An Engineering Model for Short-Channel MOS Devices

KAI-YAP TOH, MEMBER, IEEE, PING-KEUNG KO, AND ROBERT G. MEYER, FELLOW, IEEE

Abstract — **An engineering model for short-channel MOS devices which includes the effect of carrier drift velocity saturation is described. Based on a piecewise carrier drift velocity model, simplified expressions for the dc drain current I_D, the small-signal transconductance g_m, and the output conductance g_{ds} in the saturation region are derived. For a given gate voltage, the expressions depend only on the threshold voltage V_T and the dimensions of the device whose desired values are normally known.**

I. INTRODUCTION

MUCH has been published on the various short-channel effects in MOS devices from the device physics point of view, but there is no analytical dc model of short-channel MOS devices which circuit designers can use for hand calculations, such as in estimating the dc bias conditions and ac parameters. Common practice is to extend the use of the long-channel square-law model for the short-channel devices, but this approach gives considerable discrepancies in both the dc and ac solutions between the calculated and simulated results and can lead to incorrect conclusions in regard to circuit performance. On the other hand, elaborate models provided in circuit simulators involve a large number of model parameters whose values must be determined or derived from device measurements, often with the help of automated parameter extraction tools. It is difficult to develop an intuitive understanding of the device electrical behavior from such a list of parameters and it is not suitable for hand calculations. A simple engineering model for short-channel MOS devices that relates the terminal voltages to the drain current, much like the well-known square-law I–V relationship for the long-channel devices, is therefore needed. The purpose of this paper is to describe such a model. The model provides a simple picture for the essential electrical behaviors of the short-channel MOS device from the circuit designer's perspective. It will not only be useful for circuit design and analysis, it will also be useful to device designers who need to relate device and process parameters to circuit parameters.

Based on a piecewise carrier drift velocity model, simple closed-form I–V relationships between the terminal voltages and drain current I_D, and the transconductance g_m are derived, with $V_{GS}-V_T$ as the independent variable. I_D and g_m essentially depend on only the device width, the channel length, the device threshold voltage V_T, the gate oxide thickness t_{ox}, and the source/drain junction depth x_j, whose desired values are normally known. While I_D and g_m are sufficient for most digital MOS circuit calculations, the device g_{ds} in the saturation region is often required in analog MOS circuit calculations. Accurate modeling of g_{ds} is difficult, even in elaborate computer models. However, based on engineering approximations, an expression for estimating g_{ds} is derived and is found to correlate fairly well with experimental data.

The dc model presented here is a modification and extension of that discussed in [1]. In this paper, a different carrier mobility model [3] is incorporated and the electrical channel length L_e is used when the device is biased into saturation. This leads to a more accurate prediction of the device g_m. The mobility model can be used to estimate the effective mobility μ_{eff} of a device under any gate bias without taking device measurements. The use of L_e leads to a slightly more complicated calculation procedure but is necessary to extend the validity of the model down to 1-μm effective channel length L_{eff}. This is because the drain depletion width X_d takes up a significant portion of the L_{eff}, particularly for devices with channel length ≤ 1 μm and operated at high V_{DS}. To make the model simple for hand calculations and yet reasonably accurate, it is necessary to introduce empirical constants to replace complicated terms in the expressions that have only second-order effects on the accuracy of the model. Devices with L_{eff} from 3 μm down to 1 μm and with different t_{ox} and x_j have been characterized and found to be in good agreement with the model.

II. CARRIER DRIFT VELOCITY AND μ_{eff} MODELS

The main reason that the electrical characteristics of short-channel devices deviate from those of the long-channel devices is the dependence of the carrier drift velocity v

Manuscript received June 18, 1987; revised November 5, 1987. This work was supported in part by the National Science Foundation under Grant DCI-8501488, by a grant from the Semiconductor Research Corporation, and by an IBM Faculty Development Award to P.-K. Ko.

K.-Y. Toh was with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720. He is now with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598.

P.-K. Ko and R. G. Meyer are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720.

IEEE Log Number 8821667.

0018-9200/88/0800-0950$01.00 ©1988 IEEE

on the longitudinal channel field E. For long-channel devices, $v = \mu_{\text{eff}} E$, where the effective carrier mobility μ_{eff} is assumed to be a constant. For short-channel devices, μ_{eff} is no longer a constant and is a function of the transverse field E_t in the inversion layer. Increasing the transverse field will reduce the value of μ_{eff}. Also, v is no longer directly proportional to E due to high field effects. Increasing E will reduce v. A detailed examination of the relationship between v and E is therefore in order.

Several expressions used for v have been discussed in [1]. The carrier drift velocity model depicted in (1) is commonly used in computer modeling:

$$v_1 = \frac{\mu_{\text{eff}} E}{\sqrt{1 + \left(\frac{E}{E_0}\right)^2}}$$

where

$$E_0 \equiv \frac{v_{\text{sat}}}{\mu_{\text{eff}}}. \tag{1}$$

Unfortunately, the use of (1) does not lead to analytical expressions for the I–V characteristics of the short-channel devices. However, simple analytical expressions can be obtained if the carrier drift velocity is approximated by the piecewise model depicted in (2):

$$v = \begin{cases} \dfrac{\mu_{\text{eff}} E}{1 + \dfrac{E}{E_c}}, & \text{for } E \leqslant E_c \\ v_{\text{sat}}, & \text{for } E \geqslant E_c \end{cases}$$

