The Field Effect Transistor

By G. C. DACEY and I. M. ROSS

(Manuscript received May 31, 1955)

Previous work on field-effect transistors considered the performance of the device when operated with electric fields in the channel below the critical field, E_c, where the mobility of carriers becomes dependent on field. This work is reviewed and it is shown that, in this range of operation, the frequency cut-off, f, and transconductance, g_m, of the device increase with increasing values of electric field.

New theory is derived for the performance with electric fields greater than E_c, where the mobility is proportional to $E^{-1/2}$. It is shown that, although both f and g_m continue to increase with electric field in this range, the corresponding increase in the power dissipated is so rapid that such designs are unattractive. It is concluded that a good compromise is to operate with the average channel field equal to E_c. The performance in this particular case is considered in detail and the results summarized in a design nomograph. It is found that f is inversely proportional to the "pinch-off" voltage. The pinch-off voltage cannot, however, be made indefinitely small because the gate junction must be in the saturated condition. A reasonable estimate of the minimum voltage is ½ volt and this leads to a maximum value of f of 1,000 mc/s.

A description is given of the fabrication and performance of several field-effect transistors operating in both the constant and non-constant mobility ranges. It is shown that the performance of these units is in agreement with theory. One of these units had a frequency cut-off of 50 mc/s with a transconductance of 1.6 ma/v when operated at 40 volts and 40 ma.

1. INTRODUCTION

Two papers have already appeared on the field-effect transistor,[1, 2] in which the basic features of the theory were presented and their experimental verification discussed. This present paper describes certain additions to the theory and the experimental verification thereof which arise when attempts are made to realize the ultimate in high-frequency performance. In particular the effect of non-constant mobility at high

electric fields is analyzed in some detail and shown to be a governing consideration for some designs.

An attempt has been made to make the present paper sufficiently complete that it will not be necessary to refer to the previously published literature. For that reason certain of the previously published results are quoted as a starting point for the extended theory. The paper is divided into two parts. The first part presents the complete theory as of its present state of development, and the second part describes the most recent advances in the experimental verification of this theory.

Part I, Theory

2. A QUALITATIVE THEORY

In essence, a field-effect transistor can be regarded as a structure containing a semi-conducting current path, the conductivity of which is modulated by the application of a transverse electric field. In particular consider the structure shown in Fig. 1. The device as shown consists of a slab of n-type semi-conductor with an ohmic contact at each end, and two p-type contacts on opposite sides. These p-type regions are called the "gates". Consider what happens if the gates are shorted to the left hand end of the n-type slab and a positive potential, V_0, applied to the right hand end. A current, I_0, will flow between the ohmic contacts and will consist of a flow of electrons from left to right. The left hand contact

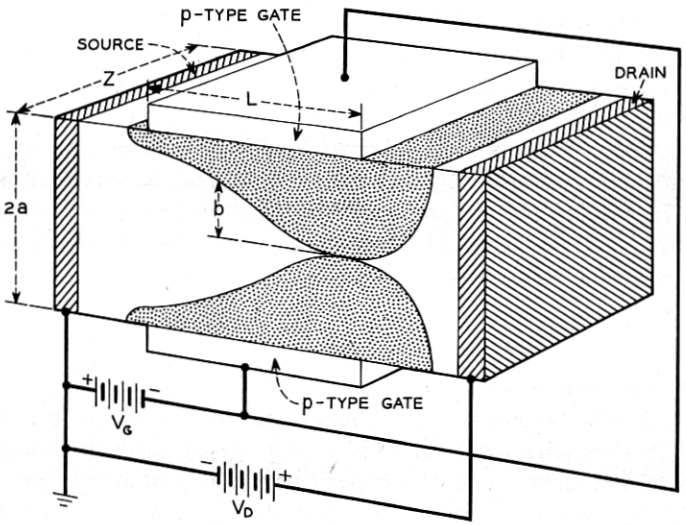

Fig. 1 — Schematic diagram of the field-effect transistor.

is the source of these electrons and the right hand contact drains them out of the material. The contacts are therefore referred to as "the source" and "the drain" and are marked accordingly in the figure. Because the n-type material has resistance, the flow of current will produce an IR drop, the potential becoming more positive toward the drain; but, since the gates are connected to the source, the same potential will appear across the p-n junctions and will bias them in the reverse direction. Space-charge regions will penetrate into the n-type material, becoming wider towards the drain since higher bias results in thicker space-charge regions. In consequence the current is confined to flow in a wedge-shaped region which is called "the channel". As the drain voltage is increased, the channel becomes narrower and the source to drain resistance higher, until at a voltage W_0 the condition is reached in which the space-charge regions from opposite gates meet. This is the condition represented in Fig. 1 where the channel is shown white and the space-charge regions are dotted. This is called "the pinch-off" condition. Beyond pinch-off the current is essentially saturated at a value I_{D0}, further increase of drain voltage resulting in only small increases in current, because most of the increased voltage appears across the space-charge region near the drain. The drain characteristic is thus of the form shown in the upper curve of Fig. 2.

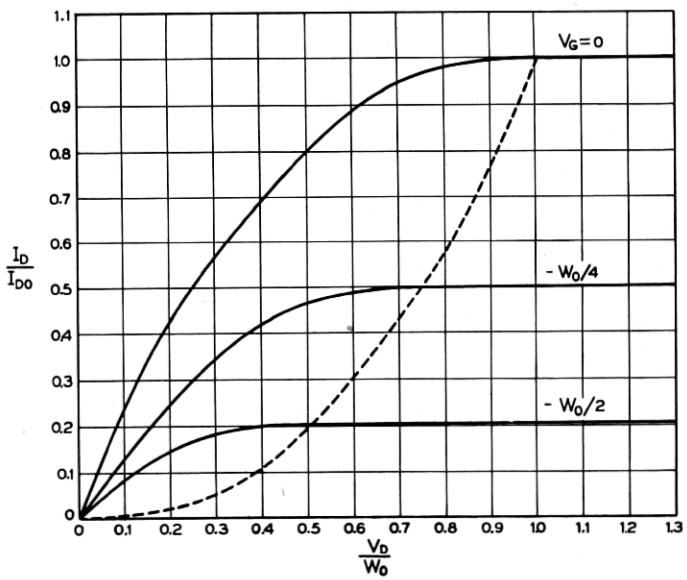

Fig. 2 — Theoretical drain characteristics for the constant-mobility case.

Now suppose that the gates are biased negatively with respect to the source with a voltage V_G. Under these conditions the magnitude of the IR drop necessary to produce pinch-off will be smaller since part of this voltage is already supplied by the bias and current saturation will accordingly occur at lower values of drain voltage and current. The drain characteristics with gate voltage as a parameter are therefore of the form shown in Fig. 2. These characteristics are similar to those of a pentode tube, and the field-effect transistor can be similarly used to obtain amplification.

The operation of the field-effect transistor depends essentially on the presence of just one type of carrier, the majority carrier. For this reason it may be called a "unipolar" transistor. The junction transistor, however, depends for its operation on both types of carrier and is hence a bipolar transistor. In the field-effect, current is carried by the majority carrier drifting in an electric field, while for a junction transistor the current is carried by the minority carrier diffusing in an essentially field-free region. Since drift velocities can be very much higher than diffusion velocities, the transit time for similar dimensions can be shorter in the field-effect than in the junction type. Hence we would expect that with similar dimensions the field-effect device would be capable of operating at higher frequencies. This conclusion is borne out by the theory.

3. THE IDEAL THEORY

The basic theory of the field-effect transistor has been presented by W. Shockley[1] and discussed by the authors.[2] For clarity and completeness the results will be given here.

Let the dimensions of the specimen and the applied voltages be as shown in Fig. 1. Let σ_0 be the conductivity of the n-type material and ρ_0 the donor charge-density. Shockley has shown that, if the wedge-shaped channel referred to above narrows sufficiently slowly, then, for drain voltages less than the saturation value, V_D, the current (per unit length in the Z direction) is given by*

$$I_D = (1/L)\,[J(V_D - V_G) - J(V_s - V_G)] \tag{3.1}$$

where the function J is given by

$$J(x) = 2\sigma_0 a x[1 - (\tfrac{2}{3})\,(2xK/\rho_0 a^2)^{1/2}] \tag{3.2}$$

in which x has the dimensions of voltage and K is the dielectric constant.

* W. Shockley has discussed the case of a p-type channel. For the n-type channel appropriate changes of sign have been made.

For germanium

$$K = 1.42 \times 10^{-12} \text{ farads/cm}$$

Above the saturation voltage, V_D, the current is substantially constant.

It will be convenient to introduce certain natural parameters defined as follows

$$\begin{aligned} W_0 &= \rho_0 a^2/2K \\ E_0 &= \rho_0 a/K = 2W_0/a \\ g_0 &= 2\rho_0\mu_0 a = 2\sigma_0 a \\ I_0 &= g_0 E_0 = 2\rho_0^2\mu_0^2/K \\ \tau_0 &= a/\mu_0 E_0 = K/\mu_0\rho_0 = K/\sigma_0 \end{aligned}$$

Note that W_0 is the bias voltage required between gate and channel to produce a space-charge region of thickness a. Hence, with the gates shorted to the source, W_0 is the saturation value of V_D. g_0 is the conductivity of the channel for zero gate and drain voltages. For small applied voltages the resistance, R_0, of the channel is

$$R_0 = L/2a\sigma_0$$

In Fig. 2 we have shown in terms of these reduced parameters the drain voltage-current characteristics as obtained from (3.1) and (3.2). From (3.1) and (3.2) we obtain the transconductance g_m:

$$\begin{aligned} g_m &= \left.\frac{\partial I_D}{\partial V_G}\right|_{V_{D-\text{const}}} \\ &= (2\sigma_0 a/LW_0^{1/2})[(V_D - V_G)^{1/2} - (V_s - V_G)^{1/2}] \end{aligned} \tag{3.3}$$

Now if in particular we take $V_s = 0$ and $(V_D - V_G) = W_0$ corresponding to grounded source and operation in the pinch-off region, (3.3) reduces to

$$\begin{aligned} g_{mG} &= (2\sigma_0 a/L)\,[1 - (-V_G/W_0)^{1/2}] \\ &= g_{m0}[1 - (-V_G/W_0)^{1/2}] \end{aligned} \tag{3.4}$$

where we have introduced

$$g_{m0} = \frac{2\sigma_0}{L} \tag{3.5}$$

the maximum transconductance.

The saturation drain current I_{DG} which flows for any gate voltage

can be obtained from (3.1) and (3.2) by setting $V_s = 0$, $(V_D - V_G) = W_0$. Doing this yields

$$I_{DG} = (g_{m0}W_0/3)\,[1 + (V_G/W_0)\,(3 - 2\sqrt{-V_G/W_0})] \tag{3.6}$$

We see from (3.6) that the current is completely cut off for a gate voltage of $V_G = -W_0$. Maximum current is

$$I_{D0} = g_{m0}W_0/3 \tag{3.7}$$

obtained for zero bias on the gate.

