2D and 3D TCAD Simulation of III-V Channel FETs at the End of Scaling

P. Aguirre, M. Rau, and A. Schenk

Integrated Systems Laboratory, ETH Zurich, Gloriastrasse 35, 8092 Zurich, Switzerland

Email: adelia@iis.ee.ethz.ch

***Abstract*—Quantum drift diffusion corrections and a simple ballistic mobility model are used to simulate $I_{\mathrm{D}}V_{\mathrm{GS}}$-characteristics of scaled 2D and 3D III-V channel FETs. The sub-threshold swing of double-gate ultra-thin-body geometries is extracted for different gate lengths, and the semi-classical results are compared with those from the quantum transport simulator QTx. The ballistic mobility recovers the QTx transfer curves of the gate-all-around nanowire FETs, except the on-currents in the linear regime. It is shown that source-to-drain tunneling sets a limit to scaling at a gate length of about 10 nm.**

***Index Terms*—III-V-MOSFET, Quantum Drift Diffusion, Quantum Transport, DG-UTB FET, GAA Nanowire.**

I. Introduction

The high electron mobility and injection velocity of III-V-compounds-based FETs make them promising candidates to replace n-type strained Si MOSFETs at future technology nodes with gate lengths shorter than 20 nm [1]. The aggressive scaling causes quantum effects which have a critical influence on the device performance. For instance, geometrical quantum confinement in the body with thickness (t_{body}) below 12 nm leads to a shift in the threshold voltage [2]. Moreover, at gate lengths shorter than 20 nm, the potential barrier between source and drain becomes thin enough, so that source-to-drain tunneling (STDT) deteriorates or even determines the sub-threshold swing (*SS*) [3]. Quantum transport (QT) simulators start to find their way into industrial environments, however, they are computational expensive for large and complex 2D and 3D devices [4]. In this paper, it is systematically shown how quantum drift diffusion (QDD) tools [6] in combination with a TCAD-friendly ballistic mobility model [12] can be used to simulate the described quantum effects. A reasonable agreement with QT simulation results is achieved for *SS*, but I_{ON} is still too high in the linear regime.

II. Simulation Approaches

First, 2D simulations of the double-gate ultra-thin-body (DG UTB) transistor shown in Fig. 1 (a) were performed using the quantum transport code QTx [4]. The same geometry with $t_{\mathrm{body}} = 7$ nm was simulated for different L_{G} ranging from 10 nm to 25 nm. Next, 3D simulations using QTx were done for gate-all-around nanowire (GAA NW) FETs. Their design parameters are the same as for the DG UTB FETs. The dimensions are given in Tabel I. The $I_{\mathrm{D}}V_{\mathrm{GS}}$-characteristics simulated with QTx were used as reference to calibrate the QDD simulation setup of the commercial simulator Sentaurus-Device (S-Device) [5]. Effective masses

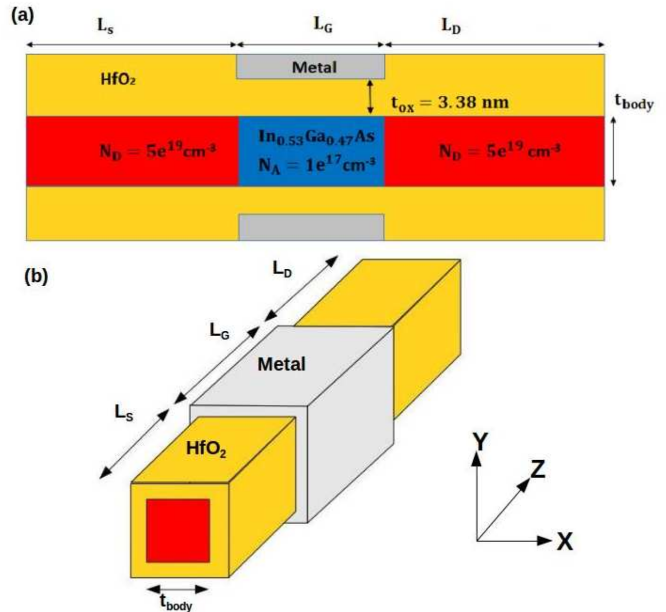

Fig. 1. Schematic of (a) $In_{0.53}Ga_{0.47}As$ double-gate ultra-thin-body (DG UTB) FET and (b) $In_{0.53}Ga_{0.47}As$ gate-all-around nanowire (GAA NW) FET.

