PROPERTIES OF AN ARBITRARILY DOPED FIELD-EFFECT TRANSISTOR

Thesis by

Ira Richer

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1964

(Submitted May 21, 1964)

ACKNOWLEDGMENT

I thank Dr. R. D. Middlebrook whose guidance and encouragement underlie this work. I am grateful to the National Science Foundation for their financial support under the Graduate Fellowship program.

The following companies donated some of the transistors reported on: Fairchild, Motorola, Siliconix, and Texas Instruments. Their help is appreciated.

Miss Doris A. Barnhart, who typed the manuscript, deserves special thanks for her patience and cooperation. Finally, I would like to acknowledge my colleagues at Caltech for numerous helpful discussions during the course of this research.

ABSTRACT

The properties of p-n junction field-effect transistors (FET's) are formulated on a general basis, in terms of an arbitrary doping profile (i.e. arbitrary gate-channel impurity distribution). The external behavior is shown to be quite insensitive to the doping profile, provided that the profile satisfies certain weak restrictions. Essentially all practical structures are included in the restricted theory. A theoretical basis is thus provided for the much-used empirical conclusion that widely different types of FET's exhibit similar functional behavior. More specifically, upper and lower bounds are obtained on the normalized transconductance, drain current, input capacitance, and bias point for zero temperature coefficient of the drain current, and on the voltage-dependent parts of various figures of merit. In each case the bounds represent the solutions of two analytically simple structures, a step-junction FET and a delta-junction FET. Many practical implications stem from these results.

Finally, a complete, small-signal, low-frequency equivalent circuit for an arbitrarily doped FET is developed by considering the capacitive current that flows between the channel and the gate. Beyond pinch-off a "new" element, the forward transfer capacitance, is present in the circuit. Below pinch-off the theory predicts that the output capacitance C_{22} and the reverse transfer capacitance C_{12} differ, and in fact that $(C_{22} - C_{12}) < 0$, whereas earlier theories and intuition indicate that $(C_{22} - C_{12}) = 0$. Measurements on a

wide variety of FET's substantiate these theoretical results. The frequency limitations of the equivalent circuit and, indeed, of all the results obtained are shown to arise from the breakdown of the gradual approximation.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The fundamental principle underlying the operation of field-effect devices, namely, the modulation of the current flow between two electrodes by the application of an electric field, was conceived as early as 1928 [1]. Since then, and especially in recent years, these semiconductor devices have aroused much interest because of their singular properties and diverse potentialities. In this work we shall restrict our attention to field-effect transistors (FET's), which are three-terminal structures whose external characteristics are similar to those of vacuum pentodes, and which therefore complement in some aspects the characteristics of "conventional" transistors.

In 1948 Shockley and Pearson [2] built a prototype FET by using a thin layer of semiconductor as one plate of a parallel-plate capacitor. A potential applied to the other (metal) plate controlled the longitudinal conductivity of the semiconductor by inducing charges in it. The degree of control was much less than that predicted theoretically, the difference being attributed to the presence of surface states [3] which immobilized many of the induced charges. In the first practical FET's, deleterious surface effects were avoided because the free surface was replaced by a p-n junction. More recently, insulated-gate FET's* (thin-film devices equivalent to Shockley's

---

* These FET's are sometimes referred to as metal-oxide-semiconductor (MOS) transistors or simply as thin-film transistors (TFT's).

prototype) have been fabricated. At present, however, most commercially manufactured FET's contain p-n junctions because some surface problems are still extant in the insulated-gate types.

The essential features of FET operation may be apprehended from the representative model of Fig. 1.1, in which a two-dimensional, planar device with a high-conductivity p-region is shown. If the p-n junction is reverse-biased by a voltage V_1 between terminals 1 and 3, the resulting space-charge (or charge-depletion) region penetrates preferentially into the lower conductivity n-region. When a longitudinal field is applied by means of a voltage V_2 between terminals 1 and 2, a current flows through the channel - the undepleted portion of the n-region. The current carriers flow into the channel from terminal 1, which is therefore called the source. The carriers flow out of the channel at terminal 2, the drain. Terminal 3 is the control electrode and is called the gate. Now, because of the potential drop in the direction of current flow, the space-charge region does not have uniform width, but has the general shape shown. There are effectively no free carriers in the space-charge region, so that the (static) drain-source conductance is determined wholly by the shape of the neutral channel and, consequently, depends upon the gate-source bias V_1. Therefore, the potential applied to the gate modulates the current flow between the drain and the source.

The presence of an insulating layer between the gate and the channel in insulated-gate FET's permits the biasing of these devices with gate-source voltages of either polarity. However, in this

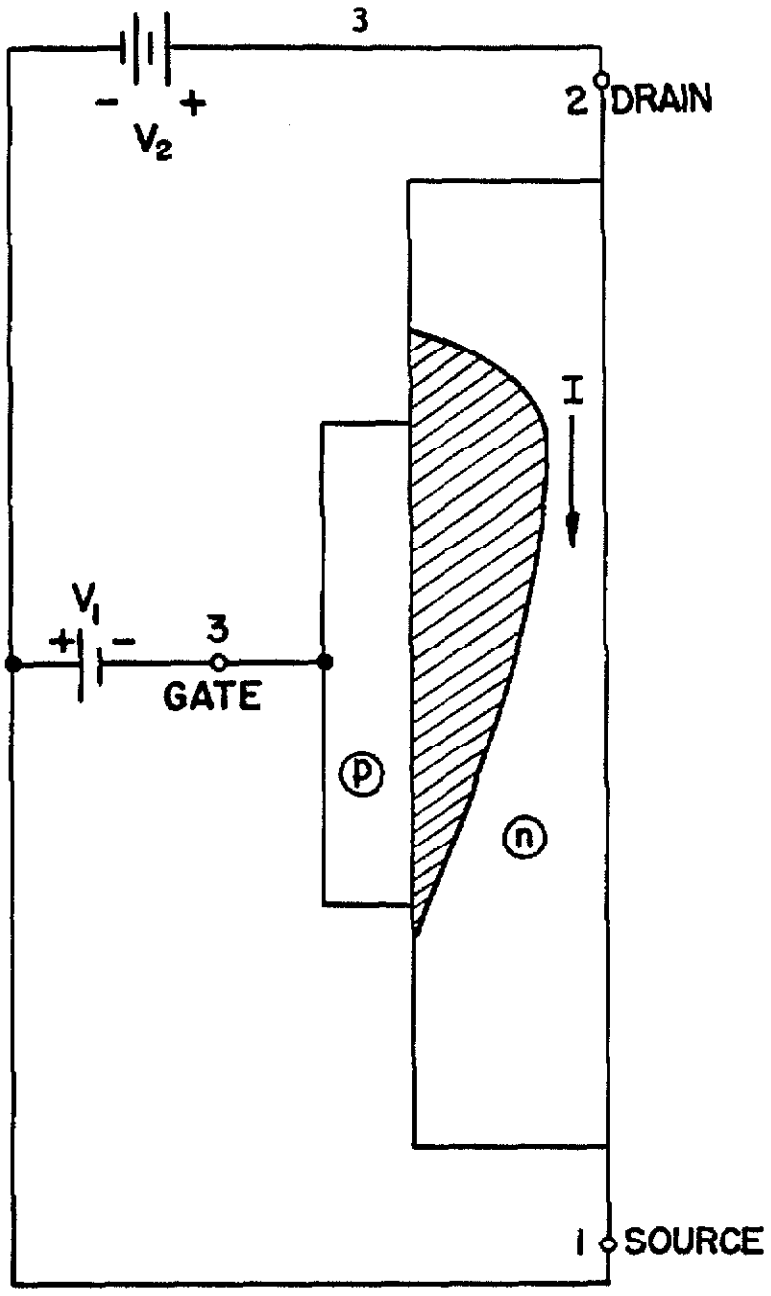


Fig. 1.1. Representative model of FET. Shaded area denotes space-charge region. The arrow indicates the direction of conventional current flow; current carriers (electrons, in this case) always flow from source to drain.

report we shall consider only the mode of operation corresponding to charge depletion within the channel. Thus, the FET's we shall investigate possess a current flow that is controlled by conductance modulation (via alteration of the shape of a conducting path), rather than by conductivity modulation.

The drain characteristics of an FET are similar to the plate characteristics of a vacuum pentode (cf. Fig. 2.7). For a given gate voltage the drain current at first increases with the drain voltage, and then, in the so-called pinch-off range, the drain current is practically independent of the drain voltage. The finite incremental output conductance that is present in the pinch-off range in an actual FET is small and is neglected in this work. Also, the gate-channel junction behaves like a reverse-biased diode, so that the input conductance is extremely small and is also neglected.

An important feature of FET's is that they are unipolar, i.e. only one type of carrier, the majority carrier, is involved in the dominant current flow mechanisms. This is in marked contrast to conventional transistors, in which both minority and majority carriers are essential to the total current flow in normal operation. The unipolar nature of the FET endows this device with several desirable properties. Most important, the current carriers travel by drift only. Hence, good frequency response, and, in particular, better frequency response than that of conventional transistors is expected. (Also, there will be no effect analogous to the minority carrier storage of the ordinary transistor.) Since recombination is not an important factor, technologically inferior materials can be

tolerated in the structure of an FET.

Although FET's have been commercially available for only a few years, many companies now produce these devices. The essential differences between various FET's are in the type of gate-channel junction, i.e. in the doping profile (impurity distribution), and in the geometry. It is easy to show that most FET geometries (excluding cylindrical) can be reduced with excellent accuracy to a single, simple, ideal structure (cf. Section 3.1). In other words, geometrical variations affect only the magnitudes of FET parameters, not their functional behavior. Doping-profile variations, however, are of a more fundamental nature. Moreover, existing profiles are almost as numerous and disparate as the manufacturing companies themselves: Some FET's contain alloy junctions; some are epitaxially grown; some contain a single diffused junction; others are double-diffused; and some contain an insulated gate. Yet, despite these very basic internal differences, the external properties of the FET's are strikingly similar. For example, particular transistors from different manufacturers may have output currents that differ by orders of magnitude at the same normalized gate bias; however, if the currents are normalized to their respective maximum values, then at any normalized gate bias these normalized currents can be so close as to be experimentally indistinguishable. The empirical results thus belie intuition: variations in the doping profile have in fact only a slight effect on the functional behavior of FET parameters.

In this work we shall attempt to establish a theoretical basis for the empirical observation that diverse types of FET's behave similarly. The problem is dealt with in Chapters II through V. The first two of these chapters provide background material. Chapter II consists in large part of a rederivation of familiar results for a step- or abrupt-junction FET; Chapter III comprises an evaluation of the approximations used in the step-junction analysis. Before the desired goal can be approached, the equations describing the operation of an FET must be written in a general form, in terms of an arbitrary doping profile. This is done in Chapter IV. Then, in Chapter V it is proved that most FET's do indeed have very similar characteristics. This result justifies certain simple approximations to the equations describing the operation of FET's, and, equally important, allows us to appreciate the range of validity of these approximations.

As another major problem in this work we investigate the equivalent circuit for an FET. Heretofore, equivalent circuits were generally semi-empirical, and if the capacitive elements were treated theoretically, they were defined in terms of voltage rates of change of the charge in transit. In Chapter VI we derive, for an arbitrary FET, the actual equivalent circuit, i.e. with capacitive as well as conductive elements expressed as ratios of a-c components of current and voltage. Two previously unpredicted capacitive elements are included in the complete circuit, and the presence of these elements

is confirmed by measurements on a wide variety of FET's. Finally, by revealing a shortcoming in one of the key approximations that is almost universally employed, the treatment of Chapter VI points out the frequency limitations of the results obtained from this and many other analyses.

CHAPTER II

ANALYSIS OF A STEP-JUNCTION FIELD-EFFECT TRANSISTOR

In this chapter we shall analyze the step-junction FET in some detail, because the physical insight gained from the solution of this specific doping profile will be helpful when we consider the general profile in Chapter IV. In fact, most of the conclusions regarding the operation of a step-junction FET may be directly applied - in a qualitative manner - to an arbitrarily doped FET.

We begin in the first two sections by setting up a general model and summarizing the assumptions to be applied to this model. Then, in the remaining sections, we restrict our attention to the step-junction device and derive and interpret the equations describing its operation.

2.1 Model

Figure 2.1 shows the model we shall use for the FET throughout this report. As mentioned in Chapter I, the principle of operation of an FET is that the potential applied to one electrode - the gate - controls the current flowing between two other electrodes - the source and the drain - by altering the shape of a space-charge or depletion region. The potential of the gate (with respect to circuit ground) is denoted by V_g. The source and the drain are separated by the distance L; the source is grounded, and the drain is biased at

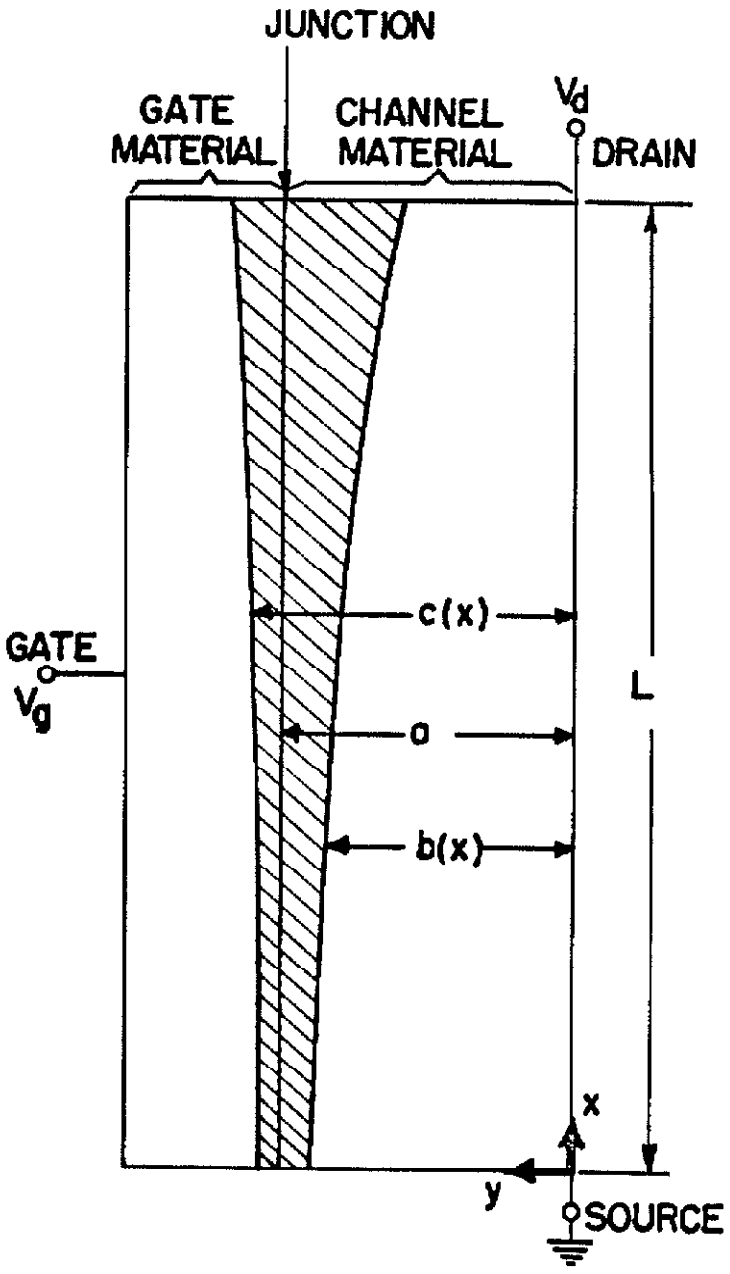


Fig. 2.1. Model of FET used in the analysis. The device has depth A perpendicular to the x-y plane. Shaded area denotes space-charge region.

V_d.* The subscript s, d, or g on any quantity indicates that the quantity pertains to the source, drain, or gate, respectively. The origin of the coordinate system is at the source electrode, with x measured in the direction of the drain and y measured in the direction of the gate. Our model is two-dimensional: we assume a uniform depth A perpendicular to the x-y plane.

The applied potentials set up a space-charge region of width (c - b). Of course, (c - b) is a function of the bias voltages, and, in general, is also a function of x. We assume that there are no free carriers in the depletion region, so that current flows between the source and drain only within the charge-neutral channel; the width of the channel is denoted by b(x).

The excess density of donors over acceptors, multiplied by the magnitude of the electronic charge, is called the doping or impurity density (or profile) and is denoted by ρ. Practical devices have one-dimensional doping profiles, so that $\rho = \rho(y)$. In n-type semiconductor $\rho > 0$, while in p-type $\rho < 0$. (The sign of ρ is the same as the sign of the charges in the depletion region.) The gate material and the channel material are separated by a junction at $y = a$. That is, $c(a) = C$. We assume that all impurities are ionized at normal operating temperatures. Then the mobile charge density, denoted by ρ_m, is equal to $-\rho$ in the charge-neutral channel.

---

* Our model is symmetrical and hence is unaffected by an interchange of source and drain. Actual devices may have some non-symmetries, so that an interchange of source and drain will modify the numerical values of device parameters.

2.2. Assumptions

In this section we list carefully the assumptions and approximations used in the subsequent analysis. The validity of these assumptions is discussed fully in the next chapter. The following list is very roughly in order of decreasing validity. The first two hypotheses will be entirely removed in subsequent chapters; the next two are generally well justified; the last three require some analysis for their justification.

Our first assumption arises because at present we are considering a step-junction FET.

(1) The doping profile is given by

$$\rho(y) = \begin{cases} +\rho_c , & 0 \le y < a \\ -\beta\rho_o, & a < y \end{cases} \tag{2.1}$$

where ρ_o and β are positive constants. We have chosen an n-channel device. For simplicity, we assume that $\beta \gg 1$, so that the conductivity of the gate material is much greater than that of the channel material. When the p-n junction is reverse biased, essentially all of the space-charge region appears in the channel material, and

$$c(x) = a = \text{constant} \tag{2.2}$$

(2) The built-in potential is negligible compared with the gate-channel potential. The difference between the Fermi levels of the p- and n-regions produces a contact or built-in potential across

the gate-channel junction. For silicon, this potential is of the order of 0.5v.

(3) The drain current is carried only by majority carriers (electrons, for the doping profile given in Eq. 2.1).

(4) Direct gate current is negligible. This current, which consists of the normal saturation current of a reverse-biased junction and possibly of recombination-generation current produced in the space-charge region, is small in comparison with the drain current.

(5) The carrier mobility is constant.

(6) The boundary between the depletion region and the channel is sharp. This assumption enables us to define accurately the channel, and to assert that the drain current flows entirely within this channel (since in this case $\rho_m = 0$ in the depletion region).

The foregoing assumptions are not uncommon in the consideration of many devices involving p-n junctions. The following two approximations, however, are peculiar to FET's and are necessary for an analytically simple solution.

(7) (a) In the channel the magnitude of the y-component of the electric field E is much smaller than the magnitude of the x-component:

$$|E_y| \ll |E_x| \quad , \quad 0 \le y \le b \tag{2.3}$$

(b) If ϵ is the permittivity of the semiconductor material, then in the space-charge region

$$\left|\frac{\partial E_x}{\partial x}\right| \ll \left|\frac{\rho}{\epsilon}\right| , \qquad b \le y \le c \qquad (2.4)$$

Assumption (7) constitutes the "gradual approximation" [4] since it implies that the channel width does not change rapidly. The motivation for this assumption arises when an FET is analyzed with $V_d \approx 0$. In this case $b(x) \approx$ constant, and the gradual approximation is well justified because the channel field is essentially one-dimensional in the x-direction, while the field in the space-charge region is almost wholly in the y-direction. In order to maintain a one-dimensional problem, the gradual approximation is applied even when V_d is not small.

In Section 2.4, one additional assumption will be made. It is not included here because some preliminary analysis is necessary before it can be properly understood.

2.3. D-C and A-C Analysis Below Pinch-Off

Now that the model has been explained and the assumptions set down, we are ready to derive the equations governing the operation of the step-junction FET. The principal results of the d-c solution are the expressions for the drain current and for the channel boundary; the a-c solution yields the small-signal equivalent circuit. Except for the capacitances, the formulae obtained in this section are given by Shockley [4], so that most of the actual analysis performed

here will be brief.

2.3.1. D-C Analysis: Drain Current and Channel Curves

Implicitly, the method we shall use here consists of solving both the space-charge region and the channel for two relations between the channel width and the potential along the channel boundary, and then matching the two relations to obtain a consistent result. Actually, we proceed by immediately incorporating the space-charge-region result into the equations describing the channel. Then the solution of the latter equations yields the quantities of interest directly.

Let $V(x,y)$ be the potential at the point (x,y). In the space-charge region, V obeys Poisson's equation. Because of the gradual approximation - in particular, Eq. 2.4 - we have

$$\frac{\partial^2 V}{\partial y^2} = - \frac{\rho(y)}{\epsilon} , \qquad b(x) \leq y \leq c(x) \qquad (2.5)$$

At the channel edge of the depletion region the electric field is zero [because of assumption (6)],

$$E_y = - \frac{\partial V}{\partial y} = 0 \qquad \text{at} \qquad y = b(x) \qquad (2.6)$$

and the potential along the gate edge of the depletion region is V_g,

$$V[x,c(x)] = V_g \qquad (2.7)$$

Equations 2.6 and 2.7 are the boundary conditions on 2.5. Using the step-junction profile Eq. 2.1 defined in the previous section, we easily integrate 2.5 to get for the potential in the space-charge region

$$V(x,y) = V_g + \frac{\rho_o}{2\epsilon}\left[(a - b)^2 - (y - b)^2\right], \quad b(x) \le y \le a \tag{2.8}$$

since $c(x) = a$ because of the high conductivity of the gate material. Again because of the gradual approximation - this time Eq. 2.3 - the current from the drain to the source flows parallel to the x-axis, and hence the potential at any point (x,y) within the channel is equal to the potential at the point $[x,b(x)]$:

$$V(x,b) = V_g + \frac{\rho_o}{2\epsilon}[a - b(x)]^2$$

$$= V_g + W_o\left[1 - \frac{b(x)}{a}\right]^2, \qquad 0 \le y \le b(x) \tag{2.9}$$

where

$$W_o = \frac{\rho_o a^2}{2\epsilon} \tag{2.10}$$

From Eq. 2.9 we see that if $V - V_g = W_o$, then $b(x) = 0$. That is, W_o is the magnitude of the reverse bias on the gate-channel junction which causes the channel width to become zero; or, more descriptively, the potential W_o causes the channel to become "pinched-off." For

this reason, W_o is referred to as the pinch-off voltage.

In order to simplify many of the formulae that we shall derive, we define several normalized parameters:

$$u(x) \equiv \frac{b(x)}{a} : \quad \text{the fractional width of the channel material occupied by the channel} \tag{2.11}$$

$$s \equiv \frac{|V_g|}{W_o} : \quad \text{the magnitude of the ratio of the source-gate potential to the pinch-off voltage}^* \tag{2.12}$$

$$d \equiv \frac{|V_g - V_d|}{W_o} : \quad \text{the magnitude of the ratio of the drain-gate potential to the pinch-off voltage}^* \tag{2.13}$$

Thus, s and d are normalized potentials that are measured with respect to the gate; below pinch-off their values lie in the range $0 \leq s,d \leq 1$. With this notation Eq. 2.9 becomes

$$\frac{V(x,b)}{W_o} = -s + (1 - u)^2 \tag{2.14}$$

---

* The built-in potential is included in the more accurate definitions of s and d given in Eqs. 3.6 and 3.7.

because $V_g < 0$ in an n-channel device. This equation, and most of the equations that follow, is valid for both n- and p-channel FET's if the appropriate sign of V is used.

Consider now the differential element of channel pictured in Fig. 2.2. The resistance between the x- and (x+dx)-planes is

$$dR = \frac{dx}{\mu\rho_o bA} = \frac{1}{\mu\rho_o aA}\frac{dx}{u}$$

where μ is the mobility of majority carriers in the channel. The voltage drop across the element is -dV, which, by 2.14 may be expressed as

$$-dV = 2W_o(1 - u)\, du$$

Hence, the magnitude of the current flowing through any section of channel is (since $du/dx < 0$)

$$I = \frac{dV}{dR} = -2G_o LW_o u(1 - u)\frac{du}{dx} \qquad (2.15)$$

in which

$$G_o = \frac{\mu\rho_o aA}{L} \qquad (2.16)$$

is the conductance of the channel in the absence of biases.

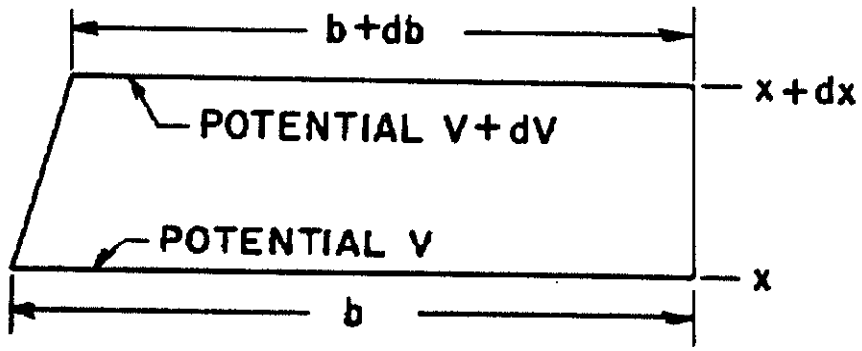


Fig. 2.2. Differential element of channel.

Because no current flows through the space-charge region, I is not a function of x, and Eq. 2.15 may easily be integrated. Also, the drain current and the source current are identical and may be unambiguously identified by the same symbol I. For convenience in later discussions we shall always consider the drain current to be a positive quantity even though I may flow into or out of the drain, depending on whether the channel is n- or p-type, respectively. From Eq. 2.14 we see that

$$u(0) = 1 - \sqrt{s + \frac{V(0,b)}{W_o}} = 1 - \sqrt{s}$$

$$u(L) = 1 - \sqrt{s + \frac{V(L,b)}{W_o}} = 1 - \sqrt{\frac{V_d - V_g}{W_o}} = 1 - \sqrt{d}$$

so that integration of 2.15 over the whole channel length yields

$$I \int_0^L dx = -2G_o L W_o \int_{1-\sqrt{s}}^{1-\sqrt{d}} u(1 - u)\, du$$

or

$$I(s,d) = I_o\left[(1 - \sqrt{s})^2(1 + 2\sqrt{s}) - (1 - \sqrt{d})^2(1 + 2\sqrt{d})\right] \qquad (2.17)$$

The multiplicative factor I_o in this equation is given by

$$I_o \equiv I(0,1) \qquad (2.18)$$

$$= \frac{G_o W_o}{3} \tag{2.19}$$

and is, as we shall show in Section 2.4, the maximum drain current that flows through the device without forward bias on the gate-channel junction. If the potential W_o were applied across a bar of material of conductance G_o, a current $I = G_o W_o$ would flow. With W_o applied across the length of the channel of an FET, the current is some factor less than $G_o W_o$ because of the presence of the depletion region. For a step junction FET (with $V_g = 0$) we see from Eq. 2.19 that this reduction factor is 3.

Consistent with the assumptions made in Section 2.2, Eq. 2.17 for the drain current is valid for voltages up to pinch-off, but not beyond; i.e. for $s < d \leq 1$, where the first inequality arises because the drain-gate potential must always be greater in magnitude than the source-gate potential under normal bias conditions (current flowing into the drain in an n-channel device). We defer discussion of Eq. 2.17 until we extend it to voltages $d > 1$ in Section 2.4.

We now derive the equation for the channel shape by integrating Eq. 2.15 over only part of the channel length, from $x = 0$, $u = u(0) = (1 - \sqrt{s})$ to the arbitrary channel boundary point (x,u):

$$I \frac{x}{L} = I_o \left[(1 - \sqrt{s})^2 (1 + 2\sqrt{s}) - u^2 (3 - 2u) \right]$$

The current I may be eliminated by use of 2.17 to give x as a function of u with s and d as the only parameters:

$$\frac{x}{L} = \frac{(1 - u)^2(1 + 2u) - s(3 - 2\sqrt{s})}{d(3 - 2\sqrt{d}) - s(3 - 2\sqrt{s})} \tag{2.20}$$

The curves drawn in Fig. 2.3 show u versus x/L with s (or V_g) fixed and d (or V_d) as a parameter, and with $d - s$ (or V_d) fixed and s (or V_g) as a parameter. The vertical marks on some channel curves indicate, for the particular case $L/a = 10$, the maximum values of x/L for which the gradual approximation is valid. These maximum values are based on the analysis of Section 3.4. Curves with no marks are valid for their entire length. These curves should be studied carefully, because a clear understanding of how the channel shape is altered as the biases are changed gives much insight into the characteristics of the FET.

2.3.2. A-C Analysis: Small-Signal Equivalent Circuit

We are now able to determine the small-signal, short-circuit admittance parameters. From Eq. 2.17 we easily obtain the (forward) transconductance and the output conductance below pinch-off. Since both s and d depend on V_g, the transconductance g_m, defined by

$$g_m = \left| \frac{\partial I}{\partial V_g} \right| \tag{2.21}$$

is

$$g_m = G_o(\sqrt{d} - \sqrt{s}) \tag{2.22}$$

which is the same as the conductance of a channel of constant frac-

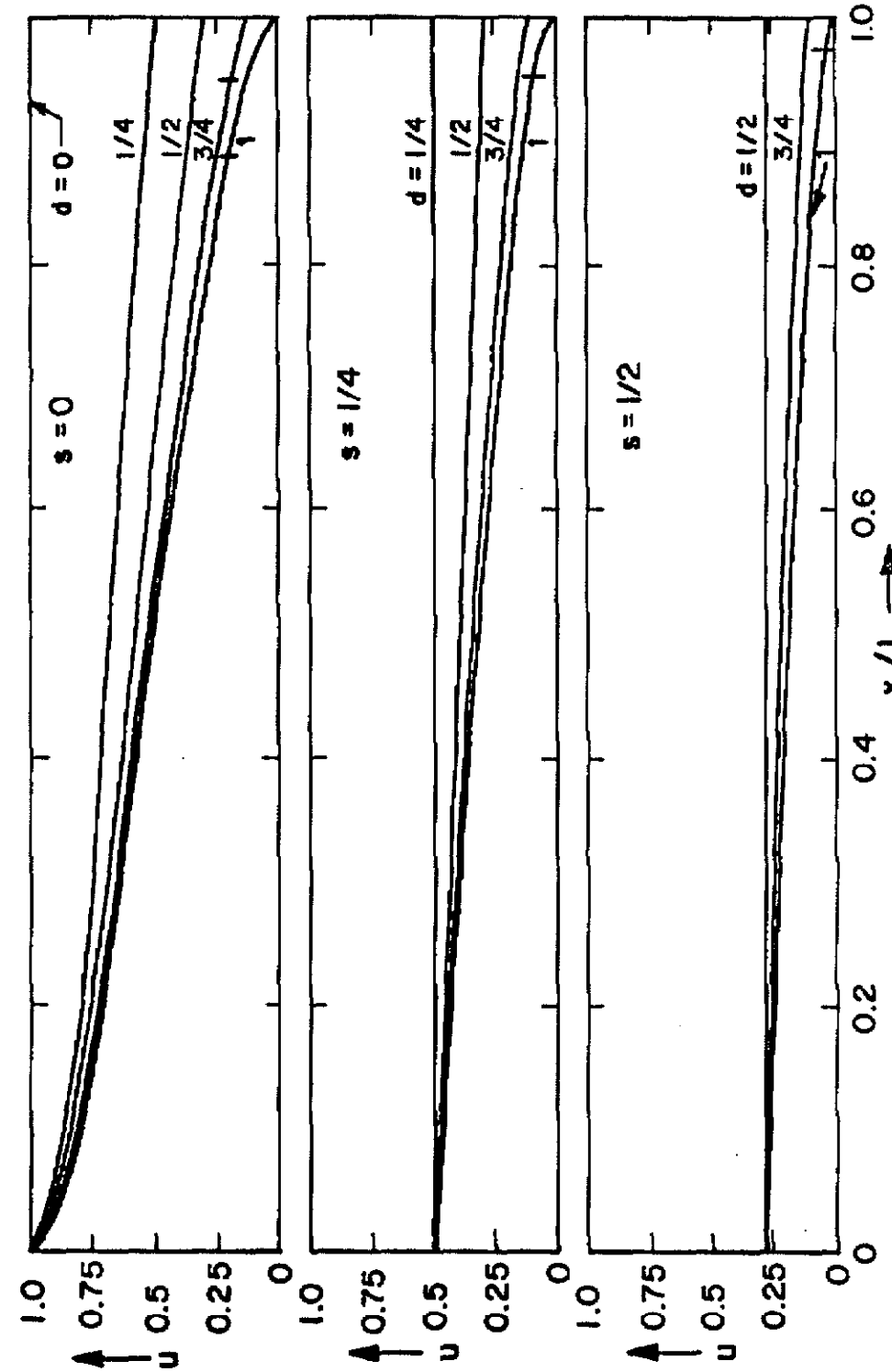


Fig. 2.3. Channel curves for the step-junction FET. The vertical marks on some curves denote, for the particular case L/a = 10, the maximum values of x/L that are consistent with the gradual approximation. Curves with no marks are accurate for their entire length.

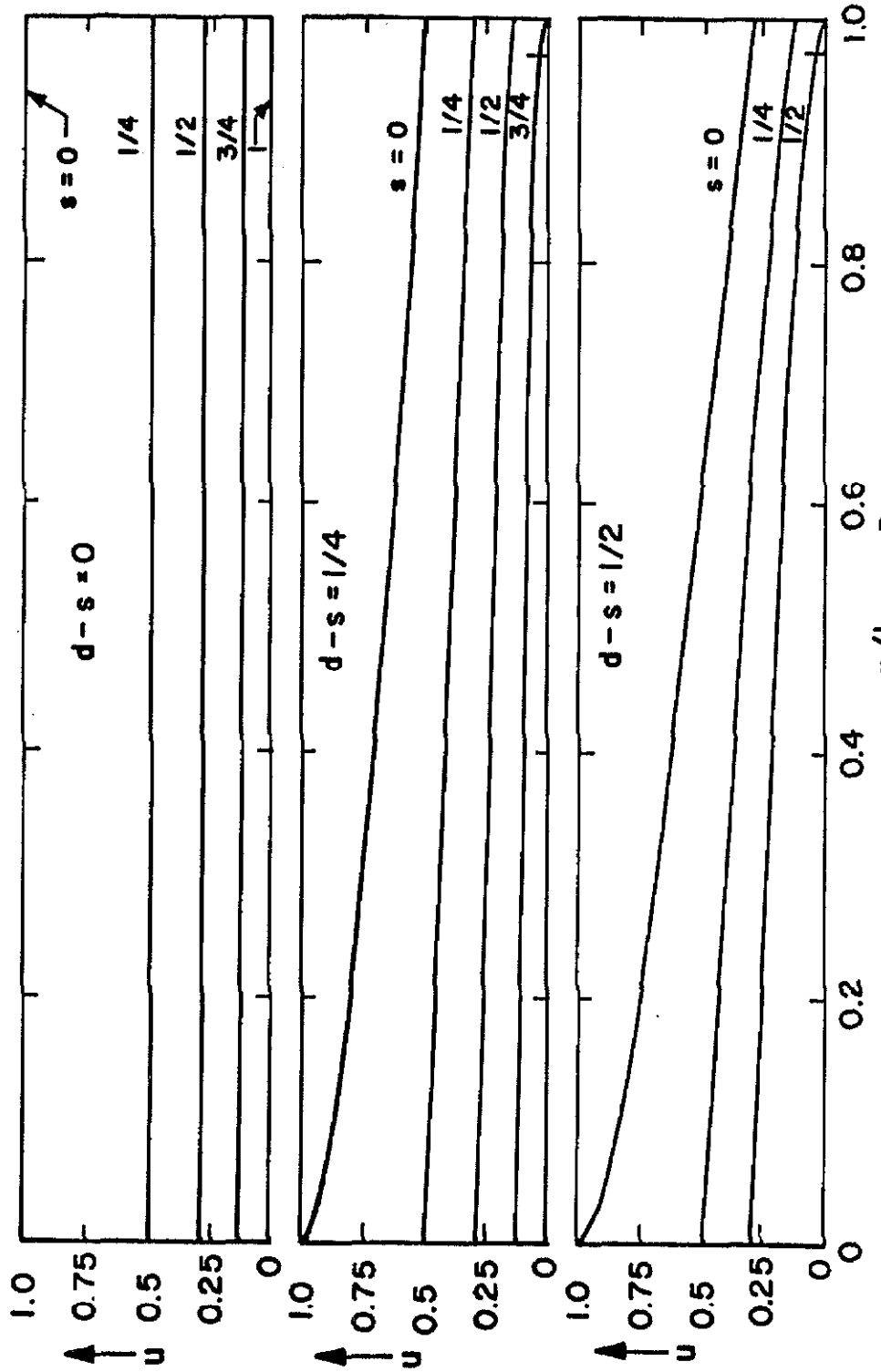


Fig. 2.3 (cont'd). Channel curves with s as the running parameter.

tional width $u(0) - u(L)$. For $d = s$, $g_m = 0$. This is logical because for these biases the channel is uniform and no current flows, and a change in the gate voltage cannot alter these conditions. The output conductance is defined by

$$G_{22} = \left| \frac{\partial I}{\partial V_d} \right| \tag{2.23}$$

and has the value

$$G_{22} = G_o(1 - \sqrt{a}) \tag{2.24}$$

We note that G_{22} is equal to the conductance of a channel of constant fractional width $u(L)$; i.e. G_{22} depends only on the drain-gate potential and goes to zero when that potential equals V_o.

The remaining elements in the equivalent circuit are capacitive. We define the short-circuit input "charge-capacitance" C_{11}^* and output "charge-capacitance" C_{22}^* in terms of the voltage rate of change of the charge Q in transit between the source and the drain*:

---

* More precisely, $C_{11}^* = dq_{in}/dV_g$ where q_{in} is the charge placed on the gate. But for every element of input charge q_{in}, an equal amount of charge goes into the space-charge region from the channel. Hence, $dq_{in}/dQ = -1$, and Eq. 2.25 is valid.

$$C_{11}^{*} = \left| \frac{\partial Q}{\partial V_g} \right| \tag{2.25}$$

$$C_{22}^{*} = \left| \frac{\partial Q}{\partial V_d} \right| \tag{2.26}$$

These quantities may differ from the "real" short-circuit input and output capacitances C_{11} and C_{22}, which are defined by

$$C_{11} = \frac{1}{\omega} \frac{1}{v_g} [\text{quadrature component of } i_g]_{v_d=0} \tag{2.27}$$

$$C_{22} = \frac{1}{\omega} \frac{1}{v_d} [\text{quadrature component of } i_d]_{v_g=0} \tag{2.28}$$

where ω is the radian frequency and v and i respectively denote the a-c components of voltage and current. The capacitances denoted by C are the real circuit parameters since they are the values that would be measured experimentally at the device terminals. The charge-capacitances C^{*} are useful quantities, however, because we shall see in Chapter VI that they do appear in the "real" equivalent circuit. In fact, C_{11}^{*} is identical to C_{11}, and C_{22}^{*} is the fraction of C_{22} that appears between the gate and the drain. Since the method of computation implied by Eqs. 2.25 and 2.26 is significantly easier than that implied by 2.27 and 2.28, and since the former approach is easier to visualize physically, we shall continue the analysis by calculating C_{11}^{*} and C_{22}^{*} and discussing them in some detail.

For a uniformly doped channel, Q is simply proportional to the channel area. Thus, since $u(0) \geq u(L)$,

$$Q = \rho_o aAL\left[u(L) + \int_{u(L)}^{u(0)} \left(\frac{x}{L}\right) du\right]$$

A straightforward calculation using 2.20 gives

$$Q = \frac{\rho_o aAL}{2}\left[1 + 3\,\frac{d(1-\sqrt{d})^2 - s(1-\sqrt{s})^2}{d(3-2\sqrt{d}) - s(3-2\sqrt{s})}\right] \tag{2.29}$$

and application of Eqs. 2.25 and 2.26 yields

$$C_{11}^* = C_o^*(\sqrt{d}+\sqrt{s})\,\frac{(1-\sqrt{d})^2 + 4(1-\sqrt{d})(1-\sqrt{s}) + (1-\sqrt{s})^2}{[3(\sqrt{d}+\sqrt{s}) - 2(d+\sqrt{ds}+s)]^2} \tag{2.30}$$

$$C_{22}^* = C_o^*(\sqrt{d}+\sqrt{s})\,\frac{(1-\sqrt{d})^2 + 2(1-\sqrt{d})(1-\sqrt{s})}{[3(\sqrt{d}+\sqrt{s}) - 2(d+\sqrt{ds}+s)]^2} +$$

$$+ C_o^*(\sqrt{d}-\sqrt{s})\,\frac{(1-\sqrt{d})(1-\sqrt{s})}{[3(\sqrt{d}+\sqrt{s}) - 2(d+\sqrt{ds}+s)]^2} \tag{2.31}$$

where we have defined

$$C_o^* \equiv C_{11}^*(0,1)$$

$$= 3\epsilon A\,\frac{L}{a} \tag{2.32}$$

These formulae are invalid near $s = d = 0$ since we have neglected

the built-in potential.

