AT&T Bell Laboratories Technical Journal
Vol. 63, No. 7, September 1984
Printed in U.S.A.

Velocity-Saturated Characteristics of Short-Channel MOSFETs

By G. W. TAYLOR*

(Manuscript received October 3, 1984)

A theory is developed for the I-V characteristics of metal oxide semiconductor field-effect transistors (MOSFETs) when the channel fields are sufficiently high to cause appreciable saturation of the carrier drift velocity. The full velocity-field curve for bulk silicon is used with the base value adjusted to account for surface scattering effects. Use of this form gave the best fit to experimental data. Using some simple expansions to reduce the rather complex integral produces a useful analytic result, which gives a continuous description from the square law results for long-channel devices throughout the whole range of velocity-saturated operation in short-channel devices. For the first time the electron temperature has been introduced as the parameter, which increases the channel charge at pinch-off, decreases the saturation voltage, and increases the channel field at the pinch-off point as the current (and hence bias voltages) is increased. The effects of series resistance and surface roughness scattering are incorporated into the analytic formulation. We compare the results with experimental submicron devices and find excellent agreement.

I. INTRODUCTION

As the metal oxide semiconductor field-effect transistor (MOSFET) evolves towards submicron channel dimensions it is found that, below gate lengths of 5 μm, its performance is modified by the effects of hot-electron scattering on carrier transport. When a device reaches a gate dimension of 0.5 μm, its complete "on" region is dominated by velocity

* AT&T Bell Laboratories.

saturation. Almost all descriptions of these phenomena in the literature are based on two-dimensional numerical solutions of total device operation.[1–4] Such an approach provides insight into device operation and the nature of its physics, but it is of little practical use from a designer's point of view, since it can realistically consider only a single device at a time. Generally, circuit simulation programs must use empirical fitting routines to generate device characteristics in the velocity-saturated mode because there is no analytical solution of this behavior. Some estimates of velocity-saturated conduction have been based on the simplified model of a constant mobility up until the saturation velocity is achieved;[5,6] although some of these results have been useful over certain particular voltage ranges with parameter fitting, the interpretation of the physics has been lacking. In fact, academically, a more rigorous treatment of the problem greatly helps us visualize the ultimate limits of device miniaturization. A proper description of device performance in velocity saturation must be established as a natural extension of the device characteristics before velocity saturation becomes significant. However, since an analytical solution can be found only for certain special cases, one is usually forced to assume a two-section model as just described. Such an approach unavoidably leads to discontinuities in the incremental parameters and, particularly, in regions of crossover. In addition, it precludes prediction of the pinch-off field, which is fixed at $\mathscr{E} = \mathscr{E}_c$ in this model. The other missing feature in previous work has been the electron temperature. Since the carrier transport is due to electrons moving with their saturated drift velocity and hence at elevated temperatures, the electron temperature must be an integral part of the representation.

In this paper we use the velocity-field relationship that most nearly fits the experimental data as a starting point to develop a model. The charge transport description is derived from basic principles and incorporates the electron temperature as a natural part of the solution. It is shown in detail how the velocity-saturated characteristics are a natural extension of the constant mobility case. We will consider the case of both the triode and the saturation regions; for the saturation case the details of channel length modulation under velocity-saturated conditions are derived in Appendix A. The effects of series contact resistance are incorporated into the theory, and the results are compared with data over a wide range of parameters.

II. THEORY

2.1 The velocity-field relationship

When the drift velocity can no longer be considered a linear function of the electric field, modifications to the metal oxide semiconductor

(MOS) theory must be introduced. Many functional forms have been used to characterize the decrease in mobility at higher electric-field strengths.[7–13] For the case of bulk Si, Conwell has shown that the assumption of a Maxwellian distribution with electron temperature, T_e, for the symmetrical part of the electron distribution function yields predictions that agree with the experiment.[14] The result for the energy-loss rate, $B(T_e)$, for phonon scattering is

$$B(T_e) = q\mu_0 \left(\frac{T_0}{T_e}\right)^{1/2} \left(\frac{v_s}{\mu_0}\right)^2 \frac{(T_e - T_0)}{T_0}, \tag{1}$$

where μ_0 is the zero field mobility, T_0 is the lattice temperature, v_s is the saturation velocity of electrons, and q is the electronic charge. Now it has been shown that the general momentum and energy relations are

$$j = qn\mu(T_e)\mathscr{E} + q\frac{d}{dx}[nD(T_e)] \tag{2a}$$

$$j\mathscr{E} = nB(T_e) + \frac{j}{q}\delta(T_e)k\frac{dT_e}{dx}, \tag{2b}$$

respectively, where $\mathscr{E}$ is the electric-field strength; $\mu(T_e)$ is the temperature, or field-dependent mobility; j is the electronic current density; k is Boltzmann's constant; n is the electron density; and $\delta(T_e)kT_e$ is the average kinetic energy transported per electron.[15] Since we are dealing with the MOS surface along which the scattering mechanisms are not well known, (1) may not be appropriate; we will continue to use it, however, to expedite the analysis. The second term on the right-hand side of (2b) is equivalent to $dS(T_e)/dx$. Here $S(T_e)$ is the flux of energy in the positive x direction; we have taken $S = -(j/q)\delta(T_e)kT_e$ only and ignored the small contribution $-K(T_e)[(dT_e)/(dx)]$ because of the thermal conductivity of the electrons. Note that δ typically is approximately a constant for a particular relaxation time relation and for acoustic phonon scattering $\delta \simeq 2$.[15] Equations (2a) and (2b) have three unknowns—$\mathscr{E}$, n, and T_e—so that the third relation of Poisson's equation for the MOS channel region is required to obtain a complete solution. The most familiar form of this equation is the one-dimensional charge equation (the gradual channel approximation) for the MOS channel, which is

$$q\int_0^\infty n\,dy = C_o(V_{GS} - V_T - V), \tag{2c}$$

where V is related to $\mathscr{E}$ by

$$\int_0^x \mathscr{E}\,dx' = V$$

and C_o is the oxide capacitance. These equations are solved by simplification of (2a). The right-hand side of (2a) is the sum of the drift and diffusion components of the device current, the drift component being characterized by a drift velocity,

$$v = \mu(T_e)\mathscr{E}. \tag{3}$$

It is well known for conduction above the threshold voltage that over most of the MOS channel, drift is the dominant component, and hence, only the first term on the right-hand side of (2a) needs to be retained. We can use this result in (2b); then by neglecting the second term on the right-hand side of (2b) and substituting from (1) for $B(T_e)$, we obtain the result

$$q\mu_0 \left(\frac{T_0}{T_e}\right)^{1/2} \left(\frac{T_e}{T_0} - 1\right) \left(\frac{v_s}{\mu_0}\right)^2 = q\mathscr{E}^2\mu.$$

Solving for T_e we find

$$T_e = T_0 \left\{ \frac{(\mathscr{E}/\mathscr{E}_c)^2}{2} \frac{\mu}{\mu_0} + \left[\frac{(\mathscr{E}/\mathscr{E}_c)^4}{4} \left(\frac{\mu}{\mu_0}\right)^2 + 1 \right]^{1/2} \right\}^2, \tag{4}$$

where $\mathscr{E}_c$ is the critical field parameter,

$$\mathscr{E}_c = v_s/\mu_0. \tag{5}$$

We can then use this result to evaluate the relative importance of the neglected term in (2b). Using (2c) in (2a) we can write the current-field relation for the MOS channel, which is

$$I = \mu(\mathscr{E})C_o(V_{GS} - V_T - V)\mathscr{E}. \tag{6a}$$

If we then use a typical mobility

$$\mu = \mu_0(T_0/T_e)^{1/2} \tag{6b}$$

such as one might find in a bulk crystal, we can use (4) and (6a) to evaluate the relative importance of the terms in (2b), and we find

$$\frac{\frac{j}{q}\delta(T_e)k\frac{dT_e}{dx}}{j\mathscr{E}} \simeq 2\delta(T_e) \frac{\frac{kT_0}{q}\left[1 + \left(\frac{\mathscr{E}}{\mathscr{E}_c}\right)^2\right]}{(V_{GS} - V_T - V)} \left(\frac{\mathscr{E}}{\mathscr{E}_c}\right)^2. \tag{6c}$$

This ratio is small until we are close to pinch-off in the channel or unless $\mathscr{E} \geq \mathscr{E}_c$. The restriction on this term comes from neglecting the diffusion term in (2a). We will assume that (4) is valid throughout the MOS channel.

Equation (4) has been derived without reference to a specific function, $\mu(\mathscr{E})$. We therefore can use (4) to determine T_e, and we must still determine $\mu(\mathscr{E})$. To determine $\mu(\mathscr{E})$, we could just assume the

bulk relation (6b), which would then yield the results

$$\mu = \frac{\mu_0}{\sqrt{1 + \left(\frac{\mathscr{E}}{\mathscr{E}_c}\right)^2}} \tag{7a}$$

and

$$T_e = T_0[1 + (\mathscr{E}/\mathscr{E}_c)^2]. \tag{7b}$$

However, there has been much speculation that this form is inadequate for the Si surface mobility,[7-14] and a variation of this form that has been used fairly widely is

$$\mu(\mathscr{E}) = \frac{\mu_0}{\left[1 + \left(\frac{\mathscr{E}}{\mathscr{E}_c}\right)^B\right]^{1/B}}, \tag{7c}$$

where B was determined by Caughey and Thomas[13] to be $B \simeq 1.1$. To illustrate the variation of $\mu(\mathscr{E})$, plots were generated and are shown in Fig. 1 for values of B ranging from 1 to ∞. The important parameters—μ_0, $\mathscr{E}_c$, and v_s—are identified in the figure.

Because the determination of B at this stage in our understanding of MOS surface physics is unavoidably experimental, we must, at the outset, choose a value for B that will allow a physical solution to be found. This is necessary because the use of an undetermined B leads to a mathematically hopeless situation. The vindication of this approach ultimately will come from an unambiguous determination of

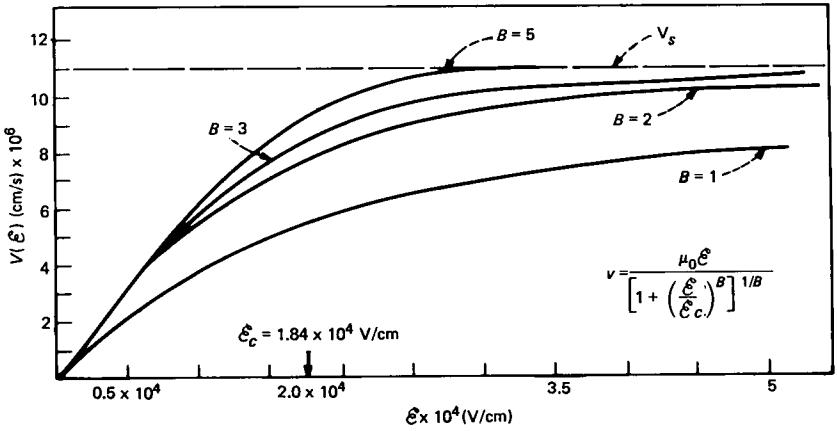

Fig. 1—Variation of $v(\mathscr{E})$ for the formula (7c) with values of B ranging from 1 to ∞. A value of $\mu_0 = 650$ cm^2/V-s was used here simply to clarify the variation of v with electric field.

the parameters μ_0, $\mathscr{E}_c$, v_s, and B from the experimental data. Since it is suspected that $B = 2$ for N channel devices (which is of main interest here), we will give the greatest attention to the treatment of this form. In our final comparison with experiment we will demonstrate that the value of $B \simeq 2$ is a good one.

In the remainder of the paper we will discuss the effects of a field-dependent mobility, $\mu(\mathscr{E})$, on the MOS I-V characteristic and the resultant modifications to the conventional device laws ($B = \infty$) for $B = 2$ generally and with some mention of the case $B = 1$. The theory will show, as is well known in short-channel devices, that when the drift velocity of carriers approaches the saturated value, the current from the conventional case, is reduced substantially.

The theory will also show the well-known experimental result that the drain saturation voltage for the velocity-saturated case is substantially less than the conventional value. It is shown that the decrease in voltage is caused by the increase in the charge in the channel at pinch-off under hot-electron conditions resulting from the elevated electron temperature.

2.2 Device characteristics

2.2.1 Triode region

Figure 2 gives a device cross section, which shows terminal voltages and possible series resistances in the source and drain leads. When the general mobility relationship (7) is used, only for the special case of $B = 1$ can a closed form solution for the current-voltage characteristic be obtained. If we use (5) and (7b) for $B = 1$, the result shown by Hoeneisen[8] is

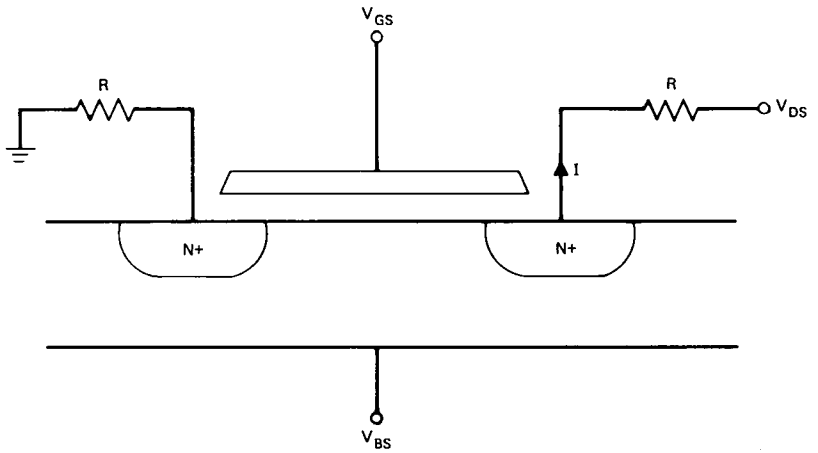

Fig. 2—Device cross section showing terminal voltages and possible parasitic series resistors.

$$I = \frac{\mu_0 C_o W}{L}\left[\frac{(V_{GS} - V_T)V_{DS} - V_{DS}^2}{1 + V_{DS}/\mathscr{E}_c L}\right]. \tag{8}$$

For the case of $B = 2$, the MOS device current, when approximated by the drift mechanism only, yields

$$I = \frac{\mu_0 W}{\left[1 + \left(\frac{\mathscr{E}}{\mathscr{E}_c}\right)^2\right]^{1/2}} C_o(V_{GS} - V_T - V)\mathscr{E}. \tag{9a}$$

By rearranging terms and changing variables, we obtain

$$\frac{IL}{\mu_0 WC_o} = \int_{V_{GS}-V_T-V_{DS}}^{V_{GS}-V_T} (z^2 - I'^2)^{1/2} dz, \tag{9b}$$

where

$$I' = \frac{I}{v_s WC_o}, \qquad z = V_{GS} - V_T - V.$$

Performing the integration, we find

$$I = \frac{\mu_0 WC_o}{2L}\left\{(V_{GS} - V_T)\sqrt{(V_{GS} - V_T)^2 - \left(\frac{I}{Wv_sC_o}\right)^2}\right.$$
$$- (V_{GS} - V_T - V_{DS})\sqrt{(V_{GS} - V_T - V_{DS})^2\left(\frac{I}{Wv_sC_o}\right)^2 - \left(\frac{I}{Wv_sC_o}\right)^2}$$
$$\left.\cdot \ln\left[\frac{V_{GS} - V_T + \sqrt{(V_{GS} - V_T)^2 - \left(\frac{I}{Wv_sC_o}\right)^2}}{V_{GS} - V_T - V_{DS} + \sqrt{(V_{GS} - V_T - V_{DS})^2 - \left(\frac{I}{Wv_sC_o}\right)^2}}\right]\right\}. \tag{10a}$$

Previously, it has been shown that (for the constant mobility case) it is possible to incorporate the effects of the bulk charge variation into V_T.[16] As the gate voltage increases into the range where velocity saturation becomes significant, the threshold voltage modulation by the drain voltage becomes relatively unimportant. We will therefore assume that we can continue to use the drain-voltage dependence of V_T that was developed for the constant mobility case. Equation (10a) describes the device characteristic exactly up until pinch-off has been achieved but, because of its implicit nature, it is of relatively little use. However, we produce useful results with some simple approximations. Consider the terms under the second square root sign. This quantity

must be positive to have a physically meaningful result, so that

$$\left(\frac{I}{Wv_sC_o}\right)^2 \leqslant (V_{GS} - V_T - V_{DS})^2.$$

Now the current on the left-hand side is a maximum and the right-hand side is a minimum at the pinch-off condition, and if the equality sign were obeyed at this point we would have

$$I = C_o W v_S \bar{Q}. \tag{10b}$$

In writing (10b) we have used the definition of $\bar{Q}$ as the channel charge at the pinch-off point, as we develop in Appendix A [see (83)]. From this result we see that the inequality rather than the equality is always obeyed, since (10b) requires that all of the charge be moving with the saturated velocity at the pinch-off point. This condition can never be reached because of the contribution of diffusion to the current flow. We therefore note that under all square root signs in (10) we will have positive quantities and also that the voltage term is generally much larger than the term I'. We may therefore make expansions of all square root quantities, and by retaining terms to the first order, we obtain

$$I \simeq \frac{1}{2}\mu_0 \frac{W}{L} C_o \left\{ (V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2 - I'^2 \ln \left[\frac{2(V_{GS} - V_T) - \dfrac{I'^2}{V_{GS} - V_T}}{2(V_{GS} - V_T - V_{DS}) - \dfrac{I'^2}{V_{GS} - V_T - V_{DS}}} \right] \right\}. \tag{10c}$$

For the moment we will neglect terms in I' in the argument of the logarithm without incurring significant error, and we may then solve the resulting quadratic equation for I to obtain

$$I = Wv_sC_o \cdot \left\{ \left[\left(\frac{\mathscr{E}_c L}{a} \right)^2 + \frac{(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2}{a} \right]^{1/2} - \frac{\mathscr{E}_c L}{a} \right\}, \tag{11a}$$

where

$$a \equiv \ln \left(\frac{V_{GS} - V_T}{V_{GS} - V_T - V_{DS}} \right). \tag{11b}$$

This expression for the current is very useful because it allows us to predict the two limiting forms of conduction that are always observed in a short-channel device. For small gate (drain voltages) the term in square brackets is small, and we may expand under the square

root sign to obtain

$$I \simeq \frac{\mu_0 C_o W}{L}\left[(V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}\right], \tag{12}$$

which is precisely the form of the constant mobility case that we expect to obtain when the effects of velocity saturation can be ignored. On the other hand, when this term becomes large, only this term needs to be retained, and we find

$$I \simeq \frac{v_s C_o W}{\sqrt{a}}\sqrt{(V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}}. \tag{13}$$

This equation applies to only a limited range, since typically there is only a small range of drain voltage in which velocity saturation is dominant before pinch-off occurs. This will become evident in our comparison of the exact and approximate solutions.

2.2.2 Saturation region

Traditionally, in MOSFET physics the terms operation in pinch-off and operation in saturation were interchangeable because it was generally thought that pinch-off was necessary for saturation to occur. For some time now it has been recognized that diffusion becomes important near the pinch-off point.[17] It is shown in Appendix A [see (79) through (83)] that the saturation voltage is reached when a particular ratio, $\bar{R}$, of drift to diffusion conduction is reached in the channel. $\bar{R}$ is derived in Appendix B [see (200) and (201)] in terms of the basic device parameters. The transfer of the conduction mechanism from drift to diffusion is the extent of any pinching effect, since beyond this point in the channel the drift component increases once again, the oxide field reverses in direction, and the current moves away from the surface into the bulk. Saturation of current occurs basically because of the occurrence of field reversal at some drain voltage. The saturation point and the field-reversal point are separated in potential in the channel by a potential, $\bar{Q}/C_o$, where $\bar{Q}$ is the channel charge (mobile charge) at the saturation point. For the constant mobility case the charge, $\bar{Q}$, is small, so that the potential difference between the two points is small in that case. Hence, there really was no need to discriminate between them. As velocity saturation becomes pronounced, the charge, $\bar{Q}$ (and thus voltage, $\bar{Q}/C_o$), grows considerably, leading to a substantial change in V_{SAT} (the drain saturation voltage) from its conventional value. Thus, the potential difference between the saturation point and the field-reversal point continues to grow, as does the physical separation between the two points, although the potential will increase more since the field is also increasing. From

this point on we will refer to the point in the channel where the diffusion current is $\bar{R}I$ as the saturation point or the pinch-off point, interchangeably. We will distinguish between it and the field-reversal point.

Under conditions of hot-electron flow (as in Appendix A) we will continue to use the Einstein relation

$$D(\mathscr{E}) = \frac{kT_e}{q}\,\mu(\mathscr{E}) \tag{14}$$

to relate the mobility and diffusion coefficients through the effective electron temperature, T_e. This has recently been shown to be a reasonable assumption if it is recognized that, for mobilities significantly below μ_0, the low field value, the electron temperature will be higher than the lattice temperature.[18] Also, according to (2a) we will consider the term

$$D(\mathscr{E})dQ/dx \gg Q\,\frac{dD(\mathscr{E})}{dx},$$

i.e., the diffusion constant is not a strong function of the electric field.

The result for the saturation voltage obtained in Appendix A [cf. (82)] is

$$\bar{V} \equiv V_{\mathrm{SAT}} = V_{\mathrm{GS}} - V_T - \left(\frac{1-\bar{R}}{\bar{R}}\right)\frac{kT_e}{q}\,\frac{(C_o + \bar{C}_s F)}{C_o}, \tag{15}$$

where F is the short-channel factor[16] at the threshold condition and $\bar{C}_s$ is the semiconductor depletion capacitance at the pinch-off point. Each quantity denoted by a bar will signify its value at the pinch-off point. From this result we can see that the charge at the pinch-off point now continues to grow as we move into the hot-electron regime, linearly with the electron temperature. Since we are using $B = 2$ and have shown for this case that T_e is given by (7b), then we have

$$\bar{Q} = \frac{(1-\bar{R})}{\bar{R}}\,(C_o + \bar{C}_s F)\,\frac{kT_0}{q}\left[1 + \left(\frac{\bar{\mathscr{E}}}{\mathscr{E}_c}\right)^2\right]. \tag{16}$$

The value of $\bar{\mathscr{E}}$, the electric field at the pinch-off position, is determined from the drift relation

$$\bar{\mathscr{E}} = \frac{I(1-\bar{R})}{W\mu\bar{Q}}. \tag{17a}$$

Using (17a) and (7a) in (16) we obtain

$$\left(\frac{\bar{\mathscr{E}}}{\mathscr{E}_c}\right)^2\left[1 + \left(\frac{\bar{\mathscr{E}}}{\mathscr{E}_c}\right)^2\right] = \left[\frac{(1-\bar{R})I}{W\mathscr{E}_c\mu_0\bar{Q}_0}\right]^2, \tag{17b}$$

which has the solution

$$\frac{\bar{\mathscr{E}}}{\mathscr{E}_c} = \left(\left\{\frac{1}{4} + \left[\frac{I(1-\bar{R})}{Wv_s\bar{Q}_0}\right]^2\right\}^{1/2} - \frac{1}{2}\right)^{1/2}, \tag{17c}$$

where

$$\bar{Q}_0 = \frac{(1-\bar{R})}{\bar{R}} \frac{kT_0}{q} (C_o + \bar{C}_s F)$$

is the value of charge at the pinch-off point in the absence of hot-electron effects.[19] This is an interesting result because it encompasses all modes of operation from the region of low currents without velocity saturation through the regime of high fields dominated by velocity saturation. For low currents we can expand the square root and (keeping only the first term) find that, without velocity saturation, we have

$$\bar{\mathscr{E}} \simeq \frac{I(1-\bar{R})}{W\mu_0\bar{Q}_0}. \tag{17d}$$

If the current (and therefore the channel field) is high (velocity saturation becomes dominant), then we have

$$\bar{\mathscr{E}} \simeq \sqrt{\frac{I(1-\bar{R})\mathscr{E}_c}{W\mu_0 Q_0}}, \tag{17e}$$

so that the field changes from a linear to a square root dependence upon the device current. We note that from (17d) and (17e) the transition from one to the other takes place at $\bar{\mathscr{E}} = \mathscr{E}_c$, as we would expect. We can also determine from (17c) that the device current that flows when $\bar{\mathscr{E}} = \mathscr{E}_c$ is

$$I = \sqrt{2}\,\frac{Wv_s\bar{Q}_0}{(1-\bar{R})}.$$

The interesting feature of this result is the independence of channel length. We should, thus, expect devices of all gate lengths to change to velocity-saturated behavior at the same value of channel current. If we now use (17c) in (16) we obtain

$$\bar{Q} = \frac{(1-\bar{R})}{\bar{R}} (C_o + \bar{C}_s F) \frac{kT_0}{q} \left[\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{I^2(1-\bar{R})^2}{(Wv_s\bar{Q}_0)^2}}\right], \tag{18a}$$

and again we can obtain the limiting cases

$$\bar{Q} = \bar{Q}_0$$

and

$$\bar{Q} = \frac{I(1 - \bar{R})}{Wv_s} \tag{18b}$$

for low and high currents, respectively. A useful approximation to (18a) is found to be

$$\bar{Q} = \bar{Q}_0 \sqrt{1 + \left[\frac{I(1 - \bar{R})}{Wv_s\bar{Q}_0}\right]^2}, \tag{18c}$$

or

$$\bar{Q} = \bar{Q}_0 + \frac{I(1 - \bar{R})}{Wv_s} \tag{18d}$$

expressions, which are mathematically more convenient. The approximation will introduce some error when the two terms under the root sign are equal, but it yields the correct value in the extremes. Using the saturation voltage (15), we may now determine the saturation current; using (8) and (11a) we obtain for the cases of $B = 1$ and $B = 2$, respectively,

$$I_{SAT} = \frac{\mu_0 C_o W}{2\bar{y}} \left[\frac{(V_{GS} - V_T)^2 - (\bar{Q}/C_o)^2}{1 + \dfrac{V_{GS} - V_T - \bar{Q}/C_o}{\mathscr{E}_c\bar{y}}}\right], \qquad B = 1 \tag{19a}$$

and

$$I_{SAT} = Wv_sC_o \cdot \left[\sqrt{\left(\frac{\mathscr{E}_c\bar{y}}{a}\right)^2 + \frac{(V_{GS} - V_T)^2 - (\bar{Q}/C_o)^2}{a}} - \frac{\mathscr{E}_c\bar{y}}{a}\right], \qquad B = 2, \tag{19b}$$

where $\bar{y}$ is the position of the pinch-off point in the channel. Since we do not have the equivalent to (6b) for the case of $B = 1$, we cannot determine T_e or, therefore, $\bar{Q}/C_o$; thus, we can proceed no further with that case.

Consider the case of $B = 2$. By substituting (18c) into (19b) and rearranging, we find

$$I^2 + 2\frac{(\mathscr{E}_c\bar{y}v_sC_oW)}{(1 + a)} I - \frac{(v_sC_oW)^2}{(1 + a)}\left[(V_{GS} - V_T)^2 - \left(\frac{\bar{Q}_0}{C_o}\right)\right] = 0, \tag{20a}$$

which has the solution

$$I_{SAT} = Wv_sC_o \cdot \left\{\left[\left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)^2 + \frac{(V_{GS} - V_T)^2 - (\bar{Q}_0/C_o)^2}{a^*}\right]^{1/2} - \frac{\mathscr{E}_c\bar{y}}{a^*}\right\}, \tag{20b}$$

where

$$a^* \equiv a + 1. \tag{20c}$$

By comparison with (11a) it can be seen that the expressions for the current in the triode region and saturation region are very similar. The currents are identical at the pinch-off position because the change from the parameter a to a^* exactly compensates for the change from $\bar{Q}/C_o$ to $\bar{Q}_0/C_o$. In using (18c) in (19b) we have omitted the factor $(1 - \bar{R})$. We must do this to obtain consistent results because (19b) was derived on the basis of drift alone.

As in the case of the triode region, (10b) can predict two limiting cases. For small gate voltages, the appropriate expansion of the square root yields

$$I \simeq \frac{\mu_0 C_o W}{2\bar{y}} \left[(V_{GS} - V_T)^2 - \left(\frac{\bar{Q}_0}{C_o} \right)^2 \right], \tag{21a}$$

as we would expect in the absence of hot-electron effects. On the other hand, in the extreme of large gate voltages we obtain the result

$$I = \frac{v_s C_o W}{\sqrt{a^*}} (V_{GS} - V_T). \tag{21b}$$

An interesting feature of this equation is the absence of $\bar{y}$ and therefore the absence of channel-length modulation effects. The interpretation is that for high enough gate voltages and/or short enough gate lengths, the velocity of carriers at the source approaches the saturation velocity. In this situation the current is determined only by mobile charge at the source, which is independent of gate length and depends only on the gate voltage. Although we can approach this situation, we could not achieve it in practice because of the onset of breakdown and punchthrough effects.

Returning to (15) and using our approximate value of $\bar{Q}$ from (18c), we find the saturation voltage to be

$$\bar{V} \equiv V_{SAT} = V_{GS} - V_T - \frac{\bar{Q}_0}{C_o} \sqrt{1 + \left(\frac{I}{W v_s \bar{Q}_0} \right)^2}. \tag{22}$$

Using (20b) in (22), we then find

$$V_{SAT} = V_{GS} - V_T - \left[\left(\frac{\bar{Q}_0}{C_o} \right)^2 \left(1 - \frac{1}{a^*} \right) + 2 \left(\frac{\mathscr{E}_c \bar{y}}{a^*} \right)^2 + \frac{1}{a^*} \right.$$

$$\left. \cdot (V_{GS} - V_T)^2 - \frac{2\mathscr{E}_c \bar{y}}{a^*} \sqrt{\left(\frac{\mathscr{E}_c \bar{y}}{a^*} \right)^2 + \frac{1}{a^*} (V_{GS} - V_T)^2 \left(\frac{\bar{Q}_0}{C_o} \right)^2} \right]^{1/2}. \tag{23}$$

This result is almost the same as one obtained in the literature using a piecewise continuous constant mobility and velocity-saturated model; i.e., from the continuity of current at a field $\mathscr{E}_c$ in the channel one has

$$\mu_0 \frac{W}{\bar{y}} C_o \left[(V_{GS} - V_T) V_{SAT} - \frac{V_{SAT}^2}{2} \right] = v_s C_o W (V_{GS} - V_T - V_{SAT}), \quad (24)$$

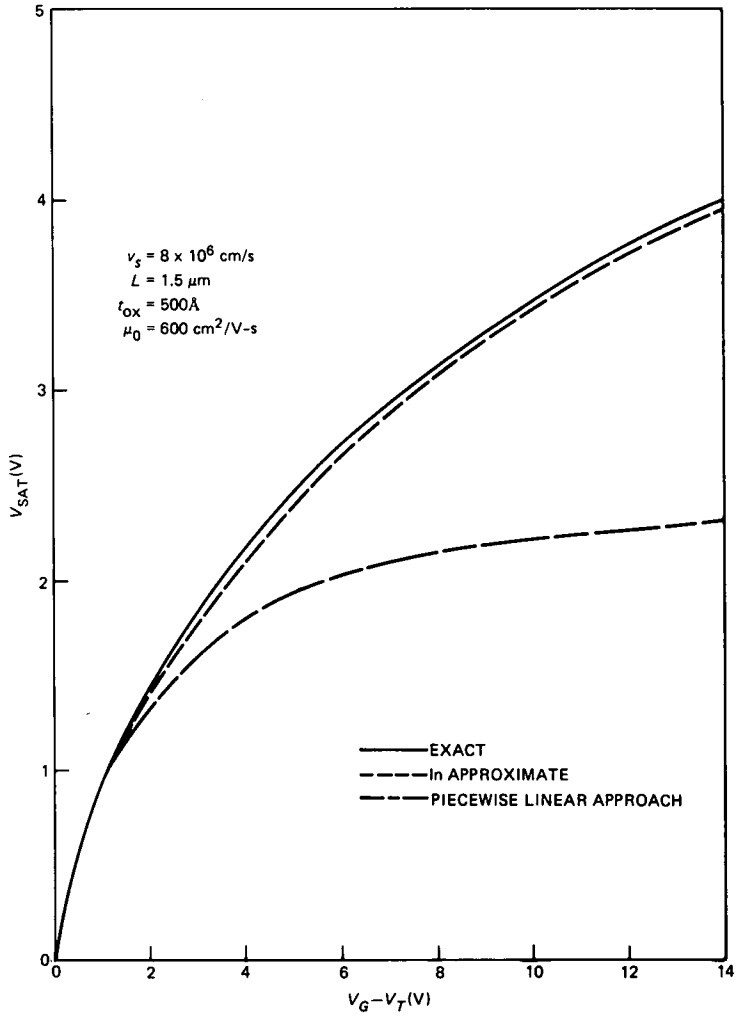

Fig. 3—Variation of V_{SAT} with gate voltage showing the complete solution using $B = 2$, the approximate solution using $B = 2$, and the piecewise linear constant mobility result.

which may be solved to yield

$$V_{SAT} = V_{GS} - V_T + \mathscr{E}_c\bar{y} - \sqrt{(V_{GS} - V_T)^2 + (\mathscr{E}_c\bar{y})^2}. \quad (25)$$

These two results are compared in Fig. 3 and it is seen that over much of the range of velocity saturation, (25) is rather inaccurate. The use of (25) to predict the current can then lead to errors of about 20 percent, as we show later in Fig. 9. By using (23) we obtain a good approximation to the exact result [the form used for a is described later by (33a)]. The differences between these two curves arise simply from the errors introduced in making the expansions to solve the original current equation (10a). The only drawback to (23) is that some iteration is required at the transition from the triode to the saturation region since a^* is a weak function of V_{SAT} through the logarithmic term. As we can see from (10b), this voltage enters only weakly into the current itself through a^*, but is necessary to determine where the saturation and triode regions meet and to provide an accurate merging of the two. In the calculation of I_{SAT} it should be noted that it is not necessary to calculate (23). Since the current is given explicitly by (20b), then (22) can be used to find V_{SAT}. It would also be possible to treat a as a constant parameter to be determined so that no iteration is required; however, it is shown later that the variation in a is important in achieving continuity of g_m.

In line with the reasoning that led to (24) it is of some interest to consider the field $\mathscr{E}_c$ as a kind of dividing line in the channel between the non-velocity-saturated and the velocity-saturated portions. Using (9a)

$$V_{GS} - V_T - V_C = \frac{\sqrt{2}I}{v_s C_o W} \quad (26)$$

is the condition in the channel when a field of $\mathscr{E} = \mathscr{E}_c$ is reached at a channel potential of $V = V_C$. We can then use (10b) to give V_C in the triode region, which is

$$V_C = V_{GS} - V_T + \frac{\sqrt{2}\,\mathscr{E}_c L}{a} - \sqrt{\frac{2}{a}(V_{GS} - V_T)^2 - \frac{2}{a}(V_{GS} - V_T - V_{DS})^2 + \left(\frac{\sqrt{2}\,\mathscr{E}_c L}{a}\right)^2}, \quad (27)$$

where a is given by (11b) and is a known function of V_{GS} and V_{DS}. We can then use this result in our expression (11a) for the current to find the position in the channel at which this field is achieved; i.e., using (26) in (11a) yields

$$y_C = \sqrt{2}$$

$$\cdot \frac{(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_C)^2 - \dfrac{(V_{GS} - V_T - V_C)^2}{2} a}{\mathscr{E}_c (V_{GS} - V_T - V_C)}, \quad (28)$$

where V_C is given by (27). If we follow the same procedure for operation in saturation then we obtain from (20b) and (26)

$$V_{GS} - V_T - V_C = \left\{ \left(\frac{\sqrt{2}\,\mathscr{E}_c \tilde{y}}{a^*} \right)^2 + 2 \left[(V_{GS} - V_T)^2 - \left(\frac{\bar{Q}_0}{C_o} \right)^2 \right] \right\}^{1/2} - \frac{\sqrt{2}\,\mathscr{E}_c \tilde{y}}{a^*}, \quad (29)$$

which is similar in form to (27) but now depends on $\tilde{y}$. The solution (28) applies to saturation or triode operation. Consider the situation $y_c = 0$, which implies either that the carriers are subjected to a field $> \mathscr{E}_c$ at all points in the channel or that they are moving with a velocity of $v_s/\sqrt{2}$ at the source end of the channel. Using (28) to find $V_{GS} - V_T - V_C$ and (11a) for I we find

$$I = \frac{W v_s C_o}{\sqrt{a^* + 1}} (V_{GS} - V_T),$$

i.e., we are very close to the velocity-saturated limit of (21b). It is then of some interest to determine the applied voltage for which this condition is achieved as a function gate length. Using (20b) we find

$$(V_{GS} - V_T) = 2\mathscr{E}_c L \sqrt{a^* + 1}. \quad (30)$$

This voltage value is plotted as a function of gate length in Fig. 4 using the values of a^* as determined by (33a) and (33b) (and shown plotted in Fig. 9). For $L = 0.5$ μm this condition is not achieved until 3V are applied to the device. These results simply point to the fact that the use of L/v_s to determine transit times usually gives values that are too small by a factor of about 2.

2.3 Discussion of results

2.3.1 I-V curves and current saturation

The interpretation of the phenomenon of the saturation of the drain current presented here is the following. Classical saturation of the current occurs when the charge in the channel has been reduced to the minimum value given by (16). This condition always occurs in the channel just before inversion in sign of the transverse oxide field occurs; i.e., pinch-off occurs at a channel potential given by (15), and when this potential has increased by another $\bar{Q}/C_o$ volts, we have zero voltage across the oxide, resulting in field inversion. Therefore, pinch-

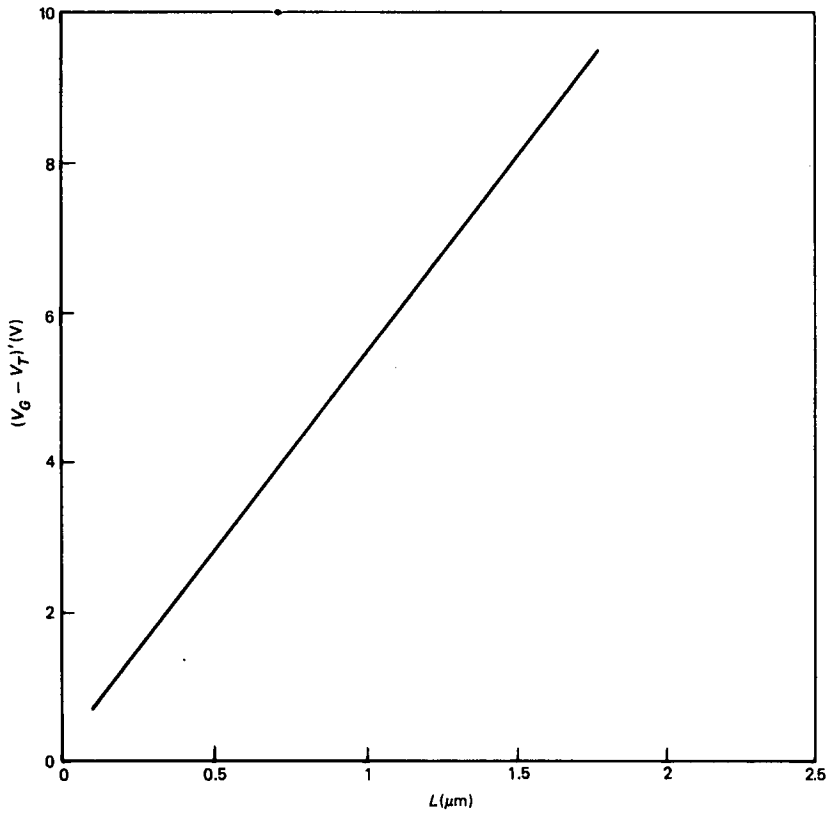

Fig. 4—Variation of $(V_{GS} - V_T)'$, the gate voltage for which the source velocity is $v_s/\sqrt{2}$, with gate length.

off in the channel will always be associated with a field inversion in the oxide. Now in the case without velocity saturation the voltage $\bar{Q}/C_o$ is typically 0.05V, which is dropped over a fairly short distance (of the order of 0.1 μm) in the channel, so that it is clear that pinch-off and field inversion occur at almost the same point in the channel. When velocity saturation becomes dominant, $\bar{Q}/C_o$ becomes a sizeable term compared to $V_{GS} - V_T$ because of the dependence of $\bar{Q}$ upon the electron temperature. This is the reason for the apparently lower drain saturation voltage in the velocity-saturated regime, as we can see from (15) and the plots in Fig. 3. Under these conditions the voltage drop in the channel between the pinch-off point and the field inversion point, and the distance between these two points may grow considerably from the constant mobility case.

The parameter a requires more attention for the computation of accurate results. From the results (12) and (21a) it is clear that a is

an important parameter only in the limit of velocity saturation. Consider the value of a at the saturation voltage for which case it has its greatest effect. Taking a in its complete form from (10c) we have

$$a \cong \ln \left[\frac{V_{GS} - V_T + (V_{GS} - V_T) \sqrt{1 - \dfrac{1}{1+a}}}{\bar{Q}/C_o + \bar{Q}_0/C_o} \right], \tag{31}$$

where we have used (18c) in the denominator and also the fact that in the limit of velocity saturation we have from (20b) or (11a)

$$\frac{I}{Wv_sC_o} \simeq \frac{V_{GS} - V_T}{\sqrt{1+a}}. \tag{32}$$

From this result we see that in the limit of a saturated drift velocity if $a < 1$, then we should be able to express a as a function of only $(V_{GS} - V_T)/(\bar{Q}/C_o)$. The result we obtain should be useful over the whole range of operation because, although determined from velocity-saturated conditions, a disappears from the expressions for the current when velocity saturation becomes unimportant. We have found by iteration that the result

$$a = 1.24 \ln \left(\frac{V_{GS} - V_T}{V_{GS} - V_T - V_{DS}} \right) \tag{33a}$$

in the triode region, which goes to

$$a = 1.24 \ln \left(\frac{V_{GS} - V_T}{\bar{Q}/C_o} \right) \tag{33b}$$

at the saturation voltage, gives excellent agreement over a wide range of parameters. The factor 1.24 was chosen to match the approximate and exact formulas for the current well into velocity-saturated operation both in the triode and saturation current regions. It was not chosen to match the exact and approximate forms of (31) itself so that it could be used in a more general way to compensate for all errors involved in the expansions in (10b). The corresponding value of a^* is found by using (33b) in (20c).

These calculations are shown for some typical devices in Figs. 5, 6, and 7 over a wide range of gate voltages, channel lengths, and gate oxide thicknesses. The approximate and exact predictions of the saturation voltage are also shown. It is noted that the logarithmic dependence of a upon drain voltage in (33) must be included to obtain an accurate result. The current therefore is a totally explicit function of the device voltages; the only iteration required is in the calculation of V_{SAT}, as we mentioned earlier. As the figure indicates, the approximations are very good.

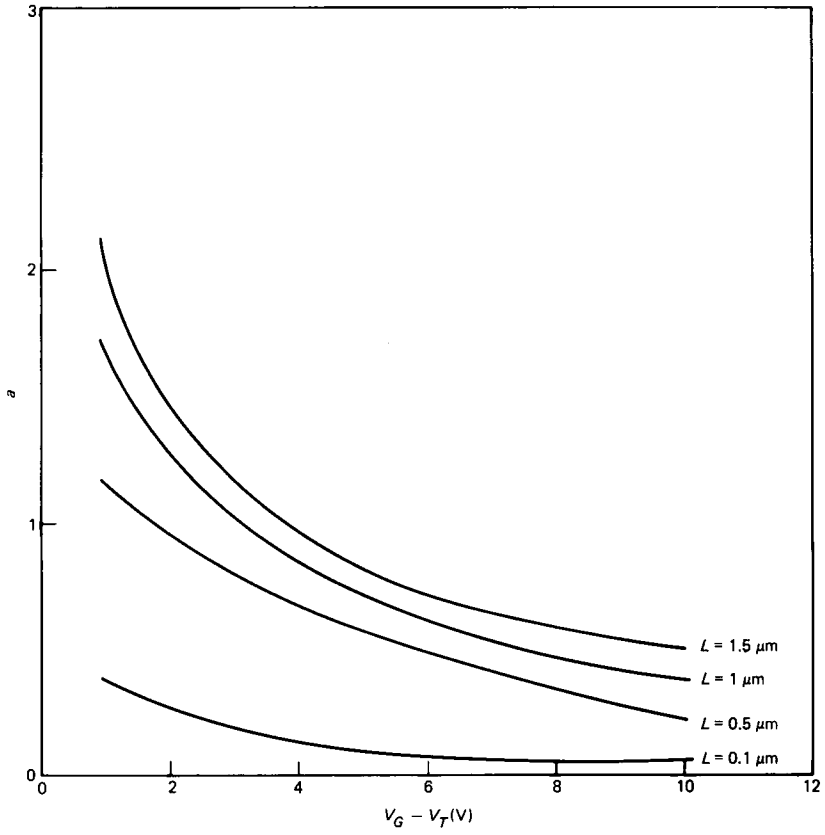

Fig. 5—Variation of the voltage parameter a at saturation with gate voltage for various gate lengths.

Figure 8 shows the conventional square law result. For the 1.5-μm device, for example, the effects of velocity saturation are relatively unimportant for $V_{GS} = 2$ but become progressively more significant as the gate voltage is increased, so that for $V_{GS} = 6$ there is a great difference between both the current and the saturation voltage. It is clear how the effects of a saturated drift velocity have reduced the current available from the device for a given supply voltage. It is for this reason that increasing the supply voltage has very little effect on gate propagation delay beyond some particular value of voltage. From (20b) we see that there is a significant departure from the square law when

$$V_{DD} \simeq V_T + \frac{\mathscr{E}_c L}{\sqrt{2a}} \quad \text{or} \quad \simeq V_T + \mathscr{E}_c L \tag{34}$$

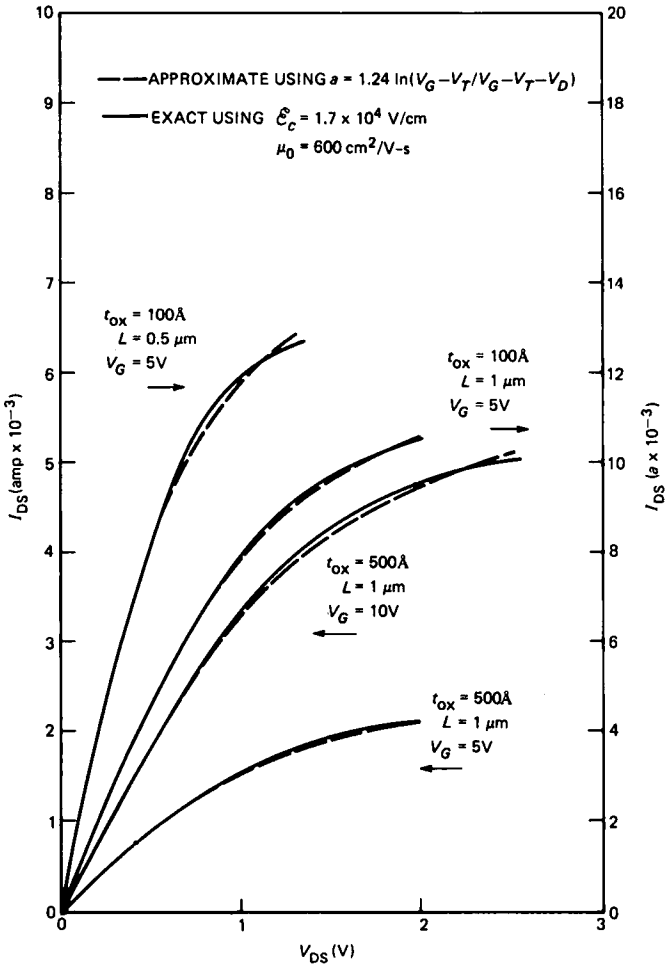

Fig. 6—Comparison of the rigorous and approximate solutions for a wide range of oxide thickness gate length and gate voltage. The voltage V_{SAT} is shown by the arrows that indicate the end of the line.

for typical values of a. We should, therefore, like to design for a supply voltage of approximately (34) since we would then have the maximum speed and the minimum power dissipation for a given channel length of the driver in an inverter for example. However, in practical applications of short-channel devices the supply voltage will be higher than this value to ensure adequate noise margins and a sufficient ratio of V_{DD}:V_T due to processing tolerances on V_T. This will definitely be the case if V_T is increased intentionally in order to suppress subthreshold leakage or if the supply voltage must be held arbitrarily at 5V to provide transistor-transistor logic (TTL) compatibility.

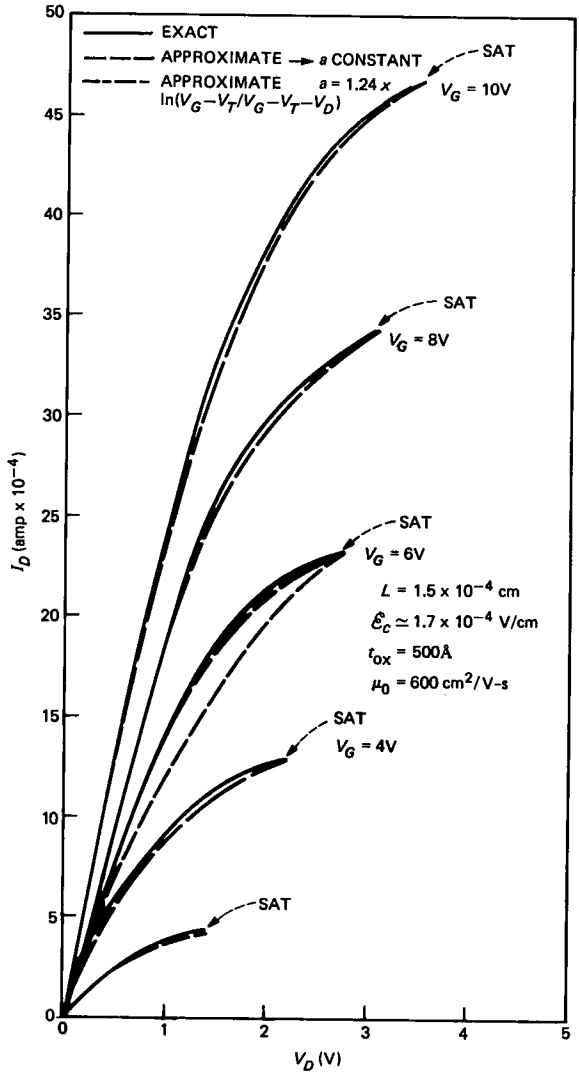

Fig. 7—Comparison of device characteristics for $L = 1.5$ μm using rigorous and approximate forms. Also, shown for $V_G = 6$V is the result of treating the ln term as constant.

The other point of interest in Fig. 8 is the comparison of the results obtained using the piecewise-linear model of the velocity-field characteristic and the more physical model presented here. The discrepancy (indicated by the arrows in the figure) grows at first quickly and then less rapidly as the gate voltage is increased; i.e., for lower gate voltages such that $V_{GS} - V_T \sim V_{SAT}$ the percent error in the current will be greater than that for $V_{GS} - V_T \gg V_{SAT}$. This is shown by the

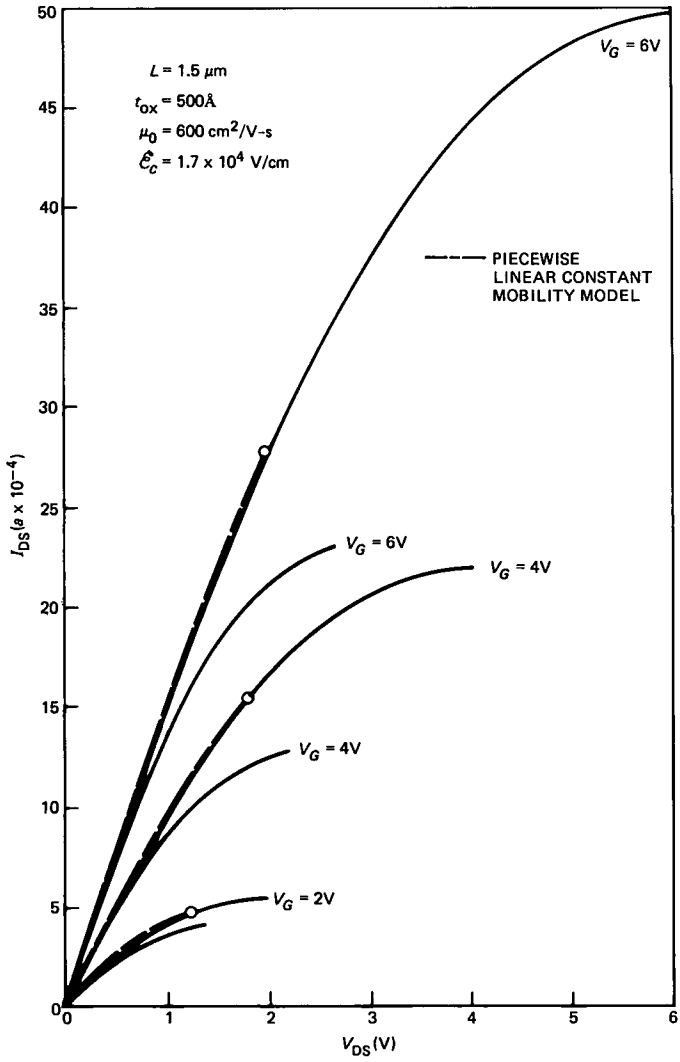

Fig. 8—Comparison of velocity-saturated model for $B = 2$ with constant mobility model. Also shown is the piecewise constant mobility model denoted by circles at saturation.

plot in Fig. 9 for the device in Fig. 8. The maximum in this curve is expected since the errors should be worst when the field at the drain is $\mathscr{E}_c$ because then the differences in the models are at maximum.

