On the Ability of the Particle Monte Carlo Technique to Include Quantum Effects in Nano-MOSFET Simulation

D. Querlioz, Jérôme Saint-Martin, K. Huet, A. Bournel, V. Aubry-Fortuna, C. Chassat, S. Galdin-Retailleau, P. Dollfus

Estimated length:

TITLE

On the ability of the particle Monte Carlo technique to include quantum effects in nano-MOSFET simulation

AUTHORS

Damien Querlioz, Jérôme Saint-Martin, Karim Huet, Arnaud Bournel, Valérie Aubry-Fortuna, Christophe Chassat, Sylvie Galdin-Retailleau, Philippe Dollfus

Institut d'Electronique Fondamentale (IEF)
CNRS – Univ. Paris-Sud (UMR 8622)
Orsay, France
e-mail: philippe.dollfus@ief.u-psud.fr

ABSTRACT

In this paper we report on the possibility to use particle-based Monte Carlo techniques to incorporate all relevant quantum effects in the simulation of semiconductor nano-transistors. Starting from the conventional Monte Carlo approach within the semi-classical Boltzmann approximation, we develop a multi-subband description of transport to include quantization in ultra-thin body devices. This technique is then extended to the particle simulation of quantum transport within the Wigner formulation. This new simulator includes all expected quantum effects in nano-transistors and all relevant scattering mechanisms which are taken into account the same way as in Boltzmann simulation. This work is illustrated by analyzing the device operation and performance of multi-gate nano-transistors in a convenient range of channel lengths and thicknesses to separate the influence of all relevant effects: significant quantization effects occurs for thickness smaller than 5 nm and wave mechanical transport effects manifest themselves for channel length smaller than 10 nm. We also show that scattering mechanisms still have an important influence in nanoscaled double-gate transistors, both in the intrinsic part of the channel and in the resistive lateral extensions.

KEYWORDS

MOSFETs, semiconductor device modeling, Monte Carlo methods, quantum transport.

I. INTRODUCTION

The particle-based Monte Carlo (MC) technique is acknowledged as a powerful method for accurately describing the carrier transport in semiconductor materials and devices within the semi-classical approximation, i.e. the Boltzmann transport equation (BTE) for the distribution function. It has been developed by many groups to study a wide variety of transport problems in many kinds of devices, to such a point that it is impossible to summarize here the most significant examples of its applications. Extensive overviews of this method may be found in [1,2]. In spite of disadvantages due to large computational requirements and some limitations inherent in the finite number of simulated particles, this technique of transport simulation is turned out to be robust, versatile, essentially free from numerical difficulties and thus suitable for device simulation even in three-dimensional (3D) real space.

However, with the downscaling of field-effect transistors into the nanometer regime, the semi-classical approximation resulting in the BTE fails to capture rigorously the quantum effects occurring in modern devices. For instance, in ultra-thin body (UTB) MOSFET as FD-SOI-FET, Fin-FET or Double-Gate (DG) MOSFET which are considered as the most promising device architectures likely to overcome short-channel effects that dramatically affect conventional bulk-MOSFET, Si channel thickness as small as 5 nm has to be considered in the near future. It yields a strong quantization of the electron gas in the direction perpendicular to the gate stack, which results in a significant change in the space and energy distributions of particles and may be reflected on the device operation and characteristics. Furthermore, for gate length aggressively downscaled in the sub-10 nm range, the wave-like nature of electrons may give rise to source-drain tunneling through the channel barrier and to quantum reflections in the channel. In this context it may be considered as meaningless to still use point particles in the transport description including properly all relevant quantum effects. In fact, this paper aims at showing the ability of the particle MC method to accurately describe quantum effects in self-consistent device simulation in such a way that the transition between classical and quantum transport appears clearly.

Today, most quantum simulators make use of a recursive technique to compute either the Green's function [3,4] or the wave function itself [5]. These methods, initially developed in the ballistic approximation, are now reaching a good level of maturity by including scattering affects, at least electron-phonon scattering, with different levels of approximation

[4 and references therein, 5 and references therein, 6]. However, including all relevant scattering mechanisms in 2D or 3D simulation of MOSFET still remains a difficult problem within these approaches. Instead of computing the wave function or the Green's function, one can equivalently use the Wigner's function (WF) that is defined in the phase space as a Fourier transform of the density operator [7]. In the classical limit this function reduces to the classical distribution function. The dynamical equation of the WF, i.e. the Wigner transport equation (WTE), is very similar to the BTE, except in the influence of the potential whose rapid space variations generate quantum effects. Again, in the classical limit of slow potential gradients the WTE becomes nothing but the BTE. The strong analogy between Wigner and Boltzmann formalisms makes possible and appealing to adapt the standard MC technique to solve the BTE by just considering the WF as an ensemble of pseudo-particles [8,9]. Scattering effects may be easily included by using the same collision operator as in the BTE.

The description of quantization effects in standard Boltzmann MC simulation (BMC) is often considered through so-called "quantum corrections" aiming at empirically reproducing the repulsive effects of interfaces related to the vanishing of the normal component of wave function, while keeping a 3D description of particles [10,11]. Within the effective mass approximation, this quantization effect can be rigorously accounted for by coupling self-consistently the standard algorithm with the 1D Schrödinger equation at each channel position [12]. This multi-subband mode-space approach is very appropriate to the simulation of ultra-thin body transistors [13] where the coupling between subbands is essentially due to inter-subband phonon scattering [14].

Starting from our BMC simulator MONACO [15,16], we recently developed a multi-subband approach (MSBMC) for the simulation of DG MOSFET [13] and extended the MC algorithm to solve the WTE in 1D structures as RTDs [17]. By mixing these two versions we have generated a multi-subband Wigner MC code (MSWMC) which gives a full quantum description of transport in thin-film transistors including all relevant scattering mechanisms, i.e. by phonons, impurities and rough interfaces [18]. In this paper we use successively BMC, MSBMC and MSWMC and compare the results obtained for different designs of unstrained UTB MOSFETs to highlight the main relevant physical effects according to channel thickness and length. Section II is devoted to BMC simulation of multigate MOSFETs with channel length and thickness of 19 nm and 8 nm, respectively. We introduce MSBMC in Section III to detail the quantization effects occurring in DG MOSFET when the body thickness is reduced and MSWMC in Section IV to discuss the quantum transport effects and the ballistic

approximation in DG transistors with channel length smaller than 10 nm.

II. THIN FILM AND MULTIGATE ARCHITECTURES – SEMI-CLASSICAL APPROACH

This section is dedicated to 2D and 3D BMC simulation of multigate MOSFETs on SOI substrate. Such structures with undoped thin-film channel and appropriate design are expected to provide high performance thanks to quasi-ballistic transport and reduced short-channel effects [19] without the need of aggressively small equivalent gate oxide thickness [16]. We consider here a physical gate length $L_G = 21$ nm for three device architectures: FD-SOI MOSFET with a single-gate (SG), double-gate MOSFET (DG) and gate-all-around MOSFET (GAA). Simulated structures are schematized in Fig. 1. Source and drain extensions are 15 nm-long and doped to 10^{20} cm^{-3}. Source/channel and drain/channel junctions are assumed to be abrupt with a gate overlap of 1 nm. The effective channel length is thus $L_{ch} = 19$ nm. The channel body and equivalent oxide thicknesses are $T_{Si} = 8$ nm and $T_{OX} = 1$ nm, respectively. The work function $\Phi_M = 4.56$ eV is assumed for the metallic gate material. In SG the buried oxide is 20 nm thick and ground biased. In GAA the body width W_{Si} is equal to the thickness T_{Si}.

A. The model of BMC simulation

In order to make clear the relationship between Boltzmann and Wigner formalisms in Section IV, we briefly summarize here the basic principles of semiclassical transport of electrons used in standard particle MC simulation. This theory is based on the BTE for the distribution function $f(\mathbf{r},\mathbf{k},t)$ defined as proportional to the density of particles in phase space

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla_r f + \frac{1}{\hbar}\mathbf{F}\cdot\nabla_k f = C\,f \tag{1}$$

where $\mathbf{v} = \hbar^{-1}\nabla_k E$ is the group velocity, $\mathbf{F} = e\nabla_r V$ is the classical electrostatic force derived from the potential $V(\mathbf{r})$ and C is the collision (or scattering) operator. In the particle MC approach the distribution function is conveniently represented by an ensemble of electrons as

$$f(\mathbf{r},\mathbf{k},t) = \sum_i \delta\left(\mathbf{r} - \mathbf{r}_i(t)\right)\delta\left(\mathbf{k} - \mathbf{k}_i(t)\right) \tag{2}$$

where $\mathbf{r}_i(t)$ and $\mathbf{r}_i(t)$ are the position and wave vector, respectively, of the ith electron.

In device simulation the electron trajectories are determined self-consistently with Poisson's equation which is solved at regular time steps Δt. The potential distribution $V(\mathbf{r})$ being approximated as constant during one time step, within an energy band $E(\mathbf{k})$ each electron obeys the classical Newton's equation of motion

$$\begin{cases} \dfrac{d}{dt}\mathbf{r} = \dfrac{1}{\hbar}\nabla_k E(\mathbf{k}) \\ \dfrac{d}{dt}\mathbf{k} = \dfrac{e}{\hbar}\nabla_r V(\mathbf{r}) \end{cases} \tag{3}$$

to be integrated during each time step by including possible scattering events likely to change instantaneously the particle's wave vector. Scattering events are considered statistically through scattering rates calculated according to the Fermi rule [1]. According to the scattering rates, the Monte Carlo game consists in selecting randomly the times of free flight between two successive scattering events, the type and the effect of scattering at the end of flight.

The BMC simulator used here was described in previous works [15,16] and extensively applied for device analysis. It couples the MC algorithm of transport [20] with either a 2D or a 3D Poisson solver, depending on whether we simulate a device with infinite (SG, DG) or finite width (GAA). The time step between two solutions of Poisson's equation is typically Δt = 0.1 fs. We consider an analytical conduction band structure of silicon consisting of six ellipsoidal non parabolic Δ valleys located along the [100] directions at 85% of the Brillouin zone edge [20]. The energy-dependent scattering rates are calculated prior to the simulation and stored in a look-up table used throughout the simulation. We consider all relevant scattering mechanisms, i.e. electron-phonon, electron-impurity and roughness scattering [20]. Throughout this work, we will assume bulk phonon energies and the same coupling constants as in bulk Si without including possible effects related to ultra-thin layers, which is still an unsolved problem [21,22]. A particle multiplication technique [23] is used to statistically enhance the number of rare events in pre-defined crucial regions, which is especially useful here to compute subthreshold currents. A particle entering such a region is duplicated M-1 times, and a charge weight 1/M is assigned to resulting sub-particles. A sub-particle leaving this region is kept in the device with probability 1/M and with initial charge weight, which ensures statistically the conservation of charge, total energy and momentum. It should be noted that for 3D simulation we systematically consider sub-particles having a weight less than unity. It allows us to reduce the particle noise to get an average number of particles per

cell greater than one, which is crucial to stabilize the solution of Poisson's equation in quasi-equilibrium regime.

B. Simulation results for SG, DG and GAA devices

Figure 2 shows the I_D-V_{GS} characteristics obtained for the three devices at $V_{DS} = 0.9$ V. To make easier the comparison, for GAA the drain current is here divided by the channel width W_{Si} although other definitions of current normalization may be used. The intrinsic gate delay will be conveniently used later to provide an unambiguous figure of merit for the comparison of transistors. The subthreshold behaviour of SG turns out to be poor with unacceptably high subthreshold slope S = 124 mV/dec. It is an expected consequence of the quite small form factor $L_{ch}/T_{Si} = 2.375$ which does not allow to control short channel effects in a single-gate device. In contrast, both DG and GAA exhibit good subthreshold slope of 67 mV/dec and 64 mV/dec, respectively.

In Figs. 3 and 4 we plot the I_D-V_{GS} characteristics for DG and GAA, respectively, at both V_{DS} values of 0.05 V and 0.9 V. By defining the threshold voltage V_T for a constant current $I_0 = 10$ µA/µm at low V_{DS}, the same value $V_T = 0.42$ V is obtained for these two devices. While the drain-induced barrier-lowering (DIBL) is 70 mV in DG, it is significantly reduced to 20 mV in GAA thanks to an excellent electrostatic gate control of the channel.

However, as shown in Fig. 2, the on-current I_{ON} of GAA is not 2 times higher than that of DG. So, as already observed for other designs of multigate transistors [16] the drive current is not proportional to the number of gates, even if the current is simply normalized to the channel width.

The intrinsic delay $C_G V_{DD} / I_{ON}$ is plotted in Fig. 5 as a function of I_{OFF} for the three devices and compared to ITRS targets for the technological nodes LSTP 32 nm and HP 50 nm [24]. The value of I_{OFF} is tuned by just changing the gate work function Φ_M. Because of too high I_{OFF} SG cannot reach any ITRS specifications. The GAA structure takes advantage of very small DIBL to get smaller I_{OFF} than DG but the I_{ON} enhancement in GAA is too limited to compensate the gate capacitance increase. Finally, DG appears as the best compromise in terms of intrinsic delay.

We now examine the electron transport in the channel of GAA and DG in the on-state ($V_{GS} = V_{DS} = V_{DD}$). We plot in Fig. 6 the fraction of electrons which have crossed the channel from the source-end to the drain-end as a function of the number of scattering events they have experienced in this interval. In both cases the curve decays quasi-exponentially but less

rapidly in GAA, which may be attributed to slightly more effective roughness scattering than in DG. For these 19 nm-long channel devices the intrinsic ballisticity B_{int}, i.e. the fraction of ballistic electrons, reaches 38% in GAA and 43% in DG. In inset is shown the parameter B_{int} as a function of channel length for DG devices with a body thickness $T_{Si} = 5$ nm. The ballisticity is negligible for $L_{ch} > 100$ nm and reaches 48% for $L_{ch} = 15$ nm. The value obtained here for $L_{ch} > 100$ nm and $T_{Si} = 8$ nm (closed circle in the inset) matches very well the curve for $T_{Si} = 5$ nm, which indicates that B_{int} is weakly dependent on the film thickness in this range.

From the results obtained here using standard BMC, different questions come to light regarding the electron transport in more advanced CMOS generations with channel length reduced below 15 nm or even 10 nm and channel thickness smaller than 5 nm in order to prevent from short channel effects. These questions are

- on the quantization effect related to channel thickness reduction below 5nm
- on the quantum transport effects related to channel length reduction below 10 nm
- on the ballistic approximation for channel length below 10 nm

We will examine these relevant problems in the next sections by making use of appropriate transport models.

III. Reduction of film thickness – Quantization effects

In this Section we focus our interest on DG devices with a channel length L_{ch} of 15 nm and a channel thickness T_{Si} of either 4 nm or 3 nm. We intend to analyze the quantization effects related to T_{Si} shrinking by making use of multi-subband Boltzmann MC simulation, whose results will be compared with that of standard BMC simulation. In the devices simulated using MSBMC, four subbands are taken into account in unprimed valleys (with a quantization mass of $0.916\, m_0$) and two subbands are considered in primed valleys (with a quantization mass of $0.19\, m_0$).

A. The multi-subband approach (2D-k electron gas, MSBMC)

This approach is based on the mode-space approximation which decouples the gate-to-gate z direction and the xy plane parallel to interfaces. Assuming the potential V to be y-independent, the formation of uncoupled subbands resulting from reduced channel thickness may be simply deduced from the effective 1D Schrödinger's equation within parabolic band approximation

$$-\frac{\hbar^2}{2m_z}\frac{\partial^2}{\partial z^2}\xi_n(x,z) - eV(x,z)\,\xi_n(x,z) = E_n(x)\,\xi_n(x,z) \tag{4}$$

where $E_n(x)$ and $\xi_n(x,z)$ are the quantized energy and envelope function, respectively, associated with *n*-th subband.

According to the real-space meshing used for solving Poisson's equation, Eq. (4) has to be solved self-consistently at each position x_i in the channel to determine the subband profile $E_n(x)$ to be used as potential energy for the particle transport along the source-to-drain axis in *n*-th subband. For such a 2D device, the Monte Carlo procedure is then applied for each occupied subband in a phase space reduced to three dimensions (x, k_x, k_y), instead of five dimensions in standard BMC simulation (x, z, k_x, k_y, k_z). To solve the 2D Poisson's equation after each time step, the electron density $n(x,z)$ is updated by distributing along the confinement direction *z* the charge of each electron present in subband *n* at position x_i according to the appropriate probability density $|\xi_n(x_i,z)|^2$.

However the requirement of solving Schrödinger's equation (4) increases the complexity of the overall algorithm. In particular, updating $E_n(x)$ and $\xi_n(x_i,z)$ at each time step considerably increases the computation time compared with standard BMC. Alternatively, it is possible to solve rigorously Eq. (4) each *N* time step only and to update the eigenvalues $E_n(x)$ at other time steps according to the first order correction

$$E_n(x_i) = E_n^j(x_i) + \left\langle \xi_n^j(x_i,z) \middle| -e\Delta V(x_i,z) \middle| \xi_n^j(x_i,z) \right\rangle \tag{5}$$

where $\Delta V(x_i,z))$ is the difference between the new electrostatic potential and that used for the last *j*-th Schrödinger solution. There is no need to correct the envelope function between two Schrödinger solutions, and for a Poisson time step Δt = 0.1 fs, solving Eq. (4) each 10 fs, i.e. choosing N = 100, turns out to be correct [14], which drastically reduces the computation time by a factor of about 15.

It should be mentioned that in this work we do not apply any non-parabolocity correction to the solution of Eq. (4) [12], which may modify high energy levels, and we assume the envelope function to vanish at Si/SiO_2 interfaces. The problem of the influence of wave function penetration into the oxide layers on the propagation of particles along the channel is not trivial and should be addressed in further works.

The MC procedure makes use here of scattering rates for 2D electron gas to be calculated according to the envelope functions $\xi_n(x_i, z)$ whose dependence on time and position generates an additional difficulty. In contrast to the case of standard BMC it is no longer possible to store the scattering rates in a look-up table prior to the simulation. They have to be regularly up-dated throughout the simulation, which contributes to weigh down the simulation time. Phonon and impurity scattering rates are derived as in Ref. [25] where 2D electron mobility in Si/SiGe heterostructures was calculated in good agreement with experimental results. The Si/SiO_2 roughness rate is calculating by considering both the classical effect of electrostatic potential fluctuations [26] and the quantum effect on eigen energies [27] which becomes significant for Si film thickness smaller than 5 nm [Esseni 03]. Standard parameters, i.e. root-mean-square Δ_m = 0.5 nm and correlation length L_c = 1.5 nm, are used to characterize the surface roughness.

B. Comparison between 3D-k and 2D-k descriptions of electron gas

First of all we plot in Fig. 7 the effective mobility computed in undoped long channel structures as a function electron sheet density for both film thicknesses considered in this section. As expected, in the case of 3Dk calculation (dashed lines) the mobility is weakly dependent on the Si thickness and very similar to that obtained in bulk Si inversion layers [Aubry 05]. In contrast, from 3Dk calculation (solid lines) we observe an enhancement for $T_{Si} = 4$ nm (circles) and a reduction of mobility for $T_{Si} = 3$ nm (squares). This trend was previously observed from transport calculation based on the momentum relaxation time approximation [28]. The mobility reduction in the thinnest structure is due to the increasing influence of roughness scattering while the small increase obtained for $T_{Si} = 4$ nm is attributed to a redistribution of most electrons in the first unprimed subband where the effective mass along transport direction is small. On a small range of film thickness, the latter effect may surpass the roughness influence. However this mobility behavior does not seem to strongly influence the characteristics of short-channel device, as shown below.

For the DG devices investigated here the gate length L_G and the channel length L_{ch} are both equal to 15 nm. The oxide thickness is now $T_{OX} = 1.2$ nm and the gate work function is $\Phi_M = 4.36$ eV. We will consider two values of body thickness T_{Si}, 4 nm and 3 nm, respectively. The power supply is $V_{DD} = 0.7$ V.

Figure 8 shows the I_D-V_{GS} characteristics obtained for both devices at $V_{DS} = 0.05$ V by BMC simulation (3Dk, triangles) and by MSBMC simulation (2Dk, circles). In each case we

essentially observe a shift between 3Dk and 2Dk results, this shift being greater for $T_{Si} = 3$ nm (Fig. 8b) than for $T_{Si} = 4$ nm (Fig. 8a). It illustrates the expected V_T-shift induced by quantization of the electron gas, which can be predicted from electrostatics considerations.

However, the situation is quite different at $V_{DS} = V_{DD} = 0.7$ V, as illustrated in Fig. 9a ($T_{Si} = 4$ nm) and Fig. 9b ($T_{Si} = 3$ nm). In such drain bias condition, similar shift is observed at low V_{GS} but an unexpected crossing of curves is observed at V_{GS} greater than about 0.5 V, which yields higher current in 2Dk gas than in 3Dk gas. This effect is very significant in the thinnest structure.

It is clearly illustrated by the behavior of the gate capacitance C_G as a function of gate voltage, which is obtained by differentiating the total charge in the device with respect to V_{GS}. At low V_{DS}, C_G evolves the same way as the current I_D (not shown) but at high V_{DS} (Fig. 10), a crossing of 3Dk and 2Dk curves is still observed. Again, at given high V_{GS} the difference of C_G between 3Dk and 2Dk cases is stronger for $T_{Si} = 3$ nm (squares) than for $T_{Si} = 4$ nm (circles). All other things being equal, this effect is certainly a consequence of carrier repopulation among the different subbands by phonon scattering.

For $T_{Si} = 3$ nm and $V_{GS} = 0.7$ V, Fig. 11 shows the distribution of electron density among the most important subbands (E_1, E_2 and two E'_1 subbands) along the sourc-drain axis in both cases of low V_{DS} (dashed lines) and high V_{DS} (solid lines). It illustrates the hot electron transfer from the lowest subband E_1 to higher subbands. At the drain-end of the channel (x = 15 nm) the density in primed subbands is even higher than in unprimed subbands, which locally modifies the shape of density along quantization axis according to the envelope functions. It should be noted that under such high gate bias, the electron transfer starts to occur in the source access. It is a consequence of the electric field which is built in this resistive region, as shown in Fig. 12 that is a plot of the subband energy profiles for $V_{GS} = V_{DS} = 0.7$ V. This heating effect in the source extension is not observed at smaller gate bias (not shown here).

Finally, we would like to mention that the mode-space approximation turns out to be a good approximation in such thin devices. Indeed, we have checked that the coupling potentials between subbands are always smaller than 0.5 meV, which is less than all energies involved in this transport problem.

IV. REDUCTION OF CHANNEL LENGTH – QUANTUM TRANSPORT AND BALLISTIC LIMIT

In this Section we study the quantum transport in DG transistors with a channel thickness fixed to the value of 3 nm and a channel length reduced to 9 nm and 6 nm, respectively. The multi-subband approach is still used as in the previous section but the Wigner formalism is developed for the transport along source-to-drain direction. The results obtained using this MSWMC simulation will be compared to MSBMC calculation to analyze the actual impact of quantum transport on device performance and characteristics. They will be also compared to ballistic Green's function calculation to emphasize the important part that scattering mechanisms still take in such small silicon devices at room temperature.

A. The Wigner Monte Carlo approach of quantum transport

To deal with quantum transport along a single real-space direction *x,* the Wigner formalism is based on the Wigner function f_w defined for each subband in the phase-space (x,k) as a Fourier transform of the density matrix [7] (for simplicity we omit here subscript *x* for k_x). Although it is not positive definite, this function has similar properties to that of a distribution function via its relation with relevant physical quantities of the system as density, energy, current… The dynamical equation of the WF is the WTE

$$\frac{\partial f_w}{\partial t} + v_x \frac{\partial f_w}{\partial x} = Q\, f_w + C\, f_w \tag{6}$$

which is very similar to the 1D form of the BTE (1). The only difference comes from the term $Q\, f_w$ that is the quantum term including the non-local effect of the potential, defined by

$$Q\, f_w(x,k) = \frac{1}{2\pi\hbar}\int dk'\, V_w(x,k')\; f_w(x,k+k') \tag{7}$$

where the Wigner potential V_w in the *n*-th subband of potential energy $E_n(x)$ is given by

$$V_w(x,k) = \int dx' \sin(kx')\left[E_n\left(x+\frac{x'}{2}\right) - E_n\left(x-\frac{x'}{2}\right)\right] \tag{8}$$

An alternative form for the quantum term $Q\, f_w$ leads to an expansion in powers of $\hbar$ and high order derivatives of the potential energy [7]. For slowly varying potential the first order approximation of this form is nothing but the effect of the classical force on the

distribution function. Thus, the BTE can be seen as the classical approximation of the WTE if the same collision operator is used in both cases, i.e. if quantum collision effects are neglected. Here, in our MSWMC simulator we will use the same scattering rates as in MSBMC presented in the previous section, which makes very clear the analogy between both formalisms.

To extend the particle MC algorithm to the BTE we use the affinity technique initially proposed by Shifren et al. [8]. It consists in describing the WF as an ensemble of pseudo-particles defined by a position x_i and a wave vector k_i and weighted by a new quantity, the affinity A_i, that contains the full wave properties of particles. The WF takes then the form

$$f_w(x,\mathrm{k},t) = \sum_i A_i(t)\,\delta\left(x - x_i(t)\right)\delta\left(\mathrm{k} - \mathrm{k}_i(t)\right) \tag{9}$$

to be compared to the particle formulation (2) of the classical distribution function. Such pseudo-particles behave and scatter as classical particles, except that the potential does no longer influence the wave vector but only the affinity through the quantum evolution term $Q\,f_w(x,k)$. The wave vector can change only after a scattering event. Inside a mesh $M(x,k)$ of the phase space the particle coordinates obey the following motion equations during a free flight

$$\begin{cases} \dfrac{d}{dt}x_i = \dfrac{1}{\hbar}\nabla_k E(k) \\ \dfrac{d}{dt}k_i = 0 \\ \displaystyle\sum_{i\in M}\frac{d}{dt}A_i = Q\,f_w(x,k) \end{cases} \tag{10}$$

The affinity can thus take negative values, in accordance with the fact that the Wigner function may be negative too. Concretely, the affinity evolution equation is applied at the end of each Poisson time step. Finally, the simulation of 2D transistors studied in the present work consists in computing in a self-consistent manner the following loop: Wigner MC trajectories of particles for one time step, particle density $n(x,z)$, electrostatic potential $V(x,z)$ through 2D Poisson's equation, subband energies $E_n(x)$ and envelope functions $\xi_n(x,z)$ through 1D Schrödinger's equation (4) at each position x_i in the channel, Wigner function (9), Wigner potential (8) and quantum evolution term $Q\,f_w(x,k)$ (7).

A critical point of this technique is the particle injection in the device. In BMC

simulation, thermal particles are only injected at ohmic contacts to ensure the neutrality and quasi-equilibrium conditions in the cells adjacent to the contacts. In WMC it is of course necessary to still inject particles of affinity equal to 1 to get charge neutrality at contacts but also to inject particles with zero-affinity in the device to properly compute the Wigner function in the full phase space. Indeed, to correctly apply the affinity evolution equation in (10) and to ensure the conservation of the total affinity of particles, each mesh $M(x,k)$ of the phase space must always contain at least one pseudo-particle. To fulfill this requirement, we showed that it is necessary to additionally inject zero-affinity particles in all empty mesh where the quantum evolution term $Q\, f_w(x,k)$ is not null [17]. This condition has to be paid by a quite large number of simulated particles, i.e. typically between 300 000 and 700 000 according to the applied bias.

Using such a procedure, there is a full compatibility between Boltzmann and Wigner algorithms that can be applied in the same device depending on whether we consider the classical access regions or the quantumly active regions.

B. The Green's function approach of quantum transport

The Green's function formalism provides a general approach to describing the quantum transport in devices [29]. It has been widely developed this last years to study different kinds of mesoscopic and nanodevices. The description of this formalism is outside of the scope of this work devoted to particle MC technique. It will be used here in the next subsection to compare our diffusive Wigner approach to a model rigorously developed for fully ballistic transport calculation. Of course both Wigner and Green formalisms are strongly connected. It may be shown that the less-than Green's function $G^<$ is proportional to the density matrix. The Wigner's function is thus a Fourier transform of $G^<$ [29,30].

Our Green's function simulator has been initially developed using the tight-binding formulation to treat 1D ballistic transport problems including self-consistent coupling with Poisson's equation [31]. It has been recently extended to the simulation of 2D structures as the DG-MOSFET in the multi-subband mode-space approximation.

C. Comparison between Boltzman, Wigner and Green simulations

As already mentioned the DG devices investigated in this section have a gate length L_G and a channel length L_{ch} which are both equal to 9 nm or 6 nm, with a channel thickness T_{Si} fixed to 3 nm. The oxide thickness is now $T_{OX} = 1$ nm and the gate work function is

$\Phi_M = 4.36$ eV. The power supply is still $V_{DD} = 0.7$ V. The source and drain extensions are 15 nm-long and doped to $N_D = 5\times10^{19}$ cm^{-3}.

The computed drain current I_D is plotted as a function of gate bias at high V_{DS} in Fig. 13 for $L_{ch} = 9$ nm and in Fig. 14 for $L_{ch} = 6$ nm. The results obtained from the three transport models are compared, i.e. from MSBMC (Boltzmann), MSWMC (Wigner) and ballistic Green simulation.

It is remarkable that Wigner and ballistic-Green results closely coincide in the subthreshold regime, which confirms that scattering mechanisms have a very small impact in this regime. In contrast, they strongly differ at high V_{GS}, which shows that scattering cannot be neglected to evaluate the drive current of nanotransitors, even for such small channel length.

When comparing Boltzmann and Wigner results, one can observe quite different subthreshold behaviors. A small shift is obtained for $L_{ch} = 9$ nm but by reducing the channel length to 6 nm a change of slope is perceptible and the shift increases. The additional drain current resulting from MSWMC is obviously due to the contribution of source-drain tunneling through the channel barrier, which of course is not accounted for in MSBMC. For $L_{ch} = 6$ nm the tunneling current increases I_{OFF} by a factor of about 5. At high V_{GS} Wigner and Boltzmann results are much closer, the Wigner current being even slightly smaller than the Boltzmann one, which may be explained by the arising of quantum reflections in the channel induced by the steep potential drop [Querlioz 06].

Figure 15 illustrates the effect of scattering in the channel and in access regions for $L_{ch} = 6$ nm. We compare the I_D-V_{GS} curve obtained by Wigner simulation (squares) with that obtained by switching the scattering off in the channel only (circles). The significant discrepancy in on-state clearly shows that scattering in the channel still influences the transport for 6nm-long silicon devices. As shown in inset, the effective ballisticity $I_D / I_{D_{ball}}$ decreases to 86% in on-state. Furthermore the comparison with fully ballistic Green calculation (diamonds) shows that the role of scattering in resistive access regions is even more important and may limit the intrinsic performance of ultra-thin transistors.

We now discuss the electrical performance of this device ($L_{ch} = 6$ nm) in terms of I_{ON}-I_{OFF} ratio and intrinsic delay $C_G V_{DD} / I_{ON}$. In Figs. 16 and 17 we plot I_{ON} and the delay, respectively, as function of I_{OFF}. Again, I_{OFF} is artificially changed here by tuning the gate work function. We compared the results obtained from the three transport models with the ITRS target for the HP 16 nm technology node [24]. Even by using the optimistic ballistic

calculation it seems impossible to reach the targets. Including all quantum and scattering effects, i.e. using Wigner simulation, makes even the results very far from the objectives. Actually, the limitation of I_{ON} by strong access resistances in ultra-thin extensions seems to be redhibitory. Technological solutions should be developed in order to drastically reduce such access resistance, as for instance smaller extensions with higher doping or strain engineering.

IV. CONCLUSION

We have shown that the particle Monte Carlo technique can successfully handle all quantum effects in the self-consistent simulation of electron transport in semiconductor devices. Provided that appropriate formalism and particle treatment are used it is possible to include quantization effects in the simulation of thin-film transistors within the semiclassical Boltzmann approach and to rigorously extend this technique to the simulation of quantum transport within the Wigner approach. Both approaches are fully compatible and may be used together in the same device. If quantum collision effects can be neglected these approaches can easily include all relevant scattering mechanisms as in the standard Monte Carlo procedure. This work has been illustrated by the analysis of device operation and performance of nano-transistors in the appropriate range of typical lengths and thicknesses to separate the influence of all relevant transport effects: quantization for thickness smaller than 5 nm and wave mechanical transport for channel length smaller than 10 nm. We have also shown that scattering mechanisms still have an important influence at nanometer scale, both in the intrinsic part of the channel and in the resistive source extension.

ACKNOWLEDGMENTS

The authors would like to thank Do Van Nam for providing Green's function calculation and for fruitful discussion. This work was supported by the European Community through Network of Excellence SINANO (IST-506844) and Integrated Project PULLNANO (IST-4-026828), and the Agence Nationale de la Recherche through project MODERN (ANR-05-NANO-002).

REFERENCES

[1] C. Jacoboni and P. Lugli, "The Monte Carlo method for semiconductor device simulation", Springer-Verlag, Wien-New York, 1989.

[2] C. Jungemann and B. Meinerzhagen, "Hierarchical device simulation: The Monte Carlo perspective", Springer, Wien-New York, 2003.

[3] R. Venugopal, M. Paulsson, S. Goasguen, S. Datta, and M.S. Lundstrom, "A simple quantum mechanical treatment of scattering in nanoscale transistors", J. Appl. Phys. vol. 93, pp. 5613-5625, 2003.

[4] S. Jin, Y.J. Park, and H.S. Min, "A three-dimensional simulation of quantum transport in silicon nanowire in the presence of electron-phonon interactions", J. Appl. Phys., vol. 99, 123719, 2006.

[5] M.J. Gilbert, R. Akis, and D.K. Ferry, "Phonon-assisted ballistic to diffusive crossover in silicon nanowire transistors," J. Appl. Phys., vol. 98, 094303, 2005.

[6] R. Lake, G. Klimeck, R.C. Brown, and D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices", J. Appl. Phys. vol. 81, pp. 7845-7869,1997.

[7] C. Jacoboni, R. Brunetti, P. Bordone, and A. Bertoni, "Quantum transport and its simulation with the Wigner-function approach," International Journal of High Speed Electronics and Systems, vol. 11, pp. 387-423, 2001.

[8] L. Shifren, C. Ringhofer, and D. K. Ferry, "A Wigner function-based quantum ensemble Monte Carlo study of a resonant tunneling diode," IEEE Trans. Electron Devices, vol. 50, pp. 769-773, 2003.

[9] M. Nedjalkov, H. Kosina, S. Selberherr, C. Ringhofer, and D.K. Ferry, "Unified particle approach to Wigner-Boltzmann transport in small semiconductor devices", Phys. Rev. B, vol. 70, 115319, 2004.

[10] D. K. Ferry, R. Akis, and D. Vasileska, “Quantum effects in MOSFETs : Use of an Effective Potential in 3D Monte-Carlo simulation of ultra-short channel devices”, IEDM Tech. Dig., pp. 287-290, 2000.

[11] B. Winstead and U. Ravaioli, “A quantum correction based on Schrödinger equation applied to Monte Carlo device simulation”, IEEE Trans. Electron Dev., vol. 50, pp. 440-446, 2003.

[12] M.V. Fischetti and S.E. Laux, "Monte Carlo study of electron transport in silicon inversion layers", Phys. Rev. B, vol. 48, pp. 2244-2274, 1993.

[13] J. Saint Martin, A. Bournel, F. Monsef, C. Chassat, and P. Dollfus, "Multi sub-band Monte Carlo simulation of ultra-thin Double Gate MOSFET with 2D Electron Gas," *Semicond. Science Technol*., vol. 21, pp. L29-L31, 2006.

[14] J. Saint-Martin, D. Querlioz, A. Bournel, and P. Dollfus, "Efficient multi sub-band Monte Carlo simulation of nano-scaled Double Gate MOSFETs", in Proc. SISPAD 2006, IEEE Catalog Number 06TH8887, pp. 216-219, 2006.

[15] J. Saint Martin, A. Bournel, and P. Dollfus, "On the ballistic transport in nanometer-scaled double-gate MOSFET", IEEE Trans. Electron Devices, vol. 51, pp. 1148-1155, 2004.

[16] J. Saint Martin, A. Bournel, and P. Dollfus, "Comparison of multiple-gate MOSFET architectures using Monte Carlo simulation", Solid-State Electron., vol. 50, pp. 94-101, 2006.

[17] D. Querlioz, P. Dollfus, V. N. Do, and V. L. Nguyen, "An improved Wigner Monte-Carlo technique for the self-consistent simulation of RTDs", in Proc. IWCE 11, 2006, J. Comp. Electron., in press.

[18] D. Querlioz, J. Saint-Martin, V. Nam Do, A. Bournel, and P. Dollfus, "Fully Quantum self-consistent study of ultimate DG-MOSFETs including realistic scattering using a Wigner Monte-Carlo approach", IEDM Tech. Dig., 2006, pp. 941-944.

[19] J-P. Colinge, "Multiple-gate SOI MOSFETs", Solid-State Electron., vol. 48, pp. 897-905, 2004.

[20] V. Aubry-Fortuna, P. Dollfus, and S. Galdin-Retailleau, "Electron effective mobility in strained $Si/Si_{1-x}Ge_x$ MOS devices using Monte Carlo simulation", Solid-State Electron., vol. 49, pp. 1320-1329, 2005.

[21] D. Esseni, A. Abramov, L. Selmi, and E. sangiorgi, "Physically based modeling of low field electron mobility in ultra-thin single- and double-gate SOI n-MOSFETs", IEEE Trans. Electron Devices, vol. 50, pp. 2445-2455, 2003.

[22] L. Donetti, F. Gámiz, J.B. Roldan, A. Godoy, "Acoustic phonon confinement in silicon nanolayers: effect on electron mobility", J. Appl. Phys., vol. 100, 013701 (2006).

[23] M.V. Fischetti and S.E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects", Phys. Rev. B, vol. 38, pp. 9721-9745, 1988.

[24] International Technology Roadmap for Semiconductors, http://public.itrs.net/.

[25] F. Monsef, P. Dollfus, S. Galdin-Retailleau, H.-J. Herzog, T. Hackbarth, "Electron transport in Si/SiGe modulation-doped heterostructures using Monte Carlo simulation", J. Appl. Phys. 95 (2004) 3587-3593.

[26] S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy, and O. L. Krivanek, "Surface roughness at the Si(100)-SiO2 interface", Phys. Rev. B, vol. 32, pp. 8171-8186, 1985.

[27] H. Sakaki, T. Noda, K. Hirakawa, M. Tanaka, and T. Matsusue, "Interface roughness scattering in GaAs/AlAs quantum wells" Appl. Phys. Lett., vol. 51, pp. 1934-1936, 1987.

[28] D. Esseni and E. Sangiorgi, "Low field electron mobility in ultra-thin SOI MOSFETs: experimental characterization and theoretical investigation", State Electron., vol. 48, pp. 927-936, 2004.

[29] S. Datta, "Electronic transport in mesoscopic systems", Cambridge University Press, Cambridge, 1995

[30] D.K. Ferry, D. Vasileska, and H.L. Grubin, "Quantum transport in semiconductor devices", International Journal of High Speed Electronics and Systems, vol. 11, pp. 363-385, 2001.

[31] V. Nam Do, P. Dollfus, V. Lien Nguyen, "Transport and noise in resonant tunneling diode using self-consistent Green's function calculation", J. Appl. Phys. vol. 100, 093705, 2006.

FIGURE CAPTIONS

Figure 1. Schematic view of simulated devices

Figure 2. I_D-V_{GS} characteristics at V_{DS} = 0.7V in SG, DG and GAA

Figure 3. I_D-V_{GS} characteristics at V_{DS} = 0.05V and V_{GS} = 0.7V in DG

Figure 4. I_D-V_{GS} characteristics at V_{DS} = 0.05V and V_{GS} = 0.7V in GAA

Figure 5. Intrinsic delay $C_G V_{DD} / I_{ON}$ As a function of I_{OFF} for SG, DG and GAA and for various gate work functions indicated through the difference $\Delta\Phi$ with a midgap gate.

Figure 6. Fraction of electrons flowing from the source-end to the drain-end of DG and GAA as a function of the number of scattering events experienced (On-State). Inset: intrinsic ballisticity as a function of channel length.

Figure 7. Effective mobility in long-channel DG devices for two body thicknesses including (solid lines) or not (dashed lines) quantization effects.

Figure 8. I_D-V_{GS} characteristics at V_{DS} = 0.05V for T_{Si} = 4 nm (a) and T_{Si} = 3 nm (b)

Figure 9. I_D-V_{GS} characteristics at V_{DS} = 0.7V for T_{Si} = 4 nm (a) and T_{Si} = 3 nm (b)

Figure 10. Gate capacitance as a function of gate voltage at V_{DS} = 0.7V for T_{Si} = 4 nm (circles) and T_{Si} = 3 nm (squares).

Figure 11. Average sheet density in the first subbands at V_{GS} = 0.7V and V_{DS} = 0.05V (dashed lines) and V_{DS} = 0.7V (solid lines). T_{Si} = 3 nm.

Figure 12. First subband energies along the *x*-axis for V_{GS} = V_{DS} = 0.7V.(T_{Si} = 3 nm).

Figure 13. I_D-V_{GS} characteristics at V_{DS} = 0.7V for L_{ch} = 9 nm using three types of simulation

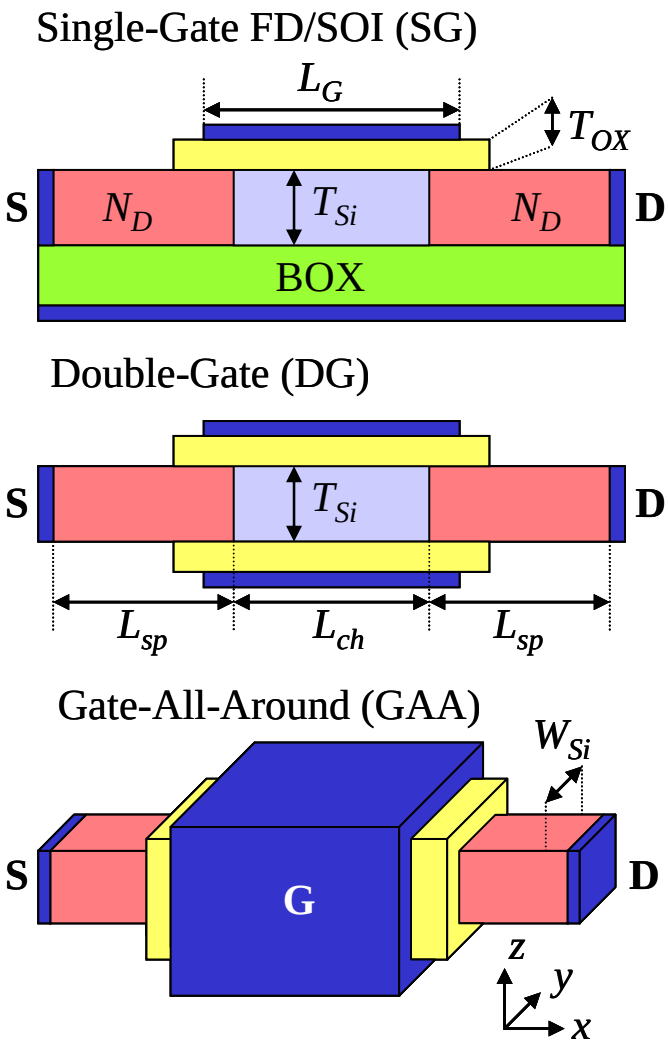
Figure 14. I_D-V_{GS} characteristics at V_{DS} = 0.7V for L_{ch} = 6 nm using three types of simulation

Figure 15. I_D-V_{GS} characteristics at V_{DS} = 0.7V for L_{ch} = 6 nm using Wigner including or not scattering in the channel. Comparison with Green's function calculation.

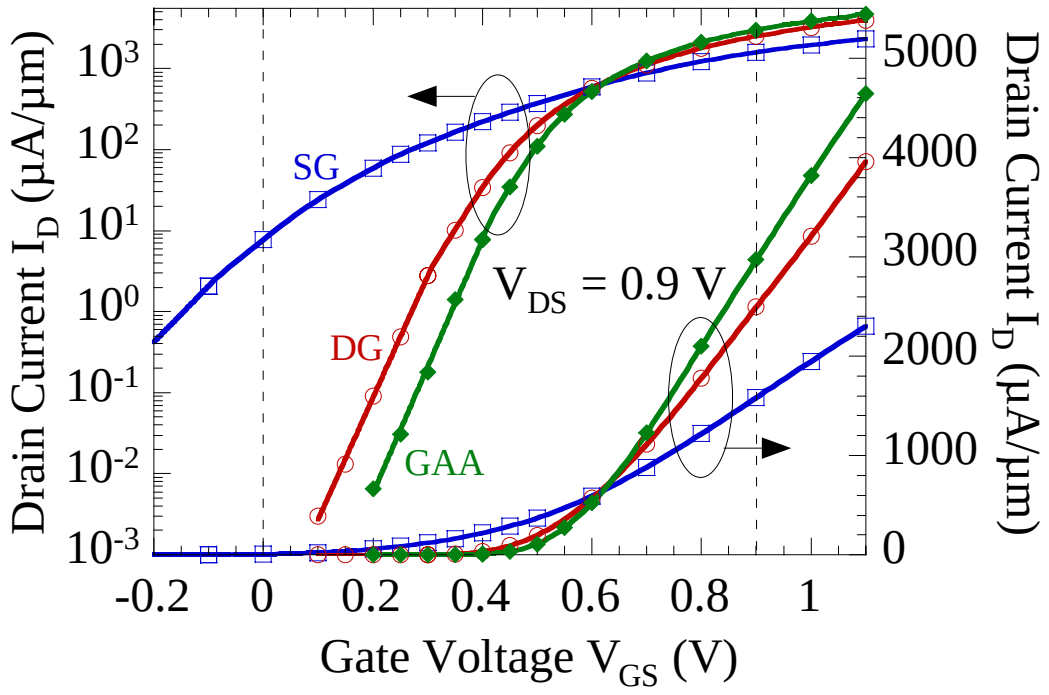
Figure 16. I_{ON}-I_{OFF} plot for various gate work functions indicated through the difference $\Delta\Phi$ with a midgap gate (L_{ch} = 6 nm).

Figure 17. Intrinsic gate delay as a function of I_{OFF} for various gate work functions indicated through the difference $\Delta\Phi$ with a midgap gate (L_{ch} = 6 nm).

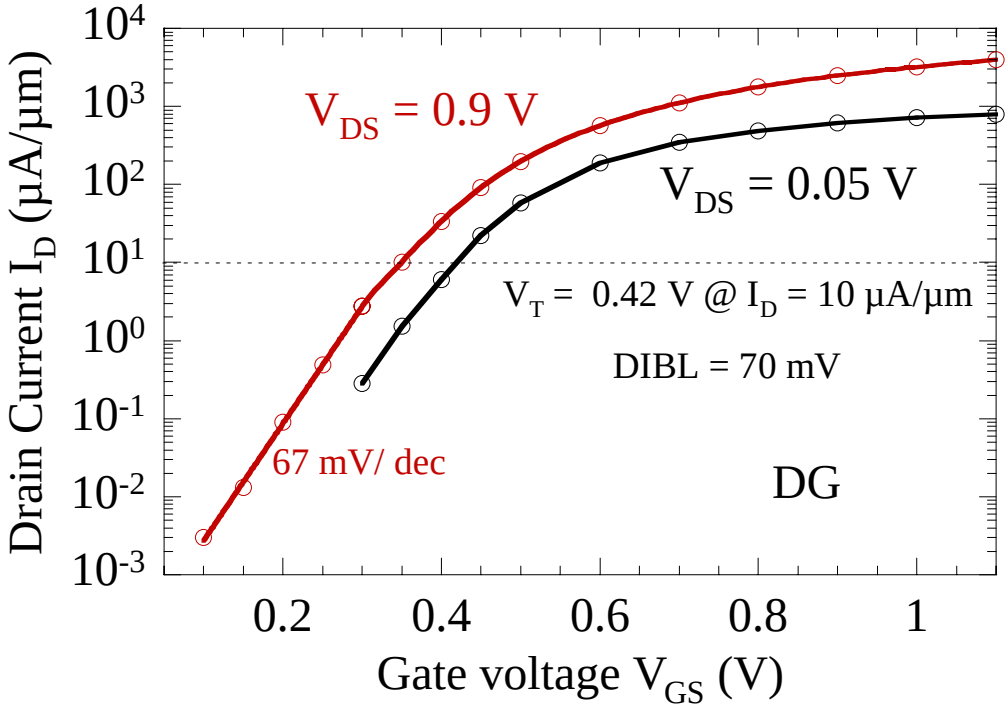
Querlioz et al. **Figure 1**



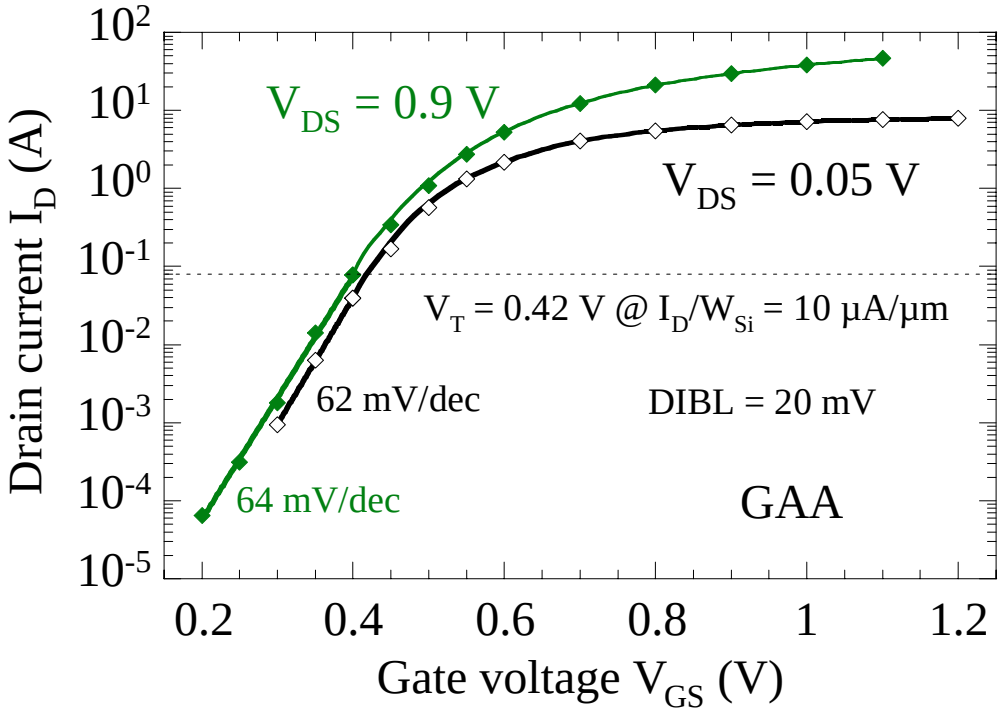
Querlioz et al. **Figure 2**



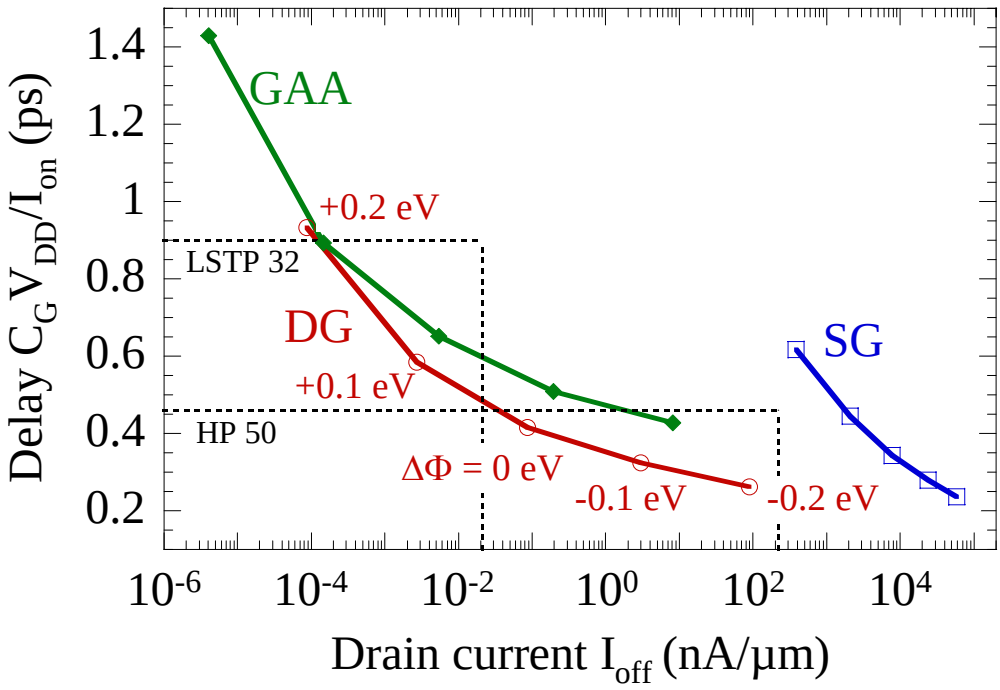
Querlioz et al. **Figure 3**



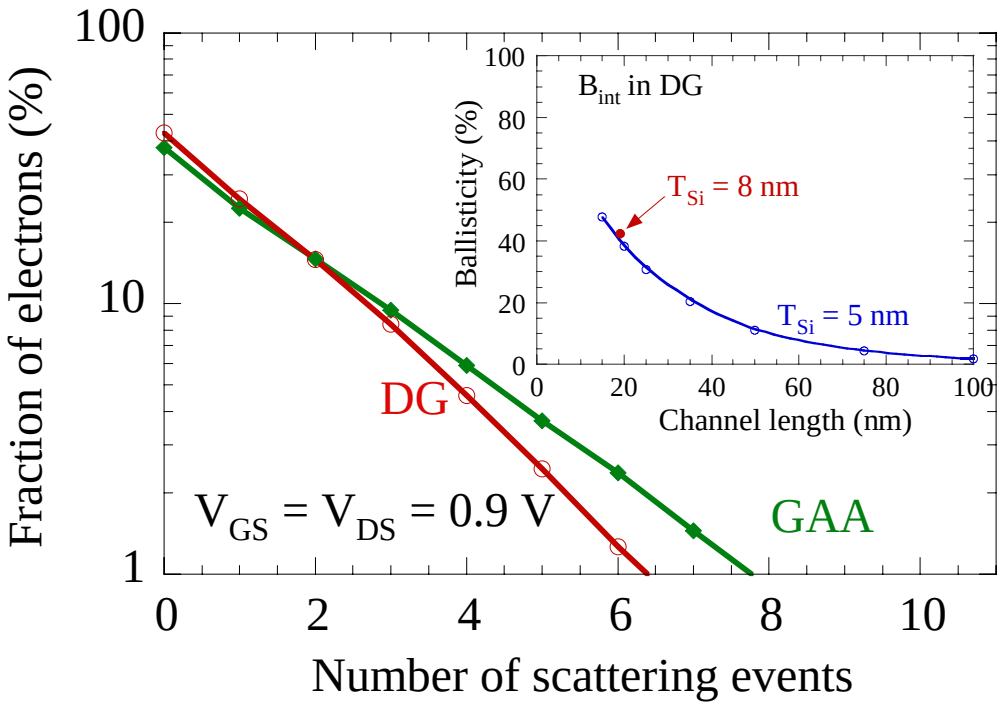
Querlioz et al. **Figure 4**



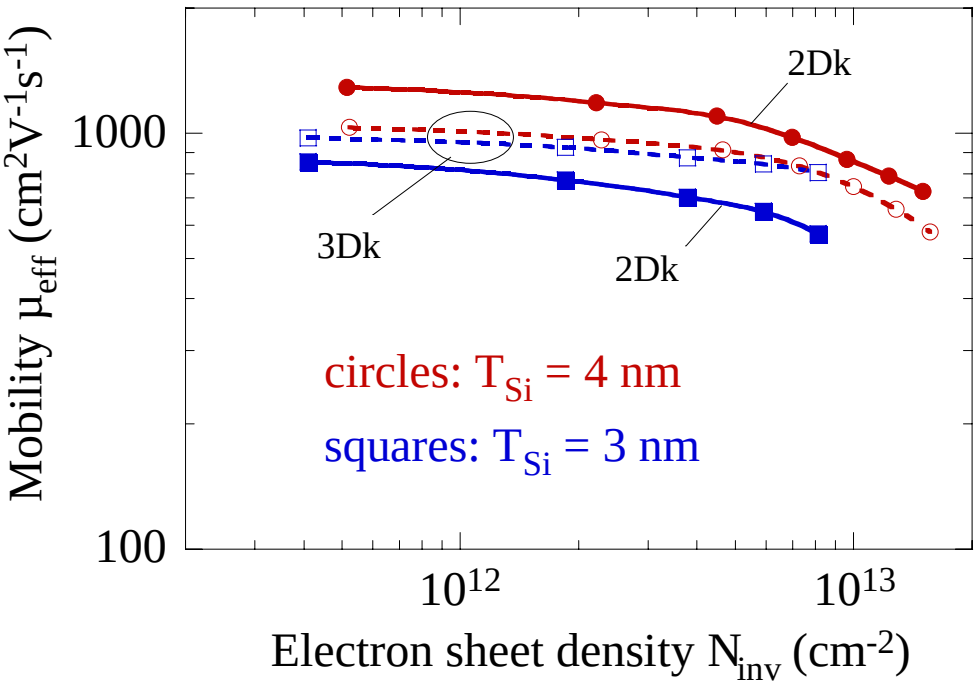
Querlioz et al. **Figure 5**



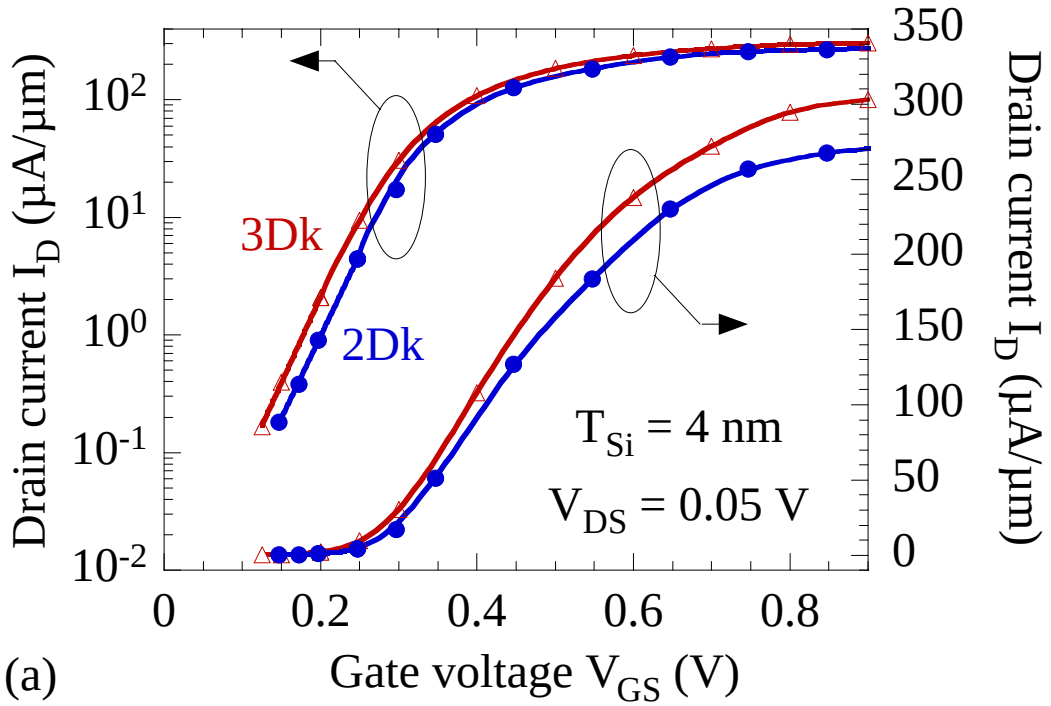
Querlioz et al. **Figure 6**



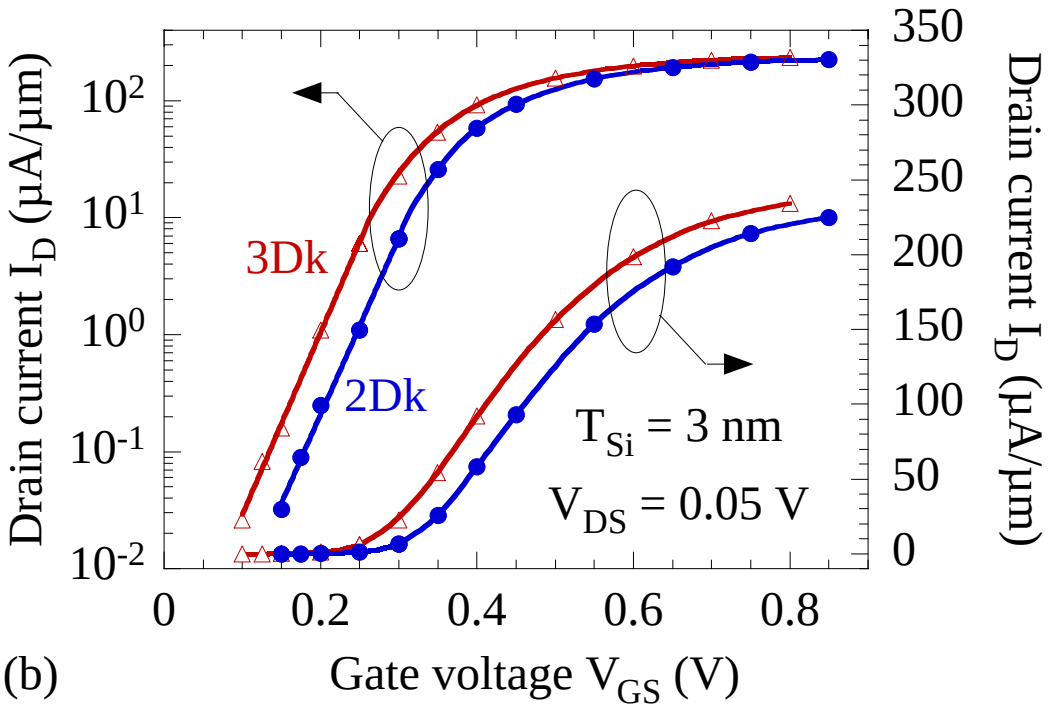
Querlioz et al. **Figure 7**



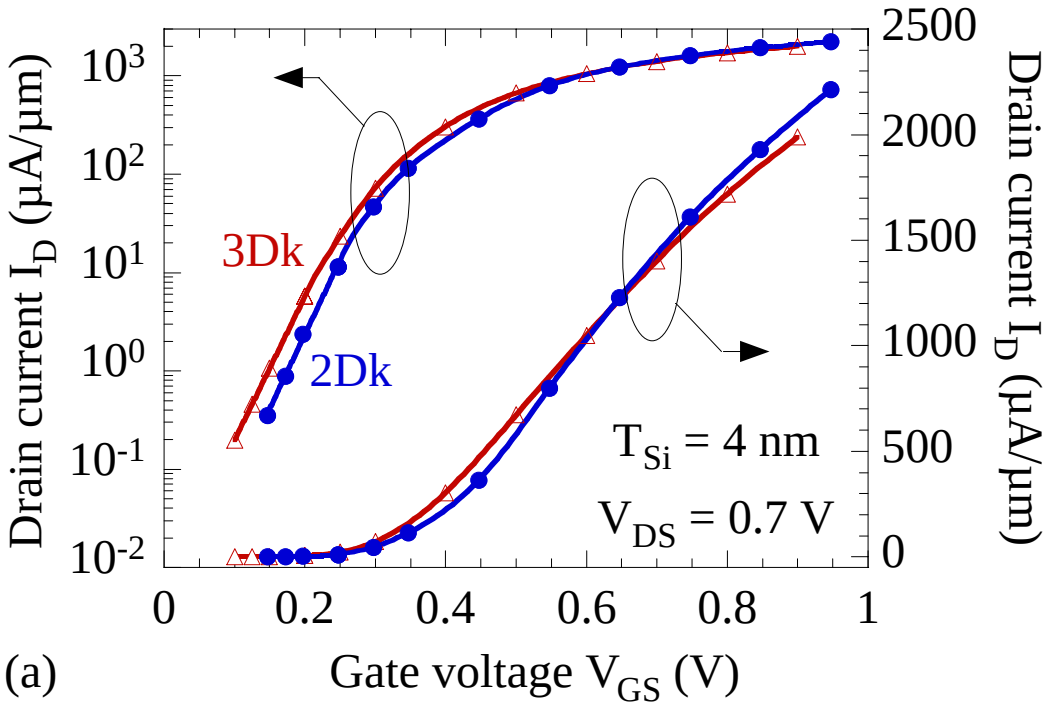
Querlioz et al. **Figure 8**



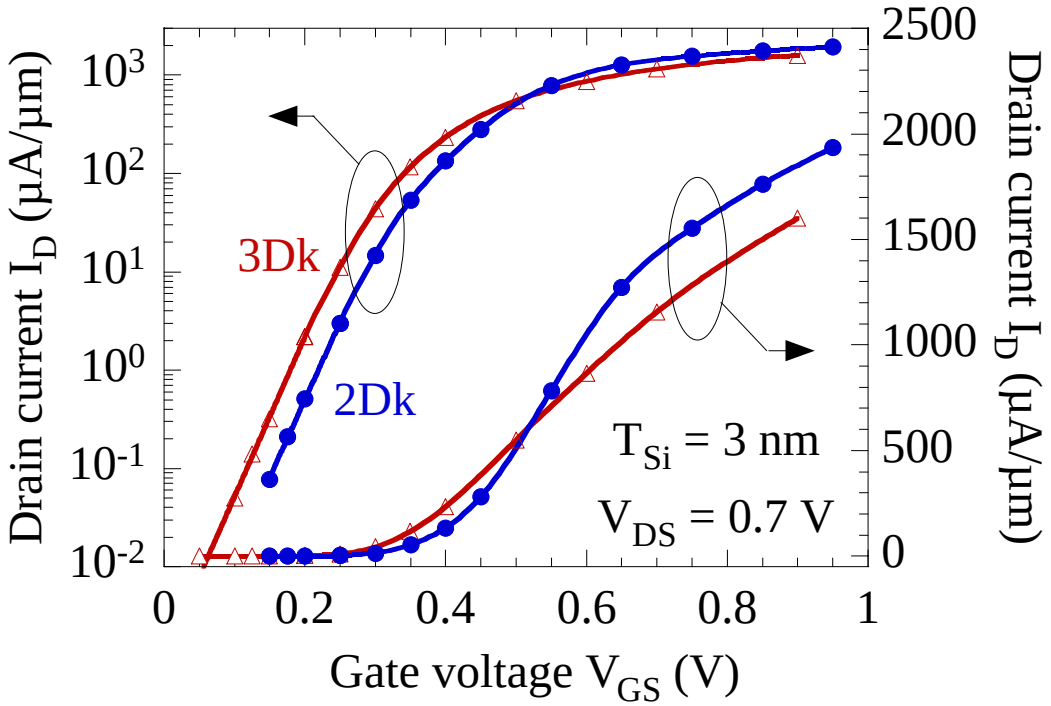
(a)



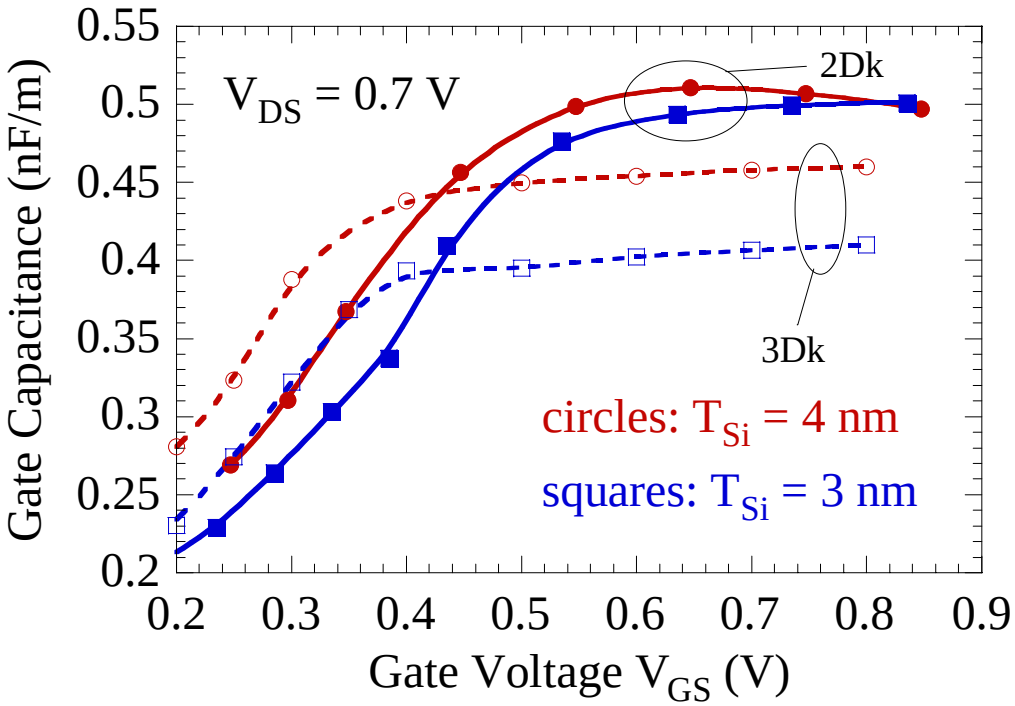
(b)


Drain current I_D (µA/µm)
Drain current I_D (µA/µm)
3Dk
2Dk
T_{Si} = 4 nm
V_{DS} = 0.7 V
Gate voltage V_{GS} (V)

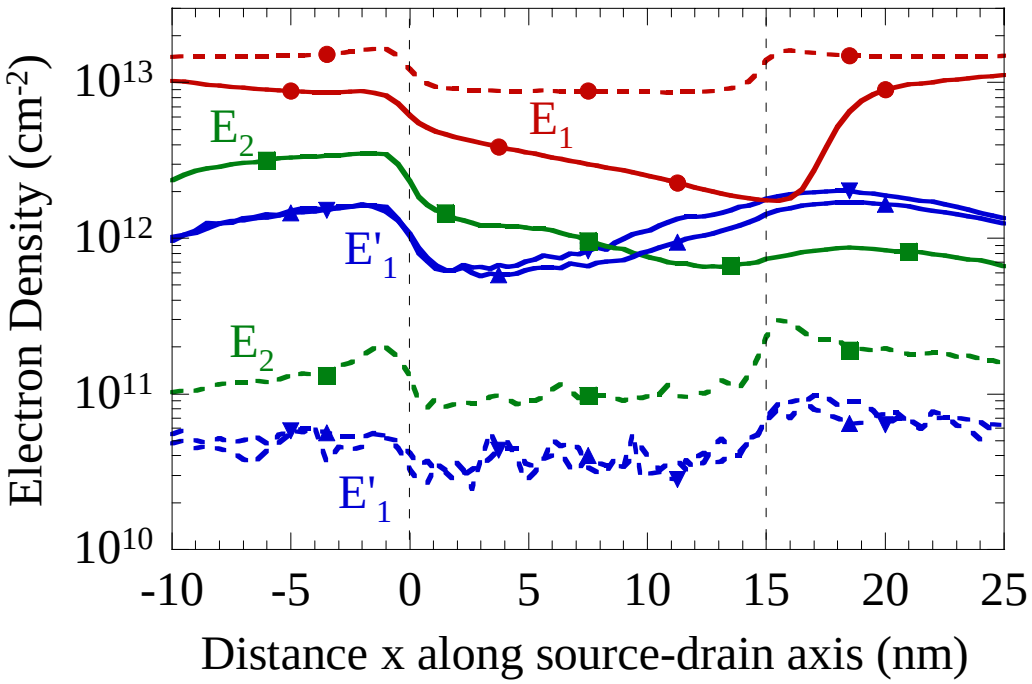
(a)


Drain current I_D (µA/µm)
Drain current I_D (µA/µm)
3Dk
2Dk
T_{Si} = 3 nm
V_{DS} = 0.7 V
Gate voltage V_{GS} (V)

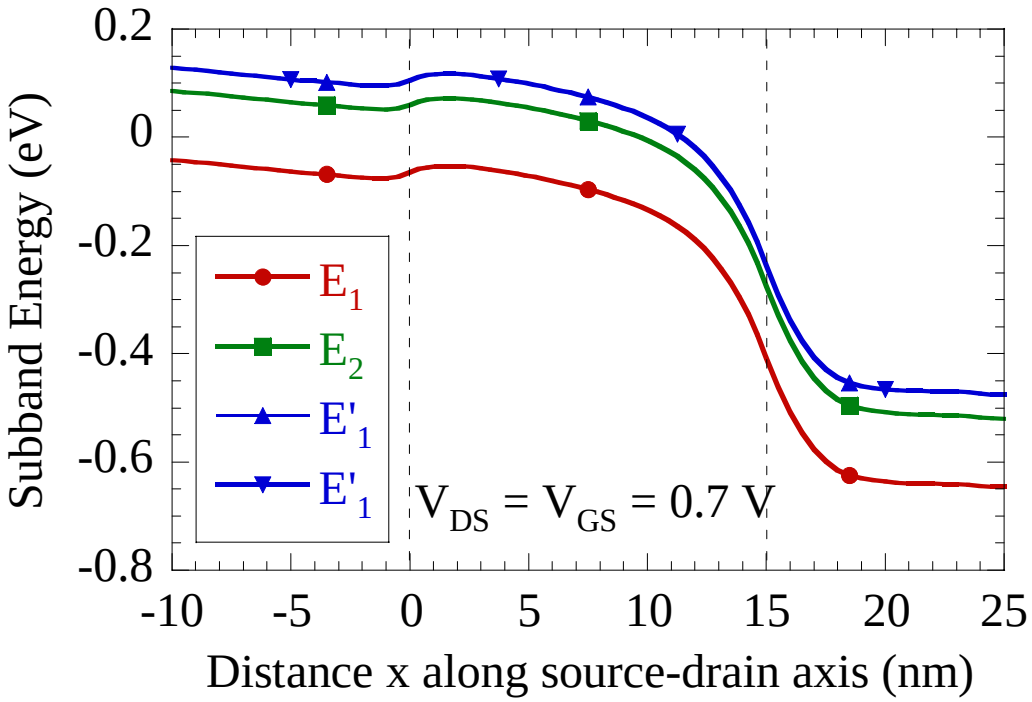
Querlioz et al. **Figure 10**



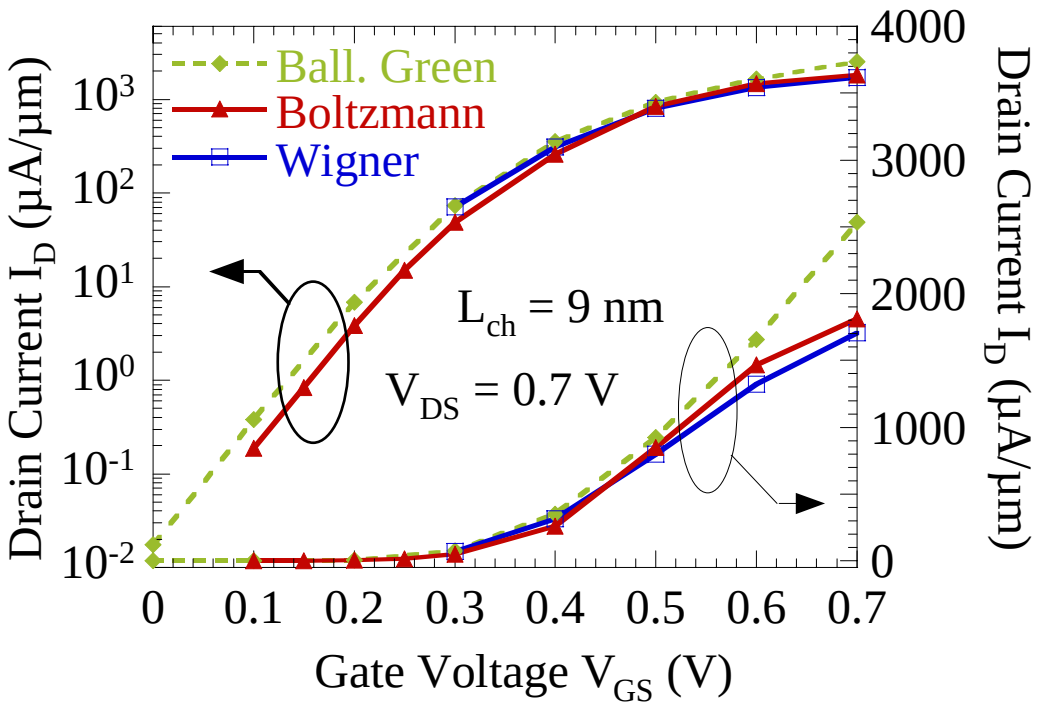
Querlioz et al. **Figure 11**



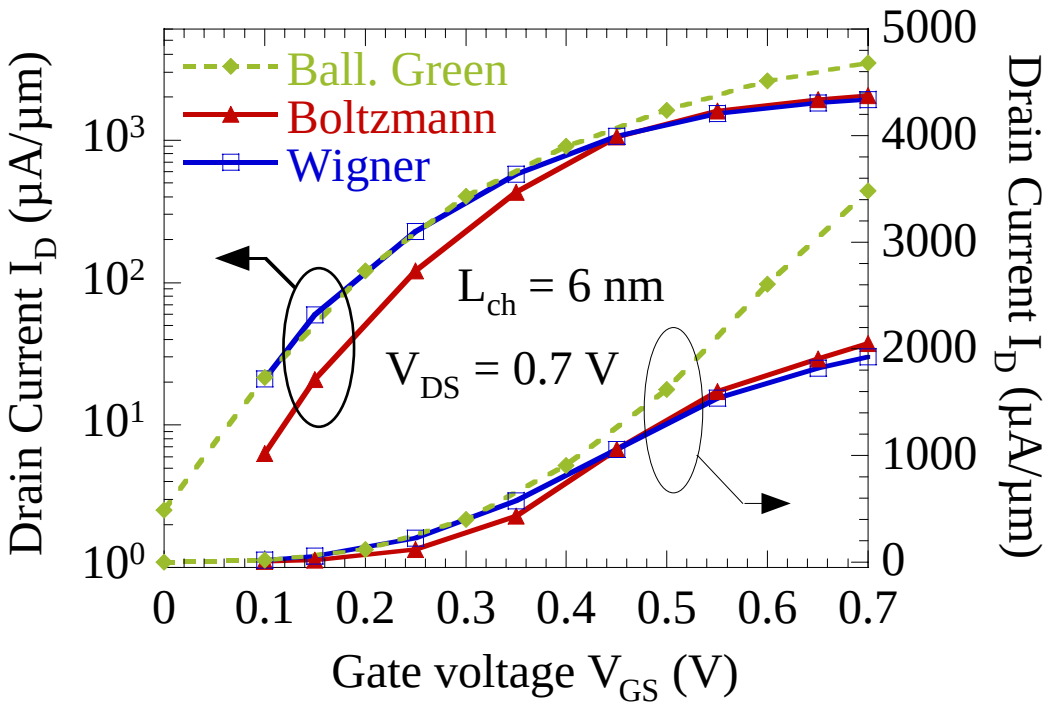
Querlioz et al. **Figure 12**

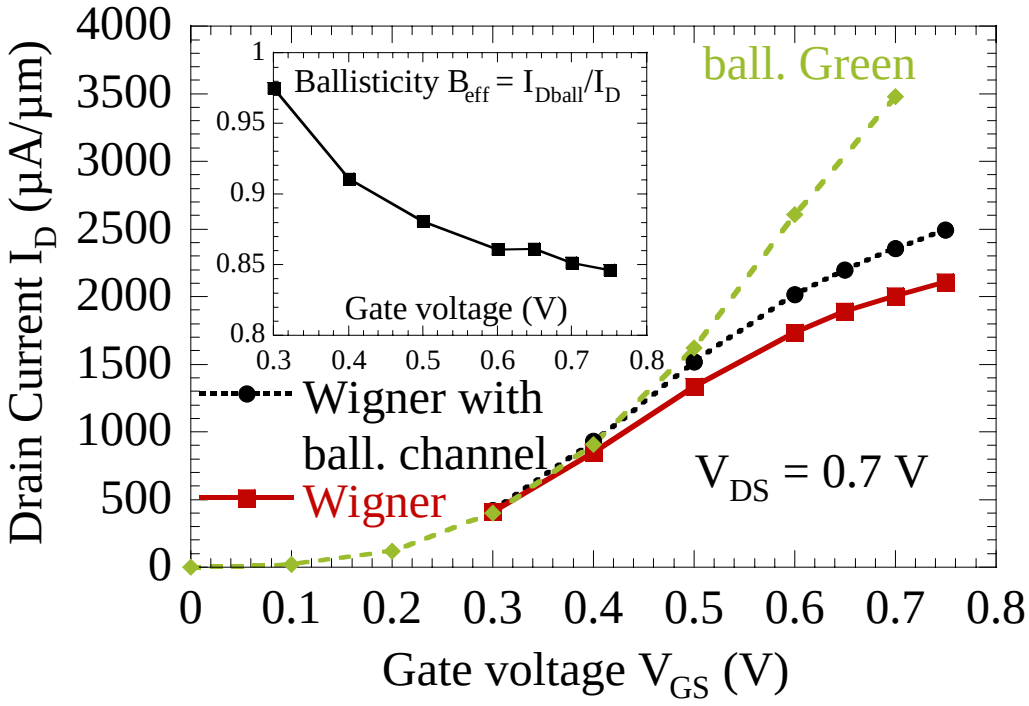


Querlioz et al. **Figure 13**

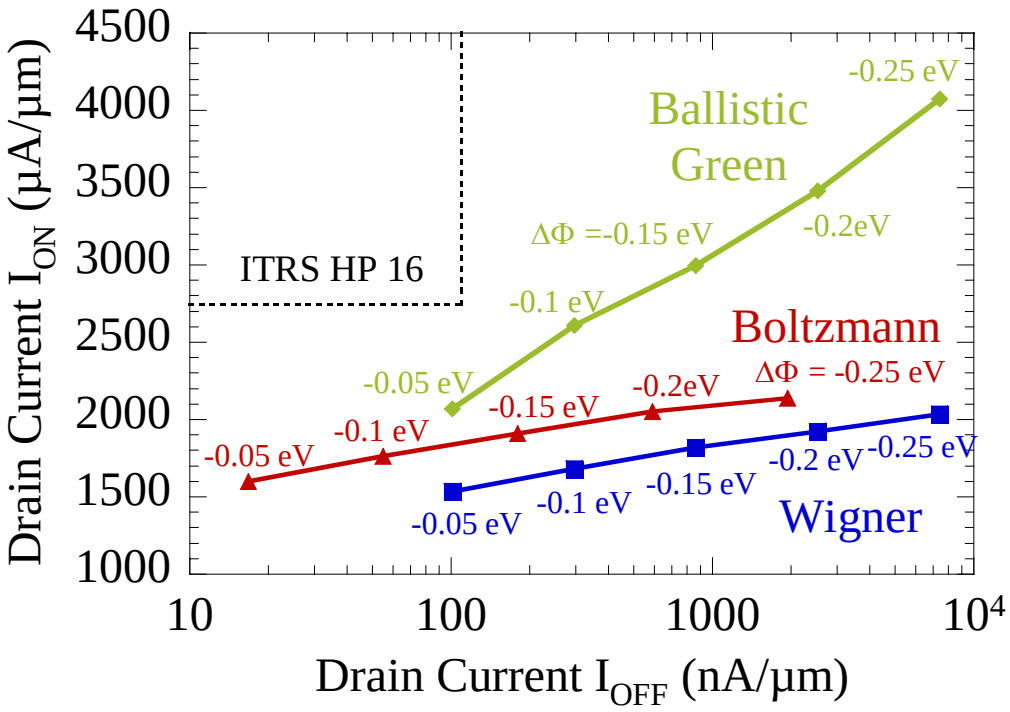


Querlioz et al. **Figure 14**




Drain Current I_D (µA/µm)
Gate voltage V_{GS} (V)
Ballisticity $B_{eff} = I_{Dball}/I_D$
Gate voltage (V)
ball. Green
Wigner with ball. channel
Wigner
V_{DS} = 0.7 V

Querlioz et al. **Figure 16**



Querlioz et al. **Figure 17**