A New Wide-Band Amplifier Technique

BARRIE GILBERT, MEMBER, IEEE

Abstract—**Precision dc-coupled amplifiers having risetimes of less than a nanosecond have recently been fabricated using the monolithic planar process. The design is based on a simple technique that has a broad range of applications and is characterized by a stage gain accurately determined by the ratio of two currents, a stage-gain–bandwidth product essentially equal to that of the transistors, and a very linear transfer characteristic, free from temperature dependence.**

I. Introduction

AMPLIFIERS having wide bandwidths, in excess of 100 MHz, with low amplitude distortion and stable gain, are of increasing importance in modern electronics. By far the most prevalent circuit configuration is the emitter-degenerated amplifier, which suffers a familiar problem: the current- and temperature-dependent emitter impedance causes distortion [1], [2] and gain instability. Negative feedback techniques over more than two stages are rarely applicable at these bandwidths due to the excess phase shift around the loop.

A further characteristic of such amplifiers is that voltage swings at the signal frequencies are present throughout the chain, and, consequently, the parasitic capacitances have a part in limiting the bandwidth. This problem is especially severe in monolithic (junction-isolated) circuits. In circuits using discrete components, inductive elements are often employed [3] to improve the bandwidth, but no satisfactory way of incorporating inductive elements in a monolithic circuit is available. Another approach is to make these parasitic capacitances part of a lumped constant delay line (the distributed amplifier [4]), but this, too, is not applicable to microcircuits. Interestingly, however, some of the amplifiers described later behave very much like distributed amplifiers in that each stage works at an f_t-limited bandwidth, and all stages contribute to the total output capability.

Suggestions have been made from time to time [5]–[7] to include compensating diodes in the collector load circuit to mitigate the nonlinearities introduced by the emitter diode, or even eliminate the linear impedance at this point altogether [7]. This is a useful technique, but has several limitations, and the motivation is still to produce a stage having *voltage gain*.

The main objective of the work reported here was to develop a cascadable circuit form (a "gain cell") that could provide dc-coupled temperature-insensitive sub-nanosecond *current gain* with the virtual absence of voltage swings, and a theoretically perfect transfer characteristic, having constant slope between the upper and lower overload limits. Conversion between voltage and current can then be made only where needed—at the input and output terminals.

It became apparent that, in addition to meeting these goals, the principle developed had several other useful properties. For example, the gain of each stage can be electronically controlled with precision over a wide range at nanosecond speeds, a property exploited in a new four-quadrant multiplier [8], [9]. The analysis also shows that the stage gain of certain configurations is independent of beta, even at gains *close to beta* in magnitude. This is something that, to the author's knowledge, no other configuration permits.

II. Foundations of the Technique

Two very common circuits, shown in Fig. 1, were married to produce a new configuration. The first, (a), is the "differential pair," now widely used as a multiplier [10]–[13]. The second, (b), is the "current source" found in practically every linear IC. The married couple are shown in Fig. 3; this marriage has proven unexpectedly fruitful, and the prolific offspring will be described in this and later papers.

First, we will consider Fig. 1(a) in more detail. It can be used as a multiplier because the transconductance from base to collector is *proportional to the emitter tail current* I_E. However, it is far from being a precise element, because this transconductance is both nonlinear and temperature dependent.

In this paper, we will frequently use the junction-diode expressions [14].

$$I = I_s \exp \frac{qV}{mkT} \quad \text{or} \quad V = \frac{mkT}{q} \log \frac{I}{I_s} \tag{1}$$

where

I = forward conduction current, $\gg I_s$
I_s = reverse saturation current
V = voltage across the junction
q = charge on the electron
m = a constant near unity [14]
T = absolute temperature.

The quantity mkT/q is about 26 mV at 300°K. Applying (1) to Fig. 1(a), and solving for the variable

Manuscript received June 28, 1968; revised September 16, 1968. This paper was presented at the 1968 ISSCC.
The author is with Tektronix, Inc., Beaverton, Ore.

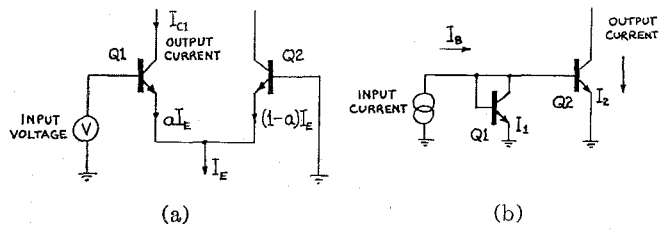


Fig. 1. Two common circuits. (a) The "differential amplifier." (b) The "current source."

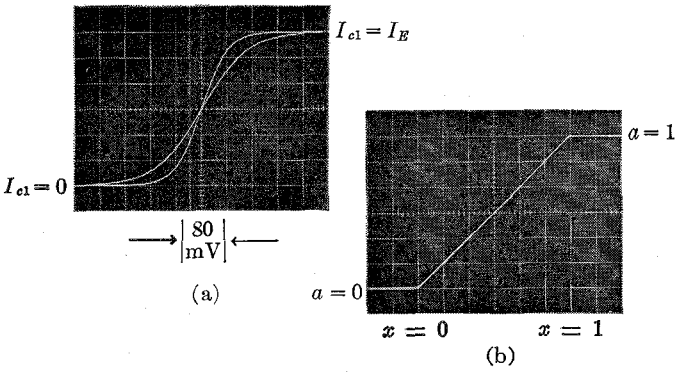


Fig. 2. Comparison of transfer curves for (a) Fig. 1(a) amplifier; (b) new circuit shown in Fig. 3. In each case, the curves for −50°C and +125°C are compared; in (b), they are virtually co-incident.

a we find

$$a = \frac{\exp \frac{qV}{mkT}}{\lambda + \exp \frac{qV}{mkT}}, \tag{2}$$

independent of I_E, where

$$\lambda = \frac{I_{s2}}{I_{s1}} = \frac{\text{area of } Q2 \text{ emitter}}{\text{area of } Q1 \text{ emitter}}, \tag{3}$$

and I_{s1}, and I_{s2} are the reverse saturation currents of the emitter diodes of $Q1$ and $Q2$. This area ratio is more usually expressed as an equivalent offset voltage,

$$V_0 = \frac{mkT}{q} \log \lambda, \quad \text{or} \quad \lambda = \exp \frac{qV_0}{mkT}. \tag{4}$$

Thus, in terms of V_0, we have

$$I_{c1} \simeq aI_E = \frac{I_E}{1 + \exp \frac{q}{mkT}(V_0 - V)}. \tag{5}$$

Characteristic transfer curves are shown in Fig. 2(a) for two temperatures. The transconductance is clearly a nonlinear, temperature-dependent quantity. Notice that an emitter-area mismatch merely shifts the transfer curve by an amount V_0, *without changing its shape*. This matter of area mismatch will be raised again in connection with the improved circuits, where the effects are very different.

The circuit of Fig. 1(b) can also be used as an amplifier, in which case the signal input is the *current* I_B. With reference to the figure it will be apparent that

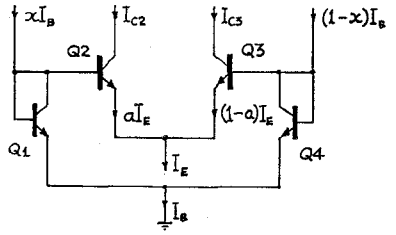


Fig. 3. The new circuit form. The text shows that the factors a and x are equal, for any I_E and at all temperatures.

$$\frac{mkT}{q} \log \frac{I_1}{I_{s1}} = \frac{mkT}{q} \log \frac{I_2}{I_{s2}} \tag{6}$$

or

$$I_2 = \frac{I_{s2}}{I_{s1}} \cdot I_1 = \lambda I_1. \tag{7}$$

This is a linear, temperature-insensitive relationship. By scaling the area of $Q2$ emitter relative to that of $Q1$, a stage gain can be realized; now *emitter-area mismatches cause a gain-error,* but do not jeopardize linearity.

Amplifiers built from directly cascaded stages like this have been successfully tested, and have the same high-speed properties as the circuits to be described—there is very little voltage swing, and, hence, f_t is the dominant factor in determining bandwidth. However, they lack the advantages of the differential configuration and have a fixed gain.

III. The Improved Circuit Principle

We can now examine Fig. 3, which can be considered as a differential amplifier in which the base-drive voltages are derived from a pair of junctions that, like the diode of Fig. 1(b), are current-driven. However, in this case there are two drive currents (the signal input), which are of the form

$$\begin{aligned} I_{B1} &= xI_B \\ I_{B2} &= (1 - x)I_B \end{aligned} \tag{8}$$

where $0 < x < 1$, and will be termed the "modulation index" of the bias current I_B.

Temporarily ignoring the effects of junction area differences, finite beta and ohmic resistances, and summing the emitter voltages around the $Q1$–$Q4$ loop, we have

$$\frac{mkT}{q} \log \frac{xI_B}{I_s} - \frac{mkT}{q} \log \frac{aI_E}{I_s} - \frac{mkT}{q} \log \frac{(1 - x)I_B}{I_s} + \frac{mkT}{q} \log \frac{(1 - a)I_E}{I_s} = 0 \tag{9}$$

which collapses to

$$a = x. \tag{10}$$

Thus, the magnitudes of the output currents are simply

$$\begin{aligned} I_{C2} &= xI_E \\ I_{C3} &= (1 - x)I_E, \end{aligned} \tag{11}$$

and the dc stage gain is

$$G_0 = \frac{I_{C2}}{I_{B1}} = \frac{I_E}{I_B}. \tag{12}$$

Typical transfer curves for this circuit are shown in Fig. 2(b).

Notice the extreme insensitivity to temperature and the sharp overload points, making the whole dynamic range useful.[1] We will now discuss the effects of departures from the simple theory due to area mismatch, ohmic resistance, and beta.

A. Area Mismatches

Equation (9) assumed all the diodes had equal areas, hence, the same I_s. In practice, however, this will never be exactly the case. We will define the variable

$$\gamma = \frac{I_{s2}I_{s4}}{I_{s1}I_{s3}}. \tag{13}$$

Reevaluating (9) to include γ, we find that

$$a = \frac{x\gamma}{1 + x(\gamma - 1)} \tag{14}$$

which is *no longer linear* with respect to x, unless $\gamma = 1$, that is, unless the emitter areas of the inner and outer pairs of transistors are *mutually* equal.

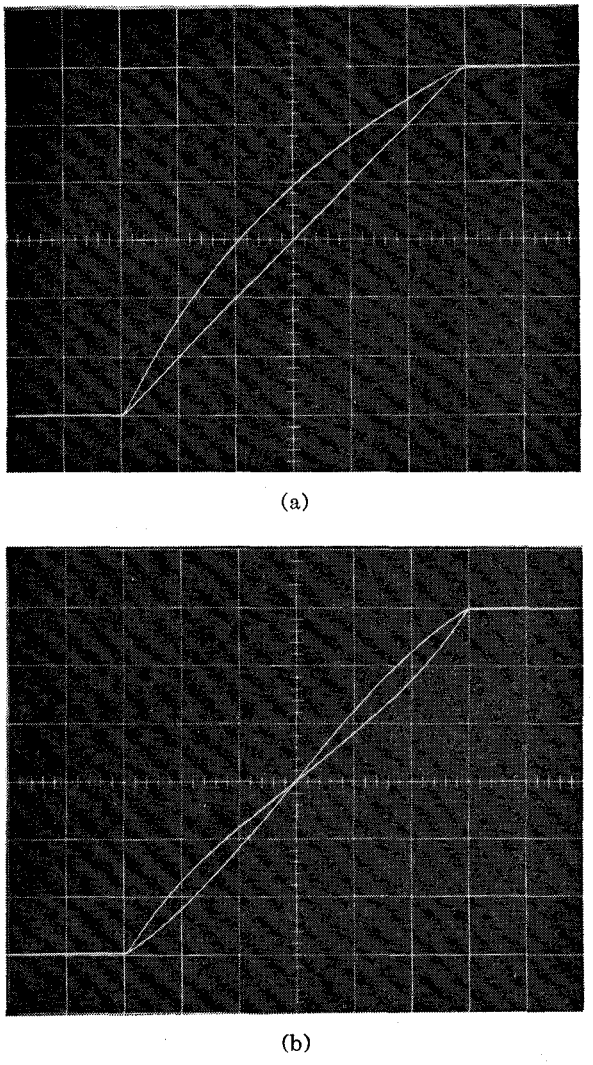
Since we are concerned with achieving low distortion in the transfer function, the effects of even small departures from the ideal case must be examined. From (14),

$$\frac{da}{dx} = \frac{\gamma}{[1 + (\gamma - 1)x]^2}. \tag{15}$$

At the extremes of the dynamic range, $x = 0$ and 1. The incremental slopes at these points are thus γ and $1/\gamma$, respectively. To a first approximation, then, small departures from $\gamma = 1$ cause the slope of the transfer curve to vary linearly over the dynamic range. For example, when $\gamma = 1.1$, corresponding to a total offset around the $Q1$–$Q4$ loop of 2.7 mV at 300°K, the gain will be 10 percent low at $x = 0$, rising to 10 percent high at $x = 1$.

The distortion caused by $\gamma = 2$ (for example, one of the four emitters has an area twice that of the others) is shown in Fig. 4(a). To confirm that area ratios are synonymous with offset voltages, this photograph also shows that linearity can be restored by introducing an error voltage of 18 mV (26 mV times log 2) into the loop.