(m_{e}) and non-parabolicity parameters were calculated from a full-band version of the QT code.

To simulate the geometrical confinement perpendicular to the transport direction in the 2D and 3D structures two models were used:

(i) *The Density Gradient (DG) Model* which adds a quantum potential (Λ) in the computation of the carrier density (n). This model depends on n, m_{e} and a fitting parameter γ. The latter can be found by a Schrödinger-Poisson solver [4] using the 1D electron density profile along a vertical cut in the middle

TABLE I
Dimensions of the GAA NWs according to future technology nodes as described in [7].

Node	L_{G} [nm]	t_{ox} [nm]	t_{body} [nm]	m_{e}/m_0
A	15	3.75	7	0.0642
B	10.4	3.25	5.5	0.0674

of the device.

In the 3D simulations of the GAA NWs we applied the anisotropic DG model which uses an attenuation matrix with diagonal elements α_l, α_v in order to scale Λ in longitudinal and vertical direction, respectively. The element α_v serves to reproduce the effect of geometrical confinement perpendicular to the transport direction, whereas the α_l-value lowers the height of the energy barrier between source and drain, which mimics the effect of STDT. Since the latter was simulated directly by the Nonlocal Tunneling (NLT) model of S-Device, $\alpha_\mathrm{l} = 0$ was chosen. The fitting parameter γ for both transverse directions (x and y) was calibrated using the 1D density profile in the center of the GAA NW along the y-direction in the middle yz-plane (compare Fig. 1). The QTx electron density profile along the same line was taken as reference. By matching the densities in the sub-threshold regime, the electrostatics (i.e. the threshold voltage) can be fitted very well, *independently of the mobility model used*.

(ii) *The Modified Local Density Approximation (MLDA)*, without any fitting parameter [10].

To include STDT, the Nonlocal Tunneling (NLT) model [6] was used in combination with the MLDA and the anisotropic DG model, respectively. The tunneling mass (m_c) was set to the value of m_e.

In a first instance, to mimic the ballistic QTx case, a constant and artificially high diffusive mobility (μ_d) of 2.3×10^4 $\mathrm{cm^2/Vs}$ was used in the simulations of the DG UTB FETs. In the analysis of the GAA NW FETs we applied a parameter-free ballistic mobility model μ_b [13] in order to improve the agreement with the QTx characteristics. In 1D it has the form

$$\mu_\mathrm{b}(x) = \frac{v_\mathrm{b}(x)}{\psi'_\mathrm{n}(x)} \; , \tag{1}$$

where the gradient of the quasi-Fermi potential (QFP) $\psi'_\mathrm{n}(x)$ is the driving force of the carriers in the channel and $v_\mathrm{b}(x)$ their mean velocity given by

$$\begin{aligned} v_\mathrm{b}(x) &= v_\mathrm{th}\sqrt{\tanh^2\left(\frac{qV_\mathrm{DS}}{2k_\mathrm{B}T}\right) + \frac{2q\psi_\mathrm{n}(x)}{k_\mathrm{B}T}} \\ &\approx v_\mathrm{th}\sqrt{1 + \frac{2q\psi_\mathrm{n}(x)}{k_\mathrm{B}T}} \; . \end{aligned} \tag{2}$$

Here, $\psi_\mathrm{n}(x) = \phi(x) - \frac{k_\mathrm{B}T}{q}\ln(n)$ is the QFP, $v_\mathrm{th} = \sqrt{k_\mathrm{B}T/m_\mathrm{c}}$ the 1D mean thermal velocity, q the elementary charge, and k_B the Boltzmann constant. The second line in Eq. (2) turns out as soon as the source-drain voltage exceeds a few $k_\mathrm{B}T$ which makes μ_b literally parameter-free.

III. Results

Fig. 2 shows the fitted $I_\mathrm{D}V_\mathrm{GS}$-characteristics for the DG UTB FETs for three gate lengths computed with the combination of anisotropic DG model and NLT. Note that the mobility is constant and large here, without any ballistic correction. The *SS* of the transistor with the shortest L_G (where the effect of STDT is strongest) is best reproduced by the anisotropic DG model. Fig. 3 presents the $I_\mathrm{D}V_\mathrm{GS}$-characteristics for the DG UTB FETs using the combination MLDA + NLT. The slopes are similar to the previous case.

