CARLETON ELECTRONICS RESEARCH LABORATORY

Department of Electronics
Carleton University
Ottawa, Canada K1S 5B6

PERFORMANCE CHARACTERIZATION OF AM/PM CONVERSION IN MICROWAVE SOLID STATE AMPLIFIERS

By
Professor J.S Wight
Miss I.M. Streibl

Technical Report No: **CERL-82-02**

For Department of Communications, Ottawa, Ontario
Under Department of Supply and Services Contract Serial No: OSU 81-00279
Sept. 18, 1981 to March 31, 1982.

PERFORMANCE CHARACTERIZATION OF AM-to-PM CONVERSION IN MICROWAVE SOLID STATE AMPLIFIERS

Submitted according to the terms of

Contract No. OSU81-00279

(RN 36100-1-0258)

(FC 4113-16350-2202)

Prepared by: Dr. J.S. Wight
(Principal Investigator)

Miss I.M. Streibl

Carleton University

March 1982

1. INTRODUCTION

This report outlines the investigations made under Contract No. OSU81-00279 into the characterization and measurement procedure for AM-to-PM distortion in microwave transistor amplifiers. The document is divided into three sections: the first concerning microwave field effect transistors, the second covering microwave bipolar transistors, and the last dealing with the measurement technique.

Part I first introduces microwave field effect transistors and presents a tutorial review of the operation of both junction field effect and metal semiconductor field effect transistors. This is followed in Chapter 3 with an analysis leading to an expression for the AM-to-PM distortion in GaAs FETs. This is based on a power series analysis of the non-linear device elements: the Schottky-barrier junction gate capacitance, the gate-voltage-dependent transconductance, the output-voltage-dependent drain conductance and the output-voltage-dependent drain capacitance. A second analysis leading to an expression for the third-order intermodulation in GaAs FETs using a Volterra series analysis on the same non-linear elements is presented in Chapter 4. Part I is concluded with a discussion of the relationship between non-linear distortion and device doping profiles.

Part II of this report first introduces microwave bipolar transistors followed by a review of considerations required in the design of microwave power devices. This is followed in Chapter 7 with an analysis leading to an expression for the third-order intermodulation in bipolar transistors using a Volterra series analysis of the non-linear device elements: the collector current, the emitter-base voltage, and the base current.

Part III deals with a systems level analysis of cascaded linear/non-linear networks which link the parameters of differential gain and differential phase to that of AM-to-PM distortion. Since the microwave link analyzer allows very accurate measurement of a device's differential gain and differential phase, this technique provides an alternative approach to the analyses of Chapters 3, 4 and 7 in evaluating the AM-to-PM distortion. The extraction of the AM-to-PM distortion coefficient k from the differential gain and differential phase measurements is discussed in detail. Also provided here is an analysis and procedure for extending this measurement beyond the dynamic range limitations of the microwave link analyzer.

Finally, concluding remarks are given in Chapter 9.

PART I

MICROWAVE FIELD EFFECT TRANSISTORS

2. MICROWAVE FIELD-EFFECT TRANSISTORS

Field-effect transistors (FETs) at microwave frequencies consist of a long narrow gate with a long narrow source closely spaced on one side and a long narrow drain closely spaced on the other. This arrangement is analogous to the long narrow emitter with a long narrow base closely spaced on either side in the planar microwave bipolar transistors. Similarly, since f_{max} for a bipolar transistor is inversely proportional to the emitter strip width S, we shall see that f_{max} for a FET depends inversely on the narrow dimension (gate length) L, of the gate electrode. Both types of device are therefore critically dependent for microwave performance on the ability to define electrodes with high geometrical precision on the semiconductor surface.

In a FET, the current flow is carried by only one type of charge carrier, hence it is a unipolar transistor. The unipolar FET has several advantages over the bipolar junction transistor:

i) It may have voltage gain in addition to current gain.

ii) Its efficiency is higher than that of a bipolar transistor.

iii) Its noise figure is lower than that of a bipolar transistor.

iv) Its operating frequency extends to X-band.

v) Its input resistance is very high, up to several Mohms.

Three FET structures are possible:

i) p-n junction FET or JFET

ii) Schottky-barrier FET or MESFET (metal-semiconductor FET)

iii) insulated gate FET or IGFET (MOSFET).

Microwave FETs have been made in both Si and GaAs. JFETs and MESFETs have been operated up to 14 GHz; IGFETs are still limited to operation below 2 GHz and consequently are not discussed in this report.

2.1 Junction Field-Effect Transistor (JFET)

The JFET was originally proposed by Shockley [1]. An n-channel FET, shown in Fig. 1, consists of N-type material sandwiched between two highly doped layers of p^+ type material. If the middle part is a P-type and it is sandwiched between N^+ layers, the FET is a p-channel device as shown in Fig. 2. In the N-channel FET, the P-type regions are called gates. Each end of the N-channel has an ohmic contact. The contact end which supplies the flowing electrons is the source; the end that collects the electrons is the drain. The source electrode is generally grounded, and the gate voltage V_g (or V_{gs}) and drain

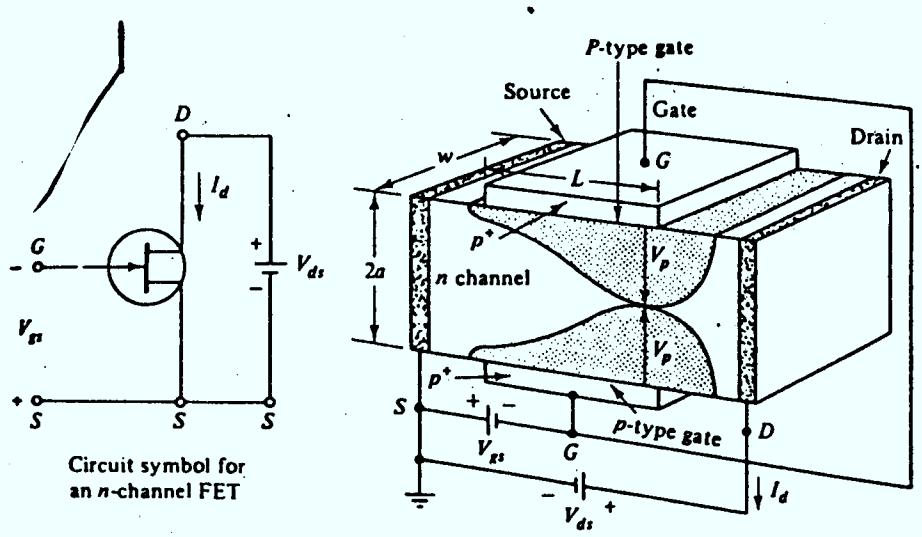


Fig. 1. Schematic diagram and circuit symbol for an n-channel FET.

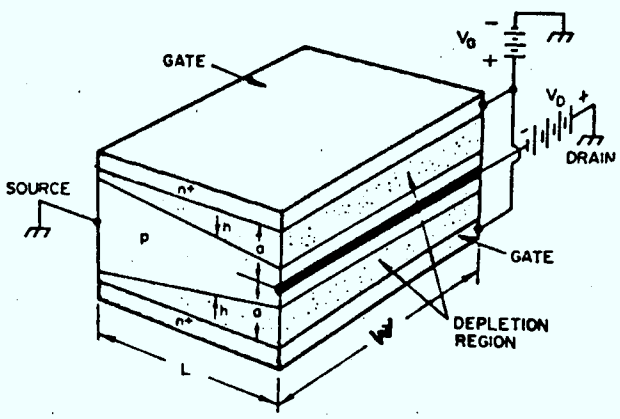


Fig. 2. Schematic diagram of a p-channel junction field effect transistor (JFET) with channel length L, channel width Z, and channel depth 2a. The source electrode is taken as the difference. Under normal operation, the gate voltage has opposite polarity as compared to source of the drain.

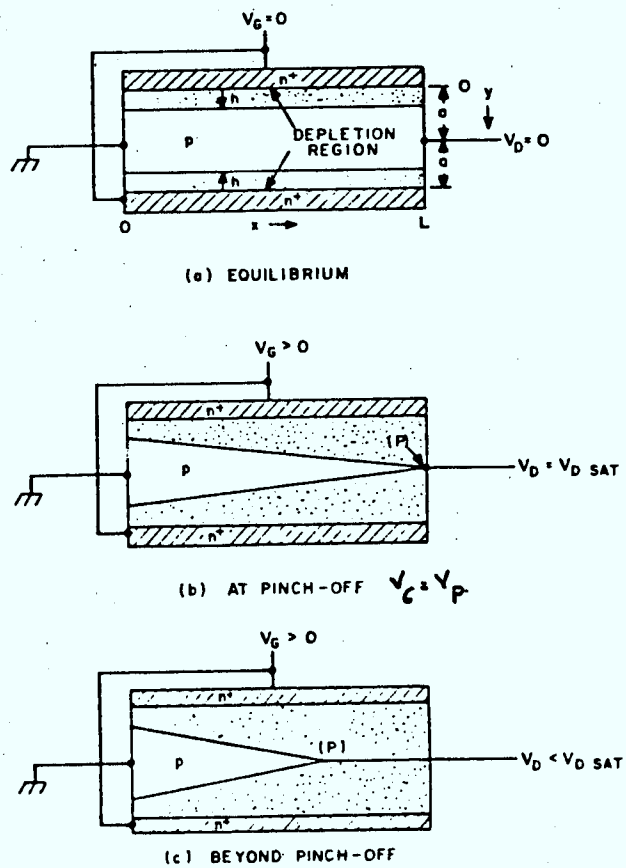


Fig. 3. Cross-sectional views of a JFET.
a) equilibrium condition
b) at pinch-off point where the depletion layers penetrate into the channel and meet at the drain end, and
c) beyond pinch-off, the point (P) moves toward the source.

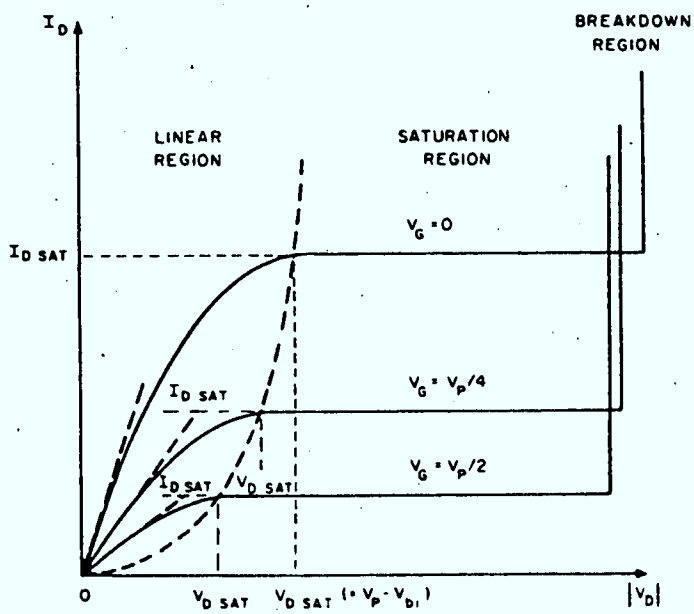


Fig. 4. Basic I-V characteristics of a JFET which includes the linear region, saturation region, and breakdown region. V_p is the pinch-off voltage. For a given V_G, the current and voltage at the point where saturation occurs are designated by I_{Dsat} and V_{Dsat}, respectively.

voltage V_d (or V_{ds}) are measured with respect to the source.

Consider a P-channel FET of channel length L, channel width W, half-channel depth at zero bias "a", and depletion layer width "h", as shown in Fig. 3. The I-V characteristics can be explained as shown in Fig. 4.

Shockley made some simplifying assumptions in his original analysis of the JFET. First, he considered the part of the channel which is not pinched off, and the pinched-off region separately. Then he made a "gradual channel approximation". In a FET, a voltage gradient exists in the x-direction in the channel because of the current flowing from source to drain. At the same time, a voltage gradient will exist in the y-direction in the gate depletion layer. The gradual approximation assumes that the x-component of the field in the depletion layer and the y-component of the field in the channel are quite small and may be neglected. From symmetry, only half the device need be considered.

Shockley's analysis leads to the following fundamental equation for the JFET given by Eqn. (1).

$$I_D = \frac{2W\mu}{\varepsilon_s L} \int_{y_1}^{y_2} [Q(a)-Q(h)]h\rho(h)dh \qquad (1)$$

From (1), we can derive the transconductance g_m as well as the channel or drain conductance g_D.

$$g_m \triangleq \frac{\partial I_D}{\partial V_G} = \frac{\partial I_D}{\partial y_1} \cdot \frac{\partial y_1}{\partial V_G} + \frac{\partial I_D}{\partial y_2} \cdot \frac{\partial y_2}{\partial V_G}$$

$$= \frac{2W\mu}{L} [Q(y_2)-Q(y_1)] \tag{2}$$

$$g_D \triangleq \frac{\partial I_D}{\partial V_D} = \frac{2W\mu}{L} [Q(a)-Q(y_2)] \tag{3}$$

Note from (2) and (3) that

$$g_{D_o}(V_D \to 0) = g_{ms}(|V_D| >> |V_p|) = \frac{2W\mu}{L} [Q(a)-Q(y_1)]$$

$$= g_{max}[1 - \frac{Q(y_1)}{Q(a)}] \tag{4}$$

where

$$g_{max} = \frac{2W\mu}{L} Q(a) \tag{5}$$

Consider a uniformly doped P-channel with doping N_A. Then

$$Q(h) = qN_A h \tag{6}$$

Here

$$I_D = \frac{2W\mu q^2 N_A^2}{\varepsilon_s L} [\frac{a}{2}(y_2^2-y_1^2) - \frac{1}{3}(y_2^3-y_1^3)] \tag{7}$$

with depletion layer widths

$$y_2 = \left[\frac{2\varepsilon_s(V_D+V_G+V_{bi})}{qN_A}\right]^{1/2} \tag{8}$$

and

$$y_1 = \left[\frac{2\varepsilon_s(V_{bi}+V_G)}{qN_A}\right]^{1/2} \tag{9}$$

where

$$V_{bi} = \frac{2kT}{q}\,\ell n\left(\frac{N_A}{n_i}\right) \tag{10}$$

for an abrupt n^+-p junction.

Normally, $V_G>0$ and $V_D<0$. However, in (8),(9) and the following equations use $|V_D|$ for V_D. Then (7) gives

$$I_D = g_{max}\left(V_D - \frac{2}{3a}\sqrt{\frac{2\varepsilon_s}{qN_A}}\left((V_D+V_G+V_{bi})^{3/2}-(V_{bi}+V_G)^{3/2}\right)\right) \tag{11}$$

where

$$g_{max} = \frac{2W\mu qN_A a}{L} \tag{12}$$

For a given V_G, I_{Dsat} occurs when the channel is pinched off, that is at $y_2 = a$.

$$I_{Dsat} = I_p\left[1-3\left(\frac{V_{bi}+V_G}{V_p}\right) + 2\left(\frac{V_{bi}+V_G}{V_p}\right)^{3/2}\right] \tag{13}$$

where I_p and V_p are the pinch-off current and pinch-off voltage (including V_{bi}) given by

$$I_p \triangleq \frac{2W\mu q^2N_A^2a^3}{6\varepsilon_s L} \tag{14}$$

$$V_p = \frac{qN_A a^2}{2\varepsilon_s} \tag{15}$$

Thus to increase I_p, a designer has the options to increase W/L, increase μ, or decrease N_A. The transconductance is

$$g_m = \frac{2W\mu qN_A}{L}(y_2 - y_1)$$

$$= \frac{2W\mu}{L}\sqrt{2\varepsilon_s qN_A}\left[\sqrt{V_D+V_G+V_{bi}} - \sqrt{V_{bi}+V_G}\right] \tag{16}$$

The drain conductance is

$$g_D = \frac{2W\mu qN_A}{L}(a - y_2)$$

$$= \frac{2W\mu}{L}\sqrt{2\varepsilon_s qN_A}\left[\sqrt{V_p} - \sqrt{V_D+V_G+V_{bi}}\right] \tag{17}$$

The saturation voltage is given by

$$V_{Dsat} = V_p - V_{bi} - V_G$$

$$= \frac{2N_A a^2}{2\varepsilon_s} - \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right) - V_G \tag{18}$$

In the linear region ($V_D \to 0$),

$$g_{D_o}(V_D \to 0) = g_{max}\left[1 - \sqrt{\frac{2\varepsilon_s(V_{bi}+V_G)}{qN_A a^2}}\right] \tag{19}$$

This theory predicts that above $V_D = V_{Dsat}$, the device will have $I_D = I_{Dsat}$. However, in practical devices, I_D still exhibits a slight increase with increasing $V_D > V_{Dsat}$ leading to a non-zero value for g_D. This effect comes about due to the modulation of the effective channel length L by the space-charge region near the drain, in a manner analogous to the base-width modulation effect in bipolar transistors.

Equations (1)-(5) can also be solved for any arbitrary N(y). It can be shown that for all potential profiles

$$I_{Dsat} = I_p \left[1 - \left(\frac{V_G + V_{bi}}{V_p}\right)\right]^n \tag{20}$$

where $2<n<2.25$. For this reason, the square-law approximation (n=2) is often made to the JFET transfer characteristics.

Parameters of Operation

The static parameters of operation that are important include the breakdown voltage, the input resistance, parasitic resistance, and temperature effects on mobility.

Breakdown occurs at the drain end of the channel due to avalanche breakdown of the gate-to-channel diode.

$$V_B = |V_D| + V_G \tag{21}$$

The input resistance is calculated from the current in the

reverse-biased gate-to-channel junction:

$$I_G = I_s[\exp(\frac{qV_G}{\eta kT})-1] \tag{22}$$

where $\eta=1$ for ideal current and $\eta=2$ for generation-recombination current. Hence

$$R_{in} = \frac{1}{g_{in}} = (\frac{\partial I_G}{\partial V_G})^{-1} = \frac{\eta kT}{q(I_G+I_s)} \tag{23}$$

Typical values range from 100 MΩ to 300 MΩ.

The parasitic resistances at the source R_S and at the drain R_D cannot be modulated by the gate voltage. Thus in the above equations the substitutions

$$\begin{aligned} V_D &\rightarrow V_D - I_D(R_S+R_D) \\ V_G &\rightarrow V_G - I_D R_S \end{aligned} \tag{24}$$

should be made to account for the IR voltage drops. Also, in the linear region, the following substitutions need to be made:

$$\begin{aligned} g_{D_o} &\rightarrow \frac{g_{D_o}}{1+(R_S+R_D)g_{D_o}} \\ g_m &\rightarrow \frac{g_m}{1+(R_S+R_D)g_m} \end{aligned} \tag{25}$$

If the doping concentration is essentially constant over some temperature range, V_p is constant. However, μ also varies with temperature so that I_p varies inversely with the square of the temperature. Also I_D decreases with increasing temperature due to temperature effects on μ and V_{bi}.

The dynamic parameters of operation that are important consist of the cut-off frequency and noise sources. These are outlined below.

A finite time interval is required for carriers to travel from source to drain. Assuming the channel mobility to be constant, the transit time is

$$\tau = \frac{L}{\mu E_x} \approx \frac{L}{\mu V_D} \tag{27}$$

For a large electric field, $v = v_{SL}$ (scattering-limited velocity) and

$$\tau \approx \frac{L}{v_{SL}} \tag{28}$$

There is also an RC time constant associated with the input capacitance C_{in} and transconductance g_m.

An equivalent circuit of a JFET in common-source configuration is shown in Fig. 5. Here g_{GD} and C_{GD} are feedback elements. Under normal operating conditions, C_{in} and $g_m V_G$ are the most important terms.

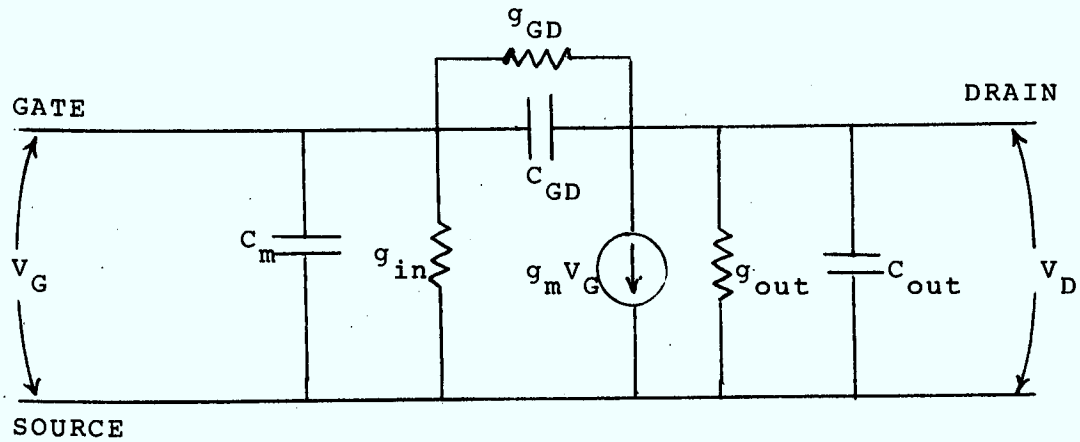


Fig. 5. Common source JFET equivalent circuit.

The maximum operating frequency is defined as the frequency at which the current through C_{in} is equal to $g_m V_G$.

$$f_m = \frac{g_m}{2\pi C_{in}} \leq \frac{\left(\frac{2W\mu q N_A a}{L}\right)}{2\pi \left(\frac{\varepsilon_s}{a} WL\right)} = \frac{q N_A \mu a^2}{\pi L^2 \varepsilon_s} \tag{29}$$

For high operating frequencies, L should be minimized. Also, for a silicon device, an N-channel is preferred since $\mu_n > \mu_p$.

The noise sources of importance in a JFET include shot noise, thermal noise and flicker noise. Shot noise from the gate leakage current is given by

$$\overline{I}^2_{sh} = 2qI_G B \tag{30}$$

where B is the bandwidth and I_G the gate current. Under reverse bias, $I_G \simeq 10^{-10}$ A, so that shot noise is small. Thermal noise in the conductive channel is

$$\overline{V}^2_{th} = \frac{4kTB}{g_m} \tag{31}$$

Increasing g_m reduces thermal noise.
Generation-recombination noise (flicker noise) due to surface effects is

$$\overline{V}^2_{th} \quad \frac{f_N}{f} \tag{32}$$

and is not important at microwave frequencies.

The above discussion is based on Shockley's gradual channel approximation. This approximation fails when the gate depletion layer width approaches zero, when the pinch-off condition is approached, and when $L/a \lesssim 10$. Two-dimensional solutions of Poisson's equation, with drift velocity saturation of carriers in the channel, have been obtained for small channel length FETs. They show that a boundary condition of the gradual channel approximation, namely that the field normal to the channel boundary is zero, can be satisfied by constructing a modified channel boundary with a reduced depletion layer width along the channel.

2.2 Metal Semiconductor Field Effect Transistors

The Shottky-barrier FET or metal-semiconductor FET was originally proposed by Mead[2]. A diagram of a MESFET using an N-type channel in GaAs is shown in Fig. 6.

The device has an interdigitated structure, fabricated by using an N-type GaAs epitaxial film about 0.15-0.35 μm thick on a semi-insulating substrate. The N channel is doped with either sulfur or tin in a doping concentration N between $8x10^{16}$ and $2x10^{17}$ cm^{-3}. The electron mobility in the channel is in the range of 3000-4500 $cm^2/v.s$. The Shottky-barrier gate is evaporated aluminum. The source and drain contacts are Au-Ge, Au-Te, or Au-Te-Ge alloys. A contact metallization pattern of gold is used to bring the source, drain, and gate contacts out to bonding pads over the semi-insulating substrate.

The I_D-V_D characteristic is similar to that for a JFET. However, there are two important differences between a JFET and a MESFET:

i) The MESFET can be made in semiconductors (such as CdS) in which p-type doping is difficult.

ii) The formation of the metal-semiconductor contact can be achieved at much lower temperatures than those required for a P-N junction. At

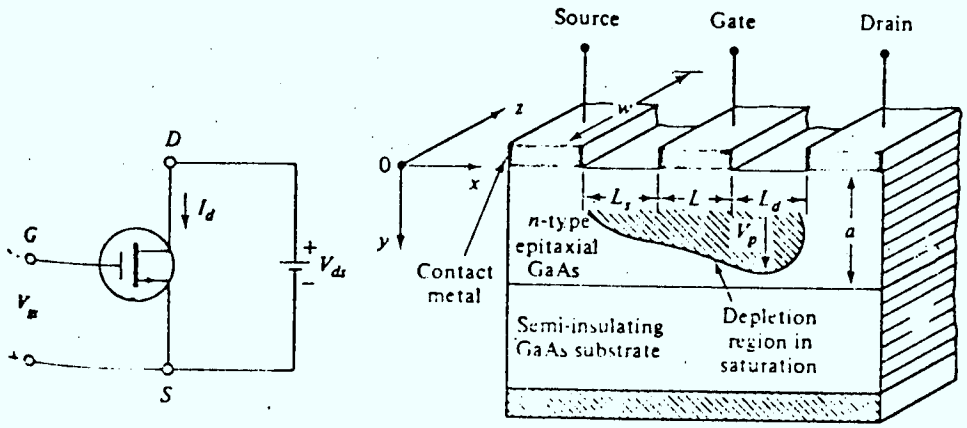


Fig. 6. Schematic diagram and symbol of a MESFET.

present, the GaAs MESFET gives the best power and noise performance among various types of GaAs transistors.

MESFET operation is similar to JFET operation. The source and drain electrodes are biased so that an electron current flows in the N-type epitaxial layer from the source, through the channel beneath the gate, to the drain electrode. The current I_D in the channel induces a voltage drop along its length with the consequence that the gate electrode becomes more reverse-biased towards its drain end. This causes a charge-depletion region to be set up in the channel to support the voltage, as required by Poisson's equation. Even with $V_G=0$, the charge-depletion region may extend at the drain end almost to pinch-off the channel against the semi-insulating substrate. At pinch-off, the current saturates and then remains almost constant for further increase of drain voltage. This process is shown in Fig. 7.

