RC 22539 (W0208-025) August 9, 2002
Electrical Engineering

IBM Research Report

Ballistic FET Modeling using QDAME: Quantum Device Analysis by Modal Evaluation

Steven E. Laux, Arvind Kumar, Massimo V. Fischetti
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Ballistic FET Modeling using QDAME: *Q*uantum *D*evice *A*nalysis by *M*odal *E*valuation

Steven E. Laux, Arvind Kumar, and Massimo V. Fischetti

Abstract

We present an algorithm for self-consistent solution of the Poisson and Schrödinger equations in two spatial dimensions with open boundary conditions to permit current flow. The algorithm works by discretely sampling a device's density of states using standing wave boundary conditions, decomposing the standing waves into traveling waves injected from the contacts to assign occupancies, and iterating the quantum charge with the potential to self-consistency using a novel hybrid Newton-Broyden method. A double-gate FET is simulated as an example, with applications focused on surface roughness and contact geometry.

I. Introduction

The quantum ballistic limit, in which electrons propagate as waves without scattering, is an important one for understanding device behavior as feature sizes are scaled to lengths comparable to the electron wavelength. Although simple analytical transport models can provide useful insight in this regime [1], [2], a complete understanding for realistic device geometries and operating conditions requires a self-consistent solution of the Poisson and Schrödinger equations with open boundary conditions to permit the flow of current. Self-consistent solution of the Poisson and open-boundary condition Schrödinger equations is complicated due to issues such as discretization of the energy continuum, wavefunction normalization, the presence of quasi-bound states, population of the electron states by the reservoirs, and finally numerical iteration to convergence [3], [4].

In this paper we introduce QDAME, an efficient and physically intuitive algorithm for solving the Poisson and open-boundary condition Schrödinger equations self-consistently for devices in two spatial dimensions with ellipsoidal parabolic energy bands, employing a formulation generalized from the one-dimensional framework in Refs. [5], [6]. In contrast to earlier approaches based on the nonequilibrium Green's function [3], [4], QDAME is based on the Quantum Transmitting Boundary Method (QTBM) [7] which, through use of a finite-element mesh, is capable of modeling devices of arbitrary geometry. Although in its present embodiment QDAME finds purely ballistic solutions, we envision that it may eventually be used as part of a Master equation approach that includes scattering, as in Refs. [5], [6]. After a discussion of the basic theory underlying QDAME, we present simulation results from a double-gate FET as an example.

II. Theory

The algorithm begins with an initial guess for the potential, which can be obtained either from a classical solution or a solution at a previous bias point. The solution for the electron charge in the potential is found using a modified version of the QTBM[7], in which a two-dimensional device geometry is connected to infinitely long leads which supply and receive current in discrete transverse modes (i.e., confined in the transverse direction and plane wave in the longitudinal direction) as illustrated in Fig. 1(a). Just as periodic boundary conditions are used to count states in a bulk solid, we impose standing wave boundary conditions at the device-lead interfaces, giving rise to an eigenvalue problem whose solutions are a discrete sampling of the continuum of energies allowed by the open boundary conditions. For each inequivalent alignment of the conduction band masses (e.g., 3 times for (100) Si), we generate two such discrete samplings – which we call the device "normal modes" – by imposing zero-value and zero-normal-derivative standing-wave boundary conditions on the wavefunction at the device-lead interfaces:

$$\text{sine} : \Psi_n^{(s)} = 0 \Longrightarrow \text{eigenvalues} \;\; E_n^{(s)}, \tag{1}$$

$$\text{cosine} : \widehat{\eta_i} \bullet \nabla \Psi_n^{(c)} = 0 \Longrightarrow \text{eigenvalues} \;\; E_n^{(c)}, \tag{2}$$

where $\widehat{\eta_i}$ denotes the outward normal at lead i. Although either Eq. (1) or Eq. (2) by itself is a complete discretized sampling of the device DOS, we must use both conditions (as a minimum) to be able to represent well both the electron distribution's value and its normal derivative at the device-lead interfaces.

Because the normal mode states cannot carry current, we must decompose each normal mode eigenfunction $\Psi_n^{(s,c)}(x,y)$ into states we call "traveling eigen-components" $\Phi_{n,i}^{(s,c)}$:

$$\Psi_n^{(s,c)}(x,y) = \sum_i \Phi_{n,i}^{(s,c)}(x,y), \tag{3}$$

The authors are with the IBM Semiconductor Research and Development Center (SRDC), T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598 USA (email: laux@us.ibm.com).

where the sum is taken over all injecting leads i. Analogous to a scattering state in quantum theory, a traveling eigen-component $\Phi_{n,i}^{(s,c)}$ is an injection "experiment" from a single lead i, containing a specific superposition of transverse modes of the lead i, at the eigen-energy $E_n^{(s,c)}$ of the sine (s) or cosine (c) normal mode. For each transverse mode p in lead i, the energy of the normal mode $E_n^{(s,c)}$ can be expressed as the sum of a transverse and a longitudinal component

$$E_n^{(s,c)} = E_p^i + \frac{\left(\hbar k_\eta^{p,i}\right)^2}{2m_\eta^i} \tag{4}$$

where E_p^i and m_η^i are respectively the transverse confinement energy of mode p and the longitudinal effective mass in lead i. Given the normal mode energy and the injection coefficients, the traveling eigen-components can be computed using a modified version of the QTBM, as explained later. The fact that either Eq. (1) or Eq. (2) gives rise to a complete discretized sampling of the device's DOS together with the assumption that the leads all inject incoherently establishes a normalization condition for the traveling eigen-components making up a normal mode:

$$\int\int_\Omega dx\, dy \left(\sum_i |\Phi_{n,i}^{(s,c)}(x,y)|^2\right) = \frac{1}{2} \tag{5}$$

where the integration extends over the device domain Ω and the factor 1/2 arises because we use two complete samplings.

The QTBM must be modified to use the supercell boundary conditions shown in Fig. 1(b), where the solution in one device is effectively "mirrored" across the device-lead interface. This modification is necessary for two reasons. First, coupling of the device to infinitely long leads results in solutions which are dominated by unphysical lead-device resonances; in contrast the sampling of the DOS is device-driven, not lead-driven, when supercell boundary conditions are used. Second, the use of supercell boundary conditions allows us to treat quasi-bound states in the device, corresponding to energies classically forbidden in the leads, analogously to propagating states: the traveling waves comprising the longitudinal part of the solution in the leads at classically allowed energies simply become growing and decaying exponentials at classically forbidden energies.

Figure 2 illustrates two example "cosine" normal modes and their traveling eigen-components for the SOI MOSFET shown schematically in the inset of Fig. 2(a). Well below the top of the energy barrier, the normal modes are localized on either the source or the drain side, and are predominantly composed of just one traveling eigen-component which is almost entirely reflected; fig. 2(a) illustrates such a case where the normal mode is localized on the drain side and the drain eigen-component consists of injection in primarily the lowest two transverse modes of the drain lead. In contrast, the normal mode illustrated in Fig. 2(b) at an energy well above the barrier peak involves a substantial mixing of the traveling eigen-components injected from both leads.

The total electron density is the sum of contributions from all ladders arising from the different possible alignments of the conduction band masses; for simplicity we concentrate here on the electron density from a single conduction band valley. The electron density $n(x,y)$ is the sum of the electron densities from the sine sampling $n^{(s)}(x,y)$ and cosine sampling $n^{(c)}(x,y)$, found by summing over all traveling eigen-component densities multiplied by a thermal occupancy factor at temperature T:

$$n^{(s,c)}(x,y) = \sum_{n,i} |\Phi_{n,i}^{(s,c)}(x,y)|^2 \sum_p c_{p,n,i}^{(s,c)} \rho_{p,i}(k_D^i, T, E_n^{(s,c)}) \tag{6}$$

Here the thermal occupancy factor is a sum of occupancy factors $\rho_{p,i}(k_D^i, T, E_n^{(s,c)})$ with coefficients $c_{p,n,i}^{(s,c)}$ representing the fractional weight of each transverse mode p in lead i making up the traveling eigen-component. We use a drifted Fermi-Dirac occupancy factor $\rho_{p,i}(k_D^i, T, E)$ for transverse mode p given by

$$\rho_{p,i}(k_D^i, T, E) = \left(\frac{8m_z^i k_B T}{h^2}\right)^{1/2} F_{1/2}\left(\frac{E_F^i - E - (\hbar^2/2m_\eta^i)k_D^i(k_D^i - 2k_\eta^{p,i})}{k_B T}\right) \tag{7}$$

where m_z^i is the out-of-plane effective mass in lead i, E_F^i is the Fermi level in lead i, $F_{1/2}$ is the Fermi-Dirac integral of order 1/2, $k_\eta^{p,i}$ is defined by Eq. (4), and the drift momentum $\hbar k_D^i$ at lead i is found from the requirement of current continuity between lead and device [5], [6]. Since the Poisson equation is solved with boundary conditions that enforce charge-neutrality at the leads, both important conditions of flux conservation and charge-neutrality near the contacts are met.

Once the quantum charge is computed for a given input potential, it must be iterated with the Poisson equation to obtain a self-consistent solution. A major stumbling block in this outer loop, which requires solution of the changes in potentials and drift momenta at the leads, is in the evaluation of the Jacobian matrix describing the nonlocal variation of the quantum charge with respect to potential. Figure 3 compares convergence of a typical problem using three different approaches: the Broyden method [8], [9], which constructs an approximate Jacobian based on the previous iterations, the Newton method with the derivative of the charge evaluated semiclassically, and a hybrid method combining these two techniques. While the Newton method with semiclassical derivative is seen to offer marked improvement over the Broyden method, the best results are obtained by a novel hybrid Newton-Broyden approach in which successive guesses to the solution are chosen based on a combination of the approximate semi-classical derivative evaluated at the present guess together with the guidance based on the solution history provided by the Broyden method.

III. Example: Double-Gate FET

Figure 4(a) shows a schematic of a double gate FET used to illustrate QDAME's capabilities. For simplicity, the gates are treated here as insulator contacts, although treatment as current-carrying leads to model gate leakage is straightforward. Figures 4(b-c) show the self-consistent conduction band edge and electron density. Although qualitatively similar to their classical counterparts, distinct quantum-mechanical features are clearly visible. Of particular note is the seamless approach in the contact regions to the source and drain leads, arising from the *a priori* setting of the lead potentials by charge neutrality. Figure 5 shows transfer characteristics for double gate structures of various effective channel lengths L_{eff} (separation between metallurgical junctions) and silicon thicknesses $t_{\rm Si}$. For well-chosen L_{eff} and $t_{\rm Si}$, a nearly ideal subthreshold swing is obtained, as anticipated, while short-channel effects become apparent for less optimal values. Figure 6 shows the evolution of the electron density from the subthreshold regime through the on-state (points (A)-(D) in Fig. 5): the electron density goes from being concentrated near the channel center in subthreshold to having peaks close to the oxide interfaces well above threshold [10].

In this high-field regime, we expect roughness along the gate oxide interface to play an important role. Although realistic roughness also varies in the out-of-plane direction, we can gain qualitative insight into how roughness affects the device characteristics by using a one-dimensional roughness potential; we choose one having an exponential autocorrelation function with parameters Δ and Λ characterizing the roughness height and correlation length, respectively [11]. In contrast to the smooth distribution in Fig. 6, Fig. 7 shows the electron density in a device with roughened Si-SiO_2 interfaces characterized by $\Delta = 0.24$ nm and $\Lambda = 0.72$ nm. A comparison of the current density distribution between a smooth versus rough case shows that the roughness not only redirects the current but also degrades it. Figure 8(a) compares transfer characteristics for two devices with roughened Si-SiO_2 interfaces having different roughness parameters with those of a smooth device. As expected, the current degradation increases with increasing gate voltage and is more pronounced in the device with greater roughness. Note that we expect some variation in transfer characteristics over many devices characterized by the same roughness parameters but having different realizations of the roughness; in Fig. 8(b) we show an average over five different realizations having the same roughness parameters so that the error bars give an idea of the variation expected over a large ensemble of devices.

Since QDAME is not restricted to rectilinear meshes, we can look at how the contact geometry affects the approach of the current into the device. Figure 9 compares the current density distribution in a thin (3 nm) double-gate FET attached to contact regions that are as wide as the Si channel (straight case), that taper gently away from the channel (tapered case), and that have an abrupt transition (dog bone case); the source (drain) lead is the full width of the left (right) extreme of the device (5 nm wide, including the oxides, for the straight case; 9 nm wide for the tapered and dog bone cases). Not surprisingly, the current has a similar distribution in the contact region as in the channel for the straight case. In the tapered case, the current enters and exits the channel following paths that hug the angled edges of the contact regions. In sharp contrast, the abrupt edges in the dog bone case result in strong back-reflections which cause the transmitted current to be largely concentrated at the center of the contact regions.

IV. Conclusion and Future Work

In conclusion, we have described an algorithm for self-consistent solution of the Poisson and open-boundary condition Schrödinger equations in two spatial dimensions and demonstrated its utility through relevant examples. Problems left for future work include improving numerical convergence through a more accurate Jacobian, incorporation of non-parabolic band structure, the inclusion of scattering through a Master equation approach, and eventually the extension to three spatial dimensions.

References

[1] K. Natori, "Ballistic metal-oxide-semiconductor field effect transistor," *J. Appl. Phys.*, vol. 76, pp. 4879-4889, Oct. 1994.

[2] F. Assad, Z. Ren, D. Vasileska, S. Datta, and M. Lundstrom, "On the performance limits for Si MOSFETs: a theoretical study," *IEEE Trans. Electron Devices*, vol. 47, pp. 232-240, Jan. 2000.

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

[4] A. Svizhenko, M.P. Anantram, T.R. Govindam, B. Biegel, and R. Venugopal, "Two-dimensional quantum mechanical modeling of nanotransistors," *J. Appl. Phys.*, vol. 91, pp. 2343-2354, Feb. 2002.

[5] M.V. Fischetti, "Theory of electron transport in small semiconductor devices using the Pauli master equation," *J. Appl. Phys.*, vol. 83, pp. 270-291, Jan. 1998.

[6] M.V. Fischetti, "Master-equation approach to the study of electron transport in small semiconductor devices," *Phys. Rev. B*, vol. 59, pp. 4901-4917, Feb. 1999.

[7] C.S. Lent and D.J. Kirkner, "The quantum transmitting boundary method," *J. Appl. Phys.*, vol. 67, pp. 6353-6359, May 1990.

[8] D.D. Johnson, "Modified Broyden's method for accelerating convergence in self-consistent calculations," *Phys. Rev. B*, vol. 38, pp. 12807-12813, Dec. 1988.

[9] V. Eyert,"A comparative study on methods for convergence acceleration of iterative vector sequences," *J. Comp. Phys.*, vol. 124, pp. 271-285, 1996.

[10] L. Ge and J.G. Fossum,"Analytical modeling of quantization and volume inversion in thin Si-film DG MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, pp. 287-294, Feb. 2002.

[11] S.M. Goodnick, D.K. Ferry, C.W. Wilmsen, Z. Liliental, D. Fathy, and O.L. Krivanek, "Surface roughness at the Si(100)-SiO_2 interface," *Phys. Rev. B*, vol. 32, pp. 8171-8186, Dec. 1985.

Figure Captions

Figure 1: Schematic of device geometry solvable using QTBM (a) with infinitely long leads; (b) with supercell boundary conditions imposed at device-lead interfaces.

Figure 2: Normal modes (top) decomposed into traveling eigen-components from the source (middle) and drain (bottom) leads for the SOI MOSFET shown in the inset at an energy (a) well below the top of the barrier; (b) well above the top of the barrier.

Figure 3: Comparison of convergence behavior using three different numerical techniques: Broyden method, Newton method with charge derivative evaluated semiclassically, and hybrid Newton-Broyden method.

Figure 4: Double-gate nFET used in simulations: (a) schematic; (b) conduction band edge; and (c) electron density.

Figure 5: Transfer characteristics for double-gate nFETs of varying effective channel length L_{eff} and Si thickness $t_{\rm Si}$. The oxide thickness is $t_{ox} = 1$ nm and the drain-source bias is 0.1 V. Note A-D refer to Fig. 6.

Figure 6: Four contours of electron density (2 contours/decade starting from $1.26 \times 10^{18}\,{\rm cm}^{-3}$) for the bias points shown in Fig. 5 ($L_{eff} = 25$ nm, $t_{\rm Si} = 10$ nm, t_{ox} =1 nm).

Figure 7: Contours of electron density in a double-gate nFET with interfaces roughened as shown using parameters $\Delta = 0.24$ nm, $\Lambda = 0.72$ nm. Here $L_{eff} = 25$ nm, $t_{\rm Si} = 10$ nm, t_{ox} =1.5 nm, the gate-source bias is 0.4 V, and the drain-source bias is 0.1 V.

Figure 8: (a) Transfer characteristics for smooth device and devices roughened with two different sets of roughness parameters. Here $L_{eff} = 25$ nm, $t_{\rm Si} = 10$ nm, t_{ox} =1.5 nm, and the drain-source bias is 0.1 V. (b) Transfer characteristics for smooth device and average of five devices with interfaces characterized by the same roughness parameters but having different realizations of the roughness. Error bars give an idea of variation expected over a large ensemble of devices.

Figure 9: Device layouts (above) and contours of current density magnitude (below) for double-gate nFETs having straight, tapered, and dog bone contact geometries. In the layouts, regions of piecewise constant doping are shaded and oxide regions are in black. In the contours, dark (light) represents areas of low (high) current density.

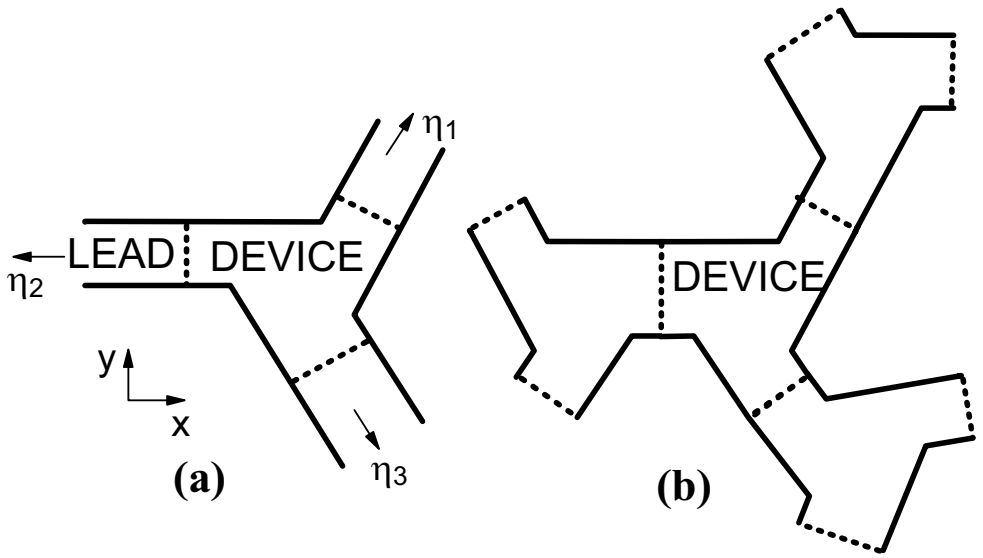

Figure 1

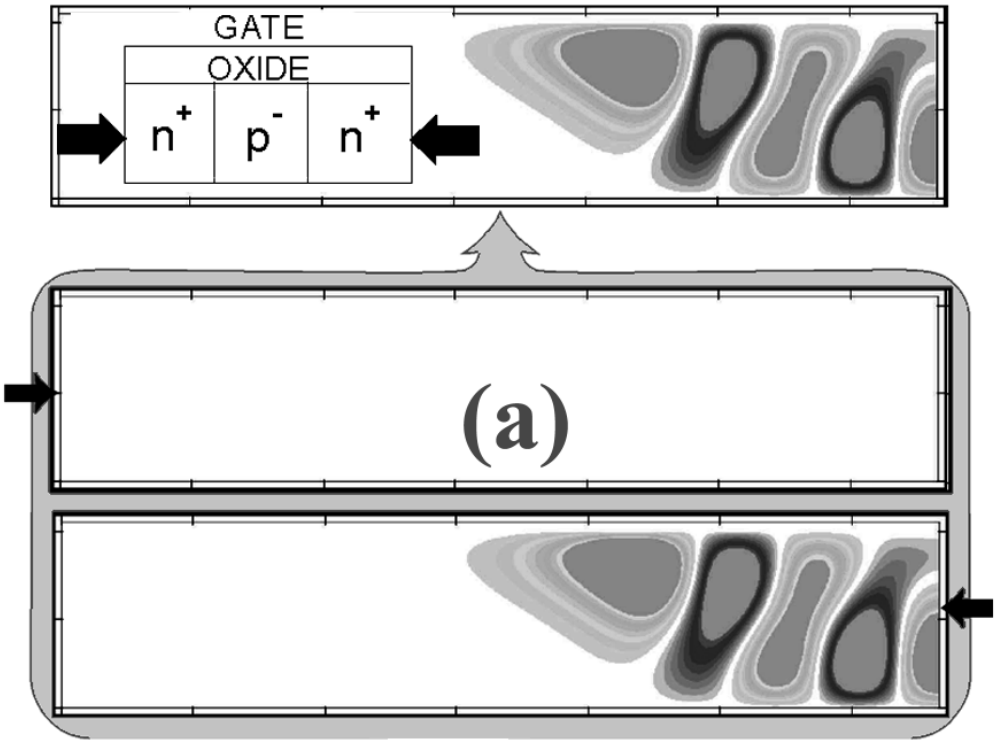

Figure 2a

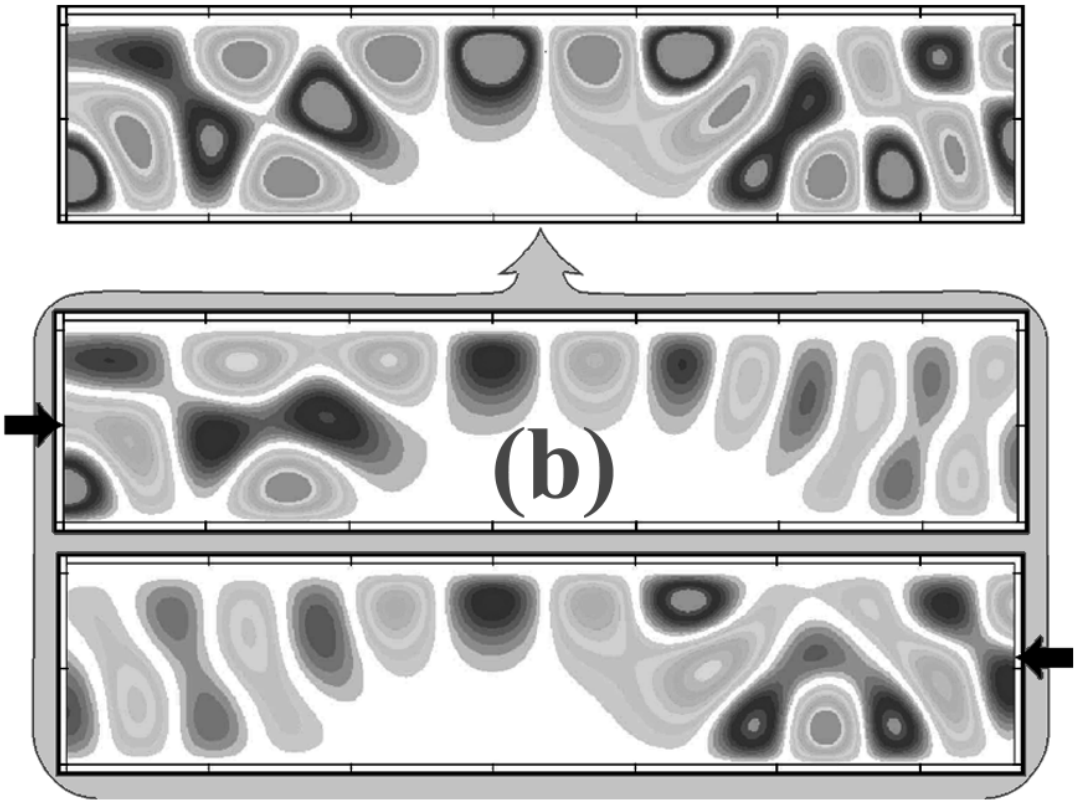

Figure 2b

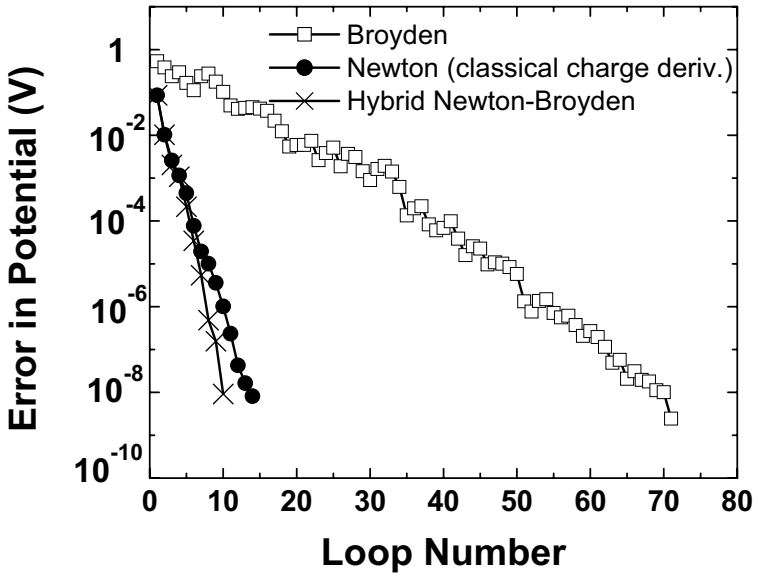
Broyden
Newton (classical charge deriv.)
Hybrid Newton-Broyden
Error in Potential (V)
1
10-2
10-4
10-6
10-8
10-10
0
10
20
30
40
50
60
70
80
Loop Number

Figure 3

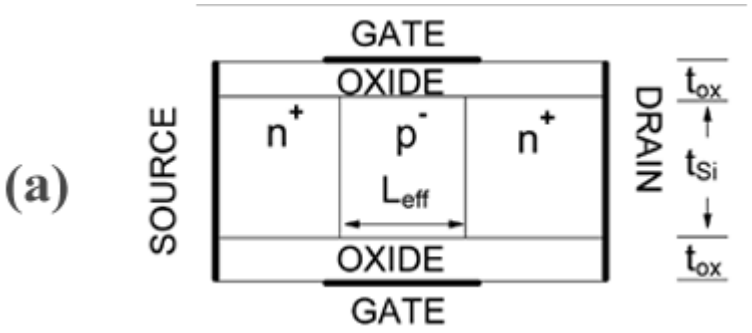

Figure 4a

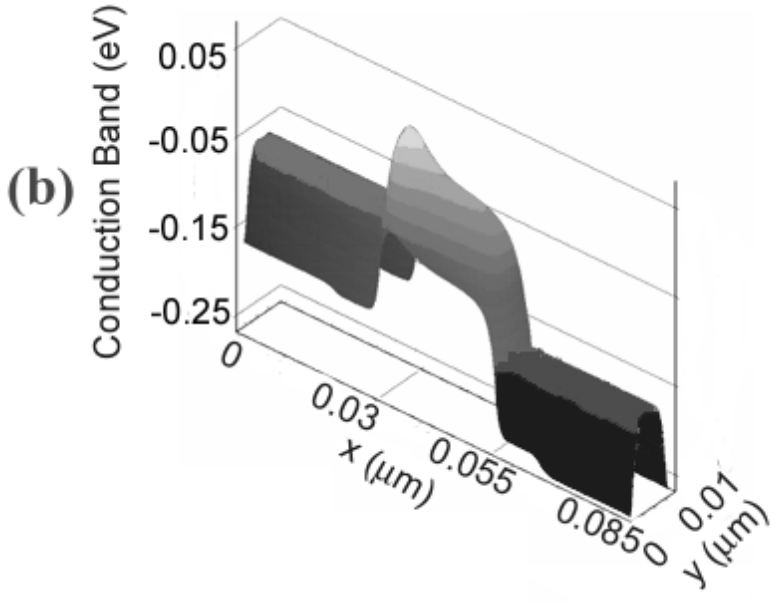

Figure 4b

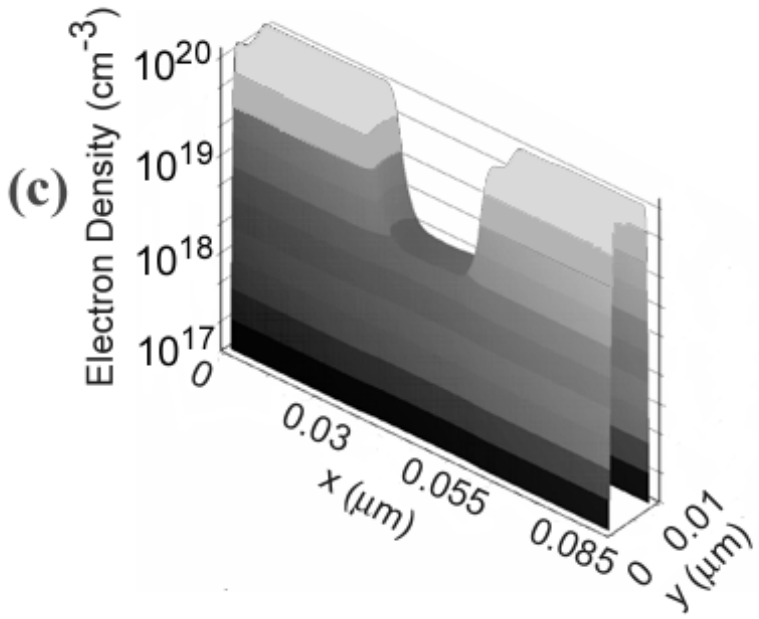
(c)
Electron Density (cm-3)
10^20
10^19
10^18
10^17
0
0.03
0.055
0.085
x (μm)
0
0.01
y (μm)

Figure 4c

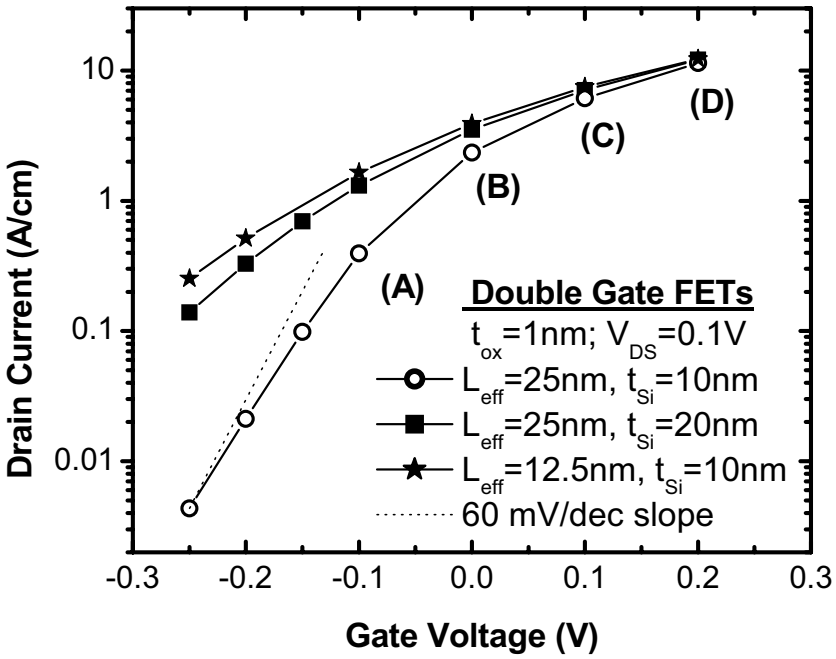
Drain Current (A/cm)
10
1
0.1
0.01
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
Gate Voltage (V)
(A)
(B)
(C)
(D)
Double Gate FETs
t_ox=1nm; V_DS=0.1V
L_eff=25nm, t_Si=10nm
L_eff=25nm, t_Si=20nm
L_eff=12.5nm, t_Si=10nm
60 mV/dec slope

Figure 5

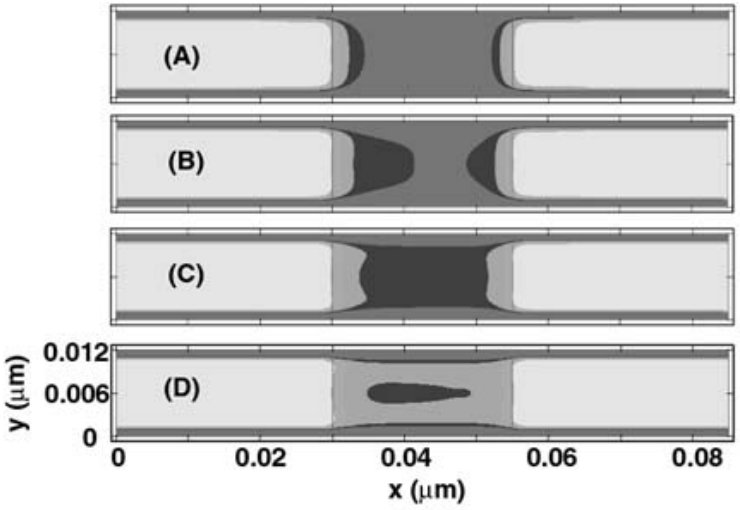

Figure 6

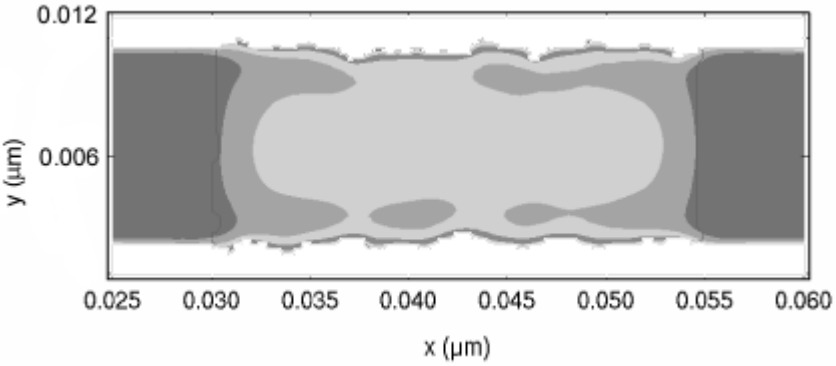

Figure 7

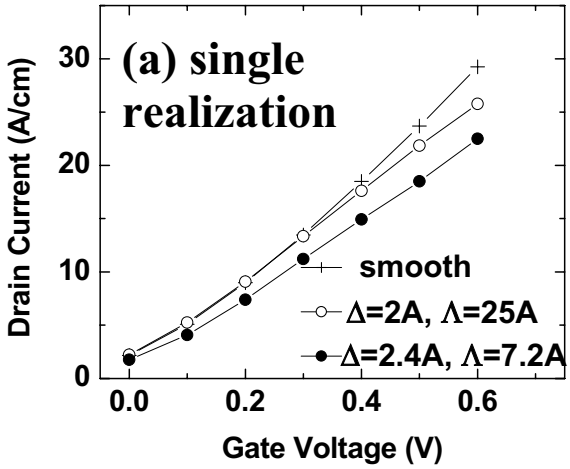
(a) single realization
Drain Current (A/cm)
Gate Voltage (V)
0
10
20
30
0.0
0.2
0.4
0.6
smooth
Δ=2A, Λ=25A
Δ=2.4A, Λ=7.2A

Figure 8a

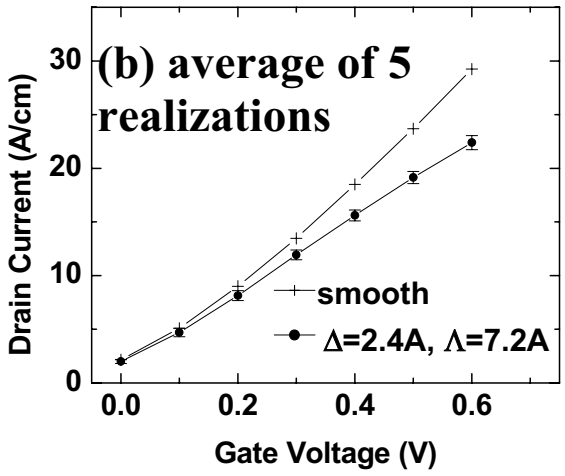
(b) average of 5 realizations
Drain Current (A/cm)
Gate Voltage (V)
smooth
Δ=2.4A, Λ=7.2A
0
10
20
30
0.0
0.2
0.4
0.6

Figure 8b

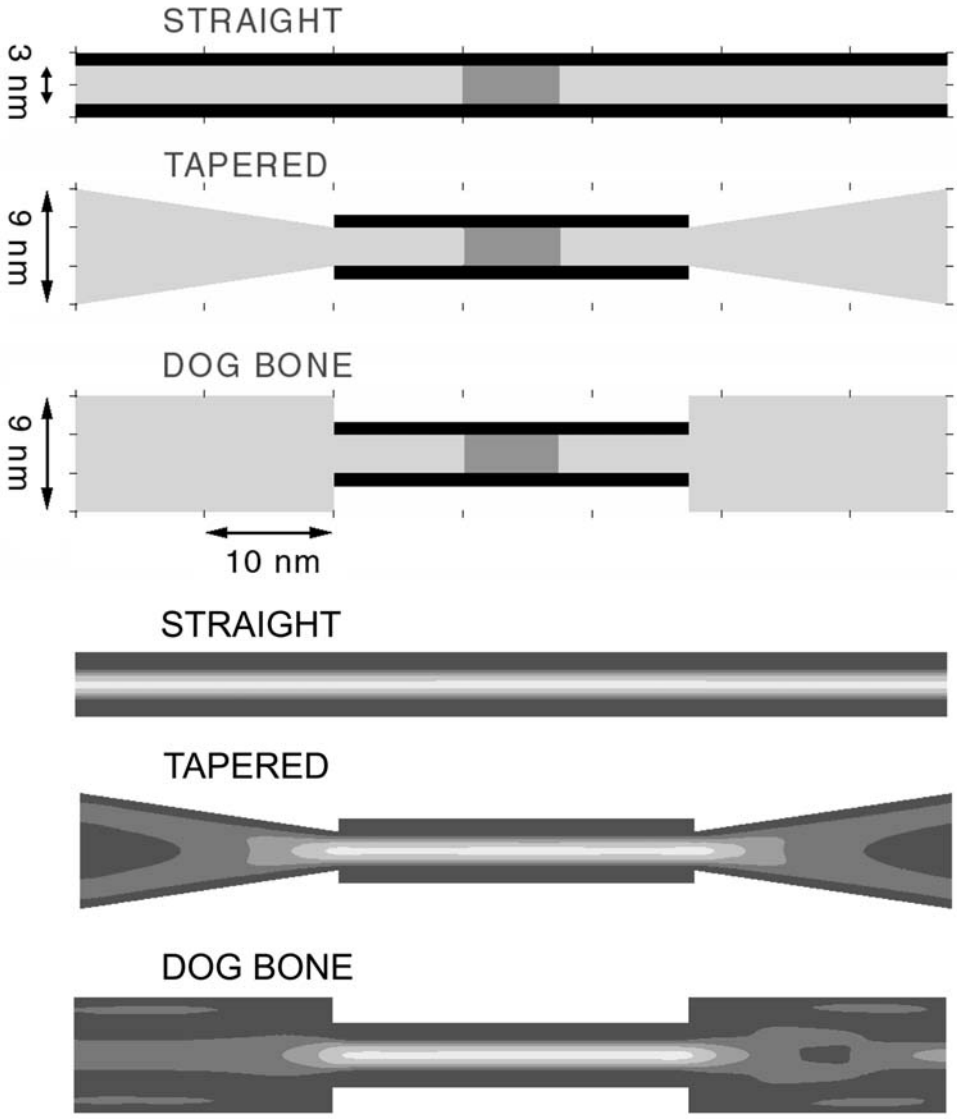
STRAIGHT
3 nm
TAPERED
9 nm
DOG BONE
9 nm
10 nm
STRAIGHT
TAPERED
DOG BONE

Figure 9