Threshold Voltage Modeling and the Subthreshold Regime of Operation of Short-Channel MOSFET's

Tor A. Fjeldly, *Senior Member, IEEE*, and Michael Shur, *Fellow, IEEE*

Abstract—**We present a new analytical model for the subthreshold regime of operation of short-channel MOSFET's and develop expressions for the threshold-voltage shift associated with the Drain-Induced Barrier Lowering (DIBL) caused by the application of a drain bias. The model is based on the concept of charge sharing, where the depletion charge under the gate is identified with counter charges on the gate electrode and in the source and drain contacts. In particular, we estimate the amount of drain-bias-induced depletion charge in the channel and develop an expression for the distribution of this charge along the channel based on a simplified two-dimensional solution of Poisson's equation. From this distribution, it is possible to find the lowering of the potential barrier between the source and the channel, and the corresponding threshold-voltage shift. The results are compared with experimental data for deep-submicrometer NMOS devices. Expressions for the subthreshold current and for a generalized Unified Charge Control Model (UCCM) for short-channel MOSFET's are presented. We expect that our theory is applicable to deep-submicrometer devices (with gate length larger than 0.1 μm). The model is suitable for implementation in circuit simulators.**

I. Introduction

When the gate electrode of a MOSFET is biased below the threshold voltage, such that the conducting channel is weakly inverted, the MOSFET is said to be in the subthreshold regime. This regime is very important since it is one of the two stationary operating states of MOSFET's in digital applications. The *on*-state of the MOSFET, which corresponds to a gate bias above threshold, allows a significant drain current to pass through the device, while the *off*-state, which corresponds to a subthreshold gate bias, should ideally block all drain current. In practice, there will always be some leakage current in the *off*-state owing to a finite amount of mobile charge at the semiconductor–insulator interface and a finite injection rate of carriers from the source into the channel. In the subthreshold regime in short-channel devices, a drain voltage induces lowering of the energy barrier between the source and the channel. This so-called Drain-Induced Barrier Lowering (DIBL) [1] causes excess injection of charge carriers into the channel and gives rise to an increased subthreshold current. Since this current has deleterious effects on the performance of digital circuits in terms of increased power dissipation and a possible shift in logical levels, it is important to develop reliable models for the subthreshold regime of the MOSFET.

Our objective is to develop a new analytical model for the threshold voltage and for subthreshold currents which can adequately describe submicrometer devices and, at the same time, is simple enough to be implemented in circuit simulators.

In the following, the electrical properties of MOSFET's in the subthreshold regime will be discussed. The standard theory for subthreshold current of MOSFET's, discussed in Section II, works well for long-channel devices and can, with minor adjustments, even account for some-channel effects, as discussed for example, in [2], [3]. However, in order to explain important new features that occur in short-channel devices with gate lengths of the order of a micrometer or less, it is necessary to consider additional effects such as those described by the charge sharing concept [4], as discussed in Section III. (For a discussion of the short-channel concept for MOSFET'S, see Sze [2].) In Section IV, the charge-sharing model is further developed to include DIBL effects and the concomitant drain-bias-induced threshold voltage shift. In Sections V and VI, the results are used to obtain generalized expressions for the subthreshold current and for the Unified Charge Control Model (UCCM) where important short-channel effects are included. For simplicity, we restrict the discussion to n-channel devices, but the results apply equally well to p-channel transistors.

The new important features of our model include an analytical treatment of the DIBL effects, threshold-voltage variation with length, and subthreshold current in deep-submicrometer devices. These features make it possible to implement this model in circuit simulators.

Manuscript received June 2, 1992; revised July 24, 1992. This work was supported by the Royal Norwegian Council for Scientific and Industrial Research, the NATO Scientific Affairs Division, Virginia Center for Innovative Technology, and the Office of Naval Research. The review of this paper was arranged by Associate Editor Y. Nishi.

T. A. Fjeldly is with the Department of Physical Electronics, Norwegian Institute of Technology, N-7043, Trondheim, Norway, and the Department of Electrical Engineering, University of Virginia, Charlottesville, VA 22903-2442.

M. Shur is with the Department of Electrical Engineering, University of Virginia, Charlottesville, VA 22903-2442.

IEEE Log Number 9204701.

II. Long-Channel Theory for Subthreshold Current

A MOSFET in the subthreshold regime will normally carry only a small drain current, even at high drain–source bias. In fact, the source and drain contacts, together with the weakly inverted or depleted channel, have much in common with a Bipolar Junction Transistor (BJT) where the channel acts as a long base, the n^+ source acts as the emitter, and the drain plays the role of the collector. Most of the applied "collector–emitter" voltage will drop across the reverse-biased junction which, in the MOSFET, corresponds to the drain–channel junction. Accordingly, away from the source and drain contacts, the central portions of the channel itself will experience very little lateral potential variation. Instead, the potential distribution in the channel will be governed almost entirely by the gate electrode. This means that the electric-field component perpendicular to the semiconductor–insulator interface is much larger than the longitudinal component, allowing the channel to be analyzed within the framework of the Gradual Channel Approximation (GCA). As a consequence of the small potential drop along the channel, the drain current is dominated by diffusion instead of drift for drain–source voltages above a few thermal voltages.

A standard theory for subthreshold current in MOSFET's can be derived by considering the channel current density as the sum of a drift term and a diffusion term. For an n-channel device, the current density J_n can be written as

$$J_n = q\left(-n\mu_n \frac{d\psi}{dx} + D_n \frac{dn}{dx}\right) = qD_n\left(-\frac{n}{V_{\text{th}}}\frac{d\psi}{dx} + \frac{dn}{dx}\right) \tag{1}$$

where x is the position in the channel measured from the source, n is the density of electrons in the channel, μ_n is the low-field electron mobility, D_n is the electron diffusion coefficient, and $\psi(x, y)$ is the potential of the channel region referred to the potential of the source electrode. In (1), it is assumed that the longitudinal electric field in the channel, $F_x = -d\psi/dx$, is sufficiently small (except for the junction region near drain) such that velocity saturation can be neglected. Likewise, the diffusion coefficient can be expressed in terms of the Einstein relation as $D_n = \mu_n V_{\text{th}}$, where $V_{\text{th}} = k_B T/q$ is the thermal voltage.

Fig. 1 shows qualitatively the band diagram and the potential distribution at the interface in the channel, $\psi_s(x) \equiv \psi(x, 0)$, for two values of drain–source bias, $V_{DS} = 0$ and $V_{DS} > 0$. At the interface, the channel consists of three regions, the source–channel junction with length x_s, the drain–channel junction with length x_d, and the middle region of length $L - x_s - x_d$. At $V_{DS} = 0$ V, the interface potential in the central part of the channel ψ_s^0 can be taken to be approximately constant.

A drain–source bias gives rise to a positive contribution, $V(x)$, to the channel potential, as indicated in Fig. 1. As a consequence, the minimum in the interface potential, $\psi_s(x_{\text{min}}) = \psi_s^0 + V(x_{\text{min}})$, will be localized at the source side of the channel, at $x = x_{\text{min}} \approx x_s$. Associated with the shift in the potential minimum, we have a reduction in the interface energy barrier between the source and the channel by $-qV(x_{\text{min}})$. This is the so-called Drain-Induced Barrier Lowering (DIBL).

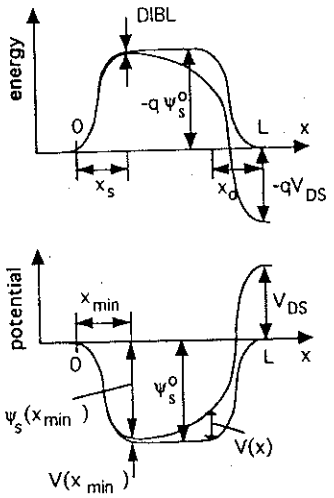

Fig. 1. Band diagram (top) and potential profile (bottom) at the semiconductor–insulator interface of an n-channel MOSFET. The symmetrical profiles correspond to $V_{DS} = 0$, and the asymmetrical profiles to $V_{DS} > 0$. The figure indicates the origin of the Drain-Induced Barrier Lowering (DIBL). The various symbols are defined in the text.

Multiplying (1) by the integrating factor $\exp(-\psi/V_{\text{th}})$, the right-hand side of this equation can be made into an exact derivative. A subsequent integration from source to drain yields (assuming that the current density remain independent of x)

$$J_n = qD_n \frac{n(L)\exp\left(-\frac{\psi(L, y)}{V_{\text{th}}}\right) - n(0)\exp\left(-\frac{\psi(0, y)}{V_{\text{th}}}\right)}{\int_0^L \exp\left(-\frac{\psi(x, y)}{V_{\text{th}}}\right) dx} \tag{2}$$

where $n(L) = n(0)$ equals the source and drain contact doping density N_c (neglecting degeneracy). With the source contact as a potential reference, we have $\psi(0, y) = 0$ at the source and $\psi(L, y) = V_{DS}$ at the drain, where V_{DS} is the intrinsic drain–source voltage.

When the device length is not too small, the channel potential can be taken to be independent of x over a portion of the channel length. We write the potential in this region as $\psi(x, y) = \psi^0(y)$. The integral in the denominator of (2) is dominated by the contribution from this portion of the channel. As can be seen from Fig. 1, the length of this section is approximately equal to $L - x_s - x_d$, and the current density can be expressed as

$$J_n = -\frac{qD_n N_c}{L - x_s - x_d}\exp\left(\frac{\psi^0(y)}{V_{\text{th}}}\right)\left[1 - \exp\left(\frac{V_{DS}}{V_{\text{th}}}\right)\right]. \tag{3}$$

For long-channel devices, where $L >> x_s - x_d$, we can neglect x_s and x_d in this expression.

The drain current is obtained by integrating the current density over the cross section of the conducting channel, yielding

$$I_d \approx \frac{qW\delta D_n N_c}{L} \exp\left(\frac{\psi_s^0}{V_{\text{th}}}\right)\left[1 - \exp\left(\frac{V_{DS}}{V_{\text{th}}}\right)\right] \quad (4)$$

where W is the width of the gate, δ is the effective channel thickness, and $\psi_s = \psi_s^0$ is the constant potential at the semiconductor–insulator interface. We again emphasize that ψ_s^0 is defined relative to the source electrode. Hence, although the interface potential relative to the interior of the p-type substrate, $V_s = \psi_s^0 + V_{bi}$, is positive, ψ_s^0 will be negative for n-channel MOSFET's. Here, V_{bi} is the built-in potential between the source contact and the substrate. Note that (4) could also be obtained directly by making the analogy between the MOSFET in the subthreshold regime and a long-base BJT, as discussed above (see [2], [3]).

At threshold, the interface potential in the channel relative to the source can be expressed as

$$\psi_{sT}^0 = V_{sT} - V_{bi} = 2\varphi_b - V_{bi} \quad (5)$$

where V_{sT} is the potential relative to the interior of the substrate at threshold and φ_b is the position of the Fermi potential relative to midgap in the neutral region of the doped substrate (see Section III). For simplicity, we have assumed that the substrate is shorted to source. (The effects of a substrate–source bias V_{BS} are found simply by replacing V_{bi} by $V_{bi} - V_{BS}$. Of course, such a replacement is only valid for negative or small positive values of V_{BS}. A positive V_{BS} comparable to V_{bi} would lead to a large substrate leakage current.)

Below threshold, the interface potential can be written as

$$\psi_s^0 = \psi_{sT}^0 + \frac{C_i}{C_i + C_{\text{dep}}} V_{GT} = -V_{bi} + 2\varphi_b + V_{GT}/\eta \quad (6)$$

where $V_{GT} = V_{GS} - V_T$, V_T is the threshold voltage, V_{GS} is the intrinsic gate–source voltage, C_i is the insulator capacitance, C_{dep} is the gate depletion capacitance of the semiconductor, and η is the subthreshold ideality factor. The ideality factor reflects the gate voltage division between the insulator-layer capacitance and the depletion-layer capacitance. Generally speaking, η is dependent on V_{GT} [2], but is usually treated as a fitting parameter which is easily obtained from subthreshold I_d versus V_{GS} characteristics [5]. Equations (4) and (6) show that the subthreshold drain current decreases nearly exponentially with decreasing V_{GT}. Ideally, for $V_{DS} > 3V_{\text{th}}$, this current should be practically independent of the drain–source voltage.

