Oxide roughness induced fluctuation effects in nanoscale MOSFET devices

PETRU ANDREI
Department of Electrical and Computer Engineering
Florida State University and Florida A&M University
2525 Pottsdamer St., Tallahassee FL 32309
USA
http://www.eng.fsu.edu/~pandrei

Abstract: - Oxide roughness induced fluctuation effects in Metal-Oxide-Semiconductor Field-Effect-Transistors (MOSFET) are analyzed by using the linearization (perturbation) technique. The oxide thickness is considered a random variable in the framework of the transport equations and all fluctuating quantities are linearized around their average values. Numerical results for a MOSFET device with simplified structure are presented for the fluctuations of the terminal currents and threshold voltage. Quantum mechanical effects are taken into consideration by using the Density-Gradient model.

Key-Words: - linearization technique, sensitivity analysis, oxide thickness fluctuations, MOSFET.

1 Introduction

It is well-known that parameters of nanoscale semiconductor devices are very susceptible to random variations of the oxide thickness and asperities of the oxide/semiconductor interfaces [1]. Random fluctuations of oxide thickness are especially pronounced in small devices where the spatial scales of these fluctuations are more or less comparable with device dimensions. For this reason, an accurate analysis of random oxide thickness induced effects is very important for further progress in the area of semiconductor device technology.

There are two conceptually different approaches to the analysis of random oxide thickness induced fluctuations in semiconductor devices. The first approach [2], known as the Monte Carlo approach, is based on generating numerous realizations of the oxide/semiconductor interface and solving the transport equations for each of such realization. Statistics of different parameters of interest are then accumulated and used to evaluate the average values and variances of those parameters. These methods are computationally very expensive since the same device-level simulations have to be performed many times. The second approach [3]-[5], known as the linearization technique is based on linearization of the transport equations with respect to the fluctuating quantities and allows the computation of the standard deviation of intrinsic parameters of semiconductor devices very efficiently. The linearization technique is much more computationally efficient than the Monte Carlo method because it circumvents computations for many devices. In addition, the linearization technique provides information on the sensitivity of parameter variances to locations of the random oxide thickness fluctuations, which makes it instrumental in the design of fluctuation resistant structures. In this article we develop a linearization technique for the analysis of fluctuations of terminal currents and threshold voltages in MOSFET devices.

The article is structured as follows. The transport model used in the simulations is briefly presented in Section II. In Section III we present the linearization technique and we apply it to the analysis of fluctuations of terminal currents and threshold voltage of MOSFET devices. Numerical simulations results are presented in Section IV, which is followed by conclusions.

2 Transport model

The electron and hole concentrations in nanoscale MOSFET devices can be described by the following system of partial differential equations [6]:

\[
\nabla \cdot (\varepsilon \nabla \varphi) + q(p - n + D) = 0 ,
\]

\[
\frac{\partial n}{\partial t} = \frac{1}{q} \nabla \cdot (j_n) - R(\varphi, n, p) ,
\]

\[
\frac{\partial p}{\partial t} = \frac{1}{q} \nabla \cdot (j_p) - R(\varphi, n, p) ,
\]

\[
j_n = -n\mu_n \nabla \varphi + D_n \nabla n - 2\mu_n n \frac{\nabla \cdot (h_n \nabla \sqrt{n})}{\sqrt{n}} ,
\]
\[ j_p = -p\mu_p \nabla \phi - D_p \nabla p + 2\mu_p p \nabla \left( \frac{\nabla \cdot (b_p \nabla \sqrt{p})}{\sqrt{p}} \right), \quad (5) \]

where \( R(\phi, n, p) \) is the net electron-hole recombination rate, \( \phi \) is the electric potential, \( n, \mu_n, D_n \) and \( p, \mu_p, D_p \) are the concentrations, mobilities, and diffusion coefficients of the electrons and holes, respectively, and \( j_e \) and \( j_p \) are the electron and hole current densities. The Quantum mechanical effects are “controlled” by the parameters [7]:

\[ b_n = \frac{\hbar^2}{4r_m n^2}, \quad b_p = \frac{\hbar^2}{4r_m p^2}, \quad (6) \]

where \( m^*_n \) and \( m^*_p \) denote the effective masses of the electrons and holes, while \( r_n \) and \( r_p \) are dimensionless parameters that account for the statistics of electrons and holes in semiconductor devices. The values of \( r_n \) and \( r_p \) vary asymptotically from 1, when only the lowest energy subband is occupied (e.g. at low temperature), to 3 when other subbands become populated as well (e.g. at high temperature).

In our simulations the electron and hole mobilities are described by the model presented in [8]. Shockley-Read-Hall and Auger recombination processes are considered in the current continuity equations (2) and (3). The net recombination term is assumed:

\[ R(\phi, n, p) = R_{SRH}(\phi, n, p) + R_{AU}(\phi, n, p), \quad (7) \]

where the Shockley-Read-Hall term is given by:

\[ R_{SRH}(\phi, n, p) = \frac{n_p - n_i^2}{\tau_n (n + n_i) + \tau_e (p + p_i)}, \quad (8) \]

where \( n_i = n_i e^{\frac{E_{Bu}}{kT}}, \quad p_i = n_i e^{\frac{E_{Bu}}{kT}}, \quad E_{Bu} \) is the difference between the defect level causing the trap and the intrinsic level, \( \tau_n \) and \( \tau_p \) are the electron and hole lifetimes. In our simulations, \( \tau_n \) and \( \tau_p \) depend on the doping concentration according to the Scharfetter relation:

\[ \tau_{n,p} = \tau_{n,p}^{\text{min}} + \tau_{n,p}^{\text{max}} - \tau_{n,p}^{\text{min}} \frac{N_{ref}^{n,p} + N_{ref}^{p,n}}{1 + \gamma_{n,p} + \gamma_{p,n}}, \quad (9) \]

where \( \tau_{n,p}, \tau_{n,p}^{\text{min}}, \tau_{n,p}^{\text{max}}, \gamma, \) and \( N_{ref}^{n,p} \) are parameters that calibrate the electron and hole lifetimes. The Auger recombination term is given by:

\[ R_{AU}(\phi, n, p) = C_n n + C_p p \left( n_p - n_i^2 \right), \quad (10) \]

where \( C_n \) and \( C_p \) are parameters that depend on the electron and hole concentrations. For a detailed description of the recombination terms and the values of the calibration parameters we recommend [8].

### 3 Technical approach

In general, any intrinsic parameter \( A \) of the device such as terminal currents, small-signal parameters, or threshold voltages can be written as a function of oxide thickness vector \( t \):

\[ A = A(t). \quad (11) \]

In equation (11), we have considered that the oxide thickness is a function of position. Hence, if the region of the oxide thickness is spatially discretized into \( N_{ox} \) cells, the oxide thickness can be specified as an \( N_{ox} \)-dimensional column vector whose components are the local values of the oxide thickness at different locations on the semiconductor-oxide interface (see Fig. 1):

\[ t = [t_1, t_2, \ldots, t_{N_{ox}}]. \quad (12) \]

![Fig. 1. Cross-section through the oxide layer.](image)

The oxide thickness is a random quantity and can be written as the sum of its average values \( t_0 \) and fluctuations \( \tilde{t} \):

\[ t = t_0 + \tilde{t}, \quad (13) \]

where by definition, the expected value \( \tilde{t} \) is equal to zero. The basic idea for the computation of fluctuations of parameter \( A \) is to linearize equation (11) with respect to the oxide thickness. In the first-order approximation, the fluctuations of parameter \( A \) can be written as follows:

\[ \Delta A = \frac{\partial A}{\partial t} \tilde{t} = \sum_{j} \gamma_{ij} \tilde{t}_j, \quad (14) \]

where \( \gamma_{ij} = \frac{\partial A}{\partial t_j} \) are the so-called oxide thickness
superposition coefficients. These coefficients show how sensitive parameter $A$ is to the fluctuations of oxide thickness at specific locations on the oxide/semiconductor interface. The last equation allows us to derive the following expression for the variance of parameter $A$:

$$\sigma_A^2 = \sum_{i,j} \gamma_i^T \cdot \gamma_j ACF(i,j),$$

(15)

where $ACF(i,j)$ is the autocorrelation function of the oxide thickness, which is defined as follows:

$$ACF(i,j) = \langle (t_i - \langle t_i \rangle) \cdot (t_j - \langle t_j \rangle) \rangle = \langle \tilde{t}_i \cdot \tilde{t}_j \rangle.$$  

(16)

where $\langle \tilde{t} \rangle$ stands for the average oxide thickness. This function can be measured directly by using atomic force microscopy experiments. In most cases, $ACF(i,j)$ can be approximated by an exponential or Gaussian distribution function.

It should be noted that the superposition coefficients are mesh dependent. Therefore, it is convenient to introduce the oxide thickness “sensitivity coefficients,” $(S^i_A)$ which are defined as:

$$S^i_A = \frac{\gamma_i^T \cdot \gamma_j ACF(i,j)}{\Delta S^i},$$

(17)

In formula (17), $\Delta S^i$ is the transversal area of the $i$th discretization cell on the oxide/semiconductor interface. Equation (15) now reads as follows:

$$\sigma_A^2 = \sum_{i,j} \gamma_i^T \cdot \gamma_j ACF(i,j),$$

(18)

### 3.1 Computation of the superposition coefficients of the terminal currents

The analysis of current fluctuations is extremely important for both digital and analog applications. Next we introduce the method for the computation of terminal currents superposition coefficients. In order to make the method suitable for numerical implementation on standard semiconductor device simulators, we present it in compact matrix form.

For the sake of brevity we will consider that

$$\frac{\partial n}{\partial t} = \frac{\partial p}{\partial t} = 0 \quad \text{(steady state)}$$

and write the transport equations (1)-(3) in discretized form [9]:

$$F(X,t) = 0,$$

(19)

where $F$ is a nonlinear vector function of the unknown “state” vector $X$ and oxide thickness vector $t$. The state vector consists of the mesh-point values of the electrostatic potential $\varphi$, electron and hole concentration $n$, and $p$, electron and hole density currents $j_n$ and $j_p$: $X = [\varphi, n, p, j_n, j_p]$. The last equation denotes the transpose of the vector.

Fluctuations of the oxide thickness $\tilde{t}$ induce fluctuations in the state vector $\tilde{X}$ that can be computed by solving the following linear system of equations [5]:

$$\tilde{F}_X \tilde{X} + \tilde{F}_t \tilde{t} = 0,$$

(20)

where $\tilde{F}_X$ and $\tilde{F}_t$ are the derivatives of $F$ with respect to $X$ and $t$, respectively. These derivatives are computed at the given dc bias point and by assuming constant (non-fluctuating) values of the doping concentration.

Now, let us denote the current through the drain terminal of the MOSFET device by $I_{\text{Drain}}$. In order to compute the superposition coefficients of $I$ it is convenient to write it as a explicit function of state vector $X$ and oxide thickness $t$:

$$I_{\text{Drain}} = I(X,t).$$

(21)

The fluctuations of terminal current can be found by linearizing (21) with respect to the fluctuating quantities:

$$\tilde{I} = (I_X) \cdot \tilde{X} + (I_t) \cdot \tilde{t},$$

(22)

where $I_X$ and $I_t$ are the derivatives of $I$ with respect to the state variable and doping concentration. By solving equations (20) and (22) for the fluctuations of the terminal current, one gets:

$$\tilde{I}_{\text{Drain}} = -\left[ g' \cdot \tilde{F}_p - (I_t) \right] \cdot \tilde{t},$$

(23)

where $g'$ is the transpose of column vector $g$, which is the solution of the following linear system of equations:

$$\tilde{F}_X g = I_X,$$

(24)

where $\tilde{F}_X$ denotes the transpose of matrix $\tilde{F}_X$. By comparing (14) and (23), it can be inferred that the superposition coefficients of the terminal current are given by the following formula:

$$\gamma_i = -(g' \cdot \tilde{F}_p - (I_t)),$$

(25)

The standard deviation of terminal current can be calculated now by using: $\sigma^2 = \sum_{i,j} \gamma_i^T \cdot \gamma_j ACF(i,j)$.

The most expensive computational task related to the calculation of these coefficients is to solve linear system (24). For 2-D simulations this system can be solved numerically by using the classical Gauss-Seidel method. However, for 3-D simulations the computation time and memory requirements increase considerably if the Gauss-Seidel method is
used, so other procedures are needed to solve (24). In our simulations, we took advantage of the diagonal dominance of matrix $\hat{F}_X$ and solved it efficiently by using the Successive Over-Relaxation (SOR) method.

### 3.2 Computation of the superposition coefficients of the threshold voltage

The technique for the computation of superposition of threshold voltage closely mimics the technique for the computation of superposition coefficients of terminal currents presented in the previous section. The discretized transport equations are now written as [9]:

$$F(X,t,V_g) = 0,$$  \hspace{1cm} (26)

where $F$ the transport equations have been written as functions of the “state” vector $X$, oxide thickness $t$, as well as gate voltage $V_g$. If $\hat{X}$ denotes the fluctuations of the state variable and $\hat{V}_g$, the fluctuations of the gate potential, we can write in the first-order approximation that:

$$\hat{F}_X \hat{X} + \hat{F}_t \hat{t} + F_{V_g} \hat{V}_g = 0,$$  \hspace{1cm} (27)

where $\hat{F}_X$, $\hat{F}_t$, and $F_{V_g}$ are the derivatives of $F$ with respect to $X$, $t$, and $V_g$, respectively.

At threshold voltage the drain current is constant and the fluctuations of the drain current are equal to zero:

$$0 = \hat{I} = I'_d \hat{X},$$  \hspace{1cm} (28)

where $I'_d$ is the transpose of the gradient of $I$ with respect to $X$. Equations (27) and (28) are coupled equations with unknowns $\hat{X}$ and $\hat{V}_g = \hat{V}_g$. In order to decouple them, we multiply equation (27) from the left by $I'_d F'_{\hat{X}}$ and use the constraint (28). After a few rearrangements we obtain the following equation for the fluctuations of threshold voltage:

$$\hat{V}_t = -\frac{g' \cdot \hat{F}_t}{g' \cdot F_{V_g}} \hat{t} .$$  \hspace{1cm} (29)

By comparing equations (29) and (23) it follows that the superposition coefficients of the threshold voltage are given by the equations:

$$\gamma'_{V_t} = -\frac{g' \cdot \hat{F}_t}{g' \cdot F_{V_g}} .$$  \hspace{1cm} (30)

### 4 Numerical results

The technique presented in the previous section was numerically implemented and used to compute the fluctuations of the threshold voltage in MOSFET devices induced by random oxide roughness. 2-D and 3-D DG models have been assumed throughout the simulations, with the value of effective electron mass $m^*_e = 0.21m_0$. For n-channel devices, the hole current is very small, hence, the total current is not very sensitive to the value of the effective hole mass. In our simulations, we assumed that $m^*_p = 0.49m_0$ (see Ref. 7).

Unless otherwise stated, the MOSFET device used in our simulations has a simplified structure with retrograde doping profile. The channel doping concentration decreases from $5 \times 10^{18}$ cm$^{-3}$ at 20 nm (and deeper) to $5 \times 10^{17}$ cm$^{-3}$ at the surface according to a truncated Gaussian distribution function. The source and drain profiles have a Gaussian distribution with n-type peak surface concentration of $10^{20}$ cm$^{-3}$ and vertical struggles of about 8.2 nm that correspond to a junction depth of about 20 nm. The lateral source and drain struggles (in the directions parallel to the conduction channel) are about 1.34 nm and the source and drain extensions under the gate are 4.4 nm. The metallurgical channel length is 30 nm and this corresponds to an effective channel length of about 34 nm (the effective channel length is defined by the points where the source-drain doping concentrations fall to $2 \times 10^{19}$ cm$^{-3}$). The thickness of the oxide is 2 nm and the width of the device is 40 nm. In the reported simulations, one of the above parameters is usually varied, while the other ones are held constant.

Recent measurements made with the help of atomic force microscopy (AFM) show that the autocorrelation function of the oxide thickness fluctuations can be described by an exponential distribution function:

$$ACF(r) = \Delta^2 e^{-r/L_c},$$  \hspace{1cm} (31)

where $\Delta$ is the roughness of the surface, $L_c$ is the correlation length of the fluctuations, and $r$ is the spatial distance measured at the surface of the oxide. Roughness can be accurately measured experimentally and the reported values range between 0.15 nm and 0.5 nm for SiO2 surfaces. Since the standard deviation of $V_t$ increases linearly with $\Delta$, we assume that roughness has a constant value of $\Delta = 0.15$ nm; for any other $\Delta$, the standard deviation of $V_t$ can be easily computed by using...
appropriate scaling. The correlation length depends mostly on the fabrication process and it is more difficult to measure experimentally. The values of $L_c$ measured by using AFM vary from 10 nm to 25 nm and the simulations presented in this section take this uncertainty into account by presenting results for the whole range of variation of $L_c$.

Figures 2(a) and 2(b) present the computed standard deviations of $V_T$ as a function of the channel length for two groups of devices. In the first group [Fig. 2(a)], devices are scaled down by using the constant field scaling rule [1], according to which the device dimensions and the doping are scaled proportionally, by the same factor. In the second group [Fig. 2(a)], devices are scaled down by using the constant voltage scaling rule according to which, if dimensions are decreased by a factor $k$, the doping concentration is increased by $k^2$. We observe that the values of the standard deviation of $V_T$ are smaller in the case of constant field scaling because of lower doping concentration in the channel. The same effect was observed by Nishinohara et al. [10] for random doping induced fluctuations of $V_T$.

In the case of long channel MOSFET devices, threshold voltage increases linearly with the oxide thickness. For correlation lengths that are large in comparison with the device dimensions, the standard deviation of the threshold voltage can be computed by using the “inversion” definition of $V_T$:

$$\sigma_{TV} = \frac{\Delta}{\varepsilon_O} \sqrt{\frac{4e_N kTN_a \ln \left( \frac{N_u}{n_j} \right)}{\varepsilon_O^2}}$$

where $N_u$ is the average doping concentration in the channel. The standard deviation of $V_T$ computed by using (32) is represented in figures 2(a) and 2(b) by dash line. Numerical simulations correctly predict the theoretical value of $\sigma_{TV}$ for long channel devices, which proves that the linearization method presented in the previous section calculates the variance of $V_T$ for long devices accurately.

For small MOSFET devices we compare our results with those published in Ref. [2]. For a 30 nm channel length device with abrupt junctions, $N_u = 5 \times 10^{18}$ cm$^{-3}$, $t_{ox} = 1.05$ nm, $x_j = 7$ nm, and by assuming constant electron and hole mobilities, as well as long correlation lengths of the oxide thickness fluctuations, our computations give the value $\sigma_{TV} = 41.5$ mV. This value is in perfect agreement with those reported in Ref. [2], which was obtained by using the Monte-Carlo technique:

$$\sigma_{TV} = (41 \pm 2) \text{ mV}.$$
Fig. 3. Standard deviation of threshold voltage as a function of the average doping concentration in channel (a), oxide thickness (b), and channel length (c), respectively.

5 Conclusion
Random oxide roughness induced fluctuations are becoming increasingly important in ultra small semiconductor devices, where the spatial scales of these fluctuations are comparable with the device dimensions. Significant deviations from the average values have been found for the threshold voltage. For a MOSFET device with doping concentration of the order of $10^{18}$ cm$^{-3}$ and channel length of 50 nm the fluctuations of the terminal currents and threshold voltages represent between 5% and 15% of the average values of these parameters. However, fluctuations increase considerably for smaller device dimensions.

6 Acknowledgements
This work is supported by the Army High Performance Computing Research Center (AHPCRC) under the auspices of the Department of the Army, Army Research Laboratory (ARL) under Cooperative Agreement number DAAD19-01-2-0014.

References: