Medina Bailon, C., Padilla, J. L., Sadi, T., Sampedro, C., Godoy, A., Donetti, L., Georgiev, V. P. , Gamiz, F. and Asenov, A. (2019) Multi-subband ensemble Monte Carlo analysis of tunneling leakage mechanisms in ultrascaled FDSOI, DGSOI, and FinFET devices. *IEEE Transactions on Electron Devices*, (doi:10.1109/TED.2019.2890985)

This is the author's final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/178345/

Deposited on: 22 January 2019

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk

Multi-Subband Ensemble Monte Carlo Analysis of Tunneling Leakage Mechanisms in Ultrascaled FDSOI, DGSOI and FinFET Devices

Cristina Medina-Bailon, José L. Padilla, Toufik Sadi, Carlos Sampedro, Andrés Godoy, *Member, IEEE,* Luca Donetti, Vihar Georgiev, Francisco Gámiz, *Senior Member, IEEE,* and Asen Asenov, *Fellow, IEEE.*

***Abstract*—Leakage phenomena are increasingly affecting the performance of nanoelectronic devices, and therefore, advanced device simulators need to include them in an appropriate way. This paper presents the modeling and implementation of direct source-to-drain tunneling (S/D tunneling), gate leakage mechanisms (GLM) accounting for both direct and trap assisted tunneling, and nonlocal band–to–band tunneling (BTBT) phenomena in a multi-subband ensemble Monte Carlo (MS-EMC) simulator along with their simultaneous application for the study of ultrascaled fully depleted silicon on insulator, double-gate silicon on insulator, and FinFET devices. We find that S/D tunneling is the prevalent phenomena for the three devices, and it is increasingly relevant for short channel lengths.**

***Index Terms*—Band–to–band tunneling (BTBT), direct source-to-drain tunneling (S/D tunneling), double-gate silicon on insulator (DGSOI), FinFET, fully depleted silicon on insulator (FDSOI), gate leakage current, Multi–Subband Ensemble Monte Carlo (MS-EMC)**

I. INTRODUCTION

THE aggressive reduction of device dimensions has increased the importance of short-channel effects (SCEs) and leakage mechanisms as relevant agents degrading the device performance and leading, for example, to the loss of gate control over the channel and the increase of the drain influence. The variation of the threshold voltage (V_{th}) as the channel length decreases is one of the main effects that needs study, without losing sight of the fact that SCEs do not only affect V_{th} but also the subthreshold characteristics contributing to off-state current degradation.

The inclusion of additional physical phenomena is thus required in the modeling of new technological nodes, in order to look for solutions to the aforementioned problems [1]–[3]. In the first place, the direct source-to-drain tunneling (S/D tunneling) arises as a downscaling limit because electrons experience a non-negligible probability of going through the narrow potential barrier located between the source and the drain [4], [5]. These electrons increase the drain current and degrade the gate control on the electrostatic performance of the devices. In the second place, the high electric field across the ultra-thin gate insulator leads to the possibility of carriers crossing the dielectric barrier, resulting in substrate-to-gate tunneling through the thin gate oxide [2]. This tunneling mechanism is known as the gate leakage mechanism (GLM) and it accounts for both the direct tunneling (DT) and the trap-assisted tunneling (TAT). In the third place, the last leakage phenomenon modeled in this paper is related to the generation of electron-hole pairs in the depletion region due to band-to-band tunneling (BTBT). In this case, high electric fields across a reverse-biased pn junction (such as the drain-to-channel region in Fig. 1) cause significant currents to flow through the forbidden energy barrier due to tunneling of electrons (resp. holes) from the valence (conduction) band of the p (n) region to empty full states in the conduction (valence) band of the n (p) region, respectively [2].

In this paper, S/D tunneling, GLM, and BTBT models have been implemented in a multi-subband ensemble Monte Carlo (MS-EMC) simulator, and their effects assessed individually as well as jointly on ultra-scaled devices. For that purpose, the selected devices are a single gate fully-depleted silicon-on-insulator (FDSOI) transistor, which has been recognized as an alternative to bulk technology, and two double gate devices: a planar double-gate silicon-on-insulator (DGSOI) transistor and a FinFET. The additional gate increases the electron confinement and, thus, improves the immunity of these devices to SCEs [6]. The main difference between the DGSOI and the FinFET is the gate orientation, which is parallel and perpendicular to the standard wafer orientation, respectively.

The structure of this paper is organized as follows. Section II gives a general overview of the MS-EMC simulation framework, describes the parameters and orientations for the proposed devices, and provides a thorough description of the additional simulation blocks needed for the implementation of the aforementioned tunneling leakage phenomena. Section III outlines the main results and their discussion. Finally, conclusions are drawn in Section IV.

This work was supported in part by the European Community's Seventh Framework Programme Marie Curie Action under Grant Agreement No. 291780 (Programme Andalucía Talent Hub); in part by the Horizon 2020 projects REMINDER (Grant No. 687931), WAYTOGO-FAST (Grant No. 662175), and in part by the Spanish Ministry of Economy, Industry and Competitivity under grant agreements TEC2014-59730-R and TEC2017-89800-R.