It is appropriate, here, to discuss an error that is committed by several authors [5]. A planar junction of area $A\,dx$ with a depletion-region thickness $(a - b)$ has a capacitance $dC = (\epsilon A dx)/(a - b)$. True, we might be tempted to assert that the input capacitance of an FET should be

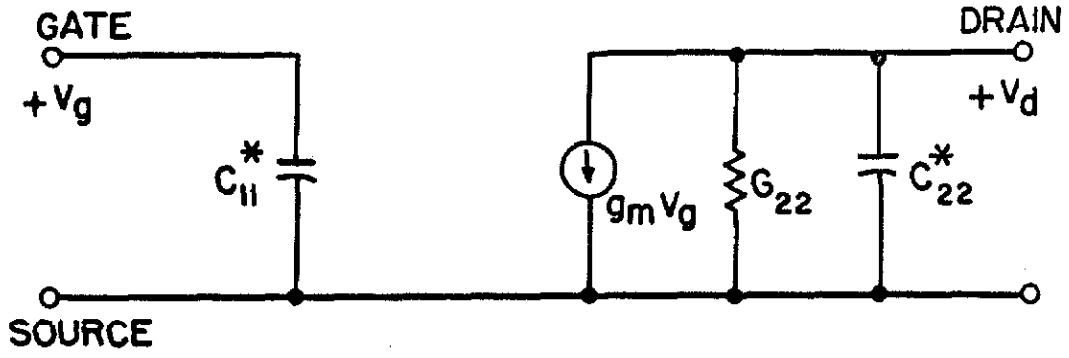
$$\frac{\epsilon A}{a}\int_0^L \frac{dx}{1 - u} = C_o^* \frac{(1 - \sqrt{d}) + (1 - \sqrt{s})}{3(\sqrt{d} + \sqrt{s}) - 2(d + \sqrt{ds} + s)}$$

where dx was calculated from 2.20. This expression differs from the previous expression 2.30 for the input charge-capacitance. The misunderstanding is that by integrating, we have tacitly assumed that the differential, parallel-plate capacitors are all connected in parallel. This is not true. The capacitors do have one common terminal (the gate), but the current flowing through the resistance of the channel results in the other terminals having a difference of potential.*

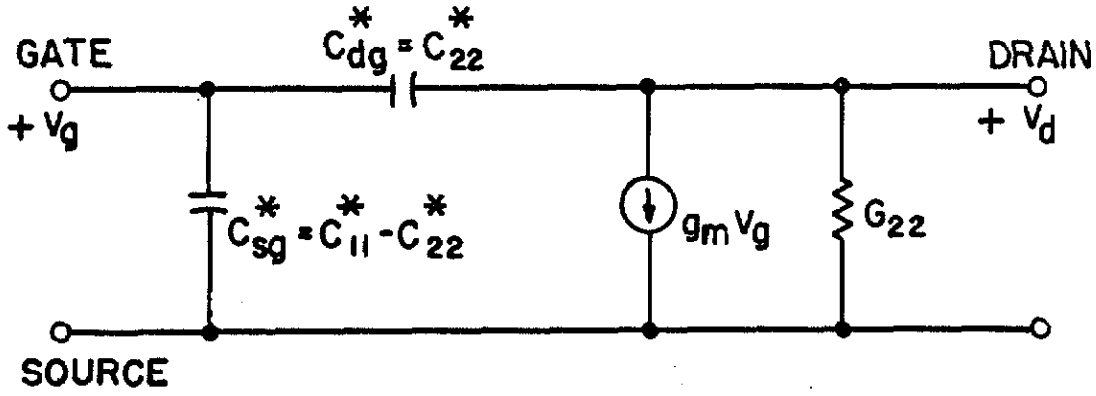
In Fig. 2.4a we show the equivalent circuit whose element values are computed in Eqs. 2.22, 2.24, 2.30, and 2.31. These elements, by definition, are short-circuit quantities, so that part of them may appear between the gate and the drain. Actually, the geometry of the FET suggests that C_{22}^* should appear entirely between the gate and the drain, for both C_{11}^* and C_{22}^* arise only from the field in the

---

* This statement is true except when $s = d$. In this case we expect the above formula to be correct, and indeed, it is.



(a)



(b)

Fig. 2.4. Small-signal equivalent circuits for the step-junction FET: (a) General circuit; (b) Modified circuit suggested by arguments in the text.

space-charge region. But all the field lines in the space-charge region emanate from the gate; i.e. none of these field lines exist between the source and the drain. Therefore, no capacitance should appear between the source and the drain.*

This intuitive conclusion is supported by two theoretical arguments. First consider the case when $s = d$ $(V_d = 0)$. Then the channel is of uniform width $b = a(1 - \sqrt{s})$, and the input charge-capacitance from Eq. 2.30 is

$$C_{11}^* = \frac{\epsilon LA}{a\sqrt{s}}, \qquad (s = d)$$

This is merely the capacitance of a reverse-biased, planar p-n junction of area LA with depletion region thickness $(a\sqrt{s})$, and is the result expected. But in our model, when $V_d = 0$, the source and the drain are indistinguishable, so we must have equal capacitances between each of these electrodes and the gate. If C_{dg}^* represents the drain-gate charge-capacitance, then the source-gate charge-capacitance is $C_{sg}^* = (C_{11}^* - C_{dg}^*)$ and we have

$$(C_{11}^* - C_{dg}^*) = C_{dg}^*, \qquad \text{or} \qquad C_{11}^* = 2C_{dg}^*$$

---

* Of course, whenever current flows between the source and the drain, there are field lines between these two terminals. These field lines generate the drain-source interelectrode capacitance, which, since it is an extrinsic parameter, is not significant in this discussion.

By Eq. 2.31, when $s = d$ the output charge-capacitance is

$$C^*_{22} = \frac{1}{2}\frac{\epsilon LA}{a\sqrt{s}} = \frac{C^*_{11}}{2} = C^*_{dg}$$

indicating that the actual equivalent circuit is that shown in Fig. 2.4b. (Since the input resistance is infinite, there is no ambiguity in the position of C_{22}.)

Now consider the general case. We observe from Eqs. 2.30 and 2.31 that C^*_{11} may be expressed as

$$C^*_{11}(s,d) = C^*_{22}(s,d) + C^*_{22}(d,s) \qquad (2.33)$$

This is an interesting result. An instructive interpretation of this relation may be obtained by reference to Fig. 2.5. In parts (a) and (b) we schematically show two n-channel transistors with biases (a) $V_g = -sW_o$, $V_d = (d - s)W_o > 0$, and (b) $V_g = -dW_o$, $V_d = (s - d)W_o < 0$. Interchanging s and d does not signify interchanging the gate and drain voltages: the drain voltage is given by $(d - s)W_o$. Interchanging s and d reverses the orientation of the channel boundary; the channel <u>shape</u> is maintained.* From Eq. 2.33 we see immediately that the short-circuit input charge-capacitance is the same in case (a) and in case (b). Thus, C^*_{11} is independent of the direction of current flow, but

---

* This is obvious since with s and d interchanged the gate-source and the gate-drain potentials are interchanged.

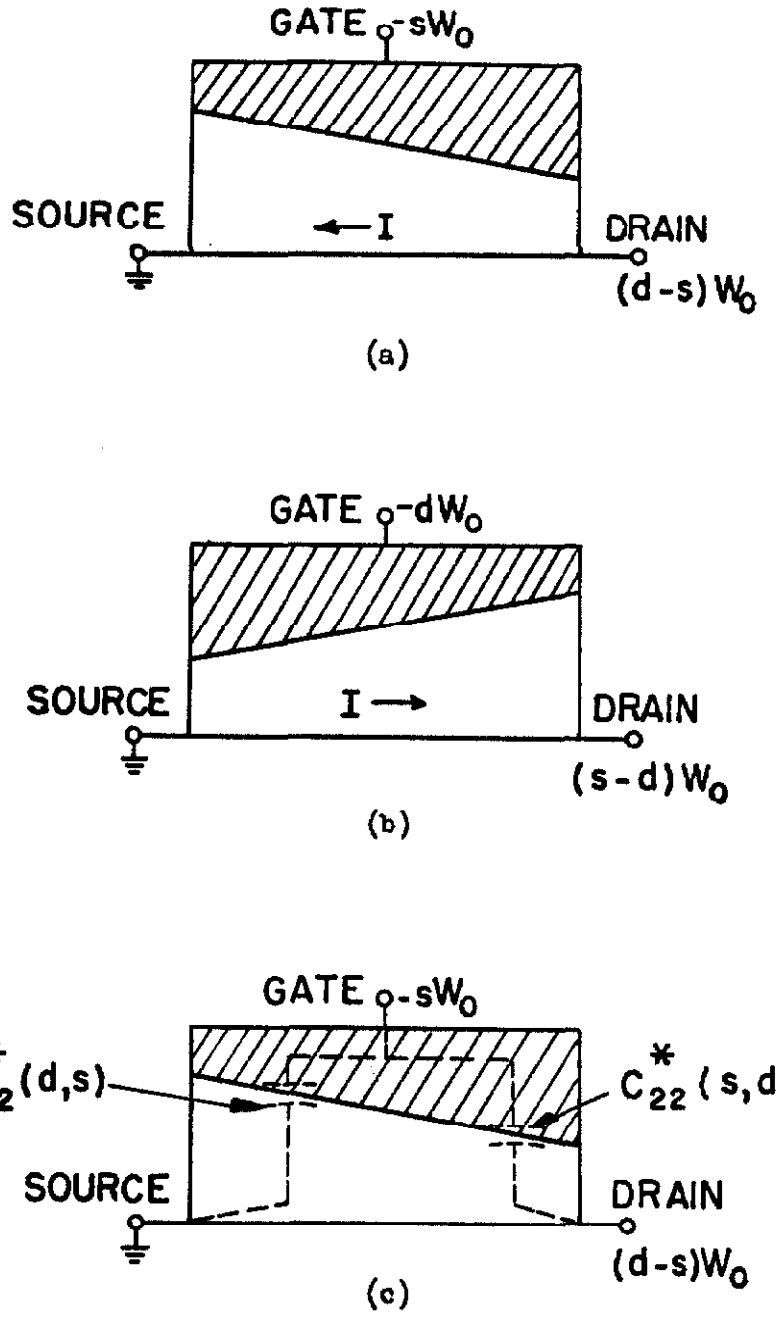


Fig. 2.5. Pertinent to the derivation of the equivalent circuit of Fig. 2.4b. Shaded area denotes space-charge region.

depends somehow upon the overall channel shape. On the other hand, C_{22}^{*} depends both upon the direction of current flow and upon the channel shape. Or, equivalently, we may say that C_{22}^{*} depends only upon the channel shape "near" the drain.* These remarks and the antisymmetry of Figs. 2.5a and 2.5b suggest the interpretation of the input and output charge-capacitances shown in Fig. 2.5c. This particular interpretation satisfies Eq. 2.33 and also predicts correct results under an interchange of s and d. The equality of C_{22}^{*} and C_{dg}^{*} follows at once from Fig. 2.5c, thereby indicating the general validity of the modified equivalent circuit of Fig. 2.4b.

The description of the step junction FET developed in this section is only partially complete, since we have not yet considered the important pinch-off range of operation. However, all the necessary derivations have been performed because, as we now show, values beyond pinch-off are obtained very simply from values below pinch-off.

2.4. Extension of Results Beyond Pinch-Off

As mentioned earlier, the results of Section 2.3 are valid for potentials satisfying $s \leq d \leq 1$. In order to extend the formulae to

---

* In normal operation - as in Fig. 2.5a - the channel is narrower near the drain. Capacitance varies inversely with the width of the space-charge region. Therefore, the capacitance associated with the part of the channel near the drain is smaller than the capacitance associated with the part of the channel near the source. These arguments, coupled with the fact that $C_{22}^{*}(d,s) \geq C_{22}^{*}(s,d)$ (which follows easily from 2.31) justify the above assertion.

$d > 1$, we make the one further assumption, in addition to the seven listed in Section 2.2. This new assumption is fundamental in all the first-order theory to be developed in this work:

(8) (a) At pinch-off ($d = 1$) all the previously derived results are valid.

(b) The effect of d for $d > 1$ may be neglected. Crude as this approximation may seem, it does give an accurate representation of the operation of an FET. The drain is at potential dW_o with respect to the gate. If we temporarily assert that the channel is pinched-off at any point whose potential is W_c (or greater) with respect to the gate, then, in effect, (b) implies that the finite voltage drop $(d - 1)W_o$ appears across a zero length of space-charge region. Actually, the channel never really pinches-off, because then no drain current would be able to flow through the resulting space-charge region. This physical fact is allowed for mathematically by a breakdown of the gradual approximation at points in the channel where the potential with respect to the gate is near W_o. Instead of pinching off ($u = 0$), the channel merely becomes very narrow ($u \approx 0$, but $u > 0$), and the electric field is very high. But a high electric field means that a large voltage drop occurs in a relatively short distance, and the potential $(d - 1)W_o$ appears across a very small length of channel. The above-mentioned implication of assumption (b) is, therefore, not so farfetched as it seemed at first glance. Furthermore, we should note that (b) follows almost directly from (a). For we see from Eqs. 2.24 and 2.31 that

when $d = 1$, $G_{22} = 0$ and $C^*_{22} = 0$: the drain-current - drain-voltage curves are horizontal and the incremental equivalent circuit contains no drain-voltage dependent elements. These heuristic arguments are borne out empirically because the characteristics of FET's are practically independent of drain voltage in the pinch-off range.

2.4.1. Drain Current

Applying assumption (8), we obtain the drain current beyond pinch-off by substituting $d = 1$ into Eq. 2.17:

$$I(s,d) = I_o(1 - \sqrt{s})^2(1 + 2\sqrt{s}) , \qquad d \geq 1 \qquad (2.34)$$

The normalized current obtained from Eqs. 2.17 and 2.34 is plotted as a function of s and d in Fig. 2.6. Because our model is symmetrical, $I(s,d) = -I(d,s)$; that is, the surface of Fig. 2.6, if extended, would be antisymmetric about the line $s = d$. The dashed curves on the surface represent the drain current for constant $(d - s)$ (i.e. constant V_d); along these curves the current decreases as s increases, until at $s = 1$ the current is zero. The solid curves represent the current for constant s (i.e. constant V_g). These curves, when drawn from a common origin, are the normalized drain characteristics as shown in Fig. 2.7. The dashed curve in Fig. 2.7 gives the current for $d = 1$. It separates the pentode-like characteristics into two regions: below pinch-off ($d \leq 1$), where the current curves are rising, and beyond pinch-off

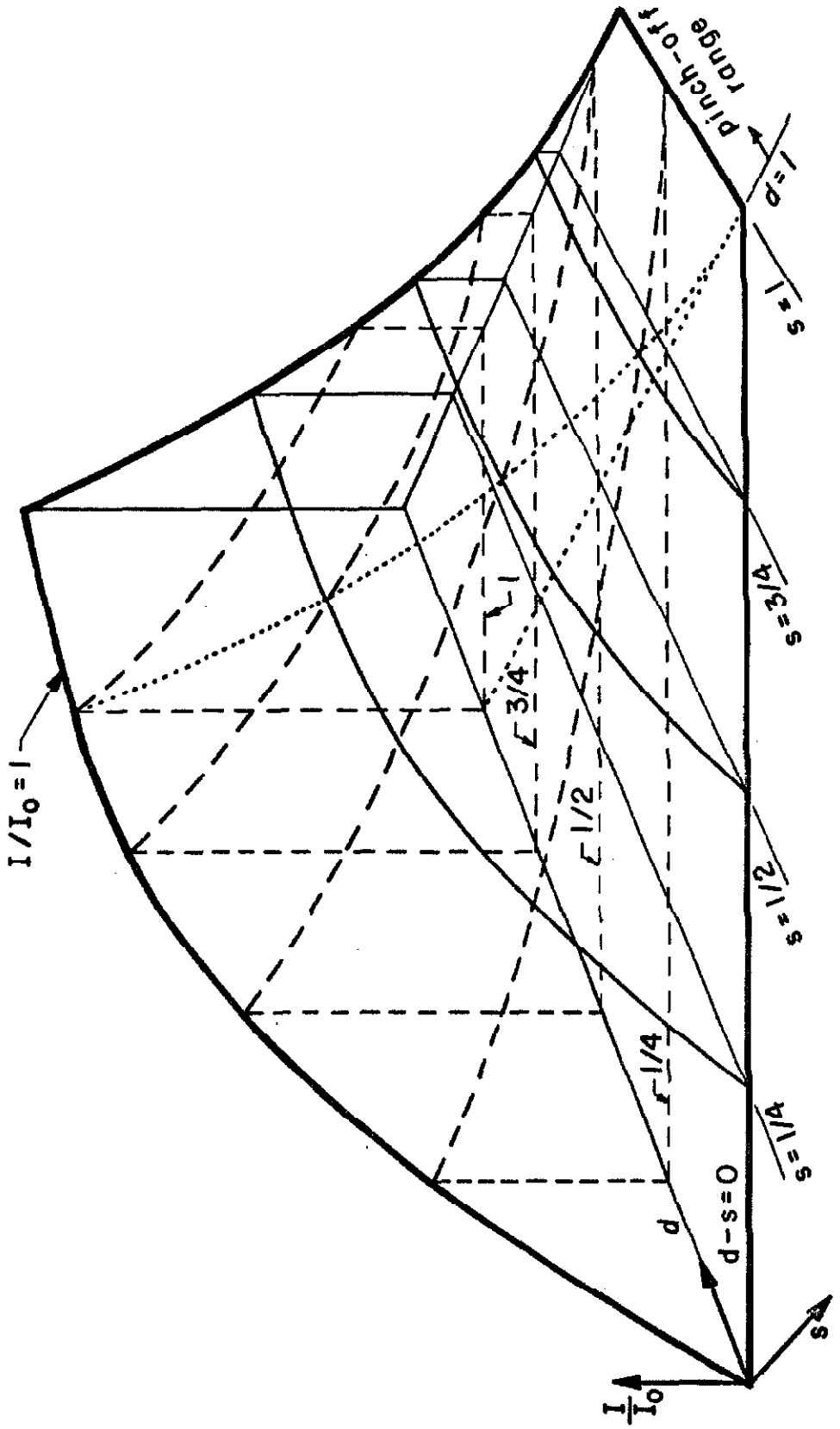


Fig. 2.6. Drain current as a function of the source-gate and the drain-gate potentials.

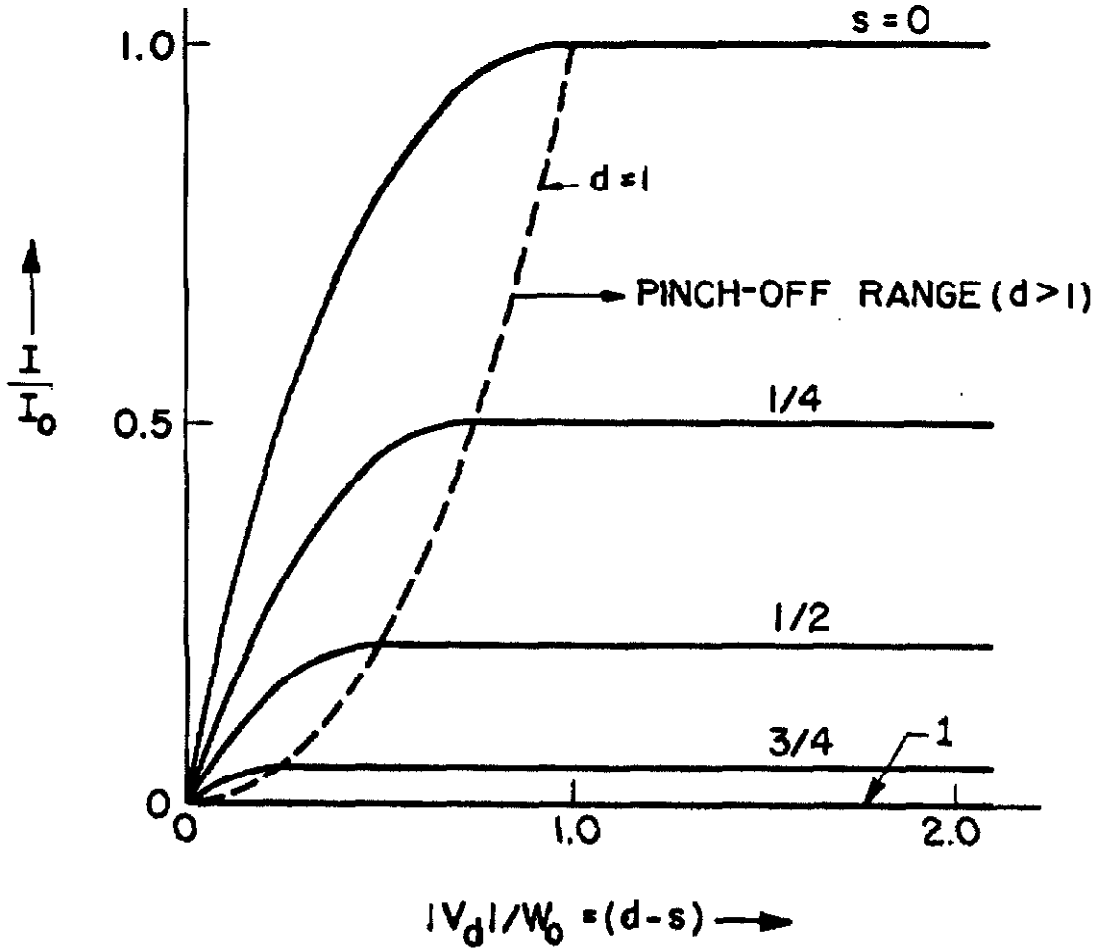


Fig. 2.7. Normalized drain characteristics of a step-junction FET.

$(d > 1)$, where the curves are flat and the current is independent of the drain voltage. (This latter region is sometimes referred to in the literature as the "current-saturation" region.)

2.4.2. Small-Signal Equivalent Circuit

The first-order equivalent circuit beyond pinch-off is similarly obtained from the appropriate equations of Section 2.4:

$$g_m = G_o(1 - \sqrt{s}) \tag{2.35}$$

$$C_{11}^* = C_o^* \frac{1 + \sqrt{s}}{(1 + 2\sqrt{s})^2} \tag{2.36}$$

$$C_{22}^* = G_{22} = 0 \tag{2.37}$$

Equations 2.35 and 2.36, along with the transfer characteristics $I(s,1)$, Eq. 2.34, are drawn in Fig. 2.8 with each quantity normalized to its value at $s = 0$. Figure 2.9 shows the equivalent circuit expressed by Eqs. 2.35 - 2.37. It is interesting to note that this circuit contains only those elements which, according to the charge-control approach [6], are essential to the operation of the device.

An error seemingly is apparent in Eq. 2.36, for that equation gives $C_{11}^* = (2/9)C_o^*$ at $s = 1$, while if we substitute $d = s$ in Eq. 2.30, the original formula for C_{11}^*, and then let $s = 1$, we arrive at $C_{11}^* = (1/3)C_o^*$. Mathematically, this discrepancy arises because the point $s = d = 1$ is a singularity of Eq. 2.30, and hence we may realize different limits upon approaching this point from

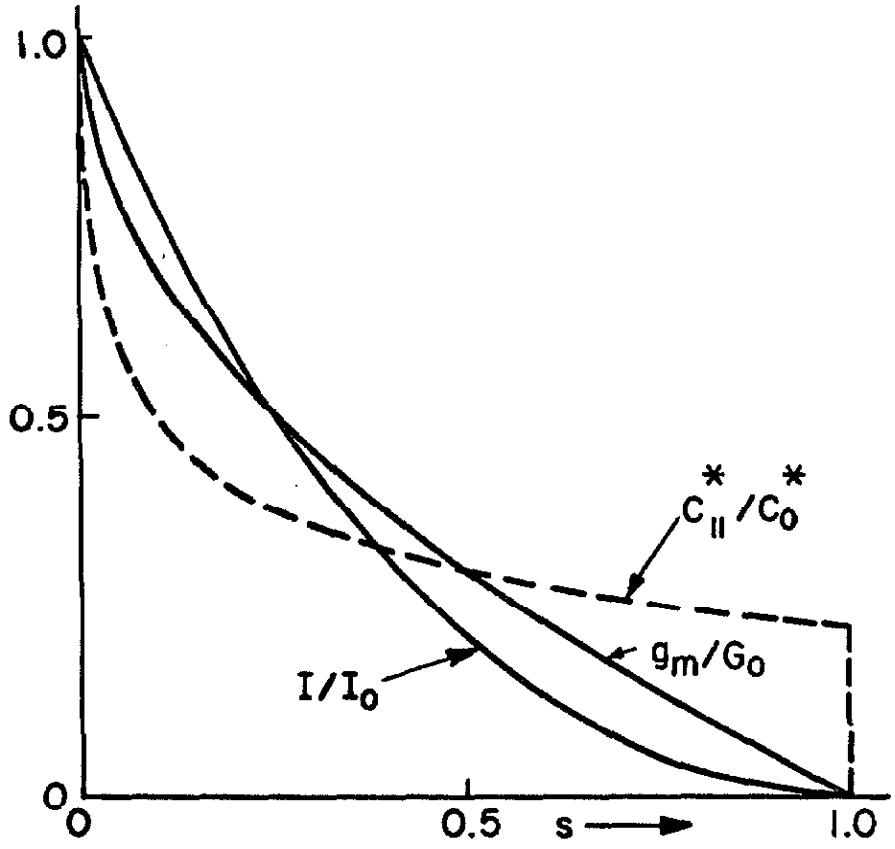


Fig. 2.8. Plots of normalized I, g_m, and C_{11}^* vs. s in the pinch-off range (d > 1).

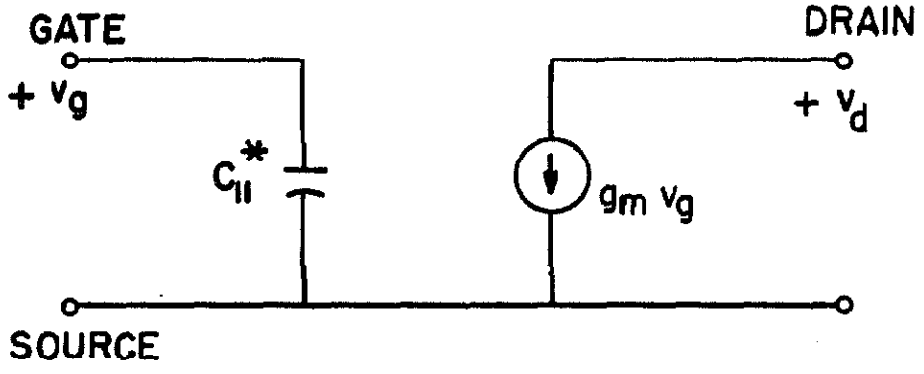


Fig. 2.9. Equivalent circuit beyond pinch-off.

different directions. Physically, we expect unusual behavior because when $s = 1$ the space-charge region occupies all the channel material. Since no free carriers remain to be incorporated in the depletion region, a further increase in s cannot add any more space-charge to the channel, and the input capacitance must vanish for $s > 1$. Both the mathematical and the physical viewpoints may be better comprehended by reference to Fig. 2.10. There we show a drawing of the theoretical surface that represents the input charge-capacitance C_{11}^*/C_o^* as a function of s and d. Although normal device operation implies $s \leq d$, the range $s > d$ is included in the diagram for completeness. Our model is symmetrical, so that $C_{11}^*(s,d) = C_{11}^*(d,s)$; i.e. the surface is symmetrical about the line $s = d$. For $s,d < 1$, $C_{11}^*(s,d)$ is given by Eq. 2.30; for $s \leq 1$, $d \geq 1$, $C_{11}^*(s,d) = C_{11}^*(s,1)$; for $s,d > 1$, $C_{11}^*(s,d) = 0$. The point $s = d = 1$ is a singularity, and, as mentioned above, the value of C_{11}^* at this point depends upon the direction of approach. The dashed curve in Fig. 2.10 represents the direction $s = d$, along which C_{11}^* approaches the value $(1/3)C_o^*$; the curves in the directions $d = 1$ and $s = 1$ approach the value $(2/9)C_o^*$; curves in intermediate directions approach intermediate values. Figure 2.11 shows several cross sections of the surface of Fig. 2.10. The solid curves represent cuts by planes parallel to the d-axis, and the dashed curve is a cut in the direction $s = d$ $(V_d = 0)$.

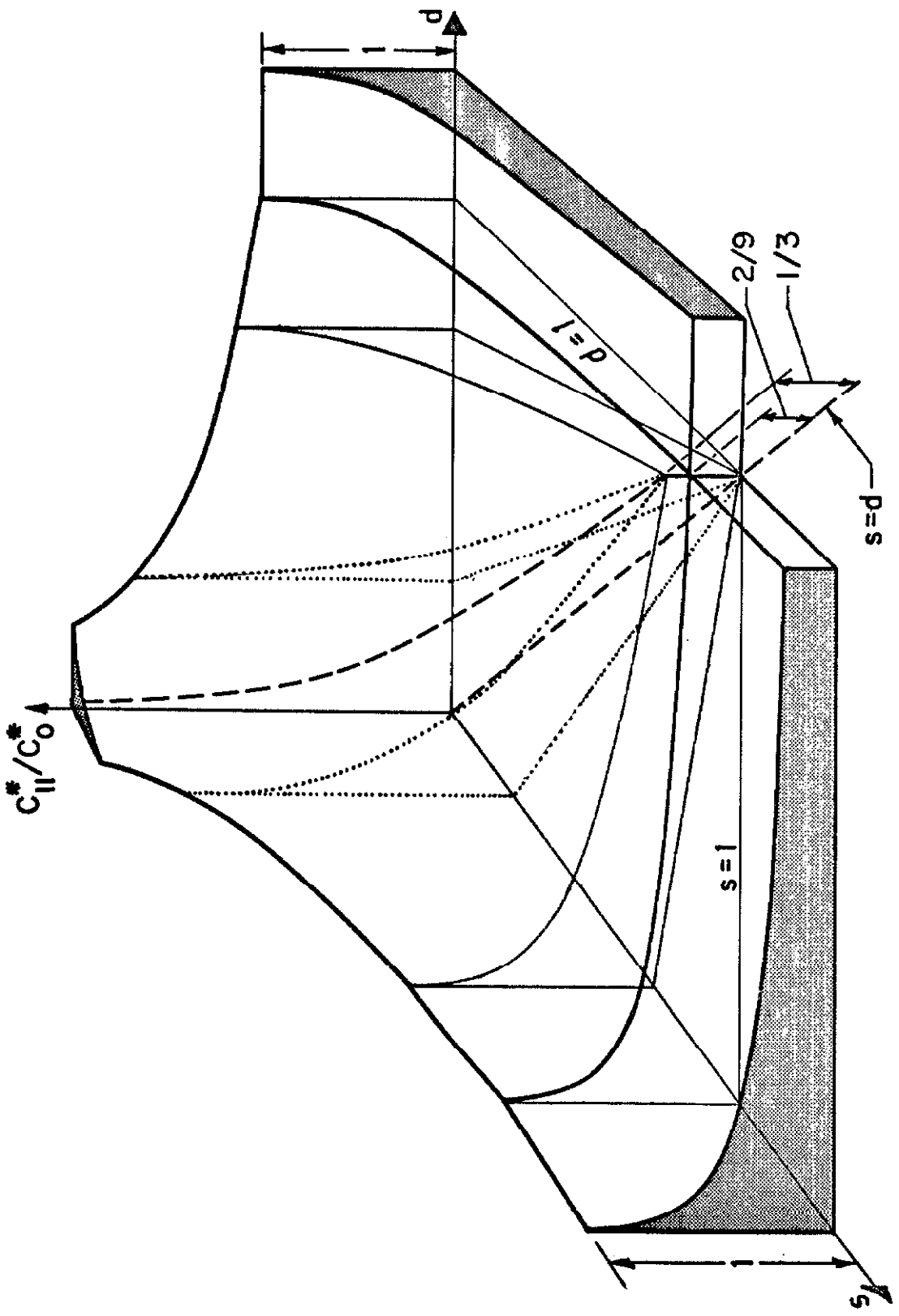


Fig. 2.10. Surface representing the input charge-capacitance.

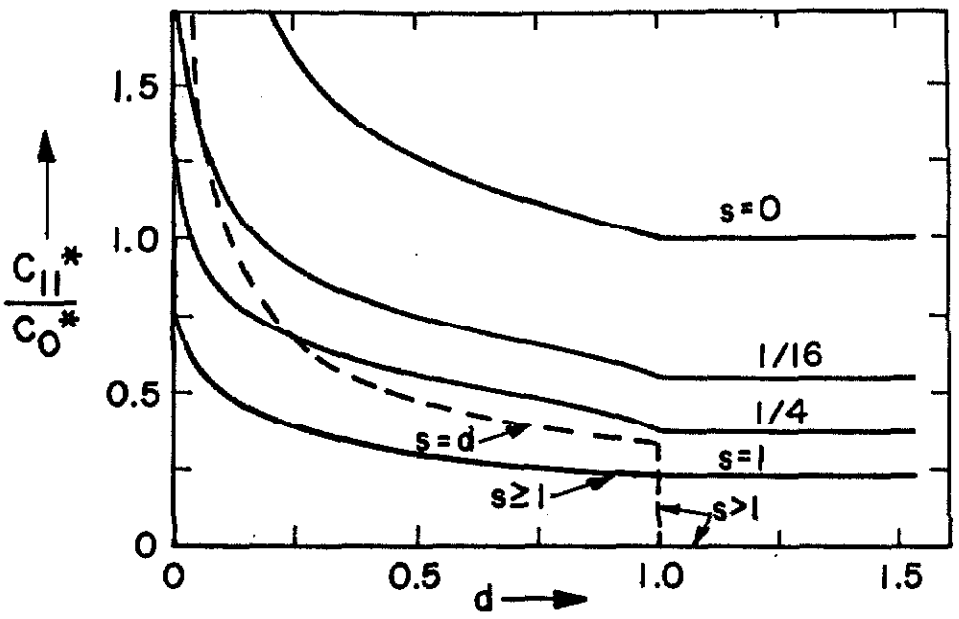


Fig. 2.11. Cross sections of the surface of Fig. 2.10.

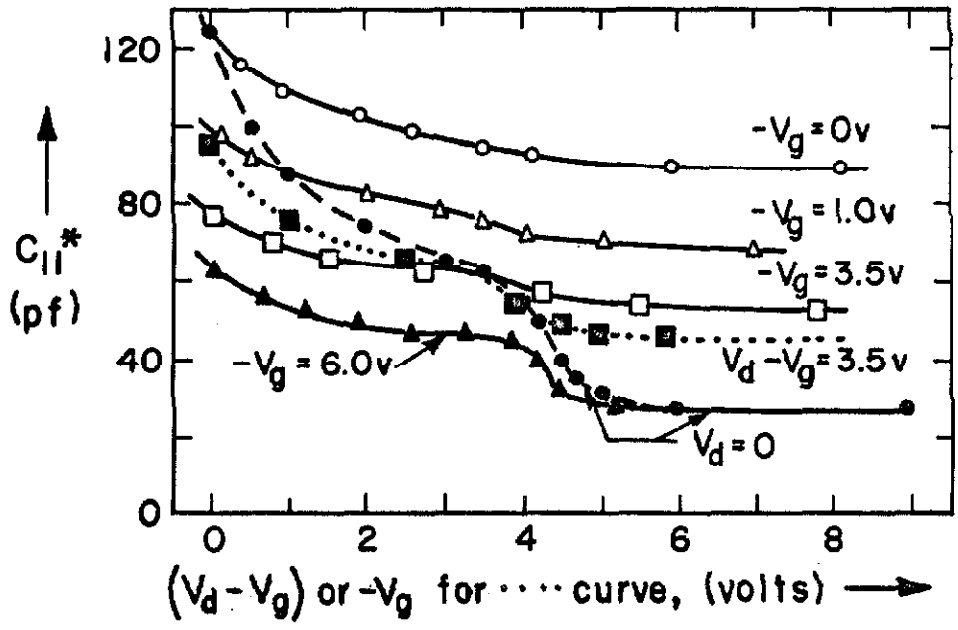


Fig. 2.12. Experimental curves analogous to those of Fig. 2.11.

Most commercial FET's do exhibit noticeable jumps in C_{11}^* when the channel becomes completely pinched-off.* Of course, these jumps will not be infinitely steep, but some units, notably those manufactured by Motorola, do give exceptionally sharp drops. Experimental curves analogous to the theoretical curves of Fig. 2.11 are presented in Fig. 2.12. (Data are from type MM765, epitaxial-junction FET; measuring setup is shown in Appendix B.) The presence of a finite built-in potential means that $V_g = 0$ does not correspond to $s = 0$, and the steep slope in C_{11}^* near $s = d = 0$ is not displayed. From the abrupt drop in the $V_d = 0$ curve we see that the pinch-off voltage is roughly 5 volts. The finite value of capacitance beyond pinch-off ($-V_g > 5v$ with $V_d = 0$ and $V_d - V_g > 5v$ with $V_g = -6.0v$) is due to interelectrode and other spurious effects, and since this capacitance is essentially independent of bias, it may be considered merely as stray capacitance. As expected, the singularity at $s = d = 1$ has been smoothed out. The ratio of the drop in C_{11}^* along the line $V_g = -6.0v$ to the drop along the line $V_d = 0$ theoretically should equal 0.67, but experimentally equals approximately 0.59. The error may be attributed in part to the smoothness of the drop-off and the resulting inaccuracies in computing the jumps.

---

* Strictly speaking, measured input capacitances should be denoted by C_{11} rather than by C_{11}^*. In Chapter VI, however, these two quantities are shown to be identical, so that for clarity, only the symbol C_{11}^* is used here.

The dotted curve represents a cut parallel to the s-axis (C_{11}^* vs. $-V_g$ with $V_d - V_g = 3.5v$) and theoretically should be identical to the corresponding cut parallel to the d-axis (C_{11}^* vs. $V_d - V_g$ with $V_g = -3.5v$). The difference between the two curves arises because the actual FET lacks the symmetry assumed in the model.

2.5. Conclusions

In this chapter we have given a rather complete, albeit terse analysis of the step-junction FET. Although the results for I and g_m are now standard, the solutions for C_{11}^* and C_{22}^* have not been given before. For this reason these quantities were discussed in detail, especially in their relation to the small-signal equivalent circuit of the FET. A not-uncommon pitfall in the calculation of the input capacitance was mentioned, and some of the singular properties of C_{11}^* were discussed, along with experimental verification of the predicted conclusions.

This chapter has provided the necessary background for the treatment of an arbitrarily doped FET. Before we begin this general treatment, however, we shall, in the next chapter, investigate the validity of the assumptions used in the preceding derivations.

CHAPTER III

VALIDITY OF THE ASSUMPTIONS USED IN THE STEP-JUNCTION ANALYSIS

Our primary purpose in this chapter is to determine the usefulness and the limitations of the step-junction analysis by an examination of the assumptions and the approximations of the previous chapter. Despite the simplicity of the step-junction treatment, the theoretical results do agree with the experimental measurements on most units tested (see also published data, e.g. in [7], [8], and [9]). Nevertheless, some FET's that were tested in the present investigation differed qualitatively in g_m from the results of Chapter II. Near gate pinch-off (where $g_m = 0$) the curve of transconductance versus gate voltage theoretically should give a slope that is much shallower than the small-bias slope; some units, however, showed an almost uniform slope. Our secondary purpose in this chapter is to explain this deviation.

The most fundamental hypothesis of the previous chapter was the choice of a model for the FET. In Section 3.1 we show the suitability of the ideal model by comparing it with commercial structures. In that section we also dispense quickly with four of the eight listed assumptions of Chapter II. The remaining four assumptions are treated in the subsequent sections. The procedure given in Section 3.2 shows how the built-in potential may be easily included in previous formulae; in Sections 3.3. and 3.4 we affirm that the constant-mobility assumption and the gradual approximation are

reasonably well justified in practice; by means of an approximate analysis in Section 3.5, we conclude that the boundary on the space-charge region may be considered abrupt.

3.1. Deviations from Idealized Geometry

Although some commercial FET's (e.g. Motorola units [10], [11]) closely approximate the geometry of our model, others (especially diffused types) apparently differ significantly from this ideal. We shall examine one typical case (three-terminal, single-diffused Texas Instruments units; no type number available) and prove that the use of the idealized geometry is an excellent approximation. Most other non-ideal transistors can be treated in a similar manner with a similar conclusion.*

A cross section of the device to be considered is shown in Fig. 3.1. The ratio $A/L \approx 40$, so that three-dimensional effects are unimportant. The source and drain contacts are not negligible in size, but have length $L' \sim L$. This condition raises doubts as to the applicability of our previous results and is therefore discussed below.

---

* FET's with concentric geometries (such as certain Crystalonics units) are described by equations with functional dependences identical to the dependences in the corresponding planar-geometry equations. The exact formulae for the concentric devices may be obtained from the planar formulae with the replacements $(L/A) \rightarrow (1/2\pi)\ln(r_2/r_1)$ and $(LA) \rightarrow \pi(r_2^2 - r_1^2)$, in which r_1 and r_2 are respectively the inner and outer radii of the channel (corresponding to $x = 0$ and $x = L$).