Since the mobile charge density in the channel has an exponential dependence on the interface potential ψ_s^0, the effective channel thickness δ can be estimated as the distance from the interface where the potential has changed by V_{th}. For this estimate, we need to find the vertical component of the electric field at the interface. According to Gauss' law, this field is equal to $|q_{\text{dep}}/\epsilon_s|$ in the subthreshold regime, where $q_{\text{dep}} = -[2\sigma_s qN_a(\psi_s^0 + V_{bi})]^{1/2}$ is the sheet depletion charge density under the gate, N_a is the acceptor doping density in the substrate, and ϵ_s is the dielectric permittivity of the semiconductor. Hence, the effective channel thickness becomes

$$\delta \approx V_{\text{th}} \sqrt{\frac{\epsilon_s}{2qN_a(2\varphi_b + V_{GT}/\eta)}}. \quad (7)$$

We note that this expression is only valid when $-2\varphi_b < V_{GT}/\eta < 0$, i.e., in the weak inversion and the depletion regime. Still, this condition is fulfilled for values of the drain current that are many orders of magnitude smaller than the threshold current.

III. Short-Channel Effects and Charge Sharing

At short gate lengths, the depletion widths associated with the source–channel and the drain–channel junctions may represent a significant fraction of the total gate length. In this case, the integral in the denominator of (2) should also reflect the variation in the channel potential in the junction depletion zones. The GCA, however, breaks down in these regions owing to the two-dimensionality of the field pattern. But, since the potential increases rapidly and the integrand becomes very small in the junction depletion zones, the contribution to the integral from these regions will be negligible. The net effect is, as discussed above, that the gate length L in (4) should be replaced by

$$L \to L - x_s - x_d. \quad (8)$$

The widths x_s and x_d can be expressed approximately in terms of conventional one-dimensional depletion widths for abrupt p-n junctions [2]

$$x_s \approx \sqrt{\frac{2\epsilon_s}{qN_a}(-\psi_s^0)} \quad (9)$$

$$x_d \approx \sqrt{\frac{2\epsilon_s}{qN_a}(V_{DS} - \psi_s^0)}. \quad (10)$$

Making this replacement in (4) and using (6) and (7) for the interface potential and the effective channel thickness, respectively, leads to the following expression for the subthreshold drain current:

$$I_d = \frac{WD_n V_{\text{th}}}{L - x_s - x_d}\sqrt{\frac{\epsilon_s qN_a}{2(2\varphi_b + V_{GT}/\eta)}} \times \exp\left(\frac{V_{GS} - V_T}{\eta V_{\text{th}}}\right)\left[1 - \exp\left(-\frac{V_{DS}}{V_{\text{th}}}\right)\right]. \quad (11)$$

Here, we have also used $V_{bi} = V_{\text{th}} \ln (N_c N_a/n_i^2)$ (disregarding degeneracy in the contacts) and $\varphi_b = V_{\text{th}}$

$\ln (N_a/n_i)$, where n_i is the intrinsic carrier density. We note that this expression for the drain current has gained additional dependence on the drain–source bias through the dependence on V_{DS} of the drain–channel junction depletion width x_d. This produces a finite output conductance in the subthreshold regime at all drain biases.

In the above analysis, it was assumed that the entire drain–source voltage drops across the drain–channel junction, and that the interface potential everywhere else in the channel is independent of V_{DS}. However, this assumption breaks down for short-channel devices. Experiments and detailed numerical calculations (see, for example, Fichtner and Potzl [6]) show that MOSFET's with such short gate lengths will experience additional short-channel phenomena such as DIBL, and a corresponding threshold-voltage shift. Since the current in the subthreshold regime is exponentially dependent on the height of the source–channel barrier, it is clear that even a small lowering of this barrier will have dramatic effects on the charge transport properties of the device.

An explanation for the shift in the threshold voltage of short-channel devices was proposed in terms of the so-called charge sharing (or charge conservation) concept, first introduced by Yau [4]. The model considers the shared depletion charge in the regions where the gate depletion zone overlaps with the depletion zones of the source and drain contracts, as indicated in Fig. 2. The shared charge is balanced by counter charges distributed between the gate electrode and the source and drain contacts. In the long-channel case, the threshold voltage can be written as [2], [3]

$$V_T = V_{FB} + 2\varphi_b + \frac{Q_{\text{tot}}}{C_i} \tag{12}$$

where V_{FB} is the flatband voltage and $Q_{\text{tot}} = LWq_{\text{dep}} = LW(4\epsilon_s qN_a\varphi_b)^{1/2}$ is the total gate depletion sheet charge at threshold, neglecting charge sharing (and assuming zero drain–source and substrate–source bias). The charge sharing causes a reduction in Q_{tot} by a fraction of the shared charge (shaded regions in Fig. 2). According to (12), this leads to an averaged shift in the threshold voltage which can be estimated as

$$\Delta V_T = -\frac{2\kappa W(x_s + x_d)\sqrt{\epsilon_s qN_a\varphi_b}}{C_i} \tag{13}$$

where κ is a constant (on the order of 0.5) which accounts for the charge sharing. In (13), it was assumed that the source and drain depletion widths in the charge-sharing region are x_s and x_d, respectively, and that the gate depletion width at threshold is $d_{\text{dep}} = (4\epsilon_s\varphi_b/qN_a)^{1/2}$.

We emphasized that the above derivation of ΔV_T is equivalent to distributing the portion κ of the shared charge evenly under the gate to reduce uniformly the depletion charge density in the gate area. This may lead to a reasonable estimate of the threshold-voltage shift at zero drain–source bias where the charge and potential distribution in the channel is quite symmetric. However, with an applied drain–source voltage, the charge distribution becomes asymmetric and (13) becomes less reliable. Typically, at large V_{DS}, we have $x_d >> x_s$ and (13) predicts that the additional threshold-voltage shift induced by the drain–source bias depends on V_{DS} as $[(V_{DS} - \psi_s^0)^{1/2} - (-\psi_s^0)^{1/2}$. However, as will be shown in Section IV, a more accurate analysis predicts a linear variation of V_T with V_{DS}, in better agreement with experimental data.

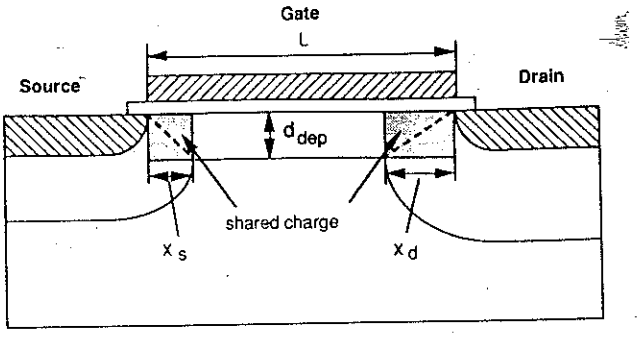

Fig. 2. Schematic illustration of the depletion zones associated with the source, the drain, and the gate for an applied drain–source voltage. The regions of charge sharing are indicated by the shaded areas near source and drain.

By assuming that the ideality factor η does not change significantly with bias conditions, the shift in the interface potential can be evaluated from (6) as

$$\Delta\psi_s = -\frac{\Delta V_T}{\eta}. \tag{14}$$

IV. Drain-Induced Barrier Lowering and Threshold-Voltage Shift

Based on a simple consideration of charge sharing, the model of Section III gives averaged estimates of some important short-channel effects in the subthreshold regime of MOSFET's. This is achieved by distributing the effective gate depletion charge evenly along the channel in order to estimate the threshold-voltage shift. While this may be a good approximation for $V_{DS} = 0$, it will fail to accurately predict the effect on V_T of an applied drain–source voltage. The reason is that a portion of the additional depletion charge induced by the drain–source bias will be distributed nonuniformly from source to drain, as indicated in Fig. 3. Likewise, the drain–source bias will induce a nonuniform shift $V(x)$ in the interface potential along the channel which increases from $V(0) = 0$ at the source to $V(L) = V_{DS}$ at the drain.

Here, we proceed to develop a model for the distribution of the induced shift $V(x)$ in the interface potential along the channel as a result of the applied drain–source bias. From such a model, it is possible to calculate the interface potential near its minimum, which defines the barrier for charge injection into the channel (see Fig. 1). An accurate estimate of the shift in the potential minimum is especially important since the channel current is exponentially dependent on the barrier height. In principle, this involves the solution of a two-dimensional Poisson's equation for the whole device, using proper boundary

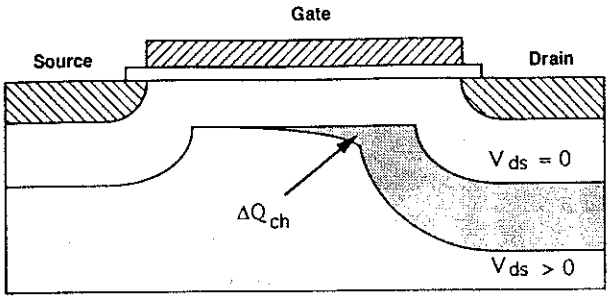

Fig. 3. Distribution of depletion charge induced by an applied drain–source bias, indicated by the shaded region. ΔQ_{ch} is the part of the induced charge which is located in the central channel region and which has its counter-charge on the gate electrode.

conditions. However, this requires extensive numerical calculations. Instead, we perform a simplified analytical calculation.

We start by considering the two-dimensional Poisson's equation for the depletion region under the gate, away from the source and drain contact depletion regions. In the subthreshold regime, the influence of the charge carrier on the electrostatics of the channel can be neglected, and we write the two-dimensional Poisson's equation as

$$\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial_y} = -\frac{qN_a}{\epsilon_s} \tag{15}$$

where F_x and F_y are the longitudinal and perpendicular components of the electric field, respectively. Integrating this equation with respect to y from the semiconductor–insulator interface through the depletion region yields

$$\frac{\partial F_x}{\partial x} d_{\text{dep}} - F_y(0) = -\frac{q}{\epsilon_s} N_a d_{\text{dep}} \tag{16}$$

where $\langle \partial F_x/\partial x \rangle$ is the average of $\partial F_x/\partial x$ over the thickness of the depletion region. This thickness can be estimated approximately from a one-dimensional theory as

$$d_{\text{dep}} \approx \sqrt{\frac{2\epsilon_s}{qN_a}(\psi_s + V_{bi})}. \tag{17}$$

The vertical component of the electric field $F_y(0)$ at the semiconductor–channel interface can be found by requiring the electric displacement to be continuous across the interface, i.e.,

$$F_y(0) = \frac{c_i}{\epsilon_s}(V_{GS} - V_{FB} - \psi_s - V_{bi}) \tag{18}$$

where $c_i = d_i/\epsilon_i$ is the insulator capacitance per unit area, d_i is the insulator thickness, ϵ_i is the electrical permittivity of the insulator, and V_{FB} is the flatband voltage. In the presence of a drain–source bias, the interface potential can be written as

$$\psi_s = \psi_s^0 + V(x). \tag{19}$$

Away from the source and drain contacts, we can assume that $\langle \partial F_x/\partial x \rangle = 0$ when $V_{DS} = 0$. We now consider (16) with and without an applied drain–source bias and express the net effect of the drain–source bias by taking the difference, i.e.,

$$\frac{\partial^2 V}{\partial x^2} d_{\text{dep}}(x) - \frac{c_i}{\epsilon_s} V(x) = \frac{q}{\epsilon_s} N_a [d_{\text{dep}}(x) - d_{\text{dep}}^0] = \frac{q}{\epsilon_s} N_a d_{\text{dep}}^0 \cdot \left(\sqrt{1 + \frac{V(x)}{\psi_s^0 + V_{bi}}} - 1 \right) \tag{20}$$

where d_{dep}^0 is the depletion width for $V = 0$. In (20), we have replaced $\langle \partial F_x/\partial x \rangle$ by $-\partial^2 V/\partial x^2$, assuming that $V(x)$ inside the gate depletion region is relatively weakly dependent on the distance from the interface. The second term on the left-hand side of (20), which is equal to the difference $F_y(0) - F_y^0(0)$, where $F_y^0(0)$ is the value of $F_y(0)$ for $V = 0$, is obtained by using (17) and (18). Since both $V(x)$ and its x-derivatives are small outside the depletion region of the drain contact, all terms in (20) can be expanded to first order in V to give

$$\frac{\partial^2 V}{\partial x^2} - \frac{V}{\lambda^2} = 0 \tag{21}$$

where

$$\lambda = d_{\text{dep}}^0 \left(1 + \frac{\epsilon_i}{\epsilon_s} \frac{d_{\text{dep}}^0}{d_i} \right)^{-1/2}. \tag{22}$$