It might seem that serious distortion would arise in this kind of amplifier due to the inevitable mismatches,



(a)

(b)

Fig. 4. Demonstration of distortions caused by (a) area mismatch ($\gamma = 2$), linearity is restored by a correction voltage; (b) ohmic resistances.

but measurements are surprisingly reassuring. Slope variations of less than ±2 percent are typical, and intermodulation products of −60 dB have been measured for two equal signals at 500 kHz having a peak-combined amplitude of 80 percent of the dynamic range. It should be pointed out that prevalent differential amplifiers of similar gain and bandwidth exhibit far greater distortion for *full-scale* signal swings.

B. Ohmic Resistances

In addition to the diode voltages described by (1) there will be components of voltage due to the ohmic (bulk) resistances of the base and emitter diffusions, metalization paths, and contact interfaces. These resistances are also current dependent due to crowding effects and to an extent dependent on the device geometry.

All of these resistive components can be referred to the emitter circuit as an *equivalent ohmic emitter resistance*. Also, we will assume that this resistance scales with emitter area (a safe assumption for many "stripe" geometries). With these resistances inserted into the

[1] These results were obtained using a single-sided input configuration (see Section III-G) driven from the oscilloscope sweep output of 100-volt amplitude via a large resistor. The input voltage variations over the dynamic range were thus negligible, and a true current-drive condition obtained.

circuit of Fig. 3 and (9) suitably modified, we find

$$\frac{mkT}{q}\log\left\{\frac{x(1-a)}{(1-x)a}\right\} = R_E\{AI_B(1-2x) - I_E(1-2a)\} \quad (16)$$

where R_E = equivalent ohmic emitter resistance of each inner transistor

A = area ratio of inner to outer pairs of emitters.

When the right-hand side of this equation vanishes, $x = a$. This occurs when $A = G_0 = I_E/I_B$, that is, *the areas are scaled in proportion to the drive currents.* If $A \neq G_0$, distortion will arise. Equation (16) has no general explicit solution for a, but a good estimate of the magnitude of the distortion can be readily obtained by a small-signal analysis.[2]

At the quiescent point ($x = 0.5$), the mean normalized slope of the transfer curve is

$$\left.\frac{da}{dx}\right|_{0.5} = \frac{\frac{mkT}{q} + AR_EI_B}{\frac{mkT}{q} + R_EI_E}. \quad (17)$$

Also, putting $x = 0, x = 1$ in (16) yields $a = 0, a = 1$, respectively. Therefore, the *mean* normalized slope is still unity. Thus, when $A > G_0$ (area ratio too big) the transfer curve will have a *higher* slope at the center than at the extremes, and vice versa. Fig. 4(b) shows transfer curves for $I_B = 6$ mA, $I_E = 6$ mA $\pm$ 3 mA using four small-geometry transistors of the same emitter area.

C. Effects of Beta

For monolithic transistors, the match of the current gain between adjacent transistors is very good and will be assumed to be perfect. A further assumption is that the beta is not a function of I_E over the current range of interest. In the presence of finite beta, the emitter currents of $Q1$ and $Q4$ are modified so that

$$\begin{aligned} I'_{B1} &= xI_B - (1-\bar{\alpha})aI_E \\ I'_{B2} &= (1-x)I_B - (1-\bar{\alpha})(1-a)I_E \end{aligned} \quad (18)$$

where $\bar{\alpha}$ = large-signal common-base dc current gain.

Substituting these values into (9) and reducing the logarithmic terms to products and quotients as before, we have

$$\frac{\{xI_B - (1-\bar{\alpha})aI_E\}(1-a)I_E}{\{(1-x)I_B - (1-\bar{\alpha})(1-a)I_E\}aI_E} = 1 \quad (19)$$

which again reduces to $a = x$. This is an astonishing result, since it implies that no matter how low the beta is, the gain *to the emitter circuit* is still G_0. Of course, I_B must still be sufficient to supply the base currents of $Q2$ and $Q3$; any excess flows into $Q1$ and $Q4$ to establish the correct drive voltages for linear operation. Since betas of 100 are typical, the gain reduction from emitter to collector is of the order of 1 percent and first order compensation can be readily made.

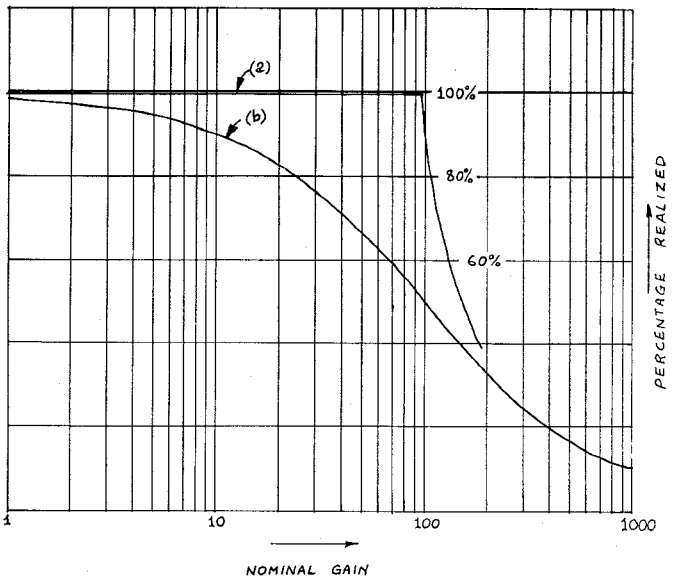


Fig. 5. Gain errors due to beta. Curve (a) shows that accurate gain up to beta is achieved with the new circuit. All prevalent transistor amplifiers have gain errors as shown in curve (b), calculated for $\beta = 100$.

The significance of the above discussion is that stage gains close to beta in magnitude can be realized with very high accuracy. It will be appreciated that this is not possible with any of the standard circuit forms (including Fig. 1(b)) where the actual gain is given by an expression of the form

$$G'_0 = \frac{\beta}{\beta + (1+G_0)}\cdot G_0, \quad (20)$$

where G_0 is now the nominal current gain (for example, the ratio of base to emitter impedances in an emitter-degenerated circuit, or the ratio λ in the circuit of Fig. 1(b)).

Measurements on an amplifier like Fig. 3 in which the betas of $Q2$ and $Q3$ were 95 showed a gain substantially equal to the ratio I_E/I_B right up to $I_E/I_B = 92$, with no degradation of linearity. The results are shown graphically in Fig. 5.

D. Collector Saturation Resistance

The input transistors are connected as diodes and thus operate with $V_{CE} = V_{BE} \approx 0.8$ volts. However, the internal collector bias is reduced to $V_{BE} - I_C R_{sat}$, and the transistor will not be a well-behaved diode when this voltage approaches zero. Since V_{BE} decreases with temperature but R_{sat} increases, the maximum usable current falls sharply with temperature. Circuits involving high-current operation must use a sufficiently large-geometry device with a buried-layer collector.

[2] Another method of analyzing this kind of distortion is used in [9], where expressions for the form and amplitude of the nonlinearities are derived.

E. Thermal Effects

The analyses so far have assumed that all transistors operate at the same temperature, which need not be the case. For example, the inner pair may be at a higher temperature than the outer pair if they are in a separate package and dissipating considerable power. It can be shown by inserting the appropriate temperatures into (9) that the variable a becomes

$$a = \frac{x^n}{(1-x)^n + x^n} \tag{21}$$

where

$$n = \frac{\text{temperature of } Q1 \text{ and } Q4 \text{ in °K}}{\text{temperature of } Q2 \text{ and } Q3 \text{ in °K}}. \tag{22}$$

Another potential source of nonlinearity is thus revealed. A temperature excess of 60°K will put $n = 1.2$ and introduce significant distortion, similar in form and magnitude to the effects of bulk resistances shown in Fig. 4(b).

In monolithic structures having close thermal coupling this problem does not arise, but *transient* thermal distortion can be a problem in critical applications. Following a step change of input current, the dissipation of each device in the quad changes, and some time is taken for the circuit to reacquire thermal equilibrium. During this time, the output signal will suffer a transient distortion whose magnitude, waveshape, and duration depend on the circuit-to-sink and transistor-to-transistor thermal impedances, and, of course, the power dissipation.

Measurements show that these effects are typically less than 1 percent of the step amplitude for the circuit being described. The gain cell described later, has less distortion than this.

F. DC Stability

Equation (14) showed that emitter-area mismatches introduce a shift in the output at the quiescent point ($x = 0.5$), but this shift is not temperature dependent. By adjusting the drive-current balance, the output can be zeroed, and remains this way over the temperature range. This is in contrast with the behavior of prevalent differential amplifiers, where the offset is corrected by a voltage, leaving a drift of 85 log λ μV/°K.

G. Unbalanced Drive Currents

The use of *complementary* drive currents is the key to linear large-signal operation of this type of circuit. However, a constant-current *offset* on one (or both) of the inputs will not impair linearity, but there will be a change in gain, and dc unbalance.

An amplitude *ratio* between the two inputs is more serious—for example, I_{B1} varies from 0 to 1 mA as I_{B2} varies from 1.5 mA to zero. This is equivalent to an area mismatch, as (14) will reveal, in which case λ represents the input ratio. If the ratio is known, linearity can be restored by an adjustment in the emitter area of one of the transistors.

The need for a pair of complementary currents can be eliminated altogether by connecting the $(1 - x)I_B$ input to ground and supplying a constant current of I_B to the junction of $Q1$ and $Q4$ emitters. Clamping diodes should also be added to the input point to control the input voltage during overload conditions ($x > 1$ or $x < 0$). This input point is a useful current-summing node close to ground potential since,

$$V_{in} = 26 \log \frac{x}{1-x} \text{ mV at } 300 \text{ °K} \tag{23}$$

thus

$$|V_{in}| < 36 \text{ mV} \quad \text{for} \quad 0.2 < x < 0.8.$$

This circuit no longer possesses the unique beta immunity of the original form and a small offset term ($a \neq 0.5$ at $x = 0.5$) arises. However, for typical betas, the errors in gain and dc balance are negligible.

IV. Alternative Configurations

Many variants of the basic circuit just discussed have been designed. Space permits only a few of these to be described here, but the generalized statement of the principle, given in the Appendix, will point the way to further forms. These examples will serve to illustrate the versatility of the principle.

A. Inverted Input Diodes

The input diodes $Q1$ and $Q4$ can be inverted and driven from current sinks (instead of current sources), as shown in Fig. 6. Neglecting the effects of beta, area mismatches, and bulk resistances, we can write[3]

$$xI_B aI_E = (1-a)I_E(1-x)I_B \tag{24}$$

or

$$a = 1 - x. \tag{25}$$

The circuit thus has a polarity reversal over the original form. It has the advantage that the input currents are *reusable* at the collectors of $Q1$ and $Q4$, a feature that is put to use in the gain cell described later.

This configuration lacks the beta immunity of the original form, because the base current of $Q2$ and $Q3$ *add* to the input currents. This, together with the fact that the input that is receiving the *smaller* fraction of the base-current drives the transistor that conducts the *larger* fraction of the emitter current, causes significant beta dependence.

By considering these currents, it is found that the output modulation index a is reduced to

$$a = \theta(1-x) + \frac{G_0}{\beta} \tag{26}$$

where

$$\theta = \frac{\beta}{\beta + (1 + 2G_0)}. \tag{27}$$

[3] The Appendix shows how these quantities can be equated from an inspection of the circuit.

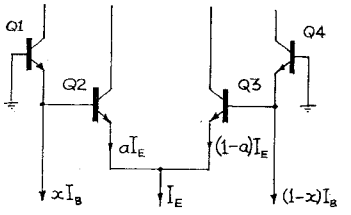


Fig. 6. Inverted form of circuit. Performance is similar, except for phase change in output ($a = 1 - x$).

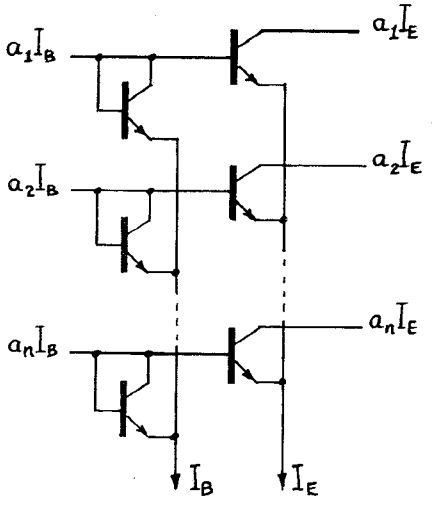


Fig. 7. Multiple-input version of Fig. 3. This circuit is useful to standardize the absolute magnitude of analog signals.

Linearity is not impaired, but the actual stage gain (to the collector circuit) is reduced to

$$G_0' = \theta G_0, \tag{28}$$

and the output swing is reduced.

For example, with $G_0 = 2$ and $\beta = 50$, G_0' is 1.8, 10 percent below the nominal gain, and the limiting values of a are 0.96 and 0.04.

The "inverted" circuit is identical in behavior to the earlier configuration as regards the effects of area mismatches, ohmic resistances, and temperature. It may also be driven by a single-sided signal, by a similar rearrangement to that suggested for Fig. 3. This time, the collectors and bases of $Q1$ and $Q4$ are joined and taken to a positive current source equal to I_B, and the emitter of $Q4$ and base of $Q3$ grounded. The input is then applied to the emitter of $Q1$, and must lie in the range 0 to I_B. Clamping diodes are needed to constrain the input voltage outside of this range.

