LARGE-SIGNAL
ELECTRONICALLY VARIABLE GAIN
TECHNIQUES

by

Max Wolff Hauser

B.S., University of California at Berkeley
(1979)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

MASTER OF SCIENCE IN
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1982

Signature of Author _____

Department of Electrical Engineering and
Computer Science, December 1, 1981

Certified by _____

James K. Roberge
Thesis Supervisor

Accepted by _____

Arthur C. Smith
Chairman, Department Committee on Graduate Students

LARGE-SIGNAL

ELECTRONICALLY VARIABLE GAIN

TECHNIQUES

by

MAX WOLFF HAUSER

Submitted to the Department of Electrical Engineering and Computer Science on December 1, 1981 in partial fulfillment of the requirements for the Degree of Master of Science in Electrical Engineering and Computer Science.

ABSTRACT

This work treats the generic problem of constructing analog gain blocks whose signal gain (or loss) can be modulated electronically. It consists of two major parts. A systematic survey of the known approaches forms the background for an investigation of a proposed refinement in bipolar-transistor variable-gain circuits.

In the survey portion, some important applications of variable-gain elements are studied, and desirable control-law and dynamic-range properties identified. The basic methods for performing this function -- based on solid-state devices and circuits -- are then examined and compared. This section considers, in a unified manner, many of the practical examples that have appeared in the literature. Circuits based on field-effect transistors and bipolar junction transistors are treated in depth.

Large-signal variable-gain blocks based on the logarithmic junction law of bipolar transistors (and related devices) exhibit errors due to imperfect logarithm conformance over wide ranges of current. The origins of these errors are reviewed. Experimental results demonstrate the errors obtained from small-area transistors operated at collector currents in the low milliampere range.

A proposed correction strategy leads to a class of correction circuits having a four-transistor core, with junction current densities that are related by a ratio parameter. As the current-density ratio grows large, the circuits can correct increasingly well for high-order effects like current crowding and high-level injection in the base. These correction circuits are potentially useful in a variety of situations where log conformance is important -- including, but not limited to, variable-gain circuits. One attractive form of correction circuit uses transistors with ratioed emitter areas.

Applications and design considerations are explored. Experi-

ments with a simple form of the correction circuit, applied to a variable-gain amplifier, show an order-of-magnitude improvement in signal-path linearity at high collector currents.

Thesis Supervisor: James K. Roberge

Title: Professor of Electrical Engineering

TABLE OF CONTENTS

Page

Page

Page

CHAPTER 1

PRELIMINARIES

1.1 Introduction

This thesis is virtually two works in one. It considers the broad question of how to construct a gain path that can be adjusted electronically, and the uses for such a function. It also investigates the particular problem of reducing static nonlinearity in a bipolar-transistor variable-gain circuit operated over a wide range of currents.

The duality of purpose divides the thesis roughly into two parts. Chapters 2 through 6 deal with variable-gain techniques and applications. Chapters 7 and 8 treat a class of second-order effects in bipolar transistors, their consequences in variable-gain circuits, and ways to mitigate them.

The purpose of chapters 2-6 is both to put the specific investigation into perspective and to survey systematically the sundry variable-gain devices and circuits presently known. The many contributions in this area have come from diverse sources. It seems worthwhile to bring them all together, identify common themes, and consider their performance in relation to typical applications.

The emphasis throughout is on large-signal approaches with frequency response including DC (i.e., baseband). In addition, small-signal, narrowband, and AC-only techniques are treated briefly.

1.2 Brief History of the Subject

Large-signal variable-gain elements (VGEs) are based on active devices or phenomena whose large-signal mathematical descriptions provide a direct means for varying the transmission of a signal. This differs from the small-signal case, in which the signal path is linear only to the extent that the signal amplitude is infinitesimal.

Before the widespread use of semiconductors, large-signal VGEs were a rarity. For one thing, there are rather few basic mechanisms that yield the large-signal multiplication of an input with a controllable factor; whereas small-signal gain control may be obtained, in principle, from any device with a static nonlinearity. A second reason was that the large-signal linearity was not often needed for the tasks performed by the technology of the day.

With the advent of transistors and the increasingly high-performance analog systems they engendered, it became both convenient and necessary to use large-signal, baseband methods for controlling gain. It could not always be assumed that the signal was narrowband and that subsequent filtering would remove distortion and noise introduced by the VGE. Wideband AGC, analog multiplication, remote amplitude control and other applications demanded accurate, predictable control of gain that was in turn made possible by field-effect transistors, PIN diodes, and, later, bipolar transistors in suitable circuit configurations.

Since about 1970 the bipolar-transistor large-signal variable-gain circuits have had the broadest utility (except for high frequencies, where PIN diodes dominate). Such circuits enjoy the economics

of planar integrated-circuit fabrication. Because they depend on an accurate logarithmic junction law, good static performance requires attention to parasitic effects in the transistors and often restricts the available gain range. Emerging applications, such as open-loop remotely-programmable amplifiers, are taxing the limits of the existing circuits.

