III-V-MOS
Technology CAD for III-V Semiconductor-based MOSFETs

Project number: 619326
Start date: 01/11/2013
Call (part) identifier: FP7-ICT-2013-11
Funding scheme: Collaborative project

D4.4
Report on mobility models for TCAD

Nature: Report
Dissemination Level: Public (PU)
Due date: M24
Actual submission date: 11/02/2016
Participants involved: IUNET, SYNOPSYS
Authors: A. Gnudi (IUNET-BO), E. Gnani (IUNET-BO), S. Reggiani (IUNET-BO), S. Carapezzi (IUNET-BO), E. Caruso (IUNET-UD), A. Erlebach (SYNOPSYS)
Edited by: A. Gnudi (IUNET-BO)
Executive summary ................................................................................................................ 3

1. Low-field mobility model for InGaAs-on-InP MOSFETs ........................................... 4
  1.1. Introduction ............................................................................................................. 4
  1.2. Analytical mobility model ...................................................................................... 4
  1.3. Mobility model calibration and validation ............................................................ 5

2. Quasi-ballistic corrections to the drift-diffusion low-field mobility models ............. 7
  2.1. Introduction ............................................................................................................. 7
  2.2. Model structure ....................................................................................................... 7
  2.3. Model implementation in Sentaurus ................................................................. 10
  2.4. Model verification and results ............................................................................. 11

Conclusions & Outlook ........................................................................................................ 17
Executive summary

This document reports on the compact mobility models suitable for TCAD which have been devised in subtasks 4.4.2 and 4.4.3. In particular, in subtask 4.4.3 a low-longitudinal field (low $V_{DS}$) electron mobility model suitable for thin-body In$_{0.53}$Ga$_{0.47}$As-on-InP MOSFETs has been set up, calibrated, and implemented through PMI (Physical Model Interface) within Sentaurus Device. The model is empirical but physically based and accounts for the main scattering sources occurring in such devices, namely Coulomb centers, phonons and surface roughness. The thin-body effective thickness, which impacts the phonon scattering and is modulated by the transverse (vertical) electric field at varying $V_{GS}$ through quantum effects, is also taken into account. The model has been calibrated on the experimental data collected on In$_{0.53}$Ga$_{0.47}$As-on-InP MOSFETs fabricated at IMEC [Alian2013, see also deliverable D3.2] with body thicknesses as low as 5 nm. The simulations carried out with the calibrated model through PMI implementation accurately reproduce the $I_{DS}$–$V_{GS}$ experimental curves at $V_{DS} = 50$ mV.

In subtask 4.4.2 a physically sound model to account for quasi-ballistic effects typical of short channel devices within a standard drift-diffusion approach has been developed. In particular, as in subtask 4.4.3, the focus has been on the low longitudinal field (low-$V_{DS}$) problem, with the intent of defining a suitable procedural model able to correct the standard mobility models used in TCAD tools, thus extending their validity to short channels. Indeed, it is well known that in short channels at low $V_{DS}$ the drift-diffusion model tends to overestimate the current, due to its inherent inability of taking into account the finite injection velocity from the contact reservoirs. The proposed correction model is based on the calculation of a “ballistic” mobility term, which is combined locally with the long channel “scattering” mobility through a Mathiessen-like rule, effectively introducing an additional current-limiting factor. The procedure has been implemented using Sdevice PMI tool and additional MATLAB and shell scripts, in a form suitable for DG-UTB FETs. A validation phase has been carried out, using the 15-nm DG-UTB template FETs of WP5 and the IUNET-Udine MSMC simulator as a reference tool. Devices with two channel lengths have been considered, namely 15 and 30 nm. The results show that the procedure is able to provide currents in good agreement with MSMC for the 30-nm case, while some differences emerge in the 15-nm FET. The reasons and implications of these discrepancies are discussed.

The proposed models of both subtasks will be tested and reviewed using the TCAD simulation setups to be developed in the project for future use, and serve as basis for additional developments.
1. Low-field mobility model for InGaAs-on-InP MOSFETs

1.1. Introduction

This section reports on the low-longitudinal field electron mobility model developed by IUNET in subtask 4.4.3 and which is suitable for thin-body InGaAs-on-InP MOSFETs. The model is physically based and accounts in a semi-empirical way for the main scattering sources occurring in such structures, namely Coulomb centers, phonons and surface roughness. The thin-body effective thickness, which is modulated by the transverse (vertical) electric field through quantum effects, is also taken into account. The model is calibrated on the experimental data collected on In0.53Ga0.47As-on-InP MOSFETs fabricated at IMEC [Alian2013, D3.2] and body thicknesses as low as 5 nm have been considered. It is shown that the simulations carried out with the calibrated model accurately reproduce the $I_{DS}-V_{GS}$ experimental curves at low $V_{DS}$. This activity led to a joint IUNET-IMEC publication [Betti2015], where more details can be found.

