RC 21956 (98701) 8 February 2001 **Electrical Engineering**

IBM Research Report

Physically Based, Mixed Mode Compact Model for the Double Gated FET

P. M. Solomon
IBM Research Division
Thomas J. Watson Research Center
P. O. Box 218
Yorktown Heights, NY 10598

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

Physically Based, Mixed Mode Compact Model for the Double Gated FET.

P. M. Solomon

IBM T.J. Watson Research Center, Yorktown Hgts, NY 10598.
solomonp@us.ibm.com

ABSTRACT

The double gated FET (DGFET), in which current may or may not flow close to either gate, exhibits complex behavior as a function of voltages applied to either gate and to the drain. This behavior is explored and analytical models derived to simulate drain current, capacitances and the short-channel effect under arbitrary bias conditions. A truncated hyperbolic velocity vs. field model has been used to model the linear region, and the space charge region has been modeled using a modified Suzuki scale length. The channel configuration in a DGFET may change, along the length of the FET, from dual to single due to the increasing channel potential, and this is simulated by modeling the transistor as two FETs in series giving rise to the name "*mixed-mode*" model. Short channel effects are simulated by introducing an equivalent short-channel charge which simulates a bias and electrode dependence short channel effect. The model has been implemented as FORTRAN subroutines and ASX simulations give realistic results with reasonable running times of 30μs/point on a 355MHz RS6000 computer.

A: Introduction

The silicon CMOS industry has followed 'Moore's law' down the scaling path to the point where the limits of conventional 'bulk' CMOS (transistors made on bulk silicon substrates) are only one or two generations in the future. The industry is actively pursuing alternatives to the bulk transistor, and a promising candidate is the double gate field effect transistor (DGFET) where a thin silicon body is gated from both sides, as shown in Fig. 1. The advantages of the DGFET have been pointed out by numerous authors in recent years [1-3], the principle one being the superior carrier confinement in such a structure which leads to better short-channel performance and thus the ability to scale the transistor to shorter channel lengths given the same restrictions on gate oxide thickness as in the bulk case. In addition to this key attribute, other important attributes are the steeper sub-threshold slope for long channel devices, since the sub-threshold gate potential is no longer divided between the channel and the substrate, and the ability of the transistor to be controlled by different voltages on the two gates [4]. The DGFET may be regarded as a four terminal device where the front and back gates can, in principle be applied separately. Such is the case also for the bulk FET where the substrate (or doped well in the more general case) acts as the fourth terminal. Control of substrate potential has long been a parameter of MOS design, and active control of the substrate has been touted as a means of dynamic threshold voltage control both to control standby power [5] and to improve performance [6].

Circuit models have long existed for FETs which take substrate bias into account [7-9] and are commonly used in the industry [10]. These models handle the case of a single FET channel, where the channel may be modulated by either gate. Modeling is also straight forward for the symmetrical double gate, where the structure is symmetrical about a plane passing between the two gates, and both gates are at the same voltage [11-14] and very sophisticated models exist which include effects of quantization and non-equilibrium transport [14].

Modeling is more complicated, and has yet to be attempted, for the general case of the asymmetric DGFET where arbitrary voltages may be applied to each of the gates. Such a model would be needed if the DGFET finds widespread application. Simpler existing models may be used to model the DGFET over a limited application range, yet the flexibility of the DGFET in allowing arbitrary voltages on all four of its terminals, with the possibility of different transport properties at either of its interfaces, and with arbitrary thickness ratios for top and bottom gate dielectric layers and semiconductor body thickness, makes a more general model most desirable. A realistic model for capacitance, including gate-to-gate capacitance is also needed for modeling loading and coupled circuit noise in arbitrary circuit configurations.

In this article we explore such a general model for the DGFET. The model is tied as closely as possible to the physics of the FET although drastic simplifications are made to yield tractable analytic equations. Where equations have to be solved numerically, close enough limits are derived to achieve convergence in just a few iterations. This model has its shortcomings as in its rather crude approach to modeling short-channel charge, and its lack of inclusion of quantized channel effects, yet it is shown to be very versatile and powerful and amenable to the inclusion of such effects in the future.

B: Approach and Simplifying Assumptions

Our approach is to convert the DGFET into an equivalent 'canonical' n-channel single gate FET (SGFET), to use a fairly conventional model (although with important new formulation of the space charge region) to derive currents and capacitances for the SGFET, and then re-apportion them to the terminals of the DGFET. Since channel configuration may change along the length of the FET, we at times model the DGFET as a DGFET in series with a SGFET giving rise to our *mixed mode* model. The emphasis is to derive a useful model for circuit simulation and, as a price of simplification, our model needs to be calibrated against hardware for accurate results. While the parameters of our model are closely tied to the physical geometry and transport properties of the DGFET, they may be adjusted to account for physics not included in the model such as non-equilibrium carrier transport.

a: DG FET configuration.

Consider first an FET having a single channel and two gates (front and back) on either side of a silicon body, as show in Fig. 1. The body is assumed to be undoped. The front and back gates have capacitances C_F and C_B to the channel respectively, which are assumed to be constant *i.e.* independent of gate voltage or position along the channel. Infinite density of states in the Si is assumed which results in an infinitely thin channel at the silicon-silicon dioxide interface and surface potential pinned to the conduction band. Work function differences are assumed to be zero, but may be allowed for by a simple shift in gate voltage. Allowance is made for finite inversion layer thickness[a1] by adjustment of effective thicknesses, and for an effective gate voltage to account for subthreshold conduction and gate capacitance. It is well known that for a very thin body region the picture of two separate channels is not correct but they merge into a single channel. At high concentrations and large perpendicular electric fields this happens only for an extremely thin body, as is illustrated by quantized channel simulations of Ieong *et al*.[15], shown in Fig. 2. Our separate channel approach is realistic for body thicknesses of greater than 10nm, but even for thinner channels workable results are obtained.

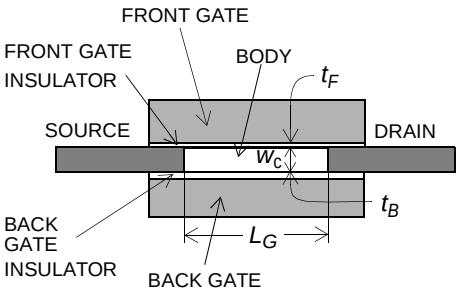

Fig. 1. Schematic cross-section of double-gate FET including definition of principal dimensions used in model.

1. Inversion layer is a misnomer here since the channel is undoped, a more rigorous term would be an electron or hole gas interface layer, yet we retain 'inversion layer' for clarity.

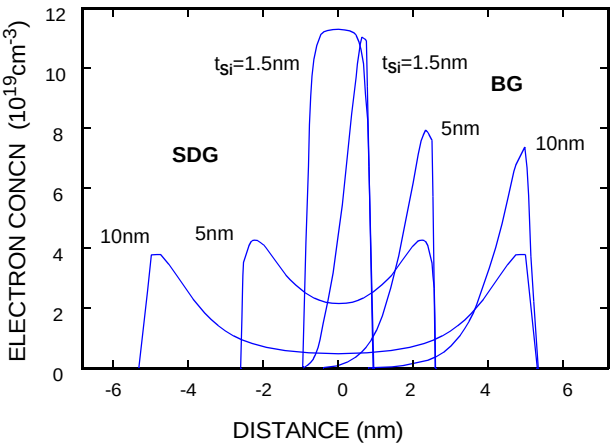

Fig. 2. Charge distribution across body region after Ieong *et al*.[15].

b: Equivalent Gate Voltage and Capacitance.

Given potentials V_F, V_B and $V(x)$ on front gate, back gate and channel respectively, the charge $Q_{CH}(x)$ induced in the channel by the gates is

$$Q_{CH}(x) = C_F[V_F - V(x)] + C_B[V_B - V(x)] \quad , \quad (1)$$

where x is the position along the channel. 1 may be expressed in terms of equivalent gate voltage and capacitance V_{GE}, and C_{GE},

$$Q_{CH}(x) = C_{GE}[V_{GE} - V(x)] \, , \quad (2)$$

where

$$C_{GE} = C_F + C_B \, , \quad (3)$$

and

$$V_{GE} = \frac{C_F V_F + C_B V_B}{C_{GE}} \, . \quad (4)$$

For the symmetrical DGFET, $V_{GE}=V_F=V_B$, and $C_{GE}=2C_F=2C_B$.

c: Mobility vs. Perpendicular Field.