where

$$E_c \equiv \frac{2 v_{\text{sat}}}{\mu_{\text{eff}}}. \tag{2}$$

The value of μ_{eff} is crucial to the accuracy of the device model. As reported in [2] and [3], μ_{eff} can be estimated as follows:

$$\frac{1}{\mu_{\text{eff}}} = \frac{1}{\mu_1} + \frac{1}{\mu_2}$$

where

$$\mu_2 = \mu_1 \left(E_1 / E_{t,\text{eff}}\right)^n. \tag{3}$$

μ_1, E_1, and n are empirical constants listed in Table I. $E_{t,\text{eff}}$ is the effective transversed field at the surface and may be approximated as follows (see Appendix A):

$$E_{t,\text{eff}} \simeq \frac{V_{GS} - V_T}{6 t_{ox}} + \frac{V_T + V_a}{3 t_{ox}} \tag{4}$$

where, semi-empirically, $V_a \simeq 0.5$ V for typical n^+ polysilicon gate devices.

TABLE I
COEFFICIENTS FOR μ_1, E_1, AND n TO BE USED WITH (3)

	μ_1 (cm^2/V-s)	E_1 (MV/cm)	n
electron	670	0.67	1.6
hole	290	0.35	1.0

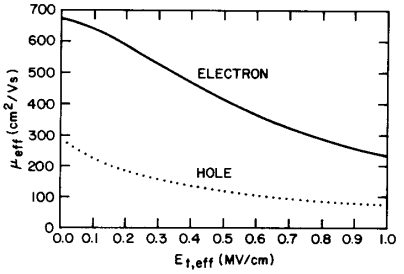


Fig. 1. Computed μ_{eff} versus $E_{t,\text{eff}}$ for electron and hole carriers.

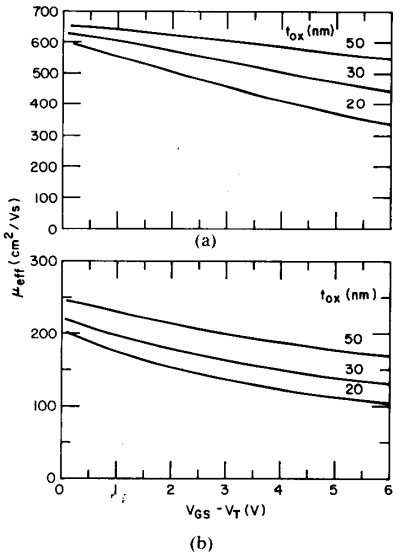


Fig. 2. μ_{eff} of electron and hole carriers versus $V_{GS} - V_T$ for $t_{ox} = 50$, 30, and 20 nm. (a) NMOS devices. (b) PMOS devices.

Calculated values of μ_{eff} versus $E_{t,\text{eff}}$ for both electron and hole carriers are shown in Fig. 1. Variations of μ_{eff} versus $V_{GS} - V_T$ for typical values of t_{ox} are also shown in Fig. 2(a) and (b) for the NMOS and PMOS devices, respectively. These curves may be used to estimate the value of μ_{eff} needed in the dc drain current expression to be discussed next.

III. LARGE-SIGNAL BEHAVIOR OF SHORT-CHANNEL MOS DEVICES

The engineering model to be described here uses the piecewise carrier drift velocity model. It can be shown [1] that the drain currents I_D in the ohmic and saturation

TABLE II
FABRICATED NMOS AND PMOS DEVICE PARAMETERS

Device	Type	L_{eff} (μm)	t_{ox} (nm)	V_T (V)	x_j (μm)
A	NMOS	3.0	50	0.71	0.5
B	NMOS	2.0	30	0.65	0.3
C	NMOS	1.0	20	0.73	0.3
D	PMOS	3.3	50	-0.70	0.5
E	PMOS	2.3	30	-0.7	0.5
F	PMOS	1.3	20	-0.57	0.5

regions are given as follows:

$$I_D = \begin{cases} \dfrac{\mu_{\text{eff}} C_{ox} W}{L_{\text{eff}}} \dfrac{1}{1+\dfrac{V_{DS}}{E_c L_{\text{eff}}}} \left(V_{GS} - V_T - \dfrac{1}{2} V_{DS}\right) V_{DS}, \\ \qquad\qquad \text{for } V_{DS} \leqslant V_{D\,\text{sat}} \\ v_{\text{sat}} C_{ox} W \left(V_{GS} - V_T - V_{D\,\text{sat}}\right) \equiv I_{D\,\text{sat}}, \\ \qquad\qquad \text{for } V_{DS} \geqslant V_{D\,\text{sat}} \end{cases} \tag{5}$$

where

C_{ox} gate capacitance per unit area,
W device width,
L_{eff} effective channel-length between the source and the drain,
V_{DS} drain-to-source voltage,
V_{GS} gate-to-source voltage,
$I_{D\,\text{sat}}$ drain current in the saturation region, and
$V_{D\,\text{sat}}$ saturation drain voltage, defined here as the drain voltage at which the carrier drift velocity saturates.