1.2. Analytical mobility model

A model is proposed in which each scattering mechanism (phonons, Coulomb centers, and surface roughness) is related to a specific mobility term, the different terms being combined through Mathiessen’s rule. Compact analytical formulas are used for the various terms, which depend on the electric field $F(x)$ and on the electron concentration $n(x)$ pointwise, in order to provide a “local” mobility model, an important feature for TCAD simulators. Besides, a direct dependence on the InGaAs body thickness $T_B$ is accounted for.

**Coulomb-limited mobility**

This contribution is modelled as

$$\mu_C = \frac{q \tau_C}{m^*}, \quad \tau_C = \tau_{C0} \left( \frac{n}{n_0} \right)^\gamma$$

where $m^* = m$, $m_0$ is the electron effective mass, $\tau_{C0}$ and $\gamma$ are fitting constants, and $n_0$ is a normalization concentration. The dependence on $n$ accounts for the electrostatic screening due to the electron population. It should be noted that the effective mass depends on the body thickness $T_B$ due to quantization and band non-parabolicity effects. This dependence is taken from the band structure calculations of WP2 [Caruso2015, Zerveas2016].

**Surface roughness-limited mobility**

It is modelled as

$$\mu_{SR} = \frac{q \tau_{SR}}{m^*}, \quad \tau_{SR} = \frac{\tau_{SR0}}{m_r} \left( \frac{F}{F_0} \right)^{-\delta}$$

where $\tau_{SR0}$ and $\delta$ are fitting parameters and $F_0$ is a normalization field.

**Phonon-limited mobility**

The surface and bulk phonon-limited mobilities are empirically modelled as
respectively, where \( \tau_{SPH0} \), \( \tau_{BPH0} \) and \( \eta \) are fitting parameters and \( T_{B0} \) is a normalization thickness. The form of the adopted expressions was previously proved effective for Si thin-body FETs [Reggiani2007]. The expected validity range is at least from near-threshold to well above threshold (see calibration in the next section). The “local” effective body thickness \( T_{B, eff,L} \) is a function of the local electric field and accounts for the effective thickness (inverse form factor) modulation due to quantization. It is computed according to the method initially proposed in [Reggiani2007] for SOI and double-gate Si MOSFETs. In essence, first the “global” effective thickness \( T_{B, eff} \) of the semiconductor layer is computed by means of a preliminary numerical solution of the coupled 1D Poisson-Schrödinger equations on a vertical cross-section of the channel for different values of \( V_{GS} \) and \( T_B \). It is defined as

\[
T_{B, eff} = \frac{\sum_i \left[N_{S,i} \left( \sum_{j \leq i} I_{FF,i,j} \right)^{-1} \right]}{\sum_i N_{S,i}}
\]

\[
I_{FF,i,j} = \int_0^T |\Psi_i(x)|^2 |\Psi_j(x)|^2 dx
\]

where the summations extend over all subbands (higher subband indexes correspond to higher subband bottom energies), \( N_{S,i} \) is the electron population of the \( i \)-th subband, \( I_{FF,i,j} \) defined in (6) is the inverse form factor between the \( i \)-th and the \( j \)-th subbands, and \( \Psi \) is the \( i \)-th wave function. Then, \( T_{B, eff} \) is plotted as a function of the effective (\( n \)-weighted average) electric field as in Fig. 1.1 and modelled through a suitable analytical expression, whose details are reported in [Betti2015], as a function of the effective field and \( T_B \). Finally, in such expression the effective field is replaced with the “local” field \( F(x) \), in order to obtain the effective body thickness \( T_{B, eff,L} \) used in (4).

### 1.3. Mobility model calibration and validation

The calibration is obtained through comparison with the experimental effective mobility \( \mu_{\text{eff, EXP}} \) extracted from the current-voltage characteristics of IMEC long-channel (\( L = 10 \mu m \)) devices at \( V_{DS} = 50 \text{ mV through} \)

\[
\mu_{\text{eff, EXP}} = \frac{I_{DL}}{q (V_{DS} - R_S I_D) W N_S}
\]