The mobility vs. effective field (μ_E vs. F_E) model is based on the universal mobility curve and follows the formulation of Villa *et. al* [16] the model includes phonon, surface roughness and coulomb terms. (Recent work has shown that the universal curve can by used even for very thin channels [17]. The terms are shown in Fig. 3, where the dashed curves from Villa *et. al*

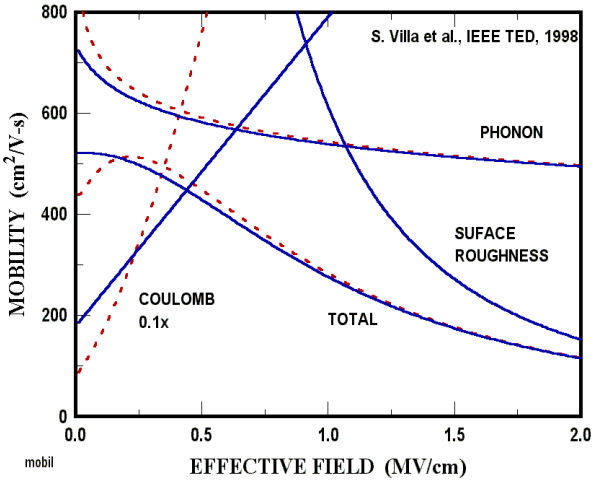

Fig. 3. Mobility model after Villa et al. [16].Dashed curves are from their model and solid curves are after modifications (see text for details).

use their default parameters for phonon and surface roughness terms and with values of oxide charge and interface states of 2×10^{17}cm^{-3} and 2×10^{11} cm^{-2} respectively. The effective perpendicular field is defined [18] as

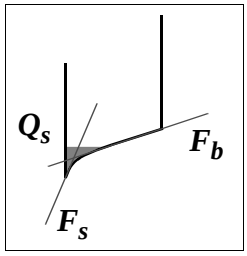

Fig. 4. Effective perpendicular field in a DGFET. In our model Q_S has zero width.

$$F_E = \frac{F_s}{2} + F_b \qquad (5)$$

where F_s is the surface (at the silicon-silicon dioxide interface) and F_b the bulk field in the silicon (behind the silicon channel). The solid curves show the mobility function used in our model where we have offset the field in the phonon term to eliminate the zero field singularity in the derivative and have added an artificial linear 'coulomb' term to achieve a zero field derivative of zero at zero field. Mobility is assumed to be the zero field mobility at negative effective fields. Additional constant mobility terms μ_F and μ_B allow for differing mobilities at the two interfaces. All terms are added using Mathiessen's rule with the μ_F and μ_B terms weighted according to the fraction of the charge at each interface:

$$\frac{1}{\mu} = \frac{1}{\mu_{SR}} + \frac{1}{\mu_{PH}} + \frac{1}{\mu_C} + \frac{1}{Q_F + Q_B}\left(\frac{Q_F}{\mu_F} + \frac{Q_B}{\mu_B}\right). \qquad (6)$$

d: Short-channel Effects.

In attempting to model threshold voltage shifts which vary in a consistent way with changes in the channel configuration within the DGFET we model them as due to an induced charge. The potential in the body of a DGFET, below threshold, has a saddle point due to the opposing tendencies of the gate vs. S/D voltages. Laplace's equation in two dimensions, $\partial^2\psi/\partial x^2 + \partial^2\psi/\partial y^2 = 0$, is replaced by Poisson's equation in one dimension by an induced charge density, $\rho_{sc} = \varepsilon_S\partial^2\psi/\partial x^2$, where ψ is the electrostatic potential. An empirical equation [19],

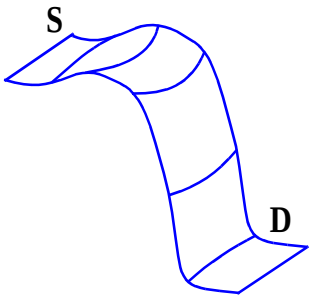

Fig. 5. Idealized potential in the body of a DGFET in the sub-threshold region showing the saddle-point created by the crossed potentials of the gates, raising, and the S/D, lowering the body potential.

$$Q_{sc} = Q_{sc0}\, e^{-\frac{L_G}{\lambda_{sc}}}\sqrt{1 + \frac{V_D}{V_{dibl}}}, \qquad (7)$$

relates the short-channel induced charge to the channel length and drain voltage using the scale length λ_{sc} of Eq.15. Here we do not tie the parameters Q_{sc0} and V_{dibl} to physical parameters of the FET here, although as its name implies, V_{dibl} is related to the drain induced barrier lowering effect (*dibl*).

The short-channel charge will be used below to generate gate offset voltages. Note that the short channel induced field subtracts from the interface field and, according to the mobility model, increases mobility at high gate voltages in short FETs.

C: Single Gate FET Model

The SGFET model has the following features: An hyperbolic velocity vs. transverse field relation with a truncated maximum, a space charge region at the end of the channel accounting for channel shortening, a vertical field dependent mobility, and voltage dependent gate/source and gate/drain capacitances. Threshold voltage is not a parameter of this model but the effective gate voltage is calculated ahead of the model which takes threshold voltage and threshold voltage shifts into account. Zero field mobility is also calculated ahead of the model using the effective field at the source end of the channel. This gives a conservative estimate of device performance in that it neglects reduction in field along the channel caused by the channel potential

a: Velocity vs. Lateral Field

The model velocity vs. field curve obeys the equation

$$v = \begin{cases} \dfrac{v_{sat}}{1+\dfrac{F_C}{F}} & F < F_C/k_{sat} - 1 \\ V_C/k_{sat} & F \geq F_C/k_{sat} - 1 \end{cases}, \qquad (8)$$

where v_{sat} is the saturation velocity, k_{sat} is the velocity truncation parameter ($k_{sat} > 1$), $F_C = v_{sat}/\mu_0$, with μ_0 the low field mobility, and $F_{CM} = F_C/(k_{sat}-1)$ the truncation field. The reason for choosing the hyperbolic velocity is that it gives simple expressions for the drain current in the linear regime, is a reasonable approximation to the velocity *vs.* field curve in silicon [8], gives self-consistent analytic expressions for the gate/source and gate/drain capacitances, and gives a smooth transition to saturation. The reason for truncating the velocity is to terminate the drain field at a finite value allowing for the formation of a drain space charge region. While the choice of k_{sat} is ad-hoc in this model, it should be related to point at which the curvature of the lateral vs. perpendicular electric fields become comparable [7,9]. k_{sat} approaches unity as the FET enters further into the velocity saturated regime and as the channel thickness decreases. Throughout this paper we use a value of 1.25 for k_{sat}.

b: Drain Current and Capacitance in Linear Region

The drain current J_D, per unit width, in the linear region is the standard derivation [9] based on the gradual channel approximation and is given by:

$$J_D = \frac{\mu_0 C_{GE}}{L_G} \cdot \frac{V_D(V_{GE} - V_D/2)}{1 + V_D/(F_C L_G)} \qquad (9)$$

where V_{GE} is the effective gate voltage, V_D the drain voltage and L_G the gate length. Solution of the same equations yield expressions for the capacitances:

$$\frac{C_{GD}'}{C_G} = \frac{(3-u_D)(2u_C - 2u_C u_D - u_D^3)}{3u_C(2-u_D)^2}, \qquad (10)$$

and

$$\frac{C_{GS}'}{C_G} = \frac{2u_C(3-3u_D+u_D^2)+u_D^3}{3u_C(2-u_D)} - (1-u_D)\frac{C_{GD}'}{C_G}, \qquad (11)$$

where $u_C = F_C L_G/V_{GE}$ and $u_D = V_D/V_{GE}$. The reason for the primed notation for C_{GD} and C_{GS} is that these will be further modified below to account for the sub-threshold region.

c: Drain Current in Saturation.

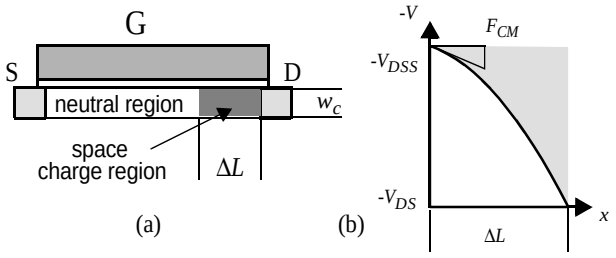

Fig. 6. (a) Schematic diagram showing position of space charge region and (b) the channel potential in the space charge region.

For the drain current, saturation occurs when the lateral electric field at the end of the channel exceeds F_{CM}. At the onset of saturation is $V_D = V_{DSS0}$. As the drain voltage increases above V_{DSS0} a space charge region of length ΔL develops at the end of the channel causing the effective channel length (L_G-ΔL), and hence the saturation voltage ($V_{DSS} \leq V_{DSS0}$), to decrease. In general

$$V_{DSS} = \frac{V_{GE}}{(2-k_{sat})}\left[-(k_{sat}-1+u_C) + \sqrt{(k_{sat}-1+u_C)^2 + 2u_C(2-k_{sat})}\right], \qquad (12)$$

where $u_C = F_C(L_G - \Delta L)/V_{GE}$, and V_{DSS0} is obtained by setting $\Delta L = 0$.