By equating the I_D expressions in (5) at $V_{DS} = V_{D\,\text{sat}}$, it can be shown that

$$V_{D\,\text{sat}} = (1-K)(V_{GS} - V_T) \tag{6}$$

where

$$K \equiv \frac{1}{1 + E_c L_e / (V_{GS} - V_T)}$$

and

$$L_e \equiv L_{\text{eff}} - X_d. \tag{7}$$

In the above equations, L_e is the device electrical channel length and X_d is the depletion width into the channel from the drain when $V_{DS} > V_{D\,\text{sat}}$. X_d may be calculated as follows [4]:

$$X_d \simeq \frac{1}{A} \ln\left[\frac{A\left(V_{DS} - V_{D\,\text{sat}}\right) + E_d}{E_c}\right]$$

where

$$\frac{E_d}{E_c} = \sqrt{1 + \left[\frac{A\left(V_{DS} - V_{D\,\text{sat}}\right)}{E_c}\right]^2} \tag{8}$$

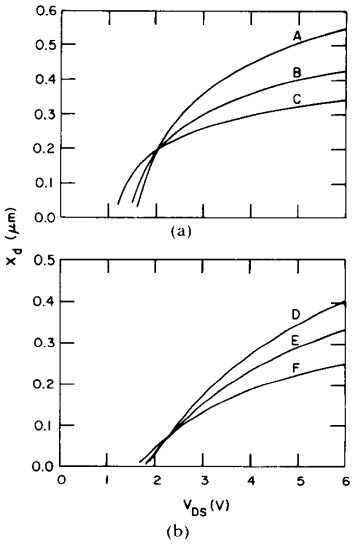


Fig. 3. X_d versus V_{DS} for the MOS devices listed in Table II. (a) NMOS devices. (b) PMOS devices.

and

$$A^2 \simeq \frac{\epsilon_{ox}}{\epsilon_{si}} \frac{1.5}{x_j t_{ox}}$$

in which 1.5 is a semi-empirical constant. $I_{D\,\text{sat}}$ can also be expressed in terms of K and $V_{GS} - V_T$ by substituting (6) into (5), which leads to

$$I_{D\,\text{sat}} = K v_{\text{sat}} C_{ox} W (V_{GS} - V_T). \tag{9}$$

This is a useful expression for $I_{D\,\text{sat}}$ in which all the short-channel effects are modeled by the factor K. As will be shown later, K may be regarded as a constant under certain conditions.

Calculated values of X_d at $V_{GS} - V_T = 2$ V for both the NMOS and PMOS devices listed in Table II are plotted in Fig. 3(a) and (b), respectively (the μ_{eff} model described in Section II has been used and $v_{\text{sat}} = 7 \times 10^6$ cm/s is assumed for both the NMOS and PMOS devices). Devices listed in Table II are fabricated with essentially the same P-well CMOS process except that the gate oxide thicknesses, implant doses, and source/drain drive-in times are different. The value of K in (7) may be calculated with the value of X_d estimated from Fig. 3 or with one round of

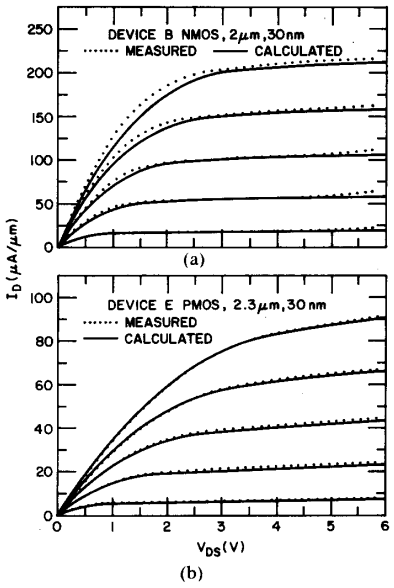


Fig. 4. Measured and calculated device $I-V$ characteristics for $V_{GS}-V_T$ from 1 to 5 V at steps of 1 V. (a) An NMOS device. (b) A PMOS device.

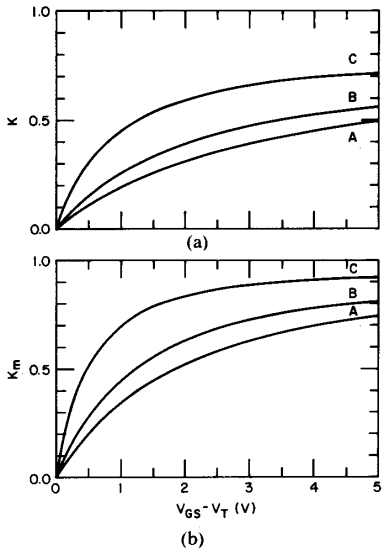


Fig. 5. Calculated (a) K and (b) K_m versus $V_{GS}-V_T$ for the NMOS devices listed in Table II.

iteration between (6) and (8). As a rule of thumb, the typical range of X_d is 0.2–0.3 μm for a 1-μm channel-length NMOS device with a 20-nm gate oxide thickness and $V_{DS}-V_{D\,\mathrm{sat}}$ in the range of 1–3 V. As an illustration, the measured and calculated $I-V$ characteristics of a 2-μm NMOS device (device B) and a 2.3-μm PMOS device (device E) are shown in Fig. 4(a) and (b), respectively. The agreements are better than ±5 percent over most of the voltage range. The mismatches at low V_{DS} are due to the lower carrier drift velocity modeled in v [1].