C. Medina-Bailon is with the Device Modelling Group, School of Engineering, University of Glasgow, G12 8LT Glasgow, Scotland, U.K., and also with the Nanoelectronics Research Group, Departamento de Electrónica y Tecnología de los Computadores, Universidad de Granada. 18071 Granada, Spain (e-mail: Cristina.MedinaBailon@glasgow.ac.uk).

V. Georgiev and A. Asenov are with the Device Modelling Group, School of Engineering, University of Glasgow, G12 8LT Glasgow, Scotland, U.K.

J.L. Padilla, C. Sampedro, A. Godoy, L. Donetti and F. Gamiz are with the Nanoelectronics Research Group, Departamento de Electrónica y Tecnología de los Computadores, Universidad de Granada. 18071 Granada, Spain.

T. Sadi is with the Department of Neuroscience and Biomedical Engineering, Aalto University, P.O. Box 12200, FI-00076 AALTO, Finland.

II. METHODOLOGY

The starting point of the simulation framework is a 2-D MS-EMC code [7], [8] which is based on the mode-space approach [9]. The system is regarded as decoupled between the confinement direction, where the 1-D Schrödinger equation is solved; and the transport plane, where a solution of the 2-D Boltzmann transport equation (BTE) is obtained using the Monte Carlo method (Fig. 1). Both equations are coupled self-consistently with the Poisson Equation solved in the 2-D simulation domain every time step t_n, as depicted in Fig. 2. This simulation scheme presents two main advantages with respect to the full-quantum approach: the first one is its affordable computational time and the second one is the inclusion of the quantum transport phenomena in a separate manner so that they can be switched ON and OFF to check their effect. Furthermore, despite the fact that the FinFET is a 3-D structure, this code can properly simulate it: a 2-D description (which assumes height much higher than thickness) can be appropriate for a FinFET with a sufficiently high aspect ratio [10].

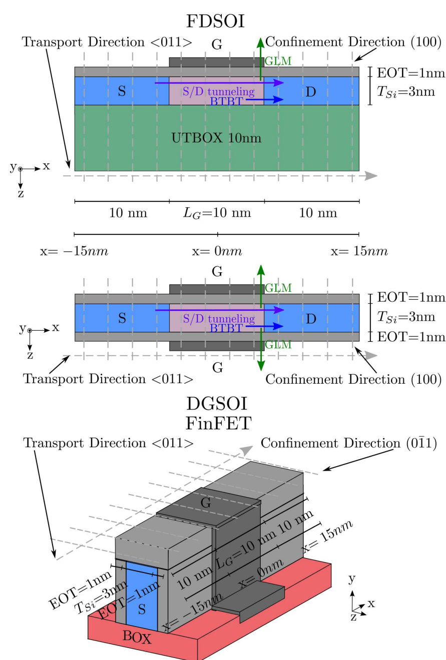

Fig. 1: FDSOI, DGSOI and FinFET structures analyzed in this paper with L_G=10nm. The 1-D Schrödinger equation is solved in the confinement direction for each grid point and the BTE is solved by the MC method in the transport plane.

A. Description of the simulated devices

The simulated FDSOI, DGSOI and FinFET are schematically depicted in Fig. 1 along with their orientations. The considered confinement direction on standard wafers changes from (100) for both planar FDSOI and DGSOI to (0$\bar{1}$1) for FinFETs, whereas the transport direction remains constant to <011> for the three devices. The difference in the confinement direction modifies the corresponding carrier transport and confinement masses, m_x and m_z, respectively [11]. Table

Device	Valley	m_x	m_z
FDSOI & DGSOI	Δ_2	$m_t = 0.198m_0$	$m_l = 0.916m_0$
(100)<011>	Δ_4	$\frac{2m_l m_t}{m_l+m_t} = 0.326m_0$	$m_t = 0.198m_0$
FinFET	Δ_2	$m_t = 0.198m_0$	$m_t = 0.198m_0$
(0$\bar{1}$1)<011>	Δ_4	$\frac{m_l+m_t}{2} = 0.557m_0$	$\frac{2m_l m_t}{m_l+m_t} = 0.326m_0$

TABLE I: Transport (m_x) and confinement (m_z) effective masses in silicon for the FDSOI, DGSOI and FinFET devices herein analyzed.

I summarizes the values of the masses in silicon for each device, where m_l and m_t are the longitudinal and transverse effective masses, m_0 is the free electron mass, and the subindex of Δ represents the degeneracy factor associated with the conduction band valley. It is noteworthy that the change in the confinement direction alters the electron distribution, modifying the lower energy subband from Δ_2 in both FDSOI and DGSOI transistors to Δ_4 in the FinFET.

