DGSOI versus Bulk: A Quantum-Ballistic Study of 25 nm nMOSFETs

Andreas Schenk[1†], Frederik Ole Heinz[2], and Bernhard Schmithüsen[1]

[1] *Integrated Systems Laboratory, Swiss Fed. Inst. of Technology, Gloriastr. 35, CH-8092 Zürich, Switzerland*

[†] *Synopsys Switzerland LLC, Affolternstrasse 52, CH-8050 Zürich, Switzerland*

[2] *Intel Corporation, Mailstop RA3-254, 2501 NW 229th Avenue, Hillsboro, Oregon 97124, U.S.A.*

e-mail: schenk@iis.ee.ethz.ch, Tel: +41 1 263 6689, Fax +41 1 263 1194

Abstract— **In this paper we study the performance of bulk and DGSOI nMOSFETs with 25 nm gate length in the quantum-coherent limit. The self–consistent wave functions are computed using a multi–sub-band scattering matrix formalism which allows to retain their full dimensionality and therefore eliminates the need for the adiabatic decomposition of the Schrödinger equation. We find that source-drain tunneling is negligibly small in both devices. Since the current is almost exclusively thermionic, the observed increase of the off-current with increasing drain bias can be attributed to drain-induced barrier lowering. The quantum-ballistic currents are by a factor of 2-3 larger than the quantum-drift-diffusion currents. The quantum-ballistic sub-threshold slope is almost the same for bulk and DGSOI MOSFET.**

I. INTRODUCTION

The progress in silicon technology continues at a breath-taking pace even beating Moore's law, but the end of MOSFET scaling can be anticipated for the year 2015 or so, since fundamental physical limits pose virtually impenetrable barriers to tera-scale integration [1]. The 2003 edition of the ITRS [2] forecasts a minimum feature size of 25 nm, a physical gate length of 10 nm, and a transistor density on chip of 2151 million for the year 2015. The switching charges will then contain only a few hundreds of electrons. This size is at the physical limit (quantum effects, non-deterministic behavior of small currents), at the technological limits (power dissipation, design complexity, tunneling leakage currents), and at the economical limit (estimated cost of a silicon fab in 2015 $\sim$ 100 billion \$).

In the tunneling regime (gate length shorter than 15 nm), the off-state current will be determined by the transmission probability of the source-drain barrier and thus will be independent on inelastic scattering events up to a possible phonon-assistance of the source-drain tunnel current. However, the on-current will be influenced both by the contact resistances and the few scattering events inside the channel. In the channel region the on-state current will be a mixture of quantum-coherence and inelastic scattering. The development of predictive TCAD tools for ultimately scaled CMOS devices is taking two directions: the inclusion of quantum effects into PDE-based device simulators (quantum-drift-diffusion (QDD) model), or the inclusion of dissipation in the quantum-ballistic (QB) transport model [3]. State-of-the-art QDD simulators are successful in reproducing one-dimensional quantum confinement, but fail e.g. in the prediction of source-drain tunneling and in any situation far from equilibrium. The second method looks upon the device as an electron waveguide and yields, therefore, an accurate description of all tunneling processes, but needs to be complemented by de-phasing and inelastic scattering. The pure QB case based on the Landauer-Büttiker formalism [4] might offer a kind of "best case" evaluation of device behavior. In this paper we use the latter approach to draw a comparison between the performance of bulk and DGSOI nMOSFETs with 25 nm gate length in the QB limit. This comparison is inspired by the two current trends in industry - bulk versus SOI.

