[15] S. Ogura, C. Codella, N. Rovedo, J. Shepard, and J. Risman, "An optimized half micrometer device using double implanted LDD structure," in *Proc. Symp. VLSI Tech.*, (Kanagawa, Japan, Sept. 1-3, 1982), pp. 42-45.
[16] E. Takeda, H. Kume, T. Toyobe, and S. Asai, "Submicrometer MOSFET structure for minimizing hot-carrier generation," *IEEE Trans. Electron Devices*, vol. ED-29, pp. 611-618, Apr. 1982.
[17] —, "Submicrometer MOSFET structure for minimizing channel hot electrons," in *Proc. Symp. VLSI Tech.*, (Maui, HA, Sept. 9-11, 1981), pp. 22-23.
[18] P. E. Cottrell, R. R. Troutman, and T. H. Ning, "Hot electron emission in n-channel IGFET's," *IEEE Trans. Electron Devices*, vol. ED-26, pp. 520-533, Apr. 1979.
[19] H. Matsumoto, K. Sawada, S. Asai, M. Hirayama, and K. Nagasawa, "The effect of gate bias on hot electron trapping," *Japan. J. Appl. Phys.*, vol. 19, pp. L574-L576, Oct. 1980.
[20] T. H. Ning, P. W. Cook, R. H. Dennard, C. M. Osburn, S. E. Schuster, and H. N. Yu, "Hot electron design constraints for 1-μm IGFET's," in *IEDM Tech. Dig.* (Washington, DC, Dec. 4-6, 1978), pp. 472-475.
[21] S. A. Abbas and E. E. Davidson, "Reliability implications of hot electron generation," in *Proc. 14th Ann. Reliability Phys. Symp.*, (Las Vegas, NE, Apr. 20-22, 1976), pp. 18-22.
[22] E. Sun, J. Moll, J. Berger, and B. Alders, "Breakdown mechanism in short-channel MOST's," in *IEDM Tech. Dig.*, (Washington, DC, Dec. 4-6, 1978), pp. 478-482.
[23] C. M. Wu, and K. W. Yeh, "Safety drain voltage avoiding avalanche breakdown in MOSFET," *IEEE Electron Device Lett.*, vol. EDL-3, pp. 245-247, Sept. 1982.
[24] W. Muller, L. Risch, and A. Schutz, "Short-channel MOS transistors in the avalanche region," *IEEE Trans. Electron Devices*, vol. ED-29, pp. 1778-1781, Nov. 1982.
[25] F. C. Hsu, R. S. Muller, and C. Hu, "Characteristics of short-channel MOSFET's," in *IEDM Tech. Dig.*, (San Francisco, CA, Dec. 13-15, 1982), pp. 282-285.
[26] F. C. Hsu, P. K. Ko, S. Tam, C. Hu, and R. S. Muller, "An analytical breakdown model for short-channel MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-29, pp. 1735-1740, Nov. 1982.
[27] M. Shiono, O. Nakajima, and C. Hashimoto, "Interface trapped charge generation due to substrate hot electron injection," *J. Electrochem. Soc.*, vol. 129, pp. 1760-1764, Aug. 1983.
[28] D. J. DiMaria, in *The Physics of SiO_2 and Its Interfaces*, S. T. Pantelidis, Ed. New York: Pergamon, 1978, pp. 160-178.
[29] E. Takeda and N. Suzuki, "An empirical model for device degradation due to hot carriers," *IEEE Electron Device Lett.*, vol. EDL-4, pp. 111-113, Apr. 1983.
[30] W. K. Meyer and D. L. Crook, "Model for oxide wearout due to charge trapping," in *Proc. Int. Rel. Phys. Symp.*, (Phoenix, AZ, Apr. 5-7, 1983), pp. 242-247.

The Effect of High Fields on MOS Device and Circuit Performance

CHARLES G. SODINI, MEMBER, IEEE, PING-KEUNG KO, MEMBER, IEEE, AND JOHN L. MOLL, FELLOW, IEEE

Abstract—A simple analytical model for the MOS device characteristics including the effect of high vertical and horizontal fields on channel carrier velocity is presented. Analytical expressions for the drain current, saturation drain voltage, and transconductance are developed. These expressions are used to examine the effect of scaling the channel length, the gate dielectric thickness, and the bias voltage on device characteristics. Experimental results from various geometry MOS devices are used to verify the trends predicted by the model. Using the physical understanding provided by the model, we examine the effect of device geometry scaling on circuit performance. We suggest that for gate capacitance-limited circuits one should reduce the channel length, and for parasitic capacitance-limited circuits one should reduce the gate dielectric thickness to improve circuit performance.

Manuscript received October 25, 1983; revised May 14, 1984. This work was partially supported by a grant from Analog Devices Corporation, the Semiconductor Research Corporation under Contract 83-01-033, and equipment donation from Hewlett-Packard Corporation.

C. G. Sodini is with the Massachusetts Institute of Technology, Cambridge, MA 02139.
P.-K. Ko was with Bell Laboratories, Holmdel, NJ 07733. He is now with the University of California, Berkeley, CA 94720.
J. L. Moll is with Hewlett-Packard Laboratories, Palo Alto, CA 94304.