B. Multiple-Input Configurations

It is not necessary to limit the input and output to a *pair* of complementary currents. Fig. 7 shows how n inputs may be accepted to produce n outputs. This is useful when the inputs are known to be in a certain *ratio* but have an absolute value which may vary widely, and it is desired to standardize the signals to a known amplitude. An example might be in connection with lateral p-n-p transistors as the drivers and level shifters.

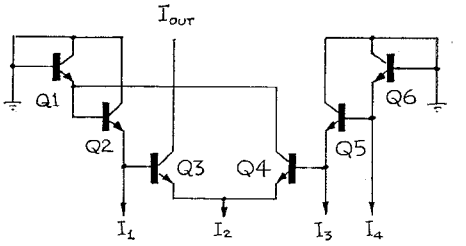


Fig. 8. Product-quotient circuit. By suitable imputs, various functions can be generated.

This circuit also possesses the beta immunity referred to earlier, and can be inverted to the Fig. 6 form.

C. Product-Quotient Configuration

Fig. 8 shows an interesting variant able to produce an output equal to the products or quotients of several inputs at nanosecond speeds. The significant feature of this circuit is the feedback connection of $Q4$'s output to the emitter of $Q1$. Neglecting second order effects, we have

$$(I_2 - I_{out})I_1I_{out} = (I_2 - I_{out})I_3I_4 \tag{29}$$

or

$$I_{out} = \frac{I_3I_4}{I_1}.$$

The magnitude of the emitter current, I_2, although not present in the expression for I_{out}, determines the maximum value the output may assume.

The circuit may be expanded to accept any odd number of inputs to generate power terms such as x^2, x^3, x^3/yz, etc. Of course, there is a practical limit to the circuit complexity determined by the stacking of mismatch errors.

By putting extra diodes in series with the emitters of the center transistors, $Q3$ and $Q4$, and in $Q1$ and $Q6$, we get

$$I_{out} = I_4\sqrt{\frac{I_3}{I_1}}. \tag{30}$$

The number of circuits that can be devised to perform functions of this kind is legion.

V. The "Gain Cell"

We will now describe what is probably the most useful circuit to have arisen from this work, certainly as far as incorporation into cascaded amplifiers is concerned. It is shown in Fig. 9, and is similar to the "inverted" circuit of Fig. 6, except that the input currents that reappear at the collectors of $Q1$ and $Q4$ are added in phase with the outputs of $Q2$ and $Q3$. The gain is thus

$$G_0'' = 1 + G_0 = \frac{(I_B + I_E)}{I_B}. \tag{31}$$

This form is attractive for several reasons. Firstly, the inner stage can operate with a gain less than unity, yet still achieve a net stage gain greater than unity; we

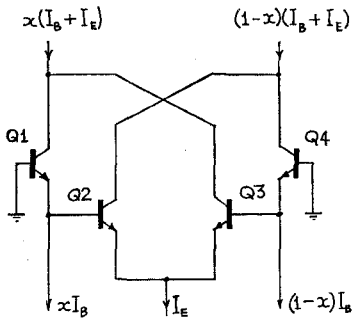


Fig. 9. The "gain cell."

would therefore expect to find this circuit faster for a given gain. Secondly, the cell is well suited to cascading, because the output of one stage can drive the next directly. This leads to a further advantage: all of the current injected at each emitter node contributes to the total output swing. Finally, the bias-voltage circuit for each stage has to supply only the base current for that stage, which is not signal dependent. Hence a low-power diode or resistor string can be used; only 0.5–1 volt per stage is needed.

A. More Exact Gain Expression

Equation (31) omits the effects of beta. It was previously shown that other parameters, such as ohmic resistances and area mismatches, did not affect the *mean* gain over the full dynamic range, but only introduced distortion. Thus, the accuracy limitations determined by beta are of most concern.

In a cascaded amplifier, the inputs of all stages, except the first, are simply the collector currents from the previous stage. Thus, starting with an input bias current, I_B, a modulation index, x, and an emitter supply current I_E, we can calculate the effective values of I_B' and x' for the input to the following stage, the stage gain G_0''', and the output swing ΔI_C.

The results are

$$\begin{aligned} I_B' &= \alpha I_B + I_E \\ \Delta I_C &= \alpha(I_B + \theta I_E) \\ G_0''' &= \alpha(1 + \theta G_0) \\ x' &= \frac{1}{2} + \frac{G_0'''}{\alpha + G_0}(x - \tfrac{1}{2}). \end{aligned} \tag{32}$$

The practical significance of these expressions is demonstrated by the photographs of Fig. 10, which show the swept gain of a gain cell using (a) transistors with a beta of 40, and (b) transistors with a beta of 200. For these measurements, $I_B = 200\ \mu$A and I_E was swept from 0 to over 1 mA. The input was modulated to a depth of 80 percent at 1 kHz. At more practical operating currents, the effects of bulk resistances must be taken into account, preferably by scaling the emitter areas.

B. Thermal Distortion

In many exacting applications, for example, oscilloscope vertical amplifiers, the transient thermal distortions mentioned earlier can be very disturbing. It can be

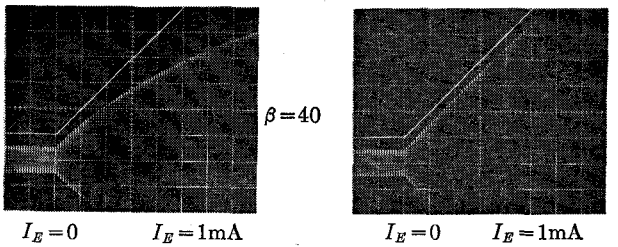


Fig. 10. By sweeping I_E, the effects of beta in limiting the accuracy of gain are demonstrated.

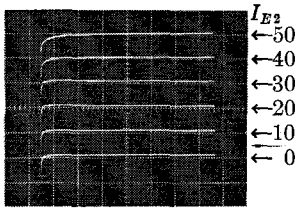


Fig. 11. Thermal distortion of gain cell. Vertical scale expanded to 1 percent/div. $I_E = 20$ mA.

shown that for the gain cell no thermal distortion arises if the total power dissipated in the outer and inner pairs of transistors is the same. Because the inner pair operates at a higher voltage (by one V_{BE}) than the outer pair, I_E must be less than I_B, so that to eliminate thermal distortion a gain of less than two must be used.

It can be simply shown that for no distortion

$$I_E = \frac{V_{CB}}{V_{CB} + V_{BE}} \cdot I_B, \tag{33}$$

where V_{CB} is the bias voltage supplied to each stage and appears across the collector-base junctions of $Q1$ and $Q4$. Typically, with $V_{CB} = 1.6$ volts, $V_{BE} = 0.8$ volts, we must use a stage gain of 1.67.

In practice, it proves very difficult to induce appreciable distortion even under high-power conditions. Fig. 11 shows the step response of a single integrated gain cell operated at very high currents with the vertical scale expanded to 1 percent per division.

VI. Transient Response

An accurate large-signal model for the transistor does not exist, and it is, therefore, not possible to determine the exact step response of this type of circuit. Various small-signal analyses have been made, however, and these indicate, not surprisingly, that the response time is dominated by the f_t of the devices. Further, if the geometries are scaled to the currents, the effects of r_b can be eliminated, and for fairly large stage-gains (greater than two or three), an adequate approximation for the stage response of the Fig. 3 circuit is obtained by treating it as a network with a single pole at f_t/G_0.

The gain-cell response is more complex, and can be approximated by a pole at $f_t/2$, flanked by a pole–zero pair. For $G_0 = 1$, ($G_0'' = 2$), the pole pair are co-incident and cancel, leaving a single-pole response with a gain–bandwidth product equal to f_t. For $G_0 < 1$, the zero moves toward the origin, while the second pole moves out,

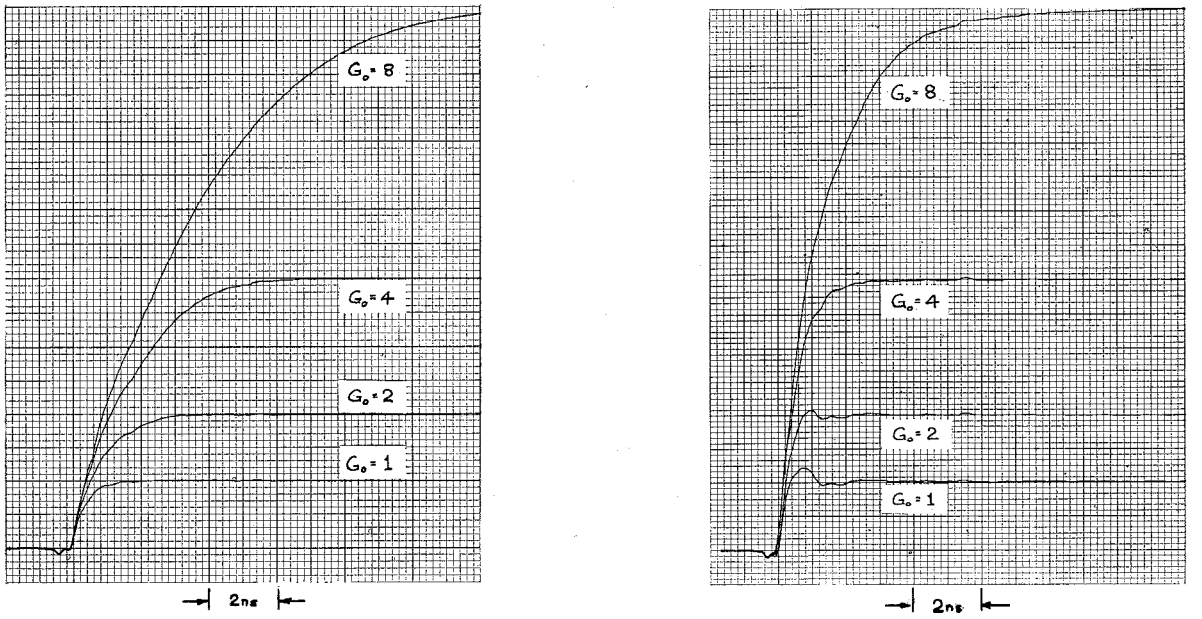


Fig. 12. Measured transient responses of Fig. 3 circuit.

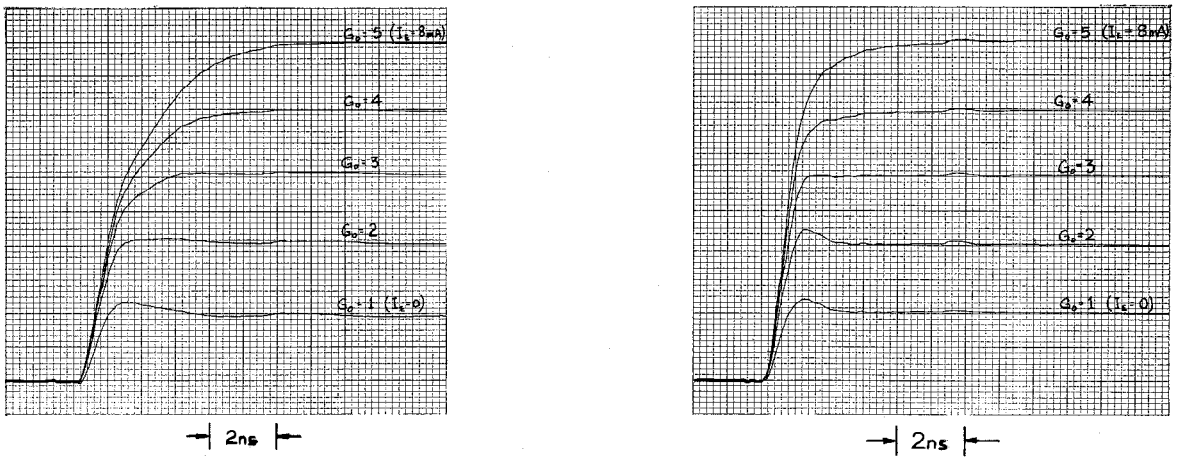


Fig. 13. Measured transient response of the gain cell.

causing an overshot response. For $G_0 > 1$, the reverse situation arises, and the response consists of an initial fast rise followed by a slower time constant. For $G_0 \gg 1$, the second pole dominates and is at approximately f_t/G_0.

Measurements on integrated forms of Fig. 3 and Fig. 9 have been made, using both the standard process (200 ohms-per-square base diffusion) and a shallower 450 ohms-per-square process. The device characteristics are summarized below.

Parameter	200-Ohm Process	450-Ohm Process	Test Conditions
f_t	600 MHz	1200 MHz	V_{CE} = 2 volts, I_C = 10 mA
C_{CS}	2 pF	2 pF	V_{CS} = 10 volts
C_{CB}	0.7 pF	0.7 pF	V_{CB} = 0 volts
β	40	200	V_{CE} = 2 volts, I_C = 5 mA
r_b	15 ohms	27 ohms	I_{EB} = 10 mA, I_C = 0

Response time was measured in a test system having an overall risetime of less than 75 ps. The waveforms were plotted on an X–Y plotter, which provided better resolution than photography for comparing several collated responses.

The results confirm the predictions of the small-signal analyses. In Fig. 12, the responses of the Fig. 3 circuits are shown, using (a) the 600-MHz devices, and (b) the 1200 MHz devices. The form of the response closely approximates a single time constant proportional to gain. Fig. 13 shows the more complex "gain cell" response for (a) the 600-MHz devices and (b) the 1200-MHz devices.