II. DEVICE DESCRIPTION

The device structures and doping profiles were defined for benchmark purposes during the EU project SINANO (*Silicon-based Nanodevices*) [5] and can be considered to be representative for the high-performance MOSFET of the current 65 nm technology node. The selected doping profiles and oxide thickness try to comply with the available specifications of the ITRS 2003 and have been tuned in order to meet the specification for the maximum allowable leakage drain current in off-state. Both devices have an equivalent oxide thickness (EOT) of 0.9 nm, a physical oxide thickness t_{ox} =1.6 nm, a permittivity $\epsilon_{\mathrm{ox}} = 7$, and a supply voltage V_{DD}=1.1 V. The DGSOI MOSFET has a silicon body thickness of t_{Si} =12 nm and a gate electrode work function of 4.60 eV (metal gate). The body is lowly p-doped ($1\times10^{15}\mathrm{cm}^{-3}$). The Gaussian n-type source/drain extension profiles have a standard deviation of $\sigma_{\mathrm{y}} = 5.64\times10^{-3}\,\mu$m and the Gaussian n-type source/drain contact profiles have a peak concentration of $1\times10^{20}\mathrm{cm}^{-3}$ and a standard deviation of $\sigma_{\mathrm{y}} = 1.12\times10^{-2}\,\mu$m. The bulk MOSFET has a gate electrode work function of 4.05 eV (n-poly), a p-type substrate with constant doping concentration of $3\times10^{18}\mathrm{cm}^{-3}$ and p-type source/drain halos of Gaussian shape with a peak concentration of $8\times10^{18}\mathrm{cm}^{-3}$ and a standard deviation of $\sigma_{\mathrm{y}} = 1.77\times10^{-2}\,\mu$m. The n-type profiles are the same as for the DGSOI MOSFET.

III. QUANTUM-BALLISTIC TRANSPORT MODEL

QB conductances and currents were obtained with *SIMNAD* (SImulator for NanoDevices) – a quantum-mechanical 3D simulator for semiconductor devices developed at ETH Zürich [6]. For QB transport simulations through quantum

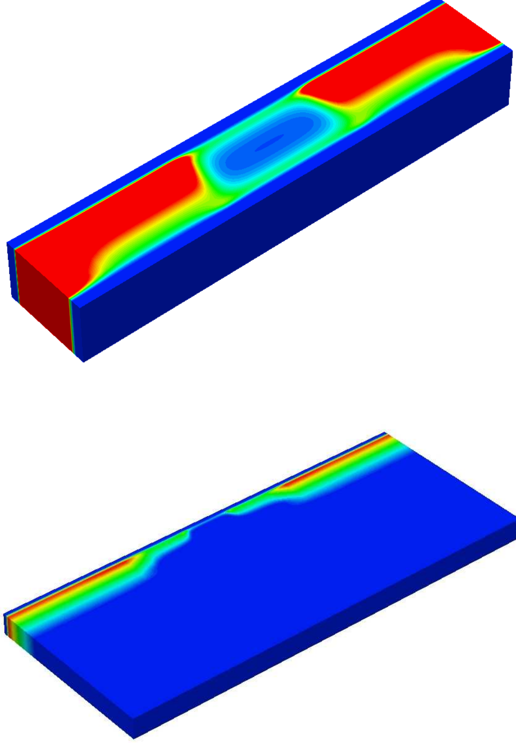

Fig. 1. Quantum-mechanical charge density in the 25 nm DGSOI (upper) and bulk MOSFET (lower).

wires/wells, the electron-electron interaction is treated on the level of a mean field approach and the Schrödinger equation with scattering boundary conditions for injection of electrons from the source (or drain) contact are solved by means of a scattering matrix approach (SMA) including sub-band mixing. This formalism allows to retain the full dimensionality of the wave functions and eliminates the need for the often applied adiabatic decomposition of the Schrödinger equation. Along the transport direction, piecewise analytical wave functions are used. This has the advantage that the discretization grid can have much fewer grid planes than are needed with a finite differences scheme such as is commonly used by the NEGF community. The QB currents in the 2D MOSFETs are computed by a 2D Landauer-Büttiker-type formula [7]

$$\begin{aligned} I^{2D} &= \frac{2}{h}\sqrt{\pi}W\frac{\sqrt{2m_\perp^* k_B T}}{h} \times \\ &\times \sum_i \Bigg(\int_{\epsilon_i^{(0)}} d\epsilon \Big(1 - \sum_j R_{\mathrm{src},i\to j}(\epsilon)\Big)\, \mathfrak{F}_{-\frac{1}{2}}\Big(\frac{\epsilon_F^{(\mathrm{src})}-\epsilon}{k_B T}\Big) - \\ &- \int_{\epsilon_i^{(n_{\max})}} d\epsilon \sum_j T_{\mathrm{drn}\to\mathrm{src},i\to j}(\epsilon)\, \mathfrak{F}_{-\frac{1}{2}}\Big(\frac{\epsilon_F^{(\mathrm{drn})}-\epsilon}{k_B T}\Big)\Bigg), \end{aligned} \tag{1}$$