CHAPTER 2

VARIABLE-GAIN ELEMENTS: FUNDAMENTALS

2.1 Terminology and Static Characteristics

Figure 2.1 shows an arbitrary, memoryless device or system with two inputs S_a, S_b and one output S_o. The output is a continuous, instantaneous function of both inputs. This device might be described by a function

$$S_o = f(S_a, S_b) \tag{2.1}$$

or, under certain conditions, by a power series of the form

$$\begin{aligned} S_o = {} & C_{00} && \text{[constant term]} \\ & + C_{10}S_a + C_{20}S_a^2 + \ldots && [S_a \text{ feedthrough terms}] \\ & + C_{01}S_b = C_{02}S_b^2 + \ldots && [S_b \text{ feedthrough terms}] \\ & + C_{11}S_aS_b && \text{[multiplicative term]} \\ & + C_{12}S_aS_b^2 + C_{13}S_aS_b^3 + \ldots \\ & + C_{21}S_a^2S_b + C_{22}S_a^2S_b^2 + \ldots && \text{[higher-order terms]} \\ & + C_{31}S_a^3S_b + \ldots \end{aligned} \tag{2.2}$$

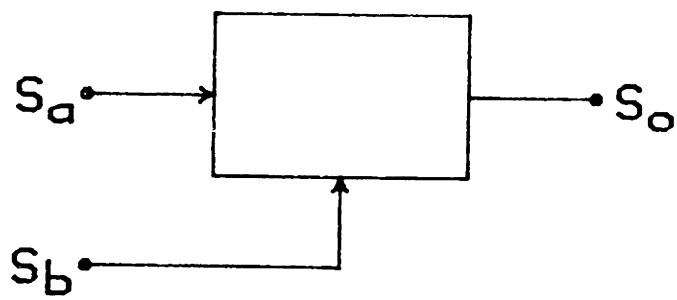


Figure 2.1 — Memoryless 2-input operator.

INPUTS AND OUTPUT MAY BE VOLTAGES, CURRENTS, ETC.

This thesis is concerned with practical realizations of a subclass of such systems, in which (2.1) has the special form

$$S_o = S_a g(S_b) \tag{2.3}$$

In other words, the output depends linearly on one of the inputs, termed the signal input, and has (in general) otherwise arbitrary dependence on the other input (the control input).

Thus, between the signal (S_a) input and the output there is linear gain (or loss) whose value depends on the control (S_b) input. This generic structure subsumes many practical devices, where the inputs and output are voltages, currents or some combination. Some typical names are voltage-controlled amplifier, voltage-controlled attenuator, current-controlled amplifier, analog multiplier, variable-gain amplifier, and so on. All of these terms describe special cases of what will be called a variable-gain element (VGE).

The gain of the VGE is $S_o/S_a = g(S_b)$. The notation $g(\cdot)$ will be used in reference to the whole function (as opposed to its value for a particular argument S_b). $g(\cdot)$ is the gain function of the VGE, also referred to as the "control law" or "gain law."

In most cases of practical interest, $g(\cdot)$ will be a monotonic function. Certain specific forms are of particular importance:

1. $g(S_b) = KS_b$, the class of analog multipliers
2. $g(S_b) = \frac{K}{S_b}$, the class of analog dividers

3. $g(S_b) = K_1 e^{K_2 S_b}$, the class of exponential VGEs

Note that, in principle, one gain function may be transformed into another by performing suitable operations on the control input.

Because the two inputs of the VGE have very different significances, it is customary to treat the block not as a two-input system but as a one-input system with a parameter, the control input. Consequently the characterization (with respect to signal distortion, overload limits, etc.) is not very different from that of an amplifier or attenuator. For example, harmonic signal distortion in a practical VGE might be approximated by

$$S_o \cong [S_a + c_2 S_a^2 + c_3 S_a^3]\, g(S_b)$$

In principle, the terms involving S_a^2 and S_a^3, along with an expansion of $g(S_b)$, may be put in the very general power-series form of (2.2). However, when dealing with observables like signal distortion, it is preferable to separate out the gain function whenever possible and treat the remainder of the system description in the same manner as an ordinary two-port.

2.2 Dynamic Characteristics

Similarly, issues of frequency response, transient response, frequency-dependent distortion and so on are treated by the familiar methods, with the control input as a parameter, as long as the control input is "static". A variable-gain element with a rapidly-varying control input will function as a modulator or mixer. This function,

indispensable in communications systems, requires a different kind of characterization, since the dynamic response of the output with respect to both inputs is important.* However, the present work is concerned with the usual implication of "variable gain"; that is, cases where only the signal-input frequency response is of concern.

The usual categories of small-signal/large-signal, broadband/narrowband, and so on, apply to VGEs, and refer to the signal input. Any device with a large-signal nonlinearity may be used as a small-signal VGE; the incremental slope of a driving-point or transfer characteristic is modulated by shifting the operating point. See section 4.1.

