MOSFETs in the Sub-threshold Region (i.e. a bit below V_T)

Clifton Fonstad, 10/28/09

In the depletion approximation for n-channel MOS structures we have neglected the electrons beneath the gate electrode when the gate voltage is less than the threshold voltage, V_T. We said that it is only when the gate voltage is above threshold that they are significant, and that they are then the dominant negative charge under the gate. Furthermore, we say that above threshold all of the gate voltage in excess of V_T induces electrons in the channel; thus our model is that the sheet charge density under the gate, q_N^*, is

$$q_N^* = \begin{cases} 0 & \text{for} \quad v_{GC} \leq V_T \\ \dfrac{\varepsilon_{ox}}{t_{ox}}\left(v_{GC} - V_T\right) & \text{for} \quad V_T \leq v_{GC} \end{cases} \tag{1}$$

As MOS integrated circuit technology has evolved to exploit smaller and smaller device structures, it has become increasingly important in recent years to look more closely at the minority carriers present under the gate when the gate voltage is less than threshold, i.e. in what is called the "sub-threshold" region. These carriers cannot be totally neglected, and play an important role in device and circuit performance. At first they were viewed primarily as a problem, causing undesirable "leakage" currents and limiting circuit performance. Now it is recognized that they also enable a very useful mode of MOSFET operation, and that the sub-threshold region of operation is as important as the traditional cut-off, linear, and saturations regions of operation.

To begin our study of the sub-threshold region, we will first quickly review the electrostatics of the MOS capacitor, and the electrostatic potential profile predicted by the depletion approximation model. Then we will use this result to derive a more accurate expression than that in Equation 1 for q_N^* below threshold, and use the resulting expression to, among other things, assess the assumption that the contribution of the mobile electrons underneath the gate to the net charge density in the depletion region is negligible compared to the contribution from the ionized acceptors. Finally we will look at the current-voltage characteristic of a MOSFET operating in the sub-threshold region, and merge it with our earlier model so that we then have a model in which the mobile electron charge is taken into account and the drain current is no longer identically zero when v_{GS} is less than V_T.

The Electrostatics of the MOS Capacitor with $v_{BC} = 0$

Consider the MOS capacitor with $v_{BC} = 0$ illustrated in Figure 1, the same structure we used when we first looked at the MOS capacitor using the depletion approximation. In the depletion
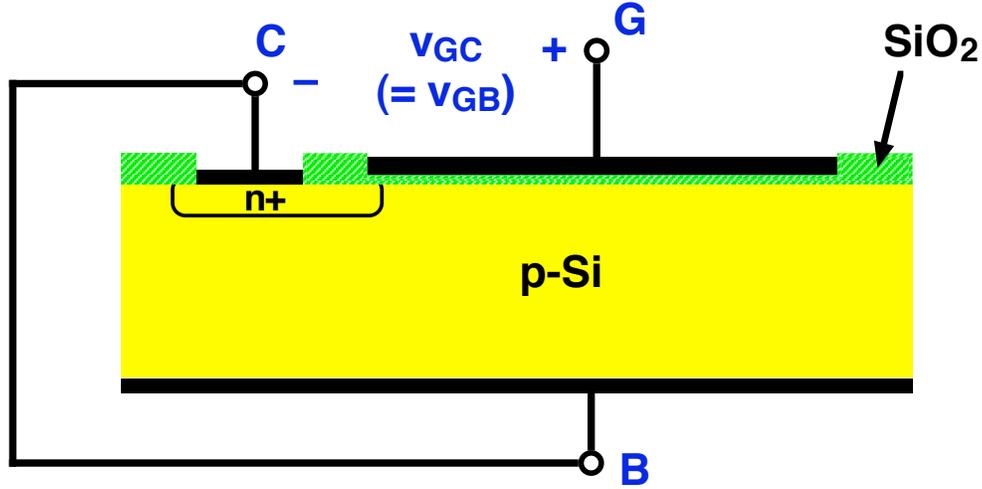

FIGURE 1
A MOS capacitor connected as a two-terminal capacitor with $v_{GC} = v_{GB} = 0$.

approximation, we assume that Equation 1 holds and that the net charge density profile, $\rho(x)$, under the gate for $V_{FB} < v_{GC} < V_T$ can be approximated as:

$$\rho(x) = \begin{cases} qN_A & \text{for} \quad 0 \le x \le x_D \\ 0 & \text{for} \quad x_D \le x \end{cases} \tag{2}$$

With this assumption, we found that the electrostatic potential profile is:

$$\phi(x) = \begin{cases} \phi_p + \dfrac{qN_A(x - x_D)^2}{2\varepsilon_{Si}} & \text{for} \quad 0 \le x \le x_D \\ \phi_P & \text{for} \quad x_D \le x \end{cases} \tag{3}$$

This expression is plotted in Figure 2, which also continues the plot through the oxide to the gate, from which we can also get the expression relating the depletion region width, x_D, to v_{GB} and V_{FB}:

$$v_{GB} - V_{FB} = \frac{qN_A x_D t_{ox}}{\varepsilon_{ox}} + \frac{qN_A x_D^2}{2\varepsilon_{Si}} \tag{4}$$
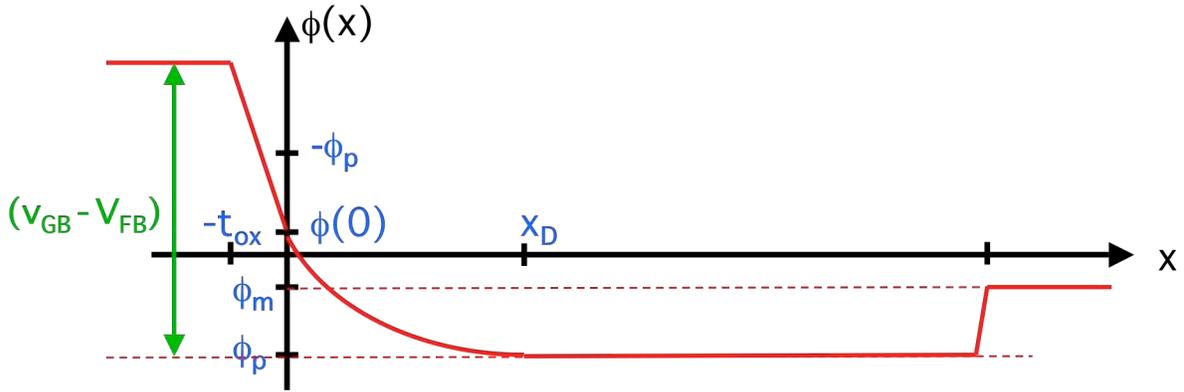
FIGURE 2
A sketch of $\phi(x)$ from the metal on the left, through the oxide, and into the p-type semiconductor in an n-channel MOS capacitor for an applied gate bias, v_{GB}, in the weak-inversion, sub-threshold region.