The general solution of (21) can be written on the following form:

$$V(x) = A \exp \frac{x}{\lambda} + B \exp \left(-\frac{x}{\lambda} \right) \tag{23}$$

where the coefficients A and B are determined from the boundary conditions. Without much error, we can assume that (23) is also valid through the source–channel junction region (i.e., $x < x_s$), in which case we have the boundary condition $V(x = 0) = 0$, which gives $A = -B \equiv V_0/2$, such that (23) can be written as

$$V(x) = V_0 \sinh \left(\frac{x}{\lambda} \right). \tag{24}$$

Here, V_0 is a constant that remains to be determined. We note that the shift $-qV(x_{\min}) \approx -qV(x_s)$ in the condition band at the channel side of the source–channel junction is identical to the DIBL (see Fig. 1).

In order to find the voltage V_0, however, it is necessary investigate more closely the electrostatics of the entire channel. In particular, we have to consider the additional charges induced in the gate electrode and in the substrate as a result of the applied drain–source voltage. We start by taking an inventory of the induced charges and counter charges, as shown schematically in Fig. 4.

From the principle of charge sharing discussed previously, we can pair charges and countercharges as indicated in Fig. 4. According to this scheme, a fraction of

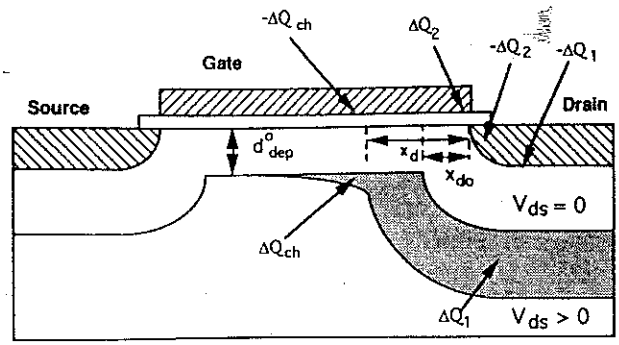

Fig. 4. Schematic overview of the drain-bias-induced charges and counter charges according to the principle of charge sharing. ΔQ_{ch} and $-\Delta Q_{ch}$ are the induced charges in the channel. The remaining charges and counter charges are between the drain and the substrate (ΔQ_1 and $-\Delta Q_1$) and between the drain and the gate (ΔQ_2 and $-\Delta Q_2$).

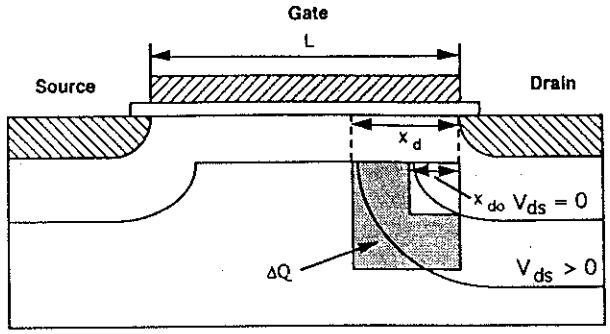

Fig. 5. Simplified model of the drain-bias-induced charge ΔQ in the substrate under the gate (shown as an approximate estimate of the depletion charge under the gate between the depletion boundary for $V_{DS} > 0$ and that for $V_{DS} = 0$). The induced channel charge ΔQ_{ch} is taken to be a fraction of ΔQ, according to the principle of charge sharing.

the drain-bias-induced depletion charge, marked ΔQ_{ch} in the figure, will have its countercharge on the gate electrode. In order to be consistent with the potential variation along the channel [see (24)], the corresponding sheet charge distribution $\Delta q(x)$ along the channel has to be as follows:

$$\Delta q(x) = \frac{V(x)}{c_{dep}(x)} \approx \frac{V_0 \epsilon_s}{d^0_{dep}} \sinh\left(\frac{x}{\lambda}\right) \tag{25}$$

where we have invoked the gradual channel approximation and expanded to lowest order in $V(x)$. Here, $c_{dep}(x)$ is the channel depletion capacitance per unit area. Assuming, for simplicity, that (25) is valid over the range $x_s \le x \le L - x_d$, we obtain the following expression for V_0 by requiring that the integral of $\Delta q(x)$ over this range equals ΔQ_{ch}

$$V_0 \approx \frac{\Delta Q_{ch} d^0_{dep}}{W\lambda\epsilon_s}\left[\cosh\left(\frac{L - x_d}{\lambda}\right) - \cosh\left(\frac{x_s}{\lambda}\right)\right]^{-1}. \tag{26}$$

It now remains to determine the induced channel depletion charge ΔQ_{ch}. A rough estimate of this charge can be obtained as indicated in Fig. 5. The shaded region in the substrate indicates roughly the amount of additional depletion charge $\Delta Q \approx qWN_a(x_d^2 - x_{d0}^2)$, induced under the gate by the drain–source bias. Here, $x_{d0} \approx x_s$ is the depletion width of the drain–channel junction at zero drain–source voltage. From the concept of charge sharing, ΔQ_{ch} can be taken to be some fraction χ of ΔQ, i.e.,

$$\delta Q_{ch} \approx \Delta Q \approx \chi W q N_a(x_d^2 - x_{d0}^2) = 2\chi W \epsilon_s V_{DS}. \tag{27}$$

Typically, χ is of the order of 0.5. However, the value of this parameter can be adjusted to account for the shape and doping profiles in the drain junction and substrate. In other words, this fitting parameter is technology-dependent. In some cases (e.g., for Low Doped Drain (LDD) MOSFET's and ion-implanted devices), not only χ but also N_a can be adjusted to a suitable value.

The parameter V_0 can be obtained by substituting (27) into (26)

$$V_0 \approx 2\chi V_{DS}\frac{d^0_{dep}}{\lambda}\left[\cosh\left(\frac{L - x_d}{\lambda}\right) - \cosh\left(\frac{x_s}{\lambda}\right)\right]^{-1}. \tag{28}$$

Substituting (28) in (24) and setting $x = x_{min} \approx x_s$, the lowering of the injection barrier is found to be

$$\text{DIBL} \approx qV(x_s) \approx qV_0 \sinh\left(\frac{x_s}{\lambda}\right)$$

$$\approx 2\chi q V_{DS}\frac{d^0_{dep}}{\lambda}\frac{\sinh\left(\frac{x_s}{\lambda}\right)}{\cosh\left(\frac{L - x_d}{\lambda}\right) - \cosh\left(\frac{x_s}{\lambda}\right)}. \tag{29}$$

We note that the barrier lowering predicted by (29) decreases exponentially with increasing gate length for $L - x_d > 3\lambda$. For sufficiently small gate lengths or sufficiently high drain–source bias such that $L \approx x_d + x_s$, the DIBL diverges and (29) is no longer valid. This condition corresponds to severe punchthrough in the device

As a consequence of the barrier lowering, we have a drain-bias-induced shift in the threshold voltage. From (14) and (29), we find

$$\Delta V_T = -\eta\Delta\psi_s(x_s) = -\eta V(x_s) = -\sigma V_{DS} \tag{30}$$

where

$$\sigma \approx \frac{2\eta\chi d^0_{dep}}{\lambda}\frac{\sinh\left(\frac{x_s}{\lambda}\right)}{\cosh\left(\frac{L - x_d}{\lambda}\right) - \cosh\left(\frac{x_s}{\lambda}\right)}. \tag{31}$$

In Fig. 6, we show experimental data of ΔV_T versus V_{DS} for two short-channel NMOS devices (from Chung *et al.* [7]). It is evident that the experimental data are in good agreement with the linear relationship of (30), and the slopes of the straight lines give the parameter σ. A linear

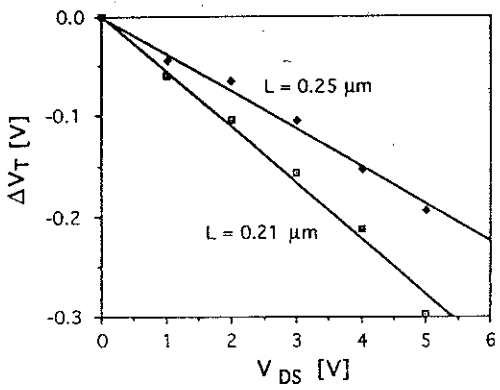

Fig. 6. Experimental threshold-voltage shift versus drain–source voltage for two NMOS devices with effective gate length of 0.21 and 0.25 μm. Equation (30) is fitted to the two data sets yielding $\sigma = 0.056$ ($L = 0.21$ μm) and $\sigma = 0.038$ ($L = 0.25$ μm). (Experimental data from Chung *et al.* [7].)

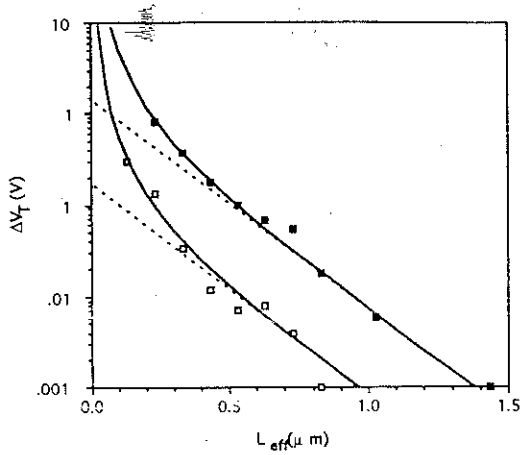

Fig. 7. Experimental values (symbols), fitted model calculations (solid lines), and exponential approximation (dotted lines) of shift in threshold voltage versus effective gate length, in a semilog plot. Upper curves: $T = 85$ K, lower curves: $T = 300$ K. (Experimental data from Liu *et al.* [8].)

relationship was also observed experimentally by Liu *et al.* [8], except for an enhanced shift in V_T at very small values of V_{DS}, typically $V_{DS} \leq .3V_{th}$. This deviation from linearity is expected since V_{DS} becomes comparable with or less than the subthreshold saturation voltage ($V_{SAT} = 2V_{th}$). Below saturation, drift current dominates over diffusion in the MOSFET channel and the applied drain–source voltage is distributed more evenly along the channel, resulting in a more pronounced DIBL than predicted above.

It is interesting to notice that the dependence of σ on N_a and L predicted by (31) is similar to that obtained by de Graaff and Klaassen [9] in cases where the hyperbolic functions can be expanded. Also, the present model expressions for the shift in threshold voltage with gate length is in good qualitative agreement with experimental data by Liu *et al.* [8], as shown in Fig. 7. Like many others, Liu *et al.* found that ΔV_T varies close to exponentially with the effective gate length L_{eff}. However, as indicated in the figure, an accelerated shift in the threshold voltage is observed at very small values of L_{eff} (typically less than 0.4 μm for the technology used), in agreement with the present theory. As can be seen from the figure, our theory reproduces the experimental data to $L_{eff} < 0.2$ μm. We expect that this model can be valid down to effective gate lengths of 0.1 μm.

The above estimates are simplified and partly empirical. They do not take directly into account, for example, the effect of the contact junction depth r_j on the short-channel effects, which may be important. In fact, experimental studies show that a transition from long- to short-channel behavior takes place when

$$L < L_{min}(\mu m) = 0.4\,[r_j(\mu m) d_i(\text{Å})(W_d + W_s)^2(\mu m^2)]^{1/3} \quad (32)$$

where W_d and W_s are the drain–substrate and source–substrate depletion widths, respectively (see Brews *et al.* [10]). This expression indicates the importance of the contact depth. Indirectly, however, this behavior may be accounted for by a judicial choice of the adjustable parameter x_s.

V. Generalized Subthreshold Current

The current–voltage characteristics in the presence of DIBL can now be derived from the drift-diffusion expression in (2). Alternatively, when the channel length becomes sufficiently small, we can obtain the characteristics in terms of the thermionic emission current across the energy barrier. In reality, however, pure thermionic emission current only takes place in a region of a few mean free paths of the maximum of the barrier, and drift-diffusion transport dominates in the rest of the channel. Hence, both processes are always present although one of them will normally dominate. Here, we calculate the drain current based on each of the two transport mechanisms separately, and propose a unified expression in which both are included, following the approach proposed by van der Ziel *et al.* [11] for short n-i-n diodes.

In the drift-diffusion theory, only the region $x_s \leq x \leq L - x_d$ will contribute significantly to the integral in the denominator on the right-hand side of (2). Besides, since the integrand is exponentially dependent on the potential, we can safely expand the hyperbolic sine function in (24) to first order in the distance, i.e.,

$$\begin{aligned}\int_{x_s}^{L-x_d} \exp\left(-\frac{\psi(x)}{V_{th}}\right) dx &\approx \exp\left(-\frac{\psi^0}{V_{th}}\right) \int_{x_s}^{L-x_d} \exp\left(-\frac{V_0 x}{V_{th}\lambda}\right) dx \\ &= \frac{\lambda V_{th}}{V_0} \exp\left(\frac{\psi^0 + V(x_s)}{V_{th}}\right) \\ &\quad \cdot \left\{-\exp\left[-\frac{V_0(L - x_s - x_d)}{V_{th}\lambda}\right]\right\}. \end{aligned} \quad (33)$$

Following the derivation used for the long-channel case, we obtain for the drift-diffusion drain current

$$I_d(\mathrm{dd}) = \frac{qN_c W\delta D_n V_0}{\lambda V_{\mathrm{th}}} \cdot \frac{\exp\left(\dfrac{\psi_s^0 + V(x_s)}{V_{\mathrm{th}}}\right)\left[1 - \exp\left(-\dfrac{V_{DS}}{V_{\mathrm{th}}}\right)\right]}{1 - \exp\left[-\dfrac{V_0(L - x_s - x_d)}{V_{\mathrm{th}}\lambda}\right]}. \tag{34}$$

This expression reduces to (11) for small values of the applied V_{DS} and/or for large values of L, since in both cases V_0 and $V(x_s)$ become very small.

We now instead assume that the drain current is governed by thermionic emission. At zero drain–source bias, the net drain current is zero, and the equilibrium thermionic emission current from source to drain, $I^0_{s\to d}$, and from drain to source, $I^0_{d\to s}$, are given by

$$I^0_{s\to d} = I^0_{d\to s} = I_0 \exp\left(\frac{\psi_s^0}{V_{\mathrm{th}}}\right) \tag{35}$$

where $I_0 = W\delta A^* T^2$ and A^* is Richardson's constant. With a finite drain–source voltage, the current contributions become (including the effect of DIBL)

$$I_{s\to d} = I_0 \exp\left[\frac{\psi_s(x_s)}{V_{\mathrm{th}}}\right] = I_0 \exp\left(\frac{\psi_s^0 + V(x_s)}{V_{\mathrm{th}}}\right) \tag{36}$$

$$I_{d\to s} = I_{s\to d} \exp\left(-\frac{V_{DS}}{V_{\mathrm{th}}}\right) \tag{37}$$

and the net thermionic emission current can be written as

$$I_d(\mathrm{te}) = I_{s\to d} - I_{d\to s} = I_0 \exp\left(\frac{\psi_s^0 + V(x_s)}{V_{\mathrm{th}}}\right)\left[1 - \exp\left(\frac{V_{DS}}{V_{\mathrm{th}}}\right)\right]. \tag{38}$$

We now wish to find a unified expression for the I–V characteristics which includes both drift-diffusion and thermionic emissions. A complete theory to this end would, however, be quite involved and impractical. But, as observed above, the two mechanisms act predominantly in a serial fashion, each representing a different part of the channel. Accordingly, the drain current will be dominated by the current-limiting mechanism which will be the one that corresponds to the highest impedance. This is quite analogous to a series combination of two admittances where the current is dominated by smaller of the two. In line with this analogy, we propose a unified expression for the subthreshold I–V characteristics as the following combination of the two currents in (34) and (38):

$$I_d \approx \left[\frac{1}{I_d(\mathrm{dd})} + \frac{1}{I_d(\mathrm{te})}\right]^{-1}. \tag{39}$$

It is clear from the above results that the current-limiting mechanism will be thermionic emission at small gate length and drift diffusion at larger gate lengths.

VI. Unified Charge Control Model for Short-Channel MOSFET's

According to the Unified Charge Control Model (UCCM) [12], [13], the inversion charge n_s in a MOSFET is related to the gate voltage swing V_{GT} and channel Fermi potential V_F as follows:

$$V_{GT} - \alpha V_F \approx \eta V_{\mathrm{th}} \ln\left(\frac{n_s}{n_0}\right) + a(n_s - n_0) \tag{40}$$

where α is a parameter accounting for the body effect (i.e., the dependence of the threshold voltage of the strong inversion on the channel voltage [13]), $a \approx q/c_i + q\Delta d/\epsilon_s$, and Δd is a quantum correction to the effective insulator thickness. In (40), we have not considered explicitly the effect of DIBL on the threshold voltage. However, by including the threshold-voltage shift determined in (30), UCCM can be generalized to read

$$V_{GT0} + \sigma V_{DS} - \alpha V_F \approx \eta V_{\mathrm{th}} \ln\left(\frac{n_s}{n_0}\right) + a(n_s - n_0) \tag{41}$$

where V_{GT0} is the threshold voltage swing at zero drain–source bias.