Narrowband applications also admit a wide range of variable-gain devices, both small-signal and large-signal. In a narrowband system, the presence of filtering relaxes the constraints on the VGE because it tends to remove noise and distortion that the VGE may introduce. This is the case usually treated in communications literature.

2.3 Feedthrough Issues

When the control input corresponds to zero gain the output should indeed be zero; similarly, the output ideally will contain no vestige of the control input when the signal input is zero and the control input is varied. Any deviation from zero output is referred to, respectively, as signal feedthrough or control feedthrough.

---

*So important that in many high-frequency mixers the use of minority-carrier devices of any kind is precluded, so none of the techniques in chapters 6, 7 and 8 are very useful.

Control feedthrough often occurs because of an imbalance in the circuit, either inherent (in the case of very simple VGE designs) or accidental (from component mismatch). Where component mismatch is the cause, there may be an effective input offset; that is, a DC signal input other than zero renders the output identically zero. Control feedthrough may be tolerated in those applications where control and signal inputs occupy well-separated frequency regimes, since the undesired component in the VGE output is readily filtered out. For this reason, the AGC elements in receiver circuits are often designed without regard to control feedthrough.

Signal feedthrough may be due to imbalances in a cancellation scheme or to the existence of parasitic paths around the variable-gain mechanism. The feedthrough of high-frequency signals can be especially troublesome; it can arise through capacitive parasitics and limit the frequency response of an otherwise attractive design. The worst form of high-frequency feedthrough occurs when there is capacitive coupling around an inverting gain stage. The composite transfer function will then include one or more right-half-plane zeros. These limit the attenuation at high frequencies and, because the two forward paths have opposing phase, lead to rapid phase variation that can be fatal in closed-loop applications.

This problem figures in the design of wideband VGEs from bipolar transistors; see sections 6.4.1 and 6.4.2.

2.4 Generic Techniques

The practical large-signal VGE approaches may be divided roughly into two categories. The "direct" techniques yield a signal

modulation because of their elementary structure, while the "algorithmic" methods synthesize it from simpler devices using some kind of mathematical identity.

Most "direct" VGEs work by varying the resistance of a conducting region through control over the number of available carriers. Examples are the channel of a FET in the nonsaturation mode, the photoresistor in a lamp/photocell unit, and the intrinsic region of a PIN diode. The two classical "algorithmic" approaches use quarter-square multiplication

$$AB = \frac{1}{4}\left[(A+B)^2 - (A-B)^2\right]$$

and logarithmic multiplication

$$AB = \exp(\ln A + \ln B)$$

In practice, square-law and (especially) logarithmic devices can be used in other convenient ways (this is discussed in chapters 5 and 6).

2.5 VGEs and Analog Multipliers

The analog multiplier is, formally speaking, only a special case of a variable-gain element. Nevertheless the terms carry distinct connotations and it is convenient to treat a multiplier as a different device with different applications.

2.5.1 Differing Objectives

Analog multipliers were originally designed exclusively for analog-computer use, although, like operational amplifiers, their use

has broadened, due largely to I.C. economics. They are employed in signal-manipulating circuits that require an explicit nonlinearity, such as RMS operators and graphic display drivers. They are used in communications circuits as high-performance balanced mixers up to high VHF or low UHF frequencies. Many of the VGE schemes in subsequent chapters, useful for applications other than multiplication (i.e., with a nonlinear control law), arose in the search for accurate, inexpensive, wideband multipliers.

One pivotal specification for a multiplier is the set of admissible input polarities. From the region in the plane formed the two input variables, it is conventional to refer to one-, two- or four-quadrant multiplication. Various tricks of offsetting or switching permit one- or two-quadrant versions to operate with additional input polarities. Variable-gain applications, by contrast, usually call for two-quadrant operation (bipolar signal, unipolar gain). Circuits for such applications are designed accordingly; with rare exceptions (see section 6.4.3), they are not constructed of one-quadrant cells.

Devices designed as general-purpose multipliers are rarely ideal for gain control; two specific factors weigh against them. First, such multipliers (which usually are four-quadrant devices) are designed with more-or-less interchangeable inputs and are optimized for minimum multiplication error with full-scale inputs. They do not perform well in applications requiring consistent signal-path performance over a wide gain range.

The second reason is the linear control law that multiplication implies. Only for a small subset of VGE applications is a linear law convenient. The reader will find more details about this in chapter 3, which treats typical applications.

2.5.2 Changing the Control Law

If a linear gain function is not desired, it is always possible in principle to modify this function by predistorting the control input. Satisfactory techniques exist for producing common nonlinearities (e.g., exponential, power or root) with bipolar transistors. Such an approach may be applied to any type of VGE.

