



Geometrical variability impact on the gate tunneling leakage mechanisms in FinFETs^{☆,☆☆}

C. Medina-Bailon^{ID*}, J.L. Padilla, L. Donetti^{ID}, C. Navarro, C. Sampedro^{ID}, F. Gamiz

Nanoelectronics Research Group and CITIC, Universidad de Granada, Granada, Spain

ARTICLE INFO

Keywords:

Geometrical variability
Gate leakage mechanism
Direct oxide tunneling
Trap assisted tunneling
Leakage current
MS-EMC
FinFET

ABSTRACT

Given the critical role that quantum tunneling effects play in the behavior of nanoelectronic devices, it is essential to investigate the influence and restraints of these phenomena on the overall transistor performance. In this work, a previously developed gate leakage model, incorporated into an in-house 2D Multi-Subband Ensemble Monte Carlo simulation framework, is employed to analyze the leakage current flowing across the gate insulator. The primary objective is to evaluate how variations in key geometrical parameters (specifically, gate oxide and semiconductor thicknesses dimensions) affect the magnitude and bias dependence of tunneling-induced leakage. Simulations are performed on a representative FinFET structure, and the results reveal that tunneling effects become increasingly pronounced at low gate voltages in devices with thinner oxides and thicker semiconductor thickness. These findings underscore the relevance of incorporating quantum tunneling mechanisms in predictive modeling of advanced transistor architectures.

1. Introduction

Since transistors have reached the nanoscale regime, the description of quantum tunneling is now an absolute necessity due to the expected alterations to appear in the device performance [1]. In particular, the use of ultra-thin insulators, where a high electric field is induced, increases the probability of tunneling through the gate oxide, modifying the electrical characteristics and thus the device reliability [2]. These tunneling phenomena are labeled as gate leakage mechanisms (GLM) [3] and involve both tunneling directions (from/to metal gate to/from semiconductor) as well as direct (DT) and elastic/inelastic trap assisted tunneling (TAT).

This study aims to thoroughly investigate the influence of the device geometry on these effects in silicon and, subsequently, on the device performance. In that sense, it is worth highlighting that this variability analysis is associated with those structural and geometrical device parameters whose variation affects these GLM instead of those connected to conventional sources of degradation. In particular, we assess the roles of: (i) the semiconductor thickness (T_{Si}), as it modifies the charge distribution along the confinement direction (altering the

particles position with respect to the insulator-semiconductor interface); (ii) the oxide thickness (T_{SiO_2}), due to its direct impact on the tunnel probability; and (iii) the gate length (L_G), as it affects the overall short channel effects in the transport direction. This has been achieved by simulating for several drain voltages a wide number of different geometrical configurations of an ultrascaled FinFET device using a 2D Multi-Subband Ensemble Monte Carlo (MS-EMC) tool.

2. Methodology

The simulation frame is based on a 2D MS-EMC code [4] based on the mode-space approach for quantum transport (Fig. 1). The simulator solves the 1D Schrödinger equation in the confinement direction and the 2D Boltzmann Transport Equation (BTE) using the Monte Carlo method in the transport plane. The system is coupled by solving Poisson equation in the 2D simulation domain. In general, the main advantages of this simulation scheme with respect to purely quantum approaches are the affordable computational time and its modular nature, which

[☆] This work was supported in part by MICIU/AEI under projects PID2020-119668GB-I00 and PLEC2022-00938, +QCHIP TSI-069100-2023-0003 (Secretaría de Estado de Telecomunicaciones e Infraestructura Digital, Ministerio de Transformación Digital), HORIZON-JU-Chips-2023-RIA-CPL-2 (FAMES ID 101182279), and Juan de la Cierva Incorporación Fellowship scheme under grant agreement No. IJC2019-040003-I. Funding for open access charge: Universidad de Granada / CBUA.

^{☆☆} The review of this paper was arranged by Kuniyuki Kakushima.

* Corresponding author.

E-mail address: cmedba@ugr.es (C. Medina-Bailon).

Table 1

Device parameters and biases considered in this work for the FinFET simulations.