For microwave FETs which have very short channel lengths, velocity saturation of the charge carriers occurs in the channel before the minimum channel width (the order of a Debye length $\sqrt{\varepsilon_s kT/q^2N}$ since at saturation the transition between channel and depletion region is no longer abrupt) is reached. The channel width at the constriction at the onset of velocity saturation is given by a (1-u) where

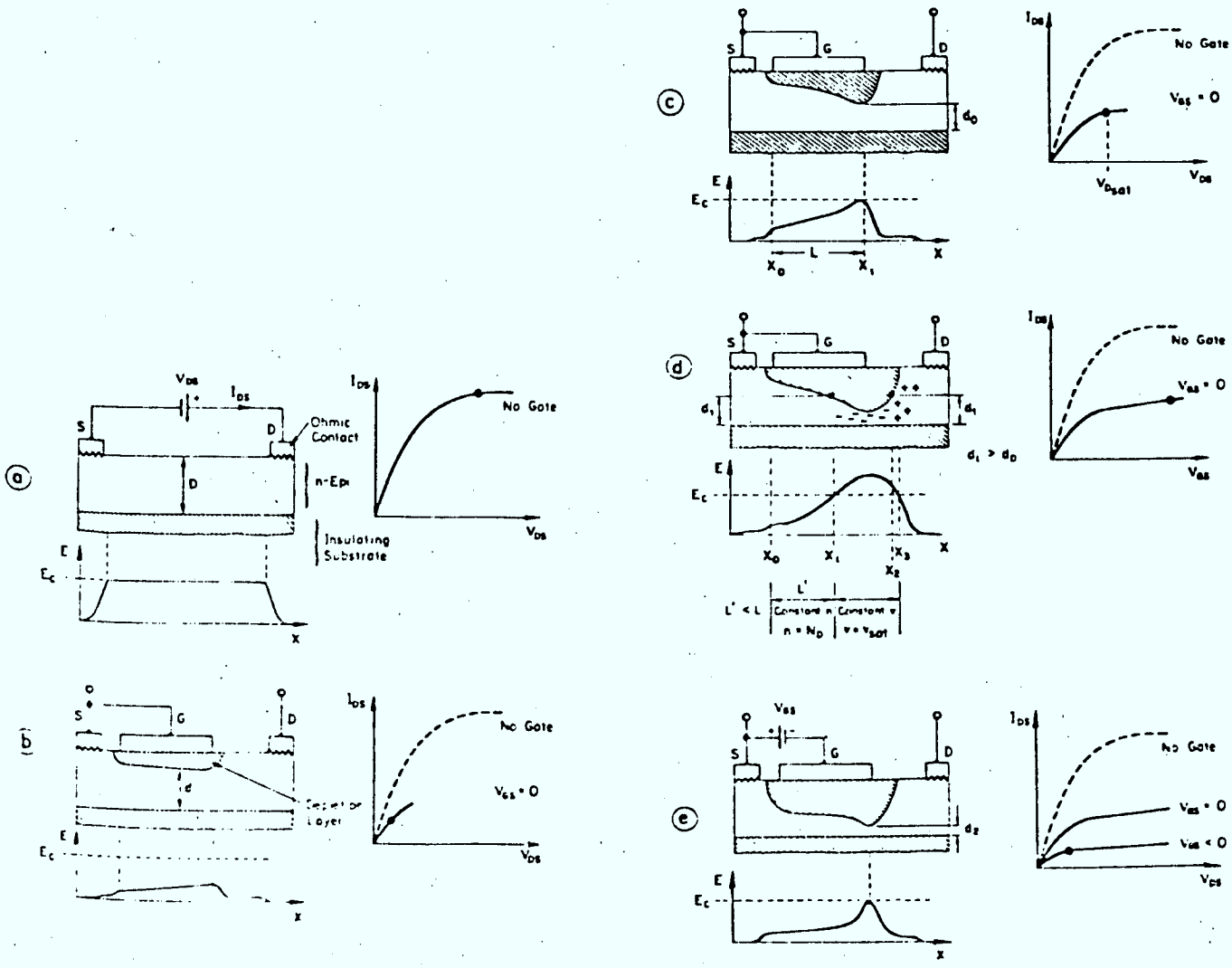


Fig. 7. a) Shows the I-V characteristic of an n-type silicon layer with two ohmic contacts. The current saturates because the electrons reach a maximum drift velocity at the critical field E_c. In (b)-(d), the current is controlled by the depletion layer under a Schottky gate, shorted to the source. In (c), the current starts to saturate at $V_{D\,sat}$ and (d) shows the formation of a stationary dipole layer in the channel for $V_{DS} > V_{D\,sat}$, [Kl] (e) Illustrates the condition for a negative gate bias. The depletion layer is wider, it constricts the conductive cross-section further, and causes the current to saturate at a lower level.

$$u^2 = \frac{V_D - V_G}{V_p} \tag{33}$$

where V_D is voltage between drain end of channel and the source, $V_G = V_{GS}$, and V_p is the voltage required to pinch-off the channel. The current flowing through the channel is then

$$I_m = I_o(1-u_m) \tag{34}$$

where subscript m refers to saturation and

$$I_o = qNv_s aW \tag{35}$$

I_o is related to I_p, the saturation current for the non-saturated velocity (Shockley) case with $V_G=0$, by

$$I_o = \frac{3I_p}{\eta} \tag{36}$$

where

$$I_p = \frac{V_p qN\mu aW}{3L} \tag{37}$$

and

$$V_p = \frac{qNa^2}{2\varepsilon_s} \tag{38}$$

Also

$$\eta = \frac{\mu_o V_p}{v_s L} = \frac{v}{v_s} \tag{39}$$

is the normalized drift velocity v with respect to the saturation drift velocity v_s. The drift velocity in the channel is

$$v = \frac{\mu_o E_x}{1 + \frac{\mu_o E_x}{v_s}} \tag{40}$$

which predicts that the velocity saturates at v_s. μ_o is the low-field mobility and E_x is the channel field magnitude. Fig. 8 shows the velocity saturation for GaAs and Si.

Combining (33) and (40) gives the reduced drain current

$$\frac{I_D}{I_p} = \frac{3(u^2-t^2)-2(u^3-t^3)}{1+\eta(u^2-t^2)} \tag{41}$$

where

$$u^2 - t^2 = \frac{V_D}{V_p} \tag{42}$$

is the reduced drain-to-source voltage and

$$t^2 = -\frac{V_g}{V_p} \tag{43}$$

is the reduced gate-to-source voltage.

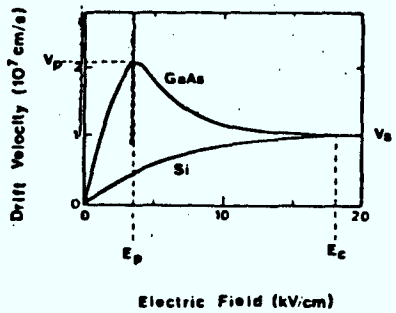


Fig. 8. Equilibrium electron drift velocity versus electric field in GaAs and silicon

Parameters of Operation

The parameters that are important in the operation of a MESFET are the transconductance, the gate capacitance, the cut-off frequency, the transit time, the drain conductance, the maximum frequency of operation, and the noise sources.

The transconductance in the saturation region is reduced from its low field value g_{m_o}:

$$g_m = g_{m_o} \frac{u_m - t}{1+\eta(u_m^2 - t^2)} \tag{44}$$

with

$$g_{m_o} = \frac{I_o \eta}{V_p} = \frac{qNa\mu W}{L} \tag{45}$$

The gate capacitance is given by

$$C_g = \frac{dQ_D}{d|V_G|} = C_{g_o} \left(\frac{\eta}{3}\right)\left(\frac{u-t}{1-u}\right)\left[\frac{4(u^3-t^3)-3(u^4-t^4)}{(1-u)(u^2-t^2+\frac{1}{\eta})} - 6t\right] \tag{46}$$

where

$$C_{g_o} = \frac{\varepsilon_s WL}{a} = \frac{qNWLa}{2V_p} \tag{47}$$

is the gate capacitance of the fully depleted channel.

The cut-off frequency of the lumped RC network of the channel and gate is

$$f_{co} = \frac{g_m}{2\pi C_g} = \frac{3}{\pi}\frac{v_s}{L\eta}(1-u)^2\left[4(u^3-t^3)-3(u^4-t^4)-6t(1-u)\right.$$

$$\left.(u^2-t^2+\frac{1}{\eta})\right]^{-1} \tag{48}$$

The transit time is

$$\tau = \frac{qNWLa-Q_D}{I_m} = \frac{L}{v_s}\frac{[1-u_m-\eta/6(u_m^4-4u_m t^3+3t^4)]}{(1-u_m)^2} \tag{49}$$

The drain conductance under saturated drift velocity conditions is reduced from its low field value.

$$g_d \simeq \frac{I_D}{E_o L}\left(\frac{1}{1+\eta(u_m^2-t^2)}\right) \tag{50}$$

where E_o is the absolute value of the channel field at the beginning of the restricted channel region.

The maximum frequency of oscillation is given by:

$$f_{max} = \frac{f_{co}}{2}\left(\frac{g_m}{g_d}\right)^{1/2} \tag{51}$$

Using (44), (48), (50) gives

$$f_{max} = \frac{f_{co}}{2}\left[\frac{\mu E_o}{v_s}\frac{(u_m-t)}{(1-u_m)}\right]^{1/2} \tag{52}$$

For $t = 0$ (i.e., $V_G = 0$) and $n>>1$, so that $f_{co} = v_s/4\pi L$ and $u_m \simeq (3/\eta)^{1/3}<<1$, (52) reduces to

$$f_{max} = \gamma\frac{v_s}{L}\left(\frac{3}{\eta}\right)^{1/3} \tag{53}$$

where $\gamma = 0.14$ for $\mu E_o/v_s = 13$, and $\gamma = 0.18$ for $\mu E_o/v_s$ in the case of GaAs.

It has been found empirically for GaAs FETs with gate length L<10 μm that

$$f_{max} \simeq \frac{33}{L} \quad \text{(GHz)} \tag{54}$$

where L is in μm.

Equation (53) shows that for small gate length transistors, where drift velocity saturation is important, $f_{max} \alpha\ 1/L$ rather than $f_{max} \alpha\ 1/L^2$ as determined from Shockley's theory. f_{max} is similar to the cut-off frequency f_τ, determined by the limiting value of the transit time $\tau = L/v_s$ from (49).

$$f_\tau = \frac{1}{2\pi\tau} = \frac{v_s}{2\pi L} \tag{55}$$

Thus, a GaAs FET has a better figure-of-merit than the Si FET for an X-band application since v_s(GaAs) = $2\text{x}10^7$ cm/s at E = 3 KV/cm, while v_s(Si) = $8\text{x}10^6$ cm/s at E = 15 KV/cm. This velocity saturation comparison is shown in Fig. 8.

A common-source equivalent circuit for a MESFET including noise sources is shown in Fig. 9. Two sources of noise in the intrinsic FET are thermal noise in the channel and induced noise in the gate. The channel noise is

represented by

$$\overline{i_{n_d}^2} = 4\ kT\ \Delta f\ g_{m_o}\ P \tag{56}$$

where g_{m_o} is the low-frequency magnitude of the transconductance, and P is a factor depending upon bias conditions.

A disturbance in channel voltage induces gate noise

$$\overline{i_{n_g}^2} = 4\ kT\ \Delta f\ \left(\frac{\omega^2 C_{sg}^2}{g_{m_o}}\right) R \tag{57}$$

where ω is the angular frequency and R is a factor depending upon bias conditions.

The two noise currents i_{n_d} and i_{n_g} have the same origin and have some correlation factor C depending on bias conditions:

$$jC = \frac{\overline{i_{n_g}^* \cdot i_{n_d}}}{\sqrt{\overline{i_{n_g}^2}} \cdot \sqrt{\overline{i_{n_d}^2}}} \tag{58}$$

Here velocity saturation has been neglected. In small gate-length FETs, the high E-field in the channel will produce carrier velocity saturation near the drain end. Using a piecewise-linear approximation for the velocity-field curve, Baechtold has calculated the effect of the increased electron temperature on the intrinsic noise sources in silicon FETs. In silicon, the electron

noise temperature is approximately T_n where

$$\frac{T_n}{T_o} = 1 + \gamma\left(\frac{E}{E_{sat}}\right)^2 \tag{59}$$

where T_o is ambient temperature, E_{sat} is saturation field (1500 KV/m) and $\gamma = 2.3$.

In GaAs,

$$\frac{T_n}{T_o} \simeq 1 + \gamma\left(\frac{E}{E_{sat}}\right)^3 \tag{60}$$

where $E_{sat} = 300$ KV/m, $\gamma = 6$.

Thus, P, R, C in (56)-(58) should be modified to account for velocity saturation. The minimum noise figure becomes

$$F_{min} = 1 + \frac{T_{n\,min}}{T_o}$$

$$= 1 + 2\sqrt{PR(1-C^2)}\,\frac{f}{f_T} + 2g_m R_i P\left(1-C\sqrt{\frac{P}{R}}\right)\left(\frac{f}{f_T}\right)^2 \tag{61}$$

where f_T (or f_{co}) is the frequency at unity current gain. Normally the MESFET is operated at $f < f_T$ so that the third term is negligible. For an optimum drain current, f_T is close to maximum while C approaches unity, thus i_{n_g} and i_{n_d} undergo considerable cancellation.

Extrinsic noise sources raise actual noise figures above intrinsic F_{min}. Gate and source resistances, R_g and

R_s, produce thermal noise

$$\overline{i^2_{n_T}} = 4kT_n \, \Delta f \, \frac{1}{R_{ext}} \qquad (62)$$

where R_{ext} represents the particular extrinsic resistance. The Miller capacitance C_{dg} and the source bonding wire inductance have a feedback effect which reduces noise figure slightly, but, as in the case of the bipolar transistor, lossy parasitic elements associated with the header will increase the noise figure.

In the MESFET, the influence of the lossy extrinsic elements is relatively stronger than in bipolar transistors, because the intrinsic noise figure of a MESFET is so low (6.6 dB at 10 GHz, 3 dB at 5 GHz).

Equivalent Circuit

At microwave frequencies, the MESFET has a very short channel length and its velocity saturation occurs in the channel before reaching the pinched path. The microwave characteristics depend not only on the intrinsic parameters such as g_m, G_D, C_{sg}, C_{dg}, R_i, etc. but also on extrinsic parameters R_g, R_s, R_d, C_{sd}, R_{gp} and C_{gp}, R_{dp}, C_{dp} (pad parasitics). Note that when extrinsic resistances are included, the highest frequency for power gain with matched input and output is:

$$f_{max} = \frac{f_{co}}{2} \left(\frac{R_d}{R_s+R_g}\right)^{1/2} \tag{63}$$

where

R_d = drain resistance

R_s = source-to-gate resistance

R_g = gate metallization resistance.

A small signal equivalent circuit given by Liechti is reproduced in Fig. 9.

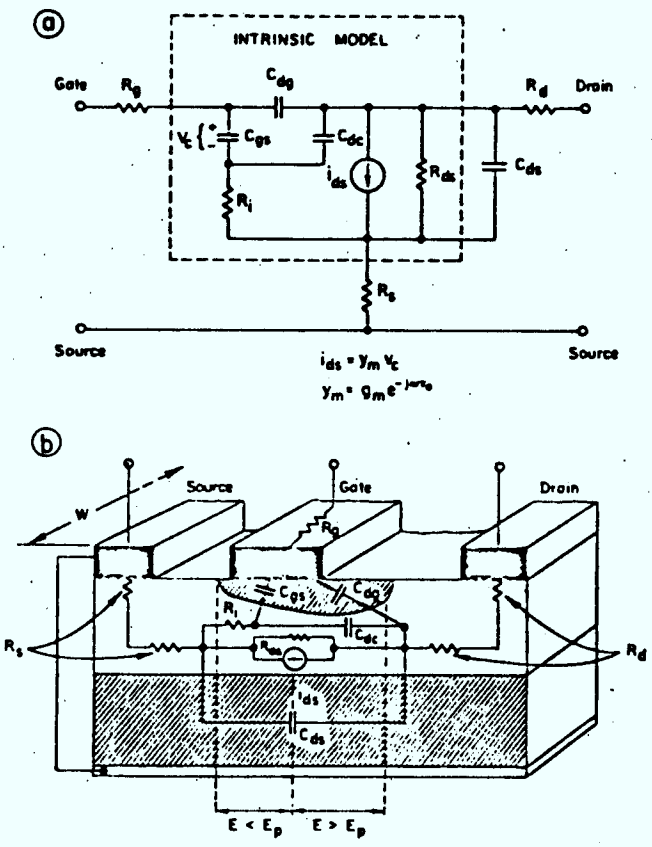


Fig. 9. (a) Is the equivalent circuit of a MESFET. Typical element values are listed in Table III.
(b) Shows the physical origin of the circuit elements.

3. MATHEMATICAL MODEL OF AM-TO-PM DISTORTION IN GaAs FETs BASED ON POWER SERIES ANALYSIS OF NON-LINEAR DEVICE ELEMENTS

3.1 Introduction

Recently there has been a strong trend towards the replacement of conventional traveling-wave-tube power amplifiers with solid-state units based on GaAs FETs. The analysis and design of these solid-state amplifiers requires that the non-linear properties of the active devices are well characterized. An analytical procedure based on either a non-linear circuit model or an analytical model of the GaAs FET can give a good understanding of these mechanisms of distortion.

Tucker [3] has investigated the gain compression and AM-to-PM conversion properties of a single stage GaAs FET amplifier based on a non-linear unilateral circuit model. This circuit model incorporates non-linearities in the gate, the transconductance and at the output or drain of the device. Gain compression is accounted for and Tucker's analysis thus applies for both small- and medium-signal input power levels. Analytical expressions are obtained which relate the gain compression to the load admittance and input power level. Normalized contours of constant gain compression and AM-to-PM conversion on the load admittance plane are given.

Fig. 10 shows the unilateral non-linear device circuit model connected as an amplifier with an input (gate) termination admittance Y_s and a load admittance Y_L. Since a unilateral model for the FET is assumed, drain-to-gate feedback elements are neglected. This enables one to consider first the input and output portions of the circuit separately and then to study their combined effect. It is assumed that there are four non-linear circuit elements: the Schottky-barrier junction capacitance at the gate C_g, the gate-voltage-dependent transconductance G_m, the output-voltage-dependent drain conductance G_d, and the output-voltage-dependent drain capacitance C_d. The first two of these, C_g and G_m, are considered part of the input circuit since the current in both of these elements depends solely on the gate voltage v_i. The elements G_d and C_d are considered part of the output circuit since the currents i_g and i_c depend solely on the output voltage v_o.

The non-linearities in the FET can be accounted for by introducing appropriate variations in the equivalent circuit elements as functions of the instantaneous drain and gate voltages. The dominant contributions to non-linear response can be expected from the variation in transconductance G_m with gate voltage and in the drain conductance G_d with drain voltage. Additional contributions arise from

the voltage dependence of the gate-source capacitance C_g. It will be assumed that the static capacitance between the source and drain is much larger than the variable contribution, so that the drain capacitance may be assumed to contribute very little IMD.

Mathematically, it is convenient to represent the variation in the equivalent circuit elements in a Taylor series around the operating point. By this approach, the transconductance can be written

$$G_m(V) = G_{m1} + G_{m2}V_g + G_{m3}V_g^2 + G_{m4}V_g^3 + \ldots \tag{64}$$

and therefore, the device RF drain current would be given by the terms

$$I_d(t) = \int_o^{V_g(t)} G_m(V)\,dV$$

$$= G_{m1}V_g + G_{m2}\frac{V_g^2}{2} + G_{m3}\frac{V_g^3}{3} + G_{m4}\frac{V_g^4}{4} \tag{65}$$

where

$$V_g = V_{gs}(t) - V_{g0} \tag{66}$$

specifies the instantaneous deviation of the gate-source voltage from the gate bias V_{g0}. In determining the expansion coefficients G_{mn}, a polynomial fit is made to the calculated variation in the transconductance from a forward gate voltage of 0.5 V to a reverse voltage corresponding to

pinch-off. The variation in G_m with drain bias is neglected by averaging this variation over the typical range of drain voltages.

A corresponding expression can be written for the drain output conductance as a function of the drain voltage

$$G_d(V) = G_{d1} + G_{d2}V_d + G_{d3}V_d^2 + G_{d4}V_d^3 + \ldots \tag{67}$$

where

$$V_d = V_{ds}(t) - V_{d0} \tag{68}$$

represents the instantaneous deviation in the drain voltage from the bias voltage V_{d0}. The dependence of G_d on gate bias has been neglected by averaging this variation over the gate voltage range. The drain voltage range used in the calculation of the expansion coefficients extends from half of the saturation voltage to twice the drain bias voltage.

An analogous expression to (64) and (67) expresses the capacitance C_g as a function of the instantaneous signal voltage $V_g(t)$:

$$C_g(V) = C_{g1} + C_{g2}V_g + C_{g3}V_g^2 + C_{g4}V_g^3 \ldots \tag{69}$$

The effective impedance of this capacitance is derived from the equation

$$I_g(t) = \frac{d}{dt}[Q_g] = \frac{d}{dt}\left[\int_o^{V_g(t)} C_g(V).dV\right] \tag{70}$$

and from this relationship the IMD current levels due to the non-linearity of $C_g(V)$ can be derived. However, the distortion due to this source is generally found to be small. Therefore, the main effect of the capacitance variation with voltage is seen to be a detuning effect with increasing signal drive level which gives rise to the sometimes observed gain expansion effects.

The expansion coefficients of $C_g(V)$ are derived from static calculations. As expected, calculations show that all the expansion coefficients such as G_{mn}, G_{dn}, and C_{gn} are dependent on the doping profile of the active layer. This dependence will be discussed in detail in the following section.

Following Tucker's analysis, we note that the gain compression of a GaAs FET is virtually independent of the admittance Y_s at the input port. The transfer function of the input circuit can thus be written in a form which lumps together the non-linear effects of C_g and G_m. It is expressed as a power series with order-dependent time delays [4]:

$$i_x(t) = \sum_{\ell=1}^{N} g_{m1} v_i^1(t-\tau_1) \tag{71}$$

where v_i and i_x are as shown in Fig. 10. g_{m1} are real expansion coefficients, and τ_1 are time delays. It has been found that for many devices N = 3 gives an adequate approximation for small- and medium-signal operation.

With the non-linearity of the input circuit characterized, it is now necessary to determine the response of the circuit to an input signal. If the transistor is driven simultaneously by two closely spaced equal-amplitude unmodulated carriers (as in the so-called two-tone test), then the input voltage v_i can be written in the form

$$v_i(t) = A(\cos \omega_1 t + \cos \omega_2 t) \tag{72}$$

where A is a real constant. This signal gives rise to components of the current i_x (Fig. 10) both at the two carrier frequencies ω_1 and ω_2 and at the two third-order IMD frequencies $(2\omega_1-\omega_2)$ and $(2\omega_2-\omega_1)$. These currents are, in complex notation,

$$I_{xc} = AH_1(1 + \frac{9}{4} A^2 H_3 H_1^{-1}) \quad \text{(carrier)} \tag{73}$$

and

$$I_{xd} = \frac{3}{4} A^3 H_3 \quad \text{(distortion)} \tag{74}$$

where

$$H_1 = g_{m1}\underline{/-\omega_1 \tau_\ell}$$

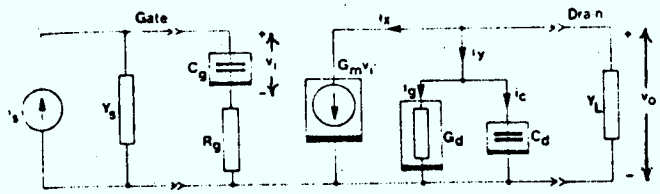


Fig. 10. Non-linear circuit model of single-stage GaAs FET amplifier.

The first term in (73) represents linear device gain while the second term accounts for gain compression or expansion in the input circuit.

The non-linear drain conductance G_d and the non-linear drain capacitance C_d are represented in power series form as follows:

$$i_g(t) = \sum_{l=1}^{N} g_l v_o^l(t) \tag{78}$$

and

$$i_c(t) = \sum_{l=1}^{N} c_l \frac{dv_o^l(t)}{dt} \tag{79}$$

where g_l and c_l are real expansion coefficients of the conductance and capacitance, respectively. As with the input circuit, distortion components of order higher than 3 are assumed to be negligible.

Components of third-order distortion due to G_d and C_d can be attributed to both the square-law and cubic terms in (78) and (79). Carrier-frequency components and third-order IMD components present in the output voltage v_o give rise to third-order distortion components in the currents i_g and i_c via the cubic terms. Similarly, second-harmonic distortion components of the carriers, at frequencies $2\omega_1$ and $2\omega_2$, in the output voltage v_o give rise to third-order distortion when combined with carrier-frequency signals and

third-order IMD signals in the square-law terms. It has been found experimentally that second-harmonic distortion components have negligible effect on third-order distortion for GaAs FETs. Thus the second-order expansion coefficients g_2 and c_2, which give rise to second-harmonic distortion, can be assumed to be zero.

In complex notation, the carrier-frequency component and the third-order IMD component of the output voltage is V_{oc} and V_{od}, respectively. The corresponding components of the current I_y (Fig. 10) are

$$I_{yc} \simeq V_{oc} J_1 (1 + \frac{9}{4} J_3 J_1^{-1} |V_{oc}|^2) \tag{80}$$

and

$$I_{yd} \simeq V_{od} J_1 + \frac{3}{4} V_{oc} |V_{oc}|^2 J_3 . (1+4V_{od} |V_{oc}|^{-1}) \tag{81}$$

where

$$J_1 = |J_1| \angle \sigma_1 = g_1 + j\omega_1 c_1 .$$

3.2 Distortion Characteristics

It is the objective to obtain an analytical expression for the AM-to-PM conversion characteristic in terms of the input power level and the load admittance. The carrier-frequency component of output voltage V_{oc} can be expressed in terms of the input voltage by applying Kirchoff's current

law at the device output. Using (73) and (80), the carrier-frequency output voltage is

$$V_{oc} = -\frac{AH_1 + \frac{9}{4} A^3 H_3}{Y + \frac{9}{4} J_3 |V_{oc}|^2} \tag{82}$$

The admittance Y is given by

$$Y = Y_L + J_1$$

where J_1 is the small-signal output admittance of the device. The deviation of device gain from its small-signal value is characterized here by the parameter k, defined as the ratio of V_{oc} as given by (82) to V_{oc} under ideal linear conditions where $H_3 = J_3 = 0$.

$$k = \frac{-YV_{oc}}{AH_1} \tag{83}$$

This parameter represents the ratio of the amplifier voltage gain at an arbitrary input signal level to the voltage gain at small-signal levels. In terms of power gain the corresponding parameter is $|k|^2$ which is referred to here as "gain compression". If $|k|^2<1$ the gain is reduced to a value below its small-signal value.

The device carrier-frequency output power (per carrier for two-tone inputs) P_{out}, written in terms of the output voltage and load conductance, is

$$P_{out} = |V_{oc}|^2 G_L \tag{84}$$

where $Y_L = G_L + jB_L$. Combining (84) and (83), the output power can be written in the form

$$P_{out} = \frac{A^2 |H_1|^2 G_L}{|Y|^2} \; |k|^2 \tag{85}$$

Now

$$P_{out} = P_{in} G_p |k|^2 \tag{86}$$

where P_{in}, which is proportional to A^2, is the input power per carrier and G_p is the small-signal (linear) power gain of the device. The small-signal power gain is a function of the load admittance and is given by

$$G_p = G_{pm} \frac{4g_1 G_L}{|Y|^2} \tag{87}$$

where G_{pm} is the maximum small-signal power gain. The gain compression $|k|^2$ is a function of both the load admittance and the input power. From (85) to (87), the maximum small-signal power gain is

$$G_{pm} = \frac{A^2 |H_1|^2}{4g_1 P_{in}} \tag{88}$$

An expression for k can be obtained from (82), (83) and (88):

$$k = \frac{1+3P_{in}B}{1+3P_{in}D|k|^2|Z|^2 Z} \tag{89}$$

where

$$B = 3g_1 G_{pm} H_3 H_1^{-1} |H_1|^{-2} = |B| \angle \omega_1(\tau_1 - \tau_3)$$

$$D = 3g_1 G_{pm} J_3 = |D| \angle \sigma_3$$

and $Z = R + jX = Y^{-1}$. Note that in (89), k is defined in terms of four independent device parameters, namely the magnitude and phase of B and D. The numerator of (89) represents the contribution of the input circuit to the value of k, and the denominator represents the corresponding contribution of the output circuit. It is thus convenient to write (89) in the form

$$k = k_i k_o \tag{90}$$

where

$$k_i = |k_i| \angle \phi_i = 1 + aP_{in}B \tag{91}$$

$$k_o = |k_o| \angle \phi_o = (1+|k_o|^2 |z_n|^2 \bar{z}_n)^{-1} \tag{92}$$

and z_n is a normalized impedance given by

$$z_n = r_n + jx_n = (3P_{in}|k_i|^2)^{1/3} |D|^{-2/3} DZ. \tag{93}$$

The gain compression due to the input circuit $|k_i|^2$ is given by (91), a linear equation in P_{in}. Similarly, the gain compression due to the output circuit $|k_o|^2$ is obtained from the non-linear (92). Since (92) is expressed in terms of the normalized impedance z_n, universal contours of constant $|k_o|^2$ can be obtained on the z_n plane and then applied to any transistor.

Some contours of constant $|k_o|^2$ are shown on the normalized impedance plane z_n in Fig. 11. Only the upper half of the z_n plane is shown since the contours are symmetrical about the r_n axis. Values of $|k_o|^2$ ranging from +3 dB (a gain increase or expansion) to -3 dB (a gain reduction) are given. As would be expected from (92), there is 0 dB of gain compression at the point $z_n = 0$ which corresponds either to a short circuit at the drain terminals of the device or to zero P_{in}. Also shown in Fig. 11 is the direction of the R and X axes of the unnormalized impedance plane Z. This illustrates that there is an axis rotation associated with the magnitude normalization in (93).

The instantaneous AM-to-PM conversion α of an amplifier is defined as the rate of change of phase angle of the carrier-frequency output signal with changing input power level in decibels. For an FET with known terminations, the AM-to-PM conversion is found by differentiating the phase angle of k with respect to the logarithm of P_{in}:

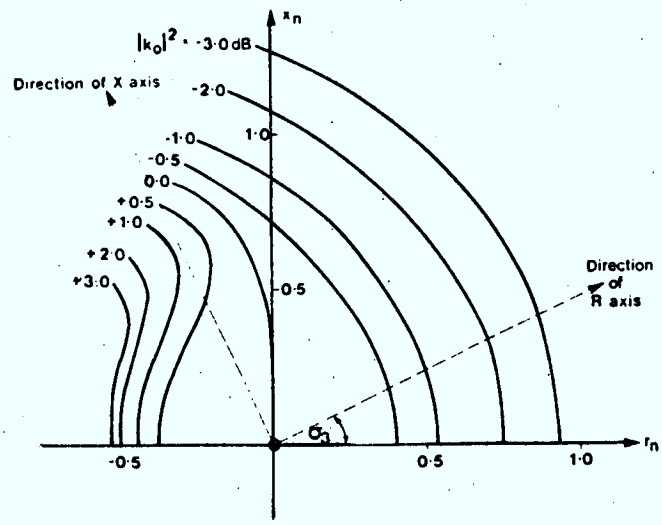


Fig. 11. Contours of constant output circuit gain compression $|k_o|^2$ on the normalized impedance plane z_n.

$$\alpha = \frac{180}{\pi} \cdot \frac{d\phi}{d(10 \log P_{in})} \quad \text{(degrees/dB)} \tag{94}$$

where

$$\phi = \phi_i + \phi_o$$

$$\phi_i = \tan^{-1} \left(\frac{3wP_{in}}{1+3vP_{in}}\right)$$

$$\phi_o = -\sin^{-1} \left(|k_o|^3 |z_n|^2 x_n\right)$$

and

$$B = v + jw.$$

From (90) to (94) it can be shown that

$$\alpha = |k_i|^{-2}\{13 \cdot 19 \cdot 3wP_{in} + Q(|k_i|^2 + 6vP_{in})\} \quad \text{(degrees/dB)} \tag{95}$$

where

$$Q = \frac{13 \cdot 19 |k_o|^2 |z_n|^2 x_n}{1+4r_n |k_o|^2 |z_n|^2 + 3|k_o|^4 |z_n|^4} \quad \text{(degrees/dB)}.$$

The first term in (95) represents the contribution of the input circuit to the overall AM-to-PM conversion while the second term represents the contribution of the output circuit. If the device parameter B is zero and thus there is no distortion in the input circuit, then the AM-to-PM conversion is equal to Q. Fig. 12 shows some contours

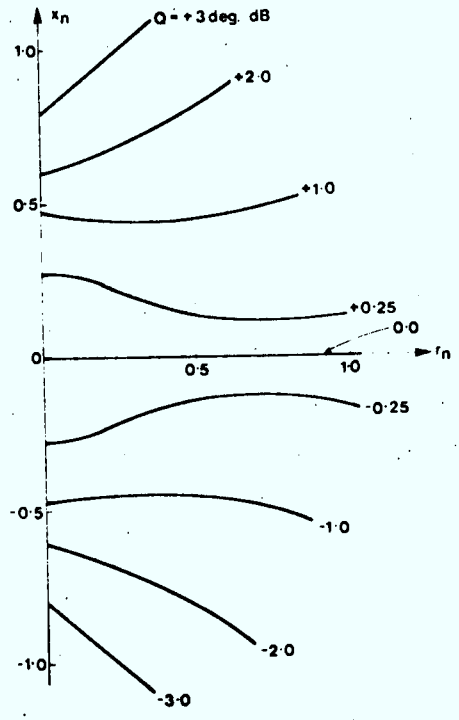


Fig. 12. Contours of constant Q on the normalized impedance plane z_n.

of constant Q on the right-half normalized impedance plane z_n. As with the contours of Fig. 11, these curves can be applied to any transistor. It is interesting to note that the values of Q given in Fig. 12 are quite small, indicating that the output circuit has good AM-to-PM conversion characteristics.

4. MATHEMATICAL MODEL OF THIRD-ORDER INTERMODULATION IN GaAs FETs BASED ON VOLTERRA SERIES ANALYSIS OF NON-LINEAR DEVICE ELEMENTS

4.1 Introduction

The microwave GaAs MESFET has found application in wideband highly linear power amplifiers. In communication systems, the device intermodulation characteristic becomes an important consideration. This section discusses distortion in MESFET amplifiers using the simple non-linear device model which is analyzed by a systematic procedure based on the Volterra series or non-linear transfer function approach. The Volterra series expansion allows a detailed representation of device characteristics including reactive effects.

The intermodulation analysis of MESFETs employing simple power series descriptions rely on an a priori knowledge of the controlling voltage. As well, it ignores the effects of out-of-band terminations. A more general analysis procedure is based on the Volterra series, and includes interactions between the non-linear parameters and spectral components at intermodulation and harmonic frequencies. The transistor model used, incorporating gate, transconductance, and drain non-linearities, is again that shown in Fig. 10 of the previous section. The feedback capacitance C_D is absent and thus the circuit is unilateral.

This model can represent MESFET behaviour up to frequencies at which feedback becomes significant. For example, for a 1 μm gate length MESFET, the model is adequate to around 8 GHz.

The transconductance non-linearity and the gate-to-source channel capacitance are again represented by the power series expressions given by Eqns. (64) and (69) of the preceding section. Following Minasian's analysis [5], the drain output conductance is represented by

$$G_d(V) = \frac{a}{(V+b)^c} \tag{96}$$

where a, b and c are experimentally determined constants. Since the distortion predictions are sensitive to the Taylor series coefficients used in analysis, good measurement accuracy is required for their determination (particularly the dominant third degree G_{m3} transconductance coefficient). The non-linearities are represented as voltage-controlled current generators using a three-term Taylor series expansion of the characteristic about the operating point

$$i_d = \sum_{n=1}^{3} G_{mn} v_g^n \tag{97}$$

$$i_g = \frac{d}{dt} \sum_{n=1}^{3} C_{gn} v_g^n \tag{98}$$

$$i_o = \sum_{n=1}^{3} G_{dn} v_{ds}^n \tag{99}$$

where i and v are difference quantities and G_{mn}, C_{gn} and G_{dn} are the Taylor series coefficients. The first-order terms in (97) to (99) represent the linear part of the model, whereas the higher order terms in the equations represent the non-linearities.

4.2 Volterra Analysis

It is assumed that the non-linear circuit under consideration can be represented by a Volterra series expansion. This enables the output $v_o(t)$ to be expressed in terms of the input signal $v_s(t)$ by the functional series

$$\begin{aligned} v_o(t) = & \int_{-\infty}^{\infty} h_1(\tau) v_s(t-\tau) d\tau \\ & + \int_{-\infty}^{\infty}\int h_2(\tau_1,\tau_2) v_s(t-\tau_1) v_s(t-\tau_2) d\tau_1 d\tau_2 \\ & + \int_{-\infty}^{\infty}\iint h_3(\tau_1,\tau_2,\tau_3) v_s(t-\tau_1) v_s(t-\tau_2) \\ & \cdot v_s(t-\tau_3) d\tau_1 d\tau_2 d\tau_3 + \ldots \end{aligned} \tag{100}$$

where $h_n(\tau_1,\ldots,\tau_n)$ is the nth-order Volterra kernel, whose Fourier transform $H_n(\omega_1,\ldots,\omega_n)$ is the corresponding nth-order non-linear transfer function in the frequency domain. In the case of low-distortion amplifiers, the non-linearities

of interest are mild, and hence only the first three terms of the Volterra series are used to characterize the transistor.

The transfer functions (denoted $H_{nc}(\omega_1,\ldots,\omega_n)$) relating the controlling capacitor voltage v_g to the input v_s are obtained as an intermediate step in deriving the transfer functions (denoted $H_n(\omega_1,\ldots,\omega_n)$) relating output v_o to the input v_s. The first-order transfer functions are found to be

$$H_{1C}(\omega) = \frac{Y_S(\omega)}{Y_i(\omega)} \tag{101}$$

$$H_1(\omega) = -\frac{G_{m1}Y_S(\omega)}{Y_0(\omega)Y_i(\omega)} \tag{102}$$

where

$$Y_S(\omega) = \frac{1}{Z_G(\omega)+\overline{R}_i} \tag{103}$$

$$Y_i(\omega) = j\omega C_{g1} + Y_S(\omega) \tag{104}$$

$$Y_0(\omega) = G_{d1} + j\omega C_{ds} + \frac{1}{Z_L(\omega)} \tag{105}$$

Equations (101) and (102) express the linear response of the circuit in the frequency domain.

The second-order transfer functions are obtained in terms of the first-order functions

$$H_{2C}(\omega_1,\omega_2) = \frac{-j\omega' C_{g2} H_{1C}(\omega_1) H_{1C}(\omega_2)}{Y_i(\omega')} \tag{106}$$

$$H_2(\omega_1,\omega_2) = \frac{-H_{1C}(\omega_1) H_{1C}(\omega_2)}{Y_0(\omega')} \cdot \left[\frac{-G_{m1} j\omega' C_{g2}}{Y_i(\omega')} + G_{m2} + \frac{G_{d2} G_{m1}^2}{Y_0(\omega_1) Y_0(\omega_2)} \right] \tag{107}$$

where

$$\omega' = \omega_1 + \omega_2. \tag{108}$$

The third-order transfer function for the output is

$$\begin{aligned} H_3(\omega_1,\omega_2,\omega_3) = -\frac{1}{Y_0(\omega'')} \Bigg[& H_{1C}(\omega_1) H_{1C}(\omega_2) H_{1C}(\omega_3) \\ & \cdot \left(G_{m3} - \frac{G_{m1} j\omega'' C_{g3}}{Y_i(\omega'')} - \frac{G_{d3} G_{m1}^3}{Y_0(\omega_1) Y_0(\omega_2) Y_0(\omega_3)} \right) \\ & + \overline{H_{1C}(\omega_1) H_{2C}(\omega_2,\omega_3)} \left(- \frac{G_{m1}^2 j\omega'' C_{g2}}{Y_i(\omega'')} + 2G_{m2} \right) \\ & + 2G_{d2} \, \overline{H_1(\omega_1) H_2(\omega_2,\omega_3)} \Bigg] \end{aligned} \tag{109}$$

where

$$\omega'' = \omega_1 + \omega_2 + \omega_3 \tag{110}$$

and the overbar indicates symmetrization.

Intermodulation is defined for the case of two equal amplitude sinusoid signals at frequencies ω_1 and ω_2 applied to the MESFET input:

$$v_s(t) = V_s\cos\omega_1 t + V_s\cos\omega_2 t. \tag{111}$$

The in-band third-order intermodulation products are generated at frequencies $2\omega_1-\omega_2$ and $2\omega_2-\omega_1$.

In terms of the non-linear transfer functions, the first-order output in Z_L at the fundamental frequency ω_1 is

$$v_{o1} = V_S|H_1(\omega_1)|\cos[\omega_1 t+\underline{/H_1(\omega_1)}]. \tag{112}$$

The third-order intermodulation output at frequency $2\omega_1 -\omega_2$ is

$$v_{o3} = \frac{3}{4}\,V_S^3|H_3(\omega_1,\omega_1,-\omega_2)|$$

$$.\cos[(2\omega_1-\omega_2)t+\underline{/H_3(\omega_1,\omega_1,-\omega_2)}]. \tag{113}$$

The output power delivered to the load impedance by each spectral component may readily be evaluated from (112) and (113). Third-order intermodulation distortion (IM_3) is defined as the ratio of the distortion output power at $2\omega_1-\omega_2$ to the fundamental frequency or desired signal power at ω_1 in the load. The case corresponding to practical two-tone measurements of amplifiers occurs when the frequency separation between the two exciting input

signals is very small:

$$\omega_1 \simeq \omega_2 \triangleq \omega. \tag{114}$$

Then the load seen by the fundamental signal and the distortion product $(2\omega_1-\omega_2\simeq\omega)$ is virtually the same, and IM_3 may be expressed in terms of the amplifier transfer functions:

$$IM_3 = 20 \log\left[\frac{3}{4} V_S^2 \frac{|H_3(\omega_1,\omega_1,-\omega_2)|}{|H_1(\omega_1)|}\right]. \tag{115}$$

Using (101)-(110) in (115), an expression for IM_3 in terms of the MESFET amplifier model parameters can be derived:

$$\begin{aligned} IM_3 = 20 \log \frac{3}{4} V_S^2 \Bigg| & \left|\frac{Y_S(\omega)}{Y_i(\omega)}\right|^2 \Bigg| \frac{G_{m3}}{G_{m1}} \\ & + \frac{1}{Y_0(\omega)}\left(\frac{\frac{4}{3} G_{02}G_{m2}}{Y_0(\omega_d)} + \frac{G_{m1}^2}{|Y_0(\omega)|^2}\left(\frac{\frac{4}{3} G_{02}^2}{Y_0(\omega_d)} - G_{03}\right)\right) \\ & + \frac{2j\omega C_{g2}}{3Y_i(2\omega)}\left(\frac{2j\omega C_{g2}}{Y_i(\omega)} - \frac{2G_{m2}}{G_{m1}}\right) - \frac{j\omega C_{g3}}{Y_i(\omega)} \\ & + \frac{\frac{2}{3} G_{02}}{Y_0(2\omega)Y_0^*(\omega)}\left(\frac{-G_{m1}2j\omega C_{g2}}{Y_i(2\omega)} + G_{m2} + \frac{G_{02}G_{m1}^2}{Y_0(\omega)}\right)\Bigg| \end{aligned} \tag{116}$$

where $\omega_d = \omega_1 - \omega_2$, the asterisk denotes the complex conjugate, and it is assumed that $Z_G(\omega_d)$ is small as is often required for stability at low frequencies. Equation (116) reveals the influence of device parameters, frequency, and

termination admittances on distortion. Intermodulation distortion depends on admittances at the harmonic and difference frequencies as well as at the fundamental frequency. The expression contains terms arising from the third degree non-linear parameters (G_{m3},C_{g3},G_{03}) as well as second degree coefficients (G_{m2},C_{g2},G_{02}) which give rise to third-order terms by the interaction of the first and second degree kernels. In practice some simplification of the expression is possible because not all terms are significant.

5. RELATIONSHIP BETWEEN NON-LINEAR DISTORTION AND DEVICE DOPING PROFILE

5.1 Introduction

The GaAs field effect transistor has emerged as a highly attractive device in power amplifier applications through the demonstration of multiwatt power output levels at X-band frequencies with power added efficiencies of up to 40 percent. Initial explorations in the design of linear amplifiers have shown that the behaviour of non-linear distortion, characterized by third-order intermodulation distortion (IMD) levels, is complex and needs to be better understood by both the device designer and the systems user. The use of "graded" profiles has been suggested to improve the IMD behaviour of GaAs FETs. This section of the report is aimed at providing a more complete understanding of the sources of IMD in a FET by outlining their relationship to various doping profiles, and showing how a better profile with respect to improved linearity may be obtained by ion implantation.

Practically all modelling of GaAs FETs has used analytical expressions derived from Shockley's [1] early work. Pucel et al [6] have included the effects of velocity saturation in a more detailed model. These analyses of FET devices depended upon the assumption of a flat profile

of a certain thickness for purposes of simplification. More recently, Higgins and Kuvas [7] have introduced a model shown in Fig. 13 which can be used to calculate the effects of non-flat profiles. The model deals with an arbitrary profile by dividing up the thickness of the active layer into 150 laminar layers. The passage of an electron under the gate region is modelled as having a region of saturated velocity following the initial short section where velocity is proportional to electric field. By observing the necessary boundary conditions in the directions along and normal to the charge flow, current may be established as a function of bias conditions. Also, gate capacitance, transconductance, and output conductance may be calculated over any range of bias conditions.

The most common technique for preparing GaAs FET active layers is vapour phase epitaxy. This approach offers some flexibility with regard to carrier profile tailoring, and has provided power FETs with outstanding performance. Therefore, it is of interest to compare the theoretical performance of FETs with implanted and epitaxial profiles in order to assess whether the intrinsic advantage of ion implantation in terms of reproducibility and uniformity can be effectively utilized in the fabrication of highly linear transistors.

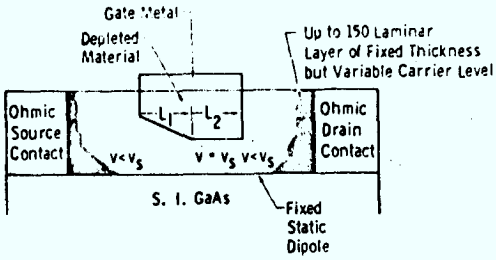


Fig. 13. Laminar layer models used in the modelling of GaAs FETs with variable carrier concentration profiles.

This type of assessment has been carried out by Higgins and Kuvas by comparing the relative merits of an idealized epitaxial profile with abrupt doping transition toward the substrate, and the profile resulting from a 500 keV Se implant compensated at the surface by a shallow 40 keV Be implant. These profiles are shown in Fig. 14 along with a conventional flat doping profile, which will serve as a reference for the predicted performance. The devices modelled in this comparison were 500 μm wide gates of 1 μm length with ohmic contact resistance of 10^{-6} $\Omega \cdot cm^2$.

The modified Pucel model was used to derive the G_m and G_d polynomial coefficients (discussed in the previous section) for these three profiles. The results are shown in Tables I and II reproduced from Higgins et al. It is important to note that the magnitude of the fifth-order (G_{m5},G_{d5}) coefficients is least for the ion implanted profile, promising a better high power IMD performance. It is also noteworthy that in the case of transconductance, G_{m5} is larger than the G_{m3} coefficient for the flat profile device.

The corresponding coefficients for the gate-source capacitance are given in Table III for the flat profile and the Se + Be implant. A comparison of the magnitude of the coefficients for each order reflects the smaller variation in this capacitance versus gate voltage for the

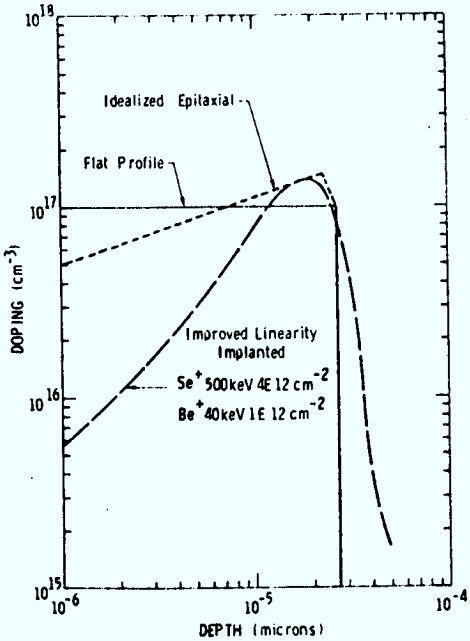


Fig. 14. Implanted and epitaxial doping profiles used in calculating the IMD characteristics of GaAs FET.

TABLE I
Polynomial Coefficients of Transconductance
$$G_m(V) = G_{m1} + G_{m2}V + G_{m3}V^2 + G_{m4}V^3 \cdots$$

Coefficient	Flat Epi	Improved Epi	Ion Implant Se + Be
G_{m1}	0.035	0.0355	0.031
G_{m2}	0.0058	0.004	0.0033
G_{m3}	-0.00045	-0.0007	0.00075
G_{m4}	0.00033	0.00058	-0.000054
G_{m5}	0.00146	0.0009	-0.0002
G_{m6}	0.00005	-0.0001	0.000042
G_{m7}	-0.0002	-0.0001	0.0000448
G_{m8}	0.000005	0.000015	-0.0000047

TABLE II
Polynomial Coefficients of Drain Output Conductance
$$V_{ds} = 10\text{ V},\ G_d(V) = G_{d1} + G_{d2}V + G_{d3}V^2 + G_{d4}V^3 \cdots$$

Coefficient	Flat Epi	Improved Epi	Ion Implant Se + Be
G_{d1}	1.64 E-4	1.99 E-4	4.28 E-4
G_{d2}	-3.07 E-5	-4.19 E-5	-4.90 E-5
G_{d3}	1.13 E-5	5.18 E-6	-2.23 E-6
G_{d4}	-1.37 E-6	7.62 E-7	1.23 E-7
G_{d5}	-3.92 E-7	-2.25 E-7	1.26 E-7
G_{d6}	5.80 E-8	-8.47 E-9	-2.02 E-8
G_{d7}	6.20 E-9	5.40 E-9	2.00 E-9
G_{d8}	-8.21 E-10	-3.41 E-10	-1.08 E-10

TABLE III

POLYNOMIAL COEFFICIENTS OF GATE-SOURCE CAPACITANCE (pF/V^{n-1}), $C_g(V) = C_{g1} + C_{g2}V_g + C_{g3}V^2 + C_{g4}V^3 \ldots$

Coefficient	Flat Epi	Ion Implant Se + Be
C_{g1}	0.2786	0.2482
C_{g2}	0.0467	0.0245
C_{g3}	0.0127	0.00029
C_{g4}	0.0004	0.0005
C_{g5}	-0.0016	-0.0003
C_{g6}	0.0025	0.0002
C_{g7}	0.0012	-0.00004
C_{g8}	-0.0008	-0.00004

TABLE IV

THE (ω_1) CURRENT COMPONENTS RESULTING FROM DRIVING A NONLINEAR CONDUCTANCE $G(V) = G_1 + G_2V + G_3V^3 \ldots +$ WITH A TWO-TONE SIGNAL $V = A\cos\omega_1 t + B\cos\omega_2 t$, $I(t) = \int G(V)dV$

Source	Component	Component if B = 0
G_1V	A	A
G_3V^3 +3	$0.25A^3 + 0.5A^2B$	$0.25A^3$
G_5V^5 +3	$0.125A^5 + 0.75A^3B^2 + 0.375AB^4$	$0.125A^5$

TABLE V

THE CURRENTS CONTRIBUTED AT REPRESENTATIVE INTERMODULATION FREQUENCIES BY THE INDIVIDUAL COEFFICIENTS OF A NONLINEAR CONDUCTANCE $G(V)$ WHERE $I(t) = \int G(V)dV$ AND $G(V) = G_1 + G_2V + G_3V^2 \ldots$ AND $V = A\cos(\omega_1 t) + B\cos(\omega_2 t)$[a]

IMD Source	Two-Tone Test Intermodulation Products $2f_1 - f_2$	$3f_1 - 2f_2$	$4f_1 - 3f_2$
G_3V^3 +3	$0.25A^2B$		
G_5V^5 +5	$0.25A^4B$	$0.125A^3B^2$	
	$0.375A^2B^3$		
G_7V^7 +7	$0.234A^6B$	$0.234A^5B^2$	$0.078A^4B^3$
	$0.937A^4B^3$	$0.3127A^3B^4$	
	$0.468A^2B^5$		

[a] For $G_3 \neq 0$ $G_5 \neq 0$ $G_{n>5} = 0$; $I_{2f_1-f_2} = 0.25G_3A^2B + 0.25G_5A^4B + 0.375G_5A^2B^3$

implanted profile, which should reduce gain expansion effects and minimize IMD contributions from this source.

5.2 Distortion Characteristics

The calculated IMD is based on the normal two-tone method. The calculations assume that the IMD voltage levels are less than 10 percent of the voltage levels of the two carrier tones. This assumption restricts the range of validity of the calculated IMD levels to less than 20 dB below the carrier level. This ratio is quite adequate for analyzing linear amplifiers, as the operating point of interest will be below this level. In the calculations of signal power gain and IMD products, fixed values of R_s, the source resistance, and R_L, the load resistance, are used. The value of load resistance R_L is chosen to be 180 Ω rather than the value $(G_{d1})^{-1}$. This choice reflects the usual condition for large-signal tuning, where the output impedance of the device is much lower than $(G_{d1})^{-1}$. The load impedance presented at the harmonic bands, i.e., frequencies much higher than the signal frequency is assumed to be very small.

Only third-order products have been calculated. Normally, it is assumed that the third-order coefficient (G_{m3}, for instance) is much larger than the fifth- or seventh-order coefficients (G_{m5},G_{m7}); from Table I, this

is evidently not so. From Tables I and II, it is seen that the fifth-order coefficients can contribute substantially to third-order IMD products. In fact, for moderate to high signal levels, the transconductance contributes mainly from its fifth-order (G_{m5}) term. This analysis contributes new insight into IMD generation by taking into account the contribution to third-order IMD products from the higher order terms in the non-linear devices.

Tables IV and V give some indication of a) how the various coefficients contribute to the manner in which the conductance (and susceptance) levels change as the signal power level rises (Table IV), and b) how each expansion term adds to the various intermodulation frequencies (Table V). The implications of Table IV are that optimum tuning and loading conditions change with signal power level because of the corresponding changes in the admittance matrix of the device. The IMD levels would be sensitive to tuning and loading conditions because they depend upon the peak RF voltage levels. Therefore, different IMD performance may be expected for large-signal and small-signal tuning conditions.

Another point to be made is that in the low power signal region, the IMD contributions of the drain conductance $G_d(V)$ dominate; and, as the signal level rises, the $G_m(V)$ contributions to IMD products become larger.

The intermodulation produced from the non-linearities due to the transconductance and the output conductance can be expected to be partially correlated as a result of the coherence between the gate and the drain voltages. This correlation is commonly observed as a cancellation effect giving dips in the IMD versus input level curve. These dips are accompanied by a change in the rate of rise of the IMD power levels for a given increment of change in input power levels.

The sign of the fifth-order relative to the third-order coefficient may be quite important in determining the IMD products at low signal levels. In fact, cancellation effects can occur in third-order IMD products from the drain (or gate) alone due to sign differences of the different coefficients.

The equivalent circuit in Fig. 10 has been used in conjunction with the information in Tables I to V to calculate the gain and the third-order intermodulation products of the three profiles shown in Fig. 14. The results are given in Fig. 15.

It is observed that the gains at small-signal levels of both epitaxial devices are about 1 dB greater than the corresponding gain of the implanted transistor. This calculation is for a 10 GHz test where the signals (two tones) are separated by only a few MHz. The gain of the

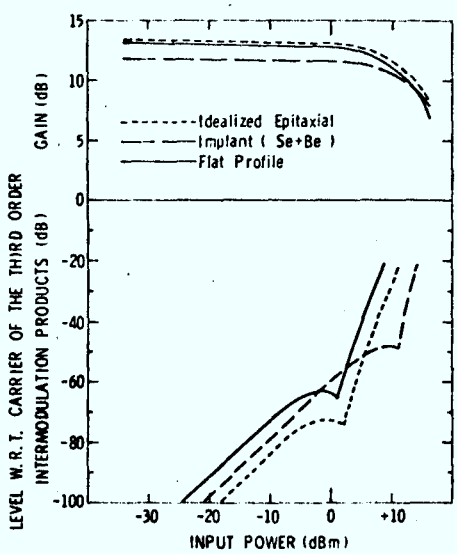


Fig. 15. Calculated gain and third-order intermodulation products versus input power using the active layer profiles in Fig. 3.

ion implanted device is less but saturates more slowly at high output power levels.

The IMD products display three distinct regions. At very low signal level, the third-order products rise 3 dB for a 1 dB increase in input signal level. This behaviour is explained by the dominant role of G_{d3} in this range. Then comes the intermediate signal level region where the cancellation effects are generally seen. In the large-signal region, the contribution from the transconductance G_m generally dominates, and the rate of rise of the IMD product is greater than 3 dB/1 dB increase in signal level. This strong increase is caused by the large contribution from G_{m5} which may equal or exceed G_{m3}. Ion implantation shows a considerable advantage in this area because G_{m5} is much less than G_{m3} for this case and is lower than equivalent values for the other profiles; this results in the IMD product for the ion implanted FET continuing to have lower level with respect to the carrier up to higher input power levels. Thus the transition from drain side (G_d) dominated IMD to gate side dominated (G_m) IMD is postponed longer in the case of the implanted FET.

At the point of saturation, where output power is considerable and gain is falling to low values, the IMD products of the simple model become inaccurate. The rate of rise of the IMD products should fall off rather than

continuing to rise at the higher rate shown in Fig. 15. The reason for this discrepancy is that the assumption of small IMD voltages becomes inaccurate and the total power is rapidly diverted into an increasing number of unwanted IMD products other than just the third-order products. This area of the IMD versus input power level is tractable to computation using methods that a) account for many frequencies of non-negligible voltage level, b) account for the total power distribution, and c) extend the power series representation for the equivalent circuit elements to a sufficiently high order.

The dependence of IMD on drain bias level has been calculated by Higgins and Kuvas for the implanted profile in Fig. 14. The principal result of a rise of drain bias level is a general reduction of the G_d coefficients. The resulting IMD products have been calculated and are shown in Fig. 16. The IMD products rise with a drop in drain bias, both at the low input power end and at moderately high input power levels. A more pronounced cancellation notch is noted at the lower drain bias because of the increase in the IMD products due to G_d.

5.3 Fabrication

The primary goal is to realize highly linear GaAs power FETs by tailoring of ion implanted profiles to

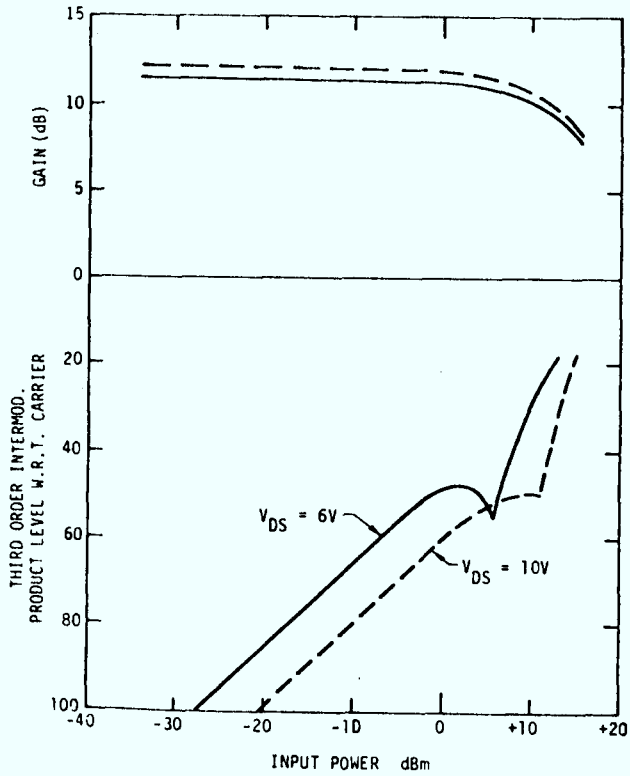


Fig. 16. Calculated gain and third-order intermodulation product versus input power for implanted GaAs FET with the drain voltage as a parameter.

optimize the doping in the active layer. Therefore, major effort in the area of device fabrication has been devoted to preparation of suitable active layers. The profiles obtained by Higgins and Kuvas are presented in this section.

The theoretical calculation indicated that surface compensated Se + Be implants can provide suitable layers for power FETs with good linearity. Active layers for experimental measurement of IMD were made from both single dose Se implants and from the combined Se + Be implant scheme of Section 2.

A typical profile resulting from such a combined implant schedule is illustrated in Fig. 17. The compensation level in this implant is relatively modest in that the doses of the Se and Be implants were 4×10^{12} and 0.5×10^{12} cm^{-2}, respectively. The pinch-off voltage of such layers is generally about 6 V and from calculations, one may deduce a carrier concentration at the surface region of about 3×10^{16} cm^{-3}.

The unit transistor cell consists of 6 gate fingers which are nominally 1 μm long and 150 μm wide for a total cell periphery of 900 μm. The source-to-drain spacing is 5 μm. The device geometry including the 1 μm long gate structures are defined by conventional photolithographic techniques. The mask set also contains diagnostic patterns for determining ohmic contact resistance, gate metal resistance, doping profile, and mobility.

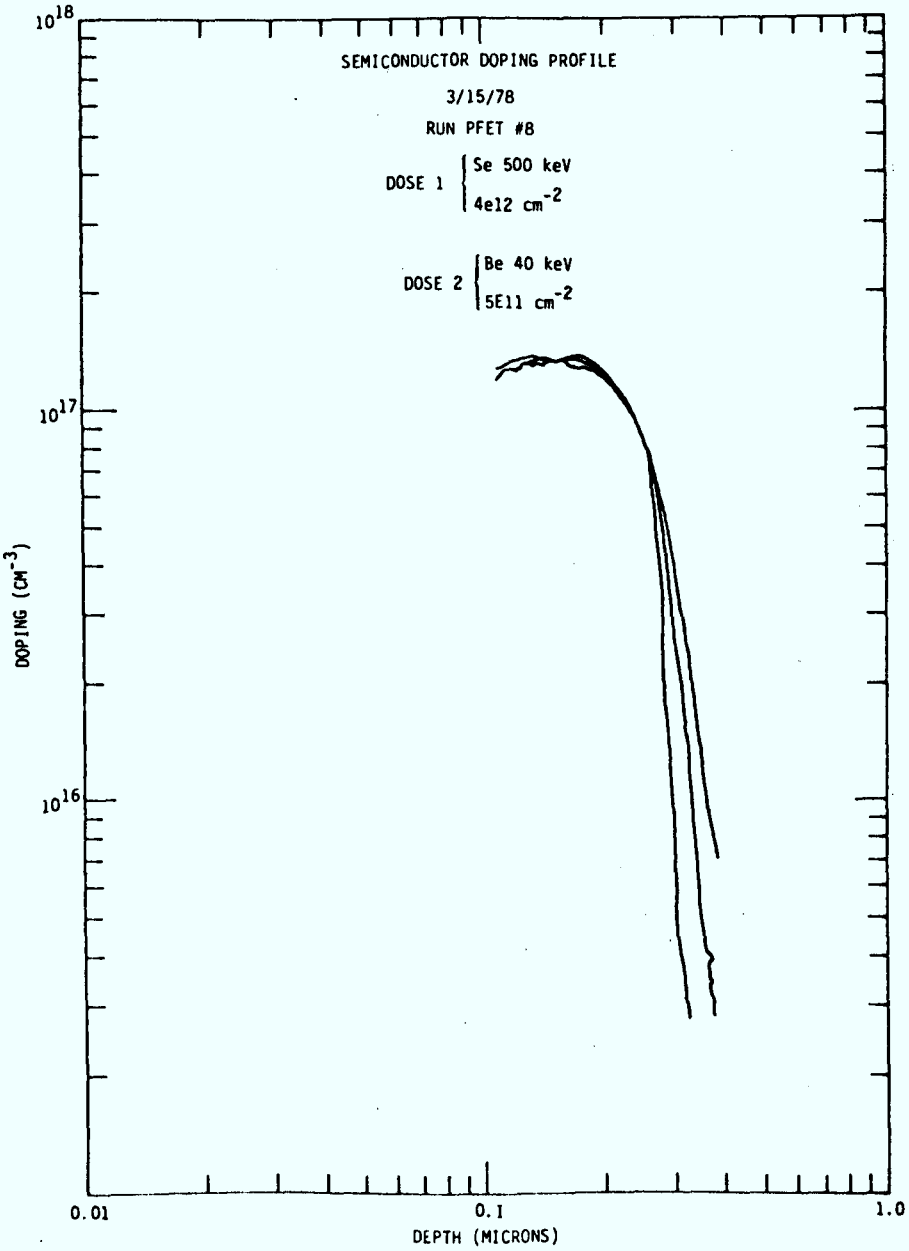


Fig. 17. Measured doping profile for combined 500 keV Se ($4x10^{12}$ cm^{-2}) and 40 keV Be ($0.5x10^{12}$ cm^{-2}) implants.

PART II

MICROWAVE BIPOLAR TRANSISTORS

6. MICROWAVE BIPOLAR TRANSISTORS

6.1 Introduction

The principles of operation of a microwave bipolar transistor are similar to those for the low-frequency device, but requirements for dimensions, process control, heat sinking, and packaging are much more severe. All microwave transistors are now planar in form and almost all are of the Silicon NPN type. Planar technology (oxide masking, dopant diffusion into epitaxial layer) is capable of one μm width definition. The preference for Silicon (even though GaAs has superior properties for microwave perforamcne such as higher electron mobility, lower dielectric constant) is due to Silicon having a natural stable passive oxide, SiO_2, which can be used as the diffusion mask, and very fine patterns can be etched in this oxide. Also, the diffusion properties of As, P, B in Si are controllable to within limits of 0.1 μm in depth as required for narrow base width (and short base transit time). GaAs, however, is not so amenable to planar processing technology and p-type dopants diffuse appreciably faster than n-type ones, so that it is difficult to make an NPN GaAs transistor with a very narrow base width required for microwave operation.

The operation of transistors at high frequencies depends on shrinking physical dimensions (mainly base width to reduce transit time, and junction areas to reduce capacitance), good processing control, and control of wafer and package parasitics.

Early [8] derived a figure-of-merit for a high-frequency transistor as

$$\text{Power Gain (Bandwidth)} = \frac{1}{4\pi\sqrt{r_b C_c \tau_{ec}}} \tag{1}$$

where

r_b = base resistance

C_c = collector capacitance

τ_{ec} = emitter-to-collector signal delay time.

Consequently, for microwave operation, r_b, C_c, and τ_{ec} must be minimized.

All microwave transistors are now planar in form and almost all are of the silicon NPN type. Transistor models can be device, measurement, or circuit-oriented. For example, h-, Y-, or S-parameters can be used for analysis.

We also know that the dc current gains are:

$$\alpha_o = h_{FB} = \frac{\Delta I_C}{\Delta I_E} \tag{2}$$

and

$$\beta_o = h_{FE} = \frac{\Delta I_C}{\Delta I_B} = \frac{\alpha_o}{1-\alpha_o} \tag{3}$$

Similarly, for small signals, we can define small-signal current gains:

$$\alpha = h_{fb} \triangleq \left.\frac{\partial I_C}{\partial I_E}\right|_{\Delta V_{CB}=0} = \alpha_o + I_E \frac{\partial \alpha_o}{\partial I_E} \tag{4}$$

$$\beta = h_{fe} \triangleq \left.\frac{\partial I_C}{\partial I_B}\right|_{\Delta V_{CE}=0} = \beta_o + I_B \frac{\partial \beta_o}{\partial I_B} = \frac{\alpha}{1-\alpha} \tag{5}$$

Then the following equivalent circuits shown in Fig. 1 can be used for analysis. In this figure, we have

r_e = emitter resistance

$r_b = r_o \frac{S}{L}$ where S is the emitter strip width and L is the emitter strip length (6)

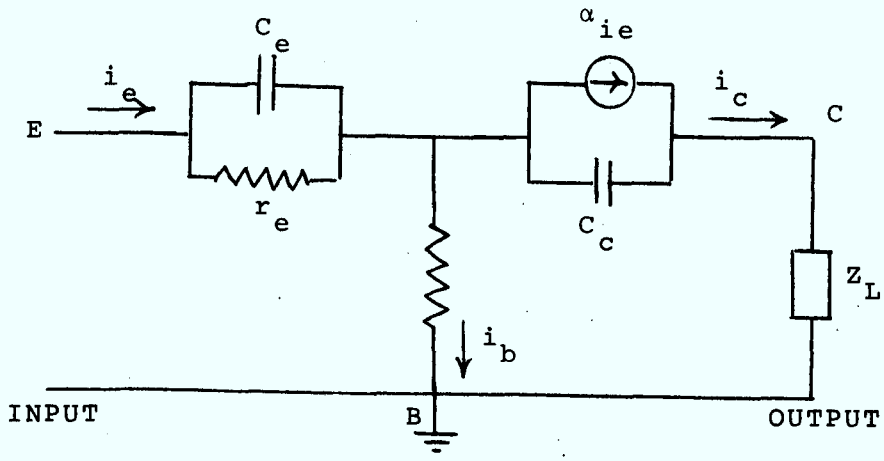
$r_o \simeq \rho_B/W_B$ where ρ_B is the average resistivity of base layer and W_B is the base width (7)

C_e = emitter depletion-layer capacitance

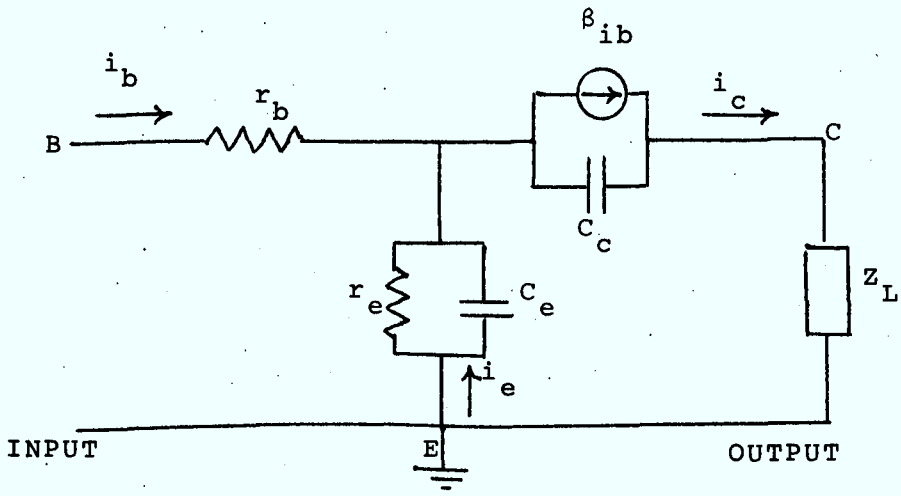
$C_c = C_D SL$ is the collector depletion-layer capacitance where C_D is the collector capacitance per unit area. (8)

The above expressions for r_b and C_c are for the stripe-base geometry shown in Fig. 2.

Important transistor parameters are cut-off frequency, gain, and noise figure and are discussed below.

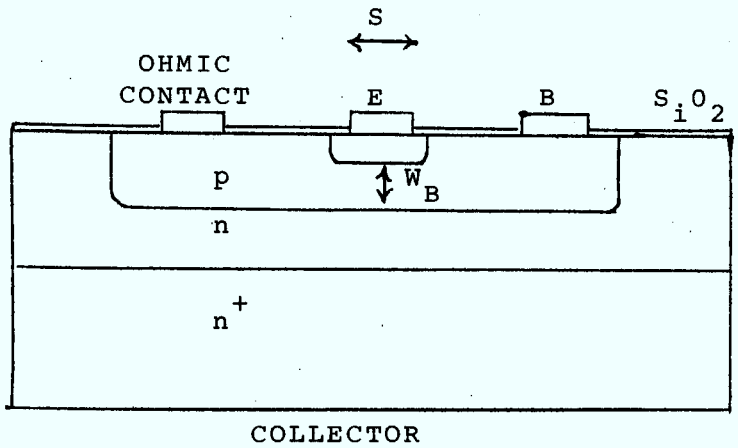


(a)



(b)

Fig. 1. Bipolar transistor equivalent circuits.
a) Common Base
b) Common Emitter



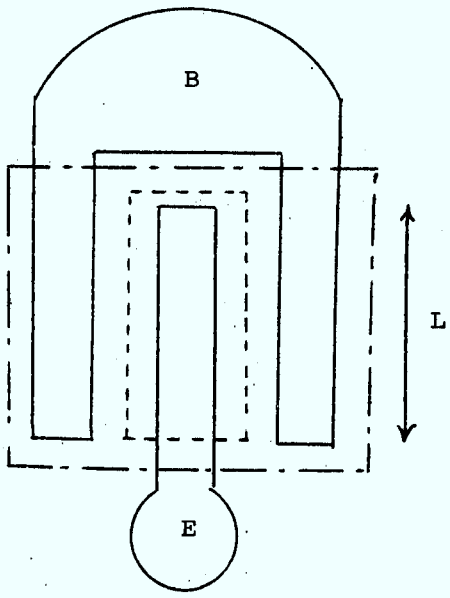


Fig. 2. Strip-base geometry bipolar structure.

Cut-off Frequency

The charge-carrier transit-time cut-off frequency f_T is the frequency at which the CE short-circuit current gain h_{fe} is unity. The cut-off frequency is related to the emitter-collector delay time τ_{ec}; the average time for an charge carrier moving at average velocity v to traverse the emitter-collector distance.

$$f_T = \frac{1}{2\pi \tau_{ec}} \tag{9}$$

Four principal regions of delay or attenuation are encountered. Consider the one-dimensional view of a NPN transistor shown in Fig. 3. Here, the emitter-base junction depletion layer capacitance C_e shunts the active emitter region. A charging time τ_E is required to charge C_e. Also carriers cross base region W through drift and diffusion. A base transit time τ_B is required during which the signal is attenuated. Carriers next cross the collector depletion layer, X_{dc} wide, under influence of an electric field (no diffusion), during the collector depletion layer transit time τ_x. Finally, the depletion collector depletion-layer capacitance C_c and resistance r_c between the collector depletion layer and external circuit terminal, produce a final $\tau_c = r_c C_c$ charge-time delay. Thus,

$$\tau_{ec} = \tau_E + \tau_B + \tau_x + \tau_c \tag{10}$$

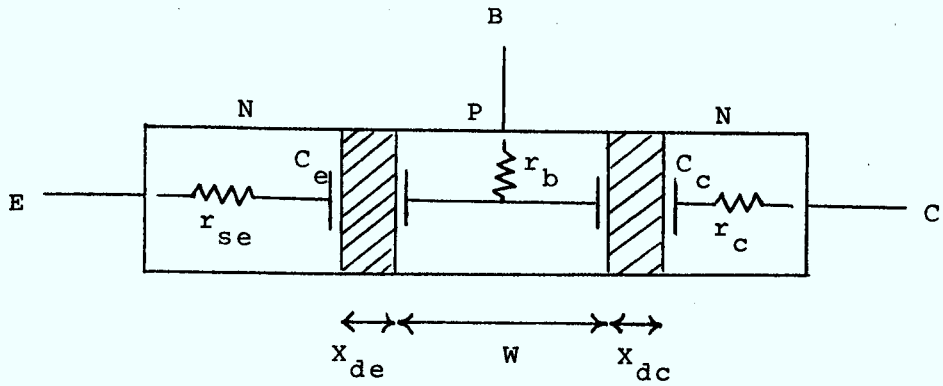


Fig. 3. One-dimensional NPN transistor.

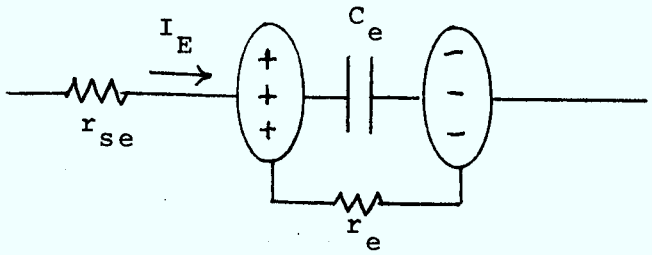


Fig. 4. Emitter base junction.

These four charging times and transit times will be discussed in some detail as follows.

In forming the emitter depletion layer charging time τ_e, the terminal emitter current divides between C_e and the space-charge resistance r_e. Hole current flows through r_e, is injected into the base and is amplified. The current through C_e is majority carriers in base and emitter. The collector capacitance C_c must also be charged as well as any other parasitic capacitance C_p connected to the base. Hence,

$$\tau_E = (r_{se}+r_e)(C_e+C_c+C_p) \tag{11}$$

Since in forward bias

$$I_E = I_{ES}(e^{qV_{EB}/kT}-1) \simeq I_{ES}\, e^{qV_{EB}/kT}$$

$$\frac{1}{r_e} = \frac{\partial I_E}{\partial V_{EB}} = I_E(q/kT) \text{ or } r_e \simeq \frac{kT}{qI_E} \tag{12}$$

The series resistance r_{se} is the emitter and emitter metallization resistance and typically is 0.5 to 1.0 Ω.

The second charging time which determines the cut-off frequency is the base layer charging time, τ_B. If the charge carriers cross the base by diffusion alone (no field), the carriers undergo a phase shift as well as attenuation as they cross the base region. Thus, the

transport factor α_T becomes complex:

$$\alpha_T = \text{sech}\ [(\frac{W}{L_B})^2 + j\omega t_B]^{1/2} \tag{13}$$

where t_B, the transit time of electrons across the P-type base, is given by:

$$t_B = \frac{W^2}{D_B} \tag{14}$$

The base cut-off frequency ω_B is the frequency $\omega_B = 1/\tau_B$ at which $|\alpha_T|$ is $1/\sqrt{2}$ below its low-frequency value. From (13), this occurs when

$$\omega_B t_B = 2.43$$

Thus

$$\tau_B = \frac{L_B^2}{D_B} = \frac{1}{\omega_B} = \frac{W^2}{2.43 D_B} \tag{15}$$

If a built-in drift field, E_{bi}, is also present in the base, due to a concentration gradient of doping through the base region,

$$E_{bi} = -\frac{kT}{q}\frac{1}{N(X)}\frac{dN(X)}{dx} \tag{16}$$

then the transit time of electrons through the base will be reduced since they will be accelerated by E_{bi}. This effect can be taken into account by modifying (15) to

$$\tau_B = \frac{W^2}{\eta D_B} \tag{17}$$

For an exponential doping N(X), η is given by

$$\eta = \frac{m^2}{m-1 + \exp^{-m}} \tag{18}$$

for m not too small where

$$m = \frac{EW}{kT/q} \approx \ln \left(\frac{N_{BE}}{N_{BC}}\right) \tag{19}$$

where N_{BE} and N_{BC} are the base impurity concentrations near the emitter and collector, respectively. For typical silicon diffused base NPN microwave transistors, $4 \leq m \leq 7$. In this range,

$$\tau_B \simeq \frac{(m-1)W^2}{m^2 D_B} \tag{20}$$

within 2%. Cooke [9] approximated η as (1.6+0.92m) so that

$$\tau_B \simeq \frac{W^2}{(1.6+0.92m)D_B} \tag{21}$$

within 1%.

Because of non-zero charge density n(W) at the collector edge of the base junction and base modulation (Kirk effect), the fluctuation of this stored charge in response to the input signal introduces an additional time delay. Thus

$$\tau_B = \frac{W^2}{\eta D_B} + \frac{1}{v(W)} \int_0^W \frac{N(W)}{N(X)}\, dx \tag{22}$$

where N(X) is the base doping, and v(W) is the electron velocity at x=W.

The collector depletion-layer transit time, τ_x, is the third term which determines the cut-off frequency. Because the collector-base junction is normally in reverse bias, there is a well-defined depletion layer $X_{dc} = X_c - W$ wide. Although the junction is formed by diffusion, it has been found that the step-junction equations still apply. Hence

$$X_d \simeq \left[\frac{2\varepsilon_r \varepsilon_o (V+\phi)}{qN}\right]^{1/2} \tag{23}$$

For example, a typical silicon microwave transistor built on 5 Ω-cm (i.e. $N = 1\text{x}10^{15}$ cm^{-3}) epitaxial material, and operated at 10 V will have $X_d \sim 3.6$ μm. Hence, the average E field on the depletion layer is

$$E_{AV} = 10.7/(3.6\text{x}10^{-6}) = 3\text{x}10^6 \text{ V/m}.$$

The electrons will then travel at the scattering limited velocity $v_{SL} \simeq 8\text{x}10^4$ m/s over most of the depletion layer since velocity saturates at about 10^6 V/m.

The transit time across the depletion layer is then

$$\tau_m = \frac{X_{dc}}{v_{SL}} \tag{24}$$

The electrons cross the base region at $v \ll v_{SL}$ so that τ_m is insignificant at low frequencies. However, at microwave frequencies, $X_d \gg W$, so that for state-of-the-art devices

$$\tau_m \simeq \tau_B \tag{25}$$

However, (25) does not include the effect of displacement current across C_e. A depletion layer transport factor can be defined as

$$\beta_m = \frac{\text{current leaving depletion layer}}{\text{current entering depletion layer}} \tag{26}$$

Since collector multiplication is assumed to be zero, the sinusoidal response is

$$\beta_m \simeq 1 - j\omega \frac{\tau_m}{2} \tag{27}$$

The signal delay time corresponding to carrier transit time τ_m is then $\tau_m/2$. Thus, the collector depletion-layer delay time is

$$\tau_x = \frac{\tau_m}{2} = \frac{X_{dc}}{2V_{SL}} = \frac{X_c - W}{2V_{SL}} \tag{28}$$

A microwave transistor is generally designed so that the collector space charge layer fully depletes the epitaxial collector region of width W_{epi} (therefore, $X_{dc} \simeq W_{epi}$).

The final charging time which determines the cut-off frequency of a bipolar microwave transistor is the collector RC charging time, τ_c. When transistors were first

constructed, the substrate had uniform high resistivity (in the order of 1-10 Ω-cm) in order to obtain reasonable breakdown voltages. Since the collector depletion layer did not extend across the whole substrate, there was a large high-resistivity region, modelled by r_c, through which the collector capacitance must be charged. Thus

$$\frac{I_{c\ out}}{I_{c\ in}} = \frac{1}{1+j\omega\ \tau_c} \tag{29}$$

where

$$\tau_c = r_c C_c \tag{30}$$

Present-day microwave transistors have a thin high-resistivity epitaxial layer on a lower resistivity substrate, greatly reducing r_c, and thus τ_c can be neglected usually.

Combining the expressions for the four charging and transit times yields the following expressions for the cut-off frequency of a bipolar microwave transistor.

$$f_T = \frac{1}{2\pi\ \tau_{ec}} = \{2\pi[\frac{kT(C_e+C_c+C_p)}{q\ I_c} + \frac{W^2}{\eta D_B} + \frac{X_{dc}}{2V_{SL}} + r_c C_c]\}^{-1} \tag{31}$$

To increase f_T, the transistor should have very narrow base thickness W, a narrow collector region, and should be operated at high-current level. As collector

width decreases, however, breakdown voltage decreases. Hence compromises must be made for high-frequency and high-voltage operation. Also, at large I_c, high-level injection produces the Kirk effect or widening of base width. This increases τ_{ec} as shown in Equation (22) and reduces f_T.

Parasitics

The predominant parasitics in microwave bipolar transistors are the base-resistance, the emitter-base junction capacitance, the collector-base junction capacitance, bonding pad capacitances and package parasitics. Each of these is discussed in detail below.

Base resistance is the distributed "spreading" resistance r_b between the internal base b' and the base contact b. It consists of four parts:

1) base resistance under the emitter

$$r_{b1} = \rho_1 S_1 / 12 W_1 \ell \tag{32}$$

2) base resistance between the emitter and p^+ region

$$r_{b2} = \rho_2 S_2 / 2 W_2 \ell \tag{33}$$

3) base resistance in the p^+ region itself

$$r_{b3} = \rho_3 S_3 / 12 W_3 \ell \tag{34}$$

4) contact resistance between metal and p^+ region R_c.

In the above equations, S_1, S_2, S_3 are widths of the emitter, emitter-to-p^+ region, and p^+ opening, respectively and ρ_1, ρ_2, ρ_3 are resistivities of the emitter, emitter-to-p^+ region, and p^+ opening, respectively. The total base resistance for n emitters is then

$$r_b = \frac{1}{n} (r_{b1}+r_{b2}+r_{b3}+R_c) \tag{35}$$

The effects of base resistance are:

1) loss of signal in base lead
2) de-biasing of the region under the emitter. Because of the low resistivity of the emitter region, the emitter side of the EB junction is essentially at the same potential. However, base current flowing through the spreading resistance r_b produces a voltage drop which is a function of distance.

Thus, those parts of the emitter junction furthest from the base contact will be operating at a lower bias than the parts nearest to the base contact. This acts as negative feedback since an increase in signal level tends to turn the transistor off.

3) increased noise figure.
4) other low-frequency effects.

The second predominant parasitic is the emitter-base junction capacitance C_E. Since the emitter-base junction is normally under forward bias, we have

$$C_E \simeq A_E \left[\frac{\varepsilon_r \varepsilon_o \, qN_{BE}}{2(V+\phi)} \right]^{1/2} \tag{36}$$

where $\simeq$ 0.7 V for silicon. Usually N_{BE} sets r_b and h_{FE} so only the emitter area A_E can be reduced to reduce C_E.

There is also a diffusion capacitance C_{DE} when an ac voltage is applied to a forward-biased EB junction. To modulate the stored base charge if no drift field is present, we have

$$C_{DE} = \frac{\Delta q_B}{\Delta V_{BE}} = \frac{qI_E}{kT} \left(\frac{W^2}{2D_B}\right) \tag{37}$$

If a drift field is present then (for η not too small),

$$C_{DE} = \frac{qI_E}{kT} \left(\frac{W^2}{2D_B}\right) \left(\frac{\eta - 1 + e^{-\eta}}{\frac{1}{2}\eta^2}\right) \tag{38}$$

The collector-base junction capacitance, C_c, consists of three capacitances:

1) C_1 under the emitter, between base and collector
2) C_2 between the base and collector in the outer region of the base
3) C_3 between the p^+ base and collector.

The capacitance C_1 is most important to C_c.

$$C_c = A_c \left[\frac{\varepsilon_r \varepsilon_o \; q \; N_C}{2(V+\phi)} \right]^{1/2} \tag{39}$$

with

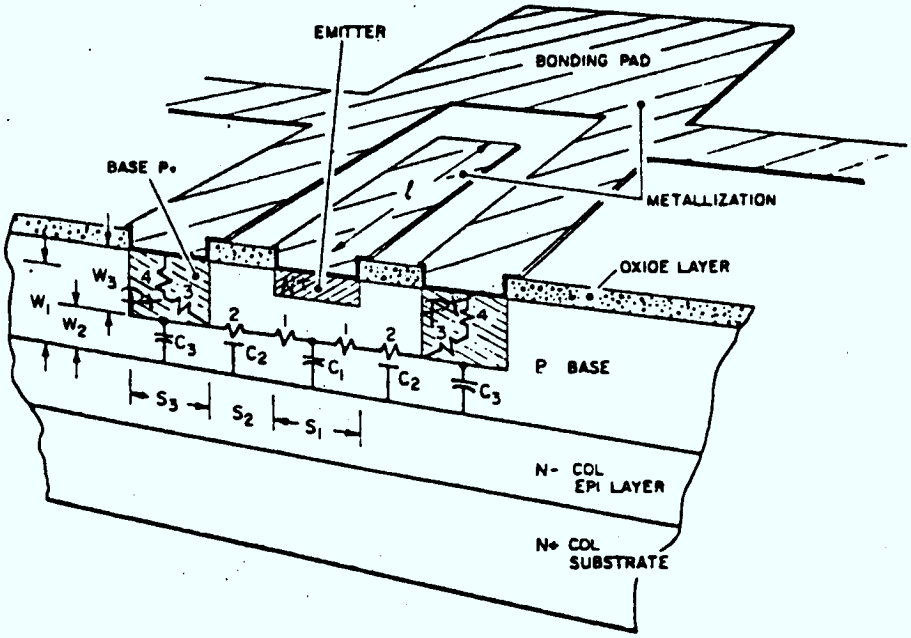
$$\rho_{collector} = \frac{1}{q\mu N_C} \tag{40}$$

The diffusion capacitance is not important here. Since $N_{BC} > N_C$, the depletion layer lies almost entirely in the collector epitaxial region and the base width does not change significantly with V_{CB} except near saturation conditions.

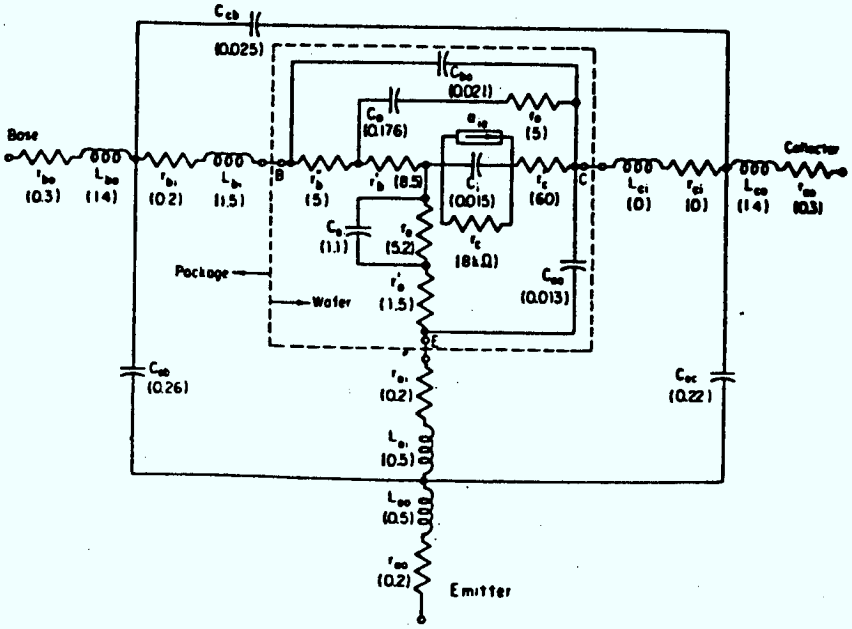
Bonding-pad parasitics consist predominantly of capacitances. On the CE configuration, there are bonding pad capacitances C_{bo} and C_{eo} which depend on pad area, and vary inversely with SiO_2 layer thickness. Pad capacitance is typically .05-0.1 pF/mil^2.

Finally, package parasitics consist of lead inductances, lead resistance, and inter-terminal capacitances.

An equivalent circuit for a microwave bipolar transistor is shown in Fig. 5. Here, the collector barrier capacitance $C_c = C_i + C_o$ where C_i is that part of C_c that occurs directly under the emitter. C_{ec} and C_{bc} are inter-terminal capacitances. L_{bo}, L_{eo}, L_{co} are external lead inductances. r_{bo}, r_{eo}, r_{co} are external lead resistances.



(a)



(b)

Fig. 5. a) Cross-section of a planar transistor showing r_b' and capacitances. 1, 2 and 3 are parts of r_b' and 4 is the contact resistance.

b) Equivalent network for transistor shown in Fig. 5a. Resistances are in ohms; inductances in nanohenries; and capacitance in picofarads.

L_{bi}, L_{ei}, L_{ci} are internal lead inductances. r_{bi}, r_{ei}, r_{ci} are internal lead resistances. The remaining parameters are due to the active wafer itself.

Power Gain-Frequency Limitations

Pritchard [10] originated the following figure-of-merit for microwave bipolar transistors:

$$\sqrt{G}\ B = \sqrt{\frac{\alpha_o\ f_T}{8\pi\ r_b\ C_c}} \tag{41}$$

where G is the power gain, B is the bandwidth, f_T is the frequency where $B = h_{fe} = 1$, r_b is the base resistance and C_c is the collector capacitance. Early modified this using $f_T = 1/2\pi\ \tau_{ec}$ so that

$$\sqrt{G}\ B = \frac{1}{4\pi\sqrt{r_b C_c \tau_{ec}}} \tag{42}$$

Another figure-of-merit is the unilateral power gain U. The unilateral gain is independent of header reactances and common lead feedback elements, and is therefore a unique measure of the intrinsic device performance. The U represents the forward power gain in a feedback amplifier when the reverse power gain has been set to zero by adjustment of a lossless reciprocal feedback network around the transistor. U can be written as

$$U \simeq \left(\frac{f_{max}}{f}\right)^2 \tag{43}$$

where f_{max} is the frequency where the power gain = 1. From (41), then the maximum frequency becomes

$$f_{max} = \sqrt{\frac{\alpha_o f_T}{8\pi r_b C_c}} \tag{44}$$

Thus (43) and (44) give

$$U = \frac{\alpha_o f_T}{8\pi f^2 r_b C_c} \tag{45}$$

Assuming that the transistor emitter and base electrode pattern is defined in terms of electrode width and spacing S, and length L, and using r_o and C_o as the base resistance and collector capacitance per unit area, respectively, then

$$r_b = r_o \frac{S}{L} \quad \text{and} \quad C_c = C_o SL \tag{46}$$

so that (44) and (45) become

$$f_{max} = \frac{1}{S}\sqrt{\frac{\alpha_o f_T}{8\pi r_o C_o}} \tag{47}$$

$$U = \frac{\alpha_o f_T}{8\pi f^2 S^2 r_o C_o} \tag{48}$$

Thus U decreases with frequency at a rate of 6 dB/octave. Both f_{max} and U are increased by decreasing the product $r_b C_c$ or equivalently, the emitter strip width S.

Equations (44) to (48) ignore the effects of package parasitics. If the capacitances of the emitter and base

bonding pads (C_{eo} and C_{bo}) become significant with respect to C_c, and the real part of the impedance due to the inner emitter and base-lead inductances (L_{ei} and L_{bi}) become significant with respect to r_b, the equivalent circuit should be analyzed numerically to determine f_{max} and U. For the values given in the equivalent circuit, we have:

Table I

	f_{max} (GHz)	f_T (GHz)	U at 1.5 GHz (dB)
Approximation	20.7	8.7	22.8
Network Analysis	10.0	6.0	16.5
Measured Value	-	5.6	15.0

Noise Figure

The noise figure of a transistor is defined as the total mean square noise voltage at output of transistor divided by the mean square noise voltage at output resulting from thermal noise source resistance R_g.

$$NF = \frac{S/N|_{input}}{S/N|_{output}} \tag{49}$$

The two main noise sources in microwave bipolar transistors are thermal noise and shot noise. Flicker noise, due to the variation of leakage current and surface recombination velocity with surface properties, is proportional to 1/F and is not important in microwave transistors.

Thermal, 'white', or Johnson noise is caused by thermal agitation of the current carriers in the bulk material of the transistor giving them random motion. The mean square noise voltage due to the base resistance r_b in bandwidth B is

$$\overline{e}_n^2 = 4kT\ r_b\ B \tag{50}$$

Thermal noise consists of the summation of short random current pulses and is uniform with frequency.

Shot noise arises since the transistor is not in thermal equilibrium under bias conditions and additional noise arises from the flow of electron and hole currents within the device. Minority charge transport across the base is mainly by diffusion. Hence there will be many fluctuations due to collisions with the lattice crystal and majority carriers. In addition, there is random generation and recombination of minority charge carriers in the base. Shot noise power from base current (partition noise) is given by

$$P_n = \frac{2eI_bB}{4g} \tag{51}$$

where g = equivalent conductance. Hence at the output

$$P_n = \frac{2e(1-\alpha_o)I_cB}{4Re\{Y_{out}\}} \tag{52}$$

Although noise arising from randomness in current conduction across the emitter and collector junctions is

a shot noise, it has been shown to be equivalent to a thermal noise source of resistance $r_e/2$. Combining these results, the noise figure for CE or CB configuration is

$$NF = 1 + \frac{r_b}{R_g} + \frac{r_e}{2R_g} + \frac{(r_b+r_e+R_g)^2}{2\alpha_o r_e R_g}\left[\left(\frac{1-\alpha_o}{\alpha_o}\right)+\left(\frac{f}{f_\alpha}\right)^2 + \frac{I_{CBO}}{I_E}\right] \quad (52)$$

where f_α is the frequency where $\alpha = \alpha_o/\sqrt{2}$ {i.e., $\alpha=\alpha_o/(1+j(f/f_\alpha)]$}.

For $F<<f_\alpha$, NF is constant determined by r_b, r_e, $(1-\alpha_o)$ and R_g. At microwave frequencies, to have low noise r_b and r_e must be minimized while f_α should be high. At lower frequencies, it is important to have high $\beta_o = \alpha_o/1-\alpha_o$.

For minimum NF, there is also an optimum R_g given by $d(NF)/dR_g = 0$. The result is

$$R_{g(opt)} \simeq (r_b+r_e)^2 + \sqrt{\frac{\alpha_o r_e(2r_b+r_e)}{(\frac{f}{f_\alpha})^2 + (\frac{1-\alpha_o}{\alpha_o})}} \quad (53)$$

and typically 10-20 Ω or so. Noise figure performance with frequency is shown in Fig. 6. Conditions for minimum NF are not necessarily the same as the conditions for maximum power gain and some compromise is usually necessary.

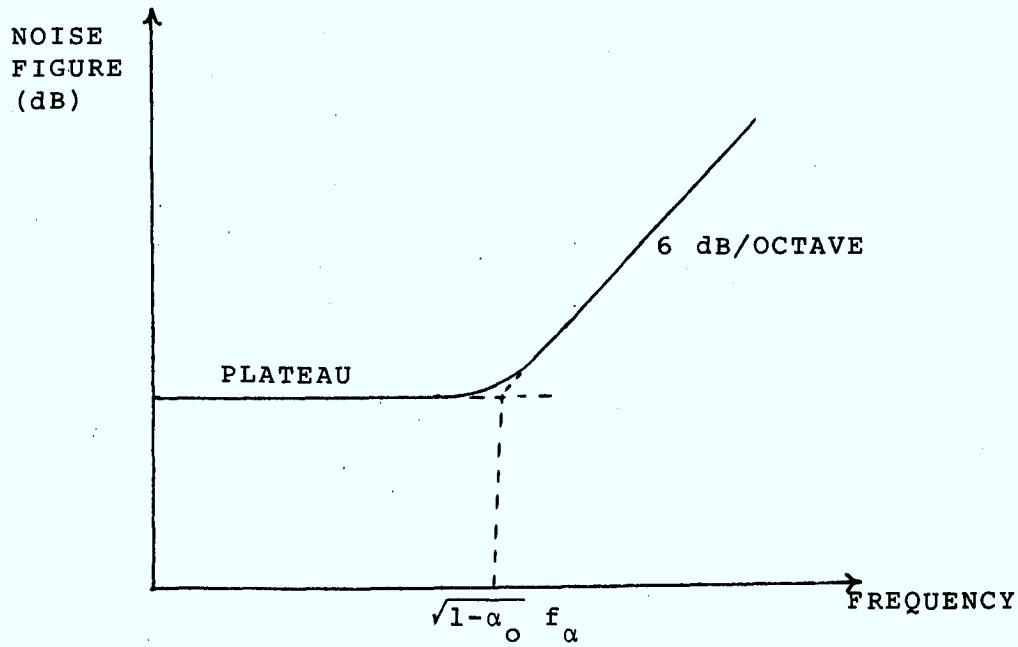


Fig. 6. Noise figure performance with frequency.

6.2 Power Transistors

Power gain and efficiency are the prime considerations for a power transistor. Power transistors must be able to handle higher currents than small-signal transistors. Since

$$GF^2 = \frac{\alpha_o f_T}{8\pi S^2 r_o C_o} \tag{54}$$

is independent of area, it should be possible to scale up the transistor area to handle any value of current, provided the device has a heat sink which will dissipate the heat generated within the active part of the transistor. However, there are physical limits to large area devices; L or N can not be increased at will. It is difficult to produce large area fault-free epitaxial silicon and control fabrication processing over a large area. In addition, small-signal models do not provide any indication of performance optimization in the large-signal region (for example, it doesn't predict saturation where CB junction becomes forward biased). Hence empirical results must be used. There are three parameters of interest; the power output, the power gain and the efficiency.

Power output is determined by the current and voltage handling capacity of the transistor. Current handling capability is determined by the emitter periphery and epi-

taxial layer resistivity (because of current crowding to the edges of the emitter and the Kirk effect, respectively). The voltage handling capability is determined by the breakdown voltage BV_{CBO} which is limited by the resistivity of the epitaxial layer and the curvature at the CB junction.

Power gain has been shown previously to be given by $U = (f_{max}/f)^2$. For high power gain, a high f_{max} is required.

Efficiency is equal to the ratio of P_{out} to P_{in} and determines the heat dissipation within the device. If the collector circuit presents a real load at the drive-signal frequency (the collector load can be real at a harmonic in a multiplier mode) and an open or short circuit at harmonic frequencies, energy will be transferred from the dc power supply and output power will be developed at the fundamental frequency. The power dissipated within the device is proportional to the voltage across the CB junction during the flow of pulsed current. High efficiency can be obtained only if current flow occurs when the total collector voltage is low. The carrier transit time must be a small fraction of an RF cycle (hence small W is required) if the high efficiency phase condition is to be maintained.

A limit exists on the maximum current in the base or maximum total mobile space charge traversing the base. When mobile charge is comparable to the fixed charge of the base, then base widening occurs, reducing efficiency and limiting high-frequency performance.

Power Frequency Limitations

Power-frequency limitations are inherent in the scattering-limited velocity of carriers in semiconductors as well as the maximum fields attainable in semiconductors without the onset of avalanche multiplication. Johnson [11] derived four basic equations for the power-frequency limitations on microwave power transistors based on the following three assumptions:

i) There is a scattering-limited velocity of carriers in a semiconductor [$v_s \sim 6\times10^6$ cm/s for electrons and holes in Si and Ge].

ii) There is a maximum electric field E_m that can be sustained in a semiconductor without having avalanche breakdown [$E_m \sim 10^5$ V/cm in Ge, $E_m \sim 2\times10^5$ V/cm in Si].

iii) The maximum current that a microwave power transistor can carry is limited by the base width.

Johnson's **first** equation gives a voltage-frequency limitation:

$$V_m f_T = \frac{E_m v_s}{2\pi} = \begin{cases} 2\times10^{11} \text{ V/s for Si} \\ 1\times10^{11} \text{ V/s for Ge} \end{cases} \tag{55}$$

where

$f_T = 1/2\pi\,\tau_{ec}$ is the charge carrier transit time cut-off frequency

τ_{ec} = L/V is the average time for a charge carrier moving at an average velocity $\overline{V}$ to traverse the emitter-collector distance L

$V_m = E_m L_{min}$ is the maximum allowable applied voltage between E and C

v_s = the scattering-limited drift velocity

E_m = the maximum electric field.

f_T cannot be increased without limit by reducing L and hence τ_{ec}, since E_m will eventually be exceeded. Attainable f_T is lower than the value predicted by (55) since E and v_s are not uniform across the device. Also, fabrication proceeds limit L to greater than 25 μm for overlay and matrix devices, and greater than 250 μm for interdigital devices.

Johnson's second equation provides a current-frequency limitation:

$$(I_m X_c) f_T = \frac{E_m v_s}{2\pi} \tag{56}$$

where

$I_m = V_m/X_c$ = maximum collector current of the device

$X_c = 1/2\pi f_T C_c$ is the reactive output impedance

C_c = the collector-base capacitance.

In practice, the area of the device limits I_m.

Johnson's third equation yields a power-frequency limitation. Maximum output power $P_m = I_m V_m$ is obtained when the load resistance R_L matches X_c. Multiplying (55) and (56) gives

$$\sqrt{P_m X_c}\ f_T = \frac{E_m v_s}{2\pi} \tag{57}$$

Thus for a given X_c, P_m must decrease as f_T is increased. Also, at a particular frequency, P_m is inversely proportional to X_c. To lower X_c requires increasing C_c thereby increasing junction areas.

Johnson's fourth limitation yields a power gain-frequency limitation:

$$\sqrt{G_m V_{th} V_m}\ f_T = \frac{E_m v_s}{2\pi} \tag{58}$$

where

G_m = maximum available power gain (MAG) which occurs when input and output of transistor are conjugately matched

$$G_m = \left(\frac{f_T}{f}\right)^2 \frac{Z_{out}}{Z_{in}} \tag{59}$$

$V_{th} = \frac{kT}{e}$ is the thermal voltage

k = Boltzmann's constant

T = absolute temperature in °K

e = electron charge

Neglecting series resistance,

$$\frac{Z_{out}}{Z_{in}} = \frac{C_{in}}{C_{out}} \tag{60}$$

where

$$C_{in} = C_{de} \simeq \frac{Q_m}{V_{th}} = \frac{I_m \tau_B}{V_{th}} \tag{61}$$

and

$$C_{out} = C_c = \frac{I_m \tau_o}{V_m} \tag{62}$$

In the above expressions, simplifying assumptions are on the optimistic side, so it is unlikely that the maximum values V_m, I_m, P_m can be reached. Nevertheless, the expressions can be used to relate device geometry to performance. In the planar transistor, a lateral current flows between the emitter and base contacts, producing a voltage drop that serves as a bias to cut off current flow from portions of the emitter (emitter de-biasing). Consequently, almost all injection occurs from the edge of the emitter (the central region contributes capacitance but little current). Also base widening occurs when injected current exceeds 40-60 μA per μm of emitter periphery. Thus,

$$I_m \simeq 4\text{x}10^{-5}\ \text{EP} \tag{63}$$

where EP is the emitter periphery in μm.

Also C_c for a typical microwave NPN silicon power transistor with 28V bias is

$$C_c \simeq 4.7\text{x}10^{-18}\ \text{BA} \tag{64}$$

where BA is the base area in μm^2 ($X_{dc} \simeq 2\ \mu m$). Combining (57, 63 and 64) gives

$$\sqrt{P_m X_c}\ f_T \simeq 125M \tag{65}$$

where

$$M = \frac{EP}{BA}\quad (\mu m/\mu m^2) \tag{66}$$

is a geometrical figure-of-merit and f_T is in GHz.

Design of Microwave Power Transistors

Because of emitter de-biasing, large P_{out} and hence large I_m requires large periphery EP of the emitter. At the same time, emitter area EA must be reduced to minimize C_e and reduce shunting of the EB junction resulting in loss of current injected into the base. Thus, a large EP/EA ratio is required.

Similarly, base area BA must be kept small to reduce shunting effect of C_c on the load resistance which reduces

load current and output power. Thus, a large EP/BA = M ratio is required.

Thirdly, base pad area PA must be small to minimize the parasitic capacitance C_{bc} which shunts C_c. C_{bc} can be reduced by increasing the oxide thickness between the metal pad and the collector region, and maximizing the EP/PA ratio.

Several geometries are used to maximize EP/EA, EP/BA, EP/PA, the three basic designs are interdigitated, overlay, and matrix (also called mesh or emitter grid). These geometries are shown in Fig. 7. State-of-the-art transistor fabrication limits emitter width to about W ~ 1 μm, base thickness to t ~ 0.2 μm, emitter length L ~ 25 μm.

A comparison of performance of power transistors is given in Table II.

For a given collector doping profile, the operating voltage increases with E_g since breakdown voltage increases with E_g. The maximum junction temperature is the temperature at which the base region becomes intrinsic. Thus, GaAs is superior to Si and Ge. Silicon has the highest thermal conductivity and best heat dissipation. The $\sqrt{P_m X_c}\ f_T$ product is about the same for silicon and GaAs, so silicon is preferred.

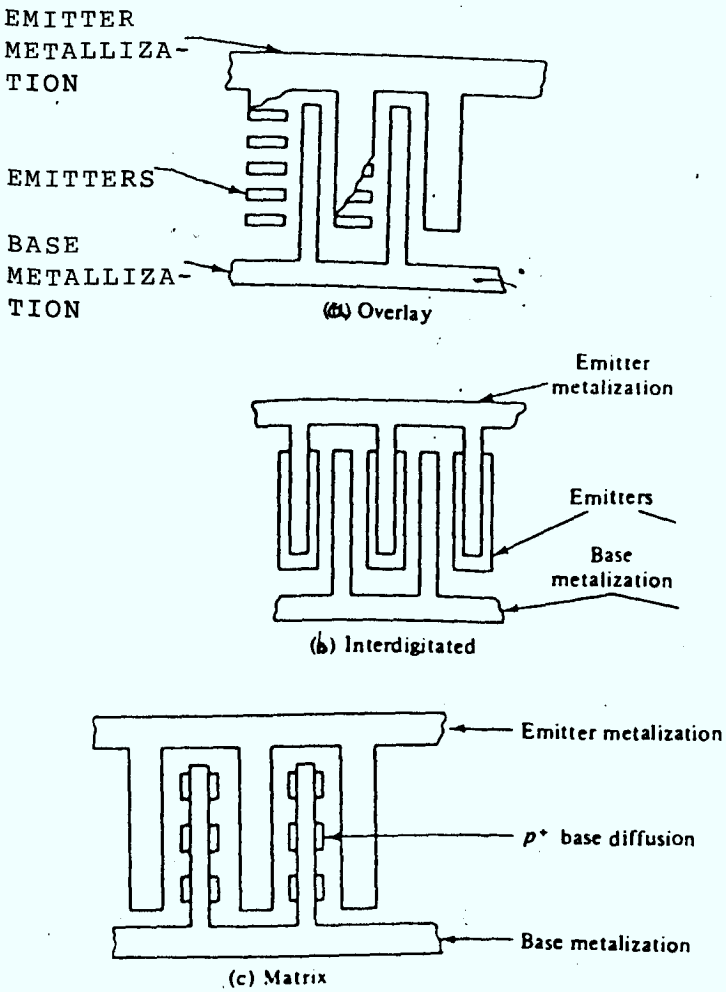


Figure of Merit (M) of Various Surface Geometries. (From Sobol and Sterzer [12]; reprinted by permission of the IEEE, Inc.)

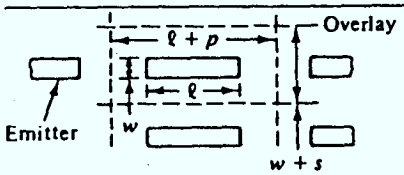
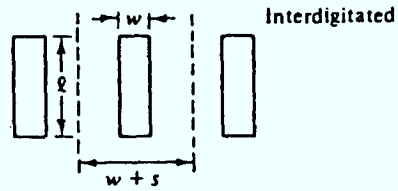
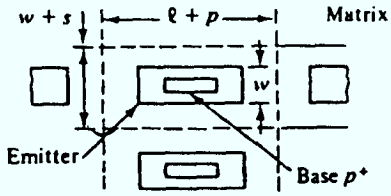
Surface geometry and unit cell	$M = \frac{EP}{BA}$
	$\frac{2(\ell + w)}{(w + s)(\ell + p)}$
	$\frac{2(\ell + w)}{\ell(w + s)}$
	$\frac{2(\ell + w)}{(w + s)(\ell + p)}$

Fig. 7. Surface geometrics of microwave power transistor [12]. (After Sobol et al.; reprinted with permission from the IEEE, Inc.).

Table II

PERFORMANCE OF POWER TRANSISTORS

Quality	Ge (*p-n-p*)	Si (*n-p-n*)	GaAs (*p-n-p*)
Band Gap (eV) at 300°K	0.80	1.12	1.43
Operating Voltage (volts)	20	50	55
Max. Junction Temp. (T_j)	100°C	200°C	450°C
Thermal Cond. (κ)	0.5 :	1 :	0.3
$\sqrt{\text{Power} \times \text{Impedance}}\, f \simeq \frac{\mathcal{E}_m v_{sl}}{2\pi}$ (V/sec)	2×10^{11}	4×10^{11}	4.6×10^{11}

For power amplifiers, class C operation is preferred. In this mode, fewer problems relating to instability and thermal runaway arise than do when operation is in class A or AB mode. In class C mode, both EB and CB junctions are reverse-biased and no current flows in the absence of an applied signal. When an RF voltage of sufficient magnitude is applied to the EB junction, it becomes forward-biased for a fraction of the RF cycle. The flow of energetic electrons injected during the time of forward EB bias represents a pulse of current in the collector circuit. If the carrier transit time is a small fraction of the RF cycle, the current is extracted when the collector voltage is low resulting in high efficiency.

For linear power amplifiers, class A or class AB mode must be used. These modes require that the EB junction be forward-biased and current flows even without the presence of RF signals. A microwave transistor operated under this condition is subject to many problems as described below.

Low-frequency instabilities associated with "second breakdown" can occur, whereas in class C operation this is not likely to occur since the transistor is shut off during each cycle.

"Second breakdown" of the collector junction is a transition from a high voltage low-current condition at the collector to a lower voltage high-current condition.

Second breakdown is partly due to thermal effects and partly due to avalanche injection in the collector at the N-N^+ transition between the epitaxial layer and the substrate. In particular, thin, lightly-doped epitaxial collector regions are susceptible to second breakdown.

In addition, several emitters may hog current (due to material or process variations over large-area transistors), producing thermal runaway or hot-spot generation which ultimately lead to device failure. The negative temperature coefficient of the V_{EB} leads to increased current injection at these sites, resulting in thermal runaway. Hot spots also contribute to second breakdown.

These problems can be minimized in two ways. First, several small devices, individually biased, may be used at power levels below the normal class C rating. Hot spots are usually not formed in small devices. Separately biasing several devices is equivalent to external ballasting an amplifier. Another, more elegant method is to ballast the transistor internally. Uniform injection, uniform temperature profiles, and resistance to second breakdown are obtained by the use of emitter ballast resistors. The resistors can be used in series with each metallization finger or can be used directly in each emitter site. The voltage drop across the ballast resistor which is passing current to a potential thermal runaway emitter site will

cancel the effect of the negative temperature coefficient of V_{EB} in that site and prevent a hot-spot from forming. Titanium ballast resistors are often used in the metallization to the multiple emitter sites, and are usually graded with higher values in the centre of the pattern than at the edges to maintain a more uniform temperature over the entire emitter area. Individual emitter site ballasting is more efficient and this is accomplished with the use of a high resistivity polysilicon layer vapour-deposited between the emitter sites and the emitter metal.

To achieve maximum bandwidth, efficiency and power gain, microwave power transistors are often made with an electrical network inside the package, next to the transistor chip, to transform the device input and output impedance to match a 50 ohm microwave transmission line. These internal matching networks are designed to include the inductance of the bond wires to the transistor chip and the parasitic capacitances of the package. The bond wire length must be accurately controlled and a different length for each cell of a multi-cell chip may be required. Aluminum or gold wires are used depending on the metallization on the transistor chip. Aluminum wire is more rigid and will hold in place; gold wires require support in mid-span.

Another possible failure mode arises when the output circuit is detuned at full power. The resulting mismatch

causes a high VSWR condition at the collector, subjecting the transistor to instantaneous voltage peaks many times the supply voltage. Avalanching then takes place in the collector depletion region. The problem can be minimized by collector ballasting, which is provided by a thick undepleted collector layer. Transistors with both emitter and collector ballasting can usually withstand high VSWR loads at any phase angle.

7. MATHEMATICAL MODEL OF THIRD-ORDER INTERMODULATION IN BIPOLAR TRANSISTORS BASED ON VOLTERRA SERIES ANALYSIS OF NON-LINEAR DEVICE ELEMENTS

7.1 Introduction

Non-linear distortion is a critical problem in many applications of bipolar transistors. In particular, intermodulation distortion is a significant problem in a long-haul solid-state analog communication system. The per channel mile cost can be significantly reduced by increasing the bandwidth of the system, and distortion presents one of the most severe limitations on increasing the bandwidth of such a system.

It is important that the frequency-dependent nature of this distortion be well understood. One reason is that distortion is more of a problem at higher frequencies where feedback is limited. Another is that in an amplifier configuration, each transistor often sees the frequency-dependent source and load impedances. A third reason arises since the input speech signal to the amplifier can be best represented by a bandlimited Gaussian noise. A fourth reason is that in an amplifier configuration, frequency shaping is used in order to compensate the frequency-dependent loss of the cable. Finally, the last and most important reason is due to a particular type of third-order product, namely, an $f_1 + f_2 - f_3$ product, which is most troublesome as it tends to add in-phase.

In this section, a very comprehensive frequency-dependent model is used and Volterra series is employed to analyze frequency-dependent distortion. This discussion follows that of Narayanan and Poon [12]. This approach has resulted in simple distortion expressions that are valuable for the circuit designer. The integral charge control model, which includes many high-level effects, forms the basis for the analysis. For a transistor with resistive terminations, the analysis leads to simple expressions for transistor Volterra transfer functions; these expressions clearly bring out the experimentally observed frequency-dependent nature of transistor distortion. These expressions are extremely simple. The asymptotic low- and high-frequency expressions give valuable insight into the nature of distortion.

Extremely accurate modelling of the transistor is necessary to compute precisely its second- and third-order distortion at different bias points and frequencies. The integral charge control model (ICM) is well suited for this purpose. For the present analysis, the three basic equations describing ICM can be conveniently expressed in functional form as follows:

$$J(Q_b, I_c, V_{eb}, V_{cb}) = 0 \tag{67}$$

$$G(Q_b, I_c, V_{eb}, V_{cb}) = 0 \tag{68}$$

$$H(I_c, V_{eb}, V_{cb}) = I_b \tag{69}$$

where Q_b is the total stored charge in the base of the transistor, I_c is the collector current, V_{eb} is the emitter-to-base terminal voltage, V_{cb} is the collector-to-base terminal voltage, and I_b is the base current (Fig. 8). Equation (67) describes the collector current. For a transistor biased in the active region, the dominant part of the collector current can be described by a charge control relationship, which states that the collector current is proportional to the exponential of V_{eb} and inversely proportional to Q_b. Equation (68) describes the total stored charge in the base of the transistor. It is through modelling of Q_b that many high-level effects, i.e., base widening, conductivity modulation, and Early effect, can be incorporated. Equation (69) which describes the base current empirically, consists of two non-ideal components and one ideal component.

The input and output equations for the common emitter configuration are

$$I_c = \frac{1}{Z_L(f)} * [V_{cc} - V_{cb} + V_{eb}] \tag{70}$$

$$\dot{Q}_b = \frac{1}{Z_g(f)} * [V_g + V_{eb}] - I_b \tag{71}$$

where * denotes operator notation, Z_L and Z_g are the com-

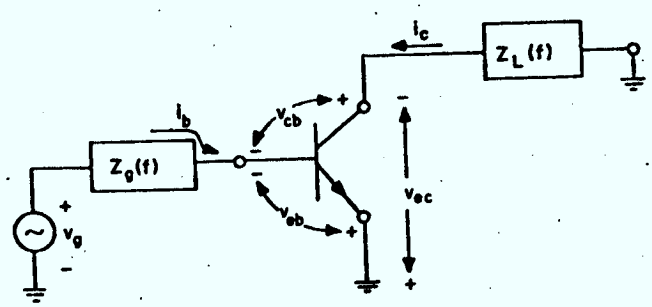


Fig. 8. Common emitter configuration.

plex output and input impedances, and V_g is the generator voltage.

The intrinsic equations (67)-(69) are independent of frequency. Using (67) and (68), any two of the variables V_{eb}, V_{cb}, I_c and Q_b can be expanded in a bivariate power series in terms of the other two. This approach conveniently separates modelling and the circuit analysis program such that minor change in the model does not affect the circuit analysis program.

The analysis outlined in this section is limited to real load and source impedances. For real load impedance, we can eliminate I_c in (67)-(69) using (70). Thus v_{eb}, v_{cb} and i_c can be expressed as one-dimensional Taylor's series of q_b (where lowercase letters denote small signal quantities). In order to bring out the physical significance of the results, we will resort to the following expansions:

$$i_c = \frac{v_{ce}}{R_L} = i_{01}q_b + i_{02}q_b^2 + i_{03}q_b^3 \tag{72}$$

$$\frac{v_{eb}}{R_L} = i_{ex1}q_b + i_{ex2}q_b^2 + i_{ex3}q_b^3 \tag{73}$$

$$i_b = i_{r1}q_b + i_{r2}q_b^2 + i_{r3}q_b^3 \tag{74}$$

where we shall call i_{on}, i_{exn}, and i_m (n = 1,2,3) the

non-linear coefficients. i_{on} expresses the f_T non-linearity and

$$i_{01} = \frac{\partial I_c}{\partial Q_b} = 2\pi \hat{f}_T \tag{75}$$

where $\hat{f}_T$ defined here is the cut-off frequency under the circuit constraint as shown in Fig. 8.

The higher-order coefficients i_{02} and i_{03} are related to the first and second derivatives of $\hat{f}_T$ versus I_c. Physically, the $\hat{f}_T$ - I_c characteristic depends on the emitter capacitance at low current and on the base-pushout effect at high current. For a transistor to be used in an amplifier circuit, it is usually the base-pushout effect that is dominant. Thus i_{02} and i_{03} are mostly determined by the detailed doping profile in the epitaxial collector region.

i_{exn} expresses the "exponential non-linearity" (also called the "input non-linearity"), which is especially important for low-bias currents. For example, expressing the collector current of an n-p-n transistor to have the form

$$I_c = I_s e^{qV_{be}/nkT} \tag{76}$$

then

$$i_{ex1} = \frac{kT}{qR_g} \frac{n}{I_c} \hat{\omega}_T \tag{77}$$

where n is the conventionally so-called emission coefficient that depends on the effects of base resistance, base widening, and conductivity modulation. Similarly, the higher order coefficients i_{ex2} and i_{ex3} are related to the derivatives of i_{ex1} with respect to collector current.

The coefficient i_{rn} expresses the non-linearity in current gain (recombination non-linearity). In particular,

$$i_{r1} = \frac{\partial I_b}{\partial Q_b} = \frac{\hat{\omega}_T}{\beta} \tag{78}$$

where β is the small signal current gain. Physically, β depends on recombination properties of the device, but also on impact ionization, base widening, conductivity modulation, and Early effect. The higher order coefficients i_{r2} and i_{r3} depend on the second and third derivatives of β and $\hat{\omega}_T$.

The non-linear coefficients i_{exn}, i_{on}, and i_m can be analytically derived; they can also be calculated using a numerical technique.

The small signal version of (71) is given below:

$$\dot{q}_b = \frac{1}{R_g}(v_g + v_{eb}) - i_b \tag{79}$$

In this equation, v_{eb}/R_g and i_b can be expressed as functions of q_b using (72) and (74).

7.2 Volterra Analysis

If q_b is expressed as a Volterra series of v_g, the kernels of q_b can be successively determined from (79). The Volterra series relating q_b to v_g is given below:

$$q_b(t) = \int_{-\infty}^{\infty} a_1(\tau) v_g(t-\tau) d\tau + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a_2(\tau_1,\tau_2) \prod_{i=1}^{2} \cdot v_g(t-\tau_i) d\tau_i + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a_3(\tau_1,\tau_2,\tau_3) \cdot \prod_{i=1}^{3} v_g(t-\tau_i) d\tau_i + \ldots \tag{80}$$

where $a_1(\tau)$, $a_2(\tau_1,\tau_2)$, and $a_3(\tau_1,\tau_2,\tau_3)$ are the first-, second-, and third-degree Volterra kernels.

Equating linear terms in (79), we have

$$A_1(f) = \frac{1}{R_g} \frac{1}{j2\pi f + (i_{r1} - i_{ex1})} \tag{81}$$

The second-order transfer function [two-dimensional transform of $a_2(\tau_1,\tau_2)$] is

$$A_2(f_1,f_2) = \frac{(i_{r2} - i_{ex2})}{j2\pi f_s + (i_{r1} - i_{ex1})} \prod_{i=1}^{2} A_1(f_i) \tag{82}$$

where $f_s = f_1 \pm f_2$.

The third-order transfer function [three-dimensional transform of $a_3(\tau_1,\tau_2,\tau_3)$] is

$$A_3(f_1,f_2,f_3)$$

$$= \frac{(i_{r3}-i_{ex3}) \prod_{i=1}^{3} A_1(f_i)+(i_{r2}-i_{ex2})\overline{[A_1(f_i)A_2(f_2,f_3)]}}{j2\pi f_t+(i_{r1}-i_{ex1})} \tag{83}$$

where $f_t = f_1 \pm f_2 \pm f_3$ and the overbar denotes symmetrizing operation.

From (73), we can now derive the Volterra transfer functions relating i_c to v_g. Let the Volterra transfer functions relating i_c to v_g be denoted by $B_1(f)$, $B_2(f_1,f_2)$, $B_3(f_1,f_2,f_3)$, etc. These functions are given below:

$$B_1(f) = i_{01}A_1(f) \tag{84}$$

$$B_2(f_1,f_2) = i_{01}A_2(f_1,f_2) + i_{02} \prod_{i=1}^{2} A_1(f_i) \tag{85}$$

$$B_3(f_1,f_2,f_3) = i_{01}A_3(f_1,f_2,f_3) + i_{03} \prod_{i=1}^{3} A_1(f_i)$$

$$+ 2i_{02} \cdot [\overline{A_1(f_1)A_2(f_2,f_3)}] \tag{86}$$

The expression for the transfer function between i_c and v_g ($B_1(f)$) is given by (84) and is easily seen to be the following one-pole expression:

$$B_1(f) = \frac{x}{\gamma+j2\pi f} \tag{87}$$

where

$$x = \frac{i_{01}}{R_g}, \quad \gamma = i_{r1} - i_{ex1}.$$

The 3 dB point will be close to beta cut-off frequency, depending upon the load impedance. However, the above expression, because of the inherent assumption of ICM, is not valid for frequencies close to $\hat{f}_T$ of the device.

The second-order Volterra transfer function $B_2(f_1,f_2)$, as given by (85), simplifies to

$$B_2(f_1,f_2) = \left[\frac{C+jDf_s}{\gamma+j2\pi f_s}\right] \prod_{i=1}^{2} \left[\frac{1}{\gamma+j2\pi f_i}\right] \tag{88}$$

where

$$f_s = f_1 \pm f_2$$

$$C = \frac{1}{R_g^2}\left[(i_{r2}-i_{ex2})i_{01}+i_{02}(i_{r1}-i_{ex1})\right]$$

$$D = \frac{2\pi}{R_g^2}\, i_{02}.$$

Comparing (87) and (88), frequency-dependent functions appear as $\gamma+j2\pi f$, in the denominator of both, where γ is a linear parameter. In the numerator of (88), the frequency dependence is given by $C+jDf$, where both C and D involve second-order non-linear coefficients. The above expression for $B_2(f_1,f_2)$ can be represented pictorially as a cascade of a linear system $1/(\gamma+j2\pi f)$, a memoryless

squarer, and a linear system $(C+jDf)/(\gamma+j2\pi f)$, as shown in Fig. 9.

The simplified expression for the third-order Volterra function $B_3(f_1,f_2,f_3)$ is given below:

$$B_3(f_1,f_2,f_3) = \left(\frac{E+jFf_t}{\gamma+j2\pi f_t}\right) \prod_{i=1}^{3} \left(\frac{1}{\gamma+j2\pi f_i}\right)$$

$$+ \overline{\left[\frac{P+jQf_t}{(\gamma+j2\pi f_t)(\gamma+j2\pi f_{si})}\right]} \prod_{i=1}^{3} \left[\frac{1}{\gamma+j2\pi f}\right] \tag{89}$$

where the overbar denotes a symmetrical Volterra transfer function,

$$E = \frac{1}{R_g^3} \left[(i_{r3}-i_{ex3})i_{01}+i_{03}(i_{r1}-i_{ex1})\right] \tag{90}$$

$$F = \frac{2\pi}{R_g^3} i_{03} \tag{91}$$

$$P = \frac{2}{R_g^3} \left[(i_{r2}-i_{ex2})i_{01}+i_{02}(i_{r1}-i_{ex1})\right](i_{r2}-i_{ex2}) \tag{92}$$

$$Q = \frac{4\pi}{R_g^3} \left[i_{r2}-i_{ex2}\right]i_{02}. \tag{93}$$

The first term in (89) is due to third-order non-linearities only. The denominator involves the function $\gamma+2\pi f$ only; the parameters E and F are functions of third-order non-linear coefficients. The first term can be visualized as a cascade of a linear system $[1/(\gamma+j2\pi f)]$, a memoryless cuber, and a linear system $(E+jFf)/(\gamma+j2\pi f)$, as shown in Fig. 10.

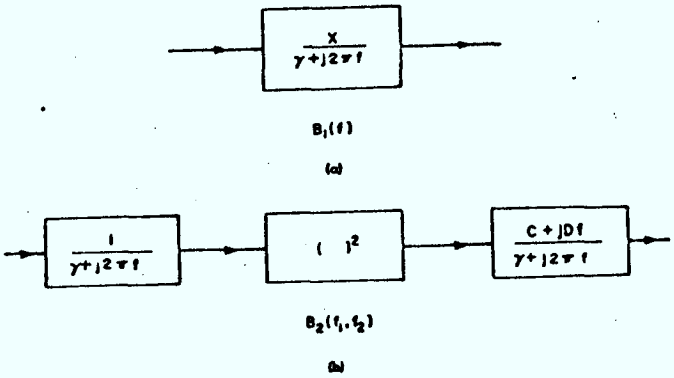


Fig. 9. (a) Linear transfer function.
(b) Second-order transfer function.

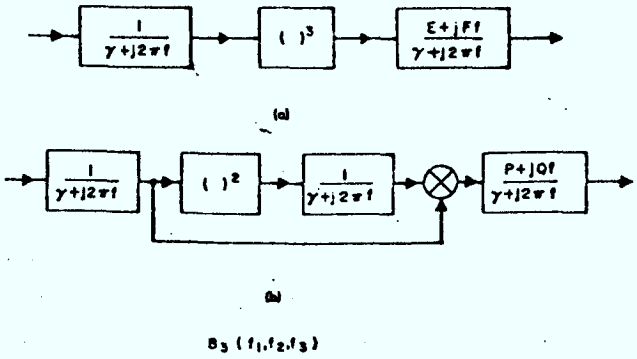


Fig. 10. Third-order transfer function.
(a) Third-order non-linearity only.
(b) Second-order interaction term.

The second term in (89) represents the contribution due to interaction of second-order non-linearities; the constants P and Q depend only on second-order non-linearities. The second term is represented by a linear system $1/(\gamma+j2\pi f)$ followed by two parallel sections (Fig. 10). One section consists of a memoryless squarer and a linear system; the other section is an identity system. The outputs of the parallel sections are multiplied and passed through a linear system $[P+Qf)/(\gamma+j2\pi f)]$.

From the above Volterra transfer functions, the distortion coefficients can be derived. The simplified expression for the third-order distortion coefficient is given below:

$$M_{3E} = 10 \log \left[\frac{d(1+ef_t^2)}{(1+cf_t^2)} + h(f_t, f_{si})\right] \tag{94}$$

where

$$d = \left[\frac{10^{-3}}{2R_L i_{01}^2} \left[\frac{i_{r3}-i_{ex3}}{i_{r1}-i_{ex1})} + \frac{i_{03}}{i_{01}}\right]\right]^2$$

$$e = \left[\frac{2\pi i_{03}}{i_{01}(i_{r3}-i_{ex3})+(i_{r1}-i_{ex1})i_{03}}\right]^2$$

$$h(f_t, f_{si}) = \frac{10^{-3}}{x^3 2R_L} \frac{1}{\gamma^2+(2\pi f_t)^2}$$

$$\cdot \left[\left| E+jFf_t + \overline{\frac{P+jQf_t}{\gamma+j2\pi f_{si}}} \right| - E^2 - F^2 f_t^2 \right]$$

The first term in the above expression is due to third-order non-linearities only and it depends only on third-order product frequency. The second term is due to the interaction of second-order non-linearities and it is a function of both the second- and third-order product frequencies. If the second-order distortion is small, this term can be neglected, as it involves the product of second-order non-linear coefficients.

Fig. 11 shows both computed and experimental results of M_{3E} as a function of third-order product frequency at 80 mA and 130 mA bias currents. It is seen that the first term of the expression (94) tracks the experimental results well. 10 log [d] and 10 log [de/c] give the low- and high-frequency values, respectively. Fig. 12 shows computed results of M_{3E} at lower bias currents. It is seen that at 35 mA, the shape of the M_{3E} versus frequency curve is not the same as would be predicted by the first term of (94).

Further insight into the nature of distortion can be obtained by focusing attention on the low- and high-frequency asymptotic limits where very simple expressions are obtained. As noted earlier, at higher bias currents the common emitter low- and high-frequency M_{3E} expressions are given by 10 log(d) and 10 log(de/c), respectively.

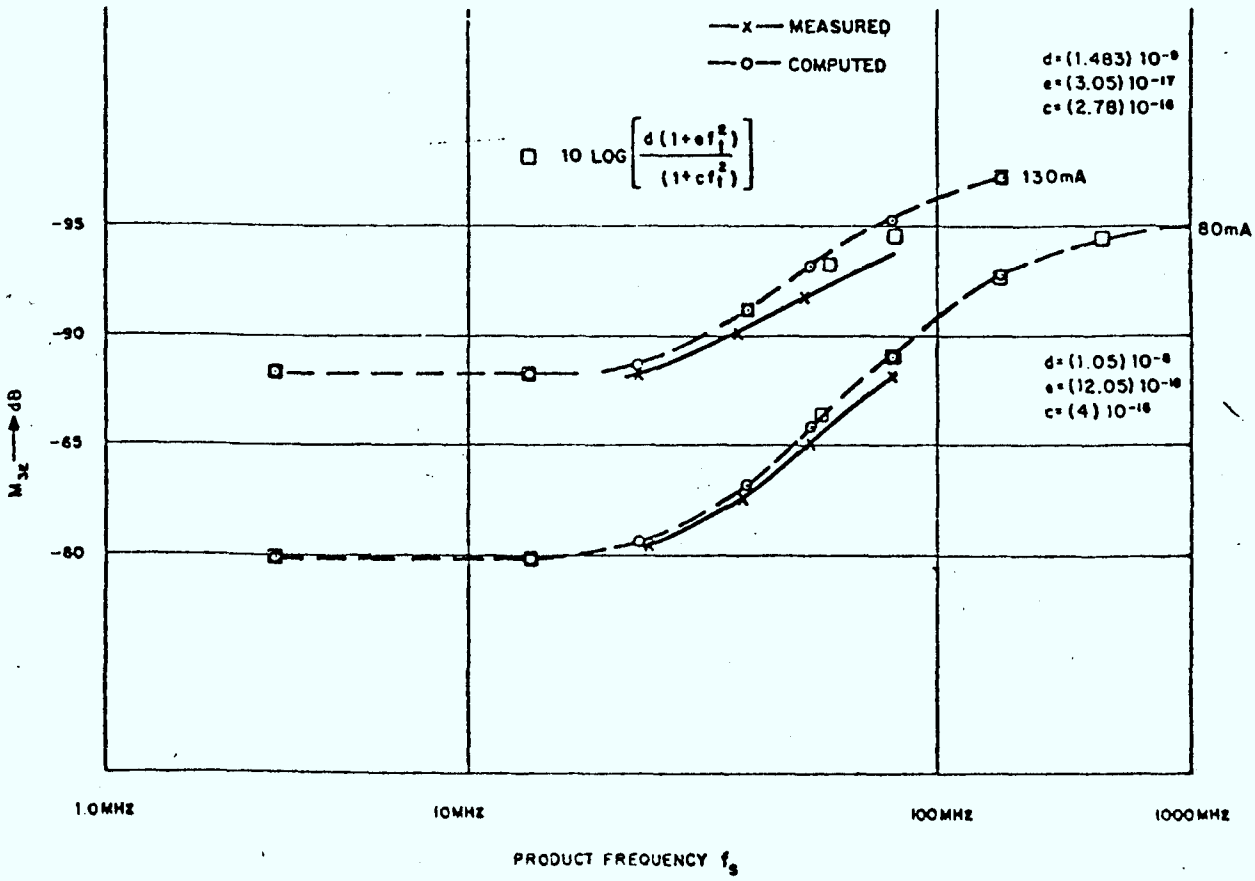


Fig. 11. Common emitter M_{3E} variation with frequency.

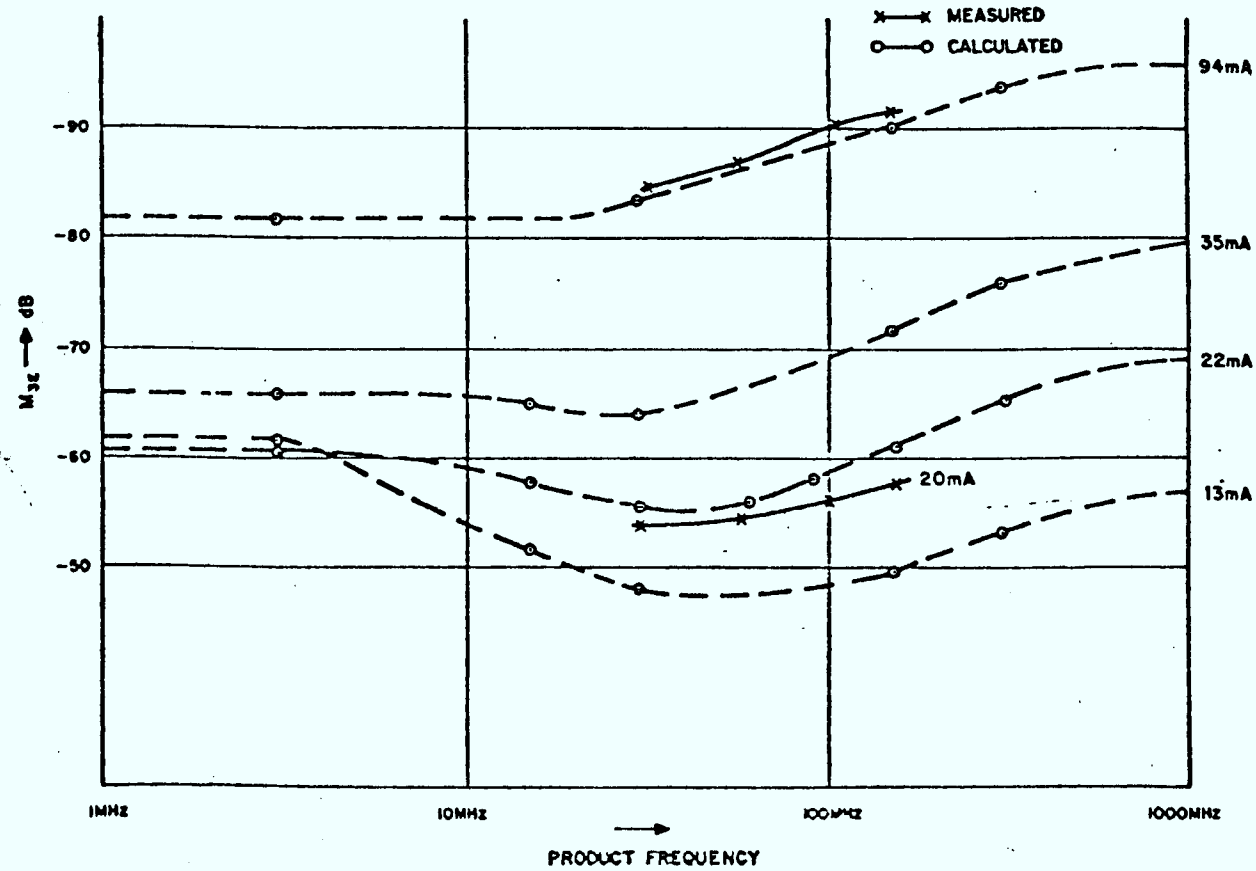


Fig. 12. M_{3E} variation with frequency 1-tone.

Using the non-linear coefficients i_{on}, i_{exn}, and i_m, the expressions for M_{3E} in the low- and high-frequency limits can be written as

$$M_{3E} \text{ (low frequency)} = 20 \log(g) \left[\frac{i_{r3}-i_{ex3}}{i_{r3}-i_{ex1}} \frac{1}{i_{01}^2} + \frac{i_{03}}{i_{01}^3}\right] \tag{95}$$

$$M_{3E} \text{ (high frequency)} = 20 \log(g) \left[\frac{i_{03}}{(i_{01})^3}\right] \tag{96}$$

where

$$g = \frac{10^{-3}}{2R_L}$$

A meaningful physical interpretation of (95) and (96) can be obtained by writing them in terms of $\hat{\omega}_T$, β, n, and their derivatives with respect to I_c. In particular, by considering a specific transistor of interest (i.e., the one whose results are shown in Figs. 10 and 11) and by dropping terms that are numerically small for that transistor at a dc operating point (i.e., I_c = 100 mA and V_{ce} = 15.0 V), the M_{3E} expressions are simplified to be

$$M_{3E} \text{ (low frequency)} = 20 \log_{10} \left[g \frac{kT}{6qR_g} \frac{\frac{2n}{I_c^3} - \frac{2n'}{I_c^2}}{\frac{1}{\beta} + \frac{kT}{qR_g}\frac{n}{I_c}} \right] \tag{97}$$

$$M_{3E} \text{ (high frequency)} = 20 \log_{10} \left[g \frac{\hat{\omega}''_T}{6\hat{\omega}_T} \right]. \tag{98}$$

Even though these simplified expressions are accurate only up to 1 or 2 dB, they illustrate the physical mechanism that is causing the distortion for this particular transistor under this particular biasing condition. In the low-frequency limit, the factor $(2n/I_c^3 - 2n'/I_c^2)$ is a result of exponential non-linearity and it is this effect that controls the low-frequency distortion. In the high-frequency limit, it is the second derivative of $\hat{f}_T$ that controls the distortion. Thus to have minimum third-order distortion, the $\hat{f}_T$ versus I_c curve should be as linear as possible.

PART III

AM-TO-PM MEASUREMENTS USING THE MICROWAVE LINK ANALYZER

8. RELATIONSHIP BETWEEN AM-TO-PM DISTORTION AND DIFFERENTIAL GAIN/DIFFERENTIAL PHASE FOR CASCADED LINEAR/NON-LINEAR NETWORKS REPRESENTING MICROWAVE FIELD EFFECT AND MICROWAVE BIPOLAR TRANSISTORS

8.1 Introduction

In the foregoing sections of this document, topics regarding the operation of both Field Effect Transistors and Bipolar Transistors were discussed. Emphasis centred on deriving intermodulation distortion parameters based on non-linear elements of the device geometry. In order to estimate the intermodulation distortion parameters of interest, in specific, the AM-to-PM distortion, it was necessary to first measure the non-linear behaviour of these device elements and then to apply either a Volterra series analysis or a power series analysis.

In the following sections, an alternative systems level method for determining the AM-to-PM distortion of a transistor will be outlined, based on the measuring capabilities of the Hewlett-Packard Microwave Link Analyzer. With this technique, it is not necessary to measure the non-linear behavour of specific device elements (the gate capacitance, transconductance, drain conductance and drain capacitance for a field effect transistor, and the collector current, base current and emitter-base voltage for a bipolar transistor). Use is made, however, of knowledge of the cascaded nature of these non-linearities.

The Microwave Link Analyzer provides a ready measurement of the differential gain/differential phase of a two-port system. Operation of the analyzer is provided in references 14 and 15. As the following analysis demonstrates, it is possible to relate this measurement to that of a two-port system's AM-to-PM distortion in an intriguing manner.

The development of expressions for differential gain/differential phase in terms of two-port network parameters (such as amplitude and phase characteristics, AM-to-PM conversion constant) follows that reported in [13].

8.2 Analysis

Differential gain and differential phase are parameters defined for a transmission system with baseband input and output ports. The input signal is composed of a sinusoidal test tone with frequency ω_m and amplitude V_m superimposed on a slowly varying sweep signal V_s.

$$v_{in} = V_s + V_m \cos \omega_m t \tag{1}$$

Varying the sweep signal amplitude, the gain and the phase of the test tone are found to be dependent on V_s due to the system non-linearities; therefore, at some amplitude $V_s = x$,

$$v_{out} = A(x)V_m \cos[\omega_m t - \phi(x)] \tag{2}$$

Using the characteristics A(x) and $\phi(x)$, the differential gain and the differential phase are defined by the following expressions:

$$DG(x) = \frac{A(x)}{A_o} \tag{3}$$

$$DP(x) = \phi(x) - \phi_o \tag{4}$$

Introducing the complex differential response, we have

$$D(x) = \frac{A(x)}{A_o} e^{-j[\phi(x)-\phi_o]} \tag{5}$$

Here the differential gain and differential phase are related to the complex differential response by the expressions:

$$\begin{aligned} DG(x) &= |D(x)| \\ DP(x) &= \text{Arc } D(x) \end{aligned} \tag{6}$$

For a system with no distortion, the differential gain is unity and the differential phase is zero. Any difference from these values will show up as a source of intermodulation distortion. In this definition, (x) is some value of the sweep signal amplitude. However, it could also be defined as the deviation from the centre frequency of the carrier by the sweep signal.

To calculate the differential gain and differential phase contributions of an FM system, a carrier phase modulated by a test-tone frequency could be assumed at the input of the system.

$$v_{ci} = V_{ci}\, e^{j[\omega_c t + \Delta p \cos \omega_m t]} \tag{7}$$

The method for measuring differential gain and phase is to sweep the carrier (ω_c) around the band centre frequency (ω_o). Since (x) can be defined as the deviation from the centre frequency of the carrier by the sweep signal, we have:

$$\omega_c = \omega_o + x \tag{8}$$

The phase deviation Δp has a small value and therefore a first-order approximation for the exponential is valid.

$$v_{ci} = V_{ci}\, e^{j\omega_c t}\, [1 + j\Delta p \cos \omega_m t] \tag{9}$$

Expressing the cosine function with exponentials:

$$v_{ci} = V_{ci} \left[e^{j\omega_c t} + \frac{j}{2} \Delta p e^{j(\omega_c + \omega_m)t} + \frac{j}{2} \Delta p e^{j(\omega_c - \omega_m)t} \right] \tag{10}$$

This phase modulated signal is expressed by the carrier and two sideband components. The transmission path will change the relative amplitudes and phases of the sideband

components, resulting in phase modulation distortion and also in phase-to-amplitude modulation conversion. In order to evaluate both these effects, a carrier with both phase and amplitude modulations has to be considered.

Denote by p and m the phase and amplitude modulation indexes, and by υ_p and υ_m the modulation phases. With these notations, a carrier having both PM and AM modulations can be written as:

$$v_c = V_c[1 + m\cos(\omega_m t-\upsilon_m)]e^{j[\omega_c t + p\cos(\omega_m t-\upsilon_p)]} \tag{11}$$

$$\simeq V_c[1 + m\cos(\omega_m t-\upsilon_m)+j\,p\cos(\omega_m t-\upsilon_p)]e^{j\omega_c t} \tag{12}$$

To express the modulation with two sideband components, we introduce the complex modulation amplitudes for AM and PM as follows:

$$\begin{aligned} M &= me^{-j\upsilon_m} \\ P &= pe^{-j\upsilon_p} \end{aligned} \tag{13}$$

Using the complex amplitudes M and P, the modulated carrier can be expressed as follows:

$$v_c = V_c\left[e^{j\omega_c t} + \frac{M+jP}{2}e^{j(\omega_c+\omega_m)t} + \frac{M^*+jP^*}{2}e^{j(\omega_c-\omega_m)t}\right] \tag{14}$$

where * denotes the complex conjugate.

Equation (14) gives a unique representation of a carrier modulated by a single test-tone frequency. Denote by U_u and U_l, the relative complex amplitudes of the upper and lower sideband components respectively, normalized to the carrier amplitude V_c.

$$v_c = V_c[e^{j\omega_c t} + U_u e^{j(\omega_c+\omega_m)t} + U_l e^{j(\omega_c-\omega_m)t}] \qquad (15)$$

Given U_u and U_l, the complex modulation amplitudes P and M can be determined by comparing Equations (10) and (11).

$$P = -j[U_u - U_l^*]$$
$$M = U_u + U_l^* \qquad (16)$$

Equations (15) and (16) give the basic relationships for calculating the differential phase and gain contributions of an FM system.

Mathematical Model for a Single Linear Network

First consider the case of a single linear network in the FM path, with gain characteristic $A(\omega)$ and phase characteristic $\psi(\omega)$. An input signal with frequency ω and phasor v_1 will produce an output v_2 given by:

$$v_2 = A(\omega)e^{-j\psi(\omega)}v_1 \qquad (17)$$

The relative amplitude and phase characteristics $\alpha(\omega)$ and $\phi(\omega)$ normalized to the complex gain at the nominal carrier

frequency ω_o will be used as follows:

$$e^{\alpha(\omega)-j\phi(\omega)} = \frac{A(\omega)e^{-j\psi(\omega)}}{A(\omega_o)e^{-j\psi(\omega_o)}} \tag{18}$$

The differential characteristics are measured with the input signal given by Equation (10). The output spectrum is calculated by multiplying each component by the corresponding complex gain factor. It should be noted that this factor is different for each sideband. Using the notation introduced above:

$$v_2 = A(\omega_c)e^{-j\psi(\omega_c)} V_{ci} \left[e^{j\omega_c t} + j\,\frac{\Delta p}{2} \frac{e^{\alpha(\omega_c+\omega_m)-j\phi(\omega_c+\omega_m)}}{e^{\alpha(\omega_c)-j\phi(\omega_c)}} e^{j(\omega_c+\omega_m)t} + j\,\frac{\Delta p}{2} \frac{e^{\alpha(\omega_c-\omega_m)-j\phi(\omega_c-\omega_m)}}{e^{\alpha(\omega_c)-j\phi(\omega_c)}} e^{j(\omega_c-\omega_m)t} \right] \tag{19}$$

In the general case, the output has both phase and amplitude modulations. The complex modulation amplitudes P and M can be determined using Equation (16).

$$P_2 = \left\{ \frac{e^{\alpha(\omega_c+\omega_m)-j\phi(\omega_c+\omega_m)}}{2e^{\alpha(\omega_c)-j\phi(\omega_c)}} + \frac{e^{\alpha(\omega_c-\omega_m)+j\phi(\omega_c-\omega_m)}}{2e^{\alpha(\omega_c)+j\phi(\omega_c)}} \right\} \Delta p \tag{20}$$

$$M_2 = \left[\frac{je^{\alpha(\omega_c+\omega_m)-j\phi(\omega_c+\omega_m)}}{2e^{\alpha(\omega_c)-j\phi(\omega_c)}} - \frac{je^{\alpha(\omega_c-\omega_m)+j\phi(\omega_c-\omega_m)}}{2e^{\alpha(\omega_c)+j\phi(\omega_c)}} \right] \Delta p \tag{21}$$

For evaluating the differential response contribution, an ideal PM demodulator is assumed at the output of the linear network, suppressing the amplitude modulation and reproducing the phase modulation component. The complex gain of the demodulated test-tone frequency component is given by the bracketed term in Equation (20). Substitution of the sweep variable (x) from Equation (8) into this term will yield the complex differential response as follows:

$$D(x) = \frac{e^{\alpha(\omega_o+x+\omega_m)-j\phi(\omega_o+x+\omega_m)}}{2e^{\alpha(\omega_o+x)-j\phi(\omega_o+x)}} + \frac{e^{\alpha(\omega_o+x-\omega_m)+j\phi(\omega_o+x-\omega_m)}}{2e^{\alpha(\omega_o+x)+j\phi(\omega_o+x)}} \tag{22}$$

From Equation (22), the differential gain and differential phase characteristics can be calculated by using the relationship given in Equation (6).

The expression derived for the complex differential response as given in Equation (22) can be simplified considerably. The differential characteristic is given by terms with differences from the normalized amplitude and phase responses in the exponentials. These differences can be approximately expressed using the first and second derivatives of the respective characteristics.

$$\alpha(\omega_o+x\pm\omega_m) - \alpha(\omega_o+x) \simeq \pm\alpha'(\omega_o+x)\omega_m + \alpha''(\omega_o+x)\frac{\omega_m^2}{2} \quad (23)$$

$$\phi(\omega_o+x\pm\omega_m) - \phi(\omega_o+x) \simeq \pm\phi'(\omega_o+x)\omega_m + \phi''(\omega_o+x)\frac{\omega_m^2}{2} \quad (24)$$

The approximations are valid when the test-tone frequency is not too high with respect to the curvature of the amplitude and phase responses. Using Equations (23) and (24), the expression for D(x) becomes:

$$D(x) = \frac{1}{2}\exp\{[\alpha'(x)-j\tau(x)]\omega_m+[\alpha''(x)-j\tau'(x)]\frac{\omega_m^2}{2}\}$$

$$+\frac{1}{2}\exp\{[-\alpha'(x)-j\tau(x)]\omega_m+[\alpha''(x)+j\tau'(x)]\frac{\omega_m^2}{2}\} \quad (25)$$

where $\tau(x)$ denotes the derivative of the phase, i.e., the group delay characteristic.

For practical systems with small transmission deviations in the passband, the arguments of the exponential terms are much smaller than unity. Hence,

$$D(x) \simeq 1 - j[\tau(x)]\omega_m + [\alpha''(x)]\frac{\omega_m^2}{2} \quad (26)$$

The differential gain and phase are calculated from Equation (6). Since the contributions from the second and third terms in Equation (26) are much smaller than unity, the absolute value and the phase of D(x) will be given by the real and imaginary parts respectively:

$$DG(x) = 1 + [\alpha''(x)] \frac{\omega_m^2}{2} \tag{27}$$

$$DP(x) = [\tau(x)]\omega_m \tag{28}$$

In the DG(x) expression, the term $\tau'^2(x) \frac{\omega_m^4}{8}$ should be added as this is often significant. The DG(x) term then becomes

$$DG(x) = 1 + [\alpha''(x)] \frac{\omega_m^2}{2} - \tau'^2(x) \frac{\omega_m^4}{8}. \tag{29}$$

Mathematical Model for a Linear Network Followed by an AM/PM Converter

In FM systems, modulation distortion is originated by two different mechanisms. The kind of distortion produced by linear networks has been discussed in the preceding paragraphs. There is, however, a second type of distortion resulting from the coupled effects of linear and non-linear circuits in the path of the phase modulated signal. This second type of distortion can be considered as a series of cascaded linear and non-linear circuits, as shown in Fig. 1. In the cascade, frequency dependence is assumed for the linear part only, whereas the non-linear circuit is taken to be independent of frequency within the sidebands of the modulated carrier. In practice, this assumption is usually justified.

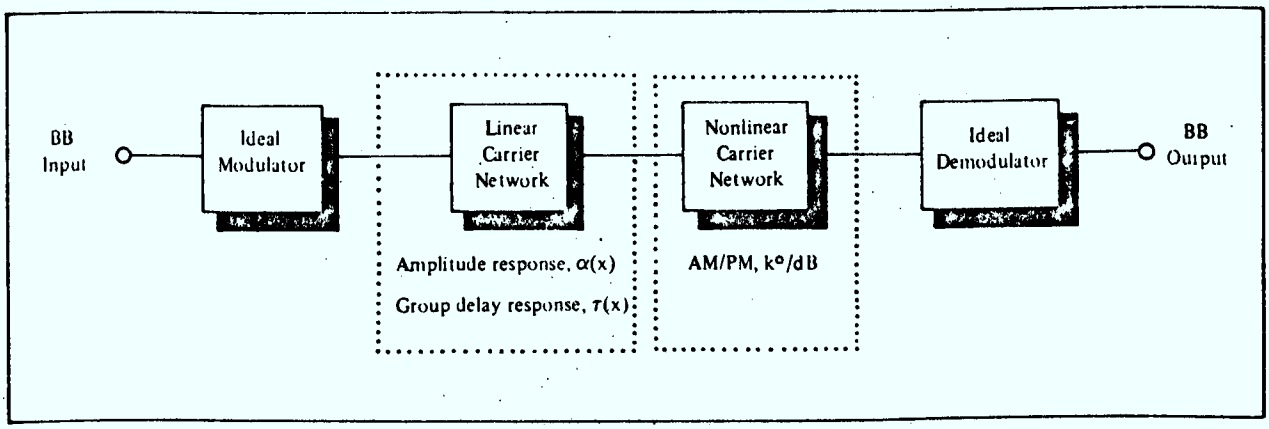


Fig. 1. Single cascaded linear/non-linear network.

A carrier signal with pure phase modulation will be transmitted by the non-linear network without distortion. As the non-linear circuit is assumed to be independent of frequency, the phase modulation index will be unchanged and there will be no conversion to amplitude modulation. However, when the non-linear network is exposed to amplitude modulation, two forms of distortion can be produced. First, the amplitude modulation index may be changed resulting in AM compression (denoted by γ). Second, the AM to PM conversion may occur whereby the AM may be partly converted into PM (denoted by k).

In practical circuits, both γ and k appear to be independent of the modulating frequency. Using the parameters γ and k, the modulation transmission through the non-linear circuit can be described in terms of the modulation amplitudes P and M. Denoting by P_2 and M_2 the input and by P_3 and M_3 the output modulation amplitudes, then:

$$P_3 = P_2 + k\, M_2 \tag{30}$$

$$M_3 = \gamma M_2 \tag{31}$$

The overall characteristics of the cascaded linear and non-linear circuits in Fig. 1 can be determined by substituting Equations (20) and (21) into Equations (30) and (31). Then for the complex amplitudes P_3 and M_3:

$$P_3 = \left[\frac{1+jk}{2}\,\frac{e^{\alpha(\omega_c+\omega_m)-j\phi(\omega_c+\omega_m)}}{e^{\alpha(\omega_c)-j\phi(\omega_c)}} + \frac{1-jk}{2}\,\frac{e^{\alpha(\omega_c-\omega_m)+j\phi(\omega_c-\omega_m)}}{e^{\alpha(\omega_c)+j\phi(\omega_c)}}\right]\Delta p \tag{32}$$

$$M_3 = \left[\frac{j\gamma}{2}\,\frac{e^{\alpha(\omega_c+\omega_m)-j\phi(\omega_c+\omega_m)}}{e^{\alpha(\omega_c)-j\phi(\omega_c)}} - \frac{j\gamma}{2}\,\frac{e^{\alpha(\omega_c-\omega_m)+j\phi(\omega_c-\omega_m)}}{e^{\alpha(\omega_c)+j\phi(\omega_c)}}\right]\Delta p \tag{33}$$

The complex differential response is given by the coefficient of Δp in Equation (32):

$$D(x) = \frac{1+jk}{2}\,\frac{e^{\alpha(\omega_o+x+\omega_m)-j\phi(\omega_o+x+\omega_m)}}{e^{\alpha(\omega_o+x)-j\phi(\omega_o+x)}} + \frac{1-jk}{2}\,\frac{e^{\alpha(\omega_o+x-\omega_m)+j\phi(\omega_o+x-\omega_m)}}{e^{\alpha(\omega_o+x)+j\phi(\omega_o+x)}} \tag{34}$$

The amplitude modulation given by M_3 has no significance if the output is connected directly to the demodulator which is assumed to be insensitive to AM. If the cascade is representing only a part of the FM system and is followed by similar cascades of linear and non-linear networks, then the amplitude modulation component has to be considered.

As in the case for a single linear network, the expression for the complex differential response for a linear network followed by an AM/PM converter can be simplified. Again, using Equations (23) and (24), the expression for D(x) becomes:

$$D(x) = \frac{1+jk}{2} \exp\{[\alpha'(x)-j\tau(x)]\omega_m+[\alpha''(x)-j\tau'(x)] \frac{\omega_m^2}{2}\}$$

$$+ \frac{1-jk}{2} \exp\{[-\alpha'(x)-j\tau(x)]\omega_m+[\alpha''(x)+j\tau'(x)] \frac{\omega_m^2}{2}\} \tag{35}$$

where $\tau(x)$ again denotes the derivative of the phase, i.e., the group delay characteristic.

For practical systems with small transmission deviations in the passband, the arguments of the exponential terms are much smaller than unity. Hence,

$$D(x) \simeq 1 - j[\tau(x)-k\alpha'(x)]\omega_m + [\alpha''(x)+k\tau'(x)] \frac{\omega_m^2}{2} \tag{36}$$

The differential gain and phase are calculated from Equation (6). Since the contributions from the second and third terms in Equation (36) are much smaller than unity, the absolute value and the phase of D(x) will be given by the real and imaginary parts respectively:

$$DG(x) = 1 + [\alpha''(x)+k\tau'(x)] \frac{\omega_m^2}{2} \tag{37}$$

$$DP(x) = [\tau(x) - k\alpha'(x)]\omega_m \tag{38}$$

In the DG(x) expression, the term $\tau'^2(x)\ \frac{\omega_m^4}{8}$ should be added as this is often significant. The DG(x) term then becomes:

$$DG(x) = 1 + [\alpha''(x) + k\tau'(x)]\ \frac{\omega_m^2}{2} - \tau'^2(x)\ \frac{\omega_m^4}{8} \tag{39}$$

Mathematical Model for a Cascaded Linear/Non-Linear Network

A general system configuration with multiple AM/PM converters is shown in Fig. 2. Each of the linear transmission networks in the system will produce some spurious PM and AM components from the input PM according to Equations (20) and (21). The spurious components are sufficiently low so that superposition can be applied. Therefore, the direct DG and DP contributions of the linear networks, being independent of k in Equations (37) and (38) will add at the output.

The AM component produced by a linear network will be affected by all the AM/PM converters placed after the linear network. However, the AM will not be the same at the input of each converter since it will be reduced by the AM compression factors of any circuits placed between the distortion source and the actual AM/PM converter. Therefore $keff_r$, the effective conversion factor of the non-linear circuit following the r-th linear network is given

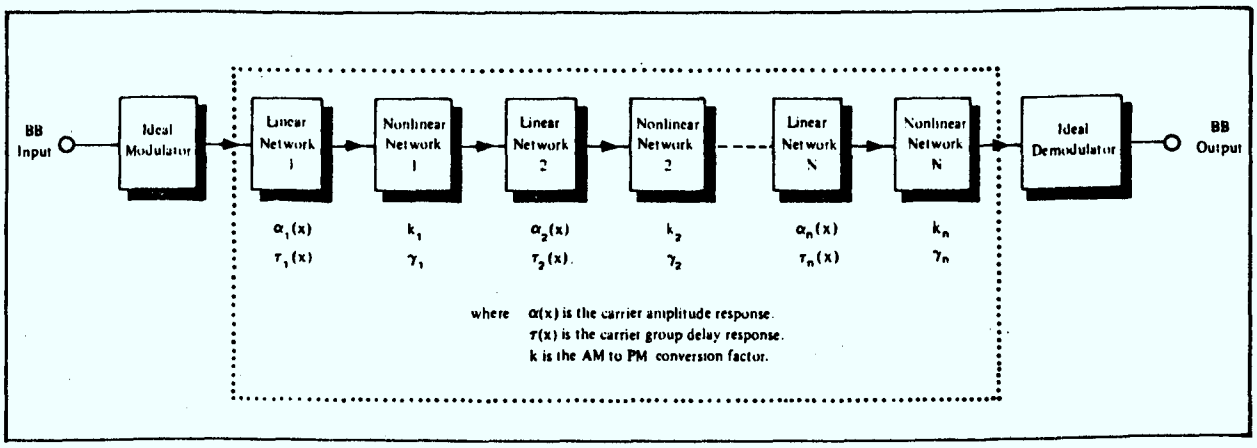


Fig. 2. Multiple cascaded linear/non-linear networks.

by the relationship:

$$keff_r = k_r + \gamma_r k_{r+1} + \gamma_r \gamma_{r+1} k_{r+2} + \ldots + \gamma_r \ldots \gamma_{n-1} k_n \ldots \tag{40}$$

where k_r and γ_r stand for the conversion and compression factors of the r-th non-linear network.

The AM produced by the linear networks is small, therefore each contribution can be calculated separately and added at the output.

Thus, for the overall DG and DP characteristics of the system in Fig. 2:

$$DG(x) = 1 + \sum_{r=1}^{n} [\alpha''_r(x) + keff_r \tau'_r(x)] \frac{\omega_m^2}{2} - \tau'^2(x) \frac{\omega_m^4}{8} \tag{41}$$

$$DP(x) = \sum_{r=1}^{n} [\tau_r(x) - keff_r \alpha'_r(x)] \omega_m \tag{42}$$

For the specific case of two cascaded pairs of linear/non-linear networks, we obtain:

$$DG(x) = 1 + [\alpha''_1(x) + keff_1 \tau'_1(x) + \alpha''_2(x) + keff_2 \tau'_2(x)] \frac{\omega_m^2}{2} - [\tau'^2_1(x) + \tau'^2_2(x)] \frac{\omega_m^4}{8} \tag{43}$$

and

$$DP(x) = [\tau_1(x) - keff_1\alpha_1'(x) + \tau_2(x) - keff_2\alpha_2'(x)]\omega_m \qquad (44)$$

Since the following relations hold true:

$$\alpha_1''(x) + \alpha_2''(x) = \alpha''_{total}(x)$$

$$\tau_1(x) + \tau_2(x) = \tau_{total}(x)$$

$$\tau_1'(x) + \tau_2'(x) = \tau'_{total}(x)$$

then the overall AM-to-PM distortion coefficient k_{total} can be determined by a comparison of Equations (42) and (38), and also by a comparison of Equations (43) and (37). These yield respectively:

$$keff_1\tau_1'(x) + keff_2\tau_2'(x) = k_{total}(\tau_1'(x) + \tau_2'(x))$$

and (45)

$$keff_1\alpha_1'(x) + keff_2\alpha_2'(x) = k_{total}(\alpha_1'(x) + \alpha_2'(x))$$

Generalization of the above deducation yields the conclusions that the model for a linear network followed by an AM/PM converter can be used for the overall performance of cascaded pairs of linear/non-linear networks irrespective of the contributions arising from the various pairs.

8.3 Identification of a Device's Linear/Non-Linear Network Representation and Extraction of Its AM-to-PM Distortion Coefficient

The equations which govern the differential gain/differential phase behaviour of a network are Equations (28) and (29) for a linear network, Equations (38) and (39) for a linear network followed by an AM/PM converter and Equations (43) and (44) for two cascaded pairs of linear/non-linear networks. We can see from these equations that the DP(x) and DG(x) vary with the modulating radian frequency as

$$DG = a + b\,\omega_m^2 + c\,\omega_m^4 \tag{46}$$

and

$$DP = d + e\omega_m \tag{47}$$

These relationships hold for all three possible networks, and at all carrier frequencies, x. The difference in the above polynomials lies in the circuit dependence of the coefficients as shown below.

$$a = 1 \quad ; \text{ all cases} \tag{48}$$

$$b = \begin{cases} \frac{1}{2}\,\alpha''(x) & ; \text{ single linear network} \\ \frac{1}{2}\,[\alpha''(x)+k\tau'(x)] & ; \text{ one pair of linear/non-linear networks} \\ \frac{1}{2}\,[\alpha''_{total}(x) + k_{total}\tau'_{total}(x)] & ; \text{ two cascaded pairs of linear/non-linear networks} \end{cases} \tag{49}$$

$$c = -\frac{\tau'^2(x)}{8} \quad ; \text{ all cases with the subscript applied for the two cascaded pairs} \tag{50}$$

$$d = 0 \quad ; \text{ all cases} \tag{51}$$

$$e = \begin{cases} \tau(x) & ; \text{ single linear network} \\ \tau(x) - k\alpha'(x) & ; \text{ one pair of linear/non-linear networks} \\ \tau_{total}(x) - k_{total}\alpha'_{total}(x) & ; \text{ two cascaded pairs of linear/non-linear networks} \end{cases} \tag{52}$$

The procedure for determining the AM-to-PM distortion coefficient consists of performing the differential gain/differential phase measurements using the Microwave Link Analyzer at the desired centre frequency x_o and using several available modulating frequencies, ω_m such as 92 kHz, 277 kHz, 555 kHz, 2.40 MHz, 3.58 MHz, 5.60 MHz and 8.20 MHz.

A quadratic minimum squared error curve fit is then applied to the data for the differential gain measurements at the various values of ω_m^2, in order to determine the coefficients a, b and c.

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \frac{1}{|S_{ij}|} \begin{bmatrix} C_{ij} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \tag{53}$$

where

$$t_1 = \sum_{i=1}^{N} DG_i$$

$$t_2 = \sum_{i=1}^{N} (DG_i)(\omega_{m_i})^2$$

$$t_3 = \sum_{i=1}^{N} (DG_i)(\omega_{m_i})^4$$

and

$$S_{11} = N$$

$$S_{12} = S_{21} = \sum_{i=1}^{N} (\omega_{m_i})^2$$

$$S_{13} = S_{22} = S_{31} = \sum_{i=1}^{N} (\omega_{m_i})^4$$

$$S_{23} = S_{32} = \sum_{i=1}^{N} (\omega_{m_i})^6$$

$$S_{33} = \sum_{i=1}^{N} (\omega_{m_i})^8$$

and C_{ij} are the cofactors of S_{ij}.

The summation is taken over the number of modulating angular frequencies used to measure the differential gain and phase.

Similarly, a linear minimum squared error curve fit is applied to the data for the differential phase measurements at various values of ω_m, in order to determine the

coefficients d and e.

$$\begin{bmatrix} d \\ e \end{bmatrix} = \frac{1}{|T_{ij}|} \begin{bmatrix} D_{ij} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{54}$$

where

$$u_1 = \sum_{i=1}^{N} DP_i$$

$$u_2 = \sum_{i=1}^{N} (DP_i)(\omega_{m_i})$$

and

$$T_{11} = N$$

$$T_{12} = T_{21} = \sum_{i=1}^{N} \omega_{m_i}$$

$$T_{22} = \sum_{i=1}^{N} (\omega_{m_i})^2$$

and D_{ij} are the cofactors of T_{ij}.

In addition to the coefficients a through e, values for the slope of the amplitude characteristic, $\alpha'(x_o)$, and the group delay characteristic $\tau(x_o)$ must be determined before the AM-to-PM distortion constant k can be found using Equation (52). Note that if the coefficient e is identically equal to the value of τ, then k = 0 and the network measured is a single linear network.

Similarly, if $\alpha''(x_o)$ and $\tau'(x_o)$ are determined, then the AM-to-PM distortion constant k can be found using Equation (49). Also if the coefficient b is identically equal to the value of $\alpha''(x_o)$, then k = 0 and the measured network is a single linear network.

One technique to evaluate $\alpha'(x_o)$, $\tau(x_o)$ or $\alpha''(x_o)$, $\tau'(x_o)$ is to use the microwave link analyzer to measure the amplitude response and group delay distortion at the desired centre frequency x_o. Using this procedure, it is obvious that more accurate estimates for $\alpha'(x_o)$, $\tau(x_o)$ are possible than for $\alpha''(x_o)$, $\tau'(x_o)$ and hence Equation (52) should be used to calculate k.

A second and possibly more accurate technique is based on manipulating the set of Equations (48) through (52). Each of these equations can be written in a single form irrespective of the network it represents. Hence:

$$a_o = 1 \tag{48}$$

$$b_o = \frac{1}{2}\,[\alpha''(x_o)+k\tau'(x_o)] \tag{49}$$

$$c_o = -\frac{\tau'^2(x_o)}{8} \tag{50}$$

$$d_o = 0 \tag{51}$$

$$e_o = \tau(x_o) - k\alpha'(x_o) \tag{52}$$

Here the subscript o indicates performance at the centre frequency x_o.

It should be noted that for a single linear network k will be equal to zero and Equations (49) and (52) revert to their previous forms. Also, for two cascaded pairs of linear/non-linear networks, all that must be done is to re-install the subscript "total" in Equations (49) and (52). Further, it should be noted that since there is no possibility of determining the individual coefficients (α_1, τ_1, k_1, α_2, τ_2, k_2) there is no longer any point in making the distinction between one pair of linear/non-linear networks and cascaded pairs.

The second procedure is as follows. The differential phase measurement is repeated at a centre frequency x_1 slightly removed from x_o, and new coefficients d_1 and e_1 are calculated. From Equation (52), we deduce

$$e_1 = \tau(x_1) - k\alpha'(x_1) \tag{53}$$

Subtracting Equation (52) from the above, and dividing through by $x_1 - x_o$ yields

$$\frac{e_1-e_o}{x_1-x_o} = \frac{\tau(x_1)-\tau(x_o)}{x_1-x_o} - \frac{k\alpha'(x_1)-k\alpha'(x_o)}{x_1-x_o} \tag{54}$$

Letting the right-hand side of Equation (54) approach the limit, while the left-hand side remains a finite difference, yields:

$$\frac{\Delta e_o}{\Delta x} \simeq \tau'(x_o) - k\alpha''(x_o) \tag{55}$$

Combining (55) with (49) to remove $\alpha''(x_o)$ yields:

$$b_o = \frac{1}{2}\,[\tau'(x)[k + \frac{1}{k}] - \frac{1}{k}\,\frac{\Delta e_o}{\Delta x}] \tag{56}$$

Inserting Equation (50) to remove $\tau'(x)$ yields:

$$\sqrt{-8c_o}\;\;[\frac{k^2+1}{k}] - \frac{1}{k}\,\frac{\Delta e_o}{\Delta x} - 2b_o = 0 \tag{57}$$

Rearranging this expression produces a quadratic equation:

$$k^2 - \left[\frac{b_o}{\sqrt{-2c_o}}\right]k + \left[1 - \frac{1}{\sqrt{-8c_o}}\,\frac{\Delta e_o}{\Delta x}\right] = 0 \tag{58}$$

The solution for this completes the algebraic manipulations.

$$k = \frac{b_o}{2\sqrt{-2c_o}} \pm \sqrt{\frac{b_o^2}{-8c_o} + \frac{1}{2\sqrt{-2c_o}}\,\frac{\Delta e_o}{\Delta x} - 1} \tag{59}$$

where

$$\frac{\Delta e_o}{\Delta x} = \frac{e_1 - e_o}{x_1 - x_o}$$

It should be noted that to determine the AM-to-PM conversion factor k, all that is required is the determination of the polynomial coefficients b, c and e at the centre frequency x_o, and the polynomial coefficient e at a frequency x_1 slightly removed from x_o.

8.4 Practical Discussions

It is difficult to estimate which of the two approaches (that of measuring $\alpha'(x_o)$ and $\tau(x_o)$ or that of using Equation (50)) will provide the best accuracy without undertaking experimental testing. It should be noted that both techniques appear to be experimentally simple but require some involved calculations. However, these calculations can easily be handled with a programmable hand-held calculator having approximately 200 lines of program memory.

With both of the above approaches, it should be a relatively quick measurement to determine the bias level dependence of the AM-to-PM distortion constant, k. It becomes a bit more involved to determine the input signal level dependence of k when the dynamic range limits of the link analyzer are exceeded. The amplifier-under-test must first be embedded in a chain consisting of a characterized amplifier at its input and an attenuator at its output. The characterized amplifier boosts the power level from the link analyzer to the desired input level of the amplifier-under-test, while the attenuator drops the output level from the amplifier-under-test to a value at which the link analyzer receiver operates. This chain of circuits is shown in Fig. 3.

From Fig. 3, we see that we now have a pair of cascaded linear/non-linear networks representing the two

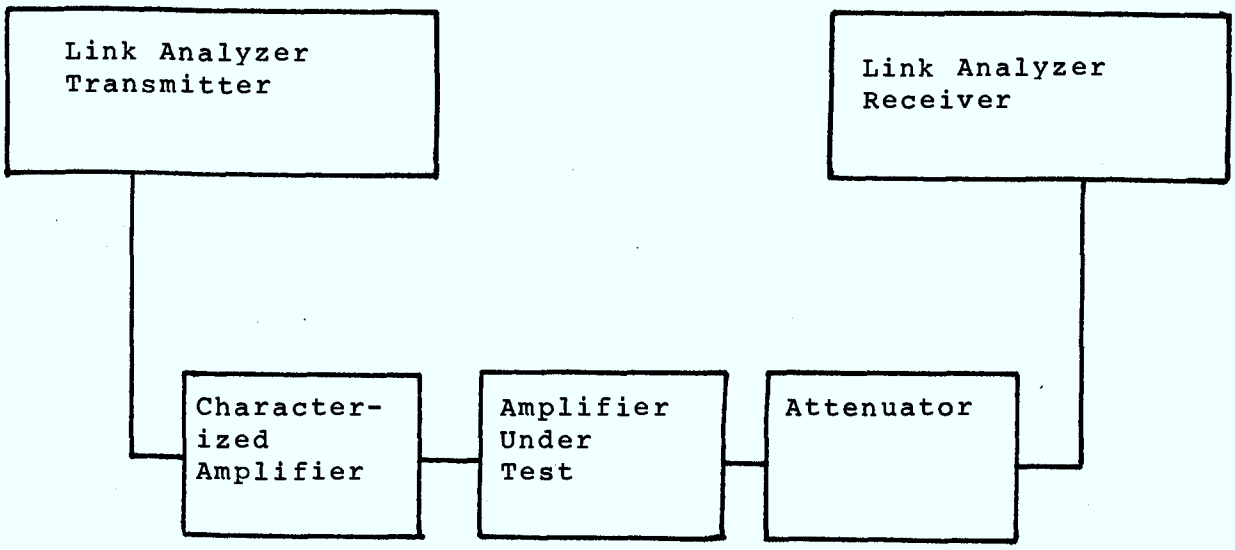


Fig. 3. Extended dynamic range measurement.

amplifiers. The attenuator is assumed to be frequency independent and to be perfectly linear. Use must now be made of Equations (43) and (44) for two cascaded pairs of linear/non-linear networks where the subscript 1 represents the first amplifier and the subscript 2 represents the second. For this extended range test, the first amplifier must be characterized alone before the cascaded amplifiers are tested together. Hence, all factors with a subscript 1 will be known.

Writing the expressions for the coefficients a through e for the differential gain/differential phase test on the two cascaded amplifiers yields:

$$a_o = 1 \tag{60}$$

$$b_o = \frac{1}{2}\,[2B_o+\alpha_2''(x_o) + k_2\tau_2'(x_o)] \tag{61}$$

$$c_o = -\frac{1}{8}\,[-8C_o+\tau_2'^{\,2}(x_o)] \tag{62}$$

$$d_o = 0 \tag{63}$$

$$e_o = [E_o+\tau_2(x_o)-k_2\alpha_2'(x_o)] \tag{64}$$

where the constants B_o, C_o, E_o are determined by the first amplifier performance, and are in fact just the polynomial coefficients obtained for the first amplifier alone.

$$B_o = \frac{1}{2}\,[\alpha_1''(x_o)+k_1\tau_1'(x_o)] \tag{65}$$

$$C_o = -\frac{\tau_1'(x_o)}{8} \tag{66}$$

$$E_o = \tau_1(x_o) - k_1\alpha_1'(x_o) \tag{67}$$

We see that Equations (60) through (64) are similar to the expressions in Equations (48) through (52) and consequently the techniques described in the preceding section apply here as well.

In specific, the procedure which led to Equation (59) for the AM-to-PM distortion constant k now leads to:

$$k_2 = -\left[\frac{B_o-b_o}{2\sqrt{2C_o-2c_o}}\right] \pm \sqrt{\frac{(B_o-b_o)^2}{8(C_o-c_o)} - \frac{(E_1-E_o-\frac{\Delta e_o}{\Delta x})}{2\sqrt{2C_o-2c_o}} - 1} \tag{68}$$

where

$$\frac{\Delta e_o}{\Delta x} = \frac{e_1-e_o}{x_1-x_o}$$

It should be recalled that B_o, C_o and E_o are the polynomial coefficients from the experimental data on the first amplifier alone, at frequency x_o. Also E_1 is the polynomial coefficient from the experimental data on the first amplifier alone at frequency x_1 near x_o. Similarly, b_o, c_o and e_o are the polynomial coefficients from the experimental data on the two cascaded amplifiers at frequency x_o.

Finally, e_1 is the polynomial coefficient from the experimental data on the two cascaded amplifiers at frequency x_1 near x_o.

Using the above expression in Equation (68) allows the calculation of the AM-to-PM distortion constant k at power levels exceeding the normal operating levels of the microwave link analyzer. Hence, it allows the experimental determination of input signal level dependence of k on microwave high power amplifiers.

9. CONCLUSIONS

The following concluding remarks concerning the investigations undertaken during this study can be made. A tutorial review of microwave field effect and bipolar transistors has been compiled. Mathematical analyses of AM-to-PM distortion and third-order intermodulation distortion have been outlined for both field effect and bipolar transistors, and using both power series and Volterra series procedures. A brief relationship between device doping profile and non-linear distortion for field effect transistors has been presented.

The relationship between AM-to-PM distortion and differential gain/differential phase for cascaded linear/non-linear networks has been outlined. Based on this relationship, a procedure for extracting the AM-to-PM distortion constant k from measured differential gain/differential phase values has been developed. This procedure is based on repeating the differential gain/differential phase measurements at two closely-related center frequencies, using several modulating frequencies, and incorporating minimum squared error curve fitting. Finally, a related procedure for extracting the AM-to-PM distortion constant k for a device which must be driven beyond the dynamic range capabilities of the microwave link analyzer has been presented.

9.1 Future Work

The first obvious task would be to provide experimental verification of the techniques developed in Part III for determining the AM-to-PM distortion constant. Verification would be provided by a comparison of the value obtained using this procedure with that obtained based on the non-linear analyses of Chapters 3, 4 and 7. It should be noted that the power series expansions required for these non-linear analyses are not easily obtainable.

Once verification is established, a technically very rewarding study would be a comparison of the AM-to-PM performances of devices manufactured by several industrial sources, which exhibit comparable performance in terms of maximum power and maximum frequency. This study could lead to conclusions regarding which device designs and which fabrication processes are superior.

REFERENCES

1. Shockley, "A Unipolar Field Effect Transistor", Proc. IRE, Vol. 40, pp. 1365-1376, Nov 1952.

2. Mead, "Schottky Barrier Gate Field Effect Transistor", Proc. IEEE, Vol. 54, pp. 307-308, Feb 1966.

3. R.S. Tucker, "Third-Order Intermodulation Distortion and Gain Compression in GaAs FETs", IEEE Trans., MTT, Vol. 27, pp. 400-408, May 1979.

4. G.L. Heifer, "Characterization of Non-linearities in Microwave Devices and Systems", IEEE Trans., MTT, Vol. 21, pp. 797-805, Dec 1973.

5. R.A. Minasian, "Intermodulation Distortion Analysis of MESFET Amplifiers Using the Volterra Series Representation", IEEE Trans., MTT, Vol. 28, pp. 1-8, Jan 1980.

6. R.A. Pucel, H.A. Hans, and H. Statz, "Signal and Noise Properties of Gallium Arsenide Microwave Field Effect Transistors", in Advances Electron and Electron. Physics, Vol. 38, New York: Academic Press, 1975, pp. 195-265.

7. J.A. Higgins, R.L. Kuvas, "Analysis and Improvement of Intermodulation Distortion in GaAs Power FETs", IEEE Trans., MTT, Vol. 28, pp. 9-17, Jan 1980.

8. Early, "Structure-Determined Gain-Based Product of Junction Triode Transistors", Proc.IRE, Vol. 46, pp. 1924-1927, 1958.

9. H.F. Cooke, "Microwave Transistors: Theory and Design", Proc. IEEE, Vol. 59, pp. 1163-1181, Aug 1971.

10. Pritchard, "Frequency Response of Grounded-Base and Grounded-Emitter Transistors", AIEE Winter Meeting, New York, 1954.

11. Johnson, "Physical Limitations on Frequency and Power Parameters of Transistors", RCA Review, Vol. 26, pp. 163-177, June 1965.

12. S. Narayanon, H.C. Poon, "An Analysis of Distortion in Bipolar Transistors Using Integral Charge Control Model and Volterra Series", IEEE Trans. Circuit Theory, Vol. 20, pp. 341-351, Jul 1973.

REFERENCES

1. Shockley, "A Unipolar Field Effect Transistor", Proc. IRE, Vol. 40, pp. 1365-1376, Nov 1952.

2. Mead, "Schottky Barrier Gate Field Effect Transistor", Proc. IEEE, Vol. 54, pp. 307-308, Feb 1966.

3. R.S. Tucker, "Third-Order Intermodulation Distortion and Gain Compression in GaAs FETs", IEEE Trans., MTT, Vol. 27, pp. 400-408, May 1979.

4. G.L. Heifer, "Characterization of Non-linearities in Microwave Devices and Systems", IEEE Trans., MTT, Vol. 21, pp. 797-805, Dec 1973.

5. R.A. Minasian, "Intermodulation Distortion Analysis of MESFET Amplifiers Using the Volterra Series Representation", IEEE Trans., MTT, Vol. 28, pp. 1-8, Jan 1980.

6. R.A. Pucel, H.A. Hans, and H. Statz, "Signal and Noise Properties of Gallium Arsenide Microwave Field Effect Transistors", in Advances Electron and Electron. Physics, Vol. 38, New York: Academic Press, 1975, pp. 195-265.

7. J.A. Higgins, R.L. Kuvas, "Analysis and Improvement of Intermodulation Distortion in GaAs Power FETs", IEEE Trans., MTT, Vol. 28, pp. 9-17, Jan 1980.

8. Early, "Structure-Determined Gain-Based Product of Junction Triode Transistors", Proc.IRE, Vol. 46, pp. 1924-1927, 1958.

9. H.F. Cooke, "Microwave Transistors: Theory and Design", Proc. IEEE, Vol. 59, pp. 1163-1181, Aug 1971.

10. Pritchard, "Frequency Response of Grounded-Base and Grounded-Emitter Transistors", AIEE Winter Meeting, New York, 1954.

11. Johnson, "Physical Limitations on Frequency and Power Parameters of Transistors", RCA Review, Vol. 26, pp. 163-177, June 1965.

12. S. Narayanon, H.C. Poon, "An Analysis of Distortion in Bipolar Transistors Using Integral Charge Control Model and Volterra Series", IEEE Trans. Circuit Theory, Vol. 20, pp. 341-351, Jul 1973.

13. "Differential Phase and Gain at Work", Hewlett-Packard Application Note 175-1, Nov 1975.

14. "H50 Microwave Link Analyzer", Hewlett-Packard Operating Manual, Nov 1973.

15. "MLA Demonstrations", Hewlett-Packard Application Note, Aug 1972.

BIBLIOGRAPHY

Field Effect Transistor Non-Linearities

1. R.S. Tucker, "Third Order Intermodulation Distortion and Gain Compression in GaAs FETs", IEEE Trans. MTT, Vol. 27, pp. 400-408, May 1979.

2. R.A. Minasian, "Intermodulation Distortion Analysis of MESFET Amplifiers Using the Volterra Series Representation", IEEE Trans. MTT, Vol. 28, pp. 1-8, Jan 1980.

3. J.A. Higgins and R.L. Kuvas, "Analysis and Improvement of Intermodulation Distortion in GaAs Power FETs", IEEE Trans. MTT, Vol. 28, pp. 9-17, Jan 1980.

4. C. Rouscher and H.A. Willing, "Simulation of Non-Linear Microwave FET Performance Using a Quasi-Static Model", IEEE Trans. MTT, Vol. 27, pp. 834-840, Oct 1979.

5. R.E. Williams and D.W. Shaw, "Graded Channel FETs: Improved Linearity and Noise Figure", IEEE Trans. Electron Devices, Vol. 25, pp. 600-605, June 1978.

6. H.Q. Tserng, V. Sokolov, H.M. Mackey, W.R. Wisseman, "Microwave Power GaAs FET Amplifiers", IEEE Trans. MTT, Vol. 24, pp. 936-943, Dec 1976.

7. A.M. Khadr and R.H. Johnston, "Distortion in High Frequency FET Amplifiers", IEEE Journal of Solid State Circuits, Vol. 9, pp. 180-189, Aug 1974.

8. D.A. Cowan, P. Mercer, A.B. Bell, "Low-Noise and Linear FET Amplifiers for Satellite Communications", IEEE Trans. MTT, Vol. 25, pp. 995-1000, Dec 1977.

9. R.S. Tucker and C. Rauscher, "Modelling the Third-Order Intermodulation-Distortion Properties of a GaAs FET", Electron. Letters, Vol. 13, pp. 508-510, Aug 18, 1977.

10. R.A. Minasian, "Large Signal GaAs MESFET Model and Distortion Analysis", Electron. Letters, Vol. 14, pp. 183-185, March 16, 1978.

11. R.A. Minasian, "Simplified GaAs MESFET Model to 10 GHz", Electron. Letters, Vol. 13, pp. 549-551, Sept 1, 1977.

12. R.A. Pucel, "Profile Design for Distortion Reduction in Microwave Field Effect Transistors", Electron. Letters, Vol. 14, pp. 204-206, March 16, 1978.

13. C. Rauscher and R.S. Tuckér, "Method for Measuring Third-Order Intermodulation Distortion in GaAs FETs", Electron. Letters, Vol. 13, pp. 701-702, Nov 10, 1977.

14. H.A. Willing, C. Rauscher, P. de Santis, "A Technique for Predicting Large Signal Performance of a GaAs MESFET", IEEE Int. Microwave Symp., pp. 132-134, June 1978.

15. E.W. Strid and T.C. Duder, "Intermodulation Distortion Behaviour of GaAs Power FETs", IEEE Int. Microwave Symp., pp. 135-137, June 1978.

16. J.A. Higgins, "Intermodulation Distortion in GaAs FETs", IEEE Int. Microwave Symp., pp. 138-141, June 1978.

Bipolar Transistor Non-Linearities

1. S. Narayanan and H.C. Poon, "An Analysis of Distortion in Bipolar Transistors Using Integral Charge Control Model and Volterra Series", IEEE Trans. Circuit Theory, Vol. 20, pp. 341-351, Jul 1973.

2. Y.L. Kuo, "Distortion Analysis of Bipolar Transistor Circuits", IEEE Trans. Circuit Theory, Vol. 20, pp. 709-716, Nov 1973.

System Non-Linearities

1. G.L. Heiter, "Characterization of Non-Linearities in Microwave Devices and Systems", IEEE Trans. MTT, Vol. 21, pp. 797-805, Dec 1973.

2. J.J. Bussgang, L. Ehrman, and J.W. Graham, "Analysis of Non-Linear Systems with Multiple Inputs", Proc. IEEE, Vol. 62, pp. 1088-1119, Aug 1974.

3. E. Bedrosian and S.O. Rice, "The Output Properties of Volterra Systems (Non-Linear Systems with Memory) Driven by Harmonic and Gaussian Inputs", Proc. IEEE, Vol. 59, pp. 1688-1707, Dec 1971.

4. S.M. Perlow, "Third-Order Distortion in Amplifiers and Mixers", RCA Review, Vol. 37, pp. 234-266, June 1976.

5. "Differential Phase and Gain at Work", Hewlett-Packard Application Note 175-1, Nov 1975.

6. R.G. Meyer, M.J. Shensa and R. Eschenbach, "Cross Modulation and Intermodulation in Amplifiers at High Frequencies", IEEE Journal of Solid State Circuits, Vol. 7, pp. 16-23, Feb 1972.

7. S. Narayanan, "Transistor Distortion Analysis Using Volterra Series Representation", Bell Syst. Tech. Journal, pp. 991-1024, May-June 1967.

8. N.M. Blackman, "The Effect of a Large Signal Upon a Small Signal in a Memoryless Non-Linear Bandpass Amplifier", IEEE Trans. Communications, Vol. 29, pp. 72-73, Jan 1981.

9. P.Y. Chang and O. Shimbo, "Effects of One Large Signal on Small Signals in a Memoryless Non-Linear Bandpass Amplifier", IEEE Trans. Communications, Vol. 28, pp. 739-743, May 1980.

10. C.M. Chie, "A Modified Barrett-Lampard Expansion and its Application to Bandpass Non-Linearities with Both AM-AM and AM-PM Conversion", IEEE Trans. Communications, Vol. 28, pp. 1859-1866.

11. H. Kusaka and F. Nishida, "More on Distortion in FDM-FM Transmission Systems", IEEE Trans. Communications, Vol. 25, pp. 1227-1233, Oct 1977.

Microwave Transistors

1. Shockley, "A Unipolar Field Effect Transistor", Proc. IRE, Vol. 40, pp. 1365-1376, Nov 1952.

2. Mead, "Schottky Barrier Gate Field Effect Transistor", Proc. IEEE, Vol. 54, pp. 307-308, Feb 1966.

3. Early, "Structure-Determined Gain-Band Product of Junction Triode Transistors", Proc. IRE, Vol. 46, pp. 1924-1927, 1958.

4. H.F. Cook, "Microwave Transistors: Theory and Design", Proc. IEEE, Vol. 59, pp. 1163-1181, Aug 1971.

5. Marshak, "Optimum Doping Distribution for Minimum Base Transit Time", IEEE Trans. Electron Devices, Vol. 14, pp. 190-194, 1967.

6. Pritchard, "Electrical Characteristics of Transistors", McGraw-Hill, 1967.

7. Johnson, "Physical Limitations on Frequency and Power Parameters of Transistors", RCA Review, Vol. 26, pp. 163-177, June 1965.

WIGHT, J.S.
Performance characterization of AM-to- PM conversion in microwave solid state amplifiers

DATE DUE
DATE DE RETOUR

SEP	07 1984		