University of Glasgow

Medina-Bailon, C., Sadi, T., Sampedro, C., Padilla, J.L., Godoy, A., Donetti, L., Georgiev, V. , Gamiz, F. and Asenov, A. (2017) Assessment of Gate Leakage Mechanism Utilizing Multi-Subband Ensemble Monte Carlo. In: 2017 Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSOI-ULIS), Athens, Greece, 03-05 Apr 2017, pp. 144-147. ISBN 9781509053148.

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/144279/

Deposited on: 14 July 2017

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

Assessment of Gate Leakage Mechanism utilizing Multi-Subband Ensemble Monte Carlo

C. Medina-Bailon*†, T. Sadi‡, C. Sampedro†, J.L. Padilla†, A. Godoy†, L. Donetti†, V. Georgiev‡, F. Gamiz† and A. Asenov‡

† Nanoelectronics Research Group, Universidad de Granada, 18071 Granada, Spain.

‡ School of Engineering, University of Glasgow, Glasgow G12 8LT, Scotland, UK.

* e-mail: cmedba@ugr.es

***Abstract*—The inclusion in advanced device simulators of quantum effects different than standard confinement becomes mandatory to describe device behavior as technology approaches the nanometer scales. This work presents a model to include the gate leakage mechanism considering direct and trap assisted tunneling in Multi-Subband Ensemble Monte Carlo (MS-EMC) simulators. The tool is used for the study of FDSOI and FinFET devices.**

***Index Terms*—gate leakage mechanism; direct tunneling; trap assisted tunneling; MS-EMC; FDSOI; FinFET.**

I. Introduction

As electronic devices have been scaled down, new transistor architectures are considered to replace standard technology. Currently, FinFETs are being recognized as an alternative to Fully-Depleted Silicon-On-Insulator (FDSOI) devices thanks to their immunity to short channel effects because the inclusion of an additional gate increases the electron confinement. In general, the differences in the orientation of both devices modify the electron distribution and the carrier confinement effective mass (m_z^*). From a modeling point of view, to explain its behavior, it has been mandatory to include additional phenomena not needed in previous technological nodes [1]. In particular, the use of ultra-thin insulators, where a high electric field appears, increases the probability of tunneling through the gate oxide [2]. This effect is known as the gate leakage mechanism (GLM) and can be divided into two categories: first, the intrinsic events are always present because they involve the direct tunneling from the substrate to the metal gate; second, the extrinsic ones are related to the existence of defect states and the tunneling path is calculated according to the trap position. This work presents the development and implementation of this phenomenon in a Multi-Subband Ensemble Monte Carlo (MS-EMC) simulator. The model includes direct and trap assisted tunneling and is applied to compare the performance of FDSOI and FinFETs.

II. Methodology

The starting point of the simulation frame is the MS-EMC code presented elsewhere [3], [4]. The tool is based on the mode-space approach of quantum transport [5], where the 1D Schrödinger equation is solved in different slices along the confinement direction, whereas the 2D Bolzmann Transport Equation is solved in the transport plane as showed in Figure 1. Both equations are coupled with the 2D Poisson Equation to keep the self-consistency of the simulator.

Device parameters and orientations for the structures herein analyzed are outlined in Figure 1. Recall that, the considered confinement direction of these devices on standard wafers changes from (100) for the planar FDSOI to $(0\bar{1}1)$ for FinFET, whereas the transport direction remains constant <011>. In this study, the channel thickness ranges from $T_{Si} = 3nm$ to $T_{Si} = 5nm$ and the rest of the parameters remain constant: the gate length $L_G = 15nm$, the gate oxide of SiO_2 with Equivalent Oxide Thickness $EOT = 1nm$, and the gate work function of 4.385eV. For the FDSOI device, a Back-Plane with a $UTBOX = 10nm$, Back-Bias polarization of $V_{BB} = 0V$, and Back-Plane work function of 5.17eV have been chosen.