It should be mentioned that the test jig incorporated neutralizing capacitances and balun transforms at critical sites to eliminate preshoot and other aberrations in the responses, and those components markedly influenced the

apparent risetimes at the lower gains. The improved performance obtained from the totally integrated amplifier (Section VII-D) supports this conclusion.

VII. Cascaded Amplifiers

A repetitive problem in dc amplifier design is that of cascading stages without "level shifting," that is, the buildup of supply–voltage requirements as each stage is added. Several techniques are used to overcome this problem. They include the use of

1) complementary *p-n-p–n-p-n* devices between two supply rails,
2) zener-diode or resistive level-shifting circuits,
3) very low bias voltages for each stage.

For monolithic designs, the third method is by far the most attractive, but is possible only where voltage swings are small, and hence is eminently suited to the circuits under discussion. Several cascaded forms will be presented that use this approach. In some cases, stages can be cascaded indefinitely without level shifting.

Other factors of interest in a cascaded amplifier are gain stability, drift, overload recovery, dinearity, noise level, and transient response. Typical results for these parameters will be given where appropriate.

A. Bandwidth of Cascaded Stages

It was shown earlier that each stage of an amplifier using the diode transistor circuit can be treated as a single pole at f_t/G_0, where G_0 is the low-frequency current gain of the stage.

The overall bandwidth, F, of N cascaded stages is thus

$$F = \frac{f_t}{G_0}(\sqrt[N]{2} - 1)^{1/2}. \tag{34}$$

Also, if the total current-gain, G, is shared equally over the N stages we can write

$$\frac{F}{f_t} = \frac{(\sqrt[N]{2} - 1)^{1/2}}{\sqrt[N]{G}}. \tag{35}$$

This "bandwidth-shrinkage" function is plotted in Fig. 14. It shows three things of interest. Firstly, there is an optimum number of stages required to maximize the bandwidth. Secondly, the maxima are well defined for low gains but vague at high gains. Finally, the bandwidth of a large number of stages is relatively insensitive to overall gain.

In discrete designs, it is frequently necessary to use less than the optimum number of stages for economic reasons. The possibility of integrating the entire amplifier removes this limitation.

B. A Practical Cascaded Amplifier

The first cascaded form we will discuss is made by taking the Fig. 3 circuit and connecting the collectors of one stage to the bases of the next, as shown in Fig. 15. The amplifying transistors operate with $V_{CB} = 0$, just as the diode-connected transistors do. Currents must be supplied at each interface, and these, together with the emitter currents, determine the gain. The overall gain of N stages is

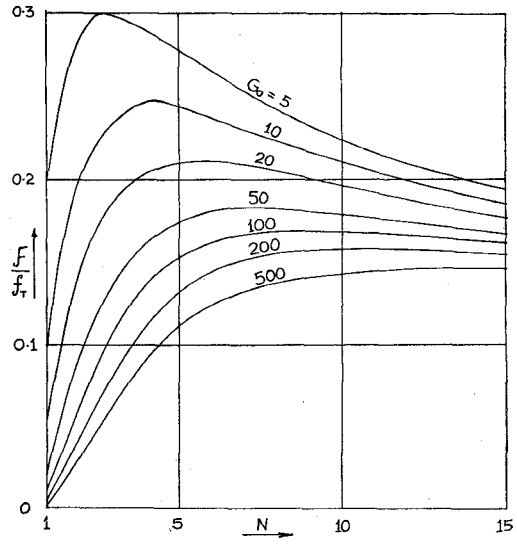


Fig. 14. Curves showing number of stages required to maximize bandwidth.

$$G = \frac{I_2}{I_1}\cdot\frac{I_4}{I_3 - I_2}\frac{I_{N+2}}{I_{N+1} - I_N}. \tag{36}$$

Consequently, the gain can be swept over a very wide range (zero to about β^N), but the gain is a sensitive function of the difference terms in the denominators of (36). Methods of overcoming this problem have been devised, using dependent current sources for the emitters. Diodes (not shown) prevent the transistors saturating at the extremes of the dynamic range.

A limited investigation of this form of amplifier has been made. Some results for a four-stage high-gain (about X1000) design are shown in Fig. 16. They demonstrate the linearity and overload recovery to a 3-μs ramp (Fig. 16(a)), and the transient response and noise level (Fig. 16(b)). In the second photograph, the responses of 600-MHz and 1200-MHz transistors are compared.

C. Cascaded Gain Cells

A series of three gain cells is shown in cascade in Fig. 17. The power is supplied as currents (to the emitter nodes) which set both the gain and the output swing capability. A low-power bias string, shown here as pairs of diodes, sets the operating voltage of each stage. The balanced nature of the amplifier results in very small signal currents in the base circuit, and permits a common undecoupled line to supply all the bases in a multistage amplifier. The overall gain, for sensible values of beta, is

$$G' = \frac{I_{E1} + I_{E2} + I_{E3} + I_{E4}}{I_{E1}}. \tag{37}$$

With this method of cascading, the correct static bias conditions to handle the steadily increasing signal ampli-

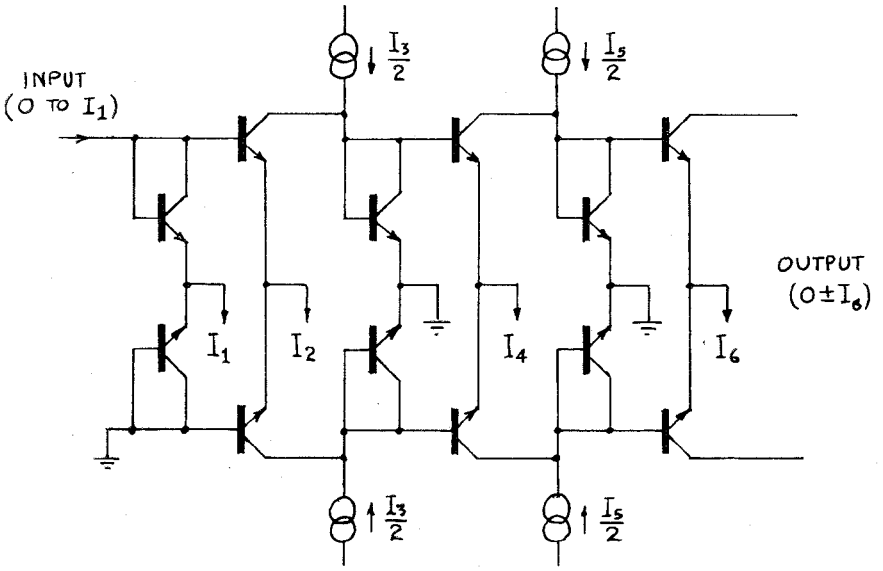


Fig. 15. A cascaded amplifier.

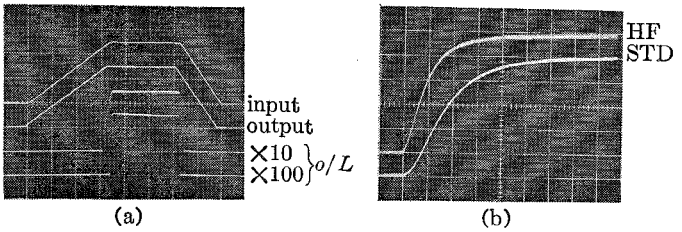


Fig. 16. Performance of amplifier similar to Fig. 15. In (a), a 2.5-μA input current (top trace) is accurately reproduced at × 1000 (second trace). When overloaded ten or a hundred times, (lower traces) recovery is rapid. The top portion of the waveform is within the linear range. Time scale is 1 μs/div. Waveforms (b) show transient response at × 1000, to a 1-μA input step using 600-MHz (top) and 1200-MHz (bottom) transistors. Time scale is 20 ns/div.

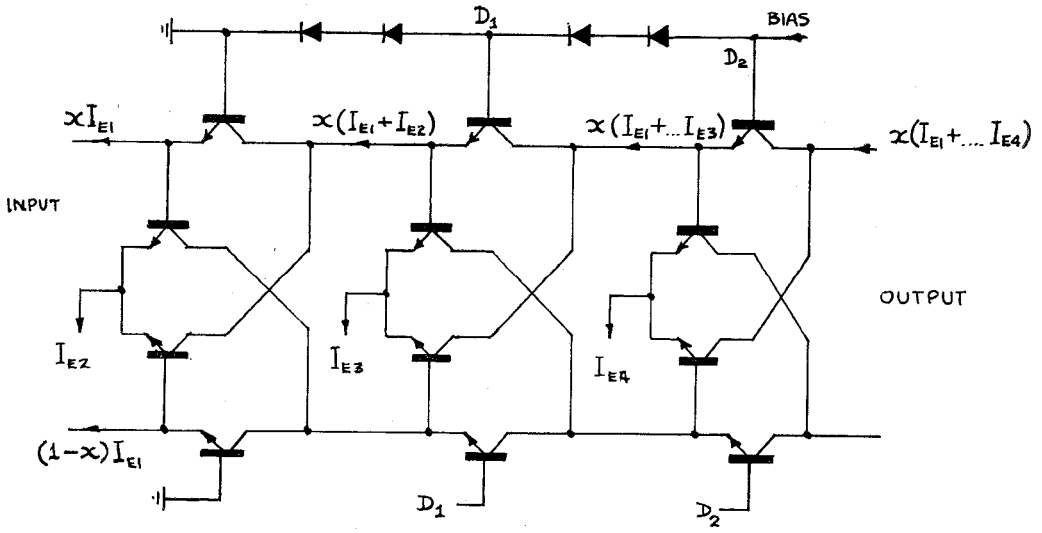


Fig. 17. Cascaded gain cells.

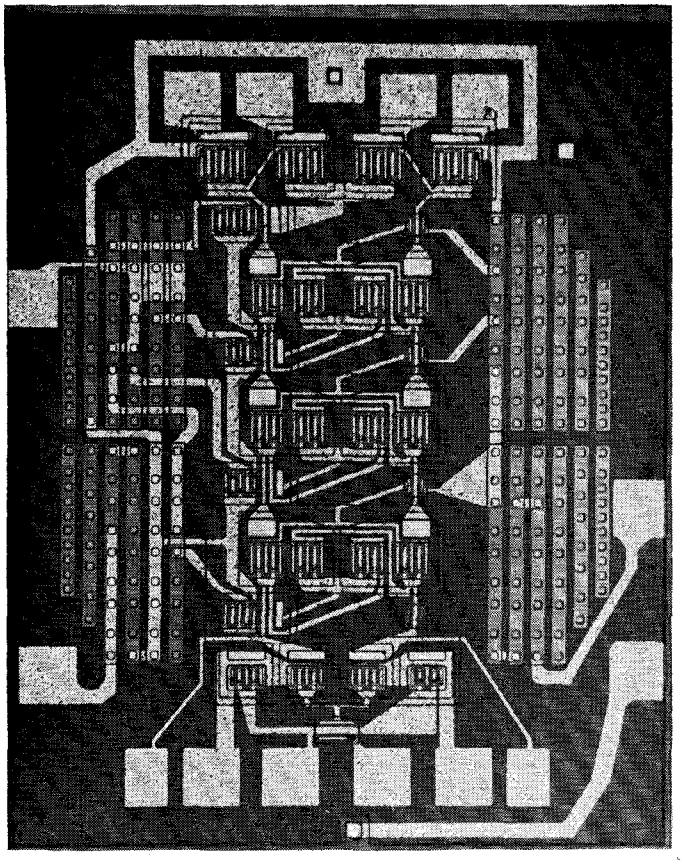


Fig. 18. Photomicrograph of integrated gain-cell amplifier. The emitter currents are supplied by the transistors just left of center, and the signal path is from bottom to top.

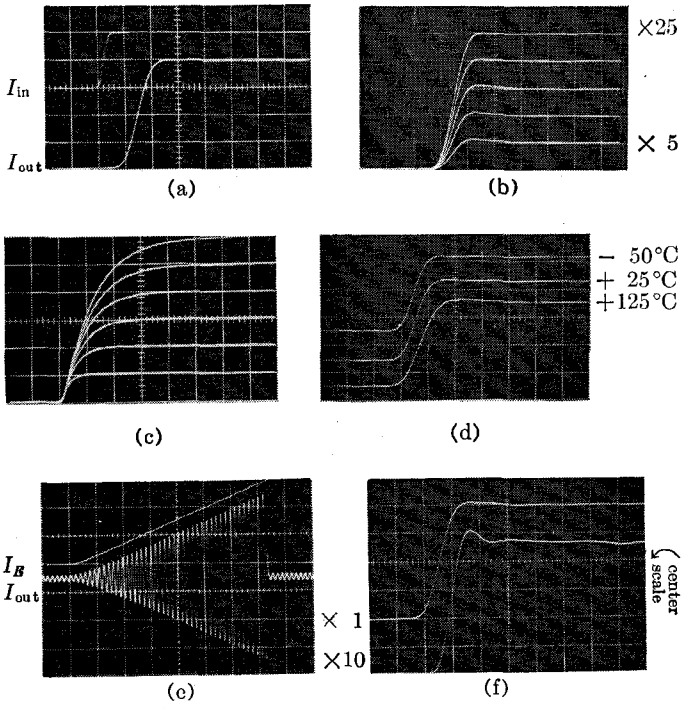


Fig. 19. Performance of the integrated amplifier. (a) Input/output delay at ×20, 1 ns/div. (b) Transient-response variations for gains of ×5 to ×25, 1 ns/div. (c) For gains ×50 to ×300, 5 ns/div. (d) Shows transient response for ×25 at −50°C, 25°C, and 125°C. (e) Demonstrates swept-gain linearity, ×1 to ×30. (f) Shows recovery from ×10 overload.

tude along the chain are automatically present, because *signal and bias currents increase at the same rate.* This is in contrast with conventional amplifiers where care must be taken to provide adequate range at each stage.

A corollary of this is that all stages overload at the same point on the overall transfer characteristic, when $x < 0$ or $x > 1$. Furthermore, at overload, no transistors saturate, and, because of the modulation reduction discussed above, the input to each successive stage never quite drops to zero. Consequently, the recovery characteristics are good.

D. An Integrated Gain-Cell Amplifier

Fig. 18 is a photomicrograph of a five-stage amplifier on a 50 × 60 mil die, using the 1200-MHz transistors. It comprises an input stage of the type shown in Fig. 3 followed by four cascaded gain-cell stages each with a current-source transistor, and a patchwork of resistors which can be used to set the base bias voltages and emitter currents.

The mask layout is of considerable importance in achieving optimum performance. For example, in order to maximize the f_t of each stage, the transistors must be increased in size from input to output, thus maintaining a constant current density. Also, to reduce collector-substrate capacitance, the gain-cell transistors are paired into two common-collector isolations, which also serve to provide a crossover region for the base circuit. Because of the low impedance level of the circuit, meticulous attention to balancing the metallization resistances is essential to achieve low distortion.

In this circuit, the inner and outer transistors have equal area. For minimum distortion due to bulk resistances a stage gain of two must be used, producing an overall gain of 16. (The input stage operates at a gain of unity.) However, quite large variations from this value can be tolerated without introducing significant distortion.

A summary of the performance of the integrated amplifier is given below, and in the waveforms of Fig. 19. The system risetime used for these measurements was 0.4 ns, and, accordingly, some corrections were necessary to the measured risetimes.

The input to the amplifier was supplied by a push–pull 50-ohm pulser driving $Q1$ and $Q4$ via 1-kΩ resistors, to ensure current-drive conditions. Also, both outputs were connected to the sampling oscilloscope via 50-ohm lines terminated at both ends. Under these conditions, no power gain was realized for current gains below 40. It is, however, entirely practical to match both input and output to 50 ohms.

By suitable choice of current ratios, a wide range of gain from zero to ×1000 could be obtained, although acceptable high-speed performance was possible only over a limited range. Some idea of the effect of gain on transient response is provided by Fig. 19(b) and (c). Considerable care had to be exercised to eliminate ringing and preshoot, particularly in respect to ground-plane

currents. A bifilar transformer was required to balance the input drive. It is interesting to note the improvement that resulted from putting all the stages on one die. Results published earlier [8] showed a large amount of ringing. These waveforms were for an amplifier in which individually packaged gain cells were breadboarded together. Series damping resistors were required between stages to bring the ringing down to even this level, and these slowed the response.

Here is a summary of the performance of the fully integrated amplifier.

Current Gain	Corrected Risetime (ns)	Overshoot (percent)	Bandwidth (MHz to −3 dB)	Gain–Bandwidth Product (GHz)
5	0.58	12	>500	–
10	0.63	8	>500	–
15	0.69	3	500	7.5
20	0.81	–	420	8.4
25	0.92	–	370	9.2
50	2.0	–	170	8.5
100	4.0	–	92	9.2
200	8.5	–	45	9.0
300	12.0	–	31	9.3

The effect of temperature on transient response was also checked, and, as expected, the risetime decreases at low temperatures, due to increases in f_t. Also checked were the accuracy of gain in response to a linearly swept current (Fig. 19(e)) and overload recovery (Fig. 19(f)). The latter shows the response for an output equal to 80 percent of the dynamic range, and that for a ×10 increase in the input.

The CW response is shown in Fig. 20, for current gain from ×5 to ×50, and for a signal equal to 50 percent of the dynamic range. Other performance details, for a gain of ×25, are

Output current swing: ±60 mA into two 25-ohm loads
Output voltage swing: 3 volts peak–peak, differentially
Dissipation: 320 mW (64 mA at 5 volts)
Gain stability: Within ±0.35 dB over −35°C to +125°C
Zero drift: <0.4 percent of full scale over −35°C to +125°C

These results are very encouraging. Subsequent improvements in device technology will almost certainly permit faster gain-cell amplifiers to be fabricated.

VIII. Summary

This paper has described what is believed to be a new technique for the design of very linear transistor amplifiers especially suited to monolithic planar fabrication. It demonstrates that useful dc-coupled wide-band amplifiers can be made using transistors alone, operating in a strictly current-gain mode. A commercial integrated circuit giving a controllable power gain of 0 to 20 dB with a dc to 500-MHz response, and designed to be cascadable without additional components, is now fully practical. This building-block concept could be expanded

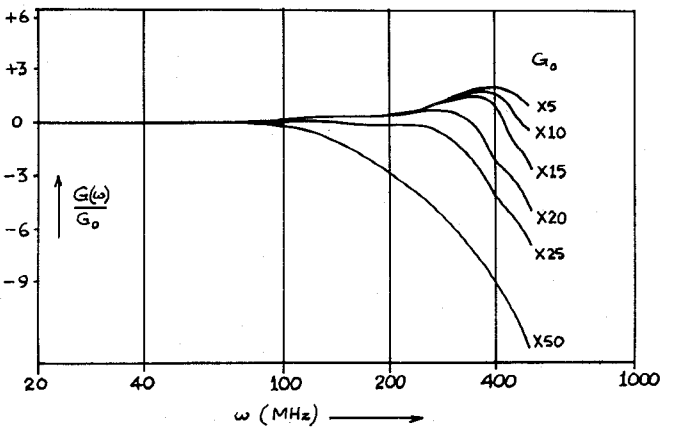


Fig. 20. CW response of integrated amplifier.

to include precision high-impedance input stages, power amplifiers, four-quadrant multipliers, etc., and would fill a long neglected need. Some of these topics will be taken up in subsequent papers.

Appendix

The Generalized Principle

The key equation for these circuits is of the form given in (9), an algebraic summation of an even number of logarithmic terms set equal to zero, having a common factor, mkT/q, which is, thus, of no consequence. The logarithmic arguments may be compounded into a single product-quotient term, and finally the antilogarithm of both sides is taken, leaving the product-quotient term equal to unity, for example

$$\frac{I_{D1}I_{D2}I_{S3}I_{S4}}{I_{S1}I_{S2}I_{D3}I_{D4}} = 1. \tag{38}$$

The saturation currents of the diodes in the loop can be extracted as the ratio

$$\gamma = \frac{\prod I_s(+)}{\prod I_s(-)} \tag{39}$$

where the plus sign indicates all diodes connected in the same direction around the loop, and the minus sign indicates those in the opposite direction. Since I_s is proportional to area, the γ ratio can also be calculated as the ratio of the *areas* of the emitter diodes in the loop. Also, every I_s varies with temperature in the same way, and thus temperature drops out of the analysis at this point, too, explaining the excellent freedom from temperature effects observed in these amplifiers. This reduces the key equation to the form.

$$\prod I(+) = \gamma \prod I(-). \tag{40}$$

Stated at length, in a closed loop containing an even number of perfect exponential diode voltages (not necessarily two-terminal devices) arranged in cancelling pairs, the currents are such that the product of the currents in diodes whose voltage polarities are positive with respect to a node in the loop is exactly proportional to the

product of the currents in diodes whose voltages are negative with respect to that node, the constant of proportionality being the ratio of the product of the saturation currents of the former set of diodes to that of the latter set.

Acknowledgment

During the course of this work, the discussions held with L. Larson of the Advanced Instruments Group and G. Wilson of the Integrated Circuits Group proved stimulating and helpful. The experimental work of E. Traa is also gratefully acknowledged.

References

[1] A. E. Hilling and S. K. Salmon, "Intermodulation in common-emitter transistor amplifiers," *Electron. Engrg.*, pp. 360–364, July 1968.
[2] L. C. Thomas, "Eliminating distortion in broadband amplifiers," *Bell Sys. Tech. J.*, p. 315, March 1968.
[3] J. S. Brown, "Broadband amplifiers," in *Amplifier Handbook*. New York: McGraw-Hill, 1966, sec. 25, pp. 25–57.
[4] L. F. Roeshot, "U.H.F. broadband transistor amplifiers," *EDN Mag.*, January–March 1963.
[5] W. R. Davis and H. C. Lin, "Compound diode-transistor structure for temperature compensation," *Proc. IEEE (Letters)*, vol. 54, pp. 1201–1202, September 1966.
[6] A. Bilotti, "Gain stabilization of transistor voltage amplifiers," *Electron. Letters*, vol. 3, no. 12, pp. 535–537, 1967.
[7] A. M. VanOverbeek, "Tunable resonant circuits suitable for integration," *1965 ISSCC Digest of Tech. Papers*, pp. 92–93.
[8] B. Gilbert, "A dc-500 MHz amplifier multiplier principle," *1968 ISSCC Digest of Tech. Papers*, pp. 114–115.
[9] ——, "A precise four-quadrant multiplier with subnanosecond response," this issue, pp. 365–373.
[10] H. E. Jones, "Dual output synchronous detector utilizing transistorized differential amplifiers," U. S. Patent 3 241 078, June 18, 1963.
[11] A. Bilotti, "Applications of a monolithic analog multiplier," *1968 ISSCC Digest of Tech. Papers*, pp. 116–117.
[12] W. R. Davis and J. E. Solomon, "A high-performance monolithic IF amplifier incorporating electronic gain-control," *1968 ISSCC Digest of Tech. Papers*, pp. 118–119.
[13] G. W. Haines, J. A. Mataya, and S. B. Marshall, "IF amplifier using C-compensated transistors," *1968 ISSCC Digest of Tech. Papers*, pp. 120–121.
[14] C. T. Sah, "Effect of surface recombination and channel on *P-N* junction and transistor characteristics," *IRE Trans. Electron Devices*, vol. ED-9, pp. 94–108, January 1968.

A Precise Four-Quadrant Multiplier with Subnanosecond Response

BARRIE GILBERT, MEMBER, IEEE

***Abstract*—This paper describes a technique for the design of two-signal four-quadrant multipliers, linear on both inputs and useful from dc to an upper frequency very close to the f_t of the transistors comprising the circuit. The precision of the product is shown to be limited primarily by the matching of the transistors, particularly with reference to emitter-junction areas. Expressions are derived for the nonlinearities due to various causes.**

I. Introduction

AN IDEAL FOUR-quadrant multiplier would perfectly satisfy the expression

$$Z = \text{constant}, XY \tag{1}$$

for any values of X and Y, and produce an output having the correct algebraic sign. Ideally, there would be no limitation on the rate of variation of either input.

All practical multipliers suffer from one or more of the following shortcomings.

1) A nonlinear dependence on one or both of the inputs.
2) A limited rate of response.
3) A residual response to one input when the other is zero (imperfect "null-suppression").
4) A scaling constant that varies with temperature and/or supply voltages.
5) An equivalent dc offset on one or both of the inputs;
6) A dc offset on the output.

In the field of high-accuracy medium-speed multipliers, the "quarter-square" technique has gained favor [1]. This method makes use of the relationship

$$XY = \tfrac{1}{4}[(X + Y)^2 - (X - Y)^2] \tag{2}$$

and employs elements having bipolar square-law voltage–current characteristics, together with several operational amplifiers.

Much work has been put into harnessing the excellent exponential voltage–current characteristics of the junction diode for multiplier applications, either by using single diodes (or transistors) in conjunction with operational amplifiers [2], or, more recently, pairs of transistors connected as a differential amplifier [3]–[6]. In the majority of cases, the strong temperature dependence of the diode voltage proved a problem, and at least two commercially available multipliers are equipped with an oven to reduce this dependence.

Manuscript received June 28, 1968; revised September 16, 1968. This paper was presented at the 1968 ISSCC.
The author is with Tektronix, Inc., Beaverton, Ore.

Another problem of the "differential-amplifier" multiplier, analyzed in [7], is the nonlinear response with respect to the base-voltage input. To achieve useful linearity, the dynamic range on this input must be restricted to a very small fraction of the full capability, leading to poor noise performance and worsened temperature dependence, including poor zero stability.

The problems associated with this type of multiplier can be largely overcome, however, by using diodes as current–voltage convertors for the base inputs, thus rendering the circuit entirely current controlled, theoretically linear, and substantially free from temperature effects. This paper is concerned mainly with the determination of the magnitude of the nonlinearities in a practical realization, and the analysis draws heavily on the groundwork laid in [7]; some mathematical expressions will be quoted directly from this paper without proof here.

II. The Basic Circuit

The basic scheme is shown in Fig. 1. It is comprised of two pairs of transistors, $Q2$-$Q3$ and $Q5$-$Q6$, having their collectors cross-connected, driven on the bases by a further pair of transistors, $Q1$-$Q4$, connected as diodes. It is the addition of this pair of diodes that linearizes the circuit. The X signal input is the pair of currents xI_B and $(1-x)I_B$. The Y signal is yI_E and $(1-y)I_E$, where x and y are dimensionless indexes in the range zero to unity.

It was previously shown [7] that the *ratio* of the emitter currents in the $Q2$-$Q3$ and $Q5$-$Q6$ pairs is the same as that in the $Q1$-$Q4$ pair and independent of the magnitudes of I_B and I_E (neglecting second order effects). We can thus write

$$\begin{aligned} I_{C2} &= xyI_E \\ I_{C3} &= (1-x)yI_E \\ I_{C5} &= x(1-y)I_E \\ I_{C6} &= (1-x)(1-y)I_E. \end{aligned} \tag{3}$$

The *differential*[1] output is

$$I_{out} = I_{C2} + I_{C6} - I_{C3} - I_{C5}. \tag{4}$$

Thus, the normalized output Z is

$$\begin{aligned} Z = \frac{I_{out}}{I_E} &= xy + (1-x)(1-y) \\ &\quad - (1-x)y - (1-y)x \\ &= 1 - 2y - 2x + 4xy. \end{aligned} \tag{5}$$

It is seen that the circuit is balanced when x and y are equal to 0.5. If we apply bias currents such that *bipolar* signals X and Y can be used as the inputs, and

[1] The output may also be taken as a single-sided signal from the collectors of $Q2$ and $Q6$, in which case it is $Z = \frac{1}{2}(1 + XY)$.

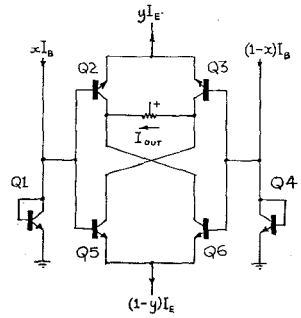


Fig. 1. The basic four-quadrant multiplier.

substitute

$$\begin{aligned} X &= 2x - 1 \\ Y &= 2y - 1 \end{aligned} \tag{6}$$

where X and Y are in the range -1 to $+1$, the output is

$$Z = XY. \tag{7}$$

This is an exact large-signal analysis, and makes no assumptions about temperature. It did, however, assume that the transistors had 1) perfectly matched emitter diodes, 2) perfect exponential characteristics (no ohmic resistance), and 3) infinite betas.

The extent to which departures from this ideal case impair the linearity will now be analyzed.

III. Distortion Due to Area Mismatches

In [7] it was found that "offset voltage"—the voltage required to balance the emitter currents of a pair of transistors in a differential amplifier—could be expressed more conveniently as a ratio of the saturation currents (or areas) of the two emitter junctions. For the four-transistor amplifier "cell" discussed in that paper, the mismatch ratio

$$\gamma = \frac{I_{S2}I_{S4}}{I_{S1}I_{S3}} \tag{8}$$

was defined. It was then shown that for $\gamma \neq 1$ (imperfect matching), the output currents were no longer simply in the same ratio x as the input currents, but had the form

$$a = \frac{\gamma}{1 + x(\gamma - 1)} \cdot x. \tag{9}$$

This can be expressed in a form that shows the nonlinearity due to area mismatches as a separate term D_A

$$a = x + D_A = x + \frac{x(1-x)(1-\gamma)}{1 + x(\gamma - 1)}. \tag{10}$$

For $\gamma \approx 1$, this simplifies to

$$D_A \approx x(1-x)(1-\gamma). \tag{11}$$

This is a parabolic function of x having a peak value $\hat{D}_A$ of $0.25(1-\gamma)$, which leads to the useful rule of thumb

$$\hat{D}_A(\text{percent}) \approx V_0(\text{mV}) \text{ at } 300^\circ\text{K} \tag{12}$$

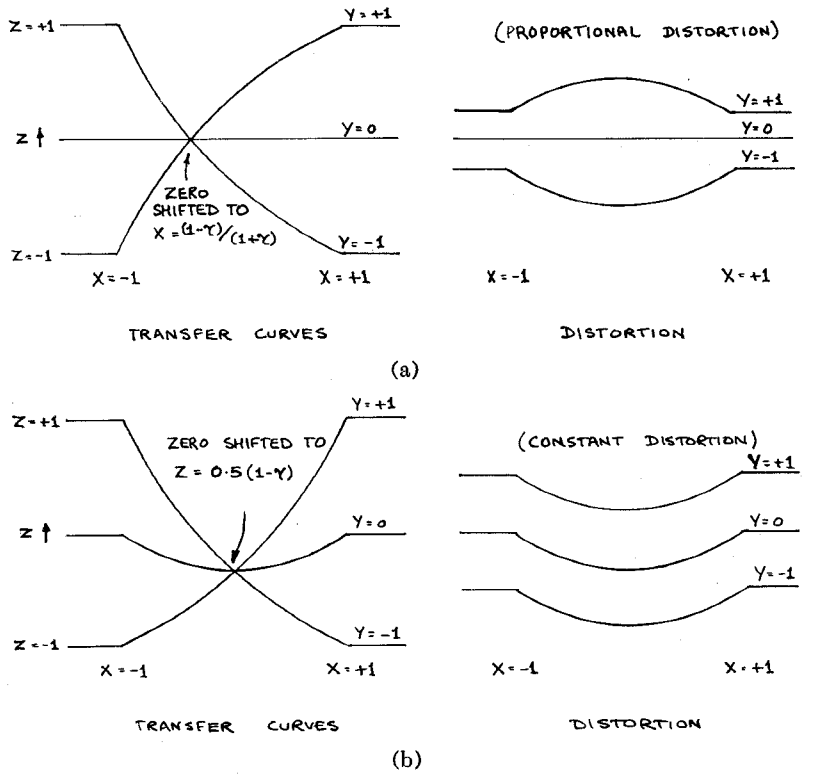


Fig. 2. Distortion introduced by area mismatches (exaggerated). (a) $\gamma_1 = \gamma_2$. (b) $\gamma_1 = 1/\gamma_2$.

where $V_0 = (kT/q) \log \gamma$, the total loop offset voltage. However, notice that $\hat{D}_A$ is not a function of temperature.

In the case of the four-quadrant multiplier, there are two such circuits working in conjunction, so we must define two area ratios

$$\gamma_1 = \frac{I_{S2}I_{S4}}{I_{S1}I_{S3}} \quad \text{and} \quad \gamma_2 = \frac{I_{S5}I_{S4}}{I_{S1}I_{S6}}. \tag{13}$$

The total distortion (with respect to the x-input) will now be a function of y. For example if $Q1$-$Q2$-$Q3$-$Q4$ match perfectly ($\gamma_1 = 1$) but $Q1$-$Q5$-$Q6$-$Q4$ do not ($\gamma_2 \neq 1$), there will be no distortion when $y = 1$, becoming maximal when $y = 0$.

The output can be expressed as

$$Z = XY + 2yD_{A1} - 2(1 - y)D_{A2} \tag{14}$$

where

$$D_{A1} \approx x(1 - x)(1 - \gamma_1)$$

and

$$D_{A2} \approx x(1 - x)(1 - \gamma_2). \tag{15}$$

It will be seen that the linearity of Z with respect to the y input is not affected by area mismatches.

For the purposes of demonstration we can consider the cases where $Q1$ and $Q4$ match perfectly, but

1) $Q2$ and $Q3$ have the same mismatch as $Q5$ and $Q6$, that is $\gamma_1 = \gamma_2$;
2) $Q2$ and $Q3$ have the opposite polarity mismatch of $Q5$ and $Q6$, but the same magnitude, that is $\gamma_1 = 1/\gamma_2$, or $\gamma_1 \approx -\gamma_2$.

In the first case,

$$Z = XY + 2x(1 - x)(1 - \gamma)(2y - 1). \tag{16}$$

When the y input is balanced, $Y = 0$, $y = \frac{1}{2}$. Thus $Z = 0$ for all values of X. Stated differently, the null suppression with respect to the X input is unaffected by this mismatch situation.

The general form of the transfer curves and distortion products for this case is shown in Fig. 2(a), which also shows that the common point of intersection P (where $dZ/dy = 0$) is shifted to $X = (1 - \gamma)/(1 + \gamma)$, $Z = 0$. Notice also that the nonlinearity is always of the same sign as the output slope, and *varies in proportion* to it.

For the second case

$$Z = XY - 2x(1 - x)(1 - \gamma) \tag{17}$$

which corresponds to a constant parabolic distortion component *added* to the signal. In this case, when the Y input is balanced, there is a *residue* on the output of peak amplitude $0.5(1-\gamma)$. The point P is thus at $X = 0$, $Z = -0.5(1-\gamma)$, as shown in Fig. 2(b).

The general co-ordinates of P are

$$P(X, Z) = \left(\frac{1 - \sqrt{\gamma_1\gamma_2}}{1 + \sqrt{\gamma_1\gamma_2}}, \frac{\gamma_2 - \sqrt{\gamma_1\gamma_2}}{\gamma_2 + \sqrt{\gamma_1\gamma_2}}\right). \tag{18}$$

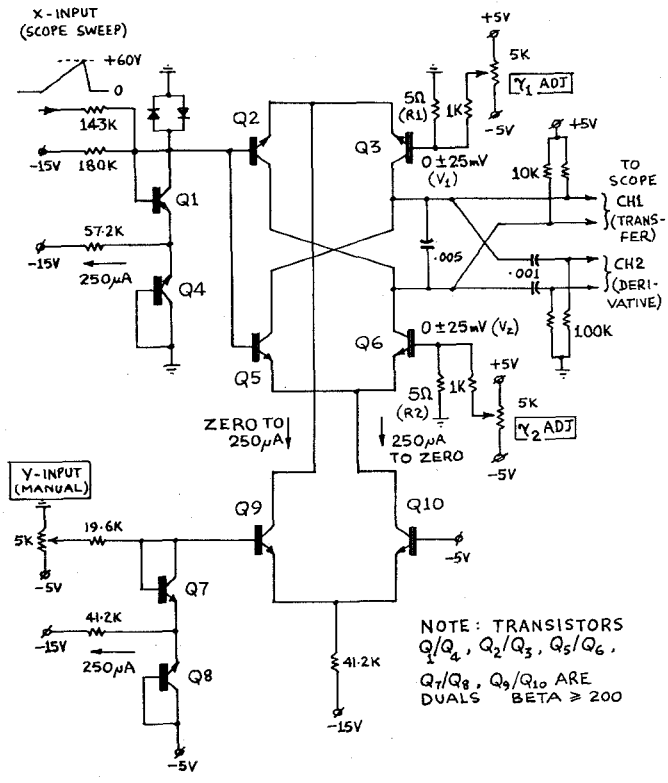


Fig. 3. Experimental circuit for investigation of nonlinear effects.

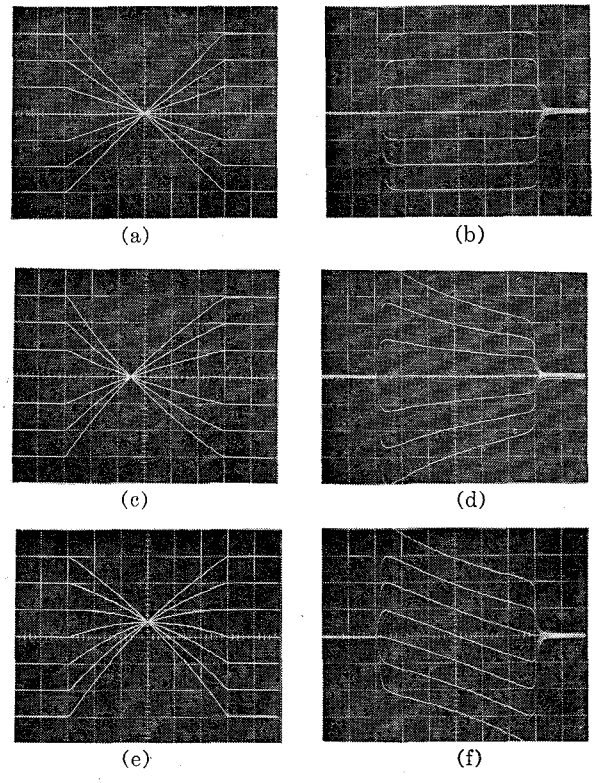


Fig. 4. Demonstration of distortion due to area mismatches. Scales are arbitrary.

Verification

To verify the above theory, and demonstrate the two cases discussed, a circuit was built as shown in Fig. 3. Use was made of the equivalence of area mismatches and offset voltage. The equivalent areas of $Q2$, $Q3$, $Q5$, and $Q6$ could be varied by the bias voltages V_1 and V_2, giving

$$\gamma_1 = e^{qV_1/kT}$$

and (19)

$$\gamma_2 = e^{qV_2/kT}.$$

The devices were operated at low powers ($I_B = I_E =$ 250 μA, $V_C = 2.5$ volts) so that the junction temperatures were close to 300 °K. The use of low operating currents also eliminated the distortion due to ohmic resistances, discussed later.

To demonstrate the nonlinearity more clearly, a linear ramp was used as the X input, and a simple R–C differentiator produced a waveform corresponding to the incremental slope of the transfer function. This technique provides a very convenient sensitive measurement of distortion, and became a valuable tool during the investigation of improved multiplier designs, without which it would have been necessary to resort to tedious point-by-point DVM measurements to reveal the nonlinearities.

Fig. 4(a) shows the transfer curves with V_1 and V_2 adjusted for minimum distortion, and Fig. 4(b) are the derivatives. Seven static values of Y, from -1 to $+1$, are shown. These demonstrate the excellent linearity that can be achieved with well-matched transistors. The departure from constant slope is within $+0$ -1 percent over 75 percent of the dynamic range. In terms of the nonlinearity term D_A (which is a measure of the deviation from the ideal line), this amounts to less than 0.3 percent at any point.

With $V_1 = V_2 = -10$ mV, ($\gamma_1 = \gamma_2 = 1.47$) the theoretical point of intersection is shifted to $X = -0.19$. The actual point is at -0.18, as shown in Fig. 4(c). Notice that the slope [Fig. 4(d)] falls as X varies from -1 to $+1$, starting 30-percent high and finishing 30-percent low. The deviation from the ideal line is now about 8 percent; of course, an offset voltage this large would be exceptional.

With $V_1 = +10$ mV, $V_2 = -10$ mV ($\gamma_1 = 1.47$, $\gamma_2 = 0.68$) the theoretical value of Z at $X = 0.0$ should be -0.197. This is close to the value of -0.185 measured from the waveforms of Fig. 4(e). An interesting feature of the derivatives shown in Fig. 4(f) is the one for $Y = 0$. Its linear form confirms the parabolic shape of the distortion term.

IV. Distortion Due to Ohmic Resistances

Fig. 5 shows the circuit with the addition of linear resistances in the emitters of all the transistors. These represent all the bulk resistances of the diffusions, particularly the base resistance, referred to the emitter circuit. In fact, these elements will be current dependent, due both to crowding effects and beta nonlinearities. However, if the device geometry is such that the cur-

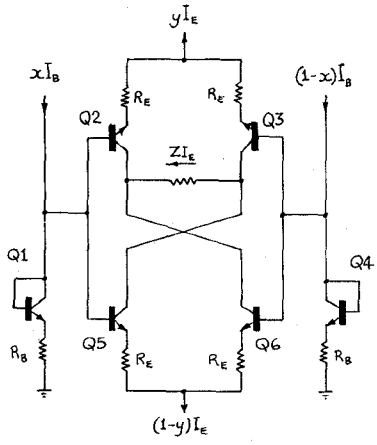


Fig. 5. Circuit having ohmic emitter resistances.

rent-density distribution is equalized in the appropriate sets of devices, the current dependence can be neglected.

Using the variables shown in Fig. 5, the loop equation for the quad $Q1$-$Q2$-$Q3$-$Q4$ under these conditions becomes

$$\frac{kT}{q}\log\left[\frac{x(1-a)}{a(1-x)}\right] = I_BR_B(1-2x) - yI_ER_E(1-2a) \quad (20)$$

which has no explicit solution for a in terms of the other variables. However, guessing that the distortion will be small, we will make the substitutions

$$a = x + D_R \quad (21)$$

and

$$\log(1-D_R) \approx -D_R$$

where D_R is the fractional distortion due to resistances. Equation (20) simplifies to

$$\frac{kT}{q}\frac{D_R}{(x+D_R)(1-x)} = yI_ER_E(1-2x-2D_R) - I_BR_B(1-2x). \quad (22)$$

Solving for the distortion term, and changing input variable from x to X,

$$D_R = \frac{q}{kT}(y\phi_E - \phi_B)\,\Delta(X) \quad (23)$$

where

$$\phi_E = I_ER_E$$

and

$$\phi_B = I_BR_B, \quad (24)$$

these representing the extra voltages in the emitter and base circuits due to resistances, and

$$\Delta(X) = \tfrac{1}{4}X(X^2-1) \quad (25)$$

which describes the form of the distortion, and has peak values of ± 0.096 at $X = \pm 0.577$, and zeros at $X \pm 1$ and 0.

Equation (23) makes the reasonable assumption that ϕ_E and ϕ_B are small compared to kT/q. For example, assume $R_E = 1$ ohm and $I_E = 2.5$ mA, giving $\phi_E = 2.5$ mV, about 10 percent of kT/q at 300°K.

Using the above approximate analysis, we can state a rule of thumb for the peak magnitude of D_R, for the quad $Q1$-$Q2$-$Q3$-$Q4$:

$$\hat{D}_{R1} \approx \pm 0.37(y\phi_E - \phi_B) \quad (26)$$

for ϕ_E, ϕ_B in millivolts, at 300°K, and $\hat{D}_{R1}$ in percent. Similarly, for the $Q1$-$Q5$-$Q6$-$Q4$ circuit, we have

$$\hat{D}_{R2} \approx \pm 0.37\,\{(1-y)\,\phi_E - \phi_B\}. \quad (27)$$

The *net* nonlinearity will come from both circuits, and vary with the y input. The outputs of each quad (and hence the distortion terms) are also weighted by y and connected out of phase. Thus

$$\begin{aligned}\hat{D}_R &= \hat{D}_{R2} - \hat{D}_{R1}\\ &= \pm 0.37[\{y\phi_E - \phi_B\}y - \{(1-y)\phi_E - \phi_B\}(1-y)]\\ &= \pm 0.37(\phi_E - \phi_B)Y \quad \text{(percent)}, \end{aligned} \quad (28)$$

with the substitution of $Y = 2y-1$. The nonlinearities introduced by *balanced* emitter resistances can be summarized as follows.

1) The distortion with respect to the X input has a symmetrical form and is a fixed percentage of the output Z.
2) There is no distortion with respect to the Y input.
3) The common point of intersection of the transfer curves is always at $X = Y = Z = 0$.
4) No distortion arises when $\phi_E = \phi_B$.

Thus, quite large ohmic resistances can be tolerated (that is, it is possible to use devices with high base resistance and/or low beta), provided that the base and emitter voltage terms are balanced. By scaling the device geometrics in the ratio I_E/I_B, the closeness with which $\phi_E = \phi_B$ is then a matter of device matching.

In practice the resistors labeled R_B in Fig. 5 will not be equal. It can be shown that under these conditions there will be a residue in the multiplier output for $Y = 0$, having the S-shaped form described by (25), and having a peak amplitude (at 300°K) of

$$\hat{D}_R \approx \pm 0.05\,I_E(R_{E2} + R_{E3} - R_{E5} - R_{E6}) \quad (29)$$

where $\hat{D}_R$ is in percent, I_E in milliamperes, R_E in ohms. Notice that this residue term is independent of ϕ_B, a fact that has been experimentally confirmed. It will be apparent that linear emitter resistances reduce the output swing capability because in the limit (when the diode voltages are small compared to the "ohmic" voltages) the circuit becomes completely cancelling for all values of X or Y. Also, the case where these resistances are unbalanced will give rise to an equivalent offset on one or both of the inputs.

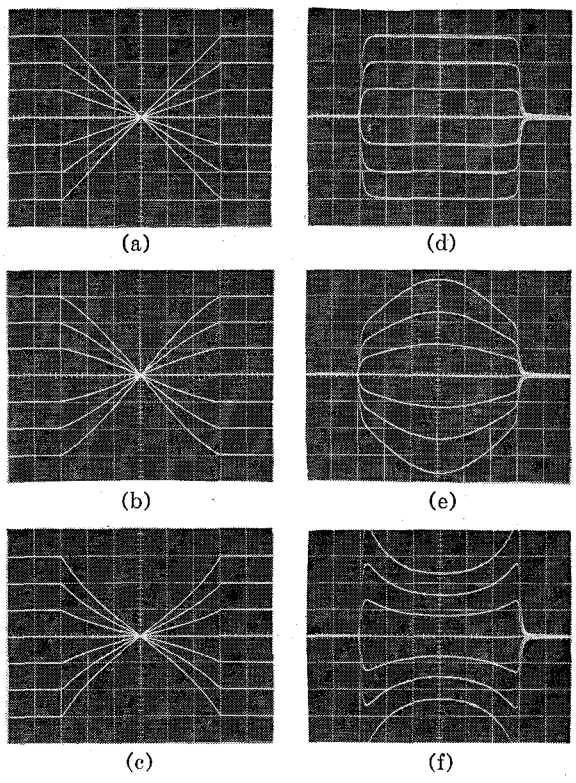


Fig. 6. Demonstration of distortion due to ohmic resistances.

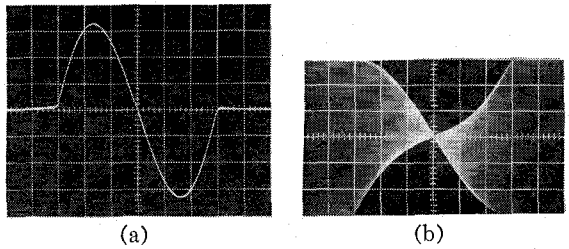


Fig. 7. Characteristic distortion due to mismatched resistances. (a) Vertical scale expanded to 0.33 percent/div. (b) 3.3 percent/div.

Verification

Using the test circuit of Fig. 3, to which emitter resistors were added, these nonlinearities were demonstrated. In Fig. 6(a), all resistors were 50 ohms and $I_B = I_E = 250$ μA; thus $\phi_E = \phi_B = 12.5$ mV. The derivative waveform, shown in Fig. 6(b), shows little degradation of linearity over the full dynamic range.

By omitting the resistors in the $Q2$-$Q3$ and $Q5$-$Q6$ emitters, a net error of $\phi_B = 12.5$ mV remains. Equation (28) predicts a nonlinear term of ± 4.5 percent at $Y = \pm 1$. The measured value is ± 3.3 percent. (Due to the approximations, (28) will err on the high side when ϕ_B or ϕ_E become comparable with kT/q). See Figs. 6(c) and (d). By omitting the resistors in the $Q1$-$Q4$ emitters, the distortion is of the opposite polarity, as Figs. 6(e) and (f) demonstrate.

The most typical distortion is due to the case where the ohmic voltages do not match, due probably to mismatches in r_b and beta. This can be demonstrated, too, by inserting the 50-ohm resistors in just the $Q2$-$Q3$ pair, when (29) predicts a peak distortion of ± 1.25 percent of full scale with the Y input balanced. The measured nonlinearity is shown in Fig. 7(a), in which the display was expanded vertically 50 times and the distortion has peak values of ± 1.1 percent. Fig. 7(b) gives the appearance of the distortion when the Y input was modulated to a depth of about 20 percent.

V. Distortion Due to Beta

The final imperfection to consider is that of finite beta. Three cases can be considered:

1) the transistors have identical, constant beta;
2) the transistors have differing, but still constant, beta;
3) the transistors have identical, current-dependent beta.

The first case was dealt with in [7] where it was shown that the only error is that the output current is reduced by the factor alpha (for I_E less than βI_B). In the version of the circuit driven by a single-sided input current (as, for example, the test configuration shown in Fig. 3), a small offset term also arises, and the dc output for $Y = 0$ is approximately

$$Z(X, 0) = (1 - \bar{\alpha}) \frac{I_E}{I_B} \tag{30}$$

where $\bar{\alpha}$ is the large-signal common-base current gain.

The second and third cases have not been completely analyzed, and it is doubtful whether explicit expressions involving all the betas and their nonlinearities would be of any value. Clearly, there is now the possibility for distortion terms to arise. However, the variations in beta from device to device, and over a small current range, are usually sufficiently small that no serious distortion should arise using typical transistors with betas in the neighborhood of 100.

VI. Thermal Distortion

The topic of thermal distortion in this category of circuits was dealt with in [7], where it was shown that theoretically no distortion arises due to the differential heating of devices if the power dissipation in the inner and outer pair are equal. This can usually be arranged, and, if necessary, the circuit can operate with I_E less than I_B.

In practice, using monolithic circuits the thermal distortion in response to a step input is very much less than 1 percent of the output amplitude, and persists for no more than a few microseconds.

VII. Transient Response

Because of the very small voltage swings at the inputs, and the cross connection of the transistors, the aberrations due to capacitances are very small, especially when properly balanced inputs are used. The main speed limitation is the f_t of the transistors.

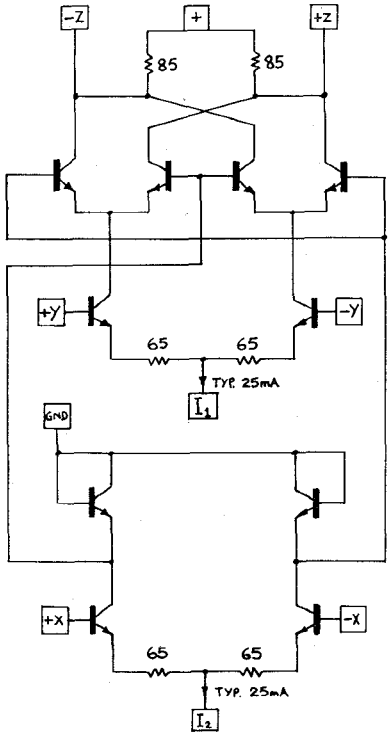


Fig. 8. Circuit used to examine high-frequency behavior.

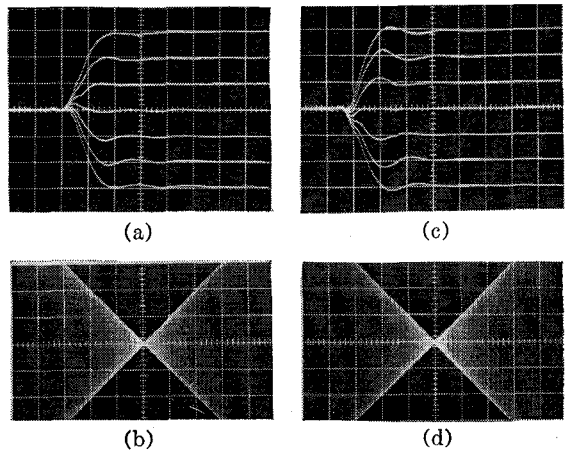


Fig. 9. Performance of integrated version of Figs. 8. (a) and (b). Transient response on X and Y inputs, respectively, at 1 ns/div. using dc control on other input. (c) and (d) 200-MHz carrier on X and Y inputs, respectively, staircase voltage on other input. Peak swing is 90 percent of full scale in all cases.

At $Y = \pm 1$, one of the transistor pairs $Q2$-$Q3$ or $Q5$-$Q6$ is producing all the output. The 3-dB bandwidth is, thus, about $f_t I_E/I_B$. At $Y = 0$, each pair receives $I_E/2$, and the bandwidth is doubled. We would, therefore, expect a risetime variation in response to a step on the X input of about two to one between these extremes. The step response to the Y input should be fairly independent of the X amplitude.

Measurements on an early integrated multiplier were made to examine the high-frequency behavior. The circuit, shown in Fig. 8, uses an "inverted" pair of input diodes [7], which are conveniently driven from pairs of emitter-degenerated stages for the X and Y inputs.

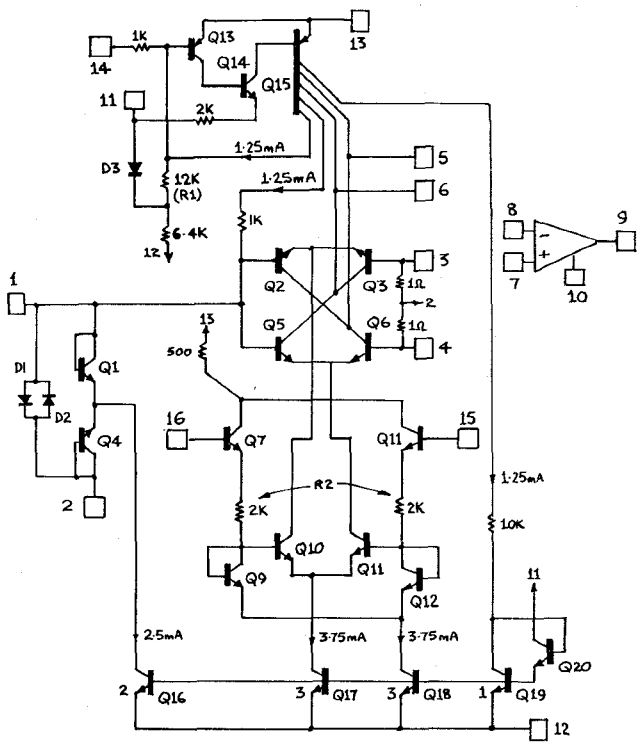


Fig. 10. Complete monolithic multiplier.

Transient response for each input is shown in Fig. 9(a) and (b). Figs. 9(c) and (d) show the CW response for a 200-MHz input, with the other input driven by the staircase output of the sampling time base in the oscilloscope used to examine the responses. The null suppression was better than 20 dB at 500 MHz.

VIII. A. Complete Monolithic Multiplier

Fig. 10 is the circuit of a complete multiplier suitable for integration. It is designed so as to be usable with a minimum of additional components to achieve medium-accuracy operation, or with extra components to perform at a higher accuracy. Wide-band operation (dc to >100 MHz) is available, or more versatility can be obtained by using the built-in operational amplifier to give division, squaring and square-rooting modes. These variations are possible by pin changes only. The X input is a single-sided current I_x into a summing point at ground potential, in the range 0 ±1 mA. The y input is a differential voltage V_Y into a high impedance (approx. 400 kΩ) in the range 0 ±5 volts. It can be shown that the output from pin 6 is

$$I_Z = \frac{I_x V_y}{5} \tag{31}$$

the scale factor being determined by the +15-volt supply and the ratio of R_1 to R_2. The diode D_3 ensures temperature stable scaling, and also makes the scaling factor proportional to the positive supply over a limited range.

Input and output current balance, and the rest of the circuit currents, are determined by the five-collector

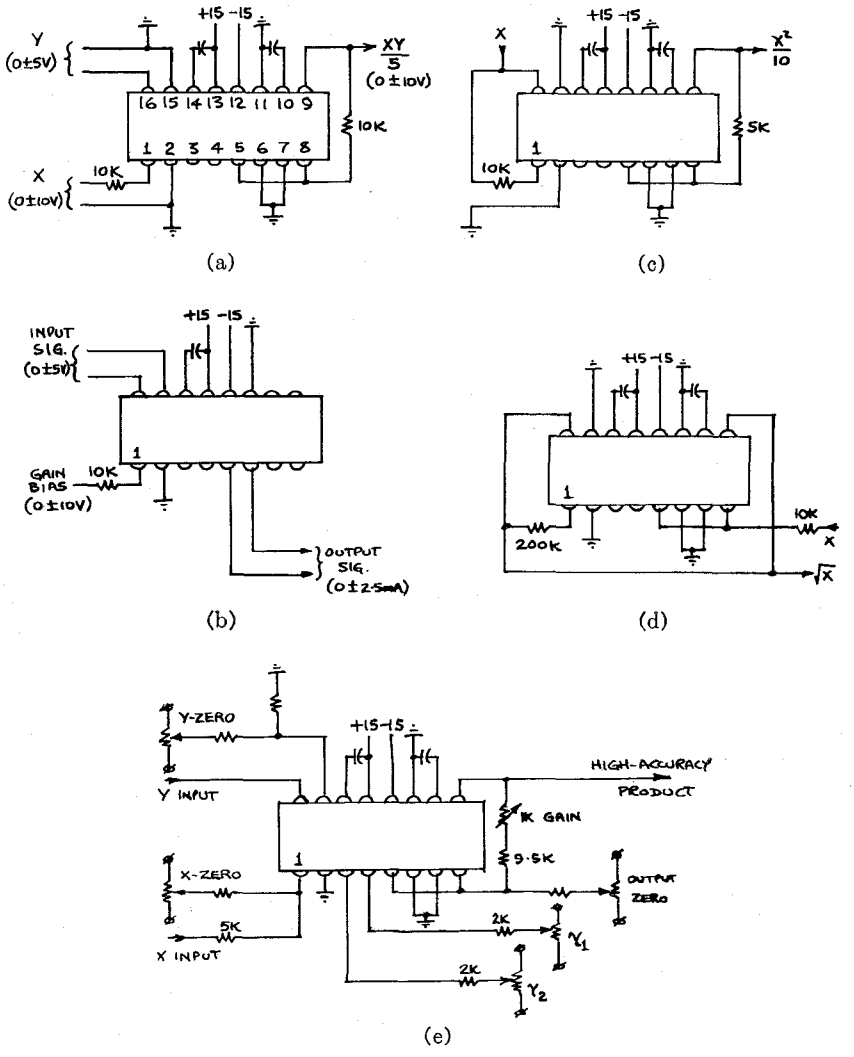


Fig. 11. Circuit of Fig. 9 connected as (a) medium-accuracy multiplier, (b) wide-band gain and polarity control, (c) squarer, (d) square-rooter, and (e) fully corrected multiplier.

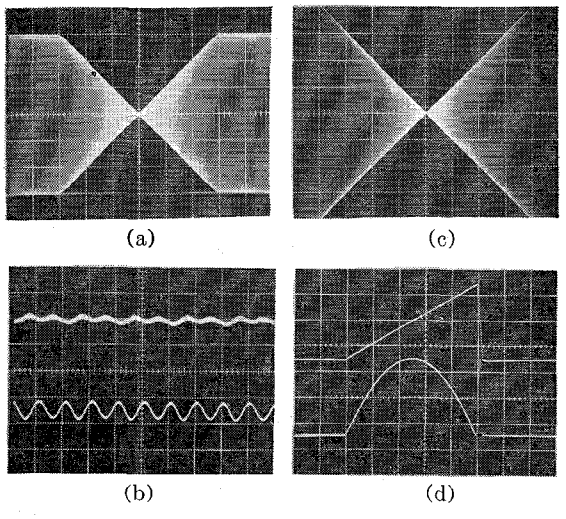


Fig. 12. Typical performance. (a) As balanced modulator, carrier frequency 5 MHz, peak output swing is 90 percent of full scale. (b) Output expanded ten times in vertical and horizontal axes—vertical now 1.67 percent of full scale/div. (c) Null suppression for full-scale 5-MHz carrier on X (upper trace) and Y (lower trace), expanded to 0.1 percent/div. (d) Offset ramp applied to both inputs produces the parabolic output, 1 μs/div.

lateral p-n-p, $Q15$. The matching of the currents to the five collectors (base diffusions) is vital to balanced operation. Measurements indicate that matching errors considerably less than ± 2 percent can be achieved. Notice that one of the collectors is connected in an operational configuration, through $Q13$ and $Q14$. This loop has to be stabilized by an external capacitor connected between pins 13 and 14. The second collector supplies a nominal 1.25 mA to balance the X input; collectors 3 and 4 supply the output balance currents; collector 5 sets up the current tails for the multiplier via $Q16$ through $Q20$.

Pins 3 and 4 give access to the bases of $Q3$ and $Q6$, allowing linearity-connection voltages to be applied. For perfect connection, these voltages should be proportional to absolute temperature. The aluminum 1-ohm resistors come close to this ideal, having a temperature coefficient of 0.38 percent per °K, slightly greater than the coefficient of kT/q at 300°K.

The operational amplifier increases the versatility of the device by permitting several modes to be imple-

mented. Pin 7 is normally grounded, and the inverted output from the multiplier, pin 5, is connected to pin 8; pin 6 is also grounded. Hence, the multiplier block $Q2$, $Q3$, $Q5$, $Q6$ works with a collector-base voltage of 0 $\pm$100 mV (the base voltage swing). The overload diodes, $D1$ and $D2$, must, therefore, be Schottky-barrier diodes having negligible conduction for most of the working voltage range at the X-input summing point, but being able to conduct heavily before the collector diodes of $Q2$ and $Q5$ under overload conditions. These may now be fabricated along with the standard silicon circuitry.

Fig. 11(a) through (e) illustrate the versatility of the circuit. The waveforms in Fig. 12 show linearity and null suppression at 5 MHz, and the output in the squaring configuration.

IX. Summary

A technique has been described that overcomes the inherent temperature dependence and nonlinearity of a transistor four-quadrant multiplier, and the feasibility of producing a complete monolithic multiplier with a worst-case linearity error of the order of 1 percent on either input has been demonstrated. Better linearity is possible by adjustment of transistor offset voltages. Bandwidths of over 500 MHz have been measured.

Acknowledgment

Thanks are due to the Integrated Circuits Group at Tektronix for the fabrication of many experimental circuits, and to G. Wilson and E. Traa [8] for helpful discussions.

References

[1] G. A. Korn and T. M. Korn, *Electronic Analog Computers*. New York: McGraw-Hill, 1956, pp. 281–282.
[2] G. S. Deep and T. R. Viswanathan, "A silicon diode analogue multiplier," *Radio and Electron. Engrg.*, p. 241, October 1967.
[3] H. E. Jones, "Dual output synchronous detector utilizing transistorized differential amplifiers," U. S. Patent 3 241 078, June 18, 1963.
[4] A. R. Kaye, "A solid-state television fader-mixer amplifier," *J. SMPTE*, p. 605, July 1965.
[5] W. R. Davis and J. E. Solomon, "A high-performance monolithic IF amplifier incorporating electronic gain-control," *1968 ISSCC Digest of Tech. Papers*, pp. 118–119.
[6] A. Bilotti, "Applications of a monolithic analog multiplier," *1968 ISSCC Digest of Tech. Papers*, pp. 116–117.
[7] B. Gilbert, "A new wide-band amplifier technique," this issue, pp. 353–365.
[8] E. Traa, "An integrated analog multiplier circuit," M.Sc. thesis, Oregon State University, Corvalis, June 1968.

Applications of a Monolithic Analog Multiplier

ALBERTO BILOTTI, SENIOR MEMBER, IEEE

***Abstract*—A fully balanced analog multiplier using differential transistor pairs is briefly described. Several circuit functions usually required in communication systems can be derived from the basic circuit. In particular, the different modes of operation leading to FM detection, suppressed carrier modulation, synchronous AM detection, and TV chroma demodulation are discussed. Experimental data obtained with a monolithic analog multiplier are also presented.**

I. Introduction

MANY CIRCUIT functions required in communication systems can be derived by way of analog multiplication. Fig. 1 shows a functional block consisting of an analog multiplier, symmetrical input limiters, and an optional output low-pass filter. When properly combined with passive networks, this block can perform FM detection, phase comparison, synchronous AM detection, amplitude modulation, and other functions based on frequency translation. This paper considers an integrated circuit that can be represented by the block diagram in Fig. 1.

Manuscript received June 5, 1968; revised September 25, 1968. This paper was presented at the 1968 ISSCC.

The author was with Sprague Electric Company, North Adams, Mass. He is now with the Faculty of Engineering, University of Buenos Aires, Buenos Aires, Argentina.

II. Basic Circuit

Fig. 2 shows the fully balanced arrangement of the three differential transistor pairs Q_1, Q_2, and Q_3 forming an analog multiplier block. The essential features of this circuit have been discussed elsewhere [1], [2] and the objective here is to achieve an understanding of the modes of operation possible, and through these modes, to consider a number of possible system applications [3]. Assume for a moment low-level driving at the inputs of V_1 and V_2. The current I_b established in the current source transistor is split in proportion to the applied voltage in transistor pair Q_1. This current division determines the bias and, therefore, the gain of the pairs Q_2 and Q_3. The output collector current summed in the load resistor R is proportional to the product of the two applied signals V_1 and V_2. The circuit topology is such that if $V_1 = 0$, the output currents due to a signal V_2 are of equal magnitude and opposite instantaneous polarity, giving a zero net output. The same is true for an applied