Figs. 4 and 5 present the transfer curves in the linear regime (V_DS = 50 mV) for GAA NWs from node A and B, respectively. Figs. 6 and 7 show the corresponding curves in the saturation regime (V_DS = 0.63 V). Again, the combinations DG + NLT and MLDA + NLT with a constant high mobility were used, but now also NLT in combination with the ballistic mobility Eq. (1). The following remarks have to be made: (i) Using the ballistic mobility together with the DG model never converges. Therefore, the electrostatics was corrected by a simple shift of the work function (WF) in this case. (ii) The apparent misfits in the threshold voltages between the DG and the QTx curves in Figs. 4 and 6 are not caused by the electrostatics, but due to the uncorrected constant mobility. (iii) Figs. 5 and 7 contain a real misfit in the threshold voltages between the DG and the QTx curves since the fitted γ-value for t_body = 5.5 nm ($\gamma = 2$) prevented convergence, and $\gamma = 1$ was used instead.

The DG curves in Figs. 5 - 7 exhibit a pronounced bump around the onset of inversion. The strength of this effect increases with decreasing cross section of the GAA NW and increasing source-drain bias. It is attributed to the breakdown of the DG method when the channel quantization goes away and only geometrical confinement remains. The DG model had been developed for the former case, but is unable to cover the latter correctly. As an artifact one can observe that in the case of extremely thin bodies a huge quantum potential is computed even in the flat-band regions of the semiconductor (also when $\alpha_\mathrm{l} = 0$). An empirical workaround to mitigate this effect is to remove the gate oxide in small parts near the source/drain contacts [11] which eliminates the "density gradient" between insulator and semiconductor there. However, this was not done in the present study. Another serious issue is related to the modeling of the density of states (DOS). In S-Device the DOS model is that of a 3D electron gas. The only way to

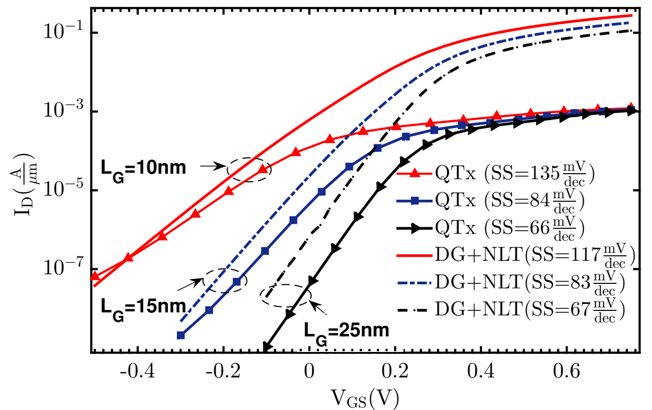

Fig. 2. $I_\mathrm{D}V_\mathrm{GS}$-characteristics obtained from the combination of DG + NLT for a 7 nm DG UTB FET with different gate lengths. Parameters: V_DS = 0.05 V, m_c = 0.0516 m$_0$, μ_d = 2.3×10^4 $\mathrm{cm^2/Vs}$, γ = 0.7, $(\alpha_\mathrm{l}, \alpha_\mathrm{v})$ = (0,1), and WF = 4.8 eV.

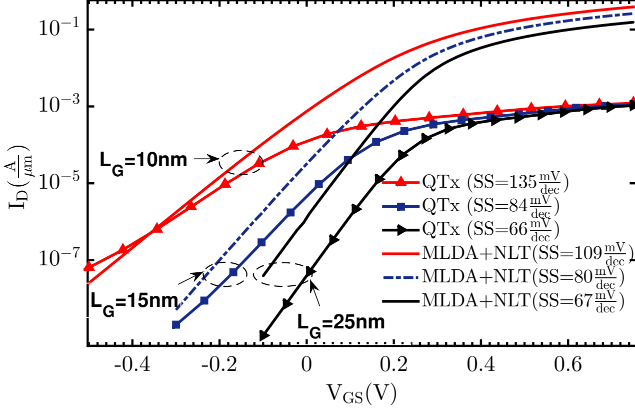

Fig. 3. $I_D V_{GS}$-characteristics obtained from the combination of MLDA + NLT for a 7 nm DG UTB FET with different gate lengths. Parameters: V_{DS} = 0.05 V, m_c = 0.0516 m_0, μ_d = 2.3 × 10^4 cm^2/Vs, γ = 0.7, (α_l, α_v) = (0,1), and WF = 4.8 eV.