In the space charge region (SCR) the potential in the body is assumed to vary mainly in the x direction with the net space charge $\rho(x)$ given by:

$$\rho(x) = \frac{C_{sc}}{\alpha w_c}[V(x) - V_{GE}'] + \rho_{sc}, \qquad (13)$$

where C_{sc} is the unit capacitance of the space charge region, assumed to be the capacitance from the center of the channel to the gates, V_{GE}' the effective gate voltage (the prime denotes that the capacitance is now C_{sc}), α the fraction of the body thickness occupied by the space charge, and ρ_{sc} the space charge due to current flow, which is assumed to be constant throughout the SCR, since the carrier velocity is saturated at v_{max}. At the beginning of the SCR, the net space charge is zero, since this region is governed by the gradual channel approximation, and the space charge due to current flow is balanced by the charge induced by the gate. Since Eq.13 is linear in V, and $\rho(x)=0$ at $V(x) = V_{DSS}$, it may be simplified to:

$$\rho(x) = \frac{C_{sc}}{\alpha w_c}[V(x) - V_{DSS}]. \qquad (14)$$

Solving Poison's equation yields a scale length

$$\lambda_{sc} = \sqrt{\frac{\alpha w_c \varepsilon_S}{C_{sc}}}, \qquad (15)$$

with ε_S the permittivity of the semiconductor body (This scale length is the same as Suzuki's [20]. when $\alpha = 1$ and the effective thickness for calculating C_{sc} is $t_{ox} + w_c/4$, where t_{ox} is the oxide thickness for a symmetric structure, consistent with a parabolic potential profile in the body region.). The solution for the length of the space charge region is

$$\Delta L = \lambda_{sc} \ln\left(u + \sqrt{1+u^2}\right), \qquad (16)$$

where $u=(V_D-V_{DSS})/V_{sc}$, $V_{sc}=F_s\lambda_{sc}$ and F_s is the electric field at the start of the SCR ($F_s = F_{CM}$ under normal saturation conditions.). At the drain end of the SCR the field is:

$$F_D = \sqrt{(F_{CM}^2 + (V_D - V_{DSS})^2/\lambda_{sc}^2)} \ . \qquad (17)$$

For certain cases we need an explicit solution for the current, for instance under a punch through, or shorted condition, where $\Delta L = L_G$. In these cases a more general form of Eq.14 is used which retains the space charge term:

$$\rho(x) = \frac{C_{sc}}{\alpha w_c}[V(x) - V_{DSS}] + \frac{J_D - J_{DSS}}{v_{max}\alpha w_c}, \qquad (18)$$

where $J_{DSS} = v_{max} C_{GE} (V_{GE}-V_{DSS})$ is the saturation current prior to punch through. Solving Poisson's equation with Eq.18 gives:

$$\Delta L = \lambda_{sc} \ln\left(\frac{u - u_s + \sqrt{1 + u^2 - 2uu_s}}{1 - u_s}\right), \qquad (19)$$

where $u_s = (J_{DSS}-J_D)/v_{max}C_{sc}V_{sc}$, and solving this for J_D gives

$$J_D = \frac{v_{max}C_{sc}}{(\eta - 1)^2}[2\eta(V - V_{DSS}) + V_{sc}(1 - \eta^2)], \qquad (20)$$

where $\eta = \exp(\Delta L/\lambda_{sc})$.

The saturated drain current is given by setting $V_D = V_{DSS}$ in Eq.9. The implicit equations of saturation Eqs. 12 and 16, are solved iteratively (typically taking 3-5 iterations) using the secant method [21] with an upper and lower bounds for ΔL obtained by setting $V_{DSS}=0$ and $V_{DSS}=V_{DSS0}$ in Eq.16. For the shorted FET, $V_{DSS}=0$, and we assume $F_s=F_{CM}$ since in such an FET a self-consistent field would be set up at the source accelerating the carriers to the saturation velocity.

d: Capacitances in Saturation.

In saturation, contributions to the capacitance come both from the unsaturated (linear) region of the FET and from the space charge region. The total charge on the gate Q_G, can be apportioned between these regions as Q_{lin} and Q_{sc} so that the capacitances may, in principle, be obtained by solving the equation:

$$\delta Q_G = \frac{\partial Q_{lin}}{\partial V_i}\delta V_i - \frac{\partial Q_{lin}}{\partial L_{sc}}\delta L_{sc} + \frac{\partial Q_{sc}}{\partial V_i}\delta V_i + \frac{\partial Q_{sc}}{\partial L_{sc}}\delta L_{sc}, \qquad (21)$$

where V_i is shorthand for V_G or V_D, and $L_{sc} = \Delta L$. While solving the explicit voltage dependent terms (first and third in Eq.21) is straight forward, the $\partial Q_{lin}/\partial L_{sc}$ term, involving the simultaneous solution of the linear and space-charge regions, cannot be solved explicitly. Noting that charge lost to the gate from the linear region tends to be added back in the SCR, we assume that the two terms in δL cancel. The linear region term for C_{GS} is obtained by solving Eq.10 at $V_D = V_{DSS}$. This cannot be done for C_{GD} since one would be left with a large 'unsaturated' capacitance. Instead C_{GD} is allowed to fall to zero according to Eq.11 by using the untruncated velocity vs. field curve.

The space charge terms, as denoted by the subscript, may be derived from the space charge equations by noting that $Q_{sc} = C_{sc}\,\Delta L\,(V_G - V_{AV})$ where V_{AV} is the average body potential so that

$$C_{scG} + C_{scD} = C_{sc}\Delta L \qquad (22)$$

and

$$C_{scD} = \frac{\lambda_{sc}C_{sc}(V_D - V_{DSS})}{\sqrt{1 + (V_D - V_{DSS})^2}}. \qquad (23)$$

e: Sub Threshold Conduction and Capacitance

Sub-threshold conduction due to the Fermi statistics of the carriers in the channel may be approximated [19] by transforming the effective gate voltage V_{GE} into an effective *thermal* gate voltage V_{GTE} through the Fermi function integral i.e.

$$V_{GTE} = \frac{2kT}{q}\ln\left[1 + \exp\left(\frac{qV_{GE}}{2kT}\right)\right], \qquad (24)$$

where q is the electronic charge, k is Boltzmann's constant and T is the absolute temperature. The factor of 1/2 in the argument of the exponent is compensated for by the square law dependence of the saturated drain current ($V_{DSS} \rightarrow 0$ as $V_{GTE} \rightarrow 0$) giving the proper sub-threshold slope. When comparing the sub-threshold currents predicted by Eqs. 9 and 24 with the diffusion-limited sub-threshold current $(qN_C w_c \mu/L_G)\exp[qV_{GE}/kT]$, where N_C is the conduction band density of states, an extra voltage shift, $\Delta V_{ST} = (kT/q)\ln(2C_{GE}kT/q^2N_C w_c)$, is needed to bring the two into correspondence. This

amounts to ~0.1V negative threshold voltage shift for typical values of the parameters. Note though that ΔV_{ST} does introduce an artificial dependence of sub-threshold current on C_{GE}. The fermi transform of the gate voltage (replacing V_{GE} by $V_{GTE)}$ stretches the capacitance into the sub-threshold region as well. When doing this the capacitances have to be multiplied by the factor [1-exp(qV_{GE} / $2kT$)] in order to conserve charge.

f: Sample solutions:

Solutions for the FET model under front gate operation (V_B<0) are shown in Fig. 7. The slope of the output characteristic, beyond saturation, is partly contributed by the *dibl* effect (Eq. 7) and partly by channel length modulation (Eqs 16 &20). Simulated characteristics of the 25nm FET are shown in Fig. 8. This shows transition between linear, saturated and shorted regions as indicated in the figure by the length (L_G-ΔL) of the neutral region. Note that by using physically derived equations for the space charge region we ensure smooth and continuous transitions.

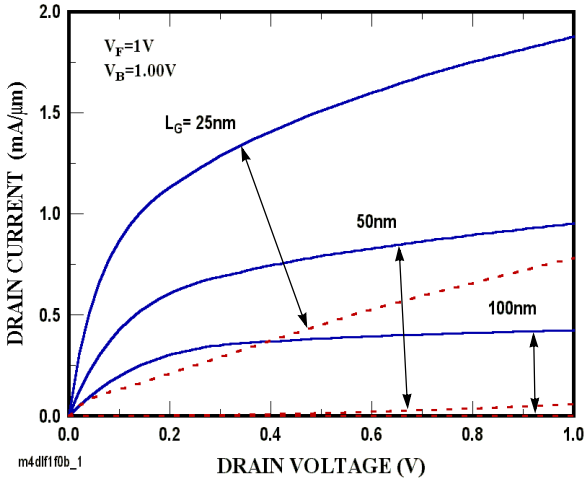

Fig. 7. Drain characteristics at channel lengths of 25, 50 and 100nm. The 25nm FET operates under punch-through conditions.

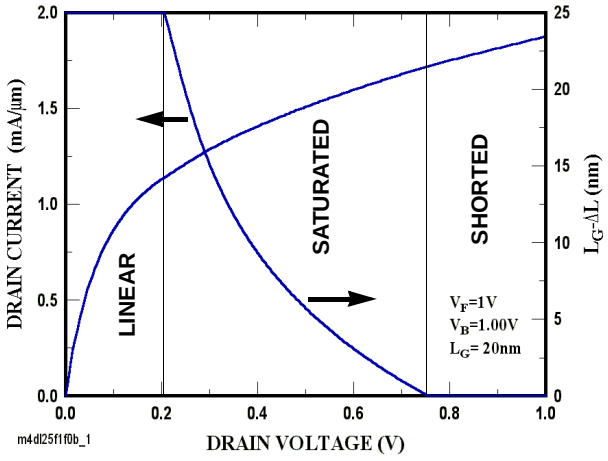

Fig. 8. Drain current (left), and effective channel length of the neutral region (right) of a 25nm FET showing transitions between linear, saturated and shorted regions.

D: Conversion of DGFET to Canonical form.

To minimize the number of special cases, the general DGFET, with arbitrary potentials, oxide thicknesses and dielectric constants, gate work functions and carrier type (*p* or *n*) is converted into a 'canonical' DGFET with *n* type conduction, positive drain voltage, zero contact potential differences and a front gate voltage larger than the back gate voltage. This is done by subtracting out contact potential differences, accounting for short-channel charge, and transposing terminals. Dielectric thicknesses are expressed as equivalent silicon thicknesses i.e. $t_F' = (\varepsilon_I/\varepsilon_S)\, t_F$ and $t_B' = (\varepsilon_I/\varepsilon_S)\, t_B$, where ε_I and ε_S are the permittivities of the gate dielectric and body semiconductor respectively.

The canonical DGFET is then converted into a canonical SGFET where the gate voltages and capacitances are converted into an effective gate voltage and capacitance (see Eqs. 3 & 4), where, in the case of a single channel, $C_F = \varepsilon_S/(t_F' + t_{inv})$ and $C_B = \varepsilon_S/(t_B' + w_c - t_{inv})$, and t_{inv} is the inversion layer thickness. Note that short-channel effects according to Eq.7 are introduced prior to the SGFET model but channel shortening, which also uses λ_{sc}, is part of the SGFET model.

E: DGFET - Mixed Mode Model

In the case of a DGFET having a single channel, or in the case of a symmetric double gate, the conversion to canonical SGFET form is straightforward, as explained above. A more complicated situation arises in an FET with two unequal channels, as illustrated in Fig. 9b. In this case the potential developed in the channel, which increases toward the drain, causes the back channel (the weaker channel according to our canonical arrangement) to pinch off before the front channel at a channel potential $V_C(x) = V_B$. While two channels are present the potential separating them is small being only the self-confinement potential (assumed to be zero under our assumption of infinite density of states). As soon as the back channel pinch-off condition is exceeded, an electric field is generated between the two channels diverting the current from the back into the front channel. This situation, depicted in Fig. 9b, shows that the DGFET under these conditions can be decomposed into two FETs in *series*, with the FET near the source, FET-1 of length L_1, having a dual channel and that nearer the drain, FET-2 of length L_2, a single channel, such as is shown in Fig. 10. There is a temptation to represent the dual channel DG FET as two FETs in parallel, because of the two parallel channels, but the strong electrostatic coupling between the two

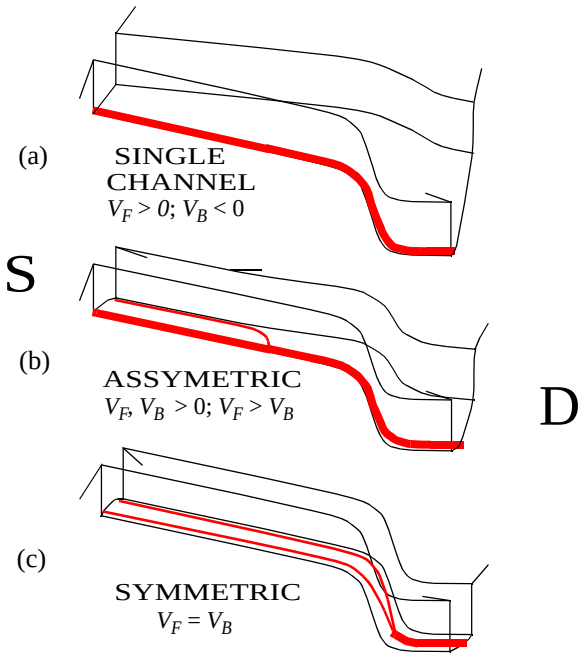

Fig. 9. Visualizations of potential along the channel of a DGFET showing the current paths a) under large negative back gate and large positive front gate bias, b) under large and small positive front and back gate biases respectively and c) under equal positive front and back gate biases.

channels, especially in the case of the thin undoped body, precludes this.

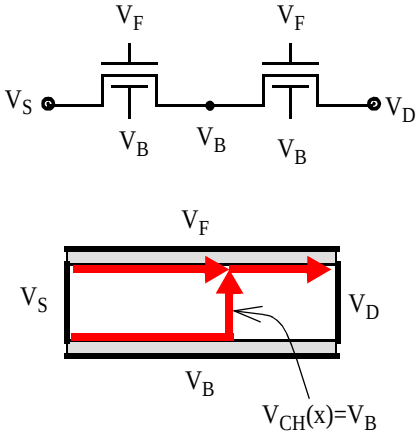

Fig. 10. *Mixed mode* decomposition a of DGFET with dual channels near the source into two separate FETs with dual channels near the source and a single channel near the drain.

The *Mixed Mode* algorithm represents the DGFET as two FETs in series with the Drain/Source intermediate node being at potential V_B. The condition therefore, for mixed mode operation, is:

$$V_F > V_B > 0 \qquad (25)$$

and

$$V_B < V_D . \qquad (26)$$

These conditions are sketched in Fig. 12 which shows the V_B / V_D half-plane. Note that the canonical condition reduces the parameter space since $V_F \geq V_B$ by definition. Under mixed mode conditions channel lengths, L_1 and $L_2=L_G - L_1$, of the two series FETs are adjusted so as to equalize the drain currents. The regions of operation for the DGFET may also be shown in the V_F /V_B plane.

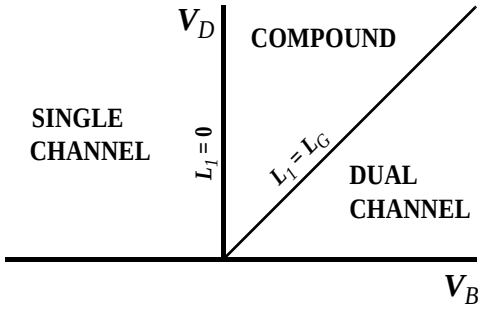

Fig. 11. Partition of the canonical half-space according to mode of operation of the DGFET.

The need for the compound algorithm is shown in Fig. 13,

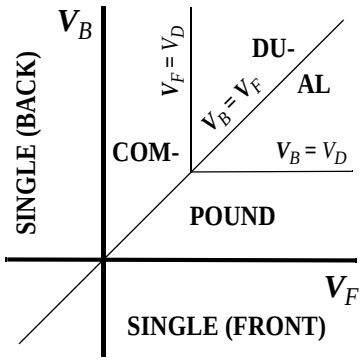

Fig. 12. Partition of the VF /VB plane according to region of operation of the DGFET.

where an FET is switched from a back-channel only mode to a dual channel mode by increasing the front-gate voltage and curves are shown with and without using the compound mode. Discontinuities in the drain current result from the fact that the transformation from dual to single channel does not explicitly

involve drain voltage so that currents cannot be matched at arbitrary drain voltages.

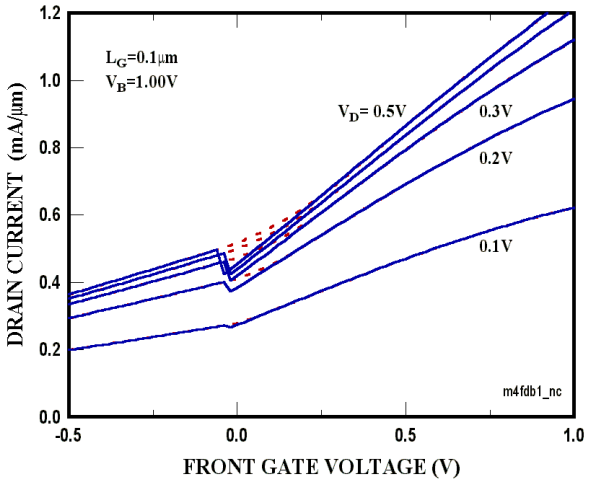

Fig. 13. Simulation of DGFET, without (solid) and with (dashed) the compound algorithm.

a: Treatment of the FET segments.

To ensure a smooth transition it is important to have an FET model based on an integral solution of the transport equations since the two FET sections are equivalent to two ranges of integration. This is the case for our hyperbolic *v-F* model below saturation, and is the reason we insisted on an integral solution (see Eqs.14-20) above saturation as well. When both FETs are in the linear regime equalizing the currents is relatively simple, and no discontinuities are encountered as one proceeds continuously from the case of a single channel FET (V_B=0, L_1 =0) to a dual channel FET ($V_B = V_{DD}$, L_2 = 0). In the case of saturation the situation is considerably more complicated and is illustrated in Fig. 14. At low V_B ($V_B \leq 0$) there is only a single channel (FET-2). As V_B is increased ($0 < V_B \leq V_{DSS}$) the length of FET-1 increases until it reaches the edge of the SCR corresponding to punch through of FET-2. For FETs having equal mobilities, this also corresponds to the onset of saturation of FET-1 ($V_B = V_{DSS}$) if $F_s = F_{CM}$, since the SCR parameters of both FETs are identical. For $V_{DSS} < V_B \leq V_D$ the boundary between FET-1 and FET-2 is within the SCR and FET-2 is punched through. Above $V_B = V_{DSS}$ (Fig. 14f) the FET is in the dual channel mode.

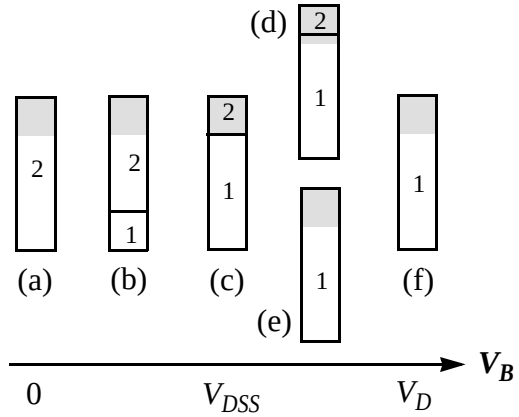

Fig. 14. Progression of FET partition for FET in saturation at positive V_F as V_B is increased. Cross-hatching indicates space charge region. Numbers refer to FET-1, (dual channel), and FET-2, (single channel).

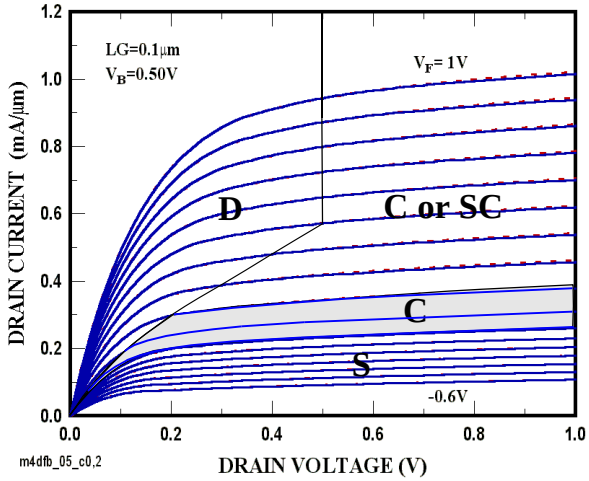

Fig. 15. Drain characteristics of DGFET showing operation in single, compound and dual modes. Steps in front gate voltage are 0.1V. Dashed curves correspond to simplified compound (SC) operation. Single, dual, compound and SC regions are indicated S, C, SC and D. The shaded area is the C region remaining after SC is applied.

To ensure that FET-1 and FET-2 act as a single unit, some parameters have to be passed between them. Firstly, in the subthreshold transformation (Eq. 24) the gate voltage is referred to the source voltage of FET-1, and this voltage has to be passed to FET-2 so that they both use the same transformation. Secondly, when FET-1 is in saturation, the source field, F_S, of FET-2 is set equal to the drain field of FET-1 as given by Eq.19.

$$F_S = \sqrt{(F_{CM}^2 + (V_{D1} - V_{DSS1})^2 / \lambda_{sc}^2)} \qquad (27)$$

where V_{D1}=V_B, and V_{DSS1} pertain to FET-1. The secant method is used here as well to solve FET-1 and FET-2 simultaneously. The lower bound on L_1 is $L_G V_B / V_D$ which assumes that FET-2 has a lower conductivity than FET-1 (since it has a higher gate voltage. This might not work for special cases such as a very asymmetric mobility.). For the upper bound at L_1, which is the lower bound for L_2, it is assumed that FET-2 is in the punch-

through condition and driven by the current of FET-1 with a full drain voltage and gate length. Solutions for a DGFET are shown in Fig. 15. This FET operates in single (S), compound (C) and dual (D) modes, as indicated in the figure, but note that the canonical condition interchanges V_F with V_B internally when $V_B > V_F$. Transitions between the various modes show no visible discontinuities in Fig. 15, and qualitatively correct behavior is seen in the change over from single (back) channel to compound or dual channel operation at V_F=0. Fig. 16 shows

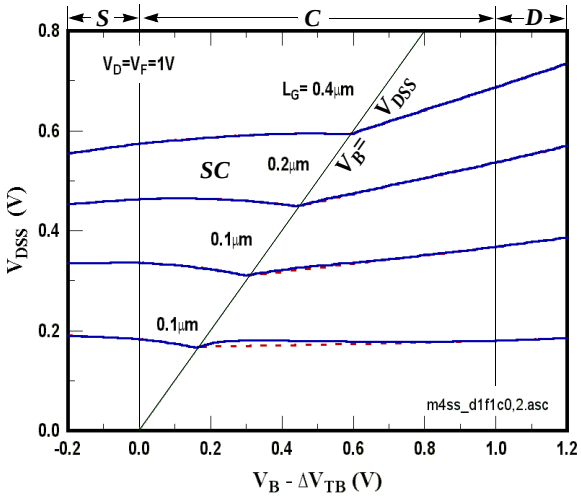

Fig. 16. Saturation voltages for the DGFET showing the transition from single channel (***S***) through compound (**C**) to dual channel (**D**) modes. The dashed lines show the **SC** model which applies for $V_{DSS} \leq V_B < V_D$.

the behavior of V_{DSS} through the various transitions showing a smooth transition from single through compound and into dual channel mode. V_{DSS} increases at large V_B because of the reduction in mobility with increasing interface field, but at shorter channel lengths this is countered by the internal field due to the short-channel effect.

b: Simplified Compound (SC) mode:

The compound mode is costly to simulate because several iterations (~5-8) are required for convergence. Referring to Fig. 14e, one can bypass part of the compound region by jumping straight to the FET-1 solution when $V_B \geq V_{DSS}$. This is possible for our SCFET model since the differential equations for the SCR do not depend on the mode. Dashed curves in Fig. 15 show that results using the SC model differ little from the full model and that the parameter space for the C mode is greatly reduced. A further analysis is shown in Fig. 17, where the boundary between FET-1 and FET-2 is shown, as well as the boundary of the space charge region. Note how the total width of the SCR region, (L_G-L_1+ΔL_2) is maintained close to the SCR width of FET-1 of the SC model confirming the self-consistency of our equations for the SCR region and showing that the SC model is a useful simplification.

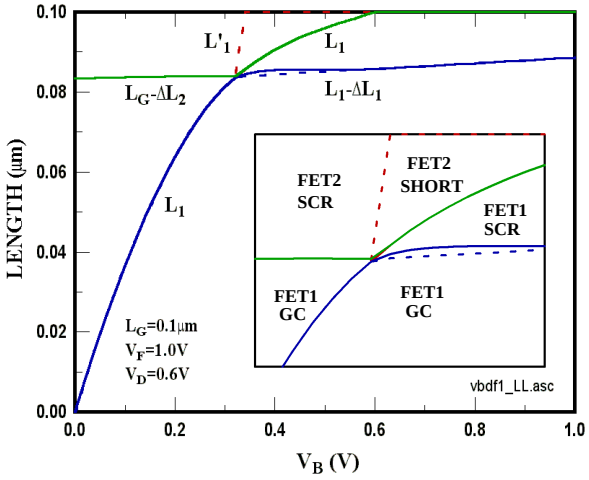

Fig. 17. Lengths of neutral and space charge regions for FET-1 and FET-2. The dashed lines refer to the SC model. The labels refer to the space charge region (SCR) and gradual channel region (GC) .

c: Double Gate Operation

An important operational mode of the DGFET when both gates are tied together, and there has been some controversy concerning the merits symmetrical vs. asymmetrical operation [22] i.e. if the channels are of equal strength or predominantly front or back (for instance if the gates have differing work functions). Such cases are shown in Fig. 18, where the gates are offset from each other by a fixed voltage and the average gate voltage increased. Transconductance (vertical spacing between curves) is ~25% higher at zero offset voltage (symmetrical operation) mainly because of increased mobility. Transconductance is reduced in the single channel mode because of reduction of charge control of the back gate but also because of reduction of mobility at low gate voltages caused by the large back field. Of course these conclusions are influenced by the choice of body and insulator thicknesses and mobility assumptions, all of which can be explored further using this model.

d: Sub-Threshold operation.

Referring to Fig. 12, the sub-threshold quadrant is at the bottom left. According to our transformation (Eq. 24 in Eq.9), the saturated sub threshold current,

$$J_{DsT} = \frac{\mu_0 C_{GE}}{2L_G}\left(\frac{kT}{q}\right)^2 e^{qV_{GE}/kT}, \qquad (28)$$

is proportional to the effective gate capacitance, which depends on the position of the channel (front or back). This dependence is an artifact of our model, but even so we must ensure a continuous current when proceeding from front to

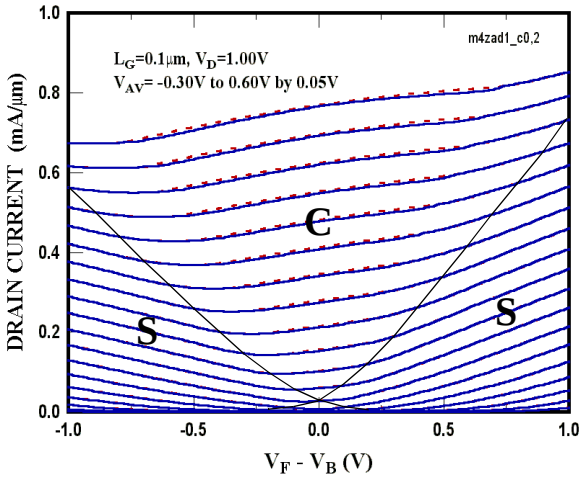

Fig. 18. Drain current of a DGFET with gates offset from each other by a constant potential, and at a given average potential with respect to the source. S and C denote single channel and compound channel regions with the demarcation lines being at V_F=0 (left) and V_B=0 (right). Dashed curves use the SC mode.

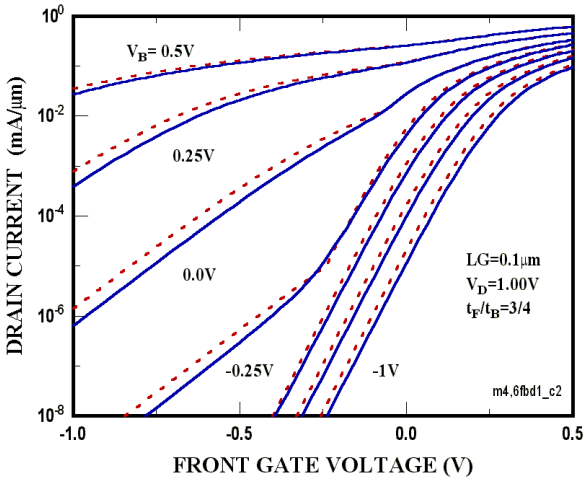

Fig. 20. Drain current vs. gate voltage, including the sub-threshold region, for a moderately asymmetrical FET (tF/tB=3nm/4nm). Solid curves include the lorentzian smoothing function with V_{inv}=0.052V.

back-channel sub-threshold operation. To do this an artificial

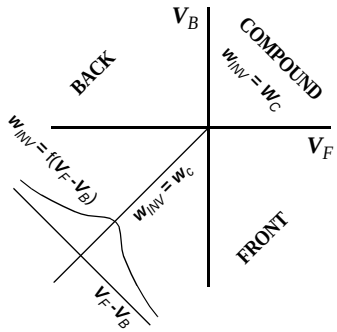

Fig. 19. Smoothing of transition between front and back channel operation.

'inversion layer thickness', w_{inv} (to be distinguished from t_{inv} above), is introduced such that $C_B = \varepsilon_{Si}/(t_b'+w_c- w_{inv})$ and C_F is unchanged. A function f (V_F-V_B) is introduced such that f(0)=w_c and f $(V_F-V_B) \rightarrow V_{inv}/F_{inv}$, where $F_{inv} = (V_F-V_B)/(t_F'+t_B'+w_c)$, as $(V_F-V_B) \rightarrow \infty$, which is applied to the entire region of single channel operation (V_F<0 and V_B<0). This function which is meant to approximate the thermal inversion layer thickness, kT/qF_{inv}, ensures equal back and front and capacitances at V_F=V_B and in the positive V_F /V_B quadrant. The capacitances are still discontinuous across the positive V_F and V_B axes, but this transition is handled by the compound mode algorithm. The lorentzian is used for the smoothing function although the precise form is not critical. Simulations of the drain current in the sub-threshold region is shown in Figs 20 and 21 for moderately and highly asymmetric cases respectively. While for the moderately asymmetric case the smoothing function is hardly needed, it clearly helps to smooth out

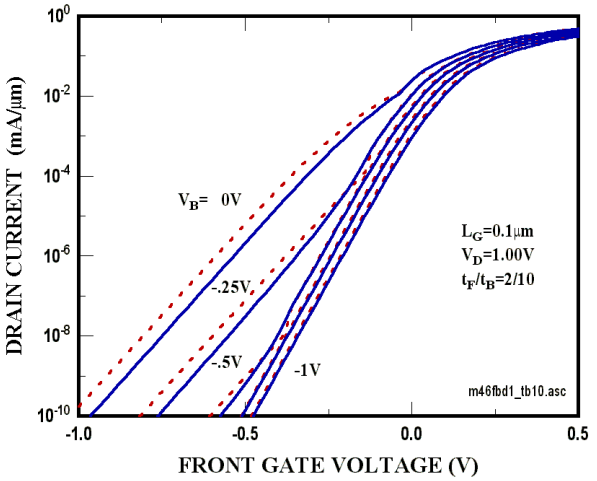

Fig. 21. Drain current vs. gate voltage, including the sub-threshold region, for a highly asymmetrical FET (t_F/t_B =2nm/10nm). Solid curves use the lorentzian smoothing function with V_{inv}=0.052V.

steps for the highly asymmetric case. This smoothing function does not introduce any adverse features into the drain characteristics above threshold, as shown in Fig. 22.

e: Capacitance Model.

For the single and dual channel modes the capacitances are evaluated according to Eqs. 10 and 11, but using the appropriate effective thicknesses depending on the channel position, and including the t_{inv} and w_c-w_{inv} terms, and on the particular gate (front or back). They are multiplied by the effective channel length, L_G - ΔL. For the space charge capacitance Eqs.22

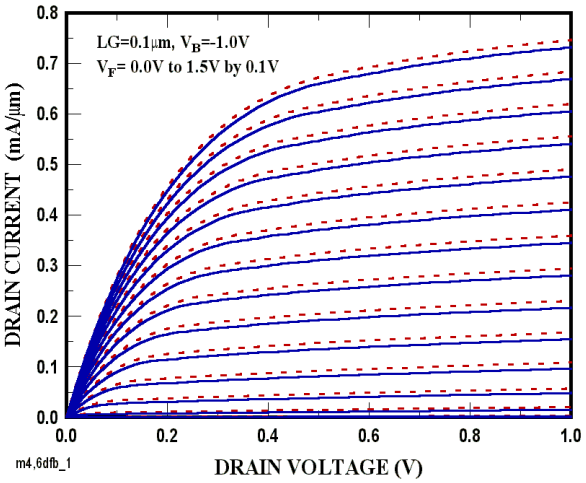

Fig. 22. Drain characteristics of a DGFET with a large back gate voltage (V_B=-1V). Solid curves are with and dashed curves without the sub-threshold smoothing function.

and 23 are used with the front of back effective thickness including the $w_c/4$ adder.

For the compound mode the capacitances of the individual

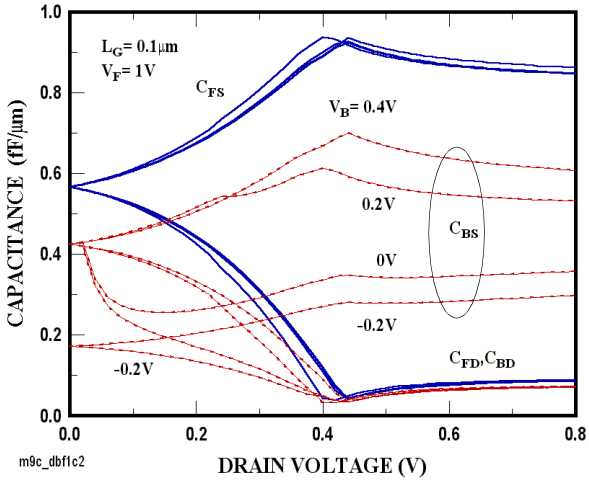

Fig. 23. Front (solid) and back gate (dashed) capacitances to source and drain (C_{FS}, C_{FD}, C_{BS}, C_{BD}) as a function of drain voltage for different back gate voltages.

FETs are not used directly, since the strongly varying channel length partition between the FETs strongly influences how the capacitances are seen from the outside. Rather the FET is re-synthesized, only for the purpose of capacitance determination, with lengths given according to the partition. Effective gate voltage and capacitance are then determined and the single FET equations are then used. Equivalent gate lengths and

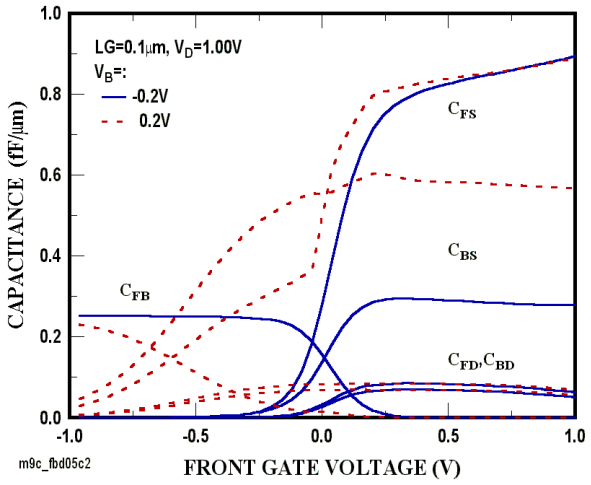

Fig. 24. Front gate capacitances to source, drain and to the back gate (C_{FS}, C_{FD}, C_{FB}) as a function of front gate voltage for back gate voltages of 0.2V (solid) and -0.2V (dotted).

thicknesses are summarized in Table 1. To conserve charge,

TABLE 1. Lengths and Thicknesses for Capacitance Model.

MODE	Region Length	Front Gate t_{eff}	Back Gate t_{eff}
Single	L_G-ΔL	$\varepsilon_r t_F$[a] $+t_{inv}$	$\varepsilon_r t_B$ $+w_c$-w_{inv}
	ΔL	$\varepsilon_r t_F$ $+t_{inv}$	$\varepsilon_r t_B$ $+w_c/4$
Dual	L_G-ΔL	$\varepsilon_r t_F$ $+t_{inv}$	$\varepsilon_r t_F$ $+t_{inv}$
	ΔL	$\varepsilon_r t_F$ $+w_c/4$	$\varepsilon_r t_B$ $+w_c/4$
Com-pound[b]	L_1	$\varepsilon_r t_F$ $+t_{inv}$	$\varepsilon_r t_B$ $+t_{inv}$
	L_G-L_1	$\varepsilon_r t_F$ $+t_{inv}$	$\varepsilon_r t$ $+w_c$-w_{inv}

a. $\varepsilon_r = \varepsilon_{Si}/\varepsilon_{ox}$

b. For the compound mode the capacitors are first apportioned by L_1:L_G-L_1 and then by L_G:L_G-ΔL.

the capacitances are multiplied by the inverse derivative of the fermi function used for the gate voltage transform as described in section E.

For the double gate FET an extra capacitance coupling front to back gate, C_{FB}, is present when the channel is absent. In the model

$$C_{FB} = \frac{(\varepsilon_{Si} L_G)[1 - \exp(qV_{ge}/2kT)]}{\varepsilon_r (t_F + t_B) + w_c}. \qquad (29)$$

Once the capacitance are determined they are then assigned to the correct external FET terminals according to the inverse transform used to obtain the canonical DGFET.

Capacitances of a 0.1mm DGFET are shown in Figs. 23 and 24. The dip in the gate to drain capacitances at the onset of saturation is an artifact of the model due to failure of the model

of the space charge region when it is very short [7]. Note the fall-off in back gate capacitance, at V_B=0, as V_D is increased. This is because of the increasing drain potential cutting off the back channel. Note also that in the sub-threshold regime the capacitance is dominated by the inter-gate capacitance C_{FB}.

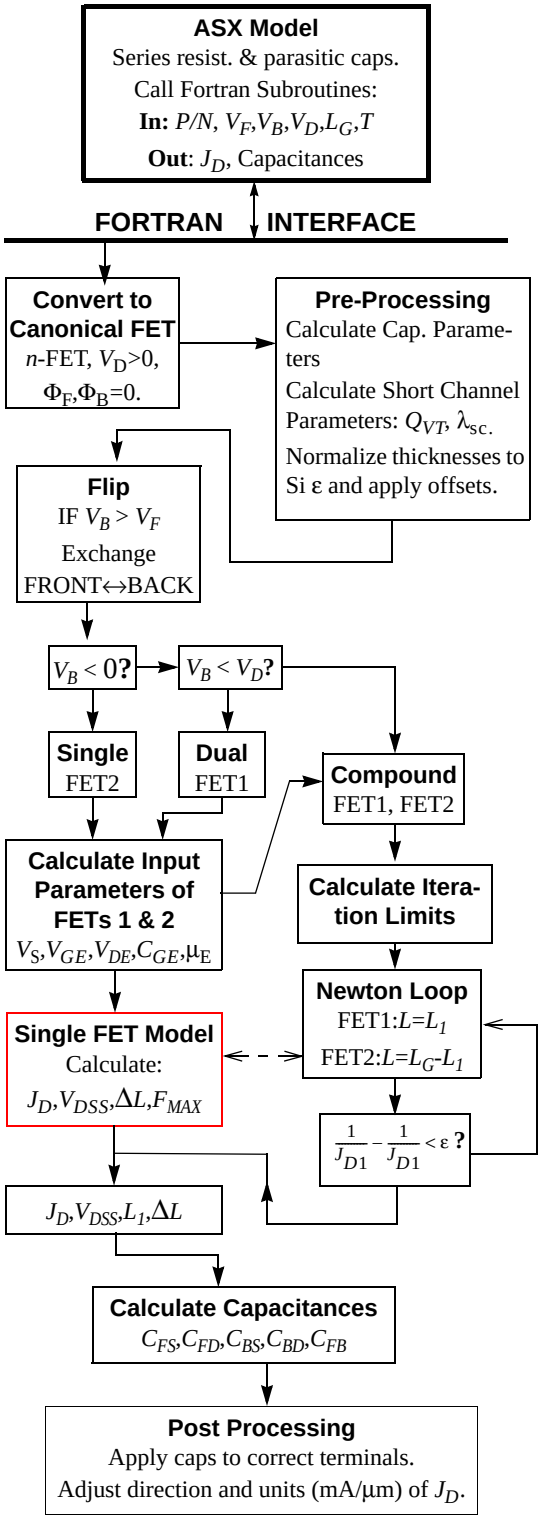

Fig. 25. Program flow for mixed mode DGFET model.

f: Program Flow.

The overall program flow is shown in the block diagram of Fig. 25. The actual device model is a 'shell' implemented in ASX ™ (a SPICE ™ like circuit simulator.) which includes external parasitic resistances and capacitances, and a call to a FORTRAN function which passes forward the FET *type* (*P* or *N*), the four terminal *voltages*, and the *temperature*, and passes back values for the drain current, J_D, and the capacitances C_{FS},C_{FD},C_{BS},C_{BD} and C_{FB}. The sub-program is compatible with the FORTRAN 77 language and consists of less than one thousand lines of code. The running time is reasonably fast ~30μs per point (averaged over an FET characteristic) on a 333MHz RS6000 processor, which enables it to be used for conventional circuit simulation.

F: Application.

a: Comparison with 2-D numerical simulations.

Our mixed-mode DGFET model has many free parameters enabling fits to almost any conceivable data given a suitable choice. In comparing it to other models or to experiment one should ask if the fitting parameters are reasonable given the physical basis for the model, and if trends in the data are reproduced in the model. It is beyond the scope of this introductory paper to do a thorough comparison, and well characterized experimental results are very sparse, so at this time we suffice to compare our model with two other works a) the DGFET design space study of Wong et al. [23], and b) a quantum-confined channel DGFET of Ieong et al. [15], the latter is a bit beyond the range of the assumptions of our model but nevertheless it is interesting to compare.

The Wong paper dealt primarily with the sub-threshold region, and here we compare our model with their data in terms of V_T shifts for the DGFET operated in symmetric mode (t_B=t_F and V_B=V_F) and ground-plane mode (t_B=1.5t_F and V_B=0). The ground plane FET has an appropriate work function or voltage applied to produce a threshold voltage for the front gate of ~0.25V. Comparison with Wong's data is shown in Fig. 26. The scale length was adjusted to 1.7λ_{sc}, rather than 2λ_{sc}, to obtain the fit for t_F=t_B=1.5nm symmetric case, and kept the same for the other cases. Our model quantitatively reproduces the trend for the 1.5nm GPFET GP case and qualitatively for the other cases. The discrepancy is largest for large w_c/t_F ratios where the approximations used in deriving the Suzuki scale length are no longer valid, nevertheless, the fit is good considering the simple assumptions used.

A fit to Ieong's data is shown in Fig. 27. This data is a 25nm gate length DGFET operated in symmetric and asymmetric modes. Thicknesses were t_F=t_B=1.5nm, and w_c=5nm. Barrier heights were adjusted to match off-currents at zero bias and were 0.29eV for the symmetric case and 0eV for the front

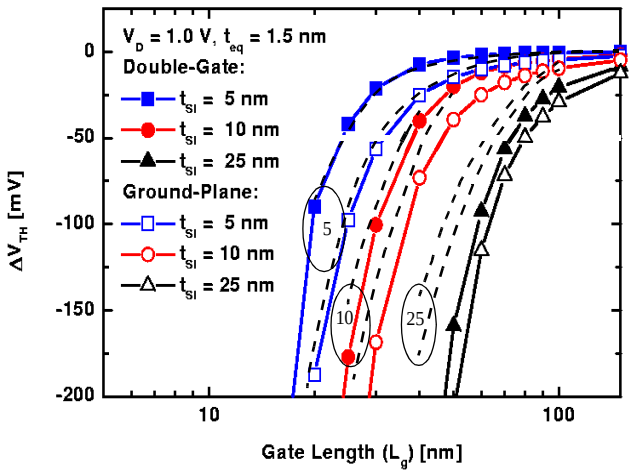

Fig. 26. Threshold voltage shift vs., gate length comparing the mixed mode model (dashed) with the data of Wong *et al.*[23]. The order of the curves is the same in both cases.

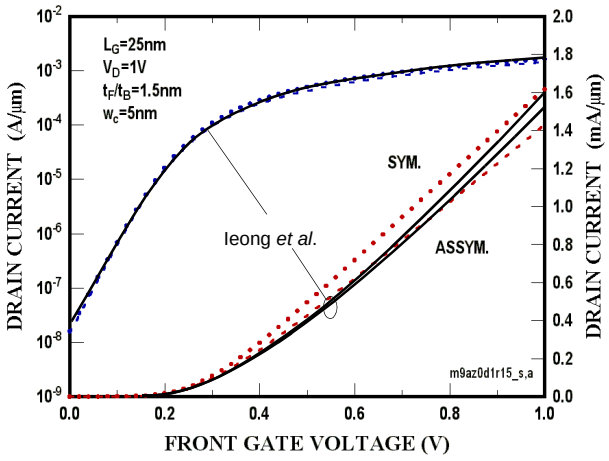

Fig. 27. Comparison of mixed mode model with 2-D numerical simulations by Ieong *et al* [15] (solid lines). Dotted lines are for a symmetric case with ϕ_F=ϕ_B =0.29eV, and dashed lined are for an asymmetric case with ϕ_F=0 and ϕ_B=0.7eV.

gate and 0.7eV for the back gate in the asymmetric case. All other parameters were nominal. At gate and drain voltages of 1V, the device current was 2.5mA/μm. This is a not unexpected current for an FET of this aggressive a design, yet it is higher by 56% than the current in Ieong's simulations. The current at V_G=1V can be made to match by adding a rather high series resistance of 150Ω-μm in series both source and drain, but such a high resistance is not indicated by data in Ieong's paper. Even when matching the current, the turn-on is rather more rounded in Ieong's case. Both the rounding and the low currents may be reproduced in our model by lowering the channel mobility to ~100cm^2/V-s. On the other hand, quantum effects may account for some of these differences because a) a lower density of states in a single sub-band channel will force the Fermi Energy to increase and b) the sub-band energies may themselves increase with bias. These shifts are expected both to lower the on-current and increase the curvature and might be incorporated into our model by using a bias dependent t_{inv}.

While our model is based on simplified physics, it yet has enough freedom to include more sophisticated effects such as quantum confinement and velocity overshoot, and mobility degradation mechanisms.

b: Delay Chain Simulations.

