NUMERICAL SIMULATION OF SEMICONDUCTOR DEVICES

Siegfried Selberherr and Erasmus Langer
Institute for Microelectronics
Technical University Vienna
A-1040 Wien / Austria

ABSTRACT

The state of the art in self-consistent numerical modeling of semiconductor devices is reviewed. The physical assumptions which are required to describe carrier transport are discussed. Particular emphasis is put on the models for space charge, carrier mobility, carrier temperature, and carrier generation-recombination. Investigations about three-dimensional effects due to the field oxide in MOS-devices analyzed with MINIMOS 5 are presented.

INTRODUCTION

Device Modeling based on the self-consistent solution of fundamental semiconductor equations dates back to the famous work of Gummel in 1964. Though since that time there has been a continuous progress on that field the shrinking dimensions of the elements of integrated circuits require even more suitable device models in view of physics and mathematics for accurate simulation. On the one hand more and more sophisticated physical models are needed (even for low ambient temperature applications [13], [17]) and on the other hand a great demand for three-dimensional simulation tools has appeared as usual two-dimensional device simulation is not applicable satisfactorily to narrow channel devices. Modern strategies for the application of simulation can be found in [4], [28].

PHYSICAL ASPECTS

Basic Equations

The model for hot carrier transport used in any numerical device simulation is based on the well known fundamental semiconductor equations (1)-(5). There are ongoing arguments in the scientific community whether these equations are adequate to describe transport in submicron devices. Particularly the current relations (4) and (5) which are the most complex equations out of the set of the basic semiconductor device equations undergo strong criticism in view of, for instance, ballistic transport [23], [35]. Their derivation from more fundamental physical principles is indeed not at all straightforward. They appear therefore with all sorts of slight variations in the specialized literature and a vast number of papers has been published where some of their subtleties are dealt with. Anyway, recent investigations on ultra short MOSFET's [37] do not give evidence that it is necessary to waive these well established basic equations for silicon devices down to feature sizes in the order of 0.1 microns [48]. Other approaches for semiconductor device simulation are based on the method of generalized stream functions [46] or on the Monte Carlo method [39].

\[
\begin{align*}
\text{div}(\varepsilon \cdot \text{grad} \psi) &= -\rho \\
\text{div} J_n = q \cdot \frac{\partial n}{\partial t} &= q \cdot R \\
\text{div} J_p + q \cdot \frac{\partial p}{\partial t} &= -q \cdot R \\
J_n &= q \cdot \mu_n \cdot n \cdot \left( \frac{E}{n} + \frac{1}{n} \cdot \text{grad} \left( \frac{n \cdot k \cdot T_n}{q} \right) \right) \\
J_p &= q \cdot \mu_p \cdot p \cdot \left( \frac{E}{p} - \frac{1}{p} \cdot \text{grad} \left( \frac{p \cdot k \cdot T_p}{q} \right) \right)
\end{align*}
\]

These equations include a set of parameters which have to be appropriately modeled in order to describe the various transport phenomena qualitatively and quantitatively correctly.

Modeling Space Charge

Poisson’s equation (1) requires a model for the space charge \( \rho \) which makes use of only the dependent variables \( \psi, n, p \) and material properties. The well established approach for this model is to sum up the concentrations with the adequate charge sign multiplied with the elementary charge (6).

\[
\rho = q \cdot (p - n + N_D^+ - N_A^-)
\]

Here a difference between room temperature and low temperature simulation becomes apparent. The doping concentration is usually assumed to be fully ionized at room temperature which intuitively does not hold for low temperature analysis. The classical way to describe partial ionization is based on the formulæ (7).

\[
\begin{align*}
N_D^+ &= \frac{N_D}{1 + 2 \cdot \exp \left( \frac{E_{fn} - E_D}{k \cdot T} \right)} \\
N_A^- &= \frac{N_A}{1 + 4 \cdot \exp \left( \frac{E_{fp} - E_A}{k \cdot T} \right)}
\end{align*}
\]

\( E_D \) and \( E_A \) are the ionization energies of the respective donor and acceptor dopant. A quite complete list of values can be found in [44]. These ionization energies are recommended to be modeled doping dependent in [12], however, it seems not to be important for MOSFET’s regarding my experience. The Fermi levels \( E_{fn} \) and \( E_{fp} \) have to be appropriately related to the dependent variables.

\[
\begin{align*}
\frac{E_{fn} - E_D}{k \cdot T} &= G_{1/2} \left( \frac{n}{N_c} \right) + \frac{E_c - E_D}{k \cdot T} \\
\frac{E_A - E_{fp}}{k \cdot T} &= G_{1/2} \left( \frac{p}{N_v} \right) + \frac{E_A - E_v}{k \cdot T}
\end{align*}
\]
\( G_{1/2}(x) \) is the inverse Fermi function of order 1/2 defined with (10).

\[
G_{1/2} \left( \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{\sqrt{y}}{1 + e^{y}} \, dy \right) = x
\]

\( N_e \) and \( N_v \) are the density of states in the conduction band and the valence band, respectively. The ratio of the density of states depends only on the ratio of the effective masses of electrons and holes. The product can be fitted to measured data of the intrinsic concentration requiring only models, e.g. [9], for the effective masses and the energy gap. With these two relations it is straightforward to compute the numerical values for the density of states. A full derivation of this partial ionization model can be found in [42].

**Modeling Carrier Mobilities**

The next set of physical parameters to be considered carefully for low temperature simulation consists of the carrier mobilities \( \mu_n \) and \( \mu_p \) in (4) and (5). The models for the carrier mobilities have to take into account a great variety of scattering mechanisms the most basic one of which is lattice scattering. The lattice mobility in pure silicon can be fitted with simple power laws.

\[
\mu_n^L = 1430 \frac{cm^2}{Vs} \cdot \left( \frac{T}{300K} \right)^{-2.18}, \quad \mu_p^L = 460 \frac{cm^2}{Vs} \cdot \left( \frac{T}{300K} \right)^{-2}
\]

(11)

The expressions (11) fit well experimental data of [3], [10] and [27].

The next effect to be considered is ionized impurity scattering. The best established procedure for this task is to take the functional form (12) of the fit provided by Cauhey and Thomas [11] and use temperature dependent coefficients.

\[
\mu_{n,p}^{LI} = \mu_{n,p}^{\min} + \frac{\mu_{n,p}^{\min} - \mu_{n,p}^{\min}}{1 + \frac{CI}{\mu_{n,p}^{\min}}}
\]

(12)

\[
\mu_{n,p}^{\min} = \begin{cases} 
80 \frac{cm^2}{Vs} \cdot \left( \frac{T}{300K} \right)^{-0.45} & T \geq 200K \\
80 \frac{cm^2}{Vs} \cdot \left( \frac{200K}{300K} \right)^{-0.45} \cdot \left( \frac{T}{200K} \right)^{-0.15} & T < 200K
\end{cases}
\]

(13)

\[
\mu_{n,p}^{\min} = \begin{cases} 
45 \frac{cm^2}{Vs} \cdot \left( \frac{T}{300K} \right)^{-0.45} & T \geq 200K \\
45 \frac{cm^2}{Vs} \cdot \left( \frac{200K}{300K} \right)^{-0.45} \cdot \left( \frac{T}{200K} \right)^{-0.15} & T < 200K
\end{cases}
\]

(14)

\[
C_n^{ref} = 1.12 \cdot 10^{17} cm^{-3} \cdot \left( \frac{T}{300K} \right)^{3.2}, \quad C_p^{ref} = 2.23 \cdot 10^{17} cm^{-3} \cdot \left( \frac{T}{300K} \right)^{3.2}
\]

(15)

\[
\alpha_{n,p} = 0.72 \cdot \left( \frac{T}{300K} \right)^{0.085}
\]

(16)

The fits (13)-(16) are from [22]. Similar data have been provided in [6] and [16].

In view of partial ionization one should consider neutral impurity scattering [41]. However, in view of the uncertainty of the quantitative values for ionized impurity scattering it seems not to be worthwhile to introduce another scattering mechanism with additional fitting parameters. Furthermore, partial ionization appears to be a second order effect even at liquid nitrogen temperature. It seems therefore justified to include partial ionization only in the space charge model and not in the carrier mobilities.

Particular emphasis has to be put on surface scattering which we model with an expression suggested by Searvey [40].

\[
\mu_{n,p}^{SURF} = \mu_{n,p}^{ref} + \left( \mu_{n,p}^{ref} - \mu_{n,p}^{ref} \right) \cdot \left( 1 - F(y) \right)
\]

(17)

\[
F(y) = \frac{1 + \exp \left( -\frac{y}{\gamma^{ref}} \right)}{1 + \exp \left( -\frac{y}{\gamma^{ref}} \right)}
\]

(18)

\( S_n = \max (0, -\frac{\partial \phi}{\partial y}), \quad S_p = \max (0, -\frac{\partial \phi}{\partial y}) \)

\( S_n^{surf} \) is assumed to be \( 7 \cdot 10^5 \frac{V}{cm} \); \( S_p^{surf} \) is \( 2.7 \cdot 10^5 \frac{V}{cm} \) and \( y^{ref} \) is 10nm.

The formulae for surface scattering are definitely not the ultimate expressions. They just fit quite reasonably experimental observations. Other approaches with the same claim can be found in, e.g., [6], [24], [31], [36]. A u-shaped mobility behavior as found in [8] has not been synthesized because we believe in a different origin than surface scattering for this experimental observation.

**Modeling Carrier Temperatures**

To describe carrier heating properly one has to account for local carrier temperatures \( T_{n,p} \) in the current relations (4) and (5) [20], [38], [47]. This can be achieved by either solving energy conservation equations self consistently with the basic transport equations [7], [30], or by using a model obtained by series expansions of the solution to the energy conservation equations [21]. We believe that the latter generally is sufficient for silicon devices although energy transport simulation will gain particular relevance for bipolar devices with shallow emitters [33].

\[
U_{n,p} = \frac{k \cdot T_{n,p}}{q} = U_t + \frac{2}{3} \tau_{n,p}^{ref} \left( \frac{\mu_{n,p}^{ref}}{\mu_{n,p}^{ref}} \right)^{3/2} \left( \frac{1 - \tau_{n,p}^{ref}}{1 - \tau_{n,p}^{ref}} \right)
\]

(19)
For the electronic voltages we have (19) as an approximation. Confirming theoretical investigations can be found in [1].

The energy relaxation times $\tau_{np}$ are in the order of 0,5 picoseconds and just weakly temperature dependent [9]. They should however be modeled as functions of the local doping concentration as motivated by the following reasoning. The product of carrier mobility times electronic voltage which symbolizes a diffusion coefficient must be a decreasing function with increasing carrier voltage (see also [9]). Its maximum is attained at thermal equilibrium. Relation (20) must therefore hold.

$$\mu_{np}^{L} \cdot U_{tn,p} \leq \mu_{np}^{L} \cdot U_{to} \tag{20}$$

Note that models for carrier diffusion coefficients are not required in the basic current relations $(4), (5)$. Substituting (19) into (20) and rearranging terms one obtains relation (21) for the energy relaxation times.

$$\tau_{np}^{f} \leq \frac{3}{2} \cdot U_{to} \cdot \frac{\mu_{np}^{L}}{(v_{np}^{L})^{2}} \tag{21}$$

In MINIMOS 5 [45] the energy relaxation times are modeled on the basis of (21) with a fudge factor $\gamma$ in the range [0,1] and a default value of 0.8.

$$\tau_{np}^{f} = \gamma \cdot \frac{3}{2} \cdot U_{to} \cdot \frac{\mu_{np}^{L}}{(v_{np}^{L})^{2}} \tag{22}$$

For vanishing doping one obtains the maximal energy relaxation times which are at 300K $\tau_{n}^{f} = 4.44 \cdot 10^{-13}$s, $\tau_{p}^{f} = 2.24 \cdot 10^{-13}$s and at 77K $\tau_{n}^{f} = 8.62 \cdot 10^{-13}$s, $\tau_{p}^{f} = 8.68 \cdot 10^{-13}$s.

Modeling Carrier Generation/Recombination

An adequate model for thermal generation/recombination even for low temperature can be found in [41]. A comment should be made on the model for the impact ionization rate which has to be supplied for the continuity equations (2) and (3). It still seems, though under heavy dispute in the scientific community, that the old Chynoweth formulation (23) of impact ionization can be used quite satisfactorily for device simulation.

$$R^{II} = -\alpha_{n,p} \cdot \frac{|J_{n}|}{q} - \alpha_{p} \cdot \frac{|J_{p}|}{q} \tag{23}$$

with:

$$\alpha_{n,p} = \alpha_{n,p}^{\infty} \cdot \exp \left( \frac{-E_{n,p}}{E} \right) \tag{24}$$

The coefficients of (24) can be modeled temperature dependent by (25) and (26) to fit experimental data [14], [15], [32]. It should be noted that there is some lack of data for liquid nitrogen temperature, cf. [43]. However it seems that this impact ionization model is probably somewhat too pessimistic for a proper quantitative prediction of substrate currents as already stated in [28], [42].

$$\alpha_{n}^{\infty} = 7 \cdot 10^{6} \cdot \left( \frac{T}{300K} \right)^{2} \tag{25}$$

$$\alpha_{p}^{\infty} = 1.58 \cdot 10^{6} \cdot \left( \frac{T}{300K} \right)^{2} \tag{26}$$

The Auger coefficients for the model of Auger recombination (27) can also be made weakly temperature dependent with (28). The fit has been made to the data of [18].

$$R^{AU} = (C_{cm} \cdot n + C_{cp} \cdot p) \cdot (n \cdot p - \tau_{i}^{2}) \tag{27}$$

$$C_{cm} = 2.8 \cdot 10^{-31} \cdot \frac{cm^{6}}{s} \cdot \left( \frac{T}{300K} \right)^{0.14} \tag{28}$$

$$C_{cp} = 9.9 \cdot 10^{-33} \cdot \frac{cm^{6}}{s} \cdot \left( \frac{T}{300K} \right)^{0.2}$$

NUMERICAL ASPECTS

The numerical solution of the semiconductor equations, i.e. discretization, linearization, and solving the resulting algebraic equation system, can not be discussed here in detail. The interested reader is referred to [2], [45]. Recent investigations of numerical algorithms can be found in [41]. The use of more general transport equations require subtle modification to the numerics, an example of which can be given in [19].

A Glimpse of Results

This chapter presents three-dimensional effects of MOSFET's due to the nonplanar nature of the field-oxide body. The investigations have been carried out by MINIMOS 5 [45] which accounts for all three spatial dimensions. Three-dimensional effects like threshold shift for small channel devices, channel narrowing and the enhanced conductivity at the channel edge have been successfully modeled. Similar investigations leading to matching results have been performed in [2], [25].

The geometry of the investigated 3-D MOSFET is given in Fig. 1: an n-channel with a 1µm x 1µm channel and gate oxide of 15ns. The oxide body of the analyzed structure can be seen in Fig. 2 (note that the oxide is between the upper and the lower plane).

In order to demonstrate the effects at the channel edge we select two different bias points. The first is near threshold with $U_{S} = U_{D} = 0.0V, U_{DS} = 1.0V, U_{GS} = 0.5V$ (the threshold voltage for this device is $U_{th} \sim 0.75V$). The potential distribution in channel length and width direction at the semiconductor/gate-oxide interface is shown in Fig. 3. (This plane penetrates into the field-oxide near the contact boundary of source and drain.) The corresponding minority carrier distribution is given in Fig. 4. A remarkable depletion region at the drain side causes the channel charge to be smaller (under certain bias conditions) than predicted by 2-D simulations.

The second bias point is far above threshold $U_{p} = U_{p} = 0.0V, U_{GS} = 1.0V, U_{GF} = 3.0V$. The corresponding potential distribution can be seen in Fig. 5. The location of the plane which the distribution is drawn for, is the same as at the previous bias condition. The high increase of the potential distribution out of the channel is due to the gate contact overlapping the
Fig.1: Investigated 3-D MOSFET structure

Fig.2: Oxide body of the structure

Fig.3: Detailed view of the surface potential at bias $U_{GS} = 0.5V$

Fig.4: Detailed view of the minority carrier density at bias $U_{GS} = 0.5V$

Fig.5: Detailed view of the surface potential at bias $U_{GS} = 3.0V$

Fig.6: Detailed view of the minority carrier density at bias $U_{GS} = 3.0V$
field-oxide. Also interesting is the minority carrier distribution (Fig. 6) which shows the enhanced conductivity at the semiconductor field-oxide interface. Note that only one half of the channel width is shown in Fig. 3 - Fig. 6; -0.5μm denotes the middle of the channel width and 0.0μm the boundary of source and drain contacts. The consequence on the device characteristics of these effects depends on the gradient of the "bird's beak" and the channel width. A high gradient in the field-oxide shape results in high parasitic current at the channel edge; this effect is less significant for low gradients. Narrow channel devices with high gradient have much higher currents than predicted by 2-D calculations while the agreement with 2-D simulations is good for wide channel devices in any case. Using a low gradient in bird's beak yields a very smooth potential distribution compared to a nearly rectangular shape.

Fig. 7 shows a comparison of two-dimensional and three-dimensional characteristics for $U_{DS} = 1$V and a rectangular approximated field-oxide.

The dependence of the drain current of n-channel devices with weak field implantation on the channel width is shown in Fig. 8. The marked points indicate the measured transistors which have been investigated at the same bias conditions where the enhanced conductivity can be seen.

ACKNOWLEDGMENT

This work is considerably supported by the research laboratories of SIEMENS AG at Munich, FRG, the research laboratories of DIGITAL EQUIPMENT CORPORATION at Hudson, USA, and the "Fond zur Förderung der wissenschaftlichen Forschung" under contract S$43/10.$ We are indebted to Prof. H. Pötal for many critical and stimulating discussions.

REFERENCES