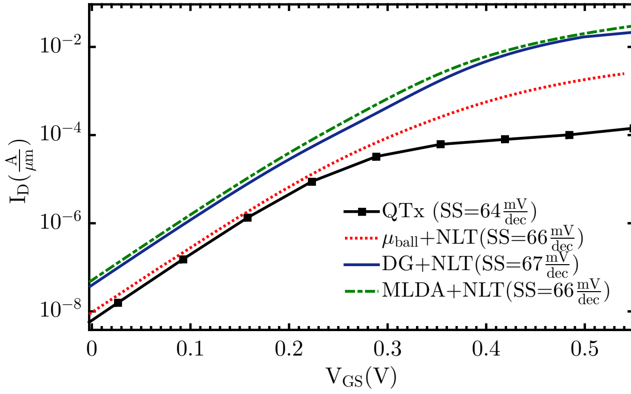

Fig. 4. $I_D V_{GS}$-characteristics obtained from the combination of DG + NLT, MLDA + NLT and μ_b for a t_{body} =7 nm GAA NW. Parameters: V_{DS} = 0.05 V, m_c = 0.0642 m_0, γ = 1.0, (α_l, α_v) = (0,1), and WF = 4.93 eV.

adjust the DOS to the value in 2D- and 1D-like devices is to scale the DOS effective mass by matching the Fermi levels in QTx and S-Device. However, for very small body thicknesses one cannot achieve convergence with the fitted value. In these cases, DOS and γ have to be changed simultaneously to match the electrostatics *and* to reach convergence.

Applying the ballistic mobility model Eq. (1) in combination with a proper WF removes the bumps and yields an overall good agreement with the QTx transfer curves, except for the on-current in the linear regime.

From the extracted *SS* values in Figs. 4 - 7 one observes that despite the superior electrostatic control of the GAA NWs, a significant leakage current caused by STDT persists. The best way to illustrate its effect is by showing the spectral current distribution in comparison to the shape of the lowest CB edge. This is done in Fig. 8 for the off-state and the on-state, respectively, comparing the gate lengths 15 nm and 10.5 nm. The on-current is basically thermionic current in both cases, i.e. STDT has no effect here. The off-current is dominated

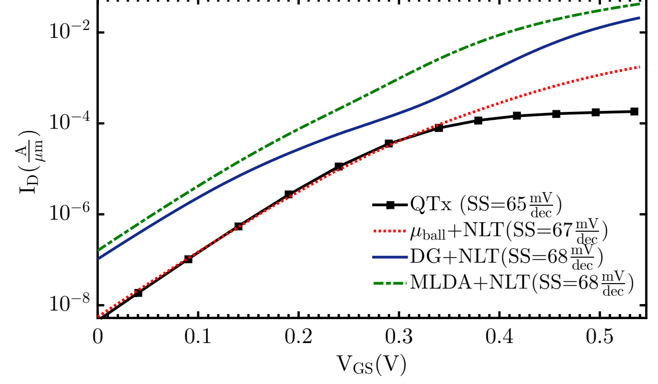

Fig. 5. $I_D V_{GS}$-characteristics obtained from the combination of DG + NLT, MLDA + NLT and μ_b for a t_{body} = 5.5 nm GAA NW. Parameters: V_{DS} = 0.05 V, m_c = 0.0674 m_0, γ = 1.0, (α_l, α_v) = (0,1), and WF = 4.98 eV.

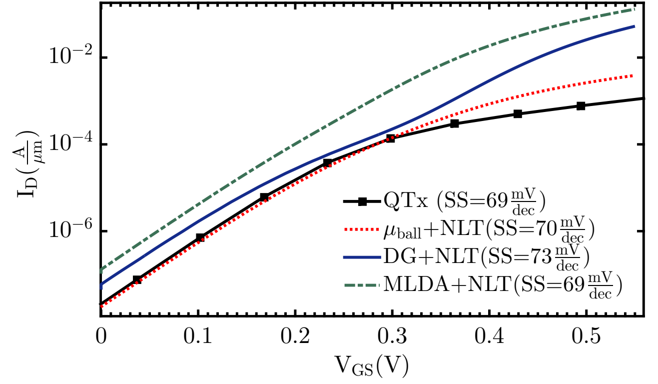

Fig. 6. $I_D V_{GS}$-characteristics obtained from the combination of DG + NLT, MLDA + NLT and μ_b for a t_{body} = 7 nm GAA NW. Parameters: V_{DS} = 0.63 V, m_c = 0.0642 m_0, γ = 1.0, (α_l, α_v) = (0,1), and WF = 4.93 eV.

by STDT - almost completely in the case of L_G = 10.5 nm. Therefore, a gate length of 10 nm can be considered as the end of scaling for III-V-channel FETs. Further scaling will also not significantly improve the on-current.

