Two-dimensional quantum mechanical modeling of nanotransistors

A. Svizhenko, M. P. Anantram,[a] T. R. Govindan, and B. Biegel
NASA Ames Research Center, Mail Stop T27A-1, Moffett Field, California 94035-1000

R. Venugopal
School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907-1285

(Received 16 July 2001; accepted for publication 9 November 2001)

Quantization in the inversion layer and phase coherent transport are anticipated to have significant impact on device performance in ''ballistic'' nanoscale transistors. While the role of some quantum effects have been analyzed qualitatively using simple one-dimensional ballistic models, two-dimensional (2D) quantum mechanical simulation is important for quantitative results. In this paper, we present a framework for 2D quantum mechanical simulation of a nanotransistor/metal oxide field effect transistor. This framework consists of the nonequilibrium Green's function equations solved self-consistently with Poisson's equation. Solution of this set of equations is computationally intensive. An efficient algorithm to calculate the quantum mechanical 2D electron density has been developed. The method presented is comprehensive in that treatment includes the three open boundary conditions, where the narrow channel region opens into physically broad source, drain and gate regions. Results are presented for (i) drain current vs drain and gate voltages, (ii) comparison to results from Medici, and (iii) gate tunneling current, using 2D potential profiles. Methods to reduce the gate leakage current are also discussed based on simulation results. © *2002 American Institute of Physics.* [DOI: 10.1063/1.1432117]

I. INTRODUCTION

Metal oxide field effect transistors (MOSFETs) with channel lengths in the tens of nanometer regime have recently been demonstrated by various research labs.[1–3] Design considerations to yield devices with desirable characteristics have been explored in Refs. 4–8. Device physics of these MOSFETs were analyzed using simple quasi-one-dimensional models.[9–13] The best modeling approach for design and analysis of nanoscale MOSFETs is presently unclear, though a straightforward application of semiclassical methods that disregards quantum mechanical effects is generally accepted to be inadequate. Quantum mechanical modeling of MOSFETs with channel lengths in the tens of nanometers is important for many reasons:

(i) MOSFETs with ultrathin oxide require an accurate treatment of current injection from source, drain and gate. Gate leakage is important because it places a lower limit on the OFF current.

(ii) Ballistic flow of electrons across the channel becomes increasingly important as the channel length decreases.

(iii) The location of the inversion layer changes from the source to the drain end, and its role in determining the $C-V$ and $I-V$ characteristics is most accurately included by a self-consistent solution of Poisson's equation and a quantum mechanical description to compute the charge density.

(iv) Approximate theories of quantum effects included in semiclassical MOSFET modeling tools are desirable from practical considerations because semiclassical methods are numerically less expensive, and much of the empirical and semiclassical MOSFET physics developed over the last few decades continues to hold true in many regions of a nanoscale MOSFET. Examples of semiclassical methods that consider some quantum mechanical aspects are the density gradient,[14,15] and effective potential[16] approaches, and quantum mechanical approximations used in the Medici package.[17] Fully quantum mechanical simulations can play an important role in benchmarking such simulators.

Central to quantum mechanical approaches to charge transport modeling is the self-consistent solution of a wave equation to describe the quantum mechanical transport, Poisson's equation, and equations for statistics of the particle ensemble. In the absence of electron–electron and electron–phonon interactions (state of the scatterer does not change), the Landauer–Buttiker formalism[18,19] is applicable. In this formalism, the wave equation is Schrodinger's equation and the statistics is represented throughout the device by the Fermi–Dirac distribution of particles incident from the contacts (source, drain and gate). In the presence of electron–phonon interaction, the Wigner function (WF) and nonequilibrium Green's function (NEGF) formalisms are used. The NEGF approach has been quite successful in modeling steady state transport in a wide variety of one-dimensional (1D) semiconductor structures.[20,21]

A number of groups have started developing theory and simulation for fully quantum mechanical *two-dimensional* simulation of MOSFETs using the: real space approach,[22–24] *k*-space approach,[25] Wigner function approach,[26] and nonequilibrium Green's function approach.[13,27,28] Other groups have taken a hybrid approach using the Monte Carlo method. The Monte Carlo approach, has been quite successful in describing scattering mechanisms in MOSFETs, in comparison

[a]Electronic mail: anant@nas.nasa.gov

 © 2002 American Institute of Physics

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

to fully quantum mechanical approaches, and can also include ballistic effects and the role of quantized energy levels in the MOSFET inversion layer in an approximate manner.[29–31] Discussing the relative merits of various approaches and quantum-corrected drift-diffusion approaches is important. In fact, such a comparison of methods using standard device structures has been initiated[32] but much work remains to be done in comparing and studying the suitability of different methods. Comparison of various methods is *not* the purpose of this paper. The purpose of this paper is to describe development of a particular approach, namely the NEGF approach, for numerical simulation of MOSFETs with two-dimensional (2D) doping profiles and charge injection from the source, drain and gate contacts. 2D simulation significantly increases computational effort over the 1D case. Nonuniform spatial grids are essential to limit the total number of grid points while at the same time resolving physical features. A new algorithm for efficient computation of electron density without complete solution of the system of equations is presented. The computer code developed is used to calculate the drain and gate tunneling current in ultrashort channel MOSFETs. Results from our approach and Medici are compared. The paper is organized as follows: formalism (Sec. II), role of polysilicon gate depletion Sec. III A, slopes of I_d vs the gate (V_g) and drain (V_d) voltages (Secs. II A–III C), role of anisotropic effective mass (Sec. III D), role of gate tunneling current as a function of gate oxide thickness and gate length in determining the OFF current (Sec. III E). It is emphasized that the calculations presented include a self-consistent treatment of two-dimensional gate oxide tunneling. Prior treatments of gate oxide tunneling in semiclassical 2D simulators incorporated 1D models.

II. FORMALISM

A. The governing equations

We consider N_b independent valleys for the electrons within the effective mass approximation. The Hamiltonian of valley b is

$$H_b(\mathbf{r}) = -\frac{\hbar^2}{2}\left[\frac{d}{dx}\left(\frac{1}{m_x^b}\frac{d}{dx}\right)+\frac{d}{dy}\left(\frac{1}{m_y^b}\frac{d}{dy}\right)+\frac{d}{dz}\left(\frac{1}{m_z^b}\frac{d}{dz}\right)\right]+V(\mathbf{r}), \tag{1}$$

where (m_x^b, m_y^b, m_z^b) are the components of the effective mass in valley b. The equation of motion for the retarded (G^r) and less-than ($G^<$) Green's functions are[19,33,34]

$$[E-H_{b_1}(\mathbf{r}_1)]G^r_{b_1b_2}(\mathbf{r}_1,\mathbf{r}_2,E) - \int d\mathbf{r}\Sigma^r_{b_1,b'}(\mathbf{r}_1,\mathbf{r},E)G^r_{b',b_2}(\mathbf{r},\mathbf{r}_2,E) = \delta_{b_1,b_2}\delta(\mathbf{r}_1,\mathbf{r}_2) \tag{2}$$

and

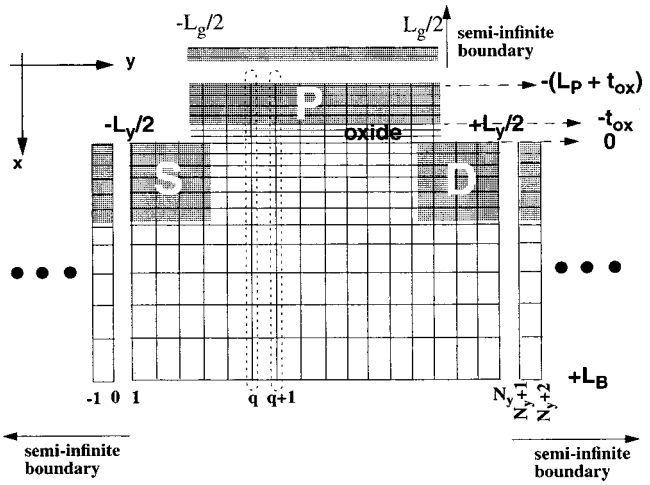


FIG. 1. The equations are solved in a 2D nonuniform spatial grid, with semi-infinite boundaries as shown. Each column q comrises the diagonal blocks of Eqs. (19) and (31). The electrostatic potential is held fixed at the begining of the semi-infinte regions closest to the device.

$$[E-H_{b_1}(\mathbf{r}_1)]G^<_{b_1,b_2}(\mathbf{r}_1,\mathbf{r}_2,E) - \int d\mathbf{r}\Sigma^r_{b_1,b'}(\mathbf{r}_1,\mathbf{r},E)G^<_{b',b_2}(\mathbf{r},\mathbf{r}_2,E) = \int d\mathbf{r}\Sigma^<_{b_1,b'}(\mathbf{r}_1,\mathbf{r},E)G^a_{b',b_2}(\mathbf{r},\mathbf{r}_2,E), \tag{3}$$

where G^a is the advanced Green's function. In the above equations, the coordinate spans only the device (see Fig. 1). The influence of the semi-infinite regions of the source (S), drain (D), and polysilicon gate (P), and scattering mechanisms (electron–phonon) are included via the self-energy terms $\Sigma^r_{b_1,b'}$ and $\Sigma^<_{b_1,b'}$. We assume that charge is injected independently from the contact into each valley. Then, $\Sigma^\alpha_{b_1,b_2,C}=\Sigma^\alpha_{b_1,C}\delta_{b_1,b_2}$, where C represents the self-energy due to contacts. Finally, the hole bands are treated using the drift-diffusion model, which is expected to be a good approximation for n-MOSFETs.

The electrostatic potential varies in the (x,y) plane, and the system is translationally invariant along the z-axis. So, all quantities $A(\mathbf{r}_1,\mathbf{r}_2,E)$ depend only on the difference coordinate z_1-z_2. Using the relation,

$$A(\mathbf{r}_1,\mathbf{r}_2,E)=\int \frac{dk_z}{2\pi}e^{ik_z(z_1-z_2)}A(x_1,y_1,x_2,y_2,k_z,E), \tag{4}$$

the equations of motion for G^r and $G^<$ simplify to

$$\left[E-\frac{\hbar^2k_z^2}{2m_z}-H_b(\mathbf{r}_1)\right]G^r_b(\mathbf{r}_1,\mathbf{r}_2,k_z,E) - \int d\mathbf{r}\Sigma^r_b(\mathbf{r}_1,\mathbf{r},k_z,E)G^r_b(\mathbf{r},\mathbf{r}_2,k_z,E)=\delta(\mathbf{r}_1-\mathbf{r}_2) \tag{5}$$

and

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

$$\left[E-\frac{\hbar^2 k_z^2}{2m_z}-H_b(\mathbf{r}_1)\right]G_b^r(\mathbf{r}_1,\mathbf{r}_2,k_z,E) - \int d\mathbf{r}\Sigma_b^r(\mathbf{r}_1,\mathbf{r},k_z,E)G_b^<(\mathbf{r},\mathbf{r}_2,k_z,E)$$

$$= \int d\mathbf{r}\Sigma_b^<(\mathbf{r}_1,\mathbf{r},k_z,E)G_b^a(\mathbf{r},\mathbf{r}_2,k_z,E), \quad (6)$$

where $Z_b = Z_{b,b}$, and for the remainder of the paper, $\mathbf{r} \rightarrow (x,y)$.

The density of states $[N(\mathbf{r},k_z,E)]$ and charge density $[\rho(\mathbf{r},k_z,E)]$ are the sum of the contributions from the individual valleys,

$$N(\mathbf{r},k_z,E)=\sum_b N_b(\mathbf{r},k_z,E)=-\frac{1}{\pi}\mathrm{Im}[G_b^r(\mathbf{r},\mathbf{r},k_z,E)], \quad (7)$$

$$\rho(\mathbf{r},k_z,E)=\sum_b \rho_b(\mathbf{r},k_z,E)=-iG_b^<(\mathbf{r},\mathbf{r},k_z,E). \quad (8)$$

B. G^r and $G^<$: Discretized matrix equations

Self-consistent solution of the Green's function and Poisson's equations requires repeated computation of the nonequilibrium charge density. This computation is often the most time consuming part in modeling the electronic properties of devices.

The common procedure to evaluate the electron density uses the expression,

$$\rho_b(\mathbf{r},k_z,E) = -iG_b^<(\mathbf{r},\mathbf{r},k_z,E) = -i\int d\mathbf{r}_1 d\mathbf{r}_2 G_b^r(\mathbf{r},\mathbf{r}_1,k_z,E) \times \Sigma_b^<(\mathbf{r}_1,\mathbf{r}_2,k_z,E)G_b^a(\mathbf{r}_2,\mathbf{r},k_z,E, \quad (9)$$

where $G^r(\mathbf{r}_1,\mathbf{r}_2,k_z,E)$ must be computed between all N_xN_y grid points and those grid points involving a nonzero Σ^α. The operation count required to solve for all elements of G^r scales as $(N_xN_y)^3$, and so the use of Eq. (9) is expensive. We have developed a new recursive algorithm to compute the electron density in systems when the discretized version of the LHS of Eqs. (5) and (6) is block tridiagonal. This algorithm requires only the evaluation of the diagonal blocks of G^r. The operation count of this algorithm scales as $N_x^3N_y$ (or $N_xN_y^3$) when the diagonal blocks correspond to lattice points in the x (or y) direction. We summarize the recursive algorithm to calculate G^r (Sec. II C) as it sets the stage for the new algorithm to compute $G^<$ (Sec. II D). We stress that Poisson's equation only requires the diagonal elements of $G^<$ [Eq. (9)]. The algorithm we develop in Sec. II C however computes the diagonal blocks of $G^<$. While this is much better than using Eq. (9) directly as discussed above, new algorithms to solve for only the diagonal elements with operation counts smaller than $N_x^3N_y$ (or $N_xN_y^3$) are very desirable.

In matrix form, Eqs. (5) and (6) are written as

$$A'G^r=\lambda \quad (10)$$

and

$$A'G^<=\Sigma^<G^a. \quad (11)$$

The self-energies due to the S, D, and P are nonzero only along the lines $y=y_S=y_1$, $y=y_D=y_{N_y}$ and $x=x_P$, respectively (see Fig. 1). The A' matrix has a dimension of N_xN_y and is ordered such that all grid points located at a particular y-coordinate correspond to its diagonal blocks. The notation adopted is that $A'_{j_1,j_2}(i,i')$ refers to the off-diagonal entry corresponding to grid points (x_i,y_{j_1}) and (x_i',y_{j_2}). The nonzero elements of the diagonal blocks of the A' matrix are given by

$$A'_{j,j}(i,i) = E' - V_{i,j} - T_{j,j}(i+1,i) - T_{jj}(i-1,i) - T_{j+1,j}(i,i) - T_{j-1,j}(i,i) - \Sigma_S^r(x_i,x_i)\delta_{j,1} - \Sigma_D^r(x_i,x_i)\delta_{j,N_y} - \Sigma_P^r(y_j,y_j)\delta_{i,1} - \Sigma^r(x_i,y_j;x_i,y_j), \quad (12)$$

$$A'_{j,j}(i\pm1,i) = T_{j,j}(i\pm1,i) - \Sigma_S^r(x_{i\pm1},x_i)\delta_{j,1} - \Sigma_D^r(x_{i\pm1},x_i)\delta_{j,N_y} - \Sigma^r(x_{i\pm1},y_j;x_i,y_j), \quad (13)$$

$$A'_{j,j}(i,i') = -\Sigma_S^r(x_i,x_{i'})\delta_{j,1} - \Sigma_D^r(x_i,x_{i'})\delta_{j,N_y}, \quad \text{for } i'\neq i,\ i\pm1, \quad (14)$$

where $E'=E-\hbar^2k_z^2/2m_z$ and $V_{i,j}=V(x_i,y_j)$. The off-diagonal blocks are

$$A'_{j\pm1,j}(i,i) = T_{j\pm1,j}(i,i) - \Sigma_P^r(y_j,y_{j\pm1})\delta_{i,1},$$

$$A'_{j,j'}(i,i') = 0, \quad \text{for } j'\neq j,\ j\pm1. \quad (15)$$

The nonzero elements of the T matrix are defined by

$$T_{j,j}(i\pm1,i) = \frac{\hbar^2}{2m^{\pm x}}\frac{2}{x_{i+1}-x_{i-1}}\frac{1}{|x_{i\pm1}-x_i|}, \quad (16)$$

$$T_{j\pm1,j}(i,i) = \frac{\hbar^2}{2m^{\pm y}}\frac{2}{y_{j+1}-y_{j-1}}\frac{1}{|y_{j\pm1}-y_j|}, \quad (17)$$

where $m^{\pm x}=2/(m_{i\pm1,j}+m_{i,j})$ and $m^{\pm y}=2/(m_{i,j\pm1}+m_{i,j})$. Nonzero elements of $\Sigma_P^r(y_j,y_y')$, where $j'\neq j$ are neglected to ensure that A' is block tridiagonal (the algorithm to calculate G^r and $G^<$ relies on the block tridiagonal form of A'). The λ appearing in Eq. (10) corresponds to the delta function in Eq. (5). λ is a diagonal matrix whose elements are given by

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

$$\lambda_{j,j}^{(i,i)}=\frac{4}{(x_{i+1}-x_{i-1})(y_{i+1}-y_{i-1})}. \tag{18}$$

C. Recursive algorithm to calculate G^r

Premultiplying Eq. (10) by λ^{-1},

$$AG^r=I, \tag{19}$$

where matrix A is a symmetric matrix for both uniform and nonuniform rectangular grids. (Note that A' is symmetric only for a uniform grid.) The recursive algorithm to calculate the diagonal blocks of the full Green's function is discussed now, using Dyson's equation for G^r, and the left-connected Green's function as in Refs. 20, 21;

(i) *Dyson's equation for* G^r: The solution to

$$\begin{pmatrix} A_{Z,Z} & A_{Z,Z'} \\ A_{Z',Z} & A_{Z',Z'} \end{pmatrix}\begin{pmatrix} G^r_{Z,Z} & G^r_{Z,Z'} \\ G^r_{Z',Z} & G^r_{Z',Z'} \end{pmatrix}=\begin{pmatrix} I & O \\ O & I \end{pmatrix}. \tag{20}$$

is

$$G^r=G^{r0}+G^{r0}UG^r \tag{21}$$

$$=G^{r0}+G^rUG^{r0}, \tag{22}$$

where

$$G^r=\begin{pmatrix} G^r_{Z,Z} & G^r_{Z,Z'} \\ G^r_{Z',Z} & G^r_{Z',Z'} \end{pmatrix},$$

$$G^{r0}=\begin{pmatrix} G^{r0}_{Z,Z} & O \\ O & G^0_{Z',Z'} \end{pmatrix}=\begin{pmatrix} A^{-1}_{Z,Z} & O \\ O & A^{-1}_{Z',Z'} \end{pmatrix}, \quad \text{and}$$

$$U=\begin{pmatrix} O & -A_{Z,Z'} \\ -A_{Z',Z} & O \end{pmatrix}. \tag{23}$$

The advanced Green's function (G^a) is by definition related to G^r by

$$G^a=G^{r\dagger}=G^{a0}+G^{a0}U^\dagger G^a \tag{24}$$

$$=G^{a0}+G^aU^\dagger G^{a0}. \tag{25}$$

Equation (21) is called Dyson's equation.[19,33]

(ii) *Left-connected retarded Green's function*: The left-connected (superscript L) retarded (superscript r) Green's function g^{rLq} is defined by the first q blocks of Eq. (19) (includes the open boundaries attached to the lattice points via the self-energy) by

$$A_{1:q,1:q}g^{rLq}=I_{q,q}, \text{ where } I_q=I_{1:q,I:q}. \tag{26}$$

g^{rLq+1} is defined in a manner identical to g^{rLq} except that the left-connected system is comprised of the first $q+1$ blocks of Eq. (19). In terms of Eq. (20), the equation governing g^{rLq+1} follows by setting $Z=1:q$ and $Z'=q+1$. Using Dyson's equation [Eq. (21)], we obtain

$$g^{rLq+1}_{q+1,q+1}=(A_{q+1,q+1}+A_{q+1,q}g^{rLq}_{q,q}A_{q,q+1})^{-1}. \tag{27}$$

Note that the last element $g^{rLN}_{N,N}$ is equal to the fully connected Green's function $G^r_{N,N}$, which is the solution to Eq. (19).

(iii) *Full Green's function in terms of the left-connected Green's function*: Consider Eq. (20) such that $A_{Z,Z}=A_{1:q,1:q}$, $A_{Z',Z'}=A_{q+1:N,q+1:N}$ and $A_{Z,Z'}=A_{1:q,q+1:N}$. Noting that the only nonzero element of $A_{1:q,q+1:N}$ is $A_{q,q+1}$ and using Eq. (21), we obtain

$$G^r_{q,q}=g^{rLq}_{q,q}+g^{rlq}_{q,q}(A_{q,q+1}G^r_{q+1,q+1}A_{q+1,q})g^{rLq}_{q,q} \tag{28}$$

$$=g^{rLq}_{q,q}+g^{rLq}_{q,q}A_{q,q+1}G^r_{q+1,q}. \tag{29}$$

Both $G^r_{q,q}$ and $G^r_{q+1,q}$ are used in the algorithm for electron density, and so storing both sets of matrices will be useful.

In view of the above equations, the algorithm to compute the diagonal blocks $G^r_{q,q}$ is given by the following steps:

(1) $g^{rL1}_{11}=A_1^{-1}$;
(2) For $q=1,2,...,N-1$, compute Eq. (27);
(3) For $q=N-1, N-2,...,1$, compute Eq. (29). Store $G^r_{q+1,q}$ if memory permits for use in the algorithm for electron density.

D. Recursive algorithm to calculate density ($G^<$)

The discretized form of Eq. (6) is

$$A'G^<=\Sigma^<G^a, \tag{30}$$

where the dimension of the matrices involved are $N=N_xN_y$. Premultiplying by λ^{-1},

$$AG^<=\Sigma^<G^a, \tag{31}$$

where $\Sigma^<$ *in Eq. (31) is equal to* λ^{-1} *times the* $\Sigma^<$ *that appears in Eqs. (3) and (30).*

Following Sec. II C, the algorithm to calculate the electron density (diagonal elements of $G^<$) is discussed in terms of a Dyson's equation for $G^<$ and the left-connected $g^{<L}$:

(i) *Dyson's equation for* $G^<$: The solution to

$$\begin{pmatrix} A_{Z,Z} & A_{Z,Z'} \\ A_{Z',Z} & A_{Z',Z'} \end{pmatrix}\begin{pmatrix} G^<_{Z,Z} & G^<_{Z,Z'} \\ G^<_{Z',Z} & G^<_{Z',Z'} \end{pmatrix}$$
$$=\begin{pmatrix} \Sigma^<_{Z,Z} & \Sigma^<_{Z,Z'} \\ \Sigma^<_{Z',Z} & \Sigma^<_{Z',Z'} \end{pmatrix}\begin{pmatrix} G^a_{Z,Z} & G^a_{Z,Z'} \\ G^a_{Z',Z} & G^a_{Z',Z'} \end{pmatrix} \tag{32}$$

can be written as

$$G^<=G^{r0}UG^<+G^{r0}\Sigma^<G^a, \tag{33}$$

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

where G^{r0} and U have been defined in Eqs. (23), and $G^<$ and G^a are readily identifiable from Eq. (32). Using $G^a = G^{a0} + G^{a0}U^\dagger G^a$, Eq. (33) can be written as

$$G^< = G^{<0} + G^{<0}U^\dagger G^a + G^{r0}UG^< \tag{34}$$

$$= G^{<0} + G^r UG^{<0} + G^< U^\dagger G^{a0}, \tag{35}$$

where

$$G^{<0} = G^{r0}\Sigma^< G^{a0}. \tag{36}$$

(ii) *Left-connected* $g^<$: $g^{<Lq}$ is the counter part of g^{rLq}, and is defined by the first q blocks of Eq. (31):

$$A_{1:q,1:q}g^{<Lq} = \Sigma^<_{1:q,1:q}g^{aLq}_{1:q,1:q}. \tag{37}$$

$g^{<Lq+1}$ is defined in a manner identical to $g^{<Lq}$ except that the left-connected system is comprised of the first $q+1$ blocks of Eq. (31). In terms of Eq. (32), the equation governing $g^{<Lq+1}$ follows by setting $Z=1:q$ and $Z'=q+1$. Using the Dyson's equations for G^r and $G^<$, $g^{<Lq+1}_{q+1,q+1}$ can be recursively obtained (derivation is presented in Appendix) as

$$\begin{aligned} g^{<Lq+1}_{q+1,q+1} = {} & g^{rLq+1}_{q+1,q+1}[\Sigma^<_{q+1,q+1} + \sigma^<_{q+1}]g^{aLq+1}_{q+1,q+1} \\ & + g^{rLq+1}_{q+1,q+1}\Sigma^<_{q+1,q}g^{aLq+1}_{q,q+1} \\ & + g^{rLq+1}_{q+1,q}\Sigma^<_{q,q+1}g^{aLq+1}_{q+1,q+1}, \end{aligned} \tag{38}$$

which can be written in a more intuitive form as

$$\begin{aligned} g^{<Lq+1}_{q+1,q+1} = {} & g^{rLq+1}_{q+1,q+1}[\Sigma^<_{q+1,q+1} + \sigma^<_{q+1} + \Sigma^<_{q+1,q}g^{aLq}_{q,q}A^\dagger_{q,q+1} \\ & + A_{q+1,q}g^{rLq}_{q,q}\Sigma^<_{q,q+1}]g^{qLq+1}_{q+1,q+1}, \end{aligned} \tag{39}$$

where $\sigma^<_{q+1} = A_{q+1,q}g^{<Lq}_{q,q}A^\dagger_{q,q+1}$. Equation (39) has the physical meaning that $g^{<Lq+1}_{q+1,q+1}$ has contributions due to four injection functions: (i) an effective self-energy due to the left-connected structure that ends at q, which is represented by $\sigma^<_{q+1}$, (ii) the diagonal self-energy component at grid point $q+1$ that enters Eq. (31), and (iii) the two off-diagonal self-energy components involving grid points q and $q+1$.

(iii) *Full less-than Green's function in terms of left-connected Green's function*: Consider Eq. (32) such that $A_Z = A_{1:q,1:q}$, $A'_Z = A_{q+1:N,q+1:N}$ and $A_{Z,Z'} = A_{1:q,q+1:N}$. Noting that the only nonzero element of $A_{1:q,q+1:N}$ is $A_{q,q+1}$ and using Eq. (34), we obtain

$$\begin{aligned} G^<_{q,q} = {} & g^{<Lq}_{q,q} + g^{<Lq}_{q,q}A^\dagger_{q,q+1}G^a_{q+1,q} + g^{<0}_{q,q+1}A^\dagger_{q+1,q}G^a_{q,q} \\ & + g^{rLq}_{q,q}A_{q,q+1}G^<_{q+1,q}. \end{aligned} \tag{40}$$

Using Eq. (35), $G^<_{q+1,q}$ can be written in terms of $G^<_{q+1,q+1}$ and other known Green's functions as

$$\begin{aligned} G^<_{q+1,q} = {} & g^{<0}_{q+1,q} + G^r_{q+1,q}A_{q,q+1}g^{<0}_{q+1,q} \\ & + G^r_{q+1,q+1}A_{q+1,q}g^{<Lq}_{q,q} + G^\dagger_{q,q+1}A^\dagger_{q,q+1}g^{aLq}_{q,q}. \end{aligned} \tag{41}$$

Substituting Eq. (41) in Eq. (40) and using Eqs. (21) and (22), we obtain

$$\begin{aligned} G^<_{q,q} = {} & g^{<Lq}_{q,q} + g^{rLq}_{q,q}(A_{q,q+1}G^<_{q+1,q+1}A^\dagger_{q+1,q})g^{aLq}_{q,q} \\ & + [g^{<Lq}_{q,q}A^\dagger_{q,q+1}G^a_{q+1,q} + G^r_{q,q+1}A_{q+1,q}g^{<Lq}_{q,q}] \\ & + [g^{<0}_{q,q+1}A^\dagger_{q+1,q}G^a_{q,q} + G^r_{q,q}A_{q,q+1}g^{<0}_{q+1,q}], \end{aligned} \tag{42}$$

where

$$g^{<0}_{q,q+1} = g^{r0}_{q,q}\Sigma^<_{q,q+1}g^{a0}_{q+1,q+1}, \tag{43}$$

$$g^{<0}_{q+1,q} = g^{r0}_{q+1,q+1}\Sigma^<_{q+1,q}g^{a0}_{g,g}. \tag{44}$$

The terms inside the square brackets of Eq. (42) are Hermitian conjugates of each other.

In view of the above equations, the algorithm to compute the diagonal blocks of $G^<$ is given by the following steps:

(1) $g^{<L1}_{11} = g^{r0}_{11}\Sigma^<_L g^{a0}_{11}$;
(2) For $q=1,2,...,N-1$, compute Eq. (39);
(3) For $q=N-1, N-2,...,1$, compute Eqs. (42)–(44).

The current density flowing between two neighboring blocks q and $q+1$ is given by

$$\begin{aligned} J(q\to q+1,k_z,E) = {} & \sum_b J_b(q\to q+1,k_z,E) \\ = {} & \frac{2e}{h}\sum_b \mathrm{Tr}[T_{q,q+1}G^<_{b;q,q+1}(k_z,E) \\ & - T_{q+1,q}G^<_{b;q+1,q}(k_z,E)], \end{aligned} \tag{45}$$

where T has been defined in Eqs. (16) and (17). The current that has leaked into the gate between any two blocks p and q is

$$J^{qp}_{\text{gate}} = \sum_b J_b(p\to p+1,k_z,E) - \sum_b J_b(q\to q+1,k_z,E), \tag{46}$$

and the total gate leakage current obtained by choosing p and q near the source and drain ends of the device.

E. Expressions for contact self-energies (Σ^r_S, Σ^r_D, and Σ^r_P)

Potential and doping profiles in the semi-infinite regions to the (a) left of "S" and right of "D" are equal to the value at $q=1$ and N_y, respectively (Fig. 1). That is, they do not vary as a function of the y-coordinate, and (b) top of the "P" is equal to the value of the topmost grid line of "P" (Fig. 1). That is, they are not a function of the x-coordinate. The retarded surface Green's functions of these semi-infinite regions are calculated from Eq. (19), when the matrices involved are semi-infinite. All diagonal submatrices of the A matrix are equal to $A_{1,1}$, A_{N_y,N_y}, and A_P, and all first upper off-diagonal matrices of the A matrix are equal to $A_{1,2}$, A_{N_y-1,N'_y}, and $A_{P-1,P}$, in the source, drain and polysilicon regions, respectively. We spell out the entire matrix for the source semi-infinite regions below (see Fig. 1):

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

$$\begin{pmatrix} \cdot & \cdot & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & 0 & 0 & 0 \\ 0 & A_{2,1} & A_{1,1} & A_{1,2} & 0 & 0 \\ 0 & 0 & A_{2,1} & A_{1,1} & A_{1,2} & 0 \\ 0 & 0 & 0 & A_{2,1} & A_{1,1} & A_{1,2} \\ 0 & 0 & 0 & 0 & A_{2,1} & A_{1,1} \end{pmatrix} \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & g_{-3,-3} & g_{-3,-2} & g_{-3,-1} & g_{-3,0} \\ \cdot & \cdot & g_{-2,-3} & g_{-2,-2} & g_{-2,-1} & g_{-2,0} \\ \cdot & \cdot & g_{-1,-3} & g_{-1,-2} & g_{-1,-1} & g_{-1,0} \\ \cdot & \cdot & g_{0,-3} & g_{0,-2} & g_{0,-1} & g_{0,0} \end{pmatrix} = \begin{pmatrix} \cdot & 0 & 0 & 0 & 0 & 0 \\ 0 & \cdot & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & I \end{pmatrix}. \quad (47)$$

The surface Green's function of these regions can be obtained by using methods in matrix algebra that transform the two-dimensional wire representing the semi-infinite contacts with N_x grid points to N_x one-dimensional wires.

The self-energies due to the contacts are (see Fig. 1)

$$\Sigma_S^r(k_z,E) = A_{1,0} g_{0,0}(k_z,E) A_{0,1}, \quad (48)$$

$$\Sigma_D^r(k_z,E) = A_{N_y,N_y+1} g_{N_y+1,N_y+1}(k_z,E) A_{N_y+1,N_y}, \quad (49)$$

$$\Sigma_P^r(k_z,E) = A_P g_P(k_z,E) A_P, \quad (50)$$

$$\Sigma_S^<(k_z,E) = -2iA_{1,0}\,\mathrm{Im}[g_{0,0}(k_z,E)]A_{0,1} f_S(E), \quad (51)$$

$$\Sigma_D^<(k_z,E) = -2iA_{N_y,N_y+1}\,\mathrm{Im}[g_{N_y+1,N_y+1}(k_z,E)] \times A_{N_y+1,N_y} f_D(E), \quad (52)$$

$$\Sigma_P^<(k_z,E) = -2iA_P\,\mathrm{Im}[g_P(k_z,E)]A_P f_P(E), \quad (53)$$

where $f_i(E)$ is the Fermi factor in contact $i \in S, D, P$.

When $\Sigma_b^\alpha(\mathbf{r}_1,\mathbf{r}_2,k_z,E)$ depends only on $E_{xy} = E - (\hbar^2 k_z^2/2m_z)$, then Eqs. (5) and (6) simplify to

$$[E_{xy} - H_b(\mathbf{r}_1)]G_b^r(\mathbf{r}_1,\mathbf{r}_2,E_{xy}) - \int d\mathbf{r}\Sigma_b^r(\mathbf{r}_1,\mathbf{r},E_{xy})G_b^r(\mathbf{r},\mathbf{r}_2,E_{xy}) = \delta(\mathbf{r}_1-\mathbf{r}_2) \quad (54)$$

(55)

and

$$[E_{xy} - H_b(\mathbf{r}_1)]G^<(\mathbf{r}_1,\mathbf{r}_2,E_{xy}) - \int d\mathbf{r}\Sigma_b^r(\mathbf{r}_1,\mathbf{r},E_{xy})G_b^<(\mathbf{r},\mathbf{r}_2,k_z,E) = \int d\mathbf{r}\Sigma_a^<(\mathbf{r}_1,\mathbf{r},E_{xy})G_b^a(\mathbf{r},\mathbf{r}_2,E_{xy}). \quad (56)$$

While solving the equations, to keep the problem two-dimensional, m_z has to be independent of (x,y). So, we assume $m_z(\mathrm{SiO_2}) = m_z(\mathrm{Si})$.

III. RESULTS AND DISCUSSION

The steady state characteristics of MOSFETS that are of practical interest are the drive current, OFF current, slope of drain current vs drain voltage, and threshold voltage. In this section, we show that quantum mechanical simulations yield significantly different results from drift-diffusion based methods. These differences arise because of the following quantum mechanical features:

(i) Polysilicon gate depletion in a manner opposite to the classical case;
(ii) Dependence of the resonant levels in the channel on the gate voltage;
(iii) Tunneling of charge across the gate oxide and from source to drain;
(iv) Quasiballistic flow of electrons.

The MIT well-tempered 25 nm device structure[35] is chosen for the purpose of discussion (MIT 25 nm device structure[35] is hereafter referred to as MIT25). The method and computer code developed can however handle a wide variety of two-dimensional structures with many terminals. We first compare the potential profiles from a constant mobility drift-diffusion solution and our quantum calculations at equilibrium. The motivation for this comparison results from the observation that the classical and quantum potential profiles should be in reasonable agreement, if the doping density is significantly higher than the electron and hole densities and the boundary conditions are the same. The doping profile of MIT25 meets this requirement in the channel region at small V_g, and we verify that the potential profiles are in reasonable agreement at $y=0$ (see "Q1 flat band" and "DD flat band" of Fig. 2). The legend "flat band" refers to the potential at $x=-t_{\mathrm{ox}}$ being fixed at the applied gate potential.

An index of abbreviations used is as follows:

Length scales: t_{ox}, oxide thickness; L_P, polysilicon gate thickness in the x-direction; L_B, boundary of substrate region in the x-direction; L_y, Poisson's and NEGF equations are solved from $-L_y/2$ to $+L_y/2$; L_g, length of polysilicon gate region in the y-direction.

Models: Q1, quantum mechanical calculations using an isotropic effective mass; Q3, quantum mechanical calculations using an anisotropic effective mass; DD, drift diffusion; flat band, potential in the polysilicon gate region is held fixed from $x=-(t_{\mathrm{ox}}+L_P)$ to $x=-t_{\mathrm{ox}}$ at the bulk value. q-poly, potential in the gate polysilicon region is held fixed at $x=-(t_{\mathrm{ox}}+L_P)$ at the bulk value, and the potential is computed quantum mechanically (self-consistently) for $x>-(t_{\mathrm{ox}}+L_P)$; c-poly, classical treatment of gate polysilicon region, as in DD.

Current and voltage: I_d, drain current; I_g, gate current; V_d, drain voltage; V_g, gate voltage.

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

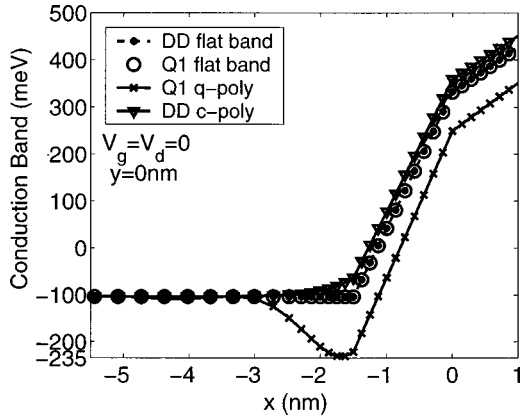


FIG. 2. Potential profile at the $y=0$ slice of MIT25, calculated by four different methods. Note the qualitative difference of the "Q1 *q*-poly" case due to electron depletion in the gate.

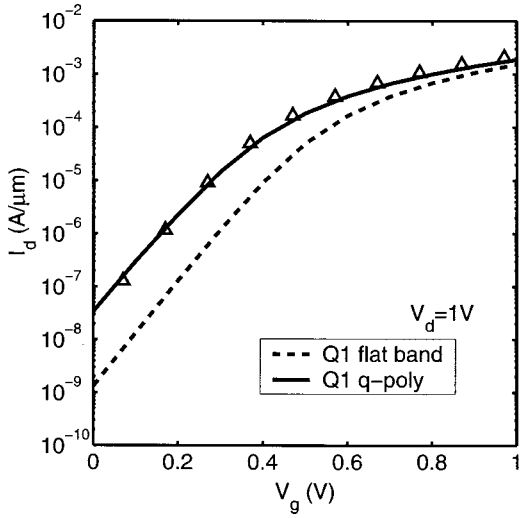


FIG. 3. Drain current vs gate voltage for $V_d=1$ V. Quantum mechanical treatment of the polysilicon gate (Q1 *q*-poly) results in much higher current.

The values of constants assumed to obtain the numerical results of this section are as follows, unless otherwise noted.

Electron effective mass of silicon: 0.3283 (isotropic), 0.19 and 0.98 (anisotropic), electron effective mass of SiO_2. $m_x=m_y=0.5$ and same as silicon in m_z direction, hole effective mass of silicon is 0.49, band gap of silicon (SiO_2): 1.12 eV (8.8 eV), energy barrier between the silicon and the oxide $\Delta E_C=3.1$ eV, dielectric constant of Si(SiO_2) is $\epsilon_{Si}=11.9$ (3.9) and $kT=0.025\,85$ eV.

A. I_d vs V_g: Effect of polysilicon depletion region

The quantum mechanically calculated electron density near the SiO_2 barrier in the polysilicon region is smaller than the uniform background doping density. This is because the electron wave function is small close to the barrier. As a result, the conduction band in the polysilicon gate bends in a direction opposite to that computed semiclassically (compare *x* and triangle in Fig. 2).[36,37] The band bending in the polysilicon gate plays a significant role in determining the threshold voltage and OFF current. To emphasize the importance of band bending, we plot the drain current vs gate voltage calculated with the gate polysilicon region treated as (i) "flat band" and (ii) "*q*-poly." We find that the computed current is larger in (ii) because quantum mechanical depletion of electrons in the polysilicon gate region close to the oxide causes lowering of the potential in the channel. The I_d vs V_g curve shifts by approximately an amount equal to the band bending in the polysilicon gate, in comparison to the flat band case. This band bending, which is measured from $-(L_P+t_{ox})$ to $-t_{ox}$ at equilibrium, is about 130 meV at the given doping density (Fig. 2). The influence of band gap narrowing has been neglected here. It must be mentioned that the band gap narrowing effect will tend to make the quantum mechanical contribution to the polysilicon band bending just discussed smaller. Future work to determine the influence of bandgap narrowing is necessary.

Computationally, a 2D treatment of the polysilicon gate region is expensive because of the additional grid points required. Note that matrix inversion depends on the cube of the matrix dimension. We point out that for highly doped polysilicon gate (in the absence of gate tunneling), a shift in the $I_d(V_g)$ curve from (i) by the equilibrium 1D built-in potential does a reasonable job of reproducing the quantum mechanical result (see triangles in Fig. 3). This approximation becomes progressively poorer with increase in gate voltage, as can be seen from the figure. This is true especially for a smaller polysilicon doping density such as 1E20 (results are not discussed here).

B. I_d vs V_g: Comparison to Medici

In the absence of gate tunneling and inelastic tunneling, the quantum mechanical current is

$$I_d=\frac{2e}{h}\int dE T_{SD}(E)[f_S(E)-f_D(E)], \tag{57}$$

where T_{SD} is the transmission probability from source to drain, and f_S and f_D are the Fermi–Dirac factors in the source and drain, respectively. The total transmission (Fig. 4) is steplike with integer values at the plateaus in-spite of the complicated two-dimensional electrostatics. In visual terms, the energies at which the steps turn on are determined by an effective "subband dependent" source injection barrier, in contrast to the source injection barrier in drift-diffusion calculations.[10] This subband dependent source injection barrier is simply the maximum energy of the subband between source and drain due to quantization in the direction perpendicular to the gate plane (*x*-direction of Fig. 1). From a practical view point, the following two issues are important in ballistic MOSFETs: (a) typically, the total transmission assumes integer value at an energy slightly above the maximum in density of states as shown in the inset of Fig. 4, and (b) the steps develop over 50 meV (twice the room temperature thermal energy). So, the shape of the steps is important in determining the value of current. Assuming a sharp step in total transmission with integer values in a calculation of current as in Ref. 9 is not quite accurate.

We compare the results from our quantum simulations with published results from quantum-corrected Medici.[35] To compare the quantum and classical results, an estimate of the

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

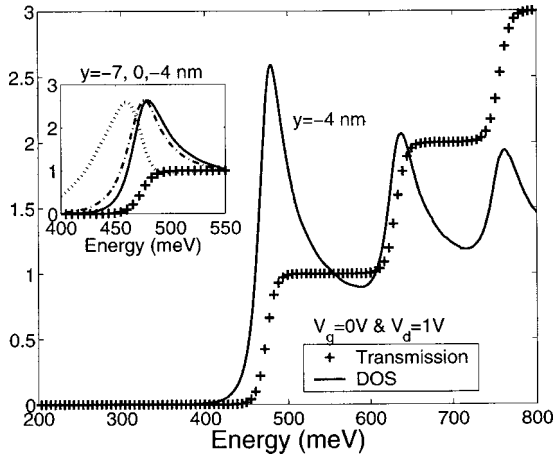


FIG. 4. Transmission (+) and density of states (DOS) vs energy at a spatial location close to the source injection barrier, at $V_g=0$ V and $V_d=1$ V. The peaks in the density of states represent the resonant levels in the channel. Inset: DOS at three different y-locations and the total transmission. The points $y=-7$ and 0 nm are to the left and right of the location where the source injection barrier is largest (close to $y=-4$ nm).

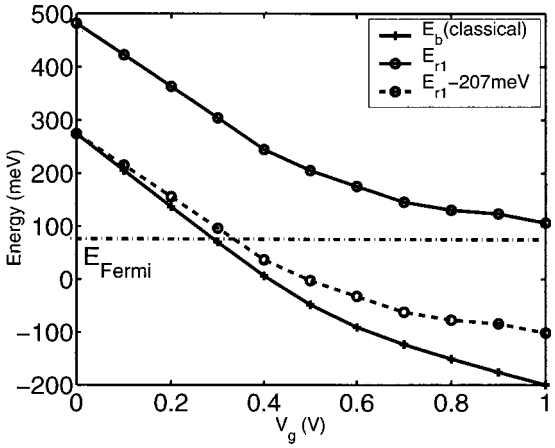


FIG. 5. Location of the first resonant level (E_{r1}) vs gate voltage and the classical source injection barrier [E_b (classical)]. Note that E_{r1} decreases slower than E_b (classical) with gate voltage due to narrowing of channel potential well.

energy of the first subband minima (E_{r1}) from Fig. 4, and the location of the classical barrier height [E_b(classical)] (Fig. 5) are useful [E_b(classical) shown is obtained from constant mobility simulations using Prophet]. The main features of this comparison are:

(a) Subthreshold region: The slope $d[\log(I_d)]/dV_g$ is smaller in the quantum case when compared to Medici (Fig. 6). Further, the current resulting from the simple intuitive expression

$$I=I_{q0}e^{(-E_r/kT)} \tag{58}$$

matches the quantum result quite accurately. I_{q0} is a prefactor chosen to reproduce the current at $V_g=0$ in Fig. 6. This match is rationalized by noting that for the values of gate biases considered, E_{r1} is well above the source Fermi energy, and E_{r2} is many kT (thermal energy) above E_{r1}. The difference in slope between the classical and quantum results can be understood from the slower variation of E_{r1} in comparison to E_b(classical) as a function of V_g (Fig. 5). We also find that the decrease of E_{r1} with increases in gate voltage is slower than the barrier height determined from the quantum potential profiles. This arises because (neglecting 2D effects) E_{r1} is determined by a triangular well (whose apex is the conduction band) that becomes progressively narrower with increase in gate voltage.

(b) Large gate biases: The drain current and slope $d[\log(I_d)]/dV_g$ are larger in the quantum case. The higher $d[\log(I_d)]/dV_g$ at large gate voltages in the quantum case can be understood from the fact that E_{r1} is above the Fermi level while E_b(classical) is below, at $V_g=1$ V (the quantum current is proportional to $\exp(-(E_{r1}-E_F)/kT)$). The mobility model assumed in the classical case also plays a role in determining the slope.

C. I_d vs V_d

The values of dI_d/dV_d and drive current are important in MOSFET applications because they determine switching speeds.[7] Figure 7 compares the drain current vs drain voltage for $V_g=0$ and $V_g=1$ V. The drive current ($V_g=1$ V) calculated using Q1 with the polysilicon region treated in the flat band and q-poly approximations is more than 100% and 200% larger than the results in Ref. 35. dI_d/dV_d in the linear region is up to three times larger in Q1. The subthreshold drain current is smaller in Q1. We however expect that with decreasing channel length, the subthreshold I_d will become larger than the Medici results due to quantum mechanical tunneling.[12]

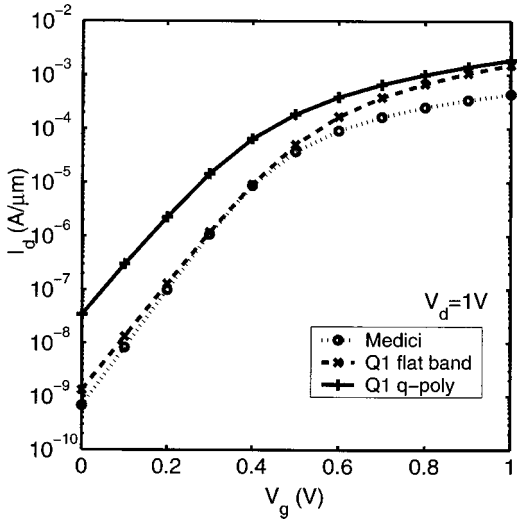


FIG. 6. Plot of drain current vs gate voltage from the quantum mechanical calculations and Medici, at $V_d=1$ V. At small gate voltages, the drain current from Medici (see Ref. 35) are comparable to the "Q1 flat band" results. The drain current from "Q1 q-poly" is, however, significantly different at large gate voltages.

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

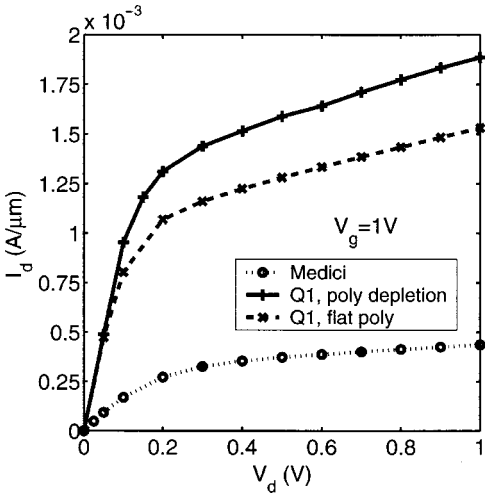


FIG. 7. Plot of drain current vs drain voltage (V_d) from the quantum mechanical calculations and Medici, at $V_g=1$ V. Note the large difference in drive current and dI_d/dV_d between Medici, (see Ref. 35) "Q1 flat band" and "Q1 *q*-poly."

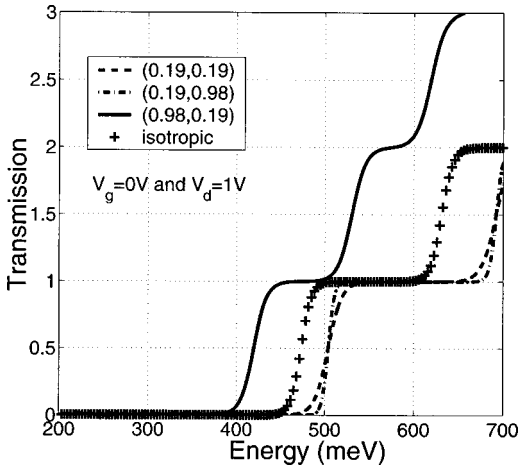


FIG. 8. Same as Fig. 4 but the anisotropic effective mass case is included. Note that the valley with the largest mass in the *x*-direction has subband energies that are about 50 meV smaller than the isotropic effective mass case even at $V_g=0$.

D. Isotropic vs anisotropic effective mass

The primary influence of anisotropic effective mass is to influence the energy of the subbands in the inversion layer. Valleys with the largest effective mass perpendicular to the oxide ($0.98m_0^*$) have subband energies that are smaller than the isotropic effective mass case. We see from the plot of transmission vs energy (Fig. 8) that the valleys with ($m_x=0.98m_0^*, m_y=m_z=0.19m_0^*$) have resonance levels that are more than 50 meV lower in energy than the isotropic effective mass case. The corresponding subthreshold current (Fig. 9) is a few hundred percent larger than the value obtained from the isotropic effective mass case. This follows by noting that the subthreshold current depends on $\exp(-E_{r1}/kT)$. The drive current (Fig. 9) from the anisotropic effective mass case is more than 25% larger than the isotropic effective mass case. Note that for large gate voltages the dependence of current on E_{r1} is subexponential. We are not aware of any calculations that compare the relative importance of the current carrying capacity of electrons in the three inequivalent valleys. We find that the valley with the largest m_x ($=0.98m_0^*$) carries 89.22% and 79.77% of the current at V_g equal to 0 and 1 V, respectively ($V_d=1$ V). Thus all three valleys are necessary for an accurate calculation of the ballistic current.

E. Gate leakage current

A major problem in MOSFETs with ultrathin oxides is that tunneling from gate to drain will determine the OFF current. The gate leakage current vs y is plotted for the MIT25 device in Fig. 10. Note that while we use a value of 3.0 for the dielectric constant of SiO_2, a value of 3.9 does not change the qualitative conclusions. At $V_g=0$ V and $V_d=1$ V, the main path for leakage current is from the polysilicon gate contact on top of the oxide to the highly doped (n^+) regions associated with the drain [source drain extension (SDE)] as shown in Fig. 10(a). At nonzero V_g, there is also an appreciable tunneling from the highly doped n^+ regions near the source to the polysilicon region on top of the gate [Fig. 10(a)]. For $t_{ox}=1.5$ nm, gate tunneling increases the OFF current by about two orders of magnitude, and for smaller oxide thicknesses, the gate leakage current is significantly larger.

We propose that the gate leakage current can be reduced by a factor of 10–100 without significantly compromising the drive current. The drive current in these ultra small MOSFETs is primarily determined by the source injection barrier, or more correctly as discussed earlier by the resonant level at the source injection barrier. So any changes that result in a reduction of the gate leakage current should not significantly alter the location of the resonant level at the

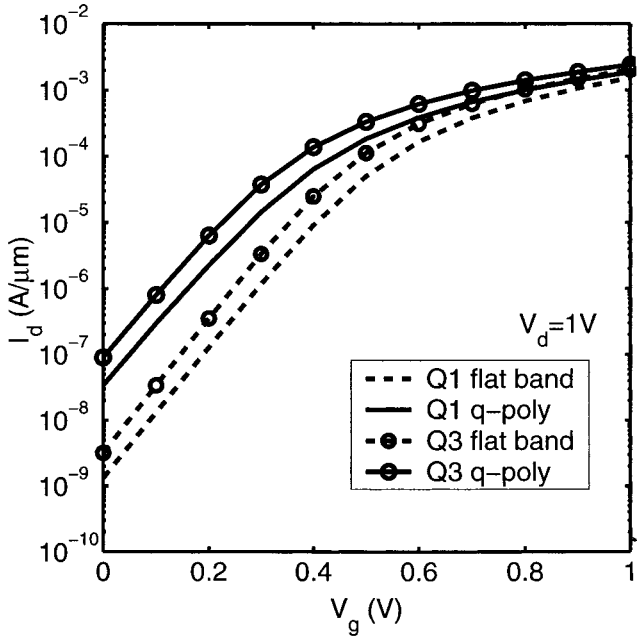


FIG. 9. Plot of drain current vs gate voltage for the isotropic and anisotropic effective mass cases, at $V_d=1$ V. The much higher current in the anisotropic effective mass case (Q3) is due to the lower suband energy shown in Fig. 8.

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

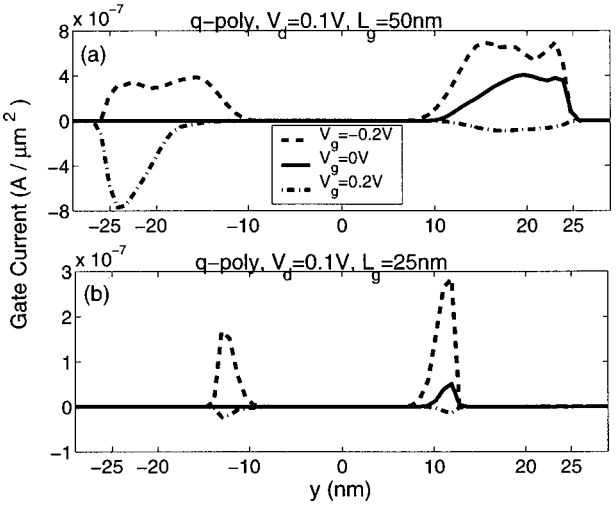


FIG. 10. Plot of gate leakage current when the device is OFF ($V_g=0$ V) as a function of the *y*-direction, from the source to drain, for L_g equal to (a) 50 nm and (b) 25 nm. Note the significant gate leakage current in the regions where the high doping in the source and drain overlap the gate in (a). A shorter gate eliminates a large fraction of the gate leakage current as shown in (b).

source injection barrier (and hence the drive current). Two methods (without regard to fabrication issues) that help in this direction are discussed below:

(i) *Shorter or asymmetric polysilicon gate region*: We propose that the gate leakage current can be significantly reduced by using shorter gate lengths. The main feature of the shorter gate lengths is a small overlap between the polysilicon gate and the n^+ region near the drain. This is pictorially represented in Figs. 11(a) and 11(b) with ''long'' and ''short'' gate lengths. To simulate the long and short gate lengths, we consider the doping profile of MIT25 with $L_g=25$ nm and 50 nm (gate length in Ref. 35). The OFF current and gate leakage current are plotted in Fig. 12. We see that the gate leakage current reduces by more than an order of magnitude, and the drive current is within two percent of the $L_g=50$ nm case, as desired (see inset of Fig. 12). The spatial profile of gate leakage current for $L_g=25$ nm is shown in Fig. 10(b). Though the gate leakage current reduces significantly, a drawback of this scheme is the requirement for very short (approximately equal to the distance between highly doped region near source and drain) polysilicon gate lengths. A polysilicon gate placed asymmetrically with respect to $y=0$ such that its overlap with the n^+ regions near the drain is small, will also serve to reduce the OFF current without compromising the drive current.

(ii) *Graded oxide*: The second proposal is to use a graded oxide, which is thinner close to the source end and thicker close to the drain end [Fig. 11(c)]. The thinner oxide near the source is not expected to alter the source injection barrier significantly, while the tunneling rate from gate to drain will be significantly smaller because of the thicker oxide in the drain-gate overlap region. We consider an oxide that is 1.5 nm thick for $y<+10$ nm and 2.5 nm for $y>11$ nm, with the thickness varying linearly in between. The polysilicon gate lengths is 50 nm. Comparison of this device to the original MIT25 with an uniform oxide and L_g

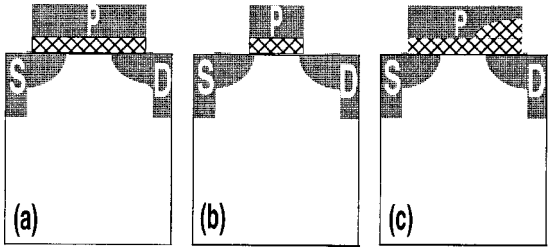


FIG. 11. Polysilicon gate and oxide configurations that could reduce the OFF current ($V_g=0$ V) significantly without drastically reducing the drive current ($V_g=1$ V). The hatched marks represent the oxide.

$=50$ nm show that while the gate leakage current decreases by one order of magnitude, the drive current decreases by only 30%. Further optimization of this device structure could yield a larger drive current, while keeping the gate leakage current small.

IV. CONCLUDING REMARKS

A modeling framework and computer code to calculate properties of ballistic MOSFETs with open boundaries at the source, drain and gate contacts have been developed. This includes an algorithm to compute the electron density using the NEGF equations that avoids solving for the entire G^r matrix even in the presence of nonzero self-energies throughout the device. Note that the simulations presented are 2D in nature and also involve self-consistency. As a result, they were numerically intensive and were typically performed on sixteen to sixty four processors of an SGI Origin machine.

The main results of this study are:

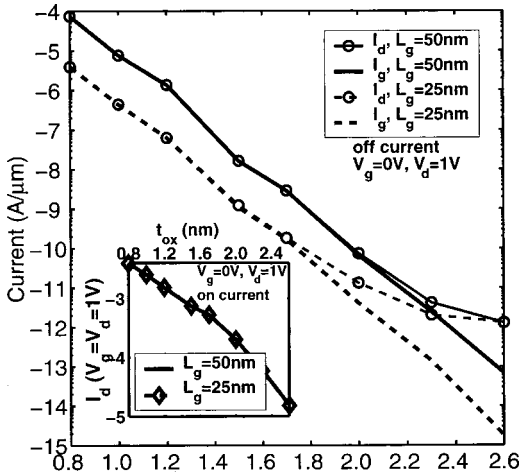


FIG. 12. Plot of drain and gate currents when the device is OFF ($V_g=0$ V) vs oxide thickness for L_g equal to 50 and 25 nm. Inset: Drain current for the the gate lengths when the device is on ($V_g=1$ V). At the larger values of t_{ox}, the gate current (I_g) is significantly smaller than the drain current (I_d), meaning that the drain current is determined by electron injected from the source to drain. At smaller values of t_{ox}, the drain current is dominated by the gate leakage current as can be seen by comparing I_d and I_g in this figure. More importantly, note that the shorter gate length ($L_g=25$ nm) gives an order of magnitude smaller drain current when the device is OFF for the smaller values of t_{ox}. The inset shows that the drive current ($V_d=V_g=1$ V) is however not effected much by the shorter gate length.

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

(a) Polysilicon gate depletion causes the conduction band close to the oxide interface to bend in a manner opposite to the semiclassical case (Fig. 2). This causes a substantial shift in the location of the conduction band bottom in the channel, which gives rise to drain currents that are different from the semiclassical case by one to two orders of magnitude. Performing quantum mechanical calculations with a flat polysilicon region, and then shifting the gate voltage axis (in I_d vs V_g) by the quantum mechanical built-in voltage shown in Fig. 2 results in an order of magnitude better agreement with results from a quantum mechanical treatment of the polysilicon region. This built-in voltage can simply be determined by 1D simulations or an analytical expression. In reality, treatment of discrete dopants in the polysilicon region will give rise to results that are in between the ''flat band'' and ''*q*-poly'' cases presented in the paper.

A quantum mechanical treatment of the polysilicon gate region results in an OFF current ($V_g=0$ V and $V_d=1$ V) that is more than 35 times larger than the OFF current from a flat band treatment of polysilicon region and published results[35] based on a sophisticated semiclassical simulator.

(b) Resonant levels in the channel the from source to drain increase the effective source injection barrier for ballistic electrons. Further, even in the ballistic limit the transmission versus energy reaches integer values over an energy range that could be many times the thermal energy. Knowledge of the detailed shape of transmission versus energy is important to accurately determine the ballistic current. The precise shape of these transmission steps depends on the details of the channel to source and drain overlap regions and the resulting 2D potential profile. Assuming a sharp steplike increases in the total transmission is incorrect.

The slope dI_d/dV_d, whose importance was emphasized in Ref. 7 and the drive current (at $V_g=1$ V) are about 300% larger than reported in Ref. 35. Further, inclusion of anisotropic effective mass in our calculation makes the quantum results deviate further from the semiclassical results as shown in Fig. 9.

(c) Tunneling of charge across the gate oxide can put a limit on the OFF current. Models of the tunnel current for thin oxide MOSFETs are important. We model the gate leakage current in two dimensions and show that significant reduction in the OFF current is possible without altering the drive current significantly. This is accomplished by changing either the gate length [Fig. 11(b)] or by introducing a graded oxide [Fig. 11(c)].

(d) Quasiballistic flow of electrons causes the slope of $d[\log(I_d)]/dV_g$ to be larger than the values obtained from drift-diffusion methods using field dependent mobility models.

This paper dealt with the modeling of steady state properties of nanoscale MOSFETs in the ballistic limit. Future work in quantum mechanical simulation of MOSFETs that are of importance include: (i) treatment of scattering mechanisms such as interface roughness and electron–phonon scattering,[38,39] (ii) treatment of discrete impurity dopants,[40,41] (iii) switching behavior of MOSFETs/time-dependent simulation,[42,43] and (iv) noise characteristics of nanotransistors.

ACKNOWLEDGMENTS

The authors would like to thank Gerhard Klimeck (JPL) and Mark Lundstrom (Purdue University) for their interest in our work and for their useful comments, and Mark Lundstrom and Kent Smith (Bell Laboratory) for recommending a procedure for faster convergence of Poisson's and NEGF equations. The calculations were performed on an SGI Origin 2000 machine located at NASA Advanced Supercomputing (NAS) Division, whom we acknowledge. We thank Bron C. Nelson of NAS/SGI for resolving some high performance computing related questions on the SGI Origin machine and Prabhakar Shatdarsham for his valuable help with Matlab. We thank Supriyo Bandyopadyay (UNL, University of Nebraska, Lincoln) and Meyya Meyyappan (NASA Ames Research Center) for arranging this collaborative effort between NASA Ames Research Center and UNL.

APPENDIX

Derivation of Eqs. (38) and (39):

Using Dyson's equation for $G^<$ [Eq. (33)], we obtain

$$g^{<Lq+1}_{q+1,q+1}=g^{r0}_{q+1,q+1}A^{\dagger}_{q+1,q}g^{<Lq+1}_{q,q+1}+g^{r0}_{q+1,q+1}\Sigma^{<}_{q+1,q+1}g^{aLq+1}_{q+1,q+1}+g^{r0}_{q+1,q+1}\Sigma^{<}_{q+1,q}g^{qLq+1}_{q,q+1}. \tag{A1}$$

Using Eq. (34), $g^{<Lq+1}_{q,q+1}$ can be expressed in terms of $g^{<Lq+1}_{q+1,q+1}$, the quantity we are solving for and known Green's functions as

$$g^{<Lq+1}_{q,q+1}=g^{<0}_{q,q+1}+g^{<0}_{q,q+1}A^{\dagger}_{q+1,q}g^{aLq+1}_{q,q+1}+g^{<Lq}_{q,q}A^{\dagger}_{q,q+1}g^{aLq+1}_{q+1,q+1}+g^{rLq}_{q,q}A_{q,q+1}g^{<Lq+1}_{q+1,q+1}. \tag{A2}$$

Substituting Eq. (A2) in Eq. (A1), we obtain

$$\begin{aligned}[I-g^{r0}_{q+1,q+1}A_{q+1,q}g^{rLq}_{q,q}A_{q,q+1}]g^{<Lq+1}_{q+1,q+1}&=g^{r0}_{q+1,q+1}\Sigma^{<}_{q+1,q+1}g^{aLq+1}_{q+1,q+1}+g^{r0}_{q+1,q+1}\Sigma^{<}_{q+1,q}g^{aLq+1}_{q,q+1}\\&+g^{r0}_{q+1,q+1}A_{q+1,q}[g^{<0}_{q,q+1}+g^{<0}_{q,q+1}A^{\dagger}_{q+1,q}g^{aLq+1}_{q,q+1}\\&+g^{<Lq}_{q,q}A^{\dagger}_{q,q+1}g^{aLq+1}_{q+1,q+1}].\end{aligned} \tag{A3}$$

Using Eq. (27) and $g^{<0}_{q,q+1}=g^{rLq}_{q,q}\Sigma^{<}_{q,q+1}g^{a0}_{q+1,q+1}$, which follows from Eq. (36), we obtain

$$g^{<Lq+1}_{q+1,q+1}=g^{rLq+1}_{q+1,q+1}[\Sigma^{<}_{q+1,q+1}+A_{q+1,q}g^{<Lq}_{q,q}A^{\dagger}_{q,q+1}]g^{aLq+1}_{q+1,q+1}+g^{rLq+1}_{q+1,q+1}\Sigma^{<}_{q+1,q}g^{aLq+1}_{q,q+1}+g^{rLq+1}_{q+1,q+1}A_{q+1,q}g^{rLq}_{q,q}\Sigma^{<}_{q,q+1}g^{aLq+1}_{q+1,q+1}. \tag{A4}$$

Noting that $g^{rLq+1}_{q+1,q}=g^{rLq+1}_{q+1,q+1}A_{q+1,q}g^{rLq}_{q,q}$, Eq. (A4) can be written as

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

$$g^{<Lq+1}_{q+1,q+1} = g^{rLq+1}_{q+1,q+1}[\Sigma^{<}_{q+1,q+1} + \sigma^{<}_{q+1}]g^{aLq+1}_{q+1,q+1} + g^{rLq+1}_{q+1,q+1}\Sigma^{<}_{q+1,q}g^{aLq+1}_{q,q+1} + g^{rLq+1}_{q+1,q}\Sigma^{<}_{q,q+1}g^{aLq+1}_{q+1,q+1}, \quad \text{(A5)}$$

where

$$\sigma^{<}_{q+1} = A_{q+1,q}g^{<Lq}_{q,q}A^{\dagger}_{q,q+1}. \quad \text{(A6)}$$

[1] J. M. Hergenrother *et al.*, International Electron Devices Meeting, 1999, IEDM Technical Digest, p. 75.
[2] G. Timp *et al.*, The ballistic nanotransistor, International Electron Devices Meeting, 1999, IEDM Technical Digest, p. 55.
[3] J. Appenzeller, R. Martel, P. Solomon *et al.*, Appl. Phys. Lett. **77**, 298 (2000).
[4] Y. Taur, D. A. Buchanan, W. Chen *et al.*, Proc. IEEE **85**, 486 (1997).
[5] Y. Taur, C. H. Wann, and D. J. Frank, 1998, IEDM Technical Digest, p. 789.
[6] H-S. P. Wong, D. J. Frank, P. M. Solomon, C. H. J. Wann, and J. L. Welser, Proc. IEEE **87**, 537 (1999).
[7] S. Thompson, P. Packan, and M. Bohr, Intel Technology Journal which can be found at developer.intel.com/technology/itj/Q31998/articles.
[8] The national technology roadmap for semiconductor technology needs, 1997 edition, Semiconductor Industry Association.
[9] K. Natori, J. Appl. Phys. **76**, 4870 (1994).
[10] M. S. Lundstrom, IEEE Electron Device Lett. **18**, 361 (1997).
[11] Z. Ren and M. Lundstrom, Superlattices Microstruct. **27**, 177 (2000).
[12] Y. Naveh and K. K. Likharev, Superlattices Microstruct. **27**, 111 (2000); F. G. Pikus and K. K. Likharev, Appl. Phys. Lett. **71**, 3661 (1997).
[13] Z. Ren, R. Venugopal, S. Datta, M. Lundstrom, D. Jovanovic, and J. Fossum, International Electron Devices Meeting, 2000, IEDM Technical Digest, p. 715.
[14] B. A. Biegel, C. S. Rafferty, Z. Yu, M. G. Ancona, and R. W. Dutton, in Proceedings of the Gigascale Integration Technology Symposium, 35th Annual Technical Meeting, Society of Engineering Science (SES35), September 27–30, 1998, Pullman, WA, p. 53.
[15] M. G. Ancona and H. F. Tiersten, Phys. Rev. B **35**, 7959 (1987).
[16] D. K. Ferry, D. K. Ferry, R. Akis, and D. Vasileska, International Electron Devices Meeting, 2000, IEDM Technical Digest, p. 287.
[17] Medici User's Manual, Avant! Corporation.
[18] M. Buttiker, Phys. Rev. B **46**, 12 485 (1992).
[19] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1997).
[20] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. Appl. Phys. **81**, 7845 (1997).
[21] G. Klimeck, D. Blanks, and R. Lake, Nanoelectronic Modeling Tool (NEMO), User's Manual and Reference Guide, Raytheon, 1998.
[22] A. Abramo, International Conference on Simulation of Semiconductor Processes and Devices, 1997, SISPAD 97, p. 105.
[23] Y. Fu and M. Willander, Superlattices Microstruct. **82**, 5227 (1997).
[24] A. S. Spinelli, A. Benvenuti, and A. Pacelli, IEEE Trans. Electron Devices **45**, 1342 (1998).
[25] A. Abramo, A. Cardin, L. Selmi, and E. Sangiorgi, IEEE Trans. Electron Devices **47**, 1858 (2000).
[26] Z. Han, N. Goldsman, and C-K. Lin, International Conference on Simulation of Semiconductor Processes and Devices, 2000, SISPAD 2000, p. 62.
[27] D. Jovanovic and R. Venugopal, International Workshop on Computational Electronics, Book of Abstracts, IWCE Glasgow, 2000, p. 30; D. Jovanovic, R. Venugopal, S. Egley, M. Lundstrom, and S. Datta, Numerical simulation of MOSFETs using nanoequilibrium field theory, the third NASA Workshop on Device Modeling, NASA Ames Research Center, Moffett Field, CA, August 26–27, 1999, Book of Abstracts.
[28] A. Svizhenko, M. P. Anantram, and T. R. Govindan, 2D quantum simulation of MOSFET using the nonequilibrium Green's function method, International Workshop on Computational Electronics, Book of Abstracts, IWCE Glasgow, 2000, p. 112; A. Svizhenko, 2D Quantum Simulations in MOSFETs, Meeting Abstracts, Volume 99-2, Abstract No. 1866, 1999 Joint International Meeting of the Electrochemical Society, Honolulu, Hawaii, October 17–22, 1999.
[29] M. V. Fischetti and S. E. Laux, Phys. Rev. B **48**, 2244 (1993).
[30] U. Ravaioli, B. Winstead, C. Wordelman, and A. Kepkep, Superlattices Microstruct. **27**, 137 (2000).
[31] D. Vasileska, W. J. Gross, and D. K. Ferry, Superlattices Microstruct. **27**, 147 (2000).
[32] Proceedings of The Third NASA Workshop on Device Modeling, Superlattices Microstruct. **27**, (2000).
[33] G. D. Mahan, *Many Particle Physics*, 2nd ed. (Plenum, New York, 1990).
[34] G. D. Mahan, Phys. Rep. **145**, 251 (1987); A number of useful relationships involving the nonequilibrium Green's functions can be found in this paper.
[35] See 25 nm device in http://www-mtl.mit.edu:80/Well/.
[36] C. Bowen, private communication based on calculations using NEMO.
[37] A. S. Spinelli, A. Pacelli, and A. L. Lacaita, IEEE Trans. Electron Devices **47**, 2366 (2000).
[38] S. Mori and T. Ando, Phys. Rev. B **19**, 6433 (1979).
[39] D. Vasileska and D. K. Ferry, IEEE Trans. Electron Devices **44**, 577 (1997).
[40] A. Asenov, G. Slavcheva, A. R. Brown, R. Balasubramaniam, J. H. Davies, A. Asenov, and S. Saini, IEEE Trans. Electron Devices **47**, 805 (2000); A. Asenov, A. R. Brown, J. H. Davies, and S. Saini, IEEE Trans. Comput.-Aided Des. **18**, 1558 (1999).
[41] D. J. Frank and H.-S. P. Wong, Simulation of stochastic doping effects in Si MOSFETs International Workshop on Computational Electronics, 2000, Book of Abstracts, IWCE Glasgow 2000, 7th International Workshop, 2000, p. 2.
[42] A-P. Jauho, N. S. Wingreen, and Y. Meir, Phys. Rev. B **50**, 5528 (1994).
[43] M. P. Anantram and S. Datta, Phys. Rev. B **51**, 7632 (1995).

Downloaded 15 Feb 2002 to 129.99.66.97. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/japo/japcr.jsp

Erratum: "Two-dimensional quantum mechanical modeling of nanotransistors" [J. Appl. Phys. 91, 2343 (2002)]

M. P. Anantram and A. Svizhenko
NASA Ames Research Center, Moffett Field, California 94035

A. Martinez
EEE Department, Rankine Building, Oakfield Avenue, Glasgow, G12 8LT, United Kingdom

(Received 5 October 2006; published online 13 December 2006)

[DOI: 10.1063/1.2393002]

A systematic error in sign which corresponds to changing $A_{i,j}$ to $-A_{j,i}$ when $i \neq j$ appeared in Sec. II. The corrected equations are:

$$g^{rLq+1}_{q+1,q+1} = (A_{q+1,q+1} - A_{q+1,q} g^{rLq}_{q.q} A_{q,q+1})^{-1}, \tag{27}$$

$$G^{r}_{q,q} = g^{rLq}_{q,q} - g^{rLq}_{q,q} A_{q,q+1} G^{r}_{q+1,q}, \tag{29}$$

$$g^{<Lq+1}_{q+1,q+1} = g^{rLq+1}_{q+1,q+1}\left[\Sigma^{<}_{q+1,q+1} + \sigma^{<}_{q+1} - \Sigma^{<}_{q+1,q} g^{aLq}_{q,q} A^{\dagger}_{q,q+1} - A_{q+1,q} g^{rLq}_{q,q} \Sigma^{<}_{q,q+1}\right] g^{aLq+1}_{q+1,q+1}, \tag{39}$$

$$G^{<}_{q,q} = g^{<Lq}_{q,q} - g^{<Lq}_{q,q} A^{\dagger}_{q,q+1} G^{a}_{q+1,q} - g^{<0}_{q,q+1} A^{\dagger}_{q+1,q} G^{a}_{q,q} - g^{rLq}_{q,q} A_{q,q+1} G^{<}_{q+1,q}, \tag{40}$$

$$G^{<}_{q+1,q} = g^{<0}_{q+1,q} - G^{r}_{q+1,q} A_{q,q+1} g^{<0}_{q+1,q} - G^{r}_{q+1,q+1} A_{q+1,q} g^{<Lq}_{q,q} - G^{<}_{q+1,q+1} A^{\dagger}_{q+1,q} g^{aLq}_{q,q}, \tag{41}$$

$$G^{<}_{q,q} = g^{<Lq}_{q,q} + g^{rLq}_{q,q}(A_{q,q+1} G^{<}_{q+1,q+1} A^{\dagger}_{q+1,q}) g^{aLq}_{q,q} - [g^{<Lq}_{q,q} A^{\dagger}_{q,q+1} G^{a}_{q+1,q} + G^{r}_{q,q+1} A_{q+1,q} g^{<Lq}_{q,q}] - [g^{<0}_{q,q+1} A^{\dagger}_{q+1,q} G^{a}_{q,q} + G^{r}_{q,q} A_{q,q+1} g^{<0}_{q+1,q}], \tag{42}$$

$$J(q \to q+1, k_z, E) = \sum_b J_b(q \to q+1, k_z, E) = \frac{2e}{\hbar} \sum_b \mathrm{Tr}\left[T_{q,q+1} G^{<}_{b;q+1,q}(k_z, E) - T_{q+1,q} G^{<}_{b;q,q+1}(k_z, E)\right]. \tag{45}$$

We also correct some other errors:

(i) $A_{i,j}$ should be changed to $-A_{j,i}$ in the Appendix.
(ii) G^{r}_{b} in the first term of Eq. (6) should be replaced by $G^{<}_{b}$.
(iii) Below Eq. (17), $m^{\pm x} = (m_{i\pm1,j} + m_{i,j})/2$ and $m^{\pm y} = (m_{i,j\pm1} + m_{i,j})/2$.
(iv) The last term of Eq. (44) should be $g^{a0}_{q,q}$ instead of $g^{a0}_{g,g}$.
(v) $\Sigma^{<}_{L}$ should be replaced by $\Sigma^{<}_{11}$ in the equation immediately below Eq. (44).

Note that the numerical implementation included the correct sign of the terms and the results of the paper are unchanged.

 © 2006 American Institute of Physics

[This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to] IP: 128.95.215.214 On: Mon, 21 Jul 2014 17:51:47