In practice, the use of a general-purpose multiplier in this way is probably not justified since the added nonlinearity does not extend the useful gain range at the same time. On the other hand, custom-designed variable-gain circuits often exhibit naturally-nonlinear gain functions of useful kinds. Techniques that give, for example, exponential and reciprocal-law gain control with wide gain range are detailed in chapter 6.

2.5.3 Linearizing the Control Law through Feedback

It is possible to obtain multiplication from arbitrary variable-gain elements without explicit predistortion of the control input. The standard servo-type linearization is shown in Figure 2.2. A feedback amplifier forces the output of one VGE to equal an external input. If both VGEs have linear signal response and are well matched, the result is an accurate multiplication regardless of the original control function. Another way to view this is that the lower VGE, in

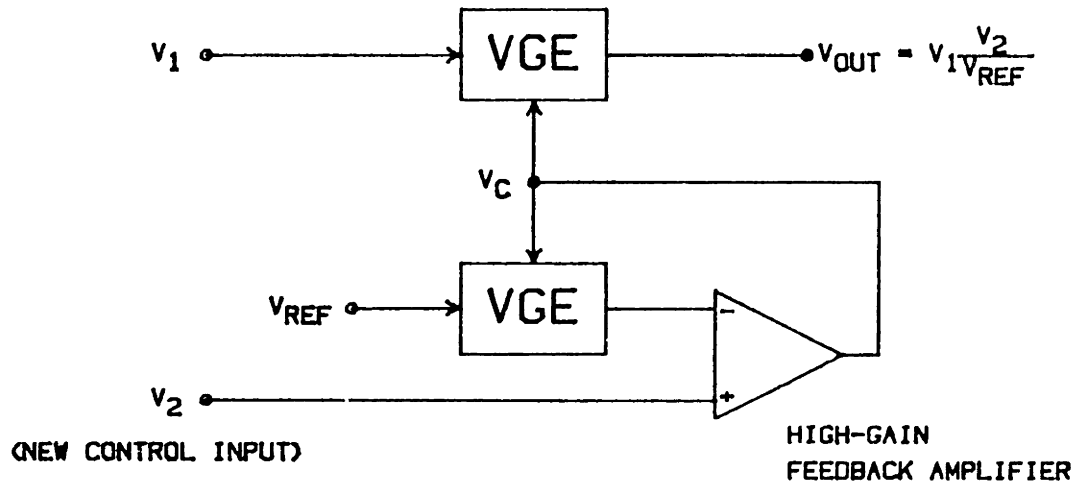


Figure 2.2 – Standard technique for linearizing the control law, using feedback and a second identical VGE.

a feedback loop, serves to predistort the control signal of the upper VGE, generating the exact inverse of the original control law.

The control law must be monotonic for this scheme to work, as it must be for any method to convert a VGE to a multiplier (otherwise the inverse of the control law is not single-valued). Bandwidth of the new control input will depend on the dynamics of the feedback loop; such systems are often slow. Accuracy of the linearized gain function depends on the static loop gain. This general structure appears in a number of published multipliers cited in later chapters. It is even possible to add additional "slave" gain paths by driving other control inputs from the output of the feedback amplifier.

CHAPTER 3

SOME APPLICATIONS AND ASSOCIATED ISSUES

In any systematic consideration of gain-control implementations, it is instructive to look at the applications for which the function is used. Although there are far too many specialized applications to be comprehensive, it is possible to say a little that applies to almost all applications and to say a lot about a few applications that, though specific, are very common.

3.1 Range-Compressing versus Range-Expanding Applications

Practical variable-gain elements exhibit maximum input levels, or input overload levels, beyond which they produce unacceptable signal distortion. The distortion may take the form of hard limiting (as in translinear circuits when the input current range is exceeded) or a gradual nonlinearity (as in certain FET circuits). In either situation, an input overload level may be defined to correspond with a given degree of acceptable distortion. This level will, in some cases, vary along with the gain of the element. In others it will not.

At the same time, many applications of VGEs fall into one of two categories. In range-compressing applications, the VGE is exploited to normalize a signal amplitude, compressing a wide range of input levels to a smaller range of output amplitudes. Automatic level control typifies this class. The second and complementary category, range-expanding applications, includes amplitude modulators,

dynamic-range expanders and the like. The salient point is that the overload level-versus-gain behavior of the particular VGE used must match the category of application in order to obtain the best performance.

In concrete terms, this means that a VGE used in a signal-modulation application should have an input overload level independent of gain (since the input amplitude will be constant), while a VGE for a signal-normalization task will ideally have an input overload level that varies inversely with gain (since the output amplitude is constant). In both of these cases the variation of maximum permissible input level with gain tracks the variation of expected input level with gain, so that at all times the VGE is operating with the maximum possible input level consistent with overload limits. A mis-tracking would result in degraded signal-to-noise ratio because of wasted input "headroom" or in a constricted gain range due to premature input overload.