with W being the width in the third dimension. The integration limits are given by the bottom of the sub-bands at the source and drain contacts, respectively, and the sum runs over all states in quantization direction. Neglecting sub-band mixing, expression (1) for a single state in quantization direction reduces to

$$\begin{aligned} I^{2D} &= \frac{2}{h}\sqrt{\pi}W\frac{\sqrt{2m_\perp^* k_B T}}{h} \times \\ &\times \int_{\epsilon_{\mathrm{sub}}}^{\infty} d\epsilon\, T(\epsilon) \left[\mathfrak{F}_{-\frac{1}{2}}\Big(\frac{\mu-\epsilon}{k_B T}\Big)\right]_{\mu=\epsilon_F^{(\mathrm{drn})}}^{\mu=\epsilon_F^{(\mathrm{src})}} . \end{aligned} \tag{2}$$

It differs from its 1D equivalent by replacing Fermi functions with Fermi–Dirac integrals, and by an additional factor of

$$\sqrt{\pi}W\frac{\sqrt{2m_\perp^* k_B T}}{h} = \sqrt{\pi}\frac{W}{\lambda_T},$$

with the (transverse) thermal wavelength $\lambda_T = h/\sqrt{2m_\perp^* k_B T}$. This factor may be interpreted as the effective number of $k_\perp$–modes that contribute to the current ($W/\lambda_T \gg 1$ must hold). For the simulation of thermionic currents the transmission probability is replaced by a step function:

$$T(\epsilon) = \Theta(\epsilon - \epsilon_{\max}) . \tag{3}$$

Here $\epsilon_{\max}$ is the maximum of the sub-band energy of the transverse mode in which the particle is injected, between the injecting and the extracting terminal. If $\epsilon \geq V_{\max}$ the particle is transmitted, otherwise it is always reflected.

The SMA may be used in a post-processing step for the computation of the current, in which case the charge density for the non-linear Poisson equation is constructed by simply populating the transverse wave functions on each slice with Fermi-Dirac statistics. This so called non-self-consistent (nsc) variant largely facilitates the evaluation of the transmission and reflection probabilities R and T. As an alternative, the full SMA may be used for the computation of the electron density inside the solution process for the non-linear Poisson equation. This self-consistent (sc) variant gains importance at threshold, where the strong injection from the source leads to noticeable quantum-mechanical effects on the electrostatics along the transport direction of the transistor, but it is computationally costly and vulnerable to convergence problems.

The quantum-mechanical charge density of both devices is shown in Fig. 1. Source/drain contacts had to be placed at the front faces because of the boundary conditions of the SMA. Several extensions of the bulk FET in y-direction (perpendicular to the channel) were tried out without any significant effect on the IV-curves. This can be explained by the decaying doping concentration in y-direction which forms a confining potential well. Both devices were meshed with a tensor-product grid of about 30'000 vertices resulting in a CPU time for a complete nsc transfer characteristics of 87 hours on a SunBlade 2000 with 1.015 GHz.

IV. Results

Fig. 2 shows the eigenenergies of the 4 lowest sub-bands in the 25 nm DGSOI MOSFET at a sheet density of $1\times10^{13}\mathrm{cm}^{-2}$ together with the band edge profile. The first number in (a,b) labels the sub-band, the second number indicates the valley pair. Here, "0" is the pair with the large effective mass in quantization direction, "1" and "2" label the degenerate valley pairs with the small effective mass in quantization direction. The splitting of the two lowest sub-bands is only 3×10^{-5} eV, whereas the splitting between the third and the fourth sub-band is 8.3×10^{-3} eV. Since the Fermi level corresponds to the energy zero, one can infer that 4 sub-bands are already sufficient for the computation of the density at 300 K. In Fig. 3 we plotted the corresponding wave functions at a sheet density of $1\times10^{13}\mathrm{cm}^{-2}$. States with (a,2) are identical to states with (a,1) – only the latter are labeled in the figure. Figs. 4 and 5 present the QB transfer characteristics. The currents of the DGSOI MOSFET were divided by a factor of 2 in order to allow for a fair comparison with the bulk FET. Curves labeled "low $\mathrm{V_{DS}}$" were obtained from the **nsc** linear-response conductance assuming that: (i) $\mathrm{V_{DS}}\sim\mathrm{k_BT}$ (injection from drain is then negligible), (ii) $\mathrm{V_{DS}}$ small enough in order not to change the transmission of the source-drain potential barrier. These assumptions make a translation from conductance to current straightforward. In fact, the expression for the linear-response conductance equivalent to Eq. (3) reads