The frequency response of the device can be estimated by the following simple argument. In order to change the gate voltage, the capacity of its p-n junctions must be charged through the resistance of the channel. This process has an associated time constant which limits the frequency response. Let us assume a wedge-shaped channel, completely pinched off at the drain end and completely open at the source end (that is $V_G = V_s = 0$). The capacity for unit length in the Z direction is approximately

$$C \simeq 4KL/a \tag{3.8}$$

The factor 4 arises because the average width of the space-charge region is approximately $a/2$ and because there are two such regions, one on either side. This capacity on the average charges through half the resistance of the channel, i.e.,

$$R = L/2a\sigma_0 \tag{3.9}$$

We would accordingly expect a limiting frequency f given by

$$f = \frac{1}{2\pi RC} = \frac{1}{2\pi}\left(\frac{a^2\sigma_0}{2L^2K}\right) \tag{3.10}$$

Another way of looking at the frequency response is to consider the transit time of a carrier along the channel. In Appendix 1 it is shown that the transit time τ is given by

$$\tau = (3/2)\,(L^2/\mu_0 W_0) \tag{3.11}$$

Substituting for W_0, we obtain

$$\tau = \frac{3KL^2}{\sigma_0^2} \tag{3.12}$$

This transit time differs from RC in equation (3.10) by about a factor of $3/2$. Thus there is essential agreement between the frequency responses as estimated from RC and from transit time.

The above theory has been summarized in the form of nomographs

which are given in Figs. 3 and 4. In Fig. 3 we have given a nomograph for the calculation of the pinch-off voltage W_0. If a straight line is drawn between the value of N on the left scale and the value of a on the right scale, it intersects the center scales at W_0 and at E, the maximum value of the electric field in the space charge region. Fig. 4 shows simultaneously the field-effect parameters, a/L, σ_0, g_{m0}, R_0, f and C. A straight line drawn across the chart intersects the various scales in a set of values which are consistent with a given design. The one remaining parameter of interest, the current, is easily obtained from (3.7).

4. MODIFICATIONS OF THE IDEAL THEORY

The theory presented in Section 3 deals with an ideal structure. In practice we shall find it necessary to modify the theory somewhat to

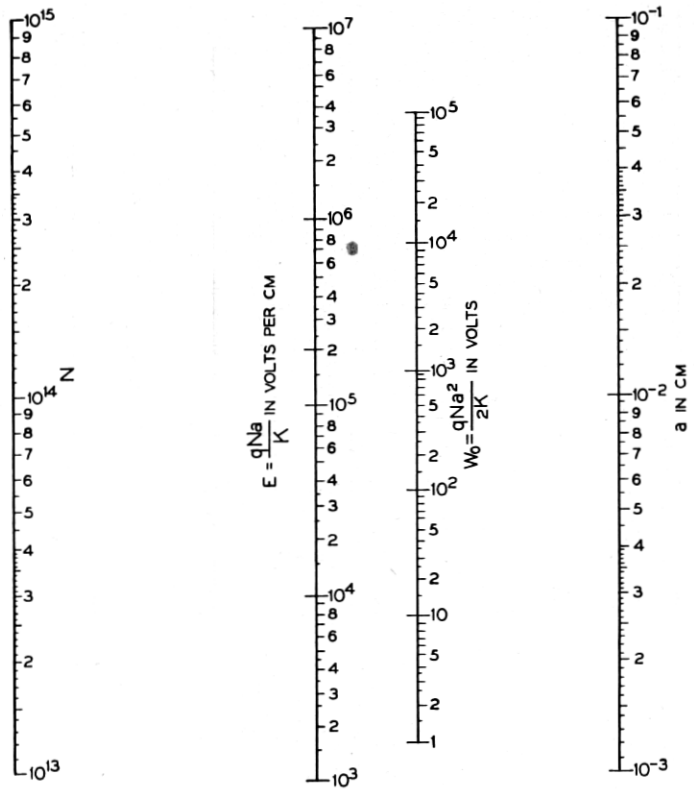

Fig. 3 — Nomograph giving the value of pinch-off voltage as a function of channel resistivity and thickness. (Constant-mobility case).

take account of special experimental conditions. In particular we shall discuss the following effects:

A. Series resistance of semiconducting paths at the source and/or drain contacts.
B. Negative resistance effects due to hole current flow into the gate.
C. Temperature effects.
D. Effects of high electric field on mobility.

A. Series Resistance

In Section 3 we have considered that the source and drain connect directly onto the channel between the gates. It is necessary in the fabrication of these units, however, to allow a small bridge of semiconductor between the actual contact and the gate. This means that a series re-

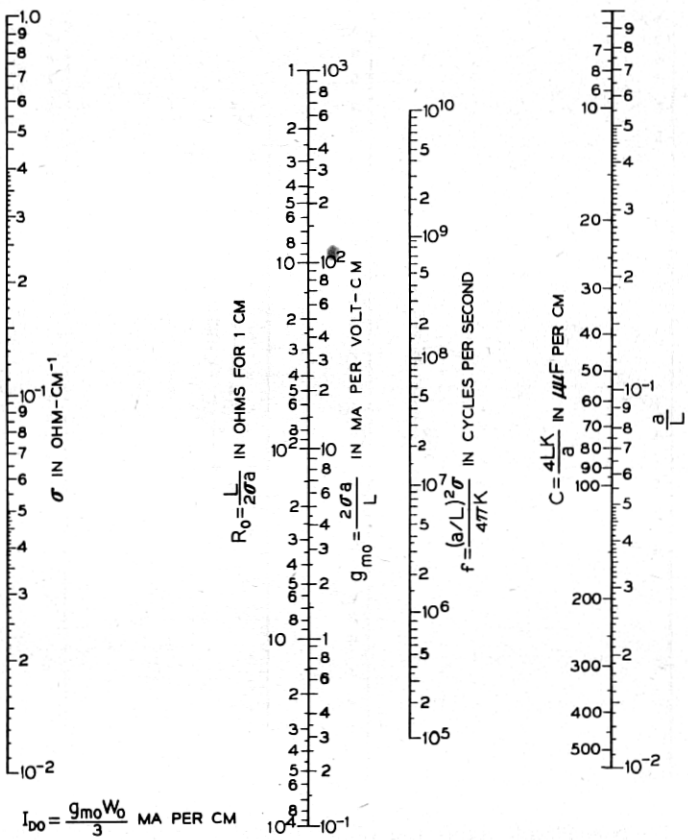

Fig. 4 — Design nomograph for constant mobility case.

sistance appears between the electrodes to which voltages are applied and the working part of the structure. It is possible to take account of these resistances by simple circuit theory.

For the calculation of g_m one must recognize two differences. Firstly, the voltages which must be inserted into (3.3) to obtain the transconductance of the working part of the structure must be changed as follows:

$$V_G \rightarrow V_G\,, \qquad V_s \rightarrow -IR_s\,, \qquad V_D \rightarrow (V_D - IR_D) \tag{4.1}$$

where R_s and R_D are the values of the source and drain resistances. Secondly, the degeneration of the source resistance must be taken into account. If the apparent transconductance is denoted by g_m' then

$$g_m' = g_m/(1 + R_s g_m) \tag{4.2}$$

which takes account of the fact that a fraction of any increase in gate voltage appears as an increased drop across the source resistance. In (4.2), of course, g_m must be calculated considering the modifications (4.1).

It is clear from (4.2) that if $g_m R_s \gg 1$, then the transconductance g_m' is simply given by $1/R_s$. If then we are to obtain the potentially high transconductance of the device, we must keep the source resistance small. The drain resistance does not have as serious consequences, the chief disadvantages being: (a) the necessity of having a higher supply voltage, and (b) $I_D^2 R_D$ heat which must be dissipated.

The source resistance also alters the frequency response. One must add R_s to the R of (3.10) in obtaining the limiting frequency.

B. Negative Gate Resistance

It is possible, in units where an appreciable fraction of the current is carried by holes, as would be the case for a channel of resistivity approaching the intrinsic value, to obtain a negative gate-resistance. This effect arises as follows: Suppose that the gate is made more positive so that it opens somewhat and allows additional electron current to flow in the channel. This additional electron current is accompanied by an increased hole current on the drain side of the gate, (within the electrically neutral region). If the unit is operating in the pinched-off region, however, these holes will not flow through the channel but will instead flow to the gate. This is the case because the electric field in the pinched-off space-charge region near the drain is directed away from the axis of the channel and is thus focussing for electrons, while tending to pull holes into the gate. From this argument we see that a positive change of gate voltage results in an increased flow of holes to the gate and thus

in a negative change in the current flowing into the gate. Hence the gate terminal presents a negative resistance.

A quantitative estimate of the negative resistance can be obtained as follows: let the fraction of the drain current which is carried by holes be f. Let the gate potential increase by V volts. If the transconductance is g_m, the drain current, I_D, will increase by $g_m V$. Then the hole current will increase by $fg_m V$. All this hole current appears as a current flowing out of the gate. Thus the change in gate current is $-fg_m V$ and the gate conductance $G_G = -fg_m$.

The holes referred to in the argument above may arise from three sources: (a) they may be thermally generated in the body of the semiconductor; (b) they may be generated at low lifetime areas on the surface; or (c) they may be injected at the contacts. It might be desirable to make a device in which the drain contact was intentionally made hole-injecting (say by the use of a p-n junction) and in which the negative resistance effects would thus be greater. However, for most applications the negative resistance is not desirable, and it is necessary to minimize the density of these holes. Source (a) can be decreased by using lower resistivity material, and source (b) by high lifetime treatment of the surface. The injection (c) of holes at the drain contact can be prevented by making the drain contact from strongly n-type material (designated n^+). Such an n-n^+ junction will not act as a source of holes.

C. Temperature Effects

As can be seen from the nomographs, the unipolar field effect transistor is a relatively high power device. Furthermore, most of the power dissipation takes place in the space-charge region near the drain. Therefore the removal of heat can be a major problem. The chief effect of heating is to increase the gate saturation current by increasing the number of thermally generated hole-electron pairs. It is true that the mobility also varies as $T^{-3/2}$ but this variation is fairly slow and serves only to vary σ_0 which enters the theory linearly.

D. Effects of Non-linear Mobility at High Fields

In germanium there is a maximum field E_c beyond which Ohm's law fails.[3] Beyond E_c the mobility decreases as the half power of the electric field, and hence the effective conductivity of the material also changes. As will be shown in Part 2, electric fields greater than E_c may well be obtained in practical designs of field-effect transistors. It is therefore necessary to consider the effects of square-root mobility on the tran-

sistor characteristics. Since the required analysis is somewhat lengthy, and the conclusions that can be drawn from it are of great interest, this analysis will be given in a separate section.

5. SQUARE-ROOT MOBILITY THEORY

The treatment which follows is similar in outline to that of W. Shockley for the constant-mobility case[1] and familiarity with the general features of that theory will be assumed. The fundamental assumption made is that the channel narrows sufficiently slowly so that locally the voltage W across the space-charge region between the channel and the gate is given by the following solution to the one-dimensional Poisson equation

$$W = qN[a - b(x)]^2/2K \tag{5.1}$$

where q is the electronic charge, N is the excess donor densitiy* in the channel material, 2a is the thickness of the material between the gates and $2b(x)$ is the thickness of the narrowed channel at a distance x from the source towards the drain. This situation is shown in Fig. 1.