Tables II and III show how the use of a diffusive μ_d without ballistic correction overestimates the on-current drastically and how μ_b from Eq. (1) improves the situation. The on-current is much better reproduced in the saturation than in the linear regime. Reasons for this are discussed elsewhere [13].

TABLE II
EXTRACTED ON-CURRENT I_{ON} AT V_{GS} = 0.5 V AND V_{DS} = 0.05 V.

Node	I_{ON}(QTx) [$\frac{A}{\mu m}$]	$I_{ON}(\mu_b)$ [$\frac{A}{\mu m}$]	I_{ON}(DG+NLT) [$\frac{A}{\mu m}$]
A	1×10^{-4}	1.56×10^{-3}	1.7×10^{-2}
B	3.7×10^{-4}	1.8×10^{-3}	1.8×10^{-2}

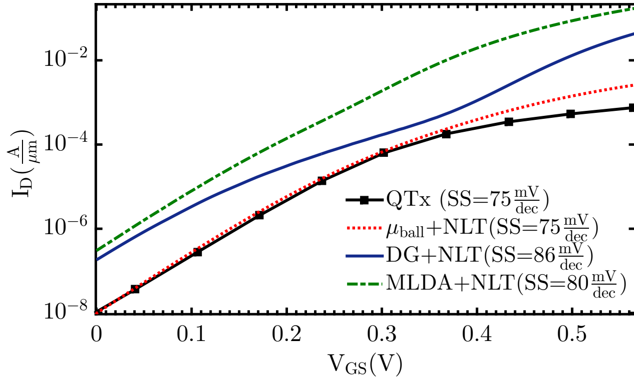

Fig. 7. $I_D V_{GS}$-characteristics obtained from the combination of DG + NLT, MLDA + NLT and μ_b for a t_{body} = 5.5 nm GAA NW. Parameters: V_{DS} = 0.63 V, m_c = 0.0674 m_0, γ = 1.0, (α_l, α_v) = (0,1), and WF = 4.98 eV.

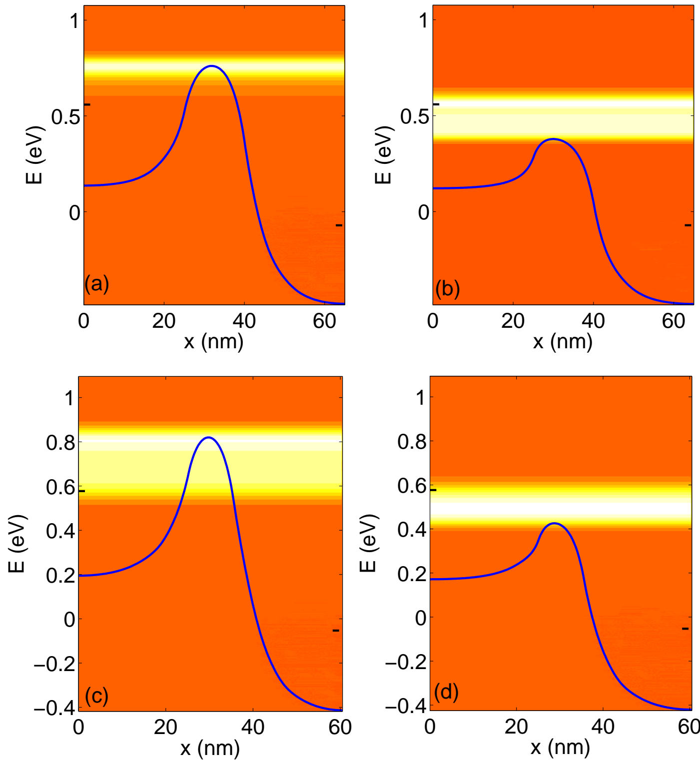

Fig. 8. Distribution of spectral currents and lowest conduction band edge for GAA NW FETs from node (A) and (B) at V_{DS} = 0.63 V. (a) Off-state, L_G = 15 nm at V_{GS} = 0.026 V. (b) On-state, L_G = 15 nm at V_{GS} = 0.48 V. (c) Off-state, L_G = 10.4 nm at V_{GS} = 0.03 V. (d) On-state, L_G = 10.4 nm at V_{GS} = 0.48 V.