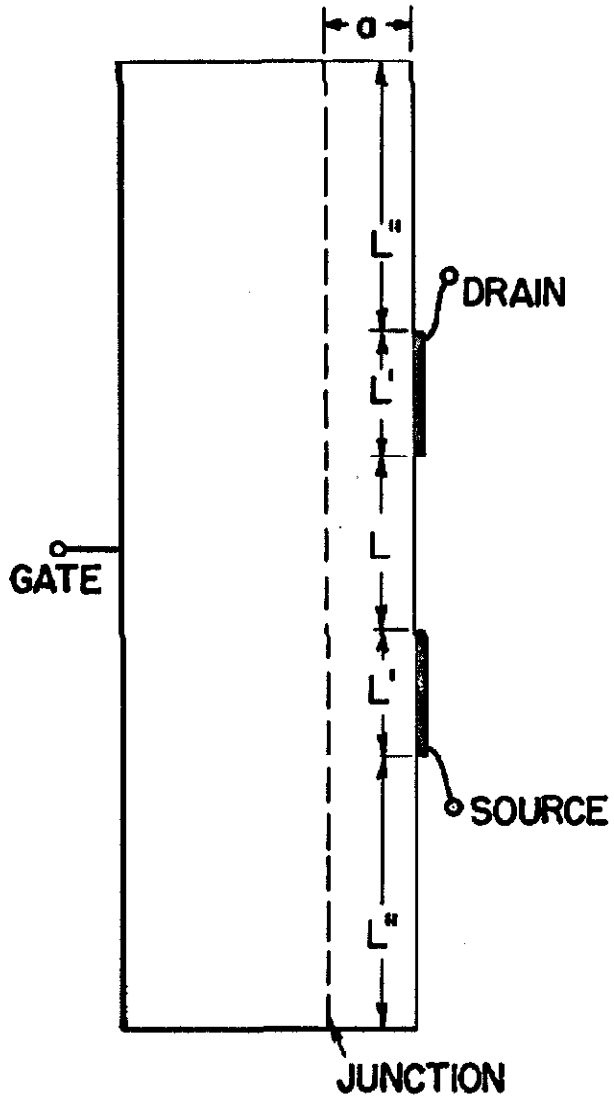


Fig. 3.1. Cross section of practical FET with non-ideal geometry.

For simplicity we consider the channel material to be a uniformly doped, unit-conductivity semiconductor. With no bias on the gate or drain, the ideal model predicts a source-drain resistance

$$R_{sd} = \frac{1}{A}\frac{L}{a}$$

We now calculate the actual source-drain resistance R'_{sd} and show that $R'_{sd} \approx R_{sd}$.

Since the distances L'' are of the order of a few L', we may, with no loss in accuracy, assume that $L'' \to \infty$. Now, if C'_{sd} is the capacitance between source and drain, then we have [12],

$$R'_{sd} = \frac{1}{C'_{sd}}$$

assuming, for convenience, that $\epsilon = 1$. The solution to the capacitance problem has been worked out in detail [13] with the result

$$C'_{sd} = \frac{A}{2}\frac{K(k')}{K(k)} \qquad (3.1)$$

where K is the complete elliptic integral, and where

$$k = \frac{\tanh\frac{\pi L}{4a}}{\tanh\left\{\frac{\pi L'}{2a} + \frac{\pi L}{4a}\right\}}, \qquad k'^2 = 1 - k^2$$

A useful approximation to Eq. 3.1 may be obtained for $L/a \gg 1$, since then

$$k \approx 1$$

$$k' \approx 2e^{-\pi L/4a} \sqrt{\frac{2 \tanh (\pi L'/2a)}{1 + \tanh (\pi L'/2a)}}$$

and [14]

$$K(k) \approx \frac{\pi}{2}$$

$$K(k') \approx \ln (4/k') \approx \ln 2 + \frac{\pi L}{4a}$$

for $L' \gtrsim a$. Thus,

$$C'_{sd} \approx \frac{A}{2} \frac{a}{(L/2) + 0.44a}$$

Although derived for $L/a \gg 1$, this approximation is within 1% even for $L/a = 2$ [15]. The desired result is then*

$$R'_{sd} \approx \frac{1}{A}\left[\frac{L}{a} + 0.88\right] \approx R_{sd} \qquad (3.2)$$

since $L/a > 10$ in practice.

As mentioned in the introduction, several of the listed assumptions of Chapter II may be dismissed briefly. The hypothesis of a step-junction doping profile will be removed in the next chapter.

---

* Equation 3.2 also obtains if $L' \to \infty$.

In practice, the direct gate current is several orders of magnitude less than the direct drain current,* and the resistivities and contact fabrication are such that the drain current is carried completely by majority carriers.** The empirical and heuristic justification for the assumption that extended the below-pinch-off formulae of Section 2.3 to voltages above pinch-off has already been given in Section 2.4. The remaining assumptions are now examined in detail.

3.2. The Built-In Potential

By means of a simple redefinition of s and d, the built-in potential may be incorporated into the previous formulae, and the results effectively remain unchanged. We shall treat this discussion on a general basis since the results are directly applicable to all doping profiles.

First, it is necessary to distinguish between applied or "external" potentials and total or "internal" potentials. The latter quantities include the built-in potential. Unless otherwise specified, henceforward the symbol V (with various subscripts) signifies an external voltage, and W signifies the magnitude of an internal voltage. We define the built-in potential to be V_c, and the applied gate voltage necessary for pinch-off to be V_p. Thus, the magnitude of the internal pinch-off voltage is

---

* The data sheet of any manufacturer may be consulted for proof of this fact.

** Qualitative discussion on this point is given in [8].

$$W_o = |V_c + V_p| \tag{3.3}$$

the magnitude of the internal gate voltage is

$$W_g = |V_c + V_g| \tag{3.4}$$

and the magnitude of the internal drain potential is simply

$$W_d = |V_d| \tag{3.5}$$

because V_d is measured from drain to source. For reverse bias on the gate-channel junction, V_c and V_g add directly, regardless of the type of channel. The presence of a finite built-in potential permits the application of V_g of sign opposite to that of V_c. These forward-bias values of V_g must be small enough to ensure that the gate current remains negligible.

All our theoretical expressions are written in terms of the internal quantitites s and d. Detailed transformation into expressions involving external voltages is unnecessary, for we observe that

$$s = \frac{W_g}{W_o} \tag{3.6}$$

$$= \frac{V_c}{V_c + V_p} + \left(\frac{1}{V_c + V_p} \right) V_g$$

$$d = \frac{W_d + W_g}{W_o} \tag{3.7}$$

$$= \frac{V_c}{V_c + V_p} + \left(\frac{1}{V_c + V_p}\right) V_g - \left(\frac{1}{V_c + V_p}\right) V_d$$

The above equations include the built-in potential and are more general than Eqs. 2.12 and 2.13. Therefore, Eqs. 3.6 and 3.7 replace these earlier definitions of s and d. Because s, for example, is linearly related to the applied voltage V_g, any curve plotted as a function of s will have the same shape if it is plotted as a function of V_g. In order to compare theory and experiment we simply add V_c to V_g, and then normalize this sum with respect to $(V_c + V_p)$ to obtain s. Excess gate current will flow if $V_g = -V_c$, so that experimental points will only lie within some fraction of the theoretical range $0 \leq s \leq 1$. Small errors in V_c are critical when the slope of the quantity of interest is steep near $s = 0$, as, for example, in the case of $C_{11}^*(s,1)$. Remarks similar to those given above apply to the internal gate-drain potential d.

3.3. Non-Constant Mobility

The hypothesis of constant carrier mobility μ will break down if the channel field is higher than the so-called critical field E_c [16]. Beyond this field the mobility first varies as $E^{-1/2}$, and then, in the velocity-limited range, varies as E^{-1}. Dacey and Ross [8] have computed I and g_m for a step-junction FET under the assumption that over the whole channel length the electric field is greater than E_c, but less than the limiting-velocity field. That is,

they assume that $\mu = \mu_c (E_c/E)^{1/2}$, in which μ_o is the low-field mobility. In the pinch-off range their results are

$$I = I_c \left(\frac{3}{2} \frac{E_c}{W_o/L} \right)^{1/2} \left[4(1 - \sqrt{s})^3 - 3(1 - \sqrt{s})^4 \right]^{1/2} \qquad (3.8a)$$

$$g_m = G_o \left(\frac{3}{2} \frac{E_c}{W_o/L} \right)^{1/2} \left[\frac{1 - \sqrt{s}}{1 + 3\sqrt{s}} \right]^{1/2} \qquad (3.8b)$$

I_o being the zero-bias drain current for the constant mobility case. The analysis leading to these equations is valid only if $W_o/L \gg E_c$, i.e. if the average channel field is considerably greater than critical. As expected, the zero-bias current and transconductance are reduced because of the lowered mobility. Aside from the normalization, Eq. 3.8a does not differ appreciably from the constant-mobility transfer characteristics 2.34; the transconductance, given by Eq. 3.8b, does differ, however, by having a slightly concave-downward shape near $s = 1$.

The above results may be applicable to some experimental FET's and, perhaps, to a few commercial models, but the parameters of most commercial FET's preclude the possibility of non-constant mobility being significant at normal operating temperatures. For example, Crystalonics type C610 and Motorola type MM764 both have n-type silicon channels, for which the critical field is 2500v/cm [16]. However, the average channel field for each of these units is roughly 500v/cm, so that Eqs. 3.8 do not apply.

3.4. Validity of the Gradual Approximation

The gradual approximation, which formed the basis of our derivations, is described by

$$|E_y| \ll |E_x|, \qquad 0 \le y \le b \tag{3.9}$$

$$\left|\frac{\partial E_x}{\partial x}\right| \ll \left|\frac{\rho}{\epsilon}\right|, \qquad b \le y \le c \tag{3.10}$$

in which $\rho = \rho_o$ and $c = a$ for the step-junction device under consideration. We now show, with actual calculations, that even under worst-case conditions these inequalities are satisfied for roughly 90% of the channel length.

First consider Eq. 3.9. When a finite drain current flows we know that the channel becomes progressively narrower in the direction of increasing x. That is, the channel boundary $b(x)$ has a negative slope everywhere. Since the channel boundary delimits the current flow, we must have $E_y \neq 0$ at $y = b$, so that, by continuity, $E_y \neq 0$ for y near b. But $E_y = 0$ at $y = 0$, and hence the maximum value of E_y within the channel occurs at $y = b$. Furthermore, the slope of the channel boundary, db/dx, gives the direction of the field at the boundary. These arguments enable us to assert that inequality 3.9 will be fulfilled if

$$\left|\frac{E_y}{E_x}\right|_{y=b} = \left|\frac{db}{dx}\right| \ll 1 \tag{3.9a}$$

is fulfilled. From Eq. 2.20 we get

$$\left|\frac{db}{dx}\right| = \frac{d(3 - 2\sqrt{d}) - s(3 - 2\sqrt{s})}{6(L/a)\, u(1 - u)} \tag{3.11}$$

Therefore, to satisfy 3.9 we need

$$u(1 - u) \gg \frac{d(3 - 2\sqrt{d}) - s(3 - 2\sqrt{s})}{6(L/a)} \equiv h \tag{3.12}$$

The right-hand side of this inequality, h, is small because the numerator is at most unity, and L/a is much greater than one.

The arguments for Eq. 3.10 run similarly. Inside the space-charge region $\partial E_x/\partial x \neq 0$. On the boundary $y = a$, however, $\partial E_x/\partial x = 0$, so that, as with 3.9, the worst violation of 3.10 occurs at $y = b$. Equation 2.8 yields $\partial E_x/\partial x$ in the space-charge region:

$$\left|\frac{\partial E_x}{\partial x}\right| = \left|\frac{\partial^2 V}{\partial x^2}\right| = 2W_o\left(1 - \frac{y}{a}\right)\left|\frac{d^2u}{dx^2}\right|$$

We evaluate d^2u/dx^2 from 3.11 to get

$$\frac{u^3(1 - u)^2}{|2u - 1|} \gg \left[\frac{d(3 - 2\sqrt{d}) - s(3 - 2\sqrt{s})}{6(L/a)}\right]^2 = h^2 \tag{3.13}$$

as the condition necessary to satisfy inequality 3.10.

Inequalities 3.12 and 3.13 are both violated near $u = 1$ and near $u = 0$. The failure of the gradual approximation near $u = 1$ does not introduce any error into the analysis because of the

relatively large channel conductance in that region. The error near $u = 0$ represents the significant breakdown of the gradual approximation. For definiteness, we now consider ">> 1" to mean "$\geq H$," where H is some number greater than one. Then (disregarding the violation near $u = 1$), the values of u for which the gradual approximation holds are given by the narrower of the two pairs of bounds

$$\frac{1}{2} - \frac{1}{2}\sqrt{1 - 4hH} \leq u \leq 1 \tag{3.14a}$$

and

$$(h^2H)^{1/3}\left[1 - \frac{1}{3}(h^2H)^{2/3} - \frac{2}{3}(h^2H)\right] \leq u \leq 1 \tag{3.14b}$$

The lower bound in 3.14a is an exact solution of 3.12, while the lower bound in 3.14b is an approximate solution of 3.13 for $h^2H \ll 1$. The relative magnitudes of h and H determine which condition is stronger.

We now consider the worst-case condition, $d = 1$, and by means of Eq. 2.20 for x/L convert Eqs. 3.14 into inequalities for x/L:

$$0 \leq \frac{x}{L} \leq 1 - \frac{1}{4}\frac{(1 - \sqrt{1 - 4hH})^2(2 + \sqrt{1 - 4hH})}{(1 - \sqrt{s})^2(1 + 2\sqrt{s})} \tag{3.15a}$$

$$0 \leq \frac{x}{L} \leq 1 - \frac{(h^2H)^{2/3}[3 - 2(h^2H)^{1/3}]}{1 - s(3 - 2\sqrt{s})} \qquad (3.15b)$$

The maximum values of x/L such that $H \geq 10$ (i.e. such that the right-hand and left-hand sides of each of the gradual-approximation inequalities 3.9 and 3.10 differ by at least an order of magnitude) are given for various values of s and L/a in the following table:

s	L/a	$(x/L)_{max}$	
0	10	0.89	Eq. 3.15a
	20	0.98	
1/4	10	0.95	
	20	0.98	
1/2	10	0.97	Eq. 3.15b
	20	0.99	

We see from this table that the gradual approximation is valid for at least 89% of the channel length, and is usually valid over a much wider range. The maximum values of x/L for $L/a = 10$ and for various values of s and d are illustrated on the channel curves drawn in Fig. 2.3.

Furthermore, with non-uniform channel doping, as is present, for example, in single- or double-diffused structures, the gradual approximation will be more accurate than with uniform doping, because in the former case the channel conductivity is (in general) larger at $y = 0$ than at $y = b$, and hence the current flow tends to be more one-dimensional.

3.5. Non-Abrupt Boundary on the Space-Charge Region

The only assumption remaining from Chapter II is that of the abrupt boundary between the space-charge region and the neutral channel. Actually, the transition from neutrality to almost complete free carrier depletion occurs in a length on the order of the Debye length L_D of the channel material [17]. Thus, we expect the abrupt-boundary assumption to break down when the channel width a is comparable with L_D. The ratio of these two quantities is

$$\frac{L_D}{a} = \frac{1}{a}\sqrt{\frac{\epsilon kT}{q\rho_o}} = \sqrt{\frac{(kT/q)}{\rho_o a^2/\epsilon}} = \sqrt{\frac{W_t}{2W_o}} \tag{3.16}$$

where $W_t = kT/q$, k is Boltzmann's constant, T is the absolute temperature, and q is the electronic charge.* For FET's with small pinch-off voltages this ratio may be large enough to necessitate the introduction of significant correction terms in the previous formulae. We shall attempt to find a second-order correction to the first-order theory. Since the problem involves second-order effects, we shall feel justified in making several broad simplifying assumptions in the following analysis.

The method we shall adopt is to assume that some mobile carriers remain on the depletion-region side of the first-order sharp boundary of the channel. We consider $a > 1$, and we divide the channel into two regions, $0 \le x \le L_1$ and $L_1 < x \le L$. In the first region

---

* The last equality in Eq. 3.16 is valid only for step-junction FET's that have very high gate conductivities, i.e. only when $W_o = \rho_o a^2/2\epsilon$.

$|V(x,0)| \leq W_o$, so that $b \geq 0$ and the conduction is due mostly to carriers in the neutral channel $0 \leq y \leq b$, but partly to carriers within $b < y < a$; the second region has $|V(x,0)| > W_o$, and hence there is no neutral channel, the conduction being entirely due to the small number of free carriers in the (partial) depletion region $0 \leq y < a$. We solve each region for the current and then eliminate L_1 from the two resulting equations. The analysis predicts normalized transconductances that are slightly larger than the first-order values for the same s. The amount of the correction increases with the ratio W_t/W_o. When compared with experimental results, we find that the theory provides the proper qualitative corrections in g_m to account for the errors in the first-order theory, but that the magnitude of the corrections is somewhat large.

We begin the analysis by assuming that the previously derived expression 2.9 for the channel boundary $b(x)$ is valid. Now, however, instead of assuming that the mobile-charge density is (in magnitude)

$$|\rho_m| = \begin{cases} \rho_o, & 0 \leq y < b \\ 0, & b < y \leq a \end{cases}$$

we assume the distribution*

---

* In this section we shall assume that the built-in potential is included in the potentials denoted by V.

$$|\rho_m| = \begin{cases} \rho_o\,, & 0 \le y \le b \\ \rho_o \exp[-|V(x,y) - V(x,b)|/W_t], & b < y \le a \end{cases} \tag{3.17}$$

The last expression represents a Boltzmann falloff of free carriers within the depletion region and is the second-order correction to the first-order abrupt falloff. See Fig. 3.2 for a sketch of the density profile 3.17.

We shall consider operation beyond pinch-off ($d \ge 1$) because the non-abruptness of the channel boundary has the greatest effect in that range. The condition $d > 1$ means that some interval of the channel material, say $L_1 < x \le L$, will have, in the $y = 0$ plane, a potential with respect to the gate greater than W_o. In this interval, then, $b = 0$, and space charge is present over the entire width $0 \le y \le a$. Furthermore, the gate potential $|V_g| = W_o$, while previously resulting in $I = 0$, now permits a finite drain current to flow. (The situation will be illustrated in Fig. 3.2.) The two-dimensional density of free carriers in a differential element of channel material in the interval $0 \le x \le L_1$ is obtained by integrating Eq. 3.17 from $y = 0$ to $y = a$. The resistance dR of this element is then $(A\,dx)/\mu$ divided by this integral, and the current is

$$I = \mu A \frac{dV(x,b)}{dx} \left[\int_0^b \rho_o \, dy + \int_b^a \rho_o \exp\{-|V(x,y)-V(x,b)|/W_t\} \, dy \right]$$

The first integral in the square brackets is the first-order contribution; the second is the additional term that results from the mobile carriers in the depletion region. By employing Eq. 2.8 for $V(x,y)$ we obtain

$$I = \mu A \rho_o \frac{dV}{dx} \left[b + \int_b^a \exp\{- \frac{W_o}{W_t}(\frac{y}{a} - u)^2\} \, dy \right]$$

$$= G_o L \frac{dV}{dx} \left[u + \phi \frac{\sqrt{\pi}}{2} \operatorname{erf}(\frac{1 - u}{\phi}) \right] \tag{3.18}$$

where we have defined[*]

$$\phi^2 \equiv \frac{2\epsilon W_t}{\rho_o a^2} = \frac{W_t}{W_o} \tag{3.19}$$

and where

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \, dt$$

is the standard definition of the error function. Using Eq. 2.14 for $V(x)$, we may integrate Eq. 3.18 over the range $0 \le x \le L_1$, $0 \le V \le W_o(1 - s)$ to get

---

* Cf. footnote page 57.

$$I = I_c \frac{L}{L_1} \left[(1 - \sqrt{s})^2 (1 + 2\sqrt{s}) + \frac{3\sqrt{\pi}}{2} \phi^3 \int_{\sqrt{s}/\phi}^{1/\phi} 2t \operatorname{erf} t \, dt \right] \qquad (3.20)$$

where I_o is the zero-bias current for the first-order abrupt-boundary case and is slightly smaller than the new zero-bias current

$$I_o \left[1 + \frac{3\sqrt{\pi}}{2} \phi^3 \int_0^{1/\phi} 2t \operatorname{erf} t \, dt \right]$$

The integral appearing in 3.20 may be solved,

$$\int_0^x 2t \operatorname{erf} t \, dt = \left(x^2 - \frac{1}{2} \right) \operatorname{erf} x + \frac{1}{\sqrt{\pi}} x e^{-x^2} \qquad (3.21)$$

but we shall retain the unintegrated expression for convenience.

We now determine another expression relating I and L_1 by considering the space-charge region defined by $L_1 < x \leq L$. In a one-dimensional treatment of the space-charge region the maximum potential that can appear across the width a is $W_o = \rho_o a^2 / 2\epsilon$. But at $y = 0$, for every $x > L_1$ the potential with respect to the gate is greater than W_o. In order to maintain a one-dimensional problem while still utilizing our simple model, we shall assume that the y-dependence of the potential in the space-charge region is unchanged, but that ρ_o is replaced by the x-dependent density ρ^*. Thus, for $L_1 < x \leq L$ the potential obeys

$$V(x,y) = V_g + \frac{\rho^*(x)}{2\epsilon} (a^2 - y^2)$$

In the interior of the channel in the first-order theory we always had $|V - V_g| < W_o$ and $b > 0$, so that no quantity analogous to ρ^* was needed. The present second-order treatment necessitates the consideration of two-dimensional effects, and ρ^* is introduced to account for these effects. Along the plane $y = 0$ (which, in the interval $L_1 < x \leq L$, corresponds to the channel boundary $y = b$ in the interval $0 \leq x \leq L_1$) the potential is

$$V(x,0) = V_g + \frac{\rho^*(x)a^2}{2\epsilon}$$

This equation is the analog of Eq. 2.9 for the potential in the neutral channel. We note that $V(L_1,0) \equiv V_g + W_o$ and that $V(L,0) = V_d$, so that $\rho^*(L_1) = \rho_o$ and $\rho^*(L) = \rho_o(V_d - V_g)/W_o = \rho_o\hat{d}$. It is the x-directed electric field resulting from $V(x,0)$ that enters into the current-flow equation.

As in the treatment for $0 \leq x \leq L_1$, we still assume a Boltzmann falloff of the free-carrier density in the y-direction, but now this falloff is from $y = 0$ rather than from $y = b$. Also, as we increase x from L_1, along the plane $y = 0$ say, we expect the free-carrier density to decrease because we are traveling away from the neutral channel. Our final assumption is that the free-carrier density experiences a Boltzmann falloff in the x-direction for $x > L_1$. Thus, in the interval $L_1 < x \leq L$ we have

$$|\rho_m| = \rho_o \exp[-|V(x,0) - V(L_1,0)|/W_t] \exp[-|V(x,y) - V(x,0)|/W_t] \tag{3.22}$$

The complete mobile-carrier density is depicted in Fig. 3.2. For $0 \le x \le L_1$, $|\rho_m|$ is given by Eq. 3.17; for $L_1 < x \le L$, it is given by Eq. 3.22. In the first-order theory $L_1 = L$, and $|\rho_m|$ is a sharp-cornered, wedge-shaped surface bounded by the surface $y = b(x)$.

Some of the above hypotheses may seem unreasonable, but, as mentioned at the beginning of this section, we are being guided by the fact that our goal is merely a correction to the almost-satisfactory first-order theory. Further, although the hypotheses themselves may seem unreasonable, the resulting mobile-carrier distribution pictured in Fig. 3.2 is quite plausible.

Combining the preceding remarks and equations for the region $L_1 < x \le L$, we obtain the drain current in the following straightforward steps:

$$I = \mu A \frac{dV(x,0)}{dx} \int_0^a |\rho_m| \, dy$$

$$= \mu A \left(\frac{a^2}{2\epsilon} \frac{d\rho^*}{dx} \right) \rho_o \exp\left[- \frac{(\rho^* - \rho_o)a^2}{2\epsilon W_t} \right] \int_0^a \exp\left[\frac{\sigma^* y^2}{2\epsilon W_t} \right] dy$$

$$= \mu a A W_o \phi \sqrt{\frac{\rho_c}{\rho^*}} \exp\left[- \frac{(\rho^* - \rho_o)}{\rho_o \phi^2} \right] \frac{\sqrt{\pi}}{2} \operatorname{erf}\left[\frac{1}{\phi} \sqrt{\frac{\rho^*}{\rho_o}} \right] \frac{d\rho^*}{dx}$$

$$\int_{L_1}^{L} I dx = 3LI_o \phi \frac{\sqrt{\pi}}{2} \exp\left(\frac{1}{\phi^2} \right) \int_1^d \sqrt{\frac{\rho_o}{\rho^*}} \exp\left(- \frac{\sigma^*}{\rho_o \phi^2} \right) \operatorname{erf}\left(\frac{1}{\phi} \sqrt{\frac{\rho^*}{\rho_o}} \right) d\left(\frac{\rho^*}{\rho_o} \right) \tag{3.23}$$

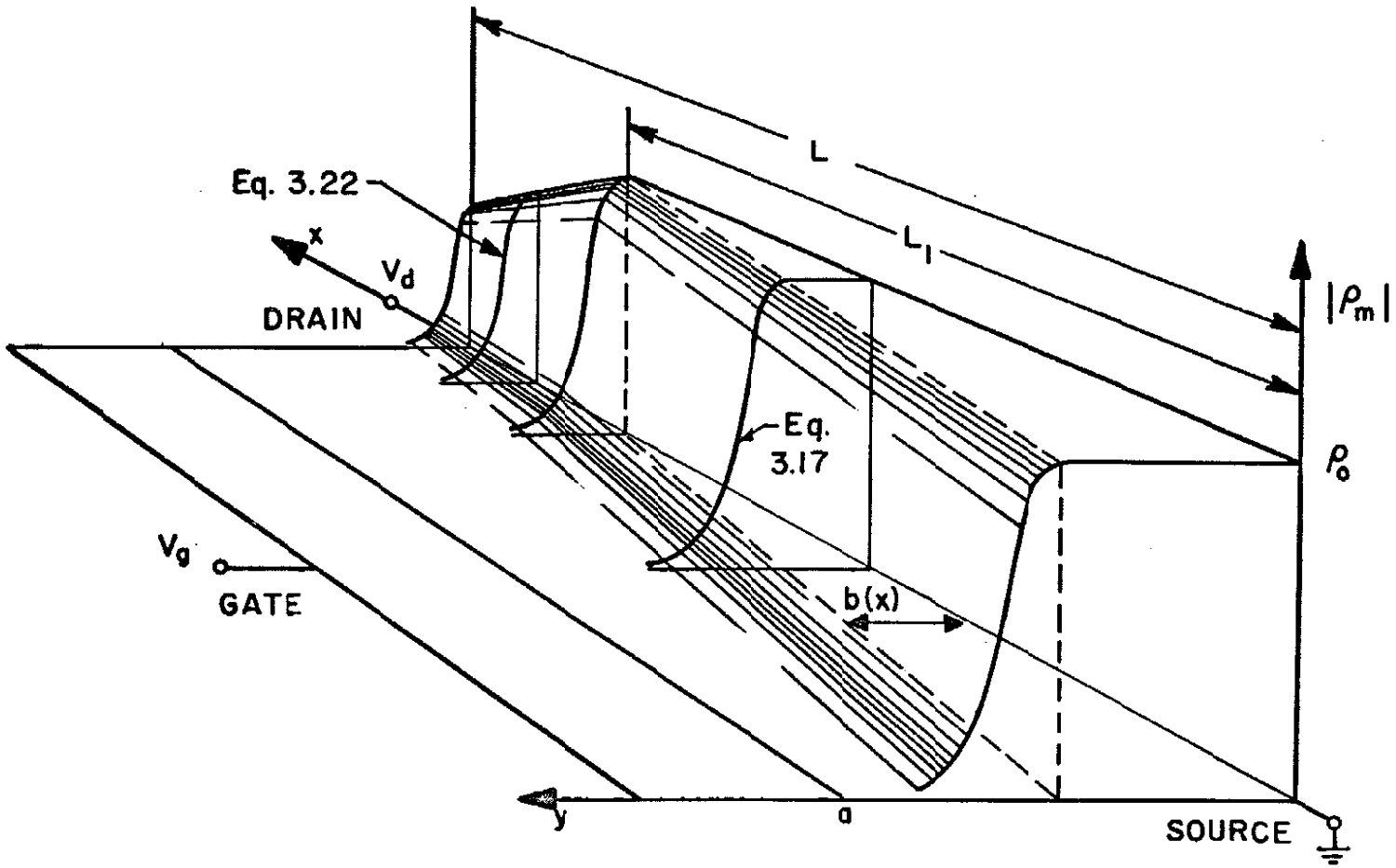


Fig. 3.2. Sketch of the mobile-carrier density in the channel material for a non-abrupt space-charge-region boundary. Incomplete depletion is present in the range $b(x) < y \leq a$. The dashed curve outlines the bounding surface of $|\rho_m|$ in the first-order theory (in which $L_1 = L$).

$$I = 3I_o \frac{L}{L - L_1} \left[d\eta^2(d) - \eta^2(1) \right] \exp\left(\frac{1}{\phi^2}\right) \qquad (3.24)$$

In Eq. 3.24 we have written

$$\eta(x) = \frac{\sqrt{\pi}}{2} \frac{\operatorname{erf}(\sqrt{x}/\phi)}{(\sqrt{x}/\phi)}$$

The function $\eta(x)$ depends parametrically on ϕ. In particular, $\eta \to C$ as $\phi \to 0$.

We now eliminate L_1 from Eqs. 3.20 and 3.24 to get

$$I = I_o \left\{ (1 - \sqrt{s})^2(1 + 2\sqrt{s}) + \frac{3\sqrt{\pi}}{2} \phi^3 \int_{\sqrt{s}/\phi}^{1/\phi} 2t \operatorname{erf} t \, dt + \right.$$

$$\left. + 3\left[d\eta^2(d) - \eta^2(1) \right] \exp\left(\frac{1}{\phi^2}\right) \right\} \qquad (3.25a)$$

The first term in the curved brackets is the first-order current; the remaining terms are small correction factors.

Equation 3.25a is valid only for $s \leq 1$. For $s > 1$ the extension is clear: Equation 3.18 gives no contribution, and we merely adjust 3.24 by integrating 3.23 from $x = 0$, $\rho^*/\rho_o = s$ to $x = L$, $\rho^*/\rho_o = d$. This gives

$$I = 3I_o \left[d\eta^2(d) - s\eta^2(s) \right] \exp\left(\frac{1}{\phi^2}\right) \qquad (3.25b)$$

Bearing in mind that d is a function of V_g, we obtain the

transconductance by differentiating Eq. 3.25 with respect to V_g:

$$g_m = \begin{cases} G_o \left\{ (1 - \sqrt{s}) + \sqrt{s}\eta(s) - \eta(d) \exp\left[- \frac{(d - 1)}{\phi^2} \right] \right\}, & s \leq 1 \qquad (3.26a) \\ G_o \left\{ \eta(s) \exp\left[- \frac{(s - 1)}{\phi^2} \right] - \eta(d) \exp\left[- \frac{(d - 1)}{\phi^2} \right] \right\}, & s > 1 \qquad (3.26b) \end{cases}$$

In silicon the built-in potential V_c is of the order of 0.5v. Since $W_o > V_c$ and since $W_t \approx 0.025v$ at room temperature, it suffices to consider $\phi < 0.2$. Then $\eta(x) \approx (\sqrt{\pi}/2)(\phi/\sqrt{x})$ for $x > 0.25$, and if we restrict our discussion to values $V_d > 1.2\ W_o$, Eqs. 3.25 and 3.26 may be simplified to

$$I \approx I_o \left\{ (1 - \sqrt{s})^2(1 + 2\sqrt{s}) + 3 \left[\left(1 + \frac{\phi^2}{2}\right)\eta(1) - \left(s - \frac{\phi^2}{2}\right)\sqrt{s}\eta(s) - \frac{\phi^2}{2}\sqrt{s} \exp\left(- \frac{s}{\phi^2}\right) \right] \right\}, \quad s \leq 1 \qquad (3.27a)$$

$$I \approx 3I_o \phi^2 \eta(s) \exp\left[- \frac{(s - 1)}{\phi^2} \right], \quad s > 1 \qquad (3.27b)$$

$$g_m \approx \begin{cases} G_o \left[(1 - \sqrt{s}) + \sqrt{s}\eta(s) \right], & s \leq 1 \qquad (3.28a) \\ G_o\, \eta(s) \exp\left[- \frac{(s - 1)}{\phi^2} \right], & s > 1 \qquad (3.28b) \end{cases}$$

Since $\eta(x) = 0$ when $\phi = 0$, it is clear from the last set of equations that I and g_m approach their first-order values as ϕ becomes small. Figure 3.3 shows normalized g_m versus s for several values of ϕ. Equation 3.27 is not plotted because it does not differ appreciably from the first-order Eq. 2.34. The sharp corner in g_m at $s = 1$ arises because of the approximations used; in practice we expect this corner to be smoothed out. We see that the general shape of g_m for $\phi > 0$ is similar to the $\phi = 0$ shape, but that the corrected values are everywhere greater than the first-order values. In particular, we note that the correction to the g_m curve, once the sharp corner is removed, is in the proper direction to account for the experimental discrepancies mentioned earlier.

In order to correlate experimental data with the theory we must calculate ϕ. An experimental plot of I/g_m vs. V_g, as described in Section 5.7, provides the means for determining the external pinch-off voltage V_p. The built-in potential may be determined from capacitance measurements: a plot of $(1/C_{11})^2$ vs. V_g with $V_d = 0$ is linear and yields V_c as an intercept. The values for the alloy-junction unit Crystalonics C610 (#2) are

$$V_p = 9.33v$$

$$V_c = 0.55v$$

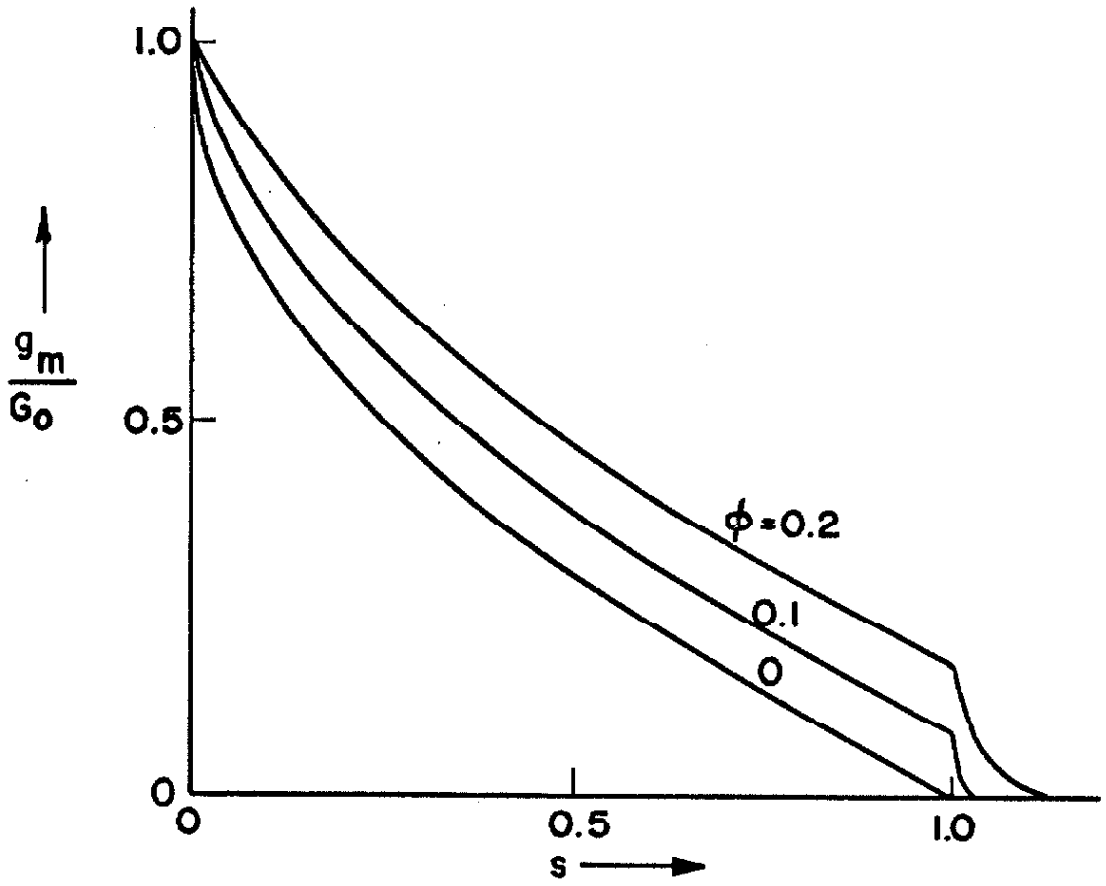


Fig. 3.3. Normalized transconductance versus normalized gate voltage. The parameter ϕ provides a measure of the importance of the non-abrupt space-charge-region boundary; $\phi = 0$ corresponds to an abrupt boundary.

Thus, $W_o = V_c + V_p = 9.88v$, and at 20°C we have

$$\phi = \sqrt{\frac{2.52 \times 10^{-2}}{9.88}} = 0.051$$

Curves of g_n for $\phi = 0$ and $\phi = 0.051$ are shown with the experimental values in Fig. 3.4a. All quantities have been normalized to their values at $s = 0.25$ because $g_m(.25) = \frac{1}{2}\left[g_m(0) + g_m(1)\right]$, and thus experimental errors tend to be averaged out. With a rounding of the corner on the $\phi = 0.051$ curve, the new theory agrees well with the data.

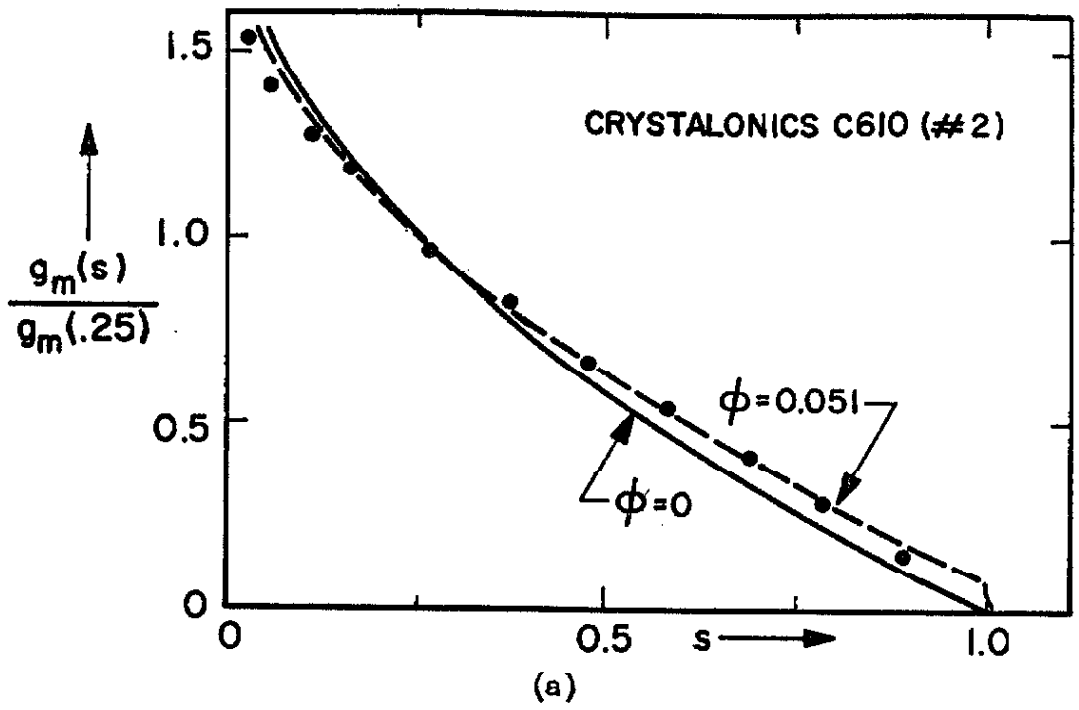
The theory was also applied to another unit, with less successful results. For this transistor - Motorola type MM764 (#3), epitaxial junction - the parameters are

$$V_p = 2.74v$$

$$V_c = 0.70v$$

$$\phi = \sqrt{\frac{2.52 \times 10^{-2}}{3.44}\ (2)} = 0.121$$

The factor of 2 is present under the square root sign because the gate and channel dopings are equal in this unit [11], and hence $W_o = 2(e_c a^2/2\epsilon)$. (Cf. footnote page 57.) Since ϕ is larger in this example, we expect greater deviations from the first-order theory. Yet the plots Fig. 3.4b show good agreement with that theory.



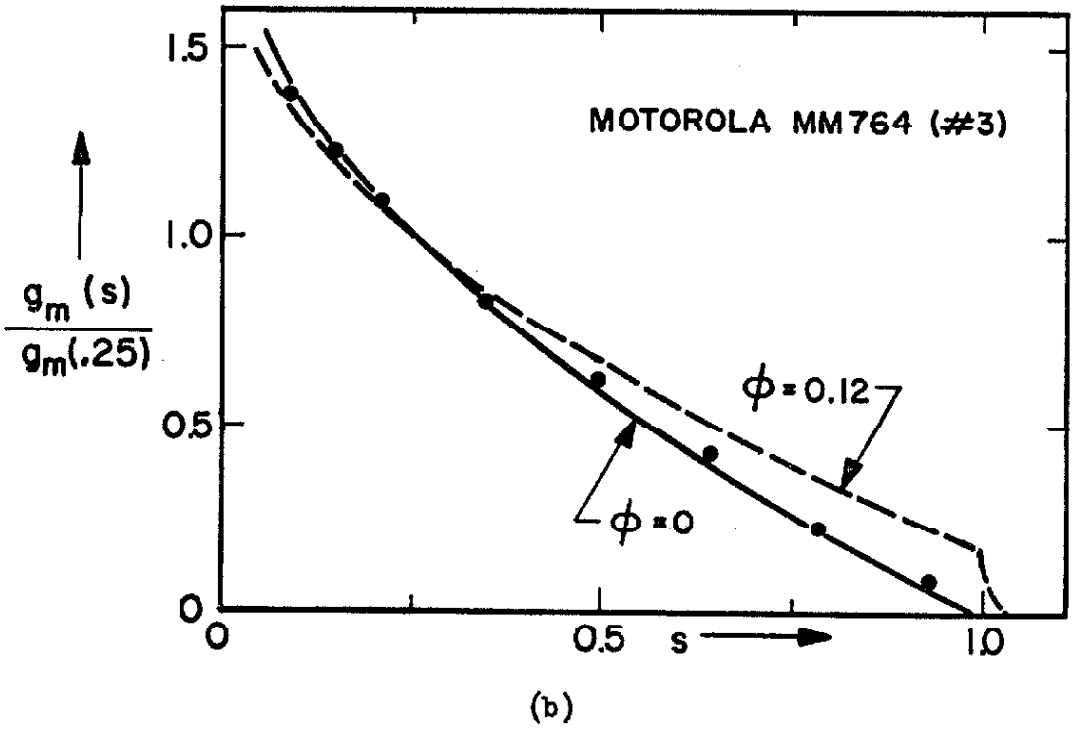


Fig. 3.4. Correlation between theoretical and experimental values of g_m. The dashed curves are based on the theory developed for non-abrupt space-charge-region boundary; the solid curves are based on the abrupt-boundary theory.