2.3.2 Substrate bias dependence

In the triode region the substrate bias totally enters through the threshold voltage as the body effect. It is also known that the threshold

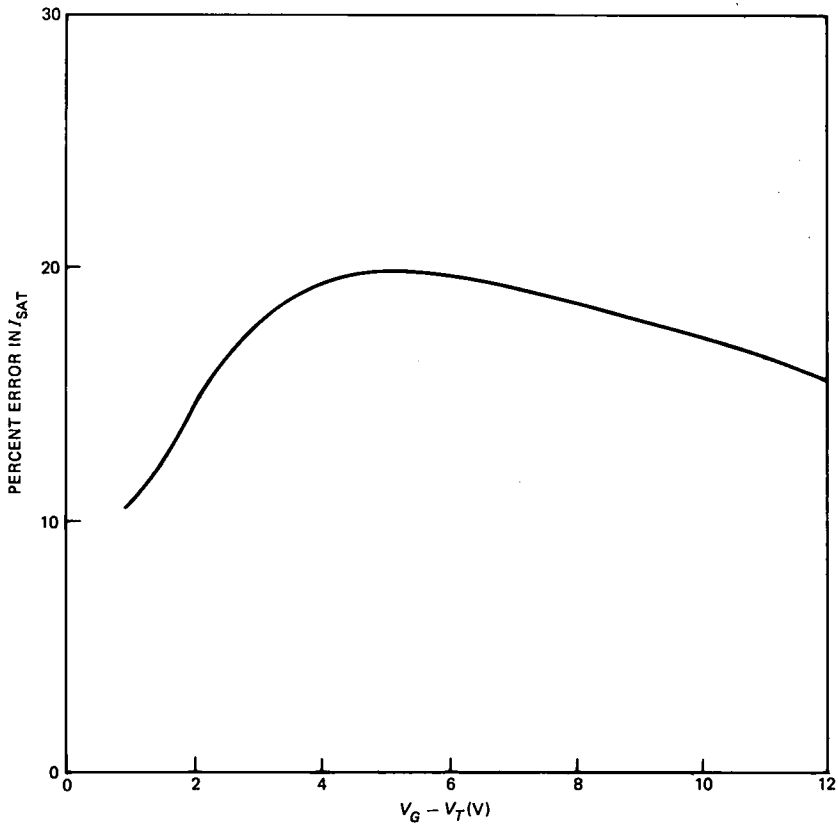

Fig. 9—Error incurred using the piecewise constant mobility model as a function of gate voltage.

voltage is dependent on drain voltage in a short-channel device and this dependence is increased by the application of substrate bias. These effects will be considered in the discussion of experimental results.

2.3.3 Effects of contact resistance and a gate-voltage-dependent mobility

In a real short-channel device, unavoidable series contact resistances can introduce great differences between actual and expected device currents. Also the mobility parameter, μ_0, exhibits a gate voltage dependence due to the increase in surface scattering with gate bias.[20] We will consider both of the effects here since, although unrelated, they affect the device current in the same way. We will consider a series resistance, R, in the drain and the source leads as shown in Fig. 2, and a mobility dependence of

$$\mu_0 = \frac{\mu_{00}}{1 + \theta(V_{GS} - V_T)}. \tag{35}$$

The parameter θ is an empirical constant that determines the dependence of the mobility upon the normal channel field[20] and is thus expected to have some substrate bias dependence. With these additions, one can show from (10c) that the triode region result is modified to

$$\frac{I_{TR}}{Wv_sC_o} = \sqrt{\left[\frac{\mathscr{E}_c'L}{a} + \frac{2R'}{a}\left(V_{GS} - V_T - \frac{V_{DS}}{2}\right)\right]^2 + \frac{(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2}{a}} - \left[\frac{\mathscr{E}_c'L}{a} + \frac{2R'}{a}\left(V_{GS} - V_T - \frac{V_{DS}}{2}\right)\right], \tag{36}$$

and the saturation region result is modified to

$$\frac{I_{SAT}}{Wv_sC_o} = \sqrt{\left[\frac{\mathscr{E}_c'\bar{y}}{a'} + \frac{2R'}{a'}\left(\frac{\bar{Q}_0}{C_o} + \frac{V_{GS} - V_T}{2}\right)\right]^2 + \frac{(V_{GS} - V_T)^2 - (\bar{Q}_0/C_o)^2}{a'}} - \left[\frac{\mathscr{E}_c'\bar{y}}{a'} + \frac{2R'}{a'}\left(\frac{\bar{Q}_0}{C_o} + \frac{V_{GS} - V_T}{2}\right)\right], \tag{37a}$$

where

$$R' = RWv_sC_o \tag{37b}$$

$$\mathscr{E}_c' = \mathscr{E}_c\,[1 + \theta(V_{GS} - V_T)] \tag{37c}$$

and

$$a' = 1 + a + 2RWv_sC_o. \tag{37d}$$

In writing (37a) we have used both (18c) and (18d) to express the linear and square root terms in $\bar{Q}/C_o$ that appear when (22) is substituted in (36). Therefore, we cannot expect a perfect match between (36) and (37a) at the saturation point, although it will be close. To achieve identical values we would have to use (18c) only and then iterate (37a) to determine I_{SAT}.

The effects of the series resistance and the mobility reduction are similar since they both act to increase the first term under the square root sign in (36) and (37a). From (36d) it is seen that the mobility degradation can be interpreted simply as a movement of $\mathscr{E}_c$ to higher values on the velocity field curve, as shown in Fig. 10. This has been

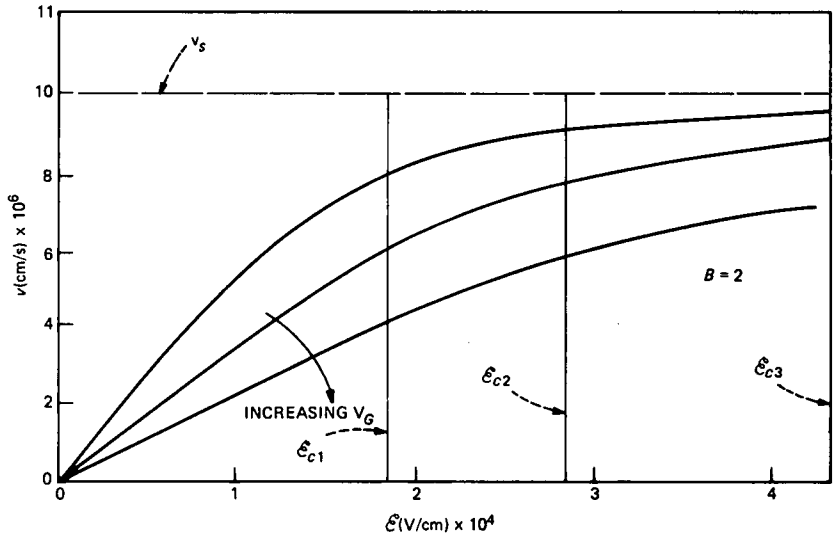

Fig. 10—Variation of $\mathscr{E}_c$ as gate voltage is increased because of the reduction in μ shifting the point of velocity saturation to higher $\mathscr{E}$.

noted elsewhere for the constant mobility case[6] and can be generalized to any velocity-field relation.

It is interesting to consider the limiting forms of these two results. From (36) by expanding the square root term in the limit of small gate and drain voltages, we obtain

$$I \simeq \frac{1}{2}\left[\frac{2(V_{GS} - V_T)V_{DS} - V_{DS}^2}{\mathscr{E}_c L + 2R'\left(V_{GS} - V_T - \dfrac{V_{DS}}{2}\right)}\right]. \tag{38}$$

If we now consider the case of small drain voltages, which describes the region in which we normally assess the dependence of μ_0 upon $V_{GS} - V_T$, we obtain

$$I = \frac{\mu_{00} C_o W}{L} \frac{(V_{GS} - V_T)V_{DS}}{1 + \left(\theta + 2R\dfrac{\mu_{00} C_o W}{L}\right)(V_{GS} - V_T)}. \tag{39}$$

From this result it is clear that one must be careful when extracting a physically meaningful value of θ from experimental data since an accurate value of R must first be known. This can be shown clearly by the data in Fig. 11 of the linear region current of three devices that are identical except for the gate length. Each curve is characterized by a section at lower gate voltages, which is linear, and a section at higher gate voltages, which is nonlinear in gate voltage. The gate voltage at

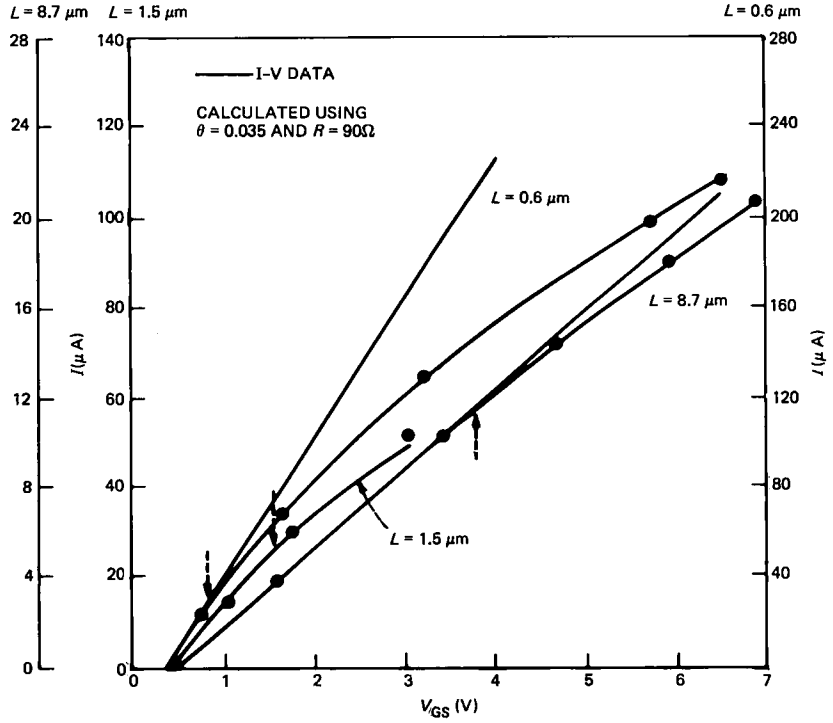

Fig. 11—Linear region current data showing the effect of R and θ. The experimental device parameters are $t_{ox} = 500\text{Å}$, $N_A = 1 \times 10^{16}$ cm^{-3}, and $\mu_{00} \simeq 560$ cm^2/V-s.

which the sections join is marked in the figure by an arrow. The most important feature of this voltage is that it moves to lower and lower values for decreasing L. From (35) we see that this breakpoint should be fixed in gate voltage—the reason for the movement to lower values is the increasing importance of the term in R in (38) for decreasing L. If we use $\theta = 0.035$ and a value of $R = 90\Omega$, then we can fit the curves as indicated. To give some idea of the importance of R, for the long-channel device the choice of R is immaterial within limits since θ is dominant. For the smallest length ($L \simeq 0.5$ μm) the choice of θ is not important because the R term is dominant. Thus, as channel lengths are reduced it becomes increasingly important to reduce the series R. Both the contact resistance and the mobility reduction will cause V_{SAT} to increase for a given V_G.

Figure 12 shows the effects of a contact resistance of 50Ω on the device current for a typical short-channel device. To write down a closed form for I, we have ignored the dependence of a upon R in arriving at (36) and (37a). The difference this makes is indicated in Fig. 12.

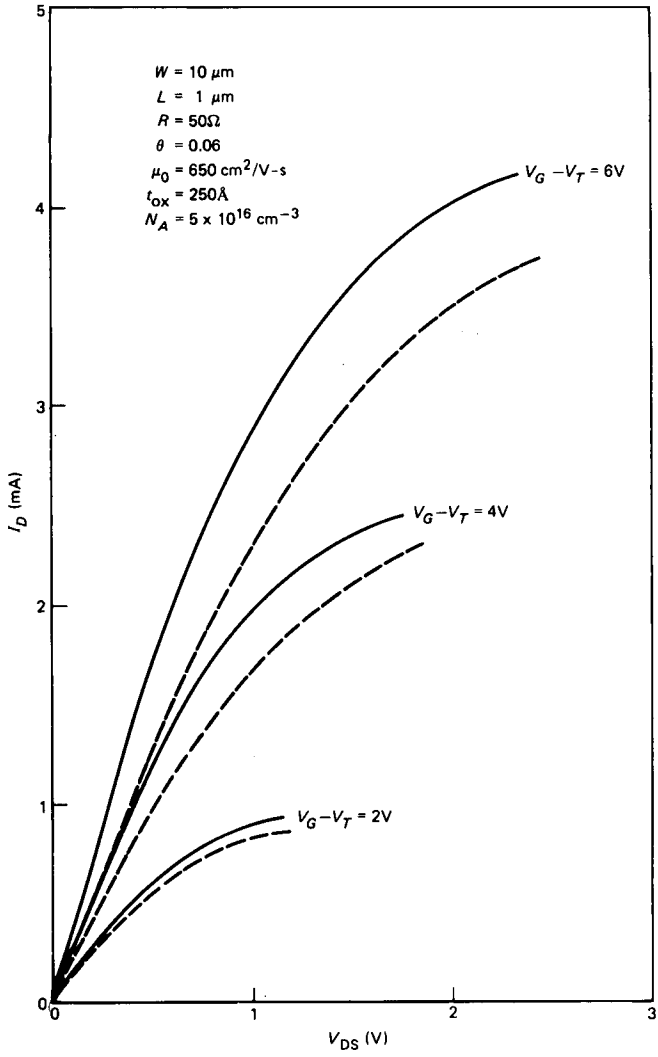

Fig. 12—Effect of series resistance on drain characteristics of a short-channel device. The curves are calculated using (36) with $R = 0$ (solid line) and $R = 505\Omega$ (dashed line). The second dashed line shows the effects of neglecting any effect of a upon R.

2.3.4 Incremental parameters and the continuity of derivatives

2.3.4.1 Current-saturated operation. Since the current as derived in (20b) is a continuous function of gate voltage in moving from non-velocity-saturated to velocity-saturated operation, then the same is true for transconductance, and we obtain

$$g_m = \frac{dI}{dV_{GS}} = Wv_sC_o\left(\left\{-\left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)^2\frac{1}{a^*}\frac{da^*}{dV_{GS}} + \frac{(V_{GS}-V_T)}{a^*}\right.\right.$$
$$\left.-\frac{1}{2}\left[\frac{(V_{GS}-V_T)^2-(\bar{Q}_0/C_o)^2}{a^*}\right]\frac{1}{a^*}\frac{da^*}{dV_{GS}} + \left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)^2\frac{1}{\bar{y}}\frac{d\bar{y}}{dV_{GS}}\right\}$$
$$\div\left[\left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)^2 + \frac{(V_{GS}-V_T)^2-(\bar{Q}_0/C_o)^2}{a^*}\right]^{1/2}$$
$$+\left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)\frac{1}{a^*}\frac{da^*}{dV_{GS}} - \left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)\frac{1}{\bar{y}}\frac{d\bar{y}}{dV_{GS}}\Bigg). \quad (40)$$

Similarly, the drain conductance is

$$g_d = \frac{dI}{dV_{DS}} = Wv_sC_o\left(\left\{-\left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)^2\frac{1}{a^*}\frac{da^*}{dV_{DS}}\right.\right.$$
$$-\frac{1}{2}\left[\frac{(V_{GS}-V_T)^2-(\bar{Q}_0/C_o)^2}{a^*}\right]\frac{1}{a^*}\frac{da^*}{dV_{DS}} - \frac{(V_{GS}-V_T)}{a^*}\frac{dV_T}{dV_{DS}}$$
$$\left.+\left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)^2\frac{1}{\bar{y}}\frac{d\bar{y}}{dV_{DS}}\right\} \div \left[\left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)^2 + \frac{(V_{GS}-V_T)^2-(\bar{Q}_0/C_o)^2}{a^*}\right]^{1/2}$$
$$+\left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)\frac{1}{a^*}\frac{da^*}{dV_{DS}} - \left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)\frac{1}{\bar{y}}\frac{d\bar{y}}{dV_{DS}}\Bigg). \quad (41)$$

We could also write these equations with a^* everywhere replaced by a and $\bar{Q}_0/C_o$ everywhere replaced by $\bar{Q}/C_o$. In that case, in the numerator of the square root terms we require the additional terms

$$-\frac{2}{a}\bar{Q}/C_o\frac{d}{dV_{GS}}\left(\frac{\bar{Q}}{C_o}\right) \quad \text{and} \quad -\frac{2}{a}\bar{Q}/C_o\frac{d}{dV_{DS}}\left(\frac{\bar{Q}}{C_o}\right)$$

in g_m and g_d, respectively.

In these equations, $\bar{y}$ is a function of both the drain and gate voltages. The parameter a^* is related to a by (20c), and in the pinch-off condition [using (22) and (11b) in 20(c)] a^* is

$$a^* = 1 + \ln\left[\frac{V_{GS}-V_T}{\frac{\bar{Q}_0}{C_o}+\frac{I}{Wv_sC_o}}\right]. \quad (42)$$

Now the value of a and its variations are unimportant for low values of V_{GS} because the conventional device equations (i.e., without velocity saturation) apply and the forms for g_m and g_d are well known. On the other hand, for operation in velocity saturation we have shown that [cf. (32)] I' is a linear function of gate voltage and so (42) will be

relatively independent of gate and drain voltage. We can, therefore, neglect the a^* terms in (40) and (41) and use the value of a^* that we obtain at the pinch-off point to obtain a fairly good approximation. However, in comparing theory with experiment for accurate results, as we shall see, we must keep the dependence of I in (42) and calculate the current by one or two iterations. Some interesting features of (40) and (41) are worth noting. In the limit of large gate voltages (i.e., velocity saturation) so that $V_{GS} - V_T \gg \mathscr{E}_c\bar{y}$, (40) predicts that

$$g_m = \frac{C_o W v_s}{\sqrt{a^*}}, \tag{43}$$

and (41) predicts that

$$g_d = C_o W v_s \cdot \left[\frac{(\mathscr{E}_c\bar{y})^2}{(V_{GS} - V_T)} \frac{1}{a^{*3/2}} \frac{1}{\bar{y}} \frac{d\bar{y}}{dV_{DS}} - \frac{1}{2} \frac{(V_G - V_T)}{a^{*3/2}} \frac{da^*}{dV_{DS}} - \frac{1}{a^{*1/2}} \frac{dV_T}{dV_{DS}}\right]. \tag{44}$$

2.3.4.2 Triode region. In the triode region the corresponding expressions for the gate transconductance and the drain conductance are

$$g_m = W v_s C_o \cdot \left\{ \frac{\left(\frac{\mathscr{E}_c L}{a}\right)^2 \frac{\dot{a}_G}{a} + \frac{(V_{GS} - V_T) - (V_{GS} - V_T - V_{DS})}{a} - \frac{[(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2]}{a^2} \dot{a}_G}{\sqrt{\left(\frac{\mathscr{E}_c L}{a}\right)^2 + \frac{(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2}{a}} + \frac{\mathscr{E}_c L}{a^2} \dot{a}_G} \right\} \tag{45}$$

and

$$g_d = W v_s C_o \cdot \left\{ - \frac{\left(\frac{\mathscr{E}_c L}{a}\right)^2 \frac{\dot{a}_D}{a} + \frac{V_{GS} - V_T - V_{DS}}{a} - \frac{[(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2]\dot{a}_D}{a^2} - \frac{V_{DS}}{a} \frac{dV_T}{dV_{DS}}}{\sqrt{\left(\frac{\mathscr{E}_c L}{a}\right)^2 + \frac{(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2}{a}} + \frac{\mathscr{E}_c L}{a^2} \dot{a}_D} \right\}. \tag{46}$$

In this case the terms in $\dot{a}$ are necessary to obtain a reasonable solution and so from (33a) we have

$$\frac{da}{dV_{DS}} = \dot{a}_D = \frac{1.24}{V_{GS} - V_T - V_{DS}},$$

$$\frac{da}{dV_{GS}} = \dot{a}_G = \frac{1.24V_{DS}}{(V_{GS} - V_T)(V_{GS} - V_T - V_{DS})}. \tag{47}$$

Again, as in the case of the saturation region, the formulas revert to their conventional forms for small values of gate and drain voltages. The main interest in considering the incremental conductance parameters is to show that continuity exists in making the transition from triode to saturation regions. We would like to show equivalence between (45) and (40) and between (46) and (41) at pinch-off conditions. By comparing (46) and (41) (written in terms of a and $\bar{Q}/C_o$) we find

$$\frac{(V_{GS} - V_T - V_{DS})}{a} = -\frac{\bar{Q}}{aC_o}\frac{d}{dV_{DS}}\left(\frac{\bar{Q}}{C_o}\right) - \frac{d\bar{y}}{dV_{DS}}\frac{1}{a}\left(\frac{I}{Wv_sC_o}\right)$$

or

$$\bar{Q}/C_o\left[1 + \frac{d}{dV_{DS}}(\bar{Q}/C_o)\right] = -\mathscr{E}_c\frac{d\bar{y}}{dV_{DS}}\left(\frac{I}{Wv_sC_o}\right). \tag{48}$$

Since the derivative on the left-hand side of (48) $\ll 1$, then (48) can be written as (173). As we show in Appendix A from (174) through (177) the derivative continuity equation (48) supplies the other condition needed together with (177) that allows determination of the parameters A_2 and h.

To illustrate these results, the drain conductance has been plotted in Fig. 13 with drain voltage as a parameter. The accuracy is quite good and continuity of the derivative is preserved in moving from the triode to the saturation regions.

2.3.5 Inclusion of the V_{DS} dependence of V_T

2.3.5.1 Triode region. In the triode region the charge-sharing formulation[16] is used in which V_T is expressed

$$V_T = V_{FB} + 2\phi_F + \frac{2\sqrt{2\epsilon qN_A}}{3C_oV_{DS}} \cdot [(V_{DS} + V_{BS} + 2\phi_F)^{3/2} - (V_{BS} + 2\phi_F)^{3/2}]F_{ON},$$

where F_{ON} is the charge-sharing factor in the triode region. Although this result may be written approximately by expanding (for small V_{DS})

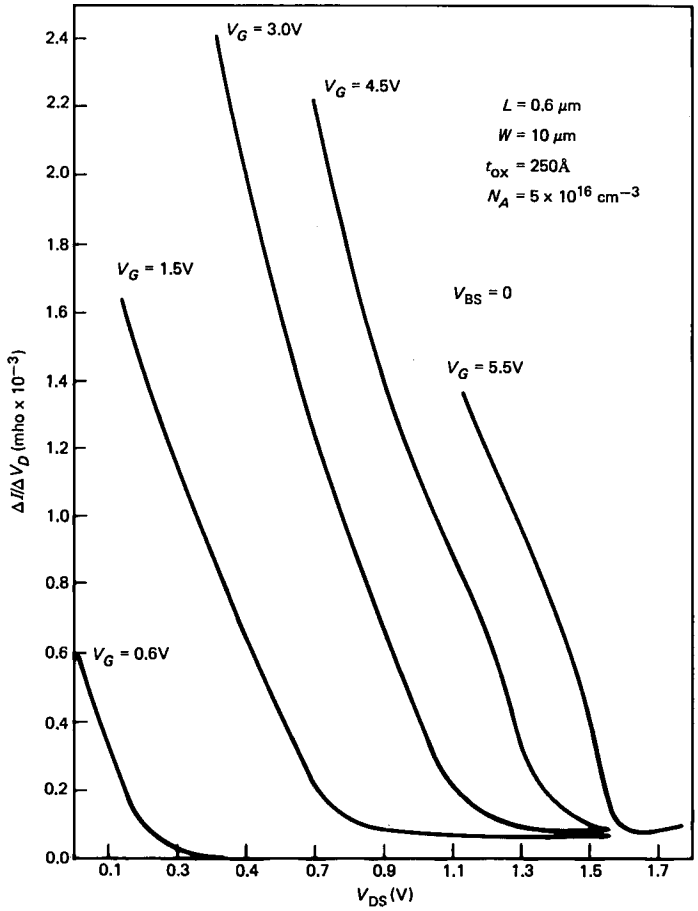

Fig. 13—Comparison of approximate and rigorous forms of the drain conductance.

as

$$V_T = V_{FB} + 2\phi_F + \frac{1}{C_o}\sqrt{2\epsilon q N_A\left(V_{BS} + \frac{V_{DS}}{2} + 2\phi_F\right)}F_{ON},$$

it is important to note that the complete form must always be used for circuit simulation purposes where discontinuities may not be allowed between the triode and saturation regions. The complete form provides a smooth transition for all voltages, whereas the approximate form introduces small glitches for some values.

2.3.5.2 Saturation region. In saturation operation the complete and approximate equations for V_T become

$$V_T = V_{FB} + 2\phi_F + \frac{2}{3}\frac{\sqrt{2\epsilon q N_A}}{C_o V_{SAT}} \cdot [(V_{SAT} + V_{BS} + 2\phi_F)^{3/2} - (V_{BS} + 2\phi_F)^{3/2}]_{SAT} \quad (49)$$

and

$$V_T = V_{FB} + 2\phi_F + \frac{1}{C_o}\sqrt{2\epsilon q N_A (V_{SAT} + V_{BS} + 2\phi_F)} F_{SAT}.$$

As in the triode region the complete form (49) must be used to avoid glitches at the transition regions. This result was developed previously.[16] The difference in this case is that V_{SAT} also depends upon the charge at pinch-off and thus the current through (22). Therefore (22), (19b), and (49) must be solved together to determine the current. To obtain results that are smooth through all transitions, iteration is required.

2.3.6 *Effective value of N_A*

Almost all practical devices have ion-implanted channels and hence nonuniform doping profiles. Since the short-channel formulation describes the device in terms of averages over the source-drain distance, a reasonable approach is to use an average value of $N_A = \bar{N}_A$ over the same distance. We can calculate this value by conserving charge in the manner

$$\bar{N}_A \bar{x}_d = \int_0^{\bar{x}_d} N_A(x)dx, \quad (50)$$

where $\bar{x}_d$ is the average depletion width between the drain and the source and has been shown to be[16]

$$\bar{x}_d \simeq \sqrt{\frac{2\epsilon}{q\bar{N}_A}} \sqrt{V_{BS} + 2\phi_F + \frac{V_{DS}}{2}}. \quad (51)$$

We can see from this result that the effective value of $\bar{N}_A$ is in general a function of both drain and substrate bias.

To determine $\bar{N}_A$ we will assume that any ion-implanted profile may be described by the parameters N_p, σ, and R_p; i.e., even after thermal processing it is assumed that N_p, σ, and an R_p may be found to give a best fit to the measured impurity profile so that

$$N_A(x) = (N_p - N_{AB})\exp\left[-\left(\frac{x - R_p}{2\sigma}\right)^2\right] + N_{AB}, \quad (52)$$

where N_{AB} is the background impurity doping. Using (52) and (51) in (50) we then have

$$\sqrt{\bar{N}_A} - \frac{N_{AB}}{\sqrt{\bar{N}_A}} = \left[\frac{q}{2\epsilon\left(V_{BS} + \frac{V_{DS}}{2} + 2\phi_F\right)}\right]^{1/2} \sqrt{\frac{\pi}{2}}\, N_p\sigma \cdot \left(\operatorname{erf}\left(\frac{R_p}{\sqrt{2}\sigma}\right) + \operatorname{erf}\left\{\frac{\left[\frac{2\epsilon}{q\bar{N}_A}\left(V_{BS} + 2\phi_F + \frac{V_{DS}}{2}\right) - R_p\right]^{1/2}}{\sqrt{2}\sigma}\right\}\right) \quad (53)$$

for the triode region and

$$\sqrt{\bar{N}_A} - \frac{N_{AB}}{\sqrt{\bar{N}_A}} = \left[\frac{q}{2\epsilon\left(V_{BS} + \frac{V_{SAT}}{2} + 2\phi_F\right)}\right]^{1/2} \sqrt{\frac{\pi}{2}}\, N_p\sigma \cdot \left(\operatorname{erf}\left(\frac{R_p}{\sqrt{2}\sigma}\right) + \operatorname{erf}\left\{\frac{\left[\frac{2\epsilon}{q\bar{N}_A}\left(V_{BS} + 2\phi_F + \frac{V_{SAT}}{2} - R_p\right)\right]^{1/2}}{\sqrt{2}\sigma}\right\}\right)$$

for the saturation region. If we use a polynomial expansion for the erf term to allow numerical computation, only a few iterations are required to obtain $\bar{N}_A$. For the most accurate results, then, a value of $\bar{N}_A(V_{DS}, V_{BS})$ is required at each bias point, although in practice only the V_{BS} dependence is very important, so an average value of V_{DS} could be used for all computation.

III. COMPARISON WITH EXPERIMENT

To establish the validity of the parameter $B = 2$, the $\sqrt{I}$ versus V_{GS} data were plotted in Fig. 14 with drain voltage $V_{DS} = V_{GS}$ for the long- and short-channel devices discussed in Fig. 11. Using all of the parameters determined from the linear region data, theoretical curves were generated and are also plotted. The drain-voltage dependence of V_T was included using the charge-sharing technique,[16] and the variation of $\bar{y}$ was taken from the drain-voltage data in order to verify the triode region model alone. It is evident that the agreement is very good. To compare with the other models, the result of using $B = 1$, and the piecewise continuous model (i.e., constant mobility) are also shown. Both are in error, one underestimating and the other overestimating the actual current.

One can identify two sections in this curve. At low voltages the $\sqrt{I}$ curve is linear in V_{GS}, indicating long-channel behavior; for higher gate voltages the $\sqrt{I}$ curve is sublinear in I, indicating the onset of the effects of velocity saturation. The change from one behavior to the

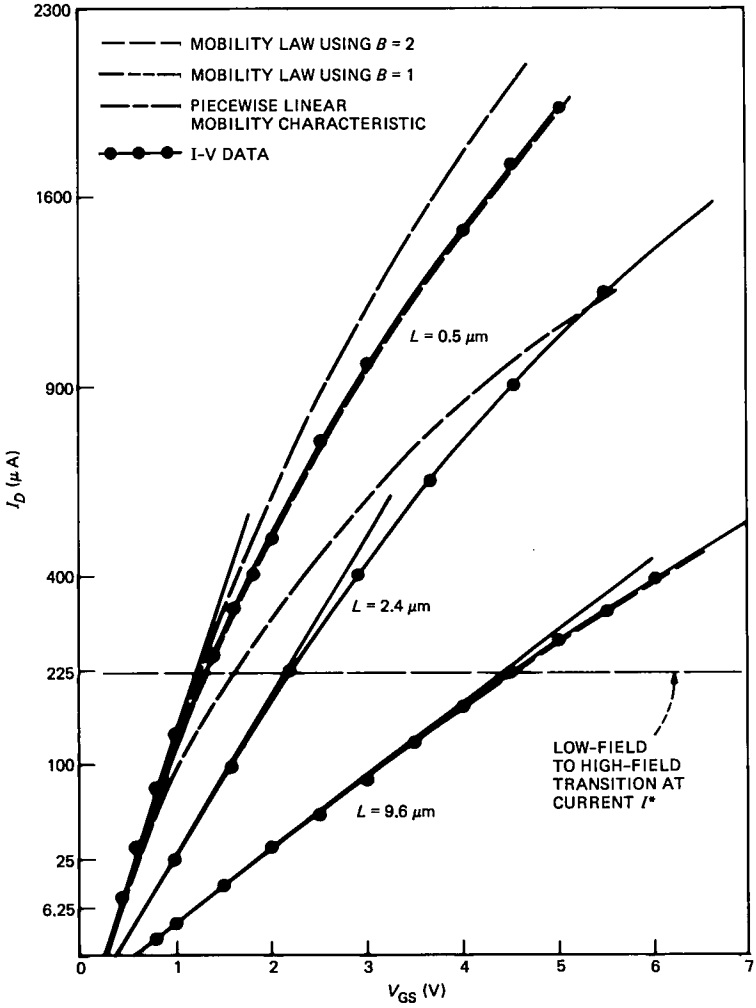

Fig. 14—Saturation region data for long- and short-channel devices plotted on a $\sqrt{I}$ scale for the condition $V_{DS} - V_{GS}$.

other is gradual and takes place when the terms under the root sign in (20b) are approximately equal; i.e., at a current level of about

$$I = v_s W C_o \frac{(V_{GS} - V_T)}{\sqrt{a^*}}.$$

This result should be independent of gate length, as we can see from (17c), which predicts the transition from $\mathscr{E} < \mathscr{E}_c$ to $\mathscr{E} > \mathscr{E}_c$ to occur for a unique value of I, which is

$$I^* = \frac{v_s W \bar{Q}_0}{(1 - \bar{R})}.$$

This behavior is observed in Fig. 14, although there is some decrease for the very shortest gate length.

As a final demonstration of the model the drain current data for two devices are shown in Figs. 15a and b. The channel length modulation term can be calculated using (169). However, this description is only valid for large gate voltages, i.e., when $\bar{\mathscr{E}}$ is $\gtrsim 10^4$ V/cm. As we approach the threshold condition, the derivation breaks down and $L - \bar{y}$ becomes anomalously large. However, near the threshold condition and in the subthreshold region, we know from other work that the simple form of

$$L - \bar{y} = \sqrt{\frac{2\epsilon}{q\bar{N}_A}(V_{DS} - V_{SAT})}$$

works well. We will, therefore, combine these two results to obtain a continuous solution

$$L - \bar{y} \frac{A_1(V_{DS} - V_{SAT})^{3/4}}{\bar{\mathscr{E}} + A_1(V_{DS} - \bar{V})^{1/4} \sqrt{\frac{qN_A}{2\epsilon}}}.$$

In Fig. 15a the fabrication parameters are $L = 0.32$ μm (1-μm coded length and 0.68-μm compensation as determined from $1/g_m$ measurements), $W = 30$ μm, $t_{ox} = 230$Å, $r_j = 0.26$ μm, and $N_{AB} = 8 \times 10^{15}$ cm^{-3}. The implanted doping parameters were determined by simulation to be $N_p = 5 \times 10^{16}$ cm^{-3}, $R_p = 0.5$ μm, and $\sigma = 0.1$ μm. Other parameters determined from low voltage data are $R = 15$ through 20Ω, $\mu_0 = 650$ cm^2/V-s and $\theta = 0.02$V^{-1}. The agreement between theory and experiment is very good for $V_{GS} > 1$V. For smaller gate voltages the characteristic is dominated by uncontrolled source-drain punchthrough, which is not described by these models.

In Fig. 15b the parameters are the same but $L = 0.85$ μm, so in this case the comparison is extended down to the threshold region (because source-drain punchthrough is not a limiting factor).

IV. SUMMARY

A model has been developed to describe the velocity-saturated characteristics of short-channel MOSFETs. The model has been based upon the velocity-field relation that most nearly fits the experimental data, and it is found that the bulk relationship based upon optical phonon emission is most appropriate if the zero field (parallel to the

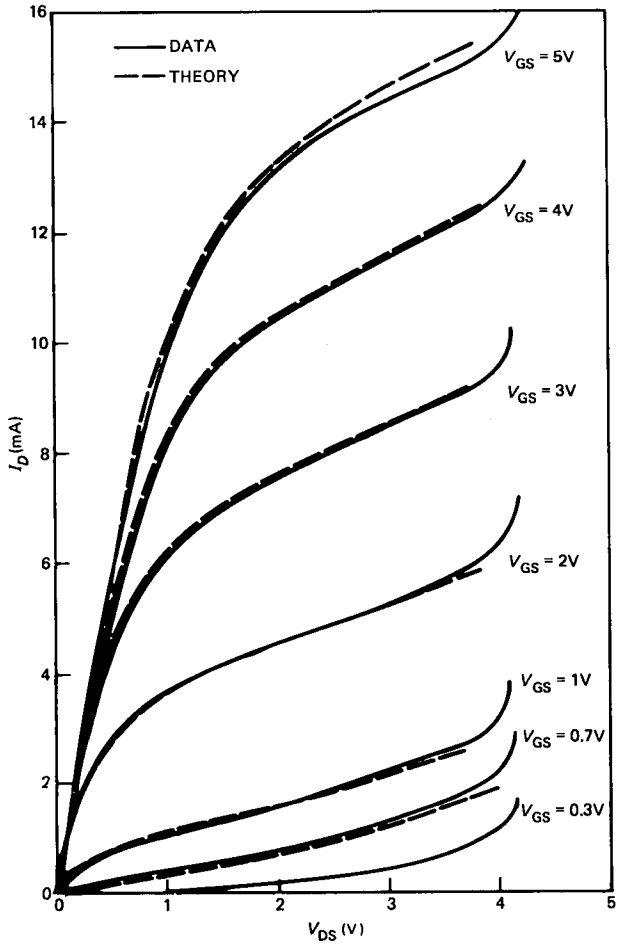

Fig. 15(a)—Drain-voltage versus drain-current data for a short-channel MOSFET showing comparison with theory. $L = 0.32$ μm.

surface) mobility is simply modified to take into account the effects of increased scattering in the potential well at the semiconductor-insulator interface. By using this approach, the electron temperature has been brought into the problem and becomes the necessary ingredient that allows a smooth transition from the long-channel to the short-channel behavior. The model is found to fit well with experimental data.

V. ACKNOWLEDGMENTS

The author would like to acknowledge helpful discussions with W. Fichtner, S. M. Sze, and G. E. Smith.

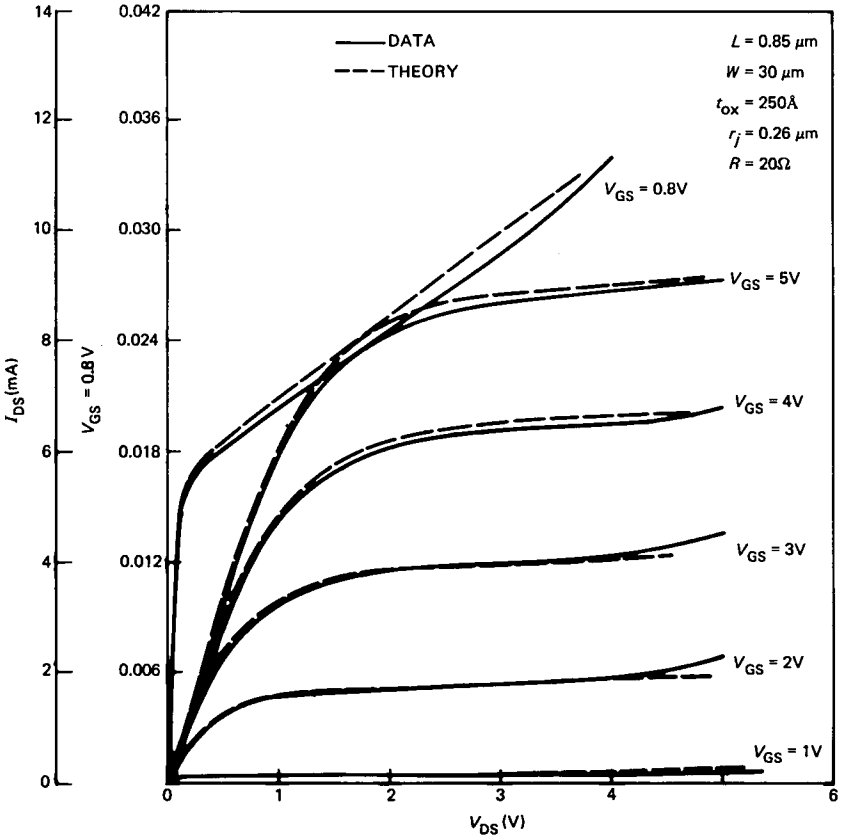

Fig. 15(b)—Drain-voltage versus drain-current data for a short-channel MOSFET showing comparison with theory. $L = 0.85$ μm.

REFERENCES

1. K. Yamaguchi, "Field-Dependent Mobility Model for Two-Dimensional Numerical Analysis of MOSFETs," IEEE Trans. Electron. Dev., *ED-26* (July 1979), pp. 1068–78.
2. T. Toyabe et al., "A Numerical Model of Avalanche Breakdown in MOSFETs," IEEE Trans. Electron. Dev., *ED-25* (July 1978), pp. 825–32.
3. H. Masuda, M. Nakai, and M. Kubo, "Characteristics and Limitations of Scaled Down MOSFETs Due to Two-Dimensional Field Effect," IEEE Trans. Electron. Dev., *ED-26* (June 1979), pp. 980–6.
4. R. Reich and D. K. Ferry, "A Two-Dimensional Simulation of a Cooled Submicrometer Indium Arsenide Schottky-Gate FET," IEEE Trans. Electron. Dev., *ED-27* (June 1980), pp. 1062–5.
5. B. Hoefflinger et al., "Model and Performance of Hot-Electron MOS Transistors for VLSI," IEEE J. Solid State Circuits, *SC-14* (April 1979), pp. 435–42.
6. T. J. Rodgers et al., "An Experimental and Theoretical Analysis of Double-Diffused MOS Transistors," IEEE. J. Solid State Circuits, *SC-10* (October 1975), pp. 322–31.
7. F. N. Trofimenkoff, "Field-Dependent Mobility Analysis of the Field-Effect Transistor," Proc. IEEE, *53* (November 1965), pp. 1765–6.

8. B. Hoeneissen and C. A. Mead, "Current Voltage Characteristics of Small Size MOS Transistors," IEEE Trans. Electron. Dev., *ED-19* (March 1972), pp. 382–3.
9. D. Roychoudhury and P. K. Basu, "A New Mobility Field Expression for the Calculation of MOSFET Characteristics," Solid State Electron., *19* (July 1976), p. 656.
10. R. J. Jardonek and W. R. Bandy, "Velocity Saturation Effects in n-Channel Deep-Depletion SOS/MOSFETS," IEEE Trans. Electron. Dev., *ED-25* (August 1978), pp. 894–8.
11. G. A. Armstrong and J. A. Magowan, "The Distribution of Mobile Carriers in the Pinch-Off Region of an Insulated-Gate Field-Effect Transistor and Its Influence on Device Breakdown," Solid State Electron., *14* (August 1971), pp. 723–35.
12. G. Merckel, J. Borel, and N. Z. Cupcea, "An Accurate Large Signal MOS Transistor Model for Use in Computer-Aided Design," IEEE Trans. Electron. Dev., *ED-19* (May 1972), pp. 681–90.
13. D. M. Caughey and R. F. Thomas, "Carrier Mobilities in Si Empirically Related to Doping and Field," Proc. IEEE, *55* (December 1967), pp. 2192–3.
14. E. M. Conwell, *High Field Transport in Semiconductors*, New York: Academic Press, 1967.
15. R. Stratton, "Diffusion of Hot and Cold Electrons in Semiconductor Barriers," Phys. Rev., *126* (June 1962), pp. 2002–14.
16. G. W. Taylor, "The Effects of Two-Dimensional Charge Sharing on the Above-Threshold Characteristics of Short-Channel IGFETs," Solid State Electron., *22* (August 1978), pp. 701–18.
17. D. P. Kennedy and P. C. Murley, "Steady State Mathematical Theory for the Insulated Gate Field Effect Transistor," IBM J. Res. Dev., *17*, No. 1 (January 1973), pp. 2–12.
18. S. C. Sun and J. D. Plummer, "Electron Mobility in Inversion and Accumulation Layers on Thermally Oxidized Silicon Surfaces," IEEE Trans. Electron. Dev., *ED-27* (August 1980), pp. 1497–508.
19. J. P. Nougier, "Identity Between Spreading and Noise Diffusion Coefficients for Hot Carriers in Semiconductors," Appl. Phys. Lett. *32*, No. 10 (October 1978), pp. 671–3.
20. P. P. Guebels and F. Van de Wiele, IEDM Tech. Digest (December 1981), p. 211.
21. K. K. Thornber, "Relation of Drift Velocity to Low-Field Mobility and High-Field Saturation Velocity," J. Appl. Phys., *51* (April 1980), pp. 2127–36.
22. H. K. J. Ihantola and J. L. Moll, "Design Theory of a Surface Field Effect Transistor," Solid State Electron., *7* (July 1964), pp. 423–30.
23. H. C. Pao and C. T. Sah, "Effect of Diffusion Current on Characteristics of Metal-Oxide (Insulator)-Semiconductor Transistor," Solid State Electron., *9* (September 1966), pp. 927–37.
24. J. A. Magowan and G. A. Armstrong, "Modelling of Short-Channel M.O.S. Transistors," Electron. Lett., *6*, No. 10 (May 1970), pp. 313–5.
25. G. W. Taylor, "Subthreshold Conduction in MOSFETs," IEEE Trans. Electron. Dev., *ED-25* (March 1978), pp. 337–50.

APPENDIX A

A.1 Introduction

In existing analytical approaches to pinch-off operation, the field at the edge of the pinch-off region is held to the value $\mathscr{E}_c$, the critical field parameter in the velocity-field relationship.[6,12,21] To avoid such an arbitrary condition, a description is presented here of the pinch-off region under conditions of hot-electron transport, which is based upon the boundary condition of a value of the electric field at the pinch-off point determined uniquely by the current and hence applied voltages. The field patterns and charge density distribution throughout the pinch-off zone are predicted.

The solution allows one to determine the effect of the gate voltage on the channel field at pinch-off through the channel current itself.

Hence, there is direct feedback between the channel current and the extent of channel length modulation. It is demonstrated that for low gate voltages the channel length modulation may be severe and the current then shows a large variation with drain voltage. However, for larger gate voltages, electrostatic feedback to the channel becomes a dominant effect, and the current shows little variation with drain voltage, i.e., the degree of channel length modulation becomes very small.

A.2 General considerations and assumptions

It is helpful initially to review the important approximations and salient results that are obtained from simplified long-channel theory. For the purposes of the discussion, Fig. 16 shows a cross section of the device, which indicates the major current flow patterns and electric-field lines.

A.2.1 Existing theories

A basic assumption of MOS theory is that the gradual channel approximation (GCA) is valid throughout the unsaturated region. In

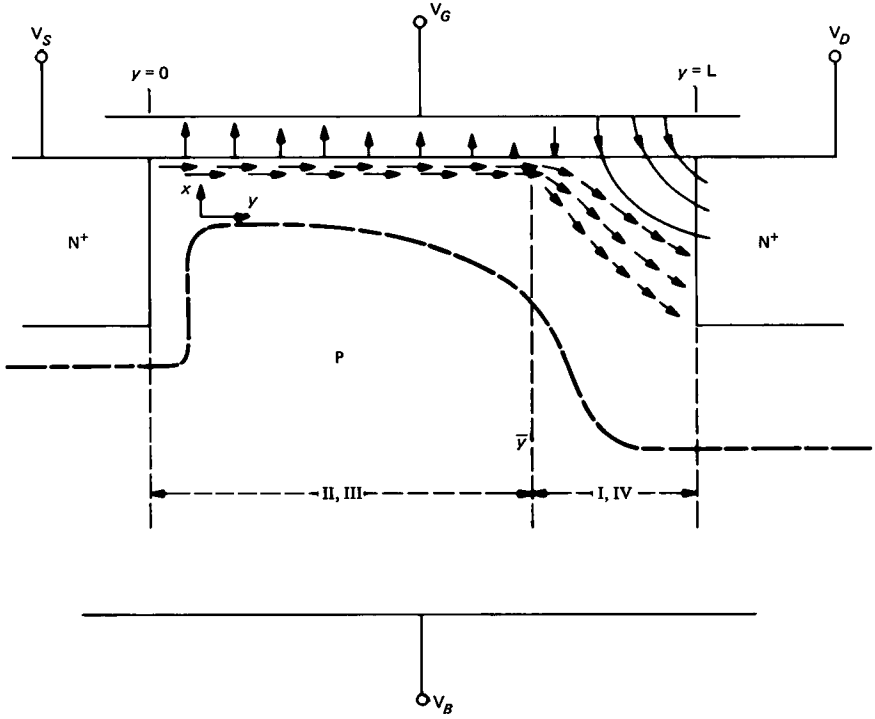

Fig. 16—MOS device cross section showing schematically the major lines of current flow (dashed arrows) and electric-field patterns (solid arrows).

a physical sense this approximation assumes that the transverse electric-field strength, $d\phi/dx$, is large compared to the longitudinal electric-field strength, $d\phi/dy$, in the channel region (ϕ is the electrostatic potential, x is the direction perpendicular to the gate, and y is the direction parallel to the channel). Hence, the charge $Q(y)$ in the channel may be determined by

$$Q(y) = C_o\,[V_G - V_{FB} - 2\phi_F - V(y)] - Q_B(y), \tag{54}$$

where C_o is the oxide capacitance, V_{FB} is the flatband voltage, $V(y)$ is the channel potential, and $Q_B(y)$ is the substrate depletion charge. The substrate charge is determined, using the depletion approximation, to be

$$Q_B(y) = \sqrt{2\epsilon q N_A[V(y) + 2\phi_F + V_{BS}]}, \tag{55}$$

where N_A is the substrate doping, q is the electronic charge, ϵ is the silicon dielectric permittivity, and

$$\phi_F = \frac{kT}{q} \ln\left(\frac{N_A}{n_i}\right). \tag{56}$$

The bulk Fermi potential is written in terms of n_i, the intrinsic silicon concentration. Another important assumption of the theory is that drift is the dominant conduction mechanism so that the continuity equation may be written

$$I = W\mu Q(y)\mathscr{E}(y), \tag{57}$$

where I is the source-to-drain current, W is the device width, μ is the channel mobility, and $\mathscr{E}(y)$ is the channel field. The channel current is obtained from (55), (56), and (57) to yield the familiar equation

$$I = \frac{\mu C_o W}{L}\left[\left(V_{GS} - V_{FB} - 2\phi_F - \frac{V_{DS}}{2}\right)V_{DS} - \frac{2\sqrt{2\epsilon q N_A}}{3C_o}\left\{(V_{DS} + V_{BS} + 2\phi_F)^{3/2} + (V_{BS} - 2\phi_F)^{3/2}\right\}\right]. \tag{58}$$

The phenomena of current saturation and pinch-off in the channel are predicted to occur when the free charge in the channel goes to zero. From a solution of (54) for

$$Q(y) = 0 \tag{59}$$

one obtains

$$V_{SAT} = V_G - V_{FB} - 2\phi_F + \frac{\epsilon q N_A}{C_o^2}\left[1 - \sqrt{1 + \frac{2C_o^2(V_G - V_{FB} + V_{BS})}{\epsilon q N_A}}\right]. \tag{60a}$$

The corresponding value of the saturated current is then found by substitution of (60a) into (58), and in the simple case, this procedure yields

$$I_{SAT} = \frac{\mu C_o W}{2L} (V_G - V_T)^2, \tag{60b}$$

where V_T is the simplified threshold value. In the general case, the more complex drain-voltage-dependent form of V_T[16] must be included in (60b), but the result is still evaluated in a straightforward manner.

A.2.2 Device characteristics in the saturation region

For voltages $V_{DS} > V_{DSAT}$, the drain depletion region extends towards the source. The position of pinch-off in the channel, occurring initially at the drain for $V(y) = V_{DSAT}$, will also move towards the source, since the condition (59) of zero charge may always be found somewhere in the channel for a function like (54), which is decreasing with voltage. Because of this feature, the potential at this position will remain constant for increasing drain voltage.

The use of (59) in the channel in saturation is, of course, an approximation that is used solely to determine the drain saturation voltage. Actually, the electron density is decreasing rapidly with distance in this region, and if we allow it to become arbitrarily small, then the electric field will have to become anomalously large if the drift component is to continue to provide continuity of current. We must therefore conclude that the drift component can no longer account for the total current flow and hence that the diffusion of carriers becomes an important conduction mechanism. This conclusion is supported by the numerical computations of other authors.[11,17] We show here that diffusion constitutes a specific fraction $\mathscr{R}$ of the current flow at the pinch-off point, and in Appendix B values for $\mathscr{R}$ are derived in the range of one third to one half.

As a basis for this work, we assume that we may represent the MOS device by four different regions of operation, as shown in Fig. 17. The quadrants are divided on the vertical axis by the subthreshold and above-threshold regions of operation and on the horizontal axis by the position of pinch-off in the channel. The quadrants are distinguished by the approximations designated in the figure, which denote for each quadrant the dominant conduction mechanism and the principle direction of the electric-field lines. In quadrants II and III the GCA applies, which may be stated

$$\frac{d^2\phi}{dx^2} \gg \frac{d^2\phi}{dy^2}. \tag{61}$$

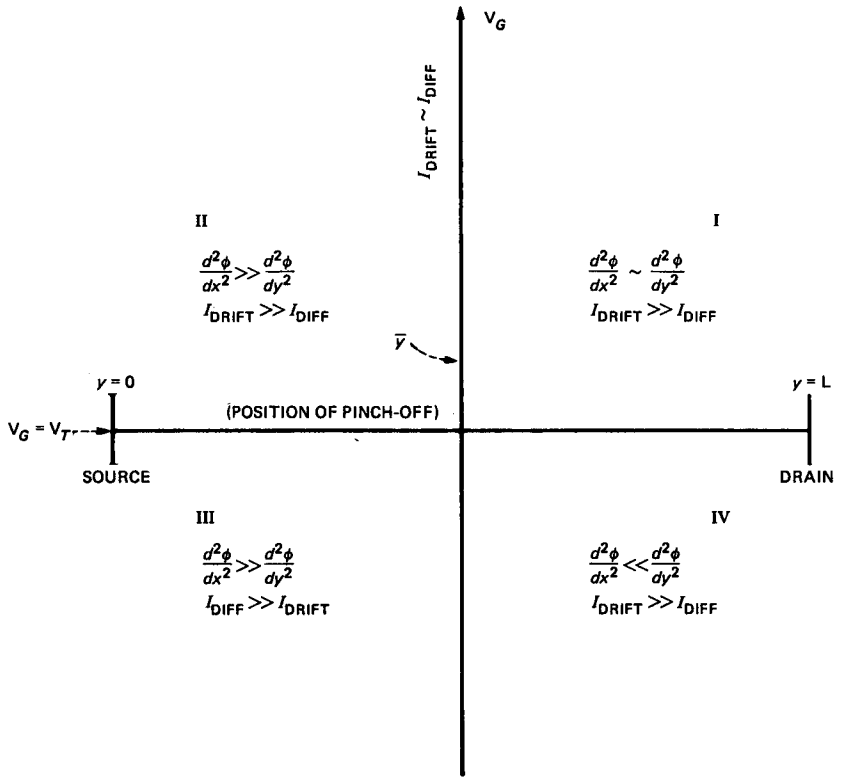

Fig. 17—Regions of operation for a MOSFET characterized by the approximations that may be used.