Fixed parameters		Varying parameters	
L_S and L_D (nm)	10	L_G (nm)	5, 7.5, 10, 12.5, 15
Distance from Si-SiO ₂ affected by GLM (nm)	1	T_{Si} (nm)	3, 4, 5
D_{Traps} (cm ⁻²)	$5 \cdot 10^{12}$	T_{SiO_2} (nm)	0.8, 0.9, 1
WF (eV)	4.385	V_{GS} (V)	[-1.5, 1.5] with $\Delta V_{GS} = 0.1$ V
Abrupt N_D and N_S doping (cm ⁻³)	10^{20}	V_{DS} (V)	0.1, 0.5, 1

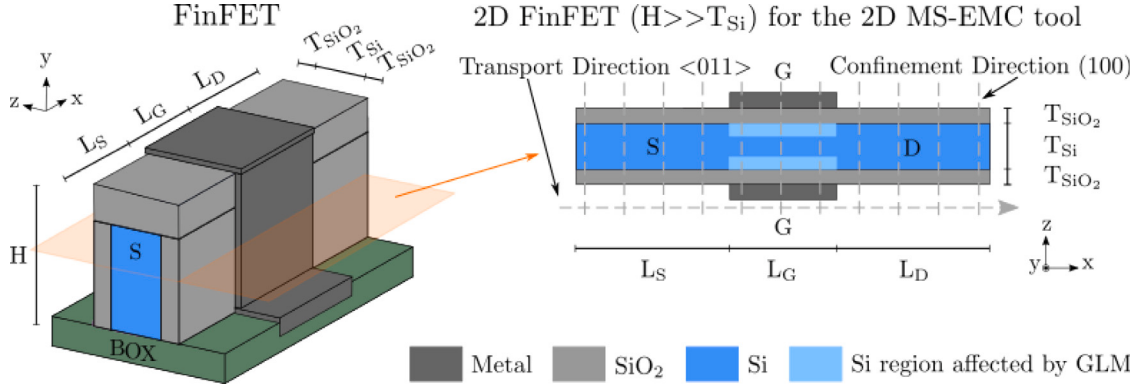


Fig. 1. Schematic FinFET device herein analyzed with confinement and transport directions (011) and <011>, respectively, and all the constant and varying geometrical parameters (see Table 1). Although FinFET is a 3D structure, it can be studied in a 2D approach considering high aspect ratio fins ($H \gg T_{Si}$) [6]. In this 2D system, x and z are the transport and confinement directions, respectively; whereas y corresponds to the infinite direction. The 1D Schrödinger equation is solved for each grid point in the transport direction, and BTE is solved by the MC method in the transport plane.

facilitates the inclusion of novel mechanisms. This tool has previously demonstrated its effectiveness in modeling electron transport, scattering, and other types of tunneling [5].

This simulation flow includes a dedicated module for self-consistency incorporating GLM [5,7], which takes into account the description of DT and both elastic and inelastic TAT events (regardless of the direction of the tunnel), their low occurrence (executing this module in larger time intervals than the generic MS-EMC loop), and its stochastic nature (by random determination of the GLM outcome of each particle). All the tunneling probabilities are computed using the WKB approximation and the trap occupation obeys the Pauli exclusion principle. The latter fact directly affects the WKB tunneling probability as it is scaled by the trap occupation factor-set to 0, 0.5, or 1-depending on whether the trap is empty, partially filled, or fully occupied, respectively. For both DT and TAT, we consider that the tunneling time of particles inside the oxide is negligible due to the narrowness of the dielectric layer. In addition, the number of traps is deterministically calculated according to the oxide dimensions and the trap density, whereas their location and energy (within a typical range of 2.9 eV to 3.9 eV below the conduction band edge of SiO₂) is randomly reckoned. In this work, the trap density (D_{Traps}) has been fixed at $5 \cdot 10^{12} \text{cm}^{-2}$ [8], a medium/high value for a good-quality gate oxide when the dielectric is SiO₂ and the wafer orientation is (100). For comparison purposes, the spatial location of the traps as well as their energies are set to be identical when simulating the same oxide geometry.

