We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

1

SEMICONDUCTOR PROCESSES AND DEVICES MODELLING

Florin Babarada
Faculty of Electronics Telecommunications and Information Technology
University Politehnica of Bucharest, Romania

1. Introduction

The advancement of knowledge in the electronic design is strongly influenced by Technology Computer Aided Design–TCAD. Here is an interesting positive feedback, because the computing power helps the designers to perform modelling, simulation, optimisation and design of the new devices with improved performance, which have the capability to increase the computing power. The chapter present the basic aspects and the state-of-the-art of processes and devices modelling completed with the new aspects presented by the author in their last few year's papers like the results of his researches.

2. Models in micro and nanoelectronics

2.1 Modelling simulation and analysis

Due to high costs and long manufacturing cycle the modelling, simulation and optimisation or simply TCAD is the foundation of micro/nanoelectronics rapidly progress.
The analysis involves the separation of the entire in component parts, characterization and judgment them and also the examination of the elements of system and the relations between them to understand.
Simulation is the imitative representation of the operation of a system or process through the operation of other's or the examination of a problem without experimentation. If the analysis can be precise about the simulation we accept the idea of approximation.
Modelling is the production of a representation or simulation of a problem, process or device, making a description or analogies to help the visualization of the aspects that can't be directly observed. Modelling is a need for analysis, simulation and design optimisation.
A model for a pure simulation, such as that produced by fitting of the curves, is usually much simpler than a model for analysis, which should reflect the physical aspects in a qualitative manner. An example is the application of Monte Carlo method, which is equivalent to producing an imitative representation of the system functioning. We must always know the limitations of a model in order to don't interpret to naive the obtained results only through the improper application of the model. Also must have experience from previous experiments and simulations.

Development of electronic devices involves many tests and scraps for manufacturing until the scope respectively the designed operation parameters are performed. Implementation of device models, simulation and analysis, can now and in the future, to decrease substantially the number of iterations during development. A rough estimation of the development effort saved by analysis and simulation is of the order of 40%. This percentage depends heavily on the conditions of each individual project. The complete elimination of the tests and scraps of development is not possible today due to the uncertainty of many parameters of the models available that are already very sophisticated and too large. It is expected that modelling of devices, especially those with high scaling factor, at which the quantum mechanical aspects become predominate, to become more important in the future. This prediction is supported by the decreased cost of computing resources and special increases the cost of experimental investigation. Numerical modelling of the devices becomes more important for miniaturized models for which only large models are the existing and imaginable tools for precise prediction and analysis of device performances.

2.2 Technological models

The technological model is a schematic or analog description of the phenomenon or system that matter for his knowledge or associated properties and could be used to the further study of its properties. When possible the models used in microelectronics are physical models that mean the modelled phenomena are represented by physical effects well understood. When the studied phenomena are unknown is calling to the empirical models. In this case the relationships between phenomenon variables are experimental determined. Quantitative empirical model in this case is a mathematical expression that fits to the experimental data. Models can contribute to technological progress as follows:

- Physical models can produce explanations or views of the phenomenon or device studied even the studied effect is not directly observable
- Physical and empirical models serve as vehicles for processes or devices studied; certain aspects of real devices or processes can be examined by studying the characteristics or the respectively model operation by simulation

Although the simulation does not replace the manufacturing, it reduces the test time and errors to a stable and optimised process. Often, a simulation allows studying windows of process to help optimise the device structure by setting the process conditions. The biggest challenge is to develop models that can quickly, cheaply and accurately to simulate processes and phenomena of advanced semiconductor devices.

3. Physical models development

The development of physical models is generally in the following stages:

- Making the model qualitatively
- Making the model quantitatively
- Solving equations of quantitative model and centering process

Achieving quality model is the design, mechanisms deduction and relationships that comprise the essence of the phenomenon observed. In this form the model can realize the visualisation of the phenomena or devices studied. This is particularly important in microelectronics, where the phenomena being investigated are not usually directly observable. An example is the description of the Bohr atom model, the nuclei surrounded by

electrons occupying orbits well defined. Another example is the description of the drift current in semiconductors as the movement of discrete particles of electric charge, respectively the electrons and holes, moving under the influence of electric field. In this case qualitative model can be translated into a set of equations or computing operations, resulting in a quantitative model. Finally the solution equations representing the quantitative model must be found and compared with experimental data to ensure that the simulation correctly emulates the phenomenon observed. In microelectronics quantitative physical models often take the form of equations with partial derivatives, Partial Differential Equations-PDE.

Simulation of micro/nanoelectronics devices and processes equations requires the evaluation model, computing operations, numerical analysis and advanced computer graphics (Rusu, 1990). Simulation helps manufacture of devices by increasing the success probability of the first experiment, for new products or processes. Existing manufacturing processes can be improved by centering process, ie finding the process combination of conditions for that we have the smallest results at the changing of the process conditions. Thus, circuit simulating behaviour modification based on conditions changing of manufacturing can identify the process tolerances.

In general, semiconductor devices can be simulated more precisely as the manufacturing processes of devices and integrated circuits because the physics of semiconductor devices it is better known. In comparison, several physical processes manufacturing of integrated circuits is still not well understood and must rely on the empirical models.

4. Empirical models

4.1 Introduction

The empirical models are only representations of experimental data and have a little or nothing physical background. Experimental data are used to create an empirical model as follows:

- Experimental results are stored in a database in the computer but are not provided information's about the interpolation results for approximation of unknown values between two points
- It used a mathematical function, which is adjusted by experimental data

To adjust the experimental data can use the next methods:

- Is used a polynomial function to pass through all points, leading to very complex functions and some experimental points may be wrong
- At one set of data graphically represented is choose a close mathematical function, which don't passes through all points and which can be adjusted

The most popular method of adjustment is the method of least squares, respectively the minimal sum of squares differences between points and the curve. If a data set match on a straight line y=a+bx, the process of finding the coefficients a and b, known as regression coefficients, is called linear regression. If it adopts a non-linear functional approximation, is used non-linear regression to find the regression coefficients.

4.2 Empirical models in semiconductor simulation

The empirical models are used to simulate semiconductor because:

- We don't have other option when the physical background is not yet known

- If incorporated as part of a program for simulation of a process or device, empirical models can serve as a tool for storing experimental data
- The simulation results for these models is fast and direct
- The empirical models can provide accurate simulations for some particular experimental conditions
- If the simulated conditions are between the experimental data, the interpolation results can be found with reasonable accuracy

Sometimes, if the individuals do not produce quantitative expressions, which constitute these models, this major limitation makes it impossible to extrapolate to conditions out of the experimented field.

Semiempirical models are the models in which phenomena are modelled by equations based on physical parameters corresponding to these phenomena. Most models used in simulation of semiconductor devices and processes are semiempirical models. Thus, the silicon oxidation in dry oxygen, at thickness of less than 350Å does not correspond Deal-Grove model. Nicollian and Reisman created a model for this area $t_{ox}=a(t+\tau)^b$ with a, b=constant, t=time of growth and τ=time required to raise an initial layer thickness x_i. Other example is the boron implantation effects arising from sewage, ie penetration of boron ions deeper than monocrystalline silicon. For modelling this effect, an exponential portion was added to Pearson IV model. The length of the falling exponential part is determined empirically to the value of 450Å. This empirical model is available in TSUPREM III and IV.

5. Design of experiment

Simulation of the manufacturing process using process simulators and extracting electrical characteristics using device simulators allow prediction of the behaviour and characteristics of the circuit from the design phase. The problem arises is that every attempt to obtain a performance model that is capable to incorporate the change effects in a broader set of parameters, is hit by hinder or even impossible to generate an analytical model that can be used effectively in design. Such as particularly important are the following aspects:

- Choosing a set of factors as more comprehensive
- Choosing the set of responses that characterize the best performance expected from the product design

The problem is usually solved iterative following the overlapping findings resulting from a series of individual experiments. Optimising a design involves finding a complete set of factors chosen so that the founding responses to have a high degree of confidence. Also, the sensitivity of the responses, given by the statistical nature of the technological implementation steps, is of particular importance in assessing the limits of tolerance of the desired response. Thus, by using based models simulation, can identify the input variables that allow the attainment of targets. More precisely, starting from the process variables such as temperature, time, energy and dose of implantation, may control the threshold voltage of the MOS transistors, parameter which influences the shape of the IV characteristics and so on the device parameters used in the circuit simulation and finally influences the circuit performances. Design of Experiment-DOE goal is to minimize the number of experiments in parallel with extraction of maximum information useful to designers. In this respect distinguish the following stages of analysis:

- Defining a set of factors that are considered to be sufficient for analysing the performance parameters of the requisite responses; the choice is based on previous knowledge
- Choosing a field of operation and a nominal value for each factor, considered acceptable in terms of tolerance of the technological process
- Defining a matrix corresponding with the DOE strategy selected
- Experimentation in selected points and collecting the results for each response
- Building a response surface model and analysis the conclusions from that study
- The revaluation of the set factors and the strategy of experimentation
- Obtaining the final response surface model

6. Response surface modelling

Design of experiments is the key point of the optimisation process design. Results of experiments are used to generate the Response Surface Model-RSM. Are taken into account three model categories:

- Linear models, where the responses are linear functions of factors
- Models of order two for higher order design, the answers are obtained as functions of parabolic factors; these models constitute the standard in RSM techniques
- Transcendental models for higher order design, which provides improved techniques for analysing data and are used in analysis of amended RSM

Linear models assume that response R_i is a linear combination of factors f_1, f_2, ... f_n. When using these models only one factor is change in each run. The experiments are chosen in star. These are easily designed and expanded to higher dimensions.

The square models guess that the response R_i is a square combination of input factors with two power grade of the factors and products of factors. In this way is take into account the interactions between factors. Strategies to design experiments in this case are different: full-factorial type and fractional-factorial type. Full-factorial strategies take into consideration all possible combinations of factors. This approach provides more information but also presents the inconvenience of requiring a long running time. Fractional-factorial strategies select a subset of the experimental points from the set full-factorial. Presents the advantage of data reuse at increasing of the problem size and easy change to full-factorial analyse.

The transcendental models assume the existence of a mechanism for transforming an initial set of factors in a modified set used for RSM.

The data obtained by experiments and those obtained by simulation are used to build a RSM, from which analysis may conclude a set of information about:

- Main effects, linear or nonlinear
- The interaction of factors
- Various factors importance in the evolution of a response
- Sensitivities of response to some factor
- Comparing the effect of a factor with the others, etc.

These results are iterative used for adjustment of coefficients, which are the input data of the RSM.

