

The Metal-Oxide-Semiconductor (MOS) Field-Effect Transistors

Acknowledgment

The author wishes to thank Mr. S. F. Guo for his guidance and careful revision of the manuscript and Mr. S. C. Wang for stimulating discussions.

L. I. Chen (陳龍英)

Chiao Tung University, Hsinchu, Taiwan

Abstract

The operation theory of MOS field-effect transistors based on some simplifying assumptions are analyzed. Then a simple method is presented to determine the effective surface carrier mobility, the parasitic resistance in the source region, and the transconductance in the saturation region. Some curves which are necessary in the method are plotted. The saturation drain current and gate voltage of these curves are normalized to the zero bias saturation drain current and pinch-off voltage respectively, thus this method is valid for different kinds of MOS transistors. The curves show that the parasitic resistance may be neglected for most of the practical MOS transistors.

I. Introduction

The field effect principle has been studied since 1930, and the research work led to the discovery of the point contact transistors in 1948 and the invention of the junction transistor thereafter. But those two types of transistors operate on the principle of minority carrier injection rather than field effect. In 1952, an active solid-state field effect device, in which a reverse biased p-n junction was used as the field effect electrode, was successfully built. But the lack of the controllability and stability of the surface prevented the progress of this junction type field effect transistor towards a practical one. The surface states were found to be highly dependent upon surface conditions, and were found to be the main source of the obstacle.

A relatively stable surface field effect active device fabricated by

silicon planar techniques has been available the past few years. The surface of this new type field effect transistor is protected by an oxide layer which reduces the concentration of surface states from the order of $10^{15}/\text{cm}^2$ to $10^{12}/\text{cm}^2$ and makes the surface more stable. Because of the device structure, the name metal-oxide-semiconductor transistor or MOS transistor was adopted.

II. The Operation Theory of MOS Field-Effect Transistor

II-1 Comparison between Field-Effect Transistor and Junction Transistor

Field-Effect Transistor	Junction Transistor
Current is carried by the majority carriers drifted by an electric field.	Current is carried by the minority carriers diffusing in an essentially field-free region
The transit time is shorter because of the higher drift velocities.	The transit time is longer because of the lower diffusion velocities.
Operating frequencies can be higher.	Operating frequencies are lower.
High input impedance	Lower input impedance
Voltage amplification	Current amplification

II-2 The Basic Structure of MOS Transistor

The Structures of MOS transistor are shown in Figs. 1(a)-(b), Fig.1(a) is the circular geometry while Fig.1(b) is the linear geometry. The source region "S" is usually internally connected to the bulk by a metal layer, otherwise the device would become a four terminal structure. The gate "G" which is a metal electrode over the oxide, is employed as the input lead, while the drain region D is usually employed as the output lead with the source as the common lead in the input and output. Metals such as aluminum, silver, gold, platinum, chromium and others are generally used, oxide thickness is of the order of 1000\AA . Thick oxide reduces the transconductance, the gain, and speed of the device, and very thin oxide makes the reproducibility of the device difficult. The substrate here acts as a support. As shown in the figure, p-type substrate, n-type source and drain regions are assumed in this paper.

II-3 The Operation Types of MOS Transistor

1. Depletion-Type Transistor

If the surface region under the gate electrode is made with same

conductivity type as the source and drain regions during fabrication process, then a conduction path which is usually called the channel will be present between the drain and source. This type of device is known as the "depletion-type" field effect transistor. This may be operated in either the depletion mode or enhancement mode.

(a) Depletion Mode:

The depletion mode is operated with the gate at reverse biased so that the current path between the source and drain electrodes can be reduced or the carriers can be depleted from the channel. Therefore maximum channel current flows in this mode of operation at zero gate voltage.

(b) Enhancement Mode:

In contrast to the depletion mode, the enhancement mode is operated with the gate at forward biased. Then carriers can be drawn into the channel so that the current can be enhanced. Minimum channel current flows in this mode of operation at zero gate voltage.

2. Induced Channel-Type Transistor

If a device is fabricated without the built-in channel, the n-type source and drain are isolated on the p-type substrate and there is no channel current at zero gate voltage. If a positive voltage is applied to the gate then holes will be depleted from the surface between the source and drain and an inversion layer of n-type will induce a further increase in gate voltage.

II-4 Elementary Physical Theory of an Electron Conduction Device

For presenting a mathematical analysis of the MOS transistor, some assumptions must be made for a simple physical model. (1) the drain current is due to drift electron channel current only, the minority carrier current, drain junction diffusion current, and the current due to generation and recombination are all negligible compared with the electron channel current. (2) the gradual channel approximation will be employed. (3) shallow channel approximation, which assumes that the total potential drop is in the oxide will be used.

When a positive voltage is applied to the gate of a MOS transistor, surface charge will be induced in the channel. With shallow channel approximation, its density is given by

$$Q_c(y) = \frac{\epsilon_{ox}}{T_{ox}} [V_G - V(y)] \quad (1)$$

where V_G is the gate voltage, $V(y)$ is the potential in the channel with distant y from the source, ϵ_{ox} is the dielectric constant of the oxide, T_{ox} is the oxide thickness.

It may happen that only part of the induced surface charges are mobile when positive gate voltage is applied, it may also happen that electron channel exists even at $V_G=0$. This phenomenon is due to the existence of the surface state charge Q_{ss} and bulk charge Q_B . Both Q_{ss} and Q_B are assumed to be independent of position along the channel y , then the mobile charge density Q_m at any point in the channel is given by

$$Q_m(y) = \frac{\epsilon_{ox}}{T_{ox}} [V_G - V(y)] + Q_{ss} + Q_B \quad (2)$$

where Q_m , Q_{ss} and Q_B are all expressed as charge per unit area. From (2) it is evident that $Q_{ss} + Q_B$ is the number of charges per unit area initially present in the gap region. There are mobile charges somewhere in the gap region if

$$\frac{\epsilon_{ox}}{T_{ox}} [V_G - V(y)] + Q_{ss} + Q_B \geq 0$$

$$\text{i.e. } V_G - V(y) \geq -\frac{T_{ox}}{\epsilon_{ox}} (Q_{ss} + Q_B) \quad (3)$$

The pinch-off voltage is defined as the potential difference across the oxide at any point in the channel at which the mobile charge will go to vanish at that point. From (3) the pinch-off voltage is

$$V_P = -\frac{T_{ox}}{\epsilon_{ox}} (Q_{ss} + Q_B) \quad (4)$$

Substituting (4), (2) becomes

$$Q_m(y) = \frac{\epsilon_{ox}}{T_{ox}} \{ [V_G - V(y)] - V_P \} \quad (5)$$

It is evident from (4) and (5) that if $Q_{ss} + Q_B < 0$, there are a large number of acceptor states or traps which must be filled by field effect before the conduction path is induced. Thus a positive gate bias must be applied to fill the acceptor-like surface states before an accumulation layer at the surface is formed. The pinch-off voltage of this induced channel-type unit is the minimum gate voltage required to turn on the drain current. On the other hand, if $Q_{ss} + Q_B > 0$, the pinch-off voltage of this depletion

type unit is the minimum voltage required to empty the donor-like surface states and turn off the surface channel.

The channel current may be obtained by using Ohm's law

$$I_D = \frac{\Delta V}{\Delta R(y)} \quad (6)$$

where ΔR is the resistance of the infinitesimal channel section of width W and length Δy , i.e.,

$$R(y) = \frac{\Delta y}{Q_m(y) \mu_n W} \quad (7)$$

Substituting (7) into (6):

$$I_D = Q_m(y) \mu_n W \frac{\Delta V}{\Delta y} \quad (8)$$

and integrating from one end of the channel to the other yields

$$I_D \int_0^L dy = W \int_{V_S}^{V_D} \mu_n Q_m(y) dV \quad (9)$$

where V_S and V_D are source and drain voltage, respectively.