IV. Conclusion

The NLT model used in combination with the anisotropic DG model can fairly reproduce the reference *SS* of the DG UTB FETs in the case of strong STDT. In GAA NWs the excellent gate electrostatic control reduces *SS* significantly compared to the DG UTB FETs with the same L_G [7].

TABLE III
Extracted on-current I_{ON} at V_{GS} = 0.5 V and V_{DS} = 0.63 V.

Node	I_{ON}(QTx) $[\frac{A}{\mu m}]$	$I_{ON}(\mu_b)$ $[\frac{A}{\mu m}]$	I_{ON}(DG+NLT) $[\frac{A}{\mu m}]$
A	7.7×10^{-4}	2.6×10^{-3}	2.7×10^{-2}
B	5.3×10^{-4}	1.5×10^{-3}	1.3×10^{-2}

Application of a ballistic mobility model yields an overall good agreement with the QTx transfer curves and decreases I_{ON} by a one order of magnitude in comparison to the simulation with the diffusive mobility μ_d. However, it is still larger than the QTx value, in particular in the linear regime. This can be due to the fact that the true 1D DOS can hardly be mimicked in S-Device. Another possibility is the model for μ_b itself, which overestimates the current in the linear regime [13].

Acknowledgment

The research leading to these results was supported by the European Commissions Seventh Framework Programme (FP7/2007-2013) under Grant Agreement No. 619326 (III-V-MOS Project). The authors would like to thank A. Erlebach (Synopsys Switzerland LLC) for valuable discussions and technical advice.

References

[1] J. A. del Alamo, Nanometer-scale electronics with III-V compound semiconductors, *Nature*, vol. 479, pp. 317-322, 2011.

[2] A. Wettstein, A. Schenk, and W. Fichtner, "Quantum device-simulation with the density-gradient model on unstructured grids", *IEEE Trans. on Electron Devices*, vol. 48, no. 2, pp. 279-284, 2001.

[3] F. O. Heinz, A. Schenk, "Self-consistent Modeling of Longitudinal Quantum Effects in Nanoscale Double-gate Metal Oxide Semiconductor Field Effect Transistors", *Jour. Appl. Phys.*, vol. 100, 084314, 2006.

[4] P. Aguirre, H. Carrillo-Nuez, A. Ziegler, M. Luisier, and A. Schenk, "Drift-diffusion quantum corrections for $In_{0.53}Ga_{0.47}As$ double gate ultra-thin-body FETs", *Proc. of Int. Conf. on Simulation of Semiconductor Processes and Devices (SISPAD)*, pp. 53-56, 2016.

[5] Sentaurus-Device User Guide, V-2016.03, Synopsys Inc., Mountain View, California, 2016.

[6] Sentaurus-Device Monte Carlo (S-Band) User Guide, V-2016.03, Synopsys Inc., Mountain View, California, 2016.

[7] M. Rau, E. Caruso, D. Lizzit, P. Palestri, D. Esseni, A. Schenk, L. Selmi, and M. Luisier, "Performance projection of III-V ultra-thin-body, FinFET, and nanowire MOSFETs for two next-generation technology nodes", *IEDM Tech. Digest*, San Francisco, pp. 758-761, 2016.

[8] A. Schenk and A. Wettstein, "Simulation of DGSOI MOSFETs with a Schrödinger-Poisson based mobility model", *Proc. of Int. Conf. on Simulation of Semiconductor Processes and Devices (SISPAD)*, pp. 21-24, 2002.

[9] M. Luisier, A. Schenk, and W. Fichtner, "Quantum transport in two- and three-dimensional nanoscale transistors: Coupled mode effects in the nonequilibrium Green's function formalism", *Jour. Appl. Phys.*, vol. 100, 043713, 2006.

[10] G. Paasch and H. Übensee, "A modified local density approximation - electron density in inversion layers", *Physica Status Solidi B*, vol. 113, pp. 165-178, 1982.

[11] A. Erlebach, Synopsys Switzerland LLC, private communication.

[12] A. Schenk and P. Aguirre, "A New TCAD Model of the Ballistic Mobility", submitted to *IEEE Trans. on Electron Devices*.

[13] P. Aguirre and A. Schenk, "Ballistic Mobility Model for QDD Simulation of Ultra-short Transistors", accepted for *Electron Devices Technology ad Manufacturing Conference (EDTM)*, Kobe, 2018.