We shall consider the case where the pinch-off voltage is so high that the field in the channel is greater than the critical field over the major fraction of its length, but not so high that the carriers reach limiting velocity.[3] A more accurate calculation would take into account the fact that near the source the field is small and would join two solutions for the ohmic and non-ohmic parts. We shall assume that in the range for which $E > E_c$, the mobility is proportional to $(E)^{-1/2}$. This law is in agreement with the data of E. J. Ryder.[3] In particular we write for the conductivity in the channel

$$\sigma_c = \mu_0 nq(E_c/E)^{1/2} = \sigma_0(E_c/E)^{1/2} \tag{5.2}$$

where μ_0 is the low field mobility, n the electron density in the channel, E_c the critical field, and σ_0 the low field conductivity.

We shall consider a unit 1 cm wide, $2a$ cm thick between gates, and with a channel length of L cm. The conductance g of the channel at any point x is proportional to the thickness of the channel there and is given by

$$g = 2b(x)\sigma_c = 2b(x)\sigma_0(E_c/E)^{1/2} \tag{5.3}$$

Making use of (2.1) we may write this in terms of W

$$g = 2\sigma_0 a[1 - (W/W_0)^{1/2}]\,(E_c/E)^{1/2} \tag{5.4}$$

* The theory is worked out here for a transistor with n-type channel and p-type gate. This is the opposite polarity to that treated by Shockley in Reference 1, but is appropriate to the specimens measured.

where $W_0 = qNa^2/2K$ has been introduced. The gradual approximation implies that $E = dW/dx$, and we may therefore write for the square of the current

$$I^2 = g^2(dW/dx)^2 = 4\sigma_0^2 a^2 E_c[1 - (W/W_0)^{1/2}]^2(dW/dx) \tag{5.5}$$

Upon integration from source to drain we obtain

$$I^2 L = g_0^2 E_c \int_{W_S}^{W_D} [1 - (W/W_0)^{1/2}]^2 \, dW \tag{5.6}$$

where we have introduced the symbol $g_0 = 2\sigma_0 a$. We shall find it convenient to introduce the integral

$$J(W) = \int_0^W [1 - (y/W_0)^{1/2}]^2 \, dy \tag{5.7}$$

in terms of which the current may be written

$$I = g_0(E_c/L)^{1/2}[J(W_D) - J(W_S)]^{1/2} \tag{5.8}$$

It is easy to evaluate $J(W)$ by the change of variable

$$u^2 = [1 - (y/W_0)^{1/2}]$$

and when this is done we find

$$J(W) = \frac{W_0}{6}(3[1 - (W/W_0)^{1/2}]^4 - 4[1 - (W/W_0)^{1/2}]^3 + 1) \tag{5.9}$$

We are now in a position to write the current from (5.8) and (5.9). We shall take

$$W_S = V_S - V_G\,; \qquad W_D = V_D - V_G \tag{5.10}$$

where V_S, V_D, and V_G are the potentials applied to the source, drain, and gate respectively. We accordingly obtain

$$\begin{aligned} I &= I_c\{3[1 - (V_D - V_G)^{1/2}/W_0^{1/2}]^4 - 4[1 - (V_D - V_G)^{1/2}/W_0^{1/2}]^3 \\ &= 3[1 - (V_S - V_G)^{1/2}/W_0^{1/2}]^4 + 4[1 - (V_S - V_G)^{1/2}/W_0^{1/2}]^3\}^{1/2} \end{aligned} \tag{5.11}$$

where we have introduced the symbol

$$I_c = g_0(W_0 E_c/6L)^{1/2} \tag{5.12}$$

This is the value of the current at pinch-off for zero gate bias. The analogous expression for the constant mobility case was[2]

$$I_{D0} = g_{m0}W_0/3 = 2W_0\sigma_0 a/3L = W_0 g_0 3L \tag{5.13}$$

Therefore, all other things being equal, the effect of non-constant mo-

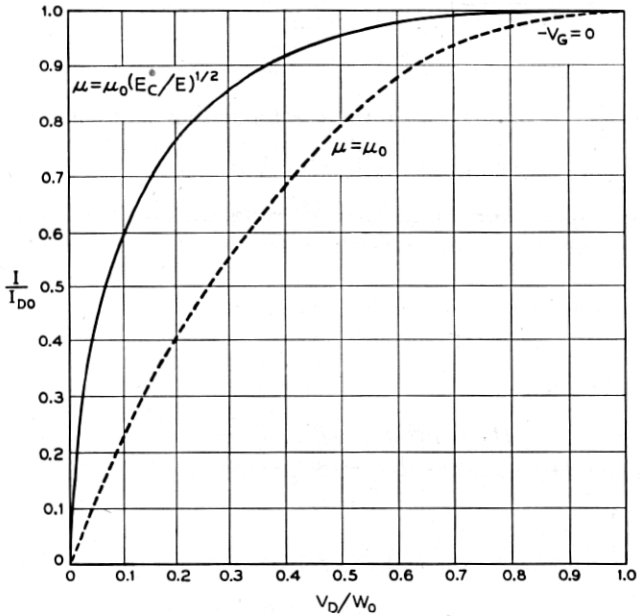

Fig. 5 — Comparison of the drain characteristics for the constant and non-constant mobility cases.

bility is to reduce the maximum current in the ratio

$$I_c/I_{D0} = (9E_cL/6W_0)^{1/2} = 1.21\ (E_cL/W_0)^{1/2} \tag{5.14}$$

It should be pointed out that this analysis is valid only for $W_0/L \gg E_c$, but within this range of validity, equation (5.14) shows that considerable reductions of current may take place.

In addition to changes in magnitude of the pinch-off current, there will be significant alterations in the shape of the I_D versus V_D characteristic. This characteristic has been calculated from equation (5.11) for the case $V_S = V_G = (V_S - V_G) = 0$ and is plotted in Fig. 5. Also shown on the same figure, for comparison purposes, is the I_D versus V_D curve for the constant mobility case. It can be seen that the non-constant mobility curve has a steeper initial slope and pinches off more gradually. The effect of this is to make the unit *appear* to have a lower pinch-off voltage. At low drain voltages, on the other hand, the unit is operating in the Ohm's law range and the observed initial resistance will agree with the slope of the curve for the constant mobility case.

It is also of interest to know the variation of the current beyond pinch-off with gate bias. This function can be obtained from (5.11) by putting

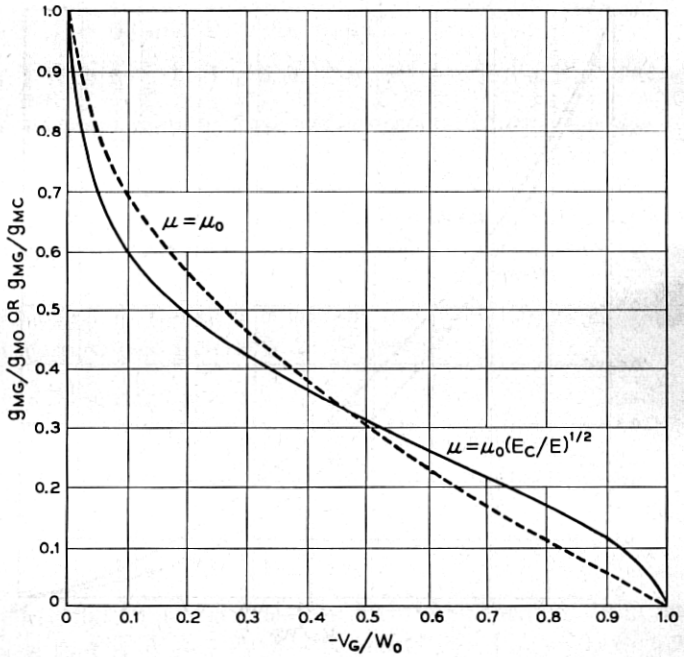

Fig. 7 — Dependence of g_m on V_g. The dotted line refers to the constant-mobility case and the solid line to the non-constant case.

Appendix 1). The conductivity however will vary along the channel as the electric field varies. We may within the accuracy of (5.23) use an average conductivity corresponding to the average field, $E_a = W_0/L$, in the channel. Making use of (5.2) we find that the decrease in frequency response caused by non-constant mobility can be estimated by

$$f_c/f = \sigma_c/\sigma_0 \approx (E_c/E_a)^{1/2} = (E_cL/W_0)^{1/2} \tag{5.24}$$

The quantity $(E_cL/W_0)^{1/2}$ is seen to be a sort of figure of comparison for the non-constant mobility case. Except for numerical factors near unity, the current, the transconductance, and the frequency response are all reduced in about this same ratio

$$I_c/I_{D0} \approx g_{mc}/g_{m0} \approx f_c/f \approx (E_cL/W_0)^{1/2} \tag{5.25}$$

Another way of looking at the frequency response is to consider the transit time across the channel. In Appendix 1 it is shown that for the constant mobility case

$$\tau = (3/2)\ (L^2/\mu_0 W_0) \tag{5.26}$$

Note that $L^2/\mu_0 W_0$ is the transit time across a distance L that we would expect from the average velocity $\mu_0 W_0/L$. On the other hand, as shown in the Appendix, for the non-constant case

$$\tau_c = 1.47\ L/\mu_0(E_c W_0/L)^{1/2} \tag{5.27}$$

Here again $L/\mu_0(E_c W_0/L)^{1/2}$ is the transit time to be expected across a distance L with the velocity $\mu_0(E_c W_0/L)^{1/2}$. Note that the ratio of (5.26) to (5.27) gives

$$\frac{f_c}{f} = \frac{\tau}{\tau_c} = \frac{1.5}{1.47}\frac{L^2}{\mu_0 W_0}\cdot\frac{\mu_0(E_c W_0/L)^{1/2}}{L} = 1.02(E_c L/W_0)^{1/2} \tag{5.28}$$

This expression is in essential agreement with (5.25) which was obtained from the RC argument.

6. DESIGN THEORY

In the previous sections we have presented the theory of field-effect transistors operating both below and above the critical field E_c. The question now arises as to what is the optimum field at which the transistor should operate. We will first consider this problem for the constant mobility range and show that both frequency response and transconductance increase with electric field. We shall then consider the square-root theory and show that, although increasing field above E_c does result in an increase in frequency response and transconductance, the corresponding increases in power dissipated are so large as to make such a design unattractive. These considerations lead to the conclusion that the optimum field for transistor operation is just the critical field, and we will present design nomographs for this condition.

A. Constant Mobility

For the gradual approximation, that is $(a/L) \ll 1$, the following equations apply

$$f = (a/L)^2\sigma_0/4\pi K \tag{6.1}$$

$$W_0 = qNa^2/2K \tag{6.2}$$

$$g_{m0} = 2\sigma_0(a/L) \tag{6.3}$$

where the symbols are as previously defined. It is seen from (6.1) and (6.3) that both the frequency response and transconductance are proportional to conductivity. Thus the designs of most interest are those with high conductivity material. The solution of the three equations

would be much simplified if it could be assumed that σ_0 is proportional to N. This is true if the density of the minority carrier is small compared to that of the majority carrier and is therefore true for material of high conductivity, corresponding to the cases of greatest interest. If we assume that

$$\sigma_0 = q\mu_0 N \tag{6.4}$$

the error introduced will be about 10 per cent for $\sigma_0 = 0.05$ ohm^{-1}-cm^{-1} and less than 1 per cent for $\sigma_0 > 0.7$ ohm^{-1}-cm^{-1}. It is therefore reasonable to apply (6.4) with the restriction that

$$\sigma_0 > 0.05 \text{ ohm}^{-1}\text{-cm}^{-1}$$

The additional restriction that the field in the channel must be less than the critical field is difficult to apply exactly. The field in the channel increases from the source to a maximum at the drain. The dependence of field upon distance down the channel has been derived by Shockley, but it is a complicated expression. A simple approximation to the limiting condition would be to stipulate that the average field in the channel, defined as $E_a = W_0/L$, be less than E_c. In the limiting case, $W_0/L = E_c$, part of the channel near the drain would be above the critical field while the source end would not. This represents approximately the desired limiting condition.

We shall now impose onto (6.1) to (6.4) the following three restrictions:

(a) Gradual approximation holds,

$$a/L \ll 1$$

(b) Conductivity is proportional to majority carrier density,

$$\sigma_0 > .05 \text{ (i.e., } \rho < 20 \text{ ohm-cm)}$$

(c) Field is less than critical,

$$W_0/L = E_a \leqq E_c$$

Firstly, consider the frequency response. Eliminating σ_0 and N from (6.1), (6.2) and (6.3) gives

$$f = (a/L)^2 \mu_0 W_0/2\pi a^2 = \mu_0 W_0/2\pi L^2$$

and substituting for W_0 from restriction iii yields

$$f = \mu_0 E_a/2\pi L \tag{6.5}*$$

* Note that this equation shows that f is proportional to $\mu_0 E_a$, the mean velocity of the electrons, divided by L, the length of the channel. Since E is always positive and increases smoothly from source to drain, the ratio $\mu E_0/L$, and hence f, will be approximately proportional to the inverse of the transit time. This result is rigorously shown in Appendix 1.

This equation shows that the frequency response is proportional to the average field and that therefore for a given length of channel the maximum frequency response will be obtained with the maximum possible value of E_a, that is, $E_a = E_c$. Another interesting relationship can be deduced from (6.5) by substituting for L from restriction (c) giving

$$f = \mu_0 E_a^2 / 2\pi W_0 \tag{6.6}$$

Thus for fixed value of E_a high values of frequency response will correspond to low values of pinch-off voltage.

Now consider the transconductance. Eliminating σ_0 and N from (6.2), (6.3) and (6.6) gives

$$g_{m0} = 4\mu_0 K W_0 / aL$$

and substituting from restriction iii for W_0/L, we obtain

$$g_{m0} = 4\mu_0 K E_a / a \tag{6.7}$$

Thus for a given thickness of channel the maximum transconductance will be obtained with the maximum possible value of E_a, that is $E_a = E_c$.

We therefore conclude that on the constant-mobility theory the best choice of average field for optimum frequency response and transconductance is the critical value, E_c.

B. Square-Root Range

Having shown that on the constant mobility theory the optimum electric field is E_c, we must next determine what, if anything, is to be gained by operating above this field; that is, in the square root mobility range. It is clear from (5.24) that, if units are operated at higher and higher W_0, the gain in frequency response is not as great as would be predicted by the constant mobility theory. In fact for a given length L, the frequency response can never exceed

$$f_t = 1/\tau_t = v_t/L \tag{6.8}$$

where v_t is the limiting drift velocity of "hot" electrons.[3] It should be pointed out however, that constant electron drift velocity is incommensurate with the usual gradual-approximation field-effect equations. This condition can obtain only at the pinched-off end of the channel, i.e., the "expop" region, for within the electrically neutral channel continuity of current and constant electron drift-velocity combine to require constant b. Fulfillment of this condition is impossible between equipotential gates since the change in W with x required to give the electric field would also require a change in b by (5.1).

As we shall subsequently show, it is not expedient to attempt to obtain high frequencies by the brute-force method of increasing W_0/L because the power goes up as W_0^2, while square-root mobility effects cause diminishing returns to set in as far as increases in frequency response are concerned. A simple way to see how much improvement in frequency response can be obtained by pushing into the nonlinear mobility range is to impose a maximum power dissipation condition. It has been found possible experimentally to dissipate by the use of cooling fins some 400 watts per cm length of gate. We will accordingly take $P_c = I_c W_0/L = 400$ as a limiting power-handling capacity. We write the square-root case equations in the following form:

$$g_{mc} = 6^{1/2}\sigma_0(L/a)^{-1}L^{1/2}W_0^{-1/2}E_c^{1/2} \tag{6.9}$$

$$f_c = \sigma_0(L/a)^{-2}L^{1/2}W_0^{-1/2}E_c^{1/2}(1/4\pi K) \tag{6.10}$$

$$P_c = I_c W_0/L = g_{mc}W_0^2/3L \tag{6.11}$$

$$W_0 = \sigma_0(L/a)^{-2}L^2\left(\frac{1}{2\mu_0 K}\right) \tag{6.12}$$

Using (6.12) and (6.9) to write P_c in terms of ρ, (L/a), and L we obtain

$$P_c = \sigma_0^{5/2}(L/a)^{-4}L^{5/2}E_c^{1/2}\left(\frac{6^{1/2}}{3}\right)\left(\frac{1}{2\mu_0 K}\right)^{3/2} \tag{6.13}$$

We now use (6.12) and (6.13) to eliminate W_0 and σ_0 from (6.10) and finally obtain

$$f_c = (1.25P_cE_c^2\mu_0^4/128\pi^5 K)^{1/5}(L/a)^{-1/5}L^{-1} \tag{6.14}$$

If, in agreement with experimental results, we take

$$P_c = 400 \text{ watts/cm}$$

$$E_c = 1000 \text{ volt/cm}$$

$$\mu_0 = 3600 \text{ cm}^2/\text{volt-sec.}$$

and

$$K = 1.4 \times 10^{-12} \text{ farad/cm}$$

(6.14) becomes

$$f_c = 1.05 \times 10^6(L/a)^{-1/5}L^{-1} \tag{6.15}$$

It should be noted that (6.15) is relatively insensitive to choices of (L/a).

For any (L/a) from 1 to say 5, (6.15) can be approximated by

$$f_c \approx 1.05 \times 10^6 L^{-1} \tag{6.16}$$

On the other hand (6.8) gives

$$f_t = 6 \times 10^6 L^{-1} \tag{6.17}$$

when Ryder's experimental value for v_t is inserted. It is clear, therefore, that the power handling capacity of a unit of given size prevents the realization of the ultimate frequency response and causes a reduction of upper frequency limit by a factor of about 6. From (6.14) we see that f_c depends on $P_c^{1/5}$ and that it is necessary to increase the power handling capacity by a factor of 32 in order to double the frequency response.

We have explored the design possibilities of units in which the critical field is never exceeded. It is shown that the limiting frequency under these restrictions is given by

$$f = E_c\mu/2\pi L = 5.7 \times 10^5 L^{-1} \tag{6.18}$$

A comparison of (6.16) and (6.18) shows that a gain of only a factor of 2 in frequency response can be obtained by operating in the non-constant mobility range. This factor is gained at a considerable cost in power dissipation, and it is therefore unprofitable to extend the design into this operating range.

Design Nomographs for $W_0/L = E_c$

Nomographs have been given, Fig. 3 and 4, representing the "field-effect" equations for the gradual approximation without any restrictions on field. It will be noted that, in order to obtain a unique set of solutions with these nomographs, three choices have to be made. For example, if σ_0 (and therefore N) and a are chosen, W_0 is determined in Fig. 3, but another choice, say (a/L), must be made to determine the values of f and g_{m0} in Fig. 4. If now we stipulate that $W_0/L = E_c$, we are left with only two choices and therefore the two nomographs can be combined in one. This new nomograph is shown in Fig. 8. In determining the scale factors $E_a = E_c$ was taken as 1,000 volts/cm. For n-type germanium at room temperature the critical field is actually 900 volts/cm but the figure of 1,000 volts/cm was used to give some simplification of the scale factors.

Of the three restrictions imposed on the analysis, restriction iii has been incorporated in the nomograph of Fig. 8. Restriction (a) states that (a/L) shall be small compared to unity. It is difficult to assess exactly

the error that will be introduced as (a/L) approaches unity but certainly the nomograph should not be used for values of (a/L) greater than ½. This region has been marked on Fig. 8. Restriction (b) states that the analysis will not be accurate for materials with $\sigma_0 < 0.05$ ohms^{-1}-cm^{-1} and this region has also been marked on Fig. 8.

A theoretical limitation which has not so far been considered is concerned with the minimum allowable pinch-off voltage. If the reverse gate bias is insufficient to saturate the junction, the gate impedance will be much lower than the value predicted from theory and will vary with gate bias. It is therefore necessary that the gate be saturated, i.e., that the reverse bias be large compared to kT/q, which, at room temperature, is approximately 0.025 volts. In practice it is found that a junction is sensibly saturated at 0.1 volts bias. However, if the gate is to be saturated

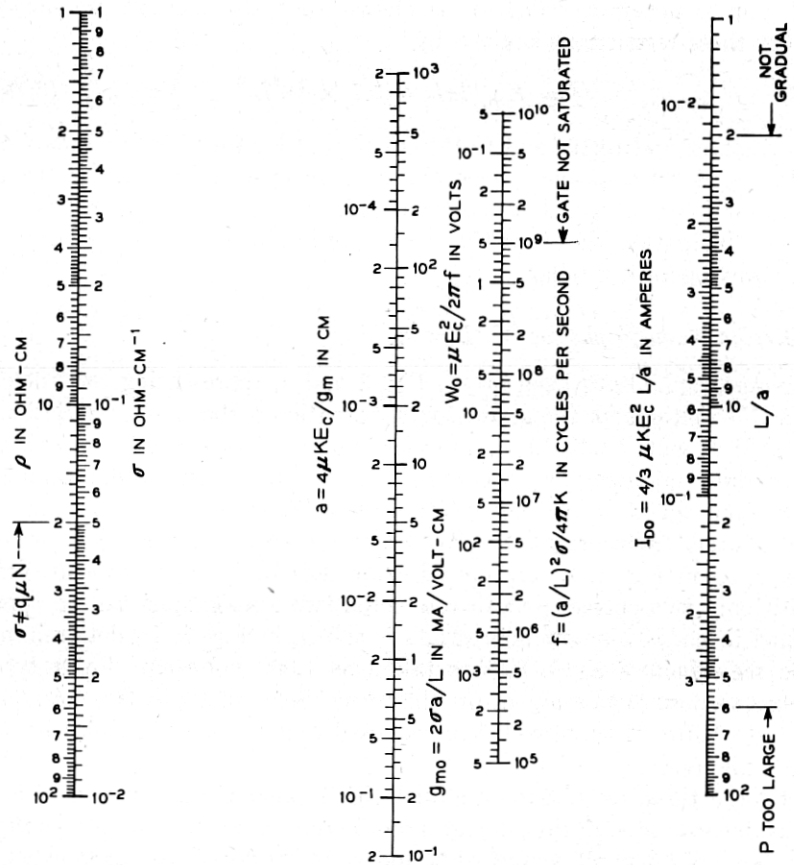

Fig. 8 — Design nomograph for average channel field equal to E_c.

over most of its length the pinch-off voltage, W_0, must be much greater than 0.1 volts. A reasonable lower limit to the value of W_0 would therefore be about 0.5 volt, and this limitation is marked on the nomograph. As a consequence of this choice, the upper limit of frequency response, as given by the nomograph, is 10^9 *c/s*.