I. INTRODUCTION

SCALING of MOS device geometries has been proposed for improved circuit performance and density; however, since the operating voltages have not been scaled proportionately, the device performance improvement is limited by high field effects [1]. This subject has been discussed by several authors [2]-[4]. Experimental results have been presented which verify that the improvement in device performance is less than expected when the lateral and vertical fields in the conducting channel become large [5]. In this paper we develop a simple model for the drain current to explain those experimental results. Some simplifying assumptions were made to obtain analytical expressions for the saturation voltage, drain current, and transconductance. We use these expressions to explain the observed deviation of the saturation voltage and transconductance from first-order theory under the high field condition. In addition, the physical mechanisms which cause these deviations are discussed. Based on this physical under-

0018-9383/84/1000-1386$01.00

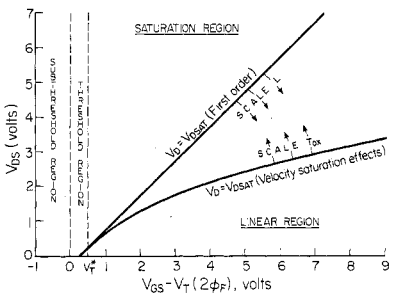

Fig. 1. Qualitative picture describing the regions of operation in a MOS transistor.

standing we propose some guidelines for the optimal use of short-channel devices in various circuit applications.

Fig. 1 shows a qualitative diagram of the various regions of operation for a MOS transistor. The first region called the subthreshold region is independent of drain voltage and is defined as the region with a gate voltage less than that which corresponds to a surface potential of twice the Fermi potential, assuming uniform doping. We call this gate voltage the classical threshold voltage. Several authors have shown that in this region the drain current is dominated by diffusion current and is exponentially dependent on gate voltage [6], [7]. A second region of operation also independent of drain voltage is the threshold region. This region is due to the nonlinear dependence of channel charge on gate voltage above the classical threshold voltage. It was first pointed out by [8] and later experimentally verified by [9]. Several authors have since studied the region [10], [11]. They showed that the drain current is composed of both drift and diffusion components and is subexponentially dependent on gate voltage. As the gate voltage is increased above V_T^* we enter the well-known linear and saturation regions. V_T^* is a qualitative threshold voltage above which we assume that the channel charge becomes linearly proportional to gate voltage. The border between the linear and saturation region occurs at $V_{DS} = V_{DSAT}$. First-order MOS theory says that $V_{DSAT} = V_{GS} - V_T$. V_T in this equation is the extrapolated threshold voltage and not $V_T(2\phi_F)$ as in Fig. 1 [12], [13].

In the following section, we present the derivation and assumptions for our model. We will show, as illustrated in Fig. 1, that the value for the saturation voltage deviates from the first-order theory value as we reduce the channel length and that it approaches the first-order theory value as we reduce the dielectric thickness. In Section III we use the simple model to explain the deviation of the dependence of transconductance on gate drive from first-order theory as we reduce the channel length or dielectric thickness. We also show that as the gate drive is made large, the transconductance approaches a constant. Our model is used to predict the dependence of the value of this constant on the device parameters of low field mobility, channel length, normal field dependence on mobility, and the saturation velocity of channel carriers. In Section IV we use the results from our model to explain the effect of reducing channel lengths and gate dielectric thicknesses on circuit performance.

Experimental results are presented to verify the trends predicted by our model. The devices used in the measurements were n-channel silicon gate transistors fabricated using a four-mask process. The transconductance was measured by observing the change in drain current with a small ($\leqslant$100 mV) change in the gate voltage bias point. The drain bias was fixed at a voltage large enough to ensure that the device was operating in the saturation region. The source-bulk voltage was set equal to zero for all cases. The gate dielectric thickness was determined by a C-V measurement of a large-area (300 μm $\times$ 300 μm) capacitor. A value of 3.9 was assumed for the relative dielectric constant of silicon dioxide. The channel lengths quoted are effective electrical channel lengths, which along with the series resistance, were measured as outlined in [14]. The device transconductance data have been adjusted to remove the effect of the series resistance.

II. The Model

The objective in the development of this model is to demonstrate the effect of high vertical and horizontal fields on the characteristics of the MOS device. We have made some simplifying assumptions which reduce the accuracy but help reveal the physics of the device operation through a simple analytical expression for the drain current.

A. Assumptions

We have assumed that drift current dominates the total current and that the mobile charge in the channel is equal to $C_{ox}(V_G - V_T)$. The threshold voltage V_T in this equation is equal to the extrapolated threshold voltage. Both of these assumptions become valid as the gate drive is increased. We have assumed that the bulk charge has a negligible effect on the threshold voltage along the channel. This assumption is reasonable for short-channel devices which are our primary interest. It is also assumed that there is no drain-induced barrier lowering. This assumption implied that the threshold voltage is independent of the drain-source voltage and is reasonable for "well-designed" devices operating with $V_{DS} \leqslant V_{DSAT}$. Finally, we have assumed that the series resistance external to the device is equal to zero.