Delay chains consisting of eleven CMOS inverters loaded with 0.5fF of inter-stage wiring capacitance were simulated using ASX circuit simulator. Device widths were 1μm for the *n*-FETs and 2μm for the *p*-FETs. Factors of 0.5 and 0.8 respectively were used in calculating mobilities and saturation velocities for the *p*-FETs, and parasitic capacitances of 0.2fF/μm were added across all terminals. Typical running times for a simulation which included ~200 time steps was ~1sec of CPU time. This is similar to the CPU time used to run our bulk CMOS models. Results for several different invertor designs are shown in Table 2. Both 0.025μm (extreme scaling) and 0.1μm designs are considered and several cases are dealt with i.e. symmetric double gate with both gates tied together (S), asymmetric double gate with strongest channel closest to front gate (AF) or back gate (AB) or ground plane design with back gate (GB) or front gate (GF) tied to ground. While this is not intended to be a design space study, merely an illustration, it does show that symmetric designs are generally the fastest, and ground plane designs the slowest. A fairly small asymmetry in the back to front gate thickness ratio (1.33) markedly decreases or decreases the delay penalty when a front or back channels respectively are used. The poor results for the ground

TABLE 2. Invertor Designs and stage Delays

	Case[a]	ϕ_F	ϕ_B	I_N	$\frac{CV}{I}$	τ_{INV}
(nm)		(eV)	(eV)	(mA)	(ps)	(ps)
L_G =25	**S**	.29	.29	1.62	.57	3.5
t_F, t_B =1.5	**AF**	0	.70	1.43	.64	3.9
w_c =5	**GB**	0	.70	0.52	1.5	8.1
L_G =100	**S**	.24	.24	.82	1.9	10.0
t_F =3	**AF**	0	.91	.63	2.2	11.9
t_B=4	**AB**	.69	0	.63	2.5	12.2
w_c =5	**GB**	0	.91	.35	3.5	17.0
	GF	.69	0	.15	8.4	37.9

a. Asymmetric; AF,AB=Asymetric, Front & Back channels; GF,GB=Ground Plane, Front or Back.

plane designs result from reduced on-current since off-current levels are maintained, but if the ground plane voltage is varied so as to reduce the threshold voltage in the active mode, performances comparable to the asymmetric double gate can be realized. Wave forms for voltage, current and capacitance for an internal stage of the delay chain are shown in Fig. 28. Capacitances are dominated by the *p*-FETs since they are wider. The rapid variations in the capacitance are not reflected in the output voltage because of the integrating nature of capacitors. No untoward glitches are seen, showing that transitions within the model are smooth.

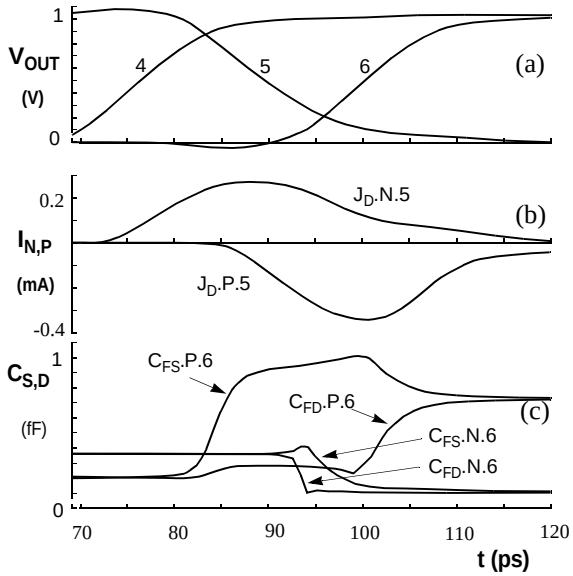

Fig. 28. Simulations of an 11 stage delay chain showing a) output voltages of stages 4,5 and 6, b) pull-down and pull-up currents of stages 5 and 6, and c) transistor capacitances at the input of stage 6.

c: Application to Mobility modulation FET.

Over the years proposals have been introduced for FETs which operate on the principle of mobility rather than charge modulation as in the case of the Mott transition [24], or for transfer from a high to low mobility quantum well [25]. A small change in gate voltage brings about an abrupt change in mobility, hence high (essentially infinite) transconductance is obtained and speeds are limited only by the intrinsic speed of the phase transition.

Since the transition is induced electrostatically, it happens at a given voltage difference between the gate and channel so, as the gate voltage is increased it starts at the source and then spreads along the channel. This effect can be modeled using our mixed mode model.

The mobility transition is simulated by having a much lower mobility (~10x) in the front vs. the back channels and with the back channel turned on (V_B=1V) to ensure there are always a large number of carriers present. The front and back-side mobility terms, μ_F and μ_B, are set to 50 and 3000 cm^2/V-s respectively The effective mobility for the SGFET is calculated in two alternative ways, firstly the mobility terms are added abruptly, simulating an abrupt phase transition, (using Mathiesson's rule) when the field at the appropriate interface is positive (V_F or $V_B > 0$), secondly the mobility is weighted by the interface charge according to Eq.6, simulating more realistic charge transfer to a low mobility interface. Simulations are shown in Figs. 29 and 30. Our model can handle this situation without undue conver-

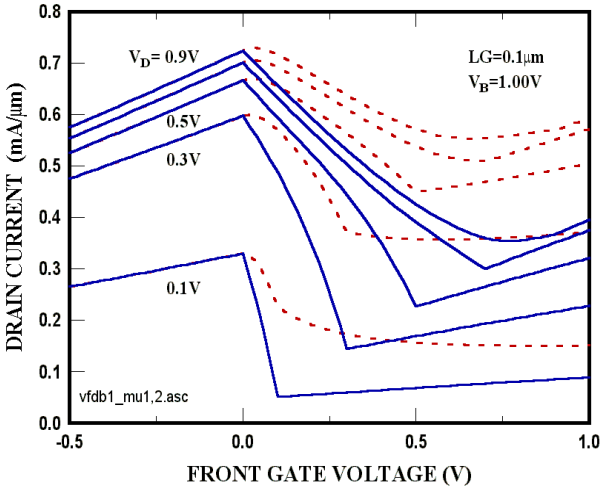

Fig. 29. Asymmetric mobility DGFET with μ_F=50 and μ_B=3000 cm2/V-s. Abrupt and charge weighted mobility models are given by given by the solid curves and dashed curves respectively.

gence problems, and IV curves are continuous through the transition. The transition is steep, but not infinitely so, the width being the drain voltage, and this limits the maximum transconductance to $Q_C(\mu_2-\mu_1)/L$, where μ_1 and μ_2 are the effective mobilities on either side of the transition and Q_C the critical charge in the channel. The switching speed of such a transistor does have an electrostatic component, as is seen from the finite transconductance, but it may be faster than the conventional transit time. The drain characteristics (Fig. 30) show how the onset of compound mode operation at $V_D > V_F$ causes the current to increase since FET-1 supplies the necessary voltage offset to push FET-2 into a high mobility regime while the length of FET-1 decreases rapidly with drain voltage to match the increasing current in FET-2.

This example illustrates the power of our mixed mode model and while characteristics such as in Fig. 30 have yet to be seen, the model could be useful to test such concepts.

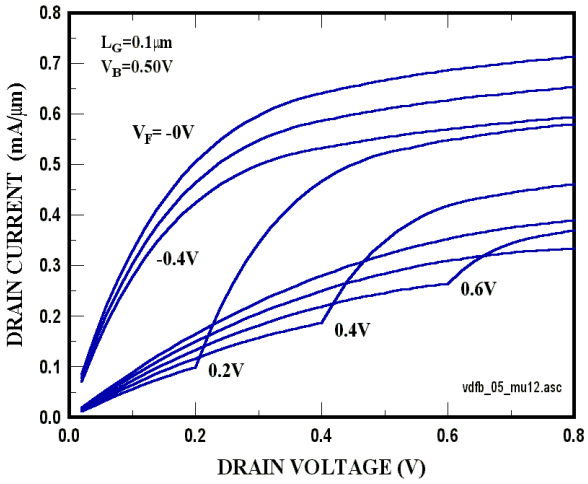

Fig. 30. Asymmetric mobility DGFET with m_F=50 and m_B=3000 cm2/V-s. Abrupt and charge weighted mobility models are given by given by the solid curves and dashed curves respectively.

G: Summary

This work presents a general and comprehensive model for the DGFET which is easy to use and where the input parameters are closely related to the physical structure of the DGFET. It handles both above threshold and sub-threshold operation, front, back and dual channel operation where mobilities of the two channels may be different, and compound operation where the channel configuration changes along the length of the device as a function of bias. New concepts introduced in this work are treatment of the compound mode as two series FETs with bias dependent gate lengths, formulation of the space charge region to allow for punched-through operation in the compound mode, and derivation of bias dependent capacitances.

The model is flexible enough to investigate novel device concepts such as the mobility modulation FET, and to test new DGFET geometries and concepts, yet simple enough to simulate complex circuits with speed and efficiency. Its ability to model details of the DGFET bias dependent behavior should facilitate accurate transient simulations including cross-coupling and circuit noise.

Some aspects of the DGFET are not addressed very well. A constant ‘inversion layer’ thickness is assumed and this should be replaced with a bias dependent value, and include quantum effects. A true self-consistent treatment of sub-threshold operation should be introduced, which should include buried channel operation caused by the short-channel charge or an n-doped channel. This model does not treat DGFETs with a thick, p-doped body. These should perhaps be modeled as two single gated FETs in parallel.

Notwithstanding these shortcomings, our expectation is that our DGFET model will prove to be of great value as DGFETs become more widely used.

Acknowledgments:

The author wishes to acknowledge the support of DARPA grant *N66001-97-1-8908* for this work, and the support of H-S. Wong, as well as valuable discussions with S. Laux and contributions from D.J. Frank.