IV. Device Transconductance g_m

For devices biased into saturation, it can be shown that from the dc model described (see Appendix B)

$$g_m = K_m v_{\mathrm{sat}} C_{ox} W \tag{10a}$$

where

$$K_m \simeq K(2-K) - B_m K(1-K) \tag{10b}$$

and

$$B_m \simeq \frac{p\left(E_{t,\mathrm{eff}}/E_1\right)^n}{1+\left(E_{t,\mathrm{eff}}/E_1\right)^n}. \tag{10c}$$

Empirically, $p \simeq 1.2$ for NMOS devices and 1.0 for PMOS devices. The first term in (10b) is the dominant factor. The second term is important only for high $V_{GS}-V_T$ and devices with $L_{\mathrm{eff}} \leqslant 1$ μm. Note that when $E_c L_e/(V_{GS}-V_T) \ll 1$, $K \to 1$, $K_m \to 1$, and from (9) and (10a)

$$I_{D\,\mathrm{sat}} \to v_{\mathrm{sat}} C_{ox} W (V_{GS}-V_T)$$

and

$$g_m \to v_{\mathrm{sat}} C_{ox} W. \tag{11}$$

These are the ultimate relationships for very short-channel MOS devices. However, for practical devices with finite channel length and operating under a practical bias condition, $K < K_m < 1$. As an example, the calculated variations of K and K_m with $V_{GS}-V_T$ for the NMOS devices listed in Table II are plotted in Fig. 5(a) and (b) respectively. It should be pointed out that K and K_m can have different values when a different value of v_{sat} is assumed, as is evident from (10a). However, other considerations, such as the accuracy of I_D and g_{ds}, lead to the conclusion that $v_{\mathrm{sat}} = 7\times10^6$ cm/s is the best choice for this engineering model. For the 1-μm NMOS device, the variations of K and K_m with respect to $V_{GS}-V_T$ may be roughly divided into three regions:

1) $V_{GS}-V_T < 0.3$ V: a linear region where the "square-law" long-channel model approximately holds;
2) $0.3\ \mathrm{V} < V_{GS}-V_T < 2.0$ V: a nonlinear region, where K and K_m should be calculated from (7) and (10); and
3) $V_{GS}-V_T > 2.0$ V: a flat region where $K \simeq 0.65$ and $K_m \simeq 0.88$.

In most digital MOS circuits, during turn-on and turn-off transients, $V_{GS}-V_T$ is typically greater than 2 V and therefore K_m and K may be treated as constants. In analog MOS circuits, $V_{GS}-V_T$ is typically within the nonlinear region and K_m and K should be calculated from the formulas. For the 3-μm device, the linear region is extended to approximately 0.7 V. The variations consist of essentially only the linear and nonlinear regions for $V_{GS}-V_T \leqslant 5.0$ V. It is apparent that even at 3-μm channel

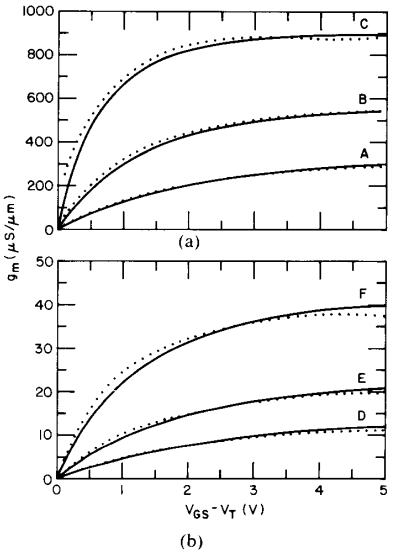


Fig. 6. Measured and calculated g_m versus $V_{GS}-V_T$, at $V_{DS}=5$ V for the MOS devices listed in Table II. (a) NMOS devices. (b) PMOS devices.

length, the electrical characteristics do not follow the long-channel behavior.

To compare the accuracy of the model, calculated and measured g_m are shown in Fig. 6(a) and (b), respectively, for the devices listed in Table II. As shown, the agreement is better than ± 6 percent over most of the gate voltage range for channel lengths from 3 μm down to 1 μm.

V. Second-Order Effects in I_D

The primary effect on I_D in the saturation region is due to the increase of X_d with V_{DS}. Second-order effects that affect I_D are not included in the dc model depicted in (5). This is because the deviations are small in comparison with I_D itself. However, the deviation may be significant in comparison with ΔI_D at high V_{DS} and constant V_{GS} and thus will add to the value of g_{ds}. It is therefore necessary to include these effects in the formulation of g_{ds}. Two second-order effects will be discussed, namely the drain-induced barrier lowering (DIBL) effect on V_T and the substrate current (I_{SUB}) effect on I_D.

For short-channel MOS devices, V_T may depend on V_{DS}, especially at high values of V_{DS}, due to the DIBL effect [5] (this can be observed as a horizontal shift in the subthreshold conduction characteristics in a plot of $\log I_D$ versus V_{GS} curves for different V_{DS}). When the DIBL effect is small, V_T may be approximated by

$$V_T = V_{T0} - \eta V_{DS}. \tag{12}$$

V_{T0} is the device threshold voltage at $V_{DS}=0$. As V_{DS} increases, V_T will decrease and $V_{GS}-V_T$ will increase, causing I_D to increase.

I_{SUB} is normally much smaller than $I_{D\text{sat}}$. Typically, $I_{\text{SUB}}/I_D < 10^{-2}$ at low $V_{GS}-V_T$ and high V_{DS}, which is the worst case. A more significant effect of I_{SUB} on I_D is the fact that I_{SUB} tends to reduce the substrate bias, which reduces V_T, and in turn increases I_D. g_{ds} can therefore increase significantly even at moderately low I_{SUB}. Unfortunately, there is no analytical form which relates this effect of I_{SUB} to I_D as yet. In addition, lightly doped drain (LDD) structures and retrograded substrate or well doping profiles reduce the effects of hot-electron current and I_{SUB}, making g_{ds} dependent on the device structure. Consequently, it is difficult to estimate g_{ds} of a short-channel MOS device accurately at large V_{DS}. Nevertheless, an approximate expression for g_{ds} can be derived which will be discussed next.

VI. Device Output Conductance g_{ds}

When the effect of I_{SUB} is omitted, it can be shown from (5) that the device conductance in saturation, denoted by g_{ds}^0, can be approximated by (see Appendix C)

$$g_{ds}^0 \simeq v_{\text{sat}} C_{ox} W K^2 \left(\frac{E_c}{E_d}\right) + \eta g_m. \tag{13}$$

For a well-designed device, the first term is the dominant factor and is due to X_d. The second term is due to the effect of DIBL and is expected to be proportional to g_m as $\Delta I_D = g_m \Delta(V_{GS}-V_T) = g_m \Delta V_T$ at constant V_{GS}. A large η will increase g_{ds}^0 and is obviously not desirable. As the channel length is reduced, g_m is increased and it is therefore important to reduce the DIBL effect or the value of η in order to maintain a reasonable device voltage gain (g_m/g_{ds}).

Equation (13) normally would give a good estimate of g_{ds} biased at low $V_{DS}-V_{D\text{sat}}$. At high $V_{DS}-V_{D\text{sat}}$, the substrate current effect should be included when better accuracy is needed. However, the calculation procedures are more involved and will be discussed in Appendix C. The final result is

$$g_{ds} \simeq g_{ds}^0 + \frac{\beta I_{\text{SUB}}}{(V_{DS}-V_{D\text{sat}})}. \tag{14}$$

From device measurement data, β varies from 2.0 to 3.0 for conventional devices (without the LDD structure). It is clear from the above expression that both η and I_{SUB} should be minimized to reduce g_{ds} (or to increase r_{ds}) and to maximize voltage gain. Ignoring the effect of I_{SUB}, (10) and (13) predict that the maximum achievable dc gain of a 1-μm channel-length device biased well into saturation is about 29 when $\eta=0$ and drops to about 18 when $\eta=0.02$, a reduction of almost 40 percent.

Fig. 7 shows the computed and measured g_{ds} of a 2-μm NMOS device (device B) versus V_{DS} for V_{DS} around and greater than $V_{D\text{sat}}$. The calculated data are based on $\eta=0.02$ and $\beta=2$. The agreement is within ± 25 percent.

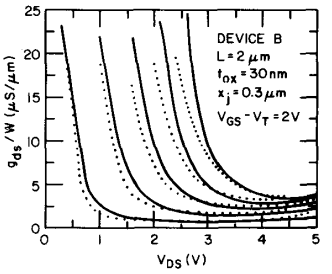


Fig. 7. Measured and calculated g_{ds} versus V_{DS} at $V_{GS}-V_T=2$ V for an NMOS device.

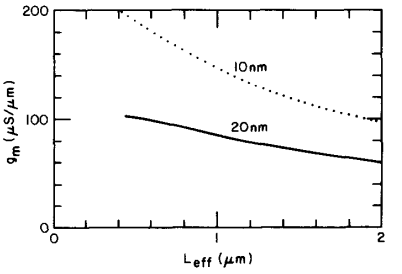


Fig. 9. Variations of g_m versus L_{eff} for NMOS devices with $t_{ox}=20$ nm (solid line) and 10 nm (dotted line).

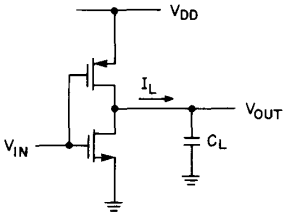


Fig. 8. Example: a CMOS inverter driving a capacitive load.

VII. Applicability to Long-Channel Regime

The closed-form expressions derived are a consequence of the simplified carrier drift velocity model depicted in (2). There is no restriction imposed on the device channel length L, assuring that the dc model described is applicable to devices with long channel length as well. For long-channel devices

$$\frac{V_{GS}-V_T}{E_cL_e} \ll 1$$

which leads to

$$K \simeq \frac{(V_{GS}-V_T)}{E_cL_e}.$$

Substituting the above and (2) into (5) and (6) will confirm the square-law I–V relationship. The regime for the so-called "long-channel" length can now be quantified by setting $(V_{GS}-V_T)/E_cL_e \leq 0.1$. For example, for long-channel NMOS devices, $\mu_{eff} \simeq 700$ cm^2/V·s, and at $V_{GS}-V_T=1$ V, L_e should be greater than 5 μm. It is therefore not surprising to observe that the 3-μm device (discussed in Section IV) does not exhibit long-channel device behavior.

VIII. Practical Examples

To demonstrate the use of the model, the peak discharge and charging current I_L of the CMOS inverter as shown in Fig. 8 are to be calculated. A different CMOS technology is deliberately chosen to check the accuracy of the proposed model. Device parameters are extracted from vendor supplied SPICE2 level-2 MOS models. For the NMOS device, $t_{ox}=25$ nm, $V_T=0.8$ V, $L_{eff}=1.05$ μm, and $x_j=0.3$ μm, and for the PMOS device, $t_{ox}=25$ nm, $V_T=-0.8$ V, $L_{eff}=1.05$ μm, and $x_j=0.4$ μm.

Case 1: Peak Discharge Current

Assume $V_{DD}=3.4$ V. $V_{in}=3.4$ V, $V_{out}=3.4$ V, the NMOS is in saturation, and the PMOS is in the OFF state. Consider the NMOS device, $V_{GS}-V_T=3.4-0.8=2.6$ V. From Fig. 2(a), $\mu_{eff} \simeq 510$ cm^2/V·s for $t_{ox}=25$ nm. From Fig. 3(a), $X_d \simeq 0.25$ μm, approximating it in between device B and device C at $V_{DS}=3.4$ V. From (7), $K \simeq 0.542$. Using these as initial values in (8) and (7) gives $X_d=0.316$ μm and $K=0.563$. The peak discharging current per unit width can now be calculated from (9):

$$I_L/W = 0.563\times 7\times 10^6 \times \frac{3.9\times 8.85\times 10^{-14}}{250\times 10^{-10}} \times 2.6$$
$$= 142\ \mu\text{A}/\mu\text{m}.$$

Case 2: Peak Charging Current

$V_{in}=0$ V, $V_{out}=0$ V, the NMOS is in the OFF state, and the PMOS is in saturation. Consider the PMOS device, $|V_{GS}-V_T|=3.4-0.8=2.6$ V. From Fig. 2(b), estimate that $\mu_{eff} \simeq 150$ cm^2/V·s for $t_{ox}=25$ nm. From Fig. 3(b), $X_d \simeq 0.17$ μm and from (7), $K \simeq 0.24$. Using these as initial values in (8) and (7) gives $X_d=0.13$ μm and $K=0.232$. Using (9) again, the peak charging current $I_L/W=58$ μA/μm. The corresponding values from BSIM [6] simulations are 141 and 58 μm/μA for the discharge and charging currents, respectively.

As another example, the short-channel model described can be used to evaluate the first-order trends in the electrical characteristics of submicrometer-channel devices. As illustrations, Figs. 9 and 10 show the trends in g_m versus L_{eff}, and K, K_m versus L_{eff}, respectively, for NMOS devices with $L_{eff} \leq 2$ μm and for $t_{ox}=20$ and 10 nm. As expected, g_m increases with decreasing L_{eff}, and K and K_m approach 1 as L_e approaches 0. It should be noted that g_m does not double in value as L_{eff} is reduced from 1 to 0.5 μm while keeping t_{ox} constant.

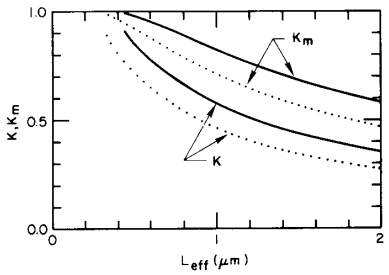


Fig. 10. Variations of K and K_m versus L_{eff} for NMOS devices with t_{ox} = 20 nm (solid line) and 10 nm (dotted line).

IX. Conclusions

The above discussions demonstrated that I_D and g_m of short-channel devices can be estimated for a given gate voltage and a given set of V_T, t_{ox}, and x_j whose typical values are usually known for a given process. g_{ds} can be estimated with the additional knowledge of η and β. Often, an educated guess in the value of η and β is sufficient. The model described is also useful for evaluating the impact of a particular short-channel technology on circuit performance before any device is available. Despite the substantial deviations from the long-channel device behavior, several simple equations are sufficient to predict the I_D, g_m, and g_{ds} of conventional short-channel MOS devices to within reasonable accuracy. For conventional MOS devices, the accuracy of the model for I_D and g_m is typically better than ± 6 percent, and that for g_{ds}, which is sensitive to the device structure and the substrate doping profile, is typically better than ± 25 percent with properly selected values of η and β. These accuracies are certainly within the acceptable range for hand-calculation purposes.

Appendix A
Derivation of $E_{t,\text{eff}}$

From [7], $E_{t,\text{eff}}$ may be approximated by

$$E_{t,\text{eff}} = \frac{1}{\epsilon_{si}}\left(\frac{Q_{INV}}{2} + Q_B\right)$$

where Q_{INV} is the inversion charge and Q_B is the bulk charge. But

$$Q_{INV} = C_{ox}(V_{GS} - V_T)$$

and

$$V_T = -V_a + \frac{Q_B}{C_{ox}} \rightarrow Q_B = C_{ox}(V_T + V_a).$$

Thus

$$E_{t,\text{eff}} = \frac{\epsilon_{ox}}{\epsilon_{si} t_{ox}}\left[\frac{V_{GS} - V_T}{2} + (V_T + V_a)\right] \simeq \frac{V_{GS} - V_T}{6t_{ox}} + \frac{V_T + V_a}{3t_{ox}}$$

as

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$$

and

$$\frac{\epsilon_{ox}}{\epsilon_{si}} \simeq \frac{1}{3}.$$

Appendix B
Derivation of g_m

From (5) and in saturation

$$g_m = \frac{\partial I_D}{\partial V_{GS}} = v_{\text{sat}} C_{ox} W\left(1 - \frac{\partial V_{D\,\text{sat}}}{\partial V_{GS}}\right). \tag{B1}$$

From (6)

$$\frac{\partial V_{D\,\text{sat}}}{\partial V_{GS}} = (1 - K) - (V_{GS} - V_T)\frac{\partial K}{\partial V_{GS}}. \tag{B2}$$

But

$$\frac{\partial K}{\partial V_{GS}} = (1 - K)^2\left(\frac{1}{E_c L_e} - \frac{a}{E_c}\frac{\partial E_c}{\partial V_{GS}} + \frac{a}{L_e}\frac{\partial X_d}{\partial V_{GS}}\right). \tag{B3}$$

From (2)–(4)

$$\frac{\partial E_c}{\partial V_{GS}} = -\frac{E_c}{\mu_{\text{eff}}}\frac{\partial \mu_{\text{eff}}}{\partial V_{GS}} = \frac{nE_c B_x}{6t_{ox} E_{t,\text{eff}}}$$

where

$$B_x \equiv \frac{(E_{t,\text{eff}}/E_1)^n}{1 + (E_{t,\text{eff}}/E_1)^n} \tag{B4}$$

and from (8)

$$\frac{\partial X_d}{\partial V_{GS}} = -\frac{1}{E_d}\left(\frac{\partial V_{D\,\text{sat}}}{\partial V_{GS}} + \frac{V_{DS} - V_{D\,\text{sat}}}{E_c}\frac{\partial E_c}{\partial V_{GS}}\right). \tag{B5}$$

Substituting (B3)–(B5) into (B2) gives

$$\frac{\partial V_{D\,\text{sat}}}{\partial V_{GS}} = \frac{(1 - K) - (1 - K)K\left[1 - \frac{V_{GS} - V_T}{E_c}\frac{\partial E_c}{\partial V_{GS}}\left(1 + \frac{V_{DS} - V_{D\,\text{sat}}}{E_d L_e}\right)\right]}{1 - K^2\left(\frac{E_c}{E_d}\right)}. \tag{B6}$$

The denominator in (B6) is due to the weak interaction between X_d and $V_{D\,\mathrm{sat}}$. Since typically $K<1$ and $E_c/E_d<1$ so that $K^2E_c/E_d \ll 1$, the value of the denominator is close to 1 to the first order of approximation. Its effect is to increase $\partial V_{D\,\mathrm{sat}}/\partial V_{GS}$ and hence decrease g_m at high field which is significant only for $L_{\mathrm{eff}} \leqslant 1$ μm and at high $V_{GS}-V_T$. For hand-calculation purposes, the following simplifications are made:

$$\begin{aligned} g_m &= v_{\mathrm{sat}}C_{ox}W\left(1-\frac{\partial V_{D\,\mathrm{sat}}}{\partial V_{GS}}\right) \\ &\simeq v_{\mathrm{sat}}C_{ox}W\left[1-(1-K)+K(1-K)(1-B_m)\right] \\ &= v_{\mathrm{sat}}C_{ox}W\left[K(2-K)-B_mK(1-K)\right] \end{aligned}$$

where

$$B_m \equiv \frac{V_{GS}-V_T}{E_c}\frac{\partial E_c}{\partial V_{GS}}\left(1+\frac{V_{DS}-V_{D\,\mathrm{sat}}}{E_dL_e}\right) \simeq pB_x.$$

Empirically, $p \simeq 1.2$ for NMOS devices and 1.0 for PMOS devices.

Appendix C
Derivation of g_{ds}

Again, from (5) and (12)

$$g_{ds}^0 = \frac{\partial I_D}{\partial V_{DS}} = v_{\mathrm{sat}}C_{ox}W\left(\eta-\frac{\partial V_{D\,\mathrm{sat}}}{\partial V_{DS}}\right). \tag{C1}$$

Similar to the derivation procedures in Appendix B, it can be shown that

$$\frac{\partial V_{D\,\mathrm{sat}}}{\partial V_{DS}} = \frac{\eta(1-K)-\eta(1-K)K\left[1-\dfrac{V_{GS}-V_T}{\eta E_dL_e}-\dfrac{V_{GS}-V_T}{E_c}\dfrac{\partial E_c}{\partial V_{GS}}\left(1+\dfrac{V_{DS}-V_{D\,\mathrm{sat}}}{E_dL_e}\right)\right]}{1-K^2\left(\dfrac{E_c}{E_d}\right)}. \tag{C2}$$

Thus from (C1) and again neglecting the denominator in (C2)

$$\begin{aligned} g_{ds}^0 &= v_{\mathrm{sat}}C_{ox}W\left(\eta-\frac{\partial V_{D\,\mathrm{sat}}}{\partial V_{DS}}\right) \\ &\simeq v_{\mathrm{sat}}C_{ox}WK^2\left(\frac{E_c}{E_d}\right)+\eta g_m. \end{aligned}$$

Substrate current I_{SUB} may be approximated by [8]

$$I_{\mathrm{SUB}} \simeq \frac{A_I}{B_I}(V_{DS}-V_{D\,\mathrm{sat}})I_{D\,\mathrm{sat}}e^{-B_I/A(V_{DS}-V_{D\,\mathrm{sat}})}$$

where A_I and B_I are related to the impact ionization coefficient of carriers α as follows:

$$\alpha = A_Ie^{-B_I/E}. \tag{C3}$$

TABLE III
COEFFICIENTS OF A_I AND B_I TO BE USED IN (C3)

	electron	hole
A_I	7.85×10^5	2.23×10^6
B_I	1.21×10^6	2.00×10^6

The values of A_I and B_I are estimated from [9] and listed in Table III. It is therefore reasonable to expect that the contribution to g_{ds} is roughly proportional to $I_{\mathrm{SUB}}/(V_{DS}-V_{D\,\mathrm{sat}})$, to the first order of approximation. From device measurement data, the proportionality factor, denoted by β, may vary from 2.0 to 3.0 for conventional devices (without the LDD structure). When $V_{DS}-V_{D\,\mathrm{sat}}$ is large, the g_{ds} of the conventional short-channel MOS device may be estimated by

$$g_{ds} \simeq g_{ds}^0 + \frac{\beta I_{\mathrm{SUB}}}{(V_{DS}-V_{D\,\mathrm{sat}})}.$$

Acknowledgment

The authors are grateful to the technical staff of the Microfabrication Laboratory, University of California, Berkeley, for their technical assistance and to J. Boernberg who provided the SPICE2 level-2 MOS model parameters and BSIM simulation results used in Section VIII.

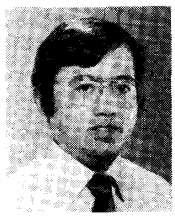
References

[1] C. G. Sodini, P.-K. Ko, and J. L. Moll, "The effect of high fields on MOS device and circuit performance," *IEEE Trans. Electron Devices*, vol. ED-31, no. 10, pp. 1386–1393, Oct. 1984.
[2] B. Hoeneisen and C. A. Mead, "Current-voltage characteristics of small size MOS transistors," *IEEE Trans. Electron Devices*, vol. ED-19, pp. 382–383, Mar. 1972.
[3] M.-S. Liang, J. Y. Choi, P.-K. Ko, and C. Hu, "Inversion-layer capacitance and mobility of very thin gate-oxide MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-33, no. 3, pp. 409–412, Mar. 1986.
[4] P.-K. Ko, "Hot electron effects in MOSFET's," Ph.D. dissertation, Univ. of Calif., Berkeley, 1982.
[5] R. R. Troutman, "VLSI limitations from drain-induced barrier lowering," *IEEE Trans. Electron Devices*, vol. ED-26, no. 4, pp. 461–468, Apr. 1979.
[6] B. J. Sheu, D. L. Scharfetter, P.-K. Ko, and M. C. Jeng, "BSIM: Berkeley short-channel IGFET model for MOS transistors," *IEEE J. Solid-State Circuits*, vol. SC-22, pp. 558–566, Aug. 1987.
[7] A. G. Sabins and T. J. Chemens, "Characterization of the electron mobility in the inverted ⟨100⟩ Si surface," in *IEDM Tech. Dig.*, 1979, p. 18.

[8] T. Y. Chan, P.-K. Ko, and C. Hu, "A simple method to characterize substrate current in MOSFET's," *IEEE Electron Device Lett.*, vol. EDL-5, no. 12, pp. 505–507, Dec. 1984.
[9] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. New York: Wiley, 1981, p. 47.

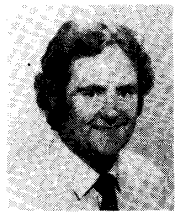
Kai-Yap Toh (M'76–S'80–M'87) was born in Singapore in 1950. He received the B.E. (with Honors) degree from Monash University, Australia, in 1974, the M.Sc. degree from the University of Pennsylvania in 1982, and the Ph.D. degree from the University of California, Berkeley, in 1986, all in electrical engineering.

He served in the Singapore Armed Forces from 1974 to 1976. He was a civilian R&D Engineer with the Defense Ministry of Singapore from 1976 to 1980 and was the head of an R&D Department in 1980. He was a Resident Consultant with the AT&T Bell Laboratories at Holmdel in 1985 working on high-frequency analog MOS circuits. He joined the IBM Research Division at the Thomas J. Watson Research Center, Yorktown Heights, NY, in December 1986 and is working on high-speed BJT circuits. His technical interests are in the design of high-speed analog and digital IC's, IC technologies, and the characterization of IC devices.



Ping-Keung Ko received the Ph.D. degree in electrical engineering in 1982 from the University of California at Berkeley.

He spent 1982 and 1983 at Bell Laboratories, Holmdel, NJ, working on high-speed MOS technologies. In 1984 he joined the University of California at Berkeley where he is an Associate Professor of Electrical Engineering and Computer Sciences. His current research areas include high-speed VLSI technologies and devices, and MOS device modeling for circuit simulation. He has authored or co-authored one book and over 70 research journal articles.



Robert G. Meyer (S'64–M'68–SM'74–F'81) was born in Melbourne, Australia, on July 21, 1942. He received the B.E., M.Eng. Sci., and Ph.D. degrees in electrical engineering from the University of Melbourne in 1963, 1965, and 1968, respectively.

In 1968 he was employed as an Assistant Lecturer in Electrical Engineering at the University of Melbourne. Since September 1968 he has been employed in the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, where he is now a Professor. His current research interests are in integrated-circuit design and device fabrication. He has acted as a consultant on electronic circuit design for numerous companies in the electronics industry. He is co-author of the book *Analysis and Design of Analog Integrated Circuits* (Wiley, 1984) and editor of the book *Integrated Circuit Operational Amplifiers* (IEEE Press, 1978).

Dr. Meyer is a past President of the Solid-State Circuits Council of the IEEE and is a former Associate Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITS and of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS.