Performance and Leakage Analysis of Si and Ge NWFETs Using a Combined Subband BTE and WKB Approach

Z. Stanojević, G. Strof, O. Baumgartner, G. Rzepa, and M. Karner
Global TCAD Solutions GmbH., Bösendorferstraße 1/12, 1010 Vienna, Austria
Email: {z.stanojevic|g.strof|o.baumgartner|g.rzepa|m.karner}@globaltcad.com

***Abstract*—We are the first to present a subband-BTE solver with a fully integrated source/drain-tunneling current calculation based on the WKB-approximation. The method is validated against ballistic NEGF calculations showing good agreement. An investigation of Si and Ge-based NWFETs is performed showing that intra-band source/drain-tunneling is not a concern for Si devices. For Ge-based PMOS devices however, tunneling leakage limits sensible L_G-scaling to around 20 nm.**

I. Introduction

As technology development progresses to the 5 nm-node and beyond, source/drain-leakage is found to be the ultimate obstacle to CMOS-scaling. In silicon, even with printed gate-lengths as low as 20 nm, intra-band source/drain-tunneling is not seen as a concern. However, this does not necessarily hold for alternative channel materials such as SiGe and Ge; both are considered candidates for the replacement of Si for high-performance CMOS at 3 nm [1, 2]. Such high-mobility materials however come at the price of increased S/D-leakage [3].

II. Methodology

The Subband-Boltzmann transport equation (SBTE) solver underlying to this work has been presented in [4]. It is commercially available as the GTS Nano Device Simulator (NDS)[5], part of GTS Framework. The solver was shown to accurately reproduce the device characteristics of highly-scaled stacked-nanowire FETs [6]. In this work, the solver is extended to cover source/drain-tunneling by implementing a WKB-based scheme operating on the complex subband structure.

A. Complex subband-structure

The complex band structure is obtained by extending the **k**-vector to the complex plane. For each subband $\mathbf{k}_0$ i.e. the minimum (for electrons, maximum for holes) is found and the Hamiltonian $H(\mathbf{k}_0 + \mathrm{i}\kappa)$ is solved to obtain the $E_n(\mathbf{k}_0 + \mathrm{i}\kappa)$-relations of the evanescent (i.e. tunneling) states (Fig. 1). For simulation of planar devices, this process has to be repeated for every **k**-grid-line parallel to the transport direction as shown in the bottom part of Fig. 1.

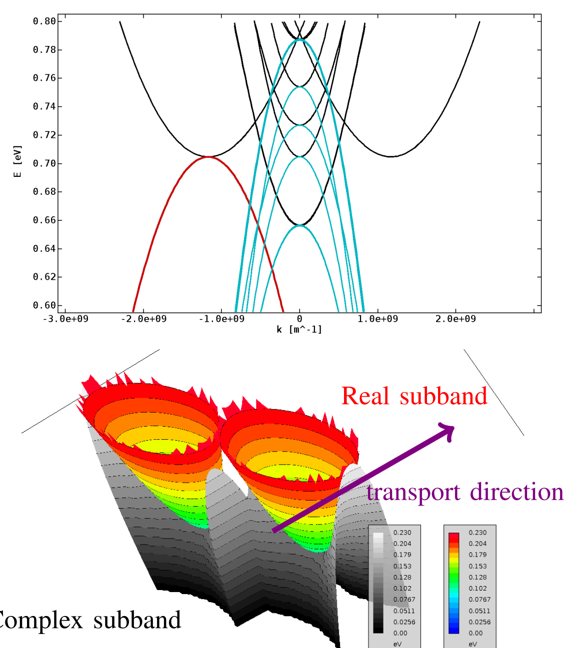


Fig. 1. Top: real (black) and complex (colored) 1D conduction band structure of a Si-NMOS-NWFET; complex subbands are expanded at the minimum of each of their real counterparts; bottom: real (colored) and complex (gray) 2D conduction band structure in a Si-NMOS-UTBFET

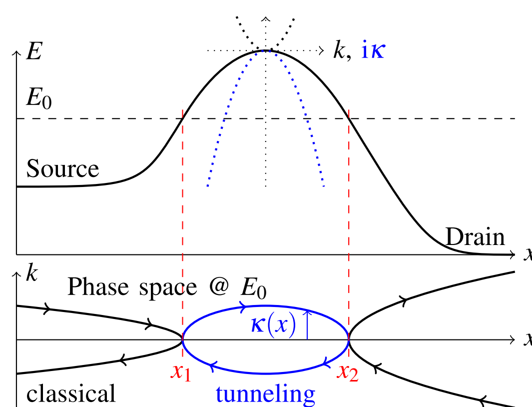


Fig. 2. Typical n-MOSFET potential in off-state along with the local real $E(x,k)$ and complex dispersion relation $E(x,\mathrm{i}\kappa)$; a cut through the $E(k,x)$ landscape at energy E_0 reveals the classical and tunneling trajectories an electron can take; the classical turning points are marked as x_1 and x_2.

© 2020 The Japan Society of Applied Physics

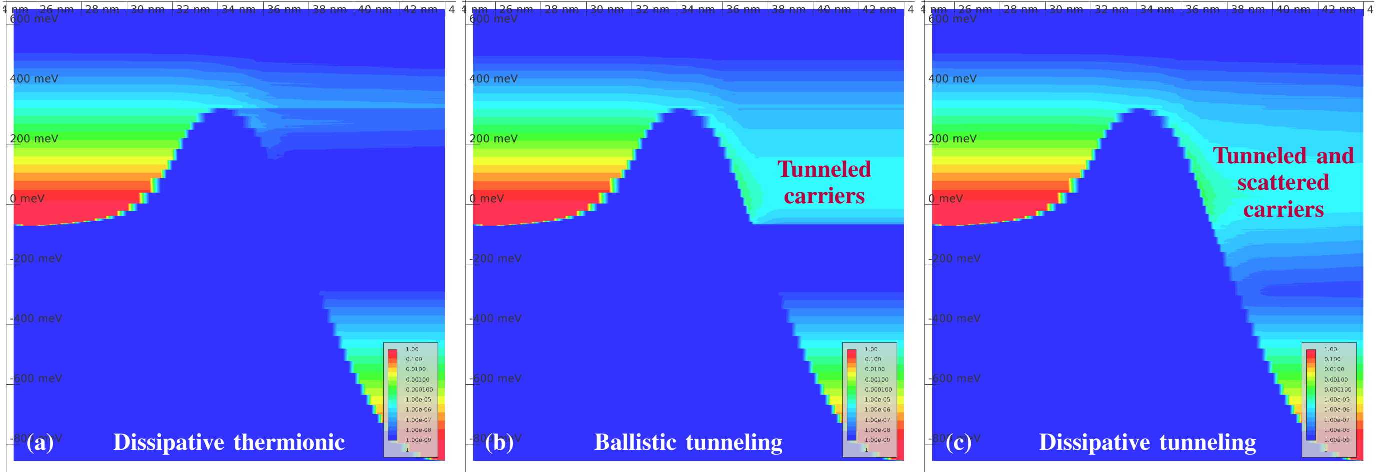


Fig. 3. Hole spectra in a Si-PMOS with L_G = 5 nm biased at V_{GS} = 0 V and V_{DS} = −0.8 V; for the thermionic-only case (a) few holes cross the barrier and are being scattered to lower energies; however the majority of the current is due to tunneling transport, shown in (b) the ballistic and (c) the dissipative cases