This issue of tracking between device input limit and expected input signal will assume particular importance in the case of trans-linear gain control using bipolar transistors. The remainder of this chapter deals with some generic and frequent applications of VGEs, both to motivate the specific VGE realizations that follow and to identify other ways in which the application context determines VGE requirements.

3.2 Automatic Gain Control

Perhaps the most well-established application of variable-gain elements is to the automatic regulation of signal amplitude. Found in radio receivers, telephony circuits, audio equipment and many other situations, this function is pervasive enough to warrant some detailed examination.* It also demonstrates the utility of a non-linear control characteristic $g(V_c)$, and it is the quintessential case of a range-compressing application (section 3.1).

This function is usually called automatic gain control (AGC). A more descriptive name would be automatic level control, and indeed the term automatic volume control (AVC) was formerly used. However, "AGC" will be retained here because of its popularity.

The objective of automatic gain control is to adjust the amplitude of a signal so as to meet certain constraints. One common criterion, and the one emphasized in the following discussion, is that the average signal amplitude be regulated to a constant value. In effect, "slow" fluctuations of the original signal's amplitdue are to be removed without distorting the "fast" level changes that constitute useful information. Other criteria are sometimes used, especially in audio applications. Dynamic range compression seeks to reduce the variation of average signal level without eliminating it altogether. Amplitude limiting passes the signal unchanged if below a certain amplitude and regulates it if that amplitude is exceeded.

---

*Consider that a typical household contains at least one (feedback-type, narrowband, small-signal) AGC system embedded in an RF receiver; although the character of the demodulated signal may seem to scarcely justify the effort, such systems clearly are ubiquitous.

The differences among such categories lie in the relationship of input signal amplitude to output (controlled) amplitude and in the dynamics (response of the system to time-varying input amplitudes). Blesser [3.1] has examined in detail the considerations involved in compression for audio applications, and Blesser and Kent [3.2] have analyzed the dynamics of feedback-type audio limiters. The remainder of this section will treat a case encountered frequently in communications and instrumentation: systems to provide a constant output amplitude, independent of input level.

3.2.1 The Basic Problem

An observable signal is assumed to be the product of a prototype signal $S(t)$ and an "envelope" $V_1(t)$, which may be slowly varying or constant (but unpredictable). $V_1(t)$ is nonnegative. $S(t)$ has an "average" amplitude (i.e., average absolute value) of unity, with "average" defined to cover a duration of interest. $S(t)$ may be a baseband signal or it may be a modulated carrier, as in the case of AM communications. The objective is to recover $S(t)$ from the product $S(t)V_1(t)$ without distortion.

It should be noted in passing that this problem can be addressed directly by the theory of multiplicative homomorphic systems, which leads to numerical methods for separating two functions in different frequency ranges that have been multiplied together [3.3].

3.2.2 Feedback AGC and "Optimal" Control Laws

Figure 3.1 shows a block diagram of a basic feedback-type AGC system. The output signal V_{OUT} is regulated by detecting its average amplitude V_2, comparing this against a desired value V_{REF},

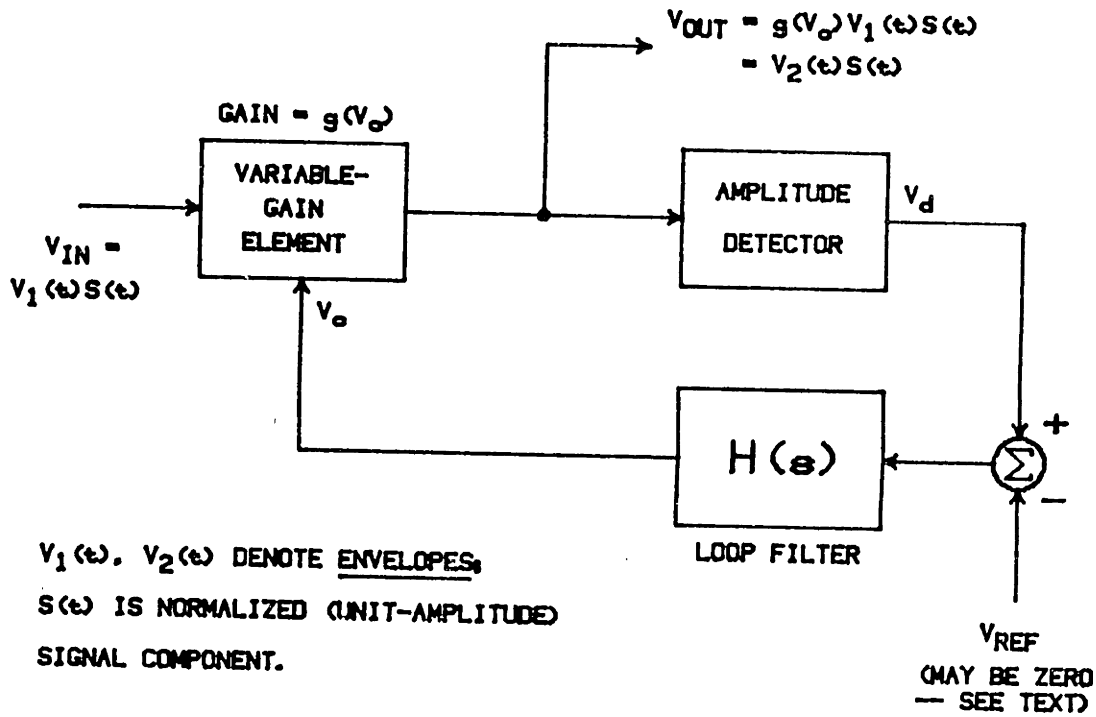


Figure 3.1 -- Simple, Generic Feedback-type AGC System

and adjusting the control input of the VGE to force V_2 towards V_{REF}. The loop filter, with a lowpass transfer function H(s), serves to define the time scale for amplitude averaging (hence the range of envelope frequencies that the loop will reject) and to filter out ripple components in the output of the amplitude detector. Both H(s) and the control law of the VGE influence the dynamics of the loop, as well as the accuracy of amplitude regulation.

The amplitude detector derives an estimate of the output amplitude $V_2(t)$. In this analysis, for convenience, H(s) is assumed to include all of the significant frequency dependence in the loop, so that the amplitude detector is memoryless. In a baseband system, the estimate will therefore consist of a rectified version of V_{OUT} and will contain substantial "ripple" due to S(t). Failure of H(s) to adequately "smooth" this estimate is one source of distortion in such a loop. More generally the amplitude detector may be a square-law or other nonlinearity, in which case some measure other than average amplitude of V_{OUT} is regulated. In a narrowband (modulated carrier) system there are additional options, such as multiple-phase rectification or quadrature (Pythagoras' law) detection [3.4], which reduce or eliminate the ripple problem. In some cases coherent (synchronous) detection may be used, with the great benefit of discriminating against wideband noise in V_{OUT}.

The AGC loop shown in Figure 3.1 is a nonlinear control system involving a mixture of control variables (V_c, V_d, V_{REF}) and modulated-singal variables (V_{IN}, V_{OUT}). In order to derive some simple small-signal results, the nonlinearity (due to the VGE) will be linearized around the large-signal operating point. To

avoid carrying the cumbersome S(t) factors, V_{IN} and V_{OUT} will be manipulated in terms of their amplitudes V_1 and V_2 respectively. If we assume

$$V_1 = V_{1q} + v_1$$

$$V_c = V_{cq} + v_c$$

$$V_2 = V_{2q} + v_2$$

where v_1, v_c and v_2 are incremental values, then the incremental (linearized) response of the VGE is

$$v_2 = v_1 g(V_{cq}) + v_c g'(V_{cq}) V_{1q} \tag{3.1}$$

where $g'(\cdot)$ is the derivative of the control function.

This linearization permits analysis of the small-signal loop properties as a function of large-signal operating point. As with any feedback system, the closed-loop dynamics are determined by a combination of H(s) and the additional gain in the loop. From (3.1), this additional gain depends heavily on the function $g(\cdot)$. The net small-signal loop gain is given by

$$T(s) = H(s) g'(V_{cq}) V_{1q}$$

The dependence of loop gain on $g'(\cdot)$ and the operating-point variables V_{cq}, V_{1q} leads to a useful criterion for choosing the function $g(\cdot)$. We require that T(s), and hence the small-signal

loop dynamics, be independent of the large-signal input level V_{1q}. This criterion is used, explicitly or implicitly, in most derivations of "optimal" control laws g(·).

There are good reasons for this choice. Obviously it is attractive to have the loop dynamics (i.e., the response to time-varying input amplitude) well defined and independent of the absolute input level. Moreover, in many situations the lowest-frequency components in the desired signal S(t) are sufficiently close in frequency to the undesired envelope $V_1(t)$ that the gain varies slightly in response to these components. Unless the loop dynamics are amplitude-independent, the result is a slight variation in the loop's group delay in response to S(t). For a narrowband signal, this gives phase modulation (PM); a spurious AM-to-PM conversion in the AGC section can be disastrous in certain communications systems.

The exact form of the "optimal" g(·) function depends on certain assumptions about the loop filter H(s).

3.2.3 Feedback AGC Systems with Averaging Loop Filters

There is a substantial body of theory applying classical control concepts to feedback-type AGC, all of it based on the tacit assumption that H(s) contains no integrations. Since H(0) has a finite value, such an AGC loop cannot achieve perfect steady-state regulation of the output amplitude. There will be a characteristic variation of V_2 as a function of V_1, with the degree of regulation depending on the control law g(·) and the zero-frequency filter transmission H(0).

A further consequence of this assumption is that V_{REF} alone does not determine the nominal output amplitude. Indeed, in early AGC systems, quite often $V_{REF} = 0$ was used and the characteristics of $g(\cdot)$ then set the output amplitude. A nonzero V_{REF} could be used to offset the control law to provide a threshold effect: the VGE operated "wide open," with no gain variation, until the input amplitude reached a certain level. This behavior, which can be very useful in certain applications, has traditionally been termed "delayed" AGC--an unfortunate usage since it has nothing to do with time delay.*

Oliver [3.5] established the basic small-signal theory of such AGC loops, in which H(s) is an "averaging" filter, typically a first-order lag. Others, such as Victor and Brockman [3.6], extended it to treat noise, time-domain effects, etc. The criterion of constant loop gain leads to an "optimal" control law of the form

$$g(V_c) = \frac{K}{V_c^n} \tag{3.2}$$

for the case of $V_{REF} = 0$ in Figure 3.1. This was demonstrated, for example, by McFee and Dick [3.7], and Vol'pyan [3.8]. If we assume for convenience that $H(0) = 1$, as in a passive "averaging" filter, then the output signal amplitude resulting from (3.2) is

$$V_2 = (KV_1)^{1/(n+1)} \tag{3.3}$$

where steady-state values are implied.

*"Terminological inexactitude"--Winston Churchill.

If the exponent n has a large value in (3.2), then $g(V_c)$ tends to drop from a large gain to zero in the vicinity of $V_c = 1$, so the loop adjusts V_{OUT} so as to make $V_c \cong 1$. A nonzero V_{REF} in this case causes an offset in the nominal value of V_{OUT}, as long as (3.2) still holds; in practice, however, $g(V_c)$ will saturate at some maximum value of gain, so that a "delayed" AGC results if V_{REF} is sufficiently positive.

The "optimal" control law of (3.2) is an example of a nonlinear and perhaps unlikely-looking gain-versus-voltage characteristic that is useful in practice. In this case the linear control law found in an analog multiplier is quite inappropriate. McFee and Dick pointed out that a common technique for small-signal gain control in bipolar transistors, using the g_m-to-I_c relation, leads to problems in radio receivers precisely because of its linearity. The small-signal gain obtained has a control law of the form

$$g(V_c) = (1-KV_c) \tag{3.4}$$

Used in a feedback-type AGC with an averaging loop filter, this yields loop dynamics that are substantially dependent on input amplitude. The loop "speeds up" under large input amplitudes and distorts the lower-frequency components of the desired signal.

Quite a different situation exists when the loop filter contains an integration.

3.2.4 Feedback AGC Systems with Integrating Loop Filters

If the loop filter transfer function H(s) contains a pole at the origin, the topology of Figure 3.1 will maintain exact steady-state regulation of output amplitude, regardless of the functional form of $g(V_c)$.* It is no longer necessary to calculate how the output level varies with steady-state input level, since it doesn't. The steady-state output level is determined entirely by V_{REF}, as the integrator will continually adjust V_c so as to force $(V_d - V_{REF})$ toward zero average value. Using the same criterion as in the previous case leads to a different "optimal" gain function--exponential rather than power-law.

This approach is generally preferred today for high-performance continuous-time AGC systems, because it is a simple way to obtain near-perfect amplitude regulation. The use of an integration in H(s) was not considered in the earlier AGC studies [3.5-3.8], perhaps because of the difficulty (or perceived difficulty) of building integrators in the technology of the day. Recent articles in the trade press by Hughes [3.9] and Porter [3.10] outlined practical design procedures for loops of both types.

The analysis of small-signal loop dynamics will be presented in this section, along with the derivation of an "optimal" control law for the VGE in the case of an integrating loop filter. The function $g(\cdot)$ will be assumed monotonically decreasing; this implies that $g'(\cdot)$ is negative, and hence (from (3.1)) a negative net loop gain is obtained without the need for an inversion in the loop

---

*Assuming that the required gain is within the gain range of the VGE, and that $g(\cdot)$ is a monotonically decreasing function.

filter. The algebraic sign of $g'(\cdot)$ is a small issue, since an inversion may easily be added elsewhere in the loop (between the amplitude detector and the VGE control input).

The small-signal analysis will use the variables defined in section 3.2.2. The amplitude detector is assumed to have unity gain, with all remaining ripple removed by the loop filter, so that $V_d = V_2$. Since V_c is the output of the loop filter, (3.1) gives

$$v_2 = v_1 g(V_{cq}) + v_2 H(s) g'(V_{cq}) V_{1q} \tag{3.5}$$

where lower-case v's denote incremental values. Rearranging gives the basic small-signal transfer function, in terms of Laplace transforms:

$$\frac{v_2}{v_1} = \frac{g(V_{cq})}{1 - H(s) g'(V_{cq}) V_{1q}} \triangleq G(s) \tag{3.6}$$

Note that $g'(V_{cq})$ is negative.