However, in quadrant I neither of these approximations is valid since the two terms are of the same order, which we write

$$d^2\phi/dx^2 \sim d^2\phi/dy^2. \tag{62}$$

The validity of (16) has been well established in MOS theory; the validity of (62) will be justified here by the solutions obtained. The other approximations indicated in Fig. 17 are that in quadrants II and IV we have

$$I_{DRIFT} \gg I_{DIFF} \tag{63a}$$

and in quadrant III we have

$$I_{DRIFT} \ll I_{DIFF}. \tag{63b}$$

In quadrant I we also have $I_{DRIFT} > I_{DIFF}$, but near the boundary we have

$$I_{\text{DRIFT}} \sim I_{\text{DIFF}}, \tag{63c}$$

which means that the components are comparable in size. The assumptions in quadrants II, III, and IV have been established by others; the validity of the assumptions in quadrant I and on its boundary will be considered here.

In the discussion of hot-electron transport we will use the mobility-field relationship[22]

$$\mu(\mathscr{E}) = \frac{\mu_0}{\left[1 + \left(\frac{\mathscr{E}}{\mathscr{E}_c}\right)^2\right]^{1/2}} \equiv \mu, \tag{64}$$

where $\mathscr{E}$ is the electric field and $\mathscr{E}_c$ is a critical field parameter that is related to the low-field mobility μ_0 by the relation

$$\mu_0 \mathscr{E}_c = v_s. \tag{65}$$

Also, the Einstein relationship will be assumed to hold; it states

$$D(\mathscr{E})/\mu(\mathscr{E}) = \frac{kT(\mathscr{E})}{q} \equiv \frac{kT}{q}, \tag{66}$$

where $D \equiv D(\mathscr{E})$ is the field-dependent diffusion coefficient and $T \equiv T(\mathscr{E})$ is the temperature of the electron in the hot-electron regime.

We will now discuss the conditions in regions I and II.

A.2.2.1 Region II. The current flow is approximately one-dimensional (parallel to the surface) because the dominant transverse field confines the carriers to a narrow potential well next to the surface. This region is described by (54), (55), and (57). Although drift is dominant in this region, it is also of interest to calculate the diffusion component. From (54) the mobile charge gradient is

$$\frac{dQ}{dy} = -(C_o + C_s)\frac{dV}{dy}, \tag{67}$$

where C_s, the space-charge capacitance, is

$$C_s \equiv C_s(y) = \sqrt{\frac{qN_A\epsilon}{2(V + 2\phi_F + V_{\text{BS}})}} \tag{68}$$

and $V \equiv V(y)$. The electric field is obtained from (57) as

$$\mathscr{E}(y) = \frac{I(1 - R)}{W\mu Q(y)}, \tag{69}$$

where R represents the fraction of the total current carried by diffusion. By using (69), (67), and (54), the diffusion component is

$$RI = I_{\text{DIFF}}$$

$$= \frac{D(\mathscr{E})I(C_o + C_s)(1 - R)}{\mu(\mathscr{E})[C_o(V_G - V_{\text{FB}} - 2\phi_F - V) - \sqrt{qN_A 2\epsilon(V + 2\phi_F + V_{\text{BS}})}]}. \quad (70)$$

A.2.2.2 Region I. As the diagram in Fig. 16 shows, the problem in this region becomes two-dimensional. In region II, the electric field is directed from the channel towards the gate but is steadily decreasing as the channel potential increases. Somewhere beyond $\bar{y}$, the pinch-off position, the field in the oxide is equal to zero, and beyond this position the electric field is directed from the gate into the silicon; hence, many of the field lines terminate on the drain electrode, as shown, since the drain potential exceeds the gate potential in this region of operation. Because the transverse field in region I no longer creates a potential well, the mobile carriers may flow away from the surface as they approach the drain, resulting in a two-dimensional current flow.

The main features of this representation are depicted in Fig. 18, which shows an expanded view of region I. Consider the flow of carriers at some position y in region I. Because of the strong electric field indicated in Fig. 18, which is directed from the gate towards the substrate in region I, the electrons are forced away from the surface. Therefore, the density of electrons will be reduced at the surface. For sufficiently large values of x (i.e., deeper in the Si) the electron density must decrease again to its substrate value, and so we conclude that the electron density must exhibit a maximum as a function of x for a given value of y. We will, therefore, represent the electron density in a general way by a function of the form

$$n(\omega, r) = N(r)\exp\{-A(r)[\omega - \omega_0^*(r)]^2 - B(r)[\omega - \omega_1^*(r)]^4 - \cdots\}, \quad (71)$$

where the higher-order terms in the distribution would be required for strong deviations from the normal case. The expansion for the electron density has been written for the generalized coordinate axes ω, r rather than for x, y to allow for the fact that the x, y system may not be the most convenient one in which to describe the distribution. The y or r dependence has been incorporated into this form through the parameters $N(r)$, $A(r)$, $B(r)$, $\omega_0^*(r)$, and $\omega_1^*(r)$. In this way, the median value, the mean position, and the width of the distribution may change with y or r. The functions $\omega^*(r)$ describe the locus of the mean of the electron distribution between the pinch-off point and the drain as illustrated in Fig. 18. In the approach taken here, we have only three physical relationships available to us for the determination of the electron distribution, so we will restrict (71) to three unknown function: i.e.,

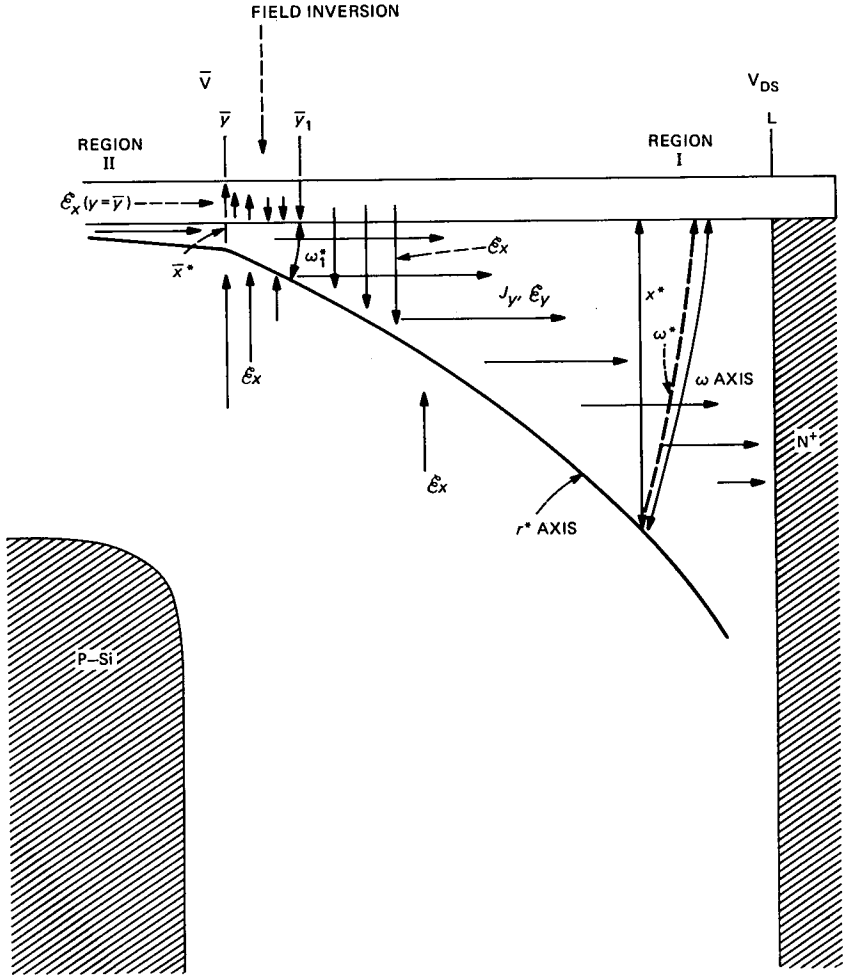

Fig. 18—Expanded view of the pinch-off zone with important physical and electrical parameters.

$$n \equiv n(\omega, r) = N(r)\exp\left\{-\frac{[\omega - \omega^*(r)]^2}{2\sigma^2(r)}\right\}, \tag{72}$$

where $A(r)$ has been rewritten as $1/2\sigma^2(r)$ and in which the functions $N(r)$, $\sigma(r)$, $\omega^*(r)$ remain to be determined from three physical relationships, which will now be described.

The first relationship is the description of the current flow mechanism and quite generally it is

$$\frac{I}{qW} = \int_0^\infty D\frac{dn}{dr}\,d\omega + \int_0^\infty \mu\mathscr{E}_r n\,d\omega \tag{73}$$

for any y between $\bar{y}$ and L where $\omega = 0$ is the silicon surface. The second relationship is the law of current continuity. Physically speaking, the current flow in the x direction must provide the current density gradient by means of which the current density in the y direction is able to change. This condition is stated formally as

$$\nabla \cdot \mathbf{J} = 0 \tag{74}$$

or, for the case being considered,

$$\frac{dJ_x}{dx} = -\frac{dJ_y}{dy},$$

where J_x and J_y are the current density components in the x and y directions, respectively. The third relationship is the two-dimensional Poisson equation

$$\nabla \cdot \mathscr{E} = \frac{q}{\epsilon} \rho, \tag{75}$$

which is written here as

$$\frac{\partial \mathscr{E}_x}{\partial x} + \frac{\partial \mathscr{E}_y}{\partial y} = -\frac{q}{\epsilon} (N_A + n), \tag{76}$$

subject to the boundary condition

$$\mathscr{E}_y = \bar{\mathscr{E}}, \qquad y = \bar{y}, \tag{77}$$

where $\bar{y}$ is the pinch-off point and $\bar{\mathscr{E}}$ is the value of longitudinal electric field in the channel at the pinch-off point. Both x and y components need to be retained in (76) since the concentration of field lines is comparable in both x and y directions. The value of $\mathscr{E}_y$ along r^* will be discussed later.

In the section of the channel between the source and the field-inversion point, (64) is a one-dimensional relationship between the channel field and the mobility. Between the field inversion point and the drain, the mobility becomes a scalar field determined by the vector

$$\mathscr{E} = \mathbf{r}\mathscr{E}_r + \boldsymbol{\omega}\mathscr{E}_\omega.$$

Since

$$\mathscr{E}^2 = \mathscr{E}_r^2 + \mathscr{E}_\omega^2,$$

(64) becomes

$$\mu = \frac{\mu_0}{\left[1 + \left(\frac{\mathscr{E}_\omega}{\mathscr{E}_c}\right)^2 + \left(\frac{\mathscr{E}_r}{\mathscr{E}_c}\right)^2\right]^{1/2}}. \tag{78a}$$

Also, the one-dimensional hot-electron temperature

$$T_e = T_0 \left[1 + \left(\frac{\mathscr{E}}{\mathscr{E}_c} \right)^2 \right] \tag{78b}$$

may be expressed

$$\beta_e = \beta_0 + \beta_0 \left(\frac{\mathscr{E}_\omega}{\mathscr{E}_c} \right)^2 + \beta_0 \left(\frac{\mathscr{E}_r}{\mathscr{E}_c} \right)^2, \tag{78c}$$

which is a scalar field, where $\beta_e = kT_e/q$. Therefore, the mobility (78a) may be expressed in terms of the temperature T_e or, equivalently, the voltage β_e using (78a) and (78c) as

$$\frac{\mu}{\mu_0} = \sqrt{\frac{\beta_0}{\beta_e}}. \tag{78d}$$

A.3 Determination of physical parameters in pinch-off operation

A.3.1 Charge, voltage, and field at the pinch-off position

The boundary between regions II and I will be defined as the pinch-off position, and all variables at this point will be designated with a bar. As the position $\bar{y}$ is approached from the source, the drift current will be a decreasing function of y and the diffusion current will be an increasing function of y. At $\bar{y}$, the drift and diffusion components of the current are

$$I_{\text{DRIFT}} = (1 - \bar{R})I \tag{79}$$

and

$$I_{\text{DIFF}} = \bar{R}I. \tag{80}$$

The method for the determination of $\bar{R}$ is outlined in Appendix B [see (200)]. It is determined only by doping and oxide thickness and has typical values of one third to one half. From (70) and (66) we have

$$C_o(V_G - V_{\text{FB}} - 2\phi_F - \bar{V}) - \sqrt{qN_A 2\epsilon(\bar{V} + 2\phi_F + V_{\text{BS}})} = \frac{1 - \bar{R}}{\bar{R}} \frac{kT}{q} [C_o + C_s(\bar{y})]. \tag{81}$$

Equation (81) may also be written approximately as

$$V_{\text{SAT}} \equiv \bar{V} = V_G - V_T - \left(\frac{1 - \bar{R}}{\bar{R}} \right) \frac{kT}{q} \left(\frac{C_o + \bar{C}_s}{C_o} \right). \tag{82}$$

The relationship (81) is identical to that obtained in the original MOS theory[23] except for the extra term on the right-hand side. In other words, rather than using (59) to determine the pinch-off voltage, we are taking account of the charge in the channel at pinch-off, and from (81) and (54) the charge is determined to be

$$\bar{Q} = \left(\frac{1 - \bar{R}}{\bar{R}}\right) \frac{kT}{q} (C_o + \bar{C}_s). \tag{83}$$

For $T = T_0$ (the lattice temperature, as is the case in long-channel devices), $\bar{Q}$ is small (provided that $\bar{R} \gtrsim 0.3$) and (82) yields the conventional result. The significance of (83)* is that we can now determine the longitudinal field in the channel at pinch-off, and from (83) and (69) it is

$$\bar{\mathscr{E}}_y = \frac{I\bar{R}}{W\mu \dfrac{kT}{q} (C_o + \bar{C}_s)}. \tag{84}$$

Equation (84) applies for any set of voltage variables in pinch-off operation. The dependence of $\bar{\mathscr{E}}$ on μ from (64) could be substituted here, and since the current at the onset of pinch-off operation (i.e., the boundary between triode and saturation regions) is known, then $\bar{\mathscr{E}}$ at the onset of pinch-off may also be expressed uniquely in terms of the applied voltages. It is therefore an ideal boundary condition for the pinch-off zone.

It is also of some interest to determine the sign and magnitude of $\mathscr{E}_x$ at the pinch-off point. From the continuity of the divergence of $\mathscr{E}$ we can relate the oxide and substrate fields as

$$\epsilon \mathscr{E}_{x_{si}} = \epsilon_{ox} \mathscr{E}_{ox}, \tag{85}$$

where $\mathscr{E}_{x_{si}}$ may be determined, as shown by Pao,[23] by a detailed solution of Poisson's equation in the x direction in which both mobile and fixed charge components are retained. An equivalent representation of $\mathscr{E}_{ox}$ is

$$\mathscr{E}_{ox} = \frac{V_G - \phi_s - V(y) - V_{FB}}{t_{ox}}, \tag{86}$$

where ϕ_s is the surface potential at the source. As is well known, ϕ_s takes the value of $2\phi_F$ at threshold and increases only slightly for additional increases in gate voltage. Therefore, using (82) in (86) we find that

$$\bar{\mathscr{E}}_{ox} = \frac{\bar{Q} + \sqrt{2\epsilon q N_A (2\phi_F + V_{BS})} + V_{FB}/C_o}{\epsilon_{ox}} \tag{87a}$$

and

$$\bar{\mathscr{E}}_{x_{si}} = \frac{\bar{Q} + \sqrt{2\epsilon q N_A (2\phi_F + V_{BS})} + V_{FB}/C_o}{\epsilon}. \tag{87b}$$

* To include short-channel effects it is necessary to replace $\bar{C}_s$ with $\bar{C}_s F$.

Hence, we conclude that both $\mathscr{E}_{ox}$ and $\mathscr{E}_{x_{si}}$ start from large positive values at the source and decrease to the much smaller values given by (87a) and (87b) at the pinch-off position. These values are practically constant with increasing gate voltage except for the small increases in ϕ_S above $2\phi_F$. The fields are directed towards the gate electrode at $\bar{y}$. These results are illustrated by the plots of Pao,[23] which are reproduced in Fig. 19. We have extended the field values past the pinch-off point

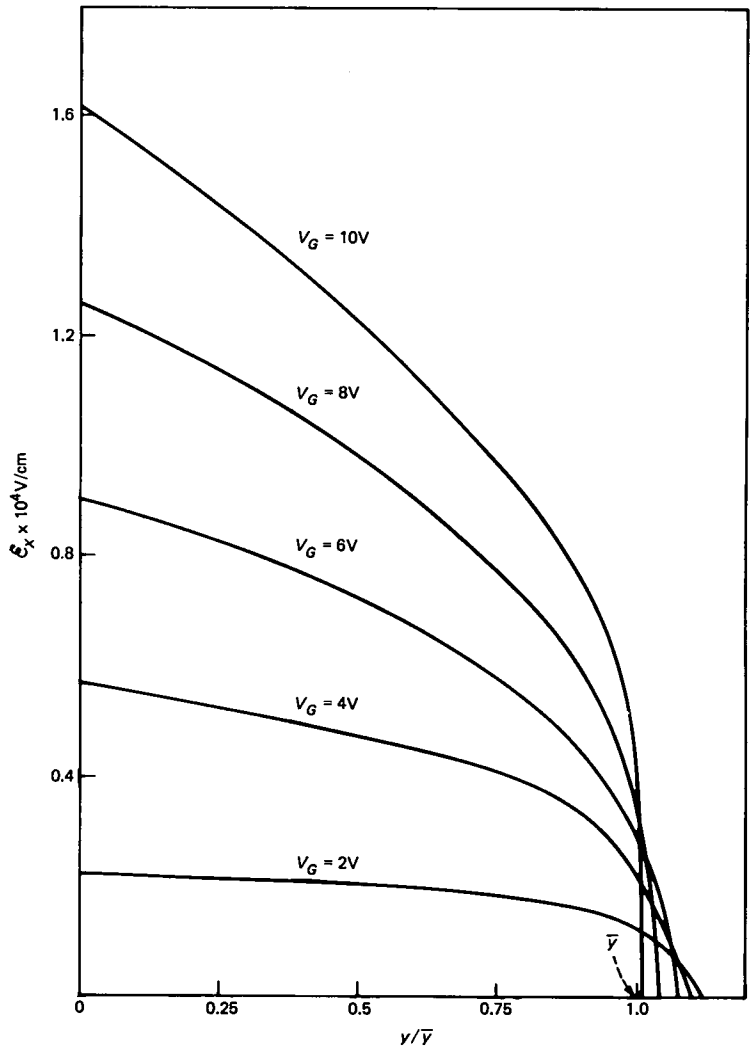

Fig. 19—Variation of transverse electric-field strength in the silicon at the surface as a function of distance along the surface from the source and through the pinch-off point parameters $t_{ox} = 2000$Å, $N_D = 4.6 \times 10^{14}$ cm^{-3}, and $L = 5$ μm.

to show that a reversal in sign of $\mathscr{E}_x$ occurs as shown by MaGowan's[24] numerical solution.[25]

A.3.2 Formulation of the differential equations beyond the pinch-off point

Using the relationships described earlier, we may now find the variation of electron density and electric fields in region I. The two-dimensional nature of the problem, which we will now discuss, applies only after field inversion has occurred. The section between pinch-off and field inversion will be considered later.

Consider the vector field for the current flow in region I,

$$\mathbf{J} = J_x\mathbf{x} + J_y\mathbf{y}, \tag{88}$$

where $\mathbf{x}$ and $\mathbf{y}$ are unit vectors in the x and y directions. The vector components are the current densities in the x and y directions and may be written generally as

$$\frac{J_y}{q} = D\frac{dn}{dy} + \mu n\mathscr{E}_y \tag{89}$$

for the y direction and

$$\frac{J_x}{q} = D\frac{dn}{dx} + \mu n\mathscr{E}_x \tag{90}$$

for the x direction. At this point we would like to choose a set of coordinate axes to most suitably represent the current flow. We know that along the oxide-silicon interface the flow is parallel to the interface and that along the streamline the flow is in the direction of the streamline. This fact suggests that we should use polar coordinates $\mathbf{r}$ and θ to represent the problem, where $\mathbf{r}$ is the vector extending from the field inversion point at the surface to some point in the pinch-off zone and θ is the angle between the vector $\mathbf{r}$ and the interface. The field inversion point is considered the origin or source point because it is the point in the channel where carriers first depart from the surface (from a charge sheet point of view). In the representation of the electron distribution by (72), then, we are taking ω to be the arc length measured from the interface for a fixed r and a variable θ. Since ω^*, σ, and N are functions of r only, the assumption becomes that we can represent the electrons by a Gaussian distribution that extends in a curvilinear fashion around a circular contour. In Fig. 20 the polar coordinates are shown schematically in an expanded view with all of the r axes emanating from the field-inversion point, y_i.

In polar coordinates (89) and (90) become

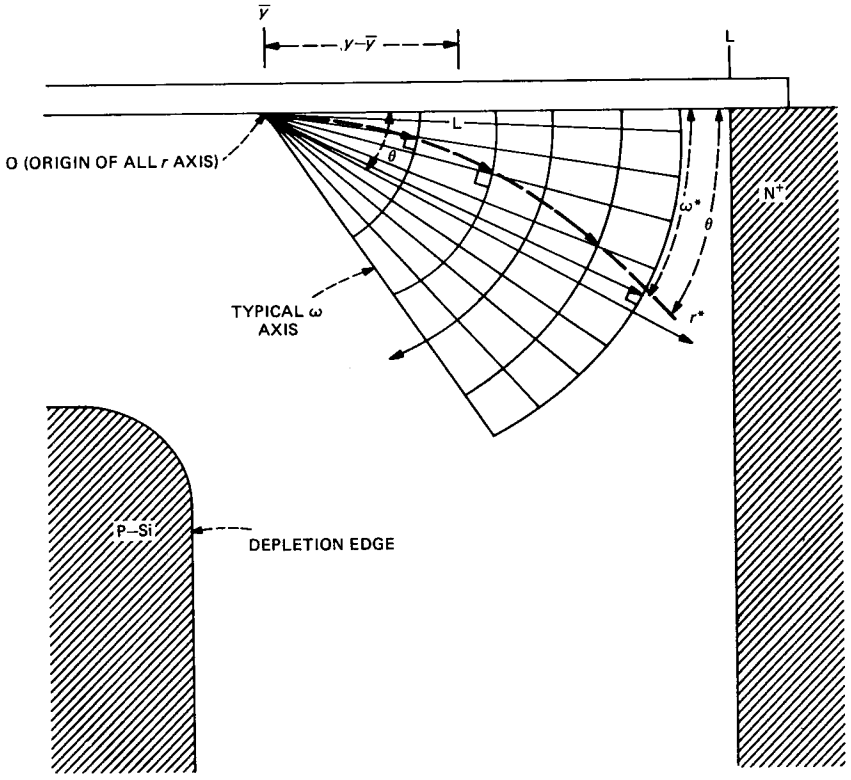

Fig. 20—Curvilinear coordinate axes in the pinch-off zone.

$$\frac{J_r}{q} = D\frac{\partial n}{\partial r} + \mu n \mathscr{E}_r \tag{91a}$$

$$\frac{J_\omega}{q} = D\frac{1}{r}\frac{\partial n}{\partial \theta} + \mu n \mathscr{E}_\omega. \tag{91b}$$

Then using n as represented by (72) in (91) and utilizing (66), we obtain for the current components

$$J_\omega = q\mu N \exp(-\eta)\left\{\mathscr{E}_\omega - \beta_e \frac{(\omega - \omega^*)}{\sigma^2}\right\} \tag{92}$$

and

$$J_r = q\mu N \exp(-\eta) \cdot\left\{\mathscr{E}_r + \beta_e\left[\frac{N'}{N} + \frac{(\omega - \omega^*)}{\sigma^2} S_\omega - \frac{(\omega - \omega^*)^2}{2}\left(\frac{1}{\sigma^2}\right)'\right]\right\}, \tag{93}$$

where

$$\eta = \frac{(\omega - \omega^*)^2}{2\sigma^2}, \qquad \mathscr{E}_\omega = -\frac{1}{r}\frac{\partial \psi}{\partial \theta},$$

$$\mathscr{E}_r = -\frac{\partial \psi}{\partial r}, \qquad S_\omega = \frac{\partial \omega^*}{\partial r} - \frac{\partial \omega}{\partial r} \tag{94}$$

and the prime denotes differentiation with respect to r and ψ in the two-dimensional scalar potential. From the divergence condition (74) we find

$$\left(\frac{1}{r}\frac{\partial n}{\partial \theta} + \frac{n}{\mu}\frac{1}{r}\frac{\partial \mu}{\partial \theta}\right)\frac{1}{r}\frac{\partial \phi}{\partial \theta} + \left(\frac{\partial n}{\partial r} + \frac{n}{\mu}\frac{\partial \mu}{\partial r}\right)\frac{\partial \phi}{\partial r} = -n\nabla^2\phi, \tag{95a}$$

where we can consider ϕ as some pseudo-potential field defined by the relations

$$-\frac{1}{r}\,\partial\phi/\partial\theta = \mathscr{E}_\omega - \beta_e\left(\frac{\omega - \omega^*}{\sigma^2}\right),$$

$$-\partial\phi/\partial r = \mathscr{E}_r + \beta_e\left[\frac{N'}{N} + \left(\frac{\omega - \omega^*}{\sigma^2}\right)S_\omega - \frac{(\omega - \omega^*)^2}{2}\left(\frac{1}{\sigma^2}\right)'\right]. \tag{95b}$$

Using (92) and (93) in (95a), we find the divergence relationship to be

$$\left\{\mathscr{E}_\omega - \beta\frac{(\omega - \omega^*)}{\sigma^2}\right\}\left\{-\frac{(\omega - \omega^*)}{\sigma^2} + \frac{1}{\mu}\frac{1}{r}\frac{\partial \mu}{\partial \theta}\right\}$$

$$+\left\{\mathscr{E}_r + \beta_e\left[\frac{N'}{N} + \frac{(\omega - \omega^*)}{\sigma^2}S_\omega - \frac{(\omega - \omega^*)^2}{2}\left(\frac{1}{\sigma^2}\right)'\right]\right\}$$

$$\cdot\left[\frac{N'}{N} + \left(\frac{\omega - \omega^*}{\sigma^2}\right)S_\omega - \frac{(\omega - \omega^*)^2}{2}\left(\frac{1}{\sigma^2}\right)' + \frac{1}{\mu}\frac{\partial \mu}{\partial r}\right] = \nabla^2\phi. \tag{96}$$