B. WKB-implementation

Based on the complex subbands, tunneling paths are constructed that allow the evaluation of the WKB-tunneling-integral,

$$\theta(E_0) = \exp\left[\int_{x_1}^{x_2} \kappa(E_0,x)\mathrm{d}x\right]. \quad (1)$$

computed for each energy E_0 between the turning points x_1 and x_2. κ is the imaginary wave number in the classically forbidden region, which for parabolic bands can be computed from

$$\kappa(E_0,x) = \frac{\sqrt{2m^*(V(x)-E_0)}}{\hbar}, \quad (2)$$

but can also be calculated for arbitrary numerical representations of the complex band structure by integrating along the E_0-contour of the complex phase-space dispersion relation, $E(x,\mathrm{i}\kappa)$. This procedure is visually represented in Fig. 2 and allows us to use both effective mass and $\mathbf{k}\cdot\mathbf{p}$ band models in this work. From the integral the first and second order approximations of the transmission coefficients (TC) can be calculated,

$$\mathrm{TC}^{1^{\mathrm{st}}} = \frac{1}{\theta^2}, \quad \mathrm{TC}^{2^{\mathrm{nd}}} = \frac{1}{(\theta + \frac{1}{4\theta})^2} \quad (3)$$

The transmission coefficients are used to derive effective tunneling rates which are entered into the SBTE system of equations at the respective turning points of each energy. The net spectral current density per subband through the barrier of a NWFET is given by the Tsu-Esaki-formula,

$$j_n(E_0) = -\frac{g_v q}{\pi\hbar}\mathrm{TC}_n(E_0)\left[f_n(E_0,x_1) - f_n(E_0,x_1)\right], \quad (4)$$

where g_v denotes valley degeneracy. In the case of a planar MOSFET with a 2D $\mathbf{k}$-space, an additional integration over the lateral $\mathbf{k}$-space component has to be performed. Since we are only concerned with steady-state simulations, it can be assumed that the the carrier tunneling happens instantaneously. Furthermore, the tunneling process is assumed to not be affected by scattering.

The seamless integration of WKB-tunneling in SBTE becomes is apparent in Fig. 3 (b) and (c), especially the latter where tunneling and scattering are shown interacting. The WKB-integrals are simple repeated 1D integrations of $\kappa(E_0,x)$ and entering the effective tunneling rates into the SBTE system does not increase its rank; this means that there is barely any impact on simulation run-time when WKB is used.

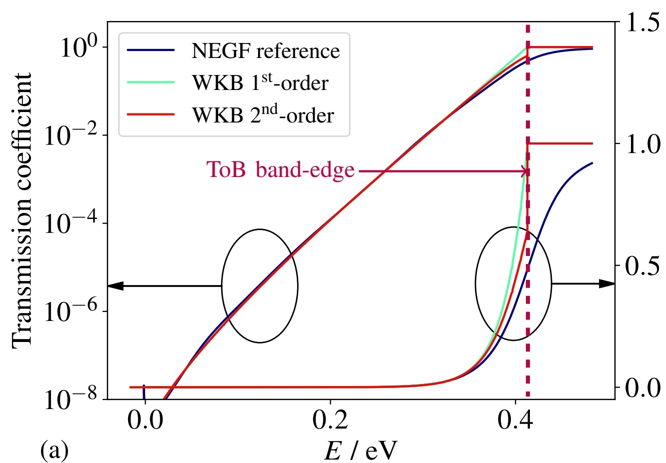
III. Results

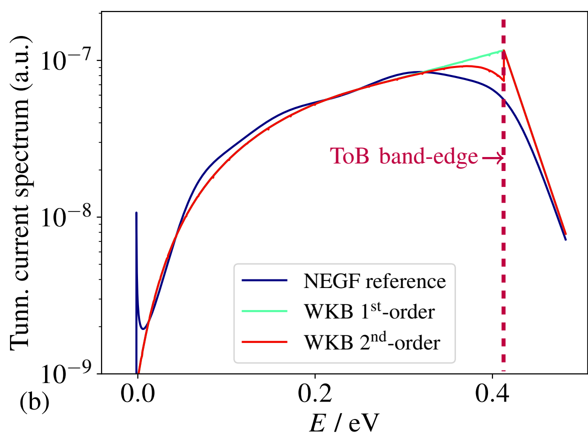
A. Validation against NEGF-reference

To validate the presented approach, we compared the results of ballistic SBTE (with and without WKB tunneling enabled) with our in-house ballistic NEGF solver, which operates in full 3D-real-space as opposed to the mode-space approach employed for the SBTE. As test case, we selected the shortest of the Si NMOS devices described in Section III-B with a gate length of 5 nm in order to emphasize the source-drain tunneling effect. Both SBTE and NEGF simulations were performed self-consistently with electrostatics and using a parabolic effective mass band structure model.

