Efficient Modeling of Source/Drain Tunneling in Ultra-Scaled Transistors

O. Baumgartner, L[a]. Filipović, and H. Kosina
Institute for Microelectronics, TU Wien
Gußhausstraße 27–29, 1040 Wien, Austria
{baumgartner|lidijafilipovic|kosina}@iue.tuwien.ac.at

M. Karner, Z. Stanojević, H.W. Cheng-Karner
Global TCAD Solutions GmbH
Landhausgasse 4/1A, 1010 Wien, Austria
{m.karner|z.stanojevic|hw.karner}@globaltcad.com

***Abstract*—In this work, a comprehensive investigation of the effect of source/drain tunneling in ultra-scaled transistors is presented. A novel approach to efficiently and accurately incorporate the quantum-mechanical effects of source/drain (S/D) tunneling in semi-classical device simulators has been developed. The ballistic quantum transport model has been implemented as part of the Vienna-Schrödinger-Poisson simulation and modeling framework. The transport formalism is based on the quantum transmitting boundary method and has been extended to provide recombination and generation rates of carriers due to the direct tunneling current across the source/drain barrier. The model has been used to investigate the effect of direct S/D tunneling on device performance in ultra-scaled double-gate and nanowire transistors. The variations in transfer and output characteristics due to the tunneling effect have been calculated for different gate lengths and channel widths. The influence on the drain induced barrier lowering is shown.**

I. Introduction

With the continued CMOS device scaling, short channel effects such as drain induced barrier lowering (DIBL) and source/drain (S/D) tunneling [1], [2] become an increasing challenge for device performance. A wide variety of devices like double-gate (DG) metal-oxide-semiconductor (MOS) or nanowire transistors [3], [4] are studied to improve the electrostatic control and enable further scaling. Technology computer-aided design (TCAD) tools aim to assist device engineers in including the relevant physics for ultra-scaled devices in their development process. Simulation studies of S/D tunneling using sophisticated quantum transport models have been carried out previously [5]. While these tools provide detailed physical insight, their computational demand is too high in TCAD applications for a wide range of devices.

We present a physically grounded modeling and simulation framework that captures the effect of source/drain tunneling in ultra-scaled MOS transistors while keeping computational costs low. The framework consists of a fast and accurate open-boundary quantum transport solver coupled to a semi-classical device simulator enabling quick performance evaluation.

II. Source/Drain Tunneling Model

Our method to simulate source/drain tunneling in nano-scaled MOS devices involves a ballistic quantum transport model for direct tunneling calculations, implemented in the fast numerical Schrödinger-Poisson solver VSP [6], [7]. It is coupled to a semi-classical device simulator (Minimos NT [8]), both being part of GTS Framework [9]. The schematic structure of the model, the simulation flow, and the data interdependencies are depicted in Fig. 1. The device simulator calculates a self-consistent potential and the quasi-Fermi levels of the contact regions. This is used by the quantum transport model in VSP to calculate the direct tunneling current through the source/drain barrier from which the recombination/generation rate follows. This information is fed back to the device simulator and enters the current continuity equations through

$$\nabla \cdot \mathbf{J}_\mathrm{n} = \mathrm{q_e} R + \mathrm{q_e} \frac{\partial n}{\partial t} + \mathrm{q_e} R_{\mathrm{tun,n}}, \tag{1}$$