Like example, for the case of MOSFET technology flow the threshold voltage VT is the output data and the input data are the following factors: oxide thickness TOX, Nsub concentration of substrate, the peak concentration of channel implant for threshold voltage

adjustment VTPEAK, the peak concentrations of LDD source and drain NLDDpeak, distance between the windows of the source and drain Lgate.

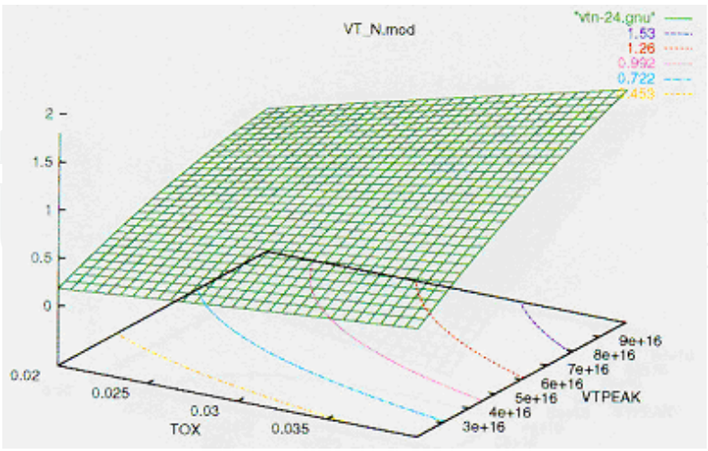

Fig. 1. RSM results for MOSFET threshold voltage VT

Were identified the main parameters that determines the threshold voltage of MOSFET like the MOS oxide thickness and the implantation dose for threshold adjustment, fig. 1.

7. Process optimisation

7.1 Introduction

TCAD software packages first need equipment models using configuration and settings as input parameters in order to obtain the process environment and process model to create the wafer data characteristics. Next using TCAD process simulator obtains the wafer state and using TCAD device simulator obtain device performance and the input data for circuit simulation. A final challenge and grand opportunity for future process modelling is to implement the accurate atomic scale reaction models respectively reaction energies, rates, products and process equipment models respectively gas flows, reactant concentrations and temperatures versus equipment settings.

In the state of the art devices small geometry effects including hot electron transport, punch-trough, avalanche multiplication, drain induced barrier lowering, oxide and junction breakdown, leakage currents, grain size effects and discrete doping elements effects are of great importance (Veendrick, 2008). Devices are also starting to exhibit significant quantum effects including gate oxide and bandgap tunnelling, inversion layer quantization, quantum transport and carrier density smoothing.

7.2 Optimisation strategies

Optimising a process technology or a device parameter involves an optimum set of factor setting such that a number of relevant results meet predefined targets. This problem is solved using the concepts of statistical Design of Experiments-DOE, for planning a number of experiments for different settings of input factors.

The simulation of process, device and circuit are performed in specific points respectively specific values for input factors for which the simulation are running. The results of

experiments are analysed for each of the responses as a function of the input factors and we obtain a response surface model-RSM. The DOE/RSM concept guarantees that with a minimum number of experiments we obtain a maximum information respectively detection of the important main effects, factor interaction effects or which factor are the most important. The RSM models are used to find factor settings that produce devices with desired specifications (Govoreanu, 2002).

7.3 Example

Process optimisation example refers to n-type MOSFET realized in 0,5μm technology using Taurus-workbench software package from Synopsys. We start with substrate <100> boron doped at $5x10^{18}$. Then epitaxial growth of 6μm silicon layer, 0,2μm oxide layer and 0,15μm nitride and in the last two layers is successively configured ISLAL and NWELL and phosphorus is implanted with $2x10^{12}$ dose and 300KeV energy. After nitride removing and oxide configuration the threshold voltage adjusting doping is performed in two steps VTN implant with boron and PUNCH implant with boron at $5x10^{11}$ dose and 50KeV energy.
The gate oxide is grown, the polysilicon gate is deposited configured and implanted with phosphorus at $5x10^{15}$ dose and 45KeV energy. After NLDD implant in the gate and source/drain area the deep implant for source/drain configuration with phosphorus at $4x10^{15}$ dose and 80KeV energy is performed.
The next process steps perform the contact and interconnection between devices and the circuit protection layer.

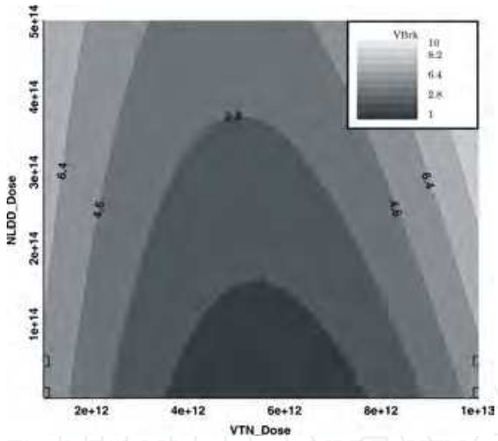

Fig. 2. RSM-V_{BRK} versus NLDD, VTN

Wafer	Units	Samp.1	Samp.2	Samp.3	Samp.4
VTN_Dose	Dose	1E+13	1.0E+13	1.0E+12	1.0E+12
PNCH_Dose	Dose	5E+11	5.0E+11	5.0E+11	5.0E+11
PNCH_Energy	Energy	50	50	50	50
tox	nm	11.9786	11.9786	11.996	11.996
NLDD_Dose	Dose	7E+12	5.0E+13	7.0E+12	5.0E+13
Xj	nm	192.945	231.634	214.388	347.719
Lchan	nm	538.034	393.185	376.42	254.542
Vt	V	0.9348	0.9338	0.3487	0.3222
IDSat	A/um	1.8E-04	3.3E-04	5.1E-04	6.7E-04
VBrk	V	5.57	5.57	5.91	5.73

Table 1. The samples parameters

Using DOE and RSM techniques the most sensitive process steps were identified respectively V_t adjusting implant and NLDD implant. These two parameters were modified successively.
The RSM results indicate a high dependence of breakdown voltage function of NLDD implant dose and a strong decreasing around $5x10^{12}$ VTN implant dose (fig. 2.), a big dependence of threshold voltage versus PUNCH implant dose and a small dependence versus VTN implant dose and a high dependence of saturation current (I_{DSS}) function of NLDD implant dose and a decreasing around $8x10^{12}$ VTN implant dose.

The increase of NLDD implant dose at Sample 2 and 4 reduces the polysilicon depletion effect, by reducing the voltage drop across the polysilicon gate and improving the device transconductance (the higher slope of transfer characteristics for Sample 2 and 4), fig. 7.

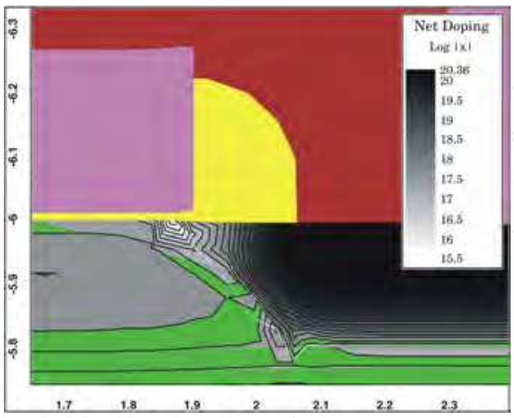

Fig. 3. Net Doping Sample 1

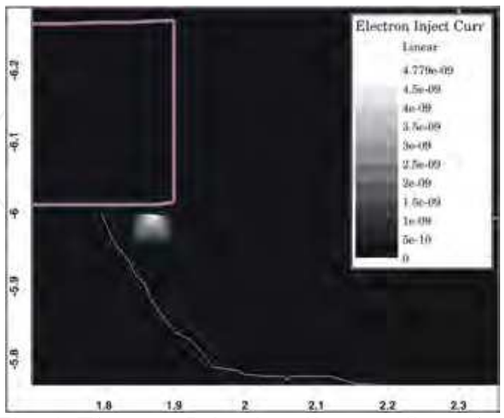

Fig. 4. Electron Injection Current Sample 2

According to Table 1 and fig. 3 to 6 the breakdown voltage is proportionally with the radius of source/drain junction (Kwong, 2002). Output resistance is reduced by decreasing the VTN boron adjusting implant dose (Sample 3 and 4) and can be explained by higher electron concentration in the channel, which allows a shorter pinchoff region. A shorter pinchoff region gives rise to a much larger magnitude of the Early voltage.

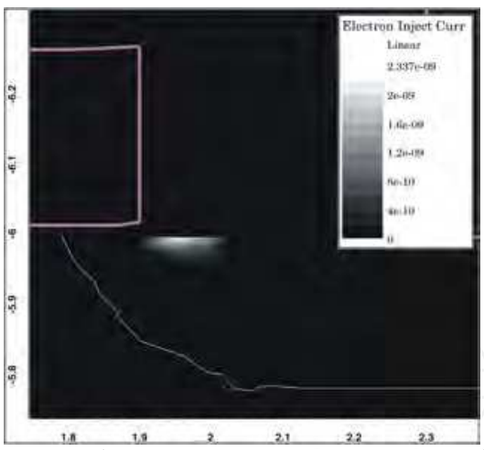

Fig. 5. Electron Injection Current Sample 3

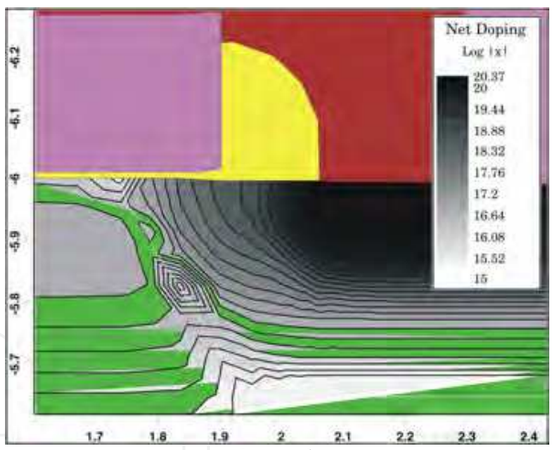

Fig. 6. Net Doping Sample 4

The decrease of NLDD dose in Sample 3, fig. 5, comparing with Sample 2 fig. 4 move the electron injection current from gate oxide to spacer decreasing the gate oxide breakdown possibility and reduces electron injection concentration which improve reliability. For the all four samples the transfer characteristics are presented in fig. 7 and the external characteristics in fig. 8 (Campian, 2003).