In this paper, the drain-bias-induced threshold voltage shift is analyzed in terms of lowering of the injection barrier between the source and the channel in the subthreshold regime. In strong inversion however, the injection barrier is reduced owing to the effect of the gate–source bias, and will eventually disappear well above threshold. Hence, the importance of DIBL will be gradually less with increasing gate voltage-source voltage and should gradually be phased out from the expression for V_T. For modeling purposes, σ should therefore include a dependence on V_{GT} as, for example, in the following empirical expression used in the unified MOSFET model discussed in [5], [13]:

$$\sigma = \frac{\sigma_0}{1 + \exp\left(\dfrac{V_{GT0} - V_{\sigma T}}{V_\sigma}\right)} \tag{42}$$

which gives $\sigma \to \sigma_0$ for $V_{GT} < V_{\sigma T}$ and $\sigma \to 0$ for $V_{GT} > V_{\sigma T}$. The voltage V_σ determines the width of the transition between the two regimes. A qualitatively similar dependence of σ on the gate voltage was obtained by de Graaff and Klaassen [9] who related this dependence to the interplay between two different physical mechanisms—DIBL (which is important in the subthreshold regime and just above the threshold) and static feedback (which is a capacitive coupling between the drain and the channel that may be important well above threshold).

REFERENCES

[1] R. Troutman, "VLSI limitations from drain-induced barrier lowering," *IEEE Trans. Electron Devices*, vol. ED-26, p. 461, 1979.

[2] S. M. Sze, *Physics of Semiconductor Devices*. New York: Wiley, 1981.

[3] M. Shur, *Physics of Semiconductor Devices*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[4] L. D. Yau, "A simple theory to predict the threshold voltage of short-channel IGFET's," *Solid-State Electron*, vol. 22, p. 701, 1974.

[5] M. Shur, T. A. Fjeldly, T. Ytterdal, and K. Lee, "Unified MOSFET model," accepted for publication in *Solid-State Electron.*, 1992.

[6] W. Fichtner and H. W. Potzl, "MOS modeling by analytical approximations, I. Subthreshold current and threshold voltage," *Int. J. Electron.*, vol. 46, p. 33, 1979.

[7] J. E. Chung, M. C. Jeng, J. E. Moon, P. K. Ko, and C. Hu, "Performance and reliability design issues for deep-submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. 38, no. 3, pp. 545–554, 1991.

[8] Z. H. Liu, J. H. Huang, J. Duster, P. K. Ko, C. Hu, M. C. Jeng, and Y. C. Chen, "Threshold voltage modeling for deep-submicrometer conventional and LDD MOSFETs at 300 K and 85 K," in *Proc. Int. Semicond. Device Research Symp., ISDRS'91* (Charlottesville, VA, Dec. 1991), pp. 411–414.

[9] H. C. de Graaff and F. M. Klaassen, *Compact Transistor Modeling for Circuit Design* (Computational Microelectronics Series), S. Selberherr, Ed. Wien, New York: Springer-Verlag, 1990.

[10] J. R. Brews, "A charge-sheet model of the MOSFET," *Solid-State Electron.*, vol. 21, pp. 345–355, 1978.

[11] A. van der Ziel, M. S. Shur, K. Lee, T. H. Chen, and K. Amberiadis, "Carrier distribution and low-field resistance in short n^+-n-n^+ structures," *IEEE Trans. Electron Devices*, vol. ED-30, no. 2, pp. 128–137, Feb. 1983.

[12] Y. Byun, K. Lee, and M. Shur, "Unified charge control model and subthreshold current in heterostructure field effect transistors," *IEEE Electron Device Lett.*, vol. 11, no. 1, pp. 50–53, Jan. 1990 (see erratum *IEEE Electron Device Lett.*, vol. 11, no. 6, p. 273, June (1990)).

[13] K. Lee, M. Shur, T. A. Fjeldly, and T. Ytterdal, *Semiconductor Device Modeling for VLSI*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

Tor A. Fjeldly (M'85–SM'88) received the M.Sc. degree in physics from the Norwegian Institute of Technology, University of Trondheim, in 1967, and the Ph.D. degree in physics from Brown University, Providence, RI in 1972.

From 1972 to 1974 he was with the Max-Planck-Institute for Solid State Physics in Stuttgart, Germany. From 1974 to 1983 he worked as a Senior Scientist at ELAB, a research organization in Trondheim, Norway. Since 1983 he has been on the faculty of the Norwegian Institute of Technology, University of Trondheim, where he is Professor of Electrical Engineering. In 1988 and 1989 he had summer research engagements with the University of Minnesota, Minneapolis. Since 1990, he has been a Visiting Professor at the Department of Electrical Engineering, University of Virginia, Charlottesville. His research interests have included fundamental studies of semiconductors and other solids, development of solid-state chemical sensors, electron transport in semiconductors, and modeling and simulation of high-speed semiconductor devices.

Dr. Fjeldly is a member of the Norwegian Academy of Technical Sciences, the American Physical Society, the Norwegian Physical Society, and the Electrochemical Society.

Michael Shur (M'78–SM'80–F'89) received the M.S.E.E. degree (with honors) from the Leningrad Electrotechnical Institute, Leningrad (now St. Petersburg), USSR (now Russia), in 1965 and the Ph.D. degree in physics from A. F. Ioffe Institute of Physics and Technology, Leningrad (now St. Petersburg), in 1967.

From 1965 to 1976 he was with A. F. Ioffe Institute of Physics and Technology. In 1976 he joined the Department of Electrical Engineering of Wayne State University, Detroit, MI. In 1978 he was with the School of Engineering of Oakland University, Rochester, MI. During 1976–1980 he also had visiting summer appointments with the School of Electrical Engineering at Cornell University, Ithaca, NY. In the summer of 1979 he was a visiting faculty member at IBM Thomas J. Watson Research Center, Yorktown Heights, NY. From 1979 to 1989 he was with the University of Minnesota, Minneapolis, MN. Since 1989 he has been John Marshall Money Professor of Electrical Engineering at the University of Virginia, Charlottesville, VA. His research has included ferroelectrics, amorphous semiconductors, solar cells, ballistic transport, high-speed semiconductor devices, and integrated circuits. He published several books, including a book on physics of semiconductor devices (with microcomputer programs) (Englewood Cliffs, NJ: Prentice-Hall, 1990). He is a member of the Editorial Board of *International Journal of High Speed Electronics*, one of the Contributing Editors of *Review of Radio Science, 1990* (published by UNESCO), and an Associate Editor of the IEEE TRANSACTIONS ON ELECTRON DEVICES.

Dr. Shur is a member of the American Physical Society, Eta Kappa Nu, Tau Beta Pi, and a member of Commission D of the International Union of Radio Science, and a Vice Chairman of the national chapter of Commission D.