The electron mobility μ_n is not constant along the channel, it is generally less than the bulk mobility due to the additional surface scattering. In order to calculate (9) it is convenient to define an average surface mobility $\bar{\mu}_n$ by

$$\bar{\mu}_n = \frac{\int_{V_S}^{V_D} \mu_n Q_m(y) dV}{\int_{V_S}^{V_D} Q_m(y) dV} \quad (10)$$

then (9) becomes

$$I_D = \frac{W}{L} \bar{\mu}_n \int_{V_S}^{V_D} Q_m(y) dV \quad (11)$$

Substitute (5) into (11)

$$I_D = \frac{W \bar{\mu}_n \epsilon_{OX}}{L T_{OX}} \int_{V_S}^{V_D} [V_G - V(y) - V_P] dV \quad (12)$$

Integrating this expression and letting $V_S=0$ yield the drain current

$$I_D = \frac{W \bar{\mu}_n \epsilon_{OX}}{2 L T_{OX}} [2(V_G - V_P) V_D - V_D^2] \quad (13)$$

Actually, there exists small value of parasitic series resistances R_{ps} and R_{pd} in source and drain respectively, therefore the limits of integration must be replaced by $I_D R_{ps}$ and $V_D - I_D R_{pd}$, then (13) becomes

$$I_D = \frac{\bar{W}\mu_n\epsilon_{OX}}{2L T_{OX}} \left\{ 2(V_G - V_P) [(V_D - I_D R_{PD}) - I_D R_{PS}] - [(V_D - I_D R_{PD})^2 - (I_D R_{PS})^2] \right\} \quad (14)$$

However, (14) is valid only for gradual channel, i.e. $V_G - V_P > V_D - I_D R_{PD}$. If $V_G - V_P \leq V_D - I_D R_{PD}$, the channel is pinch-off near the drain, and the drain current remains essentially constant as the drain voltage is further increased. The pinch-off condition may be obtained from (14) by setting $\left(\frac{\partial I_D}{\partial V_D}\right)_{V_G} = 0$, this gives

$$V_D = V_{DS} = V_G - V_P + I_D R_{PD} \quad (15)$$

Imposing the pinch-off condition yields the saturation drain current I_{DS} from (14)

$$I_{DS} = 2B \frac{V_{GP}^2}{1 + 2B R_{PS} V_{GP} + \sqrt{1 + 4B R_{PS} V_{GP}}} \quad (16)$$

where

$$V_{GP} = V_G - V_P \quad (17)$$

$$B \equiv \frac{\bar{W}\mu_n\epsilon_{OX}}{2L T_{OX}} \quad (18)$$

It should be noted that the saturation drain current I_{DS} is independent of the parasitic drain resistance R_{PD} .

If the parasitic source resistance R_{PS} is very small (16) becomes

$$I_{DS} = B V_{GP}^2 \quad (19)$$

This equation means that the transfer function is essentially parabolic and a square law device is obtained. (16) shows that the saturation drain current I_{DS} is independent of the drain voltage, and the additional increased voltage appears across the pinch-off space-charge region. This gives an infinite drain resistance. Actually, a very slight increase of the drain current does occur owing to the slight decrease of the channel length of the nonpinched-off region. This channel shortening effect gives a finite drain resistance when the drain voltage is beyond the pinch-off condition.

For most of the practical MOS transistor, the parasitic resistances are negligible. Thus, (13) and (19) can be used, and they are rewritten as

$$I_D/B = 2V_{GP} V_D - V_D^2, \text{ for } V_D < V_{GP} \quad (20)$$

$$I_{DS}/B = V_{GP}^2, \text{ for } V_D > V_{GP} \quad (21)$$

(20), (21) are shown in Fig. 2. This is the drain characteristic of an

N-channel MOS transistor.

The small signal low frequency transconductance is obtained from (14),

$$g_m = \left. \frac{\partial I_D}{\partial V_G} \right|_{V_D} = \frac{2B V_D - 2B I_D (R_{PD} + R_{PS})}{1 + 2B [V_{GP} (R_{PD} + R_{PS}) - R_{PD} V_D + I_D (R_{PD}^2 + R_{PS}^2)]} \quad (22)$$

This equation is valid only for $V_D < V_{GP}$.

If the device is operated beyond the saturation voltage, the transconductance may be obtained from (16)

$$g_{ms} = \frac{\partial I_{DS}}{\partial V_G} = 4B \frac{V_{GP}}{1 + 4B R_{PS} V_{GP} + \sqrt{1 + 4B R_{PS} V_{GP}}} \quad (23)$$

Actually, g_{ms} is the maximum value of g_m .

If the terms involving R_{PS} and R_{PD} may be neglected, then (22), (23) becomes

$$g_m = 2B V_D, \text{ for } V_D < V_{GP} \quad (24)$$

$$g_{ms} = 2B V_{GP}, \text{ for } V_D > V_{GP} \quad (25)$$

(25) may be written as

$$g_{ms} = g_{m0} \left(1 - \frac{V_G}{V_P}\right) \quad (26)$$

where g_{m0} is the zero-bias transconductance which is equal to $2B (-V_P)$.

In a first approximation, the d.c. drain resistance below saturation is computed from (20) and is given by

$$R_D \equiv \frac{V_D}{I_D} = \frac{1}{2B(V_{GP} - V_D/2)}, \quad V_D < V_{GP} \quad (27)$$

The small signal low frequency drain resistance may be obtained by differentiating (20) with V_G keeps constant,

$$r_D \equiv \left. \frac{\partial V_D}{\partial I_D} \right|_{V_G} = \frac{1}{2B (V_{GP} - V_D)}, \quad V_D < V_{GP} \quad (28)$$

The a.c. drain resistance is always less than d.c. drain resistance.

If the device is operated in the onset region where $V_D \ll V_{GP}$, then (27) and (28) become

$$r_D = R_D = \frac{1}{2B V_{GP}} \quad (29)$$

the relation of $r_D = R_D \propto \frac{1}{V_{GP}}$ will be satisfied only at low drain voltages

or at high gate voltages.

The voltage amplification factor of the device is defined as

$$u \equiv \left. \frac{\partial V_D}{\partial V_G} \right|_{I_D} = g_m r_D = \frac{V_D}{V_{GP} - V_D} \quad (30)$$

The last step is calculated by (24) and (28).

At the point of saturation the voltage amplification factor diverges due to r_D being infinite. But actually, owing to the channel length shortening effect beyond the saturation point the drain resistance is finite in the saturation region, hence the voltage amplification factor is also finite.

III. A Method for Determining the Effective Surface Carrier Mobility, the Parasitic Resistance in Source Region, and the Transconductance in Saturation Region

In this section a simple method for determining the effective surface carrier mobility, the parasitic resistance in source, and the transconductance in saturation region of MOS transistors will be introduced.

Only (16) and (23) which yield the saturation drain current and the transconductance in saturation region will be used. Since both (16) and (23) have shown good agreement in experiments by C. T. Sah, an acceptable numerical "per cent error" may safely be assumed in this method.

The depletion type will be assumed first, and then the induced-channel type will be discussed.

If equation (16) is normalized to the saturation drain current for $V_G=0$, i.e. $V_{GP} = -V_P$, the results is

$$\frac{I_{DS}}{I_{DS}|_{V_G=0}} = I_{DS}' = (1-V')^2 \left(\frac{1+K+\sqrt{1+2K}}{1+K(1-V')+\sqrt{1+2K}(1-V')} \right) \quad (31)$$

where

$$V' \equiv \frac{V_G}{V_P} \quad (32)$$

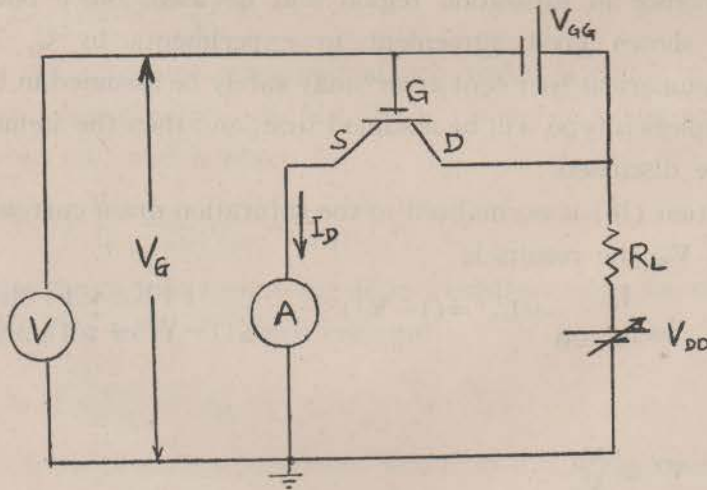
$$K \equiv -2B R_{ps} V_P \quad (33)$$

The pinch-off voltage V_P is negative for the depletion type N-channel device, and since B and R_{ps} are all positive K is also positive. If the gate voltage V_G lies in the range from zero to V_P (negative value), then the normalized gate voltage V' is positive. Eq. (31) is plotted in Fig. (3) for

several values of K . The curves of which the value of K are less than 0.1 are so closely each other, only one curve for $K=0$ is plotted to represent this range of K . This means that if K is less than 0.1, it will take no difference as if K is zero, and hence the parasitic resistance R_{ps} can be neglected in this case.