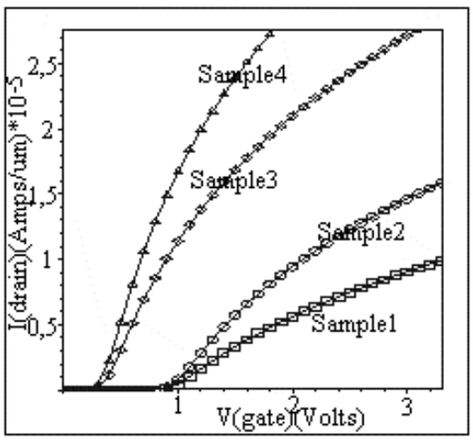

Fig. 7. I_D-V_{GS} Characteristics

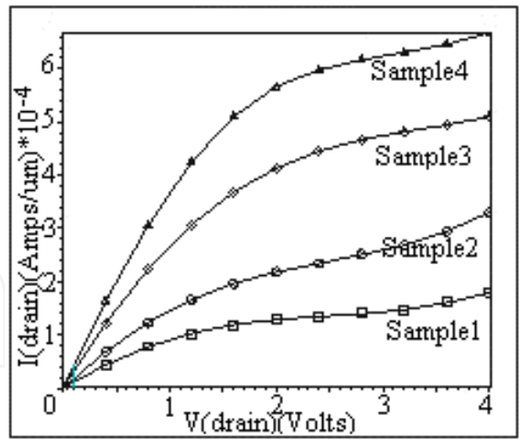

Fig. 8. I_D-V_{DS} Characteristics

The higher electron concentration in the channel gives a large Early voltage very useful in analog circuits. Increasing of lateral source/drain slope lowers also the series resistance, which improves the drive current, but for very abrupt profile junction the improvement is paid by degradation in leakage current due to more severe short channel effects.

8. MOSFET DC modelling for distortions analysis

8.1 Introduction

The scaling-down evolution of semiconductor devices will ultimately attend fundamental limits as transistor reach the nanoscale aria. In this context the MOSFET models must give the process variations and the relevant characteristics like current, conductance, transconductance, capacitances, flicker thermal or high frequency noise and distortion (Ytterdal, 2003). The new challenge of nanotechnology needs very accurate models for active devices (Scholten, 2009). The design of linear analog circuits lacks models for state-of-the-art MOS transistors to accurately describe distortion effects. This is produced by the inaccurate modelling of the second order effects induced by high vertical gate field such as mobility degradation and series resistance and second order effects induced by parallel drain field like velocity saturation in the ohmic region, channel length modulation, static feedback, weak avalanche and selfheating in the saturation region. After a rigorous description of transistor transconductance and channel conductance in ohmic and saturation region we included these effects in the MOSFET model, using a compact drain current expression for time computation reasons.

8.2 Gate induced distortions modelling

Carriers mobility degradation modelling

The channel mobility must be treated quantum-mechanically because the thickness of the inversion layer is in the order of a few Å, smaller than the De Broglie wavelength of the carriers. Quantum-mechanical calculations show that energy subbands of electrons and holes are formed in different energy valleys. The spacing of these subbands increases with the normal electric field Ex. In the weak inversion region where many subbands are occupied, quantum effects can be neglected, but in the strong inversion region where only

few subbands are occupied, quantum effects become important. In most cases more than one subband is filled and the modelling must give the right approximation of very complex scattering processes in the inversion layer. The mobility can be described by considering three mechanisms, which dominate the scattering of charge carriers in the inversion layer at the Si-Si0_2 interface.

Coulomb Scattering - μ_C

Charged centres near the Si-Si0_2 interface may be of the same charge type as the mobile inversion charge leading to Coulomb repulsion. This results in scattering, which is important for lightly inverted surfaces, high surface-charge densities or substrate doping concentrations, and less important for heavily inverted surfaces due to carrier screening. Coulomb scattering limited mobility μ_C is given by:

$$\mu_C \cong \frac{Q_{inv}}{N_A} \tag{1}$$

The above type of scattering has influence only in the weak and moderate inversion region when the drain current is dominated by the exponential dependence of inversion layer charge Q_{inv} on the gate voltage.

Phonon Scattering - μ_{ph}

Surface phonons or surfons from the quantum vibrations of the crystal lattice scatter the mobile charge carriers. Under the assumption that carriers in the inversion layer only occupy the lowest subband, the mobility determined by acoustic phonon scattering is described by:

$$\mu_{ph} \cong \left(\frac{11}{32}Q_{inv}+Q_{dep}\right)^{-1/3} \tag{2}$$

Experimentally it was found for both holes and electrons that: $\mu_{ph} \cong E_{eff}^{-1/3}$ (3)

Expression (3) deviates slightly from (2), which is ascribed to the fact that electrons occupy several subbands at intermediate values of effective field (Babarada, 2003).

Surface Roughness Scattering - μ_{sr}

The interface between the silicon crystal and the gate oxide is not atomically smooth. The above interface roughness scatters the mobile charge carriers. This type of scattering is especially important under strong inversion conditions because the strength of the interaction is governed by the distance of the carriers to the surface. The carriers, which are to the surface, will have the stronger scattering due to surface roughness. Under the assumption of single subband occupation and a Gaussian type autocorrelation function of interface roughness, μ_{sr} can be described by:

$$\frac{1}{\mu_{sr}} \cong E_{av}^2 \cdot \int_0^{\pi} (1-\cos\theta) \cdot \exp\left[-\frac{1}{2} \cdot k_w^2 \cdot L_c^2 \cdot (1-\cos\theta)\right] \cdot d\theta \tag{4}$$

where L_C is the correlation length of interface roughness, k_w is the carrier wave vector and E_{av} is the normal field averaged over the inversion layer. For a uniform doping profile the average field E_{av} can be calculated to be equal to the effective field E_{eff} with η = ½. Non-uniform doping profiles will lead to different values of η, which is an empirical parameter.

In the limit that the correlation length is much smaller than the carrier wavelength ($L_C << 1/k_w$), the mobility limited by surface roughness scattering μ_{sr} reduces to:

$$\mu_{sr} \cong E_{eff}^{-2} \tag{5}$$

The above dependence of surface roughness scattering corresponds to the experimentally found dependence of electron mobility on effective field. For larger values of correlation length L_C, mobility deviates from the inversely quadratic dependence on E_{eff} owing to the fact that the integral term in (4) depends on Q_{inv}. For holes, was found experimentally:

$$\mu_{sr} \cong E_{eff}^{-1} \tag{6}$$

The difference between equations (5) and (6) for electrons and holes, respectively, is often ascribed to the fact that at high transverse fields holes tend to congregate further away from the interface than electrons do. The larger average distance leads to a reduced influence of the interface roughness and thus to less surface roughness scattering for holes.
The above-described mechanisms can be incorporated into one channel mobility μ, as follow:

$$\frac{1}{\mu} = \frac{1}{\mu_0} + \frac{1}{\mu_{sr}} + \frac{1}{\mu_{ph}} \tag{7}$$

where μ_0 is the carrier mobility limited by ionised impurity scattering and acoustic phonon scattering in the bulk material. Equation (7) leads to:

$$\mu = \frac{\mu_0}{\theta_c + \theta_{ps}\sqrt{\theta_{ph} E_{eff}^{2/3} + \theta_{sr} E_{eff}^{2n}}} \tag{8}$$

where θ_c, θ_{ps}, θ_{ph} and θ_{sr} are empirical parameters, and n = 2 for electrons and n = 1 for holes. Expression (8) assumes that the surface roughness scattering limited mobility is more important than phonon scattering limited mobility and leads to a more accurate description of high-order derivatives.

Series resistance modelling
For DC model only the resistive element is important out of MOSFET source and drain junction parasitic components. For short-channel devices the drain current may reach very high values. At large current values the series resistance is no longer negligible and has to be taken into account using the source resistance R_S and the drain resistance R_D like in the equivalent circuit given in fig. 9.

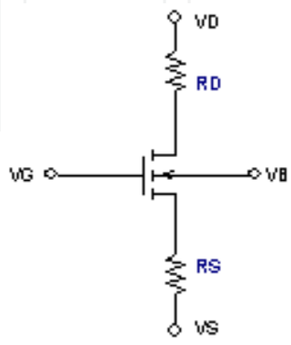

Fig. 9. The equivalent MOSFET

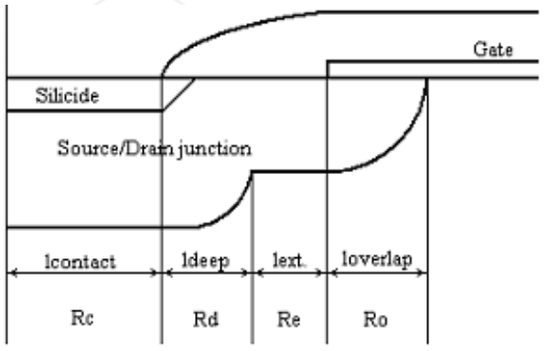

Fig. 10. Parasitic series resistance components

Including the source series resistance R_S and drain series resistance R_D in the MOSFET model we don't include additional nodes and the computation efficiency is better.

According to fig. 10, the series resistance can be divided in four components: overlap resistance R_o, extension resistance R_e, deep resistance R_d and silicide-diffusion contact resistance R_c .

Overlap resistance

With scaling-down the transistor dimensions the source-drain extension to gate overlap in the gate length is increasing around 35% and this resistance become more important and bias dependent. The overlap resistance is closely dependent on doping concentration in overlap region because the current spreading and the accumulation carrier charge density is dependent on the overlap doping concentration. The lateral slope of overlap doping profile is one of the most important parameter to controlling the short channel effect and finally the nanotransistor characteristics and for these reasons the modelling of the overlap resistance is very important for simulation and design.

The approximation of exponentially sloped doping profile and the constant accumulation charge density in the overlap region are suited for the computation reasons but usually needs fitting parameters and has good precision only near the metallurgical junction. Better results of the accumulation and spreading current are given by Gaussian doping gradient in the lateral direction of the overlap region. Because the most important component of current is distributed in the lateral junction, the vertical doping gradient of the junction is not so important. The transistor has an accumulation layer under the gate oxide in the overlap region and because the nanotransistors have very shallow junctions with high doping level and large depletion width the current is spread in the bulk region along the depletion boundary of source-drain extension junction. As a result the overlap resistance can be modelled by combination of resistances R_{OS1} series with R_{OS2} in parallel with R_{OP} as given in fig. 11. R_{OS1} and R_{OS2} are the resistances corresponding to accumulation layer in the entire surface under the gate oxide in the overlap region and R_{OP} is the spreading resistance in the neutral bulk region. The current spreading in the overlap region is modelled based on the depletion approximation and spreading resistance is calculated versus the depletion width in overlap region and the spreading angle.