Transmission coefficient and current spectrum in the off-state are shown in excellent agreement for the lowest subband in Fig. 4. A very good agreement is also achieved for the transfer characteristic at V_{DS} = 0.7 V across a wide range of gate voltages. This is surprising considering the vastly different approaches of SBTE + WKB and NEGF. No perceivable difference was noted between using the 1st and 2nd-order approximations for TC in the SBTE + WKB results.

It is worth noting that the self-consistent solutions of SBTE + WKB, which are typically calculated in a few minutes from scratch, were used to provide initial potentials for the NEGF simulations, which then took hours to complete on 20-core servers highlighting the efficiency of the presented approach.





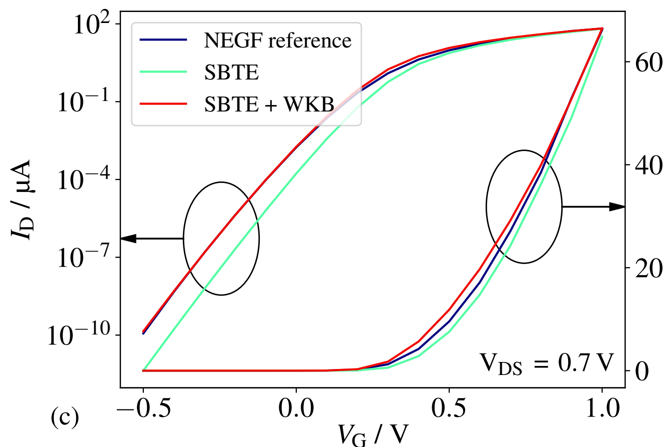
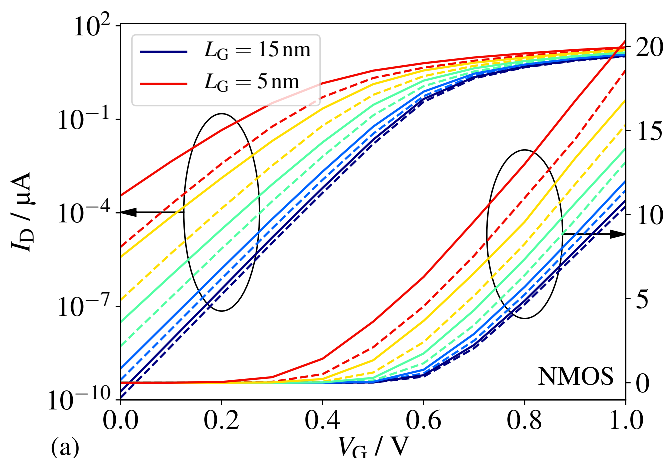
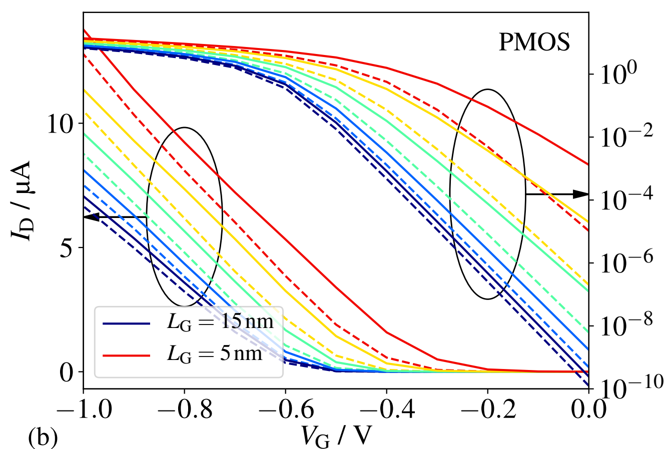


Fig. 4. Validation of ballistic SBTE (with and without WKB) against ballistic NEGF (eff. mass) using the $L_G = 5\,\mathrm{nm}$ NMOS device; graphs (a) and (b) show the comparison of transmission coefficient and tunneling current spectrum, respectively; NEGF and WKB agree very well except near the band edge, where the WKB has unity value by definition; there is also excellent agreement in the transfer characteristic in graph (c).





