Shroedinger model

The Drift-diffusion model is a good semiconductor physic approximation for most application, but for very thin devices (¡10nm) the semiconductor band structure becomes different from the simple case of infinite semiconductor, and quantum confinement phenomena could be present [13]. In order to take into account for quantum phenomena, a enhanced model must be used.

1.2.1 Shroedinger equation

The solution of Schrodinger's Equation gives a quantized description of the density of states in the presence of quantum mechanical confining potential variations. The calculation of the quantized density of states relies upon a solution of Schrödingers Equation.

$$\left(-\frac{\hbar}{2}\frac{\partial}{\partial z}\frac{1}{m_l(z)}\frac{\partial}{\partial z} + E_C(z) + \delta E_V\right)\Psi_{il} = E_{il}\Psi_{il} \quad (1.40)$$

$$\left(-\frac{\hbar}{2}\frac{\partial}{\partial z}\frac{1}{m_t(z)}\frac{\partial}{\partial z} + E_C(z) + \Delta E_V\right)\Psi_{it} = E_{it}\Psi_{it} \quad (1.41)$$

where E_{it} is the i-th eigenvalue energy for transverse mass, E_{il} is that for the longitudinal mass, Ψ_{it} is the transverse eigenfunction, Ψ_{il} is the longitudinal eigenfunction, δE_V is the energy offset for the ladder, and E_C is the

conduction band energy. This expression is for electrons but an equivalent expression exists for holes.

The Schrödinger equation is solved over a finite domain $[Z_-; Z_+]$. The endpoints of this domain is defined by the boundary condition:

$$\frac{\Psi'_{il}}{\Psi_{il}} = \pm \frac{\sqrt{2m_{il} \left| E_{il} - E_C \right|}}{\hbar} \tag{1.42}$$

is applied, where for the upper sign, all position-dependent functions are taken at Z_+ and, for the lower sign, at Z_+.

1.2.2 Statistic

Using Fermi-Dirac statistics, the discrete nature of the quantized density of states reduces the integral over energy to a sum over bound state energies. The expression for the electron concentration then becomes

$$\begin{aligned} n(x) = 2\frac{kT_L}{\pi h^2} m_t(x) \sum_{i=0}^{\infty} \left| \Psi_{il}(x) \right|^2 ln \left[1 + exp\left(-\frac{E_{il} - E_F}{kT_L} \right) \right] + \\ 4\frac{kT_L}{\pi h^2} \sqrt{m_l(x) m_t(x)} \sum_{i=0}^{\infty} \left| \Psi_{it}(x) \right|^2 ln \left[1 + exp\left(-\frac{E_{it} - E_F}{kT_L} \right) \right] \end{aligned} \tag{1.43}$$

Once the solution of Schrodingers Equation is taken, carrier concentrations calculated from Equation (1.43) are substituted into the charge part of Poisson's Equation. The potential derived from solution of Poisson's Equation is substituted back into Schrodinger's Equation. This solution process (alternating between Schrodingers and Poissons equations) continues until convergence and a self-consistent solution of Schrodinger's and Poisson's equations is reached.

1.3 Numerical technique

The physical model previously described provide us a set of partial differential equations (PDEs) to be solved. this set of PDEs must be solved on a set of spatpoints defined as a mesh that must describe the entire device. A discretized method will transform the original equations to a discrete nonlinear algebraic system that well approximate the system. This set of approximation, the mesh, and the PDEs set determine the nonlinear numerical problem to be solved.

The non-linear algebraic system is solved using an iterative procedure that refines successive estimates of the solution. Iteration continues until the corrections are small enough to satisfy convergence criteria, or until it is clear that the procedure is not going to converge. The non-linear iteration procedure starts from an initial guess. The corrections are calculated by solving linearized versions of the problem. The linear subproblems are solved by using direct techniques or iteratively.

1.3.1 Mesh definition

The first step in numerical resolution is the definition of a mesh that cover the entire device space; This mesh must contain enough points to provide the required accuracy, at the same time it's necessary to design a mesh that do not contain too many unnecessary points that impair efficiency. In order to create a good trade-off between this two rules a finite element method has been used.

The finite element method (FEM) is a numerical technique for finding approximate solutions of partial differential equations (PDE) as well as integral equations. The solution approach is based either on eliminating the differential equation completely, or rendering the PDE into an approximating system of ordinary differential equations.

The division of space into triangular spaces to create a non-uniform mesh that helps increase the number of mesh points in regions of space with high values of electric field over the mesh and relax in the less important maintaining good accuracy of the results.

To obtain good results in terms of convergence, efficiency and robustness is essential to define a good mesh that must respect certain rules [14]:

- Must contain a number of points that ensures a good accuracy.
- They must remove unnecessary points that increase simulation time.
- In order to obtain a good accuracy and robustness, it's necessary to reduce at a minimum value the number of obtuse triangles and the triangles long and thin.

- The triangle size must change smoothly in transition from a region where very small triangles must be used to a region where the use of much larger triangles.

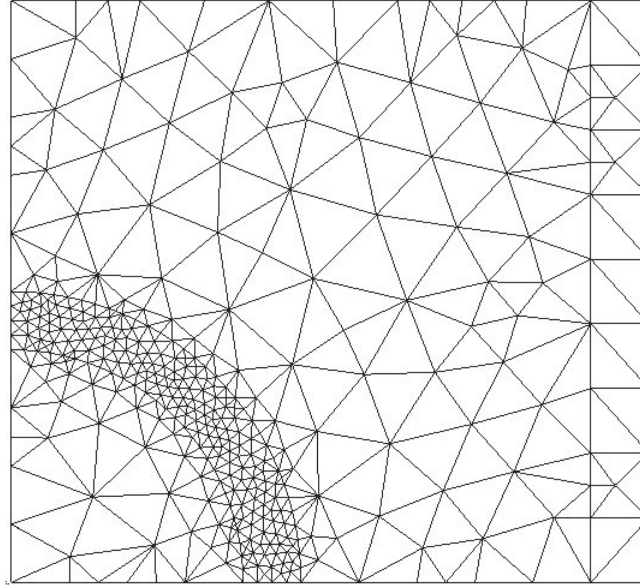


Figure 1.2: Example of 2D finite element mesh, intersections as mesh points.

1.3.2 Discretization method

Numerically solve a differential equation means to calculate approximate values of unknown functions in a sufficiently large number of points in space. so you can display the desired solution without permit detailed analysis. In particular, it is important to find an approximate method of approximation of differential operator.

Using the Euler discretization method can transform the system of partial differential equations into a set of nonlinear equations to be expressed in matrix form, which transforms the partial differential operator derived from the values of the unknown neighbors in space. Under these conditions, the Poisson equation is transformed into the following form

$$\|M\|\,\Phi = R(n,p) + \tilde{b} \tag{1.44}$$

Where $\|M\|$ is the matrix form of second order derivative operator approximation, Φ is the unknown vector, R is the space charge density, and $\tilde{b}$ is the border condition vector. The derivative operator for a one-dimensional problem with $x \in (0, a)$ can be presented as

$$\|M\| = \begin{pmatrix} \alpha_1+\beta_1 & -\beta_1 & 0 & 0 & \dots & 0 \\ \alpha_2 & \alpha_2+\beta_2 & -\beta_2 & 0 & & 0 \\ 0 & \alpha_3 & \alpha_3+\beta_3 & -\beta_3 & \ddots & 0 \\ \vdots & & \ddots & \ddots & \ddots & -\beta_{m-1} \\ 0 & 0 & 0 & \dots & -\alpha_m & \alpha_m+\beta_m \end{pmatrix} \quad (1.45)$$

where

$$\alpha_i = \frac{2\epsilon}{h_i/(h_i+h_{i+1})}, \ \beta_i = \frac{2\epsilon}{h_{i+1}(h_i+h_{i+1})} \quad (1.46)$$

and

$$\Phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_m \end{pmatrix}, \ R = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_m \end{pmatrix}, \quad \tilde{b} = \begin{pmatrix} \alpha_1\phi_D(0) \\ 0 \\ \vdots \\ \beta_m\phi_D(a) \end{pmatrix} \quad (1.47)$$

where h_i is the distance in space between i and $i-1$ mesh point, ρ_i is the space charge density evaluated at mesh point, and ϕ_D is the border condition potential at contacts.

Using the same considerations for the continuity equations we can obtain the following system matrix [15]

$$\begin{cases} \|M\|\,\Phi = \frac{1}{2\epsilon}R(n,p) + \frac{1}{2}\tilde{b} \\ \|M_n\|\,(\Phi)n = GR(n,p) + \tilde{b}_n \\ \|M_p\|\,(\Phi)p = GR(n,p) + \tilde{b}_p \end{cases} \quad (1.48)$$

Now it is important to define an iterative numerical solution to obtain the vectors Φ, n, and p. [15]

1. We assign an initial guess for the vectors of the density of charge, in fact we assign them an arbitrarily fixed but plausible values, which we denote $n^{(0)}$, $p^{(0)}$.

2. At this point begins the real iteration map. By induction suppose we have determined the vectors $n^{(k)}$, $p^{(k)}$, and $\Psi^{(k)}$. We solve the Poisson equation discretized using the values for the density $n^{(k)}$, $p^{(k)}$, and denote the solution with $(k+1)$. briefly write

$$2\epsilon\,\|M\|\,\Psi^{(k+1)} = R\left(n^{(k)}, n^{(k)}\right) + \epsilon b \quad (1.49)$$

3. We solve the equation of continuity for electrons in eq (1.48), after have eliminated the nonlinearity appropriately using the known values. in practice, if we denote by $n^{(k+1)}$ the value of the new density of electron vector, it is obtained by solving the linear equation

$$\|M\|_n \left((Phi^{(k+1)})\right) n^{(k+1)} = GR\left(n^{(k+1)}, n^{(k)}, p^{(k)}\right) + b_n \tag{1.50}$$

and by replacing

$$GR\left(n^{(k+1)}, n^{(k)}, p^{(k)}\right) = \|M\|_{n,GR}^{(k)}, n^{(k+1)} + b_{GR}^{(k)}$$

we obtain

$$\left[\|M\|_n^{(k+1)} + \|M\|_{n,GR}^{(k)}\right] n^{(k+1)} = b_{GR}^{(k)} + b_n \tag{1.51}$$

4. Similarly it is the continuity equation for holes in eq (1.48). More precisely, denote by $p^{(k+1)}$ the solution of linear problem.

$$\left[\|M\|_p^{(k+1)} + \|M\|_{p,GR}^{(k+1)}\right] p^{(k+1)} = b_{GR}^{(k)} + b_n \tag{1.52}$$

5. At this step, we have $n^{(k+1)}$, $p^{(k+1)}$, and $\Psi^{(k+1)}$ from the previous one. Now we compare two kind of errors: the absolute errors, defined as the maximum variation from the previous solution; and the right hand side error. If it appears that these errors are below a certain tolerance value we can accept vectors as a solution of the coupled system. Otherwise return to step 2.

This iterative method is successful if it ends with the convergence condition is reached, where the error tends to diverge, or are unable to obtain the solution using a reasonable number of iterations, then the simulation stops without getting a new solution, it will be necessary to modify the boundary conditions to obtain an intermediate solution, or modify the structure of the mesh to improve the accuracy of the simulation. [16]

Part I

Power MOSFET device analysis

Chapter 2

Power mosfet devices

Power electronics have been reached a crucial importance in most applications, such as automotive and aerospace applications. In order to achieve good results in terms of performance and reliability lots of studies have been made to get better electrical characteristics like on resistance, thermal dissipation, switching speed, and breakdown voltage.

Power semiconductor devices are more complicated than low-power corresponding devices in terms of operational characteristics and structure more complicated. The added complexity is due to modification made starting from low power devices made to make it ready for high power applications. The main properties that a power device must have [17] are:

1. Large breakdown voltage
2. Low on resistivity and low on voltage
3. fast switching time
4. Large power dissipation capability

In all device types, there is a trade-off between breakdown voltage and on-state resistivity. Increasing breakdown voltage will cause a higher blocking voltage for power device, but consequently will produce an increasing in on-state resistivity that causes higher power loss in device on state.

This trade off mean that there is not only one device type useful for all applications. The requirement of the specific application must be matched to the capabilities of the available devices. This will require clever and innovative design approaches.

In the next chapters, the main power devices used in this thesis work are presented and described, with particular attention to electrical characteristics and differences between p-n and MOSFET structures.

2.1 Power diode

The basic semiconductor device is made with *p-n* junction. In low power application this device is formed by the union of a n-type semiconductor with a *p-type* one in the same crystal. this kind of junction can be made by diffusing acceptor impurities into an *n-type* semiconductor crystal, or by diffusing donor impurities into a *p-type* semiconductor instead.

An important characteristic of *p-n* junction is how the doping changes from *n-type* to *p-type* region. The simplest one is defined as the abrupt change in doping level at junction position, this represent the so called abrupt junction. A more gradually change in doping density is the linearly graded junction that make linearly change in doping level near *p-n* junction. A real junction instead can be expressed with a gaussian doping profile diffusion. This doping profile came classical profile generated by a diffusion process at high temperature $(900 : 1150°C)$ [18].

Power diode device are fabricated on a heavily doped *n-type* substrate on top of which is grown a lightly doped n^- epitaxial layer of specified thickness. Finally the *p-n* junction is formed by diffusing a heavily doped *p-type* region to form the anode of the diode. A typical power diode is shown on figure 2.1 cross section. The cross section area will vary according to the amount of total current the device is designed to carry. The area A will include the Z dimension that will be very large compared with other dimensions. [19, 20]

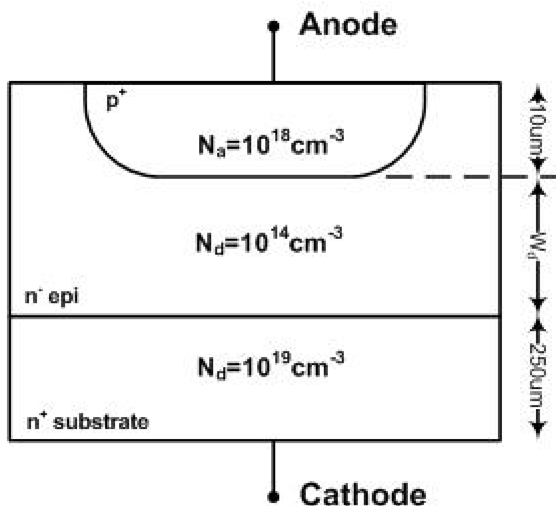


Figure 2.1: p-n-junction power mosfet diode cross section with its dimensional characteristics and doping levels.

The *i-v* characteristic of the diode is shown on figure 2.2. In forward region current grows linearly instead of a exponentially normal low power diode. This trend is due to large currents in power diode that create ohmic drops that mask the exponential *i-v* characteristic [21].

In reverse bias zone a small negative leakage current will flow until the reverse breakdown voltage V_Z is reached. When breakdown is reached, the voltage will remain essentially constant while the current increase of several orders of magnitudes, and its grow is limited only by the external circuit limit. The combination of a large current and voltage at breakdown condition leads to a very high power dissipation that cause the destruction of the device. This is the main reason that induce to use very carefully devices in that operation mode, and often it cannot be used.

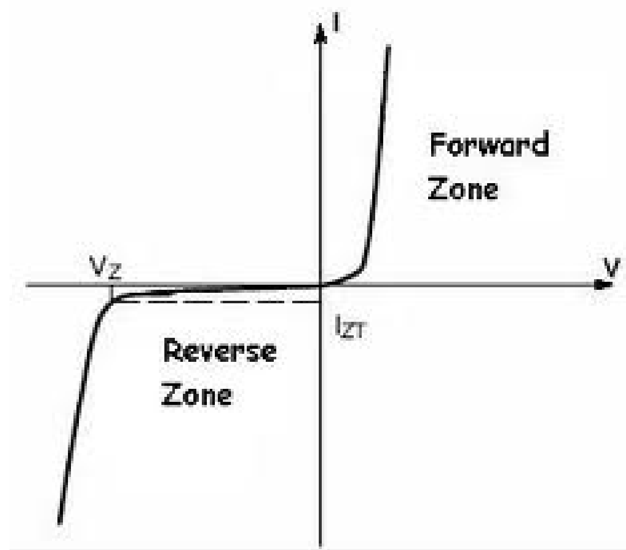


Figure 2.2: The *I-V* characteristic of a *p-n* junction diode.

2.1.1 Basic p-n physics

As first analysis, an abrupt *p-n* junction is analyzed. The abrupt junction show an immediate change in doping concentration along x direction, changing doping impurities from acceptor N_A to donor N_D. A unbalanced junction can be defined if $N_A >> N_D$. Starting from one dimensional Poisson equation we obtain

$$-\frac{\partial^2 V}{\partial x^2} = \frac{\partial \xi}{\partial x} = \frac{q}{\epsilon_s} = \frac{\rho(x)}{\epsilon_s} \left[p(x) - n(x) + N_D^+ - N_A^- \right] \tag{2.1}$$

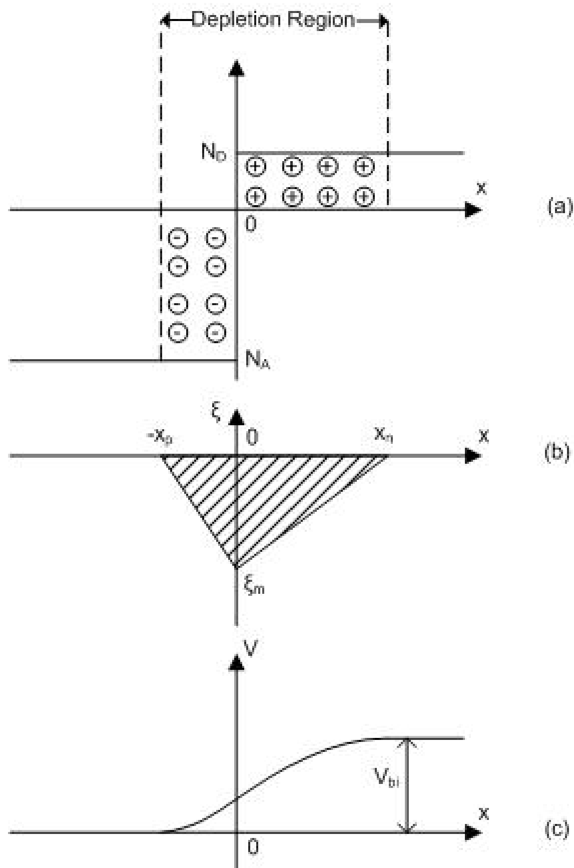


Figure 2.3: One-dimensional visualization of p-n junction in thermal equilibrium. (a) Space-charge distribution. The dashed lines indicate the majority-carrier distribution tails. (b) Electric field distribution. (c) potential variation with distance x, where V_{bi} is the built-in potential.

that becomes

$$-\frac{\partial^2 V}{\partial x^2} \approx \frac{\rho(x)}{\epsilon_s} N_A^- \quad \text{for} \quad -x_p \leq x < 0 \tag{2.2}$$

$$-\frac{\partial^2 V}{\partial x^2} \approx \frac{\rho(x)}{\epsilon_s} N_D^+ \quad \text{for} \quad 0 < x \leq x_n \tag{2.3}$$

The electric field is obtained by integrated equations (2.1) as shown in figure 2.3b

$$\xi(x) = -\frac{qN_A(x + x_p)}{\epsilon_s} \quad \text{for} \quad -x_p \leq x < 0 \tag{2.4}$$

$$\xi(x) = -\xi_m + \frac{qN_D x}{\epsilon_s} = \frac{qN_D}{\epsilon_s} \quad \text{for} \quad 0 < x \leq x_n \tag{2.5}$$

where ξ_m is the maximum electric field that exists at x=0 and is given by

$$|\xi_m| = \frac{qN_D x_n}{\epsilon_s} = \frac{qN_A x_p}{\epsilon_s} \tag{2.6}$$

Integrating equation (2.1) once again, Figure 2.3c gives the potential distribution V(x) and the built-in potential V_{bi}

$$V(x) = \xi_m \left(x + \frac{x^2}{2W} \right) \tag{2.7}$$

$$V_{bi} = \frac{1}{2}\xi_m W = \frac{1}{2}\xi_m(x_n + x_p) \tag{2.8}$$

where W is the total depletion width. Eliminating ξ_m from equation (2.8) yields

$$W = \sqrt{\frac{2\epsilon_s}{q} \left(\frac{N_A + N_D}{N_A N_D} \right) V_{bi}} \tag{2.9}$$

for a two-sided abrupt junction. For a one-sided unbalanced abrupt junction, figure (2.9) reduces to

$$W = \sqrt{\frac{2\epsilon_s V_{bi}}{qN_B}} \tag{2.10}$$

where $N_B = N_D$ or N_A depending on whether $N_A >> N_D$ or vice versa.

$$\mathbf{J_n} = q\mu_n n\xi + qD_n \nabla n \tag{2.11}$$

Where $\mathbf{J_n}$ is the electron current density, q is the elementary charge, μ_n is the electron mobility and D_n is the electron diffusivity. The same equation

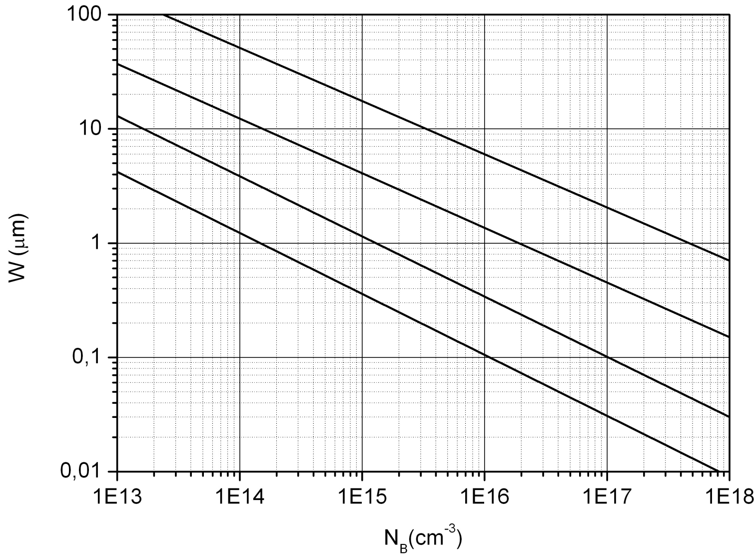


Figure 2.4: depletion-layer width as a function of doping for one-sided abrupt junction in Si for various voltage

can be written for hole current by replacing the electron concentration n, the electron mobility and diffusivity with the holes one. The equation consist of the drift component caused by the field and the diffusion component caused by the carrier concentration gradient. For a one dimensional case equation (2.11) reduce to

$$\mathbf{J_n} = q\mu_n n\xi + qD_n \frac{\partial n}{\partial x} = q\mu_n \left(n\xi + \frac{kT}{q} \frac{\partial n}{\partial x} \right) \tag{2.12}$$

At equilibrium condition, when no current flows and no voltage is applied, density current inside semiconductor is equal to zero.

Applying a reverse bias to a p-n junction a increase in the depletion regions, and a reverse low value current will be present. This reverse bias current can be approximately given by the sum of the diffusion components in the neutral region and the generation current in the depletion region [3].

2.1.2 junction breakdown

The rapid increase in current at reverse-bias voltage over p-n junction is called reverse breakdown, or avalanche breakdown. Operation of p-n junction in breakdown must be avoided because the product of a large voltage and a high current leads to very high power dissipation that will quickly destroy the junction if it's not reduced. The main mechanism that causes

breakdown are three: thermal instability, tunneling effect, and avalanche multiplication [3, 22].

Thermal instability

Thermal instability is a small effect in semiconductor with relatively small bandgap like Ge lattice. In the condition of heat dissipation due to reverse current at high reverse voltage, the junction temperature will increase its value. this temperature increase, increases the reverse current in comparison with its value at lower temperature. Because of the heat dissipation at high reverse voltage, the characteristic shows a negative differential resistance. In this case the diode will be destroyed unless some special measures such as a large series limiting resistor is used. The thermal instability is temperature dependent, in particular breakdown occurs at lower voltage as increasing room temperature.

In Silicon devices, thermal instability can be neglected due to larger bandgap that reduce drastically its effect.

Band-To-Band tunneling

The tunneling effect is a one-dimensional effect related to a square energy barrier. As can be seen in Figure 2.5, in reverse voltage condition the increase in current is generated by electrons in valence band that jump horizontally in conduction band causing a hole-electron pair generation that increases current. This phenomena is related to barrier Height E_0 and thickness W, the quantum-mechanical transmission probability T_t is given by

$$T_t = \left[1 + \frac{E_0^2 sinh^2 kW}{4E(E_0 - E)}\right]^{-1} \tag{2.13}$$

with

$$k = \sqrt{\frac{2m(E_0 - E)}{\hbar^2}} \tag{2.14}$$

where E is the carrier's energy. The probability decreases monotonically with decreasing E. When $kW \gg 1$, the probability becomes

$$T_t \approx \frac{16E(E_0 - E)}{E_0^2} exp(-2kW) \tag{2.15}$$

Finally, the tunneling current density for a p-n junction is given by

$$J_t = \frac{\sqrt{2m^*} q^3 \xi V}{4\pi^2 \hbar E_g^{1/2}} exp - \frac{q\sqrt{2m^*} E_g^{3/2}}{3q\xi\hbar} \tag{2.16}$$

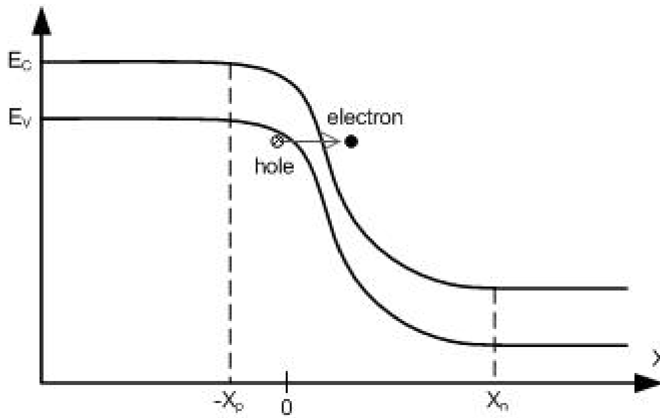


Figure 2.5: tunneling effect description on band structure

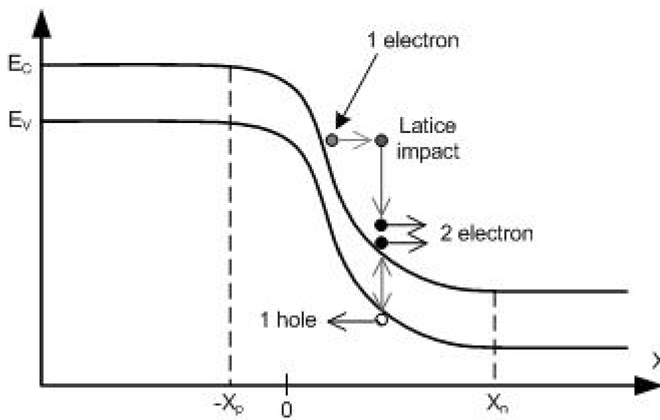


Figure 2.6: Avalanche breakdown effect on a single electron; impact with hole-electron pair generation.

where ξ is the electric field at the junction, E_g the bandgap, V is the applied voltage, and m^* the effective mass.

When the electric filed approaches 10^6 V/cm in Silicon, significant current begins to flow by means of the band-to-band tunneling process. To obtain such a high field, the junction must have relatively high impurity concentrations on both the p and n sides.

Avalanche Breakdown

Avalanche multiplication, or impact ionization, is the most important mechanism in junction breakdown, since the avalanche breakdown voltage imposes an upper limit on the reverse bias for most devices, like on the Drain

voltage of MOSFETs, and the reverse bias on diodes.