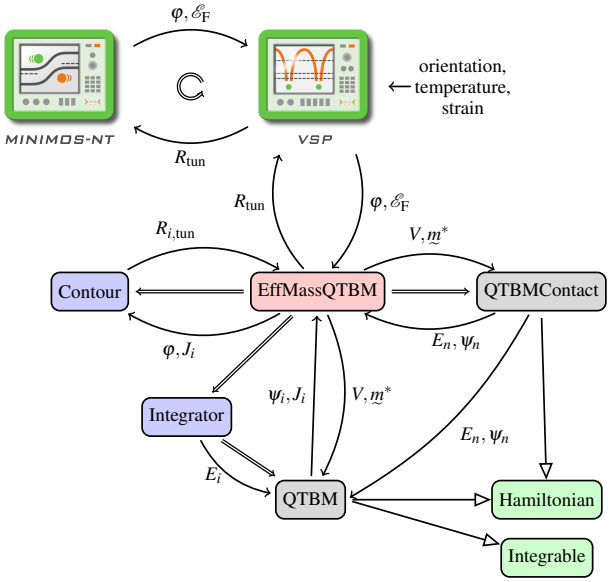


Figure 1: Schematic of the model setup. The device simulator Minimos-NT provides potential and quasi-Fermi levels to the quantum transport simulator VSP, which calls the ballistic quantum transport model to calculate the ballistic current density. The current density spectrum is used to obtain the generation/recombination rates due to S/D tunneling. The calculated rates are transfered back to the semi-classical device simulator, where they are included in the current continuity equation.

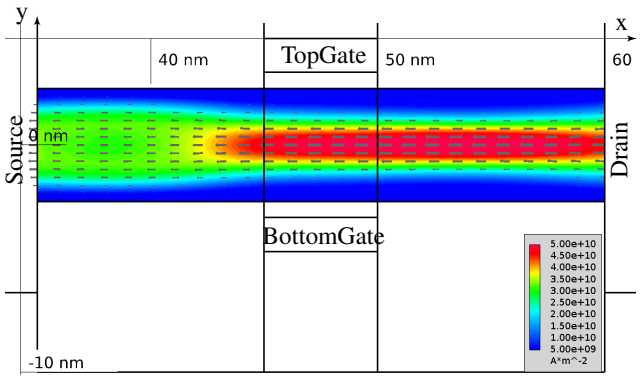


Figure 2: Ballistic electron current density in a Si nMOS double-gate transistor due to direct source/drain tunneling. The DGMOS has a width and channel length of 5 nm, respectively. The device was biased at $V_{\mathrm{DS}} = 0.8\,\mathrm{V}$ and $V_{\mathrm{G}} = 0.0\,\mathrm{V}$. Only the energy range, where direct tunneling through the source/drain barrier occurs is considered for the current calculation which greatly reduces the computational effort.

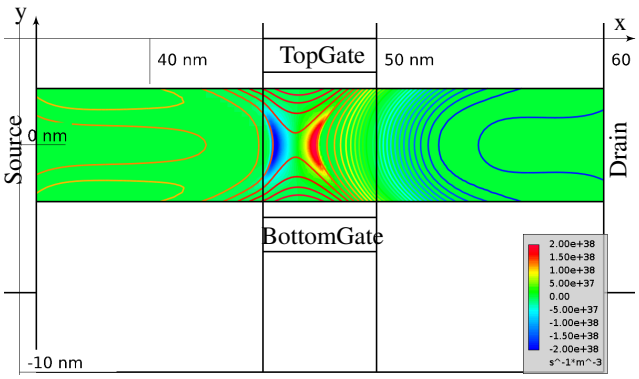


Figure 3: Generation and recombination rates in a Si nMOS double-gate transistor due to direct source/drain tunneling; Same bias as in Fig. 2; The shown iso-lines correspond to the conduction band edge in the channel region. The source/drain barrier, where the direct tunneling occurs, can be seen.

$$\nabla \cdot \mathbf{J}_{\mathrm{p}} = -\mathrm{q_e} R - \mathrm{q_e} \frac{\partial p}{\partial t} - \mathrm{q_e} R_{\mathrm{tun,p}}. \tag{2}$$

The quantum transport model is based on the quantum transmitting boundary method (QTBM) [10] and implemented in C++ within the VSP modeling framework [6], [11]. Due to the generality of the code, the model is applicable to arbitrary device dimension, unstructured grids and arbitrary crystal orientation rendering it useful for a wide range of applications. Using the QTBM formalism, we obtain the solution to

$$(\mathsf{H} - \mathscr{E}\mathbf{I} + \Sigma^{\mathrm{R}})\,|\Psi\rangle = |0\rangle\,. \tag{3}$$

To reduce computational effort, we do not solve the full system directly, but consider Σ^{R} as low-rank update of the system by solving a condensed system of equations containing the contact modes [11]. One obtains the wave functions ψ_i, transmissions T_{ij} between mode i and j, occupation, and ballistic current density, using the appropriate Fermi distribution for each contact. Solving the system for each valley ν in the conduction band and each valley μ in the valence band, we calculate the energy resolved ballistic current density for electrons $\mathbf{S}_{\mathrm{tun,n},\nu}$ and holes $\mathbf{S}_{\mathrm{tun,p},\mu}$, respectively.