There are two possible scenarios for the tunnel direction. On one side, a particle can tunnel from the channel when it is situated close to the insulator-semiconductor interface and its energy level is above that of the Fermi level in the metal (E_{FM}). Available states at any given energy above E_{FM} in each gate electrode are assumed considering that, after tunneling, the electrons thermalize. In this work, the distance from the Si-SiO₂ interfaces at which particles can be considered to tunnel from the semiconductor to the metal has been set at 1 nm for each gate. It is important to note that, due to the 2D description of the MS-EMC

code, the estimation of particles near this interface is probabilistically assigned based on their statistical distribution across the confinement direction. On the other side, the charge injected into the MS-EMC loop from the metal is determined based on the device geometry and the GLM time interval, both of which are incorporated into the calculation of the electron tunneling current density [9]. In this case, the final state of the particle in the device is determined based on the energy alignments: it can be a non-full trap, if its energy matches that of the Fermi level in the metal; or the semiconductor, if there is a subband energy in it below the Fermi level in the metal. In the latter scenario, the particles will be injected at E_{FM} into the subband that minimizes their kinetic energy.

This variability study has been conducted on a FinFET device, sketched in Fig. 1, with confinement and transport directions (011) and <011>, respectively. It is important to note that, although it is a 3D structure, it has been demonstrated that FinFETs with high aspect ratio fins ($H \gg T_{Si}$) can be studied in a 2D approach [6], which justifies the use of the 2D in-house simulator in this work. The constant as well as the varying geometrical parameters and biases are summarized in Table 1. Since the gate length and silicon thickness dimensions chosen in this work correspond to ultrascaled devices, the simulation scheme also includes the description of additional quantum phenomena to properly describe the electronic behavior. In particular, the source to drain tunneling module within the MS-EMC tool has been enabled due to the well-documented degradation of the subthreshold behavior in sub-10 nm nodes [5,10]. Moreover, the realistic conduction band structure of 2D devices has been taken into consideration by the effective electron masses for these confinement and transport directions extracted by first principle simulations [11].

3. Results

The inclusion of GLM in a simulation inherently implies the modification of the charge distribution since electrons are injected/extracted to/from the semiconductor. Fig. 2 illustrates the average number of

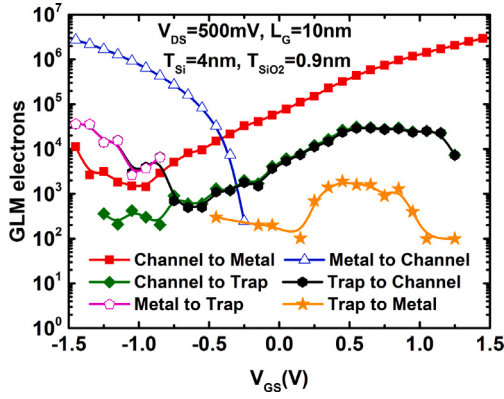


Fig. 2. Average number of injected electrons per MC time step ($t_n = 1\text{fs}$) due to GLM vs. V_{GS} , including both direct and trap assisted tunneling, for a FinFET device with $V_{DS} = 500\text{ mV}$, $L_G = 10\text{ nm}$, $T_{Si} = 4\text{ nm}$, and $T_{SiO_2} = 0.9\text{ nm}$.

electrons per femtosecond (fs) participating in the GLM processes for a particular scenario, demonstrating that the primary tunneling direction is from the metal (semiconductor) at negative (positive) gate voltage, respectively. In any case, the DT contribution is significantly larger than that of the TAT.

The total net current from channel to gate (Fig. 3) highlights the specific impact of these tunneling mechanisms, showing the same maximum level (around $1\text{--}2\text{ }\mu\text{A}/\mu\text{m}$ in absolute value) as the average number of injected electrons by both DT events are equivalent for the lowest/highest V_{GS} (Fig. 2). This net current though the $\text{SiO}_2\text{--Si}$ interface has been computed following the tunnel direction of electrons injected into the device (DT from the metal and TAT from the

oxide traps) minus the ones extracted from it (DT and TAT from the semiconductor). By doing so, the particles that are incorporated into the simulation flow (due to the tunnel from metal) increase the device current, whereas the ones leaving the semiconductor (because of the tunnel in the opposite direction) reduce it. Accordingly, this fact explains the sign of the net gate current being positive/negative for negative/positive V_{GS} , as the energy alignment between the semiconductor subbands and the Fermi level in the metal prioritizes the GLM from the metal/semiconductor, respectively.

Regarding the impact of the silicon thickness (Fig. 3.(a)), it is worth highlighting that two different trends can be observed as a function of the gate voltage. First, the GLM are more deleterious for wider silicon layers when $V_{GS} < 0\text{V}$, which can be explained by the substantial increase in the electron density due to the GLM, as illustrated by its percentage change (Fig. 4). To properly show the charge density variation in the oxide traps (and to avoid the division by zero), this percentage change is calculated as the difference between simulations with and without GLM divided by the first one. This fact is related to the energy modification when T_{Si} shrinks, making the energy difference between the Fermi level in the metal and the lowest subband larger, and thus reducing the electron tunneling current density from the metal. Second, although there is a decline in the electron density near the Si--SiO_2 interface as T_{Si} is reduced (so less particles are available for GLM from the channel), the opposite tendency can be observed for $V_{GS} > 0\text{V}$. This is explained by the fact that the DT from the semiconductor (recall that this was the dominant phenomena for positive bias according to Fig. 2) depends on the WKB probability along the confinement direction, which increases for lower and narrower potential barriers. Therefore, when T_{Si} is reduced for $V_{GS} > 0\text{V}$, the rise of the WKB probability leads to a higher number of particles undergoing this effect. On the other side, the impact of the oxide thickness (Fig. 3.(b)) is mainly due to the WKB probability itself since smaller tunnel paths give rise to larger

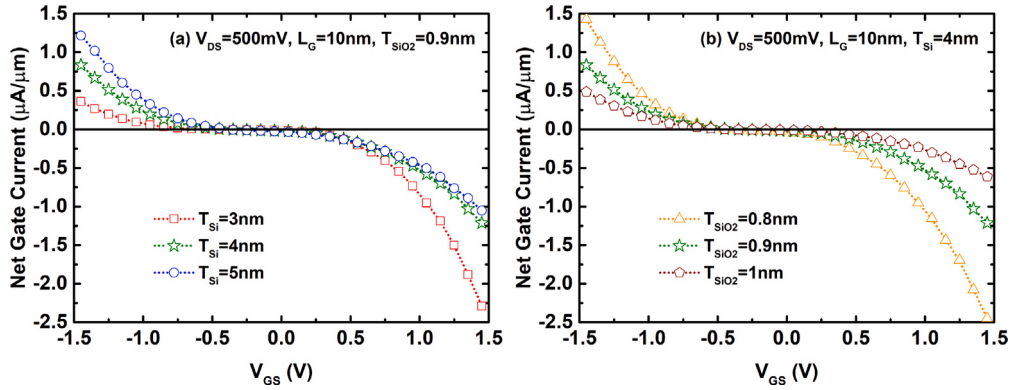


Fig. 3. Net Gate Current vs. V_{GS} for a FinFET device with $V_{DS} = 500\text{ mV}$ and $L_G = 10\text{ nm}$ as a function of (a) T_{Si} (fixed $T_{SiO_2} = 0.9\text{ nm}$) and (b) T_{SiO_2} (fixed $T_{Si} = 4\text{ nm}$). The prevailing tunneling type (to/from channel) determines the current direction (positive/negative).

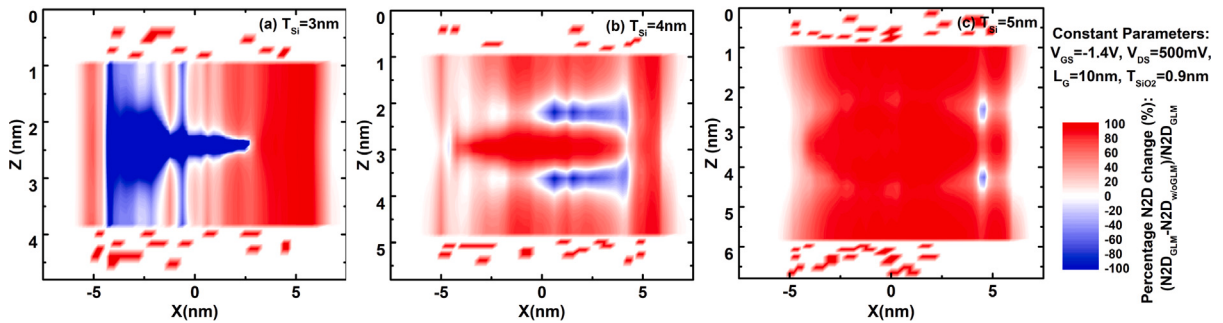


Fig. 4. Percentage change of the electron density (N2D), computed from simulations with and without GLM, along the transport (X) and confinement (Z) directions as a function of T_{Si} ((a) 3 nm , (b) 4 nm , and (c) 5 nm) for a FinFET with $V_{DS} = 500\text{ mV}$, $V_{GS} = -1.4\text{ V}$, $L_G = 10\text{ nm}$, and $T_{SiO_2} = 0.9\text{ nm}$. $X=0\text{ nm}$ corresponds to the device center.

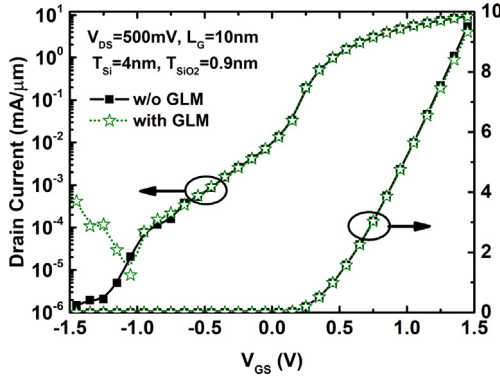


Fig. 5. I_D vs. V_{GS} with and without the GLM module for a FinFET device with $V_{DS} = 500$ mV, $L_G = 10$ nm, $T_{Si} = 4$ nm, and $T_{SiO_2} = 0.9$ nm.

probability values, thus making the GLM more detrimental for thinner oxide thicknesses.

This net gate current appearance has a direct impact on the I_D vs. V_{GS} characteristics, as depicted in Fig. 5 for the same FinFET configuration as that of Fig. 2. It is important to highlight that, despite the aforementioned fact that the maximum net currents for the lowest and highest gate voltages are comparable, the GLM impact is only noticeable in the OFF state, where the drain current has a similar magnitude as the net gate current. The slight decrease in the drain current for the ON state demonstrates that the reduction in particles due to the GLM has almost no impact on the overall drain current calculation.

The same reasoning regarding the geometrical impact on the device performance as that shown in Fig. 3 can be extended leading to Figs.

6 and 7, which show the minimum and maximum values, respectively, that the drain current can achieve in these simulations. These quantities have been chosen for each simulation considering the minimum possible drain current above the inherent stochastic noise region shown in all the Monte Carlo simulations (but selecting the same V_{GS} for the subtraction between the simulations with and without GLM) and the highest V_{GS} (i.e. $V_{GS} = 1.5$ V). Results are expressed as their percentage change for the different geometrical configurations, being computed as the difference between simulations with and without GLM divided by the first to be consistent with the results presented in Fig. 4. As can be extracted from Fig. 5, the percentage variation is much higher for the minimum current, reaching 100% in some geometrical extreme cases, compared to the change in the maximum current (which is negative due to the slightly lower drain current when considering GLM). Finally, regarding the L_G impact, it is important to highlight the reported opposite trend that can be observed due to the scaling effects: for ΔI_{MIN} , the percentile increases when the gate length decreases; whereas for ΔI_{MAX} , the percentile is more negative as L_G rises.

