NANOSCALE TRANSISTOR MODELING

R. Venugopal, Email: venugopr@ecn.purdue.edu, Phone: 505-663-1612
Purdue University, West Lafayette, IN USA 47907-1285

1. Motivation

As the dimensions of transistors shrink below 70nm, they are becoming increasingly sensitive to quantum mechanical confinement, tunneling, and to atomic scale structure effects in contacts, junction edges, oxide/silicon interfaces, and in the oxide itself. Our group at Purdue is exploring device limits and new approaches through quantum scale device simulation. This work addresses the need for the semiconductor device community to learn to think and work at the quantum scale, a key message from a Feb. 1999 Nanotransistor Workshop [1]. Two complementary approaches are being pursued. At a detailed level, the non-equilibrium Green's function (NEGF) approach provides a rigorous treatment of quantum transport on an atomic scale. At a simple level, the scattering (or transmission) approach provides a way to interpret detailed simulations and experiments and to propose new ones [2-3]. It will allow us to capture the essential physics of nanotransistors as illuminated by NEGF simulations in a way that can be communicated to the broader device community. This project is directed at the large-scale computing issues involved in simulating MOSFETs (Metal-Oxide-Semiconductor field effect transistor) by the NEGF method.

2. Background and Objectives of the Work

The objective of this work is to demonstrate 2D quantum models for sub-70 nm MOSFET's along with the development of numerical algorithms to make "full quantum" simulations (including bandstructure and scattering details) feasible. This work is being directed by Prof. Supriyo Datta, whose prior work in this area uniquely qualifies him to address this very challenging task. Earlier in this decade, along with his students McLennan, Lake and Klimeck, he showed that the non-equilibrium Green function (NEGF) formalism could be used to describe devices taking quantum interference and dissipative scattering into account [4-6]. A group at Texas Instruments subsequently developed a powerful 1D device simulation tool based on this formalism. This program, NEMO, is the standard in the field and is the basis for a new, quantum level MOSFET simulator, QUARTZ, being developed at Motorola (Jovanovic). Recently, Datta and his students Samanta and Tian have applied the NEGF formalism to the description of current flow in molecular conductors using an atomic Hamiltonian (rather than an effective mass Hamiltonian) as the starting point [7]. To extend the NEGF approach to devices (as opposed to a molecule with ~ 100 atoms) is a significant challenge. The 1999 National Technology Roadmap for Semiconductors lists 2D quantum modeling of MOSFETs as one of the "difficult challenges" for modeling and simulation below 100 nm. The development of a full quantum, 2D MOSFET simulator will require a full-scale attack on the problem by focusing on critical physics and numerical issues.

3. The NEGF Method

Transistors, in general can be classified into two categories. Those in which the primary current carriers are electrons (NMOSFETS) and those in which the primary current carriers are holes

(PMOSFETS). A simple transistor structure along with the various voltage specifications and the direction of current flow is shown in figure 1.

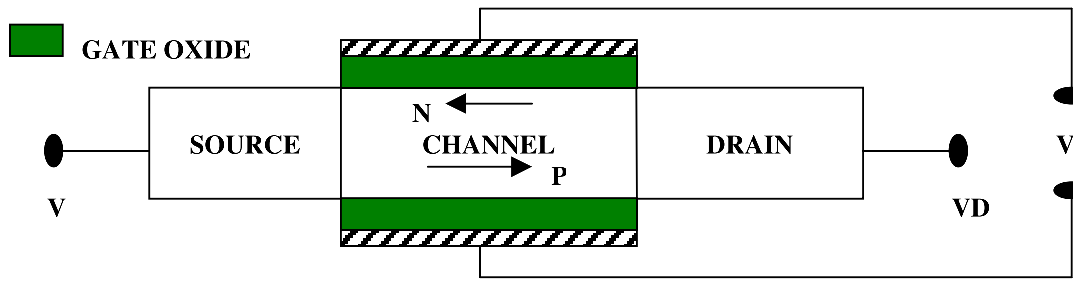


Figure 1: A dual gate MOSFET structure

The actual structure of the source/drain and channel can be quite complicated depending on the profile of the donor and acceptor dopant concentrations, engineered to meet the performance requirements specified in the National-Technology-Roadmap-For-Semiconductors (NTRS). In order to capture the physics of quantum transport (electrons/holes) in such nanoscale transistors we resort to the use of non-equilibrium quantum field theory. In this picture, any transistor can be looked upon as a system sandwiched between two reservoirs (source/drain) and driven out of equilibrium by two contacts with different Fermi levels. This representation of the transistor as well as the solution to the non-equilibrium problem is illustrated in figure 2.

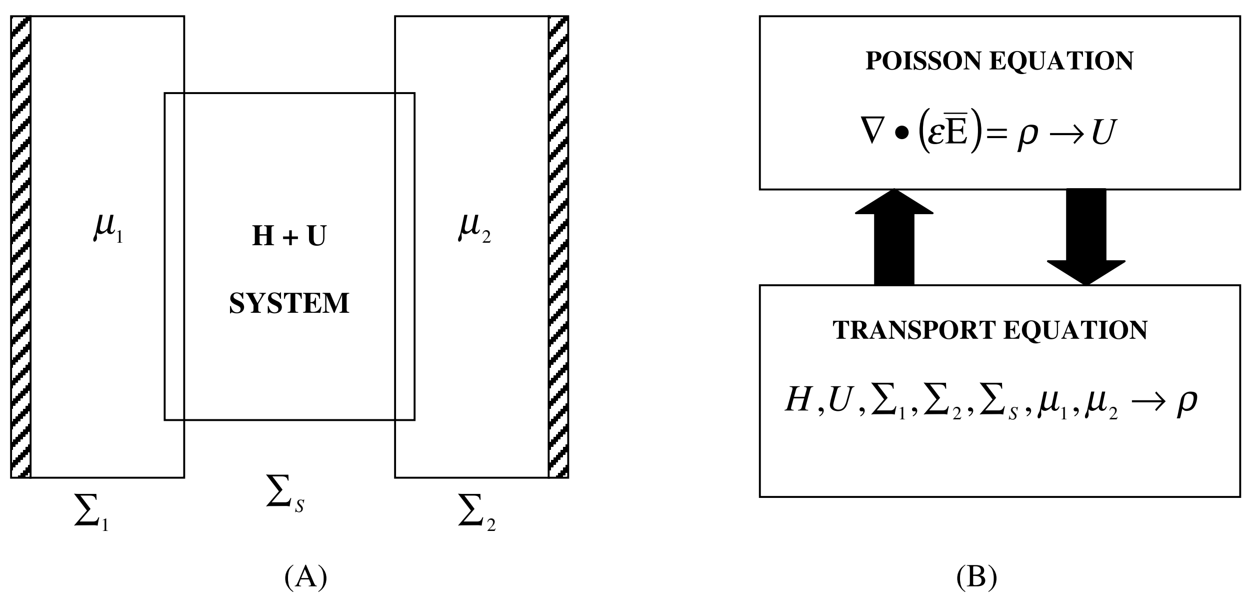


Figure 2: (A) A device driven out of equilibrium (B) Self consistent solution for the density matrix

Note that any transport equation has to be solved self consistently with Poisson's equation in order to obtain all quantities of interest such as charge density, current etc. A solution to the transport equation yields the density matrix, from which any quantity of interest can be derived. For example, the charge density as a function of position within the transistor is obtained from the diagonal entries of the density matrix in real space representation. The problem is to find the density matrix in the correct representation. For this, it is not enough for the transport equation to capture the system details through the Hamiltonian, it should also specify how the system is coupled to the contacts and the scattering processes that are effective within the system. This information is contained in the self energy functions ($\sum_1, \sum_2, \sum_S$). Thus the final discretized form of the transport equation also known as the Dyson equation is given by,

$$\left[\mathrm{EI} - \mathrm{H} - \Sigma_1 - \Sigma_2 - \Sigma_S\right]^{-1} G^R(\mathrm{E}) = \mathrm{I} \qquad (1)$$

where $G^R(E)$ is the retarded green's function. The density matrix can be obtained from the retarded Greens' function using the following relation,

$$\rho(\mathrm{E}) = \frac{i}{2\pi}\left[G^R(\mathrm{E}) - G^R(\mathrm{E})^+\right] \qquad (2)$$

All of the physical approximations (bandstructure, Coulombic interactions etc.) are contained within the Hamiltonian. The Hamiltonian is the discretized version of the following operator,

$$\mathrm{H}_{op} = \frac{-h^2}{2m}\nabla^2 + U_L + U_C \qquad (3)$$

where U_L is the lattice potential and U_C the coulomb potential of the system. Note the explicit dependence of eqns. 1 and 2 on the total energy of the system. These equations have to be solved on a fine energy grid in order to obtain an accurate internal picture. A solution to eqn. 1 self consistently with poisson's equation poses a severe computational burden as the matrices to be dealt with are unsymmetric and complex in general.

4. The Numerical Challenge

The numerical problem is essentially one of size. Consider, for example, a nanoscale MOSFET with an active area of say 10nm x 10nm. A quantum model using atomic wavefunctions as the basis will require a grid-point spacing of say 0.2 nm with ten orbitals (sp^3s^* for both cation and anion) per grid point. This means a total of ~ 2500 grid points with 10 basis orbitals at each grid point (due to spin orbit coupling). Quantum simulations in general require correlations between each pair of orbitals ($G^R(r,r')$). This means that we will need to store, recall and manipulate complex unsymmetric matrices whose dimension is ~ 25000 x 25000 (for every point in energy space), which is well beyond the capability of present-day computers. We have taken an important step towards a solution to this challenging numerical problem. It is evident that the quantum problem is complicated by the requirement of a grid size with atomic dimensions. Simpler semi-classical models by contrast require a grid-size that is determined by the spatial scale on which the potentials and carrier densities in the device vary. The obvious question is:

can quantum models be developed that work with grid-sizes determined by the length scales on which observable quantities vary, rather than the atomic length scales on which wavefunctions (particularly their phases) vary? We have shown that this is indeed possible. Starting from a fine atomic scale grid we can eliminate nodes using an exact algorithm to produce a coarser grid whose size is determined by the spatial scale of the potential variations. The final grid could also be non-uniform; for example, we could eliminate more points in the reservoir regions where potentials are relatively uniform. This ability to use a relatively coarse grid of variable size, while maintaining accuracy represents a crucial step towards the development of atomistic models for realistic devices.

5. Status of Current Work

Our initial work has been focussed on understanding the physics of quantum transport in ultrasmall transistors. To greatly reduce the computational burden during these preliminary studies, we solved the coupled NEGF and Poisson equations in a nanoscale transistor neglecting scattering and invoking a parabolic bandstructure approximation. The effect of these assumptions is to reduce the Hamiltonian for the system to a simple form given by,

$$\mathrm{H}_{op} = \frac{-h^2}{2m^*}\nabla^2 + U_C \qquad (4)$$

where the effect of the lattice potential has been absorbed into an effective mass (m*) for the carriers. Also, the simplifying assumptions enable us eliminate $\sum_S$ (a quantity that had to be self consistently computed) and the need for a basis of atomic orbitals. Instead we could work within a reduced real space framework and explore the ultimate limits of silicon technology. The result of such calculations as well as the methodology is presented in detail in the attached abstract that was submitted to the International Electron Device Meeting to be held in December 2000.

I am currently working on tackling the full problem through a collaborative effort this summer with the Computational Materials Group at Motorola (Los Alamos National Labs). Work at Motorola is focused on solving the NEGF problem invoking the parabolic bandstructure approximation and in the absence of scattering in a full real space basis instead of a reduced space basis presented in the abstract. Also, a parallel effort is being pursued to include the effect of scattering into the NEGF calculations in a full real space basis. This computational task employs a 150 node (1 node = 2processors) Linux Cluster. Note that tackling the problem is real space as opposed to orbital space reduces the matrix sizes quoted in Section 4 by a factor of 10. Despite this gain, addition of scattering complicates the problem so greatly that preliminary calculations require about 50 processors with a memory requirement of 500 MB per processor. The computational time despite the use of a Linux Cluster is about 8 hours per bias point. This simulator (QUARTZ) is the first of its kind in trying to solve the NEGF problem with scattering effects self consistently treated. By the end of the summer, I will have completed a full 2D NEGF solver for electrons and holes in the absence of scattering, and this will be the starting point for my work when I return to campus in the Fall.

6. What I hope to learn from the course

The problem that I am addressing is computationally challenging. It involves solving a set of coupled partial differential equations whose coefficients themselves need to be evaluated self consistently ($\sum_S$, U_C). The complexity is further aggravated by the need to use a fine grid in energy and to store and manipulate large matrices of complex numbers. The important tools that are needed to attack this problem include fast eigenvalue and eigenvector solvers as well as optimized inversion routines for complex, asymmetric sparse matrices. Also, sophisticated libraries to efficiently parallelize matrix manipulations and efficiently share machine memory is something that would be invaluable. Scientific visualization tools to generate 2D and 3D plots would help with debugging the large code that needs to be developed.

7. References

[1] Executive Summary, "Research Issues and Directions for Nanotransistors: Technology, Physics, and Simulation," National Institute of Standards and Technology, Gaithersburg, MD, Feb. 8-9, 1999.

[2] K. Banoo and M.S. Lundstrom, "Electron Transport in a Model Si Transistor," submitted for publication, July, 1999.

[3] Z. Ren and M. Lundstrom, "Electron Transport in Nanoscale MOSFETs: A Scattering Theory Interpretation," Presented at the NASA Ames Device Modeling Workshop, Moffett Field, CA, August 26-27, 1999.

[4] S. Datta, Electronic Transport in Mesoscopic Systems, Cambridge University Press, 1995.

[5] R. K. Lake and S. Datta, "The Non-equilibrium Green's Function Method Applied to Double Barrier Resonant Tunneling Diodes," Phys. Rev. B, 45, 6670-6685 (1992).

[6] G. Klimeck, R. K. Lake, S. Datta and G. Bryant, "Elastic and inelastic scattering in quantum dots in the Coulomb-blockade regime", Phys. Rev. B, 50, 5484-5496 (1994).

[7] W. Tian, S. Datta, S. Hong, R. Reifenberger, J. Henderson and C.P. Kubiak, "Conductance Spectra of Molecular Wires," J. Chem. Phys. , 109, 2874-2882 (1998).

The Ballistic Nanotransistor: A Simulation Study

Z. Ren, R. Venugopal, S. Datta, and M. Lundstrom
Email: zhibin@ecn.purdue.edu, Phone: 765-494-1185, Fax: 765-494-6441
Purdue University, West Lafayette, IN USA 47907-1285
D. Jovanovic, Motorola/LANL, Los Alamos, NM 87545
J.G. Fossum, University of Florida, Gainesville, FL 32611-6130

Future MOS transistors may operate near their ballistic limits [1], so it is important to understand ballistic device physics and the prospects for achieving quasi-ballistic operation. In this paper, we explore the device design and physics issues of MOSFETs at the scaling limits using semiclassical and full quantum simulations. The device we presume is a double-gate (DG) MOSFET with symmetrical (n^+/n^+), asymmetrical (n^+/p^+) or symmetric midgap gates (depending on the body thickness) to achieve suitable V_T [2]. Simulations treat either classical or quantum transport, either ballistic or with scattering, and use methods that range from Matlab scripts which run in a few seconds to large-scale codes that require a parallel supercomputer. This work establishes: 1) the physics of charge control in 10 nm scale MOSFETs, 2) the maximum source velocity achievable under ballistic conditions, 3) the influence of multi-subband transport, 4) the separate roles of quantum effects in the normal and longitudinal directions, and 5) the possibility of quasi-ballistic operation given the expected values of μ_{eff}. We then use these quantum-scale simulations to critically examine scaling limits for thin and ultra-thin body devices considering issues such as parasitic source/drain resistance, gate overlap with the source/drain, DIBL, V_T roll-off, and subthreshold slope. The paper provides a clear description of the physics of MOSFETs at the scaling limits, of the device design issues that arise for such devices, as well as a description of a suite of simulation tools that can be used to explore nanotransistors more generally.

Two classes of simulation tools have been developed for this work. The first (2D/1D) couples a 2D Poisson-solver to a 1D transport model to accurately treat MOS electrostatics and efficiently treat transport. The 1D transport model is a good approximation for DG MOSFETs (exact in the 1-subband limit). We have also implemented both semiclassical and quantum ballistic transport models, as well as conventional drift-diffusion and energy-transport models. A simple treatment of scattering has also been implemented. This work extends recent pioneering work [3] by: 1) treating quantum transport along the channel with a Green's function approach that eliminates the WKB approximation and provides a way to include scattering [4], 2) treating quantum confinement normal to the channel with self-consistent, Schrödinger-Poisson simulations, 3) including multiple subbands, 4) comparing quantum and classical transport models, and 5) examining the effects of scattering and the possibilities of quasi-ballistic transport. Because the simulations are extremely fast (complete *I-V* characteristics are produced by Matlab scripts in a few minutes), they are useful for exploring device physics and design options. The second class of simulation tools provide a full 2D solution to the quantum-transport problem using the nonequilibrium Green's function method [4, 5]. These simulations carry with them a high computational burden, but they provide a benchmark to assess the simpler 1D codes, and they include oxide tunneling effects, which were ignored, for present, in the simpler models. This suite of simulation tools provides us with an ability to realistically assess MOS devices at their scaling limits.

Several researchers have explored the role of SOI body thickness on 2D electrostatics (e.g. [6]), but in this work, we also examine the role of multi-subband conduction in these devices and its relation to body thickness. For a fully depleted device with a thick body, several subbands are occupied; the higher in-plane effective mass leads to a lower source injection velocity [7] which degrades the performance of ballistic devices. For an ultra-thin body, one subband conduction may be achieved, which significantly increases the source injection velocity. We examine these effects by 2D/1D simulations and confirm that subband occupancy effects are an important design consideration. We also confirm that for relatively thick bodies, the asymmetric gate design provides an acceptable V_T and essentially the same on-current as a hypothetical mid-gap gate [2]. For ultra-thin bodies, however, mid-gap gate is needed for an acceptable V_T. We also confirm that the source velocity that would be deduced from the terminal characteristics of such a device is, indeed, the thermal injection velocity [7]. Finally, we examine quantum effects both normal to the channel and along the channel. Quantum-confinement effects (normal to the channel) have a strong effect on V_T and the effective gate capacitance. Quantum-transport along the channel has a minor effect on above-threshold current of a ballistic nanotransistor (~ 5%), but the subthreshold swing is degraded by tunneling through the source/channel barrier.

Real devices will operate below the ballistic limit because of carrier scattering. Present-day devices operate at about 50% of the ballistic limit [8], so it is of interest to examine whether scattering can be neglected in 10 nm-scale devices. We examined this issue by extending the 2D/1D ballistic model to include scattering within a scattering theory framework [7]. The results show that the inversion-layer mobility as measured on long-channel devices, continues to be a physically relevant parameter for nanoscale MOSFETs. The results also show that for a 10nm DG MOSFET, a mobility of at least 200 cm^2/V-s is needed to achieve the on-current target [9]; such a device would operate at 60% of its ballistic limit. Finally, we examine scattering in the NEGF approach to confirm that the essential physics as gleaned from semiclassical simulations retains validity in quantum scale devices.

The paper concludes with an examination of some practical issues that may limit the performance of DG nanoscale SOI MOSFETs. The parasitic series resistances expected for such devices are established as is their role on device performance. The tip resistance, due to the finite grading of the S/D junctions, is also examined as is the issue of the required S/D overlap to achieve full on-current [10]. In summary, this paper identifies the important device physics issues for nanoscale MOSFETs and addresses them by simulations at the quantum-scale.

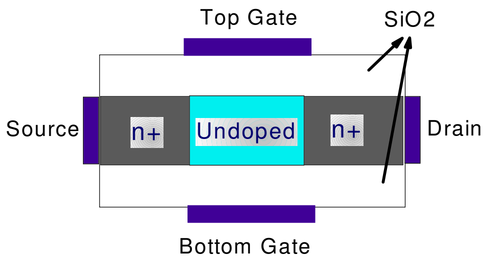
The Ballistic Nanotransistor: A Simulation Study

Ren, Venugopal, Datta, Lundstrom, Jovanovic, and Fossum

REFERENCES

[1] G. Timp, *et al.*, *Tech. Dig.*, *Int. Electron Dev. Mtg*., 55, 1999.
[2] J.G. Fossum, *et al*., presented at the Silicon Nanoelectronics Workshop, Honolulu, HI, June 10-11, 2000.
[3] Y. Naveh and K.K. Likharev, *IEEE Electron Dev. Lett*., **21**, 242, 2000.
[4] R. Lake, *et al.*, *J. Appl. Phys.*, **81**, 7845, 1996.
[5] D. Javanovic and R. Venugopal, presented at the 7th Int. Conf. On Compuational Electronics, Glasgow, Scotland, May 22 - 25, 2000.
[6] H.S.P. Wong, *et al., Tech. Dig., Int. Electron Dev. Mtg.*, 407, Dec., 1998.
[7] M.S. Lundstrom, *IEEE Electron Dev. Lett.,* **18**, 361, 1997.
[8] F. Assad, *et al.*, *Tech. Dig., Int. Electron Dev. Mtg*., 547, 1999.
[9] F. Gamiz, *et al.*, *IEEE Trans. Electron Dev*., **45**, 1122, 1998.
[10] S. Thompson, *et al.*, Intel Corp.,www.intel.se/technology/itj/q31998/articles/art_3.htm.

Fig. 1 The DG MOSFET examined in this work. Three different gate structures were examined: i) symmetric (top/bottom n^+/n^+), asymmetric (top/bottom n^+/p^+), and symmetric with a midgap gate.



An oxide thickness of 1.5 nm was used and the well-known effects of oxide tunneling were not addressed. Two different body thicknesses were examined in detail, t_{Si} = 5 nm and 1.5 nm.

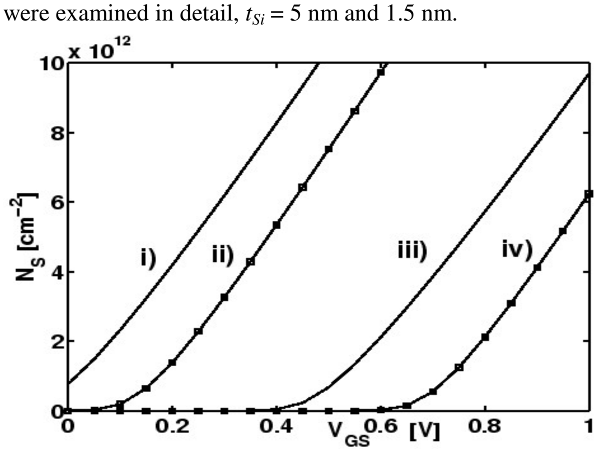


Fig. 2 Charge control in the DG MOSFET. We plot Q_{inv} vs. V_{GS} for four cases: i) n^+/n^+ gate 5 nm body, ii) n^+/n^+ gate 1.5 nm body, iii) n^+/p^+ gate 5 nm body, iv) n^+/p^+ gate 1.5 nm body. The asymmetrical gate provides an adequate V_T for the 5nm device, but too high for the 1.5 nm device. For the 1.5nm body, a midgap gate material is required.

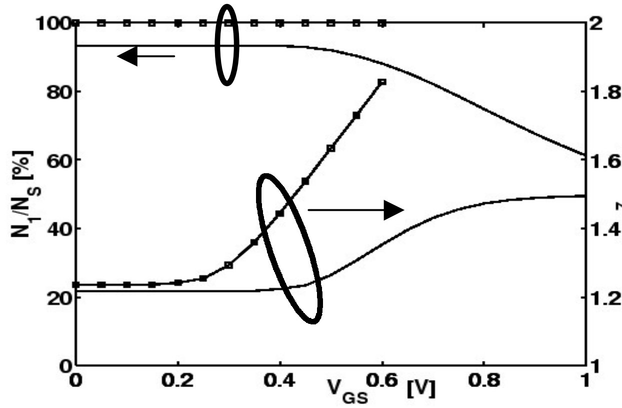


Fig. 3 Subband occupations and thermal velocity vs. gate voltage. Two cases are shown: i) n^+/p^+ gate with 5 nm body (lines) and ii) mid-gap gate with 1.5 nm body (symbols). On the left, we plot N_1/N_S and on the right, the corresponding thermal injection velocity, υ_T. For the 5 nm body, the source injection velocity is degraded at high V_{GS} because of upper subband occupation.

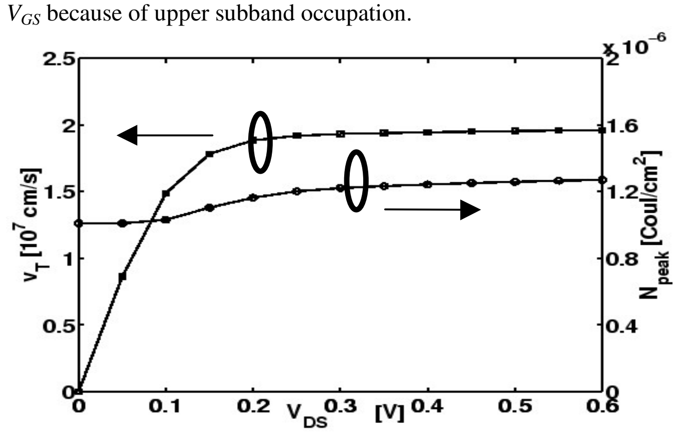


Fig. 4 Velocity and inversion layer charge at the beginning of the channel (top of the barrier) vs. V_{DS} for the midgap gate with 1.5 nm body (V_{GS} = 0.6V). These results confirm that the maximum source velocity is the thermal injection velocity shown in Fig. 3 and that the inversion layer charge is fixed by MOS electrostatics. (The slight rise with V_{DS} is the result of DIBL).

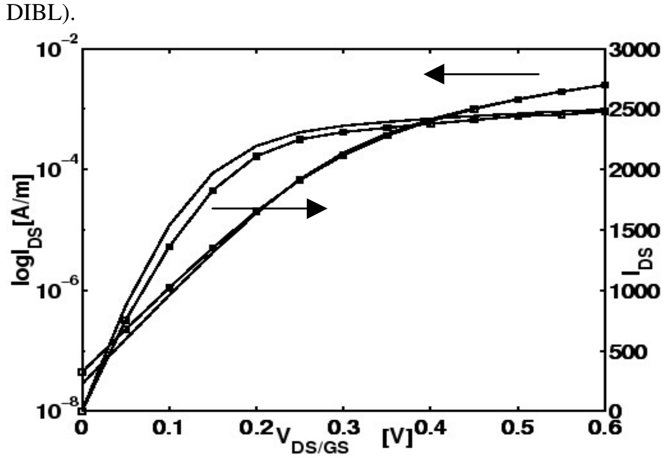


Fig. 5 Comparison of the classical and quantum simulations for the L = 10 nm, 1.5 nm body midgap gate device. On the left, we plot log I_{DS} vs. V_{GS} (for V_{DS} = 0.6V) and on the right, I_{DS} vs. V_{DS} (for V_{GS} = 0.6V). The quantum simulations (symbols) give a subthreshold swing that is slightly higher (70 mV/dec vs. 66 mV/dec) than the semiclassical simulation but it gives about the same (5% less) on-current.

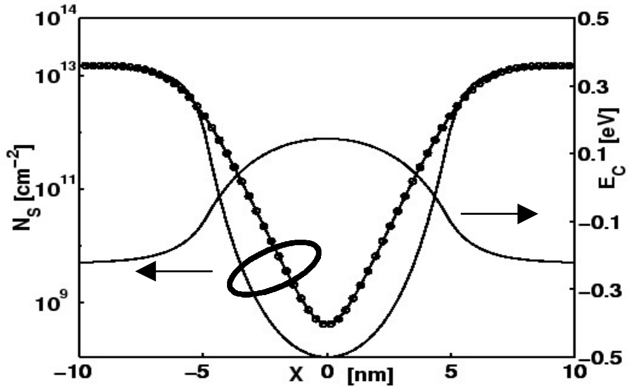


Fig. 6 Comparison of the classical and quantum carrier densities vs. position for the L = 10 nm, 1.5 nm body midgap gate device at V_{DS} = V_{GS} = 0V. Although the device displays low off-current, the effects of quantum (symbols) mechanical tunneling into the channel are clearly seen.

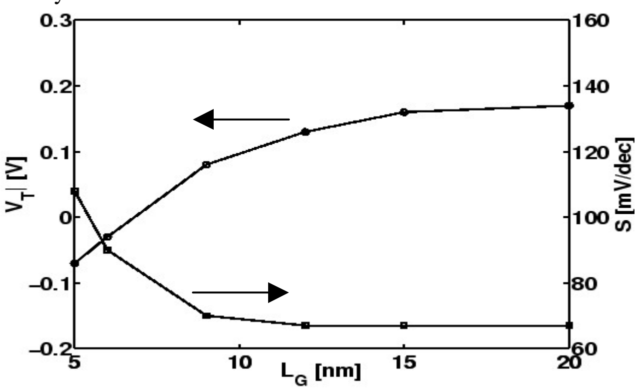


Fig. 7 Short channel characteristics vs. channel length for the 1.5 nm body device as computed by classical ballistic simulations. On the left, we plot V_T vs. L and on the right the subthreshold swing, S. The results show that below L=10nm, the thin-body performance begins to degrade.

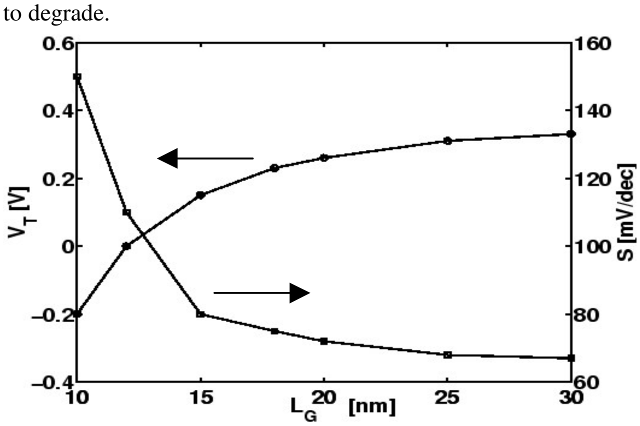


Fig. 8 Short channel characteristics vs. channel length for the 5 nm body device as computed by classical ballistic simulations. On the left, we plot V_T vs. L and on the right the subthreshold swing, S. The results show that below L=15nm, the thick-body performance begins to degrade.

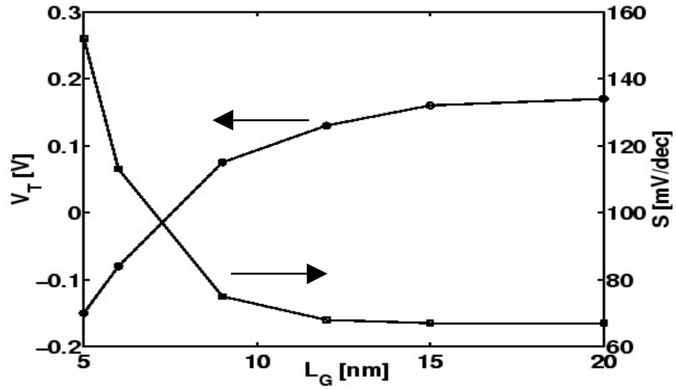


Fig. 9 A repeat of Fig. 7 but with quantum ballistic simulations. Tunneling through the source/channel barrier severely degrades the short channel performance below L = 10 nm.

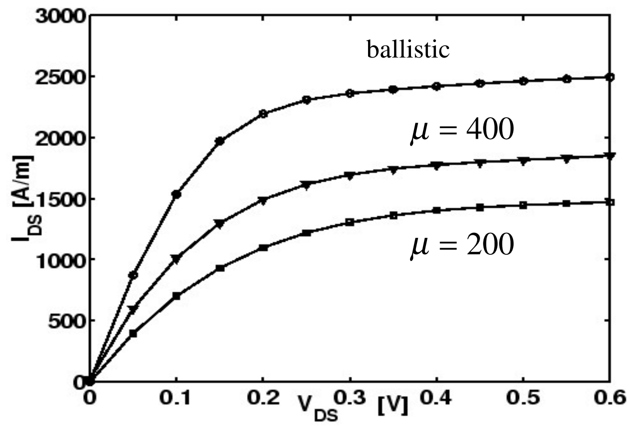


Fig. 10 I_{DS} vs. V_{DS} comparison of the (1D/2D) simulations with and without scattering for the 1.5 nm body device with midgap gates. We show three curves: i) the ballistic curve, ii) one with μ_{eff} = 400, and iii) one with μ_{eff} = 200. The results show that the on-current even in the quasi-ballistic regime, is controlled by the carrier mobility and that even with a mobility of 200, an L=10 nm device will still operate at 60% of its ballistic limit.

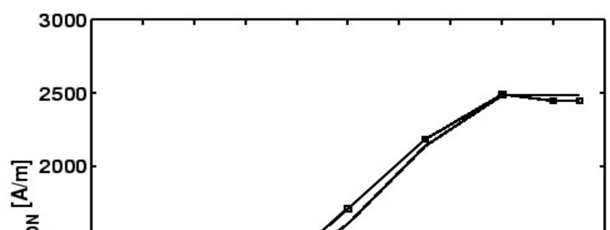


Fig. 11 On-current vs. S/D overlap with the gate. Both classical and quantum (symbols) results are shown. These simulations establish the minimum overlap needed and show essentially the same results for quantum and classical simulations. For these ideal, abrupt junctions, no overlap is necessary, but for more realistic junctions, some overlap is necessary [10].