The failure of the new theory in fitting the data of Fig. 3.4b, and the success of the old theory in fitting this and most other experimental results tend to indicate that the first-order theory is adequate for practical purposes, and that possibly the fit in Fig. 3.4a is merely fortuitous, the deviation in g_m being caused by other phenomena. The defects in the new theory may perhaps be traced to the severity of the assumptions employed. But if the assumptions were too acute, then their effect should be less than that calculated here, and hence the space-charge-region boundary may be considered abrupt.

3.6. Conclusions

The general conclusion that may be drawn from the discussions in this chapter is that the simple, first-order analysis performed in Chapter II is satisfactory. This conclusion is supported by both theoretical and empirical evidence. The few experimental deviations in g_m are small enough in quantity and in magnitude to be attributed - in the light of the work done in this chapter - to effects lying outside our model and its related assumptions. Further, the assumptions used in the analysis of the step-junction FET will be at least as accurate when applied to other practical profiles because, as mentioned in Section 3.4, these latter dopings in general have higher conductivity at $y = 0$ than at $y = b$. This results in more current flowing in the $y = 0$ plane, hence improves the

accuracy of the gradual approximation and reduces the importance of the non-abrupt boundary on the space-charge region and of the deviations from idealized geometry. Thus, we may now generalize the analysis of the step-junction profile to arbitrary profiles and have confidence in the accuracy of the results.

CHAPTER IV

ANALYSIS OF AN ARBITRARILY DOPED FIELD-EFFECT TRANSISTOR

In this chapter we shall use the same model for the FET as in Chapter II, but we now generalize the problem to relate to an arbitrary doping density $\rho(y)$. Thus, except for the assumptions on the doping and conductivity (and on the built-in potential), all the assumptions from Section 2.2 will be carried over. In particular, the gradual approximation and the assumption used to extend the results for drain-gate voltages below pinch-off $(d \leq 1)$ to voltages above pinch-off $(d > 1)$ are fundamental to the following analysis.

Section 4.1 contains an interesting derivation of approximate FET characteristics. We show in Section 4.3 that these characteristics correspond to an important type of doping profile, and hence the procedure used in Section 4.1 is of some significance. The general treatment, analogous in approach to that of Chapter II, is presented in Section 4.2. The final results for the static and small-signal parameters are quite simple and in many cases are less complicated when expressed in terms of an arbitrary doping profile than when expressed in terms of a particular profile. Further, the general equations permit broad inferences to be made about the similarity of the properties of all FET's. Two useful, illustrative applications of the general treatment are given in Section 4.3.

4.1. A Simple, Approximate Derivation of FET Characteristics [18]

In the following derivation we shall use the charge-control approach [6], [19] to show, without specifying the detailed nature of the doping profile, that beyond pinch-off the drain current and the internal gate voltage are related approximately by a square law.

The transit time τ_t is defined as the average time required for a carrier to travel from the source to the drain. Alternatively, τ_t is the time required for the charge in transit Q to flow out of the drain. But charge is flowing out of the drain at the rate of I coulombs per second. Thus,

$$\tau_t = \frac{Q}{I} \quad , \qquad \text{or} \qquad I = \frac{Q}{\tau_t} \tag{4.1}$$

where, for the purposes of this derivation, we consider Q (as well as I) to be a positive quantity. This equation is valid only if Q is approximately constant during the transit of a carrier from the source to the drain, that is, if the transit time is much less than the time required for the redistribution of charges during a change in applied potential. The redistribution time is of the order of the reciprocal of the angular frequency ω of the applied potential, and hence the following theory is applicable for frequencies such that $\omega \ll 1/\tau_t$. The fundamental charge-control relation 4.1, along with two simplifying assumptions, forms the basis of the following derivation.

The basic assumption that we shall employ in this section is that the electric field in the channel is uniform, and hence equal to W_d/L. Then the drift velocity is constant at $\mu W_d/L$, and the transit time is $\tau_t = L^2/\mu W_d$. We break up Q into two components, Q_o and Q_g, such that $Q = Q_o - Q_g$. The first component, Q_o, is the charge that would exist in the absence of biases (independent of bias, but dependent on the doping profile*). The other component is the charge induced by the gate voltage. In general, Q_g is positive, but it may be negative for forward-biased insulated-gate FET's. We may express this charge in terms of the internal potentials by means of the relation

$$Q_g = C_g(W_g + W_d/2)$$

where $(W_g + W_d/2)$ is the average gate-channel potential (since the field is constant), and where C_g represents a capacitance, which, under certain conditions, may be identified as the total gate capacitance. As a first approximation C_g may be considered constant, independent of the biases. Combination of these results gives the drain current as

$$I = \frac{\mu Q_o}{L^2}\left(1 - \frac{W_g + W_d/2}{Q_o/C_g}\right)W_d \qquad (4.2)$$

---

* In terms of the channel conductance G_o, Q_o may be written as $Q_o = G_o L^2/\mu$.

The output conductance is

$$G_{22} = \left| \frac{\partial I}{\partial W_d} \right| = \frac{\mu Q_o}{L^2} \left(1 - \frac{W_g + W_d}{Q_o/C_g} \right)$$

This quantity goes to zero when $W_g + W_d = Q_o/C_g$, and so, by analogy with the results of Chapter II, we identify the pinch-off voltage as

$$W_o = Q_o/C_g \qquad (4.3)$$

In terms of the normalized voltages $s = W_g/W_o$ and $d = (W_g + W_d)/W_o$, Eq. 4.2 becomes

$$I = \frac{\mu Q_o W_o}{L^2} \left(1 - \frac{s}{2} - \frac{d}{2} \right) (d - s), \qquad s \leq d \leq 1 \qquad (4.4a)$$

The drain current at, and, by extension, beyond pinch-off is obtained by substituting $d = 1$ in this expression. Thus,

$$I = \frac{\mu Q_o W_o}{L^2} \frac{1}{2} (1 - s)^2, \qquad d > 1 \qquad (4.4b)$$

Plots of these drain characteristics will be similar to the step-junction characteristics Fig. 2.7; only the spacing between curves will be different. Beyond pinch-off the transfer characteristics, Eq. 4.4b, are especially close to the corresponding step-junction curve (see Fig. 4.2). In Section 4.3 we shall find a particular

doping profile that yields the current-voltage dependence given by Eqs. 4.4.

The pinch-off voltage W_o defined by Eq. 4.3 is not really a constant because C_g in general depends on both W_g and W_d. Beyond pinch-off, however, C_g is essentially independent of W_d, so that in Eq. 4.4b W_o is dependent,* though only weakly so, upon W_g. Thus, the fact that C_g is a function of the biases, and the assumption that the channel field is uniform result in the square-law dependence 4.4b being only an approximation. Further discussion on the power-law nature of FET characteristics, including a theoretical explanation for the similarity between the square-law characteristics and the step-junction characteristics, is given in Section 5.7.

4.2. General Treatment

We are now ready to treat an arbitrarily doped FET. We shall use the same assumptions and the same approach as in Chapter II, except now we work with an unspecified impurity profile $\rho(y)$. First, we solve Poisson's equation in the space-charge region to obtain a relation between the channel potential and the channel width. Then we use this relation implicitly to determine the drain current I and the charge in transit Q as functions of the biases. The various derivatives of I and Q specify the small-signal equivalent circuit, and their ratio Q/I defines the transit time. Apart from

---

* Note that this bias dependence of W_o also appears in the denominator of s.

constant multiplicative factors, all of the device parameters will be expressed entirely in terms of an integral of the doping profile.

Figure 2.1 gives the device model. Equations 2.5, 2.6, and 2.7 may be carried over directly from Section 2.3; we rewrite them here for convenience:

Poisson's equation combined with gradual approximation:

$$\frac{\partial^2 V(x,y)}{\partial y^2} = - \frac{\rho(y)}{\epsilon} , \qquad b(x) \le y \le c(x) \qquad (4.5)$$

Boundary conditions:

$$\begin{cases} E_y = - \dfrac{\partial V}{\partial y} = 0 \qquad \text{at} \quad y = b(x) & (4.6) \\ \\ V[x,c(x)] = V_g & (4.7) \end{cases}$$

A first integration of 4.5, with boundary condition 4.6 gives

$$\frac{\partial V}{\partial y} = - \frac{1}{\epsilon} \int_{b(x)}^{y} \rho(y_1)\, dy_1 \qquad (4.8)$$

The boundary of the depletion region in the gate material also has $E_y = 0$, so that

$$\int_{b(x)}^{c(x)} \rho(y_1)\, dy_1 = 0 \qquad (4.9)$$

This equation merely states that in any plane perpendicular to the junction the space charge within the channel material is equal in

magnitude and opposite in sign to the space charge within the gate material. When 4.8 is integrated subject to 4.7 we get

$$V(x,y) = V_g + \frac{1}{\epsilon}\int_{y}^{c(x)}\left[\int_{b(x)}^{y_2}\rho(y_1)\,dy_1\right]dy_2 \qquad (4.10)$$

At the channel boundary we may use Eq. 4.9 and integrate 4.10 by parts to obtain

$$V(x,b) = V_g - \frac{1}{\epsilon}\int_{b(x)}^{c(x)} y\rho(y)\,dy \qquad (4.11)$$

as the potential within the channel corresponding to the step-junction expression 2.9.* At present, Eq. 4.11 is in terms of applied voltages. To convert to internal voltages we merely add V_c to V_g and take the magnitude of each side of the resulting equation:

$$W(x,b) = \left|V_g + V_c - \frac{1}{\epsilon}\int_{b(x)}^{c(x)} y\rho(y)\,dy\right| \qquad (4.12)$$

At $x = 0$, we have $b = b_s$, $c = c_s$, and $W = 0$, so that

$$W_g = |V_g + V_c| = \left|\frac{1}{\epsilon}\int_{b_s}^{c_s} y\rho(y)\,dy\right| \qquad (4.13)$$

---

* Equation 2.9 cannot be deduced directly from 4.11 because a limiting process is involved. Thus, although we let $\beta\to\infty$ in the doping profile given by Eq. 2.1, and although in that case $c\to a$, we see from 4.9 that $\beta(c - a)$ is finite and equals $(a - b)$. To obtain Eq. 4.11 we must substitute Eq. 2.1 for $\rho(y)$ and <u>then</u> take limits.

Since $W(x,b) \geq W(0,b) = W_g$, Eq. 4.12 may be written in the form

$$W(x,b) = \left| \frac{1}{\epsilon} \int_{b(x)}^{c(x)} y\rho(y)\, dy \right| - W_g \qquad (4.14)$$

This expression, together with the relation between b and c given by Eq. 4.9, is the desired function which gives the channel potential in terms of the channel width. The magnitude of the potential across the gate-channel junction is $(W + W_g)$. When the channel is pinched-off, $b = 0$, and this potential is equal to the pinch-off voltage W_o. Therefore, from Eq. 4.14

$$W_o = \left| \frac{1}{\epsilon} \int_{0}^{c_m} y\rho(y)\, dy \right| \qquad (4.15)$$

where c_m is defined as the maximum penetration of the space-charge region into the gate material and is obtained from Eq. 4.9:

$$\int_{0}^{c_m} \rho(y)\, dy = 0 \qquad (4.9a)$$

Before proceeding with the analysis of the channel, we introduce three new normalized parameters. Corresponding to $u(x) = b(x)/a$, we define

$$w(x) \equiv \frac{c(x)}{a} \qquad (4.16)$$

(and $w_m = c_m/a$). We also normalize the y-coordinate with respect

to a:

$$z \equiv \frac{y}{a} \tag{4.17}$$

To avoid confusion, we would like to use the same symbol, ρ, to denote the doping profile as a function of z. Therefore, we make the replacement

$$\rho(y) \rightarrow \rho\left(\frac{y}{a}\right) = \rho(z) \tag{4.18}$$

Note that $\rho(1) = 0$ and $\rho(u)\rho(w) \leq 0$ since the junction at $z = 1$ separates two oppositely doped materials. Finally, we define θ as the magnitude of the normalized channel potential measured with respect to the gate:

$$\theta(x) \equiv \frac{W[x,b(x)]}{W_o} + \frac{W_g}{W_c} = \frac{W[x,b(x)]}{W_o} + s \tag{4.19}$$

Many of the following results appear in a convenient, symmetrical form because θ is measured with respect to the gate. The symmetry arises, in part, from the fact that the source contact has $\theta = s$, and the drain contact has $\theta = (W_d/W_o + s) = d$, and thus s and d have equal significance, each being the "potential" at one of the output electrodes. In terms of these new quantities, the equations for the various internal potentials become

$$\theta = \frac{\int_{u}^{w} z\rho(z)\ dz}{\int_{0}^{w_m} z\rho(z)\ dz} \tag{4.20}$$

$$s = \frac{\int_{u_s}^{w_s} z\rho(z)\ dz}{\int_{0}^{w_m} z\rho(z)\ dz} \tag{4.21}$$

$$W_o = \frac{a^2}{\epsilon}\left|\int_{0}^{w_m} z\rho(z)\ dz\right| \tag{4.22}$$

subject to the conditions

$$\int_{u}^{w} \rho(z)\ dz = \int_{u_s}^{w_s} \rho(z)\ dz = \int_{0}^{w_m} \rho(z)\ dz = 0 \tag{4.23}$$

Absolute value signs are not needed in Eqs. 4.20 and 4.21 because the numerator and denominator in each expression always have the same sign. Equation 4.23 permits us to write Eq. 4.20 as

$$\theta = \frac{\int_{u}^{w} (1 - z)\rho(z)\, dz}{\int_{w_m}^{0} (1 - z)\rho(z)\, dz} \qquad (4.20a)$$

This form is especially useful when antisymmetric profiles $[\rho(2 - z) = -\rho(z), \quad 0 \le z \le 2]$ are considered.

We now solve for the drain current by examining the differential element of channel pictured in Fig. 2.2. The two-dimensional density of free carriers in this element is

$$a \int_{0}^{u(x)} \rho(z)\, dz$$

where we temporarily assume an n-type channel $(\rho > 0)$ to avoid difficulties with signs. The resistance of the element is thus

$$dR = \frac{dx}{A\mu} \left[a \int_{0}^{u(x)} \rho\, dz \right]^{-1}$$

From Eq. 4.19 we have $dW = W_o d\theta$, so that the drain current is

$$I = \frac{dW}{dR} = \mu a \Lambda W_c \frac{d\theta}{dx} \int_{0}^{u(x)} \rho\, dz \qquad (4.24)$$

which, upon integration from $x = 0, \quad \theta = s$ to $x = L, \quad \theta = d$ gives

$$I = \frac{\mu a A W_o}{L} \int_s^d \left\{ \int_0^{u[x(\theta)]} \rho \, dz \right\} d\theta \qquad (4.25)$$

since I is not a function of x. In this formula we have explicitly indicated the dependence of u on θ (as given in Eqs. 4.20 and 4.23). We obtain g_m by partial differentiation of Eq. 4.25 with respect to W_g:

$$g_m \equiv \left| \frac{\partial I}{\partial W_g} \right| = \frac{\mu a A}{L} \int_{u(d)}^{u(s)} \rho \, dz \qquad (4.26)$$

This result has also been obtained by Bockemuehl [20]. When $s = 0$ and $d = 1$ (i.e. when $u(s) = 1$ and $u(d) = 0$) the transconductance is

$$g_m(0,1) = \frac{\mu a A}{L} \int_0^1 \rho \, dz$$

We recognize this expression as the conductance G_o of the channel in the absence of any biases. Thus, in general,

$$g_m(0,1) = G_o = \frac{\mu a A}{L} \left| \int_0^1 \rho \, dz \right| \qquad (4.27)$$

where we have inserted absolute value signs in order to include p-type channels ($\rho < 0$).

We define a new function,

$$g(\theta) \equiv \frac{\int_0^{u(\theta)} \rho \, dz}{\int_0^1 \rho \, dz} \tag{4.28}$$

in terms of which many of our final results may be expressed most simply. The function $g(\theta)$ always lies within the range $g(1) = 0 \leq g(\theta) \leq 1 = g(0)$. We now remove the assumption of an n-type channel and rewrite Eqs. 4.25 and 4.26 in terms of G_o and g:

$$I(s,d) = G_o W_o \int_s^d g(\theta) \, d\theta \tag{4.29}$$

$$g_m(s,d) = G_o[g(s) - g(d)] \tag{4.30}$$

From Eq. 4.30 we may glean the primary significance of the function $g(\theta)$: It is the normalized transconductance beyond pinch-off at normalized gate voltage θ [i.e. $g(\theta) = g_m(\theta,1)/G_o$].

When $s = 0$ and $d = 1$ the drain current is

$$I_o \equiv I(0,1)$$

$$= G_o W_o \int_0^1 g(\theta) \, d\theta \tag{4.31}$$

Since $0 < g(\theta) < 1$ for $0 < \theta < 1$, $I_o < G_o W_o$, as was the case with the step-junction FET. The general form of the family of drain characteristics given by Eq. 4.29 is the same as in Fig. 2.7. For a

given s the drain current increases with $(d - s) = |V_d|/W_o$ until $d = 1$. Then, in the pinch-off range, the current remains at its $d = 1$ value. Further, as in the step-junction device, g_m in the general FET is equal to the conductance of a channel of constant fractional width $u(0) - u(L)$.

Differentiation of Eq. 4.29 with respect to W_d gives the output conductance as

$$G_{22}(s,d) = G_o g(d) \tag{4.32}$$

Again G_{22} is equal to the conductance of a channel of constant fractional width $u(L)$, is independent of the gate-source voltage, and goes to zero when $d = 1$.

It is easy to determine an antisymmetric doping profile that will produce a desired $g(\theta)$ (provided that $d^2g/d\theta^2 > 0$). The necessary equations are

$$u(\theta) = 1 - \frac{g'(1)}{g'(\theta)} \tag{4.33}$$

$$o(u) = B_1 \frac{[g'[\theta(u)]]^3}{g''[\theta(u)]}, \qquad 0 \leq u < 1 \tag{4.34a}$$

$$= B_2 \frac{1}{(1 - u)u'(\theta)}, \qquad 0 \leq u < 1 \tag{4.34b}$$

and

$$\rho(u) = -\rho(2 - u) \quad , \qquad 1 < u \leq 2 \qquad (4.35)$$

where primes denote differentiation with respect to θ, and where B_1 and B_2 are arbitrary constants. The inverse function $\theta(u)$ is obtained from Eq. 4.33 for use in Eq. 4.34. [The requirement $g''(\theta) > 0$ ensures that $u(\theta)$ is a strictly decreasing function of θ, hence ensures the existence of a unique inverse function $\theta(u)$.] Verification that $\rho(u)$ defined above does generate the function $g(\theta)$ may be obtained by substitution into Eqs. 4.20a and 4.28.

We now calculate the charge in transit and the capacitive equivalent-circuit elements. The differential channel element of Fig. 2.2 has

$$|dQ| = aA\, dx \left| \int_0^u \rho\, dz \right| = \frac{G_o L}{u}\, g(\theta)\, dx$$

free carriers, and hence

$$|Q| = \frac{G_o L}{\mu} \int_0^L g(\theta)\, dx$$

From Eqs. 4.24, 4.27, and 4.28, we have

$$dx = \frac{W_o L}{\bar{u}}\, G_o g(\theta)\, d\theta$$

so that

$$|Q| = \frac{G_o^2 L^2 W_c}{\mu I(s,d)} \int_s^d g^2(\theta)\, d\theta$$

We may substitute Eq. 4.29 for I to obtain Q entirely in terms of the function g:

$$|Q| = \frac{G_o L^2}{\mu} \frac{\int_s^d g^2(\theta)\, d\theta}{\int_s^d g(\theta)\, d\theta} \tag{4.36}$$

The input and output charge-capacitances are obtained by differentiating this relation with respect to W_g and W_d respectively:

$$C_{11}^*(s,d) = C_{22}^*(s,d) + \frac{G_o L^2}{\mu W_o} g(s) \frac{\int_s^d g(\theta)\{g(s) - g(\theta)\}d\theta}{\left[\int_s^d g(\theta)d\theta\right]^2} \tag{4.37}$$

$$C_{22}^*(s,d) = \frac{G_o L^2}{\mu W_o} g(d) \frac{\int_s^d g(\theta)\{g(\theta) - g(d)\}d\theta}{\left[\int_s^d g(\theta)d\theta\right]^2} \tag{4.38}$$

We may draw the same conclusions concerning these general charge-capacitances as we drew concerning the step-junction capacitances in Chapter II. First, we see immediately from the above equations that

$$C_{11}^*(s,d) = C_{22}^*(s,d) = C_{22}^*(d,s) \tag{4.39}$$

Figure 2.5 and the accompanying arguments indicating that C_{22}^* should appear entirely between the gate and the drain are thus relevant here. These results for the charge-capacitances, together with the results for the conductance elements g_m and G_{22}, prove that the small-signal equivalent circuit in the general case is identical to the circuit of Fig. 2.4. Above pinch-off the circuit reduces to that of Fig. 2.9 with only C_{11}^* and g_m present.

Second, when $s = d$ $(W_d = 0)$ the channel is of uniform width, and we expect C_{11}^* to equal the capacitance of a planar p-n junction of area LA with depletion-region thickness $a(w_s - u_s)$. To prove this we apply l'Hôpital's rule to Eq. 4.37 to get

$$C_{11}^*(s,s) = - \frac{G_o L^2}{\mu W_o} \frac{dg(s)}{ds}$$

It is not difficult to show that

$$\frac{dg}{ds} = - \frac{\mu \epsilon A W_o}{L a G_o} \frac{1}{w_s - u_s} = \frac{\int_0^{w_m} z\rho \, dz}{\int_0^1 \rho \, dz} \frac{1}{w_s - u_s} < 0 \qquad (4.40)$$

(We shall derive this relation in somewhat more detail in the next chapter.) Substitution of this expression into the previous formula gives

$$C_{11}^{*}(s,s) = \frac{\epsilon A L}{a} \frac{1}{w_s - u_s} \tag{4.41}$$

which is the desired result.

Finally, in Section 2.4.2 we showed that for a step-junction FET

$$\lim_{s\to 1} \left[\lim_{d\to s} C_{11}^{*}(s,d) \right] = \frac{2}{3} \lim_{s\to 1} C_{11}^{*}(s,1)$$

Application of l'Hôpital's rule to $C_{11}^{*}(s,1)$ from Eq. 4.37 and comparison with Eq. 4.41 yield the same relation in the general case. The mathematical and physical reasons for the presence of the non-unity numerical factor $2/3$ are given in Section 2.4.

Another quantity of interest is the transit time τ_t defined in the previous section. From Eqs. 4.29 and 4.36 we have

$$\tau_t = \frac{|Q|}{I} = \frac{L^2}{\mu W_o} \frac{\int_s^{\bar{d}} g^2 \, d\theta}{\left(\int_s^{\bar{d}} g \, d\theta \right)^2} \tag{4.42}$$

Now, the velocity of carriers in the channel is equal to $\mu \, dW/dx$, and hence τ_t may also be written as

$$\tau_t = \int_0^L \frac{dx}{\mu \, dW/dx} = \int_0^L \frac{dx}{\mu W_o \, d\theta/dx}$$

By variational calculus techniques we may minimize τ_t with respect

to $\theta(x)$ [with $\theta(0) = s$ and $\theta(L) = d$ fixed]. Minimum τ_t occurs when $d\theta/dx = \text{constant} = (d - s)/L$. Since the maximum operating frequency of a device varies as $1/\tau_t$, the above calculation shows that an FET with a constant channel field will have the best frequency response. The simple derivation in the previous section assumed a constant channel field and hence corresponds to this type of optimum FET.

We may summarize the separate conclusions of this section by stating that differences among FET's are merely quantitative, and that all the qualitative conclusions from the step-junction discussions of Chapter II may be applied to any other FET. In Chapter V we shall prove that even quantitative results on different FET's may be very close, but first we show the advantages of the general treatment by working out some practical examples.

4.3. Applications of the General Treatment

In this section we shall apply the previous results to two useful, general doping profiles. The first profile is antisymmetric and in the form of a power law, and includes the linearly graded junction as a special case and the step junction as one limiting case. The other limiting case of this profile, in which all the channel mobile charge is in the plane $y = 0$, is defined as a "delta" junction. The second profile has a region of zero doping (insulator) separating uniformly, but unequally doped p- and n-regions. This profile, in addition to yielding values for the limiting cases of the unsymmetrical step junction and the delta junction, provides equations

governing the operation of insulated-gate FET's. Although these examples include a wide variety of possible FET structures, we shall see that the resulting external properties are quite similar.

For convenience, before beginning the calculations we rewrite the formulae for $g(\theta)$ and for θ:

$$g(\theta) = \frac{\int_0^{u(\theta)} \rho \, dz}{\int_0^{1} \rho \, dz} \tag{4.43}$$

$$\theta = \frac{\int_{w_m}^{u} z\rho \, dz}{\int_0^{w_m} z\rho \, dz} = \frac{\int_{w_m}^{u} (1 - z)\rho \, dz}{\int_0^{w_m} (1 - z)\rho \, dz} \tag{4.44}$$

with

$$\int_u^w \rho \, dz = \int_0^{w_m} \rho \, dz = 0 \tag{4.45}$$

If $\rho(z)$ is antisymmetric about $z = 1$, then Eq. 4.45 yields

$$\left.\begin{aligned} w &= 2 - u \\ w_m &= 2 \end{aligned}\right\} , \qquad \text{(antisymmetric } \rho\text{)}$$

and the function $(1 - z)\rho(z)$ is symmetric about $z = 1$. Equation 4.44 may then be simplified to

$$\theta = \frac{\int_u^1 (1 - z)\rho \, dz}{\int_0^1 (1 - z)\rho \, dz} \quad , \qquad \text{(antisymmetric } \rho\text{)} \qquad (4.46)$$

Example 1.

We consider the antisymmetric profile given by

$$\rho(z) = \begin{cases} \rho_1(1 - z)^n & , \quad 0 \le z < 1 \\ C & , \quad z = 1 \\ -\rho_1(z - 1)^n & , \quad 1 < z \le 2 \end{cases} \qquad (4.47)$$

where ρ_1 is a positive constant and $n \ge 0$. Figure 4.1 shows $\rho(z)$ for several values of n. An antisymmetric step junction is given by $n = 0$; a linearly graded junction by $n = 1$; and a delta junction by the limit $n \to \infty$. We apply Eqs. 4.46 and 4.47 to get

$$\theta = (1 - u)^{n+2}$$

$$g = 1 - (1 - u)^{n+1}$$

The parameter u may be eliminated from these equations to give g as a function of θ:

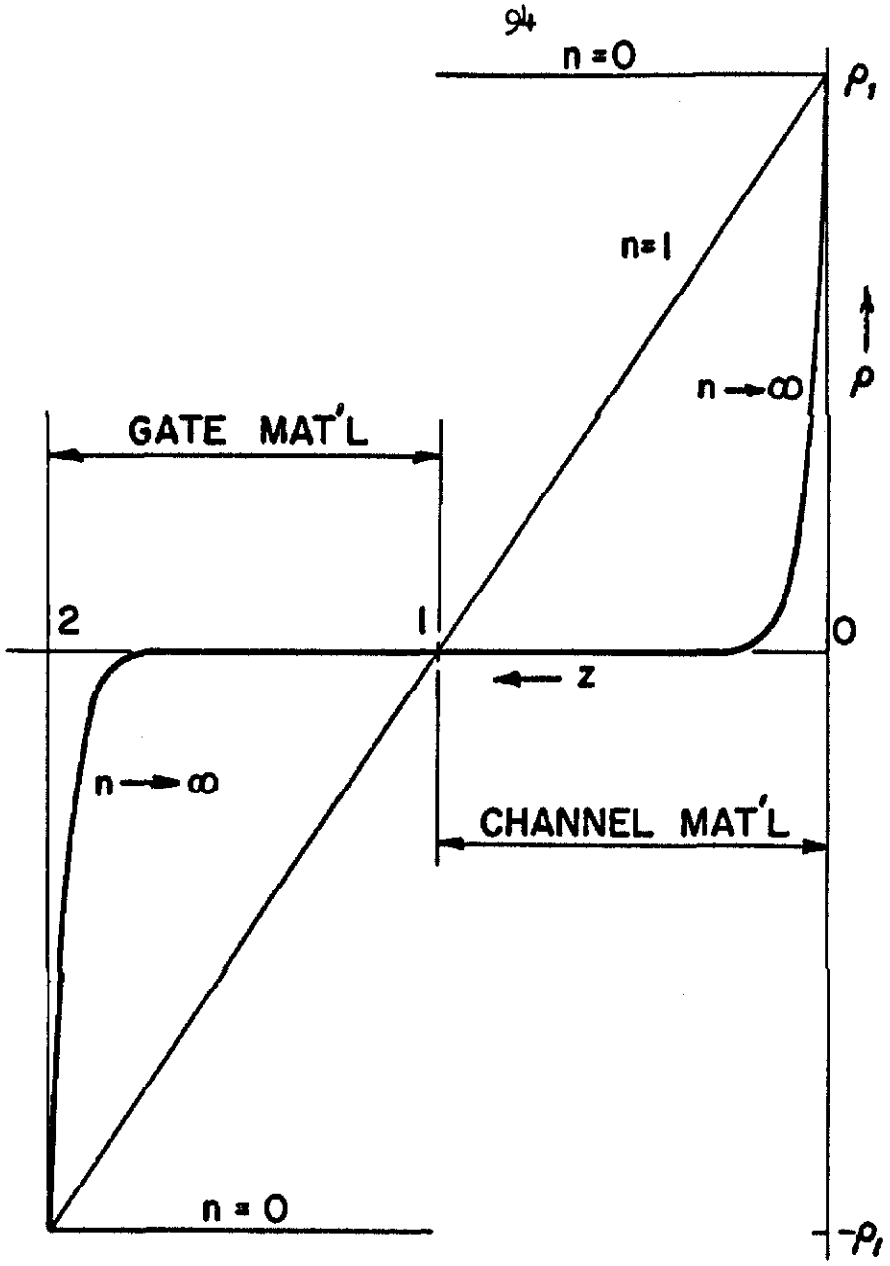


Fig. 4.1. Doping profile used in Example 1. A step junction is given by $n = 0$, a linearly graded junction by $n = 1$, and a delta junction by $n \to \infty$.

$$g(\theta) = 1 - e^{\left(\frac{n+1}{n+2}\right)} \tag{4.48}$$

Now that $g(\theta)$ is known, we may calculate the drain current and the small-signal equivalent circuit for arbitrary biases from the formulae of the previous section. We shall, however, restrict our attention to biases of greatest interest, namely, to operation beyond pinch-off. For $d = 1$ the relevant normalized equations are:

$$\frac{g_m}{G_o} = g(s) \tag{4.49}$$

$$\frac{I}{I_o} = \frac{\int_s^1 g \, d\theta}{\int_\theta^1 g \, d\theta} \tag{4.50}$$

$$\frac{C_{11}^*}{C_o^*} = g(s) \frac{\int_s^1 g(0)\{g(s) - g(0)\} \, d\theta}{\left[\int_s^1 g \, d\theta\right]^2} \frac{\left[\int_0^1 g \, d\theta\right]^2}{\int_0^1 g(1 - g) \, d\theta} \tag{4.51}$$

where

$$C_o^* \equiv C_{11}^*(0,1)$$

$$= \frac{G_o L^2}{\mu W_o} \frac{\int_0^1 g(1-g)\, d\theta}{\left(\int_0^1 g\, d\theta\right)^2} \tag{4.52}$$

(Also, $C_{22} = C_{22}^* = 0$.) Substitution of $g(\theta)$ into these equations gives

$$\frac{g_m}{G_o} = 1 - s^N \tag{4.53}$$

$$\frac{\bar{I}}{I_o} = 1 - \left(1 + \frac{1}{N}\right)s + \frac{1}{N}s^{N+1} \tag{4.54}$$

$$\frac{C_{11}^*}{C_c^*} = \frac{(1 - s^N)[1 - (2N+1)s^N + (2N+1)s^{N+1} - s^{2N+1}]}{\left[1 - \left(1 + \frac{1}{N}\right)s + \frac{1}{N}s^{N+1}\right]^2} \tag{4.55}$$

in which we have written

$$N = \frac{n+1}{n+2}$$

For $n = 0$ $(N = 1/2)$ these equations reduce to the step-junction results of Chapter II; for $n = 1$ $(N = 2/3)$ we obtain equations for a linear junction,

$$\frac{g_m}{G_o} = 1 - s^{2/3}$$

$$\frac{I}{I_o} = 1 - \frac{5}{2}s + \frac{3}{2}s^{5/3} \qquad (4.56)$$

$$\frac{C_{11}^*}{C_o^*} = \frac{(1 - s^{2/3})\left[1 - \frac{7}{3}s^{2/3} + \frac{7}{3}s^{5/3} - s^{7/3}\right]}{\left[1 - \frac{5}{2}s + \frac{3}{2}s^{5/3}\right]^2}$$

and as $n \to \infty$ $(N \to 1)$ we obtain the especially simple delta-junction formulae

$$\frac{g_m}{G_o} = 1 - s$$

$$\frac{I}{I_o} = (1 - s)^2 \qquad (4.57)$$

$$\frac{C_{11}^*}{C_o^*} = 1$$

The results for step-, linear-, and delta-junction FET's are drawn in Fig. 4.2. We note that the formulae for the delta junction are identical to those from the simple derivation of Section 4.1 (square-law for I, constant C_{11}^*).

Example 2.

We consider the unsymmetric profile

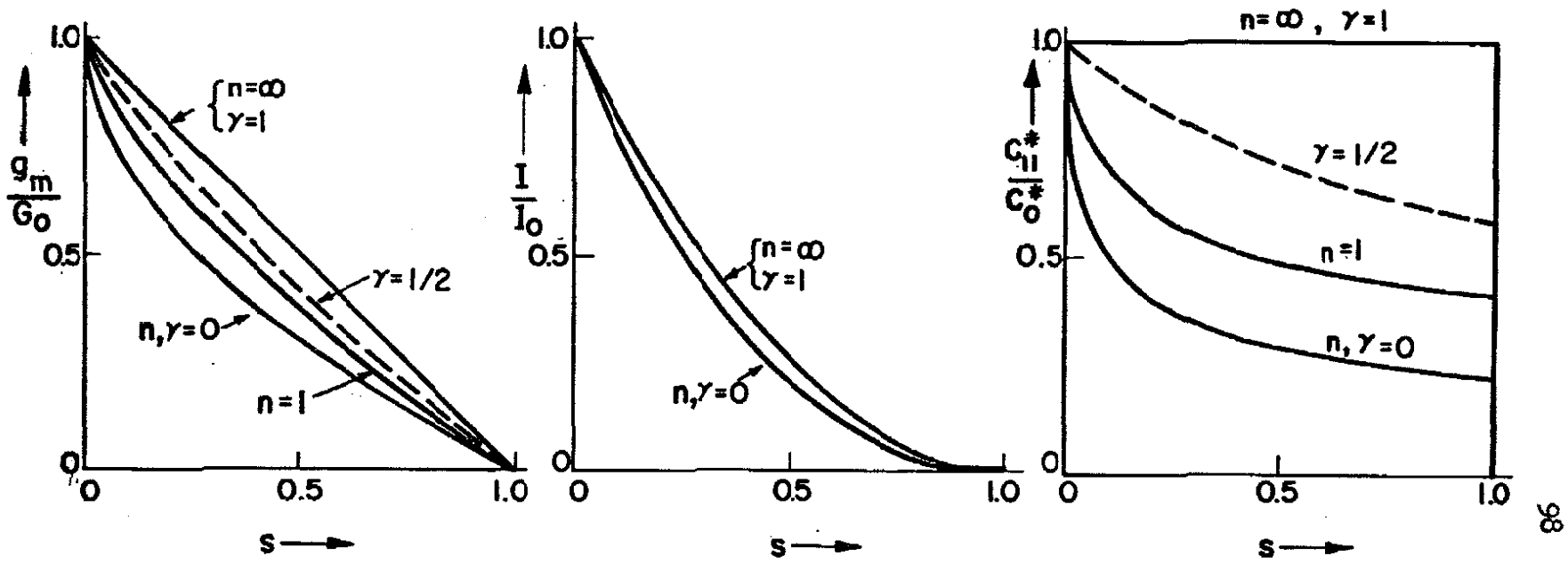


Fig. 4.2. Results for the two examples in the text. The parameter n applies to Example 1, the parameter γ to Example 2. Curves are shown for a step-junction FET ($n = 0$ or $\gamma = 0$); for a linear-junction FET ($n = 1$); for an insulated-gate FET with $\gamma = 1/2$; and for a delta-junction FET ($n = \infty$ or $\gamma = 1$). Because of the proximity of the step- and delta-junction transfer characteristics, the intermediate curves have not been drawn.

$$\rho(z) = \begin{cases} \rho_o , & 0 \le z < \alpha \\ 0 , & \alpha \le z \le 1 \\ -\beta\rho_o , & 1 < z \end{cases} \qquad (4.58)$$

where ρ_c and β are positive constants. This function generalizes the step-junction profile Eq. 2.1 by having an insulating region $\alpha \le z \le 1$ between the gate and channel materials. If $\alpha = 1$, we have an ordinary step-junction doping; if $\alpha < 1$ we have an insulated-gate FET; the limit $\alpha \to 0$ defines a delta junction. By varying α within the range $0 < \alpha < 1$ we specify different ratios of insulator thickness to channel thickness.

We proceed in the same manner as in Example 1, but now we must use Eqs. 4.44 and 4.45 to determine $\theta(u)$ since ρ is not anti-symmetric. The results are

$$g = \frac{u}{\alpha} \qquad (4.59)$$

$$\theta = \left(1 - \frac{u}{\alpha}\right)\left[1 - \frac{u}{\alpha}\left(\frac{1 - \gamma}{1 + \gamma}\right)\right] \qquad (4.60)$$

where we have written

$$\gamma = \frac{1 - \alpha}{1 + \alpha/\beta} \qquad (0 \le \gamma < 1)$$

From Eq. 4.45 we find that $w_m = 1 + \alpha/\beta$ is the maximum penetration of the depletion region into the gate material and hence is the total effective width of the device (since the material $z > w_m$ is

unnecessary for device operation). Thus, γ is the ratio of insulator width to total effective width. Since g is equal to the ratio u/α, we see from Eq. 4.60 that $g(\theta)$ depends only on the parameter γ. Therefore, FET's differing both in conductivities and in insulator thicknesses can have identical properties. [For example, a unit with a ratio of gate doping to channel doping (β) equal to 10 and a ratio of insulator thickness to channel thickness ($1/\alpha - 1$) equal to 2.2 is identical in external characteristics to a unit with a doping ratio 1 and a thickness ratio 4.] Further, ordinary step-junction FET's ($\alpha = 1$) have $\gamma = 0$, regardless of β, and the normalized characteristics of these devices are independent of the relative doping levels. Delta-junction devices ($\alpha \to 0$) have $\gamma \to 1$, and again the results are independent of β. For values of α in the range $0 < \alpha < 1$, g (and hence I and C_{11}^*) depends on both α and β. The important special case of an insulated-gate FET with a metallic gate ($\beta \to \infty$) corresponds to $\gamma = 1 - \alpha$.

We now solve for the transconductance, the drain current, and the input charge-capacitance in the pinch-off range. Elimination of u/α from Eqs. 4.59 and 4.60 gives

$$g(\theta) = \frac{1}{1 - \gamma}\left[1 - \sqrt{1 - (1 - \theta)(1 - \gamma^2)}\right] \qquad (4.61)$$

as the normalized transconductance. In solving for I and C_{11}^* we may avoid performing any integration by substituting

$$\theta_1(\theta) = 1 - (1 - \theta)(1 - \gamma^2) \tag{4.62}$$

in Eq. 4.61. Then $\varepsilon(\theta_1)$ is proportional to the step-junction function $1 - \sqrt{\theta_1}$, and $d\theta_1$ is proportional to $d\theta$. Furthermore, $\theta_1(1) = 1$, and hence (from Eqs. 4.50 and 4.51) the dependence of I and C_{11}^* on $s_1 \equiv \theta_1(s)$ will be the same as the dependence of the step-junction I and C_{11}^* on s. That is,

$$I = D_1(1 - \sqrt{s_1})^2(1 + 2\sqrt{s_1})$$

$$C_{11}^* = D_2 \frac{1 - \sqrt{s_1}}{(1 + 2\sqrt{s_1})^2}$$

where D_1 and D_2 are constants. These constants may be eliminated by normalization. When $s = 0$, $s_1 = \gamma^2$ and $I/I_o = C_{11}^*/C_o^* = 1$. Therefore, bearing in mind that s_1 must be obtained in terms of s from Eq. 4.62, we have

$$\frac{I}{I_o} = \frac{(1 - \sqrt{s_1})^2(1 + 2\sqrt{s_1})}{(1 - \gamma)^2(1 + 2\gamma)} \tag{4.63}$$

$$\frac{C_{11}^*}{C_o^*} = \frac{(1 + 2\gamma)^2}{(1 + \gamma)} \frac{1 + \sqrt{s_1}}{(1 + 2\sqrt{s_1})^2} \tag{4.64}$$

as the desired results. Curves of g_m/G_o, I/I_o, and C_{11}^*/C_o^* versus s are drawn in Fig. 4.2 for a step-junction FET ($\gamma = 0$), for a delta-junction FET ($\gamma = 1$), and for an insulated-gate FET

with $\nu = 1/2$. The delta-junction device corresponds to the simple model of the insulated-gate FET that is usually analyzed (relatively narrow conducting channel). For this case the equations derived here are valid for forward as well as reverse gate biases.