The value of K can be rapidly obtained by a simple experiment. In this experiment the drain voltage V_D is set to a sufficient high value to saturate the drain current. If the pinch-off voltage V_P is known, then the normalized gate voltage V' can easily be obtained. Some points of the normalized saturation current I_{ds}' as a function of the normalized gate voltage can be rapidly plotted. These experimental points are then compared with the curves of Fig. (3), and the value of K can be determined.

If the pinch-off voltage V_P is unknown, it can be obtained from the two terminal characteristics in which the gate electrode is connected to the drain through a battery $V_{GG} > |V_P|$. Its positive side is tied to the drain as shown in the following figure.



In these connections, the device is in the saturation region since $V_D = V_G + V_{GG} > V_G - V_P = V_{DS}$. From (16) the onset of the drain current occurs when $V_G - V_P = 0$, thus

$$V_P = V_G = V_{D0} - V_{GG} \quad (34)$$

where V_{D0} is the drain voltage which makes the onset of the drain current.

The magnitude of K may also be obtained by another method stated below.

The normalized drain currents have been plotted in Fig. (3) as function of V' for several values of K by use of (31). It can also be plotted as function of K for several values of V' . When $V'=0$, i.e. $V_G=0$, then $I_{DS}'=1$ for all values of K , it is a horizontal line with $I_{DS}'=1$. If V' increases, it becomes a curve which intersects the verticle axis ($K=0$) at $(1-V')^2$ and is monotonic increasing to a limit $(1-V')$. Finally, if $V'=1$, it becomes a horizontal line again with $I_{DS}'=0$ for all value of K . Thus the slope of the curves at $K=0$ must increase from zero to a maximum and then decrease to zero as V' changes from zero to one. The slope at $K=0$ for any V' may

be calculated from (31) as $\left. \frac{\partial I_{DS}'}{\partial K} \right|_{K=0}$ and is

$$S_0 = \left. \frac{\partial I_{DS}'}{\partial K} \right|_{K=0} = V' (1-V')^2 \quad (35)$$

From this equation it is obvious that the slope is zero when V' equals zero or one. Now the maximum value of the slope may be found by setting $\frac{ds_0}{dV'}=0$, this becomes

$$\frac{d s_0}{d V'} = (1-V') (1-3V') = 0 \quad (36)$$

thus $V' = \frac{1}{3}$, and $S_{0(max)} = \frac{4}{27}$ (37)

Therefore at $V' = \frac{1}{3}$, the variation of I_{DS}' with K will be maximum.

The curve of I_{DS}' vs. K for $V' = \frac{1}{3}$ is plotted in Fig. (4). This curve is the most convenient one to be used to determine the magnitude of K . Two other curves are also plotted in Figs. (5) and (6) with $V' = \frac{1}{2}$ and $\frac{1}{4}$.

By setting $V' = \frac{1}{3}$ and I_{DS}' being the ratio of the saturation drain current at $V_G = V_F/3$ to the saturation drain current at $V_G=0$, then from Fig. (4) the value of K can be determined. Similarly it may be obtained by Figs. (5) and (6). The mean value of these three values of K will be considered as the most suitable one.

Eq. (16) may be written as

$$\frac{I_{DS}}{2BV_p^2} = \frac{(1-V')^2}{1+K(1-V')+\sqrt{1+2K(1-V')}} \quad (38)$$

and is plotted in Fig. (7) for $V'=0$. If the value of K has been determined, then the value of B can be calculated since the saturation drain current (in this case for $V'=0$) and the pinch-off voltage are all measurable quantities and have been measured when the value of K is to be determined.

Now the effective surface carrier mobility and the parasitic resistance in source can be calculated from (18) and (33) as

$$\bar{\mu}_n = \frac{2L}{W} \frac{T_{OX}}{\epsilon_{OX}} B \quad (39)$$

$$R_{PS} = \frac{K}{-2B V_p} \quad (40)$$

(39) can also be written as

$$\bar{\mu} = \frac{2L^2}{C_{OX}} B \quad (41)$$

where C_{OX} is the capacitance of the oxide layer over the channel area, and is given by

$$C_{OX} = \frac{WL\epsilon_{OX}}{T_{OX}} \quad (42)$$

The channel length, channel width, dielectric constant and thickness of a oxide layer are all known quantities for certain MOS transistors. Thus (39) can be used to determine the effective surface carrier mobility if B is known.

The transconductance in saturation region can be obtained from (23) which is rewritten as

$$\frac{g_{ms}}{4B(-V_p)} = \frac{(1-V')}{1+2K(1-V')+\sqrt{1+2K(1-V')}} \quad (43)$$

and is plotted in Fig. (8) for several values of V' .

IV Discussion

Several simplifying assumptions have been used in analyzing the characteristic of MOS transistors. In general the gradual channel approximation is valid over a substantial portion of the channel near the source. In the drain region, space charge effects cannot be neglected, and this approximation is no longer valid.

Since the maximum potential drop across a surface inversion layer

will not exceed half the band gap potential of the semiconductor by more than a few kT , the shallow channel approximation will be valid if the potential drop between the gate and the channel is several volts or greater.

The simple theory based on these assumptions have been compared with experimental measurements by C. T. Sah, the comparison shows good general agreement with the theory of the dc characteristics, but discrepancies are found for the transconductance. The possible sources of the discrepancies may be come from the dependence of the surface state charge, the bulk charge, and the surface mobility on the gate voltage.

The common-source MOS transistors are assumed and discussed throughout this paper. They can also be analyzed and operated with common-gate and common drain.

The surface carrier mobilities of the MOS transistors not only vary along the channel but also depend upon the drift field and the gate field. Thus if the normalized gate voltage V' and the drain voltage V_D are set to the actual operation value in determining the effective surface carrier mobilities, and the transconductance in the saturation region, then it will give more accurate results.

Up to the present, only n-channel depletion type is assumed. The method presented in Section III can also be used to the induced channel type if some modifications are made: [1] Since the pinch-off voltage of an induced channel type is positive, thus K is defined as $(+)2BR_{ps} V_p$, [2] In Figs. (3) and (8) the region of V' must be changed from two to one instead of from zero to one, [3] In Figs. (4) (5) and (6) the values of V' are replaced by $5/3$, $3/2$, and $7/4$, respectively, [4] In Fig. (7) the value of V' is replaced by two, [5] In Eq. (40) plus sign is used for the parasitic resistance, [6] Two terminal characteristics method of determining the pinch-off voltage can still be used by setting $V_{gg}=0$, [7] The normalized saturation drain current is the ratio of the current at any bias to the current with bias equal to $2 V_p$.

V. Conclusion

The dc characteristics of the simple model of MOS transistors based on some simplifying assumptions have shown good agreement with experiment by C. T. Sah. But some discrepancies are found for the differential characteristics such as transconductance especially in the region below saturation. Thus the method presented in this paper can be used to determine

accurately the effective surface carrier mobility and the parasitic resistance, but the transconductances determined by this method may give some discrepancies.

Fig. (3) shows that if K is less than 0.1 the parasitic resistance can be neglected without giving large error.