A practical limitation in the design of field-effect transistors is the allowable power dissipation in the unit. This is determined by the power that can be removed without undue temperature rise. With use of a cooling fin on the gate, the power that can be removed from a unit of 1 cm width will be approximately proportional to the length, L, of the channel. Therefore, a reasonable design criterion would be the power dissipated per unit length of channel, that is, $P = W_0 I_{D0}/L$. This function reduces to

$$P = W_0 I_{D0}/L = (4/3)\mu_0 K E_c^3 (L/a)$$

Putting in numbers for μ_0, K and E_c gives

$$P = 6.7\ (L/a) \tag{6.19}$$

In the units that have been made and cooled by this method, the channel length was 0.005 inches. With the fin water-cooled a maximum power of 5 watts could be dissipated with only a small rise in temperature. If this case is taken to represent the best that can be done, then $P \doteq 400$ watts/cm and substituting in equation (6.19) gives a maximum value of (L/a) of 60. Therefore the limitation imposed by the allowable power dissipation restricts the designs to those having values of (L/a) less than 60. This region is marked on the nomograph.

Some Possible Designs of Field-Effect Transistors

In Table I are shown the properties of some transistors designed by means of the nomograph. Unit No. 1 is chosen to give the maximum theoretical frequency response, 1,000 mc/s, with the widest possible channel. It is therefore the most feasible design for the highest frequency response. However, the dimensions of this unit — channel length and distance between gates both being about ¼ mil — make fabrication difficult. If such a unit could be made, it would have some very desirable properties: frequency response of 1,000 mc/s, transconductance of 70 *ma/v* and ability to operate with a 12 ma current from a 0.5-volt supply.

Unit No. 2 is designed for maximum frequency response together with maximum transconductance. The dimensions of this unit are too small to be considered feasible with existing techniques. Unit No. 3, designed for maximum power dissipation could be made provided that

TABLE I — DESIGN PARAMETERS AND CHARACTERISTICS OF SOME FIELD EFFECT STRUCTURES

No.	Remarks	ρ	a	L	f	g_{m0}	W_0	I_{D0}	Power
		ohm-cm	*cm*	*cm*	*C/S*	*ma/v*	*Volts*	*ma*	*Watts*
1	Max. f, max. a	15	3×10^{-4}	6×10^{-4}	10^9	70	0.5	12	6×10^{-3}
2	Max. f, max. g_{m0}	0.07	2×10^{-5}	6×10^{-4}	10^9	10^3	0.5	180	9×10^{-2}
3	Max. power	20	1.1×10^{-2}	6.6×10^{-1}	8×10^5	1.7	600	400	240
4	$\rho = 20$, $W_0 = 30$	20	2.6×10^{-3}	2×10^{-2}	1.6×10^7	8	30	80	2.5

junctions to 20 ohm-cm germanium could be made with breakdown voltage greater than 600 volts. If it were possible to obtain body breakdown,[4] such junctions would stand approximately 800 volts reverse bias. In practice however, perhaps due to surface breakdown, it is difficult to exceed 100 volts.

Unit No. 4 represents about the best unit that has been made with the molten-metal process.

PART II. EXPERIMENTAL RESULTS

1.0 INTRODUCTION

In Part I we have presented the design theory of field-effect transistors; in this part we shall present experimental data which verify the design theory. All specimens to be described were made using n-type *Ge* for the channel material. The p-type gates were formed by the indium alloy-process.

In order to obtain higher frequency response the dimensions of the channel were made small. The exact choice of resistivity was determined by two considerations. It must be sufficiently low to avoid negative-resistance effects but not so low that the pinch-off voltage becomes excessive for reasonable channel thickness. It was found that 20 ohm-cm germanium was a good compromise. By the use of a confining jig during alloying, it was possible to make alloy gates as small as 5 mils long. With such dimensions and material, the power dissipation in the channel was of the order of a few watts and means had to be provided for removing heat from the specimen. This was done by attaching a cooling fin to the indium during the alloying process. As previously,[2] a grown n-n^+ junction was incorporated to provide a drain contact that would not inject

holes and a tin alloy contact was used for the source. Close source to gate spacing was achieved by alloying the source and gate contacts simultaneously in the jig. A schematic diagram of the unit is shown in Fig. 9 and a photograph of the completed structure in Fig. 10. An exploded view of the alloying jig used in making the unit is shown in Fig. 11.

In all, 9 successful units were made in this way. The salient properties of these are shown in Table II. Before discussing these properties we will describe the testing procedures used.

2. MEASUREMENT TECHNIQUES

2.1 *Measurement of Static Characteristics*

A preliminary examination of the characteristics was made using a pulse-operated *E-I* presentation unit. In this way it was possible to

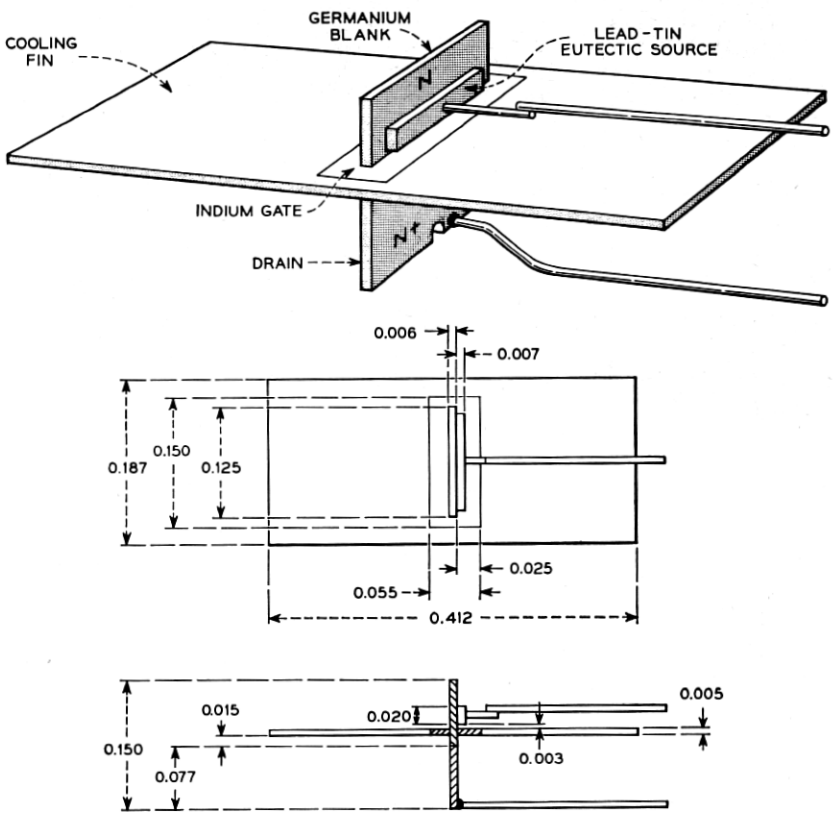

Fig. 9 — Schematic diagram of experimental transistor.

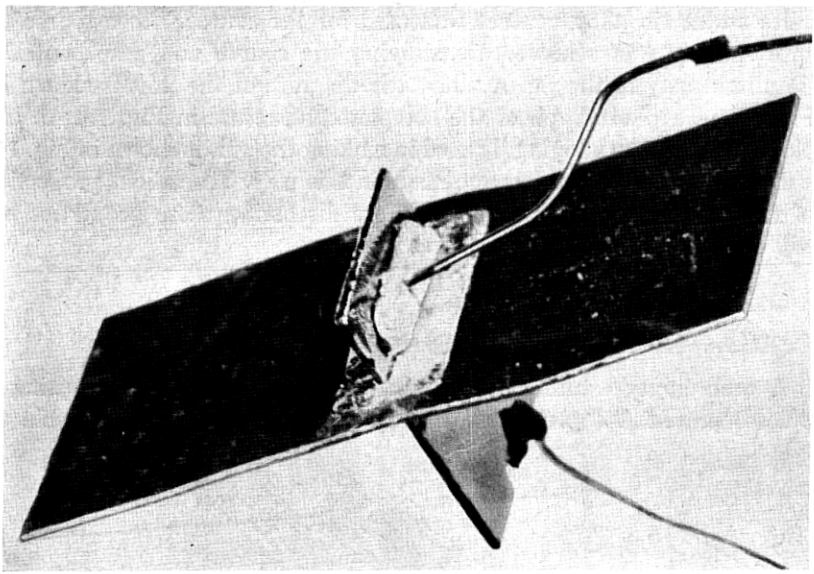

Fig. 10 — Photograph of experimental transistor. The largest component is the cooling fin which also serves as the gate contact. The upper lead goes to the source, the lower to the drain.

determine the general behavior of a specimen without risk of overheating. If the specimen appeared to be satisfactory, that is, showed pinch-off and transconductance, the dc characteristics were then measured.

2.2 *Determination of Source, Channel and Drain Resistances* R_s, R_0, *and* R_D

If a positive voltage is applied to the drain while the gate is open-circuited, then the gate will float at a potential approximately equal to that at the source end of the channel. The reason for this can be seen in the following way. The net gate current must be zero. If the gate were to become much more positive than the source end of the channel, a considerable fraction of the length of the gate would be biased in the forward direction and a large current would flow into the gate. If, on the other hand, the gate were to become much more negative than the source end of the channel, all of the gate junction would be in the reverse direction and the saturation current would flow out of the gate. The gate must therefore take up a potential very close to that of the source end of the channel. In fact the potential will be slightly more positive so that a forward current from a small portion of the gate at the source

end will neutralize the reverse saturation current of the remainder of the gate. Thus the open-circuit gate potential is closely equal to the potential drop in the source resistance R_S and, if the current I_S is measured, the value of R_S can be determined. In the experiment, the gate potential was measured with an electronic voltmeter having an input impedance of 15 $M\Omega$, so that the gate was essentially open-circuited.

To determine the value of R_D the source and drain connections were interchanged and the experiment repeated. R_0 was determined from the measured value of the sum of R_D, R_0 and R_D. The sum was determined from the ratio of V_D to I_D for values of V_D sufficiently small compared to W_0 so that there was negligible "pinching-off" in the channel.

2.3 *Frequency Response*

The cut-off frequency of many of these units was about 20 mc/s and in one case, as high as 50 mc/s. A thorough investigation of the frequency

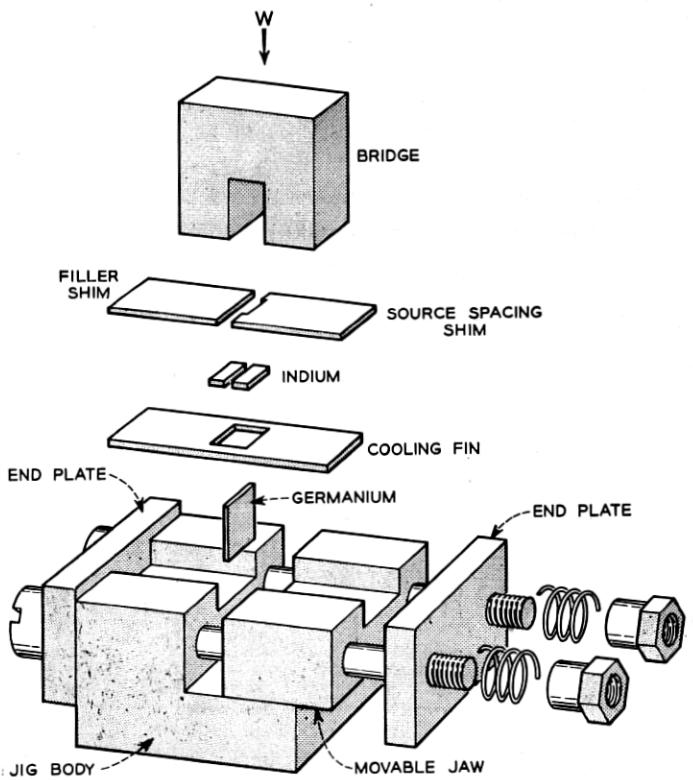

Fig. 11 — Exploded view of the jig used in fabrication.

TABLE II

Unit No.	W_0	I_{D0}	g_{m0}	R_0	R_s	f (300°K)
	Volts	*ma*	*ma/v*	Ω	Ω	*mc/s*
20	35	6.5	0.7	340	300	5
23	55	40	1.0	310	200	—
24	70	35	1.0	360	140	20
27	50	120	5.0	—	—	—
29	40	20	1.0	300	500	17
30	37	20	1.3	500	160	18
32	40	40	1.6	200	152	50
35	28	15	1.0	110	36	31
36	20	12	1.0	—	—	—

response up to these frequencies would be a long and difficult experiment and was not attempted. However, an estimate of the response can be obtained from the performance of a unit when operated as an oscillator. It is shown in Appendix 2 that the maximum frequency at which a closely-coupled feedback oscillator can be made to oscillate is approximately equal to the cut-off frequency f_1 determined by the theory. The circuit used in the experiment is shown in Fig. 12. Close coupling for the feedback was obtained by winding one coil inside the other. The presence of oscillation was detected by a loosely-coupled radio receiver. Coarse changes in frequency were made by changing coils, while fine control was obtained with the variable capacitance K. The experiment was started by making the circuit oscillate at a comparatively low frequency, say 100 kc/s. The frequency was then increased continuously until no oscillation could be excited. It was also found possible to modulate the oscillation in the circuit of Fig. 13.

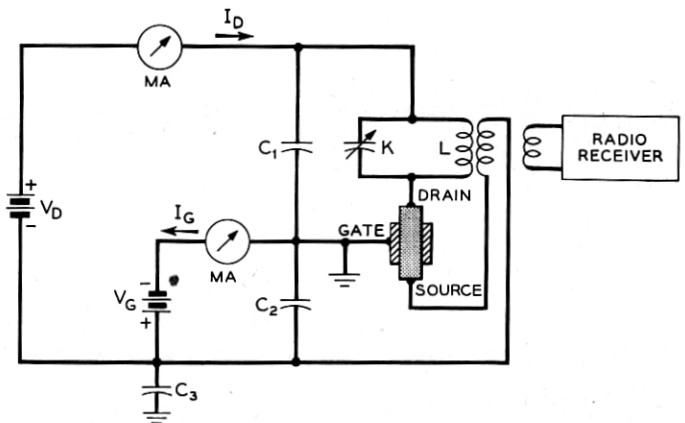

Fig. 12 — Test circuit used in estimating the frequency response.

3.0 *Experimental Results*

In all, test were made on nine specimens and the properties of these are shown in Table II. In these specimens the gate length, L, was about 7 mils, so that the pinch-off voltage at which the average field became equal to the critical field, E_c, was 20 volts. It is seen that most of the specimens had pinch-off voltages much in excess of 20 volts and would therefore not obey the constant-mobility theory. We shall describe one such unit, No. 32, in detail in Sections 3.3 and 3.4 below. However, two specimens, No. 35 and No. 36 had lower pinch-off voltages and they should approximately agree with the constant-mobility theory. Of these two, No. 35 was investigated more thoroughly and we will begin by discussing its behavior.

3.1 *Static Characteristics of No. 35*

The drain characteristics of unit No. 35 are shown in Fig. 14, and the gate characteristics in Fig. 15. It is seen that the unit had a pinch-off voltage of about 27.5 volts. With a gate length of 7 mils this gives an average field in the channel of about 1,500 volts/cm. This is 50 per cent higher than the critical field. However the square-root theory shows that no significant correction need be applied until the unit is operating much farther into the square-root range. In Fig. 16 the experimental points for $V_G = 0$ are replotted to the normalized scales of V_D/W_0 and I_D/I_{D0}. On the same figure are plotted the theoretical curves for the constant-mobility and square root cases. It is seen that the experimental curve is in close agreement with that for the constant-mobility theory.

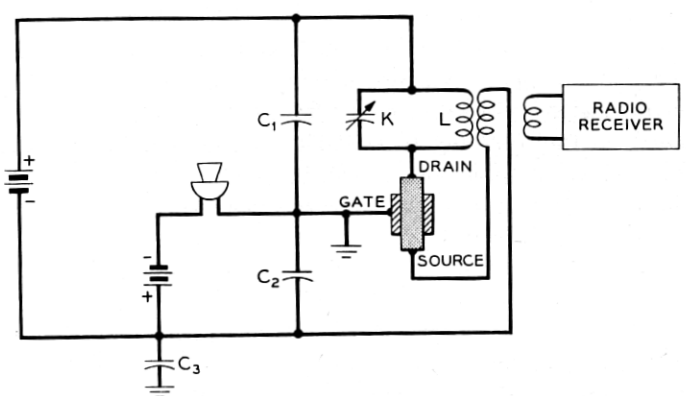

Fig. 13 — Modified circuit for demonstrating both amplitude and frequency modulation.

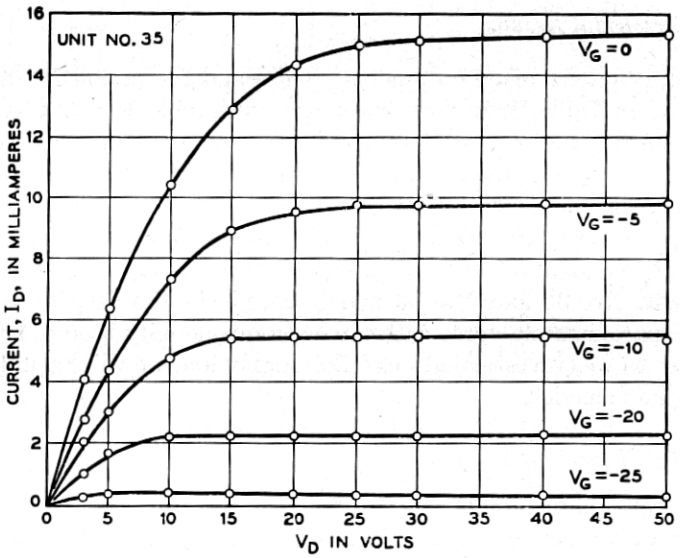

Fig. 14 — Experimental drain characteristics for unit No. 35.

In Fig. 17 a comparison has been made between the theoretical and the observed dependance of pinch-off current on the "effective gate bias," V_{Ge}, which differs from the applied bias by the voltage drop in the source resistor. The experimental points are shown as circles and the solid line is the constant-mobility curve adjusted to pass through the $V_G = 0$ point (i.e., $V_{Ge} = 0.55$). It is seen that there is good agreement.

Fig. 16 and 17 just described show that the current-voltage relationships are of the form predicted by the constant-mobility theory. It remains to determine whether the magnitudes of such characteristic parameters as I_{D0}, g_{m0}, etc., are in agreement with those calculated from the dimensions of the specimen. The accuracy of such a comparison will depend on the accuracy to which the dimensions of the specimen are known. The channel length, L, can be measured under a microscope while the channel thickness, a, can be calculated from the value of pinch-off voltage. The width Z of the channel, however, is much more difficult to determine. From sections made of other specimens, it appears that the channel cross-section tends to be elliptical rather than rectangular.* The exact influence that this departure will have on the effective width of the channel has not been determined, but certainly the effective value of Z will be less than the total thickness of the blank. A reasonable

* This effect can perhaps be minimized by the use of the 111 plane orientation of the original material during alloying.

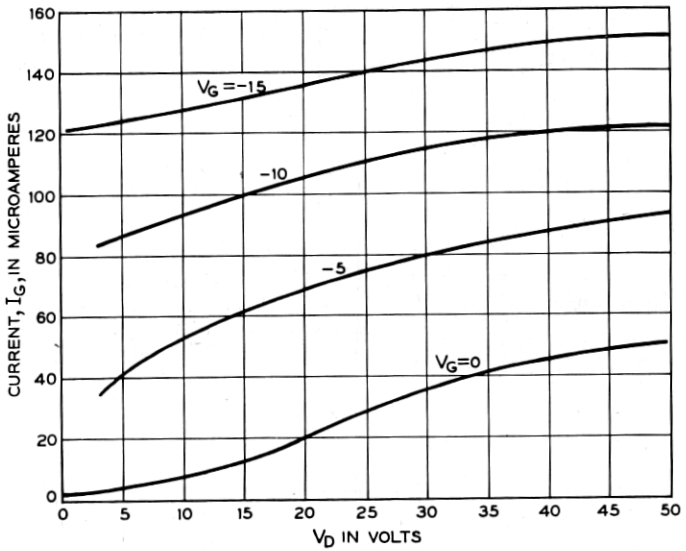

Fig. 15 — Gate characteristic for unit No. 35.

approach is to take Z as an unknown and on a basis of comparison of theory and experiment, determine its value and see if it is reasonable.

The measured transconductance at zero effective gate bias is 1.65 ma/v, whereas theory gives a value of 12 ma/v per cm in the Z direction. This comparison would suggest an effective channel width Z of 0.14 cm as compared to a blank width of 0.32 cm. Taking Z as 0.14 cm, the theory gives a value of pinched-off current at zero gate bias of 15.1 ma, and this is in good agreement with the experimental value of 15.8 ma. We conclude that the experimental values are in good agreement with the constant-mobility theory if Z is taken to be about one-half the blank thickness. This is deemed to be a reasonable supposition.

3.2 *High Frequency Performance No. 35*

The performance of the unit as an oscillator gave further confirmation of the theory. At comparatively low frequency, say 100 kc/s, the occurrence of oscillation was accompanied by a forward current at the gate. This is analogous to the grid current observed when a vacuum tube oscillates. When a negative bias was applied to the gate, the amplitude of oscillation (as observed on an oscilloscope) increased and, by adjusting the gate bias, a maximum peak to peak amplitude equal to twice the drain bias was obtained. This behavior shows that the transconductance was sufficiently high so that the amplitude of oscillation was not limited

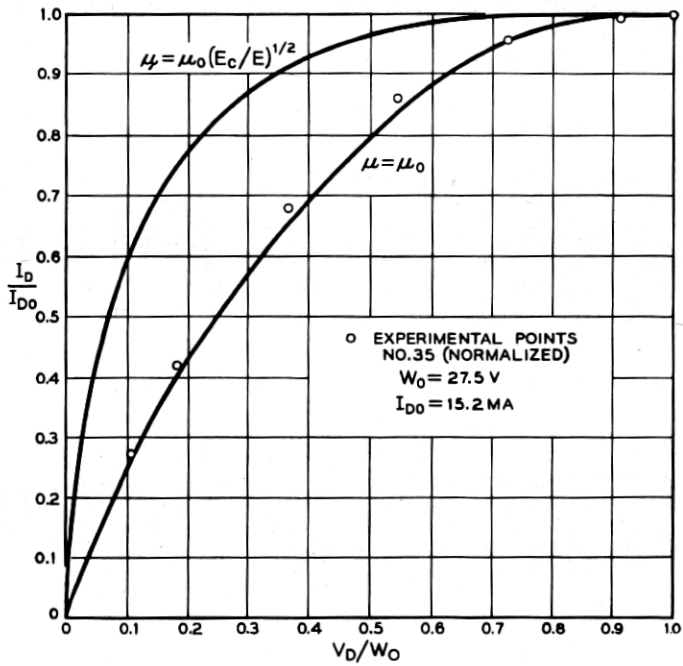

Fig. 16 — Comparison of the experimental drain characteristic for unit No. 35 with the two theoretical curves. It can be seen that the experiment points agree with the constant-mobility theory.

by it, but rather by the bias voltages. At higher frequencies the forward gate-current was not observed. Furthermore the amplitude of the oscillation was much less than twice the drain bias and decreased with frequency. This implies that at these higher frequencies the circuit had become transconductance limited. As the frequency increases so does the required transconductance and, since the maximum transconductance is obtained at very small bias, the gate bias and therefore signal amplitude must decrease.

The highest frequency at which continuous oscillation could be obtained was 31 mc/s. This figure agrees reasonably well with the value of 48 mc/s predicted from the theory. At slightly higher frequency, oscillation occurred for a few seconds after the bias voltages were applied, and then died out. This effect can be attributed to the temperature dependence of the transconductance. When the power was first turned on the transconductance was just big enough to support oscillation. After switch-on the temperature of the specimen increased slightly causing a decrease in conductivity of the germanium. Since the transconductance

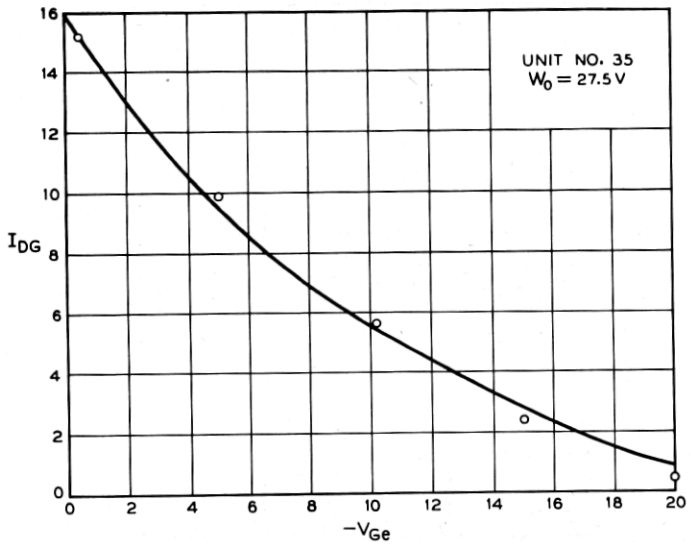

Fig. 17 — Dependence of saturated drain current on gate voltage. The solid line is derived from the constant-mobility theory while points are experimental values for unit No. 35.

is proportional to the conductivity, the transconductance fell slightly and oscillation ceased.

With this specimen it was found possible to amplitude and frequency modulate the oscillation. A telephone transmitter was used to modulate the gate bias as shown in Fig. 13. The mechanism for modulation is as follows. When the gate bias is varied both the transconductance of the device and its input capacity vary. The first effect causes amplitude modulation of the oscillations under conditions when operation is transconductance limited, that is, at the higher frequencies. Since the input impedance appears across the tank circuit, the change in input capacity causes frequency modulation under conditions when the input capacity and the tank-circuit capacity are of the same order of magnitude, that is, at the higher frequencies. It was found that both *AM* and *FM* could be obtained at frequencies close to 31 mc/s. At much lower frequencies only very slight *AM* was observed. It was not possible to look for *FM* at these lower frequencies, as the receiver was not equipped for *FM* detection in this range.

3.3 *Static Characteristics of Unit No. 32*

The static drain characteristics of unit No. 32 are shown in Fig. 18. It is seen that the apparent pinch-off voltage is about 40 volts. With a

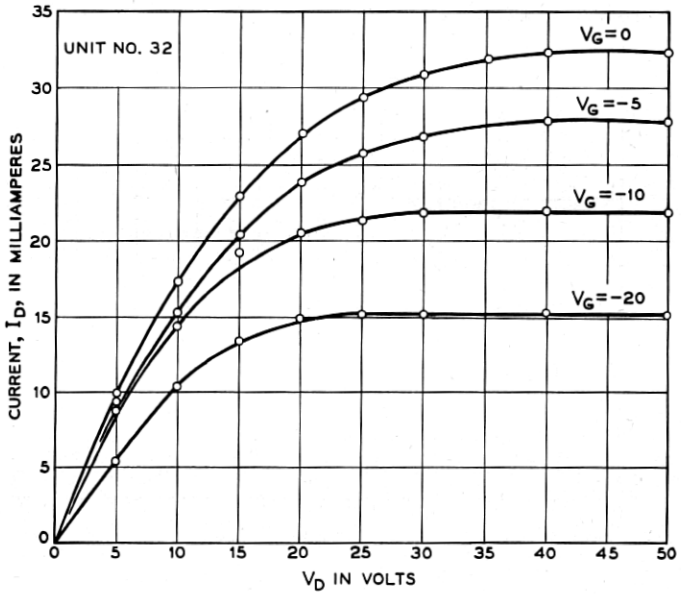

Fig. 18 — Experimental drain characteristics for unit No. 32.

gate length of 0.007″, an average field is obtained in the channel of 2300 volts/cm. This is well above the critical field and we should therefore expect departure from the constant-mobility theory.

The dependence of pinched-off drain current on effective gate voltage for unit No. 32 is shown in Fig. 19. The observed points are shown as circles. The two dashed curves are derived from the constant-mobility theory for two values of pinch-off voltage — W_0 = 40v and W_0 = 55v. It can be seen that the slopes of these curves are too steep to fit the observed points. (Compare their fit with that of Fig. 17). The *solid* line is derived from the square-root mobility theory for W_0 = 55v, and fits the observed points well. The value of 55 volts was chosen to give the best fit. Since the square-root mobility theory predicts a very gradual pinch-off it would be difficult to determine W_0 accurately from the $I_D - V_D$ characteristic, but referring to Fig. 18 one may see that a value of 55 volts is not unreasonable. Thus it seems that unit No. 32 follows the general form of the square-root mobility theory.

We shall now compare the *magnitudes* of I_{D0}, g_{m0}, etc., as determined by the two theories and the dimensions of the specimen, with the observed values. As before we assume that the effective width Z is such as to give agreement with I_{D0} and is reasonable. In Table III the observed values and values calculated from the two theories are compared. We compare four cases; A, B, C and D defined as follows:

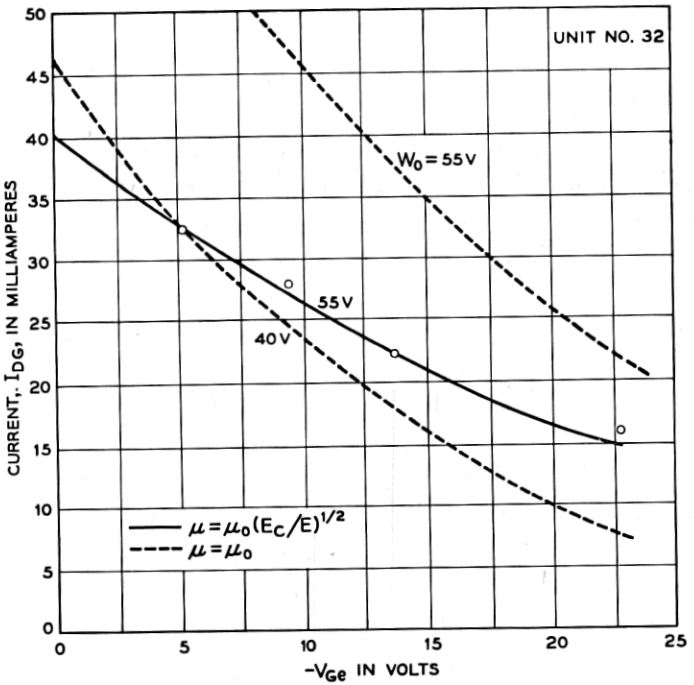

Fig. 19 — Dependence of saturated drain current on gate voltage for unit No. 32

TABLE III

	Constant μ theory		Square-root μ theory		
	A	*B*	*C*	*D*	Units
	(*obs*)	(*calc*)	(*obs*)	(*calc*)	
W_0 (assumed)	—	40	—	55	volts
W_0 (observed)	40	—	55?	—	volts
Z (assumed)	—	0.113	—	0.196	cm.
Z (observed)	0.32	—	0.32	—	cm.
g_{m0}	1.8	1.8	1.9	1.9	ma/v
I_{D0}	46	24	40	35	ma
f	50	70	50	46	mc/s

A. Observed values extrapolated to $V_{Ge} = 0$ (from $V_G = 0$ data) by using $W_0 = 40$v and the dependence of I_D and g_m on V_G given by the constant-mobility theory.

B. Values calculated from the $\mu = \mu_0$ theory using $W_0 = 40$v.

C. Observed values extrapolated to $V_{Ge} = 0$ (from $V_G = 0$ data) by using $W_0 = 55$v and the dependence of I_D and g_m on V_G given by the square-root mobility theory.

D. Values calculated from the square-root mobility theory using $W_0 = 55$v.

It is seen that, even when Z is chosen to give perfect agreement in g_{m0}, the agreement between theory and experiment is better in the square-root case than in the constant case. In addition, the value of Z that was assumed in the square-root case is a larger fraction of the blank width and therefore is more reasonable.

3.4 *High Frequency Behavior of No. 32*

Oscillations were obtained with unit No. 32 up to a frequency of 50 mc/s. This value is in reasonable agreement with the calculated value of 46 mc. The general behavior of the unit was similar to that of No. 35.

SUMMARY AND CONCLUSIONS

In the foregoing sections we have extended the design theory of field-effect transistors to include the effects of field-dependent mobility. It has been shown that when these factors are taken into account it should be possible to approach the 1,000 mc range at low power. Other possible advantages exist, however, for field-effect transistors. They can be designed for high power at lower frequencies. They can be used as direct replacements for pentode tubes without power supply alterations. It should also be possible, especially in silicon designs, to obtain very high input impedance and low noise, making the structure attractive for electrometer applications. On the debit side, the variation of g_m with V_G may lead to circuit complications in some applications. In the last analysis, of course, the eventual acceptance of the field-effect transistor hinges, as in so many cases, upon economic factors.

ACKNOWLEDGMENTS

The authors wish to acknowledge the valuable assistance of P. W. Foy and W. Wiegmann who fabricated the units described herein.