For each valley, the energy resolved net recombination rate due to direct source/drain tunneling at each energy $\mathscr{E}_i$ is obtained by calculating

$$R_{\mathrm{tun,n},\nu}(\mathscr{E}_i, \mathscr{S}) = \frac{\nabla \mathscr{E}_{\mathrm{C}}}{\|\nabla \mathscr{E}_{\mathrm{C}}\|} \cdot \frac{\mathbf{S}_{\mathrm{tun,n},\nu}(\mathscr{E}_i, \mathscr{S})}{\mathrm{q_e}} \tag{4}$$

$$R_{\mathrm{tun,p},\mu}(\mathscr{E}_i, \mathscr{S}) = \frac{\nabla \mathscr{E}_{\mathrm{V}}}{\|\nabla \mathscr{E}_{\mathrm{V}}\|} \cdot \frac{\mathbf{S}_{\mathrm{tun,p},\mu}(\mathscr{E}_i, \mathscr{S})}{\mathrm{q_e}} \tag{5}$$

along the contour $\mathscr{S}$ of the conduction band edge $\mathscr{E}_{\mathrm{C}}$ or valence band edge $\mathscr{E}_{\mathrm{V}}$ at this energy. The contour corresponds to the classical turning points. The normalized gradient of the band edge potential in that expression is used to determine whether the tunneling current enters or leaves the tunneling barrier. It defines the sign of the calculated contour weights and thereby whether recombination or generation of the carrier occurs at that specific energy and spatial position.

We integrate $R_{\mathrm{tun,n/p},\nu/\mu}(\mathscr{E}_i, \mathscr{S})$ over all energies with the adaptive *Integrator* module based on our previous work using the Clenshaw-Curtis formula [12], [13]. The total recombination rate for electrons and holes is calculated by the sum for each valley ν and μ, respectively. These quantities are transferred back to the semi-classical device simulator, where they are included in the current continuity equations.

III. Results and Discussion

We carried out simulations of double-gate (2D) and nanowire (3D) nMOS devices. Structure creation and meshing was done with the GTS Structure tool [9]. The simulations can be carried out on single work stations or on distributed computers by making use of the functionality of the GTS Framework.

A. 2D Double-Gate MOS Transistor

As 2D devices we use silicon DG nMOS transistors with highly n-doped source and drain extensions with a donor concentration of $N_{\mathrm{D}} = 1 \times 10^{20}\,\mathrm{cm}^{-3}$. First we investigate a device with a channel width of $D = 5\,\mathrm{nm}$ and physical gate length of $L_{\mathrm{G}} = 5\,\mathrm{nm}$. Figure 2 shows the ballistic quantum-mechanical current density through the source/drain tunneling barrier at a bias of $V_{\mathrm{DS}} = 0.8\,\mathrm{V}$ and $V_{\mathrm{G}} = 0.0\,\mathrm{V}$. From the spectral current density, we calculate the recombination/generation rates due to tunneling. This is illustrated in Fig. 3. The shown iso-lines correspond to the conduction band edge in the channel region. The source/drain barrier, where the direct tunneling occurs, can be seen. The calculated generation and recombination rates for electrons, having opposite signs, are to the left and right of the barrier. This means the electrons vanish on one side and reappear on the other side as expected for a non-local effect.

The transfer characteristics for the DG nMOS with a channel width and physical gate length of 5 nm are shown in Fig. 4.

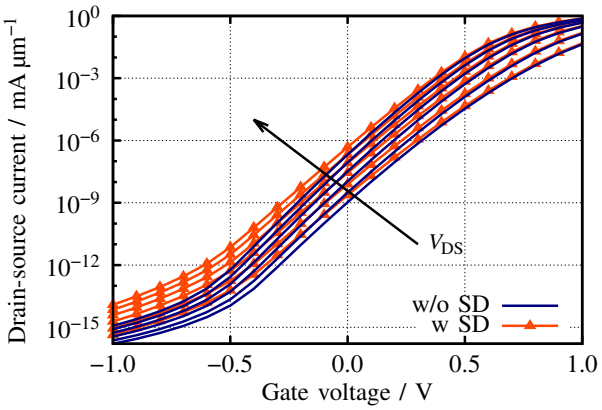


Figure 4: Transfer characteristics for a DG nMOS at V_{DS} ranging from 50 mV to 1 V with and without source/drain tunneling. The channel width and the physical gate length are 5 nm. The increased off-current and sub-threshold slope due to S/D tunneling is clearly visible.