The two examples presented here have one common property. In both cases the step- and delta-junction curves form rather close bounds on all other possible curves. The existence of these bounds might well be expected because in these examples intermediate results are produced by doping profiles that can be considered "intermediate" between step and delta junctions. The closeness of the bounds is of definite practical interest and is also of theoretical value because it provides motivation for the work of the next chapter.

4.4. Conclusions

The general treatment developed in this chapter has enabled us to write the equations describing the operation of an FET in a useful, compact form. All results are expressed in terms of an integral of the (arbitrary) impurity profile of the FET. This integral, which is equal to the normalized transconductance beyond pinch-off, is therefore of fundamental significance. The general formulation simplifies calculations because only the substitution of the impurity profile is required in order to obtain the final results, the intermediate steps having been eliminated. This is in contrast to the conventional treatment in which the impurity profile is specified at the first step, and in which the solution must be repeated from the beginning for different structures.

Furthermore, and perhaps more important, the general formulation allows us to draw general conclusions. We found that the properties of all FET's are qualitatively similar, and hence the observations made in Chapter II about step-junction FET's are true for all types of FET's. In particular, the drain characteristics always have the same overall shape, and the small-signal equivalent circuit always contains the same elements with the same physical interpretations and with analogous voltage dependences.

The examples of Section 4.3 emphasize these conclusions and adumbrate the analysis of the next chapter where, although still working with general, qualitative formulae, we obtain specific, quantitative results.

CHAPTER V

BOUNDS ON THE PARAMETERS OF AN ARBITRARILY DOPED FIELD-EFFECT TRANSISTOR

In this chapter we shall attain the goal toward which the previous analyses have been converging: a theoretical basis for the much-used empirical conclusion that widely different types of FET's exhibit similar functional behavior.

In the derivations of Section 4.2 no limitations were made on the impurity profile. The general results obtained in that section are valid for any FET. We now show that if the functional form of the doping profile of an FET is slightly restricted - but restricted in such a way that essentially all manufactured types are included in the theory - then the properties of the FET are remarkably insensitive to the particular profile used. Specifically, we prove that for any s in the pinch-off state ($d \geq 1$), the normalized transconductance, drain current, and input charge-capacitance, and the voltage-dependent parts of various figures of merit are each bounded by their step- and delta-junction values. We also find a simple formula, independent of the junction type, for the bias point necessary for zero temperature coefficient of the drain current. As a final topic we consider the power-law nature of the transfer characteristics and find limits on the exponent in the power law.

The results of this chapter have important practical implications, being applicable to a variety of problems pertaining to FET's.

5.1. Preliminary Derivations

In order to establish most of the results that will be obtained in this chapter we shall use two inequalities involving the derivatives of $g(s)$. We derive these inequalities in this section.

The relevant equations for s and g from Chapter IV are:

$$s = \frac{\int_u^w z\rho(z)\,dz}{\int_{w_m}^{u} z\rho(z)\,dz} = \frac{\int_u^w (1-z)\rho(z)\,dz}{\int_{w_m}^{u} (1-z)\rho(z)\,dz} \tag{5.1}$$

$$\int_u^w \rho(z)\,dz = \int_0^{w_m} \rho(z)\,dz = 0 \tag{5.2}$$

$$g = \frac{\int_0^u \rho(z)\,dz}{\int_0^1 \rho(z)\,dz} \tag{5.3}$$

Since the only values of u and w that are considered in this chapter are u_s and w_s, the subscript s is omitted. We bear in mind, however, that u and w always pertain to the source.

For convenience in the analysis of this section we shall assume an n-channel device. Thus,

$$\rho(z) \begin{cases} \geq C \ , & 0 \leq z < 1 \\ = C \ , & z = 1 \\ \leq C \ , & 1 < z \end{cases}$$

The final results are in no way dependent on this assumption.

The first derivative of g with respect to s is

$$\frac{dg}{ds} = \frac{dg}{du}\frac{du}{ds} = \frac{\rho(u)}{\int_0^1 \rho \, dz} \frac{\int_0^{w_m} (1 - z)\rho \, dz}{(1 - w)\rho(w)\frac{dw}{du} - (1 - u)\rho(u)}$$

From Eq. 5.2 we have

$$\rho(w)\frac{dw}{du} - \rho(u) = 0 \tag{5.4}$$

so that

$$g' = - \frac{\int_0^{w_m} (1 - z)\rho \, dz}{\int_0^1 \rho \, dz} \frac{1}{w - u} \tag{5.5}$$

where we have used a prime to denote differentiation with respect to s. The integrands in both numerator and denominator are non-negative for n-channel devices (non-positive for p-channel devices), and hence $dg/ds < 0$. Differentiation of Eq. 5.5 and use of Eq. 5.4 give

$$g'' = \frac{\left[\int_0^{w_m} (1 - z)\rho \, dz\right]^2}{\int_0^1 \rho \, dz} \frac{1}{(w - u)^3} \frac{\rho(u) - \rho(w)}{-\rho(u)\rho(w)} \qquad (5.6)$$

Since $w \geq \bar{z} \geq u$, each fraction in the above expression is strictly positive. (For p-type channels both the denominator of the first fraction and the numerator of the third fraction reverse sign.) Thus, for all doping profiles,

$$g'' > 0 \qquad (5.7)$$

This is the first of the two preliminary inequalities.

The second inequality is

$$-g' - 2sg'' \geq 0 \qquad (5.8)$$

Combination of Eqs. 5.1, 5.5, and 5.6 reveals that this relation is satisfied if

$$H(u) \equiv (w - u)^2 \frac{\rho(w)\rho(u)}{\rho(w) - \rho(u)} - \int_u^w 2(1 - z)\rho \, dz \geq 0 \qquad (5.9)$$

for n-channel devices. The ratio $\rho(u)/\rho(w)$ is always negative (or zero), hence

$$0 \leq \frac{\rho(w)}{\rho(w) - \rho(u)} = \frac{1}{1 - \rho(u)/\rho(w)} \leq 1$$

and since $w \to 1$ as $u \to 1$, we have

$$H(1) = 0$$

Thus, if $dH/du \leq 0$ for $0 \leq u \leq 1$, then $H(u) \geq 0$, and, since the range $0 \leq u \leq 1$ is equivalent to the range $0 \leq s \leq 1$, Eq. 5.8 will be established. After some cancellation, the required derivative becomes

$$\frac{dH}{du} = (w - u)^2 \frac{d}{du}\left[\frac{\rho(w)\rho(u)}{\rho(w) - \rho(u)}\right]$$

which is negative (or zero), provided that

$$\frac{d}{du}\left[\frac{1}{\rho(u)} - \frac{1}{\rho(w)}\right] \geq 0 \qquad (5.10a)$$

or (since $du/ds < 0$), provided that

$$\frac{d}{ds}\left[\frac{1}{\rho(u)} - \frac{1}{\rho(w)}\right] \leq 0 \qquad (5.10b)$$

(Note that $\rho(w) < 0$.) This expression is equivalent to the condition

$$-\frac{1}{\rho^3(u)}\left.\frac{d\rho}{dz}\right|_{z=u}+\frac{1}{\rho^3(w)}\left.\frac{d\rho}{dz}\right|_{z=w}\geq 0 \qquad (5.11)$$

which is useful if a particular pair of values for u and w is known. Therefore, inequality 5.8 will be satisfied if

$$\frac{1}{\rho(u)}-\frac{1}{\rho(w)}$$

is an increasing function* of u or, alternatively, a decreasing function of s. For p-channel FET's the contrary applies. Although this condition is sufficient, and by no means necessary, we now show that it is fulfilled by substantially all commercial profiles.

Clearly, monotonic $\rho(z)$ satisfy Eq. 5.10. Most manufactured types (alloy, epitaxial, single-diffused, and insulated-gate FET's) fall in this category. The only practically important FET's that employ non-monotonic doping profiles have double-diffused structures (with two gates).** A sketch of an n-channel double-diffused profile is given in Fig. 5.1. The plane where the two gate-channel depletion regions would meet if $W_g = W_o$ were applied is chosen as $z = 0$. (We assume that the two gates are tied together electrically.) If, for analysis purposes, the double-diffused device is split along the $z = 0$ plane, then it is easy to show that the $g(s)$ of the original

---

* The function need not be strictly increasing; i.e. a zero derivative is permissible.

** Motorola units, for example, have two gates, but are symmetrical and hence may be considered as two identical, single-gate, monotonically doped FET's in parallel.

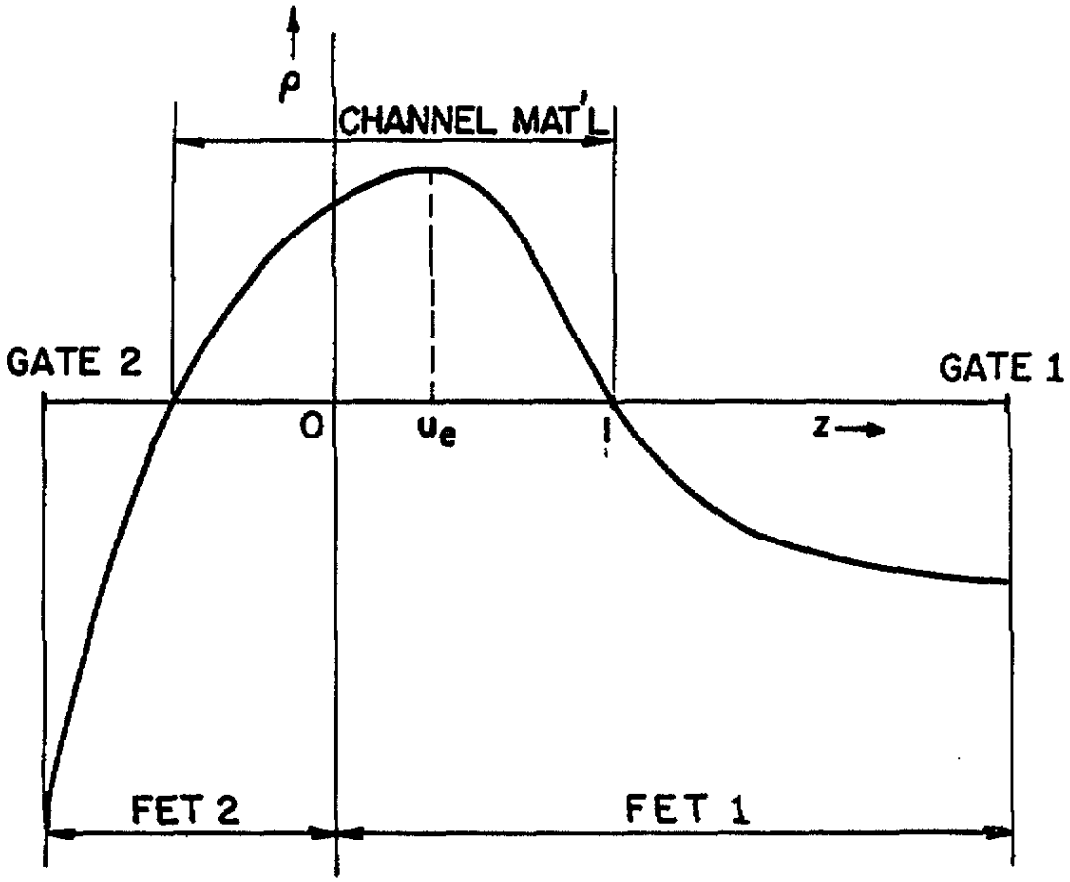


Fig. 5.1. Sketch of a double-diffused doping profile.

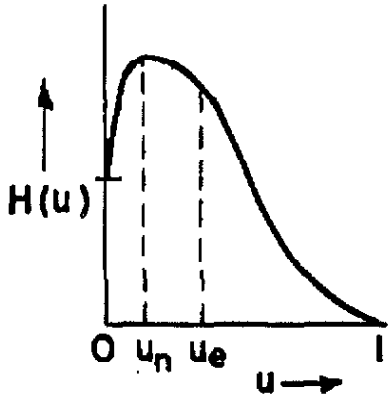


Fig. 5.2. Possible H(u) for FET #1 of Fig. 5.1.

FET is equal to a linear combination of the $[g_i(s)]$'s of the two smaller FET's; i.e. $g = \zeta g_1 + (1-\zeta)g_2$, where $0 \le \zeta \le 1$, and the subscripts 1 and 2 denote the two single-gate units indicated in Fig. 5.1. Double-diffused doping profiles have only one extremum (near $z = 0$ in the channel). Denote the location of this extremum by $z = u_e$, and, without loss of generality, assume that $u_e \ge 0$ and that $\rho(1) = 0$. Then FET #2 has a monotonic profile, and g_2 satisfies Eq. 5.8. For $z \ge u_e$ the profile of FET #1 is monotonic, and hence for this unit, $H(u)$, the function defined in Eq. 5.9, has a negative slope over the interval $u_n < u \le 1$ say, where $dH(u_n)/du = 0$. We know that $u_n < u_e$ because Eq. 5.10a implies a condition on ρ that is weaker than monotonicity. If $u_n \le 0$, then FET #1 satisfies 5.10 for its entire width, and g_1 - and also, therefore, the overall function g - satisfies 5.8. If $u_n > 0$, then $dH/du > 0$ for $0 \le u < u_n$. However, $dH/du < 0$ is only a sufficient condition - not a necessary condition - for the establishment of Eq. 5.8. Figure 5.2 shows a possible $H(u)$ for FET #1. Unless the slope of H is rather steep for $0 \le u < u_n$, H will be positive throughout the interval $0 \le u \le 1$. Furthermore, even if $H(u) < 0$ near $u = 0$, Eq. 5.8 may still be satisfied for the overall g- function because we merely need

$$\zeta\left[-g_1' - 2sg_1''\right] \ge -(1-\zeta)\left[-g_2' - 2sg_2''\right]$$

and we have already seen that the right-hand side of this inequality

is negative (or zero). Although the above arguments are merely qualitative, with reasonable certainty we may conclude that double-diffused devices obey Eq. 5.8. Thus, essentially all manufactured FET's are included in the following "restricted" theory.

To recapitulate the results of this section, we rewrite the two important inequalities with their relation to the impurity profile:

$$g'' > 0 \quad , \qquad \text{all } \rho \qquad (5.12)$$

$$-g' - 2sg'' \geq 0 \quad , \qquad \text{restricted } \rho \qquad (5.13)$$

These equations are applicable to both n- and p-channel FET's.

5.2. Transconductance

We now apply the inequalities of the previous section to prove that the normalized transconductance beyond pinch-off, $g(s)$, is bounded by the step- and delta-junction formulae

$$(1 - \sqrt{s}) \leq g(s) \leq (1 - s) \qquad (5.14)$$

The upper bound follows directly from Eq. 5.12 and hence is valid for all doping profiles. The inequality $g'' > 0$ means that $g(s)$ is concave upward. But the values $g(0) = 1$, $g(1) = 0$ are fixed. Therefore, $g(s)$ must lie below the straight line $(1 - s)$; that is,

$$g(s) \leq 1 - s$$

Actually, the equality holds only at $s = 0$ and at $s = 1$.

The lower bound derives from Eq. 5.13. If that inequality holds, then

$$\int_0^s [-g'(t) - 2tg''(t)]\, dt \geq 0$$

$$g(t) - 2tg'(t) \Big]_0^s \geq 0$$

From Eqs. 5.1 and 5.5 we see that

$$sg'(s) \propto \frac{1}{w - u} \int_u^w (1 - z)\rho(z)\, dz$$

The right-hand side of this proportionality is equal to the average value of $(1 - z)\rho(z)$ over the interval $u \leq z \leq w$. As $s \to 0$, this average value approaches zero, hence $\lim_{t \to 0} tg'(t) = 0$, and

$$g(s) - 2sg'(s) - 1 \geq 0 \qquad (5.15)$$

or

$$0 \leq s^{-3/2} g(s) - 2s^{-1/2} g'(s) - s^{-3/2}$$

$$= 2\left[-s^{-1/2} g(s) + s^{-1/2}\right]'$$

The intermediate inequality 5.15 is weaker than the initial inequality 5.13 (because the latter implies the former). However, Eq. 5.13 is convenient in later proofs, and since it holds for essentially all profiles of interest, we have used it as our starting point. Integration of the last expression from s to 1 yields

$$2\left[1 + s^{-1/2}g(s) - s^{-1/2}\right] \geq 0$$

or

$$g(s) \geq 1 - \sqrt{s} \tag{5.16}$$

Thus, the normalized transconductance for most practical FET's lies between the values for step- and delta-junction FET's. These bounds are illustrated in Fig. 4.2. The bounds that we shall subsequently obtain on other quantities are dependent only on these bounds on g and on inequalities 5.12 and 5.13 on the derivatives of g. Since we can obtain real junctions as close as we like to the ideal step and delta junctions, tighter bounds on $g(s)$ cannot be obtained without further restrictions on $\rho(z)$. However, these further restrictions would necessarily eliminate some practical profiles from the theory.

5.3. Drain Current

In this section we show that the normalized drain current satisfies

$$(1 - \sqrt{s})^2(1 + 2\sqrt{s}) \leq I(s)/I_o \leq (1 - s)^2 \qquad (5.17)$$

the upper bound being the delta-junction expression, and the lower bound the step-junction expression.

We first prove that if $g < (1 - s)$ for $0 < s < 1$, then $I/I_o \leq (1 - s)^2$. [If $g = (1 - s)$, the right-hand equality in Eq. 5.17 obtains.] From Eq. 4.50 we have

$$\frac{I}{I_o} = \frac{\int_s^1 g(t)\,dt}{\int_0^1 g(t)\,dt} = 1 - \frac{\int_0^s g\,dt}{\int_0^1 g\,dt}$$

We must show that

$$1 - \frac{\int_0^s g\,dt}{\int_0^1 g\,dt} \leq (1 - s)^2 = 1 - 2s + s^2$$

or that

$$\psi(s) \equiv \int_0^s g\,dt - \frac{\int_0^1 g\,dt}{(1/2)}\left(s - \frac{s^2}{2}\right) \geq 0 \qquad (5.18)$$

But $g < (1 - s)$ implies

$$\int_0^1 g\,dt < \int_0^1 (1 - t)\,dt = \frac{1}{2}$$

Let

$$\frac{\int_0^1 g\,dt}{(1/2)} = \lambda < 1$$

so that

$$\psi(s) = \int_0^s g\,dt - \lambda\left(s - \frac{s^2}{2}\right)$$

and

$$\psi'(s) = g(s) - \lambda(1 - s)$$

Now, $\psi(0) = \psi(1) = 0$ and $\psi'(0) = (1 - \lambda) > 0$. Thus, $\psi(s)$ is zero and has a positive slope at $s = 0$, and is zero at $s = 1$. Hence, if $\psi'(s)$ has only one zero in the range $0 < s < 1$, then $\psi(s)$ can never pass through zero and will always be positive (since ψ is continuous). This statement and the following arguments are elucidated by the curves sketched in Fig. 5.3. In order that $\psi' = 0$ we must have $g(s) = \lambda(1 - s)$. But $\lambda(1 - s)$ is a straight line, and $g'' > 0$, i.e. $g(s)$ is concave upward. Therefore, a straight line can intersect $g(s)$ at only two points. Both $g(s)$ and $\lambda(1 - s)$ are zero at $s = 1$, and since $\lambda < 1$, the remaining

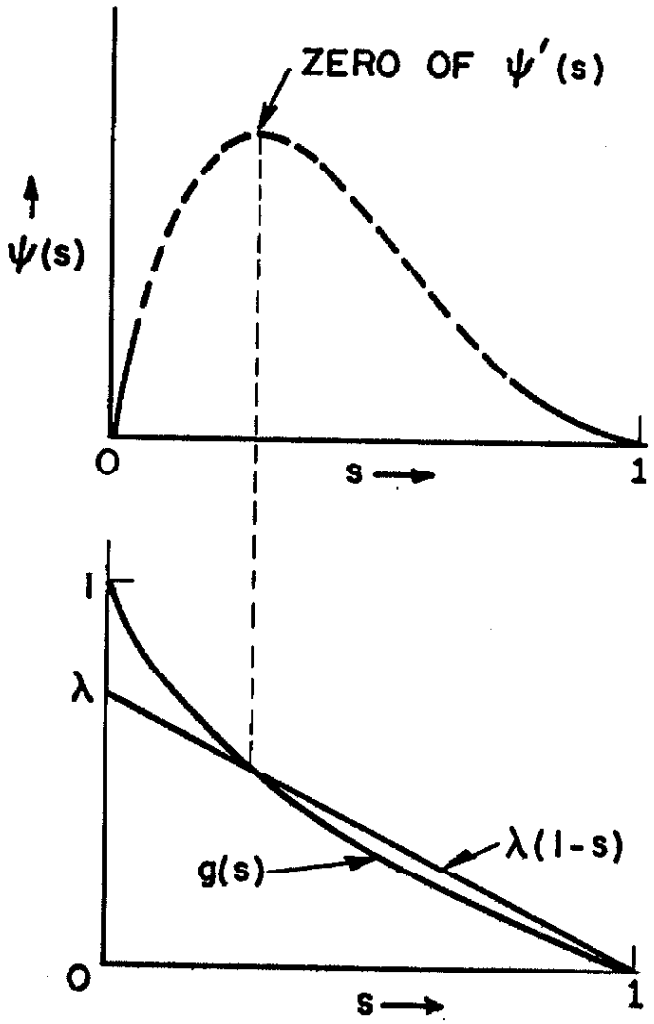


Fig. 5.3. Pertinent to the proof of Eq. 5.18.

point of intersection lies within the range $0 < s < 1$. Thus $\psi'(s)$ does have one and only one zero, and Eq. 5.18 is proved.

To complete the proof of Eq. 5.17 we now show that if $g > (1 - \sqrt{s})$ for $0 < s < 1$, and if $(-g' - 2sg'') \geq 0$, then

$$\frac{I}{I_o} \geq (1 - \sqrt{s})^2(1 + 2\sqrt{s}) = 1 - 3s + 2s^{3/2}$$

[Again, if $g = (1 - \sqrt{s})$, the equality holds.] The method of proof is similar to that for the upper bound. We must show that

$$\psi_1(s) \equiv \frac{\int_0^1 g\, dt}{(1/3)} \left(s - \frac{2}{3} s^{3/2}\right) - \int_0^s g\, dt \geq 0 \qquad (5.19)$$

Let

$$\frac{\int_0^1 g\, dt}{(1/3)} = \Lambda$$

Since $g > (1 - \sqrt{s})$,

$$\int_0^1 g\, dt > \int_0^1 (1 - \sqrt{s})\, dt = \frac{1}{3}$$

and $\Lambda > 1$. Further,

$$\psi_1'(s) = \Lambda\,(1 - \sqrt{s}) - g(s)$$

so that, as in the previous proof, $\psi_1(0) = \psi_1(1) = 0$ and $\psi_1'(0) = (\Lambda - 1) > 0$. Thus, if $\psi_1'(s)$ has only one zero in the range $0 < s < 1$, then $\psi_1(s) > 0$. To show that ψ_1' does indeed have only one zero, we change variables from s to $s_1 = \sqrt{s}$ $(0 \le s_1 \le 1)$ and show that $\psi_1'(s_1) = \Lambda(1 - s_1) - g(s_1^2)$ has only one zero. By analogy with the previous proof, we need only show that $g(s_1^2)$ is concave downward (since $\Lambda > 1$), or that

$$\frac{d^2g}{ds_1^2} \le 0$$

But

$$\frac{d^2g}{ds_1^2} = \frac{d}{ds_1}\left[2s_1 \frac{dg}{ds}\right]$$

$$= 2\frac{dg}{ds} + 4s_1^2 \frac{d^2g}{ds^2} = 2(g' + 2sg'')$$

$$\le 0$$

by hypothesis. This completes the proof of Eq. 5.19 and establishes Eq. 5.17.

Thus, practically all manufactured FET's have normalized transfer characteristics that lie between step-junction and delta-junction characteristics. The boundary curves, pictured in Fig. 4.2, are especially close, and in fact are almost indistinguishable experimentally.

5.4. Input Charge-Capacitance

Again using the inequalities of Section 5.1, we can show that the normalized input charge-capacitance lies within the range

$$\frac{1+\sqrt{s}}{(1-2\sqrt{s})^2} \leq \frac{C_{11}^*(s)}{C_c^*} \leq 1 \qquad (5.20)$$

The proof of this statement is deferred to Appendix A because a result from Section 5.5. is needed. The bounds on C_{11}^* are the step- and delta-junction formulae and are illustrated in Fig. 4.2. Because of the presence of a finite built-in potential V_c, C_o^* cannot be measured directly. Therefore, experimental points will not occupy some small interval of $0 \leq s \leq 1$ near $s = 0$. Since the step-junction curve (lower bound) has a very steep slope near $s = 0$, experimental curves will actually lie between closer bounds than the theoretical bounds of Eq. 5.20 and Fig. 4.2. (Similar remarks apply to the bounds on $g(s)$, but the effect of V_c is not so strong in this case.)

5.5. Figures of Merit

In this section we shall define and derive bounds for various figures of merit for FET's. The bounds for each figure of merit are the expressions for step- and delta-junction doping profiles, and, as in Sections 5.2 - 5.4, the bounds are close enough that the detailed nature of the impurity profile is unimportant. The parameters of importance are μ, W_o, L, and G_o.

The figures of merit that we shall consider are based on the simple, small-signal equivalent circuit beyond pinch-off (Fig. 2.9; g_m and C_{11}^* are the only elements present) and are

$$f_1 = G_o \left[g(s)\right] \tag{5.21}$$

$$f_2 = \frac{g_m(s)}{I(s)} = \frac{1}{W_o}\left[\frac{g(s)}{\int_s^1 g(t)\,dt}\right] \tag{5.22}$$

$$f_3 = \frac{1}{\tau_t} = \frac{\mu W_o}{L^2}\left[\frac{\left\{\int_s^1 g(t)\,dt\right\}^2}{\int_s^1 \{g(t)\}^2\,dt}\right] \tag{5.23}$$

$$f_4 = \frac{g_m(s)}{C_{11}^*(s)} = \frac{\mu W_o}{L^2}\left[\frac{\left\{\int_s^1 g(t)\,dt\right\}^2}{\int_s^1 g(t)\{g(s) - g(t)\}\,dt}\right] \tag{5.24}$$

To a good approximation the output admittance vanishes beyond pinch-off, so that f_1 is the gain of an FET and f_4 is the gain-bandwidth product. The ratio of the gain to the steady-state current necessary for that gain is specified by f_2. As a rather general figure of merit, f_3 gives an estimate of the maximum frequency of operation of the device [19]. In each of the figures of merit f_i, the square brackets contain a dimensionless, voltage-

dependent term which we denote by F_i. We now assume that inequality 5.13 holds [i.e. $(-g' - 2sg'') \geq 0$], and we establish the step- and delta-junction bounds for each f_i.

$\underline{F_1}$

Since $F_1 = g(s)$, we merely refer to Eq. 5.14 to assert that

$$(1 - \sqrt{s}) \leq F_1 \leq (1 - s) \tag{5.25}$$

Although $g = (1 - s)$ gives the optimum (i.e. the maximum) F_1, this particular g yields at most a factor of 2 improvement over any other g.

$\underline{F_2}$

Before demonstrating the desired results, we must prove two lemmas. We first show that if $(-g' - 2sg'') \geq 0$, then

$$g(t) \geq g(s)\frac{1 - \sqrt{t}}{1 - \sqrt{s}} \qquad \text{for } s \leq t \leq 1 \tag{5.26}$$

Equation 5.26 means that over the range $s \leq t \leq 1$ the transconductance function $g(t)$ will be greater than or equal to a function of the form $g \sim (1 - \sqrt{t})$, provided that both functions have the value $g(s)$ at $t = s$. If $s = 0$, Eq. 5.26 reduces to the lower bound for $g(s)$ (Eq. 5.16). Let us define

$$\psi_2(t) \equiv g(t) - \frac{g(s)}{1 - \sqrt{s}} (1 - \sqrt{t})$$

Then

$$\dot{\psi}_2(t) = \dot{g}(t) + \frac{1}{2\sqrt{t}} \frac{g(s)}{(1 - \sqrt{s})}$$

and

$$\ddot{\psi}_2(t) = \ddot{g}(t) - \frac{1}{4t^{3/2}} \frac{g(s)}{1 - \sqrt{s}}$$

where the dots denote differentiation with respect to t. Now, if $\dot{\psi}_2(t) = 0$, then

$$\frac{g(s)}{1 - \sqrt{s}} = -2\sqrt{t}\ \dot{g}(t)$$

and

$$\ddot{\psi}_2(t) = \frac{1}{2t} \left[2t\ddot{g}(t) + \dot{g}(t)\right] \leq 0$$

by hypothesis. Thus, any stationary point of $\psi_2(t)$ is either a relative maximum or a point of inflection. But $\psi_2(1) = \psi_2(s) = 0$. Hence, $\psi_2(t)$ is always positive over the range $s \leq t \leq 1$, and the lemma is proved.

The second lemma states that $g'' > 0$ implies

$$g(t) \leq g(s)\frac{1 - t}{1 - s} \qquad \text{for} \qquad s \leq t \leq 1 \qquad (5.27)$$

Because of the close similarity between this and the previous lemma,

its interpretation and its proof are omitted.

Substitution of Eqs. 5.26 and 5.27 into

$$F_2 = \frac{g(s)}{\int_s^1 g(t)\,dt}$$

yields directly the desired bounds

$$\frac{2}{1 - s} \leq F_2 \leq \frac{3}{(1 - \sqrt{s})(1 + 2\sqrt{s})} \tag{5.28}$$

In this case $g = (1 - \sqrt{s})$ gives the optimum figure of merit, but the largest and smallest values of F_2 differ at most by a factor of 1.5.

F_3

The voltage-dependent part of f_3 is

$$F_3 = \frac{\left(\int_s^1 g\,dt\right)^2}{\int_s^1 g^2\,dt}$$

and has the bounds

$$\frac{2}{3}\,\frac{(1 - \sqrt{s})(1 + 2\sqrt{s})^2}{(1 + 3\sqrt{s})} \leq F_3 \leq \frac{3}{4}\,(1 - s) \tag{5.29}$$

corresponding to a step and delta junction, respectively.

In order to prove the upper bound we must show that

$$\psi_3(s) \equiv 3(1 - s)\int_s^1 g^2\,dt - 4\left(\int_s^1 g\,dt\right)^2 \geq 0 \qquad (5.30)$$

Since $\psi_3(1) = 0$, it suffices to prove that

$$\psi_3'(s) = -3(1 - s)g^2(s) + \int_s^1 g(t)[8g(s) - 3g(t)]\,dt \leq 0 \qquad (5.31)$$

for $0 \leq s < 1$. Consider the integrand,

$$\psi_4(t) \equiv g(t)[8g(s) - 3g(t)]$$

appearing in Eq. 5.31. For a given $g(s)$ and a particular t, ψ_4 will have a relative maximum when

$$g(t) = \frac{4}{3}\,g(s)$$

But Eq. 5.27 indicates that

$$g(t) \leq g(s)\frac{1 - t}{1 - s} \leq g(s)$$

Thus, we can never maximize the integrand ψ_4 (for any $s \leq t \leq 1$), but we come closest to maximization if at each t we choose $g(t)$ as large as possible, that is, if we choose

$$g(t) = g(s)\frac{1 - t}{1 - s}$$

Hence,

$$\psi_3'(s) \le -3(1-s)g^2(s) + \int_s^1 g(s)\frac{1-t}{1-s}\left[8g(s) - 3g(s)\frac{1-t}{1-s}\right] dt$$

$$= 0$$

as was to be shown.

The proof of the lower bound proceeds along exactly the same lines, except that Eq. 5.26 is used in place of Eq. 5.27.

The two bounds in Eq. 5.29 are very close, differing at most by a factor of 1.125.

$\underline{F_4}$

We shall prove that

$$F_4 = \frac{\left(\int_s^1 g\ dt\right)^2}{\int_s^1 g(t)[g(s) - g(t)]\ dt}$$

is bounded by the step- and delta-junction expressions

$$\frac{2}{3}\frac{(1-\sqrt{s})(1+2\sqrt{s})^2}{(1+\sqrt{s})} \le F_4 \le \frac{3}{2}(1-s) \qquad (5.32)$$

Subtracting the reciprocal of the upper bound from $1/F_4$, and discarding the positive common denominator, we get

$$3(1 - s)\int_s^1 g(t)[g(s) - g(t)]\,dt - 2\left(\int_s^1 g\,dt\right)^2$$

as the quantity which we must show to be positive. If we substitute Eq. 5.30 and remove the factor of $(1 - s)/2 \geq 0$, then we need only show that the remaining factor,

$$\psi_5(s) \equiv \int_s^1 g(t)[2g(s) - 3g(t)]\,dt$$

is always positive (or zero). Now,

$$\psi_5'(s) = 2g'(s)\int_s^1 g\,dt + g^2(s)$$

and

$$\psi_5''(s) = 2g''(s)\int_s^1 g\,dt \geq 0$$

since $g'' > 0$. But $\psi_5''(1) = 0$, so that $[\psi_5'(s)]' \geq 0$ implies that $\psi_5'(s) \leq 0$ for $0 \leq s \leq 1$. Similarly, $\psi_5(1) = 0$, so that $\psi_5'(s) \leq 0$ implies that $\psi_5(s) \geq 0$, which establishes the upper bound.

The proof of the lower bound is algebraically a little more complicated but uses essentially the same approach. We must show that

$$\frac{3}{2}\frac{(1 + \sqrt{s})}{(1 - \sqrt{s})(1 + 2\sqrt{s})^2} - \frac{1}{F_4} \geq 0 \qquad (5.33)$$

or, equivalently, that

$$\psi_6(s) \equiv \frac{3(1 + \sqrt{s})}{(1 - \sqrt{s})(1 + 2\sqrt{s})^2}\left[\int_s^1 g\,dt\right]^2 -$$

$$-2\int_s^1 g(t)[g(s) - g(t)]dt \geq 0$$

Now, $\psi_6(1) = 0$, so if $\psi_6'(s) \leq 0$ for $0 \leq s \leq 1$, $\psi_6(s) \geq 0$. It suffices to demonstrate that

$$\psi_7(s) \equiv \frac{\sqrt{s}(1 - \sqrt{s})^2}{\int_s^1 g\,dt}\,\psi_6'(s) \leq 0$$

Performing the required calculations, we get

$$\psi_7(s) = \frac{3(-1 + 2\sqrt{s} + 2s)}{(1 + 2\sqrt{s})^3}\int_s^1 g\,dt - 6g(s)\frac{\sqrt{s}(1 - s)}{(1 + 2\sqrt{s})^2} +$$

$$+ 2\,[-\sqrt{s}g'(s)](1 - \sqrt{s})^2$$

Again, $\psi_7(1) = 0$, and hence $\psi_7(s) \leq 0$ if $\psi_7'(s) \geq 0$. From the above formula we get

$$\psi_7'(s) = \frac{6(1 - \sqrt{s})(2 + \sqrt{s})}{\sqrt{s}(1 - 2\sqrt{s})^4} \int_s^1 g\, dt - 3g(s) \frac{(1 - 3\sqrt{s} - s)}{\sqrt{s}(1 + 2\sqrt{s})^3} \cdots$$

$$+ 2g'(s) \frac{(1 - \sqrt{s})(1 + \sqrt{s} + s)}{(1 + 2\sqrt{s})^2} +$$

$$+ 2[-\sqrt{s}g'(s)]'(1 - \sqrt{s})^2 \qquad (5.34)$$

But

$$[-\sqrt{s}g']' = \frac{1}{2\sqrt{s}}[-2sg'' - g'] \geq 0 \qquad (5.35)$$

by hypothesis (Eq. 5.13). Also, Eq. 5.35 implies that

$$\int_s^1 \frac{1 - \sqrt{t}}{\sqrt{t}} [-2t\ddot{g}(t) - \dot{g}(t)]\, dt \geq 0$$

Integration gives

$$g(s) + 2\sqrt{s}(1 - \sqrt{s})g'(s) \geq 0$$

or,

$$2g'(s) \geq - \frac{g(s)}{\sqrt{s}(1 - \sqrt{s})} \qquad (5.36)$$

Further,

$$\int_s^1 g\,dt \geq \int_s^1 g(s)\frac{1-\sqrt{t}}{1-\sqrt{s}}\,dt$$

$$= \frac{1}{3}\,g(s)(1-\sqrt{s})(1+2\sqrt{s}) \qquad (5.37)$$

where use has been made of Eq. 5.26. Thus, since the coefficients of $\int_s^1 g\,dt$, $g'(s)$, and $[-\sqrt{s}g'(s)]'$ in Eq. 5.34 are all positive, we may substitute Eqs. 5.35, 5.36, and 5.37 to get

$$\psi_7'(s) \geq \frac{g(s)}{\sqrt{s}(1+2\sqrt{s})^3}\Big[2(1-\sqrt{s})^2(2-\sqrt{s}) - 3(1-3\sqrt{s}-s) - (1-\sqrt{s}+s)(1+2\sqrt{s})\Big]$$

$$= 0$$

thereby proving Eq. 5.33. As with F_1, F_2, and F_3, the two bounds on F_4 are close, their ratio at most equaling 2.25.

The existence of the close bounds on the F_i's indicates that the doping profile may sensibly be neglected in the optimization of a particular figure of merit f_i; only the parameters μ, W_o, L, and G_o need be considered.* Examination of the constant

---

* Another figure of merit is the reciprocal of the input charge-capacitance. The above conclusion applies to this quantity, whose bounds have been discussed in Section 5.4.

multiplicative factors in Eqs. 5.21 - 5.24 for $f_1 - f_4$ reveals that in general μ and G_o should be maximized and L should be minimized. (On the other hand, optimization of $1/C_{11}^*$ requires a small G_o.) Also, except for f_2, best results occur if W_o is large. Moreover, since all f_i (except f_2) are maximum at $s = 0$, the internal bias corresponding to zero external bias should be as small as possible; i.e. V_c/W_o should be small, or W_o should be large. These remarks must be qualified because other phenomena have to be taken into account in the design of an FET. Two such phenomena, whose effects we now consider briefly, are the maximum electric field in the channel and the power dissipation.

An increase in W_o/L will increase the electric field in the channel and hence will decrease the transit time. Although the frequency response will improve accordingly, Dacey and Ross [8] have shown that this improvement is accomplished at the expense of the power dissipation. They conclude that it is not worthwhile to increase W_o/L above the critical-field value, beyond which $\mu \sim E^{-1/2}$.

The power dissipated in an FET is equal to the drain current times the drain voltage. This product is of the order of the zero-bias drain current I_o times the pinch-off voltage W_c. The area available for dissipation is LA. Since $I_o/W_oG_o \sim 1$ for all impurity profiles, the power dissipation per unit area is of the order of

$$\frac{I(s)|V_d|}{LA} \sim \frac{I_c W_o}{LA} \sim \frac{W_o^2 G_o}{LA}$$

which sets upper limits on W_o and G_o and lower limits on L and A.

Thus, although the detailed nature of the doping profile $\rho(z)$ may be disregarded in the design of an FET, compromises must be reached in the choice of the other parameters.

5.6. Temperature Compensation

At a particular bias point, the drain current of an FET will vary with temperature because of the temperature dependences of the built-in potential and the channel conductivity. Fortunately, these two dependences have opposite effects, and it is possible in most transistors to choose a bias point such that the temperature coefficient of the drain current is zero [21], [22]. Below, we first show that this bias point is practically independent of the type of FET junction, depending only on the external pinch-off voltage and the type of channel semiconductor. Then we consider the delta-junction FET in more detail in order to obtain an estimate of the equivalent input drift.

Using reasoning similar to that employed in [21], we calculate the change δI in the drain current caused by a small change δT in the temperature. As the temperature increases at constant external gate-source voltage, the magnitude of the built-in potential (and hence the magnitude of the internal gate potential) decreases, and so

the drain current increases. On the other hand, the decrease in channel conductivity with increasing temperature tends to decrease the current. In the temperature range of interest, the variation of the conductivity may be taken to arise solely from the temperature dependence of the mobility. Thus, at constant V_g

$$\delta I(s) = \left[-g_m(s) \frac{\partial |V_c|}{\partial T} + \left(\frac{1}{\mu} \frac{\partial \mu}{\partial T} \right) I(s) \right] \delta T \qquad (5.38)$$

From this equation we see that $\delta I / \delta T = 0$ at $s = s_{oo}$ if

$$\frac{I(s_{oo})}{g_m(s_{oo})} = \mu \left(\frac{\partial \mu}{\partial T} \right)^{-1} \left(\frac{\partial |V_c|}{\partial T} \right)$$

or, since $\mu \sim T^{-m}$, if

$$\frac{I(s_{oo})}{g_m(s_{oo})} = \frac{T}{m} |\xi| \qquad (5.39)$$

where

$$\xi(T) \equiv - \frac{\partial V_c}{\partial T} \qquad (5.40)$$

$$\approx (\pm)\ 2.0 \times 10^{-3}\ \text{v/}^{\circ}\text{C} \quad \text{for} \left\{ \begin{matrix} \text{p-} \\ \text{n-} \end{matrix} \right\} \text{channel silicon IGFET's at room temperature}$$

For silicon, $m \approx 2.5$ [23], so that $I(s_{oo})/g_m(s_{oo}) \approx 0.25\text{v}$. (The experimental values cited in [21] give $I(s_{oo})/g_m(s_{oo}) = 0.31\text{v}$.)

Now we employ the bounds on $f_2 = g_m/I$ obtained in the previous section to get*

$$\frac{1}{16}\left(1 + \sqrt{9 - 24\,\frac{T|\xi|}{mW_o}}\right)^2 \leq \left(1 - 2\,\frac{T|\xi|}{mW_o}\right) \qquad (5.41)$$

Since W_o is the internal pinch-off voltage (i.e. V_c is included), and since $V_c \gtrsim 0.5v$ in silicon, serviceable FET's have $W_o > 1v$, which implies that $T|\xi|/mW_o \lesssim 0.25$. But for $T|\xi|/mW_o < 0.27$ the two bounds in Eq. 5.41 are within ten percent of each other, so that for experimental purposes, s_{oo} may be assumed independent of the junction type, and for further derivations only the analytically simpler delta-junction bound need be considered.

Despite this simplification, the form of Eq. 5.41 is not too useful because both sides of the right-hand inequality are functions of T. We may eliminate this drawback by solving for the external bias voltage V_{goo} corresponding to the internal normalized voltage s_{oo}. Since $W_o = |V_c + V_p|$ and $s = (V_c + V_g)/(V_c + V_p)$, the zero-temperature-coefficient bias is

$$V_{goo} = V_p - 2\,\frac{T\xi}{m} \qquad (5.42)$$

---

* Actually, if $(8/24) \leq T|\xi|/mW_o \leq (9/24)$, the step-junction limit is double-valued, and, therefore, the zero-coefficient bias point can lie in the range $0 \leq s_{oo} \leq (1/16)(1 - \sqrt{9 - 24T|\xi|/mW_o})^2$. However, as shown below, practical values of $T|\xi|/mW_o$ are smaller than $8/24$, so that only the range given in Eq. 5.41 is of importance.

Although the right-hand side of this equation depends on T, this dependence is very weak since the variations in V_p and in $T\xi$ tend to cancel. Thus, over an appreciable temperature range, V_{goo} may be considered constant; for best accuracy, however, its value should be computed at the design-center temperature T_o:

$$V_{goo} = V_p(T_o) - 2\frac{T_o\xi(T_o)}{m} \qquad (5.43)$$

This formula, as well as the other formulae in this section, is valid for both p- and n-channel devices provided that the appropriate sign of ξ is used. Since V_p cannot be determined with great accuracy, Eq. 5.43 may in practice be applied to all types of doping profiles.*

In addition to knowing the bias point for zero temperature coefficient of drain current, we should also be aware of the consequences of biasing at voltages other than this optimum. For this computation we employ the concept of equivalent input drift - the variation in applied gate voltage V_g that would produce the same effect on the drain current as a given change in temperature. To include this hypothetical variation of input voltage in Eq. 5.38, V_g may be added to V_c in the partial derivative multiplying g_m. We may obtain the end result more directly, however, by noting that (for a delta-junction FET)

---

* For $V_{goo} \approx 0$, the percentage error in V_{goo} from Eq. 5.43 may be considerable. The absolute error, however, will be small.

$$I \sim \mu(T)[1 - s(T)]^2$$

$$\sim T^{-m}[1 - s(T)]^2$$

all other quantities in the formula for I being independent of T. Thus, for temperature-independent current we must have

$$[1 - s(T)] \sim T^{m/2}$$

or

$$\frac{1 - s(T)}{1 - s_o} = \left(\frac{T}{T_o}\right)^{m/2}$$

in which the subscript o refers s to the design-center temperature T_o. Substitution for s in terms of external voltages gives

$$V_g(T) = (V_p + V_c) - V_c(T) - (V_{po} - V_{go})\left(\frac{T}{T_o}\right)^{m/2} \qquad (5.44)$$

(The sum $V_p + V_c$ is independent of T.) If the applied gate voltage varies with temperature according to the function given in this equation, then the drain current will remain constant.

In order to develop more specific results we make the practical approximation that the built-in potential varies linearly with temperature:

$$V_c(T) = V_{co} - \xi_o(T - T_o)$$

where $\xi_o \equiv \xi(T_o)$ is a constant. Also, we define ΔV_o to be the deviation of the design-center bias voltage V_{go} from the external pinch-off voltage V_{po}:

$$\Delta V_o \equiv V_{po} - V_{go}$$

These two relations permit Eq. 5.44 to be expressed in the convenient form

$$\frac{V_g - V_{go}}{\xi_o T_o} = \left(\frac{\Delta V_o}{\xi_o T_o}\right)\left[1 - \left(\frac{T}{T_o}\right)^{m/2}\right] - \left(1 - \frac{T}{T_o}\right) \qquad (5.45)$$

At room temperature the normalizing voltage $\xi_o T_o$ is of the order of 0.6v for silicon. Equation 5.45 is plotted in Fig. 5.4. The curves in Fig. 5.4a show, for various values of m, the normalized equivalent input drift (relative to the design-center bias) as a function of temperature, with $\Delta V_o/\xi_o T_o$ as a parameter. In each case the optimum curve is the one with zero slope at $T/T_o = 1$. These optimum curves give the smallest equivalent input drift over a temperature range centered at T_o; they correspond to the solution for V_{goo} (Eq. 5.43) obtained under the assumption that V_g was constant. For germanium and silicon m lies within the range $1 < m < 3$. If $m = 2$, perfect compensation can be attained at the bias $V_{goo} = V_{po} - \xi_o T_o$; for $m \neq 2$, small amounts of drift are unavoidable. Figure 5.4a indicates that FET's with smaller values of m possess two advantages over those with larger values, even though the small-m units have relatively large equivalent input drifts at

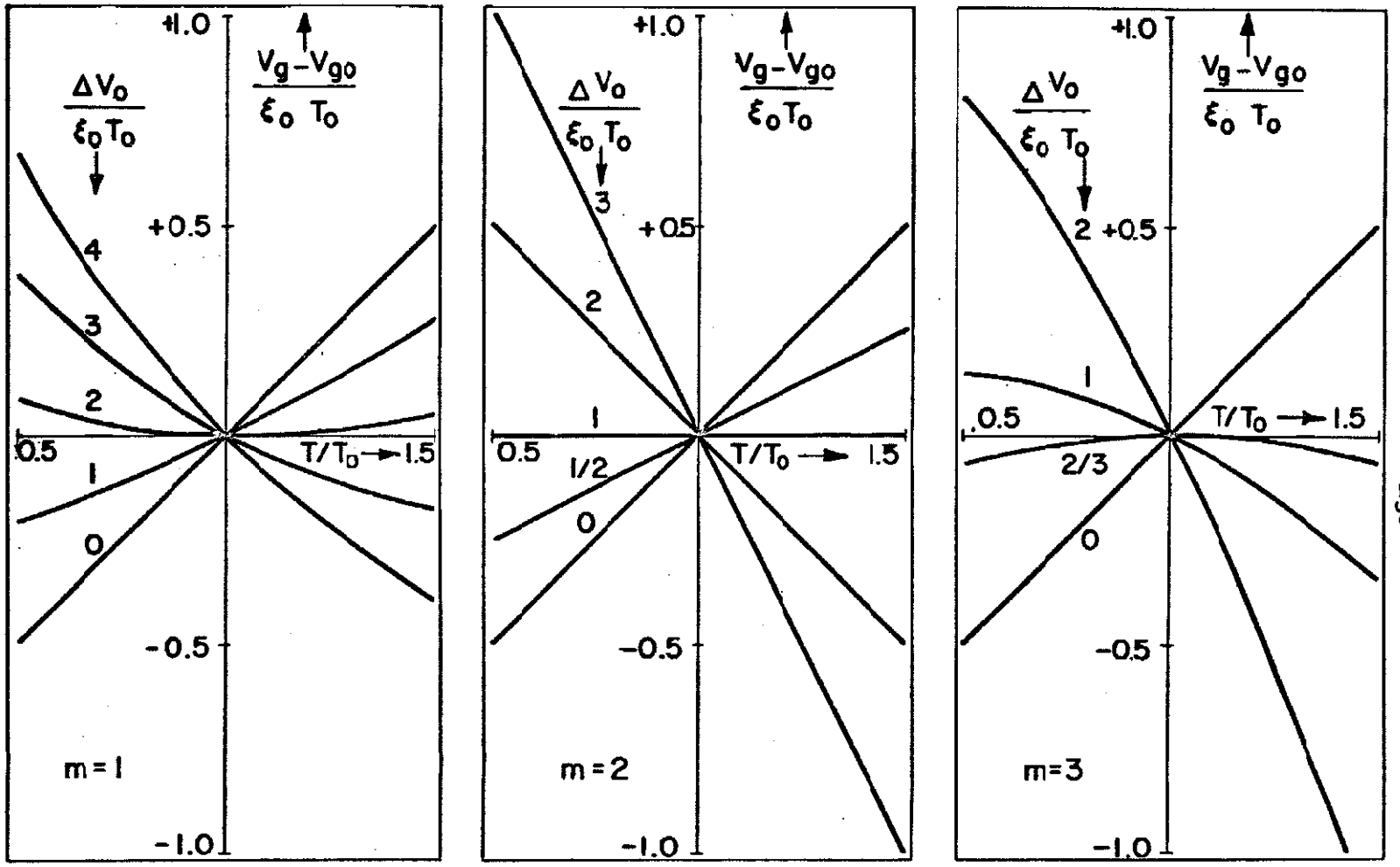


Fig. 5.4a. Normalized equivalent input drift - relative to the design-center bias V_{go} - as a function of temperature, with $\Delta V_o/\xi_o T_o$ as the running parameter.

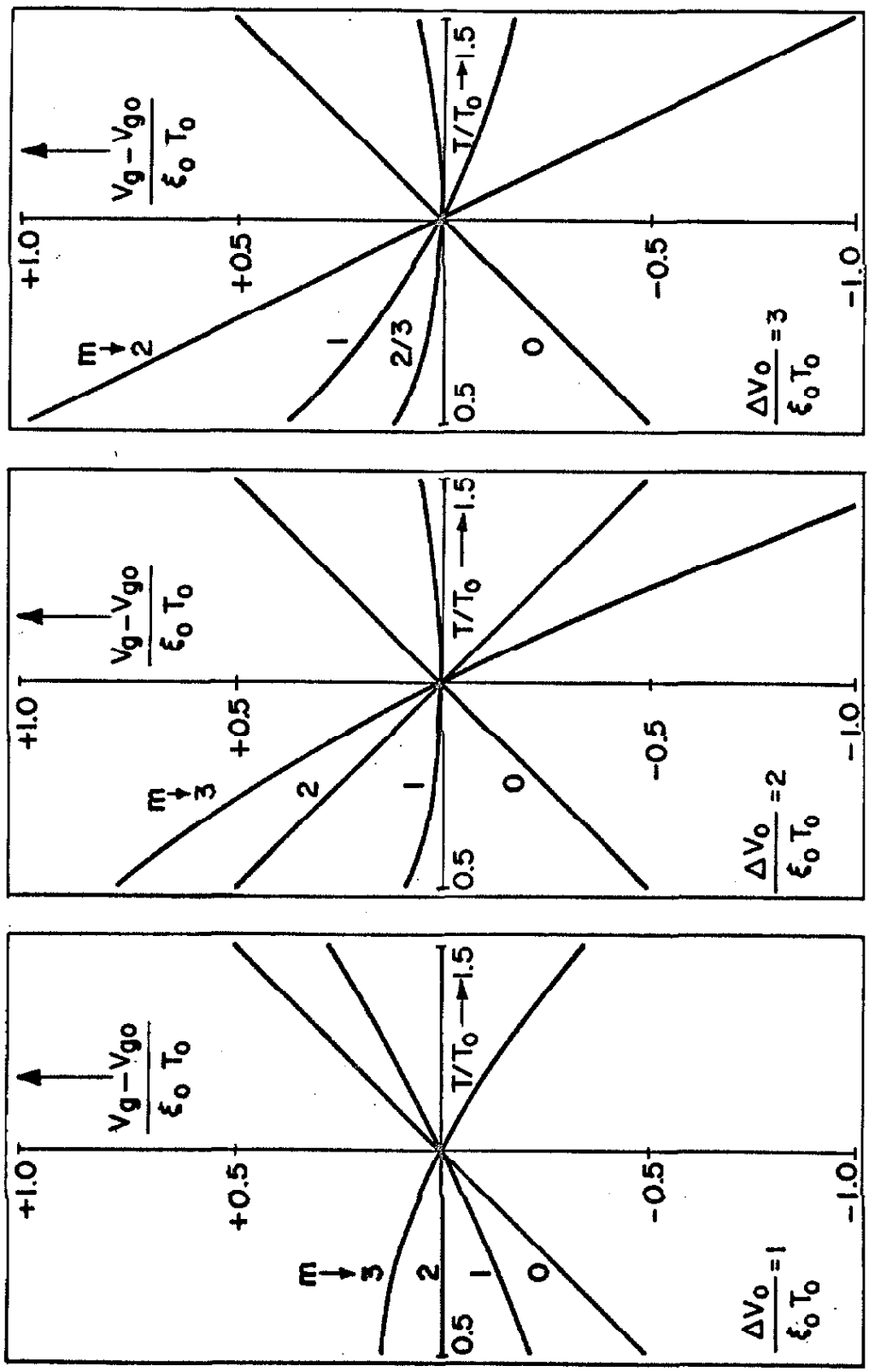


Fig. 5.4b. Normalized equivalent input drift with m as the running parameter.

the optimum bias point. First, in a unit with smaller m the optimum bias point is further from pinch-off, and hence more of the transistor's operating range may be utilized. Second, a given deviation of the actual bias from the optimum bias will result in a smaller equivalent input drift.

Figure 5.4b displays the same independent and dependent variables as Fig. 5.4a but has m as a parameter with families for several values of $\Delta V_o/\xi_o T_o$. The $m = 0$ curves are straight lines with unit slope; they correspond to the constant-mobility case, in which only the linear temperature variation of the built-in potential affects the current. The plots in Fig. 5.4b are of use in the estimation of the equivalent input drift generated because of the variation in the value of the exponent m; within each family the comparison is among FET's biased at the same voltage relative to the pinch-off voltage. Variations in m are seen to be less significant at smaller $\Delta V_o/\xi_o T_o$.

Although the variation of V_c with T is actually not exactly linear, the above quantitative results are reasonably accurate. In any case, the optimum bias V_{goo} given by Eq. 5.43 with ξ defined by Eq. 5.40 does give the smallest input drift.

5.7. Power-Law Nature of the Transfer Characteristics [24], [25]

In making experimental measurements of FET static drain characteristics in the pinch-off range, determination of the effective pinch-off voltage is not possible by direct measurement because of the presence of spurious drain current at and beyond pinch-off.

Further, indirect measurement is hampered by the absence of a theoretical straight-line function from which the pinch-off voltage may be obtained as an intercept or a slope. If the transfer characteristics could be approximated by a power law, then the pinch-off voltage could be determined easily from a straight-line plot of experimental quantities. We have seen that the upper bound on $I(s)/I_o$ is a power law with exponent $n = 2$. Since the lower bound is very close to the upper bound, it is reasonable to expect that the transfer characteristics of any FET (subject to the conditions on the doping density discussed in Section 5.1) can be described by a power law with $n \approx 2$. In this section we shall confirm this expectation theoretically by proving that the value of the exponent n is confined to the rather narrow range $2 \leq n \leq 2.25$, and experimentally by presenting representative data corroborating the theory and illustrating the usefulness of the power-law approximation.

We assume a power law of the form

$$\frac{I(s)}{I_o} = (1 - s)^n \tag{5.46}$$

The ratio I/g_m is then

$$\frac{I(s)}{g_m(s)} = \frac{W_o}{n}(1 - s) \tag{5.47}$$

which is a linear function of the gate voltage. Comparison of this expression with the theoretical bounds given in 5.17 shows that

$n = 2$ is clearly the smallest possible value of n. The largest value of n may therefore be obtained by matching Eq. 5.46 with the lower limit in Eq. 5.17.

As suggested by Eq. 5.47, experimental determination of the exponent n in Eq. 5.46 is most easily done by fitting a straight line through points of I/g_m versus V_g. Therefore, it is more consistent to obtain the upper bound by matching, not I, but I/g_m from 5.46 with that from the lower limit of 5.17. The "best match" is defined as that value of n which gives zero average difference between the exact and the approximate normalized functions $I/g_m W_o$ over the interval $s_c \leq s \leq 1$. The lower end of the interval is taken as $s_c \geq 0$ rather than $s = 0$ in order to account for the built-in potential, whose presence may limit the experimentally measurable range of s. Thus, maximum n is obtained from the solution of

$$\int_{s_c}^{1} \left[\frac{1}{3}(1 - \sqrt{s})(1 + 2\sqrt{s}) - \frac{1}{n}(1 - s) \right] ds = 0$$

The result is

$$n = 2.25 \frac{(1 + \sqrt{s_c})^2}{1 + 2\sqrt{s_c} + 1.5 s_c} \tag{5.48}$$

and shows that n is at most 2.25 (when $s_c = 0$) and decreases monotonically toward a lower limit of 2 as $s_c \to 1$.

The graphical interpretation of Eq. 5.48 is that for a step-

junction FET, n is determined by the straight line $(1/n)(1 - s)$ which best matches the universal curve $(1/3)(1 - \sqrt{s})(1 + 2\sqrt{s})$ over the range $s_c \leq s \leq 1$. The result is shown in Fig. 5.5 for $s_c = 0$ and for $s_c = 0.20$, for which $n = 2.25$ and $n = 2.15$, respectively. The value of n is quite insensitive to s_c. We see from Fig. 5.5 that the theoretical $I/g_m W_o$ curve for a step junction is best fitted by a straight line over the range $s_c \leq s \leq 1$ if s_c is greater than the value $s = 1/16$ for which $I/g_m W_o$ attains a theoretical maximum.

Remarks similar to the above apply to junctions other than step junctions, but in these cases the maximum value of n (corresponding to $s_c = 0$) will lie between the values 2.25 and 2.

Plots of I/g_m versus V_g for commerical p-n junction FET's of various structures are shown in Fig. 5.6. (Data were taken on a General Radio Vacuum-Tube Bridge type 561-D; cf. Appendix B.) The external pinch-off voltage V_p and the exponent n are obtained directly as the intercept on the voltage axis and the reciprocal slope of the straight line. The values for these quantities are given in Table 5.1. It is seen that the experimental points do indeed define straight lines quite closely, thus vindicating the use of the power-law relation 5.46. Moreover, within experimental error the values of n lie in the expected theoretical range. The departure from a straight line in some units near the pinch-off voltage is due to drain leakage current, which is the effect that prevents direct measurement of the pinch-off voltage. The expected maximum in I/g_m occurs at small, forward gate biases, but is not significant in

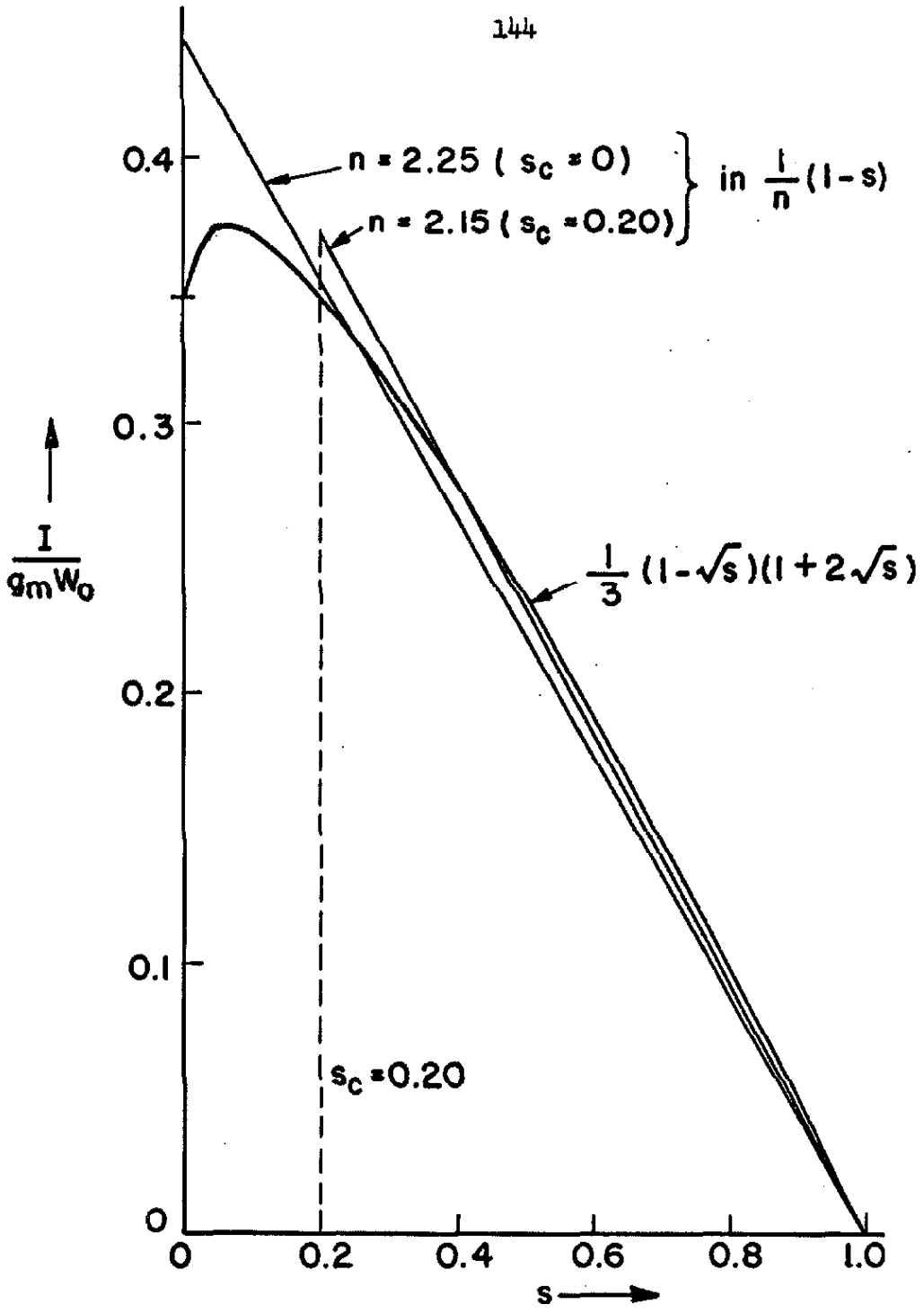


Fig. 5.5. Theoretical curve of $I/g_m W_o$ vs. s for a step-junction FET, and the best-match straight line for two values of s_c.

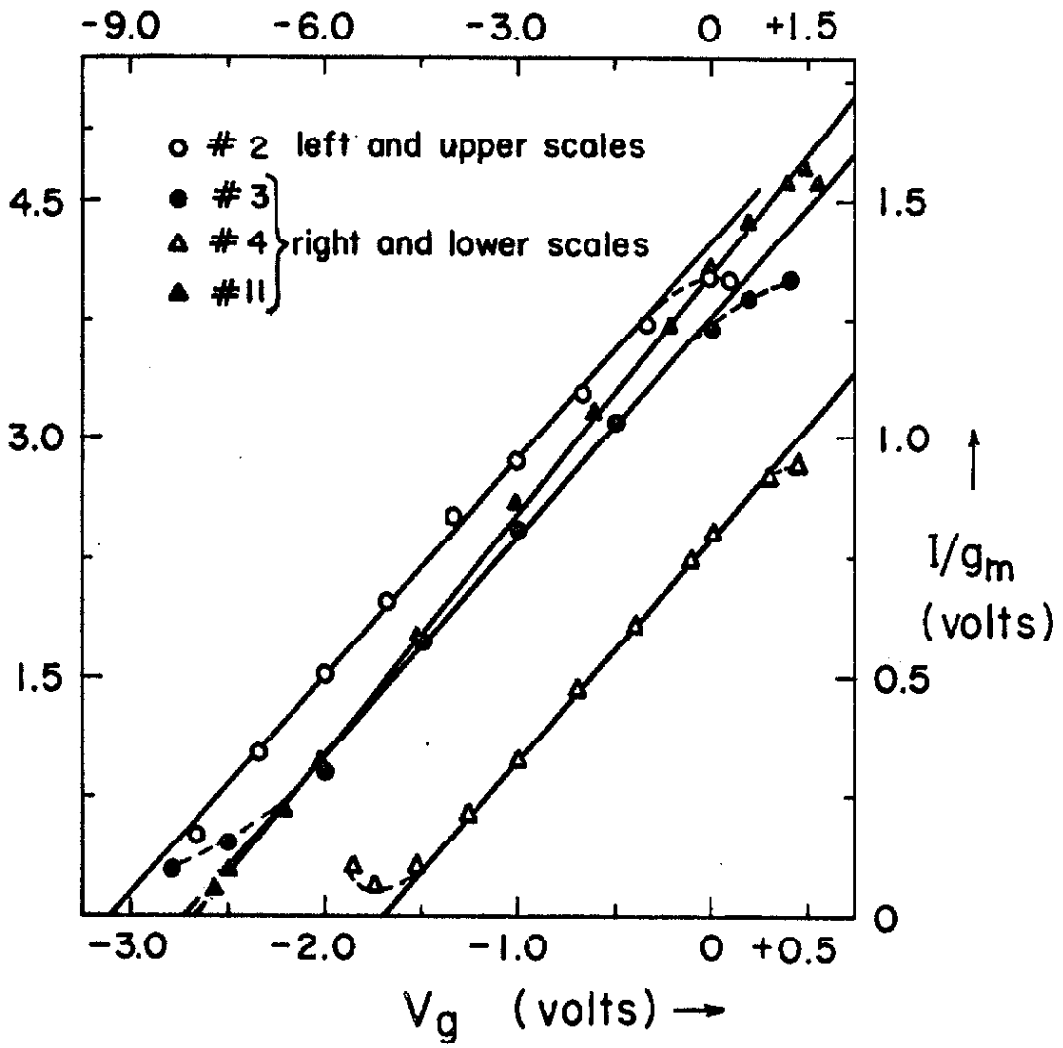


Fig. 5.6. Plots of I/g_m vs. V_g for four commercial FET's (see Table 5.1). For unit #4 (p-channel device) the sign of V_g has been reversed. The drain biases were V_d = 25v (#2), and $|V_d|$ = 10v (#3, #4, #11).

UNIT	TYPE	NATURE OF JUNCTION	V_p (volts)	n
#2	Crystalonics C610	alloy	-9.33	2.20
#3	Motorola MM764	epitaxial	-2.74	2.18
#4	Texas Instruments TIX691	diffused	+1.70	2.15
#11	Fairchild FSP401	diffused	-2.68	1.98

Table 5.1. Data from the curves of Fig. 5.6.

determining the best-fit straight line if the range $0 < |V_g| < V_p$ is considered.

It is concluded that the power-law relation of Eq. 5.46 satisfactorily represents both theoretically and experimentally the transfer characteristics of an FET in the pinch-off region and allows values of the pinch-off voltage and the exponent to be determined directly from experimental measurements. Equation 5.46 is not intended to be exact at any one point, but represents an approximation to the overall shape of the actual transfer characteristics. Further, the worst percentage errors in 5.46 occur near $s = 1$ $(V_g = V_p)$ where $I \approx 0$, i.e. where percentage errors are unimportant experimentally.

5.8. Conclusions

An upper and a lower bound have been established for each of a number of quantities describing the operation of FET's beyond pinch-off. The quantities considered were the drain current and the exponent in its power-law approximation, the equivalent-circuit elements, several figures of merit, and the bias voltage for zero temperature coefficient of drain current. The method of approach used in this chapter may, of course, be applied to other quantities. The bounds, which are applicable to substantially all practical impurity profiles, represent the solutions of two analytically simple structures, a step-junction FET and a delta-junction FET. In each case the bounds are close enough so that the solution of an analytically intractable device, such as one with a diffused junction, may

be satisfactorily approximated by either bound. Thus, burdensome or hopelessly involved computations are avoided. Further, for design and fabrication purposes the doping profile is only of secondary importance.

The proximity of the bounds on the drain current and the parabolic upper bound suggested a power-law approximation to the transfer characteristics. The theoretical and experimental conclusion that the exponent in this power law is close to the value 2 lends weight to the simple derivation performed in Section 4.1.

It must be emphasized that the results of this chapter have been derived from analysis containing some approximations based on a device model which itself contains certain idealizations; the bounds that have been obtained are therefore not necessarily absolute for all practical FET's. Nevertheless, because of the anticipated narrow limits on the device parameters, it is a reasonable approximation for most practical applications to employ the step- or delta-junction formulae.

CHAPTER VI

COMPLETE FIRST-ORDER EQUIVALENT CIRCUIT OF AN ARBITRARILY DOPED FIELD-EFFECT TRANSISTOR

The assumption in Chapter IV that the channel current I did not vary with position enabled the equation relating the current and the electric field to be solved by a simple integration. Under a-c conditions, however, I is not independent of x, since some capacitive current flows between the channel and the gate. In this chapter we extend the analysis of Chapter IV by including this current. The resulting equation is correspondingly more difficult, and only an approximate solution is developed. The present treatment, which again is based on the gradual approximation, is valid for arbitrary impurity profiles and yields as its objective the "real" equivalent circuit, that is, the one that would be measured at the device terminals (related to the a-c components of voltage and current; cf. Section 2.3.2). The complete small-signal, low-frequency equivalent circuit is obtained directly from the approximate solutions for the drain and gate currents and is limited in validity by the gradual approximation.

Two "new" elements are contained in this real equivalent circuit: a forward transfer capacitance present both below and beyond pinch-off, and a drain-source capacitance present below pinch-off. In other respects the circuit is identical to the circuit of Fig. 2.4, the remaining elements having the voltage dependences derived in Section 4.2. In particular, the source-gate and drain-gate charge-capacitances are identical to the corresponding real quantities. Data from a wide

variety of FET's substantiate the predicted circuit.

6.1. Differential Equation for the Channel Potential, Including the Effects of Gate-Channel Current

As mentioned in the foregoing introduction, the capacitive current that flows between the channel and the gate must be considered if a complete solution of the channel is to be obtained under a-c conditions. The problem of solving for the channel potential with the gate-channel current included is similar to that of solving for the potential along a non-uniform transmission line with longitudinal resistance and shunt capacitance, both of which are voltage dependent. (The assumption of only longitudinal resistance is consistent with the neglect of E_y in the channel.) In this section we shall derive the differential equation for the channel potential by a careful consideration of the flow of charges.

Figure 6.1 shows the flow of charge in a typical element of an FET at time t. The net rate of flow of charge across the element in the positive x-direction is

$$I(x+dx,t) - I(x,t) = \frac{\partial I}{\partial x}\, dx$$

Since the channel current is not uniform in the present treatment, we use the symbol I to denote values within the channel, and we use subscripts s, d, and g to refer I to the device terminals. In accordance with the sign convention given in Fig. 6.1, Eq. 4.24 for I may be used for both n- and p-type channels in the above formula if a minus sign is inserted:

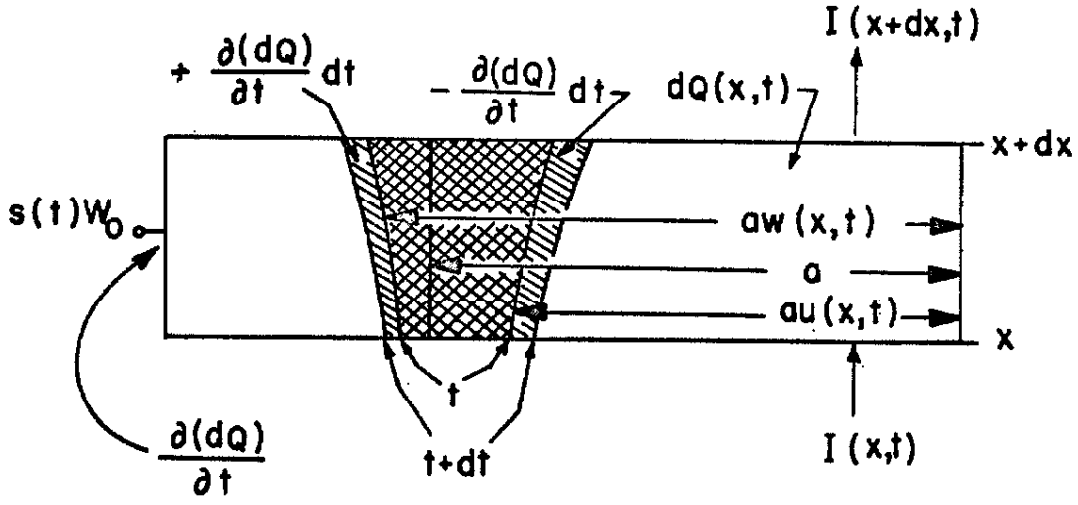


Fig. 6.1. Flow of charge in an element of an FET. The cross-hatched area represents the space-charge region at time t; the shaded area represents the space charge added in the time interval dt.

$$\frac{\partial I}{\partial x}\,dx = \frac{\partial}{\partial x}\left[-\mu a A W_o \frac{\partial Q}{\partial x}\int_0^{u(x,t)} \rho(z)\,dz\right]dx \qquad (6.1)$$

As in Chapter IV, $\theta(x,t)$ is the normalized potential in the channel measured with respect to the gate and is given by

$$\theta(x,t) = \frac{\int_{u(x,t)}^{w(x,t)} (1 - z)\rho(z)\,dz}{\int_0^{w_m} (1 - z)\rho(z)\,dz} \qquad (6.2)$$

with

$$\int_{u(x,t)}^{w(x,t)} \rho(z)\,dz = 0 \qquad (6.3)$$

Now, the mobile charge dQ contained in the differential element of channel must change (with time) if the outgoing current $I(x+dx,t)$ differs from the incoming current $I(x,t)$. In time dt the <u>amount</u> of this change is $[\partial(dQ)/\partial t]\,dt$. The <u>rate</u> of the change is thus $\partial(dQ)/\partial t$ and is, of course, equal to the rate at which charge flows into the gate terminal of the element. This latter rate in turn is equal to the rate at which charge accumulates on the depletion region on the gate side of the junction. Therefore,

$$\frac{\partial}{\partial t}\,[dQ] = \frac{\partial}{\partial t}\left[aA\int_1^{w(x,t)} \rho(z)\,dz\right]dx$$

$$= \frac{\partial}{\partial t}\left[-aA\int_{u(x,t)}^{l} \rho(z)\, dz\right]dx \qquad (6.4)$$

where use has been made of Eq. 6.3. Since the net flow of charge into the element must be zero, the right-hand sides of Eqs. 6.1 and 6.4 are equal:

$$\mu W_c \frac{\partial}{\partial x}\left[\frac{\partial \theta}{\partial x}\int_0^u \rho\, dz\right] = \frac{\partial}{\partial t}\left[\int_u^l \rho\, dz\right] \qquad (6.5)$$

We now normalize the length variable x with respect to L

$$X \equiv \frac{x}{L} \qquad (6.6)$$

define

$$\tau_o \equiv \frac{L^2}{\mu W_o} \qquad (6.7)$$

which has the dimensions of time (and is equal to $3/4$ the zero-bias transit time for a delta-junction FET, $2/3$ that for a step-junction FET), and substitute the normalized transconductance function defined in Eq. 4.28,

$$g(\theta) = \frac{\int_0^{u(0)} \rho(z)\, dz}{\int_0^1 \rho(z)\, dz}$$

so that Eq. 6.5 becomes

$$\frac{\partial}{\partial X}\left[g(0)\ \frac{\partial \theta}{\partial X}\right] = -\tau_o \frac{\partial}{\partial t}\ [g(\theta)] \tag{6.8}$$

or

$$\frac{\partial^2}{\partial X^2}\left[\int_s^{\theta} g(\phi)\ d\phi\right] = -\tau_o \frac{\partial}{\partial t}\ [g(\theta)] \tag{6.9}$$

The two spatial boundary conditions on θ are

$$\theta = s(t) \qquad \text{at} \qquad X = 0 \tag{6.10a}$$

$$\theta = d(t) \qquad \text{at} \qquad X = 1 \tag{6.10b}$$

The temporal boundary condition depends on the initial conditions that are assumed.

Equation 6.9 is the partial differential equation governing the channel potential. No restrictions have been made on ρ, on the frequency, or on the magnitude of the applied signal, but the validity of this equation is limited by the gradual approximation.

6.2. Approximate Solution for the Drain and Gate Currents

The differential equation 6.9 for the channel potential is nonlinear, and its solution is difficult even for the simple delta-junction FET for which $g(\theta) = (1 - \theta)$. We shall therefore obtain only an approximate solution valid for low frequencies and small a-c

signals. We assume that the right-hand side of Eq. 6.9 is small, and solve for the current by an iterative procedure.

As a first approximation we assume that $\partial g/\partial t = 0$. The solution of 6.9 subject to boundary conditions 6.10 is then

$$\int_s^\theta g(\phi)\, d\phi = X \int_s^d g(\phi)\, d\phi$$

$$= X \frac{I_1(s,d)}{G_o W_o} \tag{6.11}$$

where I_1 denotes the expression for the drain current developed in Chapter IV. Not unexpectedly, the first approximation to the a-c solution is the static solution of Chapter IV. We use this result to obtain a better approximation for $\partial g/\partial t$ for substitution into the right-hand side of Eq. 6.9. Equation 6.11 yields

$$g(\theta)\dot{\theta} - g(s)\dot{s} = X \frac{\dot{I}_1}{G_o W_o}$$

where dots denote differentiation with respect to time. Thus, since $\partial g/\partial t = (dg/d\theta)\dot{\theta}$, the second approximation to Eq. 6.9 is

$$\frac{d^2}{dX^2}\left[\int_s^\theta g\, d\phi\right] = -\tau_o \frac{dg}{d\theta} \frac{1}{g(\theta)} \left[X \frac{\dot{I}_1}{G_o W_o} + g(s)\dot{s}\right] \tag{6.12}$$

Consistent with this second-order approximation we may substitute

$$X = \frac{G_o W_o}{I_1} \int_s^\theta g\, d\phi$$

from Eq. 6.11 into the right-hand side of Eq. 6.12. Multiplication of both sides of the resulting equation by

$$\frac{d}{dX}\left[\int_s^\theta g\, d\phi\right] dX = g(\theta)\, d\theta$$

gives

$$\frac{1}{2}\frac{d}{dX}\left[\frac{d}{dX}\int_s^\theta g\, d\phi\right]^2 dX = -\tau_o\left[\frac{\dot{I}_1}{I_1}\int_s^\theta g\, d\phi + g(s)\dot{s}\right] dg$$

which yields

$$\frac{1}{2}\left[\frac{d}{dX}\int_s^\theta g\, d\phi\right]^2 = -\tau_o\left\{\frac{\dot{I}_1}{I_1}\int_s^\theta\left[\int_s^\phi g(\nu)\, d\nu\right]\frac{dg(\phi)}{d\phi}\, d\phi + g(\theta)g(s)\dot{s}\right\} + \frac{1}{2}h^2 \qquad (6.13)$$

where h is a constant of integration. The integral in the curved brackets may be expressed in a simpler form by partial integration:

$$\int_s^\theta\left[\int_s^\phi g(\nu)\, d\nu\right]\frac{dg(\phi)}{d\phi}\, d\phi = \int_s^\theta g(\phi)[g(0) - g(\phi)]\, d\phi$$

For brevity, we define

$$f(\theta) = \frac{\dot{I}_1}{I_1}\int_s^\theta g(\phi)[g(\theta) - g(\phi)]\, d\phi + g(\theta)g(s)\dot{s} \qquad (6.14)$$

so that Eq. 6.13 may be written as

$$\left[\frac{d}{dX}\int_{s}^{\theta} g\, d\phi\right]^2 = \left[g(\theta)\frac{d\theta}{dX}\right]^2 = h^2 - 2\tau_o f(\theta) \qquad (6.15)$$

Since the second-order solution of Eq. 6.9 differs from the first-order solution only in terms of the order of τ_o, we must have $h \sim I_1/G_o W_o$, and hence

$$h^2 \gg |2\tau_o f(\theta)|$$

Therefore,

$$\left[\frac{1}{h^2 - 2\tau_o f(\theta)}\right]^{\frac{1}{2}} \approx \frac{1}{h}\left[1 + \frac{\tau_o}{h^2} f(\theta)\right]$$

and Eq. 6.15 becomes

$$g(\theta)\left[1 + \frac{\tau_o}{h^2} f(\theta)\right] d\theta = h\, dX$$

which, upon integration and application of boundary condition 6.10a, yields

$$hX = \int_{s}^{\theta} g(\phi)\, d\phi - \frac{\tau_o}{h^2}\int_{s}^{\theta} g(\phi) f(\phi)\, d\phi$$

The other boundary condition, 6.10b, allows the determination of h:

$$h = \frac{I_1}{G_o W_o} + \frac{\tau_o}{h^2} \int_s^c gf \, d\phi$$

$$h \approx \frac{I_1}{G_o W_o} + \tau_o \left(\frac{G_o W_o}{I_1} \right)^2 \int_s^d gf \, d\phi$$

Thus, correct to the first power of τ_o, Eq. 6.15 becomes

$$g(\theta) \frac{\partial \theta}{\partial X} = \frac{I_1}{G_o W_o} + \tau_o \left(\frac{G_o W_o}{I_1} \right)^2 \int_s^c g(\phi)[f(\phi) - f(\theta)] \, d\phi$$

The current in the channel is

$$I = - \frac{\mu a A W_o}{L} \frac{\partial \theta}{\partial X} \int_0^{u(\theta)} \rho \, dz$$

Henceforth, as in the previous chapters, the d-c component of the channel current will be considered a positive quantity. (Its direction depends merely on the type of channel.) With this understanding, the above equation may be written

$$I = G_o W_o g(\theta) \frac{\partial \theta}{\partial X}$$

$$= I_1 + \tau_o G_o W_o \left(\frac{G_o W_o}{I_1} \right)^2 \int_s^d g(\phi)[f(\phi) - f(\theta)] \, d\phi \qquad (6.16)$$

The two values of I of particular interest, the drain and gate currents, are given by

$$I_d = I\big|_{X=1} \tag{6.17}$$

$$= I_1 + \tau_o G_o W_o \left(\frac{G_o W_o}{I_1}\right)^2 \int_s^d g(\phi)[f(\phi) - f(d)]\, d\phi \tag{6.18}$$

$$I_g = I\big|_{X=0} - I\big|_{X=1} \tag{6.19}$$

$$= -\tau_o G_o W_o \left(\frac{G_o W_o}{I_1}\right)[f(s) - f(d)] \tag{6.20}$$

From the definition of I_1,

$$\frac{I_1(s,d)}{G_o W_o} = \int_s^d g\, d\phi \tag{6.21}$$

we obtain

$$\frac{\dot{I}_1(s,d)}{G_o W_o} = g(d)\dot{d} - g(s)\dot{s} \tag{6.22}$$

Equations 6.18 and 6.20, together with 6.21, 6.22, and 6.14 for $f(\theta)$, are the sought-after results that will provide the complete first-order equivalent circuit. In Section 6.5 we show that this solution is valid for angular frequencies ω such that $\omega\tau_o \ll 1$.

The time dependences of I_d and of I_g are contained in the time dependences of s and of d and in the time derivatives present in f. For sinusoidal signals the time derivative terms generate components of the currents that are out of phase with the applied voltages. The above equations thus enable us to determine the real equivalent-circuit capacitances.

6.3. Equivalent Circuit Below Pinch-Off; Experimental Verification of $(C_{22} - C_{12})$ as a "New" Circuit Element

In this section we shall obtain the complete equivalent circuit for an FET by assuming small, sinusoidal voltages at the device terminals. The conductive elements in the circuit are G_{22} and g_m; the capacitive elements are C_{11}, C_{12}, C_{21} $(\neq C_{12})$, and C_{22} $(\neq C_{12})$. In addition to the source-gate and drain-gate capacitances that arise in the treatment of Chapter IV, the particular circuit configuration that we shall justify with physical arguments contains a finite (negative) drain-source capacitance $C_{ds} = C_{22} - C_{12}$ and a finite forward transfer capacitance $(C_{21} - C_{12})$. Experimental data which confirm the presence of C_{ds} are presented in this section. The capacitance $(C_{21} - C_{12})$ remains finite beyond pinch-off and is considered in the next section.

In terms of the short-circuit admittance parameters y_{ij} of a two-port network, the general equations describing the operation of an FET are

$$\left.\begin{aligned} i_g &= y_{11}v_g + y_{12}v_d \\ i_d &= y_{21}v_g + y_{22}v_d \end{aligned}\right\} \qquad (6.23)$$

with the conventional signal polarities. We may place Eqs. 6.18 and 6.20 in this form by substituting

$$s(t) = s_o + p_1(t)$$

$$d(t) - s(t) = (d_o - s_o) + p_2(t)$$

in which s_o and $(d_o - s_o)$ are the normalized d-c gate and drain biases, and where p_1 and p_2 correspond to the a-c signals v_g and v_d. We assume small, sinuscidal signals, and hence we have

$$|p_1| \ll s_o, \qquad |p_2| \ll (d_o - s_o) \tag{6.24}$$

and

$$p_1, p_2 \sim e^{j\omega t} \tag{6.25}$$

Equation 6.24 allows us to expand $I_1(s,d)$ as

$$I_1(s,d) = G_o W_o \int_s^d g(\phi)\, d\phi$$

$$\approx I_1(s_o,d_o) + [g(d_o) - g(s_o)]p_1 + g(d_o)p_2$$

Use of the above relations and the results of the previous section yields the results

$(I_d - i_d)$ = d-c drain current

$$= I_1(s_o, d_o) \tag{6.26}$$

$$y_{11} = j\omega C_{11} \tag{6.27}$$

$$y_{12} = -j\omega C_{12} \tag{6.28}$$

$$y_{21} = g_m - j\omega C_{21} \tag{6.29}$$

$$y_{22} = G_{22} + j\omega C_{22} \tag{6.30}$$

in which the circuit elements are given by

$$C_{11} = C_{12} + \tau_o G_o g(s) \frac{\int_s^d g(\theta)\{g(s) - g(\theta)\}\, d\theta}{\left[\int_s^d g(\theta)\, d\theta\right]^2} \tag{6.31}$$

$$C_{12} = \tau_o G_o g(d) \frac{\int_s^d g(\theta)\{g(\theta) - g(d)\}\, d\theta}{\left[\int_s^d g(\theta)\, d\theta\right]^2} \tag{6.32}$$

$$g_m = G_o[g(s) - g(d)] \tag{6.33}$$

$$C_{21} = C_{12} + \tau_o G_o [g(s)-g(d)] \frac{\int_s^d g(\theta) \left\{ \int_s^\theta g^2(\phi)d\phi - \int_\theta^d g^2(\phi)d\phi \right\} d\theta}{\left[\int_s^d g(\theta)\, d\theta \right]^3} \qquad (6.34)$$

$$C_{22} = G_o\, g(d) \qquad (6.35)$$

$$C_{22} = C_{12} - \tau_o G_o g(d) \frac{\int_s^d g(\theta) \left\{ \int_s^0 g^2(\phi)d\phi - \int_\theta^d g^2(\phi)d\phi \right\} d\theta}{\left[\int_s^d g(\theta)\, d\theta \right]^3} \qquad (6.36)$$

The values of s and d in Eqs. 6.31 - 6.36 should actually be s_o and d_o, but the subscripts have been dropped for convenience. All the circuit elements are positive quantities. The signs in Eqs. 6.27 - 6.30 may be established either by physical arguments or by a careful consideration of voltage polarities and current directions.

As expected, the d-c drain current, the transconductance, and the output conductance agree with the corresponding expressions derived in Chapter IV. Furthermore, the charge-capacitances satisfy the identities

$$C_{11}^* \equiv C_{11}$$

$$C_{22}^* \equiv C_{12}$$

These observations are reassuring, for they confirm that the present theory does not radically alter the conclusions of Chapter IV. The remaining quantities, $(C_{21} - C_{12})$ and $(C_{22} - C_{12})$, are new and are not predicted by previous theories or by intuition.

The above remarks and Eqs. 6.31 - 6.36 suggest the equivalent circuit shown in Fig. 6.2. This circuit satisfies two important symmetry properties inherent in the FET model under consideration. First, as discussed in Section 2.3.2, an interchange of s and d should produce an interchange of C_{sg} and C_{dg}. If $C_{sg} = (C_{11} - C_{12})$ and if $C_{dg} = C_{12}$, then this condition is fulfilled. Second, if s and d are interchanged, the magnitude of the "active" part of the forward transfer admittance should be unaltered, but the sign should be reversed. This conclusion may be established by referring to Fig. 2.5. In normal operation (Fig. 2.5a) with source-gate and drain-gate potentials (s,d), a variation in s (i.e. in V_g) with d - s (i.e. V_d) constant, changes the space-charge-region shape over the whole channel length, thereby altering the drain current. With potentials (d,s) applied (Fig. 2.5b), a variation in d changes the space-charge region shape in exactly the same manner as the variation in s did with biases (s,d) hence produces the identical effect on the drain current.* However, interchanging s

---

* In this argument we have neglected the drain current that flows through the "passive" component C_{12} of the forward transfer admittance. The magnitude of this current is modified if s and d are interchanged (cf. Fig. 2.5c).

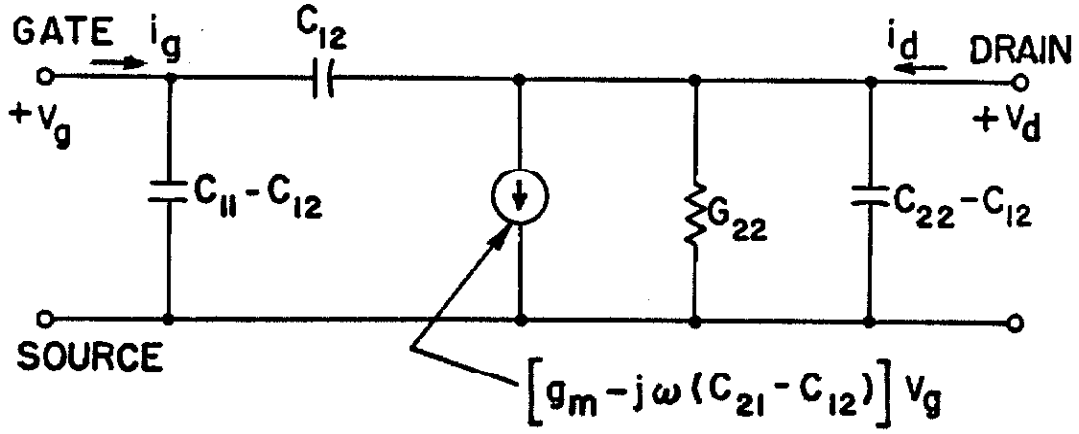


Fig. 6.2. Equivalent circuit for an FET.

and d reverses the direction of current flow. Therefore, the signs of the admittances differ in the two cases. Equations 6.33 and 6.34 immediately reveal that only the sign and not the magnitude of the active admittance $[g_m - j\omega(C_{21} - C_{12})]$ is altered if s and d are interchanged. Thus, the second symmetry condition is fulfilled by the circuit of Fig. 6.2.

Formulae for the capacitances may be obtained for any doping profile by substitution into the previous equations. Even for simple step- and delta-junction FET's, however, the resulting expressions for arbitrary biases are rather cumbersome and will not be given. Figure 6.3 shows the short-circuit capacitances C_{ij} for these two types of FET as a function of the drain-gate voltage for the particular gates biases $s = 0$ and $s = 1/4$. The corresponding curves for the equivalent-circuit capacitances $C_{sg} = (C_{11} - C_{12})$, $C_{dg} = C_{12}$, $C_{ds} = (C_{22} - C_{12})$, and $(C_{21} - C_{12})$ are given in Fig. 6.4.

Of the two "new" circuit elements C_{ds} and $(C_{21} - C_{12})$, we shall discuss the former in this section since it vanishes beyond pinch-off, and we shall discuss the latter in the next section since its visualization is easier in the pinch-off range of operation. Although the experimentally measurable capacitance C_{22} is positive, the circuit element C_{ds} is negative. The existence of this negative element is made plausible if we examine how C_{22} arises physically. Basically, the short-circuit output capacitance is defined by

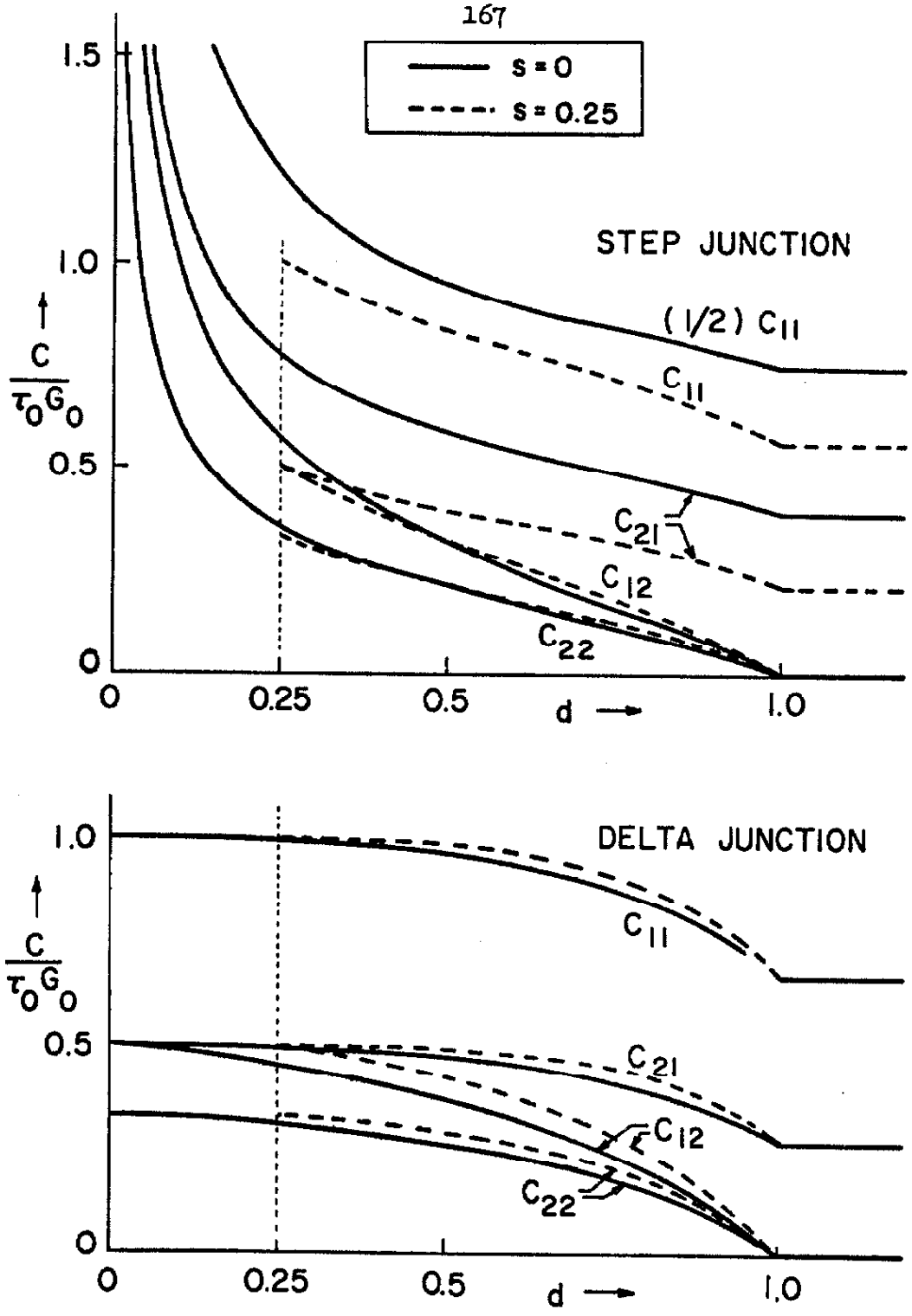


Fig. 6.3. Normalized short-circuit capacitances versus normalized gate-drain potential for gate-source potentials $s = 0$, $s = 1/4$. For $d > 1$ the capacitances are independent of d.

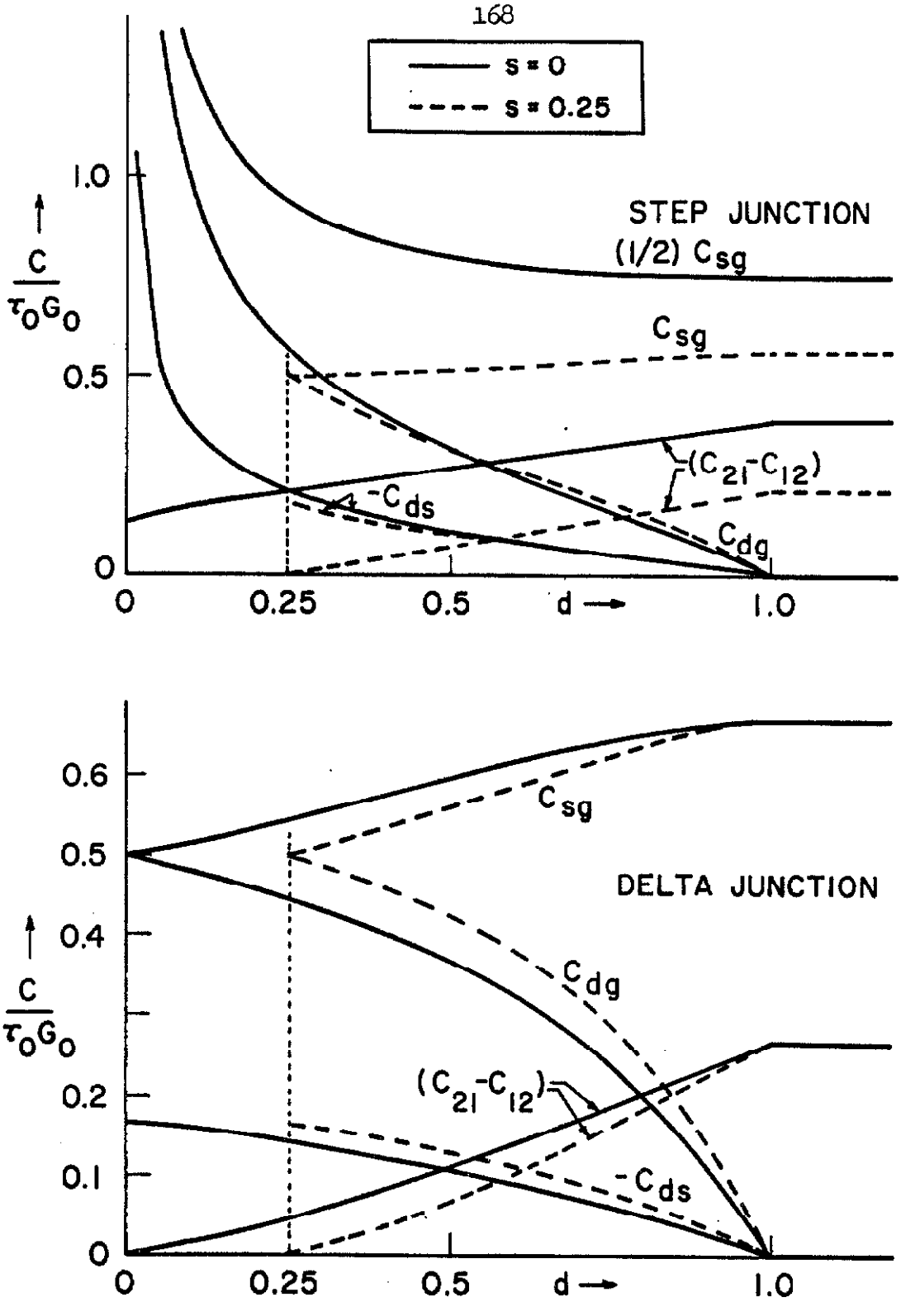


Fig. 6.4. Equivalent-circuit capacitances corresponding to the short-circuit capacitances of Fig. 6.3.

$$C_{22} = \frac{1}{\omega}\frac{1}{v_d}\left[\text{quadrature component of } i_d\right]_{v_g=0}$$

Even with $v_g = 0$, some capacitive current flows between the gate and the source because the space-charge-region boundary varies along its entire length when $v_d \neq 0$. The flow pattern of the capacitive current is shown pictorially in Fig. 6.5. At the drain terminals the capacitance $C_{22} > 0$ is observed. All the capacitive current from the drain flows into the space-charge region, but in addition some capacitive current flows from the source into the space-charge region. (Because of the structure of an FET, no capacitive current flows directly from the drain to the source. Cf. footnote page 29.) Either from physical arguments or from Eq. 6.16 for $I(x,t)$ we may infer that the quadrature components of i_s and i_d are always in opposite directions. Thus, the net quadrature component of i_d may be considered to be made up of two components, i_g in the same (spatial) direction as i_d, and i_s in the opposite direction to i_d. This latter component implies a negative drain-source capacitance.

In order to verify experimentally that the output capacitance C_{22} is smaller than the reverse transfer capacitance C_{12}, measurements were taken on a wide variety of FET's. The experimental setups used to obtain these data and the other equivalent-circuit data presented in this chapter are described in Appendix B. Consistently and without exception C_{22} was smaller than C_{12} below pinch-off. Moreover, beyond pinch-off $C_{22} = C_{12} \approx$ constant (very slightly dependent on bias), indicating that in this range these elements may

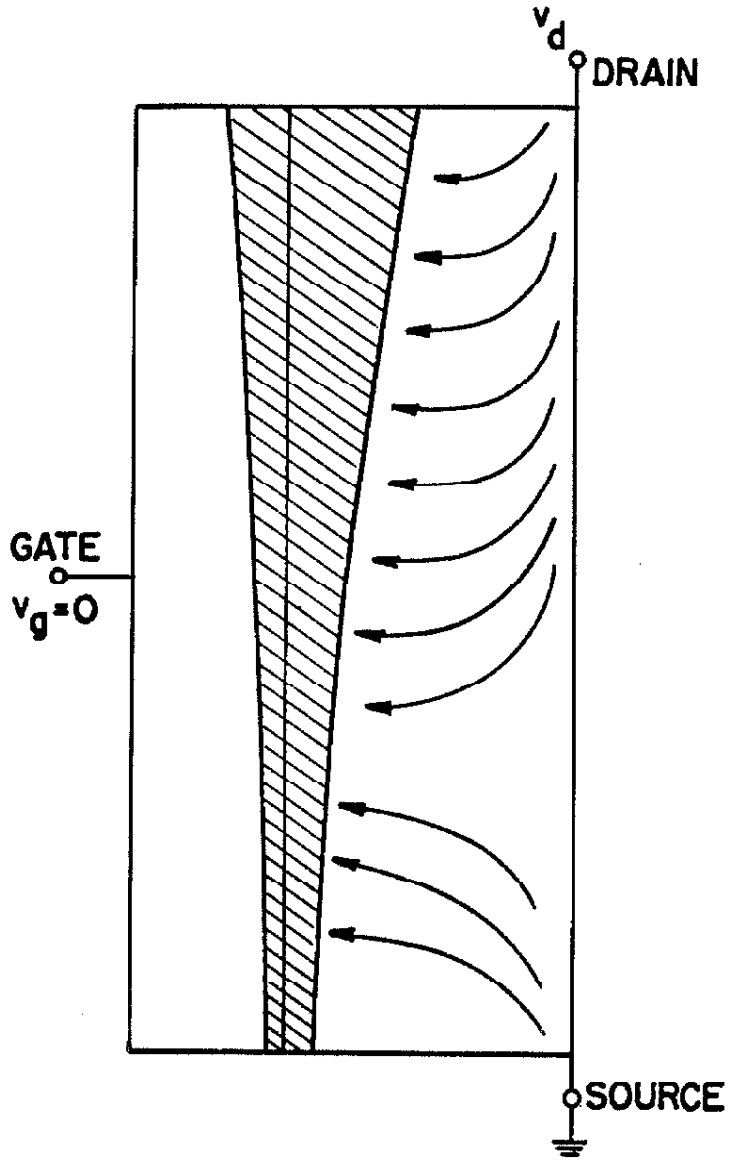


Fig. 6.5. Rough sketch of the capacitive current in an FET with $v_g = 0$.

be identified as stray capacitances. A convenient quantitative experimental test is to compare the values of C_{12} and C_{22} at $s = d$ $(V_d - 0)$. Application of l'Hôpital's rule to Eqs. 6.32 and 6.36 shows that $C_{12}(s,s) = -(1/2)\tau_o G_o g'(s)$ and that $C_{22}(s,s) = -(1/3)\tau_o G_o g'(s)$. Hence,

$$\frac{C_{22}(s,s)}{C_{12}(s,s)} = \frac{2}{3} \tag{6.37}$$

regardless of the type of junction. Use of this formula eliminates the necessity of including the built-in potential or the pinch-off voltage in the analysis of the experimental data. Table 6.1 shows some representative data for the ratio C_{22}/C_{12} with biases $V_g = V_d = 0$. (The stray capacitance was subtracted from the measured values before the ratio was computed.) Furthermore, the ratios were roughly independent of gate bias, even though the individual capacitance values varied greatly. Although some units have ratios that differ appreciably from the theoretical value of 0.67, the above statement that $C_{22} < C_{12}$ without exception below pinch-off indicates that the present theory is certainly qualitatively valid. The quantitative errors present in some units may be attributed to the idealizations used in the theoretical model and in its analysis.

The presence of $C_{ds} \neq 0$ in the equivalent circuit below pinch-off is not too important practically because this capacitance is small compared with the other circuit capacitances. However, the

TRANSISTOR	NATURE OF JUNCTION	C_{22}/C_{12} ($V_g, V_d = 0$)
Amelco FG34 (#54)	diffused	0.87
Crystalonics C610 (# 1)	alloy	0.69
Motorola MM763 (#12)	epitaxial	0.76
MM764 (# 3)		0.96
MM765 (#24)		0.92
Texas Ins. TIX691 (# 4)	diffused	0.96
3-lead, p-channel device (no model #)		0.71

Table 6.1. Representative data showing the ratio C_{22}/C_{12} with the biases $V_g = V_d = 0$. (The ratios are roughly independent of V_g.) The theoretical value is 0.67.

existence of this capacitance in actual FET's does lend weight to the validity of the foregoing theory.

6.4. Equivalent Circuit Beyond Pinch-off; Experimental Verification of C_{21} as a "New" Circuit Element

As in Chapters II and IV, the equivalent circuit of the previous section may be extended to drain-gate voltages greater than pinch-off by substituting $d = 1$ in the appropriate equations. The only non-zero elements are

$$C_{11} = \tau_o G_o g(s) \frac{\int_s^1 g(\theta)[g(s) - g(\theta)]\, d\theta}{\left[\int_s^1 g(\theta)\, d\theta\right]^2} \tag{6.38}$$

$$g_m = G_o g(s) \tag{6.39}$$

$$C_{21} = \tau_o C_o g(s) \frac{\int_s^1 g(\theta)\left\{\int_s^\theta g^2(\phi)\, d\phi - \int_\theta^1 g^2(\phi)\, d\phi\right\} d\theta}{\left[\int_s^1 g(\theta)\, d\theta\right]^3} \tag{6.40}$$

and the resulting equivalent circuit is given in Fig. 6.6. The input capacitance C_{11} and the transconductance g_m have already been discussed. The forward transfer capacitance C_{21}, however, is new and is not present in the general circuit of Fig. 2.9. The physical explanation of C_{21} may be evinced by considering the flow pattern of capacitive current in the channel. The pattern will be similar to

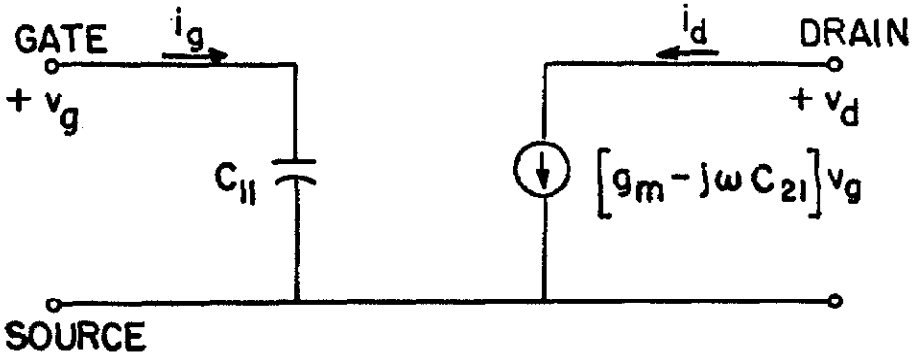


Fig. 6.6. Equivalent circuit for an FET beyond pinch-off.

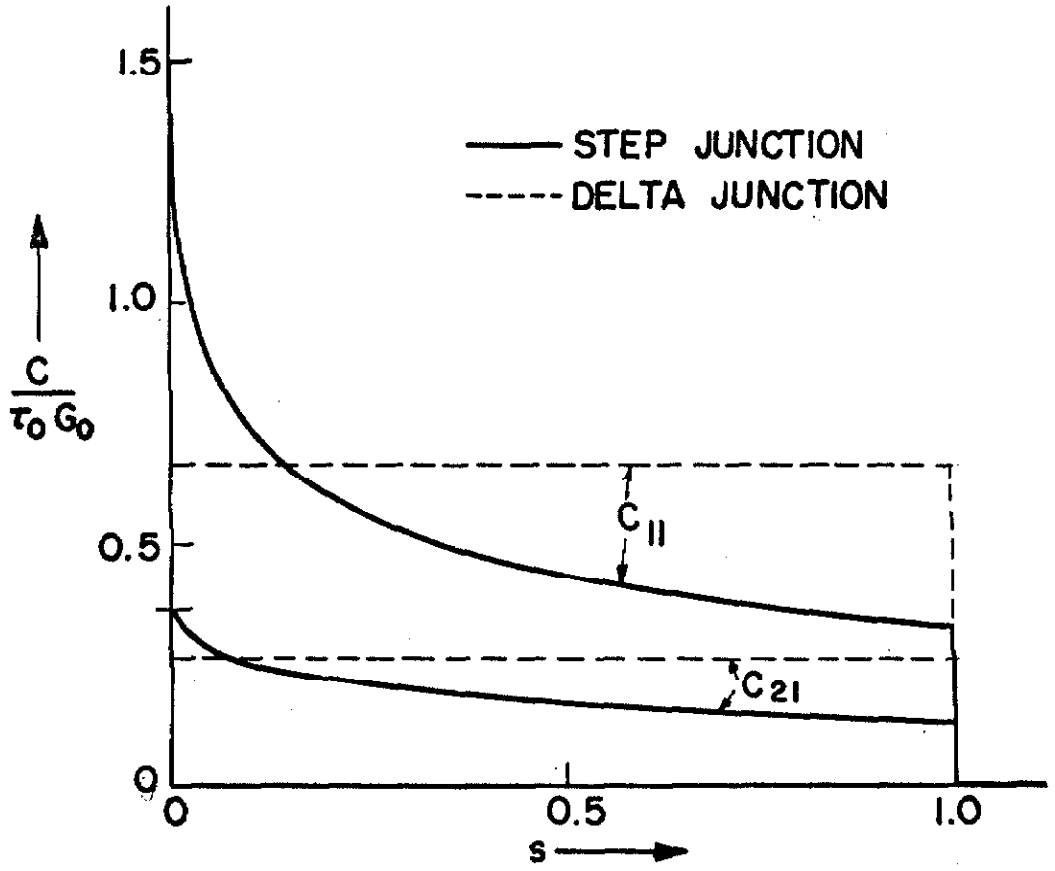


Fig. 6.7. Equivalent-circuit capacitances beyond pinch-off for step- and delta-junction FET's.

that given in Fig. 6.5, except that beyond pinch-off a variation v_d in the drain voltage has no influence on the channel boundary and hence cannot affect the drain or gate currents. However, a variation v_g in the gate voltage will alter the channel boundary along its entire length. The capacitive current flowing toward the drain (as well as the capacitive current flowing toward the source) will therefore be affected. The variation in the drain current produces C_{21}.

The step- and delta-junction expressions for the forward transfer capacitance are

$$C_{21} = \frac{9}{70}\,\tau_o G_o\,\frac{3 + 15\sqrt{s} + 10s}{(1 + 2\sqrt{s})^3}\,, \qquad \text{step junction} \qquad (6.41)$$

$$C_{21} = \frac{4}{15}\,\tau_o G_o\,, \qquad \text{delta junction} \qquad (6.42)$$

In Fig. 6.7 these equations are plotted along with the curves for the input capacitances.

For $s > 1$ all capacitances vanish, so that both C_{11} and C_{21} should theoretically exhibit a sharp drop at $s = 1$. (This phenomenon has been discussed in Section 2.4.2.) At the singular point $s = d = 1$, the magnitudes of these drops depend on the direction of approach, but for $d > 1$ the drops are independent of the biases. In fact, application of l'Hôpital's rule to Eqs. 6.38 and 6.40 yields

$$\left.\frac{C_{21}}{C_{11}}\right|_{\substack{s=1\\d>1}} = \frac{2}{3} \tag{6.43}$$

independent of the type of FET.

Measurements were taken on a large sampling of FET's and the predicted circuit was substantiated. Most important, a bias-dependent forward transfer capacitance was observed in all transistors tested. The output and reverse transfer capacitances were equal and essentially constant, and hence their presence may be attributed to interelectrode effects; that is, the equivalent circuits of the devices themselves have $C_{12} = C_{22} = 0$. (Further justification for this statement is the fact that beyond gate pinch-off C_{21} was equal to C_{12} and C_{22}.) A finite output conductance was also measured, but this parameter is considered a second-order effect because it is small and because its explanation requires drastic modification of the simple model that is being used.

Curves of C_{11} and C_{21} versus V_g for two transistors - Crystalonics C610 (#1) and Motorola MM763 (#12) - are shown in Fig. 6.8.* Although in every unit tested, C_{21} was definitely dependent on s, the amount of actual variation of this quantity (for $|V_g| < |V_p|$) did not always agree with the predicted variation (as

---

* The absence of sharp drops in C_{11} and C_{21} makes it difficult to fit the Crystalonics curves to the theory; the small voltage range on the Motorola unit and the fact that three "arbitrary" parameters must be chosen (V_p, V_c, and $\tau_o G_o$) render a theoretical fit on this unit unavailing.

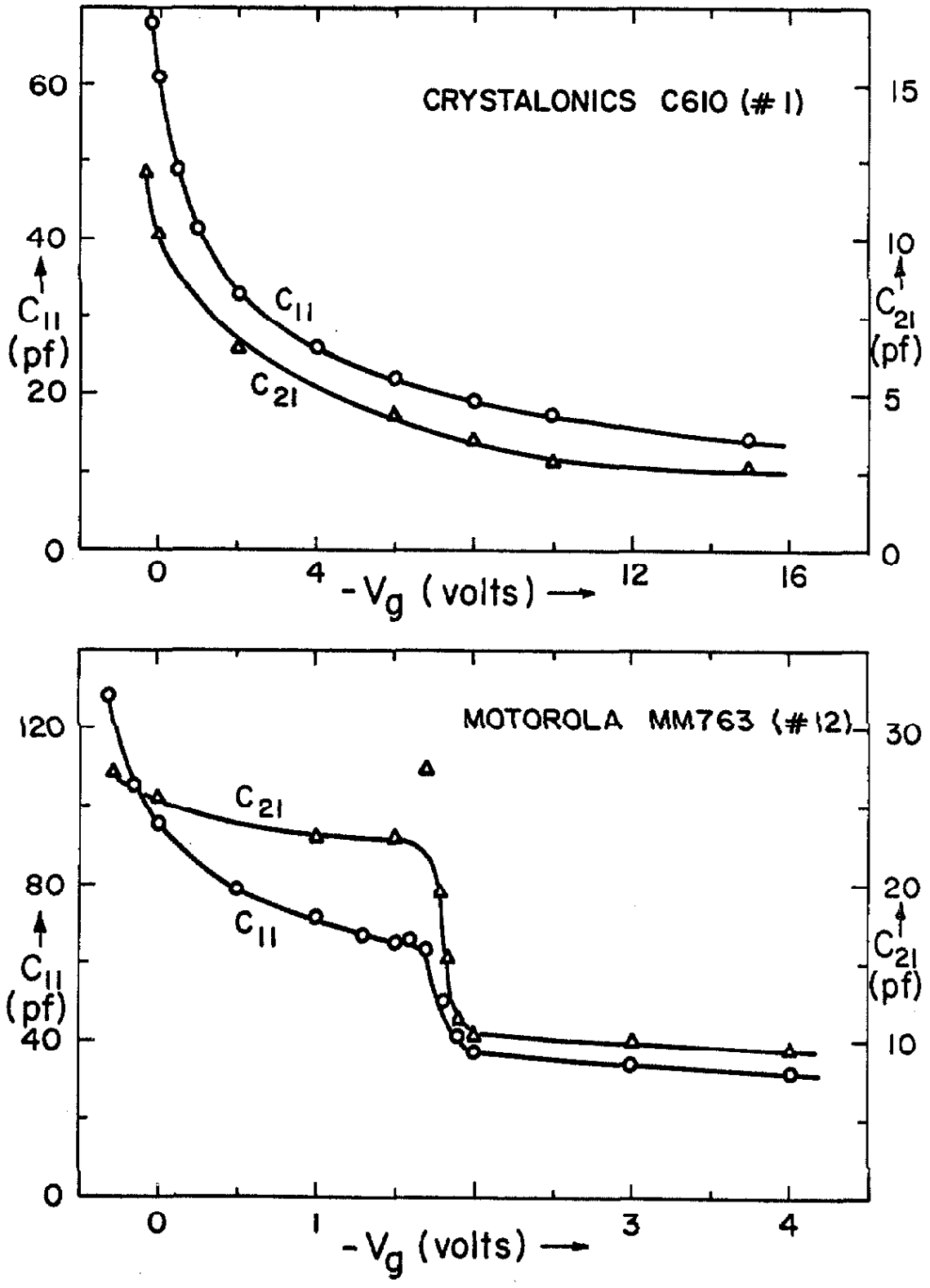


Fig. 6.8. Experimental curves for the equivalent-circuit capacitances beyond pinch-off.

in the C_{21} curve for the Motorola unit of Fig. 6.8.) This discrepancy is partly due to the fact that measurements could not be taken near $s = 0$ $(V_g = -V_c)$, where C_{21} has a steep slope. Nevertheless, units which exhibited reasonably sharp drops in C_{11} and C_{22} at gate pinch-off and which therefore permitted calculation of the ratio C_{21}/C_{11} of Eq. 6.43 yielded values close to the theoretical value 0.40. These data are presented in Table 6.2.

The experimental verification of the existence of C_{21} serves to confirm the validity of the theory developed in this chapter. Of practical importance is the fact that although the susceptance ωC_{21} appears together with the usually large conductance g_m, the deleterious effect of C_{21} will be noticeable at higher frequencies and also at gate biases near pinch-off (where g_m is small). In addition, the mere presence of C_{21} in the equivalent circuit accounts theoretically for the previously unexplained high-frequency falloff in g_m.

6.5. Validity of Results

The equivalent circuit that has been developed from an approximate solution of the differential equation 6.9 is valid for "small signals" and "low frequencies." Equation 6.24 defines "small-signal" by requiring that the a-c components of the terminal voltages be small compared with the d-c components. In this section we shall show that our solution is valid if $\omega\tau_0 \ll 1$, and hence "low frequency" means $\omega \ll 1/\tau_0$.

First consider the case when $s = d$ $(V_d = 0)$. With these

TRANSISTOR		NATURE OF JUNCTION	C_{21}/C_{11} s=1, d>1
Motorola		epitaxial	
MM763	(# 2)		0.48
	(#12)		0.42
MM764	(# 3)		0.43
MM765	(#24)		0.42
Texas Ins.		diffused	
TIX691	(#13)		0.47
	(# 4)		0.45

Table 6.2. Experimental values for the ratio. C_{21}/C_{11} at gate pinch-off. The theoretical value is 0.40.

biases the meaning of the solution of the initial iteration, Eq. 6.11, is not clear, and the accuracy of the final solution, Eqs. 6.18 and 6.20, is in doubt. To a first approximation, however, the channel is uniform, so that the channel current is equal to the current through a reverse-biased junction and is

$$I(x,t) = j\omega v_g(t) \frac{\epsilon A}{a(w-u)}\left(x - \frac{L}{2}\right)$$

$$= -j\omega v_g G_o \tau_o g'(s)\left(\frac{x}{L} - \frac{1}{2}\right) \qquad (6.44)$$

where use has been made of Eqs. 4.40 for $g'(s)$ and 6.7 for τ_o. This current causes a voltage drop Δv between the center of the channel and the drain and source terminals. Thus, the channel is not really uniform, but varies in width because the gate-channel potential is not constant. Figure 6.9 depicts an FET under these conditions. If the channel voltage drop Δv is small compared with the depletion-region voltage drop V_g, then the uniform-channel approximation is satisfactory. The channel resistance (per unit length) is given by

$$R = \left[\mu a A \int_0^u \rho \, dz\right]^{-1}$$

$$\approx [LG_o g(s)]^{-1} = \text{constant} \quad (\text{independent of } x)$$

so that

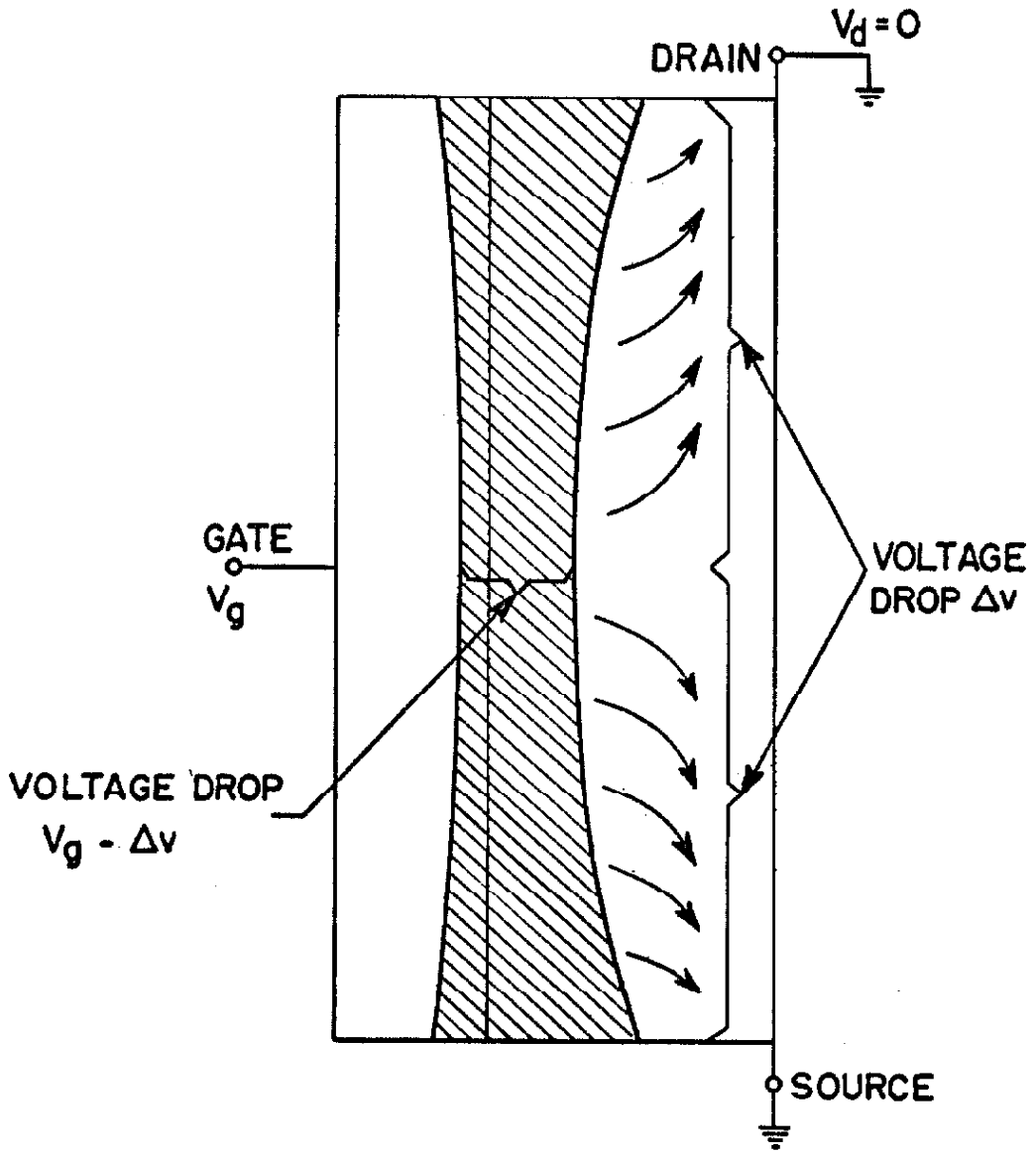


Fig. 6.9. FET with bias $V_d = 0$. The capacitive current across the gate-channel junction is indicated by the curved arrows and causes the voltage drops Δv in the channel.

$$\Delta v = R \int_{L/2}^{L} I(x,t)\, dx$$

$$= -\frac{1}{8}\, j\omega v_g \tau_c \frac{g'(s)}{g(s)}$$

and the channel may be considered uniform if

$$\left|\frac{\Delta v}{V_g}\right| = \frac{1}{8}\,\omega\tau_o \left|\frac{v_g}{V_g}\right| \left|\frac{g'(s)}{g(s)}\right| \ll 1$$

For the doping profiles considered in Chapter V, it is easy to show from Eqs. 5.26 and 5.27 that

$$\frac{1}{1-s} \leq \frac{|g'(s)|}{|g(s)|} \leq \frac{1}{2\sqrt{s}(1-\sqrt{s})}$$

the right-hand side being the value for a step-junction device. Therefore, for practical FET's we merely need

$$\frac{1}{8}\,\omega\tau_o \left|\frac{v_g}{V_g}\right| \frac{1}{2\sqrt{s}(1-\sqrt{s})} \ll 1 \qquad (6.45)$$

Since small-signal operation entails $|v_g/V_g| \ll 1$, inequality 6.45 is violated only for $s \approx 0$ and $s \approx 1$. The violation near $s = 0$ arises because of the small voltage drop across the depletion region and may be disregarded because we have neglected the built-in potential; i.e. $s = 0$ can never be attained in practice. The important violation, near $s \approx 1$, occurs because the channel is

then relatively narrow and the voltage drop Δv correspondingly large. We conclude that the channel is uniform if $\omega\tau_c < 1$, except for gate biases very close to pinch-off.

But, for a uniform channel, integration of Eq. 6.44 gives

$$I_d = -\int_{L/2}^{L} I(x,t)\, dx$$

$$= +\frac{1}{2} j\omega v_g G_o g'(s)$$

$$I_g = -2I_d = -j\omega v_g G_c g'(s)$$

and these equations are identical to those obtained by taking the limit $d \to s$ in the results of Section 6.3:

$$I_d = -j\omega v_g C_{12}(s,s)$$

$$= +\frac{1}{2} j\omega v_g G_o g'(s)$$

$$I_g = j\omega v_g C_{11}(s,s)$$

$$= -j\omega v_g G_o g'(s)$$

Thus, if $\omega\tau_o < 1$, the approximate solutions for I_d and I_g are valid for $s = d$, unless s is very close to 1. By extension, then, if $\omega\tau_o < 1$, the solutions are also valid for $s \approx d$.

Now consider the case when $s \not\approx d$. We have approximately solved the differential equation

$$\frac{\partial^2}{\partial x^2}\left[\int_s^{\vartheta} g(\phi)\, d\phi\right] + \tau_o \frac{\partial}{\partial t}\left[g(\theta)\right] = 0 \qquad (6.46)$$

by assuming that the magnitude of each of the two terms on the left-hand side was large in comparison with the magnitude of the sum of the two terms. We know that our solution is accurate to terms of order $\omega\tau_o$, and that the resulting left-hand side of Eq. 6.46 is actually not zero, but is of order $(\omega\tau_o)^2$. The individual terms, however, are each of order $\omega\tau_o$. Therefore, the above assumption and our approximate solution are valid if*

$$\omega\tau_o \ll 1, \qquad \text{or} \qquad \omega \ll \frac{1}{\tau_o} \qquad (6.47)$$

The second inequality above defines "low-frequency." From Eqs. 5.24 and 5.32 we see that

$$\frac{1}{\tau_o} = \frac{\mu W_c}{L^2} \sim \left.\frac{\varepsilon_m}{C_{11}}\right|_{\substack{s=0 \\ d>1}}$$

Some typical values are given in the next section.

---

* There is another viewpoint from which this condition may be derived: An "exact" solution of Eq. 6.46 could be written as a power series in $\omega\tau_o$; terms of order higher than $(\omega\tau_o)^1$ could be neglected if $\omega\tau_o \ll 1$. Furthermore, since the next term in the approximate solution would introduce components $\pi/2$ out of phase with the components resulting from the terms in $\omega\tau_o$, a sufficient condition for validity might be $\omega < 1/\tau_o$ rather than $\omega \ll 1/\tau_o$. We have been conservative, however, and have used the latter inequality.

Further approximations to the solution of differential equation 6.46 would give rise to additional equivalent-circuit elements. These elements would become important at frequencies higher than those which satisfy Eq. 6.47. But Eq. 6.46 is based on the gradual approximation which itself breaks down at higher frequencies when larger I_g flows; that is, when, in addition to the longitudinal channel resistance and the gate-channel capacitance, the "transverse" channel resistance must be included in the analysis. Thus, the theory developed in this chapter represents the limit of applicability of the gradual approximation.

6.6. Some Further Practical Considerations

In order to inject additional practical feeling for the analysis conducted in the chapter, we shall, in this section, first indicate a more useful form for the forward transfer admittance of an FET and then display typical numerical values for some of the equivalent-circuit parameters.

Beyond pinch-off the forward transfer admittance appearing in the FET equivalent circuit (Fig. 6.6) is*

$$y_{21} = g_m - j\omega C_{21} \tag{6.48}$$

$$= g_m(1 - j\omega C_{21}/g_m)$$

---

* Arguments similar to those which follow apply below pinch-off to the circuit of Fig. 6.2.

A more instructive form for this admittance is the single-pole approximation

$$y_{21} = \frac{g_m}{1 + j\frac{\omega}{\omega_g}} \tag{6.49}$$

with

$$\omega_g \equiv \frac{g_m}{C_{21}}$$

being the radian frequency (dependent on the bias point) at which the magnitude of y_{21} is 3-db below its d-c value. By using Eq. 6.49 rather than 6.48 we can consider g_m to be the only element in the forward transfer admittance. Then Eq. 6.49 provides the visualization and interpretation for the frequency dependence of g_m.

The break frequency ω_g is somewhat greater than the frequency

$$\frac{1}{\tau_o} = \frac{\mu W_c}{L^2} \sim \left.\frac{g_m}{C_{11}}\right|_{\substack{s=C \\ d>1}} \tag{6.50}$$

beyond which, as was shown in the previous section, the circuit of Fig. 6.6 may no longer be valid.* Thus, the single-pole approximation might be inaccurate near ω_g because the assumptions used in

---

* We may interpret the breakdown of the low-frequency equivalent circuit in either of two ways. At high frequencies we can expect additional elements to be of significance, or equivalently, we can consider the existing elements to be frequency dependent.

developing the approximation may become untenable at these higher frequencies. The circumstances are not unlike those occuring in the low-frequency approximation to the collector-emitter current gain α of a conventional transistor

$$\alpha = \frac{\alpha_o}{1 + j\frac{\omega}{\omega_\alpha}}$$

This single-pole approximation, valid at low frequencies, is invalid near the break frequency, since, in particular, the incorrect phase is predicted at $\omega = \omega_\alpha$ [26].

Typical measured values of the external pinch-off voltage, the drain current, and the equivalent circuit parameters beyond pinch-off are presented in Table 6.3 for several commercial FET's. All data are with $V_g = 0$ and with $V_d > V_p$. ($V_g = 0$ is not equivalent to $s = 0$.) For each transistor, the listed pinch-off voltage, together with the knowledge that the built-in potential lies within the approximate range $\sim 0.5v < |V_c| < \sim 1.0v$, provides an estimate of the value of s that corresponds to $V_g = 0$. The capacitance values are those of the "intrinsic" devices themselves; the constant stray capacitances that remain beyond gate pinch-off have been subtracted. The break frequency ω_g is equal to g_m/C_{21}. The ratio $g_m/2\pi C_{11}$ is of the order of the validity frequency $1/2\pi\tau_o$ and, incidentally, is equal to the gain-bandwidth product. As expected on theoretical grounds, ω_g is greater than g_m/C_{11} for each FET.

TRANSISTOR	NATURE OF JUNCTION	V_p (v)	I_d (ma)	g_m (μmho)	C_{11} (pf)	C_{21} (pf)	$\omega_g/2\pi$ (Mc)	$g_m/2\pi C_{11}$ (Mc)
Amelco FG34 (#54)	diffused	-3.7	4.17	3190	91	13.2	38	5.6
Crystalonics C610 (# 1)	alloy	-10.3	0.614	204	43.5	7.4	4.4	0.75
Motorola MM763 (#12)	epitaxial	- 1.9	1.64	1710	61	15.5	18	4.5
MM764 (# 3)		- 2.7	2.90	2360	92.5	15	25	4.0
MM765 (#24)		- 4.5	6.50	3160	61	15	34	8.2
Texas Ins. TIX691 (# 4)	diffused	+ 1.7	3.24	4030	29.5	8	80	22

Table 6.3. Values for some FET parameters at $V_g = 0$. Except for the unit TIX691, the particular values shown for the incremental parameters were measured at 1Mc; for TIX691 the 100kc values are listed (see text). The drains were biased beyond pinch-off ($V_d = 25v$ for Amelco and Crystalonics FET's, $|V_d| = 10v$ for Motorola and Texas Ins. FET's). Experimental setups are shown in Appendix B.

On all the transistors listed in the table, g_m and C_{11} were constant up to approximately 2Mc, the upper measuring frequency of the test equipment. On all the units except Texas Instruments TIX691, C_{21} was also constant up to 2Mc. Of these units, all except Crystalonics C610 have $g_m/2\pi C_{11} > 2Mc$; i.e., as predicted, the equivalent circuit is valid at low frequencies. Unit C610 has $g_m/2\pi C_{11} < 2Mc$, indicating that for this unit the low-frequency circuit may be employed well beyond the validity frequency $1/\tau_o$. (This statement might apply to the other units as well, but measurements could not be taken at sufficiently high frequencies to obtain verification.) The applicability of the low-frequency circuit at high frequencies is not inconsistent with the derivations of the previous section, since they predict validity below $1/\tau_o$, not invalidity above $1/\tau_o$.

On TIX691, C_{21} was not constant with frequency. (The tabulated capacitance was measured at 100kc; the 50kc and 100kc values were approximately equal.) However, the magnitudes of g_m and C_{11}, both of which were constant, predicted a validity frequency $1/2\pi\tau_o \gg 2Mc$. Theoretically, then, C_{21} should have been constant. Although no firm explanation can be given for this apparent contradiction, there were two potential sources of appreciable error in the measurements on this unit. First, the fact that the small capacitance C_{21} appears in parallel with the large conductance* g_m precludes measurements

---

* The data in Table 6.3 shows that these values are respectively small and large both relative to each other and relative to other units.

at lower frequencies. Thus, measurements were confined to the narrow range 50kc - 2Mc (limit of equipment), the balance at 50kc (where the conductance to susceptance ratio $g_m/\omega C_{21}$ was greater than 10^3) not being very sharp. Second, on this unit - and only on this unit - the interelectrode capacitance shunting the intrinsic capacitance C_{21} was large in comparison with C_{21}. In fact, the interelectrode capacitance was almost double the intrinsic capacitance. Hence, a small error in the total forward transfer capacitance would produce a large error in C_{21}.

The data in Table 6.3, though representative, only cover part of the wide numerical ranges available from commercial FET's. It should be clear, however, that the conclusions derived in this chapter apply with equal force to all parameter values within these ranges.

6.7. Conclusions

We have set up and solved by an approximate iterative method the differential equation for the channel potential in an arbitrarily doped FET. The chief result of the solution is the small-signal, low-frequency equivalent circuit, "low-frequency" meaning frequencies satisfying $\omega \ll 1/\tau_o$, where τ_o is roughly equal to the transit time τ_t at zero gate bias (in the pinch-off range). The equivalent circuit contains a drain-source capacitance and a forward transfer capacitance in addition to the expected source-gate and drain-gate capacitances, forward transconductance, and output conductance. Although a cursory investigation of FET operation does not indicate the existence of these two "new" elements, the hindsight gained from

the analysis performed in Section 6.2 suggested physical explanations that are quite simple. Measurements on numerous types of FET's corroborated the theoretical results and prove that these elements are indeed present in the equivalent circuit. Bearing in mind the simplicity of the gradual approximation, the idealizations inherent in the model that was used, and the generality of the analysis, we may conclude that the qualitative and quantitative experimental results are highly satisfactory.

In addition to other effects, the breakdown of the gradual approximation at higher frequencies necessitates a more detailed approach if a high-frequency equivalent circuit is to be developed Thus, new methods must be evolved if further analysis is to be performed, the utility of the gradual approximation having been exhausted.

CHAPTER VII

CONCLUSIONS

In this work the operation of FET's has been considered from a completely general standpoint in order to ascertain and compare the properties common to all FET's.

The principal feature of FET performance that has been established theoretically is the qualitative and quantitative similarity in external characteristics manifested by transistors possessing considerable differences in internal characteristics. The final results have been developed in a sequence of connected analyses. First, background material was presented in Chapter II, where a step-junction FET was investigated. The salient aspects of FET behavior were educed, including a detailed consideration of the charge-capacitances, since these elements hitherto had not been discussed. Second, prior to the development of the general treatment for arbitrary impurity profiles, the validity of the step-junction analysis was examined, and the conclusion drawn that the simplifying assumptions and approximations used in Chapter II are satisfactory for practical purposes. This conclusion reaffirmed the usefulness of the step-junction theory and provided confidence for the forthcoming general theory.

With the preparatory analysis of Chapter II, the transition to an arbitrarily doped FET in Chapter IV was accomplished in a fairly straightforward manner. The primary advantage of the general

formulation is that it makes general properties apparent. The important parameters describing FET operation have been written in a compact, symmetrical form entirely (apart from constant multiplicative factors) in terms of the normalized transconductance function $g(\theta)$ [defined by Eq. 4.28] and the normalized gate-source and gate-drain potentials s and d. From these equations it was easy to deduce that all FET's exhibit similar qualitative behavior. Hence, the key results derived for a step-junction FET - such as the form of the drain characteristics, the configuration of the equivalent circuit, and the properties of the charge-capacitances - can be applied directly to an FET with an arbitrary junction.

Finally, in order to obtain quantitative comparisons between different types of FET's, some weak restrictions are assumed in Chapter V on the previously arbitrary impurity profile. The resulting restricted theory yields bounds on some important FET parameters beyond pinch-off: the transconductance, the drain current, the input charge-capacitance, several figures of merit, and the bias point for temperature-stable drain current. The conditions imposed on the impurity profile are weak enough that essentially all manufactured FET's (except cylindrical) are embraced by the restricted theory. Moreover, the derived bounds, which for each parameter are the step- and delta-junction formulae, cannot be tightened without eliminating some practically important profiles from the theory. Thus, the quantitative conclusion is that for practical purposes FET characteristics may be considered independent of the doping profile.

This conclusion has many implications. Technologically, it is useful because it means that design and fabrication efforts can be concentrated on the other FET parameters: the geometry, the channel conductance, and the pinch-off voltage. For application purposes the conclusion is significant because it justifies the approximation of the equations governing an arbitrary device by those governing either of two simple devices - a step-junction or a delta-junction device - or, if slightly better accuracy is required, by a power-law. Theoretically, the conclusion is important because it obviates the solution of analytically intractable devices; also, the fact that the delta-junction results are a good approximation to the results for all other FET's justifies, for conceptual purposes, the simple charge-control derivation of Section 4.1, in which the nature of the gate-channel junction is not considered.

The other major objective of this work is the development of the real equivalent circuit of an FET. This objective was achieved in Chapter VI, where the general treatment of Chapter IV was extended by including the capacitive gate-channel current. The approximate solution that was obtained yields the first-order equivalent circuit - valid for small signals and low frequencies - and adds two elements, a drain-source capacitance and a forward transfer capacitance, to the general circuit of Chapter IV. The predicted circuit, including most notably the two new elements, has been substantiated by measurements on a wide variety of FET's.

The equivalent circuit of Chapter VI breaks down at higher frequencies, apparently because of the inexact nature of the solution to the "exact" differential equation. However, the underlying cause for the breakdown is the failure of the gradual approximation at higher frequencies. Thus, since the approximate solution of Chapter VI contains all the information that was obtained in the general analysis of Section 4.2, all the results that have been derived are strictly valid only at low frequencies ($\omega \ll 1/\tau_o = \mu W_o/L^2$). High-frequency results can only be achieved with revision or replacement of the gradual approximation.

The general approach used in Chapters IV - VI, and, in particular, the techniques employed in Chapter V to obtain bounds on the important FET parameters may be fruitfully applied to areas other than those that have been explored here. A specific problem that warrants solution is a general formulation of the properties of cylindrical FET's, with the intent of obtaining limits analogous to those derived for planar FET's. With the properties of both planar and cylindrical devices in general terms, the relative advantages and disadvantages of each type of structure could be easily determined.

An important parameter that has not been derived quantitatively in a satisfactory manner is the finite output conductance G_{22} beyond pinch-off. Although G_{22} is usually a second-order quantity, in some circumstances its effect can be significant. The analysis beyond pinch-off that has been performed by Shockley [4] is based on some rather severe assumptions that limit its practical applicability. In addition, the solution is difficult to use even in simple cases

because involved calculations are required. The simple model that has been used throughout this work is inadequate for the explanation of G_{22}; either additional principles must be introduced or some of the first-order assumptions must be discarded. (Although a non-abrupt boundary on the space-charge region implies a finite G_{22} beyond pinch-off, the magnitude of this G_{22}, as predicted by the analysis of Section 3.5, is much too small to account for the observed conductances.) A possible starting point might be based on the approach Grosvalet, et. al. [27] used to explain the conditions actually prevailing in the channel when an FET is operating in the pinch-off state. Their technique was to incorporate the concept of velocity-limiting fields that are supposedly present where the channel is very narrow. If this approach is practicable for the explanation of G_{22}, a solution must be developed that would be accurate enough to be generally applicable, yet simple enough to avoid burdensome computations.

One further point that should be more firmly established involves the inclusion of double-diffused structures in the restricted theory of Chapter V. The plausibility arguments that were presented in Section 5.1 may be sufficient, but a more rigorous foundation would be desirable.

APPENDIX A

BOUNDS ON THE NORMALIZED INPUT CHARGE-CAPACITANCE

Here we prove the assertion made in Section 5.4 that the normalized input charge-capacitance lies between the step- and delta-junction bounds

$$\frac{1 + \sqrt{s}}{(1 + 2\sqrt{s})^2} \leq \frac{C_{11}^*}{C_0^*} \leq 1 \tag{A.1}$$

if

$$(1 - \sqrt{s}) \leq g(s) \leq (1 - s) \tag{A.2}$$

We have shown in Section 5.5 that

$$\frac{2}{3} \frac{(1 - \sqrt{s})(1 + 2\sqrt{s})^2}{(1 + \sqrt{s})} \leq \frac{g_m}{C_{11}^*} \leq \frac{3}{2} (1 - s)$$

The lower bound in this equation has the value $2/3$ at $s = 0$, the value 0 at $s = 1$, and is concave downward in the range $0 \leq s \leq 1$. The upper bound is a straight line passing through $3/2$ at $s = 0$ and 0 at $s = 1$. Therefore, if each bound is normalized to its value at $s = 0$, we clearly have

$$(1 - s) \leq \frac{(g_m/C_{11}^*)}{(G_o/C_o^*)} \leq \frac{(1 - \sqrt{s})(1 + 2\sqrt{s})^2}{(1 + \sqrt{s})} \tag{A.3}$$

If the identity $g_m/G_o \equiv g$ and Eq. A.2 are inserted in A.3, Eq. A.1 results.

APPENDIX B

EXPERIMENTAL SETUPS

Figures B.1 through B.4 show the circuit configurations that were used to measure the short-circuit admittances y_{ij} of the FET's. (In the symbol for the FET the arrow is located near the source and indicates the direction of conventional current flow.) A Wayne Kerr Radio Frequency Bridge type B601 was used in conjunction with an oscillator and a null detector. The particular settings on the bridge multiplier switches and the particular bridge output tap that were chosen depended on the magnitude of the admittance being measured.

The relatively large drain current (several ma) at small V_g on some units caused an appreciable voltage drop in the bridge transformers (on all measurements except y_{11}); when this occurred, the drain bias was adjusted in order to maintain the drain-source voltage at the desired value. As a check on the possible effects of the direct current flowing through the bridge transformers, several different biasing arrangements were tried; the results from each arrangement were in excellent agreement.

Because the forward transconductance of an FET has a sign opposite to that of a passive conductance (cf. Fig. 6.2; positive v_g results in positive i_d), the bridge was modified internally in order to read a "negative" conductive component in the y_{21} measurement. Thus, R_{21} in Fig. B.3 is shown connected to the "common" terminal

of the bridge. For slightly better accuracy (and as a check), the transconductance, as well as the output conductance, was also measured at 1kc on a General Radio Vacuum-Tube Bridge type 561-D.

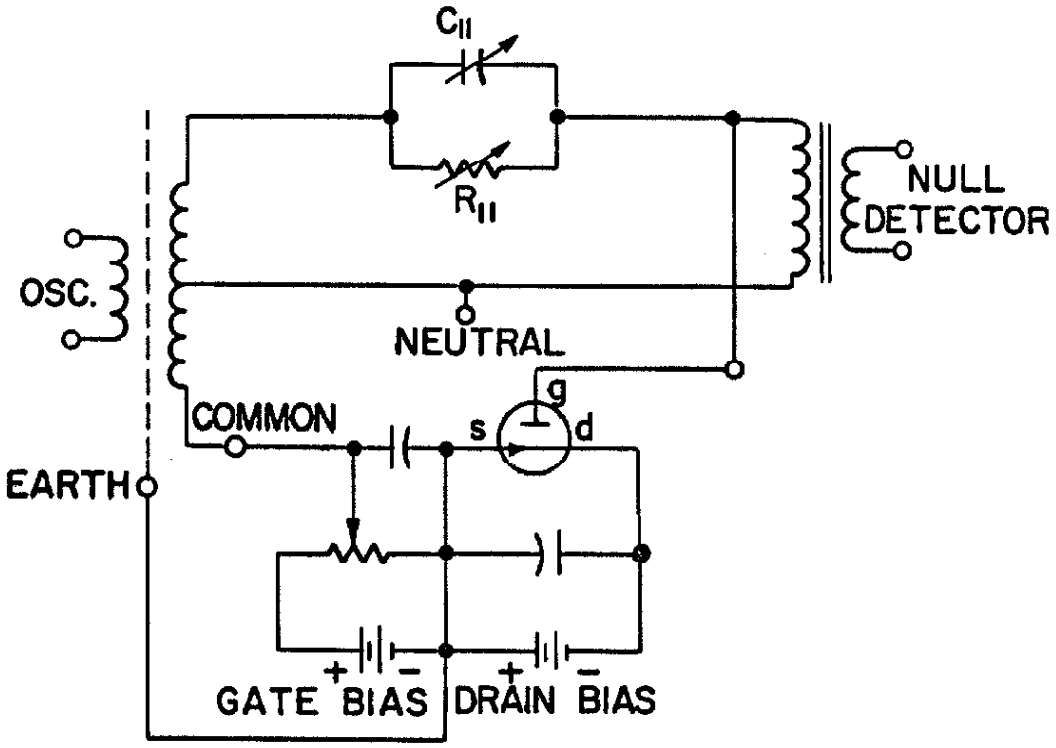


Fig. B.1. Experimental setup for measuring y_{11}. Polarities are shown for a p-channel FET.

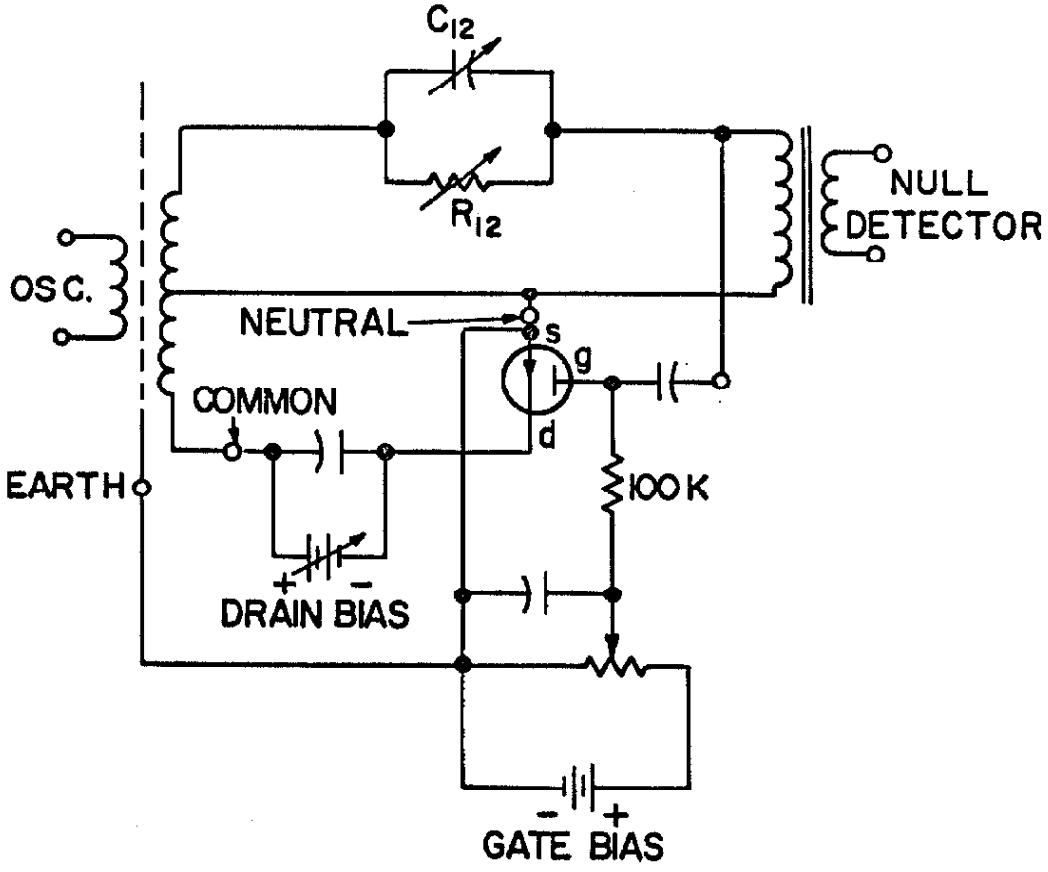


Fig. B.2. Experimental setup for measuring y_{12}. Polarities are shown for a p-channel FET.

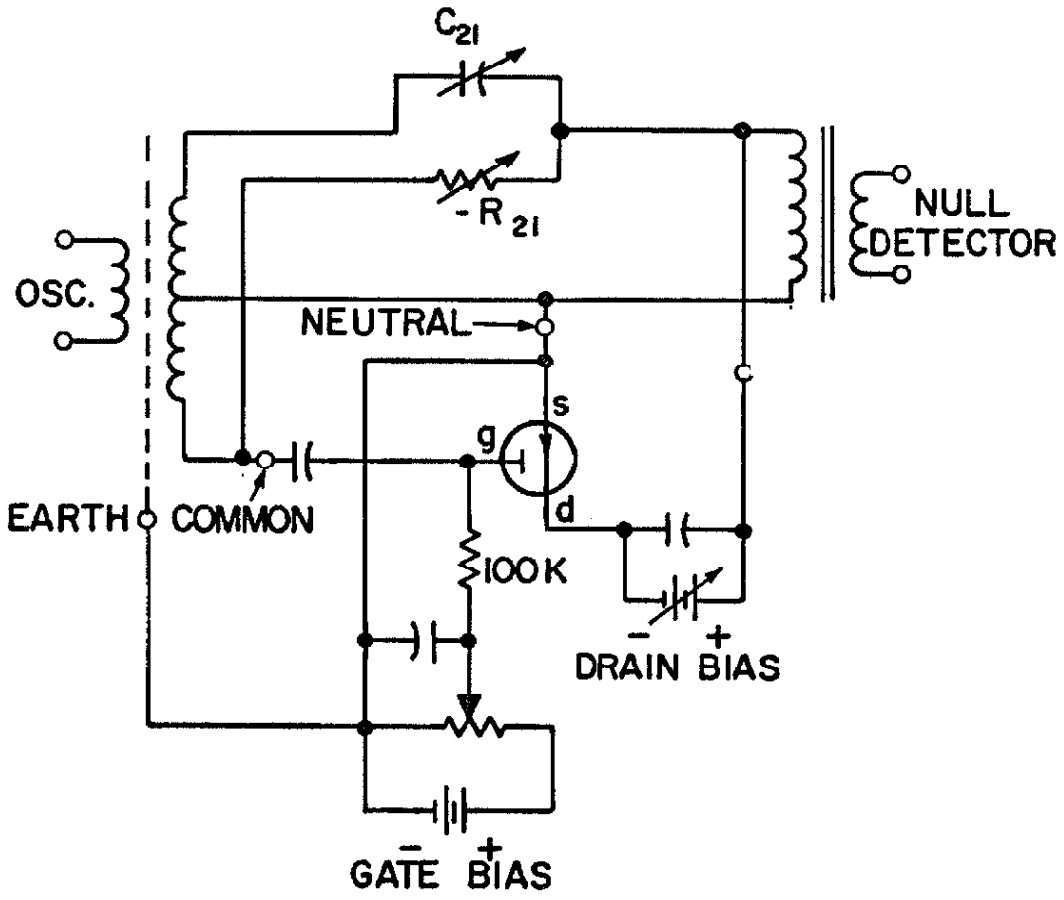


Fig. B.3. Experimental setup for measuring y_{21}. Polarities are shown for a p-channel FET.

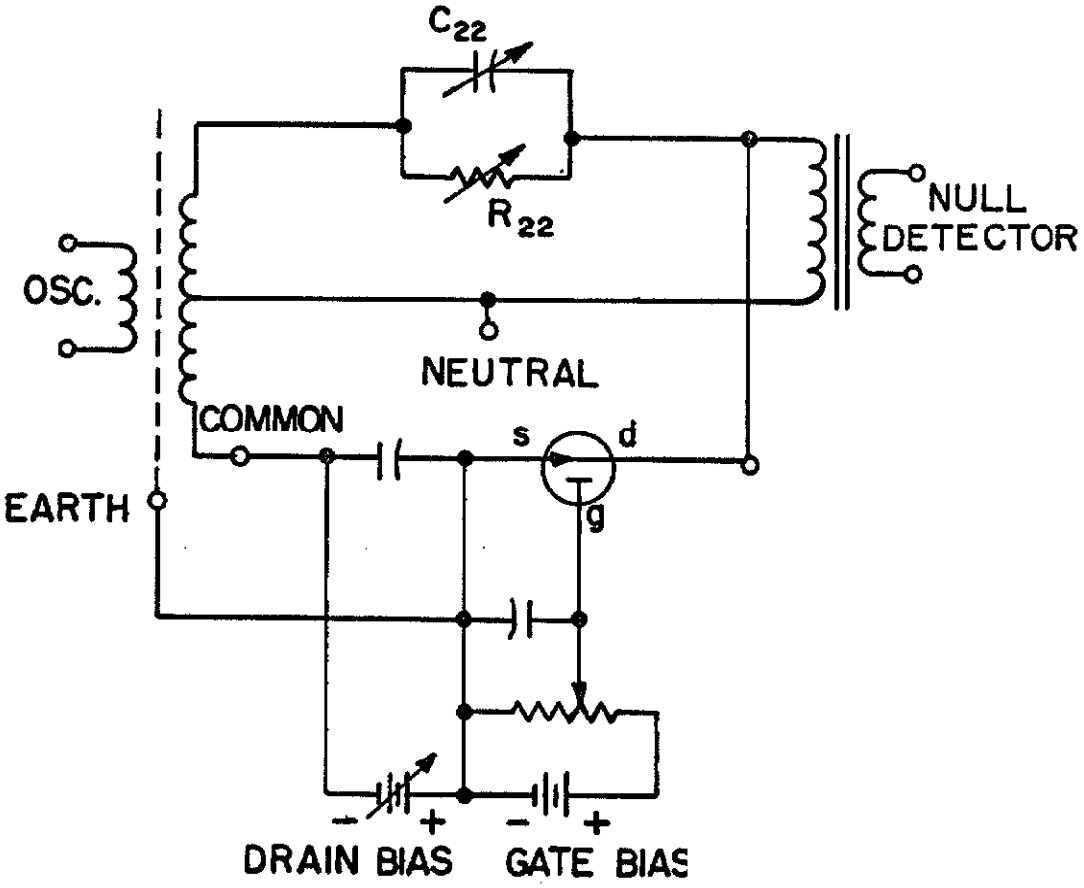


Fig. B.4. Experimental setup for measuring y_{22}. Polarities are shown for a p-channel FET.

LIST OF PRINCIPAL SYMBOLS

The subscript s, d, or g pertains to the source, drain or gate.

			Section in Which Symbol is Defined.
A	=	Depth of FET.	2.1
C (C_{ij})	=	Capacitance (at terminals ij).	2.3
C^* (C^*_{11},C^*_{22})	=	Charge-capacitance (at input, at output).	2.3
C^*_c	=	$C^*_{11}(s,d)$ for s = 0, d = 1.	2.3
E	=	Electric field.	2.2
G_o	=	Channel conductance in the absence of biases.	2.3
G_{22}	=	Output conductance.	2.3
I	=	Channel current; also, drain current under the assumption of zero gate current.	2.3
I_o	=	Maximum drain current, (provided that the gate-channel junction is reverse biased).	2.3
$I_{\perp}$	=	D-C drain current.	6.2
L	=	Distance between source and drain (length of channel).	2.1
Q	=	Charge in transit between source and drain.	2.3
T	=	Absolute temperature.	3.5
V	=	"External" potential (neglects built-in potential).	3.2
V_c	=	Built-in potential.	3.2

V_p	=	External pinch-off voltage.	3.2
W	=	Magnitude of internal potential.	3.2
W_o	=	Magnitude of internal pinch-off voltage.	3.2
a	=	Width of channel material.	2.1
b	=	Width of channel; space-charge-region boundary in the channel material.	2.1
c	=	Space-charge-region boundary in the gate material.	2.1
c_m	=	Maximum penetration of space-charge region into gate material (corresponds to b = 0).	4.2
d	=	Magnitude of normalized internal gate-drain potential.	3.2
g	=	Normalized transconductance function; normalized transconductance beyond pinch-off.	4.2
g_m	=	Transconductance.	2.3
i	=	A-C component of current.	2.3
s	=	Magnitude of normalized internal gate-source potential.	3.2
u	=	b/a	2.3
v	=	A-C component of voltage.	2.3
w (w_m)	=	c/a (c_m/a)	4.2
x,y	=	Coordinates.	2.1
z	=	y/a	4.2
ϵ	=	Permittivity.	2.2
ϑ	=	Magnitude of normalized gate-channel potential.	4.2
μ	=	Mobility of majority carriers in the channel.	2.3

ρ	=	Excess density of donors over acceptors, multiplied by the magnitude of the electronic charge.	2.1
ρ_m	=	Mobile charge density.	2.1
τ_o	=	$L^2/\mu W_o$	6.1
τ_-	=	Average transit time.	4.1
ω	=	Radian frequency.	2.3

REFERENCES

[1] J. E. Lilienfeld, "Device for controlling electric current," U.S. Patent No. 1,900,018.

[2] W. Shockley and G. L. Pearson, "Modulation of Conductance of Thin Films of Semiconductors by Surface Charges," Phys. Rev., vol. 74, pp. 232-233; July 15, 1948. Also, see W. Shockley, Electrons and Holes in Semiconductors, D. Van Nostrand Co., Inc., Princeton, N.J., pp. 29-34; 1950.

[3] J. Bardeen, "Surface States and Rectification at a Metal Semiconductor Contact," Phys. Rev., vol. 71, pp. 727-727; May 15, 1947.

[4] W. Shockley, "A Unipolar 'Field-Effect' Transistor," Proc. IRE, vol. 40, pp. 1365-1376; November, 1952.

[5] See e.g., H.A.R. Wegener, "The Cylindrical Field-Effect Transistor," IRE Trans. on Electron Devices, vol. ED-6, pp. 442-449; November, 1959; and S. Teszner, "Sur les Nouvelles Structures du Tecnetron," Onde Elect., vol. 41, pp. 307-320; April, 1961.

[6] R. D. Middlebrook, "A Modern Approach to Semiconductor and Vacuum Device Theory," Proc. IEE, vol. 106, Part B, Supplement No. 17, pp. 887-902; May, 1959.

[7] G. C. Dacey and I. M. Ross, "Unipolar 'Field-Effect' Transistor," Proc. IRE, vol. 41, pp. 970-979; August, 1953.

[8] G. C. Dacey and I. M. Ross, "The Field Effect Transistor," Bell Syst. Tech. J., vol. 34, pp. 1149-1189; November, 1955.

[9] I. Richer, "Input Capacitance of Field-Effect Transistors," Proc. IEEE, vol. 51, pp. 1249-1250; September, 1963.

[10] G. C. Onodera, personal communication.

[11] G. C. Onodera, W. J. Corrigan, and R. M. Warner, Jr., "Silicon Field-Effect Transistor with Internal Epitaxial Channel," Proc. IRE, vol. 50, p. 1824; August, 1962.

[12] W. R. Smythe, Static and Dynamic Electricity, McGraw-Hill Book Co., Inc., New York, N.Y., p. 234; 1950.

[13] H. R. Kaiser and P. S. Castro, "Capacitance Between Thin Film Conductors Deposited on a High-Dielectric-Constant Substrate," Proc. IRE, vol. 50, pp. 2142-2143; October, 1962.

[14] H. B. Dwight, Tables of Integrals and Other Mathematical Data, The Macmillan Co., New York, N.Y., p. 171; 1957.

[15] P. N. Wolfe, "Capacitance Calculations for Several Simple Two-Dimensional Geometries," Proc. IRE, vol. 50, pp. 2131-2132; October, 1962.

[16] E. J. Ryder, "Mobility of Holes and Electrons in High Electric Fields," Phys. Rev., vol. 90, pp. 766-769; June 1, 1953.

[17] W. Shockley, "The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors," Bell Syst. Tech. J., vol. 28, pp. 435-489; July, 1949.

[18] R. D. Middlebrook, "A Simple Derivation of Field-Effect Transistor Characteristics," Proc. IEEE, vol. 51, pp. 1146-1147; August, 1963.

[19] E. C. Johnson and A. Rose, "Simple General Analysis of Amplifier Devices with Emitter, Control and Collector Functions," Proc. IRE, vol. 47, pp. 407-418; March, 1959.

[20] R. R. Bockemuehl, "Analysis of Field Effect Transistors with Arbitrary Charge Distributions," IEEE Trans. on Electron Devices, vol. ED-10, pp. 31-34; January, 1963.

[21] J. A. Hoerni and B. Weir, "Conditions for a Temperature Compensated Silicon Field Effect Transistor," Proc. IEEE, vol. 51, pp. 1058-1059; July, 1963.

[22] L. J. Sevin, "Behavior of Field-Effect Transistor Characteristics with Temperature," Technical Memo, Texas Instruments, Inc.; July 1, 1963.

[23] E. M. Conwell, "Properties of Silicon and Germanium: II," Proc. IRE, vol. 46, pp. 1281-1300; June, 1958.

[24] R. D. Middlebrook and I. Richer, "Limits on the power-law exponent for field-effect transistor transfer characteristics," Solid-State El., vol. 6, pp. 542-544; September, 1963.

[25] I. Richer and R. D. Middlebrook, "Power-Law Nature of Field-Effect Transistor Experimental Characteristics," Proc. IEEE, vol. 51, pp. 1145-1146; August, 1963.

[26] R. B. Hurley, Junction Transistor Electronics, John Wiley and Sons, Inc., New York, N.Y., p. 241; 1958.

[27] J. Grosvalet, C. Motsch, and R. Tribes, "Etude de Phénomènes Fondamentaux Déterminant les Caractéristiques des Eléments à Effet de Champ," Ann. Radioélect., vol. 17, pp. 265-279; October, 1962.