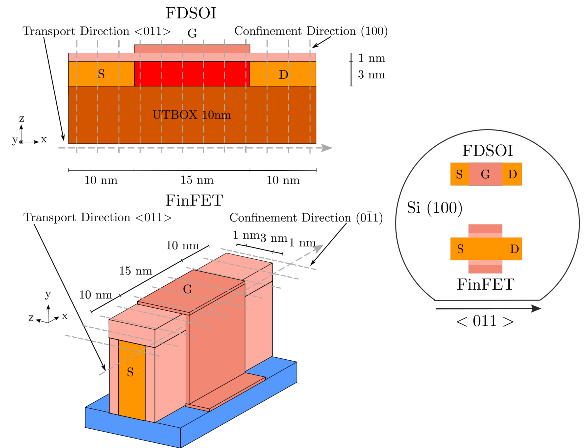

Fig. 1. FDSOI and FinFET structures analyzed in this work with L_G=15nm. 1D Schrödinger equation is solved for each grid point in the transport direction and BTE is solved by the MC method in the transport plane

Before starting the Monte Carlo iterations, some initial characteristics must be fixed for including GLM. First, the number of traps is deterministically calculated according to the oxide dimensions and the trap density, which in turn depends on the material and the wafer orientation. Second, the location and energy level of each trap is randomly reckoned. In this particular work, the trap location inside the gate insulator and its energy have been set to the same value in both devices for comparison purposes, and the number of traps is estimated considering a trap density of $10^{12}cm^{-2}$ in the SiO_2 oxide.

Third, it is indispensable to keep in mind that the MS-EMC code makes use of a 2D description in which the particle motion is characterized in the transport direction whereas it is unknown in the confinement one. Accordingly, the number of particles near the dielectric is required, given that the distance between their location and the interface modifies the tunnel probability [6]. The simulated particles are distributed along the whole device and hence the percentage of the ones near the interface (n_{intf}) is estimated with respect to the total number of particles ($n(x,z)$):

$$n_{intf} = \frac{\sum_{intf}\sum_{x} n(x,z)}{\sum_{z}\sum_{x} n(x,z)}, \quad (1)$$

where $intf$ represents the region near the interface in the z direction. In this study, $intf$ is chosen as the 10% of the T_{Si} for each gate. Moreover, an electron can be trapped only when it is located near a trap location in the oxide with 3D coordinates. Therefore, the dimension of the trap is defined as a cube with the same sizes in the three directions and the percentage of charge n_{perz} that can be located in that cube is estimated.

The last step required before starting the Monte Carlo iterations is the calculation of the tunneling probabilities for each mechanism. In general, the WKB approximation [7] is considered for the estimation of the tunneling probabilities of the electrons similarly to S/D tunneling [8]:

$$T(E) = \exp\left\{-\frac{2}{\hbar}\int_{a}^{b}\sqrt{2m_z^*(E_{CB}(x,z)-E)}\,\mathrm{d}z\right\} \quad (2)$$

where a and b are the starting and ending points, E and m_z^* are the energy and the confinement effective mass of the electron, respectively, and $E_{CB}(x,z)$ corresponds to the energy of the conduction band at the point (x,z). For the GLM, this transmission coefficient does not depend only on the barrier thickness, height, and structure, but also on some specific factors characteristic of each mechanism [9].

Direct tunneling probability is given directly by the WKB approximation. A Fermi-Dirac distribution of the electrons and available states at any given energy in the gate electrodes is assumed considering that, after tunneling, the electrons thermalize (as for the tunnel from the trap to the metal gate).

Trap assisted tunneling must generally consider the trap occupation obeying the Pauli exclusion principle so that this probability allows for two particles in a trap as a maximum. In the case of the inelastic tunnel into a trap, an additional factor must be included in order to transfer to a phonon the difference in energy. When the particle energy is higher (resp. lower) than the trap energy, a phonon is emitted (resp. absorbed) and the probability is multiplied by $(1+n(\omega))$ (resp. $n(\omega)$), $\hbar\omega$ being the difference between the particle and the trap energy, and $n(\omega)$ the Bose–Einstein factor. Finally, the elastic tunnel from the trap to the substrate is forbidden if the energy trap state is lower than the minimum subband. Otherwise, the new energy level of the particle is chosen minimizing to the lower kinetic energy.

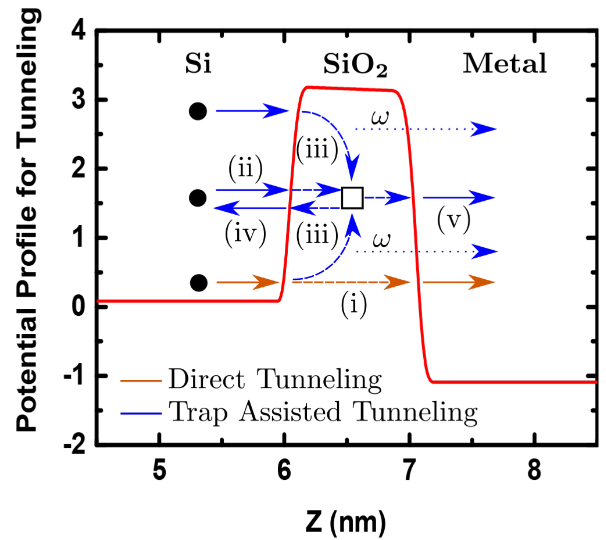

Fig. 2. Schematic band diagram of a MOS structure with metal gate and silicon substrate where the GLM implemented in the MS-EMC simulator are described: (i) direct tunneling, (ii) elastic tunneling and (iii) inelastic tunneling into a trap emitting or capturing a phonon with energy $\hbar\omega$, (iv) detrapping to the substrate, and (v) tunneling from the trap to the gate.

When all the initial parameters of the system are introduced, the Monte Carlo iterations begin. The implementation of the model and its inclusion in the MS-EMC flowchart are shown in Figure 3. According to the fundamentals of the free-flight technique of an electron used in Monte Carlo algorithms, its new position and properties in the transport direction is calculated after a random flight time. Then, before estimating the probability of undergoing a tunnel process and due to the low frequency of GLM, the particles can only undergo this type of tunneling according to a certain period of occurrence (Δt_{GLM}) instead of after each integration step.

Two different scenarios determined by the particle location can arise as depicted in Figure 3: when it is in the channel or it is trapped. On the one hand, if the particle is located in the channel, the first step is to decide if the particle is located near the substrate-dielectric interface. Subsequently, this scenario can be divided in turn into two different sub-scenarios. If the particle is located near the interface and near some trap in the dielectric (always bearing in mind its 3D coordinates), it can undergo both direct and trap assisted tunneling. Otherwise, if the particle is near the interface but far away for any trap, it can only undergo direct tunneling through the insulator. On the other hand, if the particle is located in a trap, the first step is to determine whether the particle can go back to the substrate, leave the device by tunneling to the gate contact, or remain in the trap.

The tunneling path for all the mechanisms needs now to be described. Due to the negligible tunneling time caused by the thin dielectric involved in the gate leakage mechanism and the low frequency of these events, it is reasonable to assume that the electron goes directly from the starting point to the ending point on the same time step. Apart from that, the charge

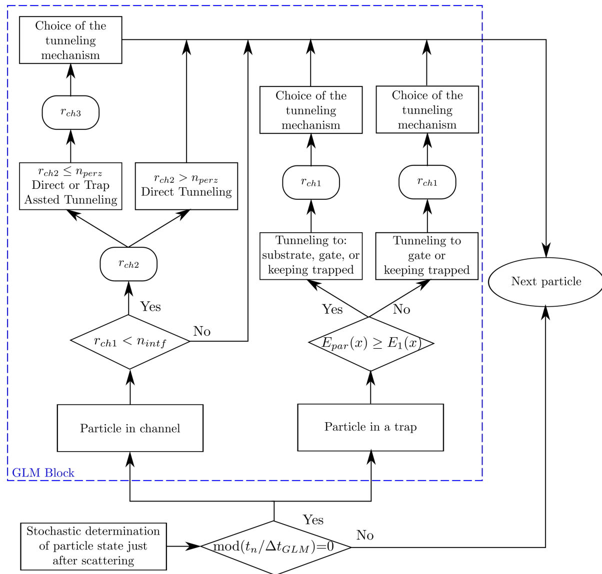

Fig. 3. Flowchart of the gate leakage mechanism (GLM) code included in the MS-EMC simulator where N_{sim} is the considered particle, subscript n stands for the iteration number, Δt_{GLM} is the time step where GLM is calculated, r_{ch1}, r_{ch2}, or r_{ch3} are uniformly distributed random numbers, n_{intf} is the percentage of particles near the interface between the substrate and the oxide, n_{perz} is the percentage of charge that can be located near a trap taking into account the 3D direction, $E_{par}(x)$ and $E_1(x)$ are the particle and the lowest subband energies in the transport direction (x), respectively.

trapped is dynamically included in the 2D Poisson solution in order to preserve the self-consistency during the simulation time.

III. Results and Discussion

All the tunneling mechanisms described above and estimated by means of the WKB approximation are negatively affected by longer tunneling paths, higher potential barriers, or bigger m_z. Considering that both devices have the same potential barrier between the substrate and the metal gate due to the dielectric, only the difference in the effective confinement mass modifies the tunnel probability. In this study, m_z^* for the most populated valley is higher for the FDSOI ($m_l = 0.916 m_0$) compared to the FinFET ($\frac{2 m_l m_t}{m_l + m_t} = 0.326 m_0$). As a result, the transmission coefficient is higher for FinFETs.

However, this fact is compensated by the volume inversion effect which concentrates the charge distribution in the center of the channel. It can be appreciated in Figure 4 where the electron distribution is shown along the transport and confinement directions. The bigger the number of particles near the interface between the substrate and the oxide in the FDSOI, the higher the tunneling probability.

This effect results in the reduction of the number of particles that can experience any GLM as shown in Figure 5, where different remarks must be made. First, the direct tunneling through the oxide is the dominant phenomenon in both devices due to the ultra-thin oxide. Second, the reduction of the total charge near the interface decreases the probability of suffering

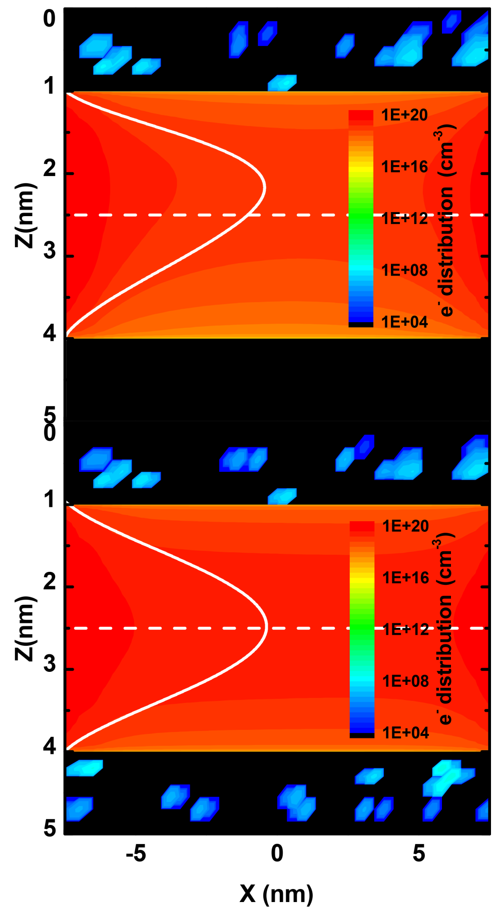

Fig. 4. Electron distribution in cm^{-3} along the transport (X) and confinement (Z) directions in the 15nm device including GLM for FDSOI (top) and FinFET (bottom) with $T_{Si} = 3nm$, $V_{GS} = 0.6V$ and $V_{DS} = 100mV$. Recall that $X = 0$nm corresponds to the center of the device. The solid line is the charge distribution in the Z direction in arbitrary units.

either direct or trap assisted tunneling through the oxide. Therefore, the number of affected electrons is reduced in the FinFET compared to the FDSOI. Third, the probability of a trapped electron to return to the substrate should generally be lower than the probability of tunneling to the gate contact because of the required available energy states. Furthermore, notice that the available states decrease as the gate bias increases until this type of tunneling becomes forbidden. Due to the lower charge near the interface in the FinFET, this fact starts at lower gate bias. Eventually, the slope in the average number of electrons affected by GLM is higher for the FinFET. In other words, the number of particles undergoing direct tunneling tends to be similar for the FDSOI and FinFET as the gate bias increases.

Finally, Figure 6 shows the impact of GLM on the relative variation of drain current compared to the ideal case without leakage ($\Delta I_D / I_{D,w/o}$) as a function of V_{GS}. Two different scenarios must be commented. The first one is when the gate bias is lower than the threshold voltage. In this region, ΔI_D is positive because the number of particles owing to GLM lowers the current. The second one is when the gate bias is

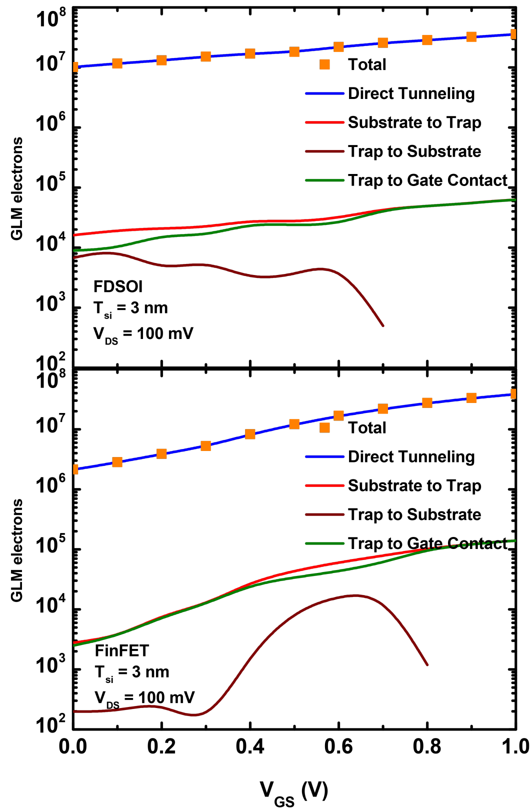

Fig. 5. Average number of electrons in arbitrary units affected by the total gate leakage mechanism, and by each one individually as a function of V_{GS} in the 15nm device and $T_{Si} = 3nm$ at low drain bias for FDSOI (top) and FinFET (bottom).

higher than the threshold voltage, ΔI_D being slightly positive because the reduction in the potential barrier increases the thermionic current, and so the impact of GLM in comparison with $I_{D,w/o}$ is very small. The appreciable onset shift between both scenarios is due to the quantum confinement of the electrons near the interface. The I_{OFF} reduction and the low effectiveness on the I_{ON} due to this phenomenon can be advantageous if a device with the highest I_{ON}/I_{OFF} ratio is demanded.

When the channel thickness increases, the percentage of electrons near the interface decreases. Therefore, the number of affected particles is reduced and the $\Delta I_D/I_{D,w/o}$ decreases. We observe that the presence of a single gate in FDSOI reduces the drain current and thus the impact of GLM on the drain current is larger for this device. In addition, the impact of the GLM on the relative variation of drain current is more persistent in FDSOI because it remains for higher V_{GS} values at any channel thickness.

IV. Conclusions

This work presents the implementation of the gate leakage mechanism (GLM) including direct and trap assisted tunneling in a MS-EMC tool for the comparison of its effect in ultrascaled FDSOI and FinFET devices. Three different assumptions of the motion of a particle undergoing trap

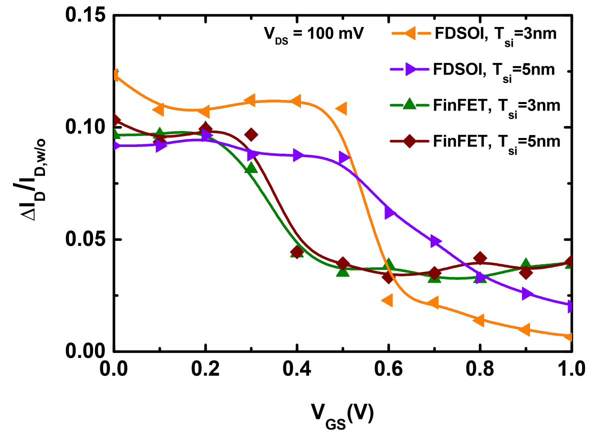

Fig. 6. Ratio of the drain current variation to the drain current without taking into account GLM ($\Delta I_D/I_{D,w/o}$) as a function of V_{GS} for FDSOI and FinFET at low V_{DS}.

assisted tunneling have been explained: tunneling into a trap, or out of a trap to the substrate, and tunneling of electrons out of a trap to the gate electrode. Our calculations show that the main difference in our simulations concerning both devices undergoing this quantum effect comes from the electron confinement near the interface. The number of electrons tunneling through the oxide is higher in the single gate FDSOI in contrast with the vertical FinFET. In conclusion, the FinFET is a better candidate for future ultra-low power applications, with minimized gate leakage currents.

V. Acknowledgment

The authors are grateful for the support given by the Spanish Ministry of Science and Innovation (TEC2014-59730-R), H2020 - REMINDER (687931) and H2020 - WAYTOGO-FAST (662175).

References

[1] H.-S. Wong, "Beyond the conventional transistor," *IBM Journal of Research and Development*, vol. 46, pp. 133–168, March 2002.
[2] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI devices*. New York: Cambridge University Press, 2009.
[3] 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.
[4] 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.
[5] 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.
[6] N. Yang, W. Henson, J. Hauser, and J. Wortman, "Modeling study of ultrathin gate oxides using tunneling current and capacitance-voltage measurement in MOS devices," *IEEE Transactions on Electron Devices*, vol. 46, pp. 1464 – 1471, July 1999.
[7] D. J. Griffiths, "The WKB approximation," in *Introduction to Quantum Mechanics*. New Jersey: Prentice Hall, 1995, ch. 8, pp. 274–297.
[8] C. Medina-Bailon, C. Sampedro, F. Gámiz, A. Godoy, and L. Donetti, "Confinement orientation effects in S / D tunneling," *Solid-State Electronics*, vol. 128, pp. 48 – 53, 2017.
[9] G. C. Jegert, "Modeling of leakage currents in high–κ dielectrics," Ph.D. dissertation, Technische Universität München, München, March 2012.