Equation 4 is useful because it can be solved explicitly for x_D, and the result can be used to obtain an expression for $\phi(x)$ as a function of v_{GB}. However, it will turn out that what is most important to us is $\phi(0)$, the value of the potential at the interface, and $\phi(0)$ is much easier to relate to v_{GB} than is $\phi(x)$ at an arbitrary x. To do so we first find x_D in terms of

$$\phi(0): x_D = \sqrt{\frac{2\varepsilon_{Si}\left[\phi(0) - \phi_p\right]}{qN_A}} \tag{5}$$

Using this in Equation 4 gives us an equation relating $\phi(0)$ and v_{GB} that will be useful to us shortly:

$$v_{GB} - V_{FB} = \frac{t_{ox}}{\varepsilon_{ox}}\sqrt{2\varepsilon_{Si}qN_A\left[\phi(0) - \phi_p\right]} + \left[\phi(0) - \phi_p\right] \tag{6}$$

Sub-threshold Electron Sheet Charge Density, $v_{GC} = 0$

Returning to our original goal, which was to find the electron population density, n(x), under the gate, and then the electron sheet charge density, $q_N{}^*$, we note that the Boltzman relationship between the electrostatic potential and carrier population holds under the gate of the MOS capacitor in Figure 1 because the current in the x-direction is zero. Thus we have:

3

$$n(x, v_{GB}) = n_i e^{\phi(x, v_{GB})/\phi_t} \tag{7a}$$

$$= \frac{n_i^2}{N_A} e^{\left[\phi(x, v_{GB}) - \phi_P\right]/\phi_t} \tag{7b}$$

$$p(x, v_{GB}) = n_i e^{-\phi(x, v_{GB})/\phi_t} \tag{8a}$$

$$= N_A e^{-\left[\phi(x, v_{GB}) - \phi_P\right]/\phi_t} \tag{8b}$$

In these equations we use ϕ_t for the thermal voltage, kT/q, and we have indicated the dependency on v_{GB} to emphasize that these populations depend on the gate voltage as well as on position, x. To obtain the Equations 7b and 8b, we have used $p_o = N_A = n_i \exp(-\phi_p/\phi_t)$ to get expressions explicitly including the quantity $[\phi(x,v_{GB}) - \phi_p]$, which also appears in Eqs. 5 and 6. Note: In many texts, $[\phi(0,v_{GB}) - \phi_p]$ is identified as $V_B(v_{GB})$, the voltage drop between the silicon bulk and the oxide-silicon interface, i.e. $V_B(v_{GB}) \equiv [\phi(0,v_{GB}) - \phi_p]$.

We can calculate the electron sheet charge density, q_N^*, by multiplying $n(x)$ by $-q$ and integrating with respect to x from the interface, $x = 0$, into the silicon until $x = x_i$, where x_i is defined as the depth at which $\phi(x) = 0$, and thus where we have $n(x_i) = p(x_i) = n_i$:

$$q_N^*(v_{GB}) = \int_0^{x_i(v_{GB})} -q\, n(x, v_{GB})\, dx \tag{9a}$$

$$= -q n_i \int_0^{x_i(v_{GB})} e^{\phi(x, v_{GB})/\phi_t}\, dx \tag{9b}$$

$$= \frac{-q n_i^2}{N_A} \int_0^{x_i(v_{GB})} e^{\left[\phi(x, v_{GB}) - \phi_P\right]/\phi_t}\, dx \tag{9c}$$

The end-point $x = x_i$ is used for the integration because for $x > x_i$, $p(x) > n(x)$ and the material is still p-type, while for $x < x_i$, $n(x) > p(x)$ and the material net n-type and is said to be "weakly inverted."   This is actually a minor point, however, and not worth fretting about, because $n(x)$ falls off very rapidly with increasing x, and the main contributions to the integral come from the region near the interface (i.e. small x) where $\phi(x)$ is near $\phi(0)$.   The integral itself will be significant only when $\phi(0)$ approaches $-\phi_p$, which further reduces the importance of the tail, and how far from the interface one integrates.

The next step in calculating q_N^* would seem to be to replace $\phi(x)$ in Eq. 9c with an explicit function of x so we can do the integral. We can do this using Eq. 3, and we find

$$q_N^*(v_{GB}) = -\frac{q\,n_i^2}{N_A} \int_0^{x_i(v_{GB})} e^{qN_A[x-x_D(v_{GB})]^2 / 2\varepsilon_{Si}\phi_t}\, dx \tag{10}$$

This integral is clearly difficult to evaluate, however, without resorting to numerical techniques, a less than optimum situation. An alternative approach is to make use of our earlier observation that the main contribution to the integral occurs near x = 0, and to further note that near x = 0 the potential variation with x is nearly linear. A bit of algebra gives us

$$\phi(x, v_{GB}) = \phi(0, v_{GB}) - \frac{qN_A}{2\varepsilon_{Si}}[2x_D(v_{GB}) - x] \cdot x \tag{11a}$$

$$\approx \phi(0, v_{GB}) - \frac{qN_A x_D(v_{GB})}{\varepsilon_{Si}} \cdot x \tag{11b}$$

$$\approx \phi(0, v_{GB}) - \sqrt{\frac{2qN_A[\phi(0, v_{GB}) - \phi_p]}{\varepsilon_{Si}}} \cdot x \tag{11c}$$

Using this approximation in the Eq. 9c yields an analytical expression which does not obscure the dependences on material properties

$$q_N^*(\phi) \approx -q\frac{n_i^2 \phi_t}{N_A} \sqrt{\frac{\varepsilon_{Si}}{2qN_A[\phi(0, v_{GB}) - \phi_p]}}\, e^{[\phi(0, v_{GB}) - \phi_p]/\phi_t} \quad \text{for } \phi(0, v_{GB}) \le -\phi_p \text{ i.e } v_{GB} \le V_T \tag{12}$$

Note that in the range validity, $\phi(0, v_{GB}) \le -\phi_p$, corresponds to gate voltages such that $v_{GB} \le V_T$.

To see if the sub-threshold charge is large or small relative to the channel charge which develops above threshold [i.e., $q_N^*(v_{GB}) = C_{ox}^*(v_{GB} - V_T)$], one can evaluate Eq. 12 for values of $\phi(0, v_{GB})$ near threshold, say for example $(-\phi_p - 10\phi_t)$ to $-\phi_p$. The corresponding values of v_{GB} can be found using Eq. 6, and then one can plot q_N^* verses v_{GB} in the vicinity of $v_{GB} = V_T$; the result is shown in Figure 3 for a MOS capacitor with $N_A = 10^{18}$ cm^{-3} and $t_{ox} = 3$ nm. Note that the weak-inversion charge does not increase further above threshold, but instead is pinned at the value it reaches when $v_{GB} = V_T$. (This value is 7.2×10^{-9} Coul/cm^2 for the specific values of N_A and t_{ox} used in Figure 3).

The first thing to note in Figure 3 is that the range of gate voltages where the sub-threshold charge is significant in this example is from about 50 mV below threshold to threshold. Note also that neglecting the sub-threshold charge only leads to a 6 mV error in the threshold, V_T, that
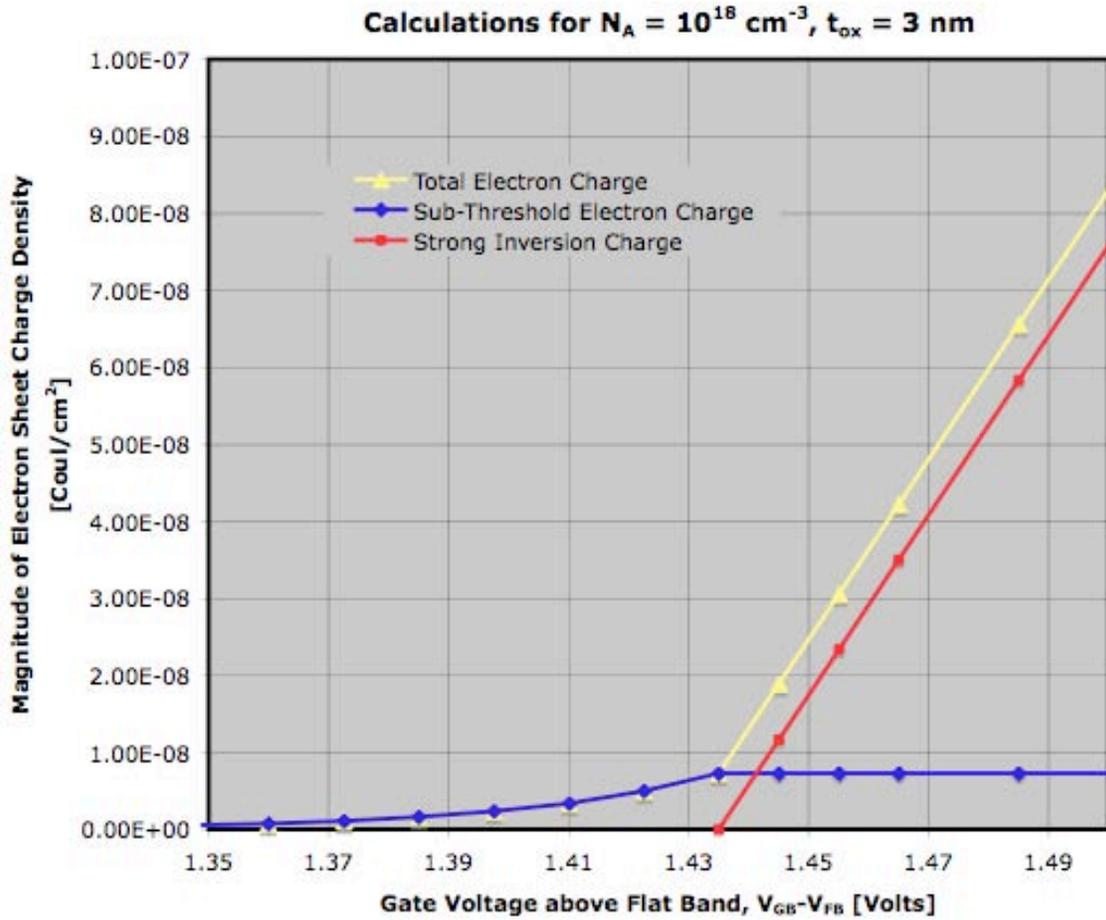
FIGURE 3
The electron sheet charge density under the gate with a gate voltage in the vicinity of threshold. The blue curve corresponds to the sub-threshold weak-inversion charge [Eq. 12], and the red curve is the strong inversion charge from traditional depletion approximation modeling [Eq. 1]. The sum is plotted in the yellow curve.

would be extrapolated from a C-V measurement; thus, taking the sub-threshold electron charge into account makes only a very minor correction to the predicted value of 1.43 Volts.

Another way to get a feel for the relative significance, or insignificance, of the sub-threshold electron charge is to compare it to the sheet charge in the depletion region at threshold, i.e. $qN_AX_D = (2\varepsilon_{Si}|2\phi_p|qN_A)^{1/2}$. A quick calculation shows that this charge is 5.2×10^{-7} coul/cm^2, which is about 70 times larger than $q_N^*(V_T)$.

Returning to our expression for the charge under the gate below threshold, Eq. 12, we next take the derivation one step further to express q_N^* explicitly in terms of v_{GB}. One way to do this is to return to Eq. 6 and solve it for $\phi(0,v_{GB})$, which yields

6

$$\phi(0, v_{GB}) = \phi_p + \frac{2\varepsilon_{Si}qN_A}{(\varepsilon_{ox}/t_{ox})^2}\left[\sqrt{1 + \frac{v_{GB} - V_{FB}}{2\varepsilon_{Si}qN_A/(\varepsilon_{ox}/t_{ox})^2}} - 1\right]^2 \tag{13}$$

Putting this into Eq. 12 clearly won't lead to an equation that gives one much insight, so it seems worthwhile to look into making simplifying approximations before proceeding, particularly since we know from Figure 3 that the most important region is the small range of voltages below $v_{GB} = V_T$, i.e. $\phi(0) = -\phi_p$. Perhaps $\phi(0)$ can be modeled satisfactorily as varying linearly with V_T.

To see if a linear approximation is realistic, it is instructive to use Eq. 12 to plot $\phi(0, v_{GB})$ vs v_{GB}, which has been done in Figure 4. Referring to this figure, it seems clear that a linear approximation will be pretty good when the surface potential is much larger that the bulk potential, and that it is reasonable to approximate $\phi(0, v_{GB})$ near $v_{GB} = V_T$ as

$$\phi(0, v_{GB}) - \left(-\phi_p\right) \approx \left(v_{GB} - V_T\right)\left.\frac{d\phi(0, v_{GB})}{dv_{GB}}\right|_{\phi(0, v_{GB}) = -\phi_p} \tag{14}$$

To evaluate the derivative in this expression we first use Eq. 12 to calculate $dv_{GB}/d\phi$,
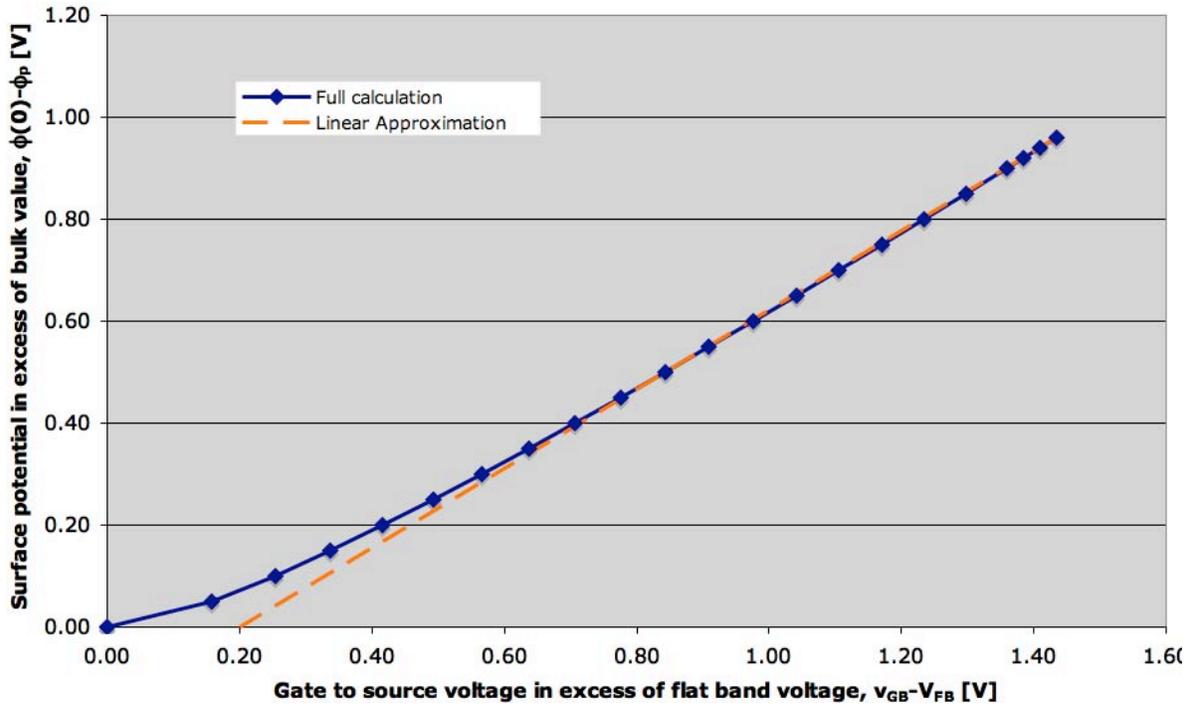

FIGURE 4
Comparing the variation of the surface potential, $\phi(0)$, with gate voltage, v_{GB}, in flat band on the left to the vicinity of threshold on the right. The relationship is clearly very linear in the weak inversion sub-threshold region.

$$\frac{dv_{GB}}{d\phi(0)} = 1 + \frac{t_{ox}}{\varepsilon_{ox}}\sqrt{\frac{\varepsilon_{Si}qN_A}{2[\phi(0)-\phi_p]}} \tag{15}$$

Then we evaluate it at $\phi(0,v_{GB}) = -\phi_p$, labeling the result "n":

$$\left.\frac{dv_{GB}}{d\phi(0)}\right|_{\substack{\phi(0)=-\phi_p \\ v_{GB}=V_T}} = 1 + \frac{1}{C_{ox}^*}\sqrt{\frac{\varepsilon_{Si}qN_A}{2[-2\phi_p]}} \equiv n \tag{16}$$

Notice that in writing this expression we have replaced ε_{ox}/t_{ox} with its equivalent, C_{ox}^*. Inverting Eq. 16 and using it in Eq. 14 yields the approximate linear expression we seek

$$\phi(0,v_{GB}) \approx (-\phi_p) - \frac{1}{n}(V_T - v_{GB}) \tag{17}$$

Finally, we insert Eq. 17 into Eq. 12 getting a much more manageable and instructive expression for $q_N^*(v_{GB})$ than we would have gotten using Eq. 13 in Eq. 12 directly. Using Eq. 17 we find

$$q_N^*(v_{GB}) \approx -q\frac{n_i^2\phi_t}{N_A}\sqrt{\frac{\varepsilon_{Si}}{2qN_A[-2\phi_p - (V_T - v_{GB})/n]}}\, e^{[-2\phi_p - (V_T - v_{GB})/n]/\phi_t} \quad \text{for } v_{GB} \leq V_T \tag{18}$$

where $V_T = V_{FB} + |2\phi_p| + (2\varepsilon_{Si}qN_A|2\phi_p|)^{1/2}/C_{ox}^*$.

Equation 18 can be simplified further by recognizing that $e^{-2\phi_p} = (N_A/n_i)^2$. Making this substitution yields

$$q_N^*(v_{GB}) \approx -\phi_t\sqrt{\frac{qN_A\varepsilon_{Si}}{2[-2\phi_p - (V_T - v_{GB})/n]}}\, e^{-(V_T - v_{GB})/n\phi_t} \quad \text{for } v_{GB} \leq V_T \tag{19}$$

Still further simplification can be achieved by $(V_T-v_{GB})/n$ is typically much smaller than $|2\phi_p|$, so we can say $[-2\phi_p - (V_T-v_{GB})/n] \approx -2\phi_p$. Making this approximation, we see that the square root term in Eq. 19 is the same as the one in Eq. 16 defining n and that it is approximately $(n-1)C_{ox}^*$. Consequently, we can write $q_N^*(v_{GB})$ as.

$$q_N^*(v_{GB}) \approx -(n-1)C_{ox}^*\phi_t\, e^{-(V_T - v_{GB})/n\phi_t} \quad \text{for } v_{GB} \leq V_T \tag{20}$$

This result concludes our derivation of $q_N^*(v_{GB})$ for two-terminal MOS capacitors. Summarizing our results thus far, we find that accounting for the electrons under the gate below threshold has an arguably negligible impact on the electrostatics of the MOS capacitor. Its

primary impact is to require a minor correction in the threshold voltage calculation. The situation will be quite different in the case of the sub-threshold drain current of a MOSFET, which we will study shortly. First, however, we must extend our results to include three-terminal MOS capacitors, i.e., to the case where $v_{BC} \neq 0$.

Sub-threshold Electron Sheet Charge Density in a 3-terminal MOS Capacitor, $v_{BC} \neq 0$

To model the sub-threshold drain current of a MOSFET, we have to extend our model of the electron population density, n(x), under the gate to include a bias between the adjacent n+ region and the channel, i.e. to allow for a non-zero v_{BC}, as shown in Figure 5. To understand how this bias, v_{BC}, effects the electron population under the gate, it is instructive to first think about the role the n+ region played in establishing the electron population in the two-terminal MOS capacitor situation we just looked at. We in fact did not even mention the n+ region during our derivation, but the n+ region does play a very important role because it supplies the electrons that are under the gate below threshold, i.e., $q_N^*(v_{GB}<V_T)$. The n+ region is a reservoir of electrons, and those electrons spill out of it into the region under the gate when the electrostatic potential energy barrier confining the electrons to the reservoir, $-q\Delta\phi$, is lowered by a positive gate voltage. The barrier is lowered most near the oxide-silicon interface, and most of the carriers spill across near that interface, however, at any depth that the barrier is reduced, carriers can spill across. What we calculated when we found $q_N^*(v_{GB})$ is the total number of electrons that can
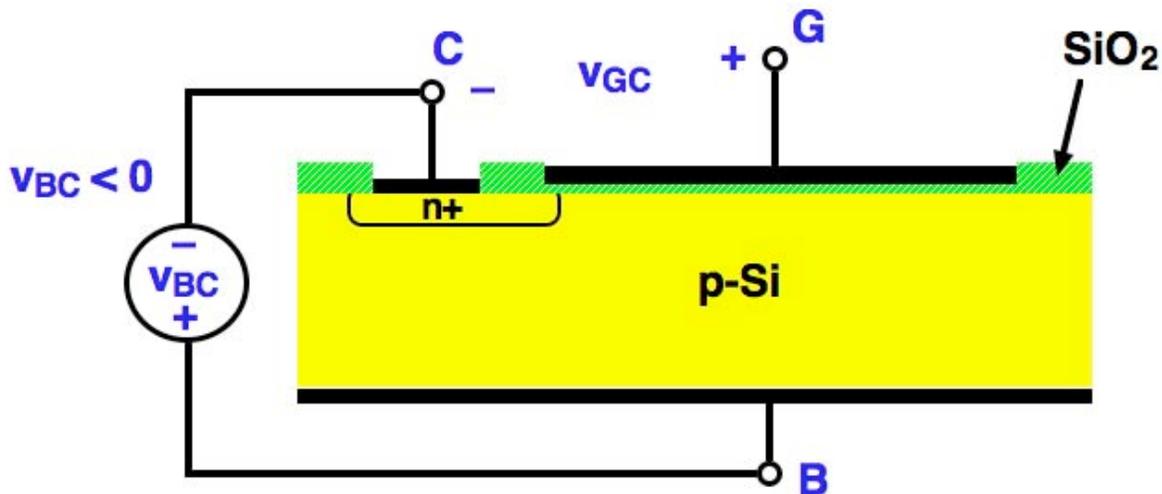

FIGURE 5
A MOS capacitor connected as a bias applied between the substrate and n+ region, i.e. $v_{BC} \neq 0$.

spill across and into the region under the gate when the voltage on the gate is v_{GB}.

If the n+ region is biased positive relative to the substrate, i.e. when $v_{BC} < 0$, then the potential energy of the carriers in the reservoir is lower than when $v_{BC} = 0$ V. As a result fewer carriers can spill out of the source and into the region under the gate; quantitatively, the number that can spill into this region decreases exponentially with $|v_{BC}|$, i.e. e^{v_{BC}/ϕ_t}. Thus Eq. 12 becomes:

$$q_N^*\left(v_{GB}\right) \approx -q\phi_t \sqrt{\frac{\varepsilon_{Si}}{2qN_A\left[\phi(0,v_{GB})-\phi_p\right]}}\ e^{\left[\phi(0,v_{GB})+v_{BC}\right]/\phi_t}\quad \text{for}\ \ \phi(0,v_{GB}) \le -\phi_p - v_{BC} \qquad (21)$$

and the equivalent of Eq. 20 is:

$$q_N^*\left(v_{GB}\right) \approx -\left[n(v_{BC})-1\right]C_{ox}^* \phi_t\ e^{-\left[V_T(v_{BC})-v_{GC}\right]/n\phi_t}\quad \text{for}\ \ v_{GB} \le V_T(v_{BC}),\ v_{BC} \le 0 \qquad (22)$$

The substrate bias, v_{BC}, does not appear explicitly in Eq. 22, but it does play an important role because now both V_T and n are now functions of v_{BC}:

$$n(v_{BC}) = 1 + \frac{1}{C_{ox}^*}\sqrt{\frac{\varepsilon_{Si}qN_A}{2\left[-2\phi_p - v_{BC}\right]}} \qquad (23)$$

and

$$V_T\left(v_{BC}\right) = V_{FB} + \left|2\phi_p\right| + \left|v_{BC}\right| + \frac{1}{C_{ox}^*}\sqrt{2\varepsilon_{Si}qN_A\left(\left|2\phi_p\right| + \left|v_{BC}\right|\right)} \qquad (24)$$

When $v_{BC} < 0$, the depletion region is wider near threshold, so n is smaller (near to 1) and the threshold voltage, V_T, is larger than when $v_{BC} = 0$.

With the derivation of Eq. 22 we are ready to find the drain current of a MOSFET biased below threshold and in weak inversion, i.e. in the sub-threshold region.

The MOSFET Drain Current in the Sub-threshold Region

In the gradual channel approximation modeling of MOSFET terminal characteristics we have done thus far we have said that the drain current is identically zero when the gate voltage is less than the threshold voltage, i.e. when $v_{GS} \le V_T$. As we will soon see, the population of mobile electrons we have just calculated under the gate provides a mechanism for charge flow between the drain and source even when $v_{GS} \le V_T$, and thus there is in fact a small, non-zero drain current through a MOSFET biased below threshold. This is significant because on an integrated circuit chip with millions of transistors which are supposed to be off and therefore not dissipating any

power, a little current flowing through each transistor can easily add up to be a significant power drain, and be a source of serious heating. Furthermore we will find that a MOSFET's characteristics are well defined in the sub-threshold region, and are in no way anomalous or parasitic. In fact, they have such interesting properties that it is highly advantageous in a large set of applications to design circuits that operate specifically in the sub-threshold region.

Consider the MOSFET illustrated in Figure 6, and suppose that it is biased with a positive drain voltage, i.e. $v_{DS} \geq 0$, and a negative substrate voltage, i.e. $v_{BS} \leq 0$. Suppose also that the
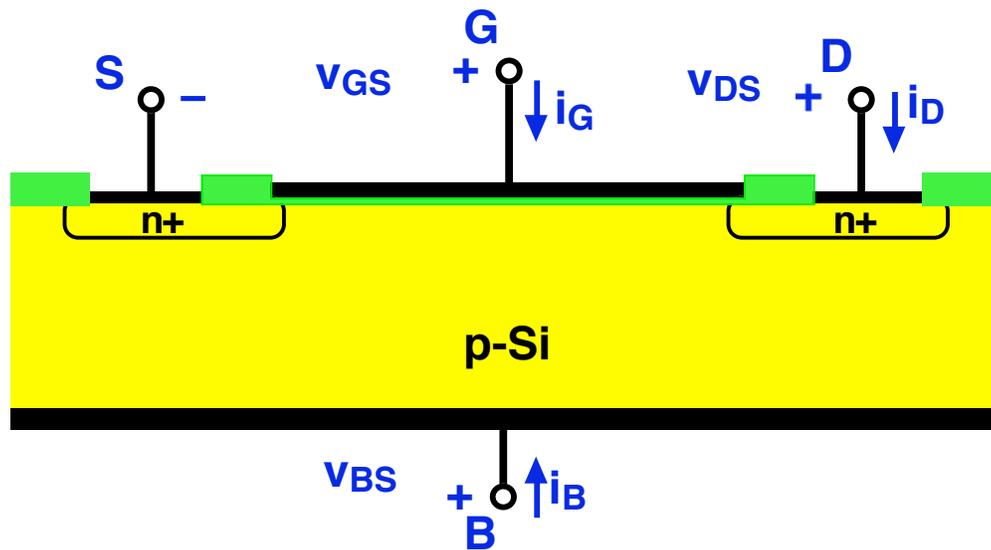

FIGURE 6
A cross-sectional view of an n-channel MOSFET. The sub-threshold region corresponds to having the device biased with $v_{BS} \leq 0$, $v_{DS} \geq 0$, and v_{GS} near, but below threshold, V_T.

gate is biased positively and a bit less than threshold. At the source end of the gate, electrons will spill into the region under the gate as we have just calculated. The same will be true at the drain end of the gate, and if $v_{DS} = 0$ the electron population under the gate will be uniform because the situation will be equivalent to that in a 3-terminal MOS capacitor, and the steady state current will be identically zero. If $v_{DS} > 0$, however, the density of electrons that can spill under the gate at the drain end will be smaller than at the source end by a factor of e^{-v_{DS}/ϕ_t}. Rather than electrons flowing under the gate from the drain, the electrons spilling into the region under the gate from the source will flow all the way under the gate and into the drain n+ region when they reach it. As a result there will be a net positive drain current. To calculate the size of this drain current, we need to model how the electrons flow under the gate.

First, consider drift. There is a voltage difference between the source and the drain, but the electric field laterally between them is small because the voltage between the back contact and the source, v_{BS}, falls across the depletion region between the n+ source and the p-type substrate, and the voltage between the back contact and the drain, v_{BD}, falls across the drain-substrate depletion region. There is thus negligible lateral field in the channel region below threshold to drift any electrons there from source to drain.

Recognizing that drift is negligible below threshold, we see that the mechanism behind the sub-threshold drain current must be diffusion driven by the electron concentration gradient going from the source to drain:

$$i_D = -W \cdot D_e \frac{dq_N^*}{dy} \tag{25}$$

Since i_D is effectively uni-directional, i.e. wholly y-directed, and does not depend on L, the concentration gradient must be constant and equal to the difference between the end-point concentrations divided by L, giving us

$$i_D = -W\,D_e \frac{q_N^*(0) - q_N^*(L)}{L} \tag{26}$$

Using Eq. 21, we see that because the barrier the electrons see at the drain is v_{DS} larger than the barrier at the source, $q_N^*(y = L)$ and $q_N^*(y = 0)$ must be related by:

$$q_N^*(L) = q_N^*(0)e^{-v_{DS}/\phi_t} \tag{27}$$

Using this in Equation 26 we obtain

$$i_D = \frac{W}{L} D_e \left[n(v_{BS}) - 1 \right] C_{ox}^* \phi_t \, e^{-[V_T(v_{BS}) - v_{GS}]/n\phi_t} \left[1 - e^{-v_{DS}/\phi_t} \right] \tag{28}$$

Replacing D_e with $\mu_e \phi_t$, Eq. 27 becomes

$$i_D(v_{GS}, v_{DS}, v_{BS}) = \frac{W}{L} \mu_e C_{ox}^* \left[n(v_{BS}) - 1 \right] \phi_t^2 e^{-[V_T(v_{BS}) - v_{GS}]/n(v_{BS})\phi_t} \left[1 - e^{-v_{DS}/\phi_t} \right] \tag{29}$$

To simplify things further, and get a final working expression for i_D, we recognize the leading terms are K_o, leave out the explicit dependence of n and V_T on v_{BS}, and reorder v_{GS} and V_T:

$$i_D = K_o (n - 1) \phi_t^2 e^{(v_{GS} - V_T)/n\phi_t} \left(1 - e^{-v_{DS}/\phi_t} \right) \tag{30}$$

Finally we can identify the pre-factor $K_o(n-1)\phi_t^2$ as the sub-threshold saturation current, $I_{S,s-t}$, and write this expression as:

$$i_D = I_{S,s-t}\,e^{(v_{GS}-V_T)/n\phi_t}\left(1 - e^{-v_{DS}/\phi_t}\right) \quad \text{with} \quad I_{S,s-t} \equiv K_o(n-1)\,\phi_t^2 \qquad (31)$$

This expression is plotted in Figure 7 on a log-linear graph, i.e. log i_D vs v_{DS}, which results in equally spaced drain current traces for equal increments in v_{GS}-V_T. Exactly the same plot would be obtained for a BJT if log I_C was plotted vs. v_{CE}, for equal increments in v_{BE}, except that rather than increasing by a factor of 10 for each 60 mV increment in v_{BE}, which is the case for i_C, i_D increases by a factor of 10 for each $n \times 60$ mV increment in v_{GS}.
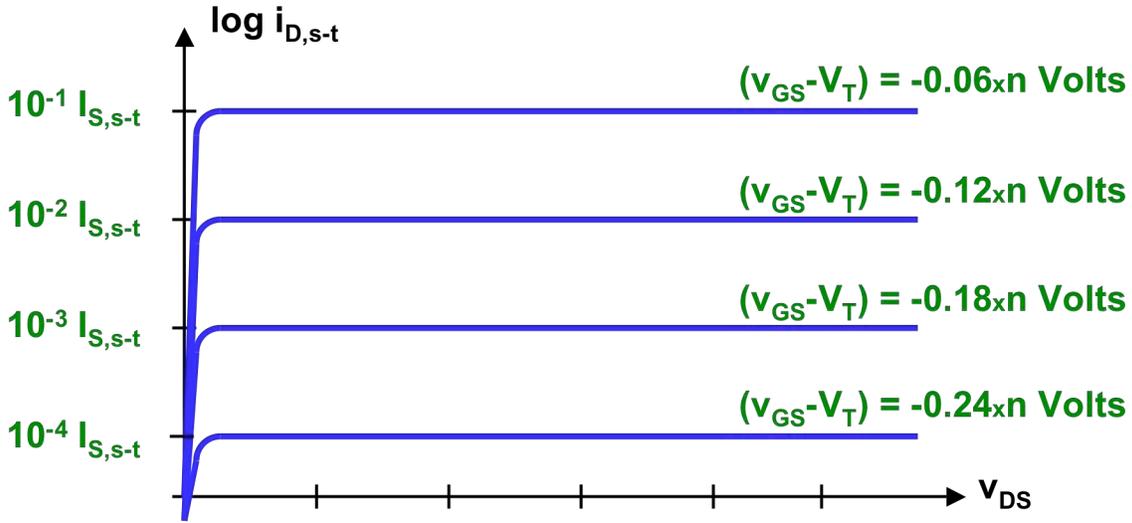

FIGURE 7
The output characteristics of an n-channel MOSFET operating in the sub-threshold region in a log-linear plot. The drain current is normalized to the sub-threshold drain saturation current, $I_{S,s-t}$.

Connecting the Sub-threshold Diffusion and GCA Strong Inversion Drift Models

The sub-threshold drain current expression, Eq. 31, was derived assuming $v_{GS} \le V_T$, and it reaches its maximum value, $I_{S,s-t}$, at $v_{GS} = V_T$. For larger v_{GS}, according to the depletion approximation model, the width of the depletion region, x_D, and the variation of the electrostatic energy profile with x, stay fixed. This implies that the diffusion current between the drain and source saturates at its peak value, $I_{S,s-t}$, and the additional drain current, i_D, for $v_{GS} > V_T$ is due to carrier drift in the inversion region (i.e., in the channel). Strictly speaking, then, we should write the drain current of an n-channel MOSFET as follows:

$$i_D(v_{GS}, v_{DS}, v_{BS}) = \begin{cases} I_{S,s-t} e^{(v_{GS}-V_T)/n\phi_t} & \text{for} \quad \dfrac{1}{\alpha}\left[v_{GS}-V_T\right] < 0 < v_{DS} \\[2mm] I_{S,s-t} + \dfrac{K_o}{2\alpha}\left[v_{GS}-V_T\right]^2 & \text{for} \quad 0 < \dfrac{1}{\alpha}\left[v_{GS}-V_T\right] < v_{DS} \\[2mm] I_{S,s-t} + K_o\left\{v_{GS}-V_T-\alpha\dfrac{v_{DS}}{2}\right\}v_{DS} & \text{for} \quad 0 < v_{DS} < \dfrac{1}{\alpha}\left[v_{GS}-V_T\right] \end{cases} \quad (32)$$

Notice that when writing Eq. 32 we assumed that $v_{DS} > 3\phi_t$, so $(1-e^{v_{DS}/\phi_t}) \approx 1$.

In practice, however, we seldom include $I_{S,s-t}$ in the expressions for the drain current in the linear and saturation regions. To understand why this is reasonable it is useful to compare the magnitudes of the terms $I_{S,s-t}$ and $K_o(v_{GS}-V_T)^2/2\alpha$. To do this in a general way we can first notice that both of these terms depend on K_o, so it makes sense normalize them relative to K_o, and thus to compare $i_{D(sub-threshold)}/K_o$ and $i_{D(strong\ inversion)}/K_o$. We will look in the vicinity of V_T, and vary $(v_{GS}-V_T)$. That way we don't have to specify either K_o or V_T, and our results will be more general.

In the sub-threshold region we have

$$\frac{i_{D,sub-threshold}}{K_o} = (n-1)\,\phi_t^2\,e^{(v_{GS}-V_T)/n\phi_t} \quad (33)$$

and above threshold, in strong inversion and saturation, we have

$$\frac{i_{D,sub-threshold}}{K_o} = \frac{1}{\alpha}\left(v_{GS}-V_T\right)^2 \quad (34)$$

For the sake of discussion, let us say that we have an n-channel MOSFET with the same MOS capacitor we used earlier, with $N_A = 10^{18}$ cm^{-3} and $t_{ox} = 3$ nm, and that $v_{BS} = 0$ and $v_{DS} \gg \phi_t$. The α (and n) of this device is 1.25, and the maximum normalized diffusion current, $i_{D,s-t}/K_o$, is 1.56 x 10^{-4} V^2. For comparison, the normalized drift current when v_{GS} is just 60 mV above threshold, i.e. $(v_{GS} - V_T) = 0.06$ V, is almost an order of magnitude larger, 1.5 x 10^{-3} V^2, and it is more than two orders of magnitude larger, 1.6 x 10^{-2} V^2, when $(v_{GS} - V_T) = 0.2$ V.

If we plot the sum of the normalized diffusion and drift currents in the range $V_T \pm 0.5$ V, as shown in Figure 8, we can barely see the diffusion current on a linear plot. We have to look much more closely to V_T, well within $V_T \pm 0.1$ V, to see it, as Figure 8b illustrates.   The same currents are presented in a log-linear plot in Figure 9. In this plot, the exponential nature of the sub-threshold diffusion current is clear, and so too is the n x 60 mV/decade sub-threshold slope.
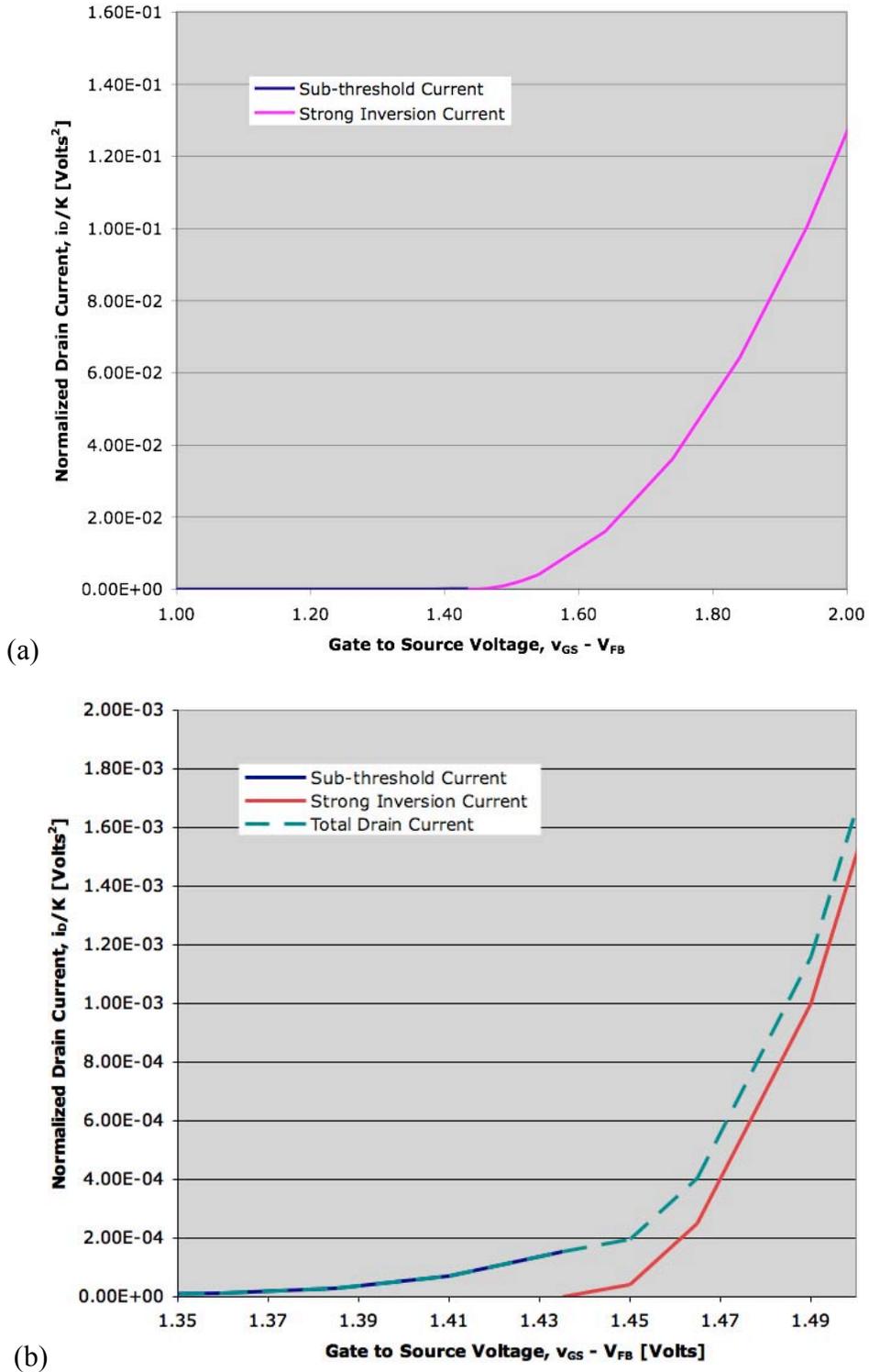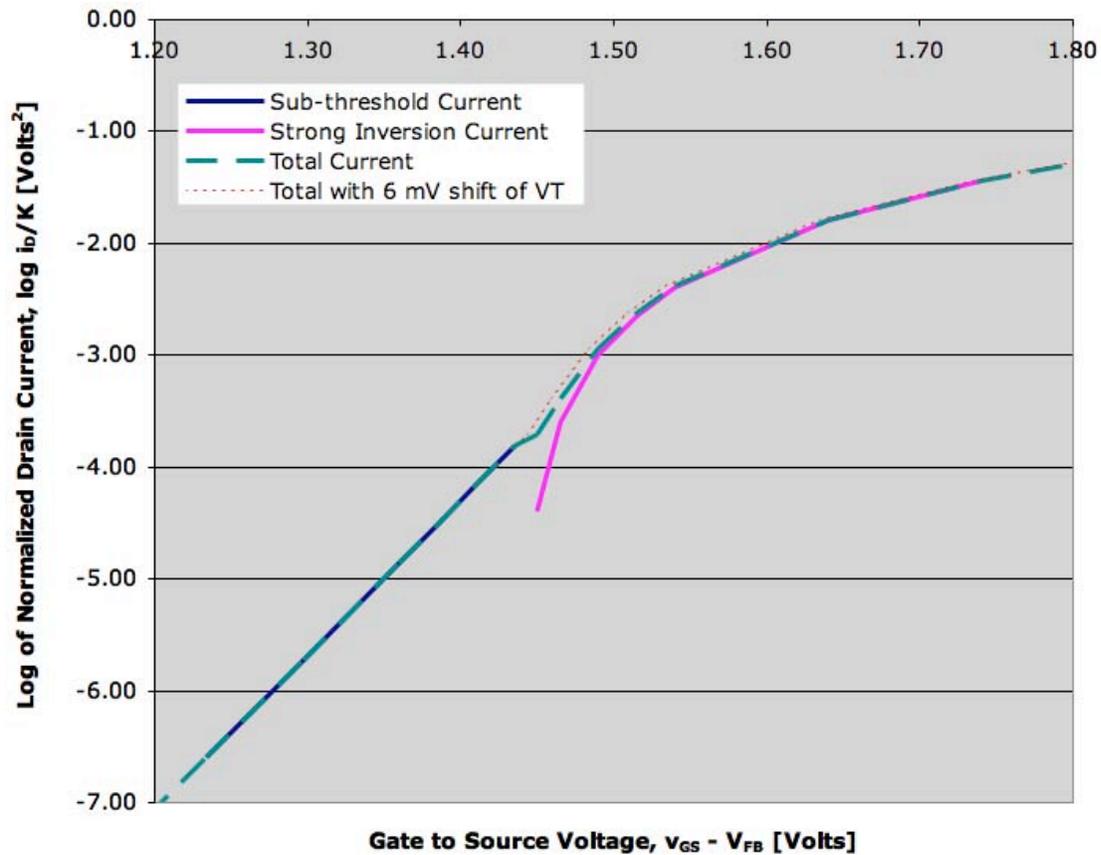
(a)

(b)

FIGURE 8

The drain current, i_D, for gate biases from just below to just above threshold illustrating the relative sizes of the diffusion (sub-threshold) and drift (strong inversion) components of i_D. Figure 8 a shows the situation with a course voltage scale, while Figure 8b has smaller increments and thus focuses in more on the curves near V_T.

FIGURE 9
The drain current, i_D, for gate biases from just below to just above threshold as in Figure 8b, but now plotted on a log-linear graph so the nature of the small, sub-threshold current and its variation with input voltage, v_{GS}, is more clear.

The simple depletion approximation model we are using, our neglect of drift currents below threshold, and our use of a saturated diffusion current above threshold, are all reasonable approximations, and are better the further we are above or below threshold. We should expect them to have limitations very near threshold, however, and in particular we should not demand too much from them as we pass from $v_{GS} < V_T$ to $v_{GS} > V_T$. The curves look rather smooth in Figure 8 b, but we see a slight kink on the log scale of Figure 9. Interestingly, this kink can be smoothed nicely by using the adjusted threshold in the saturation current expression, as shown in Fig. 9, but is largely coincidence as the 6 mV shift came from looking at the total gate charge in the sub-threshold region, not the inversion layer in strong inversion. Just the same, our simple models accurately reflect the physics, and do an outstanding job no matter how one looks at it.

6.012 Microelectronic Devices and Circuits
Fall 2009