4. Conclusions

In this work, we have thoroughly investigated the impact of geometrical variability on GLM in ultrascaled FinFET devices using a 2D MS-EMC simulation tool. The results demonstrate that tunneling effects (both direct tunneling and trap-assisted tunneling) exert a significant influence on device performance, particularly at low gate voltages. Our analysis reveals that thinner oxide layers and increased semiconductor thickness enhance the detrimental impact of GLM. In particular, the gate leakage becomes more pronounced under negative gate voltages due to higher injection from the gate, whereas at positive gate voltages, leakage from the semiconductor dominates. Additionally, although the influence of GLM on the ON-state current is minimal, the OFF-state performance is considerably degraded,

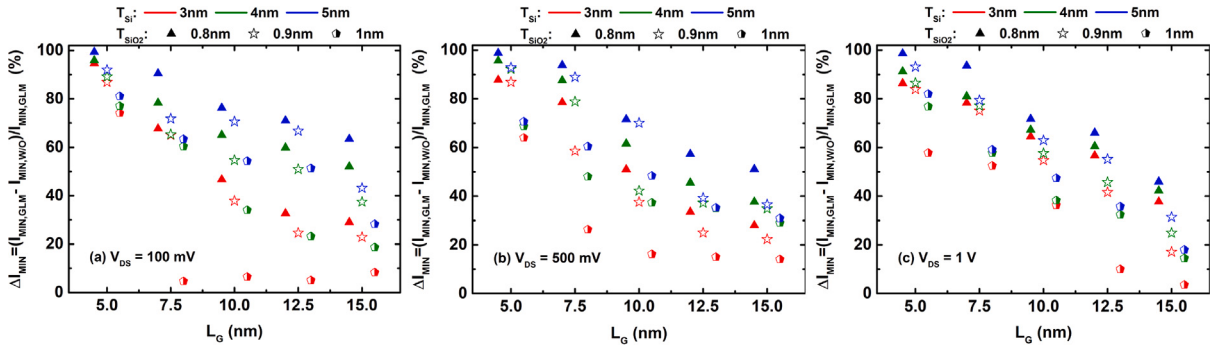


Fig. 6. Percentage change of the minimum drain current (ΔI_{MIN}), computed from simulations with and without GLM, vs. L_G for a FinFET with T_{SiO_2} and T_{Si} combinations as a function of V_{DS} ((a) 100 mV, (b) 500 mV, and (c) 1 V).

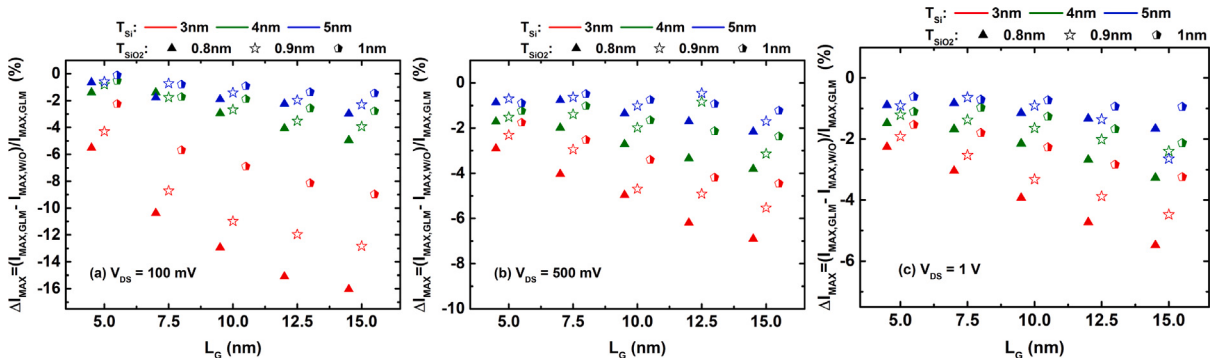


Fig. 7. Percentage change of the maximum drain current (ΔI_{MAX}), computed from simulations with and without GLM, vs. L_G for a FinFET with T_{SiO_2} and T_{Si} combinations as a function of V_{DS} ((a) 100 mV, (b) 500 mV, and (c) 1 V).

with the minimum drain current experiencing substantial variations depending on the device geometry. Furthermore, we observed that the gate length plays a critical role: shorter gate lengths amplify the effects of GLM on the minimum current, whereas longer gate lengths slightly reduce the maximum current due to quantum confinement and source-to-drain tunneling interactions. These findings underscore the importance of accounting for geometrical variability when modeling and designing next-generation nanoscale FinFETs, especially in relation to gate leakage and its implications for power consumption and device reliability.

CRediT authorship contribution statement

C. Medina-Bailon: Writing – original draft, Software, Project administration, Methodology, Funding acquisition, Formal analysis. **J.L. Padilla:** Writing – review & editing, Methodology, Formal analysis. **L. Donetti:** Writing – review & editing, Methodology, Formal analysis. **C. Navarro:** Writing – review & editing, Methodology, Formal analysis. **C. Sampedro:** Writing – review & editing, Project administration, Methodology. **F. Gamiz:** Writing – review & editing, Project administration, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- [1] Querlioz D, Nguyen H, Saint-Martin J, Bournel A, Galdin-Retailleau S, Dollfus P. Wigner-Boltzmann Monte Carlo approach to nanodevice simulation: from quantum to semiclassical transport. *J Comput Electron* 2009;8:324–35. <http://dx.doi.org/10.1007/s10825-009-0281-3>.
- [2] Padilla JL, et al. Gate leakage tunneling impact on the InAs/GaSb heterojunction electron–Hole Bilayer tunneling field-effect transistor. *IEEE Trans Electron Devices (TED)* 2018;65(10):4679–86. <http://dx.doi.org/10.1109/TED.2018.2867721>.
- [3] Roy K, Mukhopadhyay S, Mahmoodi-Meimand H. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. *Proc IEEE* 2003;91(2):305–27. <http://dx.doi.org/10.1109/JPROC.2002.808156>.
- [4] Sampedro C, et al. Multi-subband ensemble Monte Carlo simulator for nanodevices in the end of the roadmap. In: Lirkov I, Margenov S, editors. Large-scale scientific computing. Lecture notes in computer science, vol. 11958, 2019, p. 438–45. http://dx.doi.org/10.1007/978-3-030-41032-2_50.
- [5] Medina-Bailon C, et al. Multisubband ensemble Monte Carlo analysis of tunneling leakage mechanisms in ultrascaled FDSOI, DGSOI, and FinFET devices. *IEEE Trans Electron Devices (TED)* 2019;66(3):1145–52. <http://dx.doi.org/10.1109/TED.2019.2890985>.
- [6] Sampedro C, et al. 3D multi-subband ensemble Monte Carlo simulator of FinFETs and nanowire transistors. In: Proceedings 2014 international conference on simulation of semiconductor processes and devices. 2014, p. 21–4. <http://dx.doi.org/10.1109/SISPAD.2014.6931553>.
- [7] Medina-Bailon C, et al. Description of gate-to-channel tunneling leakage mechanism in a 2D Monte Carlo simulator. In: Proceedings 2023 international conference on simulation of semiconductor processes and devices. 2023, p. 309–12. <http://dx.doi.org/10.23919/SISPAD57422.2023.10319641>.
- [8] Medina-Bailon C, et al. Impact of the trap attributes on the gate leakage mechanisms in a 2D MS-EMC nanodevice simulator. In: Proceedings 2018 international conference on numerical methods and applications. Cham: Springer International Publishing; 2018, p. 273–80. http://dx.doi.org/10.1007/978-3-030-10692-8_30.
- [9] Schenk A, Heiser G. Modeling and simulation of tunneling through ultra-thin gate dielectrics. *J Appl Phys* 1997;81(12):7900–8. <http://dx.doi.org/10.1063/1.365364>.
- [10] Wang J, Lundstrom M. Does source-to-drain tunneling limit the ultimate scaling of MOSFETs? In: IEDM tech. dig. 2002, p. 707–10. <http://dx.doi.org/10.1109/IEDM.2002.1175936>.
- [11] Medina-Bailon C, et al. Quantum enhancement of a S/D tunneling model in a 2D MS-EMC nanodevice simulator: NEGF comparison and impact of effective mass variation. *Micromachines* 2020;11(2):204. <http://dx.doi.org/10.3390/mi11020204>, 1–11.