Figure 2.6 describe impact ionization for an electron in a pn junction in reverse bias condition; In particular, the free electron in the conduction band moves to the right due to electric field and increase its energy, when a collision of free electron with lattice occurs, the electron lose its exceeding energy returning to the lower limit of the conduction band. If this transferred energy is sufficiently high, it will generate an electron-hole pair, which may repeat the same sequence [23].

In order to characterize this phenomena in p-n structure, we first derive the basic ionization integral which determines the breakdown condition. Assumes that a curent I_{p0} is incident at the left-side of the depletion region with width W. If the electric field in the depletion region is high enough that electron-hole pairs are generated by the impact ionization process, the hole current I_p will increase with distance through the depletion region and reaches a value $M_p I_{p0}$ at W. Similarly, the electron current I_n will increase from $x = W$ to $x = 0$. The total current ($I = I_p + I_n$) is constant at steady state. The incremental hole current at x equals the number of electron-hole pairs generated per second in the distance dx:

$$d\left(\frac{I_p}{q}\right) = \left(\frac{I_p}{q}\right)(\alpha_p dx) + \left(\frac{I_n}{q}\right)(\alpha_n dx) \tag{2.17}$$

or

$$\frac{dI_p}{dx} - (\alpha_p - \alpha_n) I_p = \alpha_n I \tag{2.18}$$

Where α_n and α_p are respectively electron and hole ionization rates. The solution of equation (2.18) with boundary condition
that $I = I_p(W) = M_p I_{p0}$ is given by:

$$I_p(x) = \frac{I\left\{\frac{1}{M_p} + \int_0^x \alpha_n exp\left[-\int_0^x (\alpha_p - \alpha_n)\, dx'\right] dx\right\}}{exp\left[-\int_0^x (\alpha_p - \alpha_n) dx'\right]} \tag{2.19}$$

where M_p is the multiplication factor of holes and is defined as

$$M_p = \frac{I_p(W)}{I_p(0)} \tag{2.20}$$

Equation (2.19) can be written as

$$1 - \frac{1}{M_p} = \int_0^W \alpha_p exp\left[-\int_0^x (\alpha_p - \alpha_n) dx'\right] dx \tag{2.21}$$

The avalanche breakdown is defined as the voltage where M_p approaches infinity. Hence the breakdown condition is given by the ionization integral

$$\int_0^W \alpha_p exp\left[-\int_0^x (\alpha_p - \alpha_n)dx'\right]dx = 1 \tag{2.22}$$

If the avalanche process is initiated by electrons instead of holes, the ionization integral is given by:

$$\int_0^W \alpha_n exp\left[-\int_x^W (\alpha_n - \alpha_p)dx'\right]dx = 1 \tag{2.23}$$

Equations (2.22) and (2.23) are equivalent; this means that breakdown condition depends only on what is happening within the depletion region and not on the carriers that initiate the avalanche process.

For semiconductor with equal ionization rates ($\alpha_n = \alpha_p = \alpha$), equations (2.22) and (2.23) reduce to simple expression

$$\int_0^W \alpha dx = 1 \tag{2.24}$$

From the breakdown conditions described above and the field dependence of the ionization rates, the breakdown voltages, maximum electric filed, and depletion layer width can be calculated. As discussed previously, the electric field and potential in the depletion layer are determined from the solutions of Poisson's equation. Depletion layer that satisfy Equation (2.22) can be obtained numerically using an iteration method, like Runge-Kutta method. With known boundaries we obtain

$$BV = \frac{\xi_m W}{2} = \frac{\epsilon_s \xi_m^2}{2q}(N_B)^{-1} \tag{2.25}$$

for a one-sided abrupt junction where N_B is the ionized background impurity concentration of the lightly doped side, ϵ_s is the silicon dielectric constant, and ξ_m the maximum electric field.

By using breakdown voltage value obtained by formula (2.25) an solving Poisson equation it can be obtained the depletion layer width as

$$W_{BV} = 2.67e^{10}(N_B)^{-7/8} \tag{2.26}$$

and the maximum electrical field value

$$\xi_{BV} = \left(\frac{N_B}{2.41e^{-29}}\right)^{1/8} \tag{2.27}$$

The mechanism of breakdown for Silicon junctions with breakdown voltages less than about '$4E_g/q$ is found to be due to the tunneling effect. For junction with breakdown voltages higher than $6E_g/q$, the main mechanism is

the avalanche multiplication. At voltages $4E_G/q < V < 6E_G/q$, the breakdown is due to a mixture of both avalanche and band-to-band tunneling. Since the energy bandgap Eg in Silicon decrease with increasing temperature, the breakdown voltage in that semiconductor due to tunneling effect has a negative temperature coefficient; that is, the voltage decreases with increasing temperature [23].

2.2 Power MOSFET

Power MOSFET device differs too from its equivalent low power device for the same reasons expressed before. The power MOSFET structure is no more specular from source-to-drain view, since drain region must be able to support very high blocking voltage; Perhaps this structure has been modified and now is similar to power diode one. A classical power MOSFET is shown in figure 2.7. Starting from a n^+ bottom substrate a n^+ region is grown to form Drain drift region. After what a *p-type* doping implantation is made to form body region, and finally a *n-type* doping implantation inside body region is made to create source region. After doping formation, gate oxide and contact are growth on top of device, so structure is ready.

The doping in the two n^+ end layers, labeled source and drain in figure 2.7, is approximately the same in both layer and is quite large, typically $10^{19}cm^{-3}$. The *p-type* middle layer named body is the region where the channel is established between source and drain in a small layer of interface near gate oxide. This region is typically doped at $10^{16}cm^{-3}$. The n^- layer is the drain drift region and is typically doped at $10^{14} - 10^{15}cm^{-3}$. This drift region determines the breakdown voltage of the device [24].

If no voltage is applied to gate, there is no contact between source and drain, so no current can flow between the two contacts because one of *p-n* junctions (either the body-source junction or the drain-body junction) will be reverse biased by either polarity of applied voltage between the drain and source. However, an application of a voltage that biases the gate positive with respect to source will convert the silicon surface beneath the gate oxide into an *n-type* channel, connecting the source to the drain and allowing the flow of an appreciable current. The current flow vertically from drain contact at bottom device to gate oxide center, pass across body region horizontally at gate-to-body interface, and reach source contact.

The gate thickness, gate width, and the number of devices connected electrically in parallel (with every single contact connected in parallel) are important in determining how much current will flow for a given gate-to-source voltage. The previously described structure is the so called VDMOS, meaning Vertical Diffused MOSFET. The name crudely describe the fabri-

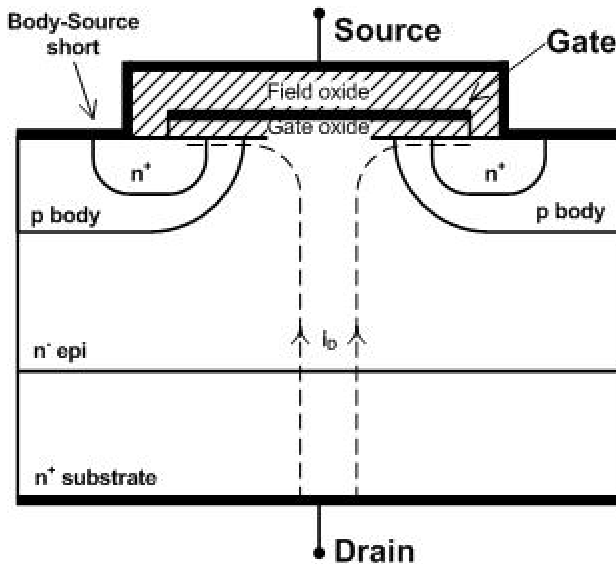


Figure 2.7: Power MOSFET vertical cross section. A single MOS device is often composed of many thousands of cells connected in parallel that replicates this structure.

cation sequence of the device.

2.2.1 *I-V* characteristic

The power MOSFET is a tree terminal device where the gate contact intended as input is electrically isolated and is used to control the flow of current between the other two output terminals source and drain. Figure 2.8 show the output characteristic, drain current i_D as a function of drain-to-source voltage V_{DS} with gate-to-source voltage V_{GS} as a parameter.

In power electronic applications, the MOSFET is used as a switch to control the flow of power to de load towards gate contact [25]. In these applications the MOSFET transverses the i_D-V_{GS} characteristics from cutoff trough the active region to the linear region as the device turns on and back again when it turns off. The cutoff, active, and linear region are shown on figure 2.8.

The MOSFET is in cutoff when the gate-source voltage is less than threshold voltage V_{th}, a typical value for threshold voltage in power MOSFET applications for an *n-type* is about one volt. The device is an open circuit and must hold off the power supply voltage applied to the circuit. When the device is driven by a large gate-source voltage, it is driven into the linear region where the drain-source voltage $\mathrm{V}_D S$ is small. In this region

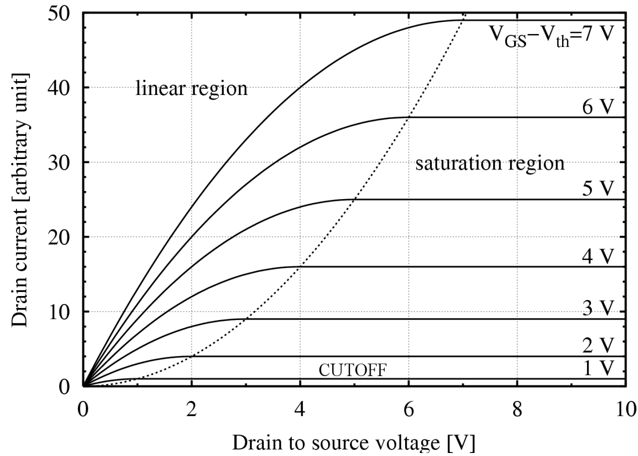


Figure 2.8: Drain current to drain voltage characteristic of an n-channel MOSFET at various gate voltages.

the power dissipation can be kept within reasonable bounds by minimizing V_{DS} even if the drain current is fairly large. The mosfet is in the linear region when

$$v_{GS} - V_{th} > v_{DS} > 0 \tag{2.28}$$

In the active region the drain current is independent from the drain-source voltage and depends only to gate-source voltage. Can be said that in this operational region, drain current is saturated. A simple first order theory predicts that in the active region the drain current is given by

$$i_D = K(v_{GS} - V_{th})^2 \tag{2.29}$$

Where K is a constant that depends on device geometry. This simple model don't take into account of some parasitic effects like:

- bjt parasitic effect related to n-p-n junctions respectively from source-body-drain.
- the presence of two devices side by side produce a jfet effect in drift drain region under the gate due to body-to drain depletion region.
- the high structure's parallelization cause a source increase in on state resistance.

This characteristics are related to VMOSFET, some other power MOSFET structures can reduce such effects.

2.3 Trench Power MOSFET

The Trench power MOSFET, or UMOSFET, represent a variation from power VDMOSFET design that have the advantage of higher cell density but is more difficult to manufacture than the planar device [26]. Trench power MOSFET device structure is presented in figure 2.9. This new device topology put the two gate channels vertically at the interface between *p-body* and gate oxide regions.

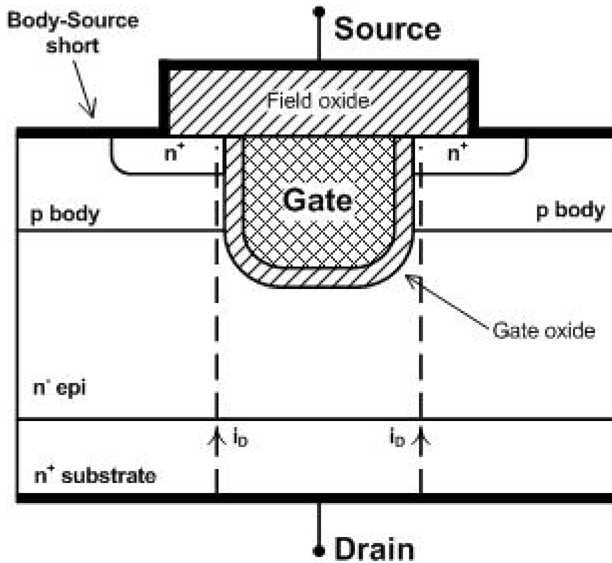


Figure 2.9: Trench power MOSFET device cross section

In fact, trench power MOSFET device fabrication start with an epitaxial growth of *n-epi* drain layer, followed by a trench formation with a etching process will be made. this process must be very accurate to ensure a low roughness on its border because a higher roughness can increase the on resistivity. After trench formation the gate oxide oxidation and poly gate deposition will be made. Subsequently a *p-type* doping implantation is made to form body region, and a *n-type* doping implantation inside body region near trench edges are made to create source regions. Finally a field oxide is put on top of the trench to electrically isolate gate to source contact and the three contacts are realized.

2.3.1 Trench power MOSFET breakdown

MOSFET devices have two voltage ratings that should not be exceeded: V_{GSmax} and BV_{DSS}. The maximum allowable gate-source voltage V_{GSmax}

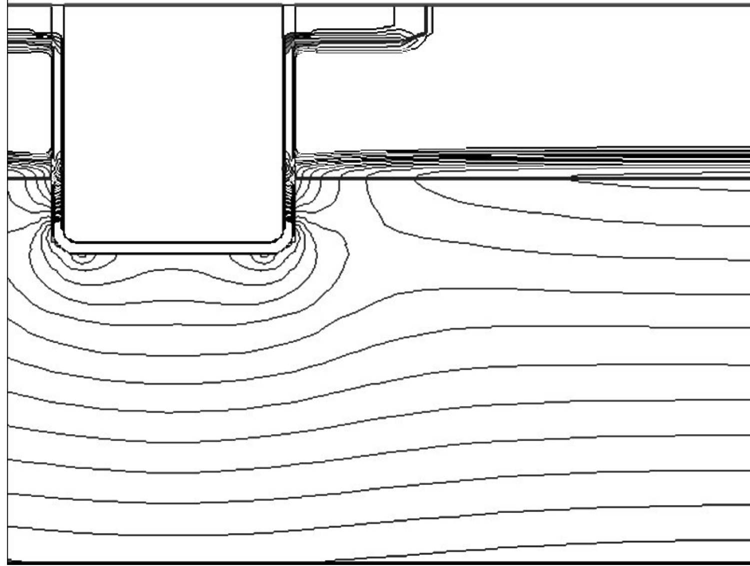


Figure 2.10: Electric field contour lines in Trench power MOSFET's cross section at breakdown condition

is determined by the requirement that the gate oxide not be broken down by large electric fields. Good-quality thermally grown SiO_2 breaks down at electric field values on the order of $5e^6 - 10e^6$ v/cm. This means that a gate oxide 100nm thick can theorically support a gate-source voltage of 50-100 V. For a Trench device the voltage value will decrease by a factor $2\pi r$ due tu the trench bottom rounding gate that create a tip effect increasing local electric field value. The maximum allowable drain-source voltage BV_{DSS} is the largest voltage the MOSFET can hold off without avalanche breakdown of the drain-body p-n junction. Large values of breakdown voltage are achieved by the use of the lightly doped drain drift region. The lightly doped drift region is used to contain the depletion layer of the reverse biased drain-body junction. The length of the drift region is determined by the desired breakdown voltage rating. The electric field maximum value on Trench device is no more on bulk-drain junction, as for a simple junction, the maximum electric field shift nearing trench's border. Figure 2.10 show electrical field cross section in a Trench power MOSFET device in breakdown condition BV_{DSS}. As can be noted, the maximum electric field is placed on trench rounding shape neighborhood [27].

Chapter 3

Trench power MOSFET simulations

As described above, using a device simulator allows to obtain experimental results on structures not yet been realized or to evaluate the internal features that otherwise would be impossible to analyze. This chapter uses the numerical techniques described above for simulations of Trench power MOSFET devices. After an initial setup and testing of the model setup and simulation were carried out a series of simulations to identify the main parameters that influence the breakdown voltage and identify the cause. Furthermore, the result was compared in terms of breakdown of the trench structure with a pn junction structure reference.

3.1 Simulation setup

To obtain good simulation results it is essential that every aspect of the simulation is well defined. In particular:

- Set up a physical model that includes the main phenomena that may occur

- The structure must be built properly in terms of size and doping

- The device mesh must show a good trade-off between a mesh with enough mesh points to allow a good degree of approximation, and one with the fewest points to help you get quick results

- Te numerical resolution method must show an adequate level of approximation of all the electrical parameters to be evaluated.

Finally, it is good practice to check that the simulation result is correct, trying to explain the physics of the device or by comparing the results obtained with a simulation of a test device. In order to obtain a high current value with low resistivity, a power MOSFET device is made by the union of a large number of MOSFET devices shorted in parallel, the single devices are very long and occupy all the die width. [26]

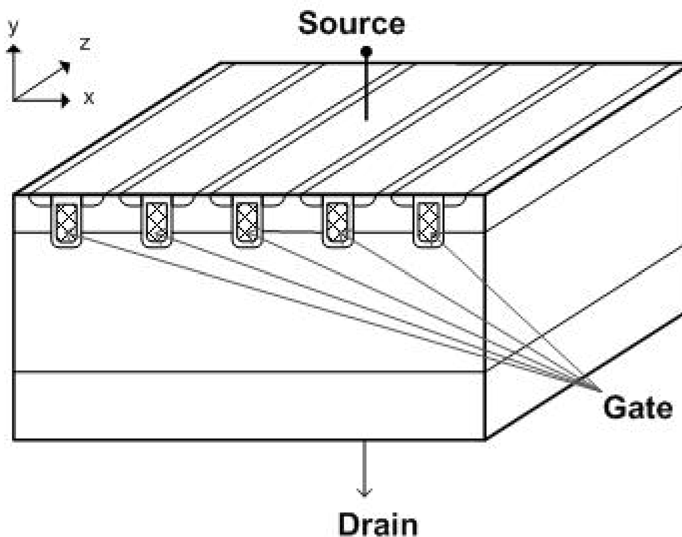


Figure 3.1: Schematic design of classical Trench power MOSFET device

The structure does not change its characteristics in the Z direction, this allow to create two-dimensional simulations assuming that along the

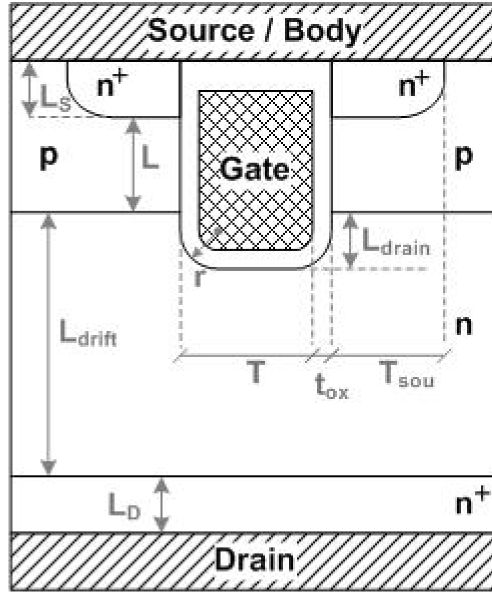


Figure 3.2: Trench power MOSFET 2D vertical cross section with its main geometrical parameters

Z direction the electrical characteristics does not change; This is a good approximation since power devices are very large [20, 28].

Trench power MOSFET structure is described in figure 3.2. The gate is located at structure's top-center, the bulk and source regions are shorted in a single contact (Source), and drain is placed at structure's bottom. Drain doping have different doping concentration from top to bottom direction; in particular there are two different regions; a drift region with a lower doping concentration and a fixed length on top, and a substrate region with higher doping level to ensure lower resistivity. Drift region doping level is not constant, a Gaussian distribution near the bottom interface with n+ region is present. This Gaussian distribution came from diffusion process that move doping atoms from substrate to drift region. The diffusion length is calculated as the distance between bulk-drain p-n junction and the depth where doping reaches the value of 2 times drift doping value, this distance is called for simplicity drift length.

The trench's shape exceed the bulk region and penetrate inside drain drift region, this length is called Ltrench and is evaluated as the vertical distance between bulk-to-drain p-n junction and the lower point in trench oxide-semiconductor interface.

The structure's main characteristics default values are shown on table 3.1. The structure was simulated using the drift diffusion model, with the

Name	Value	Description
T	$750nm$	Trench width
t_{ox}	$100nm$	Oxide thickness
t_{sou}	$250nm$	Source pad width
t_{Tot}	$7\mu m$	Domain simulation length
r	$200nm$	Trench's bottom radius
L_S	$75nm$	Source pad depth
L	$1\mu m$	Gate length
L_{drain}	$300nm$	Trench penetration into drain region
L_{drift}	$4.5\mu m$	Drain lighly doped region depth
L_D	$0.5\mu m$	Drain strongly doped region depth
L_{Tot}	$6.1\mu m$	Total simulation height
N_S	$1e^{20}cm^{-3}$	Source pad n-type doping concentration
N_G	$1e^{20}cm^{-3}$	Gate polysiliocn n-type doping conc.
N_B	$2e^{18}cm^{-3}$	Body p-type doping concentration
N_{drift}	$3e^{15}cm^{-3}$	Drain n-type lighly doped doping conc.
N_D	$1e^{20}cm^{-3}$	Drain n-type strongly doped doping conc.

Table 3.1: Default geometrical and doping concentration values

addition of impact ionization model to properly simulate the breakdown phenomenon. Figure 3.3 shows mesh description of default Trench MOS structure, the following mesh is made by the union of two different finite element method meshes: one with square mesh points, in areas with uniform and triangular elements, and one with a FEM mesh with triangular shape where the density of points increases by grade. In particular, the mesh is thickened near the region of oxide trench, regions in which we find a triangular mesh, the mesh in the remaining structure maintains the rectangular shape which allows a more simple numerical analysis. [29] Near the body-drain junction, the mesh is thickened in the y direction only, so it is not necessary to use a triangular mesh to increase the density of points. The number of simulation points to the following structure is 7088, with 7002 elements, of whom 460 triangular and 6542 rectangular, the solution to a problem of this type requires about 0.45-5.21sec.

3.1.1 pn reference structure setup

The trench structure's breakdown phenomena is somehow linked to the breakdown of a p-n structure where the p-type region is the bulk region and the n-type region is the drain region; The junction is unbalanced, since doping of the p region is much greater than the doping of the region n,

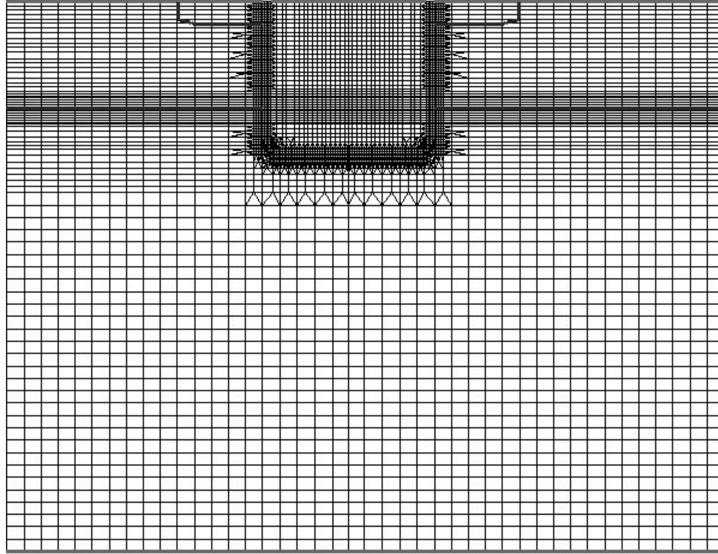


Figure 3.3: Trench structure mesh description

perhaps the drain region is the only one to be discussed. [30]

The new structure to be observed must delete the source pads and the gate's Trench, losing one of the terminals and becoming a simple pn junction. The gate terminal is then deleted, the source/body contact becomes the anode, while the Drain terminal will become the cathode. This change in structure allows you to observe a structure in which the only phenomena related to the breakdown is due to unilateral pn junction, the effect of trench structure has been completely eliminated.

The mesh of the pn reference structure is greatly simplified compared to the Trench mesh structure, because of the absence of Trench structure, it is not necessary to increase the number of simulation points in that area. Since along the X axis there are no variations of electrical parameters, the simulation can use a coarser mesh in that direction. The mesh structure is rectangular and contains 2608 points with 2432 triangles, for a simulation time of 0.05-0.3s for every step.

3.1.2 measurement setup

To obtain good results in terms of breakdown is important to use appropriate physical model. The current density during the blocking of the device allow us to use the simple drift diffusion model, in which are activated this mobility models: high field mobility, to manage mobility in drift region, the model-dependent transverse electric field, to well manage channel mobility,

and model-dependent doping. It is also triggered the Lombardi avalanche generation model [4] that well simulate the avalanche breakdown effect.

To obtain the device's breakdown voltage a voltage swing on drain terminal is made, by increasing the drain voltage until it reaches a maximum current that is uniquely recognized as current breakdown [31, 32], the resulting voltage value at the swing's end represents the breakdown voltage (BVDD).

Figure 3.4 shows two different simulations of drain current as a function of drain voltage with and without impact ionization model. The simulation without impact ionization shows a low current, exclusively related to the junction reverse current which tends to remain roughly constant with increasing reverse voltage. The simulation model that includes impact ionization displays almost identical current for voltages lower than breakdown voltage, this is due to the low influence of the generated charge with respect of minority carriers that are responsible of reverse current.

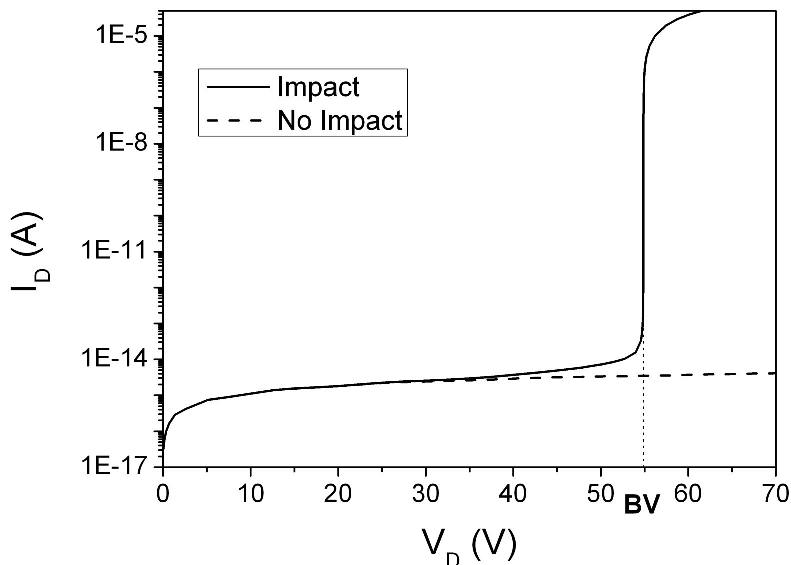


Figure 3.4: Reverse drain current versus drain voltage for simulations with and without impact ionization model.

Approaching the breakdown voltage the contribution in the current generation of impact increases its value, up to become dominant on the reverse current, causing a slight increase in drain current. Once breakdown value is reached, the impact ionization reaches its critical value increasing substantially the current value by several orders of magnitude.