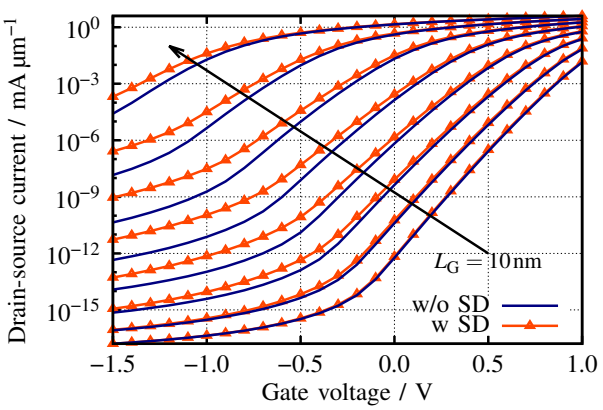


Figure 5: Current-voltage characteristics for a DG nMOS at $V_{DS} = 0.8$ V with and without source/drain tunneling. The channel width is 7 nm and the physical gate length is varied from 10 nm down to 3 nm. As the gate length decreases, the electrostatic control is reduced and the tunneling barrier gets thinner. This leads to higher tunneling current.

The increased off-current and sub-threshold slope when S/D tunneling is considered can be clearly seen.

To investigate the effect of gate length scaling on the device performance we used a DG nMOS with a channel width of 7 nm and varying gate length. The current-voltage (I/V) characteristics at $V_{DS} = 0.8$ V is depicted in Fig. 5. When reducing the physical gate length from 10 nm down to 3 nm the DIBL affects device performance clearly. When S/D tunneling is included in the calculations, the effect becomes considerably stronger. As the gate length decreases, the electrostatic control is reduced and the tunneling barrier gets thinner. This leads to the higher tunneling current, as calculated in the simulations.

B. 3D Nanowire Transistor

Similar to the case of the double-gate transistor we carried out an investigation of circular silicon nanowire transistors. The doping concentrations and device dimensions were chosen according to the 2D device configuration. The nanowire transistor has a diameter of 5 nm as well as a physical gate length of 5 nm.

Figure 6 shows the calculated generation and recombination rates due to direct S/D tunneling in the channel region of the nanowire. The current-voltage characteristics of this device with and without source/drain tunneling for a bias of $V_D = 0.8$ V are shown in Fig. 7a. For this case a small contribution of the direct tunneling leakage to the off-current can be seen. We then changed the geometry of the device and repeated the simulation. When increasing the physical gate length of the nanowire transistor from 5 nm to 7 nm the S/D tunneling current is completely suppressed (see Fig. 7b). We went on to increasing the nanowire diameter from 5 nm to 7 nm while keeping the gate length at 5 nm (cf. Fig. 7c). Compared to the smaller diameter in Fig. 7a we see a strong DIBL even without tunneling. When the source/drain direct