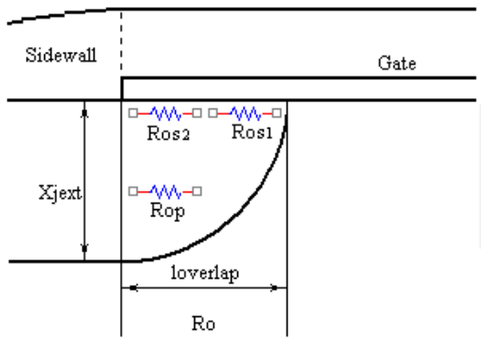

Fig. 11. Overlap resistance components

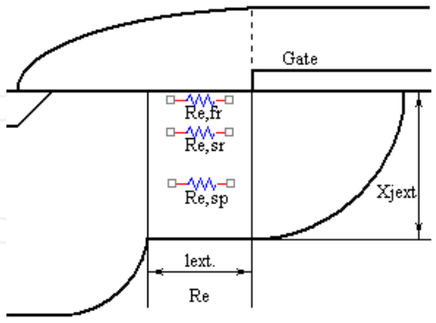

Fig. 12. Extension resistance components

Extension resistance

The extension resistance corresponds to the region from the gate end to beginning of deep source/drain junction like in fig. 12. Three component resistances connected in parallel compose the extension resistance: fringing resistance, surface resistance and spreading resistance. The fringing resistance taken into account to characterize the fringing field effect

on the surface extension region and can be calculated by iterative solution of the potential relationship of gate-sidewall-silicon system. The surface resistance characterize the uniform doping region in the extension region and is calculated in function of rezistivity, doping concentration and dimensions of these region. The spreading resistance in the extension region can be expressed function of vertically graded Gaussian doping profile and the spreading angle.

Deep resistance

The deep resistance corresponds to lateral deep diffusion of source/drain-gate structure and its contribution to series resistance is small because of heavily doped deep junction, fig. 13. In the deep region the modulation effect of the gate fringing field is negligible because the distance to gate and because the deep region is heavily doped to be modulated. The surface resistance in deep region include the lateral extension of silicide layer. The spreading resistance in the deep region can be calculated considering one similar angle like in the extension region.

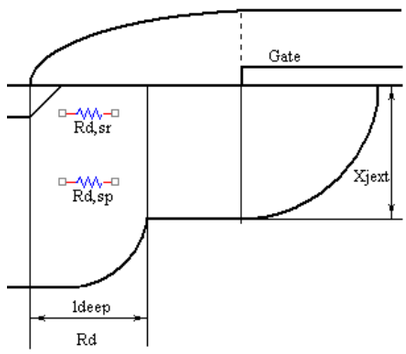

Fig. 13. Deep resistance components

Contact resistance

The contact resistance refers at resistance between the silicide contact and the diffusion layer and is strong dependent on the silicide layer thickness, the junction doping concentration and the silicide material. Under the silicide-diffusion contact the current flow increases with the increase of the silicide thickness and a significant current is pushed in the silicon region under the silicide. Because the total contact silicide-diffusion resistance has a minor effect in total series resistance we can neglect this effect. Now the source or drain series resistance is given by:

$$R_{S,D}=Ro+Re+Rd+Rc=(R_{O,S1}+R_{O,S2}\|R_{O,Sp})+(Re,fr\|Re,sr\|Re,sp)+(Rd,sr+Rd,sp)+Rc \quad (9)$$

In the ohmic region the drain series resistance value is equal with the source series resistance. At high level of lateral electric field if the carriers attend the saturated velocity at the drain end, the drain series resistance and source series resistance become different. In the saturation region the drain series resistance may reach several times higher values than the source series resistance but because the drain current is strongly dependent on drain voltage, the effect of drain series resistance is low.

The extension resistance, deep resistance and contact resistance are independent of terminal voltages and are inversely proportional to channel width. The overlap resistance due to overlap of polysilicon gate on the source/drain region, an accumulation layer is formed.

Because the accumulation layer charge is directly dependent on the gate voltage the overlap resistance is dependent on the gate voltage too.
From computing reasons we used a very compact form of series resistance versus gate voltage dependent, given by:

$$R_S = r_1 \cdot \left(1 + \frac{r_2}{r_3 + V_{GS} - V_T}\right), \tag{10}$$

where, r_1, r_2 and r_3 are empirical parameters.

The current expression
In order to include the series resistance R_S and R_D in MOS transistor model we must replace the new expressions including the resistances R_S and R_D in the drain current in the ohmic region.

$$V_{GS,1} = V_{GS} - I_D \cdot R_S \tag{11}$$

$$V_{DS,1} = V_{DS} - I_D \cdot (R_D + R_S) \tag{12}$$

$$V_{SB,1} = V_{SB} + I_D \cdot R_S \tag{13}$$

$$V_{T,1} = V_{FB} + \Phi_B + \gamma\sqrt{\Phi_B + V_{SB} + I_D \cdot R_S} \cong V_{FB} + \Phi_B + \gamma\sqrt{\Phi_B + V_{SB}} + \delta \cdot I_D \cdot R_S \tag{14}$$

V_{FB} is the flat-band voltage, Φ_B is the surface potential at onset of strong inversion, γ is the body effect coefficient $\gamma = \frac{\sqrt{2 \cdot q \cdot \varepsilon_{Si} \cdot N_A}}{C_{ox}}$, δ describe the influence of non-uniform doping concentration $\delta = \frac{\gamma}{2\sqrt{\Phi_B} + V_{SB}}$, R_S is the series resistance corresponding to drain or source resistance. The drain current in the ohmic region is given by:

$$I_D = \frac{\mu \cdot C_{ox} \cdot W}{L} \cdot \left[V_{GS,1} - V_{T,1} - \frac{1}{2}(1+\delta) \cdot V_{DS,1}\right] \cdot V_{DS,1} \tag{15}$$

The finally drain current expression is given by:

$$I_D = \frac{K\left[V_{GS} - V_T - \frac{1}{2}(1+\delta)V_{DS}\right] \cdot V_{DS}}{1 + K \cdot 2R_S \cdot \left[V_{GS} - V_T - \frac{1}{2}(1+\delta)V_{DS}\right]} \tag{16}$$

$$\text{where} \quad K = \frac{\mu \cdot C_{ox} \cdot W}{L}, \tag{17}$$

μ is the mobility given by expression (8), C_{ox} is the gate oxide capacitance per unit area $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$, W is the effective channel width and L is the effective channel length.

8.3 Drain induced distortions modelling

In the analog circuit design the MOS transistor can be driven as well by the drain terminal. By example a drain voltage driven MOSFET operating in the ohmic region like a gate voltage controlled resistor with the resistor linearity limited by the transistor drain voltage induced distortions. Other example can be the load device in amplifier circuits where the drain terminal drives the MOSFET in saturation.

Ohmic region

Drain Current

The drain current expression for strong inversion long channel devices in compact models is obtained by drift current expression function of surface potential. Using a first-order Taylor expansion around V_{SB}, with the source as reference, the drain current is:

$$I_D = \mu C_{ox} \frac{W}{L}\left[V_{GS} - V_T - \frac{1}{2}(1+\delta)\cdot V_{DS}\right]\cdot V_{DS} \tag{18}$$

where mobility μ is bias dependent due to mobility degradation. The equation (18) is quadratic expression in drain voltage, implying that the third-order derivative $\frac{\partial^3 I_D}{\partial V_{DS}^3}$ is equal to zero for all bias conditions, which is non-physical and also is non-symmetrical, which leads to a discontinuity in the higher-order derivatives at V_{DS}=0. For more accurate and symmetrical description of the higher-order derivatives drift current expression function of surface potential can be expanded around $V_{SB}+\frac{1}{2}V_{DS}$ and results (Babarada, 2005):

$$I_D = \mu C_{ox} \frac{W}{L}\cdot V_{GS\#}\cdot V_{DS} = K\cdot V_{GS\#}\cdot V_{DS} \tag{19}$$

$$V_{GS\#} = V_{GB} - V_{T0} - \frac{1}{2}(V_{DB}+V_{SB}) - \gamma\left(\sqrt{\frac{1}{2}(V_{DB}+V_{SB})+\Phi_B} - \sqrt{\Phi_B}\right) \tag{20}$$

The equation (20) is valid for both positive and negative values of V_{DS} and consequently the second-order and higher-order derivatives are continuous at V_{DS}=0. The mobility is given by expression (8). The simulation results of second and third-order harmonic amplitude, using the drain current expression (19) gives better fit to the measurements.

Velocity Saturation

For short-channel transistors equation (19) becomes inaccurate due to the effect of velocity saturation. The usual description of the carrier velocity is expressed by an empirical relation. For a more precise description of hole velocity saturation and prevent discontinuities at V_{DS}=0, used the next adjusted expression:

$$v = \frac{\mu\frac{\partial\psi_S}{\partial y}}{\sqrt{1+\left[\left(\frac{\mu}{v_c}\frac{\partial\psi_S}{\partial y}\right)^2 \Big/ \sqrt{P^2+\left(\frac{\mu}{v_c}\frac{\partial\psi_S}{\partial y}\right)^2}\right]+\left(\frac{\mu}{v_{sat}}\frac{\partial\psi_S}{\partial y}\right)^2}} \tag{21}$$

where v_c is saturation velocity limited by acoustic phonon scattering, P is a fitting parameter and v_{sat} is saturation velocity limited by optical phonon scattering. Assuming first-order approximation of $\frac{\partial\psi_S}{\partial y}$ equals with $\frac{V_{DS}}{L}$, this equation can be implemented like the empirical relation of the carrier velocity, in the drain current expression, including also the series resistance effect:

$$I_D = \frac{\mu_0\cdot C_{ox}\cdot W\cdot V_{GS\#}V_{DS}}{\sqrt{\left(\frac{\mu_0 L}{\mu}\right)^2+\left[\left(\frac{\mu_0}{v_c}V_{DS}\right)^2 \Big/ \sqrt{P^2+\left(\frac{\mu_0 V_{DS}}{v_c L}\right)^2}\right]+\left(\frac{\mu_0}{v_{sat}}V_{DS}\right)^2}+\beta_R V_{GS\#}} \tag{22}$$

where $\beta_R = 2\mu_0 C_{ox} W R_S$ and R_S is given by equation (10). The harmonic distortion results of equation (22) are accurate also for p-type transistors and have no discontinuities at V_{DS}=0.