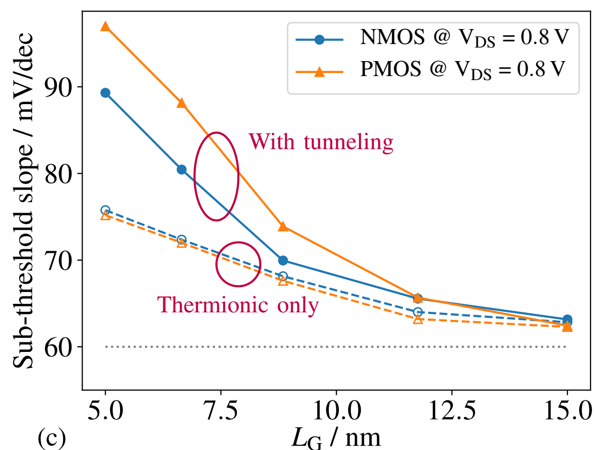


Fig. 5. Saturated I_D/V_G-curves for Si-NMOS (a) and Si-PMOS (b) for varying L_G; dashed curves represent thermionic-only transport while solid curves represent transport with tunneling enabled; SS-roll-off is shown in graph (c) for both device types, where PMOS shows slightly higher SS-degradation

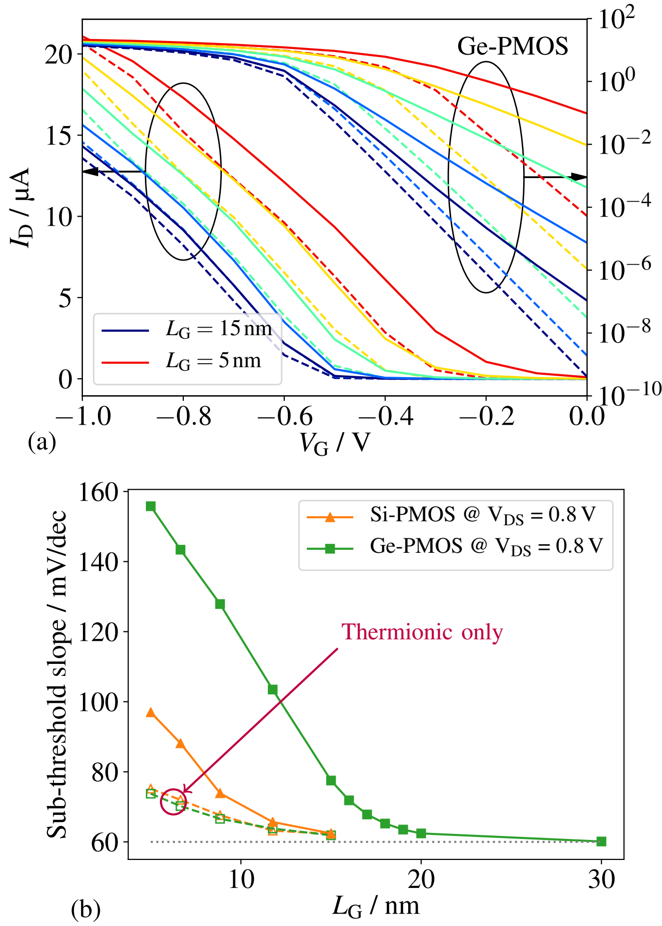


Fig. 6. Saturated I_D/V_G-curves for Ge-PMOS (a) – same as Fig. 5 (b); SS-roll-off is shown in graph (c) for Si & Ge-PMOS, where the Ge-PMOS shows severe SS-degradation for $L_G < 20$ nm