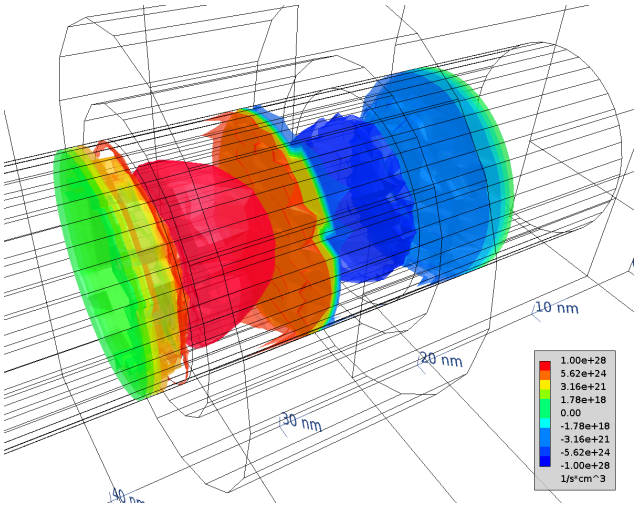
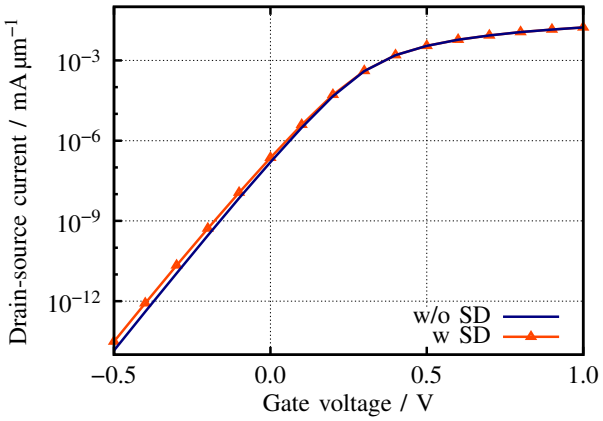


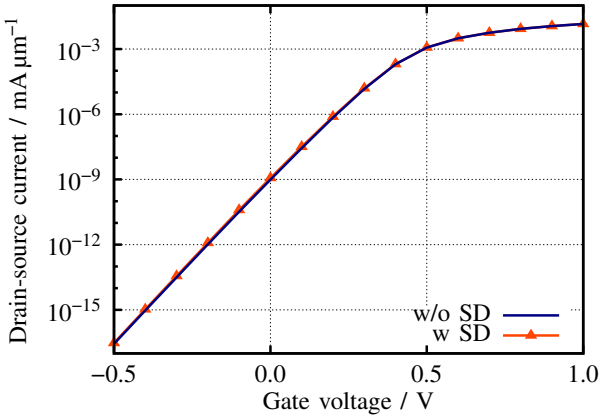
Figure 6: Generation and recombination rates in a Si nMOS nanowire transistor due to direct source/drain tunneling. The nanowire has a diameter and channel length of 5 nm. The device was biased at $V_{DS} = 0.8$ V and $V_G = 0.0$ V. The direct tunneling occurs in the center of the device. Electrons recombine on the source side of the channel (blue iso-surface) and are generated on the drain side (red iso-surface).

tunneling current is included in the calculation, the device performance further deteriorates.

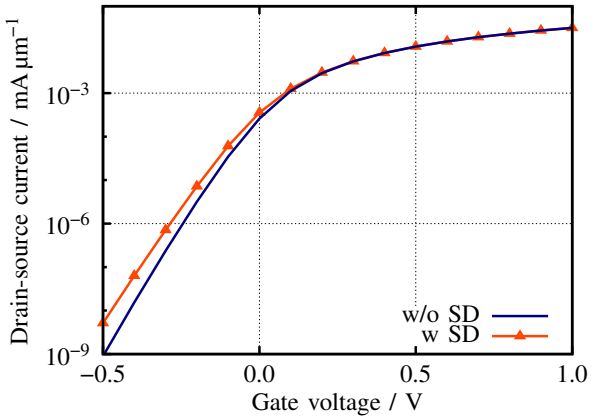
The simulations demonstrate the improved electrostatic control in the nanowire transistor compared to the double-gate MOS device. However, small changes in the device geometry such as a reduced gate length or varying diameter can have tremendous impact on the device characteristics. As shown in Figs. 7a to 7c these variations are enhanced further due to the direct tunneling source/drain leakage current.



(a) $D = 5\,\text{nm}$, $L_G = 5\,\text{nm}$



(b) $D = 5\,\text{nm}$, $L_G = 7\,\text{nm}$



(c) $D = 7\,\text{nm}$, $L_G = 5\,\text{nm}$

Figure 7: Current-voltage characteristics for a 3D silicon nanowire nMOS at $V_{DS} = 0.8\,\text{V}$ with and without source/drain tunneling for different nanowire diameters and gate lengths. The changes in electrostatic control due to changes in device geometry can be clearly seen.