These devices have been parametrized for gate lengths ranging from 5 to 20nm, whereas the rest of the technological parameters remains constant: a channel thickness T_{Si}=3nm, a SiO_2 gate oxide with an equivalent oxide thickness EOT=1nm, and a metal gate work function of 4.385eV. The additional device parameters for the FDSOI device are a back-plane with an UTBOX of 10nm and work function of 5.17eV, and back-bias polarization V_{BB}=0V.

B. Description of the model

The flowchart of the MS-EMC simulator with the additional blocks for the three tunneling mechanisms is depicted in Fig. 2. It is of note that, on the one hand, S/D tunneling is evaluated every time step t_n, the same as the other blocks involved in the generic MS-EMC loop. On the other hand, the GLM and BTBT blocks are only executed with larger time intervals Δt_{GLM} and Δt_{BTBT}, respectively, and therefore, the corresponding modifications are added to the system when these blocks finish. Fig. 2 also shows the specific stage inside the MS-EMC loop, in which each block is triggered. Both S/D tunneling and GLM are evaluated for each particle after the Monte Carlo flight, whereas BTBT is calculated after the subband profile updating. Let us now analyze the characteristics of each tunneling leakage mechanism.

First, our S/D tunneling model calculates the probability of traversing the potential barrier and, when that happens, it mimics the motion of the affected electron inside the forbidden region [12]. The physical process is described in Fig. 2 and it starts after stochastically determining the new position of the electron in the Monte Carlo procedure. Then, if the particle is located near the potential barrier and its energy (E_{par}) is lower than its maximum (E_{PB}), it would either rebound from it (experiencing backscattering) or traverse the potential barrier via S/D tunneling. In order to decide the fraction of electrons experiencing each phenomenon, the transmission

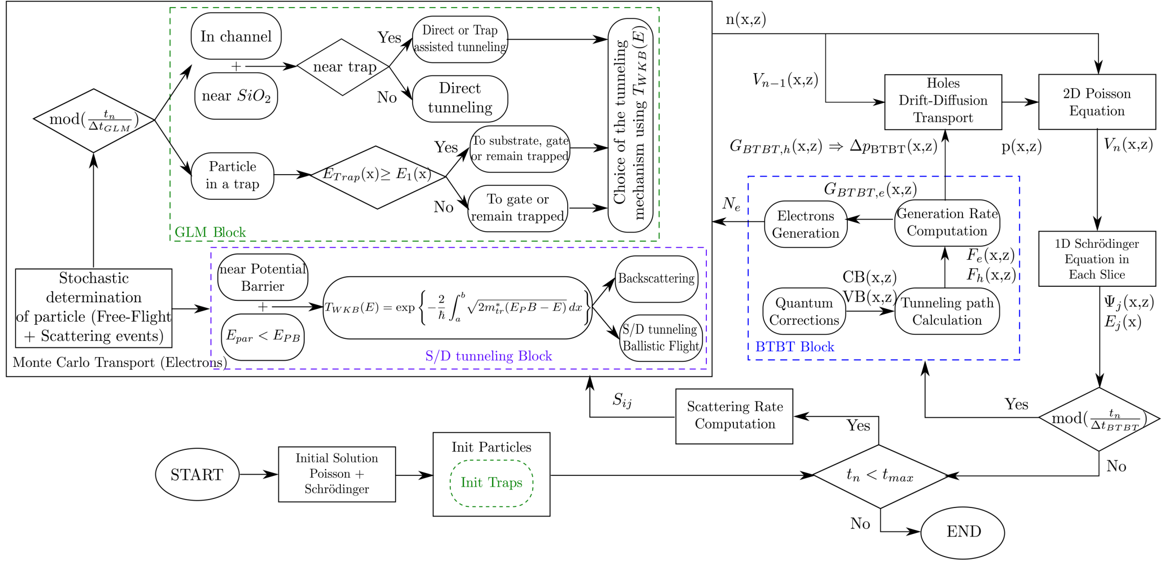

Fig. 2: Flowchart of the MS-EMC simulator with the additional blocks of the three tunneling leakage mechanisms herein implemented. x is the transport direction, z is the confinement direction, $n(x,z)$ and $p(x,z)$ are the electron and hole concentrations, respectively, $V(x,z)$ is the potential profile, $E_j(x)$ is the subband energy, $\Psi_j(x,z)$ are the subband eigenfunctions, S_{ij} are the scattering rates, the subscript n stands for the iteration number, Δt_{GLM} and Δt_{BTBT} are the time steps where GLM and BTBT are calculated, respectively. In the S/D tunneling block: E_{par} is the particle energy, E_{PB} is the potential barrier energy, $T_{WKB}(E)$ is the transmission coefficient for an energy E using the WKB approximation, a and b are the starting and ending points, and m^*_{td} is the effective mass of the electron. In the GLM block: E_{trap}(x) is the trap energy. In the BTBT block: CB(x,z) and VB(x,z) are the conduction and valence bands, respectively, $F_{e/h}(x,z)$ are the electric fields of electrons/holes, $G_{\mathrm{BTBT},e/h}(x,z)$ are the electron/hole generation rates, $\Delta p_{\mathrm{BTBT}}(x,z)$ accounts for the generated holes, and N_e is the number of generated superparticles.