where \( R_S \) is the total parasitic series resistance measured at IMEC. In (7) \( N_S \) is the free electron density per unit area in the channel, which cannot be directly obtained from C-V measurements due to the presence of a significant amount of interface and oxide trapped charge. Therefore, \( N_S \) has been calculated through dedicated Poisson-Schrödinger simulations after a preliminary calibration of the trap density on the measured C-V curves for each body thickness [see deliverable D3.2]. Fig. 1.2 shows the result of this calibration.
procedure for the 10-nm case. Similar results have been obtained for the other thicknesses. A trap density located at the InGaAs-Al2O3 interface has been assumed and it is composed of a fixed charge contribution \( N_F \) plus an energy-exponential distribution \( D_{\text{IT}}(E) \) in the conduction band, the latter mimicking the presence of border traps in the oxide. The parameters characterizing \( D_{\text{IT}}(E) \) and \( N_F \) have been adjusted to best fit the C-V curves for each thickness. Fig. 1.3 reports \( D_{\text{IT}}(E) \) for three body thicknesses, compared with experimental distributions from the literature. The agreement is satisfactory except for the thinnest sample (5 nm), for which a lower \( D_{\text{IT}}(E) \) is required to achieve a good fit of C-V data.

**Fig. 1.1.** Effective thickness (eq.(5)) vs. effective field for different body-thickness values \( t_B \). Symbols: 1D Schrödinger-Poisson simulations. Lines: analytical fitting. In the simulations the quantization mass for the InGaAs layer has been taken from WP2 band-structure calculations [Caruso2015, Zerveas2016].

**Fig. 1.2.** \( C_{\text{G}}-V_{\text{GS}} \) characteristics for a \( t_B = 10 \) nm MOSFET. Dots: measured data. Solid lines: simulations through Schrödinger (green) and DG (red) model with fitted trap density (see Fig. 1.3). Dashed line: Schrödinger simulation without traps.

**Fig. 1.3.** Energy distribution of interface traps representing the acceptor border traps in the conduction band. Lines: exponential model fitted on \( C_{\text{G}}-V_{\text{GS}} \) data. Symbols: data from literature.

**Fig. 1.4.** Effective electron mobility vs. free electron density per unit area for different body-thicknesses. Symbols: experimental data. Lines: model.

The final calibrated effective mobility curves are reported in Fig. 1.4 as a function of the channel electron density for three \( t_B \) values, together with the experimentally extracted data.
It must be noticed that the same set of parameters is used for all model curves (the full list is found in [Betti2015]). The agreement is good.

The single mobility contributions for the $T_B = 15$ nm case are shown in Fig. 1.5: surface roughness is responsible for the mobility decrease at high electron densities, while screened Coulomb scattering determines the positive slope at low densities. A donor doping concentration of $10^{19}$ cm$^{-3}$ is assumed in the channel.

As a final check of the proposed model validity, 2D simulations of the measured MOSFETs have been performed with Synopsys Sentaurus with the new mobility model implemented through the PMI tool. The resulting turn-on characteristics are shown in Fig. 1.6. The good agreement confirms the consistency of the approach, the small differences being attributable to the definition of the extracted series resistances.

![Fig. 1.5. Mobility contributions to the total mobility vs. free electron density per unit area for $T_B = 15$ nm. The represented mobilities are “local” and are taken at the point corresponding to the maximum electron density $n$ for each $N_S$.](image)

![Fig. 1.6. Simulated (lines) and measured (symbols) turn-on characteristics of the IMEC devices with $L = 10$ μm at $V_{DS} = 50$ mV for different $T_B$ values. Simulations are based on the proposed mobility model.](image)

## 2. Quasi-ballistic corrections to the drift-diffusion low-field mobility models

### 2.1. Introduction

This section describes the development, relative to subtask 4.4.2, of a physically sound model to account, within a standard drift-diffusion approach, for quasi-ballistic effects, which occur in short channel devices. In particular, as for the model described in the previous section, the focus is on the long longitudinal field problem, with the intent of defining a suitable procedure to apply corrections to the standard low-field mobility models used in TCAD tools, thus extending their validity to short channel transistors. Indeed, it is known that the drift-diffusion model tends to overestimate the current in short channel FETs at low $V_{DS}$, due to its inability of taking into account the finite injection current from the contact reservoirs.

### 2.2. Model structure

The proposed procedure relies on the theoretical background established in [Gnani2010], which in turn was based on the semi-analytical solution of the Boltzmann transport equation
(BTE) in a nanowire (NW) silicon FET, accounting for quantization in the transverse direction (formation of subbands) [Gnani2008]. In [Gnani2010] it was analytically shown that at low $V_{DS}$ the effective mobility $\mu_{\text{eff}}$, defined through

$$I_D = \mu_{\text{eff}}Q_{\text{inv}}(x_m) \frac{V_{DS}}{L_g} \tag{8}$$

where $Q_{\text{inv}}(x_m)$ is the electron charge per unit length at the top of the barrier coordinate $x_m$, can be obtained from

$$\frac{1}{\mu_{\text{eff}}} = \frac{1}{\mu_{\text{bal}}} + \frac{L}{L_g \mu_{\text{sc}}} \tag{9}$$

where $L$ is the total device length (including source/drain regions), $L_g$ the gate length, $\mu_{\text{bal}}$ the “scattering” mobility, i.e. the $\mu_{\text{eff}}$ value in the long-channel limit, and $\mu_{\text{sc}}$ the “ballistic” mobility calculated as in (8), but with all scattering mechanisms being suppressed. Eq. (9) can be interpreted as a sort of Matthiessen’s rule, slightly modified by the presence of the $L / L_g$ ratio, which takes into account the impact of the source/drain regions on the value of the drain current.

Based on the above theoretical result, the proposed procedure for the computation of the mobility including quasi-ballistic corrections at low $V_{DS}$ consists of the following steps (the procedure outlined here could be applied equally well to NW and ultra-thin-body (UTB) FETs, but the details and formulae reported hereafter are specific for a DG UTB FET):

1) Determination of the bottom energy profiles $E_{o,n}(x)$ of the subbands along the UTB longitudinal coordinate $x$. This step requires in principle the solution of the Schrödinger equation in each channel cross-section. It can be simplified though, in the spirit of a TCAD approach, through an approximate first-order perturbation solution, where in each cross-section the UTB quantum well is treated as the zero-order unperturbed Hamiltonian, and the potential energy as the first-order perturbation. Taking also into account the band non-parabolicity, the energy dispersion relation of the $n$-th subband in each cross-section is expressed as [Jin2007]

$$E_n(x; k_x, k_y) = <w_n(z)|E_C(x, z)|w_n(z)> + \frac{-1 + \sqrt{1 + 4\alpha \frac{\hbar^2}{2m} \left[ \frac{n^2 \pi^2}{t_D^2} + (k_x^2 + k_y^2) \right]}}{2\alpha} \tag{10}$$

where $z$ is the vertical (quantization) coordinate, $w_n(z)$ the $n$-th unperturbed eigenfunction of the quantum well, $E_C(x, z)$ the conduction-band bottom energy within the UTB (perturbation potential), $\alpha$ the non-parabolicity factor, $m$ the isotropic effective mass of the $\Gamma$ valley, and $t_D$ the UTB thickness. $E_C(x, z)$ is obtained by means of a TCAD simulation with quantum corrections, such as density gradient (DG) or, more suitably, modified local density approximation (MLDA) currently implemented in Synopsis SDevice tool. The $n$-th subband profile $E_{o,n}(x)$ is then obtained by setting $k_x = k_y = 0$ in (10).
2) Once the subband profiles are obtained, the ballistic current for the \( n \)-th subband is computed by considering the \((k_x, k_y)\)-state occupation of the \( n \)-th subband at the coordinate \( x_{M,n} \) corresponding to the top value of \( E_{o,n}(x) \). The total ballistic current is written as

\[
I_{bal} = \sum_n \sum_{k_y,k_x} 2q v_{x,n}(k_x,k_y) \left[ \Theta(k_x) F_{S,n}(k_x, k_y) + \Theta(-k_x) F_{D,n}(k_x, k_y) \right]
\]  

(11)

where \( v_{x,n}(k_x,k_y) \) is the \( x \)-component of the group velocity of the \( n \)-th subband, \( \Theta(\cdot) \) is the unit step function, and \( F_{S/D,n}(k_x, k_y) = f_{S/D}[E_n(x_{M,n}; k_x, k_y)] \) with

\[
f_{S/D}(E) = \frac{1}{\exp \left( \frac{E - E_{F,S/D}}{k_B T} \right) + 1}
\]

(12)

the electron Fermi energy-distribution with Fermi level given by the source/drain reservoir and \( x_{M,n} \) the abscissa at the top of the barrier of the \( n \)-th subband. Converting the summation over \((k_x, k_y)\) into an integral and using (10), equation (11) is expressed through an integration over the kinetic energy \( \epsilon \)

\[
I_{bal} = \frac{q \sqrt{2m}}{\pi^2 h^2} \sum_n \int_0^\infty \left[ f_S(\epsilon + E_{o,n}(x_{M,n})) - f_D(\epsilon + E_{o,n}(x_{M,n})) \right] \sqrt{\epsilon} \left[ 1 + 2\alpha \epsilon_{QW,n} + \alpha \epsilon \right] d\epsilon
\]

(13)

where

\[
\epsilon_{QW,n} = -1 + \sqrt{1 + \frac{4\alpha \hbar^2 n^2 \pi^2}{2m \frac{t_b^2}{\hbar^2}}}
\]

(14)

is the \( n \)-th energy level of the quantum well, which includes the non-parabolic correction. In (13) the integral is evaluated numerically and in practice \( x_{M,n} = x_{M,1} \) is used for all \( n \), i.e. the location of the top-of-the-barrier of the lowest subband is used for all other subbands. This is a good approximation as exemplified, for instance, in Fig.7.15 of [EPS2011].

It should be noticed that the expression of the ballistic current does not require the value of the mass and non-parabolicity factor for each subband, but only the bulk values \( m \) and \( \alpha \), and is therefore fully consistent with the Sentaurus MLDA approach.

Even if not used in the mobility model, we report hereafter also the expression of the electron density integrated over the thickness of the UTB at the location \( x_{M,n} = x_{M,1} \) in the ballistic case, which is calculated starting from

\[
N_T = \sum_n \sum_{k_y,k_x} 2 \left[ \Theta(k_x) F_{S,n}(k_x, k_y) + \Theta(-k_x) F_{D,n}(k_x, k_y) \right].
\]

(15)

Following a procedure similar to the one outlined above for the current, one obtains

\[
N_T = \frac{m}{2\pi h^2} \sum_n \int_0^\infty \left[ f_S(\epsilon + E_{o,n}(x_{M,o})) + f_D(\epsilon + E_{o,n}(x_{M,o})) \right] \left[ 1 + 2\alpha (\epsilon_{QW,n} + \epsilon) \right] d\epsilon.
\]

(16)

Expression (16) of \( N_T \) will be used in the discussion of results in the next section.
3) Once $I_{bal}$ is computed according to (13), the ballistic mobility is calculated as

$$\mu_{bal}(x) = \mu_{bal} \frac{I_{bal}}{q \int_0^{\epsilon_B} n(x, z) \left| \frac{\partial \phi_n(x, z)}{\partial x} \right| dz},$$  

where the electron quasi-Fermi potential $\phi_n$ and concentration $n$ are obtained again from DG or preferably MLDA simulations, and the integral is extended over the UTB thickness.

4) Finally, $\mu_{eff}$ is computed in each point using (9), combining the ballistic and the scattering mobilities. It should be noticed that, according to the definition (17), $\mu_{bal}$ is constant in each cross-section (i.e. it is z-independent), but is in principle a function of the longitudinal coordinate $x$. However, at low $V_{DS}$ the $x$-dependence is quite soft, due to the quasi-uniform channel conditions. Moreover, it is difficult, or even impossible, to implement the full $x$-dependent expression (17) in Sentaurus using only PMI and external scripts. Therefore, at least for this first experimental version, a constant $\mu_{bal}$ is used everywhere inside the semiconductor equal to $\mu_{bal}(x_{M,1})$. On the other hand, for $\mu_{sc}$ any local doping- and/or normal-field-dependent model is allowed, which is judged suitable for InGaAs UTB FETs.

2.3. Model implementation in Sentaurus

The illustrated procedure for UTB DG FETs has been implemented in Synopsys Sentaurus, for proof-of-concept and validation purposes, using PMI tool and auxiliary scripts. Some implementation details are reported hereafter.

The full procedure is handled by a bash shell script and is schematically depicted in Fig. 2.1. For each bias point, an initial drift-diffusion simulation with MLDA provides the first-guess values of $E_n(x, z, n(x, z)$ and $\partial \phi_n(x, z)/\partial x$. ASVisual TCL script exports such values, which are used to compute $E_{o,n}(x)$ through (10), the top of the barrier location $x_M$ as well as the ballistic current by means of (13-14), and the ballistic mobility (17) by using a MATLAB code. The value of the ballistic mobility is then written into the parameter file, and a new MLDA simulation is started. The simulation command file contains the keyword

```bash
emobility(Enormal(Unibo) DopingDependence(Quasi_ballistic_mobility))
```

in the Physics section, which uses the newly defined PMI mobility model “Quasi_ballistic_mobility” to compute the “bulk” mobility with ballistic corrections according to (9), while the normal field dependence is accounted for through the Unibo model with modified parameters, in order to handle the InGaAs UTB device as described in the next section. The simulation results provide the updated values of $E_C(x, z, n(x, z)$ and $\partial \phi_n(x, z)/\partial x$ and the entire procedure is iterated until convergence is achieved. The convergence is checked by monitoring the behaviour of $\mu_{bal}$ itself across successive iterations. The adopted convergence criterion is $|\mu_{bal}^i - \mu_{bal}^{i-1}| < \mu_{err}$, with $\mu_{err}$ typically set to 0.1 cm$^2$/Vs. Convergence is obtained in a few iterations (2-3): indeed, if the mobility were constant throughout the device (as is the case for a constant $\mu_{sc}$ and hence a constant $\mu_{eff}$), all the internal quantities $E_C(x, z, n(x, z)$ and $\partial \phi_n(x, z)/\partial x$ would not change with the iterations, and so neither $I_{bal}$ nor $\mu_{bal}$, and convergence would thus be achieved at the first iteration. In the general non-constant mobility case convergence remains fast anyhow.

It should be noticed that since the $L/L_g$ factor in (9) cannot be handled automatically through PMI, the parameters of the chosen mobility model $\mu_{sc}$ are adjusted directly in the parameter file prior to starting the procedure, according to the total device and gate length.
2.4. Model verification and results

In order to verify the model and the procedure, one of the two template InGaAs DG-UTB FETs defined in milestone MS8 has been chosen as benchmark device, and specifically the one corresponding to the gate length of 15 nm. The device parameters are listed in Table 1, and are equal to the ones in MS8, except for the highlighted underlap length, which has been updated to 2 nm according to the recent results of the device optimization activity in WP5 (see also D5.1), motivated by suppression of short channel effects (SS and DIBL).

Table 1: planar, DG-UTB transistor, $L_G=15$ nm

<table>
<thead>
<tr>
<th>Param.</th>
<th>Value</th>
<th>Description</th>
<th>Param.</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_G$</td>
<td>15 nm</td>
<td>Gate length</td>
<td>Oxide</td>
<td>HfO$_2$</td>
<td>Gate oxide</td>
</tr>
<tr>
<td>$L_S\times L_D$</td>
<td>24.8 nm</td>
<td>Source\Drain length</td>
<td>$T_{OX}$</td>
<td>3.84 nm</td>
<td>Oxide thickness</td>
</tr>
<tr>
<td>$T_{ch}$</td>
<td>7 nm</td>
<td>Channel thickness</td>
<td>EOT</td>
<td>0.68 nm</td>
<td>Equivalent oxide thickness</td>
</tr>
<tr>
<td>$L_{und}$</td>
<td>2 nm</td>
<td>Underlap length</td>
<td>$T_{volG}$</td>
<td>3 nm</td>
<td>Volumetric gate thickness</td>
</tr>
<tr>
<td>$S_{dop}\times D_{dop}$</td>
<td>$5\cdot 10^{19}$ cm$^{-3}$</td>
<td>Source\Drain doping</td>
<td>$V_{DD}$</td>
<td>0.63 V</td>
<td>Supply voltage</td>
</tr>
<tr>
<td>$N_{ch dop}$</td>
<td>$1\cdot 10^{17}$ cm$^{-3}$</td>
<td>Channel doping</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 2.1**: Flow chart of the implemented procedure for the quasi-ballistic mobility corrections.
A second DG-UTB FET with a longer channel \((L_G = 30 \, \text{nm})\) is also used for comparison, with parameters as in Table 1, but with no underlap, which is not necessary in this case.

The benchmark simulation tool is the MSMC developed by IUNET-Udine, which includes a comprehensive set of physically-based microscopic models of scattering mechanisms. The MSMC simulations include the \(\Gamma\) and \(L\) valleys, with parameters \(m_f = 0.043, \alpha_f = 1.5 \, \text{eV}^{-1}, \Delta E_{L-\Gamma} = 0.76 \, \text{eV}, m_{L-I} = 1.57, m_{L-L} = 0.23, \alpha_L = 0.5 \, \text{eV}^{-1}\), which are consistent with the band-structure calculations of WP2. The gate workfunction is \(WF = 4.7067 \, \text{eV}\) and the InGaAs affinity is \(\chi_{\text{InGaAs}} = 4.5 \, \text{eV}\). Scattering include polar phonons, remote phonons and interface roughness (linear model with \(\Delta_{\text{rms}} = 1.3 \, \text{nm}\) and \(\lambda = 1.5 \, \text{nm}\)).

The same \(\Gamma\) and \(L\) parameters used in MSMC have been used also in Sdevice in the “Multivalley” section, as well as \(WF\) and \(\chi_{\text{InGaAs}}\). In the external quasi-ballistic correction loop, only the \(\Gamma\) valley has been considered for simplicity in the calculation of the ballistic current (13) (its contribution is always largely dominant), with again the same bulk \(m\) and \(\alpha\) parameters, limiting the number of subbands to \(n = 10\).

The first step is the verification of the consistency of the device electrostatic properties between the two codes. This is an important step, because \(n(x, z)\) appears in the denominator of the expression of the ballistic mobility (17) and so enters the mobility correction loop. Fig. 2.2 reports the electron density integrated over the UTB thickness \(N_f\) vs. \(V_{GS}\) in a long channel device at \(V_{DS} = 0\). The first observation concerns the impact of the wave-function penetration (WFP) into the oxide, which can be accounted for in the MSMC: its effect is sizable and can be approximately quantified in a threshold voltage shift of about 0.1 V. Since the MLDA approach does not account for this effect [Penzin2013], to the purpose of verification and for all other results shown in this document the WFP has been suppressed in MSMC simulations. It is seen from Fig. 2.2 that when WFP is neglected MSMC and TCAD are in reasonable agreement, even if TCAD exhibits a slightly lower slope for \(V_{GS} \leq 0.7 \, \text{V}\), which is the range over which the currents will be compared (see \(V_{DD}\) in Table 1). The figure also reports, for verification purpose only, the result of eq. (16), which lies very close to the MSMC and TCAD curves.
The second step is the calibration of the Sdevice mobility model on the long channel MSMC mobility data, used as reference. The effective scattering mobility, defined as

$$\mu_{s_c, eff} = \frac{\int_{0}^{L} \mu_{sc}(z) n(z) \, dz}{\int_{0}^{L} n(z) \, dz},$$

(18)

and calculated with the MSMC and Sdevice is reported in Fig. 2.3 as a function of $N_T$. Since in Sdevice there is no mobility model suitable for InGaAs UTB channels, it has been decided to adapt the eNormal(Unibo) model, which is flexible enough in terms of functional dependence on the normal electric field and screening effects. The Unibo model parameters have been fitted on the MSMC data within the considered $V_{GS}$ range, corresponding in the figure to the region at the left of the vertical line. It has not been possible to extend the fitting to the outer region, probably due to the functional dependence of the model on the local normal electric field. It is seen however that the agreement within the relevant $V_{GS}$ range is satisfactory.

Fig. 2.4. Turn-on characteristics at $V_{DS} = 5 \text{ mV}$ for the 30-nm benchmark FET in linear (left) and log (right) scale calculated with: MSMC (triangles), Sdevice w/o correction loop (dashed), Sdevice with correction loop neglecting in (9) the $L/L_g$ factor (i.e. $L/L_g = 1$) (red stars) or fully taking it into account ($L/L_g \neq 1$) (green squares), ballistic current $I_{bal}$ (13) (blue squares).

Fig. 2.5. Turn-on characteristics at $V_{DS} = 5 \text{ mV}$ for the 15-nm benchmark FET in linear (left) and log (right) scale. Symbols as in Fig. 2.4.
Once the long channel mobility has been calibrated, the loop procedure for the quasi ballistic correction has been applied to the two benchmark devices with \( L_G = 15 \) nm and \( L_G = 30 \) nm. Fig. 2.4 shows the low \( V_{DS} \) turn-on characteristics for the 30-nm benchmark computed with different methods. First of all it is noticed that the uncorrected TCAD (no loop), i.e. the raw Sdevice with the calibrated Unibo model, calculates a current approximately four times larger than MSMC, which confirms the great impact of ballistic effects for such short channel III-V devices. The ballistic current (blue squares), computed through eq. (13), is in this case comparable with the uncorrected current. The combined mobility eq. (9) brings down the total current (green squares) considerably to values close to the MSMC results. It is interesting to note that if \( L/L_g = 1 \) is set in (9), i.e. if the standard Mathiessen rule is used, the resulting current (red stars) is well above the MSMC current. This confirms the need of the \( L/L_g \) correcting factor, as originally found in [Gnani2010] for Si NW FETs.

Similar considerations can be repeated for the 15-nm currents in Fig. 2.5. Here the ballistic current is, as expected, quite similar to the 30-nm case, while the uncorrected TCAD current is even larger than the ballistic one, which means that the standard TCAD drift-diffusion model is in this case clearly unphysical. The correction loop brings the current down to values similar to the 30-nm case, while the MSMC results are higher for the 15-nm than for the 30-nm case, as expected due to the channel transport being not fully ballistic. This means that the corrected TCAD current is below the MSMC current. A further look at the results reveals that the corrected TCAD current with \( L/L_g = 1 \) follows instead the expected increase going from 30-nm to 15-nm, but overestimates the current. All this can be interpreted as the onset of a criticality of the \( L/L_g \) term, combined with other difficulties related to the reliability of the scattering mobility dependence on the transverse electric field in a very short channel where two-dimensional electrostatic effects play an important role, as well as transport within the source/drain regions, which have a significant impact on the current in this case.

To better compare the different approaches, Figs. 2.6 and 2.7 report the electron density \( N_T(x) \) and average velocity \( v_T(x) \) profiles along the longitudinal coordinate of the 30-nm device for two gate biases, one slightly above threshold (\( V_{GS} = 0.4 \) V) and the other near \( V_{DD} \) (\( V_{GS} = 0.7 \) V). The \( V_{DS} = 20 \) mV value has been chosen here in order to improve the accuracy of the statistical averages in the MSMC, but it has been verified that the device still fully operates in the linear region, i.e. the current is proportional to \( V_{GS} \). The density profiles of Fig. 2.6 show that the agreement with MSMC is excellent in the channel (confirming the results of Fig. 2.2) and also in the high gradient regions, less good in the highly degenerate source/drain regions. The differences next to the source/drain contacts are due to the different treatment of the boundary conditions in the Poisson equation: the MSMC imposes Neumann conditions, Sentaurus ohmic contact conditions. However, such differences are supposed to have little impact on the drain current.

The average electron velocity in Fig. 2.7 is consistent with the currents of Fig. 2.4. At \( V_{GS} = 0.7 \) V the corrected TCAD loop slightly overestimates the velocity in the channel, much more so if \( L/L_g = 1 \) is used. At \( V_{GS} = 0.4 \) V the opposite is true, and the MSMC results lie above the corrected TCAD loop. What is the origin of these differences is difficult to say. One possible reason might be that at low \( V_{GS} \) the channel density is much less uniform in the longitudinal direction than at high \( V_{GS} \) (in Fig. 2.7 at \( V_{GS} = 0.4 \) V the velocity profile shows a sharp peak in the middle of the channel, which corresponds to a minimum of the electron density), therefore the two-dimensional field distribution and the diffusivity play a more important role at low \( V_{GS} \). All such issues would deserve a much deeper future investigation.
As a conclusion of this validation phase, we believe it is fair to say that the correction procedure behaves well for gate lengths of the order of 30 nm, in particular at high gate voltages, when a well-defined channel region exists, less well at 15-nm gate length, where probably a number of assumptions become critical. In both cases the model represents a remarkable and necessary improvement over the state of the art.

Fig. 2.6. Electron density profile per unit area along the longitudinal coordinate of the 30-nm FET at \( V_{DS} = 20 \) mV and \( V_{GS} = 0.4 \) V and 0.7 V, calculated with: MSMC (dashed lines) and TCAD with full correction loop (filled symbols).

Fig. 2.7. Average electron velocity along the longitudinal coordinate of the 30-nm FET at \( V_{DS} = 20 \) mV and \( V_{GS} = 0.4 \) V and 0.7 V computed with: MSMC (dashed lines), TCAD with correction loop with (filled symbols) and w/o (open symbols) taking into account the \( L/L_g \) factor.
References


Conclusions & Outlook

The problem of the definition of compact mobility models for III-V FETs suitable for TCAD implementation has been tackled in this deliverable. The focus has been on low-field (low $V_{DS}$) models. Within this scope, two aspects have been dealt with: 1) electron mobility model for long-channel thin-body InGaAs-on-InP MOSFETs; 2) electron mobility model with quasi-ballistic corrections for short-channel FETs.

In both cases the proposed models have a sound physical basis. Concerning point 1), a complete low-field mobility model, accounting for Coulomb, phonon and surface roughness scattering, as well as the $V_{GS}$-dependent effective body thickness due to quantum effects, has been developed and calibrated on IMEC $C-V$ and $I_{DS}-V_{GS}$ experimental data of devices with body thicknesses as low as 5 nm. The model has been implemented in Sentaurus through PMI and simulations of long-channel MOSFETs have been shown which reproduce well the experimental currents at $V_{DS} = 50$ mV.

Concerning point 2), a procedure for incorporating quasi-ballistic corrections into drift-diffusion simulations at low-$V_{DS}$ has been devised. The procedure is based on the concept of “ballistic” mobility and current, which are calculated through integral expressions involving the conduction-band energy profiles and the electron concentration consistently extracted from Sentaurus simulation results using the MLDA approach. Quantization in the channel due to vertical confinement and electric field, as well as band non-parabolicity effects, are accounted for in the ballistic corrections. It should be noticed that no ad-hoc fitting parameter is used.

The procedure, in a form suitable for planar DG-UTB FETs, has been implemented through Sentaurus PMI tool and external MATLAB and shell scripts. The MSMC simulator of IUNET-UD has been used as reference tool. The benchmark devices have been taken from the 15-nm template DG-UTB FET elaborated in WP5. First of all, the long-channel mobility model has been calibrated on MSMC data, then two benchmark devices have been simulated with 15- and 30-nm gate lengths. The results show that the model provides currents in good agreement with MSMC at least down to 30-nm, while a degradation is seen for the shortest length. In both cases the improvement over the uncorrected drift-diffusion simulations is quite significant.

Both the proposed models can now be tested and reviewed to decide about implementation into the TCAD simulation template setups to be developed in the project, and will serve as basis for future developments.