APPENDIX 1

It is of interest to calculate the transit time across the channel and the channel shape. In the constant-mobility case Shockley has shown[1] that the channel shape is

$$x = -(aI_0/I)\,[u^2/2 - u^3/3] = -6L[u^2/2 - u^3/3] \qquad (A.1)$$

where $u = b/a$. The field at any point $u(x)$ is given by

$$E(u) = I/g(u) = Ia/g_0 b = IaE_0/I_0 b = W_0/3Lu \tag{A.2}$$

so that the transit time τ is given by

$$\tau = \int_0^L dx/\mu_0 E = 18L^2 \int_0^1 [u(u - u^2)/\mu_0 W_0]\, du \tag{A.3}$$

Upon integration

$$\tau = 3L^2/2\mu_0 W_0 \tag{A.4}$$

For the non-constant mobility case we can readily calculate the channel shape as follows. From (5.3) and (5.5) we write

$$I^2 = 4\sigma_0^2 b^2 E_c (dW/db)\ (db/dx) \tag{A.5}$$

The value of dW/db can be obtained from (5.1) and is

$$dW/db = 2W_0(a - b)/a^2 \tag{A.6}$$

When this value is substituted in (A.5) we obtain

$$I^2 dx = 8\sigma_0^2 E_c W_0 a^2 u^2 (1 - u) du \tag{A.7}$$

where again $u = b/a$. Upon integration we obtain

$$I^2 x = 8\sigma_0^2 E_c a^2 W_0 (u^3/3 - u^4/4) \tag{A.8}$$

Thus if we let $I = I_c$ when $x = L$, $u = 1$ we obtain

$$I_c = \sigma_0 a (2E_c W_0/3L)^{1/2} \tag{A.9}$$

as before. Using (A.9) to eliminate I from (A.8) we finally obtain

$$x = 12L(u^3/3 - u^4/4) \tag{A.10}$$

In Fig. 20 the channel shape for the constant and square-root mobility cases as obtained from (A.1) and (A.10) are shown. It can be seen that the two channels are of similar general shape and that the non-constant mobility case would behave experimentally like a constant-mobility channel with smaller a.

The transit time τ_c for the non-linear case is readily obtained from the integral

$$\begin{aligned} \tau_c &= \int_0^L dx/\mu_0 (E_c E)^{1/2} = (12L/\mu_0)(3L/E_c W_0)^{1/2} \int_0^1 (u^3 - u^4)\, du \\ &= (3L/5\mu_0)(6L/E_c W_0)^{1/2} = 1.47L/(\mu_0^2 E_c W_0/L)^{1/2} \end{aligned} \tag{A.11}$$

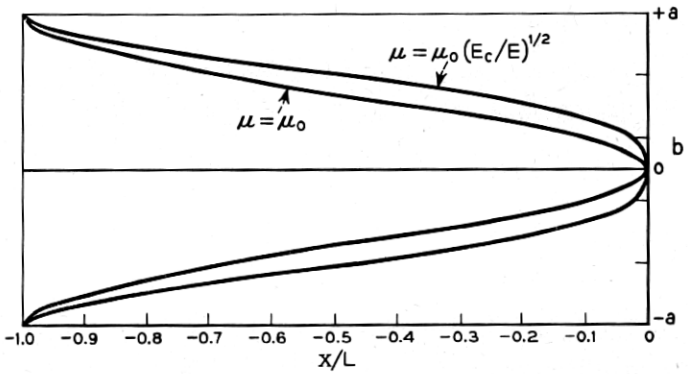

Fig. 20 — Comparison of theoretical channel-shapes for the constant and non-constant mobility cases.

The expression $(\mu_0^2 E_0/L)^{1/2}$ is the velocity corresponding to square root mobility in the average field $E_a = W_0/L$.

APPENDIX 2

Proof that in a unity coupled feedback oscillator circuit, the maximum oscillation frequency of a field-effect transistor is approximately equal to the theoretical value f_1.

We assume that the only reactive element in the transistor is the capacitance across the space-charge layer. This capacitance has to be charged through the channel resistance. Now if we represent the capacitance and resistance by lumped circuit elements, in accordance with Shockley's theory, the transistor may be considered to be as shown in Fig. 21(a). When an incremental voltage ΔV_1 is applied between gate and source, an incremental voltage ΔV_2 will appear across the capacitance C causing a change, ΔI, in drain current proportional to ΔV_2. That is

$$\Delta I \propto AV_2$$

Now at d-c, $\Delta V_2 = \Delta V_1$ and $\Delta I = g_{m0}\Delta V_1$, where g_{m0} is the maximum zero-bias transconductance. Therefore the constant of proportionality is g_{m0} and we may write

$$\Delta I = g_{m0}\Delta V_2$$

At any frequency $f = \omega/2\pi$

$$\Delta V_2 = \Delta V_1/(1 + j\omega CR)$$

$$\therefore \Delta I = g_{m0}\Delta V_1/(1 + j\omega CR)$$

and we may define an A.C. transconductance, g_ω, as

$$g_\omega = \Delta I/\Delta V_1 = g_{m0}/(1 + j\omega CR) \qquad \text{(A2.12)}$$

The input impedance at the gate is simply that of C and R in series Using this fact and the result of equation (A2.12), the transistor may be represented by the equivalent circuit of Fig. 21(b).

Now, consider the circuit of Fig. 22(a) which represents a feedback oscillator using a field-effect transistor. If we assume that the feedback coils are unity coupled, then oscillation will be just possible when V_1

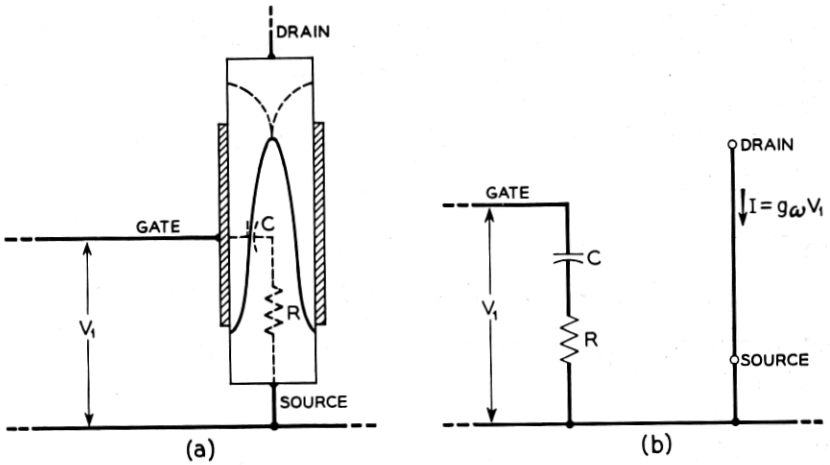

Fig. 21 — (a) Schematic diagram of equivalent circuit. (b) equivalent circuit.

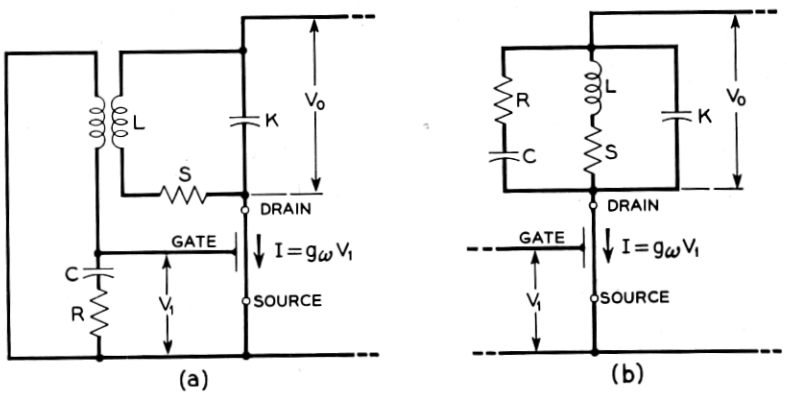

Fig. 22 — (a) oscillator equivalent circuit. (b) simplification of (a) assuming unity coupling.

and V_0 are equal in magnitude and phase. Furthermore, the input impedance of the gate will effectively appear across the tank circuit as shown in Fig. 22(b). Now the theoretical cut-off frequency of the transistor is f_1 where

$$f_1 = \frac{1}{2\pi CR} = \omega_1/2\pi \tag{A2.13}$$

At frequencies far below f_1, the shunting impedance of C and R will be high and can be neglected. If, then, the Q of the tank circuit (LKS) is sufficiently high, the circuit will oscillate. However, at frequencies close to f_1, the shunting impedance will be of the order of R, which in the units tested was about 100 Ω. At such frequencies, if the Q of the tank circuit is high, the effect of the resistance S will be small compared to that of R and may be neglected. We will now analyze the circuit neglecting S.

Then

$$V_0/V_1 = g_\omega Z$$

where Z is the impedance of the tank circuit. Substituting for g_ω and Z gives

$$\begin{aligned} V_0/V_1 &= [g_{m0}/(1 + j\omega CR)]\left[\frac{1}{j\omega L} + j\omega K + j\omega C/(1 + j\omega CR)\right]^{-1} \\ &= g_{m0}\frac{j\omega L[1 - \omega^2 LC - \omega^2 LK - j\omega CR(1 - \omega^2 LK)]}{(1 - \omega^2 LC - \omega^2 LK)^2 + \omega^2 C^2 R^2[1 - \omega^2 LK]^2} \end{aligned} \tag{A2.14}$$

This is real if

$$\omega^2 LC + \omega^2 LK = 1 \tag{A2.15}$$

Substituting this condition into equation (A2.14) and putting $V_0/V_1 = 1$ gives

$$g_{m0}\omega^2 LCR[1 - \omega^2 LK] = \omega^2 C^2 R^2[1 - \omega^2 LK]^2$$

and if $1 - \omega^2 LK \neq 0$

$$g_{m0}L = CR(1 - \omega^2 LK)$$

Substituting for $(1 - \omega^2 LK)$ from equation (A2.15) gives

$$g_{m0} = \omega^2 C^2 R^2$$

$$= \frac{\omega^2}{\omega_1^2} \cdot \frac{1}{R}$$

$$\therefore \omega^2 = \omega_1^2 g_{m0} R.$$

But

$$g_{m0} = \frac{1}{R_0} \div \frac{2}{R}. \tag{A2.16}$$

$$\therefore \omega = \sqrt{2}\omega_1 .$$

The maximum possible frequency of oscillation is approximately equal to the theoretical cut-off frequency as shown in (A2.16). This analysis was made neglecting the effects of stray capacities. However, each of the strays can be considered as equivalent to a capacity in parallel with the tank circuit and therefore simply changes the effective magnitude of K. Thus (A2.16) is valid in the presence of stray capacity. Therefore by using a closely-coupled feedback oscillator and determining the maximum frequency at which oscillations can be obtained the theoretical cut-off frequency can be approximately determined.

REFERENCES

1. W. Shockley, Proc. I.R.E., **40**, p. 1374, Nov., 1952.
2. G. C. Dacey and I. M. Ross, Proc. I.R.E., **41**, Aug., 1953.
3. E. J. Ryder and W. Shockley, Phys. Rev., **81**, p. 139, 1951; and W. Shockley, B. S. T. J. **30**, p. 990, 1951.
4. S. L. Miller, Avalanche Breakdown in Germanium. Phys. Rev. (To be published).