The simulation drain voltage swing simulation of a trench MOS device with the impact ionization model previously described requires the simu-

lation of 60-80 point steps, each with a different boundary conditions, and requires from 7 to 10 minutes in total, depending on stability, convergence, and robustness.

measurements comparison

In order to obtain good results in terms of real simulation, the reference structure has been modulated to be near to a real device structure realized by STMicroelectronics(TM). This device has been shipped as a single Trench device with 3 contacts on the interstitial space between DIEs on a silicon wafer. The measurement setup make a $I_D - V_D$ curve with a keithley 4200 SCS. The $I_D - V_D$ simulation and measurement comparison are shown on figure 3.5. Measurement device characteristic shows a similar trend to the simulation one with impact ionization model, the current increase is the same order of magnitude at breakdown condition, the increase is very sharp but less offended than the simulation, due to parasitic series resistance, and impact ionization effect on drain current show a higher effect on low voltage value. [32,33] The following comparison allow us to use the reference

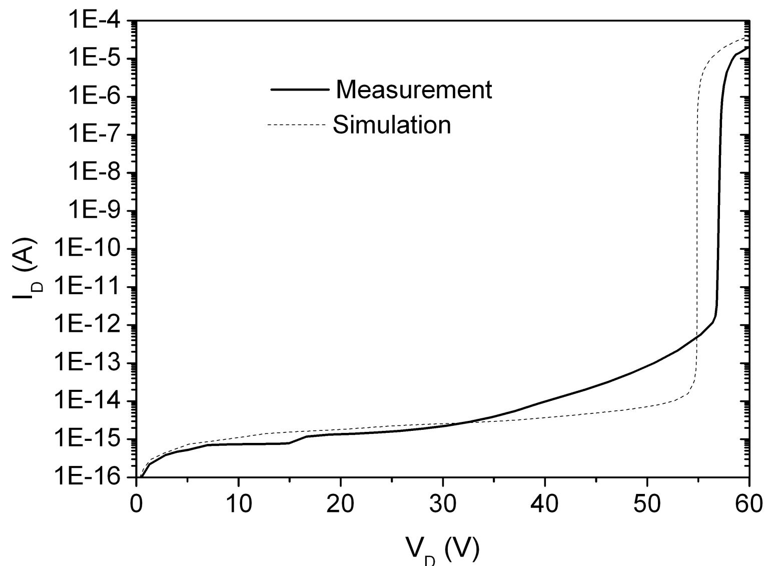


Figure 3.5: ID-VD comparison between simulation with impact ionization mdel and device measurement

structure as a starting point for subsequent simulations allowing to consider the following results as a good representation of reality.

3.2 pn structure simulations

To better study Trench devices and understand the breakdown phenomena is useful to make a comparison between the trench and an equivalent device with a good physical breakdown model in literature. [34] Figure 3.6 shows the target device after removal of the gate contact, the wells of source and gate oxide. This new structure is a pn junction sided with the n-doped less than the region p. For convenience, the Source contact will be referred as anode and Drain as cathode.

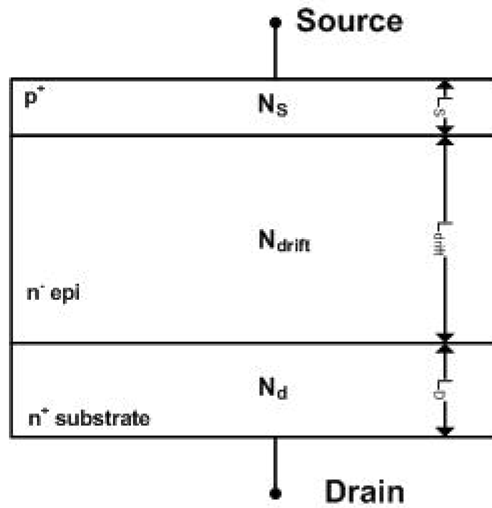


Figure 3.6: Two dimensional cross section p-n reference structure

Name	Value	Description
T_{Tot}	$2\mu m$	Domain simulation length
L_S	$1.075\mu m$	Source p-type depth
L_{drift}	$4.5\mu m$	Drain lighly doped region depth
L_D	$0.5\mu m$	Drain strongly doped region depth
N_S	$1e^{20}cm^{-3}$	Source pad n-type doping concentration
N_{drift}	$3e^{15}cm^{-3}$	Drain n-type lighly doped doping conc.
N_D	$1e^{20}cm^{-3}$	Drain n-type strongly doped doping conc.

Table 3.2: Default geometrical and doping concentration values for pn reference structure

As can be seen from Table 3.2, the reference structure characteristics maintain the same values as trench structure, this in order to realize a

device with the same characteristics. The only difference is the reduction of total device width T_{tot} to $2\mu m$ only, this reduction is possible because the device does not change of any kind along the y direction, so the width change does not affect much in terms of Breakdown analysis.

3.2.1 Parameter analysis

Once set the process to obtain the breakdown voltage, and validated the model, it is possible to perform simulations by varying the structure to evaluate the variation of breakdown voltage and sensitivity to parameters. It's required a comprehensive study of variation for all the parameters of the structure previously described, but the analysis made in Chapter 2.1.2 allow us to exclude certain parameters; In particular is not expected due to fluctuations in the width of the device (T_{tot}) , because the device does not change along the x axis. Since junction is unilateral, the source region depth and doping parameters will not affect the breakdown [22]. Finally, the heavily doped drain region is seen as a very long and heavily doped region, hence also the changes in this region will not affect the analysis made.

The previous considerations allow us to made an analysis on reference structure only in terms of variation in drift region length and drift region doping concentration. This analysis could be useful to make a comparison between trench and p-n reference structure By the use of same Ldrift and Ndrift for both structures.

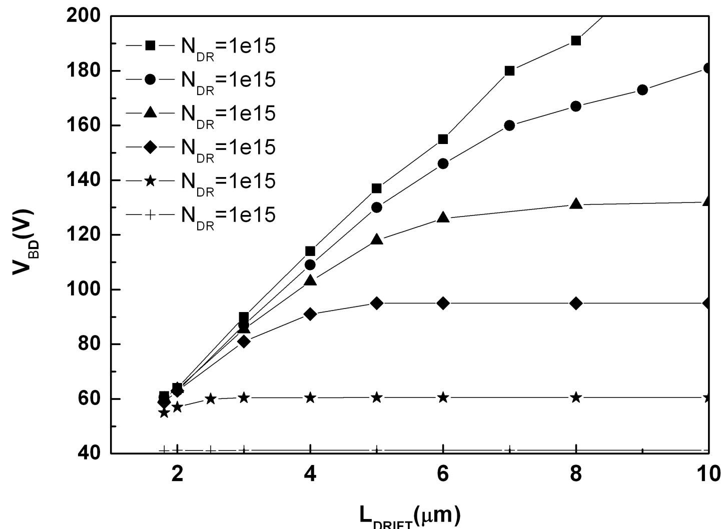


Figure 3.7: Breakdown voltage value as a function of drift length for various drift region doping for a pn reference structure.

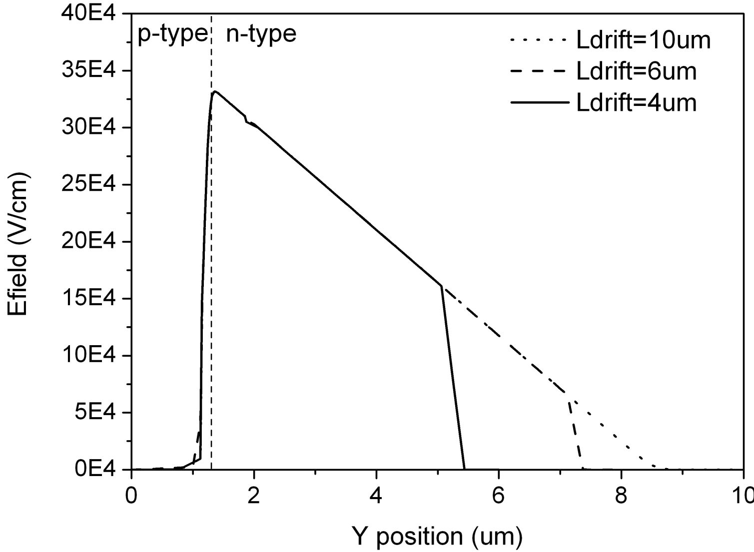


Figure 3.8: Horizontal electric field p-n structure cross section at different drift length in breakdown condition.

Figure 3.7 shows breakdown voltage in p-n reference structure for different drift region doping. For low drift region length value, breakdown value is directly related to drift region length [35], when drift region length reaches a limit value, the breakdown voltage becomes constant. This limit value will decrease as increasing drift region doping. This breakdown dependence to L_{drift} is related to electrical field shape in drift region. Figure 3.8 shows electrical field cross sections in p-n structures with different drift region lengths at breakdown condition for $N_{drift} = 3e^{15}cm^{-3}$. The maximum electrical field value is not affected by drift length and is always placed on p-n junction, the electric field shape at junction proximity is not exactly triangular and does not have an edge because the junction doping profile is not abrupt, but Gaussian, so the total doping at junction proximity is greatly reduced in both n and p sides and the smoothing produces a reduction of the slope at this simulation points. Since the depletion region lies inside drift region the electrical field profile maintain its triangular shape, when drift region length becomes smaller than depletion region maximum value, the depletion region extends into n^+ drain region, causing a variation in electrical field trend that change shape from triangular to trapezoidal with a slope variation corresponding to $n^- - n^+$ drain border and consequently a reduction in breakdown voltage. The analysis made above shows that for a given doping profile in the drift region there is a characteristic length of the depletion region in the drift region, which we will call this pa-

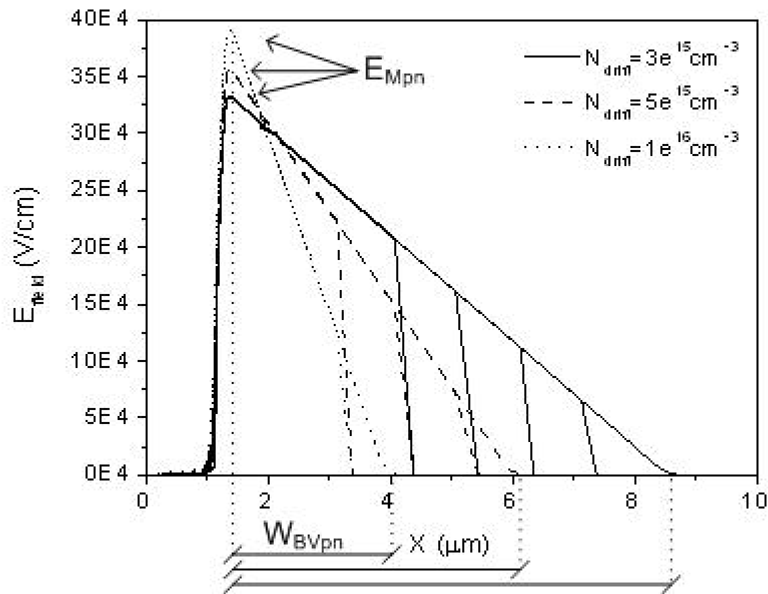


Figure 3.9: Horizontal electric field p-n structure cross section at different drift length and doping at breakdown condition.

rameter W_{BVpn}, allows us to distinguish devices in two different types: the first drift region of very large ($L_{drift} > W_{BVpn}$) that allow you to include the entire depletion region in the drift region, features that allow you to link the voltage exclusively to the doping profile of breakdown. The second drift regions with closer ($L_{drift} \geq W_{BVpn}$), which instead predict that the depletion region extends into the drain region, generating a breakdown voltage dependence from L_{drift}. Figure 3.9 shows the electric field profile for devices with different drift region doping at breakdown condition. From the previous graph can be noted that the maximum electric field value (E_{Mpn}) is related only to drift region doping, particularly is directly proportional to N_{drift}, and consequently also the maximum depletion region width (W_{BVpn}) depends only on the drift region doping in a inversely proportional manner.

3.3 Trench structure simulations

The analysis in the reference structure is simplified by the removal of the Trench structure, which reduces the analysis to only two parameters to be made. In Trench device the parameter number cannot be reduced in the same way as the complexity of the structure does not allow it. Moving to a much greater number of parameters than 2, the number of device

parameter combinations become quickly very high, making impossible even the simulation computation that the results interpretation. Consequently, it is passed to the analysis of individual parameters to assess what were the main parameter that influence the device breakdown. The analysis tends to eliminate from the study those parameters that do not affect the breakdown and to separate the effects in different categories, namely the shape of the trench, the drain region, and the maximum size of the device. As in the case of the pn structure, even in the case of a trench device the main features relating solely to the drift region, hence all the parameters related to the region of Body and Source will not be treated, in addition to the drain region doped more. This severely limits the number of parameters to be studied that are 7: T, t_{ox}, t_{Tot}, r, L_{drain}, L_{drift}, N_{drift}.

3.3.1 Pitch exclusion

As a first analysis, the variation of breakdown voltage as a function of the distance between two devices connected in parallel trench has been analyzed. This type of simulation is necessary because a single device consists of a device series connected in parallel, so their close proximity could lead to the breakdown phenomena are not negligible. The creation of a new structure that connects two devices in one trench simulation has been required, this new structure was built as in Figure 3.10 where it adds a new parameter, called the T_{pitch}, defined as the distance between a device and the other, this distance can be taken from any point of the structure to its equivalent in the adjacent structure. This kind of simulation requires almost twice number of simulation points, in particular, it changes from 7088 to 12287 points of simulation and the total time of simulation that depends exponentially on the number of points increases from 15min to 1h:15min.

In this analysis, we study the relationship between the breakdown of device with two Trench at a distance T_{pitch} normalized with the breakdown of a single Trench device with the same geometric characteristics. The analysis was carried out on two different devices with two lengths of the drift region, the first with a very large thickness that allows the inclusion of the entire depletion region in the drift region ($L_{drift} = 10\mu m$), the second with a very low thickness ($L_{drift} = 3\mu m$) and therefore depletion region exceeds in the heavily doped drain region, presenting a pinchoff effect. The simulation results, varying the distance between the trench is shown in Figure 3.11, where you can see that the structure with higher drift length value does not show breakdown voltage dependence from the distance between the two Trench and its breakdown value is very close to reference breakdown value; While in the case of the structure with very narrow region of drift, the breakdown voltage tends to increase its value from reference one with

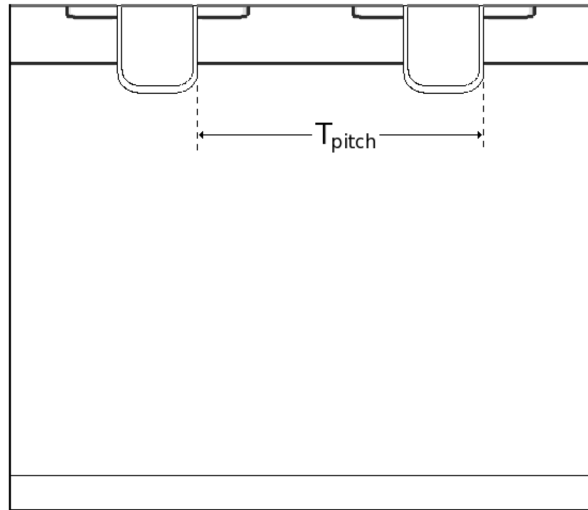


Figure 3.10: Two device simulation to evaluate neighboring effect.

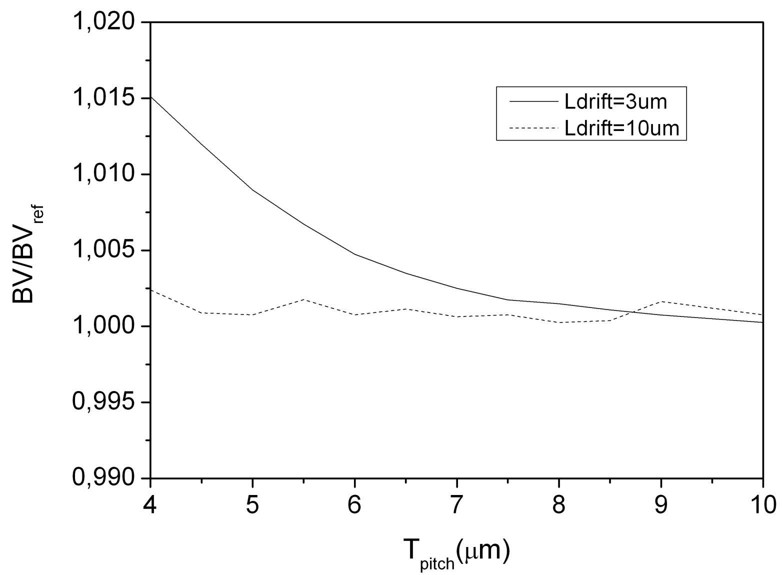
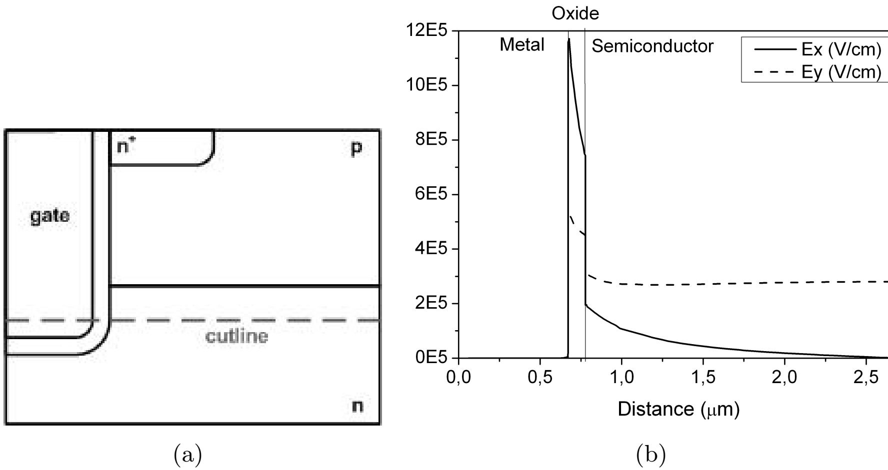


Figure 3.11: Breakdown voltage value for a double cell structure normalized over a single structure breakdown voltage as a function of pitch length.



(a) (b)

Figure 3.12: (a)Cutline representation on Trench structure. (b)Electrical field X and Y components over one-dimensional horizontal cutline described in a.

decreasing distance between the two trenches, but keeping at a very low percentage changes; In fact, even for a very small distance ($4\mu m$) does not exceed a percentage change of 1.5%.

Figure 3.12(b) shows the x and y components of the electric field. As you can see, the X component of the electric field tends to decrease rapidly expulsion oxide-semiconductor interface. Assessing the maximum distance within which you may notice a horizontal electric field comparable to the vertical electric field as X_{Ex}, we can say that if the distance between the two interfaces is higher than neighboring $2X_{Ex}$ then it will present effects of proximity. When the device proximity ensures enough space between devices, it's not required any change in breakdown voltage.

3.3.2 Drift region analysis

This section analyzes the behavior of breakdown voltage as a function of Drift region. This type of analysis is very interesting, because it is the union point between the Trench and pn-type simulations. As in the case of pn simulations, Trench structure is changed in N_{drift} and L_{drift} parameters. The changes will be made with identical parameter values to the reference pn structure ones, so as to enable comparison between the two structures to assess the differences. Figure 3.13 shows the results of breakdown voltage for various drift region length values (L_{drift}) and doping (N_{drift}). As can you see, the results are similar to the pn reference structure. It can be noted the

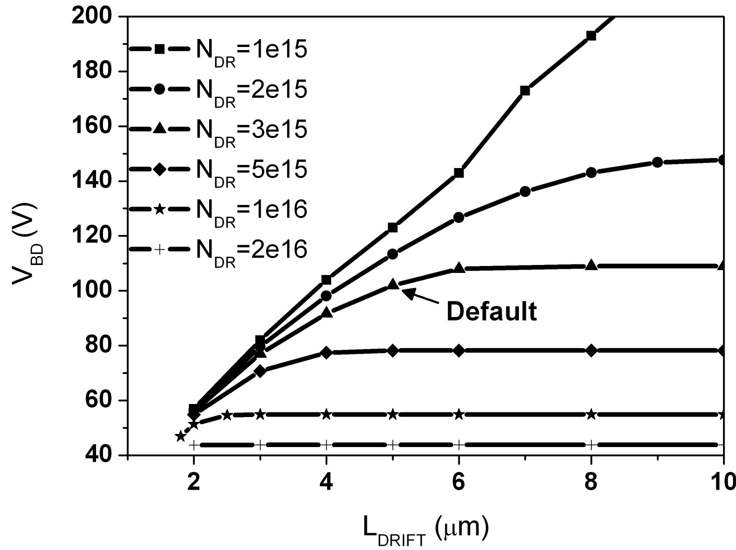


Figure 3.13: Breakdown voltage value as a function of drift length for various drift region doping for a Trench structure.

same behavior as a function of drift region length, a region that is directly dependent on the length of drift increasing the breakdown voltage until it reaches a maximum voltage value, once reached this maximum value, increasing Ldrift the voltage breakdown tends to remain constant. This result shows us how the drift region of trench structure plays a fundamental role in the evaluation of the breakdown voltage.

From the above analysis we can see how the structure created with the default parameters is in the linear region, where there is a dependence of breakdown voltage on the drift length. This condition reflects the modern power devices, since a high value would cause an increase in the on resistivity (R_{dsON}) that often creates many problems, such as reduced efficiency and increased dissipation. The L_{drift} value is still very close to the limit which separates the two regions with different characteristics, linear dependence and independent, to allow it to obtain a sufficiently high breakdown voltage, which allows you to use the device for higher voltage.

3.3.3 Trench's shape analysis

The final analysis to be carried on Trench MOSFET structure affects the shape and size of the trench itself. The study assumes that the gate oxide structure has been formed by thermal oxidation, this process ensures the formation of an oxide that is as much as pure and with the mininum defects as possible. Consequently, the oxide thickness will be about the same as the

interface of the trench bottom, rounded edges, and along the two side walls that form the gate. Consequently, the oxide thickness is taken constant on both the walls and on the bottom of the trench, and along the two rounding interfaces it follows concentric circular profiles with the same center.

Trench radius

The Trench structure curvature radius, indicated by the parameter r, represents the semicircle radius which describes the oxide-semiconductor interface joining one of the two vertical sides of the trench with the horizontal face. This parameter was analyzed separately from other Trench's shape parameters because the the Trench excavation structure does not allow to vary considerably this parameter, so it is useful to observe the structure for the radius value of curvature obtained in a real structure, this analysis is still useful to understand the structure behavior in order to modulate any new processes to improve the breakdown voltage by setting up a new radius value fixed by technology. Figure 3.14 shows the comparison between

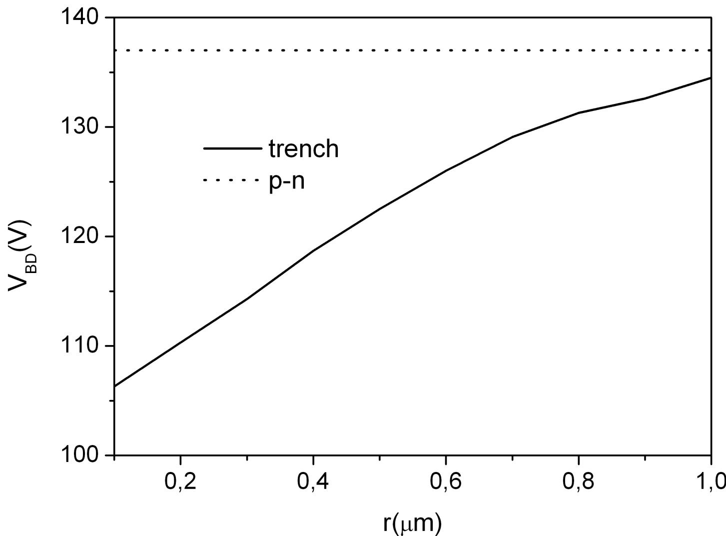


Figure 3.14: Breakdown voltage dependence from Trench corner radius in comparison to p-n reference structure breakdown value.

breakdown voltage in reference structure and trench structure as a function of curvature trench radius. As can be seen, the breakdown voltage shows a strong dependence on the curvature radius, in particular, the breakdown voltage tends to decrease with decreasing of curvature radius, which is due to the effect of the edge which tends to increase the variation of electric field near the Trench's border. The dependence is linear and the slope is equal

to 31V for each 100nm of curvature radius.

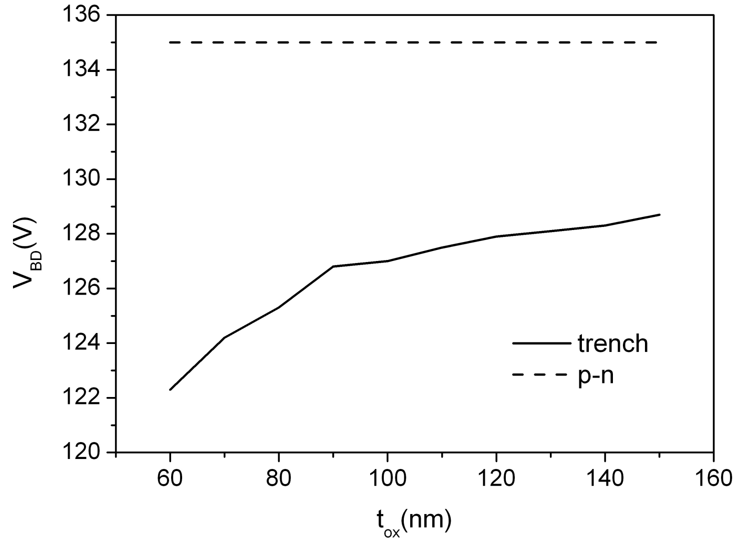


Figure 3.15: Breakdown voltage dependence from gate oxide thickness in comparison to p-n reference structure breakdown value.

Oxide thickness

The oxide thickness plays a fundamental role in the MOSFET structure, since its variation tends to change many parameters such as threshold voltage, subthreshold slope, and in case of power devices even the breakdown voltage. Changing this parameter causes the variation of the gate capacitance and therefore the amount of charge that is accumulated oxide-semiconductor interface. By decreasing the thickness of the oxide gate capacitance increases and the amount of charge at the interface increases by increasing the electric field and causing the best conditions for the impact ionization from going to decrease the breakdown voltage. Figure 3.15 shows the trend of the breakdown voltage as a function of gate oxide thickness. As you can see there is a slight dependence on the oxide thickness, in particular, has a linear directly proportional dependence of breakdown voltage over gate oxide thickness. In particular, increasing the gate oxide thickness of 10nm, the breakdown voltage will increase of 0.77V.

Trench width

As you can see in figure 3.16, the width of the trench does not affect the breakdown voltage, this is because the effect of reducing the width of the

trench is to bring the two MOS structures located on the same trench with the Metal in common, also decreasing semiconductor regions in the bottom of the trench. The gate region becomes thinner, but it causes no change in the profile of potential, will head the cabinet. The semiconductor region below the trench does not show substantial changes in the profile of its potential electric field, and two trench structures do not cause any interference that may cause a variation of breakdown voltage. The width of the trench can be changed in an arbitrary manner, will then be determined so as not to have an aspect ratio too high and to maintain the lowest possible resistivity of the device.

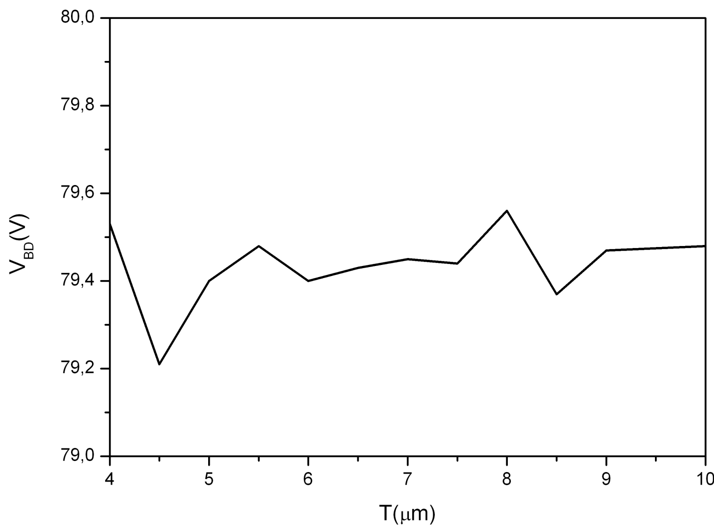
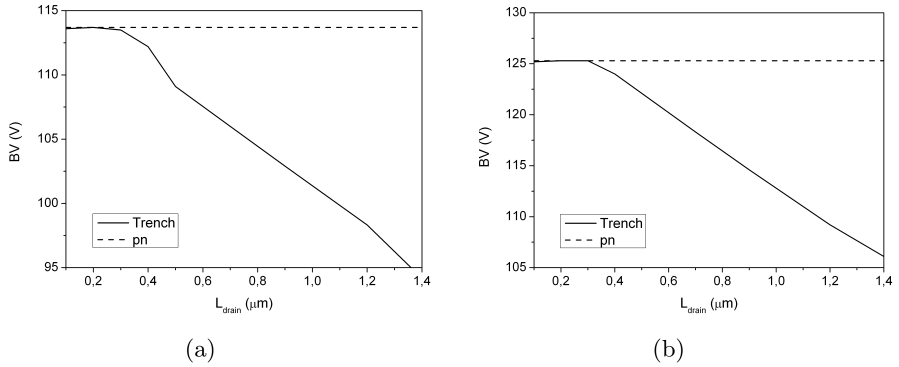


Figure 3.16: Breakdown voltage versus Trench width.

Trench length

The vertical MOS structure, as well as in the region of body, will extend several hundred nanometers in the drift region, this to ensure that the channel is formed completely and be able to connect the source to the drift region. This overlap creates a variation of potential and electric field in the drift region, near the MOS structure that alters breakdown voltage.

Figure 3.17 shows breakdown voltage of reference structure and a structure with region of drift long enough not to allow the pinchoff effect described above. As you can see, the two curves have very similar trends, in particular, the breakdown voltage is identical to the breakdown voltage of the reference structure if L_{drain} does not exceed a threshold value once exceeded this value, the voltage Breakdown tends to decrease linearly with



(a) (b)

Figure 3.17: Breakdown voltage versus Trench depth for reference structure(a) and for structure with a long drift region $L_{drift} = 10\mu m$ (b).

L_{drain} length. The slope of the two curves are approximately equal and amounted to $18.2(V/\mu m)$.

To obtain good results in terms of breakdown has deepened the study of L_{drain} variation for different values of drift region length and doping, so it was possible to observe the breakdown voltage variation in different conditions as shown in Figure 3.18. The same family of simulations was repeated for a different value of drift region doping N_{drift} and has still come to these conclusions:

- The breakdown voltage is kept constant at the value of the breakdown of the reference structure until it reaches a limit value which we call T_{limit}.
- Once this limit the breakdown voltage tends to decrease linearly with L_{drain} length.
- The linear region slope remains approximately constant when the drift region is able to completely contain the depletion region, condition verified when $L_{drift} > W_{PNmax}$.
- The slope tends to increase by decreasing the drift region than the threshold value.
- The variation of the drift region doping will affect the change of the parameters mentioned above (T_{limit}, W_{PNmax} and the breakdown slope).

From the analysis made it is clear that the device breakdown is mainly influenced by the drift region characteristics and secondly the Trench characteristics, in particular from gate oxide thickness t_{ox}, Trench's drift penetration length L_{drain}, and curvature radius r. While less significant changes

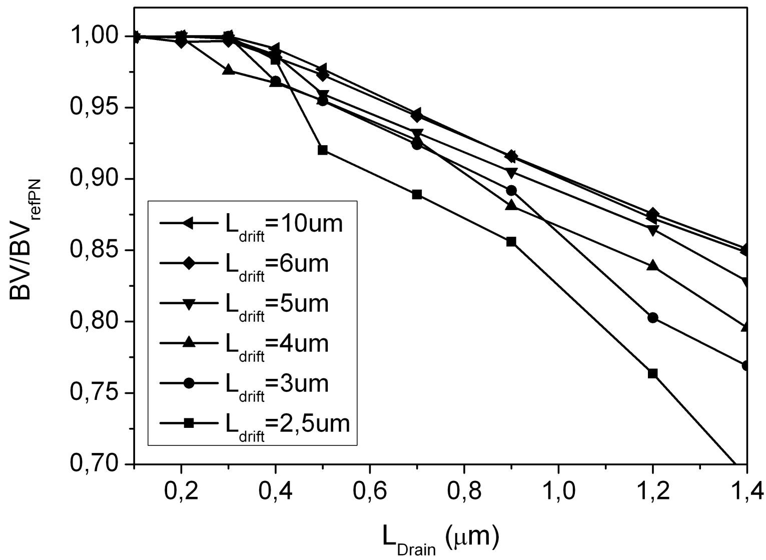


Figure 3.18: Breakdown voltage variation from reference pn structure value over trench penetration into drift region's length at different drift region length.

in breakdown voltage are shown from other characteristics like the bulk and source region characteristics, the variation of the distance between the various devices in parallel T_{pitch}, and the variation of some Trench characteristic like T and N_{Gate}.

Chapter 4

Novel Breakdown model

In previous chapters we discussed the variation of breakdown voltage according to the variation of internal parameters of the device, parameters like doping concentration and device length, based on simulations made without evaluating the device's internal electrical characteristics and the physical effect that will cause the breakdown voltage. This chapter will examine the causes that generate the breakdown voltage and try to find a physical model that directly links the breakdown voltage and the geometrical and doping parameters. As a starting point the analysis of a simplified reference structure will be made, this structure is presented as a simple unilateral pn junction with same Trench's characteristics on drift, body, and drain region, evaluation made by using several model found in literature, once obtained a good model in agreement with simulation results the model will be compared with Trench structure simulation to evaluate the differences and find a union point between the two models. Once you have some similarity will be worth using a new empirical model for calculating the breakdown voltage based on the correction of the pn device model. Finally, the model will be compared with simulations to assess the actual accuracy.

4.1 Pn breakdown model

As described previously in section 2.1.2, breakdown condition in a unilateral pn device could be expressed as the solution of ionization integral that reaches the value of 1 inside the depletion region [36, 37]

$$\int_0^W \alpha dx = 1 \tag{4.1}$$

The same condition can be used for reference p-n structure since electrical field lines have no y-component that takes to a one-dimensional problem. To make a one dimensional analysis, a vertical device cross section will be made as described in figure 4.1 and all the following analysis are made on this cross section.

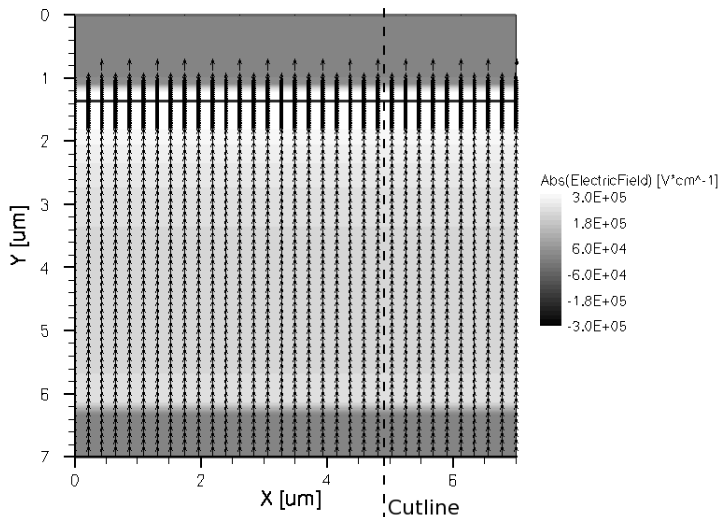


Figure 4.1: Electrical field simulation results of pn reference structure, with absolute value as grayscale and arrows as X/Y direction, pn junction is placed at black line at $y = 1.47\mu m$.

When electrical field absolute maximum value reaches this critical value in an arbitrary point inside structure, the p-n junction will increase its current value of several orders of magnitude due to avalanche breakdown. This electrical field absolute maximum value is always reached on pn junction and can be calculated by integrating the ionization integral obtaining the following formula [38]

$$E_{Cpn} = \left(\frac{N_{drift}}{KK}\right) \tag{4.2}$$

Where $KK = 2.41e-29cm5/V8$, and E_{Cpn} is the critical absolute electrical field value expressed in V/cm. This electrical field maximum value has been

calculated for an abrupt unilateral pn junction, this approximation take into account of a perfect triangular electric field shape along Y-axis. This approximation is valid if Body diffusion length is smaller enough to ensure that the triangular rounding shape can be less than electric field maximum value. Note that Electrical field is drift n-region doping dependent only and is evaluated for a unilateral pn junction. Once you get the maximum electric field value, and knowing that this value lies on the pn junction, you can use this condition as a boundary condition for calculating the voltage across the pn junction in the n-type region with

$$\begin{cases} \frac{d^2V(y)}{dy^2} = -\frac{qN(y)}{\epsilon_s} \\ V(0) = \psi_{pn} \\ \dot{V}(0) = E_{Cpn} \end{cases} \tag{4.3}$$

where ϵ_s is the silicon dielectric constant, q is the absolute value of the elementary charge, $N(y)$ is the doping concentration into drain region as function of y direction, ψ_{pn} is electrostatic potential at pn junction, and V(y) is the electrostatic potential. Equation (4.3) is a classical Cauchy problem where first equation is Poisson equation in a lightly n-doped layer with assumption that pn junction is placed at $y = 0$, and the other two equations are boundary conditions for initial value and first derivative initial value respectively. The first boundary condition requires the need to know the potential at the junction, which depends on the potential drop due to the p region, this potential drop can be neglected if the junction is unilateral, and therefore consider it equal to 0, otherwise it becomes necessary to solve the same system in the region p. In the case of study, Poisson equation needs to be solved only in drain side. By integrating two times formula (4.3) between 0 and W, that is depletion region length, we can extract the potential variation in drain region and obtain the breakdown voltage value. This breakdown value can be evaluated even if n-side depletion region exceeds drift region and extends in n^+ drain region. Figure 6 show comparison between simulation results and the analytical model, showing a good accordance even for a wider value of drift region doping and lengths. As previously described, the breakdown value shows drift region length dependence until a drift limit value length is reached; after this limit, the breakdown value will maintain a constant value. This limit value can be evaluated with formula (4.4) that represents n-side depletion region length for constant doping value at N_{drift}, where W_{BVp-n} is expressed in cm. [39]

$$W_{Vpn} = 2.67e^{10} {N_{drift}}^{-\frac{7}{8}} \tag{4.4}$$

The solution of equation (4.3) with approximation of $\psi_{pn} = 0$ in terms of electrical field absolute value in comparison to simulation results are shown

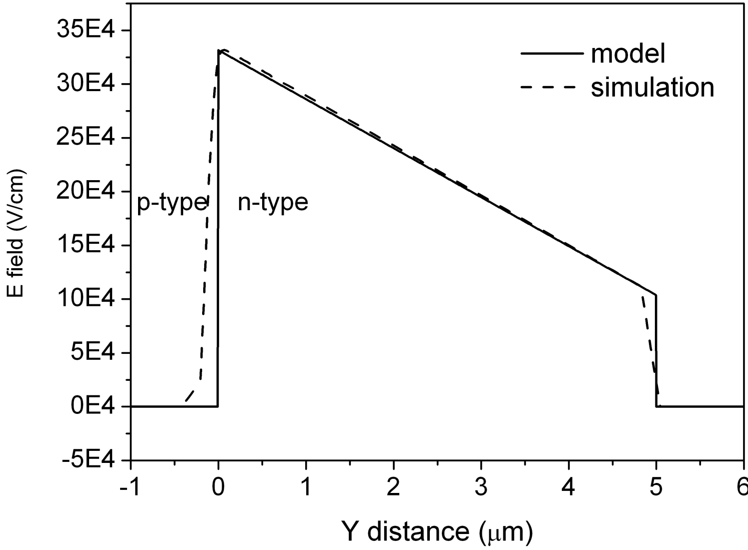


Figure 4.2: Electrical field distribution along y axis for simulation and physical model.

in figure 4.2 where solid line is physical model electrical field, and dashed line is simulated electrical field. As can be noted, the two electrostatic potential curves are very close, and its maximum value is identical, but in p-region the simulated one present an electric field contribution that is not present in the model one that approximate the electrostatic potential inside p-region. This difference will be present even in potential profile expressed in figure 4.3 where the electric field in the p-region will cause a variation in electrostatic potential that increase the potential with a square law since to reach a value that will cause a potential shift in the n-region.

In order to take into account of this effect, equation 4.3 must be solved for negative value from $-W_p$ to 0. By approximating doping concentration as a constant and that

$$W_p = \sqrt{\frac{2\epsilon_s V_{BD}}{qN_A}} \tag{4.5}$$

Where V_{BD} is the breakdown voltage that we want to calculate, this formula could not be evaluated if we don't know breakdown voltage, but it could be approximated with previous voltage value obtainet with $\psi_{pn} = 0$ approximation. All the parameters presented here has been compared to simulation results shown on figure 4.4, where different electric field vertical cross section are presented. It can be noted that the maximum electric field value E_{Cpn} is drift region doping dependent only, instead of the depletion region in p-region W_{Vpn} that depends from drift region doping and from

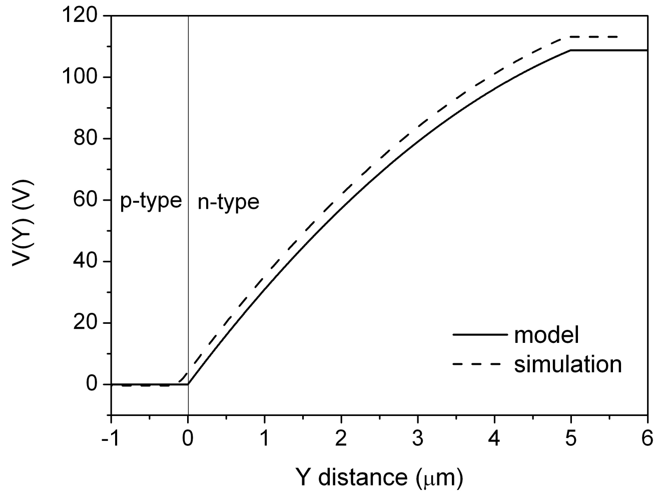


Figure 4.3: Electrostatic potential distribution along Y axis for simulation and physical model. Note that simulation results show a increasing in the p-region voltage near the pn junction due to a presence of a depletion region neglected in the model one.

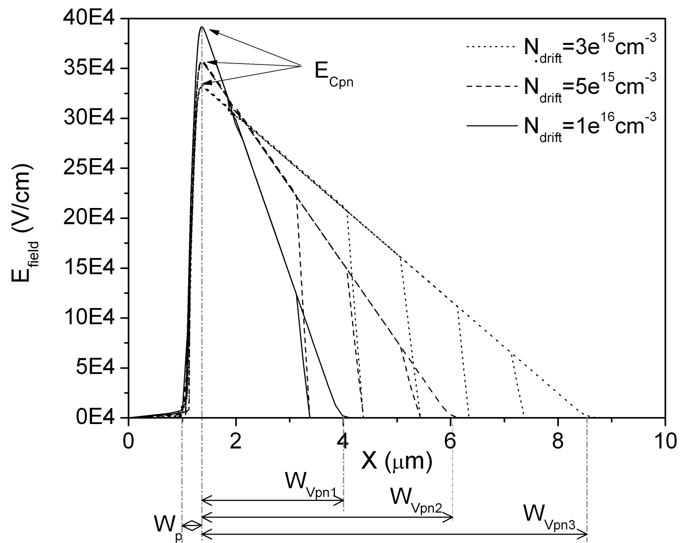


Figure 4.4: Electric field profile from various simulation with different doping concentration in drift region and for different drift lenghts at breakdown condition.

drift region length too. The depletion region in the p-side W_p is a little different between simulation one, this approximation will produce a lower voltage on the p region in the order of one volt, tis reproduces an overestimation in the breakdown voltage negligible with respect to the breakdown voltage.

Once obtained the depletion region value it's possible to solve (4.3) integrating in twice in space region between 0 and W_p obtaining

$$\psi_{pn} = \frac{qN_A W^2}{2\epsilon_s} \tag{4.6}$$

That for reference structure it results $\psi_{pn} = 4.75V$.

The model as described now doesn't take into account of built-in potential caused by the doping difference that create a built in potential given by the formula

$$V_{bi} = \frac{kT_L}{q} ln\left(\frac{N_A N_D}{n_i}\right) \tag{4.7}$$

where T_L is the lattice temperature in Kelvin degrees, n_i is the intrinsic concentration, and V_{bi} is the built in potential.

By adding this potential to the breakdown results we obtain the correct physical model for breakdown extraction.

4.1.1 Model analysis

In the previous section we obtained the correct model that well approximate the electrostatic potential and electric field inside a unilateral pn structure, we can use it to evaluate breakdown voltage for the same simulations made in section 3.2.1 in order to obtain a good comparison. Has been extracted the doping concentration Y section for every structure, and has been used as doping concentration in equation (4.4), by solving it numerically the breakdown potential has been evaluated by using the resulting potential and adding the built in potential. The calculation results are shown in figure 4.5 where simulation results show a good accordance to the model one.

In order to make a good analysis is a good idea to identify if n-side depletion region at breakdown condition lies in drift region or extends in n^+ drain region. This condition could be verified by looking at figure 4.6 and find if the device is placed above or below the curve. This condition is important because in a non-fully depleted, in the event that any altering of the structure so requires, the depletion region can be increased or decreased depending necessary to adapt to new conditions, while in the case of region drift completely emptied the depletion region does not change considerably, as the effect on the heavily doped region would be less. This is a significant

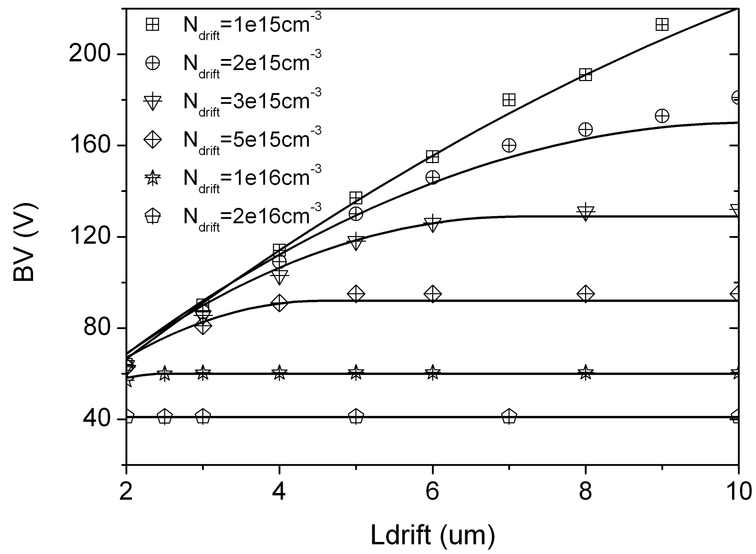


Figure 4.5: p-n reference structures breakdown voltage over drift length for simulation and physical method; Analytical model as solid lines.

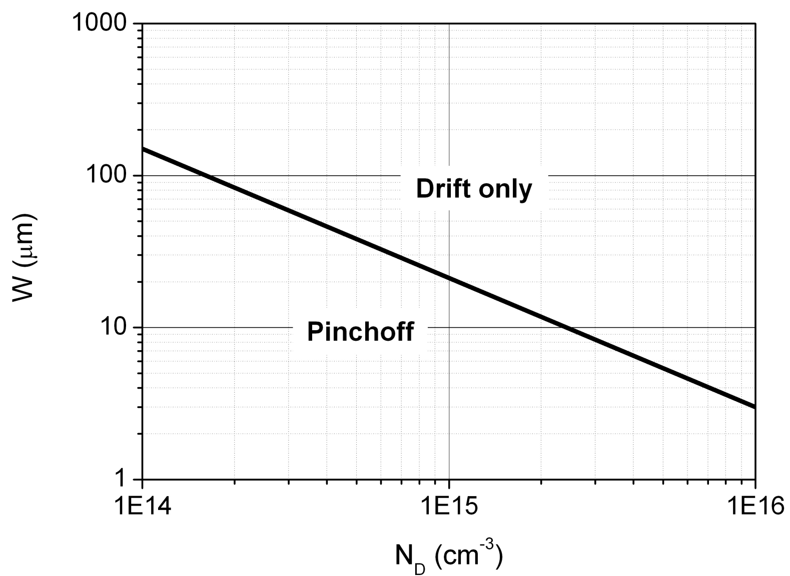


Figure 4.6: Maximum value for depletion region in n-type region for different doping value.

difference in the subsequent analysis in which this structure is modified by the inclusion of Trench structure.

4.2 Trench breakdown model

When device nature will be changed by adding a trench shape and the source pads inside p^+ region, the structure changes from a pn simplest one to a trench power MOSFET. This structure modification will affect electrical field lines that now will be created by two different effects: One component due to the p-n structure with a y oriented electrical field component, and another due to the MOS structure with a particular electric field. As it is known, a MOS structure in inverse polarization will generate a depletion region that generate an electric field perpendicular to the MOS structure. Perhaps around the trench's shape there's an additional electrical field component directed as the normal of oxide-semiconductor interface,

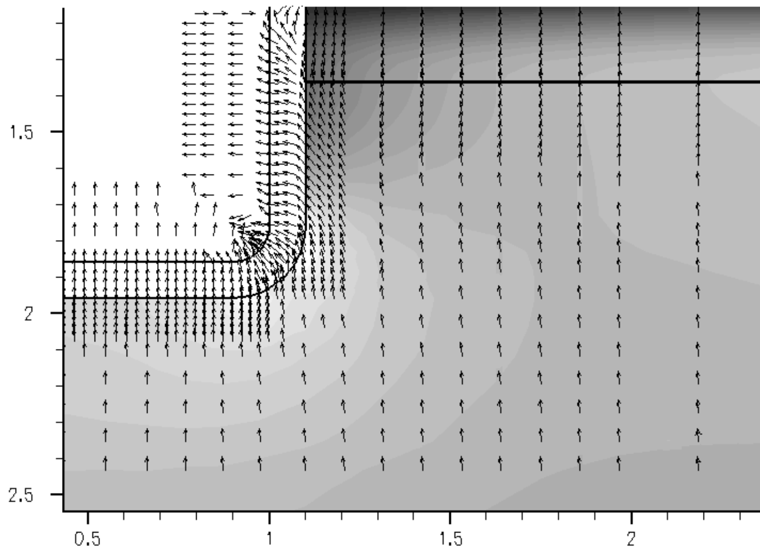


Figure 4.7: Electrical Field cross section simulation results for Trench MOSFET structure, with absolute value as grayscale and arrows as X/Y direction, pn junction is placed at black line at $y = 1.47\mu m$, the Trench MOS structure is on the top left. The plot is resulting by a zoom near the trench edge, in particular, starting from Trench's center to $2\mu m$ right, and from the middle of the channel to a middle area inside the drift region to Y direction.

this component will be y oriented on bottom of Trench, x-directed on the two sidewalls , and a a composition of x and y components to the bottom

Trench edges. [27, 40] Figure 4.7 show a Trench device cross section of Bottom right trench's detail with $V_D = 50V$ and zero voltage on source and gate terminals, the grayscale indicates the electrical field absolute value, and the arrows show the electrical field vector direction. As can be seen, the arrows far enought from the trench structure are Y-oriented because of the p-n electric field influence's only, even under the bottom trench's structure the lines are Y-oriented only, but the electric field is higher because of the addition of the bottom MOS structure, that's always Y-oriented. But near the lateral MOS structure and at the trench's angles, the electrical field lines are affected by both reverse voltage depletion and MOS depletion component with an x component different from zero, and at the semiconductor-oxide interface the electric field lines are perpendicular to the interface itself showing an X-component only on the vertical interfaces. This considerations make the problem a two-dimensional one.

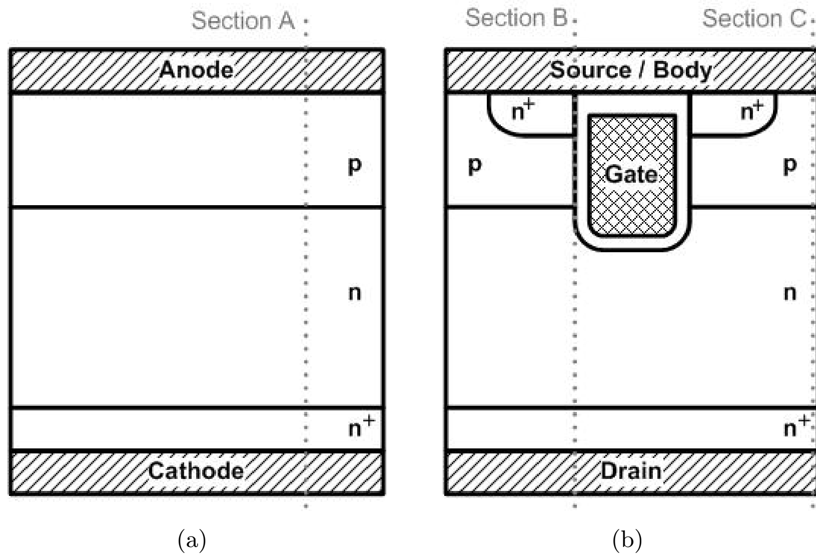


Figure 4.8: (a)Schematic representation of p-n reference structure with cross section A. (b)Schematic representation of Trench power MOSFET structure with cross section B and C. These cross section will be used for the following analysis and device comparison.

In order to evaluate Trench electrical properties a number of device cutline has been defined. In particular, the vertical cross section previously described for the p-n reference structure will be compared to some Trench MOSFET vertical cross sections. These cross sections are shown in figure 4.8, where a single label has been associated to every cutline. The reference structure cross section will be named as "section A". This cross

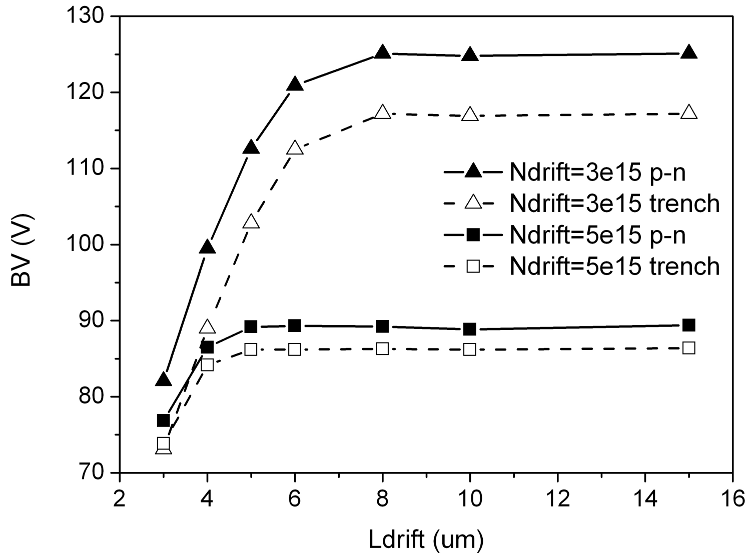


Figure 4.9: Breakdown voltage for trench and p-n reference structure versus drift length, with filled symbols as pn structure and empty symbols for Trench structure.

section is taken over an arbitrary X point inside p-n structure, the X position is irrelevant to the electrical characteristic since the p-n problem is one–dimensional. The other two cross sections are Trench structure related, the first one, so called "Section B" is a vertical cross section taken to touch one of the two vertical Oxide-Semiconductor Trench's interfaces, which of the two interfaces is taken is irrelevant, because of the device symmetry. The "section C" is taken to a fixed distance from section B, in order to ensure that we are far enough from Trench's influence and that the electric field conditions are p-n component related only.

4.2.1 Electrical consideration

As can be seen from figure 4.9, the breakdown voltage in Trench structure is lower than pn reference structure with equivalent drift region characteristics, this is caused from the different electric field trend near Trench structure, analyzed previously in Figure 4.7, which shows that the peak electric field has moved from the body-drain junction region to the rounding trench border. This displacement causes a shift in breakdown position from p-n junction to trench rounding shape neighborhood, meaning that breakdown does not follow a spatial line, but is generated on a smallest spatial region. Because of this shift of breakdown position, the electric field maximum value on pn drain to bulk junction on trench device will not reach

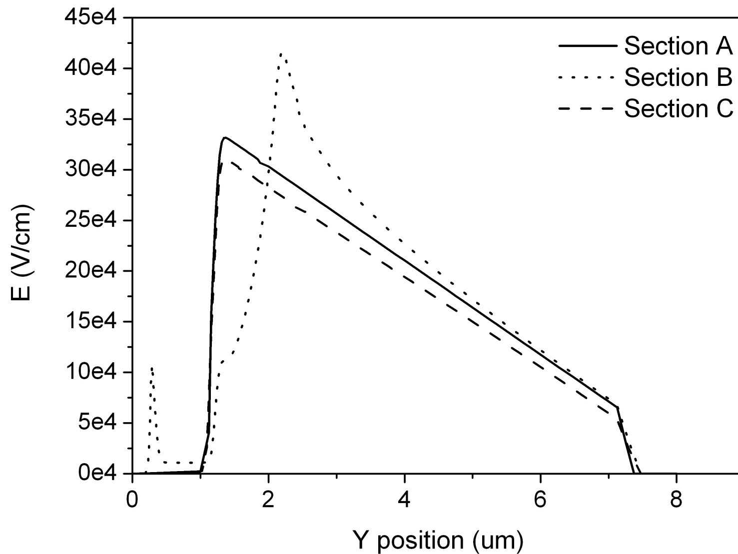


Figure 4.10: Vertical electrical field cross section of trench device and p-n reference structure device for section A, B and C as described previously. Section from top to bottom with axis origin at source's silicon-contact interface.

the critical value expressed in formula (4.2), so this formula cannot be used to evaluate electrical field maximum value reached at trench rounding corner because of higher holes concentration caused by gate electric field influence that change semiconductor local space charge near trench border.

4.2.2 novel breakdown model

Figure 4.10 shows electrical field vertical cross section value for p-n reference structure and for trench structure near gate oxide and at simulation border in breakdown condition. The electrical field cross section in p-n structure shows a classical electrical field distribution in a p-n structure in inversion mode with a maximum electrical field value in accordance to formula (4.4) at the p-n junction. Trench structure vertical cross section B shows instead trench influence on electrical field maximum value that shift electrical field maximum value position near trench bottom edge and increase its value as compared with p-n reference structure. Vertical cross section at $3.5\mu m$ from trench's center (section C) is not affected by trench influence, so the electrical field profile shows a similar shape to p-n one. This similarity is due to electric field horizontal component absence that makes the cross section C analysis a one-dimensional approach problem. The cross section A

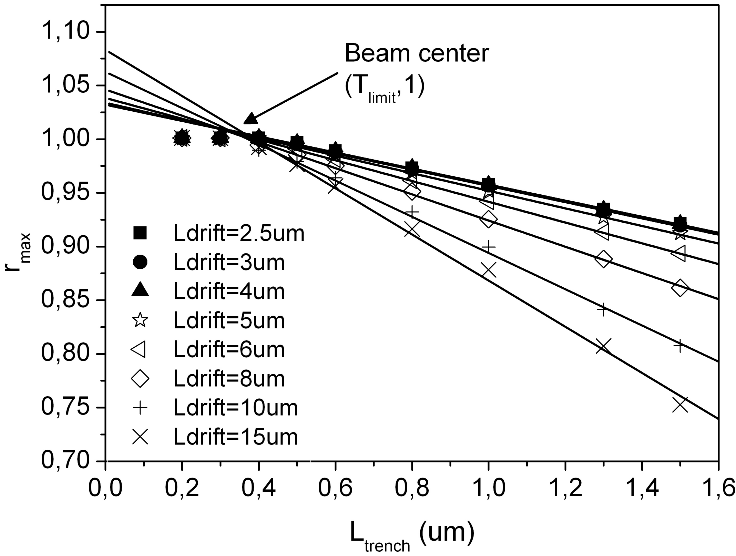


Figure 4.11: Maximum electrical field value on cross section C over maximum electrical field on p-n reference structure (E_{max}/E_{maxPN}) as function of trench length for various drift region doping at $N_{drift} = 3e^{15}cm^{-3}$.

shows electrical field cross section in p-n reference structure at breakdown condition with a maximum electrical field value higher than trench one on section C. This reduction in electrical field maximum value, that for simplicity we call electrical field ratio, can be used to correct analytical p-n model to obtain breakdown voltage for trench structure as shown in formula 4.8, where E_{maxp-n} is the maximum electrical field on p-n reference structure, E_{max} is the maximum electrical field absolute value on section C, and r_{max} is the ratio between the two values.

$$r_{max} = \frac{E_{max}}{E_{maxPN}} \tag{4.8}$$

In order to evaluate drift and trench length's r_{max} dependence a wide range simulation has been made. The electrical field ratio r_{max} was calculated as the ratio between the maximum electric field in the trench structure and the value of electric field calculated using the formula (4.2). This ratio value has been extracted for different drift length L_{drift}, drift doping N_{drift}, and trench length L_{trench}, sequent simulation results are shown for different L_{drift}, and L_{trench} values with constant N_{drift} value, and the analysis has been made again for another N_{drift} value.

First of all, the analysis with $N_{drift} = 3e^{15}cm^{-3}$ has been made. The r_{max} results are shown in figure 4.11. where scatters are simulation results

and solid lines are analytical line calculated with least mean square method for every Ldrift value, the analytical lines are calculated to have a minimal confidence bound of 95%. In the calculation of least squares were excluded all simulation points where $r_{max} \cong 1$ because as can be noted at such points the curve deviates sharply from the linear profile. As can be seen, for trench length smaller than a limit value the electric field ratio is fixed to one and is not dependent from both drift region and trench length. Increasing trench length over this limit value will introduce a dependence of electrical field ratio from trench length, this limit value has been evaluated by calculating the least mean square line for every drift length with exclusion of simulated point close to one and calculating the L_{drift} value with a r_{max} value of one. This limit value, evaluated for every drift region length, could be assumed as constant for all the lines and will be indicated as T_{Limit}. Since the electrical field ratio dependence from trench length is linear and Tlimit is placed on a single point in figure 4.11 for every drift length value, the model for electrical field ratio calculation can be expressed with a beam lines with center in the point T_{Limit} and 1 respectively as X and Y coordinates, and with a variation in line slope. A good approximation of r_{max} calculation is expressed in formula (4.9), where $m(L_{drift})$ is the slope of electrical field ratio at drift length value.

$$r_{max} = \begin{cases} 1 & for \quad L_{trench} \leq T_{limit} \\ 1 + m\left(L_{drift}\right)\left(L_{trench} - T_{limit}\right) & for \quad L_{trench} > T_{limit} \end{cases} \tag{4.9}$$

Drift region dependence of $m(L_{drift})$ slope has been analyzed in figure 4.12. It can be noted that m slope is a constant value if the depletion region lies drift region, decreasing drift region length will cause an increase in m slope value. This slope variation is related to drift region shrinking under trench formation causing the depletion region reduction. The numerical model chosen for m slope as a function of L_{drift} is given by:

$$m(L_{drift}) = m_0\left[1 - Aexp\left(-\alpha L_{drift}\right)\right] \tag{4.10}$$

the model is a exponential decay model, where m_0 is the slope for device with $L_{drift} > W_{BVpn}$, α is the decay factor inversely proportional to W_{BVpn}, and A is a multiplication factor. All the parameters has been calculated using least mean square method with a minimum confidence bound of 95% and its numerical value for $N_{drift} = 3X10^{15}cm^{-3}$ are: $A = 12.5$, $\alpha = 0.78$, and $m_0 = 0.075$.

Using the model for maximum electric field reduction in trench devices described by formula (4.8) and replacing the border condition in formula (4.3) it's possible to evaluate breakdown voltage comparison between simulation results and the novel numerical method. Comparison results are

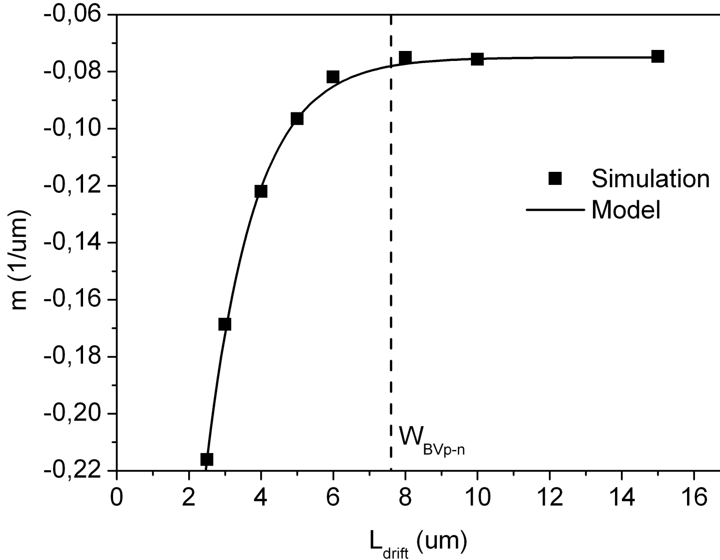


Figure 4.12: Dependence of the slope parameter m over drift length L_{drift} and its numerical approximation.

shown in figure 9 where symbol points are simulations and solid lines are numerical model results. At this point we can use the method described above to obtain the breakdown voltage for a pn junction with the r_{max} correction to maximum electric field in order to obtain the same results for Trench Power MOSFET structures. Calculations were performed for structures with identical simulation's characteristics in order to compare simulated results, the results comparison are shown in Figure XXX, where the points are the simulation results and the lines are the of the implemented model results. The model is consistent with the simulations, allowing to validate model.

4.2.3 model synthesis

To assess the variation of breakdown voltage as a function of drift region doping the following rules was repeated with enough N_{drift} values to cover the most commonly drift region doping value used in power MOS devices. The procedure was:

- Choose a N_{drift} value and make a number of simulations with various L_drift, and L_{drain} values.
- From these results has extract the maximum electric field Trench structure value at breakdown condition.

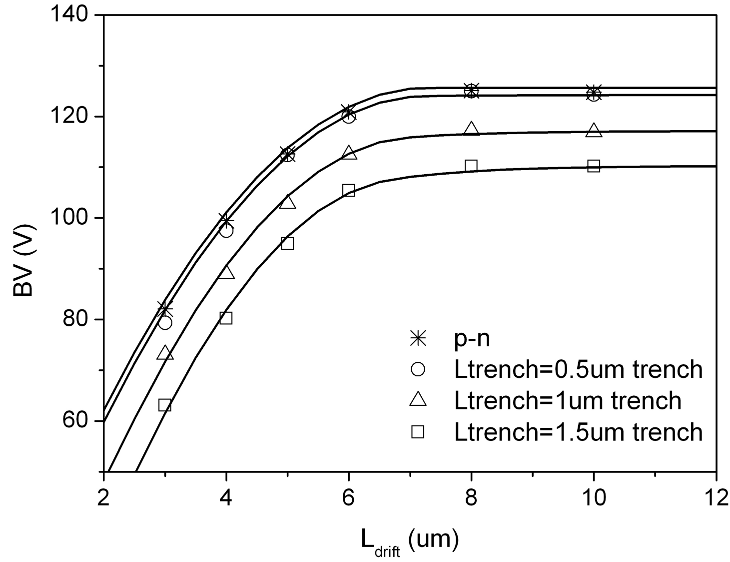


Figure 4.13: Trench structure breakdown voltage over drift length for different trench length; solid lines as Trench analytical models, and dashed line as p-n analytical model.

- Divide this values by reference pn structure electrical field maximum value evaluated with (4.2) to obtain r_{max} for each structure.
- For every L_{drift} fixed value, evaluate the m and q coefficients for the line passing through r_{max} points.
- Calculate the intersection of all lines to get T_{limit}.
- Make the fitting of the m slope with equation (4.10) in order to obtain the values of A, α and m_0.

With this procedure it was possible to show the variation of the parameters of the empirical model as a function of the doping of the drift region, the results shown in Figure 4.14(a) and (b).

As you can see, all the parameters are drift doping dependent. T_{limit} in particular tends to decrease with increasing doping, this suggests that increasing doping, the breakdown area tends to remain on the trench edge and don't move on junction area. m_0 tends to increase with N_{drift}, m_0 is the r_{max} dependence over L_{trench} under conditions of drift region not completely depleted; increasing the drift region doping, the electric field variation along y axis becomes greater, this cause an increasing in the potential difference seen from Trench bottom MOS structure producing an higher electric field on the Trench's corner and reducing the breakdown voltage. The remaining

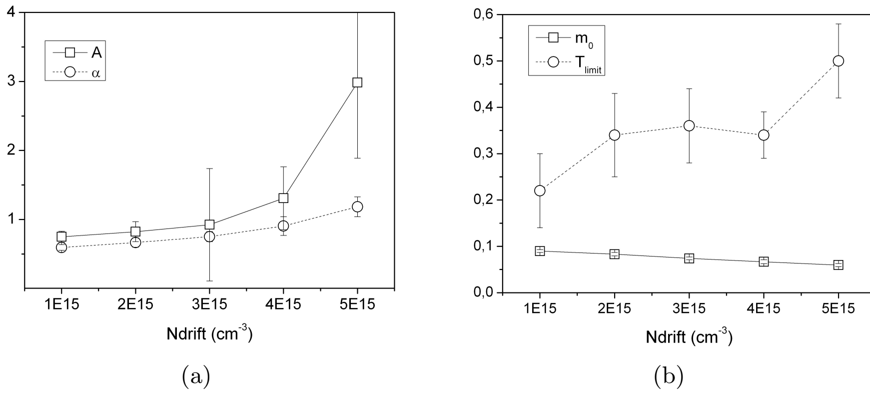


Figure 4.14: Drift region doping dependence for (a)A and α model parameters. (b)m_0 and T_{Limit} model parameters.

two parameters are somewhat related to m_0, as they represent the slope variation where the depletion region extends in the more doped drain region also, which is why they tend to increase with N_{drift}.

Part II

memory device simulations

Chapter 5

capacitorless device description

Several years ago, the classical transistor was redesigned to implement a dynamic random access memory (DRAM) cell under the name of zero-capacitor RAM or, most commonly, Z RAM [41–48]. Unlike the classical 1T/1C DRAM cells, the ZRAM memory consisted only of a silicon on insulator (SOI) MOSFET, but did not require an additional capacitor. The information in Z RAMs is stored in the SOI substrate as excess majority charge created by impact ionization and/or by band to band tunneling (BTBT) at the bulk drain junction induced by the relative high drain voltage. Indeed, the capability of storing charge in the bulk of a MOSFET and the study of its effects on transistor behavior has been discussed since the 1980's [49, 50]. The Z-RAM concept is a remarkable example of how an undesirable phenomenon, the floating body effect of SOI technology [51, 52], can be transformed into a desirable one, the storing capacitance of a DRAM cell. Nowadays, Z-RAM is drawing a lot of interest from the semiconductor industry since it offers several potential advantages respect to conventional DRAM [53, 54]:

(i) extremely high density thanks to the elimination of the additional capacitor;

(ii) low cost of fabrication, since it is implemented on a standard SOI logic process without exotic process steps;

(iii) excellent delay-power trade-off, due to the use of SOI technology;

(iv) possibility of taking advantage of multi-gate architectures, as demonstrated in Ref. [46–48].

In this chapter, we explore the physics of the bipolar operation of double gate (DG) type II Z RAM cells. These cells are sometimes operated in transient mode, with time-delay between gate and drain pulses; dc operation of the cells however is also possible and provides equivalent capabilities. We will use steady-state cell operation in this scaling study, where the excess charge is accumulated at the gate interfaces and not in the body of the SOI substrate. This reformulation simplifies cell operation and allows explicit comparison of scaling potential of generation II devices with respect to the scaling limits of type I cell [55–57].

First of all, in section 5.1 we present a summary of Random Access Memory with its characteristics. In section 5.2, we present the main characteristics of a SOI MOSFET device that will be used as ZRAM structure, and in the other sections we discuss the physics of the READ/WRITE/HOLD operation for bipolar mode operation using a SOI MOSFET reference structure taken as a long device length and describing the main characteristics of the simulation setup, like physical model and numerical technique.

5.1 Memory device concepts

In modern digital devices, with the name "memory" we indicate some device or device part that have capabilities to store information as logical level of "1" or "0", this digital device part will tke a lot of importance, and with device scaling it requires to be always more reliable, to have a lower power consumption, and require a largest bit density over chip area. This memory is indicated into Computer science as Random Access Memory (RAM). The RAM digital memories can be divided into several types, that indicate its characteristics; The RAM memories can be divided into two categories [58]:

1. The dynamic RAM (DRAM) that stores a bit of data using a transistor and capacitor pair series connected, which together comprise a memory cell. The capacitor holds a high or low charge (1 or 0, respectively), and the transistor acts as a switch that lets the control circuitry on the chip read the capacitor's state of charge or change it. As this form of memory is less expensive to produce than static RAM, it is often used as the memory block inside Von Neumann architecture because of its lower costs per bit.

2. The static RAM (SRAM). In this kind of memory a bit of data is stored using the state of a flip-flop. This memory is more expensive than other memories, but is generally faster and usually have less power consumption, in modern digital devices, because of its speed, this kind of memory is often used as cache memory between the CPU

and the memory of von Neumann architecture in order to ensure better performance for data and instruction code with low memory requirements.

Both static and dynamic RAM are volatile memories, it means that their 1 or 0 state will be lost if device is powered off. In order to have a memory device that could maintain data even without a power supply, the Read-only memory (ROM), the EEPROM, and the flash memory has been taken. In particular, the ROM devices does not allow the WRITE operation, because the data stored on it are fixed, like on a CD, or DVD, instead of flash devices that could be rewritten like magnetic Hard Disk, or USB flash drive.

5.1.1 Dynamic RAM

DRAM is made by a square array in which every element is composed by a n-MOSFET device and a capacitor, and the entire structure is composed by long horizontal lines connecting each row known as word Lines, and other long vertical lines connecting each column known as bit lines. Figure 5.1 show an example of 2x2 cell DRAM memory with two bit lines, two word lines and 4 memory block, a single memory block is connected as follows: the capacitor is connected between ground and source MOSFET contact, and the MOSFET is connected at Gate contact to bit line and drain contact at word line. With this kind of structure, the data will e written as a voltage level inside the capacitor, as an example, if the device is powered with a voltage between 0 and V_{DD} the logic state "0" could be expressed as a voltage of 0V, and the logic state "1" with the voltage V_{DD}, so in order to make a write operation it's necessary to put the correct voltage on the bit line and is necessary to put a positive voltage in the corresponding word line in order to create a contact between the it line containing the data to be stored and the correct capacitance to charge or discharge. The other devices on the same bit line are not affected by this modification because the MOSFET are not enabled, perhaps there is no connection between capacitor and bit line, but the write operation must be made for every device on the same word line, because of the activation of all the MOSFETs device on that row.

In order to read the data stored in a cell the corresponding bit line is pre-charged to a voltage exactly in-between high and low logic levels (ex. $V_{DD}/2$), this operation put the bit line parasitic capacitor at this voltage. At this point the bit line is putted in floating state, because the bit lines are relatively long, the parasitic capacitance is larger enough to maintain its voltage for a small time. In a relatively small time that allow to not lose the bit line charge, the desired word line is then activated creating a short contact between the cell capacitor and the bit line. This operation

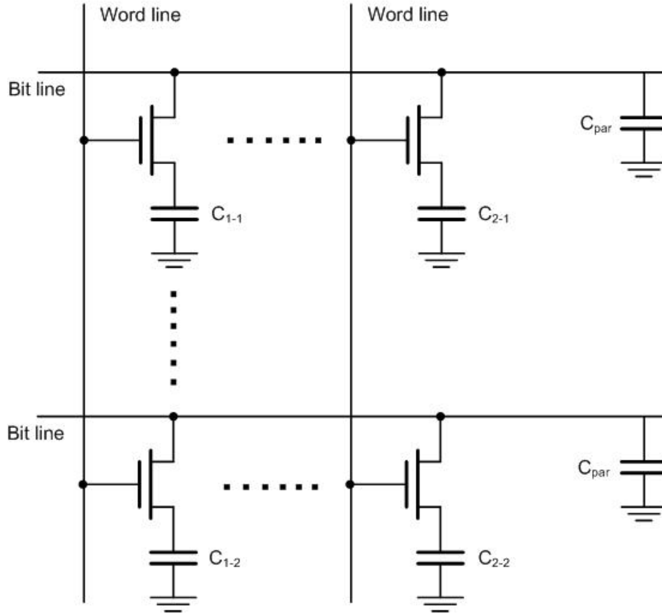


Figure 5.1: Dynamic ram schematic representation of 4 data bit on two bit line and two word line.

ensures that the cell capacitor will charge or discharge the bit line parasitic capacitance increasing or decreasing the bit voltage depending on whether voltage is stored on it. After this step, the word line is deactivated and the word line could be read. All columns are sensed in simultaneously, because the word line activate all the MOSFETs on the line that connect all the capacitor to the relative bit line, so the results are stored into a read buffer. Since the bit capacitor will be discharged by the parasitic capacitor, the read operation result destructive, perhaps after a read operation, a write operation of the same data must be made.

The hold operation simply holds off all the word lines taking all the capacitors disconnected from the bit lines. This operation is not ideal, perhaps the capacitors are connected to a large value resistor that slowly discharge the capacitor. This phenomena could lead to loss of information stored on the cell if this is not read in a fixed time. In order to solve this problem the system will refresh every row in a cycle activity that ensure that all the data are refreshed every fixed time of 64ms.

The refresh operation could be made by a simple read operation with the buffer data storing followed by a write operation of the same buffered data. Every operation requires a fixed time to be executed, but this values must be consistent to the technology limit fixed in literature [59], this limit value and its description is expressed in table 5.1

Name	Value	Description
t_{RC}	$84ns$	read or write cycle time from one cycle to another
t_{RAC}	$50ns$	Access time to valid data out
t_{RB}	$20ns$	bit/word line pre-charge time
t_{HB}	$15ns$	bit/word line for hold charge time
t_{WB}	$15ns$	bit/word line for write charge time
t_{RW}	$30ns$	word line pre-charge time
t_{HW}	$25ns$	word line for hold charge time
t_{WW}	$25ns$	word line for write charge time
t_{HM}	$64ms$	maximum hold time for data persistence

Table 5.1: time parameters for dram memory devices operating at 50ns from ITRS standard.

5.2 SOI MOSFET devices

A Silicon On Insulator wafer (SOI) consist in a silicon wafer with a buried SiO_2 insulator. Hence, the SOI wafer is constituted by three layers: a silicon substrate at the bottom, a silicon oxide layer, also called BOX; and a silicon layer at the top. The BOX and silicon layer thickness are in the order of hundreds of nanometers.

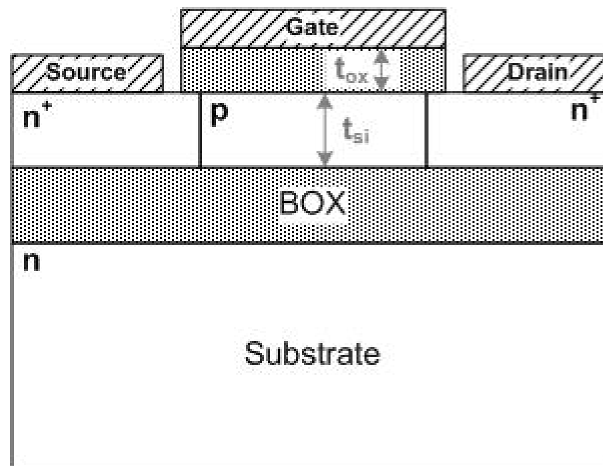


Figure 5.2: Cross section of a SOI MOSFET.

The realization of a MOSFET device on this kind of substrate involves the construction of a very thin device enclosed between two oxide layers, and therefore completely electrically isolated. The schematic cross section of a SOI device is shown on figure 5.2. This kind of isolation allows avoid-

ing the latch-up between adjacent devices, and allow to store charge in the silicon body under the gate. A SOI MOSFET device contains two MOS structures: the first is the classic MOS structure device-related and composed of the gate contact, the gate oxide, and the top layer of silicon. The second structure is a mirrored MOS down and start with the top layer of silicon as a semiconductor, the BOX as oxide, and the substrate as metal. The SOI MOSFET physics is strongly dependent on the thickness and doping concentration of silicon layer under the field oxide t_{si}. Evan based on doping concentration. This dependence is related to semiconductor MOS depletion region that for a classical MOS structure can be evaluated as

$$W_D = \sqrt{\frac{2\epsilon_s \phi_s}{qN_A}} \tag{5.1}$$

where ϕ_s is the potential at semiconductor-oxide interface. With respect to t_{si} it's possible to define two main types of SOI MOSFETs: Partially Depleted (PD) or Fully Depleted (FD) SOI devices.

In a PDSOI device, the silicon layer thickness t_{si} is larger than twice the value of W_D. In this case, there is no interaction between the depletion regions arising from the front interface and the back interface due to the buried oxide. The presence of a sufficiently large area of silicon allows the two regions to remain independent of each other, and therefore do not perceive any change in power due to the interface adjacent to

In a FDSOI device the silicon layer thickness t_{si} is smaller than W_D. In this case, the silicon layer is fully depleted at threshold. FDSOI devices in the upper and lower interfaces do not have enough space in the region of silicon to keep electrically independent of each other, therefore, any variation on one of the two interface influences the second interface. A bulk MOSFET, and a PDSOI MOSFET can be schematized with an equivalent capacitor network shown in figure 5.3(a). In such a network a variation of the gate voltage dV_G corresponds to a variation of the surface voltage $d\phi_S \leq dV_G$.

$$\frac{dV_G}{d\phi_S} = 1 + \frac{C_D}{C_{OX}} \tag{5.2}$$

This equation show as there is no dependence from bottom voltage in PDSOI. For a FDSOI device the schematic diagram changes to figure 5.3(b), and therefore the gate voltage variation $d\phi_S$ can be connected to the voltage of the bottom gate voltage using the following formula

$$\frac{d\phi_S}{dV_{G2}} = -\frac{C_{Si}C_{OX2}}{C_{OX1}\left(C_{Si} + C_{OX2}\right)} \tag{5.3}$$

where

$$C_{OX1} = \frac{\epsilon_{OX1}}{t_{OX1}}, \qquad C_{OX2} = \frac{\epsilon_{OX2}}{t_{OX2}}, \qquad C_{si} = \frac{\epsilon_{si}}{t_{si}} \tag{5.4}$$

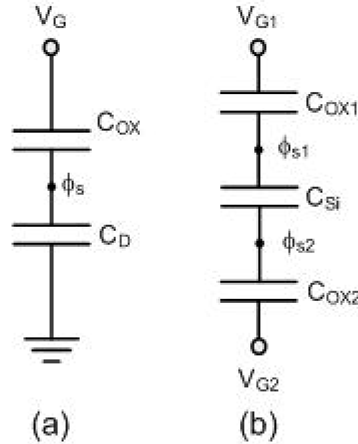


Figure 5.3: Equivalent capacitor networks for a PDSOI top interface MOSFETs (a) and a FDSOI MOSFETs (b).

In most cases equation (5.3) can be approximated by

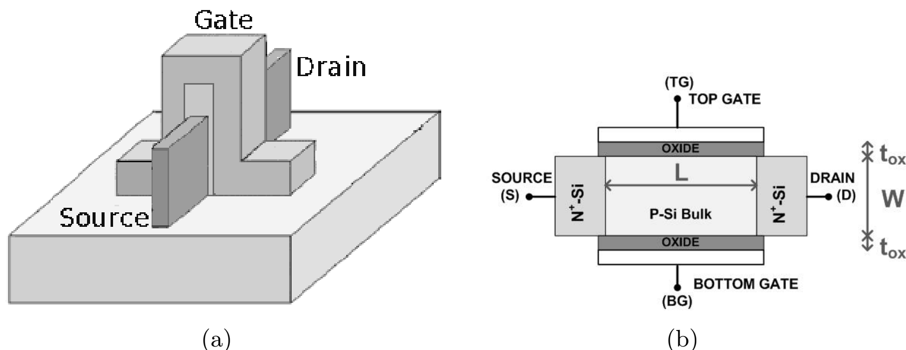
$$\frac{d\phi_S}{dV_{G2}} = -\frac{t_{OX1}}{t_{OX2}} \tag{5.5}$$

That for a SOI device, in which $t_{OX1} < t_{OX2}$ will translate in a smaller effect on ψ_S of bulk voltage.

5.3 FinFET devices

Another kind of devices that could be used to implement a ZRAM memory is the FinFET, the classical device's structure with 3 terminal is shown in figure 5.4a. The FinFET term was coined by Dams Hisamoto to describe a double gate transistor made of non-planar SOI substrate based on the design of transistor DELTA (fully Depleted Leanchannel transistor). The main feature that distinguishes a FinFET device from a planar device is that the conduction channel is formed around a silicon “Fin”, which constitutes the device's body. The fin's body is covered by the gate oxide on the two side walls, while gate oxide is substantially thicker at the top, which allows to neglect the upper MOS structure. From a perpendicular planar section to the wafer take at the FinFET's middle shown at figure 5.4b you can see that the FinFET structure is actually composed of two parallel gates that share the bulk region, the two gate oxides also have the same thickness, so the resulting structure is symmetrical, unlike the SOI MOSFET where the top gate has a greater effect than the lower one, due to the gate thickness. As can be seen from figure 5.4a, the two gates are shorted by the polysilicon

gate that surround the structure, this device characteristic is not desirable to use it as a ZRAM, perhaps the two gates must be electrically separated, this operation must be made by a etching on the device's top polysilicon.



(a) (b)

Figure 5.4: (a) Schematic representation of a FinFET device. (b) Two-dimensional FinFET device cross section, section perpendicular to Wafer plane taken at fin's middle.

5.4 Bipolar mode description

Two different types of Z-RAM configurations have been proposed. In the first generation of Z-RAMs (type I), the electron-hole pairs are created by impact ionization and/or BTBT of inversion charges at the drain end of the transistor. Instead, the stored charge in type II Z-RAM cells consists of the majority carriers injected from the source into the bulk and subsequently collected by the drain field [60–62]. The cell works like a bipolar transistor, with the source as the emitter, the bulk as the base, and the drain as the collector; The base charge is maintained by the extra charge generated by impact ionization at body-drain junction. The gate oxide and the SOI substrate prevent the leakage of the excess charge stored in the bulk by applying appropriate potentials at the electrodes. Reports in the literature suggest that type-II ZRAM have higher READ/WRITE currents, programming window, and retention time compared to the type I mode [16]. However the physical mechanism of its operation and the ultimate scalability of the devices have not been studied in details.

Figure 5.5 show the reference Z-RAM cell used in this work, this kind of device consists of a Double Gate (DG) SOI MOSFET with the two gate contacts connected to the word lines and the source and drain electrodes connected to the bit lines. The device length (L) and device width (W) are both 100nm, the gate oxide thickness (t_{ox}) is 10nm, the buried oxide

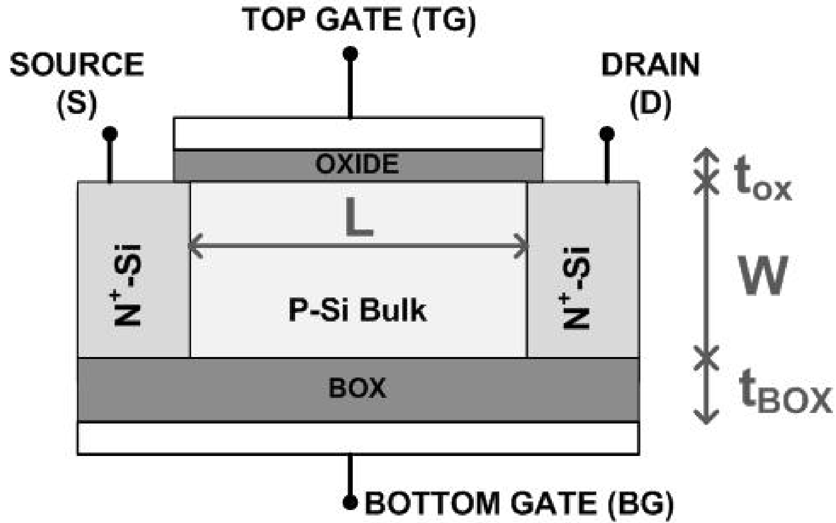
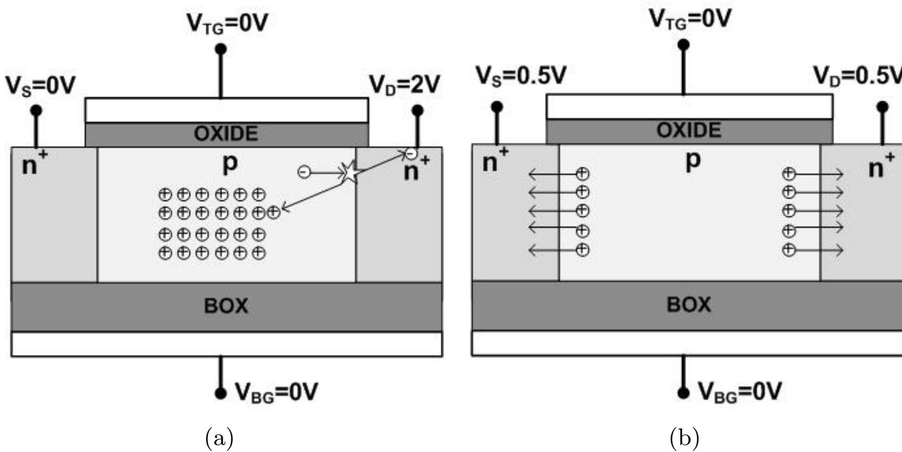


Figure 5.5: Schematic cross section of device under test, consisting of a 2D double gate SOI n-MOSFET with device length $L = 100nm$, device width $W = 100nm$, gate oxide thickness $t_{ox} = 10nm$, buried oxide thickness $t_{BOX} = 30nm$ and bulk doping $N_B = 10^{17}cm^{-3}$. Due to the short L the bulk is deeply depleted at equilibrium (all potential equal to 0) and the average hole concentration in the bulk is around 10^6cm^{-3}.

thickness is larger and is $t_{BOX} = 30nm$, and the bulk doping is $10^{17}cm^{-3}$. This structure has been tested to be a good reference structure in order of drain read/write operations. Two dimensional device simulations were performed to analyze the operation of the cell and, in addition to the typical drift diffusion transport model, impact ionization and BTBT models have been explicitly included. The drift-diffusion model has been replaced by the Poisson-Schrödinger model when device dimensions gets lower enough to present quantum confinement problems (length¡15nm). At equilibrium, the potentials of all the electrodes are set to zero (in the following we will refer to this condition as INIT phase). The device is fully depleted due to the short L and the maximum hole concentration in the bulk is around 106 cm-3. Fig. 2 shows the bias configurations applied to bit lines and to word lines in the four operation modes. Let us emphasize that no particular transient combination of the electrode potentials is required, since the electrode potentials are simultaneously changed from one operation mode to another. In the next subsections, we will discuss the physical mechanisms of WRITE ”1”, WRITE ”0”, READ and HOLD operation modes and the timing analysis of the Z-RAM cell under investigation.

5.4.1 Write operation

As illustrated in Figure 5.6a, during the WRITE “1” mode (W1), the two gates and the source are held at the same bias which is smaller with respect to the drain bias. The applied bias prevents the creation of inversion charge at the interfaces. Electrons are injected through the source-bulk energy barrier and are collected by the relatively high potential at the drain contact in a manner reminiscent of typical bipolar junction transistors. Due to the high longitudinal electric field at the bulk drain junction, excess electron-hole pairs are created via impact ionization and BTBT processes. Excess electrons are pushed out from the bulk towards the drain due to the favorable field, while excess holes are pushed towards the source and are trapped in the bulk if appropriate bias is applied to electrodes.



(a) (b)

Figure 5.6: Bias configuration and charge representation during WRITE “1” (a), and WRITE “0” (b) operations. During WRITE “1”, impact ionization and BTBT at the drain side create a lot of excess positive charge which is distributed over the entire volume of the bulk. During WRITE “0” the possibly accumulated holes are pushed out from the interfaces towards the source and the drain due to the reduced energy barriers.

Figure 5.6b illustrates the bias configuration applied to the structure in the WRITE “0” mode (W0). Source-bulk and drain-bulk energy barriers are lowered by applying negative potentials to source and drain with respect to top and bottom gates. This allows removal of the charges previously stored in a eventual WRITE “1” operation.

5.4.2 Hold operation

As illustrated in figure 5.7, during the HOLD mode (H), a negative gate voltage is applied for both gates with respect to source and drain bias that is lightly positive. This bias sets a stationary accumulation condition for the two interfaces. However, if the HOLD operation follows a WRITE “0” operation, the stationary condition is reached only after a certain delay because the bulk is deeply depleted and within this delay the cell remains in the state “0”. As will be discussed in details in section 5.5, the holes necessary to satisfy the accumulation condition are generated by the slow thermal generation processes or BTBT between the source/drain to bulk junctions.

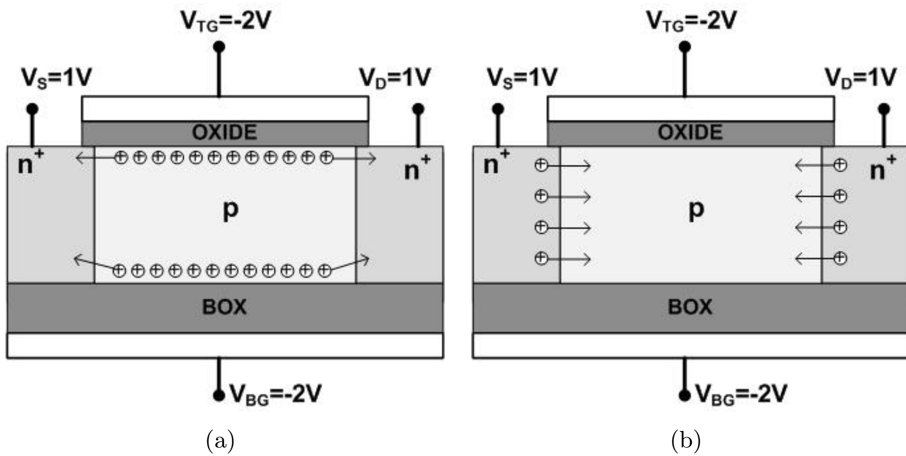


Figure 5.7: Bias configuration and charge representation during HOLD after write “1” (a), and HOLD after write “0” (b) operations. The charge accumulated at the two interfaces during the HOLD operation (a) is the charge accumulated during write “1” and it tends to slowly escape from source and drain. HOLD after write “0” (b) suffers from the leakage generated by thermal generation and/or BTBT which increases the hole concentration in the bulk. The two hold conditions show the same voltages.

On the other hand, if the HOLD operation follows a WRITE “1” operation, holes are already present in the bulk because they have been generated by impact ionization and/or BTBT so that the stationary condition is reached very fast. Most of the excess charge generated during the WRITE “1” phase is lost by leakage through the bulk to source/drain energy barriers, and the rest of the charge is accumulated at the two interfaces so that the self consistent accumulation condition imposed by the top/bottom gate to source/drain bias is satisfied. This accumulation charge represents the

information relative to the state “1”.

A number of observations can be made about the operation modes of ZRAM-II cells. First, the amount of charge representing the state “1” does not depend on the actual charge generated during WRITE “1” but only on the negative top/bottom gate to source/drain bias applied during the HOLD mode. The only role of WRITE “1” is speeding up the hole generation rate in the bulk. This means that it is not important to tune exactly the impact ionization/BTBT models in the device simulator nor it is necessary to determine the exact value of the generated excess charge. A further advantage of the proposed approach is that, because the value of the accumulated charge is self-consistent with the applied bias, in the HOLD mode the average hole concentration remains nearly constant for an infinite time. In other words, the retention associated to the state “1” is infinity. The retention of the cell information is thus limited by the leakage of state “0” as will be discussed in the subsection 5.5.

5.4.3 Read operation

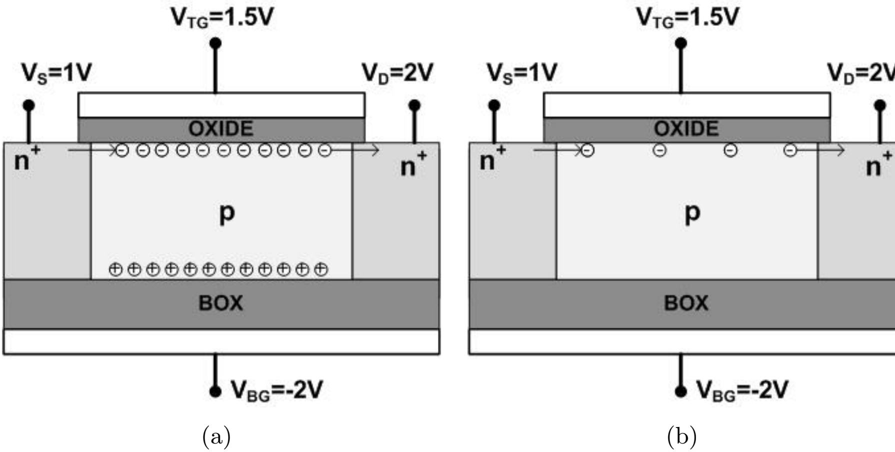


Figure 5.8: Bias configuration and charge representation during READ with a “1” stored on device (a), and during READ with a “0” stored on device (b) operations. During READ “1”, the charge stored at the bottom interface increases the bulk potential at the top interface which in turn reduces the source-bulk energy barrier allowing a high READ current. If the READ operation follows a WRITE “0” operation (b) no charge is present at the bottom interface and the high source-bulk energy barrier at the top interface causes a lower READ current.

As shown in figure 5.8, to implement the READ operation mode (R),

the two interfaces are biased asymmetrically and operated in different ways. The top interface works in a way similar to the WRITE "1" mode except that the bulk-drain reverse bias is not large enough to produce excess charge (READ disturb). On the other hand electrons are injected from the source to the bulk and collected from the drain field. The bottom interface works as in HOLD mode (i.e. an accumulation condition), since similar potentials are applied at the source and gate electrodes.

If the cell is in the state "1", the accumulation charge at the top interface is lost during READ operation, because of the reduced source-bulk barrier. The accumulation charge at the bottom interface, however, is maintained because the bottom interface is in HOLD mode. The excess charge at the bottom interface increases the potential locally and if the two gates are close to the other, the potential at the top interface is raised as well (gate coupling). The higher potential at the top interface reduces the source bulk energy barrier and more electrons are injected, which in turn increases the drain current.

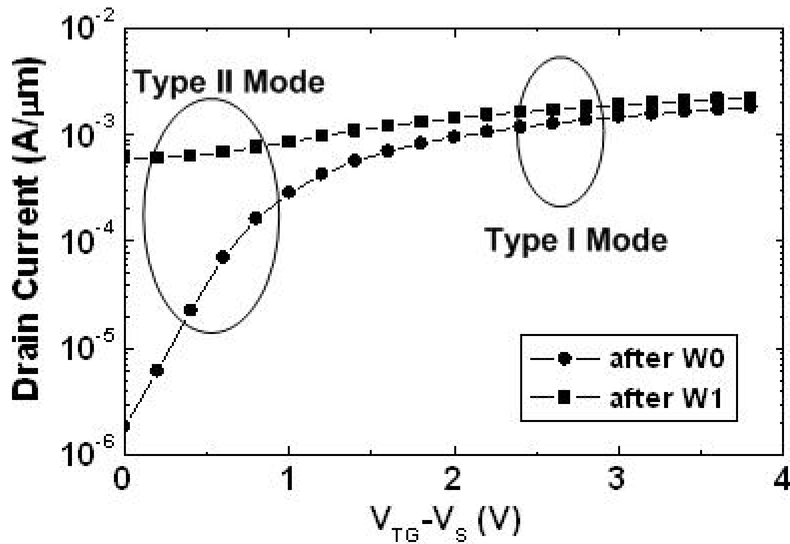


Figure 5.9: Drain current after WRITE 0 (I0) and WRITE 1 (I1) after a HOLD time of 100ms as a function of the top gate voltage respect to the source potential ($V_{TG} - V_S$). The source, drain and bottom gate voltages are the same for the READ mode.

It is interesting to compare the READ currents after WRITE "1" (I1) and after WRITE "0" (I0) between the type I and the type II operation modes (Figure 5.9). For this purpose, we evaluated I1 and I0 as a function

Operation	V_{TG}	V_{BG}	V_S	V_D
Write “1” (W1)	$0V$	$0V$	$0V$	$2V$
Write “0” (W0)	$0V$	$0V$	$0.5V$	$0.5V$
Hold (H)	$-2V$	$-2V$	$1V$	$1V$
Read (R)	$1.5V$	$-2V$	$1V$	$2V$

Table 5.2: Voltage setup for all the operation mode of ZRAM type II SOI MOSFET devices.

of the top gate voltage with all the other electrodes biased as in the READ mode. At lower (higher) VTG-VS the Z-RAM cell works in type II (type I) mode since the device is biased in the sub-threshold (inversion) regime.

It is apparent that by moving from type II towards type I operation mode, the ratio I1/I0 is reduced with the corresponding loss of the programming window. This result reflects the different physical mechanisms governing the impact of the stored holes on the READ current. In the inversion regime of the type I cells, the stored holes increase the current due to the reduction of threshold voltage through the classical body effect. Instead, in the sub-threshold regime of the type II cells, the stored holes increase the READ current by lowering the source-bulk energy barrier. Since the current depends exponentially on the bulk potential below threshold (barrier lowering effect) and sub-linearly above threshold (body effect), a higher programming window is observed for the type II operation mode.

5.5 transient analysis

All the operation that must be made to make the device a memory cell has been descried and summarized in table 5.2, it must operate with respect to the ITRS limit specified in table 5.1. In order to obtain good simulation results it must be made a good transient setup, describing the time evolution of signal input and the logical level that we must use.

In order to well approximate the cell behavior, some cell external considerations must be made. As can be seen from section 5.4 the single cell allow us to use a single device as a memory cell, perhaps we can remove the capacitor used to store the memory bit and store information inside the MOSFET directly. But the new voltage requirements lead to an increase in the number of word line and bit line, this is because of a single cell require to have voltage variations on two gates instead of one, increasing the number of word lines from 1 to 2, and the source that is no more connected to the capacitor, but need to e connected to a new bit line. The new cell configuration requires to include a parasitic capacitor for each line, one for

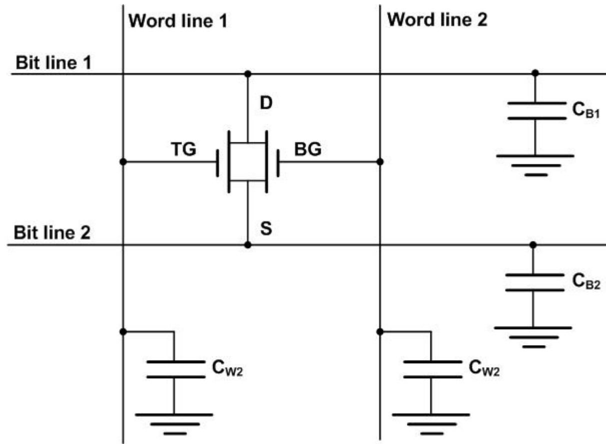


Figure 5.10: DRAM memory cell representation with ZRAM technology. Note that for a single cell there are two bit lines and two word lines, everyone with its parasitic capacitance, and everyone with its voltage potential to be applied.

every contact.

To consider the control electronics used for word and bit lines we must correctly determine the temporal sequence of the four input signals (D, S, TG, and BG) to device. Figure 5.11 shows the time course of Drain voltage input in a normal cycle of use of a memory cell ZRAM. In particular it is possible to note that in normal conditions, the Drain terminal is maintained at a voltage of hold, then since the beginning of the write voltage is adjusted to the desired voltage using a linear ramp of duration t_{WP}. Once you reach the desired voltage, the device will be able to write for a time t_{RC} required to accumulate charge in the bulk region of a scripture in the case of state 1, or to empty the residual charges in the case of writing a 0.

Once finished writing, the memory is reported in the state of hold following a further linear ramp lasting t_{HP}. The hold condition is maintained until it is required to read the data, or if it reaches the maximum residence time of the data, and must therefore proceed with the refresh operation, the worst operating conditions in the hold time must be equal to t_{HM}. After the hold period begins reading operation with the increase of drain voltage for a period of t_{RP} and maintaining the condition of writing for a t_{RC} time enabling the parasitic capacitance of the bit line to load the value voltage on the logic state stored in memory. Once the memory read operation is shown in the state of hold and the operation is repeated by writing the data.

The same operation description could be performed in the case of source terminal, while the two gate terminals (BG) and (TG) show the same operative setup with different ramp times determined respectively from the

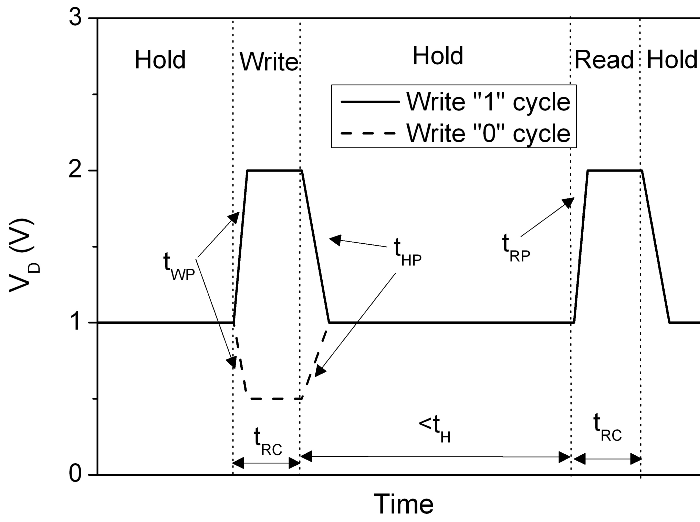


Figure 5.11: Time evolution of Drain voltage for a Write, Hold, Read cycle in both cases of Write 1 and Write 0. On the time flow are presented the main time constraints, and the rise and fall curves are approximated to an ideal linear curve.

write, read and hold time on to the word line (t_{WW}, t_{RW}, and t_{HW}).

Chapter 6

Capacitorless simulations

Once defined the main characteristics of ZRAM type I and type II, in order to verify device functionalities and to obtain the device scaling a good simulation setup with an opportune mesh definition and a good physical model must be made. Once validated our model a simple operative test has been made. Starting from this test structure has been made the scaling analysis in order to understand the length scaling limits due to short channel effects like Drain Induced Barrier Lowering, velocity saturation, and impact ionization [63]. The simulation setup make a comparison to current value in READ “1” and READ “0” values after a maximum hold time defined by ITRS [59] to verify data persistence. Within a scaling limits considerations the main leakage problems has been evaluated in order to better understand the parameters to modulate to obtain better results. As last characterization has been evaluated the functionality window to identify working devices and find the minimum technological length that limit such technology.

6.1 simulation setup

The first step to be made is to design a good device structure, as described previously, device structure must be a four terminal FinFET with the two gate oxide electrically separated, in order to recognize it, the two gates will be called top gate, and bottom gate. The device's mesh must be finer enough to ensure a good analysis of local spatial charge inside bulk area, the silicon p-type region, and finer enough near the pn junctions. The reference mesh is shown in figure 6.1, where contacts are shown in gray; in order to make faster simulations the Gate polysilicon regions has been removed assuming that gate potentials are applied directly to gate oxide by applying a potential that take into account of polysilicon built-in potential calculated as:

$$\psi_{poly} = -\frac{kT}{q} ln\left(fracN_{poly}n_i\right) \tag{6.1}$$

where N_{poly} is the polysilicon n-type doping concentration, and ψ_{poly} is the polysilicon built-in potential that must be add to gate potential.

The mesh inserts several points in the oxide regions to allow a calculation of better gate leakage current, such as Fowler-Nordheim tunneling, or direct tunneling. The simulations involve the use of the drift-diffusion for the case of large structures ($W_{min} = 25nm$) in the case of very small structures, by reducing device width over the previous limit implies the semiconductor charge confinement, this phenomena require the use of the self-consistent Schrödinger-Poisson model to better assess the reduction of the density of states along the vertical direction.

The first simulated structure complies with the characteristics described in Table 6.1, the simulation was carried out in the time domain by using the potential described in the table 5.2.

Name	Value	Description
t_{ox}	$10nm$	Gate Oxide thickness
W	$100nm$	Gate over Source/drain overlap length
L	$100nm$	Gate length
L_{DD}	$10nm$	Gate over Source/drain overlap length
$N_{S/D}$	$1e^{20}cm^{-3}$	Source and Drain pad n-type doping concentration
N_{poly}	$1e^{20}cm^{-3}$	Gate polysiliocn n-type doping conc.
N_B	$5e^{17}cm^{-3}$	Body p-type doping concentration
ψ_{poly}	$0.59V$	Gate built-in potential

Table 6.1: Geometrical and doping concentration values for starting structure

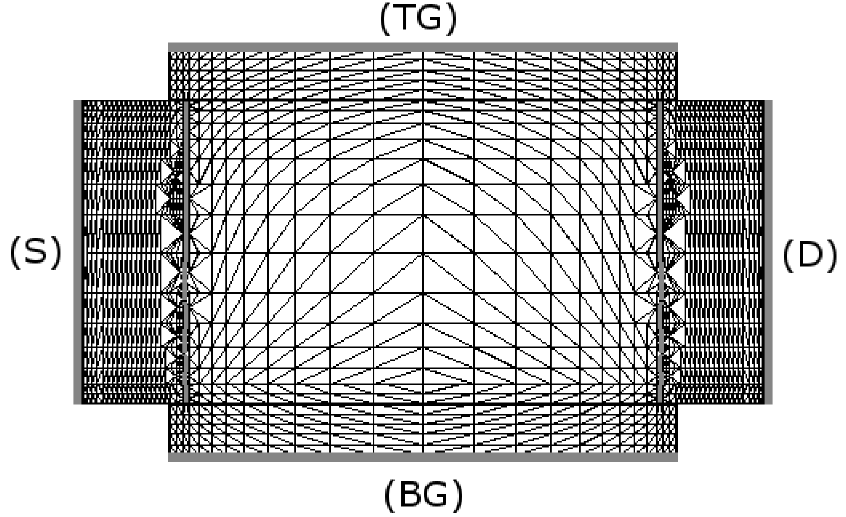


Figure 6.1: Mesh definition in reference ZRAM structure, where the top gate and bottom gate are respectively on device top and bottom, source is on the top-left, and drain is on the top-right.

To evaluate the variations of the internal device is necessary to know the amount of charge stored in the device, which is the number of gaps in the region of body. gaps must be integrated to calculate the concentration of holes in both directions. In the simulations discretized the integrals are approximated by the sum of the concentration values ??for the charge mesh spacing horizontally and vertically.

$$holes = \int_0^W \int_0^L p(x,y)dx, dy \cong \sum_{N}^{i=1} \sum_{M}^{j=1} n(i,j)\Delta X_i \Delta Y_j \tag{6.2}$$

This approximation could be made if mesh points are distributed with a rectangular order, as for figure 6.1. The concentration of charge in the hold phase, together with the drain current during the Read are the two most important parameters to consider during the analysis of the device.

6.1.1 Initial simulation

In figure 6.2, we plot as a function of time the hole concentration in the bulk (a) and the READ drain current (b) for a typical sequence of memory operations. At time $t = 0s$ the cell is biased in the INIT mode with all electrode potentials equal to zero, then we pass into the HOLD state, in this initial phase, the device is completely depleted, and the bulk charge

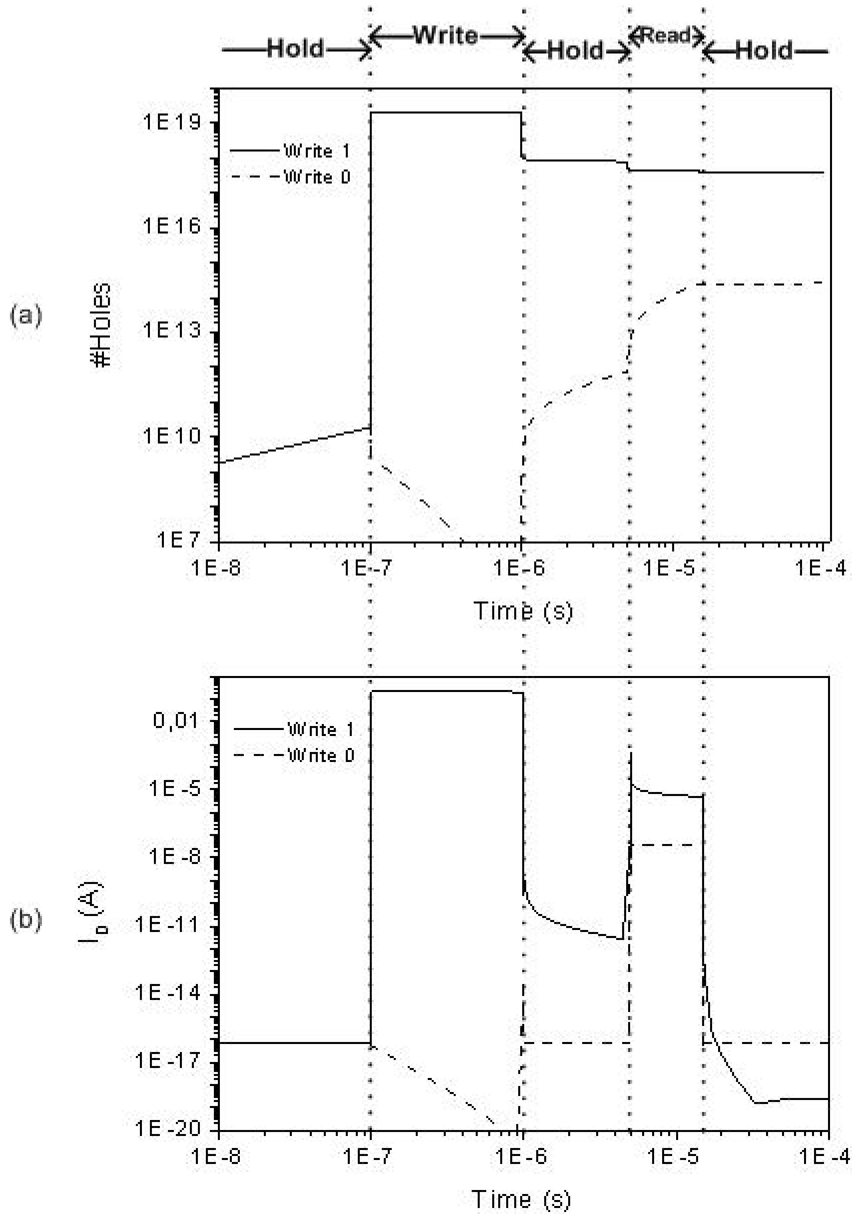


Figure 6.2: Timing of the average hole concentration in the bulk (a) and of the drain current (b) for a typical sequence of memory operations including read and write operations for both 1 and 0 data.

concentration is very low (10^{10} holes). At time $t = 10^{-7}s$, a WRITE operation is performed and device behavior is different depending on the data written.

In the write "1" case, a lot of excess charge is created (around 10^{19} bulk holes). Also the drain current is very high due to the large number of excess electrons generated. At time $t = 10^{-6s}$ the WRITE "1" bias is removed and the HOLD bias is applied. The hole concentration reduces very fast, most of the generated charge is lost by source/drain leakage, and the remaining charge is accumulated at the two gate interfaces. Note that in Figure 6.2a, the average hole charge during the HOLD mode ($t > 10^{-6}s$) remains constant independent of time because of the self-consistent accumulation condition at the two interfaces. At time $t = 5X10^{-6}s$, the READ bias is applied. The average hole concentration reduces roughly by a factor two because of the charge lost at the top interface through the reduced source-bulk energy barrier. The READ drain current is much higher with respect to the READ drain current after INIT because of the stored charge at the bottom interface which is maintained during the READ operation.

In the write "0" case, the holes stored at the interfaces rapidly escape through the reduced source bulk and drain-bulk energy barriers and the average number of holes in the bulk reduces with respect to the INIT phase. At time $t = 10^{-6s}$ the HOLD bias is applied. As already stated, the number of holes increases in the depleted bulk due to thermal generation and BTBT. Finally, at time $t = 5X10^{-6}s$, the READ bias is applied. The number of holes increases (READ disturb) because the bottom interface is in the HOLD mode and BTBT occurs. In order to reduce READ disturbs after a WRITE "0"

i) the READ pulse should be as short as possible,

ii) the source-bottom gate bias should not be too high to enhance BTBT rates.

The READ current in the state "0" is very similar to the READ current after INIT, but it is much lower with respect to the READ current after WRITE "1". This difference in current magnitude allows a high READ sensitivity and the reading mode that switches from voltage to current suppress the problem related to the parasitic capacitance word line limiting the device number on the same line.

In order to understand device reading operation a 1-D cutline has been made, the better cutline pass through one gate to another and is placed at gate center in order to visualize the maximum effect of gate charge inside body region. Figure 6.3 show the 1-D cutline previously described showing current density versus Y position, The cutline starts its zero coordinate on

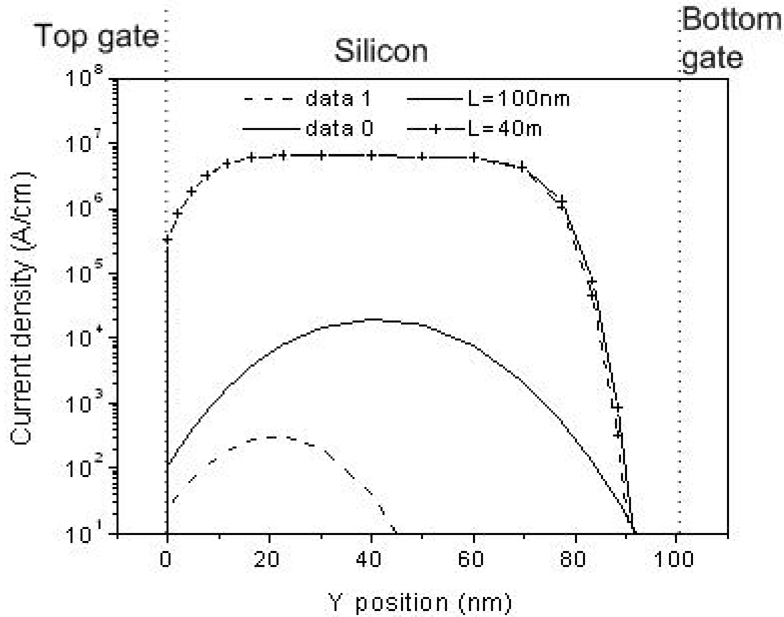


Figure 6.3: Current density cutline from top gate to bottom gate, taken at gate's middle for both conditions of 1 data and 0 data after a hold time of 100ms and for two different structures; withouth symbol for L=100nm and with symbol for L=40nm.

top gate-semiconductor interface and ends on the bottom interface. The current density is related to parasitic BJT, is higher at device's center and equal to zero at bottom gate interface due to gate potential that cause the charge retention. For a structure with $L = 100nm$, the read after a write "1" show a higher density current than the same read after a write "0". This current variation is generated by the increasing potential at fin's center caused by a higher hole concentration, as showed in figure 6.4. The behavior noted above is no longer visible if you change the length of the device bringing it to $L = 40nm$. In this case, the short channel effects tend to cancel out the difference between the two cases "1" and "0" and make the device ineffective.

The results in terms of the difference current and concentration of gaps between the two states depends strongly on the hold time between the write and read later. As shown in figure fig:Ztiming the electrical characteristics are measured after the hold operation after 50ms from the read operation beginning, the time elapsed since the reading operation start is the hold time, so we can analyze electrical characteristics as a function of hold time.

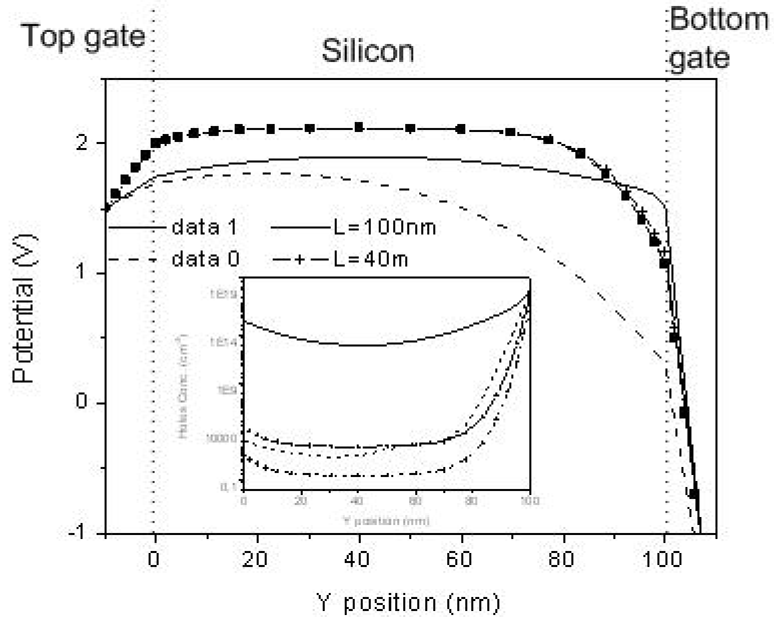


Figure 6.4: Potential cutline from top gate to bottom gate, taken at gate's middle for both conditions of 1 data and 0 data after a hold time of 100ms and for two different structures; without symbol for L=100nm and with symbol for L=40nm. Inset figure as hole concentration.

6.1.2 first result

In figure 6.6 the READ currents and the average hole concentration in the bulk after WRITE “0” and WRITE “1” are shown as a function of the HOLD time. The reduction of current (hole concentration) after WRITE “1” is due to the source/drain leakage of the residual charge stored in the bulk while the leakage associated to state “0” is due to hole injection into the bulk caused by BTBT and thermal generation. As already discussed in the previous chapter, the retention of the cell is limited by the leakage associated to the state “0”. This leakage is due to two different processes occurring during the HOLD mode: the BTBT at the source/drain and source/bulk junctions and the thermal generation processes in the bulk. In figure 6.7 device simulations have been performed switching ON and OFF the model for BTBT and the average hole concentration in the bulk is shown as function of the retention time for two different oxide thicknesses. A stronger leakage associated to the state “0” is observed when BTBT is ON, while the thermal generation is responsible of the leakage in the case of BTBT OFF. As shown in the figure 6.7 and as we will explain in the section 5.4.2, the impact of BTBT on the leakage in the state “0”

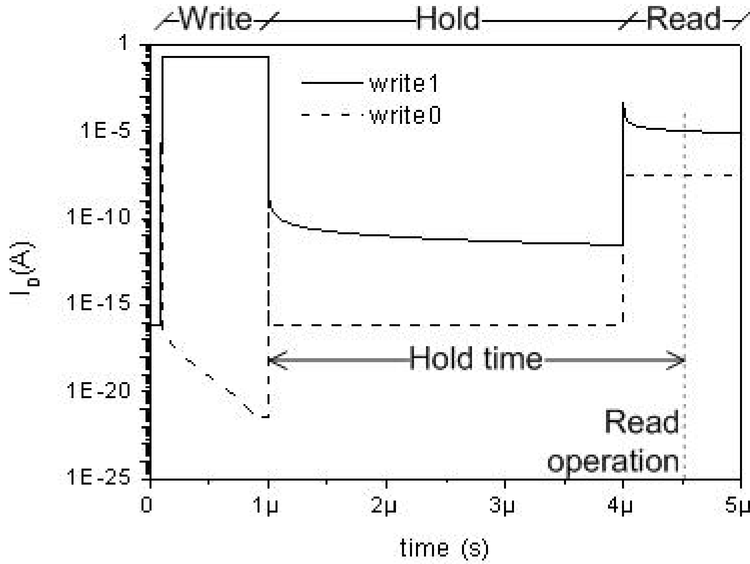


Figure 6.5: Timing sequence of the simulation setup for write "1" and write "0" conditions with particular emphasis on hold time used in sequent analysis.

increases by scaling the transversal geometries (W and/or t_{ox}). In fact in figure 6.7 it can be observed that the impact of BTBT is higher in the case $t_{ox} = 7nm$ with respect to the case $t_{ox} = 10nm$. BTBT is associated to the field created by the energy barriers in HOLD mode. A more negative top/bottom gate to source/drain bias during HOLD mode increases the BTBT, the leakage of state "0" and the READ current after WRITE "0". On the other hand, a more negative top/bottom gate to source/drain bias during HOLD mode increases the accumulation charge after WRITE "1" and the READ current after WRITE "1" so that a compromise should be chosen for electrode potentials in HOLD mode in order to maximize READ sensitivity and retention time.

6.2 Scaling study of ZRAMcell

With the basic operation of the type II Z RAM cell established, it is important to explore the scaling potential of such a cell as a function of longitudinal (L) and transverse (W, t_{ox}) geometrical parameters. The key performance metrics of this study are the average hole concentration in the bulk and the READ currents corresponding to the two different states.

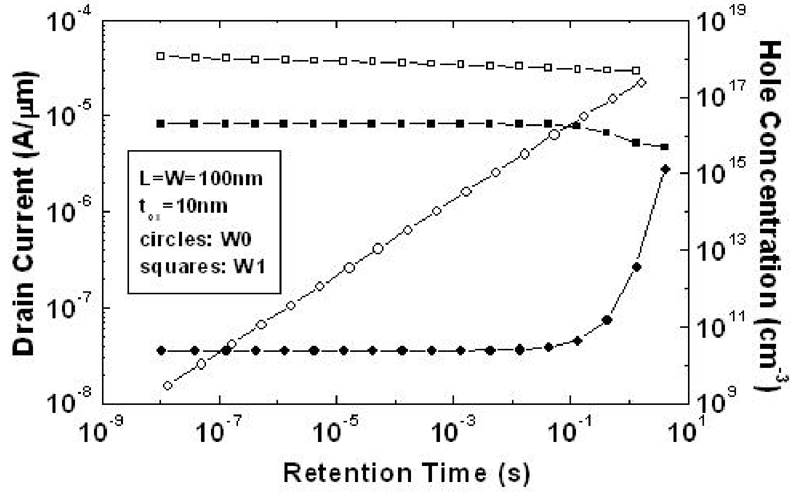


Figure 6.6: READ currents (filled symbols) and average hole concentration in the bulk (empty symbols) after a WRITE "1" and after a WRITE "0" as a function of the retention time for $L = W = 100nm$ and $t_{ox} = 10nm$. The retention of state "1" is limited by the slow source/drain to bulk leakage of the trapped charge, while the retention of state "0" is limited by BTBT and thermal generation.

6.2.1 Longitudinal scaling

In figure 6.8, we plot as a function of L the average hole concentration and the READ currents corresponding to the two different states after a retention time of $100ms$. As L is reduced, the hole concentration decreases for both the states due to the increase in bulk potential induced by enhanced short channel effects (SCEs). Moreover, since the effect of the bulk potential on the hole concentration is stronger in the state "1" due to the higher hole concentration, the difference in the hole concentration between the two states decreases with L. It is worth noting that at sufficiently small L, this hole concentration differential disappears since the steady-state condition at the interfaces does not correspond to the accumulation regime and the Z-RAM cell is unable to retain the charge at the interfaces. The READ currents for both states increase as the device length is reduced. This is due to the higher injection of electrons through the reduced source-bulk barrier at shorter device lengths. The ratio I1/I0, however, decreases rapidly due to the reduction of the difference in the hole concentration between the two states and due to the reduction in gate coupling, i.e. the influence of the charge at the bottom interface on the potential at the top interface.

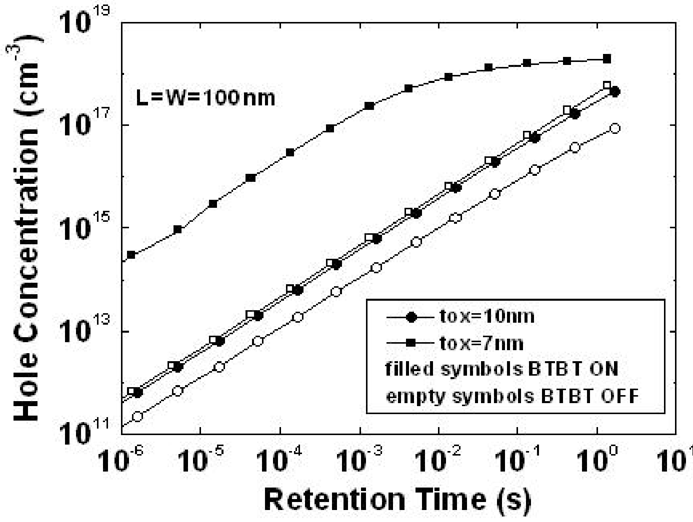


Figure 6.7: Evolution of the hole concentration in the state “0” for two different oxide thicknesses ($7nm$ and $10nm$) and by switching ON and OFF the BTBT model. A stronger leakage associated to the state “0” is observed when BTBT is ON, while the thermal generation is responsible of the leakage in the case of BTBT OFF. Moreover, the impact of BTBT on the leakage in the state “0” increases by scaling the transversal geometries (t_{ox} and t_{BOX} in this case).

Indeed the solution of the 2D Poisson equation suggests that the increased curvature of longitudinal energy bands associated with the higher SCE, produces a lower curvature of energy bands along the transverse direction, so that a charge at the bottom interface is less effective in changing the potential at the top interface (low gate coupling). For sufficiently small L, the differential of the hole concentrations vanishes and the two READ currents coincide. As we will discuss in the next subsections, this limiting value for L strongly depends on the transversal geometrical parameters, W and t_{ox}.

6.2.2 Transversal scaling

In figure 6.9 and 6.10, we plot the average hole concentration and the READ currents corresponding to the two different states after a retention time of $100ms$ as a function of W and t_{ox}, respectively. Reducing W (t_{ox}) is equivalent to increasing L, because it reduces SCE with corresponding increase in the hole concentration for both states. Both currents reduce because of the lower electron injection from the source into the bulk caused by the higher

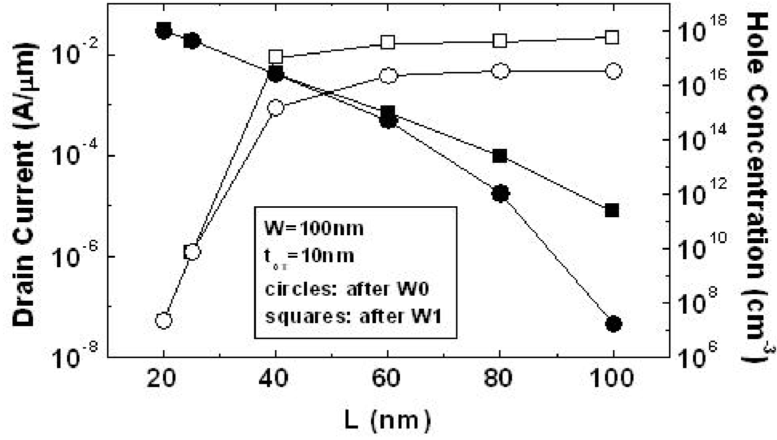


Figure 6.8: Drain current (filled symbols) and average hole concentration in the bulk (empty symbols) after a HOLD time of $100ms$ as a function of L after a WRITE “1” and after a WRITE “0”. As L is reduced, short channel effects become increasingly pronounced and the source-bulk barrier is reduced. The higher injection of electrons from the source increases both currents. The higher bulk potential associated with a lower L and enhanced SCEs reduces the average hole concentration.

source-bulk energy barrier. Moreover, in the case of W scaling, the current in state “1” reduces as well because the two gates are closely coupled and the charge stored at the bottom interface is lost during the READ mode due to the low source-bulk energy barrier at the top interface. As W (t_{ox}) is reduced, the gate coupling increases and the ratio I_1/I_0 increases as well. When W (t_{ox}) is sufficiently low, the steady state hole concentration in the bulk after WRITE “0” is so high that it is indistinguishable from the accumulation charge stored after WRITE “1” and as a result, the ratio I_1/I_0 starts to decrease. This high hole concentration is ascribed to the BTBT contribution. In fact, as W (t_{ox}) is reduced, the longitudinal field at the source/drain to bulk junctions increases thus enhancing the BTBT. Figure 6.7 shows the impact of BTBT on t_{ox} scaling (a similar effect has been observed on W). Device simulations were performed switching ON and OFF the BTBT model for $t_{ox} = 10nm$ and for $t_{ox} = 7nm$ and the average hole concentration in the bulk is plotted as function of the retention time. It is evident that the hole concentration increase due to the BTBT is stronger at lower t_{ox}.

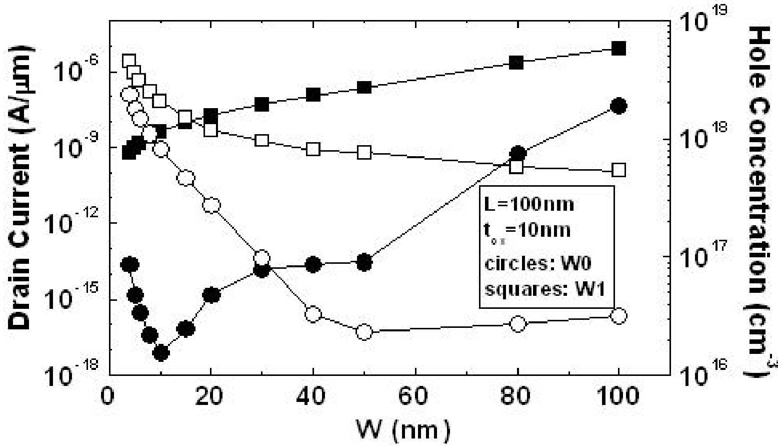


Figure 6.9: Drain current (filled symbols) and average hole concentration in the bulk (empty symbols) after a HOLD time of 100ms as a function of W after a WRITE "1" and after a WRITE "0". As W is reduced, SCEs are reduced and the source-bulk barrier increases. The lower injection of electrons from the source reduces both currents. The lower bulk potential associated with a lower W and reduced SCEs increases the average hole concentration.

6.2.3 scaling considerations

We now summarize the scaling properties associated to the DG Z RAM cell working in bipolar operation mode. Let us assume that a higher I_1/I_0 ratio can be used as a metric of higher READ sensitivity (noise margin), larger programming window, and higher retention time. Transverse scaling of device width and oxide thickness work in a similar manner in reducing SCEs: they both reduce the READ currents, but improve their ratio. Nevertheless, transverse scaling is limited by BTBT due to the high hole concentration in the state "0" which can become indistinguishable from the accumulation charge of state "1". As opposed to transversal scaling, the reduction of L increases the READ currents (I_0 and I_1), at the expense of a reduced I_1/I_0. To overcome this problem, W and/or t_{ox} must be scaled as well. The choice of the right geometry (L, W, t_{ox}) involves a tradeoff between current amplitudes and READ sensitivity (of the current sense amplifier). Figure 6.11 shows a summary of the scaling of the DG Z RAM cell in bipolar operation mode. Once L and t_{ox} are fixed, a minimum (empty symbols) and a maximum (filled symbols) W value are required to have:

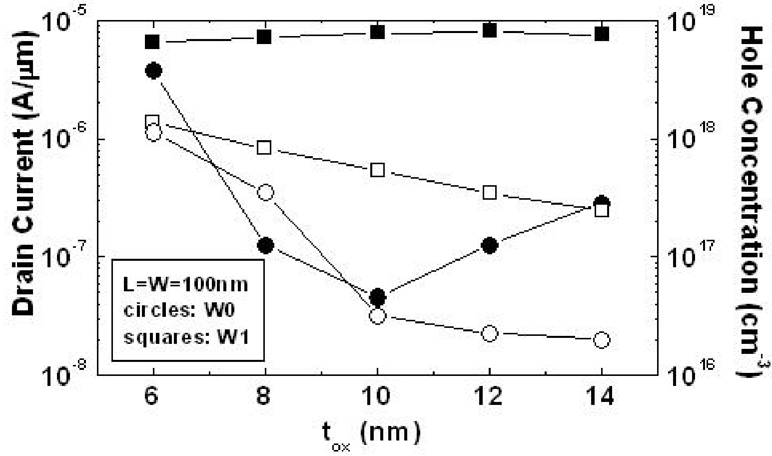


Figure 6.10: Drain current (filled symbols) and average hole concentration in the bulk (empty symbols) after a HOLD time of $100ms$ as a function of t_{ox} after a WRITE "1" and after a WRITE "0". As t_{ox} is reduced, SCEs are reduced and the source-bulk barrier increases. The lower injection of electrons from the source reduces both currents. The lower bulk potential associated with a lower t_{ox} and reduced SCEs increases the average hole concentration.

i) A retention time higher than 100 ms;

ii) A ratio $I_1/I_0 > 10$ at $t = 0s$ (just after WRITE "1").

The retention time is defined as the time necessary for I_0 (I1) to reach the average value I_M defined as

$$I_M = (I_0 + I_1)/2 \; at \; t = 0s \tag{6.3}$$

The presence of a minimum and a maximum for W are clearly evident from the discussion in the subsection 5.4.2 and from figure 6.9. Indeed as discussed in [55] regarding type I operation mode, the minimum W is expected to be imposed by the quantum limit to a few nanometers [64]. It is apparent that, as t_{ox} is scaled down, the minimum allowed L is reduced. We found that for $t_{ox} = 8nm$, a device length of $\approx 15nm$ may be acceptable. It is worth noting that the scaling limit of $15nm$ is lower respect to the $25nm$ found for type I operation mode [55]. As in the case of type I mode, the source/drain to bulk barrier lowering due to SCEs and the BTBT are the principal factors which limit the device scaling. Nevertheless, the higher

READ sensitivity, programming window and retention time of type II mode allows better scalability with respect to type I mode.

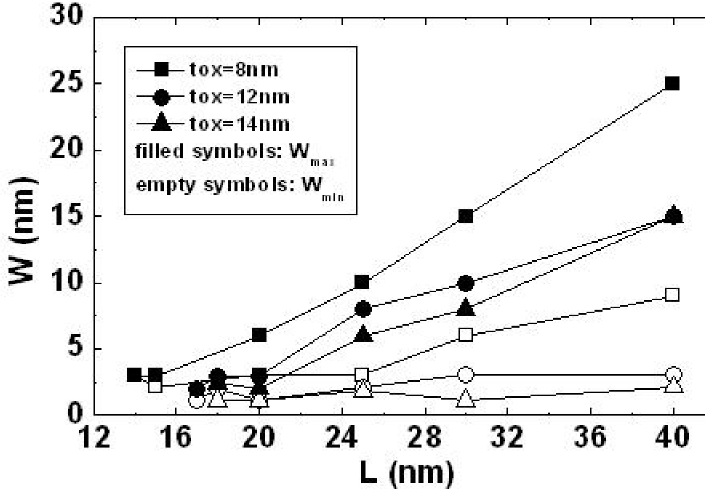


Figure 6.11: For each t_{ox} and L a minimum device width (W_{min}) and a maximum device width (W_{max}) are required to have a retention time of $100ms$. The minimum is due to the high hole concentration in the bulk after WRITE "0", associated with BTBT, and to the leakage of the bottom gate charge through the reduced source-bulk energy barrier at the top interface during the READ mode. The maximum is due to the enhanced SCEs for high W. As t_{ox} is scaled down the minimum allowed L decreases while W_{min} is not strongly affected. Let us not that W_{min} is indeed imposed by quantum limit to a few nanometers setting the minimum gate length to $\approx 15nm$.

Appendix A

FinFET simulations

Another simulation result obtained by simulation is the investigation of geometry dependence of the series resistance and the carrier mobility in FinFET devices by 3D simulations. The fin width dependence of the series resistance is investigated and extracted by a numerical method. The resistivity has been measured for different fin width, large and small.

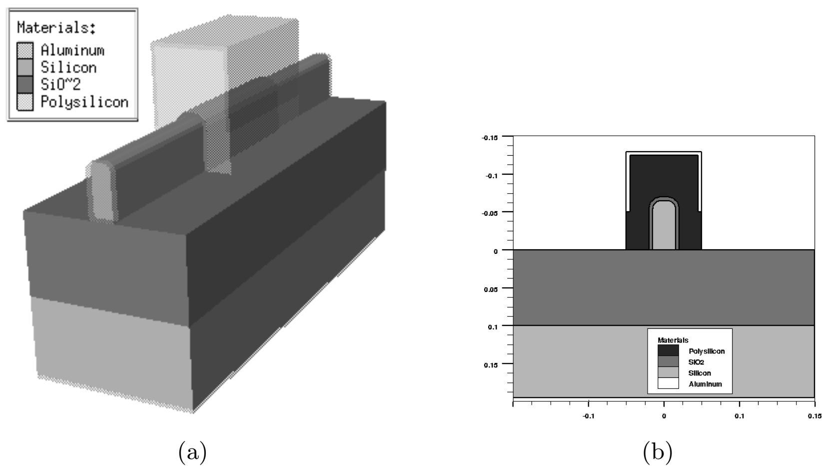


Figure A.1: (a) FinFET triple gate device under simulation. (b) Two-dimensional FinFET device cross section perpendicular to fin structure.

The characteristics of the FinFET structure have been described in chapter 5.3, the proposed structure consists of a device with triple gate FinFET, where the gate is placed on the lateral surfaces of the Fin and the upper side, the top edges are rounded, and the gate maintains a constant thickness along the edge of the fin [65]. The FinFET structure's description under simulation is described in Figure A.1.

The 3D simulations made are using the drift-diffusion model with lattice heating temperature model, the quasistationary solution of the problem has been made. Under these conditions a $id - vg$ measurements in order to calculate threshold voltage has been made. with resulting threshold voltage as $V_T = 0.347V$. Once determined threshold voltage a simulation with contacts potential value taken as $V_B = 0V$, $V_D = 1V$, $V_S = 0V$, $V_G = V_T + 1V$ has been made in order to evaluate the Source/Drain series resistance in ON conditions.

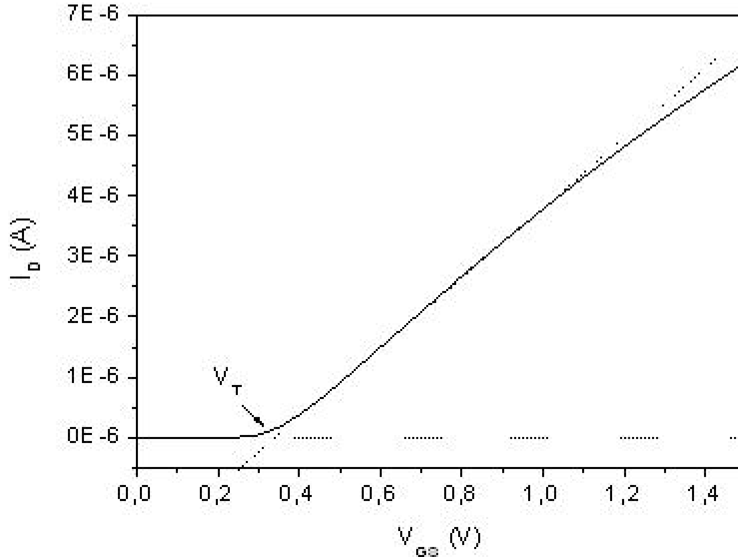


Figure A.2: $V_{GS} - I_D$ curve for FinFET triple fate structure showing a $V_T = 0.347V$

From the figure A.3 can be seen as the current density in the drain region is not constant in the device, but in the vicinity of the bulk-drain junction is thickened on the two oxide-semiconductor interfaces then open up and cover the entire fin. To calculate the resistivity of the device should be separated into two parts, one that takes into account the effect of widening the current, and the second consists simply of a block of semiconductor with resistivity constant [66].

The current is not constant within the source and drain regions, as can be seen from Figure A.3, at the Drain–Body junction, the current is concentrated at the two oxide–semiconductor interfaces occupying a constant section of the device, this effect will influence the resistivity of the material, which is calculated from this region thickness. away from the junction, the current tends to occupy a layer of silicon directly proportional to distance from the junction and along an angle *phi*. When the current occupies the half fin horizontal space, the device current can be considered constant

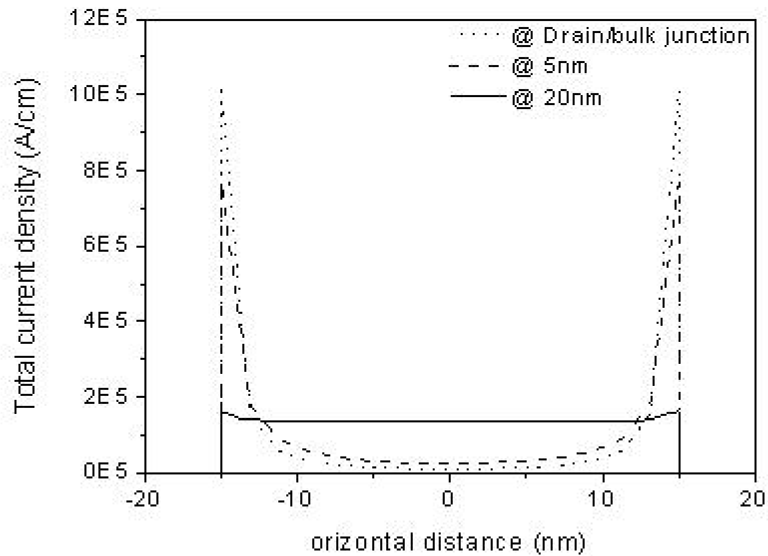


Figure A.3: total current density along a horizontal cutline from the two lateral fin's sidewalls. Section taken at Drain–Body junction, at 5nm from the same interface inside Drain region and at 25nm from the interface inside Drain region.

and the material show the classical resistivity [67]. The same analysis has been made on top interface, considering the same behaviour as the lateral sidewalls. Then the resistivity calculation must be made on a single gate-oxide two-dimensional interface and the results multiplied for the three gates width.

Using the method described above for calculating the resistivity has been possible to calculate the dependence of the resistivity as a function of range gate voltage and the device width [68, 69]. Analysis shows that increasing the gate voltage will decrease the angle phi, as results, the spatial region with current variations along fin axis will increase causing an increase in series resistance. This behavior can explain the strong dependence of series resistance from gate voltage.

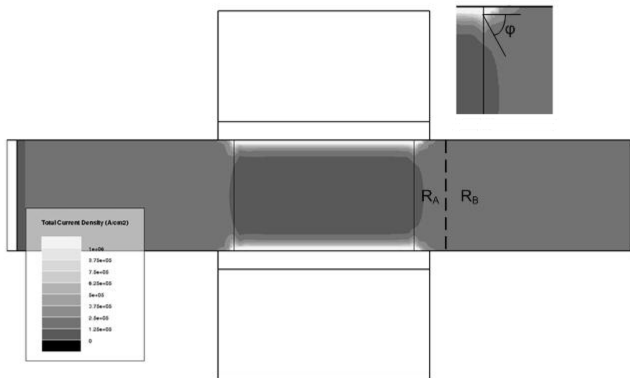


Figure A.4: total current density along a parallel plane to wafer at fin's center.

Appendix B

List of Symbols

Symbol	Description	Value	Unit
D_n	Electron diffusivity	36	cm^2/s
D_p	Hole diffusivity	12	cm^2/s
E	Energy		eV
E_C	Bottom of conduction band		eV
E_F	Fermi Energy level		eV
E_V	Top of conduction band		eV
ξ	Electric field		V/cm
G	Generation factor		cm^{-3}
h	Planck constant	$6.62X10^{-34}$	Js
$\hbar$	Normalized Planck constant	$61.05X10^{-34}$	Js
I	Current		A
J	Current density		A/cm^2
J_n	Electron current density		A/cm^2
J_p	Hole current density		A/cm^2
k	Boltzmann constant	$1.38X10^{-23}$	J/K
m_n	Electron effective mass	$4.73X10^{-31}$	kg
m_p	Hole effective mass	$1.46X10^{-31}$	kg
n	Free electron density		cm^{-3}
n_i	Intrinsic carrier density	$1.01X10^{10}$	cm^{-3}
N	Doping concentration		cm^{-3}
N_A	Acceptor doping concentration		cm^{-3}
N_C	Effective DOS in conduction band		cm^{-3}
N_D	Donor doping concentration		cm^{-3}
N_V	Effective DOS in valence band		cm^{-3}
q	Magnitude of electronic charge	$1.6X10-19$	C

Continued on next page

Symbol	Description	Value	Unit
R	Recombination factor		cm^{-3}
T_L	Lattice temperature		$\circ K$
V	Voltage		V
V_{BD}	Breakdown voltage		V
V_{th}	Thermal voltage	26	mV
W	Thickness		μm
ϵ_0	Permittivity in vacuum	$8.85X10^{-14}$	F/cm
ϵ_s	Silicon permittivity	$10.29X10^{-13}$	F/cm
μ_0	Mobility in vacuum	$1.256X10^{-8}$	H/cm
μ_n	Electron mobility	1500	cm^2/Vs
μ_p	Hole mobility	450	cm^2/Vs
ρ	Local space charge density		V/cm^3
ϕ_n	Quasi Fermi level for electrons		V
ϕ_p	Quasi Fermi level for holes		V

Conclusions

This thesis has dealt with two different problems solved with electronic device simulations (TCAD). The first relates to the Trench power MOSFET device characterization, with particular attention to device breakdown voltage as a function of the device parameters. The second one is about the simulation of Zero capacitor DRAM devices.

The trench structure analysis requires the use of a device simulator that implements the drift-diffusion model with impact ionization model. For a better analysis results, the Trench structure is compared with the equivalent pn structure which shows similar trends in breakdown voltage, but slightly higher, due to the of the trench structure's absence that introduces an additional electric field component anticipating the breakdown. For reference structure, the breakdown voltage depends only on drift region doping and length. Turning to Trench structure it can be noted that each Trench's shape parameter affects the breakdown voltage, in particular:

- The distance between two structures in the trench does not affect the calculation of the breakdown voltage.
- The Trench curvature radius and the oxide thickness are directly proportional to the breakdown voltage.
- The Trench width does not affect the breakdown voltage.
- The trench length is the most important trench's parameter in order to evaluate a good model.

A one-dimensional analytical model for pn structure has been presented using a maximum electrical field value prediction as border condition for Poisson equation. The same model has been used for trench MOS devices using a correction factor for maximum electrical field calculation based on trench penetration into drain region and on drain region length. The analytical model shows good results in comparison with simulation results for wide range simulations.

For the Zero capacitor DRAM simulations, so called ZRAM devices, we have presented a simulation study aimed at understanding the operation mode, the potential performance in terms of READ sensitivity, programming windows and retention time, and the scalability of a double gate type II Z RAM cell with respect to the type I cells. We find that the operations of a type II ZRAM cell can be implemented by changing simultaneously all electrode potentials and does not necessarily require an appropriate time sequence of bias voltages. Moreover, in the proposed operation mode, the excess charge is stored at the gate interfaces and not in the bulk body. This excess charge is a self consistent charge, created during the WRITE "1" phase by impaction ionization and BTBT at the drain side and defined by the accumulation condition imposed by the gate bias during the HOLD phase. The independence of the stored charge on the particular WRITE "1" bias configuration allows an excellent determination and tuning of device performances by experiments and device simulations. Stored data is read by an asymmetrical bias configuration of the gate interfaces where the bottom gate interface works in a manner similar to the HOLD mode while the top gate interface works in a manner similar to the WRITE "1" mode but with a lower drain bias in order to avoid drain disturbs. The charge eventually stored at the bottom interface increases the bulk potential at the top interfaces which in order reduces the source-bulk energy barrier allowing a high READ current. Because of the exponential dependence of the READ current respect to the bulk potential a higher I_1/I_0 ratio is found respect to type I operation mode allowing higher READ sensitivity, programming windows and retention times. Data retention is limited by the leakage associated to the state "0" due to BTBT at the source/drain to bulk junctions. Except for device geometries with degraded SCEs (L¡¡W and/or t_{ox}) the time retention associated to the state "1" is infinity because the stored charge is the self consistent accumulation charge dictated by gate potentials. Extensive scaling analysis is done involving longitudinal and transversal geometrical parameters. Longitudinal scaling (L) is limited by SCEs which increase the bulk potential and reduce hole density hence the retention and the READ sensitivity of state "1". To compensate this effect transversal scaling (W, t_{ox}) must be used. Moreover, longitudinal scaling increases the absolute value of READ currents. Transversal scaling reduces SCEs improving the performances of state "1", but the associated increase of BTBT reduces performances of state "0". Moreover transversal scaling reduces the absolute value of READ currents. The choice of longitudinal and transverse dimensions is a trade-off between speed, READ sensitivity, retention and programming windows. It is found that scaling limit of device length is around 15 nm which is lower respect to the 25nm found for type I operation mode.

Bibliography

[1] M. R. Pinto, S. Conor, Rafferty, and R. W. DD. "PISCES2 - Poisson and Continuity Equation Solve". *Stanford Electronics Laboratory Technical Report Stanford University*, pages 14–21, September 1984.

[2] S. Selberherr. *Analysis and Simulation of Semiconductor Devices.* Springer-Verlag, 1984.

[3] S. M. Sze. *Physics of Semiconductor Devices.* Wiley-interscience, 2 edition, 1981.

[4] L. et al. "A Physically Based Mobility Model for Numerical Simulation of Non-Planar Devices". *IEEE Trans. on CAD*, **volume 32**:1164, November 1988.

[5] W. Joyce and R. Dixon. "Analytic Approximation for the Fermi Energy of an ideal Fermi Gas". *Applied Physics Letters*, **volume 31**:354–356, November 1977.

[6] R. C. Jaeger. *Microelectronics.* McGraw-Hill, 1997.

[7] S. Selberherr. "Process and Device Modeling for VLSI". *IEEE Trans. on CAD*, **volume 24**, no. 2:225–257, 1984.

[8] A. K. Singh. *Electronic Devices And Integrated Circuits.* PHI, 1983.

[9] D. Caughey and R. Thomas. "Carrier Mobilities in Silicon Empirically Related to Doping and Field." *IEEE Proceedings*, **volume 55**:2192–2193, 1967.

[10] L. et al. "A Physically Based Mobility Model for Numerical Simulation of Non-Planar Devices". *IEEE Trans. on CAD*, **volume 26**:1164, 1988.

[11] A. G. Chynoweth. ""Ionization Rates for Electrons and Holes in silicon"". *Physical Review*, **volume 109**:1537–1540, 1958.

[12] G. Hurkx, H. de Graaf, and W. K. et. al. "A Novel Compact Model Description of Reverse Biase Diode Characteristics including Tunneling." In "ESSDERC", pages 49–52. 1990.

[13] K. Tomizawa. *Numerical simulation of submicron semiconductor devices*. Artech house, 1993.

[14] C. Price. "Two Dimensional Numerical Simulation of Semiconductor Devices". Stanford University, May 1982. Ph.D. Dissertation.

[15] W. H. Press, B. Flannery, S. A. Teukolsky, and W. T. Vetterling. "Numerical recipes". Cambridge University press, 1987.

[16] Silvaco. *Device simulation software. User's Manual.* ATLAS, 5.10.r edition, 2005.

[17] N. Mohan. *Power electronics*. Wiley-interscience, 3 edition, 2003.

[18] R. N. Hall and W. C. Dunlap. "p-n junctions prepared by Impurity Diffusion". *Phys. Rev.*, **volume 80**:467, 1950.

[19] S. R. Stiffler and A. O. Cifuentes. "The effect of trench corner shapes on local stress fields: a three-dimensional finite-element modeling study". *IEEE Transactions on Electron Devices*, **volume 40**:557–563, 1993.

[20] C. H. Xu and D. Schroder. "Modelling and simulation of power MOSFETs and power diodes". In "IEEE PESC", pages 76–83. 1988.

[21] W. Shockley. "The Theory of p-n Junctions in semiconductors and p-n Junction Transistors". *Bell Syst. Tech.*, **volume 28**:435, 1949.

[22] A. Goetzberger, B. McDonald, R. Haitz, and R. M. Scarlet. "Avalanche effects in Silicon p-n junction". *J. Appl. Phys.*, **volume 34**:1591, 1963.

[23] P. Mars. "Temperature Dependence of Avalanche Breakdown Voltage in p-n Junctions". *International Journal of Electronics*, **volume 32**, no. 1:23–27, 1963.

[24] S. Ikeda, Y. Usunaga, and H. Yoshida. "Power MOSFET for Switching Regulator". *Telecommunications Energy Conference*, pages 212–215, October 1982.

[25] A. Galluzzo, M. Melito, S. Musumeci, and M. Saggio. "A new high voltage power MOSFET for power conversion applications". In "Industry Applications Conference", pages 2966–2973. 2000.

[26] S. Suyama, T. Yachi, and T. Serikawa. "A new trench fabrication technique for silicon substrate utilizing undercutting and selective etching". *Jour. of Vacuum Sci and Tech*, pages 905–908, May 1985.

[27] S. A. Suliman, N. Gollagunta, L. Trabzon, J. Hao, R. S. Ridley, C. M. Knoedler, G. M. Dolny, O. O. Awadelkarim, and S. J. Fonash. "The dependence of UMOSFET characteristics and reliability on geometry and processing". *Semiconductor Science and Technology*, **volume 16**:447–454, 2001.

[28] T. H. Ueda D and K. G. "A new vertical power MOSFET structure with extremely reduced on-resistance". *IEEE trans. Electron Devices*, **volume 32**:2–6, 1984.

[29] J. E. Thompson, Z. U. A. Warsi, and C. W. Mastin. *Numerical Grid Generation.* Elsevier, 2001.

[30] R. Ng, F. Udrea, K. Sheng, and K. Ueno. "Lateral unbalanced super junction (USJ)/3D-RESURF for high breakdown voltage on SOI". In "Power Semiconductor Devices and ICs", pages 395–398. June 2001.

[31] C. Bulucea, M. Kump, and K. Amberiadis. "Field Distribution and Avalanche Breakdown of Trench MOS Capacitors operated in Deep Depletion". *IEEE Transactions on Electron Devices*, **volume 36**, no. 11, 1989.

[32] H. R. Chang, R. D. Black, V. A. K. Temple, W. Tantraporn, and B. J. Baliga. "Ultra low specific on-resistance UMOS FET". *IEEE International Electron Device Meeting*, **volume 32**:642–645, 1986.

[33] S. M. Sze and G. Gibbons. "Avalanche Breakdown Voltages of Abrupt and Linearly graded p-n junctions in Ge, Si, GaAs, and GaP". *Applied Physics letters*, **volume 8**:111–113, 1966.

[34] R. Felici. "R.D.T." *Bell Syst. Tech.*, 1982.

[35] A. Engelko and H. Bluhm. "Simulation of semiconductor opening switch physics". *Pulsed Power Plasma Science*, **volume 1**:318–321, 2001.

[36] Y. Okuto and C. Crowell. "Treshold Energy Effect on Avalanche Breakdown Voltage in Semiconductor Junctions". *Solid State Electronics*, **volume 18**:161–168, 1975.

[37] T. Lackner. "Avalanche multiplication in semiconductors: A Modification of Chynoweth's law". *Solid State Electronics*, **volume 34**, no. 1:33–42, 1991.

[38] H. Ballan and M. Declercq. "High voltage devices and circuits in standard CMOS technologies". *Springer*, **volume 1**:23–42, 1999.

[39] M. Bhatnagar and B. Baliga. "comparison of 6H-SiC, 3C-SiC, and Si for power devices". *IEEE Transactions on Electron Devices*, **volume 40**:645–655, 1993.

[40] L. Jong-Seok, S. Ho-Hyun, L. Han-Sin, Y. S. Man, and K. Ey-Goo. "The effect of a Shielding Layer on Breakdown Voltage in a Trench Gate IGBT". In "International Conference on Power Electronics", pages 62–65. 2007.

[41] S. Okhonin, M. Nagoga, J. M. Sallese, and P. Fazan. "A SOI capacitorless 1T-DRAM concept". *IEEE SOI Conference*, pages 153–154, 2001.

[42] T. Ohsawa, K. Fujita, T. Higashi, Y. Iwata, T. Kajiyama, Y. Asao, and K. Sunouchi. "Memory design using one-transistor gain cell on SOI". In "ISSCC Conference", pages 152–153. 2002.

[43] J. G. Fossum, Z. Lu, and V. P. Trivedi. "New Insights on Capacitorless Floating-Body DRAM Cells". *IEEE Electron Device Letters*, **volume 28**, no. 6, 2007.

[44] H. J. et Al. "A New Capacitorless 1T DRAM Cell: Surrounding Gate MOSFET With Vertical Channel (SGVC Cell)". *IEEE Transactions on Nanotechnology*, **volume 6**, no. 3, May 2007.

[45] M. G. Ertosun, P. Kapur, and K. C. Saraswat. "A Highly Scalable Capacitorless Double Gate Quantum Well Single Transistor DRAM: 1T-QW DRAM". *IEEE Electron Device Letters*, **volume 29**, no. 12:1405–1407, 2008.

[46] C. Kuo, T. J. King, and C. Hu. "A capacitorless double-gate DRAM cell". *IEEE Electron Device Letters*, pages 345–347, June 2002.

[47] E. Yoshida, T. Miyashita, and T. Tanaka. "A study of highly scalable DG-FinDRAM". *IEEE Electron Device Letters*, **volume 26**, no. 9:655–657, Sep. 2005.

[48] M. Bawedin, S. Cristoloveanu, and D. Flandre. "A Capacitorless 1T-DRAM on SOI Based on Dynamic Coupling and Double-Gate Operation". *IEEE Electron Device Letters*, **volume 29**, no. 7, July 2008.

[49] C. E. D. C. et Al. "Single-Transistor Latch in SOI MOSFET's". *IEEE Electron Device Letters*, **volume 9**, no. 2, 1988.

[50] J. Gautier, K. A. Jenkins, and J. Y. C. Sun. "Body charge related transient effects in floating body SOI NMOSFET's". In "IEDM Conference", 1995.

[51] G. G. Shahidi, A. Ajmera, F. Assaderagh, R. J. Bolam, E. Leobandung, W. Rausch, D. Sankus, D. Schepis, L. F. Wagner, K. Wu, and B. Davari. "Partially Depleted SOI Technology for Digital Logic". *ISSCC Tech. Digest*, page 426, 1999.

[52] G. G. Shahidi. "SOI technology for the GHz era". *IBM J. RES AND DEV.*, **volume 46**, no. 2, 2002.

[53] K. Kim, C. G. Hwang, and J. G. Lee. "DRAM technology perspective for gigabit era". *IEEE Transactions On Electron Devices*, **volume 45**:598–608, March 1998.

[54] J. A. Mandelman, R. H. Dennard, G. B. Bronner, J. K. DeBrosse, R. Divakaruni, Y. Li, and C. J. Radens. "Challenges and future directions for the scaling of dynamic random-access memory (DRAM)". *IBM J. RES AND DEV*, **volume 46**, no. 2, March 2002.

[55] N. Z. Butt and M. Alam. "Scaling Limit of Double-Gate and Surround-Gate Z-RAM Cells". *Transactions on Electron devices*, **volume 54**, no. 9:2255–2261, 2007.

[56] T. Tanaka, E. Yoshida, and T. Miyashita. "Scalability Study on a Capacitorless 1T-DRAM: From Single-gate PD-SOI to Double-gate FinDRAM". In "IEDM Conference", 2004.

[57] R. Ranica, A. Villaret, P. Malinge, P. Candelier, P. Masson, P. Mazoyer, and T. Skotnicki. "Modelling of the 1T-Bulk capacitor-less DRAM cell with improved performances: The way to scaling". *Solid-State Electronics*, **volume 49**:1759–1766, 2005.

[58] S. K. Moore. "Masters of memory". *IEEE Spectrum*, **volume 44**:45–49, January 2007.

[59] "International Technology Roadmap for Semiconductors", 2010.

[60] S. Okhonin, M. Nagoga, E. Carman, R. Beffa, and E. Faraoni. "New Generation of Z-RAM". In "IEDM Conference", 6, pages 925–928. 2007.

[61] S. O. et Al. "Ultra-scaled Z-RAM Cell". In "SOI Conference", 1, pages 157–158. 2008.

[62] K. W. Song. "55nm Capacitor-less 1T DRAM Cell Transistor with Non-Overlap Structure". In "IEDM Conference", 2. 2008.

[63] N. Balley and B. Baylac. "Analytical modelling of depletion-mode MOSFET with short- and narrow-channel effects". *Solid-State and Electron Devices*, pages 225–238, December 1981.

[64] S. Puget, G. Bossu, A. Regnier, R. Ranica, A. Villaret, P. Masson, G. Ghibaudo, P. Mazoyer, and T. Skotnicki. "Quantum effects influence on thin silicon film capacitor-less DRAM performance". In "IEEE International SOI Conference Proceedings", pages 157–158. 2006.

[65] J. P. Collinge. *FinFET and other multi-gate transistors*. Springer, 2007.

[66] A. Dixit, A. Kottantharayil, N. Collaert, M. Goodwin, M. Jurczak, and K. D. Meyer. "Analysis of the parasitic S/D resistance in multiple gate FETs". *IEEE Transaction on Electron Devices*, **volume 52**, no. 6:1132–1145, 2005.

[67] J. P. Campbell, K. P. Cheung, J. S. Suehle, and A. Oates. "A Simple Series Resistance Extraction Methodology for Advanced CMOS Devices". *IEEE Electron Device letters*, **volume 32**:1047–1049, August 2011.

[68] G. Hu, C. Yang, and Y. Chia. "Gate-voltage-dependent effective channel length and series resistance of LDD MOSFET's". *IEEE Transaction on Electron Devices*, **volume 34**, no. 12:2469–2475, 1987.

[69] B. J. Sheu, C. Hu, P. K. Ko, and F. C. Hsu. "Source and drain series resistance of LDD MOSFET's". *IEEE Transaction on Electron Devices*, **volume 5**, no. 9:365–367, 1984.