$$G^{2\mathrm{D}} = \frac{2e^2}{hk_\mathrm{B}T}\sqrt{\pi}\,W\frac{\sqrt{2m_\perp^* k_\mathrm{B}T}}{h}\times \int_{\epsilon_\mathrm{sub}}^{\infty} d\epsilon\, T(\epsilon)\,\mathfrak{F}'_{-\frac{1}{2}}\Big(\frac{\epsilon_\mathrm{F}^{(\mathrm{src})}-\epsilon}{k_\mathrm{B}T}\Big)\,. \tag{4}$$

Comparing (5) with (3) immediately leads to $G^{2\mathrm{D}} = eI^{2\mathrm{D}}/k_\mathrm{B}T$ under the above assumptions.

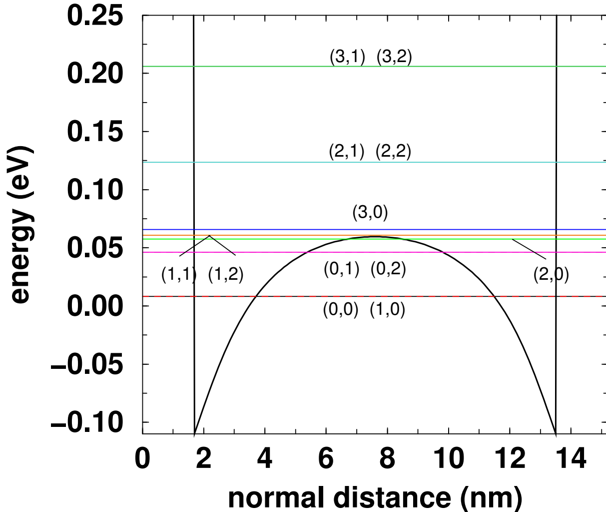

Fig. 2. Band edge profile and eigenenergies of the 4 lowest sub-bands in the 25 nm DGSOI MOSFET at a sheet density of $1\times10^{13}\mathrm{cm}^{-2}$.

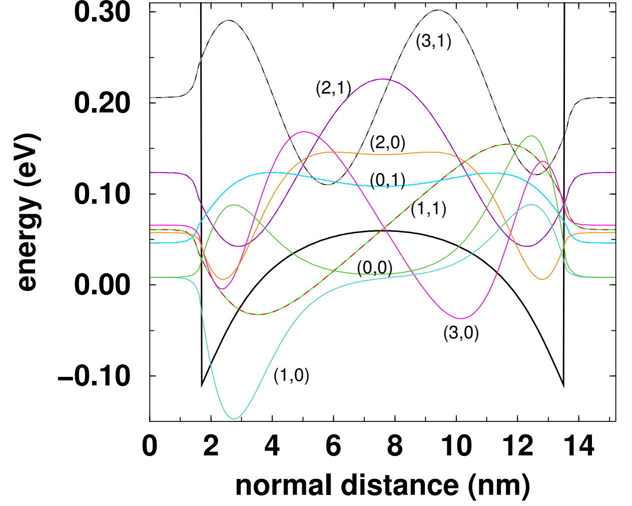

Fig. 3. Wave functions corresponding to the energy levels in Fig. 2.