probability (T_{WKB}) is integrated using the Wentzel, Kramers, and Brillouin (WKB) approximation [13]. It depends on the carrier position (starting and ending points along the transport direction) and some specific parameters related to the tunneling phenomena, such as the transport effective mass in the tunneling direction (which corresponds to m_x in Table I) and the band profiles. It is important to highlight in this stage that the particle energy being involved in the tunneling process is the total energy in the transport plane considering only the component of the kinetic energy in the direction that faces the potential barrier. Once the tunneling probability is known, a rejection technique is used to determine whether the electron will undergo backscattering or S/D tunneling. Consequently, a uniformly distributed random number is compared to the probability of tunneling through the barrier at a specific energy. Eventually, the tunneling path is established considering that electrons fly through the potential barrier during a certain period of time, following a ballistic flight inside it, evaluated according to Newton's mechanics in an inverted potential profile. The choice of this tunneling path has already shown its accuracy when compared with a ballistic transport description making use of the nonequilibrium Green's function (NEGF) formalism, especially for the degradation in the subthreshold region [14].

Second, the GLM model has been implemented including DT and both elastic TAT and inelastic TAT [15], [16]. The GLM treatment inside the simulator can be divided into two stages, as is shown in Fig. 2. The first step corresponds to the initialization of the trap-related parameters before the Monte Carlo iterations. In it, the number of traps is deterministically calculated according to the oxide dimensions and the trap density, whereas their location is randomly reckoned. The traps are considered to be neutral with a constant capture cross section of the order of σ_T=10^{-15}cm^2. Once estimated, the trap distribution is set to be identical in the three devices for comparison purposes.

The second stage of the GLM modeling is included inside the Monte Carlo loops after stochastically determining the new position of the electron as depicted in the Monte Carlo Transport block (Fig. 2). At that moment, two possible scenarios are allowed depending on the particle location: electrons can be in the channel or trapped. For both DT and TAT, we consider that the tunneling time of the particle inside the oxide is negligible due to the narrowness of the dielectric layer and the low frequency of this mechanism. On the one hand, if the particle is located in the channel, it is indispensable to know whether it is near the insulator interface or not. As the motion of the particles in our 2-D MS-EMC tool is only known in the transport direction, and considering that the simulated particles are distributed across the whole device, the percentage of those near the interface is estimated with respect to the total number of particles. Then, the choice of a particle position along the confinement direction is randomly calculated. If it is located near the dielectric and near some trap, it can undergo either DT or TAT, whereas if it is located near the dielectric but not of a trap, then it can only experience DT. On the other hand,

once the particle has been trapped, it can either return to the channel [only if the trap energy, E_{Trap}(x), is higher than the lower subband energy, E_1(x)], tunnel to the gate contact, or remain in the trap. Moreover, when the electron is trapped, its charge is dynamically included in the 2-D Poisson solution in order to preserve the self-consistency during the following Δt_{GLM}.

For determining the resulting phenomena in each scenario, the tunneling probabilities for each mechanism are calculated making use of the WKB approximation as in the S/D tunneling process but making use of the confinement effective mass (m_z in Table I). DT probability is directly estimated making use of this approximation, whereas some other considerations are needed for TAT such as the Pauli's exclusion principle. More details about how to calculate tunneling probabilities, including TAT, can be found in [16], [17].

Finally, the BTBT algorithm herein implemented calculates the nonlocal direct and phonon-assisted tunneling considering quantum confinement effects [18]. It is based on the Kane's model which translates the tunneling current into suitable generation rates [G_{BTBT}(x,z)] for both electrons and holes. Its implementation can be divided into several boxes as depicted in Fig. 2. The first box corresponds to the necessary subband corrections through the estimation of the first bound state of the conduction [CB(x,z)] and valence [VB(x,z)] bands. If this correction were not considered, the generated particles could reach forbidden states implying a violation of the energy conservation principle. The next box is related to the tunneling path calculation, which refers to the carrier motion inside the forbidden energy region. This paper computes the path following the valence band maximum gradient trajectory (F_{max}) so that the carriers move following the direction imposed by the electric field. Moreover, this tunneling path is dynamically modified in each simulation step according to the up-to-date electrostatic configuration given by our MS-EMC simulator. More details about the procedure followed to evaluate this F_{max} trajectory can be found in [18] where it is also compared to another trajectory assumption in a silicon-based n-type tunnel FET. After the determination of the starting and ending points for the tunneling process, the electric field is computed by using those two points and the distance between them. It is important to highlight in this stage that, as electrons and holes effectively follow independent paths, the electric field [F_e(x,z) and F_h(x,z), respectively] for both carriers are needed.