Saturation region

The drain current expression (22) gives an accurate description of the DC behaviour in the ohmic region, which can be extended to the saturation region by replacing the drain voltage V_{DS} by the saturation voltage V_{DSsat}. For long-channel devices the saturation voltage V_{DSsat} can be calculated in good approximation from the zero value of the first-order derivative of the drain current expression (19) to drain voltage V_{DS}. For short-channel devices the calculation of V_{DSsat} is less evident, since the saturated drain conductance may be much larger than zero. The saturated drain current can by written:

$$I_{Dsat} = -W \cdot Q_{invL} \cdot v_{sat} \tag{23}$$

where Q_{invL} is the inversion layer charge at the drain side. Using the current expression (22) and (23) and neglecting the influence of mobility degradation, the saturation voltage is:

$$V_{DSsat} = V_{DSsat\_i}\left[1 - \frac{1}{2}\frac{\left(\sqrt{\beta_n^2 + \beta_p^2} - \frac{1}{2}\beta_R\right)}{\sqrt{\left(L/V_{DSsat\_i}\right)^2 + \beta_n^2 + \beta_p^2} + \frac{1}{2}\beta_R}\right] \tag{24}$$

where $\beta_n = \mu_0 / v_{sat}$, $\beta_p = \mu_0 / \left(v_c \cdot \sqrt{P}\right)$ and V_{DSat_i} is the saturation voltage for an ideal long-channel transistor.

Channel length modulation

When V_{DS} is increased beyond V_{DSsat}, the velocity saturation point moves towards the source, causing effectively that the channel length L is shortened by a length ΔL. That made the conductance to be non-zero in the saturation region and the drain current function of I_{Dsat}, the drain current at $V_{DS}=V_{DSsat}$, and L_{eff}, the effective channel length defined by:

$$L_{eff} = \mathrm{L} + \frac{\mu}{v_{sat}} \cdot V_{DSsat} \tag{25}$$

Velocity saturation results in an effective increase of the channel length, which may become important for short-channel devices. Using the velocity saturation expression (21) the effective channel length is:

$$L_{eff} = \sqrt{L^2 + \frac{\left(\frac{\mu}{v_c} \cdot V_{DSsat}\right)^2}{\sqrt{P^2 + \left(\frac{\mu}{v_c} \cdot \frac{V_{DSsat}}{L}\right)^2}} + \left(\frac{\mu}{v_{sat}} \cdot V_{DSsat}\right)^2} \tag{26}$$

The channel length modulation ΔL expression is:

$$\Delta L = l_c \cdot ln\left(\frac{V_{DS} - V_{DSsat} + \sqrt{\left(V_{DS} - V_{DSsat}\right)^2 + E_{sat}^2 \cdot l_c^2}}{E_{sat} \cdot l_c}\right) \tag{27}$$

where E_{sat} is the lateral electric field for which the carrier velocity saturates and is assumed to be constant and lc is length parameter.

Static feedback

Because the electrostatic coupling between drain and channel region an extra mobile charge ΔQ_{inv} is injected in the channel when the drain bias is increased beyond saturation. The effect of ΔQ_{inv} on drain current can be modelled as a decrease of the threshold voltage, where σ_{sf} is a parameter:

$$\Delta V_{Tsf} = \frac{\sigma_{sf} \cdot \sqrt{V_{DSsat}} \cdot V_{DS}}{L} \tag{28}$$

Self-Heating

The effective working temperature in the inversion layer is: $T = T_0 + R_{Th} \cdot P_{dis}$ (29)

where T_0 is the ambient temperature, the dissipated power is: $P_{dis} = I_D \cdot V_{DS}$ (30)

and thermal device resistance is:

$$R_{Th} = \frac{X_{sub}}{\lambda_{Th} \cdot W \cdot L} \tag{31}$$

where X_{sub} is the thickness of the substrate and λ_{Th} is the thermal conductivity of silicon. The dependence of drain current function of temperature is:

$$I_D = I_{Dsat} / \left(1 + R_{Th} \cdot I_{Dsat} \cdot V_{DS} / T_0\right) \tag{32}$$

8.4 Smoothing function

Usual is assumed that drift current can be neglected in weak inversion and diffusion current can be neglected in strong inversion. Around threshold voltage is considered the moderate inversion region, where both the drift current and diffusion current are important.

The transition from the ohmic region to the saturation region is continuous for the drain current. For analog circuit modelling, this transition should also be continuous for the higher-order derivatives of drain current to drain voltage, or the model should be continuous. To arise these requires we replace drain-source voltage V_{DS} with an empirical function V_{DSsf} that changes smoothly from V_{DS} in the ohmic region to V_{DSsat} in the saturation region. This empirical function V_{DSsf} is usual named smoothing function.

In order to preserve the model symmetry respectively the discontinuities of higher-order derivatives at V_{DS}=0V, we choose a smoothing function for which the derivate reported to V_{DS} is equal to unity, at V_{DS}=0V:

$$V_{DSsf} = \frac{V_{DS} \cdot V_{DSsat}}{\left(V_{DS}^{\,2m} + V_{DSsat}^{\,2m}\right)^{1/2m}} \tag{33}$$

where: m is an empirical parameter, which can be integer only; V_{DS} is the drain-source voltage and V_{Dssat} is the drain-source saturation voltage.

8.5 The current expression

Distinction between the drift and the diffusion component of the drain current should be maintained in all inversion regions, for an accurate description of the moderate inversion region (Gildenblat, 2009):

$$I_D = I_{drift} + I_{dif} \tag{34}$$

For model symmetry the drift current expression is given by usual formula, and making a Taylor expansion around $V_{SB} + \frac{1}{2} V_{DS}$, like in equation (19):

$$I_{drift} = \beta \left(V_{GBeff} + \Delta V_G - \frac{\psi_{sL} - \psi_{s0}}{2} - \gamma \sqrt{\frac{\psi_{sL} + \psi_{s0}}{2}} \right) \left(\psi_{sL} - \psi_{s0}\right) \tag{35}$$

where: β is the gain factor; $V_{GBeff} = V_{GS} + V_{SB} - V_{FB}$, ΔV_G is determined by transition from drain induced barrier lowering-DIBL in weak inversion to static feedback in strong inversion; ψ_{sL}

is the surface potential at the drain side; ψ_{S0} is the surface potential at the source side; γ is the body effect coefficient. Similar the diffusion current is:

$$I_{dif}=\beta\cdot\gamma\cdot u_T\left[\sqrt{\psi_{s0}+u_T\exp\left(\frac{\psi_{s0}-V_{DS}-2\Phi_F}{u_T}\right)}-\sqrt{\psi_{s0}}-\left(\sqrt{\psi_{sL}+u_T\exp\left(\frac{\psi_{sL}-V_{DS}-V_{SB}-2\Phi_F}{u_T}\right)}-\sqrt{\psi_{sL}}\right)\right] \tag{36}$$

and the total drain current expression, taking into account: drain saturation voltage, drain induced barrier lowering, mobility degradation, series resistance, velocity saturation, channel length modulation, static feedback and weak avalanche, is given by:

$$I_D=\frac{L\left(I_{drift}+I_{diff}\right)}{\left(\mu_0/\mu\right)\left(L_{eff}-\Delta L\right)+\left(\beta_R+\beta_{Th}\cdot V_{DS}\cdot V_{DSsf}\right)V_{GS\text{-}2}}\left(1+W_{av}\right) \tag{37}$$

where: L is the channel length; μ_0 is low field bulk mobility; μ is carriers mobility (Babarada, 2003); L_{eff} is the effective channel length; ΔL is the channel length modulation; β_R includes the series resistance; β_{Th} includes self heating; W_{av} is the weak avalanche, $V_{GS\text{-}2}$ has the expressions:

$$V_{GS-2}=\frac{1}{2}V_{GS-1}+\frac{1}{2}\sqrt{V_{GS-1}^2+4\cdot\varepsilon^2} \tag{38}$$

where ε is a smoothing factor and $V_{GS\text{-}1}$ has the next expressions:

$$V_{GS\text{-}1}=V_{GS}-V_{FB}-2\Phi_F-V_{DSsf}/2-\gamma\sqrt{2\Phi_F+V_{SB}+V_{DSsf}/2} \tag{39}$$

8.6 Results

The higher-order derivatives have been performed by applying a sinusoidal signal to the terminal under investigation and by measuring the higher-order harmonics in the drain current, like in the configuration from fig. 14. The signal frequency is a few KHz in order to neglect the influence of capacitances. In this situation the distortions can be completely determined by the MOS transistor steady state.

The drain current I_D can be expanded in a Taylor series:

$$I_D=d_0+d_1\cdot v_i+d_2\cdot v_i^2+d_3\cdot v_i^3+\ldots=\sum_{i=0}^{\infty}d_i\cdot v_i^i \tag{40}$$

v_i is a sinusoidal signal $v_i=V\cdot\sin(\omega t)$ and the coefficients $d_i=\frac{1}{i!}\cdot\frac{\partial^i I_D}{\partial V_{G(D)S}^i}\Big|_{V_{Do},V_{Go},V_{Bo}}$.

The drain current expression can be rewritten in terms of sin($n\omega t$):

$$I_D=a_o+a_1\cdot\sin(\omega t)+a_2\cdot\cos(2\omega t)+a_3\cdot\sin(3\omega t)+\ldots \tag{41}$$

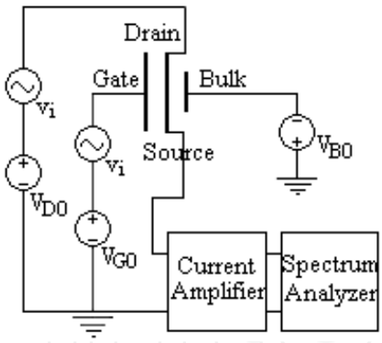

Fig. 14. Measurement circuit

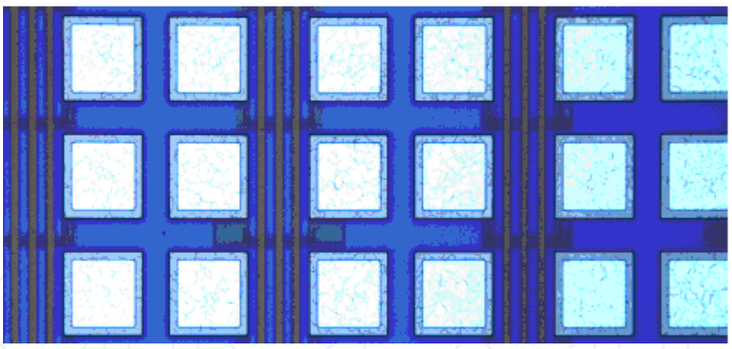

Fig. 15. Different W/L Transistors Array

The coefficients $|a_1|$, $|a_2|$ and $|a_3|$ are the signal harmonics amplitudes and were measured with the spectrum analyser, fig. 14. The coefficients can be written as:

$$\begin{aligned} a_1 &= d_1 \cdot V + \frac{3}{4} \cdot d_3 \cdot V^3 + \frac{5}{8} \cdot d_5 \cdot V^5 + \ldots \\ a_2 &= -\frac{1}{2} \cdot d_2 \cdot V^2 - \frac{1}{2} \cdot d_4 \cdot V^4 - \frac{15}{32} \cdot d_6 \cdot V^6 - \ldots \\ a_3 &= -\frac{1}{4} \cdot d_3 \cdot V^3 - \frac{5}{16} \cdot d_5 \cdot V^5 - \frac{21}{64} \cdot d_7 \cdot V^7 - \ldots \end{aligned} \tag{42}$$

For small enough values of the signal amplitude V, the coefficients (42) reduce to:

$$a_1 \approx d_1 \cdot V \;; \quad a_2 \approx -\frac{1}{2} \cdot d_2 \cdot V^2 \;; \quad a_3 \approx -\frac{1}{4} \cdot d_3 \cdot V^3 \tag{43}$$

In this way the high order derivatives reported to V_{GS} notated g_{mi} or reported to V_{DS} notated g_{di} can be calculated or extracted from the measured harmonics amplitude.
Using the harmonic amplitude notation HD_i the second-order and third-order harmonic amplitude are:

$$HD_2 = a_2 \approx \frac{1}{2} \cdot d_2 \cdot V^2 = \frac{1}{4} \cdot \frac{\partial^2 I_D}{\partial V_{G(D)S}^2} \cdot V^2 \tag{44}$$

$$HD_3 = a_3 \approx \frac{1}{4} \cdot d_3 \cdot V^2 = \frac{1}{24} \cdot \frac{\partial^3 I_D}{\partial V_{G(D)S}^3} \cdot V^3 \tag{45}$$

Fig. 15 show device geometries array with W=10μm and different L like were used in the next figures. Fig. 16 present for n-MOSFET the simulated (lines) and measured (symbols) values of gate induced distortion $g_{m3} = (\partial^3 I_D)/(\partial V_{GS3})$ at low drain bias (V_{DS}=50mV) and fig. 17 present the drain induced distortion $g_{d3} = (\partial^3 I_D)/(\partial V_{DS3})$ at V_{GS}=1V.

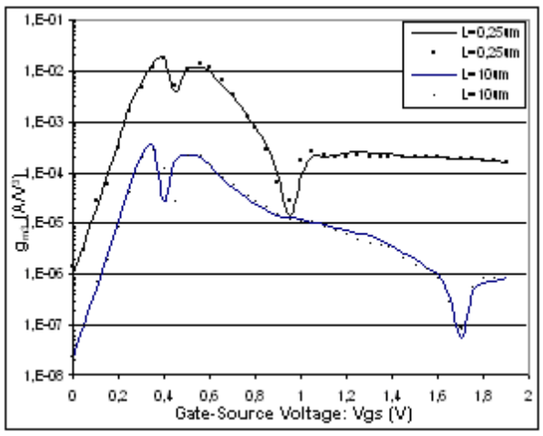

Fig. 16. g_{m3} for n-type MOS Transistor

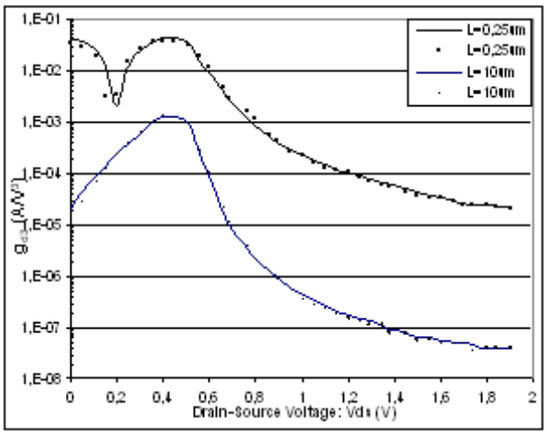

Fig. 17. g_{d3} for n-type MOS Transistor

Fig. 18 present for p-MOSFET the simulated (lines) and measured (symbols) values of gate induced distortion $g_{m3} = (\partial^3 I_D)/(\partial V_{GS3})$ at low drain bias (V_{DS}=-50mV) and fig. 19 present the drain induced distortion $g_{d3} = (\partial^3 I_D)/(\partial V_{DS3})$ at V_{GS}=-1V.

A good fit between simulated and experimental results can be observed.

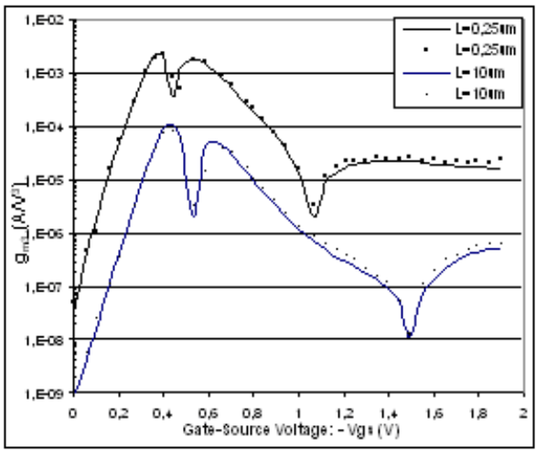

Fig. 18. g_{m3} for p-type MOS Transistor

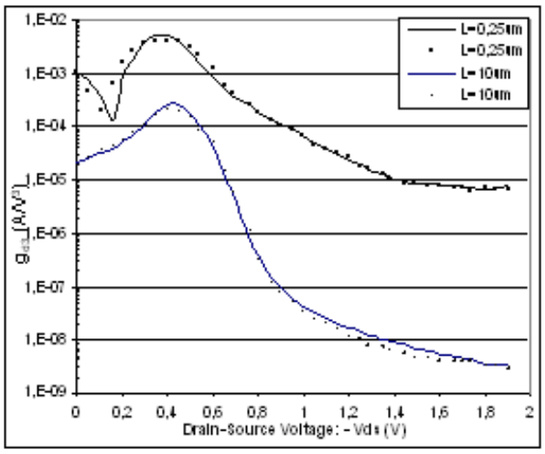

Fig. 19. g_{d3} for p-type MOS Transistor

The transistors array is performed in 0,25µm technology with transistors n and p enhanced type and n^+ polysilicon gate for transistors type n and p^+ polysilicon gate for transistors type p. The gate oxide thickness is about 4nm and the maximum allowed supply voltage V_{DD} is 2,5V. The LDD structure is no longer used and description of gate voltage dependent series resistance become redundant and the model will be a constant source and drain resistance.

8.7 Conclusions

Precise physical description of mobility degradation in circuit-level MOSFET models is essential for distortion analysis. In strong-inversion both hole and electron mobility is mainly limited by two scattering mechanisms: phonon scattering and surface roughness

scattering, but they still exhibit a different dependence on effective normal field E_{eff}. For long channel MOS transistors at low drain voltage V_{DS}, the I_D current expression and its higher-order derivatives is accurate for a large region of gate voltages and bulk voltages. The results can be extended to short channel length transistors, when the gate voltage dependency of series resistance was also incorporated in the models.
Gate induced distortion increases with decreasing channel length as series resistance becomes more important. To minimize distortion in circuits, it is often important to make the third-order derivative zero. This is possible for long channel n-type MOS transistors, where mobility degradation is more dominant than the effect of series resistance.
The DC-biasing point for which g_{m3} becomes zero is essentially determined by the presence of the E^2_{eff} term and can be accurately predicted using equation (37). So the designers of analog integrated circuits, where the sensitivity to the modelling details and the feedback between components is more important than for digital electronics can simulate, optimise and design more efficiently.
PMOS transistors are the better choice for drain terminal driven low-distortion applications, because is less affected by the harmonic distortion in the ohmic region and from weak-avalanche effects, compared with n-type transistors.
The unified model gives an accurate description of harmonic distortion in all inversion. The model has a low number of parameters for both n or p type of transistors. The parameters extraction requires the measured current-voltage characteristics, which can be performed with usual testing tools. The unified model is well scalable and the drain current and its higher order derivatives are precisely described over a large geometries range and terminal applied voltage values. The model can be used for modern CMOS technologies in which both the n-channel and p-channel transistors are of the enhancement-type and LDD-structures are no used, it was found that an adjusted expression of hole mobility and a constant series resistance have to be used.

9. High k dielectric-semiconductor interface modelling and analysing

9.1 Introduction

High-k dielectric oxides are presently investigated as alternative gate dielectric films for complementary metal oxide silicon-CMOS transistors. The continuous decrease of ultra-large-scale integration dimensions determines that SiO_2 gate dielectric attends critical dimensions of tens of Å and reaches fundamental limitations in preventing current leakage from the gate into the channel (Lo, 1997). One possible solution is to replace SiO_2 with higher permittivity insulator materials. Among the promising candidate materials investigated are HfO_2, $HfSiO_4$, $ZrSiO_4$, La_2O_3 and Y_2O_3. Presently are important issues related to the integration of these materials and that makes highly desirable to previously understand how the new materials properties affects the functionality of CMOS devices. A correct quantum-mechanical model must properly evaluate the channel charge distribution and the leakage current flowing between the gate and the channel through tunnelling. Consequently the gate, the insulator and the substrate semiconductor channel must be considered like a single space region accessible to all free charge carriers. Looking from the gate this high quality thin oxide is responsible for the continued increase of the gate leakage, which increase the power consumption of integrated circuits. The understanding of the

MOS system begins very important for research the tunnelling current in EEPROM devices and also in high performance MOS devices with ultra thin oxides (Cassan, 2000).

9.2 The gate leakage currents

The charge distribution and quantum-mechanical leakage currents in ultra thin metal-insulator-semiconductor gate stacks composed of several layers materials are very important (Yeo, 2002). Considering all the capacitor like a single quantum mechanical quantity the effective mass approximation for the electrons in the different valley and the Hartree approximation for the electron-electron interaction in inversion layer, the Schrödinger-Poisson equation can be solved. Because the insulating layer is relatively thin but the energy barriers separating the inversion layer from the gate electrode is high enough to prevent the flow of electrons to the gate, the potential well host the majority of inversion layer electrons and the channel is coupled only weakly with the gate (Magnus, 2000).

9.3 The iterative approximation method

The first fully numerical self-consistent results of the inverted MOS structure were mainly attributed to Stern. Then the self–consistent solution has been extended to holes in inverted pMOS structure by Moglestue. The quantum mechanical treatment of the MOS structure in the accumulation regime was described by Sune (Sune, 1992). The self-consistent Schrödinger-Poisson equations were applicable to an inverted structure in the next approximations: the effective mass approximation, the ideal interface semiconductor-oxide and interruption of wave function at interface semiconductor-oxide. The time-independent Schrödinger equation in 3D space, using the position vector $\mathbf{R}=(\mathbf{r}, z)$ can be formally written:

$$H\psi(\mathbf{r}, z) = E\psi(\mathbf{r}, z), \tag{46}$$

where $\psi(\mathbf{r},z)$ is the wave function, E is the eigenvalue energy, H is the system Hamiltonian, composed from kinetic energy T and potential energy W. For long channel device the potential profile is mainly one dimensional and the drain and source regions can be considered like electrons reservoirs for the inversion layer. The 1D simplification allows using the wave operator like a function of the z coordinate only:

$$\psi(\mathbf{r},z) = \phi(z)e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{47}$$

where $\mathbf{k}=(k_x,k_y)$ is the wave vector in the (x,y) plane. So the carrier are quantized in the z direction and are free to move in the $\mathbf{r}=(x,y)$ plane, with a continuous energy component. After phase transformation and imposing the constraint of vanishing for the first derivative of the wave function, the envelope 1D time-independent reduced equation (46) is:

$$-\frac{\hbar^2}{2m_{zz}}\psi''(z) + W\psi = E_z\psi(z) \tag{48}$$

where $\hbar$ is reduced Planck constant, m_{zz} is the effective masses in m_o units, W is potential energy, $\Psi(z)$ is the 1D envelope wave functions and E_z is the eingenvalue energy.

Considering the MOS structure a quantum mechanical system, an externally applied gate bias induces a potential well that confines carriers in the region of the semiconductor-oxide interface. The electrostatic potential and charge respect the Poisson equation in any z direction from silicon region:

$$\frac{d^2V(z)}{dz^2} = -\frac{1}{k_{Si}\varepsilon_0}\rho(z) \tag{49}$$

where $V(z)$ is the electrostatic potential, $\rho(z)$ is the charge density, k_{Si} is the Si relative dielectric constant. Assuming the p-type substrate with completely ionized impurities and neglecting the hole concentration in inversion can approximate the charge density:

$$\rho(z) = \rho_{depl}(z) - qn(z), \tag{50}$$

where ρ_{depl} is the depletion layer charge and $n(z)$ is the carrier's distribution.
Close to the interface the electrons have a position dependent concentration proportional with the probability density and a sum of each energy valley and subband.

$$n(z) = \sum_{i,j} n_{i,j}(z) = \sum_{i,j} N_{ij}^{(2D)}\left(E_{z,ij}, E_F\right)\left|\psi(z)\right|^2 \tag{51}$$

where $N_{ij}^{(2D)}$ is the subband population which integrates the all possible energies of a subband of the 2D density of states, $|\psi(z)|^2$ is the probability density, $E_{z,ij}$ is the solution of 1D Schrödinger equation (48) and represents the discrete bottom level of a particular energy subband j, for each valley i and E_F is Fermi energy level. The carrier's distribution can be more detailed using the valley and spin degeneracy and Fermi-Dirac statistics. The assumption that the silicon-oxide interface is ideally, was technologically realized by election the [001] surface orientation that minimizes the dangling bonds at the interface, resulting a high quality interface after passivation (Babarada, 2008).
Considering the quantization effects of silicon-insulator interface an approximate geometrical solution to calculate the charge densities and subband energy levels reduces consistently the computational complexity for leakage current evaluation. Using the same effective mass approximation the areal density of charge in the inversion layer is:

$$N_{inv} = \sum_{i,j} \oint_z N_{ij}^{(2D)}\left(E_{z,ij}, E_F\right)\left|\psi(z)\right|^2 dz = \sum_{i,j} N_{ij}^{(2D)}\left(E_{z,ij}, E_F\right) \tag{52}$$

Using the geometrical approximation of Si band bending in inversion (Muller, 1997) the energy level is:

$$E_{z,ij} = \left(\frac{\hbar^2}{2m_{z,i}}\right)^{1/3}\left(\pi q F_{ef}\frac{3}{2}\left(j+\frac{3}{4}\right)\right)^{2/3} \tag{53}$$

and the subband charge is:

$$q_{ij} = \frac{2E_{z,ij}}{3qF_{ef}}, \tag{54}$$

where F_{ef} is the $E_{z,ij}$ corresponding effective electric field. Then the inversion charge is:

$$q_{inv} = \sum_{i,j} q_{i,j}\frac{N_{i,j}^{(2D)}}{N_{inv}}, \tag{55}$$

and the total silicon surface bending:

$$\Psi_S = \Psi_D + q\frac{N_{inv}q_{inv}}{k_{Si}\varepsilon_0} + \frac{k_B T}{q} \tag{56}$$

where Ψ_S is the surface potential, Ψ_D is the drop voltage at surface due to space charge region. The last term is the influence of doping concentration to charge region (Muller, 1997). Using the charge boundary conditions the equations can be iteratively solved to attain the convergence in the next sequence:

- Guess the initial N_{inv}, Ψ_S and Ψ_D
- Consider charge boundary condition $N_{inv\text{-}bc}$
- Iterate Ψ_S with condition $N_{inv}(\Psi_S)/N_{inv\text{-}bc} \rightarrow 1$
- Iterate Ψ_D with condition $\Delta\Psi_D \rightarrow 0$

- Compute the potential distribution

We have possible loops from out to input, of step 3 and 4 and from out of step 4 to input of step 3. The method can be used also for tunnelling based leakage currents in high-k dielectric stach.

9.4 Results

For numerical simulations we used the ATLAS devices simulator software package from Silvaco. The main module program used is presented in fig. 20, in order to generate the MOS structure presented in fig. 21.

Then was calculated the gate current, fig. 22 and the capacity from gate to substrate, fig. 23, function of polysilicon doping concentrations $10^{19}cm^{-3}$, $10^{20}cm^{-3}$ and $10^{21}cm^{-3}$.

```
mesh
x.mesh loc=-0.01 spac=0.01
x.mesh loc=0.01  spac=0.01
y.mesh loc=-0.04 spac=0.001
y.mesh loc=0.02  spac=0.001
region number=1 x.min=-0.01 x.max=0.01 y.min=-0.04 y.max=-0.03  \
       material=aluminum
region number=2 x.min=-0.01 x.max=0.01 y.min=-0.03 y.max=-0.005 \
       material=poly
region number=3 x.min=-0.01 x.max=0.01 y.min=-0.005 y.max=0.0   \
       material=oxide
region number=4 x.min=-0.01 x.max=0.01 y.min=0.0 y.max=0.02     \
       material=silicon
electrode x.min=-0.01 x.max=0.01 y.min=-0.04 y.max=-0.03 name=gate
electrode bottom name=substrate
doping region=2 p.type concentration=1e19 uniform
doping region=4 p.type concentration=1e17 uniform
solve init
solve vgate=-1.5
solve vgate=-3
save outfile=mos2ex15-3V19.str
# tonyplot mos2ex15-3V19.str -set mos2ex15_BD.set
log outfile=mos2ex15_CV19.log
solve vgate=-2.8 vstep=0.2 vfinal=3.0 name=gate ac freq=1e6 previous
tonyplot mos2ex15_CV19.log -set mos2ex15_CV.set
```

Fig. 20. The main ATLAS program module

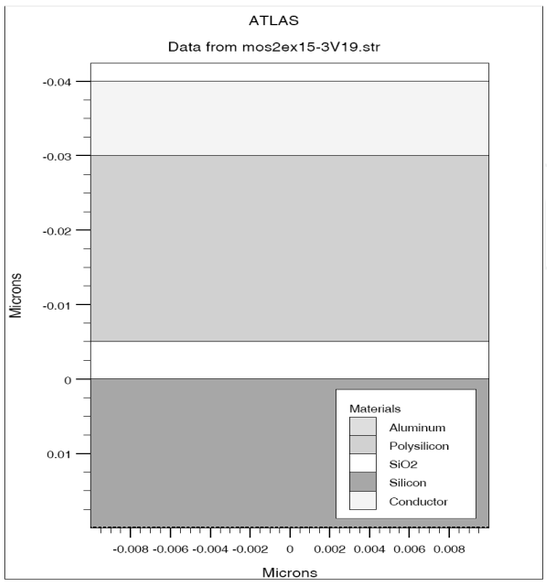

Fig. 21. The device structure

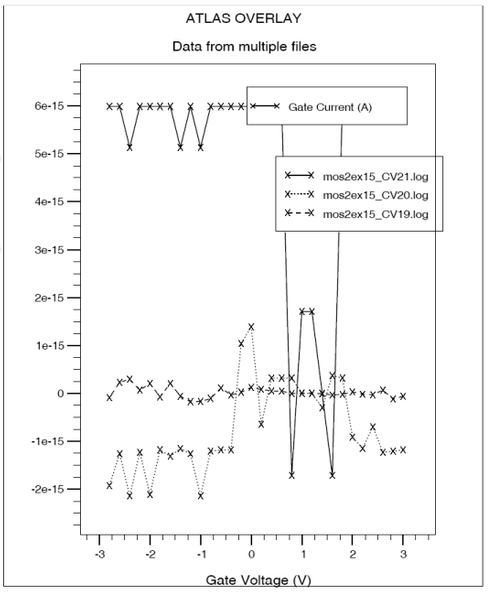

Fig. 22. The Gate Current

The first numerical simulations proves the dependence of leakage current, fig. 22 and depletion effect fig. 23, function of doping concentration like considered in chapter 9.3. Using the barrier height of 3.1eV, substrate doping $5x10^{17}cm^{-3}$, effective silicon oxide mass of $0.5m_o$ and donor poly doping $6x10^{19}$ the results of short computation iterative approximation, fig. 24, of silicon oxide current gate density calculated (1.5nm and 2nm) was in good agreement with experimental gate current density curves presented in (Yang, 2000) and noted [9] (1.41nm[9] and 1.95nm[9]).