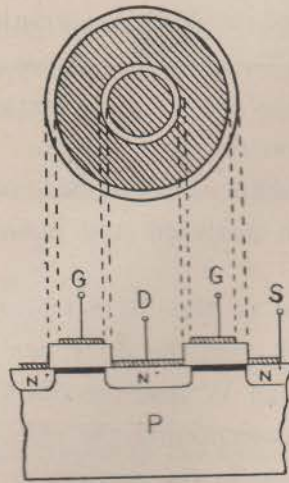
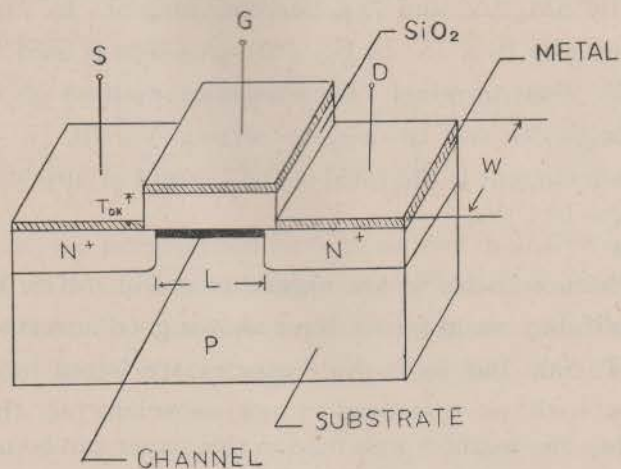


Fig. 1 (a) CIRCULAR STRUCTURE

(b) LINEAR STRUCTURE OF MOST



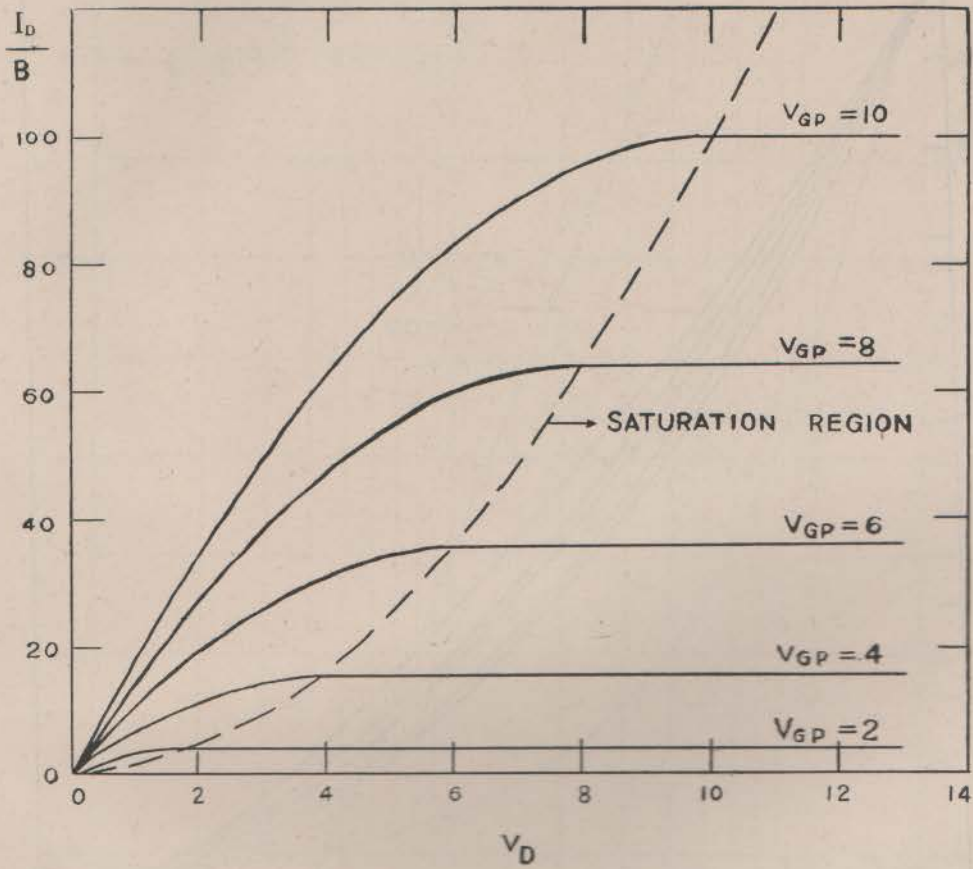
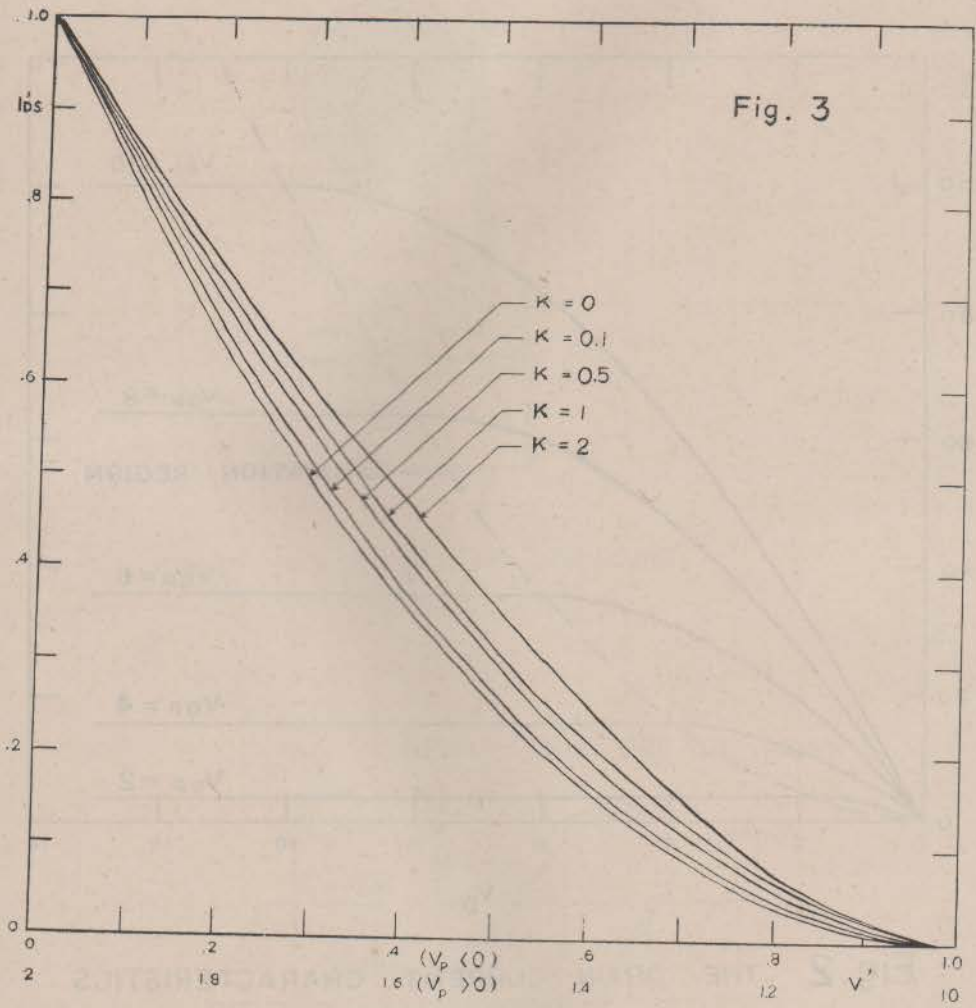
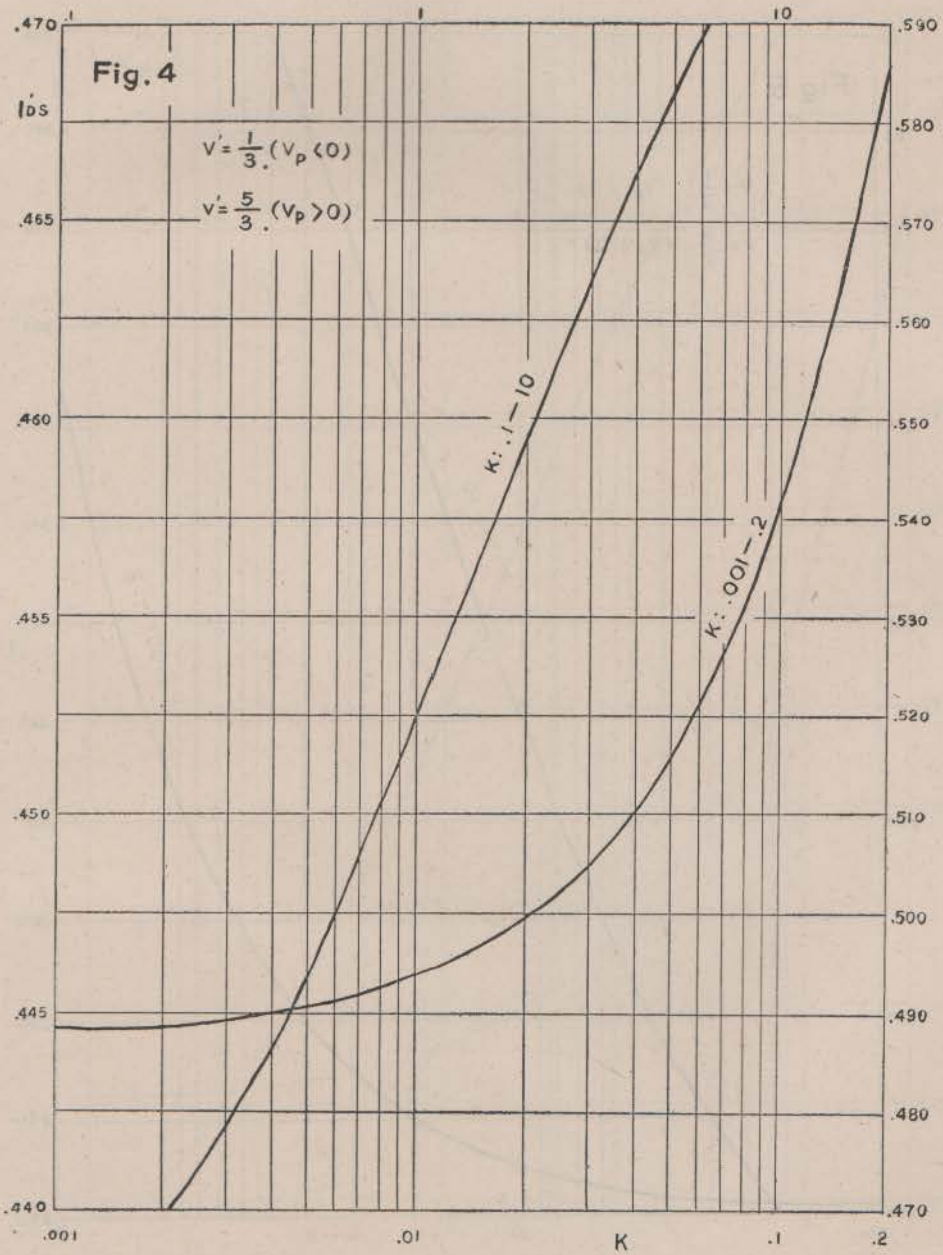
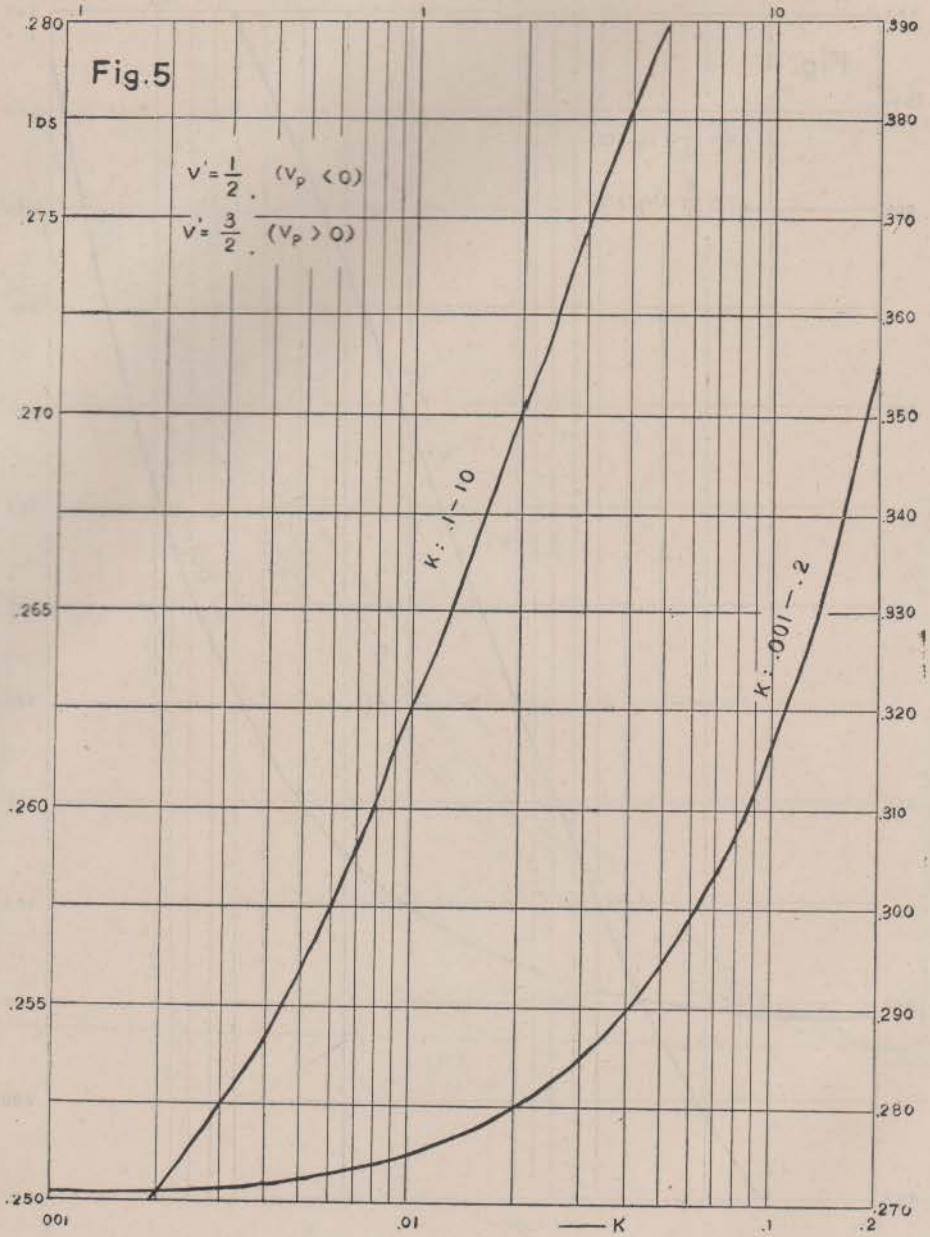
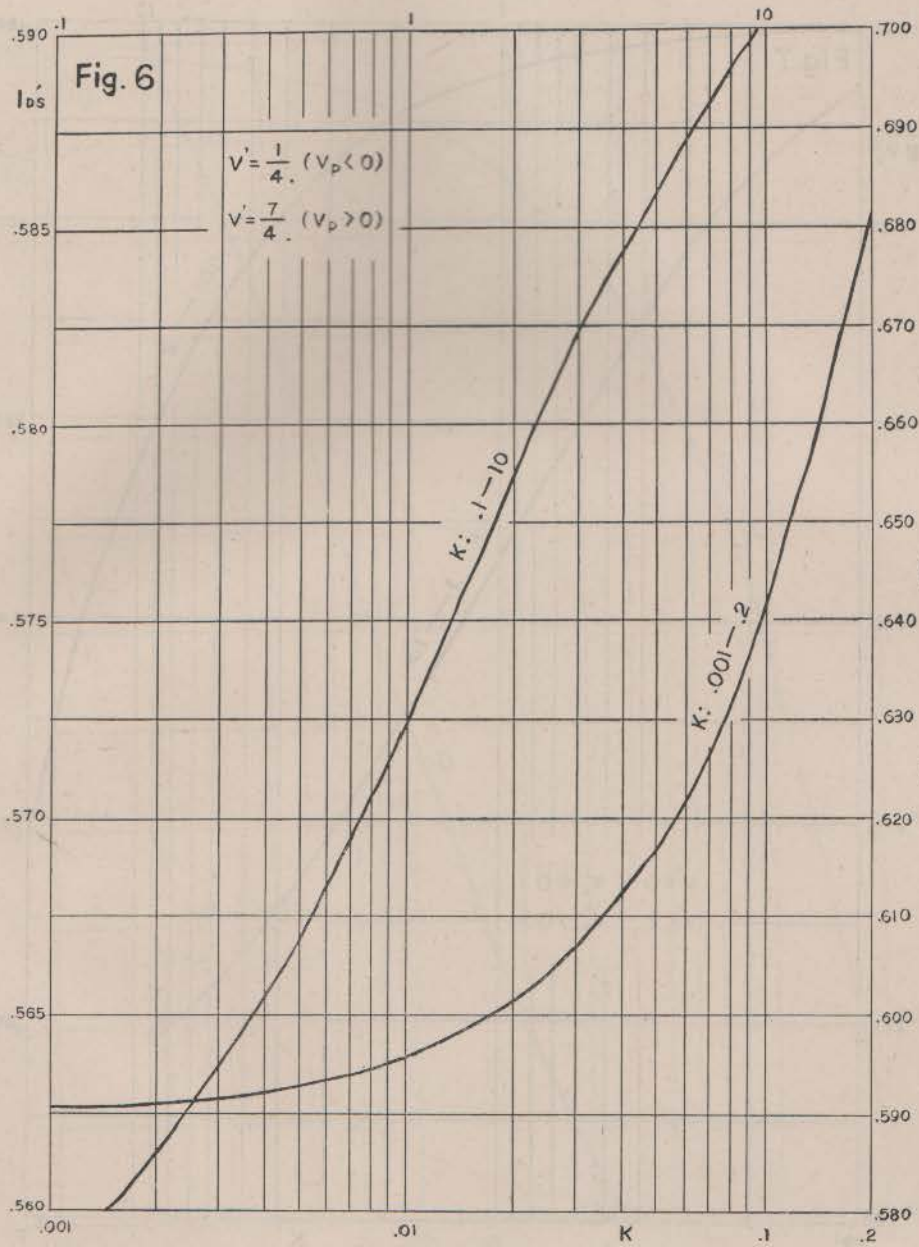


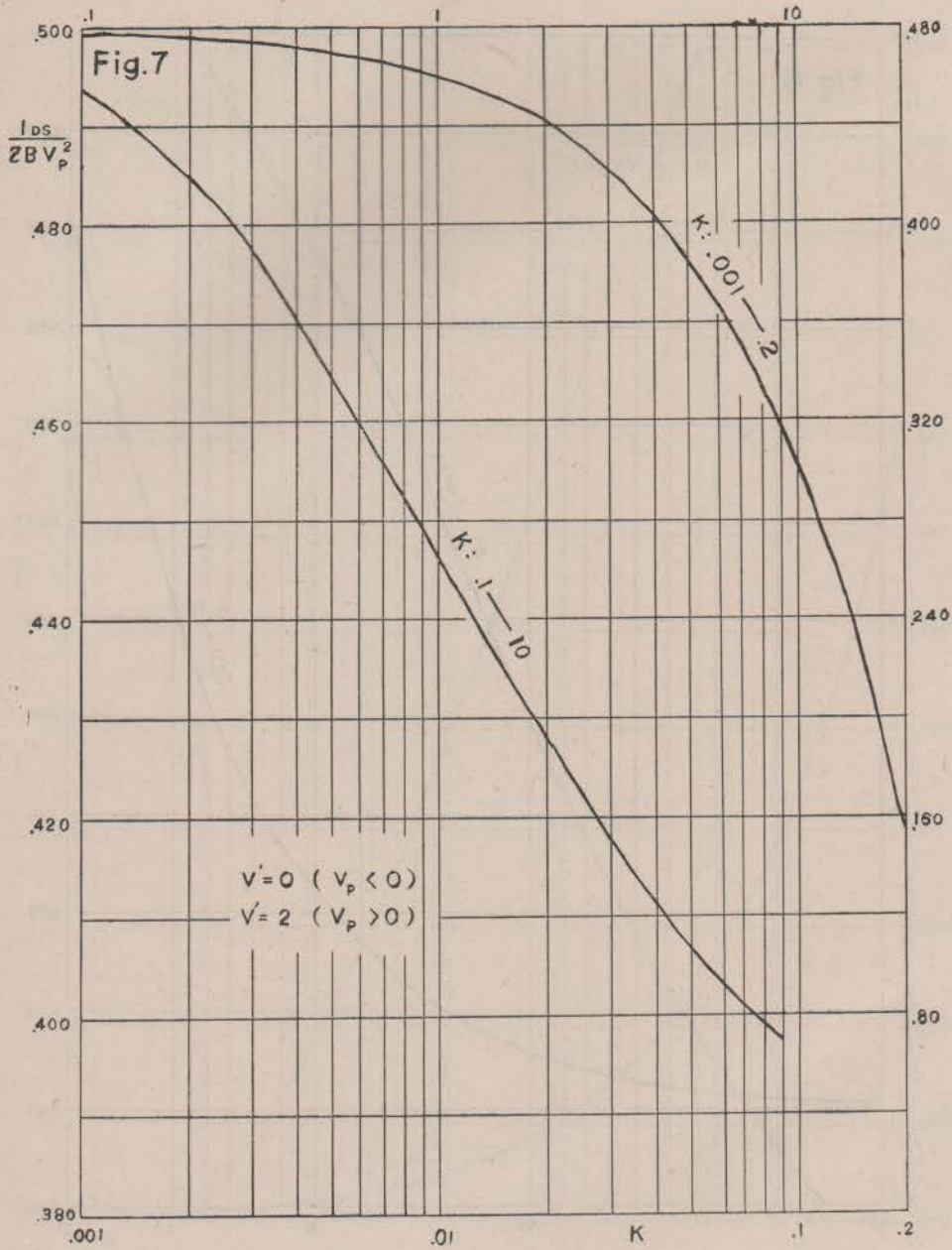
Fig. 2 THE DRAIN CURRENT CHARACTERISTICS

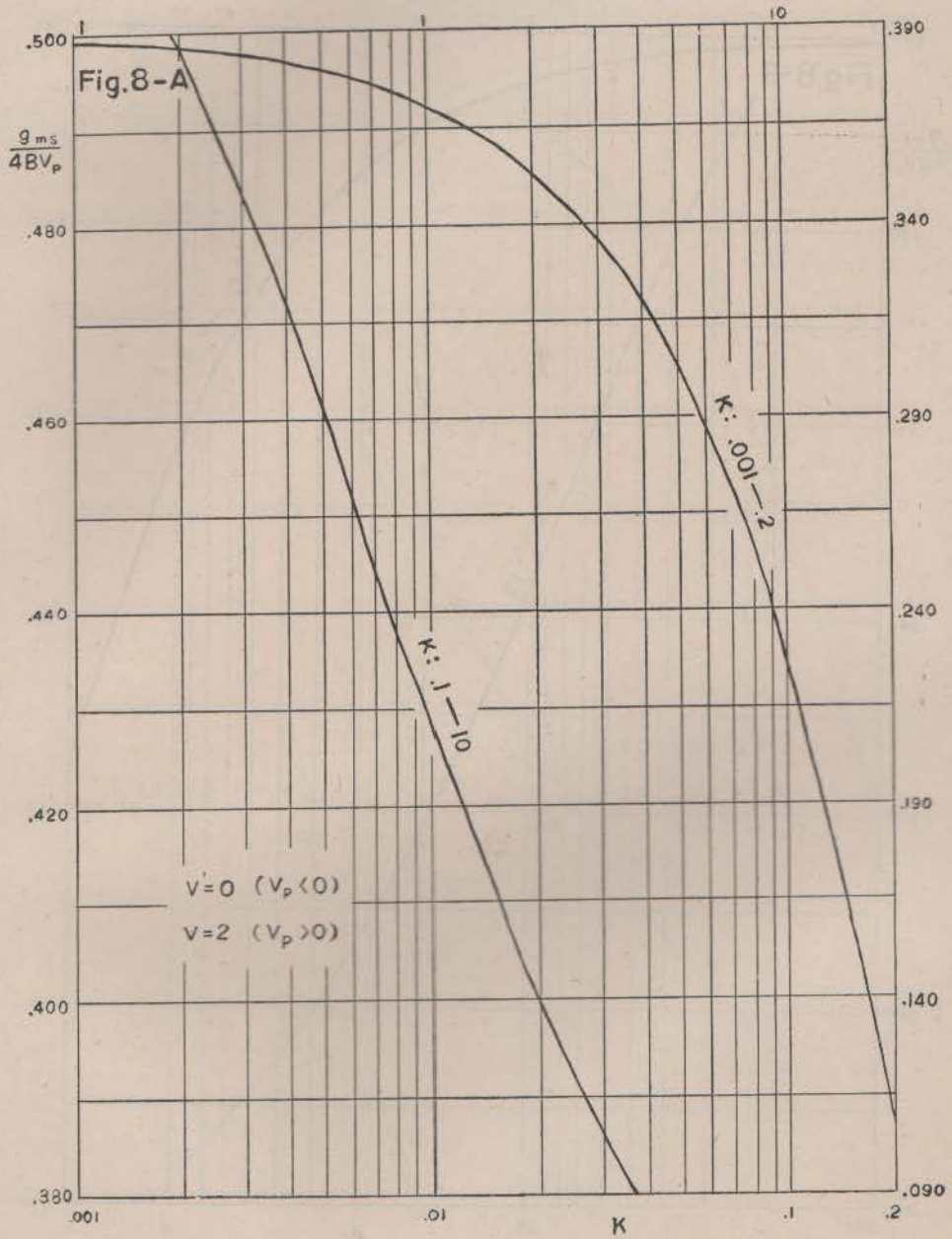


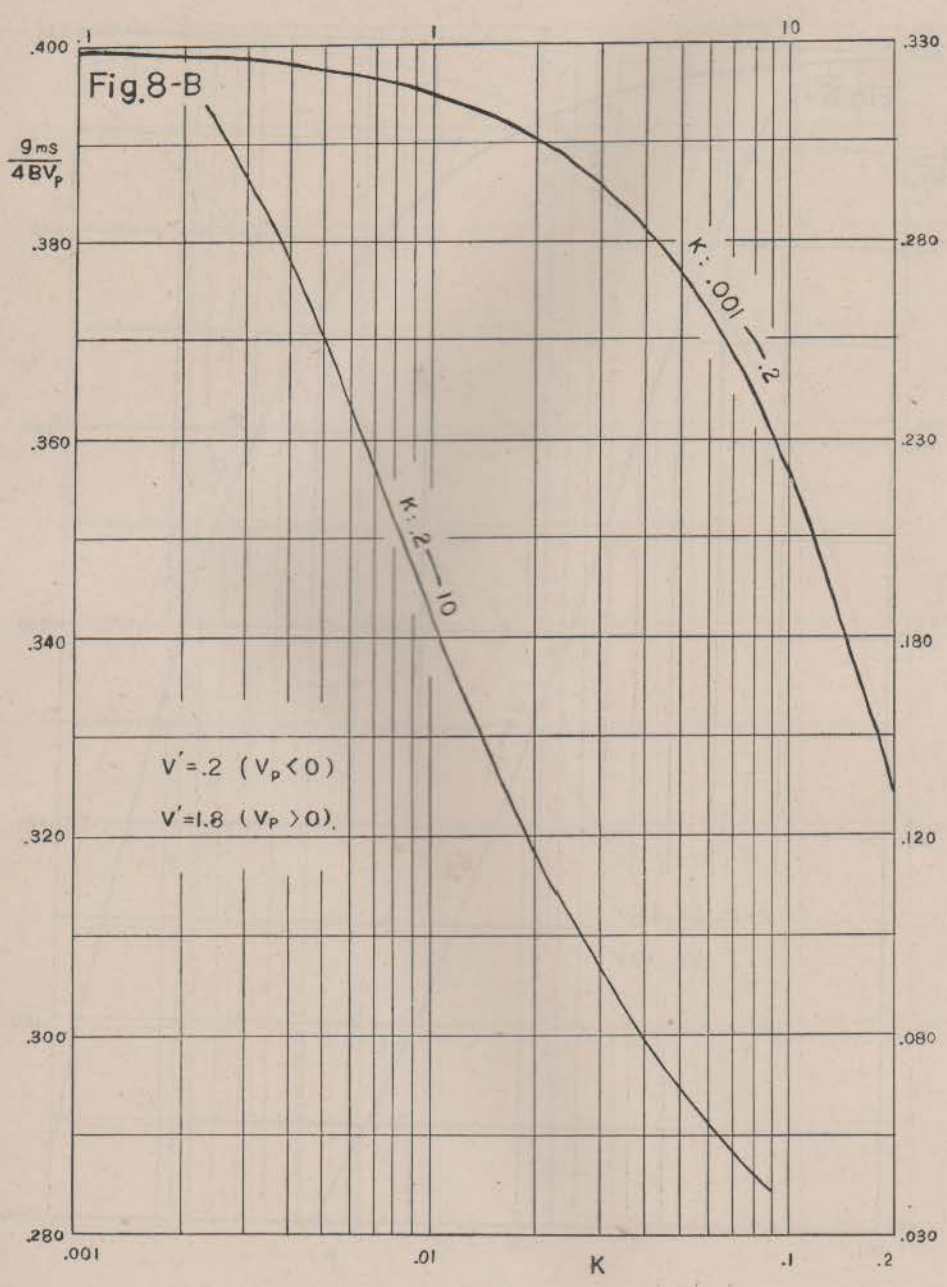


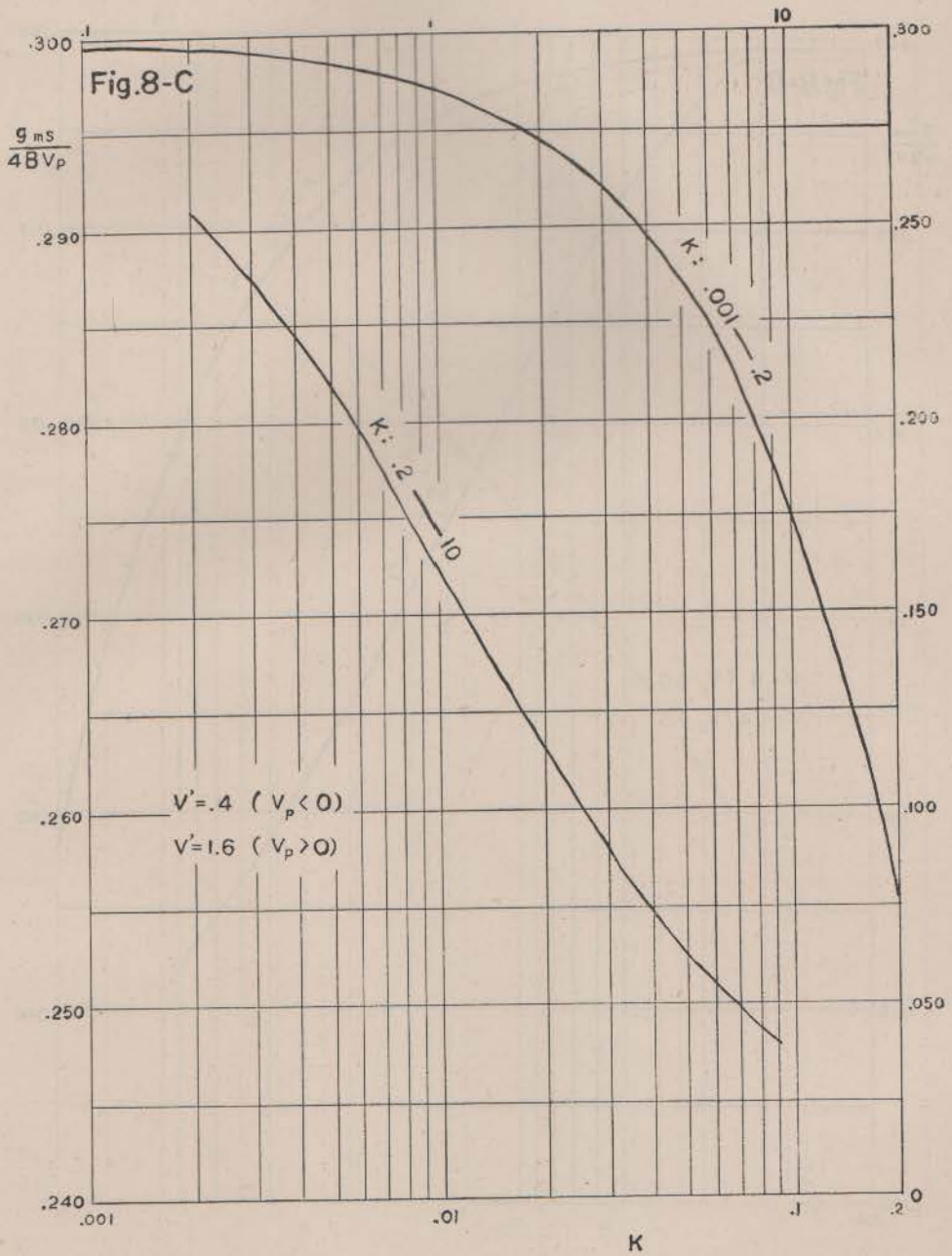


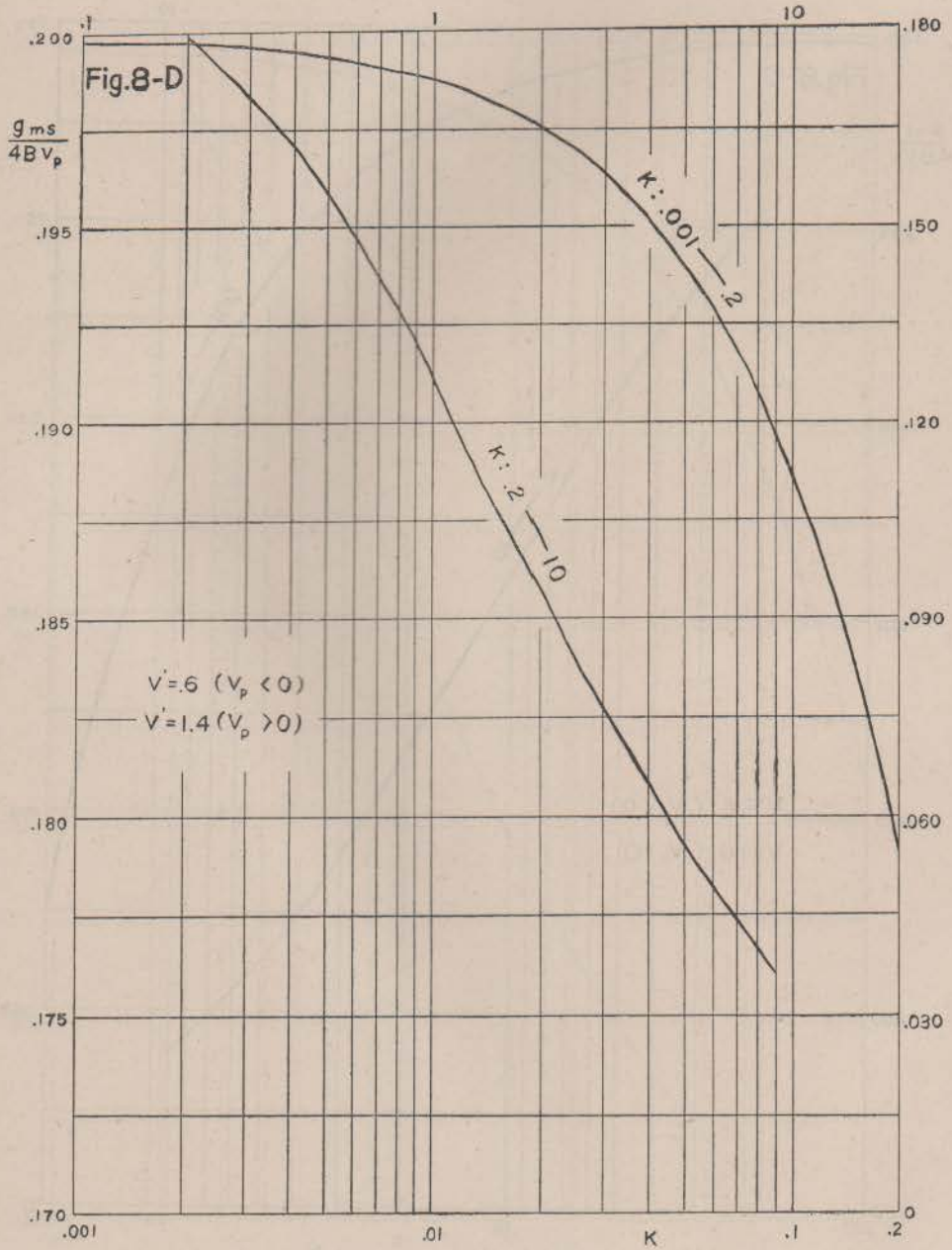


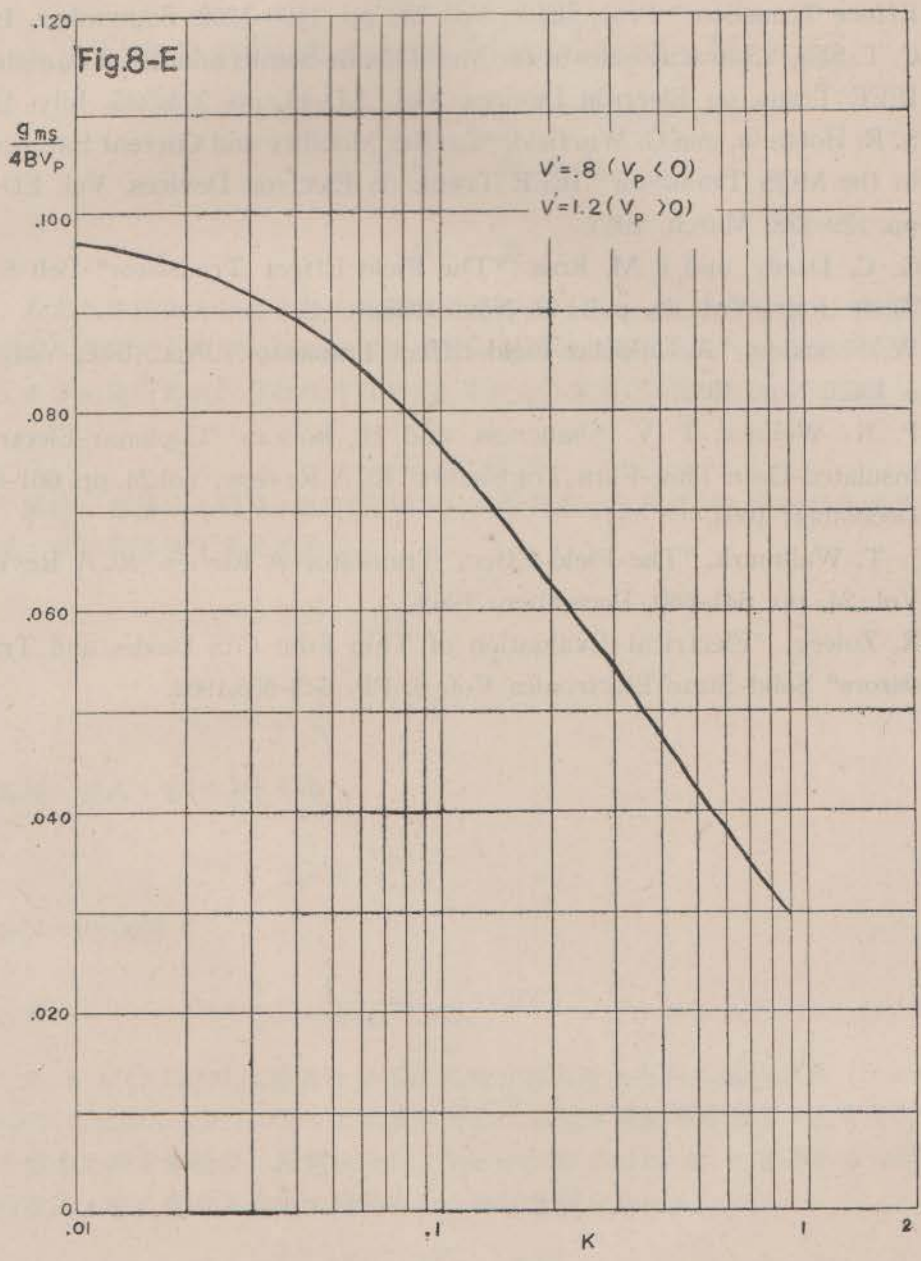












References

1. S. R. Hofstei, and F. P. Heiman, "The Silicon Insulated Gate Field-Effect Transistor" Proc. IEEE, Vol. 51, pp. 1190-1202; September, 1963.
2. C. T. Sah, "Characteristics of the Metal-Oxide-Semiconductor Transistors" IEEE Trans. on Electron Devices, Vol. ED-11, pp. 324-345; July, 1964.
3. S. R. Hofstein, and G. Warfield, "Carrier Mobility and Current Saturation in the MOS Transistor" IEEE Trans. on Electron Devices, Vol. ED-12, pp. 129-138; March, 1965.
4. G. C. Dacey, and I. M. Ross. "The Field Effect Transistor" Bell Syst. Tech. Jour., Vol. 34, p. 1149, Nov. 1955.
5. W. Shockley, "A Unipolar Field-Effect Transistor" Proc. IRE, Vol. 40, p. 1365, Nov. 1952.
6. P. K. Weimer, F V. Shallcross, and H. Borkan "Coplanar-Electrode Insulated-Gate Thin-Film Transistors" RCA Review, Vol.24, pp. 661-675, December, 1963.
7. J. T. Wallmark, "The Field-Effect Transistor-A Review"RCA Review, Vol. 24, pp. 641-660, December, 1963.
8. R. Zuleeg, "Electrical Evaluation of Thin Film CdS Diodes and Transistors" Solid-State Electronics Vol. 6, PP. 645-655,1963.