Thereupon, both generation rates, $G_{BTBT,e}$(x,z) and $G_{BTBT,h}$(x,z), are calculated as a function of the nonlocal electric field and the updated quantized band profile. The last step is to translate the generation rates into generated charge so that it can be incorporated into the simulation flow. As for holes, since they are described by a drift-diffusion approach, a correction in their concentration, $\Delta p_{\mathrm{BTBT}}(x,z)$, is simply added to account for the generated carriers. On the other hand, a number of superparticles N_e representing electrons are generated in the fundamental subband. The grid cell, in which the generated superparticle emerges once it reaches the conduction band, is calculated according to the generation rate distribution.

The BTBT model developed in this paper represents an improvement with respect to its previous version [18] as to the slice selection where the charge is injected. In the prior version, we only selected one slice and only its associated tunneling charge, resulting from integrating the generation rate across it, was injected in the selected slide. Therefore, we avoided the overestimation that would result if we injected in that slice all the tunneling charge. Nevertheless, in spite of its accuracy, this technique still implied a certain underestimation since the charge corresponding to other slices was systematically neglected. This procedure worked well as far as the charge corresponding to the selected slice resulted to be much higher than any of the others. However, in the case of several slices featuring comparable charge levels, this method needed to be refined. Thus, our improvement in this paper consists of taking several slices and injecting in each one its corresponding charge. We consider all those slices fulfilling the requirement of possessing a charge level above the 10% of the charge corresponding to the most probable slice. Then, the superparticles are injected according to the generation rate distribution. By doing so, we avoid selecting those slices whose injected superparticles would have lower weight, which in turn would increase the computational load without providing additional insight.

III. Results and Discussion

The probability of occurrence for any of the previously mentioned phenomena depends on the specific characteristic of each device, such as the electron distribution or the transport and confinement effective masses. That leads to a different number of electrons experiencing each tunneling event for the considered devices, as shown in Fig. 3.

First, the probability of tunneling through the potential barrier (S/D tunneling) depends on the tunneling path length, on the potential barrier height, and on the transport effective mass according to the WKB approximation. It is important to highlight that S/D tunneling presents the highest number of electrons for all the devices. If we focus on the comparison between devices, the larger transport effective mass of the FinFET in comparison to both SOI devices (Table I) reduces the tunneling probability, whereas the higher and larger energy profile of the DGSOI in comparison to the FDSOI reduces it. As a result, for low gate voltage, the FinFET presents a lower number of particles affected by the S/D tunneling (see Fig. 3). Nevertheless, the number of S/D tunneling particles tends to be very similar in the three devices as the gate voltage increases owing to their very small and narrow potential profile.

Second, as for the GLM, larger geometrical confinement, which tends to concentrate the charge towards the center of the channel, produces a decrease in the number of electrons experiencing this phenomenon. This volume inversion effect is more significant in the FinFET for low gate biases due to the smaller confinement mass. In general, DT is the dominant phenomenon in the GLM for all the devices due to the small oxide thickness.

Third, the average number of electrons generated by BTBT depends on the generation rate calculation which is determined according to the tunneling path. Fig. 4 shows for the

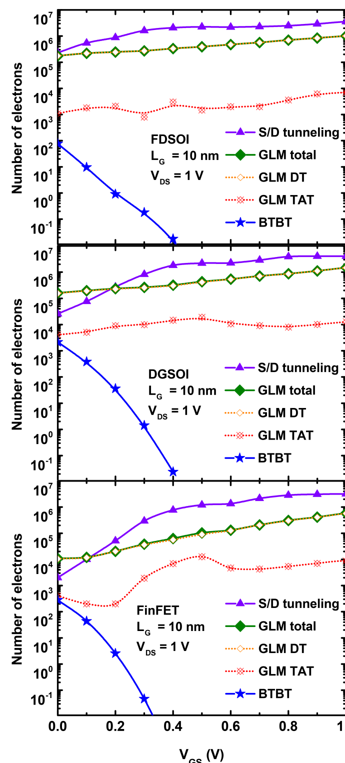

Fig. 3: Average number of electrons generated by S/D tunneling, GLM (DT+TAT), and BTBT as a function of V_{GS} for a gate length of 10nm and V_{DS}=1V.

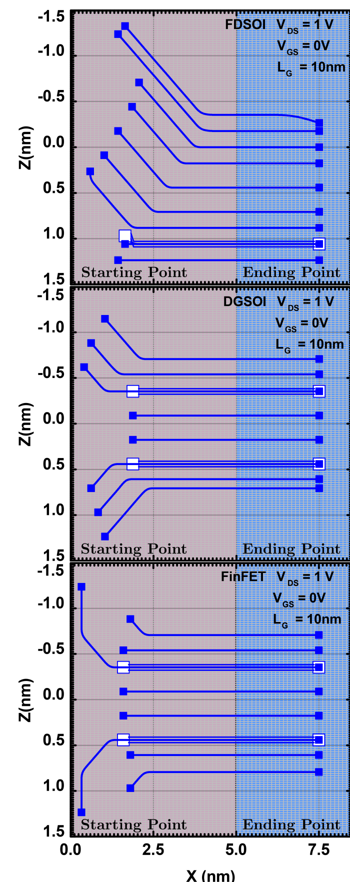

Fig. 4: Some examples of different tunneling paths estimated using the $F_{\max}$ criterion for L_G=10nm for FDSOI (top), DGSOI (middle), and FinFET (bottom) with V_{GS}=0V and V_{DS}=100mV. White symbols stand for tunneling paths whose ending point is shared with another tunneling path with different starting point. X=5nm corresponds to the limit between the channel and the drain.