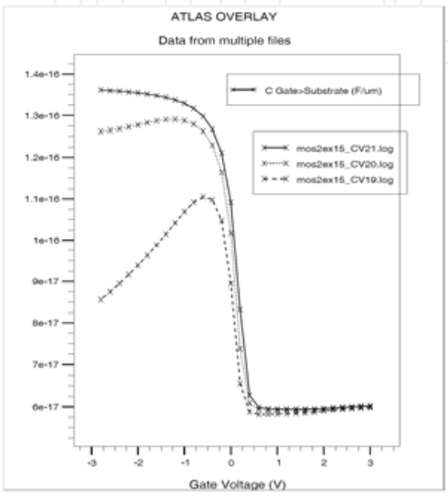

Fig. 23. The Gate-Substrate Capacity

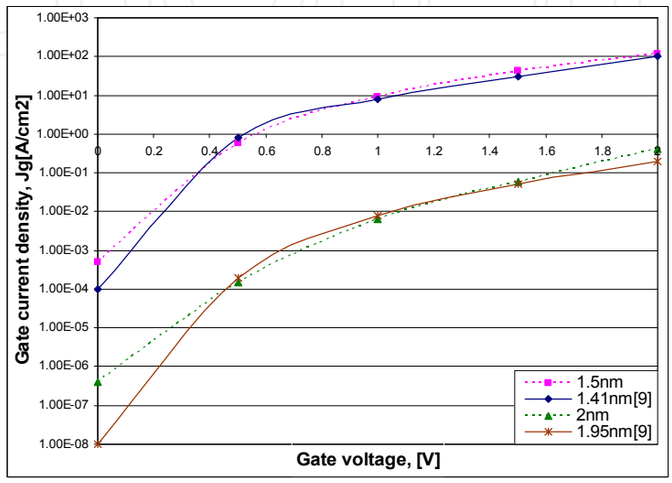

Fig. 24. Silicon oxide gate current density

A little overestimation of leakage current at high gate bias voltage is observed also in other reports (Buchanan, 2000), based of approximation of Fermi level by the value in the bulk silicon substrate.

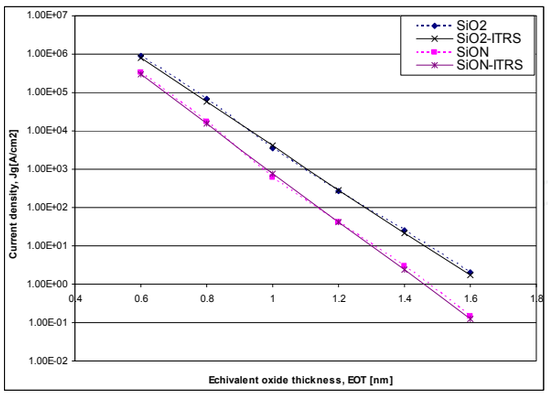

Fig. 25. Oxide and oxynitride leakage current

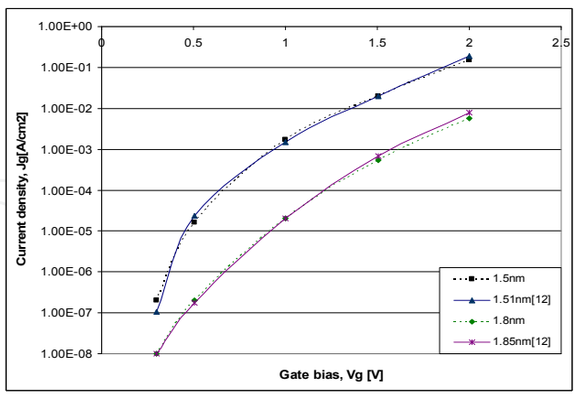

Fig. 26. Al_2O_3 high-k stacks leakage currents

The polysilicon doping level suppresses the gate leakage current for gate bias in inversion because the additional voltage drops over the depleted layer (Yang, 2000). This solution decreases the drive capacitance and the device performances. The substrate doping level affects the leakage current through the surface potential of the channel. Because increasing the physical thickness of gate dielectric affects the device parameters like drive current, a

compromise solution is to increase the dielectric constant using the SiON layer with dielectric constant up to 7.6 for Si_3N_4. The performances of SiON like gate dielectric are better than SiO_2 as in fig. 25, according with simulations at Vg=1V and ITRS. Comparing the calculated data with gate leakage current through Al_2O_3 high-k dielectric stacks presented in (Buchanan, 2000) a good fit was obtained, fig. 26.

9.5 Conclusions

High-k atomic layer deposition stacks like insulating in the metal-insulating-semiconductor structure was studied. An iterative approximate method to calculate the 1D MOS structures main electric parameters without using the Schrödinger-Poisson equations was used. This method is based on approximation of effective field function of doping parameters. The tunnelling currents can be calculated more rapidly and the study for different gate dielectric stacks can be made. The precision can be increased by 2D or 3D analysis of Schrödinger-Poisson equations. The main application is to calculate the direct tunnelling current due to the thin oxide layers. The method is extensible to high-k dielectric stacks in order to study the influence of several material parameters like the impact of layer thickness on gate leakage and the approach of gate stack scalability. The results obtained using numerical calculation show that the increase of the gate dielectric constant has a very important effect in reducing the leakage currents. Comparing the results from fig. 24 and fig. 26 for 1V gate bias and 1.5nm thickness the increase of dielectric constant to 7 reduce the leakage current with 4 order of magnitude. Other simulations show that the leakage current decrease significant when the interfacing oxide is completely eliminated. Future works will be focus of other high-k dielectric stacks like HfO_2, $HfSiO_4$, $ZrSiO_4$, La_2O_3, and Y_2O_3.

10. References

Babarada, F.; et all. (2003). Carrier Mobility and Series Resistance MOSFET Modelling, *BioMEMS and Nanotechnology*, vol. 5275-49, pp. 354-363, SPIE's, Perth, Australia

Babarada, F.; et all. (2005). MOSFET Conductance Modelling Including Distortion Analysis Aspects, *Proceedings of the International Conference*, Sinaia, România

Babarada, F.; Plugaru, R.; Rusu, A. (2008). Electrical characterization of atomic layer high-k dielectic gate for advanced CMOS devices, *Proceedings of the International Conference*, pp. 363-366, ISBN 978-1-4244-2004-9, Sinaia, România

Buchanan, D.A.; et all. (2000). 80 nm poly-silicon gated n-FET with ultra thin Al_2O_3 gate dielectric for ULSI applications, *IEDM Tech. Dig*., pp. 223-226, San Francisco, USA

Bucher, M.; et all. (2007). A Scalable Advanced RF IC Design-Oriented MOSFET Model, *Int. Journal of RF and Microwave Computer-Aided Engineering*, DOI 10.1002/mmce

Campian, I.; Profirescu, O.; Babarada, F.; Lakatos, E. (2003). MOSFET Simulation-TCAD Tools/Packages, *Proceedings of the Int. Conf.*, pp. 235-238, Sinaia, România

Cassan, E. (2000). On the reduction of direct tunnelling leakage through ultra-thin gate oxides by one-dimensional Schr.-Poisson solver, *J. Appl. Phys.* 87(11), pp. 7931-7939

Gildenblat, G.; Zhu, Z.; McAndrew, C. (2009). Surface potential equation for bulk MOSFET, *Solid-State Electronics*, 53: pp. 11-13

Govoreanu, B.; et all. (2002). On the use of Bayesian Neural Networks for TCAD Empirical Modeling, *Romanian Journal Science and Technology*, pp.329-338, vol 5, no 4, România

Kwong, M.; Kasnavi, R.; Grifin, P.; Duton, R. (2002). Impact of Lateral Source/Drain Abruptness on Device Performance, *IEEE Trans. on El. Dev.*, vol. 49, no. 11.

Lo, S.; Buchanan, D.; Taur, Y.; Wang, W (1997). Quantum mechanical modelling of electron tunnelling current from the inversion layer of ultra-thin-oxide nMOSFET, *IEEE El. Dev. Lett.*, 18(5), pp. 209-211

Magnus, W.; Schoenmaker, W.; (2000). Full quantum mechanical model for the charge distribution and the leakage currents in ultrathin metal-insulator-semiconductor capacitors, *J. Appl. Phys*. 88(10), pp. 5833-5842

Muller, H.; Schultz, M. (1997). Simplified method to calculate the band bending and the subband energies in MOS capacitors, *IEEE Trans. El. Dev*. 44(9), pp. 1539-1543

Rusu, A. (1990). *Microelectronics Active Components Modelling*, Editura Academiei Române

Scholten, A.J.; et all. (2009). The new CMC standard compact MOS model PSP: advantages for RF applications, *IEEE Journal of Solid-State Circuits*, vol. 44, no. 5, pp. 1415-1424

Sune, J.; Olivio, P.; Ricco, B. (1992). Quantum mechanical modelling of accumulation layers in MOS structures, *IEEE Trans. El. Dev.*, 39(7), pp. 1732-1739

Veendrick, H. (2008). *Nanometer CMOS ICs: From Basics to ASICs*, Springer, ISBN 978-1-4020-8332-7, Netherland

Yang, N.; Henson, W.; Wortman, J. (2000). A comparative study of gate dielectric tunnelling and drain leakage currents in n-MOSFET with sub-2-nm gate oxides, *IEEE Trans. El. Dev.* 47(8), pp. 1636-1644

Yeo, Y.; King, T.; Hu, C.; (2002). Direct tunnelling leakage current and scalability of alternative gate dielectrics, *Appl. Phys. Lett.*, 81(11), pp. 2091-2093

Ytterdal, T.; Cheng, Y.; Fjeldly, T. (2003). *Device Modeling for Analog and RF CMOS Circuit Design*, J. Wiley, ISBN 0-471-49869-6, England

IntechOpen

Semiconductor Technologies

Edited by Jan Grym

ISBN 978-953-307-080-3
Hard cover, 462 pages
Publisher InTech
Published online 01, April, 2010
Published in print edition April, 2010

Semiconductor technologies continue to evolve and amaze us. New materials, new structures, new manufacturing tools, and new advancements in modelling and simulation form a breeding ground for novel high performance electronic and photonic devices. This book covers all aspects of semiconductor technology concerning materials, technological processes, and devices, including their modelling, design, integration, and manufacturing.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Florin Babarada (2010). Semiconductor Processes and Devices Modeling, Semiconductor Technologies, Jan Grym (Ed.), ISBN: 978-953-307-080-3, InTech, Available from: http://www.intechopen.com/books/semiconductor-technologies/semiconductor-processes-and-devices-modelling

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen