A new F(ast)-CMS Algorithm for Efficient Three-Dimensional NEGF Simulations of Arbitrarily Shaped Silicon Nanowire MUGFETs

A. Afzalian, C.-W. Lee, N. Dehdashti Akhavan, R. Yan, I. Ferain, P. Razavi and J.-P. Colinge
Tyndall National Institute, University College Cork
Lee Maltings, Prospect Row, Cork, Ireland
e-mail: aryan.afzalian@tyndall.ie

***Abstract*— We present here 3D quantum simulations based on Non-Equilibrium Green's Function (NEGF) formalism using the Comsol Multiphysics™ software and on the implementation of a new Fast Coupled Mode-Space (FCMS) approach. The FCMS algorithm allows one to simulate transport in nanostructures presenting discontinuities, as the normal Coupled Mode-Space (CMS) algorithm does, but with the speed of a Fast Uncoupled-Mode Space (FUMS) algorithm (a faster algorithm that cannot handle discontinuities). Using our simulator, we also show that energy barriers resulting from cross-section variations at the gate edge of a nanowire can be optimized to improve the on/off current ratio. A subthreshold slope steeper than the kT/q.log(10) limit of classical transistors together with symmetrical source-drain operation is demonstrated for the first time using this new Variable barrier tunnel transistor (VBT) concept.**

Silicon on insulator technology, MOS devices, Quantum effect semiconductor devices, Quantum Wires, Semiconductor device modeling, Tunnel FET, steep subthreshold slope.

I. INTRODUCTION

As quantum effects are playing a crucial role in ultra-scaled devices, new device architectures and new simulations tools are needed to tame and exploit these new effects as much as possible. Among the new methods, the NEGF (Non-Equilibrium Green's Function) methods [1-2] has gained popularity. This method can, however, be very time consuming, which is a serious drawback for intensive device simulations. In an attempt to reduce simulation time, the mode space approach (MS) methods has been introduced and proved efficient owing to the strong lateral confinement in thin structures. In the MS approach, however, there is still a need to compute the wavefunctions and the energy levels in every cross-section along the transport x-direction. Up to now this has only been done efficiently enough to allow single processor operation in the FUMS (Fast uncoupled mode space) approach [2]. Due to the assumptions done, the FUMS has been shown to be valid only for nanowires with constant and small cross-sections [2]. As we will show here, however, constant cross-section transistors may not be the best option and by introducing local constrictions in the structure the on to off ratio of nanowires can be significantly improved. The more general CMS (coupled mode space approach) algorithm allows for the exploration of new device architectures for the end of the roadmap and the simulation of structures with any kind of shape, such as for example a change of cross section, the presence of tunnel barriers or discontinuities in silicon wire shape. The CMS method, however, increases simulation time by an order of magnitude and requires the use of computer clusters with parallel processing.

We have addressed this issue by developing a new fast algorithm which combines the advantages of both methods: the fast coupled mode-space approach, FCMS. It allows one to simulate non-constant cross-sections at a speed nearly equal to that of FUMS, by using only the CMS approach locally where it is needed (e.g. near discontinuities).

II. FCMS ALGORITHM:

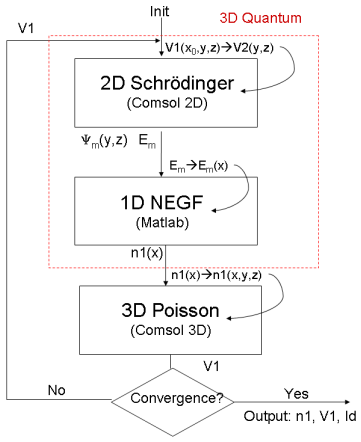

Figure 1. Computation algorithm for device simulation.

We have developed a 3D quantum simulator which is based on the non-equilibrium Green functions using the Comsol Multiphysics software and on an implementation of our fast coupled mode space (FCMS) approach. The main loop of the program computes the electrostatic potential in the device, V_1, self-consistently, solving Poisson equation, and the electron concentration, n_1, using the non-equilibrium Green functions (NEGF) [1]. As we are using a mode-space approach, the coordinate y and z of the cross section perpendicular to the

transport direction, x, of the device are replaced by the mode energies $E_m(x)$ of the electronics subband in the devices. This allows one to drastically reduce the computational budget as in practice only the first few subbands are occupied by electrons and need to be calculated [2]. If M is the number of occupied subbands, the size of the Hamiltonian matrix to invert in the MS will be $(M.N_x)^2$ instead of $(N_y.N_z.N_x)^2$ in a real space (RS). N_x, N_y and N_z are the number of mesh points in the x, y and z directions. Typical number of mesh points ranges between 50 to 100 in each directions, while M takes a typical value between 1 and 10 for nanowires with a cross-section diameter below 10nm. We have, however, to provide the subband energy profile $E_m(x)$ and the wave functions $\Psi_m(x,y,z)$ related. This can be done by solving a 2D Schrödinger problem in the cross section of the device. The computation algorithm is summarized in Fig.1.

In the FUMS algorithm, the wave function in the cross section of the device is assumed to be constant along the transport direction. One further assumes that this constant wave function is the solution of the 2D Schrödinger equation (in the cross-section) using the average value of the 3D potential, $V_1(x,y,z)$ in the x-direction, $V_2(y,z)$. In this case, we have to compute the wavefunction only once. This has been shown to be only valid for nanowires with a constant cross-section below 10 nm × 10 nm [2].

In the CMS algorithm, these assumptions are not needed. This allows for the simulation of structures with variable cross-section, tunnel barriers or other types of discontinuities in a semiconductor nanowire device. This, however, imposes to compute a full 2D Schrödinger problem for each x-mesh point (there are typically a hundred of mesh points along the x-direction of a device). Fig. 2 shows the evolution of the first wavefunction in the center of the Si body of a constant 3nm×3nm cross-section nanowire with a channel length L=10nm and a source/drain extension length L_{sd}=10nm, computed using the CMS method. The variations of the wavefunction are small enough to allow for the use of the FUMS technique in such a structure.

In Fig. 2, we can also see the evolution of the first wave functions in the center of the Si body of a nanowire with same geometry but with a thin oxide barrier (1nm) between the source extension and the body of the transistor. The wavefunction is no longer constant and, therefore, any attempt to solve this problem with a FUMS algorithm will lead to wrong results or even to non-convergence. Simulations results show, however, that the wavefunction shape varies only in the immediate vicinity of the barrier. At x=0, which corresponds to the end of the barrier and the beginning of the body, the wave function is already very close to its value far from the barrier. The same holds for x=-1.5nm, which is 0.5 nm (or 1 x-mesh point) to the left of the barrier. Simulations of different kinds of discontinuities (variation of the cross-section size, introducing a thin layer of a different material or fixed charges,…) ranging from a few tenth of nanometers to several nanometers has lead us to the conclusion that the electron wave function perturbation can be considered as quasi-local at this scale and that it is in equilibrium with the local dimensions of the structure.

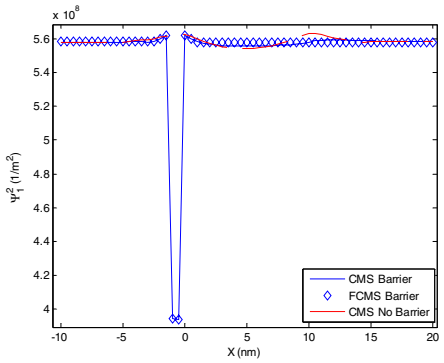

Figure 2. Evolution of the first wave function in the center of the Si body of a constant 3nm×3nm cross-section nanowire without (Resp. with) a thin oxide barrier of 1nm inserted between the source extension and the body of the transistor computed with the CMS (resp. with the CMS and FCMS). L=L_{sd}=10nm.

The FCMS method proposed here is based on this observation. The basic assumption of the FCMS method is that the variation of the wavefunction is localized around the discontinuities but is constant far from them. We can then neglect the coupling in the Hamiltonian matrix and solve one 2D Schrödinger equation with an x-averaged potential in the cross-section and derive the energy subbands in each part identified as constant as in the FUMS algorithm (*i.e.* in the example of Fig. 2, for x=[-10nm to x=-1.5nm[and for x=[0nm to x=20nm]) and take into account the coupling between different device sections in the Hamiltonian and solve in each cross-section (at each x-mesh points) around the discontinuities (i.e. for x=[1.5nm to x=0nm[) a 2D Schrödinger equation with the real potential in these cross-sections as in the CMS algorithm. In our example (see Fig. 2), with a typical x-mesh spacing a=0.5nm, we can obtain the energy and the wave function profile needed for the NEGF in FCMS by solving only five 2D Schrödinger problems instead of the sixty-one 2D Schrödinger problems requiring being solved using the CMS method.