IV. Conclusion

We present a new TCAD model to include source/drain tunneling in device simulations in an efficient and physically accurate way. For that purpose, we coupled the semi-classical device simulator Minimos-NT to the quantum-mechanical simulation framework Vienna-Schrödinger-Poisson (VSP). We derived a means to incorporate the source/drain tunneling current as a recombination/generation term in the current continuity equation. We used the model to investigate the effect of source/drain tunneling in 3D nanowire and 2D DGMOS transistors. We show the influence of S/D tunneling on off-current and sub-threshold-slope. The consequences of gate-length scaling and geometry variations on the tunneling leakage current have been shown.

Acknowledgment

This work has been supported by the The Austrian Research Promotion Agency (FFG) project n° 838551 (NeGFQTS).

References

[1] J. Wang and M. Lundstrom, "Does source-to-drain tunneling limit the ultimate scaling of MOSFETs?" in *2002 International Electron Devices Meeting (IEDM)*, Dec. 2002, pp. 707–710.

[2] S. H. Park, Y. Liu, N. Kharche, M. Jelodar, G. Klimeck, M. Lundstrom, and M. Luisier, "Performance Comparisons of III-V and Strained-Si in Planar FETs and Nonplanar FinFETs at Ultrashort Gate Length (12 nm)," *IEEE Transactions on Electron Devices*, vol. 59, no. 8, pp. 2107–2114, Aug. 2012.

[3] S. Bangsaruntip, A. Majumdar, G. Cohen, S. Engelmann, Y. Zhang, M. Guillorn, L. Gignac, S. Mittal, W. Graham, E. Joseph, D. Klaus, J. Chang, E. Cartier, and J. Sleight, "Gate-all-around silicon nanowire 25-stage CMOS ring oscillators with diameter down to 3 nm," in *2010 Symposium on VLSI Technology (VLSIT)*, Jun. 2010, pp. 21–22.

[4] S. Bangsaruntip, K. Balakrishnan, S.-L. Cheng, J. Chang, M. Brink, I. Lauer, R. Bruce, S. Engelmann, A. Pyzyna, G. Cohen, L. Gignac, C. Breslin, J. Newbury, D. Klaus, A. Majumdar, J. Sleight, and M. Guillorn, "Density scaling with gate-all-around silicon nanowire MOSFETs for the 10 nm node and beyond," in *2013 IEEE International Electron Devices Meeting (IEDM)*, Dec. 2013, pp. 20.2.1–20.2.4.

[5] S. Mehrotra, SungGeun Kim, T. Kubis, M. Povolotskyi, M. Lundstrom, and G. Klimeck, "Engineering Nanowire n-MOSFETs at $L_g < 8$ nm," *IEEE Transactions on Electron Devices*, vol. 60, no. 7, pp. 2171–2177, Jul. 2013.

[6] O. Baumgartner, Z. Stanojevic, K. Schnass, M. Karner, and H. Kosina, "VSP—a quantum-electronic simulation framework," *Journal of Computational Electronics*, vol. 12, no. 4, pp. 701–721, 2013.

[7] "Vienna Schrödinger Poisson." [Online]. Available: http://www.globaltcad.com/vsp

[8] *Minimos-NT User Manual*, Institute for Microelectronics and Global TCAD Solutions GmbH. [Online]. Available: http://www.globaltcad.com/en/products/minimos-nt.html

[9] "GTS Framework," http://www.globaltcad.com/en/products/gts-framework.html. [Online]. Available: http://www.globaltcad.com/en/products/gts-framework.html

[10] C. S. Lent and D. J. Kirkner, "The Quantum Transmitting Boundary Method," *J. Appl. Phys.*, vol. 67, no. 10, pp. 6353–6359, May 1990.

[11] O. Baumgartner, Z. Stanojevic, L. Filipovic, A. Grill, T. Grasser, H. Kosina, and M. Karner, "Investigation of quantum transport in nanoscaled GaN high electron mobility transistors," in *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), 2014*, Sep. 2014, pp. 117–120.

[12] O. Baumgartner, M. Karner, and H. Kosina, "Modeling of High-k-Metal-Gate-Stacks Using the Non-Equilibrium Green's Function Formalism," in *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sep. 2008, pp. 353–356.

[13] C. W. Clenshaw and A. R. Curtis, "A Method for Numerical Integration on an Automatic Computer," *Num. Math.*, vol. 2, pp. 197–205, 1960.