B. Scaling study of Si and Ge-based NWFETs

Due to their superior electrostatic integrity, GAA devices allow further shortening of the printed gate length without compromising their low leakage current – at least until the onset of source-drain tunneling. To investigate the limits of gate length scaling, Si NMOS and PMOS NWFETs with 6 nm diameter, 7 nm spacer length and following gate lengths were simulated: 15 nm, 11.8 nm, 8.9 nm, 6.7 nm, 5 nm (33 % decrements). Channel orientation was set to ⟨110⟩. Intrinsic channel and 10^{20} cm^{-3} source/drain-doping with abrupt junctions between gate and spacers were assumed. Two-band and six-band **k**·**p**-models were used for the conduction and valence band structure, respectively. Acoustic, optical, inter-valley phonon scattering, surface roughness scattering, and Coulomb scattering off impurities were included. The sub-threshold slopes (SS) were recorded at $|I_D| = 0.1$ nA. The results shown in Fig. 5 for Si-based NMOS and PMOS show no effect at $L_G = 15$ nm, slowly progressing to just below 100 mV at $L_G = 5$ nm.

The situation is however different when germanium is used as PMOS channel material. Holes in Ge are known to have significantly lower effective masses than in Si, which is beneficial for boosting on-current but also increases the tunneling TC, adversely affecting off current. For a quantitative assessment, we replaced Si with Ge in the PMOS devices mentioned above while performing the same scaling analysis. The results show that tunneling has a much more severe impact in Ge-base PMOS NWFETs, as shown in Fig. 6. Source-drain tunneling is so severe that it results in unacceptable SS below a printed gate lenght of 20 nm. For a more realistic device with diffused junctions and gate overlap, the minimum required L_G might be even higher.

IV. Conclusion

We presented a methodology for including intra-band source/drain-tunneling in semi-classical transport based on WKB. The methodology is shown to be of comparable accuracy to ballistic NEGF for MOSFET devices while leveraging the advanced transport modeling capabilities of the subband-BTE framework. We showed that while source/drain-tunneling is not a serious concern in Si-based devices, Ge-based PMOS NWFETs suffer from increased tunneling leakage prohibiting L_G-scaling below 20 nm.

References

[1] L. Witters, F. Sebaai, A. Hikavyy, A. P. Milenin, R. Loo, A. De Keersgieter, G. Eneman, T. Schram, K. Wostyn, K. Devriendt, A. Schulze, R. Lieten, S. Bilodeau, E. Cooper, P. Storck, C. Vrancken *et al.*, "Strained germanium gate-all-around pmos device demonstration using selective wire release etch prior to replacement metal gate deposition," in *2017 VLSIT*, June 2017, pp. T194–T195.

[2] H. Arimura, L. Witters, D. Cott, H. Dekkers, R. Loo, J. Mitard, L. . Ragnarsson, K. Wostyn, G. Boccardi, E. Chiu, A. Subirats, P. Favia, E. Vancoille, V. De Heyn, D. Mocuta, and N. Collaert, "Performance and electrostatic improvement by high-pressure anneal on si-passivated strained ge pfinfet and gate all around devices with superior nbti reliability," in *2017 VLSIT*, June 2017, pp. T196–T197.

[3] S. Jin, A. Pham, and Y. Nishizawa, "Performance evaluation of ingaas, si, and ge nfinfets based on coupled 3d drift-diffusion/multisubband boltzmann transport equations solver," in *2014 IEDM*, Dec 2014, pp. 7.5.1–7.5.4.

[4] Z. Stanojević, M. Karner, O. Baumgartner, H. W. Karner, C. Kernstock, H. Demel, and F. Mitterbauer, "Phase-space solution of the subband Boltzmann transport equation for nano-scale TCAD," in *2016 SISPAD*, Sept 2016, pp. 65–67.

[5] Global TCAD Solutions, "Nano Device Simulator," http://www.globaltcad.com/nds.

[6] M. Karner, O. Baumgartner, Z. Stanojević, F. Schanovsky, G. Strof, C. Kernstock, H. W. Karner, G. Rzepa, and T. Grasser, "Vertically stacked nanowire MOSFETs for sub-10nm nodes: Advanced topography, device, variability, and reliability simulations," in *2016 IEDM*, Dec 2016, pp. 30.7.1–30.7.4.