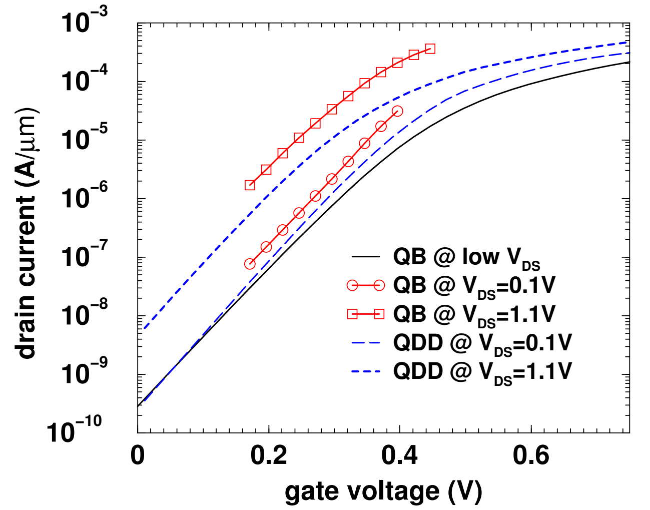

Fig. 4. Quantum-ballistic transfer characteristics of the 25 nm DGSOI MOSFET. Explanations are given in the text.

Curves with indicated $\mathrm{V_{DS}}$ labeled "QB" are the **sc** QB currents including the non-equilibrium charge injection from the contacts. A finite forward bias requires to use this mode. Although only small fractions could be obtained due to convergence problems, one observes an increasing off-current with increasing $\mathrm{V_{DS}}$. This is merely due to a DIBL effect, since the current is almost exclusively thermionic (otherwise the thinning of the source-to-drain barrier would lead to an increased tunnel current). For comparison we plotted in the same figure the quantum-drift-diffusion (QDD) currents, where transport is assumed to be dissipative and all ballistic effects are neglected. Note that QDD exactly reproduces the quantum VT shift. Below threshold the ratio between QB and QDD currents of the DGSOI MOSFET is about 3 at $\mathrm{V_{DS}}=1.1$ V and about 2 at $\mathrm{V_{DS}}=0.1$ V. The effect of source-to-drain tunneling is shown in Fig. 6. It is more pronounced for the DGSOI

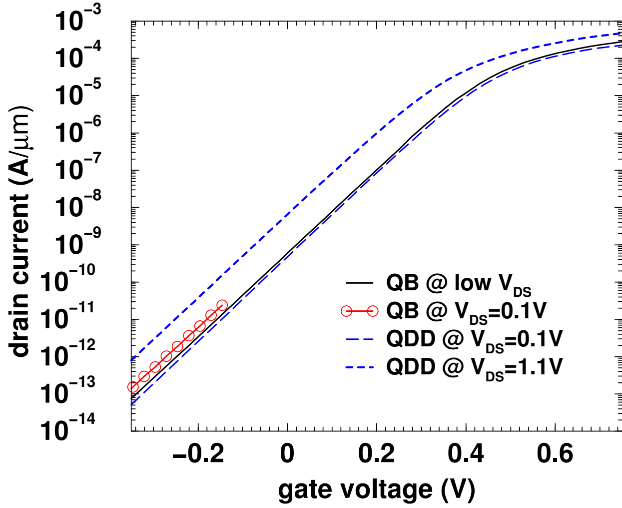

Fig. 5. Quantum-ballistic transfer characteristics of the 25 nm bulk MOSFET. Explanations are given in the text.

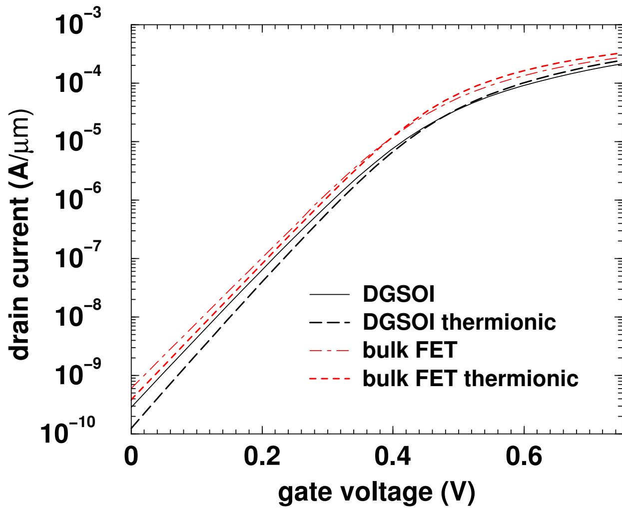

Fig. 6. Comparison of the nsc quantum-ballistic currents of DGSOI and bulk FET at low V_{DS} (see text for the meaning of "low V_{DS}"). The thermionic currents (transmission replaced by a step function) are also shown.