At this point we need a representation for the electric fields $\mathscr{E}_\omega$ and $\mathscr{E}_r$. Consider $\mathscr{E}_\omega$ first. If drift and diffusion were equal and n were described by (72), then one could show that $\mathscr{E}_\omega = \beta_e[(\omega - \omega^*)/\sigma^2)]$. We will assume that even though drift and diffusion are not equal, the functional form of $\mathscr{E}_\omega$ remains the same under conditions of current flow and is altered only in magnitude so that we may write

$$\mathscr{E}_\omega = K_1\frac{\beta_e(\omega - \omega^*)}{\sigma^2}, \tag{97}$$

where K_1 is some number that is constant throughout the pinch-off zone. We will take the representation a step further by assuming that all electric fields may be represented by these functions so that we can

write

$$\mathscr{E}_r = -K_2 \frac{\beta_e \omega^*}{\sigma^2} \quad \text{and} \quad \mathscr{E}_r^* = -K_3 \frac{\beta_e \omega^*}{\sigma^2} \tag{98}$$

for the surface and along the streamline, respectively, where K_2 and K_3 are constants in the pinch-off zone, which must be determined. From (97) and (98) we obtain

$$\frac{1}{r}\frac{\partial \mathscr{E}_\omega}{\partial \theta} = \frac{K_1 \beta_e}{\sigma^2} + K_1 \frac{(\omega - \omega^*)}{\sigma^2}\frac{1}{r}\frac{\partial \beta_e}{\partial \theta} \tag{99}$$

and

$$\frac{\partial \mathscr{E}_r}{\partial r} = -\frac{K_2 \beta_e}{\sigma^2} S_\omega - K_2 \beta_e \omega^* \frac{d}{dr}\left(\frac{1}{\sigma^2}\right) - K_2 \frac{\omega^*}{\sigma^2}\frac{\partial \beta_e}{\partial r}, \tag{100}$$

where

$$S_\omega^* = \frac{\partial \omega^*}{\partial r}.$$

Substituting (97) and (98) into (78c), we obtain for the surface (and for the region very close to the surface)

$$\beta_e = \frac{A}{2} + \sqrt{\left(\frac{A}{2}\right)^2 - \beta_0 A}, \tag{101}$$

where

$$A = \frac{(\sigma^2 \mathscr{E}_c)^2}{\beta_0 [K_1^2(\omega - \omega^*)^2 + K_2^2 \omega^{*2}]}. \tag{102}$$

In the same way, in the vicinity of the streamline we have

$$\beta_e^* = \frac{A^*}{2} + \sqrt{\left(\frac{A^*}{2}\right)^2 - \beta_0 A^*}, \tag{103}$$

where

$$A^* = \frac{(\sigma^2 \mathscr{E}_c)^2}{\beta_0 [K_1^2(\omega - \omega^*)^2 + K_3^2 \omega^{*2}]}. \tag{104}$$

Differentiating (101), we have

$$\frac{1}{r}\frac{\partial \beta_e}{\partial \theta} = \frac{1}{A}\frac{1}{r}\frac{\partial A}{\partial \theta}\left[\beta_e + \frac{\beta_0 A}{2\sqrt{\left(\frac{A}{2}\right)^2 - \beta_0 A}}\right] \simeq \frac{1}{A}\frac{1}{r}\frac{\partial A}{\partial \theta}\beta_e \tag{105}$$

and

$$\frac{\partial \beta_e}{\partial r} = \frac{1}{A}\frac{\partial A}{\partial r}\left[\beta_e + \frac{\beta_0 A}{2\sqrt{\left(\frac{A}{2}\right)^2 - \beta_0 A}}\right] \simeq \frac{1}{A}\frac{\partial A}{\partial r}\beta_e \tag{106}$$

and the approximation is good throughout most of the hot-electron regime because $A \gg \beta_0$. In (105) and (106) the derivatives of A are written

$$\frac{1}{A}\frac{1}{r}\frac{\partial A}{\partial \theta} = -\frac{2K_1^2(\omega - \omega^*)}{K_1^2(\omega - \omega^*)^2 + K_2^2\omega^{*2}} \tag{107a}$$

and

$$\frac{1}{A}\frac{\partial A}{\partial r} = \frac{-2[-K_1^2(\omega - \omega^*)S_\omega + K_2^2\omega^* S_\omega^*]}{K_1^2(\omega - \omega^*)^2 + K_2^2\omega^{*2}} - 2\sigma^2\frac{d}{dr}\left(\frac{1}{\sigma^2}\right), \tag{107b}$$

where we have used

$$\sigma' = -\frac{\sigma^3}{2}\frac{d}{dr}\left(\frac{1}{\sigma^2}\right) \tag{107c}$$

to transform the term in σ'.

Using (78c) for the mobility we now find

$$\frac{1}{\mu}\frac{1\partial \mu}{r\partial \theta} = -\frac{1}{2}\frac{1}{\beta_e}\frac{1}{r}\frac{\partial \beta_e}{\partial \theta} \tag{108}$$

$$\frac{1}{\mu}\frac{\partial \mu}{\partial r} = -\frac{1}{2}\frac{1}{\beta_e}\frac{\partial \beta_e}{\partial r}, \tag{109}$$

so that the derivatives of the mobility may be evaluated using (105) and (106).

Poisson's equation may now be reduced to a simpler form using these results. Substituting (107c) and (107b) into (105) and (106), we find

$$\frac{1}{\beta_e}\frac{1}{r}\frac{\partial \beta_e}{\partial \theta} = \frac{2K_1^2}{\omega^*(K_1^2 + K_2^2)}, \qquad \frac{1}{\beta_e^*}\frac{1}{r}\frac{\partial \beta_e^*}{\partial \theta} = 0 \tag{110}$$

and

$$\frac{1}{\beta_e}\frac{\partial \beta_e}{\partial r} = \frac{1}{\beta_e^*}\frac{\partial \beta_e^*}{\partial r} = -\frac{2S_\omega^*}{\omega^*} - 2\sigma^2\frac{d}{dr}\left(\frac{1}{\sigma^2}\right). \tag{111}$$

The Poisson equation (76) in cylindrical coordinates is

$$\frac{\partial \mathscr{E}_r}{\partial r} + \frac{1}{r}\mathscr{E}_r + \frac{1}{r}\frac{\partial \mathscr{E}_\omega}{\partial \theta} = -\frac{q}{\epsilon}(N_A + n). \tag{112}$$

Using (110) and (111) in (99) and (100) to find $\partial \mathscr{E}_\omega/\partial\omega$ and $\partial \mathscr{E}_r/\partial r$, we can reduce (112) to an equation in $(1/\sigma^2)$ for the surface ($\omega = 0$), which is

$$K_2\omega^* \frac{d}{dr}\left(\frac{1}{\sigma^2}\right) + \frac{K_2 S_\omega^*}{\sigma^2} - \frac{K_2\omega^*}{r\sigma^2} + \left(\frac{K_2^2 - K_1^2}{K_1^2 + K_2^2}\right)\frac{K_1}{\sigma^2} = -\frac{1}{\lambda_e^2}. \quad (113)$$

Using symmetrical arguments for the streamline locus we have

$$K_3\omega^* \frac{d}{dr}\left(\frac{1}{\sigma^2}\right) + K_3\frac{S_\omega^*}{\sigma^2} - \frac{K_3\omega^*}{r\sigma^2} + \frac{K_1}{\sigma^2} = \frac{-1}{\lambda_e^{*2}}. \quad (114)$$

The parameters λ_e, λ_e^* are modified Debye lengths

$$\lambda_e = \sqrt{\frac{\epsilon\beta_e}{q(N_A + Ne^{-\eta})}} \quad (115)$$

and

$$\lambda_e^* = \sqrt{\frac{\epsilon\beta_e}{q(N_A + N)}}$$

due to the presence of mobile charge. The traditional Debye length is given by

$$\lambda_0 = \sqrt{\frac{\epsilon\beta_e}{qN_A}}.$$

A useful result that we shall need later is found by subtracting (113) from (114) to give

$$K_3\left(1 - \frac{2K_1^2}{K_1^2 + K_2^2}\right) - K_2 = \frac{K_2}{K_1}\left(\frac{\sigma}{\lambda_e^*}\right)^2 - \frac{K_3}{K_1}\left(\frac{\sigma}{\lambda_e}\right)^2. \quad (116)$$

We now employ the condition

$$\oint \mathscr{E} ds = 0 \quad (117)$$

over any path in the pinch-off zone since ψ is a conservative field. Using (97), (98), and (117) we have

$$\int_0^{\omega^*} K_1\beta_e \frac{(\omega - \omega^*)}{\sigma^2}\, d\omega = \int_0^r (K_3 - K_2)\beta_e \frac{\omega^*}{\sigma^2}\, dr'. \quad (118)$$

To simplify this result, we will assume that we may ignore the dependence of β_e upon ω on the left-hand side. Then by integrating over ω, differentiating with respect to r, and using (110) and (111), we find

$$\omega^* \frac{d}{dr}\left(\frac{1}{\sigma^2}\right) = -\frac{2(K_3 - K_2)}{K_1 \sigma^2}. \tag{119}$$

Substituting (118) into (112) we find

$$\frac{d\omega^*}{dr} - \frac{\omega^*}{r} = \frac{2(K_3 - K_2)}{K_1} - \frac{K_1}{K_2}\left(1 - \frac{2K_1^2}{K_1^2 + K_2^2}\right) - \frac{1}{K_2}\left(\frac{\sigma}{\lambda_e}\right)^2. \tag{120}$$

Substituting (118) into (113) we have

$$\frac{d\omega^*}{dr} - \frac{\omega^*}{r} = \frac{2(K_3 - K_2)}{K_1} - \frac{K_1}{K_3} - \frac{1}{K_3}\left(\frac{\sigma}{\lambda_e^*}\right)^2. \tag{121}$$

Subtracting (119) from (120) would yield (116) as before. Returning to (110) we use (118) to give

$$\frac{1}{\beta_e^*}\frac{\partial \beta_e^*}{\partial r} = \frac{1}{\beta_e}\frac{\partial \beta_e}{\partial r} = \frac{2}{\omega^*}\left[2\frac{(K_3 - K_2)}{K_1} - S_\omega^*\right]. \tag{122}$$

The results (109) and (121) are then used in (108) and (109) to obtain functions of K_1, K_2, and K_3, which are

$$\frac{1}{\mu}\frac{1}{r}\frac{\partial \mu}{\partial \theta} = \frac{-K_1^2}{(K_1^2 + K_2^2)} \cdot \frac{1}{\omega^*} \qquad (\omega = 0, \text{ silicon interface}), \tag{123a}$$

$$\frac{1}{\mu^*}\frac{1}{r}\frac{\partial \mu^*}{\partial \theta} = 0 \qquad (\omega - \omega^*, \text{ streamline}), \tag{123b}$$

and

$$\frac{1}{\mu^*}\frac{\partial \mu^*}{\partial r} = \frac{1}{\mu}\frac{\partial \mu}{\partial r} = -\frac{1}{\omega^*}\left[2\left(\frac{K_3 - K_2}{K_1}\right) - S_\omega^*\right]. \tag{123c}$$

We may now return to the evaluation of (96). The right-hand side of (96) may be evaluated using (95b) to be

$$\begin{aligned}
\nabla^2\phi = {} & -\frac{1}{r}\frac{\partial \mathscr{G}\omega}{\partial \theta} + \frac{\beta_e}{\sigma^2} + \left(\frac{\omega - \omega^*}{\sigma^2}\right)\frac{1}{r}\frac{\partial \beta_e}{\partial \theta} - \frac{\partial \mathscr{G}_r}{\partial r} \\
& - \left[\frac{N'}{N} + \left(\frac{\omega - \omega^*}{\sigma^2}\right)S_\omega - \frac{(\omega - \omega^*)^2}{2}\left(\frac{1}{\sigma^2}\right)'\right]\frac{\partial \beta_e}{\partial r} - \frac{\beta_e \partial}{\partial r}\left(\frac{N'}{N}\right) \\
& + \frac{\beta_e}{\sigma^2}S_\omega^2 - \beta_e(\omega - \omega^*)S_\omega\left(\frac{1}{\sigma^2}\right)' - \beta_e(\omega - \omega^*)S_\omega\left(\frac{1}{\sigma^2}\right)' \\
& + \beta_e\frac{(\omega - \omega^*)^2}{2}\left(\frac{1}{\sigma^2}\right)'' - \beta_e\frac{(\omega - \omega^*)}{\sigma^2}\frac{\partial^2 \omega^*}{\partial r^2} - \frac{\mathscr{G}_r}{r} \\
& - \frac{\beta_e}{r}\left[\frac{N'}{N} + \frac{(\omega - \omega^*)}{\sigma^2}S_\omega - \frac{(\omega - \omega^*)^2}{2}\left(\frac{1}{\sigma^2}\right)'\right].
\end{aligned} \tag{124}$$

We need to consider (96) along the two separate loci. Along the main streamline ω^*, we have $\omega - \omega^* = 0$, and we will neglect the diffusion component N'/N compared to the drift component. Using (124) and (122) in (96) with $\mathscr{G}_r = \mathscr{G}_r^*$, we then find

$$\mathscr{G}_r^* \left[\frac{N'}{N} + \frac{1}{\mu^*}\frac{\partial \mu^*}{\partial r}\right] = \frac{\beta_e}{\lambda_e^{*2}} + \frac{\beta_e}{\sigma^2}(1 + S_\omega^2) - \frac{1}{r}\frac{2\beta_e}{\omega^*}\left[2\left(\frac{K_3 - K_2}{K_1}\right) - S_\omega\right], \quad (125)$$

where we have considered the second derivatives of β_e and ω^* to be unimportant. The second locus is along the interface ($\omega = 0$), and in that case (96), (97), (98), (110), and (123) are used to give

$$\left(-K_1\frac{\omega^*}{\sigma^2} + \frac{\omega^*}{\sigma^2}\right) \times \left(\frac{\omega^*}{\sigma^2} - \frac{K_1^2}{K_1^2 + K_2^2}\frac{1}{\omega^*}\right) + \left\{-K_2\frac{\omega^*}{\sigma^2} - \frac{\omega^* S_\omega}{\sigma^2} - \frac{\omega^2}{2}\left(\frac{1}{\sigma^2}\right)'\right\} \cdot \left\{-\frac{\omega}{\sigma^2}S_\omega - \frac{\omega^2}{2}\left(\frac{1}{\sigma^2}\right)' - \frac{1}{\omega}\left[2\left(\frac{K_3 - K_2}{K_1}\right) - S_\omega\right]\right\} = \nabla^2\phi, \quad (126)$$

which may be rewritten, using (119), as

$$(-K_1 + 1)\left(1 - \frac{K_1^2}{K_1^2 + K_2^2}\left(\frac{\sigma}{\omega^*}\right)^2\right) = \left\{K_2 + S_\omega - \left(\frac{K_3 - K_2}{K_1}\right)\right\} \cdot \left\{-S_\omega + \left(\frac{K_3 - K_2}{K_1}\right) - \left(\frac{\sigma}{\omega^*}\right)^2\left[2\left(\frac{K_3 - K_2}{K_1}\right) - S_\omega^*\right]\right\} + \left(\frac{\sigma^2}{\omega^*}\right)^2 \nabla^2\phi. \quad (127)$$

We will now assume that the electron distribution is well localized in the pinch-off zone, so that $\sigma \ll \omega^*$.

Then the term $\nabla^2\phi$ may be dropped and (127) may be reduced to

$$K_2 = \left(\frac{K_1}{K_3 - K_2}\right)\frac{\left[1 + \left(\frac{K_3 - K_2}{K_1} - S_\omega\right)^2 - K_1\right]}{\left(1 - K_1\frac{S_\omega}{K_3 - K_2}\right)}. \quad (128)$$

Another relationship between the electron distribution parameters is found from the total current (73). Using (78a) and (72) in (73),

$$\frac{I}{qW} = N\mu_0 \left\langle \frac{\mathscr{E}_r}{\sqrt{1 + \left(\frac{\mathscr{E}_r}{\mathscr{E}_0}\right)^2 + \left(\frac{\mathscr{E}_\omega}{\mathscr{E}_c}\right)^2}} \right\rangle \int_{-\infty}^{\infty} e^{-\eta} d\omega$$

$$+ D_0 \left\langle \sqrt{1 + \left(\frac{\mathscr{E}_r}{\mathscr{E}_c}\right)^2 + \left(\frac{\mathscr{E}_\omega}{\mathscr{E}_c}\right)^2} \right\rangle \frac{d}{dr} \int_{-\infty}^{\infty} Ne^{-\eta} d\omega, \quad (129)$$

where the notation $\langle\ \rangle$ represents an average over the ω axis. The assumption is that the respective arguments are weak functions of ω compared to the exponential term. Then the equation may be written in simplified form as

$$\frac{\mathscr{E}_r^*}{\beta_e} \int Ne^{-\eta} d\omega + \frac{d}{dr} \int Ne^{-\eta} d\omega = \frac{I}{qWD_0 \sqrt{1 + \left(\frac{\mathscr{E}_r^*}{\mathscr{E}_c}\right)^2}}, \quad (130)$$

where the average value of $\mathscr{E}_\omega$ is taken to be zero. These two terms are the contributions of drift and diffusion, respectively, to the total current for a position r in the pinch-off zone. Equation (119) or (120) and (130) must be solved in the pinch-off zone by using boundary conditions that are appropriate to describe the electron distribution (72). Because this distribution can only make sense after a reversal of the x field in the silicon has taken place, then we cannot use the boundary conditions (83) and (84), which were established for the pinch-off point where the $\mathscr{E}_x$ field is directed towards the gate. Instead, we must establish new boundary conditions somewhere beyond the field reversal point, which itself is beyond the pinch-off point. It is also noted that (125) and (128) are equations in K_2, K_3, and K_1 that must be solved together with another relation provided by these revised boundary conditions.

A.3.3 Revised boundary conditions

The pinch-off point to which the boundary conditions (82) through (87) apply occurs in the channel at a location closer to the source than to the field inversion point. In a long-channel device these points are fairly close; in a short-channel device the separation grows since the charge at the pinch-off point grows. The potential between the two points is $V_G - V_T - V_{SAT}$. The boundary conditions at a point y_1 in the channel just after field inversion has occurred are estimated in the folowing way.

From (69) we obtain

$$\left.\frac{d\mathscr{E}}{dy}\right|_{y=\bar{y}} = -\frac{\bar{R}}{(1-\bar{R})} \frac{\bar{\mathscr{E}}^2}{\beta_e} - \frac{I(1-\bar{R})}{Z\mu^2\bar{Q}} \left.\frac{d\mu}{dy}\right|_{y=\bar{y}}. \quad (131)$$

In calculating (131) we are considering the derivatives of R to be negligible. The second term may be evaluated using (109) with $r = y$ and the one-dimensional form (64) of μ to obtain

$$\frac{d\mu}{dy} = -\mu \frac{\beta_0}{\beta_e} \frac{\bar{\mathscr{E}}}{\mathscr{E}_c^2} \frac{d\mathscr{E}}{dy}\bigg|_{y=\bar{y}}. \tag{132}$$

Substituting (132) into (131), using (79) for β_e, and solving for $(d\mathscr{E})/(dy)$, we find

$$\frac{d\mathscr{E}}{dy}\bigg|_{y=\bar{y}} = -\frac{\bar{R}}{(1-\bar{R})} \frac{\bar{\mathscr{E}}^2}{\beta_0}. \tag{133}$$

Using (133), (99), and (110) in (76), we obtain

$$-\frac{\bar{R}}{(1-\bar{R})} \frac{\bar{\mathscr{E}}^2}{\beta_0} + \frac{\beta_e K_1}{\sigma_1^2}\left(\frac{1-(K_1/K_2)^2}{1+(K_1/K_2)^2}\right) = -\frac{q}{\epsilon}(N_A + n). \tag{134}$$

In (134) we have mixed terms for two different positions. The term in $\bar{\mathscr{E}}^2$ actually applies only at the pinch-off position, whereas the term in σ_1^2 applies to a point y_1 just beyond the field-reversal point. We will assume that we can use the term in $\bar{\mathscr{E}}^2$ at y_1, if we change $\bar{\mathscr{E}}$ to $h\bar{\mathscr{E}}$, where h is a parameter >1, which determines how much $\mathscr{E}_y$ has increased in the interval $\bar{y}$ to y_1. For almost all levels of current above threshold we can ignore the charge term on the right-hand side so that we obtain the result

$$\sigma_1 = \sqrt{K_1' \frac{(1-\bar{R})}{\bar{R}}} \frac{\beta_0}{h} \frac{1}{\bar{\mathscr{E}}} \sqrt{1 + \left(\frac{h\bar{\mathscr{E}}}{\mathscr{E}_c}\right)^2}, \tag{135}$$

where

$$K_1' = K_1 b$$

and

$$b = \frac{1-(K_1/K_2)^2}{1+(K_1/K_2)^2}. \tag{136}$$

The result predicts that as $\bar{\mathscr{E}}$ increases, σ_1 at first decreases as $\sqrt{K_1'}/\bar{\mathscr{E}}$ for small values of $\bar{\mathscr{E}}$ and then decreases as $\sqrt{K_1'}$ for larger values of $\bar{\mathscr{E}}$. Using (98) to express the continuity of the longitudinal field, we can use (135) to find the boundary condition

$$\omega_1^* = \frac{K_1'}{K_2} \frac{\beta_0}{h\bar{\mathscr{E}}} \frac{(1-\bar{R})}{\bar{R}}. \tag{137}$$

A further boundary condition is imposed by Gauss' law at the oxide-silicon interface between $\bar{y}$ and L as

$$\epsilon \mathscr{E}_\omega|_{\omega=0} = \epsilon_o \mathscr{E}_{ox}|_N, \tag{138}$$

where $\mathscr{E}_{ox}|_N$ is the normal field in the oxide. The reader may show fairly easily that this condition may be written

$$\frac{d\mathscr{E}_\omega}{dy} = \frac{C_o}{\epsilon}\mathscr{E}_y. \tag{139}$$

Now the left-hand side using (97), (111), and (118) may be expressed as K_1^2/K_2 so that we find, after considerable algebra, the relationship

$$\left(\frac{K_1}{K_2}\right) = \frac{C_o}{\epsilon}\frac{\beta_0}{h}\bar{\mathscr{E}}\left(\frac{1-\bar{R}}{\bar{R}}\right)b$$

or

$$\frac{K_1}{K_2} = \alpha b \equiv \alpha',$$

where

$$\alpha = \frac{C_o\beta_0(1-\bar{R})}{\epsilon h \bar{\mathscr{E}}\bar{R}}. \tag{140}$$

If we knew Q_1, we could obtain the boundary condition N_1 from the normalization relation

$$N_1 = \frac{qQ_1}{\sqrt{2\pi}\sigma_1}. \tag{141}$$

We will determine Q_1 later since it is not required at this point to proceed with a solution of these equations. Also, the boundary condition on $\mathscr{E}_\omega$ is, from (140),

$$\mathscr{E}_{\omega 1} = \frac{K_1}{K_2}h\bar{\mathscr{E}} = \frac{C_o\beta_0}{\epsilon}\frac{(1-\bar{R})}{\bar{R}}b. \tag{142a}$$

To establish the potential V_1 for the new boundary condition, we need two contributions, $\Delta_1 V$ and $\Delta_2 V$, determined as follows. From the pinch-off point to the field-reversal point we have $\Delta_1 V = \bar{Q}/C_o + V_T - V_{FB}$, and from the field reversal point to y_1 we have $\Delta_2 V$. In this section, as a general approximation we may say

$$\mathscr{E}_\omega \simeq \frac{\epsilon_{ox}}{\epsilon t_{ox}}\left(\int_{\bar{y}}^{y}\mathscr{E}_y dy - V_G + \bar{V}\right)$$

or, alternatively, if we integrate from y_i, the field reversal point, then

$$\mathscr{E}_\omega = \frac{\epsilon_{ox}}{\epsilon t_{ox}}\int_{y_i}^{y}\mathscr{E}_y dy. \tag{142b}$$

For $y = y_1$, the new boundary condition, we have $\int_{y_i}^{y_1} \mathscr{G}_y dy = \Delta_2 V$, which we can then use in (142b), along with (142a), to find

$$\Delta_2 V = \beta_0 \left(\frac{1 - \bar{R}}{\bar{R}}\right) b.$$

Using $\bar{Q}/C_o$ from (18c), the total voltage drop between $\bar{y}$ and y_1 is therefore

$$\Delta V \equiv V_1 - V_{SAT} = \Delta_1 V + \Delta_2 V$$
$$= V_T - V_{FB} + \beta_0 \frac{(1 - \bar{R})}{\bar{R}} b + \frac{\bar{Q}_0}{C_o} \sqrt{1 + \left(\frac{I}{W v_S \bar{Q}_0}\right)^2}. \quad (143)$$

We then have for r_1

$$r_1 = \frac{2}{(h + 1)} \frac{\Delta V}{\bar{\mathscr{G}}}, \quad (144)$$

where an average value of $\mathscr{G}$ between $\bar{y}$ and y_1 has been used. Using (143) with $b \simeq 1$, we have

$$r_1 = \frac{2}{(h + 1)\bar{\mathscr{G}}} \left[V_T - V_{FB} + \beta_0 \left(\frac{1 - \bar{R}}{\bar{R}}\right) \right.$$
$$\left. \cdot \left(1 + \frac{C_o + C_s}{C_o} \sqrt{1 + \left(\frac{I}{W V_S \bar{Q}_0}\right)^2}\right)\right]. \quad (145)$$

A.3.4 Approximate solutions to the equations

Using the integrating factor $-(1/r)$ and the boundary values ω_1^*, r_1, we may solve (120) to obtain

$$\omega^* = Cr \ln\left(\frac{r}{r_1}\right) + \frac{\omega^{*_1}}{r_1} r, \quad (146a)$$

$$S_\omega^* = \frac{\partial \omega^*}{\partial r} = C \ln\left(\frac{r}{r_1}\right) + C + \frac{\omega_1^*}{r_1}, \qquad \frac{\omega^*}{r} = C \ln\left(\frac{r}{r_1}\right) + \frac{\omega_1^*}{r_1}, \quad (146b)$$

where

$$C = 2\left(\frac{K_3 - K_2}{K_1}\right) - \frac{K_1}{K_2}\left(1 - \frac{2K_1^2}{K_1^2 + K_2^2}\right) - \frac{1}{K_2}\left(\frac{\sigma}{\lambda_e}\right)^2 \quad (146c)$$

and is considered to be constant. Used in (119), this result can be written, by using (140) and ignoring the term in σ/λ_e, as

$$\frac{1}{v}\frac{dv}{d\omega^*} = -\frac{(C + \alpha)}{S_\omega^*}\frac{1}{\omega^*},$$

where $v = (1/\sigma^2)$, which then yields

$$\sigma^2 = \sigma_1^2 \left(\frac{\omega^*}{\omega_1^*}\right)^a, \tag{147a}$$

where

$$a = \frac{C + \alpha}{S_\omega^*}. \tag{147b}$$

We now use (115); assuming that $K_1 \ll K_2$ and that $K_2 \simeq K_3$ we can write (115) as

$$K_3 - K_2 \simeq \frac{K_2}{K_1} \frac{\sigma^2}{\epsilon\beta_e} qN. \tag{148}$$

This condition on K_1 amounts to saying that the field that is pushing carriers away from the surface is much smaller than the field that is driving the current; this is always found to be the case in the on region.

To find N, we return to (130) and assume that at the position y_1 the dominant conduction component is drift. That is, at the saturation point the diffusion current accounts for about one half to one third of the current flow but at a distance, r_1, further down the channel, drift has again become dominant. If drift has become dominant, then we can say from (130)

$$Q = \frac{I}{W\mu \mathscr{E}_r^*}, \tag{149a}$$

and as velocity saturation becomes predominant, this becomes

$$Q = \frac{I}{Wv_s}. \tag{149b}$$

From (149b) we can see that the gradient of the charge along the streamline actually goes to zero in the limit of velocity saturation. Therefore, the assumption of total drift is reasonable. With this assumption (149b) gives us a boundary condition Q_1 at y_1, and therefore, N_1, which would be

$$N_1 = \frac{I}{\sqrt{2\pi}} W\mu_1^* \mathscr{E}_{r1}^* \sigma_1 = > \frac{I}{W\sqrt{2\pi}\sigma_1 v_s},$$

from the normalization of the charge, which approaches the velocity-saturated condition shown. From (149b) we can write the equation for N for any position as

$$N = \frac{I}{W\mu \mathscr{E}_r^* \sqrt{2\pi}\sigma}. \tag{150}$$

Using (150) and (103) in (148) and then the result in (146c), we find

$$C = 2\frac{K_2}{K_1^2}\frac{q}{\epsilon}\frac{\beta_0 K_3 \bar{Q}_0 \bar{\mathscr{E}}^2}{\mathscr{E}_c^4\sqrt{2\pi}}\frac{\omega_1^{3a/2}}{\sigma_1^3}\cdot\omega^{2-\frac{3a}{2}}. \tag{151}$$

Since C is independent of position, we have from (151)

$$a \simeq 4/3. \tag{152}$$

Substituting from (135), (137), and (140) we can reduce (151) to

$$C \simeq C + \alpha = \frac{\gamma}{h^2 K_1^{3/2}}, \qquad \gamma = \sqrt{\frac{2}{\pi}}\left(\frac{1-\bar{R}}{\bar{R}}\right)^{3/2}\frac{\beta_0}{\epsilon}\frac{(C_o + C_s)}{\mathscr{E}_c}. \tag{153}$$

In writing (148) we assumed that $K_3 - K_2 \ll K_3, K_2$, and we have used this fact again here. It is noted from (147) that (152) implies the result

$$S_\omega = \frac{3}{4}(C + \alpha). \tag{154}$$

We now use (150) in (125). From (150) or (149) we have

$$\frac{N'}{N} = -\left[\frac{1}{\mathscr{E}_r^*}\frac{d\mathscr{E}_r^*}{dr} + \frac{1}{\mu^*}\frac{d\mu^*}{dr} + \frac{1}{\sigma}\frac{d\sigma}{dr}\right], \tag{155}$$

and using this result and (107c) in (125), we obtain

$$\mathscr{E}_r^*\left[-\frac{1}{\mathscr{E}_r^*}\frac{d\mathscr{E}_r^*}{dr} + \frac{\sigma^2}{2}\frac{d}{dr}\left(\frac{1}{\sigma^2}\right)\right] = \beta_e\left[\frac{1}{\lambda_e^{*2}} + \frac{1+S_\omega^2}{\sigma^2}\right],$$

where we have dropped all terms in $(\sigma/\omega^*)^2$ (such as the term in β/ω^*) and $\sigma^2/r\omega^*$ because we are assuming that the Gaussian is localized well enough that $(\sigma/\omega^*)^2 \ll 1$. Using (100) for the streamline and (114), we obtain

$$K_3\frac{S_\omega}{\sigma^2} + \frac{3}{2}K_3\omega^*\frac{d}{dr}\left(\frac{1}{\sigma^2}\right) = -\left(\frac{1}{\lambda_e^{*2}} + \frac{1+S_\omega^2}{\sigma^2}\right),$$

and then substituting from (119) and (146c)

$$K_3 = \frac{1+S_\omega^2}{\frac{3(C+\alpha)}{2} - S_\omega^*}, \tag{156}$$

where the term $(1/K_2)(\sigma/\lambda_e)^2$ has been considered small.

We are going to assume that in the range of high fields in which we are interested,

$$C \gg 1, \qquad K_1$$

and $S_\omega \simeq 0$, which means that $(\partial\omega^*/\partial r) \simeq (\partial\omega/\partial r)$ [cf. (94)], so that the streamline locus is effectively a straight line. Then (128) and (156) may be subtracted, and (146c) may be used to give

$$S_\omega^* = \frac{3}{2}(C+\alpha)\frac{1}{(1+K_1)\left(\frac{C+\alpha}{2}\right)} \tag{157}$$

$$K_2 = \frac{2}{C+\alpha}\left[1+\frac{(C+\alpha)^2}{4}\right] \tag{158}$$

and

$$K_3 = \frac{2}{C+\alpha}\left[1+\frac{(C+\alpha)^2}{4}\right] + \left(\frac{C+\alpha}{2}\right)K_1. \tag{159}$$

Using (138) and (153) we then find

$$K_1 = \left(\frac{\gamma\alpha}{2h^2}\right)^{2/5}, \qquad K_2 = \left(\frac{\gamma}{2h^2}\right)^{2/5}\left(\frac{1}{\alpha}\right)^{3/5} \tag{160, 161}$$

$$K_3 = \left(\frac{\gamma}{2h^2}\right)^{2/5}\left(\frac{1}{\alpha}\right)^{3/5} + \frac{1}{2^{2/5}}\left(\frac{\gamma}{2h^2}\right)^{4/5}\left(\frac{1}{\alpha}\right)^{1/5} \tag{162}$$

$$C+\alpha = \left(\frac{\gamma}{h^2}\right)^{2/5} 2^{1/5}\left(\frac{1}{\alpha}\right)^{3/5}. \tag{163}$$

These are the constants that characterize the electric fields. They are determined by the parameter α, or, equivalently, $\bar{\mathscr{E}}$, through (140). These results are plotted in Fig. 21 over the total range of $\bar{\mathscr{E}}$ for a practical device. All of the assumptions we have made are validated by the plots. We now use these results to determine the length of the pinch-off zone $L - \bar{y}$. From (117) we have

$$\frac{\omega^{*2}}{\sigma^2} = \frac{2(K_3 - K_2)}{K_1 K_3}\frac{(V_{DS} - V_1)}{\beta_e}. \tag{164}$$

Using (103) for β_e along the streamline and substituting (147) for σ, we obtain a result for ω^*, which is

$$\omega^* = \frac{\omega_1^*}{\sigma_1^{3/2}}(C+\alpha)^{3/4}\left[\frac{(V_{DS} - V_1)\beta_0}{\mathscr{E}_c^2}\right]^{3/4}. \tag{165}$$

Now $\omega^* = r^*\theta$ and $L - y_i = r^*\cos\theta$, so that we may write the length of the pinch-off zone as

$$L - y_i = \frac{\cos\theta}{\theta}\frac{\omega_1}{\sigma_1^{3/2}}(C+\alpha)^{3/4}\left[\frac{(V_{DS} - V_1)\beta_0}{\mathscr{E}_c^2}\right]^{3/4}, \tag{166}$$

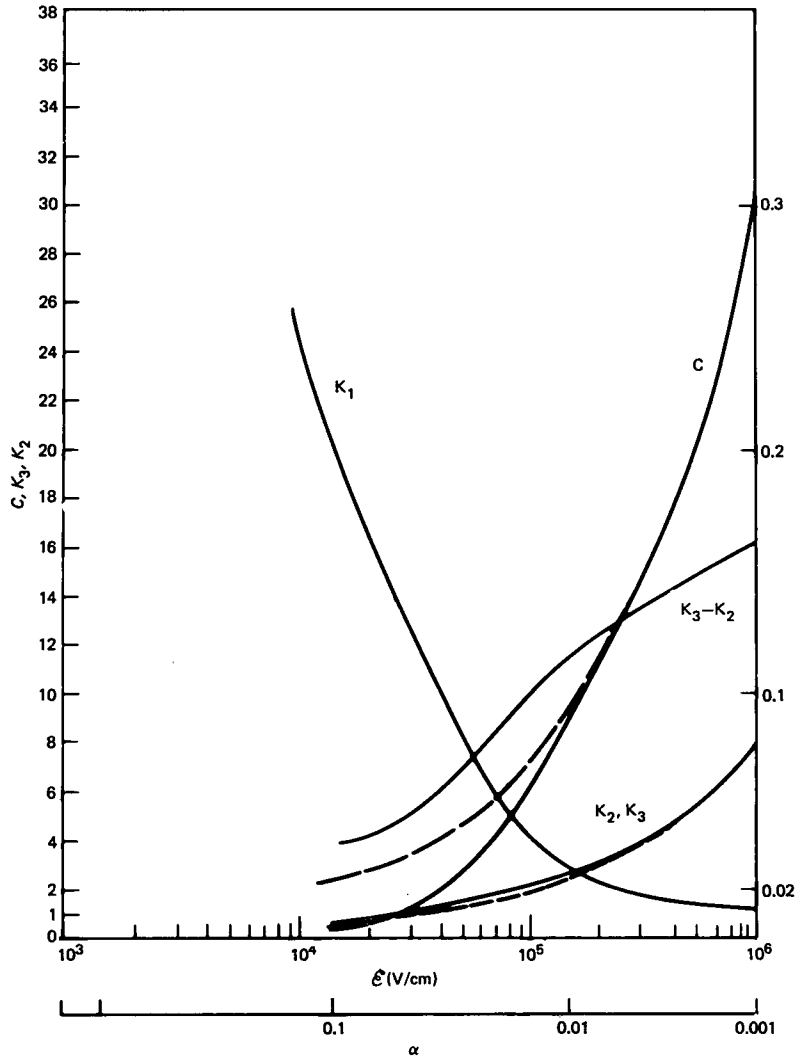

Fig. 21—Variation of the field constants with pinch-off field $\mathscr{E}$ or its normalized form, α. The device parameters are $t_{ox} = 250$Å and $N_A = 5 \times 10^{16}$ cm^{-3}.

where θ is the angle formed by the interface and the vector from the point y_1 to the streamline axis. Using (135), (137), (140), and (163) for σ_1, ω_1, and $(C + \alpha)$, respectively (under the condition $b \simeq 1$), we have

$$L - y_i = \frac{\cos\theta}{\theta} \cdot \alpha^{5/4} \cdot \frac{\epsilon}{C_o} \cdot 8^{3/4} \cdot \left(\frac{\bar{R}}{1-\bar{R}}\right)^{3/4} \beta_0^{1/4} (V_{DS} - V_1)^{3/4}. \quad (167)$$

We now require an estimate of θ. In solving for the constants K_1,

K_2, and K_3, we have assumed that $S_\omega \simeq 0$ and that S_ω^* was constant, which implies that θ can be considered constant since $\omega = \gamma\theta$, and so

$$\theta = C\left\langle \ln \frac{\gamma}{\gamma_1} \right\rangle, \tag{168}$$

where the ln term must be written as some average. We cannot actually estimate θ because the solution does not take into account the effect of the finite depth of the drain junction. We will, therefore, use a value of $\theta = 45$ degrees to expedite calculation of (167). This is a reasonable assumption based on the examination of two-dimensional numerical solutions of this region. (Note that the function $\cos\theta/\theta$ is not a strong function anyway, varying between 0.5 and 0.85 for θ between 30 and 45 degrees).

The result (167) applies for voltages $V_{DS} > V_1$ but not to the situation when $V_1 > V_{DS} > V_{SAT}$. To extend the result over the interval $\Delta V = V_1 - V_{SAT}$ and thus to extend $L - y_1$ to $L - \bar{y}$, we will write (167) approximately as

$$\Delta L_H = L - \bar{y} = \frac{h^{1/5}}{\bar{\mathscr{E}}^{3/2}} \cdot A_1[(V_{DS} - V_{SAT} + A_2\Delta V)^{3/4} - (A_2\Delta V)^{3/4}], \tag{169}$$

where

$$A_1 = \frac{\cos\theta}{\theta}\left(\frac{1}{\gamma}\right)^{4/5} 8^{1/2}\left(\frac{C_o}{\epsilon}\right)^{1/2}\left(\frac{1-\bar{R}}{\bar{R}}\right)^{3/4}\beta_0^{3/4} \tag{170}$$

and h is determined by the condition that $L - \bar{y} = r_1$ for $V_{DS} = V_{SAT} + \Delta V$ and A_2 is an additional constant that we must find. Using (145) in (169) and for the saturation condition, noting that the current in velocity saturation may be written $I = W\mu Q \cdot \bar{\mathscr{E}}^2/(1-\bar{R})\mathscr{E}_c$, we have

$$\frac{2}{(h+1)}\left(\frac{\bar{Q}_0}{C_o}\right)^{1/4}\frac{\bar{\mathscr{E}}^{1/2}}{\mathscr{E}_c^{1/2}}\cdot\frac{1}{\bar{\mathscr{E}}} = \frac{A_1 h^{1/5}}{\bar{\mathscr{E}}^{3/2}}[(1+A_2)^{3/4} - A_2^{3/4}]. \tag{171}$$

To determine A_2 and h, we require another relation, which is given by the condition of the continuity of the derivative of the current at the boundary between the triode and the saturation regions. This condition is found to be [cf. (48)]

$$\bar{Q}/C_o = -\mathscr{E}_c\frac{d\bar{y}}{dV_{DS}}\left(\frac{I}{Wv_sC_o}\right). \tag{172}$$

Since $\bar{Q} \simeq I/Wv_s$ for higher channel fields, (172) becomes

$$1 = -\frac{\mathscr{E}_c d\bar{y}}{dV_{DS}}. \tag{173}$$

From (169) we have

$$\frac{d\bar{y}}{dV_{DS}} = \frac{h^{1/5}}{\bar{\mathscr{E}}^{3/2}} A_1 \frac{3}{4} \frac{1 + dV_T/dV_{DS}}{(V_{DS} - V_{SAT} + A_2\Delta V)^{1/4}}$$

$$- \frac{3h^{1/5}A_1}{2\bar{\mathscr{E}}^{5/2}} [(V_{DS} - V_{SAT} + A_2\Delta V)^{3/4} - (A_2\Delta V)^{3/4}]. \tag{174}$$

For $V_{DS} = V_{SAT}$, (174) reduces to

$$\frac{d\bar{y}}{dV_{DS}} = -\frac{3}{4} \frac{h^{1/5}A_1}{\bar{\mathscr{E}}^{3/2}} \frac{1}{(A_2\Delta V)^{1/4}}, \tag{175}$$

where we have considered $(dV_T/dV_{DS}) \ll 1$. Using (175) in (173) gives the condition

$$\frac{3}{4} \frac{h^{1/5}A_1}{A_2^{1/4}\Delta V^{1/4}\bar{\mathscr{E}}^{1/2}} = 1. \tag{176}$$

In writing (174) we have assumed that the dependence of $h^{1/5}$ on $\bar{\mathscr{E}}$ and hence V_{DS} can be ignored. Noting that for $\bar{y} = y$ we have $L - \bar{y} = L - \bar{y}_1 = r_1$ and $V_{DS} - V_{SAT} = \Delta V$; then using (144) in (169) we have

$$\frac{2\Delta V}{(h+1)\bar{\mathscr{E}}} = \frac{h^{1/5}A_1\Delta V^{3/4}}{\bar{\mathscr{E}}^{3/2}}. \tag{177}$$

Using this result in (176) yields

$$A_2 = \left(\frac{3}{2(h+1)} \frac{\mathscr{E}_c}{\bar{\mathscr{E}}}\right)^4. \tag{178}$$

Since $h > 1$ and $\bar{\mathscr{E}} \geq \mathscr{E}_c$ in the range of interest here, A_2 will be a small number (i.e., <0.1, typically). We are, therefore, justified in treating the square brackets as 1 in (171) and so we find for h

$$(h+1)h^{1/5} = \bar{\mathscr{E}}2 \left(\frac{\bar{Q}_0}{C_o}\right)^{1/4} \frac{1}{\mathscr{E}_c^{1/2}} \frac{1}{A_1}. \tag{179}$$

If we considered a typical value of $\bar{\mathscr{E}} \simeq 10^5$ V/cm (i.e., well into the hot-electron regime) and use $A_1 \simeq 175$ [cf. (170)], then we find $h \cong 2.8$. For higher values of $\bar{\mathscr{E}}$, h would increase, although we would not expect values much higher than $\bar{\mathscr{E}} = 10^5$ V/cm in practice. Therefore, we will treat h generally as a constant with a value 2.5 through 3.0, especially since it appears in $L - \bar{y}$, the result of major interest, only as $h^{1/5}$.

A.4 Discussion of results

The result (169) is valid for high values of pinch-off field $\mathscr{E}$ so that the hot-electron approximations [cf. (105) and (106)] are valid. For low values of $\mathscr{E}$, $L - y$ becomes anomalously large according to (169) and is obviously incorrect. However, it has been shown that for gate voltages near and below V_T, a good representation of ΔL is

$$\Delta L_L = L - \bar{y} = \sqrt{2 \frac{\epsilon}{qN_A} (V_{DS} - V_{SAT})}. \tag{180a}$$

For gate voltages in the 'on' region (180) becomes notoriously gross. If we now combine the results (180) and (169) in the form

$$\frac{1}{\Delta L} = \frac{1}{\Delta L_H} + \frac{1}{\Delta L_L}, \tag{180b}$$

we obtain a representation of $\bar{y} = L - \Delta L$, which is good for all regions of operation.

We demonstrate this agreement by the curve shown in Fig. 22. The curves show data and theory for a 0.5-μm device. The lower solid theory line shows the agreement obtained using (179) to determine $\bar{y}$, which is then used in the velocity-saturated model of the companion paper. The agreement is quite good. The upper solid line shows the result that is obtained if the ΔL_L [cf. (180a)] is used alone to predict $\bar{y}$ and there is considerable error. However, there is not as much error as one might expect on the basis of $(L/L - \Delta L_L)I_{SAT}$, which is normally considered to be the case where I_{SAT} is the current at the onset of saturation. This is clear from the comparison of ΔL, ΔL_L, ΔL_H, and $\Delta L'$ shown in Fig. 23. The variation of ΔL in Fig. 23 is much more than is manifested in Fig. 22 because the velocity-saturated current [cf. (20b)]

$$I = Wv_sC_o\left[\sqrt{\left(\frac{\mathscr{E}_c\bar{y}}{a^*}\right)^2 + \frac{(V_{GS} - V_T) - (Q_0/C_o)^2}{a^*}} - \frac{\mathscr{E}_c\bar{y}}{a^*}\right]$$

loses its dependence upon $\bar{y}$. In the limit of total velocity saturation there is no dependence at all. Therefore, we are completely justified in making the approximation such as (168). In fact, the coefficient A_1 could be considerably in error without having much effect. The same may be said about the radius of curvature of the junction. The fact that we treated a vertical instead of a cylindrical junction and did not consider explicitly the junction depth is really of very little consequence. The most important feature of the result is the inverse dependence of ΔL_H upon $\mathscr{E}^{3/2}$. It is this dependence that allows ΔL to track the device geometries and applied voltages in a continuous fashion.

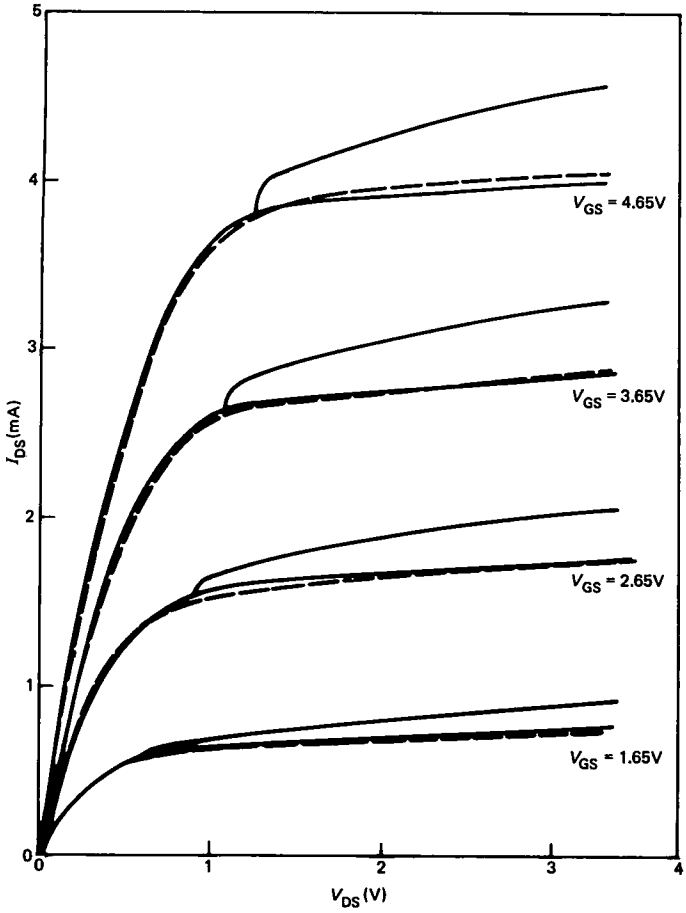

Fig. 22—Comparison of drain voltage data (dashed curve) of a velocity saturated device with theory using the composite result (128) (lower solid curve) and the simple depletion result (127) (upper solid curve). The effect of series resistance ($\simeq 50\Omega$) has been removed from the data and the device parameters are $L = 0.5$ μm, $t_{ox} = 250$Å, $W = 10$ μm, $N_A = 5 \times 10^{16}$ cm^{-3}, $r_j = 0.25$ μm, $\mu_0 = 650$ cm^2/V-s, $\theta = 0.03$, and $v_s = 10^7$ cm/s.

APPENDIX B

B.1 Introduction

In a previous work, we had examined the problem of ensuring the continuity of current in moving from the subthreshold to the above-threshold regions by introducing a voltage parameter δV_G into the expression for the saturated current so that it became

$$I_{AT} = \frac{W\mu C_o}{2\bar{y}} (V_G - V_T + \delta V_G)^2, \tag{181}$$

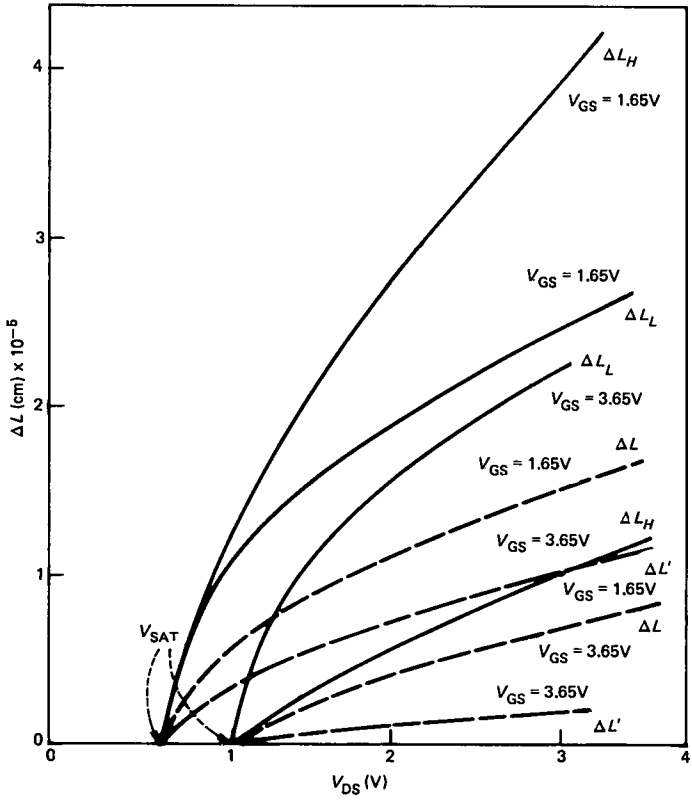

Fig. 23—Variations of ΔL_H, ΔL_L (solid curve) and ΔL (dashed curve) for the device in Fig. 11. Also indicated is the effective $\Delta L'$ if the current in saturation is $I = I_{SAT}L/L - \Delta L$, where I_{SAT} is the pinch-off value. $\Delta L'$ is less than ΔL because of the effects of velocity saturation.

where I_{AT} refers to the above-threshold current.[16] The parameter δV_G was introduced to account for the absence of diffusion in the above-threshold formulation. It was determined on the basis of a unique value of current at the threshold condition from consideration of the above-threshold and subthreshold currents. However, to be consistent in this approach and to have a useful result, we must also require continuity of the derivative of the current between the subthreshold and above-threshold regions of operation.

In this appendix we will re-examine the conditions of the continuity of current and its derivative at two important transitions in the device, namely, (1) the transition between the subthreshold and the above-threshold saturation regions, and (2) the transition between the triode region and the saturation region, both of which are above threshold. These conditions are needed to determine $\bar{R}$, the fraction of the total current carried by diffusion at the pinch-off point above threshold.

It is noted that the former approach[16] of substituting the parameter δV_G into the above-threshold result for all gate voltages is not complete. When determined self-consistently, the parameter δV_G gives only the voltage for which the currents below and above threshold are equal but says nothing about their derivatives.

The transition from the nonsaturated region below threshold to the triode region above threshold will be considered only after the results have been obtained since it only exists over a drain voltage of about 2 kT/q and hence is relatively unimportant. Also, the transition from linear to triode regions below threshold does not need to be considered since its continuity has already been established.[25]

B.2 Transition from subthreshold to above-threshold saturation conduction

Since we are concerned at this transition with currents flowing just above the threshold voltage (i.e., almost at the threshold point), then we are justified in using the above-threshold current expressions that do not include velocity saturation because the channel fields for such voltages are less than $\mathscr{E}_c$. From (83) we have

$$\frac{\bar{Q}}{C_o} = \left(\frac{1-\bar{R}}{\bar{R}}\right)\frac{kT}{q}\frac{(C_o + \bar{C}_S F)}{C_o}, \tag{182a}$$

where

$$\bar{C}_S = \sqrt{\frac{\epsilon q N_A}{2(V + V_{BS} + 2\phi_F)}} \tag{182b}$$

is the semiconductor depletion capacitance at the pinch-off point in the channel. At the threshold condition we will have $\bar{C}_S = C_S$, i.e., the same value at the source and the pinch-off point. For a constant T, $\bar{Q}$ is dependent on V_{GS} only through $\bar{C}_S$, on V_{DS} only through $F(V_{DS})$, and on V_{BS} through both of these. (It may also have some dependence upon V_{GS} and V_{DS} through the parameter $\bar{R}$.) We will assume at this point that the dependence of $\bar{Q}$ upon bias parameters is sufficiently weak that we may ignore its derivatives in the calculation of the derivatives of the current above threshold. We shall re-examine the validity of the assumption later. The result (82) suggests that we should use it in the triode region equation

$$I_{DS} = \frac{\mu C_o W}{L}\left[(V_{GS} - V_T(V_{DS}))V_{DS} - \frac{V_{DS}^2}{2}\right] \tag{183}$$

to obtain a modified form of the saturation current, which is

$$I_{SAT} = \frac{\mu C_o W}{2\bar{y}}\left[(V_{GS} - V_T(V_{DS}))^2 - \left(\frac{\bar{Q}}{C_o}\right)^2\right]. \tag{184}$$

The threshold voltage has been written as a function of drain voltage to show that the 3/2 power terms of the conventional triode expression may be represented.[16] We conclude from (184), therefore, that true square law behavior of the device may only be observed for $V_{GS} - V_T(V_{DS}) \gg \bar{Q}/C_o$.

In considering the saturation current at the transition between the two regions, we are going to assume that an offset gate voltage of η above the threshold voltage must be applied to the device in order for the above-threshold theory to predict a finite current flowing at the transition point. This situation is illustrated in Fig. 24, which shows data for a reasonably short-channel device ($L = 2.5$ μm, $t = 500$Å, and $N_A \simeq 10^{16}$ cm^{-3}) in the region near the threshold voltage. The voltage η is shown as the voltage increment from V_T to the onset of the straight section. An equivalent way of stating this condition is that the above-threshold current merges with the subthreshold current not

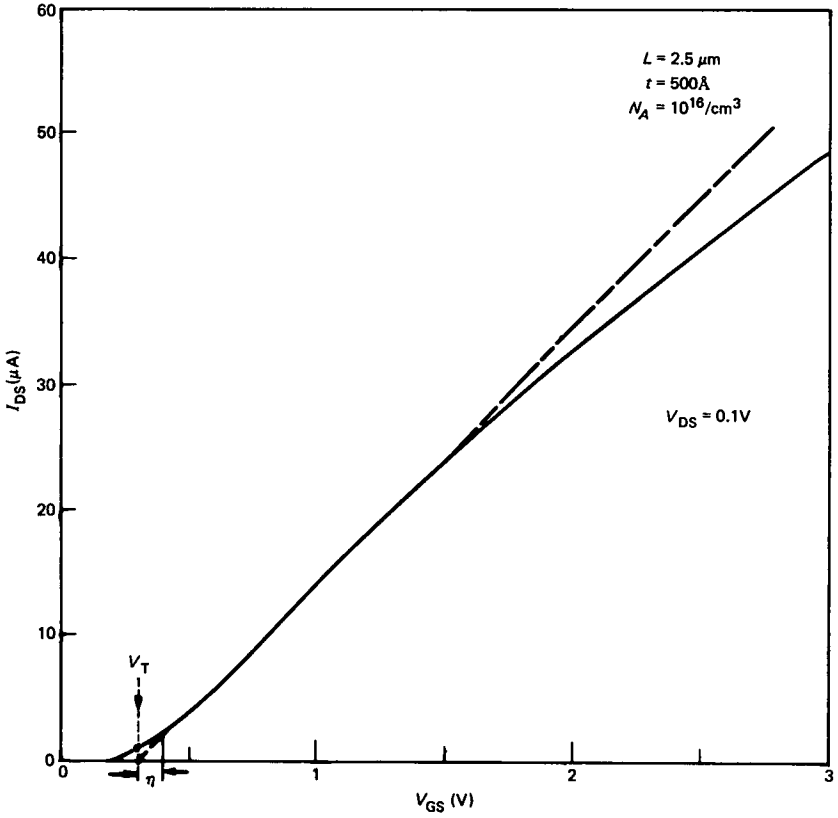

Fig. 24(a)—Variation of drain current with gate voltage showing the determination of η and its relation to V_T for linear region characteristic.

for $V_{GS} = V_T$ but for a slightly larger value of $V_{GS} = V_T + \eta$, so that (184) becomes

$$I_{SAT}|_{V_T+\eta} = \frac{\mu W C_o}{2\bar{y}}\left[\eta^2 - \left(\frac{\bar{Q}}{C_o}\right)^2\right]. \tag{185}$$

In writing the saturation current we have used $\bar{y}$ in place of L in (184) and (185), where $\Delta L = L - \bar{y}$ is the channel-length modulation in saturation operation. The relative size of ΔL to L can be significant in a short-channel device.

At the gate voltage V_T, the subthreshold theory predicts the flow of a finite diffusion current, which would correspond to an effective

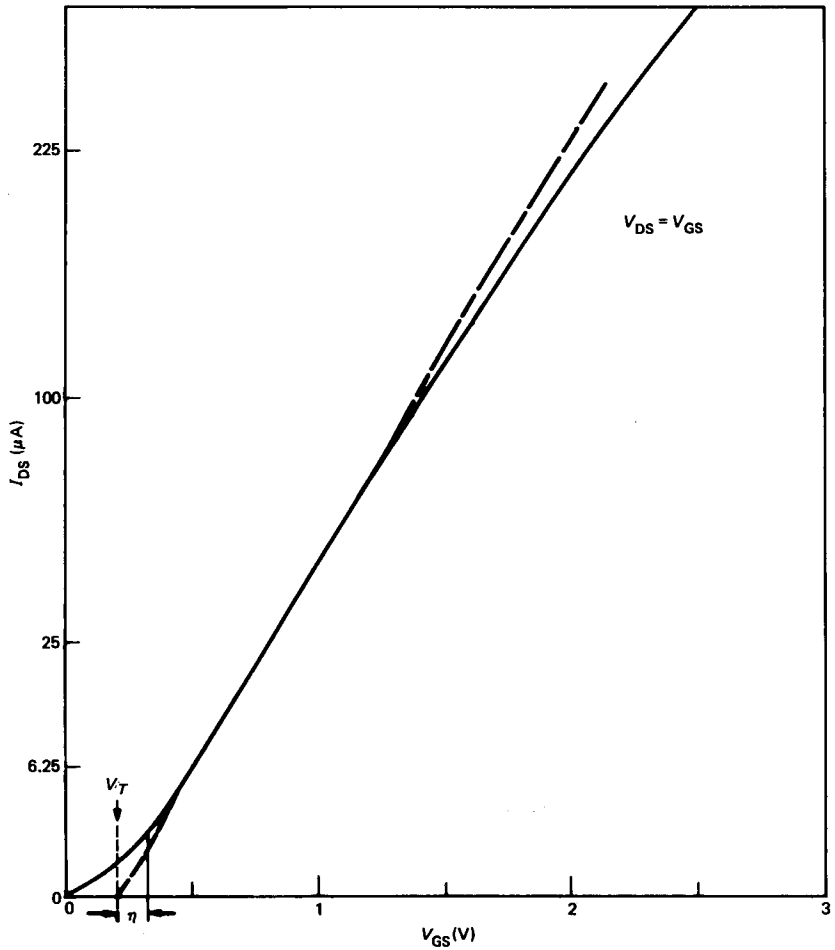

Fig. 24(b)—Variation of drain current with gate voltage showing the determination of η and its relation to V_T for saturation region characteristic.

surface potential $\hat{\phi}_S = 2\phi_F$. Since the transition between the regions takes place for $V_G = V_T + \eta$, in a similar way we should introduce a corresponding offset surface potential into the subthreshold formulation. It may be defined as the additional surface potential above the threshold value ($\hat{\phi}_S = 2\phi_F$) that is achieved when the additional voltage η is applied. The subthreshold current may be written[25]

$$I_{ST} = \frac{qWD}{\bar{y}} \sqrt{\frac{\epsilon}{2qN_A(V_{BS} + \hat{\phi}_S)} \left(\frac{kT}{q}\right)} n_i e^{\frac{-q\phi_F}{kT}} e^{\frac{q\hat{\phi}_S}{kT}}, \quad (186)$$

where $\hat{\phi}_S$ is the effective surface potential in subthreshold operation. Therefore, when a merging of the subthreshold and above-threshold currents is considered, the transition will occur not for $\hat{\phi}_S = 2\phi_F$ but rather for a slightly larger surface potential of $2\phi_F + \xi$. The voltage variation of the current is predominantly in the exponential term through the effective surface potential $\hat{\phi}_S$, which is obtained from the equation

$$V_{GS} = V_{FB} + \hat{\phi}_S + \frac{1}{C_o} \sqrt{2\epsilon q N_A (V_{BS} + \hat{\phi}_S)} \cdot F. \quad (187)$$

Therefore, we will neglect the voltage dependence of $\hat{\phi}_S$ in the preexponential term. The factor F has been defined previously for short-channel devices.[16] At the threshold condition it takes the specific form

$$F = \frac{L + r_j - \frac{1}{2}\sqrt{KV_{DS}} - \frac{1}{2}[KV_{DS} + 2r_j\sqrt{K(V_{bi} + V_{BS} + V_{DS})} + r_j^2]^{1/2} - \frac{1}{2}[2r_j\sqrt{K(V_{bi} + V_{BS})} + r_j^2]^{1/2}}{L - \sqrt{KV_{DS}}}, \quad (188)$$

where

$$K = 2\epsilon/qN_A,$$

r_j = the junction depth,

and

N_A = the doping concentration

and it is assuming that $V_{bi} \simeq 2\phi_F$.

The transition from subthreshold to above-threshold conduction as a function of drain voltage is obvious in a short-channel device because the variation with drain-source voltage of the effective $\hat{\phi}_S$ (i.e., the average ϕ_S in the channel determined from two-dimensional charge-sharing techniques) below threshold and the V_T above threshold are

quite pronounced. Actually, this transition always exists even in a long-channel device since we may never have absolutely no channel length modulation except in the limit of a device of infinite length. Hence, there will always be some curve of I versus V_{DS}, which passes from the subthreshold to the above-threshold region at some V_{DS}, so that we may consider this condition in a general way. By imposing the condition of continuity of current, we obtain from (185) and (186) the relation

$$\frac{1}{2}\left[\eta^2 - \left(\frac{\bar{Q}}{C_o}\right)^2\right] = \left(\frac{kT}{q}\right)^2 \frac{C_S}{C_o} e^{\frac{q\xi}{kT}}, \tag{189}$$

where C_S, as mentioned in the definition (182b), is the capacitance of the semiconductor depletion region. This capacitance will be approximately the same at any position between the source and the edge of the drain-depletion region (i.e., the saturation point in the channel) for gate voltages up to the threshold condition, since we are assuming a negligible field in the channel for the subthreshold formulation. The term ξ represents the additional surface potential that is required above threshold to achieve a matching of the solutions. We expect this increase in surface potential introduced in (186) to correspond to the concomitant increase in gate voltage η. The solutions are joined, therefore, not at $V_{GS} = V_T$ but at a slightly higher voltage, which is determined from (187) to be

$$V_T + \eta = V_{FB} + 2\phi_F + \xi + \frac{1}{C_o}\sqrt{2\epsilon q N_A (V_{BS} + 2\phi_F + \xi)}\, F + \frac{kT}{q}\frac{C_s}{C_o}, \tag{190}$$

where the term $[(kT)/q]/(C_s/C_o)$ is an attempt to represent the change in mobile charge in the channel between the two conditions. We have simply used $Q_n = N_A \bar{X}_d$, the charge in the channel under diffusion-limited conditions, where $\bar{X}_d$ is the effective depth of channel charge.[25] We must now impose the condition of continuity of the derivatives at the transition. From (184) we have (neglecting the derivatives of $\bar{Q}$)

$$\frac{dI}{dV_{DS}} = -\frac{I}{\bar{y}}\frac{d\bar{y}}{dV_{DS}} - \mu C_o \frac{W}{\bar{y}}\eta\frac{dV_T}{dV_{DS}}, \tag{191}$$

and from (186) we have

$$\frac{dI}{dV_{DS}} = -\frac{I}{\bar{y}}\frac{d\bar{y}}{dV_{DS}} + \frac{q}{kT} I \frac{d\hat{\phi}_S}{dV_{DS}}. \tag{192}$$

We therefore have at the transition from (191) and (192) that

$$\eta \frac{dV_T}{dV_{DS}} = -\left(\frac{kT}{q}\right)\frac{C_S}{C_o} e^{\frac{q\xi}{kT}} \frac{d\hat{\phi}_S}{dV_{DS}}. \tag{193}$$

From (187) we obtain for the derivative of the threshold voltage

$$\frac{dV_T}{dV_{DS}} = \frac{1}{C_o}\sqrt{2\epsilon q N_A (V_{BS} + 2\phi_F + \xi)}\,\frac{dF}{dV_{DS}} \tag{194}$$

or, for the surface potential at the transition

$$\left.\frac{d\hat{\phi}_S}{dV_{DS}}\right|_{\phi_S = 2\phi_F + \xi}$$

$$= -\frac{1}{(C_o + C_S F)}\sqrt{2\epsilon q N_A (V_{BS} + 2\phi_F + \xi)}\,\frac{dF}{dV_{DS}}. \tag{195}$$

Using (194) and (195) in (193) we obtain

$$\eta = \frac{kT}{q}\frac{C_S}{C_o + C_S F} e^{\frac{q\xi}{kT}}. \tag{196}$$

Using (190) and (187) at $V_G = V_T$, we find

$$\eta = \xi + \frac{1}{C_o}\sqrt{2\epsilon q N_A}$$

$$\cdot(\sqrt{V_{BS} + 2\phi_F + \xi} - \sqrt{V_{BS} + 2\phi_F})F + \frac{kT}{q}\frac{C_s}{C_o}, \tag{197}$$

and by expanding the square root term, since $\xi \ll V_{BS} + 2\phi_F$, we have

$$\eta \simeq \frac{(C_o + C_S F)}{C_o}\xi + \frac{kT}{q}\frac{C_s}{C_o}. \tag{198}$$

Combining (196) and (198) we obtain

$$\xi = \frac{kT}{q}\frac{C_S C_o}{(C_o + C_S F)^2} e^{\frac{q\xi}{kT}} - \frac{kT}{q}\frac{C_s}{(C_o + C_s F)}$$

or

$$\xi' = \frac{C_S C_o}{(C_o + C_S F)^2} e^{\xi'} - \frac{C_s}{C_o + C_s F}, \tag{199}$$

where $\xi' = q\xi/kT$. We may, therefore, determine ξ once values of oxide thickness, doping density, and substrate bias have been specified by using this transcendental relationship. An interesting feature of this parameter is its lack of dependence upon T.

By this method we have ensured continuity of the current and of its

derivative in passing from the subthreshold to the above-threshold region. It is realized that the absolute value of the predicted current will be somewhat low around $V_G = V_T$ because the subthreshold region does not incorporate drift and because the above-threshold result does not incorporate diffusion. However, this is not expected to be of any serious consequence. To calculate the current for voltages between $V_T + \eta$ and V_T, the surface potential is found from (198) and then (186) is used.

The results of calculating ξ and η are shown in Figs. 25 and 26 as a function of oxide thickness and substrate doping over a wide range of these parameters. These parameters are independent of gate bias but have a slight dependence on V_{DS} through the parameter F. In Figs. 25 and 26 a value of $F = 0.8$ was used, and to illustrate the influence of the drain bias, the curves are also shown in dashed lines for $F = 0.6$. Generally, the results show that in the range of useful device operation (say $N_A \sim 1\text{–}2 \times 10^{16}$ cm^{-3} for $t_{ox} \simeq 500$Å or $N_A \simeq 3\text{–}5 \times 10^{16}$ cm^{-3} for $t_{ox} = 250$Å), the parameters η and ξ are slowly decreasing functions of both doping and oxide thickness and may be readily determined from these simple calculations. For rather thick oxides and high doping levels, η and ξ tend to rise again, which shows that the subthreshold region is penetrating farther into the above-threshold region. One can

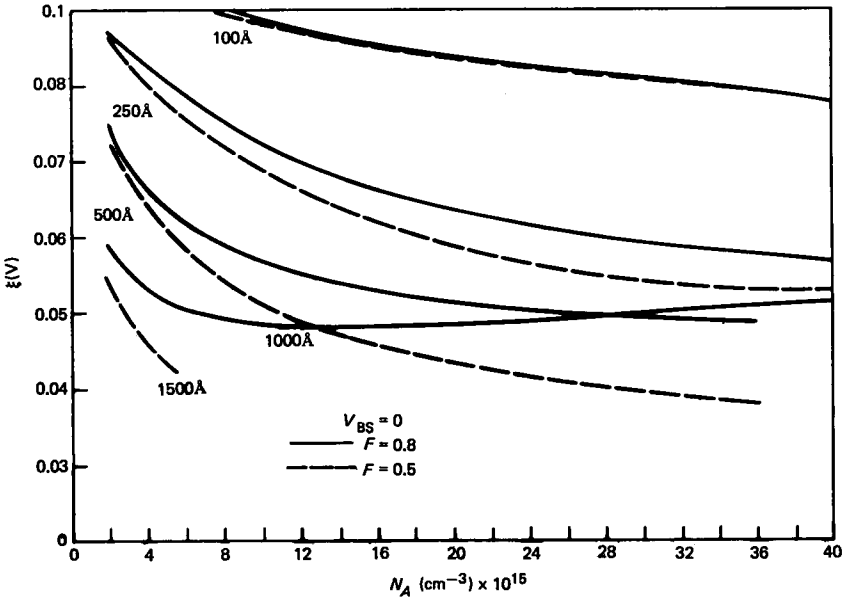

Fig. 25—Variation of ξ, the surface potential increment above threshold, as a function of substrate doping for several oxide thicknesses. The dashed curves show the effect of drain voltage through the short-channel factor F.

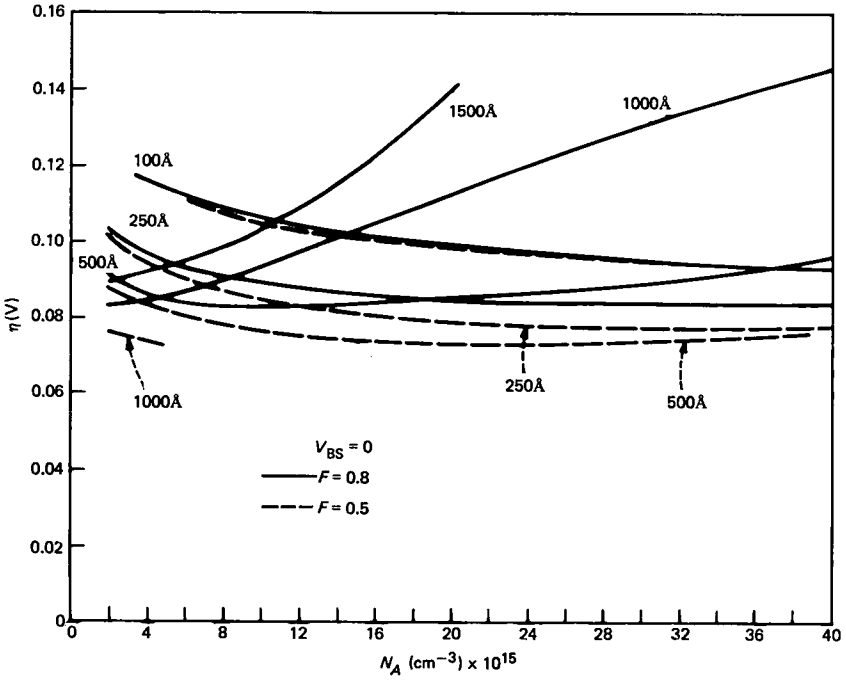

Fig. 26—Variation of η, the gate voltage above V_T for a merging of regions, as a function of substrate doping and oxide thickness. The variation with the factor F is shown by dashed lines.

think of η and ξ as offset voltage parameters that describe the protrusion of the subthreshold region into the above-threshold region.

We may now solve for $\bar{Q}/C_o$ using (196) and (189) and obtain

$$\frac{\bar{Q}}{C_o} = \frac{kT}{q}\frac{(C_o + C_S F)}{C_o}\,\xi'\sqrt{1 - 2/\xi'}. \tag{200}$$

We may then use this result, together with (182a), to determine

$$\bar{R} = \frac{1}{1 + \xi'\sqrt{1 - 2/\xi'}}. \tag{201}$$

From (200) and (199) we conclude that $\bar{Q}/C_o$ is approximately independent of bias parameters. The approximation is equivalent to ignoring the dependence of C_s upn $\bar{V}$. However, this is the same approximation we made to arrive at (191), so it is a consistent one. Within the same approximation, we see from (201) that $\bar{R}$ is also weakly dependent upon bias parameters, another fact we have used to arrive at (191). We show later in the discussion how good the approximation is.

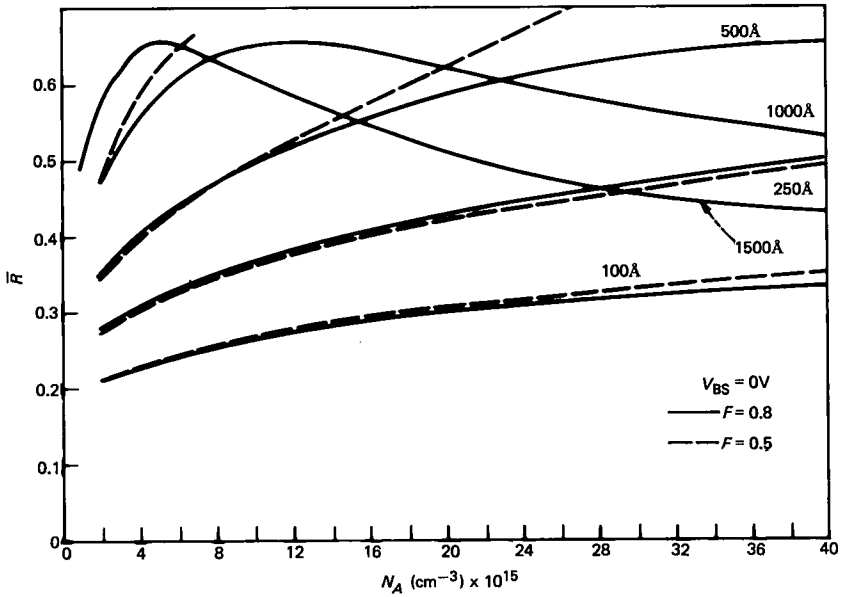

Fig. 27—Variation of $\bar{R}$, the fraction of diffusion at the pinch-off point as a function of N_A and t_{ox} for two values of F (and therefore V_{DS}).

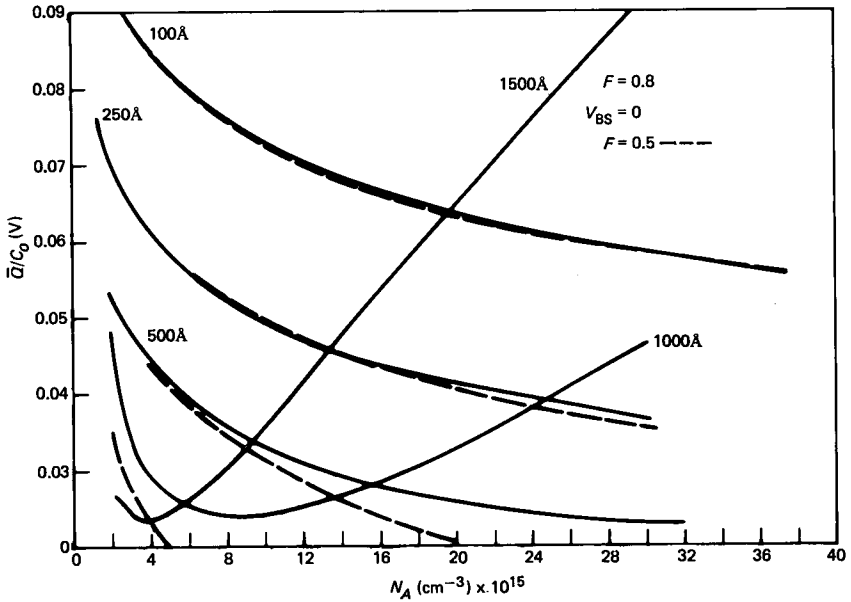

Fig. 28—Variation of $\bar{Q}$, the channel charge at the pinch-off point as a function of N_A and t_{ox} for two values of F.

The values of $\bar{R}$ and $\bar{Q}$ are plotted in Figs. 27 and 28 for substrate biases of $V_{BS} = 0, -3V$ as a function of substrate doping and oxide thickness. As for the other parameters, the effect of V_{DS} through the parameter F is not that great. Both $\bar{R}$ and $\bar{Q}$ are moderate functions of doping for a given oxide thickness. The value of $\bar{R}$ ranges from 0.2 to 0.6, but typically for $t_{ox} = 500\text{Å}$, $N_A = 10^{16}$, or $t_{ox} = 250\text{Å}$, $N_A = 5 \times 10^{16}$, we find $\bar{R} \simeq 0.4$, i.e., about one half of the current is drift and one half of the current is diffusion at the pinch-off point. As in the case of η and ξ for higher dopings and thick oxides, $\bar{Q}$ increases considerably and $\bar{R}$ drops off. Therefore, as the subthreshold transition point pushes higher above the threshold voltage, the pinch-off charge increases, and the fraction of the total current carried by drift tends to increase.

It should be noted that the parameter $\bar{R}$ has been derived from conditions near threshold in which region velocity saturation of carriers can be ignored. The form of $\bar{Q}$ [cf. (182b)] applies with or without velocity saturation; our interpretation of $\bar{Q}$ is that the effects of velocity saturation enter only through T, which becomes a hot-electron temperature under velocity-saturated conditions. From (182b) this is equivalent to $\bar{R}$ being independent of velocity saturation, and this is supported by (201) since ξ' is temperature-independent [cf. (199)]. It is, therefore, reasonable for $\bar{R}$ to be used for all gate voltages even though it was derived from conditions at the transition.

AUTHOR

Geoffrey W. Taylor, Ph.D. (Electrical Engineering), 1972, University of Toronto; AT&T Bell Laboratories, 1979—. After two years of postdoctoral studies at the University of Toronto Mr. Taylor joined the Solid State Electronics Centre at Honeywell, where he worked on the n-channel charge-coupled devices (CCDs), dynamic RAMs, and MOS device development. In 1976 he joined the Central Research Lab at Texas Instruments, where he worked for one year on the development of large-area CCDs for visible imaging and for two years on the development of metal semiconductor field-effect transistors (MESFETs) MOSFETs, CCDs, and dynamic memory structures for VLSI applications. At AT&T Bell Laboratories he has been involved in NMOS circuit design and exploratory MOS device work. He is presently investigating new device concepts in Si and GaAs.