TABLE I. COMPARISON OF 3D NEGF METHODS:

Method	time/processor *(typical, relative unit)*	Cross-Section Shape	Run on standard PC
FUMS (Fast Uncoupled Mode Space)	**1**	Constant only	**yes**
CMS (Coupled Mode Space)	5-10	**any**	Possible but long
RS (Real Space)	>100	**any**	No, need for parallel processing(typ>100 processors)
Our FCMS (Fast Coupled Mode Space)	**1.2**	**any**	**yes**

The simulations with different cross-sections, channel length, number of energy and mesh points show that the

increase in simulation time in FCMS with 10% of discontinuity along the transport (x) direction (*e.g.* a 3nm length local discontinuity in a device with a total length L_X of 30nm) does not exceed 20% while allowing to simulate discontinuities with the same precision than CMS (see for example Fig.3.Right). The FUMS approach is typically between 4-10 times faster than the CMS method on a 2Ghz 64-bit Dual-Core Pentium PC. Table 1 summarizes the different NEGF methods, their relative speed and performances for 3D simulations.

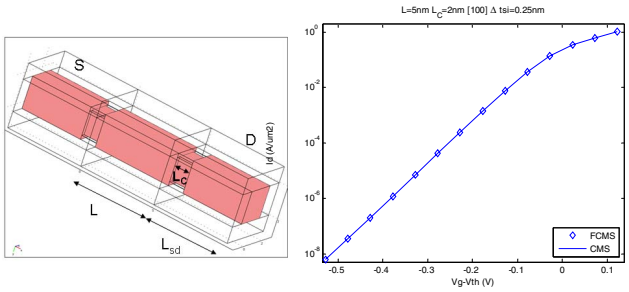

Figure 3. Schematic view of a gate-all-around (GAA) nanowire with 2 constrictions of the semiconductor cross-section at the Source extension (S)-body and body-Drain extension (D) boundaries. Right: Id-Vg curves simulated with CMS and FCMS with 2 constrictions of the semiconductor cross-section with a diameter reduction Δt_{si}=0.25 nm. L_C=2nm.

III. IMPROVEMENT OF THE SWITCHING CHARACTERISTICS WITH VARIABLE TUNNEL BARRIERS

Using our simulator we have investigated the effect of local cross-section changes in the transport properties of nanowire MOSFETs (Fig. 3). The effect of cross-section variations in a nanowire results in energy barriers in the conduction band (a N-Mosfet is considered here, but similar behavior in the valence band is obtained for a P-MOSFET) that can be interpreted as a local increase of the bandgap [3]. In narrow wires with a cross section of a few nanometers, this can strongly influence the current and the characteristics of the device (Fig. 4). A small constriction resulting in a barrier of the order of a 0.1eV (with a diameter reduction Δt_{si}=0.25 nm) is shown to be an effective way to create a tunnel barrier that can be used to improve the on/off current ratio in transistor with channel length equal or smaller than 5nm.

Two effects are responsible for the improvement of the on/off current ratio and the subthreshold slope:

In very short channel devices, L≈5nm and below, a small constriction resulting in a barrier of the order of a 0.1eV is able to reduce significantly the problem of the current that tunnels under the channel barrier and degrades the slope of the Id-Vg characteristics in subthreshold regime (Fig. 5). Increasing the height of barrier by increasing Δtsi up to 0.5 nm does not reduce much further the off-current, while resulting in a drastic reduction of the on current and, therefore, of the overall on/off current ratio (Fig. 3) [3]. The channel barrier tunneling current decreases for increasing channel length, so that above 7.5nm, this effect is not significant.

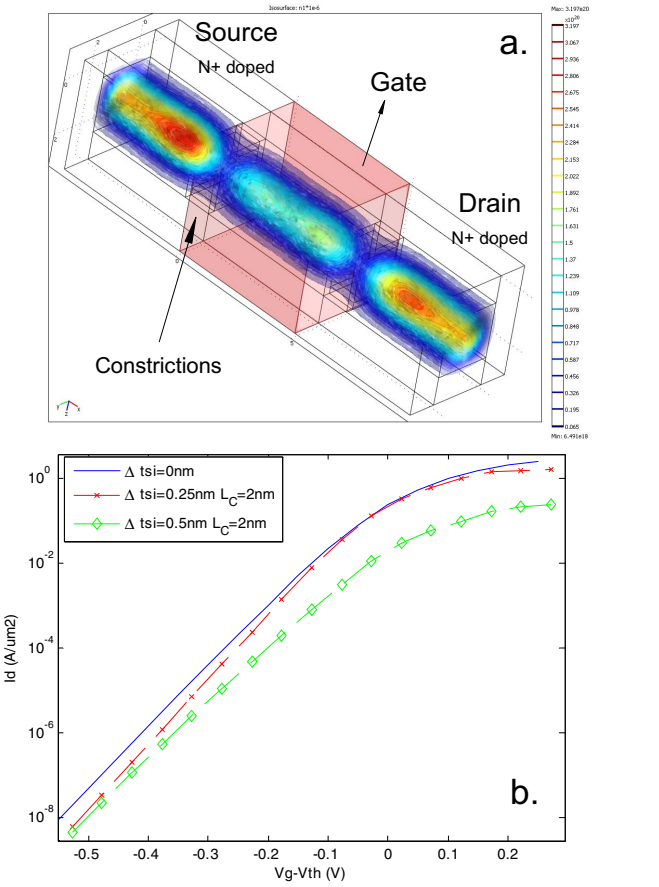

Figure 4. Simulation results obtained with FCMS in a [100] 2×2 nm^2 GAA nanowire with the 2 constrictions. L_{sd}=L=5nm. V_d=0.1V. a) Electron concentration Δt_{si}=ΔW_{si}=0.25nm. V_g-V_{th}=0.25V. The reduction of the electron concentration value in the constrictions due to the created barrier can be clearly observed. b) I_d-V_g curves without and with 2 constrictions of the semiconductor cross-section with Δt_{si}=0.25 nm and Δt_{si}= 0.5 nm. The on to off current ratio of the transistor with a constriction of 0.25nm is improved by 4.

There is however a second effect that improves the subthreshold slope in short and long transistors: By placing the constrictions at the edge of the channel, the fringing electric field modulate the height of the barrier so that the transparency of the barrier is increased when increasing the gate voltage: The electrostatics of the regions at the edge of the gate are mostly controlled by the gate below threshold, and by the source and drain voltage above threshold. As a result, the energy bands at the edges of the gate are bent upwards for negative value of (V_G-V_{th}), *i.e.* below threshold, while they are pulled downwards for V_G>V_{th}, *i.e.* above threshold. By placing a barrier such as a constriction at the edges of the gate (Fig. 6.a), the natural movement of the energy bands with gate voltage can be used to improve the control of the transparency of the barrier in such a way that the off current is blocked more efficiently that the on current. The simulation results show that by properly selecting height, width and shape of the barrier, the on-to-off current ratio and the subthreshold slope can be significantly improved, compared to a classical transistor even in transistor where there is no significant tunnel barrier tunnelling current. The subthreshold slope can even be smaller than KT/q*log(10) or 59.6mV/dec at T=300K (Fig. 6.b).

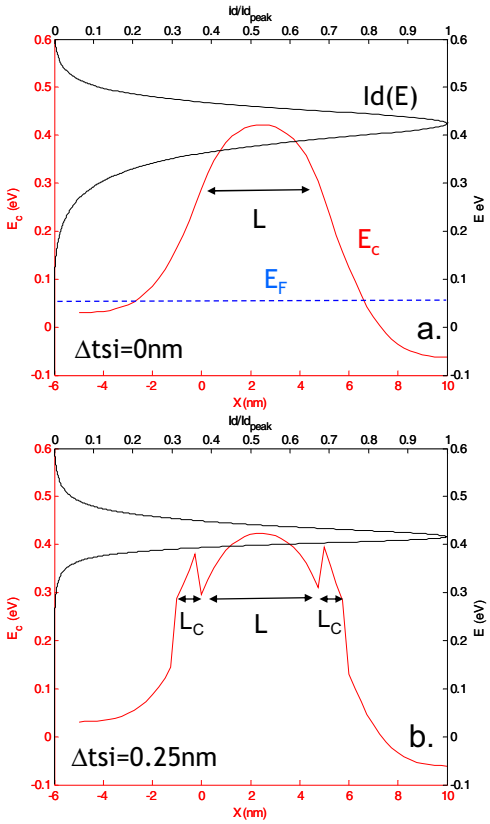

Figure 5. Conduction band vs. transport direction (x) and normalized spectral density of current in the nanowire of Fig.4 at Vg-Vth=-0.35V without (a) and with constrictions at the gate edge (Δtsi=0.25nm) (b).

The device is in fact part of a new more general concept of tunnel field-effect (TFET) transistor, that we have called the Variable-Barrier Transistor (VBT) [4], which is able to break the 60mV/dec subthreshold slope limit barrier of classical transistors. The device is based on tunnelling of electrons within the conduction band (intraband tunnelling) rather than the more familiar interband tunnelling through a reverse-biased P^+N^+ junction (with a P-type source and an N-type drain) [5-9]. This allows for symmetrical transistor TFET operation (as opposed to diode-like operation) for the first time.

IV. Conclusions

We have presented here a new 3D quantum simulation algorithm based on Non-Equilibrium Green's Function (NEGF) formalism using the Comsol Multiphysics™ software and on the implementation of a new Fast Coupled Mode-Space (FCMS) approach. The FCMS algorithm allows one to simulate transport in nanostructures presenting discontinuities, as the normal Coupled Mode-Space (CMS) algorithm does, but with the speed of a Fast Uncoupled-Mode Space (FUMS) algorithm (a faster algorithm that cannot handle discontinuities). Using our simulator, we have shown that energy barriers resulting from cross-section variations at the gate edge of a nanowire can be optimized to improve the on/off current ratio. Subthreshold slope steeper than the kT/q.log(10) limit of classical transistors together with symmetrical source-drain operation is demonstrated for the first time using this new Variable barrier tunnel transistor (VBT) concept.

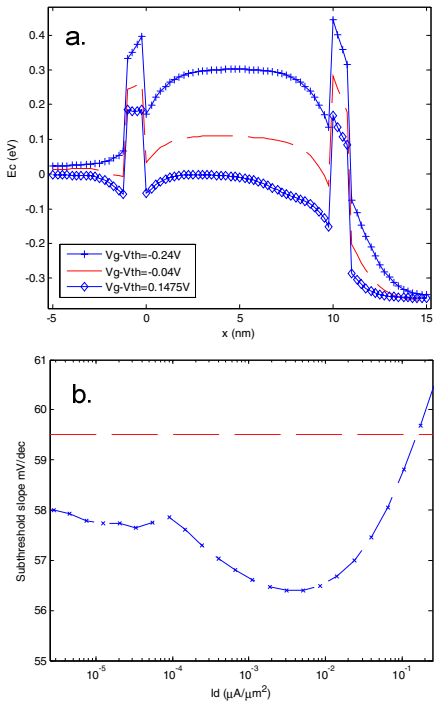

Figure 6. Evolution with V_G-V_{th} of the 1st subband *vs. x* profile in the 2nm × 2nm [100] GAA nanowire with the two constrictions of the semiconductor cross-section. $\Delta t_{si}=\Delta W_{si}=0.5$ nm and 0.65nm at Source and Drain side respectively (a). Subthreshold slope (b). L=10nm. V_d=0.4V. The length of the constrictions, L_C, is 1nm.

Acknowledgement:

This work is supported by Science Foundation Ireland under Grant 05/IN/I888.

References

[1] S. Datta, Superlattices and Microstructures, Vol. 28 (4), pp. 253-278 (2000).

[2] J. Wang *et al.*, J.A.P., Vol. 96 (4), pp. 2192-2203 (2004).

[3] A. Afzalian *et al.*, ECS Transactions, Vol. 19 (4), pp. 229-234, 2009.

[4] A. Afzalian et al., *to be published.*

[5] Reddick W, Amaratunga G., "Silicon surface tunnel transistor". Appl. Phys. Letters, 1995; 67:494–6.

[6] Koga J, Toriumi A. "Three-terminal silicon surface junction tunneling device for room temperature operation." IEEE Electron Dev. Letters, 1999; 20:529–31

[7] Qin Zhang, Wei Zhao, and Alan Seabaugh, "Low-Subthreshold Swing Tunnel Transistors", IEEE ELECTRON DEVICE LETTERS, VOL. 27, NO. 4, APRIL 2006,p. 297.

[8] J. Knoch, S. Mantl, J. Appenzeller, "Impact of the dimensionality on the performance of tunnelling FETs: Bulk versus one-dimensional devices", SOLID-STATE ELECTRONICS Vol. 51, Apr. 2007, pp. 572-578.

[9] Choi W.Y, Park B.G, Lee J.D, Liu T.-J.K., "Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec", IEEE ELECTRON DEVICE LETTERS, VOL. 28, NO. 8, AUGUST 2007, pp. 743-745.