MOSFET, but altogether very small because of the 25 nm gate. The current of the bulk FET is larger by a factor of about 1.5 over the whole range. This has nothing to do with the nsc scheme, because the influence of the injected channel charge is negligible below the threshold voltage. The sub-threshold swing is only slightly better for the DGSOI MOSFET and there is slightly more quantum reflection in the on-state in case of the bulk FET. Fig. 7 compares both devices in terms of their QDD currents. Here, the sub-threshold swing of the DGSOI MOSFET shows a significant improvement compared to the bulk FET. The DIBL at $V_{DS} = 1.1$ V is similar and the on-currents at the same forward bias are identical.

V. Conclusion

We studied the performance of bulk and DGSOI nMOSFETs with 25 nm gate length in the QB limit. The increase

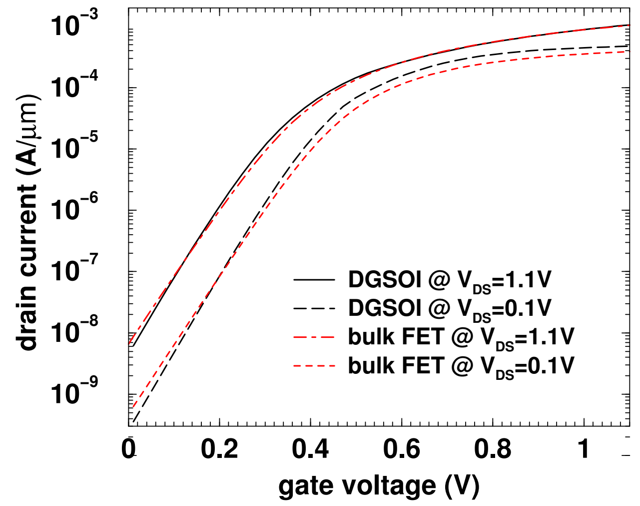

Fig. 7. Comparison of the quantum-drift-diffusion currents of DGSOI and bulk FET at $V_{DS} = 0.1$ V and $V_{DS} = 1.1$ V.

of the off-current with increasing drain bias was attributed to drain-induced barrier lowering. QB currents are by a factor of 2-3 larger than the QDD currents of the same device. The QB sub-threshold slope is almost the same for bulk and DGSOI MOSFET, whereas the QDD sub-threshold slope is significantly steeper for the DGSOI MOSFET. At low drain bias, the current of the bulk FET exceeds that of the DGSOI MOSFET by a factor of about 1.5 over the whole gate voltage range. This is in contrast to the QDD on-current which is larger for the DGSOI MOSFET. A possible explanation for the former observation is that the stronger quantization in the DGSOI MOSFET results in a smaller density of the 2DEG. The latter finding can be attributed to stronger surface-roughness scattering in the bulk FET. One may conclude that QB simulations can lead to qualitatively erroneous results when applied to the comparison of different MOSFET architectures.

VI. Acknowledgment

The first author is grateful for the financial support by Fujitsu Laboratories Ltd.

References

[1] J. D. Meindl et al., "Limits on Silicon Nanoelectronics for Terascale Integration", *Science*, vol. **293**, 2044 (2001).
[2] ITRS, 2003 Edition.
[3] R. Venugopal et al., "Quantum mechanical analysis of channel access geometry and series resistance in nanoscale transistors", J. Appl. Phys. **95(1)**, 292-305, 2004.
[4] M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas. *Phys. Rev. B*, 31(10):6207–6215, 1985.
[5] http://www.sinano.org/
[6] SIMNAD User's Manual. ETH Zürich, 2004.
[7] Frederik Ole Heinz. Simulation Approaches for Nanoscale Semiconductor Devices. Hartung und Gorre Konstanz, 2004.