B. Velocity Versus Electric Field Model

We assume the following two-region piecewise empirical model for the velocity versus electric field for electrons in the inversion layer [4].

$$v = \frac{\mu_{eff} E}{1 + E/E_C}, \quad E \leqslant E_C$$

$$= v_{SAT}, \quad E > E_C \tag{1}$$

where

E is the lateral electric field and
E_C is the critical field at which the carriers are velocity saturated and is equal to $2v_{SAT}/\mu_{eff}$.

$$\mu_{eff} = \frac{\mu_0}{1 + \theta V_G'} \tag{2}$$

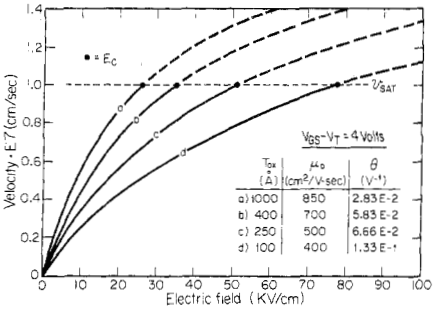

Fig. 2. Channel carrier velocity versus lateral electric field derived from (1). For $E \geq E_c$ the velocity is assumed to be v_{SAT}.

where

- μ_0 is the low field inversion layer mobility,
- V'_G is the gate-source voltage minus the extrapolated threshold voltage, and
- θ is a fitting parameter which governs the normal field dependence ($\approx 2E - 7/T_{ox}V^{-1}$) [15].

We assume the saturation velocity of the carriers in the inversion layer to be a constant, equal to 1E7 cm/s. Recent measurements using a novel time-of-flight technique have demonstrated this assumption to be reasonable [16]. Fig. 2 shows a plot of the velocity of electrons in the inversion layer versus lateral electric field for various oxide thicknesses as generated by (1). The low field mobilities were typical values derived from experimental devices fabricated with appropriate substrate doping [9], [17], [18].

We compare the model in (1) with two commonly used empirical velocity field models [19], [20]. Including the effect of the normal field the models are

$$\text{Model A} \quad v = \frac{\mu_{\text{eff}} E}{\left[1 + \left(\frac{E}{E_c}\right)^2\right]^{1/2}} \tag{3a}$$

$$\text{Model B} \quad v = \frac{\mu_{\text{eff}} E}{1 + \frac{E}{E_c}}. \tag{3b}$$

The difference between the model in (1) and model B can be seen as follows. In model B the electric field E must be much larger than the critical field E_C for the velocity to approach v_{SAT}. The piecewise linear model in (1) yields the velocity of carriers equal to v_{SAT} at the critical electric field. The result is that the piecewise linear model has a steeper slope at high fields and more precisely models the experimental data than model B.

In Fig. 3 we compare the model in (1) with models A and B and experimental data from [21]. The normal field dependence parameter θ was set equal to zero for all cases. The more complicated model A fits the experimental data quite well while the simpler model B underestimates the velocity. The model in (1) has a reasonable fit to the experimental data up to the point at which the velocity is approaching saturation and yet maintains a simple form. The most important point to understand is: *as the normal field is increased it takes a higher lateral field for the carriers to reach the same velocity.*

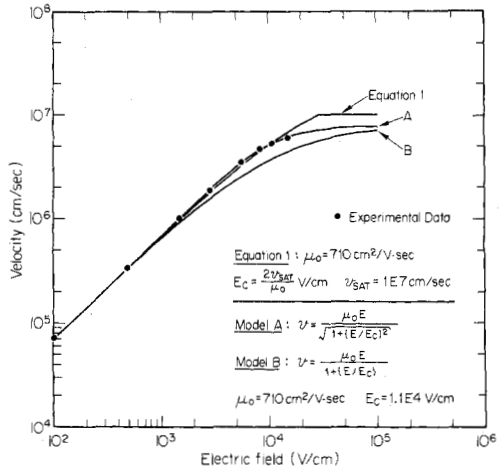

Fig. 3. Comparison of commonly used velocity-field models (models A and B) with the model in (1). Measured data points are taken from Coen and Muller [21].

This central theme will be used to explain many of the experimentally observed phenomena.

C. Derivation of Drain Current Model

To derive the current equation we first write the current at any point x along the channel as

$$I_D = I(x) = WC_{ox}(V'_G - V(x))\,v(x) \tag{4}$$

where

- W is the device width,
- C_{ox} is the gate capacitance per unit area,
- $V(x)$ is the potential difference between minority-carrier quasi-Fermi potential and equilibrium Fermi potential in the bulk at point x, and
- $v(x)$ is the velocity of carriers at point x.

From the velocity field model in (1) we can express the lateral electric field $E(x)$ as

$$E(x) = \frac{I_D}{W\mu_{\text{eff}}C_{ox}(V'_G - V(x)) - I_D/E_C} = \frac{dV(x)}{dx}. \tag{5}$$

By integrating from $x = 0$ to $x = L$ and $V(x) = V_S$ to $V(x) = V_D$ we arrive at

$$I_D = \frac{W\mu_{\text{eff}}C_{ox}(V'_G - V_{DS}/2)V_{DS}}{L(1 + V_{DS}/E_C L)} \quad V_{DS} \leq V_{DSAT}. \tag{6}$$

We define the saturation voltage to be the drain voltage at which the carriers at the drain become velocity saturated. By our model, this condition corresponds to the point at which the lateral electric field at the drain end of the channel becomes equal to the critical field E_C. Substituting the above condition into (5) we get

$$I_D = \frac{W\mu_{\text{eff}}C_{ox}(V'_G - V_{DSAT})E_C}{2}. \tag{7}$$

By equating (6) and (7) we can solve for V_{DSAT}.

$$V_{DSAT} = \frac{E_C L V'_G}{E_C L + V'_G}. \tag{8}$$

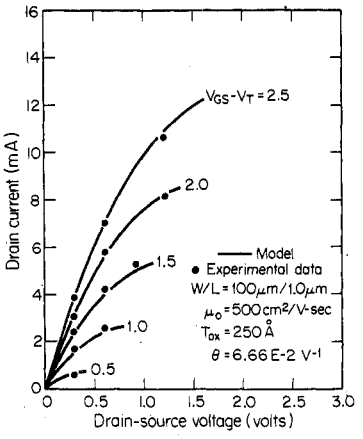

Fig. 4. Comparison of the drain current characteristics generated from the model with experimental data.

Now we have analytically defined the "linear" region for our model.

D. Deviation of V_{DSAT} from Simple Theory

We can explain qualitatively the change in V_{DSAT} as one scales the channel length and dielectric thickness that was shown in Fig. 1. As one scales the channel length it takes a lower voltage at the drain to reach the critical field E_C. Therefore, V_{DSAT} is reduced from its first-order theory value of $V_{GS} - V_T$. As one scales the gate dielectric thickness, however, the critical field E_C increases because of the reduction in carrier mobility with increasing normal field. Therefore, the drain voltage to achieve the critical field at the drain increases and the saturation voltage V_{DSAT} increases toward the first-order theory value.

E. Device Characteristics Generated by Model

Fig. 4 shows the drain current characteristics for $V_{DS} \leq V_{DSAT}$ for a 100 μm/1.0 μm device with a gate dielectric thickness of 250 Å. Also shown are experimental data which correspond closely to our model. The model shows that for low gate drives the value of V_{DSAT} is approximately equal to the gate drive. As one increases the gate drive we see that the saturation voltage is considerably less than the gate drive.

Fig. 5 shows a plot of the normalized transconductance versus gate drive for a 1.1- and a 4.1-μm derive with a gate dielectric thickness of 230 Å. The transconductance is derived from our model at $V_{DS} = V_{DSAT}$. The agreement between experiment and the model is excellent for the 4.1-μm device. The deviation between the model and the 1.1-μm experimental device is explained as follows. At a small gate drive, the effect of channel length modulation and drain-induced barrier lowering both cause the drain current to increase significantly with the drain voltage. Because these effects are not included in the model, the calculated transconductance is smaller than its experimental counterpart. At high gate bias our simple model overestimates the transconductance because the effect of the bulk charge is not included in our model.

The ultimate goal of the model is to qualitatively predict the deviation in transconductance from first-order theory as we scale the channel length and gate dielectric thickness of the MOS device. Although the model is not quantitatively accurate, it does correctly predict the trends in transconductance as one scales device dimensions. Because of the simple analytical nature of the model, we can begin to explain how the various physical parameters will affect transconductance and ultimately device performance.

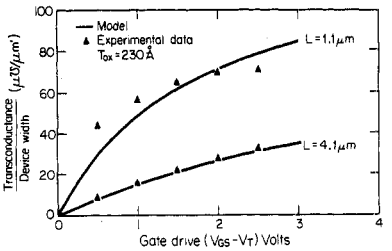

Fig. 5. Comparison of transconductance versus gate drive between the model and experimental data with $V_{DS} \geq V_{DSAT}$.

III. Source and Drain Point of View

Equation (4) of the previous section equates the drift current at any point x along the channel to the total mobile charge times the velocity of the carriers at that point. In this section, we will examine the drift current as a function of gate voltage at the source and drain end of the channel of the MOS device. All modeling results shown were obtained at $V_{DS} = V_{DSAT}$. The saturation drain voltage was calculated using (8).

A. Source Point of View

At the source end of the device the drift current is

$$I = WC_{ox} V'_G v(0) \tag{9}$$

where $v(0)$ is the velocity of carriers at the source. We know the channel charge at the source very accurately. In fact, it has been shown in [9] that we can measure that channel charge without assuming a threshold voltage. We do not know the velocity of carriers at the source accurately; however, one can quickly realize that the maximum current available for a given gate dielectric capacitance and gate drive occurs when the *carriers are velocity saturated at the source.*

By differentiating (9) we find that the small-signal transconductance is given by

$$g_m = WC_{ox}\left[v(0) + V'_G \frac{\partial v(0)}{\partial V'_G}\right]. \tag{10}$$

The maximum transconductance for a MOS device occurs when the carriers are velocity saturated at the source end. This value is

$$g_m = WC_{ox} v_{SAT}. \tag{11}$$

One should note that under this extremely high lateral field, the gradual channel approximation breaks down. We use the equation, however, only for the purpose of qualitative physical explanations.

1) Scaling Channel Length: As the lateral field increases the velocity of carriers at the source increases to a limit of v_{SAT}. First-order theory predicts that the lateral field and consequently the velocity of carriers increases linearly with gate drive and the reciprocal of channel length. Because of

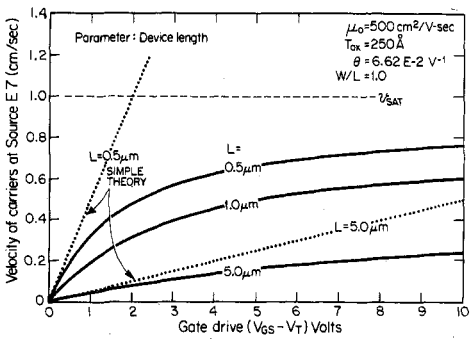

Fig. 6. The velocity of carriers at the source versus gate drive with $V_{DS} = V_{DSAT}$ and channel length as a parameter. These results are derived from (7)-(9).

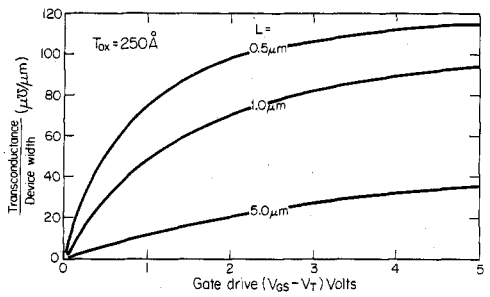

Fig. 7. Transconductance versus gate drive with channel length as a parameter derived from the model with $V_{DS} = V_{DSAT}$.

the velocity saturation limit, the velocity of carriers begins to deviate significantly from first-order theory. Fig. 6 derived from our model shows the velocity of carriers at the source versus gate drive with $V_{DS} = V_{DSAT}$, for various channel lengths. Note that at high lateral field (which corresponds to high gate drives or short channel lengths), the velocity of carriers does not increase linearly with gate drive. Also note that at high lateral field, the velocity of carriers does not increase linearly with the reciprocal of channel length for a given gate drive.

By understanding the deviation of the velocity of carriers at the source with increasing lateral field from first-order theory, one begins to see why the transconductance does not follow this theory. Fig. 7 shows a plot of transconductance versus gate drive with channel length as a parameter calculated using our model. One immediately notices that the transconductance does not increase linearly, as we scale the channel length, or with increasing gate drive. Second, for a fixed gate drive, $WC_{ox}v_{SAT}$, which in this case is 138 μS, is the maximum attainable transconductance as the channel length is reduced. Finally, for a finite channel length the transconductance approaches only a fraction of the maximum $WC_{ox}v_{SAT}$ for large gate drive. This point will be discussed in Section III-B.

2) Scaling Dielectric Thickness: It has been shown by several authors that the mobility of carriers is reduced with increasing normal field [15], [18]. In our model this is accounted for by a reduction in μ_0, the low field mobility parameter, and an increase in θ, the normal field dependence fitting parameter. The reduction in mobility corresponds to a reduction in slope of the velocity versus lateral field curve as well as an increase in the critical field for velocity saturation (see Fig. 2). Because the normal field increases with the reciprocal of dielectric thickness and with gate drive, we expect deviation from first-order theory for the velocity of carriers at the source and hence the transconductance.

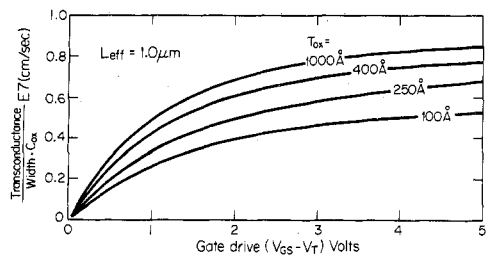

Fig. 8. Transconductance versus gate drive with the gate dielectric thickness as a parameter derived from the model with $V_{DS} = V_{DSAT}$.

To understand the dependence of transconductance on the dielectric thickness we first look at the case with a very low lateral field. This case corresponds to a long-channel device with the electric field at the source much less than the critical field for velocity saturation. Under this condition for a given gate dielectric thickness the transconductance increases less than linearly with gate drive V'_G due to the mobility reduction with increasing normal field. For a given gate drive the transconductance increases less than linearly with dielectric capacitance again due to the reduction of mobility which reduces the velocity of carriers at the source. A second case occurs when the lateral field at the source is equal to the critical field such that the carriers at the source are velocity saturated. Under this condition the transconductance is equal to $WC_{ox}v_{SAT}$ and, therefore, linearly increases with gate capacitance and remains constant with gate drive.

Device geometries of practical interest fall somewhere between these two extremes. Fig. 8 shows a plot of calculated transconductance using our model, normalized by the dielectric capacitance versus gate drive with the dielectric thickness as a parameter for a 1.0-μm channel-length device. Fig. 9 shows the experimental results displaying the same general trends. At low gate drives the transconductance increases less than linearly with gate drive and with gate capacitance. This corresponds to the low lateral field case. At higher gate drives we find that the transconductance versus gate drive becomes a constant which is a fraction of the maximum transconductance available in a MOS device. In this medium field case, we see that the transconductance is still increasing less than linearly with gate capacitance. To better understand this phenomenon we turn to the drain point of view.

B. Drain Point of View

At the drain end of the MOS device the drift current can be approximated by

$$I = WC_{ox}(V'_G - V_{DSAT})v_{SAT} \tag{12}$$

at $V_{DS} = V_{DSAT}$ assuming no bulk charge effect on the threshold voltage. It is interesting to note that at the drain end we know the velocity of carriers to be the saturated velocity; however, we do not know accurately the mobile charge content. This is the inverse of the known quantities at the source end.

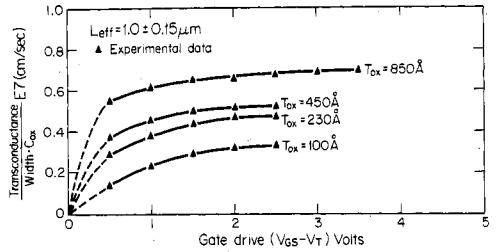

Fig. 9. Experimentally measured transconductance versus gate drive with the gate dielectric thickness as a parameter. All effective device lengths were between 0.85 and 1.15 μm and $V_{DS} \geq V_{DSAT}$.

By differentiating (12) with respect to gate voltage we obtain

$$g_m = WC_{ox}\left[1 - \frac{\partial V_{DSAT}}{\partial V_G}\right] v_{SAT}. \tag{13}$$

We can see that as the change in V_{DSAT} with respect to gate voltage goes from the first-order theory value of 1 toward 0, we approach the maximum transconductance of the MOS device. It should be noted that the carriers are always velocity saturated at the drain end of the channel when $V_{DS} \geq V_{DSAT}$ even for very long-channel devices. Therefore, unlike the source end where we studied the velocity of carriers to understand transconductance, at the drain end we must investigate the change in V_{DSAT} with respect to gate voltage.

Rewriting (13)

$$g_m = KWC_{ox}v_{SAT}, \quad 0 < K < 1. \tag{14}$$

In the previous section we saw that, for a given channel length and dielectric thickness, as the gate drive became large the transconductance approached a fraction of the maximum $WC_{ox}v_{SAT}$. Because we have an analytical expression for V_{DSAT}, (8), we can differentiate it and take the limit as the gate drive gets large. This yields a value for K as a function of device length and the dielectric thickness dependent parameters μ_0 and θ.

$$K = \frac{1}{1 + \dfrac{2v_{SAT}L\theta}{\mu_0}}. \tag{15}$$

From the simple expression in (15) we can make several qualitative observations. First, we note that as the channel length is reduced to zero the devices will exhibit its maximum transconductance. This is an expected result. Second, one can see that as the low field mobility is reduced it takes a shorter device to approach the maximum transconductance level. This observation is explained by realizing that as the low field mobility is reduced it takes a higher lateral field for the carriers to reach velocity saturation at the source. Third, as θ is increased corresponding to a greater reduction of velocity with normal field it takes a shorter device to have K approach 1. Finally, if θ is set equal to zero, corresponding to ignoring the normal field effect on mobility, K would equal 1 for all device lengths as the gate drive was made very large. Experiment has shown this not to be the case. Therefore, we have demonstrated the importance of including the dependence of velocity on normal field to understand the ultimate limits of the MOS device.

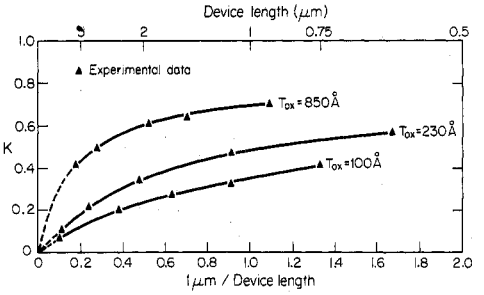

Fig. 10. The fraction of the maximum transconductance available in a MOS device K versus the reciprocal of channel length with the gate dielectric thickness as a parameter. This experimental data demonstrates that it takes a shorter device to achieve the maximum transconductance as one reduces the dielectric thickness.

The K factor calculated in (15) is subject to some question since the gate drive must be much larger than E_cL for the limit taking step to be valid. In state-of-the-art devices, these quantities are on the same order. Therefore, we believe that additional effects, which were not included in our model may also contribute to the fact that the transconductance is less than its maximum for large gate drive. Fig. 10, however, which is a plot of experimentally measured values for K versus the reciprocal of channel length with the dielectric thickness as a parameter, verifies the observations described in this subsection.

C. Summary

In the last section we looked at transconductance as a function of three parameters, namely, gate dielectric thickness, channel length, and gate drive. We saw that as the channel length was reduced there was an increase in the lateral field resulting in an increase in transconductance. We noted that the increase in transconductance was less than linear with the reciprocal of channel length and approached a maximum as the carriers at the source approached velocity saturation. For a constant gate drive, as the dielectric thickness was reduced the normal field increased causing a reduction in mobility. This in turn resulted in a less than linear dependence of transconductance on dielectric capacitance. Finally, as the gate drive was increased both the lateral and normal field increased, assuming the device is saturated. The increasing lateral field tends to increase the transconductance while the increasing normal field tends to decrease it. These compensating effects cause the transconductance to approach a constant equal to a fraction of the maximum of $WC_{ox}v_{SAT}$ as the gate drive gets large. The fraction of the maximum transconductance K qualitatively depends on the device parameters of low field mobility, channel length, normal field dependence of mobility, and the saturation velocity of channel carriers as shown in (15).

IV. Circuit Performance

When using a MOS device in a circuit, one generally is constrained to a particular power supply voltage. Therefore, in this section we have fixed the power supply voltage to study the effects of scaling on circuit performance. We chose a 2-V power supply and assume a 0.5-V threshold voltage yielding a 1.5-V gate drive. This relatively low voltage is chosen to try to minimize high field effects and thus determine the limits

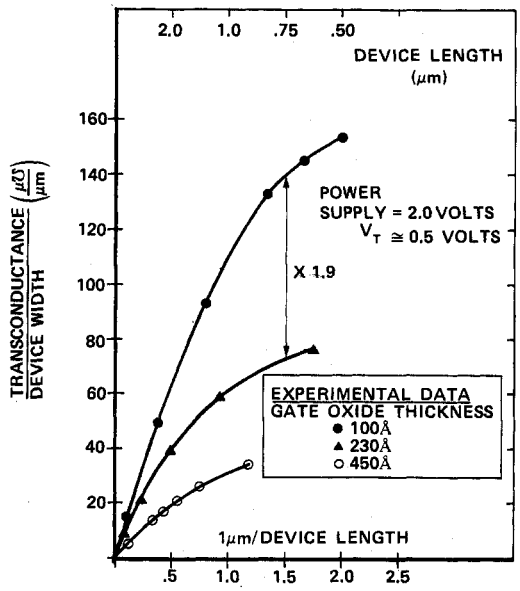

Fig. 11. Measured transconductance versus the reciprocal of channel length with a power supply voltage of 2.0 V. Gate dielectric thicknesses are 100, 230, and 450 Å.

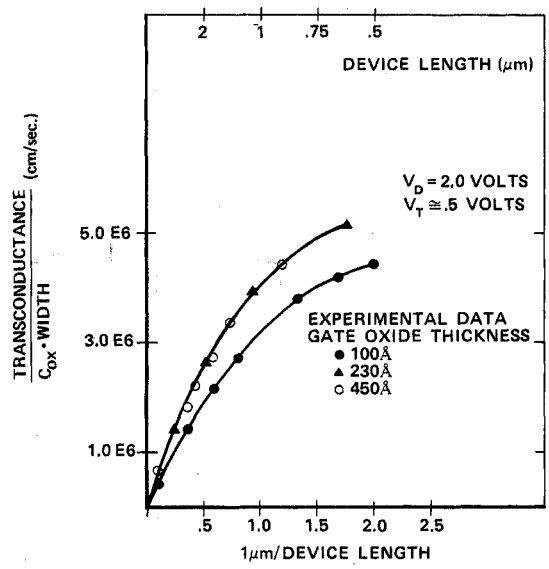

Fig. 12. Measured normalized transconductance by C_{ox} versus the reciprocal of channel length with a power supply voltage of 2.0 V. Gate dielectric thicknesses are 100, 230, and 450 Å.

of the benefits of scaling device geometries on circuit performance. With a larger power supply voltage, the same trends of deviation of transconductance from first-order theory will be observed at larger device geometries.

We divide circuits into two general classes, namely gate capacitance and parasitic capacitance limited circuits. An example of a circuit whose load capacitance is dominated by gate dielectric capacitance is a densely packed logic gate. An example of a circuit whose load capacitance is dominated by parasitic (other than gate dielectric) capacitance is a large bus line driver.

In keeping with the spirit of this paper, where we are trying to identify trends rather than produce absolutely accurate results, we propose that to speed up gate capacitance limited circuits one should maximize the quantity

$$\frac{g_m}{C_G} = \frac{g_m}{W \cdot L \cdot C_{ox}}. \tag{16}$$

On the other hand, for parasitic capacitance circuits, since one cannot in principle change the loading, one should maximize the transconductance. Strictly speaking, a large-signal transconductance, defined as an average small-signal transconductance over the voltage range of interest, should be used in (16). For a qualitative picture, however, we used the small-signal transconductance with the device biased in saturation.

In Fig. 11 we are plotting transconductance versus the reciprocal of channel length with the gate dielectric thickness as a parameter. First we notice that the transconductance approaches a constant as we decrease the channel length. This effect is caused by the channel carriers approaching velocity saturation at the source. Second, we notice that the transconductance increases with a reduction in gate dielectric thickness less than linearly. To explain this point we normalize the data in Fig. 11 by the dielectric capacitance. The result is shown in Fig. 12. Scaling the dielectric thickness from 450 to 230 Å caused no measurable change in the normalized transconductance. This result occurs since the power supply voltage was chosen low enough that the mobility reduction with normal field was not observable. In reducing the dielectric thickness to 100 Å, however, the increased normal field causes a degradation in the normalized transconductance. The magnitude of the degradation decreases with decreasing channel length because the carriers are approaching velocity saturation at the source. Under this condition the effect of the mobility reduction with normal field on transconductance is reduced.

With the understanding of the effect of scaling device geometries on transconductances, we can make some suggestions to improve circuit performance. For gate capacitance limited circuits one does not want to reduce the gate dielectric thickness, because the transconductance does not increase linearly with dielectric capacitance for practical device geometries. Instead, one should scale the channel length. One must realize, however, that the transconductance is nearly constant with decreasing channel length (for "short" devices) and the speed improvement will go as $1/L$. For parasitic capacitance limited circuits one should reduce the dielectric thickness. Although the increase in transconductance may be less than linear with dielectric capacitance, it remains significant for device geometries of practical interest as demonstrated in Fig. 11.

V. Conclusions

In this paper we have developed a simple model which includes the effect of high vertical and lateral fields on the saturation voltage, and the transconductance at that voltage, for a MOS transistor. This model predicts that the saturation voltage decreases from that predicted by first-order theory as we reduce the channel length. It also predicts that the saturation voltage increases toward the first-order theory value as we decrease the dielectric thickness.

The model predicts the following observations concerning transconductance.

1) For low vertical and lateral fields, the transconductance is proportional to gate drive $1/L$ and $1/T_{ox}$.

2) Given a channel length and gate drive such that the lateral field is less than E_C at the source, the transconductance increases less than linearly with $1/T_{ox}$ due to the reduction in mobility with normal field.

3) Given a dielectric thickness and gate drive, the transconductance increases less than linearly with $1/L$ to a maximum of $WC_{ox}v_{SAT}$. The maximum occurs when the channel carriers are velocity saturated at the source end of the device.

4) Given a channel length and dielectric thickness, the transconductance increases less than linearly with gate drive to a maximum of $KWC_{ox}v_{SAT}$. The increasing gate drive increases both normal and lateral fields. The increasing normal field increases the critical field for velocity saturation to occur at the source. This "negative feedback effect" causes the transconductance to approach a fraction of the maximum achievable in a MOS transistor.

A) Given a dielectric thickness, K approaches 1 as we reduce the channel length.

B) Given a channel length, K is reduced as we reduce the dielectric thickness. Therefore, it takes a shorter device for K to approach 1 as the dielectric thickness is reduced.

All of the trends predicted by our model have been experimentally verified by results presented in this paper or in [5], [17].

With this model we have tried to convey a physical understanding of the effect of scaling channel lengths and dielectric thicknesses on transconductance. We suggested that for gate capacitance limited circuits one should reduce the channel length, and for parasitic capacitance limited circuits one should reduce the dielectric thickness to improve circuit performance.

Acknowledgment

The authors would like to thank D. A. Antoniadis for his critical reading and helpful suggestions in the preparation of this manuscript.

References

[1] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. SC-9, p. 256, Oct. 1974.

[2] P. K. Chatterjee, W. R. Hunter, T. C. Holloway, and Y. T. Lin, "The impact of scaling laws on the choice of n-channel and p-channel for MOS VLSI," *IEEE Electron Device Lett.*, vol. EDL-1, p. 220, Oct. 1980.

[3] Y. A. El-Mansy, "MOS device and technology constraints in VLSI," *IEEE Trans. Electron Devices*, vol. ED-29, p. 567, Apr. 1982.

[4] B. Hoefflinger, H. Sibbert, and G. Zimmer, "Model and performance of hot-electron MOS transistor for VLSI," *IEEE Trans. Electron Devices*, vol. ED-26, p. 513, Apr. 1979.

[5] C. G. Sodini and J. L. Moll, "The effect of thin gate dielectrics on scaling MOS devices," in *IEDM Tech. Dig.*, p. 103, 1982.

[6] R. M. Swanson and J. D. Meindl, "Ion-implanted complimentary MOS transistor in low voltage circuits," *IEEE J. Solid-State Circuits*, vol. SC-7, p. 146, Apr. 1972.

[7] R. R. Troutman, "Subthreshold design consolidation for insulated gate field-effect transistors, *IEEE J. Solid-State Circuits*, vol. SC-9, p. 55, Apr. 1974.

[8] T. I. Kamins and R. S. Muller, "Statistical considerations in MOS calculation," *Solid-State Electron.*, vol. 10, p. 423, 1967.

[9] C. G. Sodini, T. W. Ekstedt, and J. L. Moll, "Charge accumulation and mobility in thin dielectric MOS transistors," *Solid-State Electron.*, vol. 25, p. 833, 1982.

[10] G. Baccarani and M. R. Wordeman, "Transconductance degradation in thin-oxide MOSFET's, in *IEDM Tech. Dig.*, p. 278, 1982.

[11] Y. Tsividis, "Moderate inversion in MOS devices," *Solid-State Electron.*, vol. 25, p. 1099, 1982.

[12] A. S. Grove, *Physics and Technology of Semiconductor Devices*. New York: Wiley, 1967.

[13] R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*. New York, Wiley, 1977.

[14] G. J. Chern *et al.*, "A new method to determine MOSFET channel length," *IEEE Electron Device Lett.*, vol. EDL-1, p. 170, Sept. 1980.

[15] A. G. Sabnis and J. T. Clemens, "Characterization of the electron mobility in the inverted ⟨100⟩ Si surface," in *IEDM Tech. Dig.*, vol. 18, 1979.

[16] J. A. Cooper, Jr., and D. F. Nelson, "Measurement of the high field drift velocity of electrons in inversion layers on silicon," *IEEE Electron Device Lett.*, vol. EDL-2, p. 171, July 1981.

[17] P. K. Ko, "Hot-electron effects in MOSFET's," Ph.D. dissertation, Univ. of California, Berkeley, 1982.

[18] S. C. Sun and J. D. Plummer, "Electron mobility in inversion and accumulation layers on thermally oxidized silicon surfaces," *IEEE Trans. Electron Devices*, vol. ED-27, p. 1497, Aug. 1980.

[19] F. N. Trofinenkoff, "Field-dependent mobility analysis of the field effect transistor," *Proc. IEEE*, vol. 53, p. 1765, Nov. 1965.

[20] D. M. Caughey and R. E. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proc. IEEE*, vol. 55, p. 2192, Dec. 1967.

[21] R. W. Coen and R. S. Muller, "Velocity of surface carriers in inversion layers on silicon," *Solid-State Electron.*, vol. 23, p. 35, 1980.