three devices some examples of tunneling trajectories whose common feature is that they share the same ending point in the x-direction. Notice that, for a given device, different starting points can reach the same ending point (this fact is illustrated by white symbols). The longer BTBT paths in the single gate FDSOI, due to the top-bottom asymmetry in the band curvature, decreases the generation rates compared to the other devices. Accordingly, this device shows the lowest number of electrons injected by BTBT see (Fig. 3). As for the DGSOI and the FinFET, both show symmetrical trajectories with respect to a horizontal axis at the center of the channel. However, the longer tunneling paths in the FinFET due to its lower potential barrier reduce the $G_{BTBT}(x, z)$ and, thus, the number of generated electrons (see Fig. 3). As it can be seen, this phenomenon only has visible influence at low gate biases. This can be explained by the following: as the gate bias increases, the potential barrier decreases leading to a reduction of the BTBT probability caused by longer tunneling paths. This is observed for the DGSOI and the FinFET (see Fig. 3) given that for a very low V_{GS}, the number of electrons generated by BTBT is similar to that corresponding to TAT. This result might be also observable for the FDSOI but for a negative V_{GS} owing to the aforementioned band profile asymmetry.

The impact on the threshold voltage variation (ΔV_{th}) of the three phenomena, separately as well as combined, as a function of the channel length is represented in Fig. 5. This variation quantifies the impact of each mechanism near the threshold region when the devices are scaled down. It has been calculated as the difference between a simulation including one or all mechanisms and another one without any tunneling phenomenon.

Let us now analyze ΔV_{th} when each tunneling has been separately simulated. S/D tunneling increases the drain current at any drain bias due to the contribution of the particles inside the potential barrier. In this scenario, ΔV_{th} is negative and it has a higher impact for higher V_{DS}. On the other hand, GLM and BTBT have no influence on the threshold voltage variation since the particles that leave the device and the generated electron-hole pairs, respectively, are negligible in comparison to the total particles contributing to the drain current in the threshold region.

If we perform simulations including all the mechanisms, we observe that S/D tunneling dominates ΔV_{th} for the three devices at any drain biases, being more important as the devices are scaled down. Moreover, the influence of S/D tunneling is lower in the DGSOI and the FinFET due to better gate control minimizing SCEs.

Another parameter that shows a remarkable description of

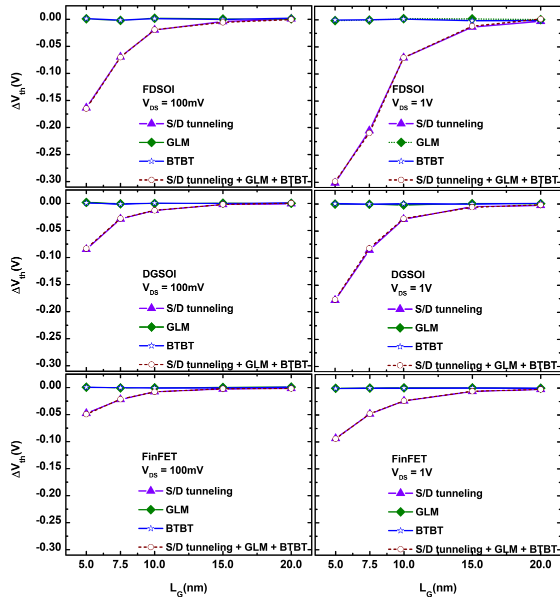

Fig. 5: Threshold voltage variation (ΔV_{th}) as a function of L_G when considering S/D tunneling, GLM, and BTBT individually, as well as simultaneously, for FDSOI (top), DGSOI (middle), and FinFET (bottom) at low drain bias (left) and saturation conditions (right).

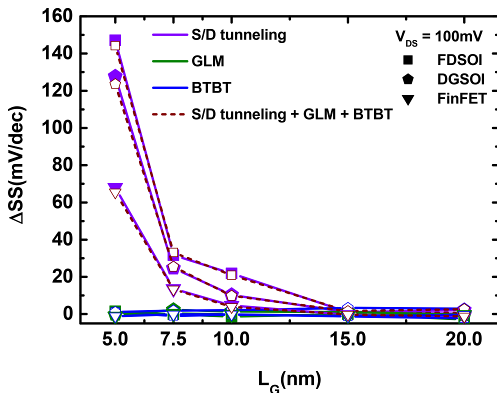

Fig. 6: Subthreshold swing variation (ΔSS) as a function of L_G when considering S/D tunneling, GLM, and BTBT individually, as well as simultaneously, for FDSOI, DGSOI, and FinFET at V_{DS}=100mV.

the device behavior in the subthreshold regime is the subthreshold swing (SS). Fig. 6 shows the SS variation (ΔSS) for a simulation with S/D tunneling, GLM, and BTBT separately as well as combined, and another without any tunneling mechanism for the FDSOI, DGSOI, and FinFET at V_{DS}=100mV. The SS has been computed as an average of SS at each gate bias point in the subthreshold regime within a range of 200mV, where the current approximately varies by several orders of magnitude. As expected, ΔSS is positive due to the higher degradation in the SS for a simulation including different mechanisms. The main findings of this figure are quite similar to those of ΔV_{th} in Fig 5: 1) S/D tunneling is the dominant tunneling leakage progress in the subthreshold region, and 2) this mechanism is more noticeable in FDSOI devices, than in DGSOIs, and least visible in FinFETs.

Fig. 7 shows the $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio as a function of the channel length for each device considering a simulation without any tunneling leakage and others with S/D tunneling, GLM, and BTBT, separately as well as simultaneously combined. In general, the $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio provides the information about the highest (I_{ON}) and lowest (I_{OFF}) attainable currents of the devices, respectively (I_{ON}=I_D when V_{GS}=V_{DS}=1V; I_{OFF}=I_D when V_{GS}=0V and V_{DS}=100mV). Notice how the FinFET features a much higher ratio than the other devices due to its very low I_{OFF} [12].

In general, for the three tunneling phenomena, the I_{ON} current does not exhibit a noticeable difference when these mechanisms are considered because the number of particles involved in each tunnel process is negligible compared with the total contributing to the drain current at high biases. Accordingly, the change in $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio is caused by the variation of I_{OFF}.

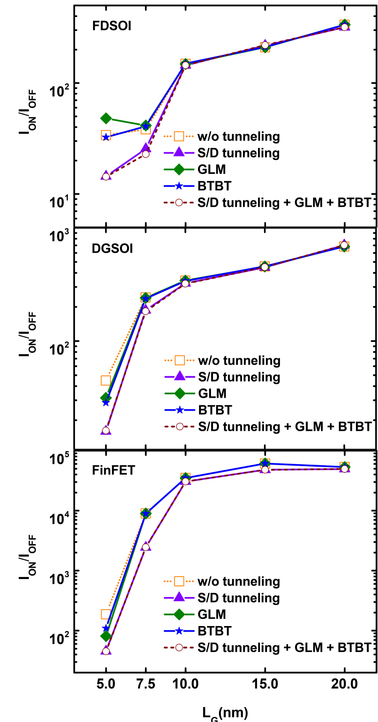

Fig. 7: $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ as a function of L_G considering a simulation w/o any tunneling mechanism, and others with S/D tunneling, GLM, and BTBT separately as well as combined for FDSOI (top), DGSOI (middle), and FinFET (bottom).

Let us analyze the change in I_{OFF} ratio for each mechanism individually. S/D tunneling decreases the $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio, because the particles located close to the potential barrier at low gate bias have an opportunity of contributing to I_{OFF}. When GLM is simulated, two different scenarios arise depending on the device. For the FDSOI, the particles that leave the device get relevance, which leads to the reduction of I_{OFF} and, therefore, an increase the $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio. On the other hand, for the DGSOI and FinFET, the effect of the particles that leave the device proves to be negligible. In consequence, the trapped charge reduces the height of the subband profiles

increasing the amount of carriers that contribute to the drain current. It results on an increase of I_{OFF} and, therefore, a reduction of the $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio. Finally, the influence of the BTBT in this ratio is very low as expected. Due to the reduced generation of electron-hole pairs at low biases, this mechanism slightly increases the I_{OFF} reducing the $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio. As a result of the above reasoning, the behavior of the $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio for the devices analyzed, including the three phenomena, is mainly determined by the prevalent phenomenon: the S/D tunneling.

IV. Conclusions

The aim of this paper is the implementation of S/D tunneling, GLM including DT and TAT, and nonlocal BTBT phenomena in an existing MS-EMC tool for the analysis of their separate and combined effects on ultrascaled FDSOI, DGSOI and FinFET devices. In general, S/D tunneling is the dominant phenomenon in the three devices due to the particles located inside potential barrier. GLM is noticeable in the OFF-state current degradation owing to the particles that leave the device in the case of the FDSOI and to the trapped charge in the gate oxide in both double gate devices. BTBT has a negligible impact on these devices, because the injected charge does not modify the electrostatics. Finally, the FinFET shows lower degradation compared to the other devices due to its larger geometrical confinement and transport effective mass. These conclusions shed light on the impact of the main tunneling mechanisms on the performance of ultrascaled FET devices.

References

[1] M. Ieong, B. Doris, J. Kedzierski, K. Rim, and M. Yang, "Silicon device scaling to the sub-10-nm regime," *Science*, vol. 306, no. 5704, pp. 2057 – 2060, 2004, DOI: 10.1126/science.1100731.

[2] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI devices*. New York: Cambridge University Press, 2009.

[3] Y. B. Kim, "Challenges for nanoscale MOSFETs and emerging nanoelectronics," *Transactions on Electrical and Electronic Materials*, vol. 11, no. 3, pp. 93 – 105, 2010, DOI: 10.4313/TEEM.2010.11.3.093.

[4] J. W. J. Wang and M. Lundstrom, "Does source-to-drain tunneling limit the ultimate scaling of MOSFETs?" *Digest. International Electron Devices Meeting,*, pp. 707–710, 2002, DOI: 10.1109/IEDM.2002.1175936.

[5] H. Iwai, "Future of nano CMOS technology," *Solid-State Electronics*, vol. 112, pp. 56–67, March 2015, DOI: 10.1016/j.sse.2015.02.005.

[6] S. Cristoloveanu, "How many gates do we need in a transistor?" 2007 International Semiconductor Conference, 2007, pp. 3–10, DOI: 10.1109/SMICND.2007.4519636.

[7] C. Sampedro, F. Gámiz, A. Godoy, R. Valín, A. García-Loureiro, and F. G. Ruiz, "Multi-Subband Monte Carlo study of device orientation effects in ultra-short channel DGSOI," *Solid-State Electronics*, vol. 54, no. 2, pp. 131–136, 2010, DOI:10.1016/j.sse.2009.12.007.

[8] C. Sampedro, F. Gámiz, and A. Godoy, "On the extension of ET-FDSOI roadmap for 22 nm node and beyond," *Solid-State Electronics*, vol. 90, pp. 23–27, 2013, DOI: 10.1016/j.sse.2013.02.057.

[9] R. Venugopal, Z. Ren, S. Datta, M. S. Lundstrom, and D. Jovanovic, "Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches," *Journal of Applied Physics*, vol. 92, no. 7, pp. 3730–3739, 2002, DOI: 10.1063/1.1503165.

[10] C. Sampedro, L. Donetti, F. Gámiz, and A. Godoy, "3D Multi-Subband Ensemble Monte Carlo Simulator of FinFETs and Nanowire Transistors." 2014 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), 2014, pp. 21–24, DOI: 10.1109/SISPAD.2014.6931553.

[11] A. Rahman, M. S. Lundstrom, A. W. Ghosh, A. Rahman, M. S. Lundstrom, and A. W. Ghosh, "Generalized effective-mass approach for n-type metal-oxide-semiconductor field-effect transistors on arbitrarily oriented wafers," *Journal of Applied Physics*, vol. 97, no. 053702, 2005, DOI: 10.1063/1.1845586.

[12] C. Medina-Bailon, J. Padilla, C. Sampedro, A. Godoy, L. Donetti, and F. Gámiz, "Source–to–Drain Tunneling Analysis in FDSOI, DGSOI and FinFET Devices by Means of Multi-Subband Ensemble Monte Carlo," *IEEE Transactions on Electron Devices*, vol. 65, no. 11, pp. 4740 – 4746, 2018.

[13] D. J. Griffiths, "The WKB approximation," in *Introduction to Quantum Mechanics*. New Jersey: Prentice Hall, 1995, ch. 8, pp. 274–297.

[14] C. Medina-Bailon, C. Sampedro, J. L. Padilla, A. Godoy, L. Donetti, F. Gamiz, and A. Asenov, "MS-EMC vs. NEGF: A comparative study accounting for transport quantum corrections." EUROSOI workshop and international conference on Ultimate Integration on Silicon (EUROSOI-ULIS), 2018, pp. 1–4, DOI: 10.1109/ULIS.2018.8354758.

[15] C. Medina-Bailon, T. Sadi, C. Sampedro, A. Godoy, L. Donetti, V. Georgiev, F. Gámiz, and A. Asenov, "Assessment of Gate Leakage Mechanism utilizing Multi-Subband Ensemble Monte Carlo." 2017 Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSOI-ULIS), 2017, DOI: 10.1109/ULIS.2017.7962585.

[16] C. Medina-Bailon, T. Sadi, C. Sampedro, J. Padilla, L. Donetti, V. Georgiev, F. Gámiz, and A. Asenov, "Impact of the Trap Attributes on the Gate Leakage Mechanisms in a 2D MS-EMC Nanodevice Simulator," *Lecture Notes in Computer Science post-proceedings, Accepted for publication*, 2018.

[17] T. Sadi, A. Mehonic, L. Montesi, M. Buckwell, A. Kenyon, and A. Asenov, "Investigation of resistance switching in SiOx RRAM cells using a 3D multi-scale kinetic Monte Carlo simulator," *Journal of Physics: Condensed Matter*, no. 8, p. 084005.

[18] C. Medina-Bailon, J. Padilla, C. Sampedro, C. Alper, F. Gámiz, and A. Ionescu, "Implementation of Band–to–Band Tunneling Phenomena in a Multi–Subband Ensemble Monte Carlo Simulator: Application to Silicon TFETs," *IEEE Transactions on Electron Devices*, vol. 64, no. 8, pp. 3084 –3091, 2017, DOI: 10.1109/TED.2017.2715403.