Reiterating, v_1 and v_2 are incremental <u>amplitudes</u> of signals at the input and output of the AGC loop. An optimal control law $g(\cdot)$, using the criterion of constant small-signal loop dynamics, is easily found from (3.6). Since the frequency dependence in (3.6) depends on the filter transfer function H(s) and the frequency-independent factor $g'(V_{cq})V_{1q}$, we require the latter to be constant (independent of large-signal operating point). But since the operating point is a steady-state condition, and since an integration in H(s) insures $V_{2q} = V_{REF}$ in steady state, we have

$$V_{1q}g(V_{cq}) = V_{2q} = V_{REF}$$

or

$$V_{1q} = \frac{V_{REF}}{g(V_{cq})} \tag{3.7}$$

Combining (3.7) with the requirement that $g'(V_{cq})V_{1q}$ be a (negative) constant yields the condition

$$\frac{g'(V_{cq})}{g(V_{cq})} = \text{CONSTANT (negative)} \tag{3.8}$$

Calling the constant $-1/V_o$, the general solution to the differential equation of (3.8) is

$$g(V) = Ke^{-(V/V_o)} \tag{3.9}$$

with V_o positive and K arbitrary.* The ideal of an exponential gain-control characteristic appears to be widely pursued in RF engineering. It was recognized by McFee and Dick [3.7], who criticized it as inappropriate (an assertion that was correct under the hypothesis of an "averaging" loop filter).

Equation (3.6) gives a general expression for the small-signal "amplitude transfer function" of the AGC loop in Figure 3.1. It is valid for both "integrating" and "averaging" loop filters.

---

*This result was first pointed out to me by David Hodsdon of the MIT Lincoln Laboratory.

In the case where the filter is a simple integrator with time constant τ, for example, a simple and reasonable transfer function results. If we take

$$H(s) = \frac{1}{\tau s} \tag{3.10}$$

then (3.6) gives

$$\frac{v_2}{v_1} = G(s) = \frac{sg(V_{cq})}{s - \frac{1}{\tau} g'(V_{cq})V_{1q}} \tag{3.11}$$

If we define $G_o = g(V_{cq})$ and $\omega_o = -g'(V_{cq})V_{1q}/\tau$ (note that ω_o is positive since g' is negative) then (3.11) becomes

$$G(s) = G_o \left(\frac{s}{s + \omega_o}\right) \tag{3.12}$$

which is a single-pole highpass transfer function. This loop is therefore highpass with respect to amplitude modulation: variations in input level at a rate much faster than ω_o are transmitted, while those at a rate much below ω_o are suppressed. The result is, of course, rigorous only for small variations about the quiescent input amplitude, but it reflects a reasonable type of behavior for an AGC system.

Equation (3.12) is also the transfer function of a single-pole RC lead circuit (or "RC differentiator") with $\omega_o = 1/RC$ (neglecting the gain factor G_o). Zelenz [3.11] examined the

effect of cascading AGC loops having this response. He noted a degradation in transient response that can have serious effects in communications systems, due basically to the highpass character of (3.12).

Little has been said here about the filter transfer function H(s), beyond the important distinction of whether or not it includes a pole at s = 0. There is considerable flexbility in choosing H(s), allowing for zeroes or an additional pole, for example. As with any feedback control system, the design of this filter is a crucial issue, determining such performance indices as sharpness of the "highpass" behavior and ability to track ramping, as well as constant, input amplitudes. A stability question will arise if H(s) contains more than one pole or if additional phase shift appears in the loop.

Moreover, there are many variations, not discussed here, on the basic loop of Figure 3.1. One of the most important is the pulsed or discrete-time AGC loop, arising whenever (as in pulsed radar) the input signal is not continuous in time. The treatment of such a signal involves modifying the circuitry so that V_c is updated periodically rather than continuously.

Such considerations are important in the design of a complete AGC system. Of more interest in the present context is the consequence of abandoning the loop format altogether, in favor of a feedforward topology.

3.2.5 Feedforward AGC Systems

A feedforward AGC is extremely simple in concept. An incoming signal's amplitude is measured, and the measurement

used to control a variable-gain element, which adjusts the signal gain appropriately (Figure 3.2). If the gain of the VGE is made to vary inversely with input amplitude, this structure will clearly yield a constant output amplitude. Perhaps more significant, the relation between input and output amplitudes may be explicitly and quite aribtrarily set (it need not even be monotonic) through the choice of the VGE control law $g(\cdot)$.

In contrast to feedback-type AGC, this approach was not much used until the advent of accurate, repeatable VGEs using bipolar transistors. As with all feedforward systems, it lacks the automatic verification of output behavior and concomitant immunity to component imperfection or drift. It also lacks stability problems, and great freedom is possible in the design of the filter transfer function H(s). Since it is no longer enclosed in a self-adjusting loop, H(s) can contain no poles at the origin. It will generally be an averaging filter that serves to define the separation between envelope and desired-signal frequencies, and to block the ripple components, if any, in the output of the amplitude detector.

If the amplitude detector renders an accurate estimate of the input amplitude V_1, and H(s) is a unity-gain averaging filter so that $H(0) = 1$, then the gain law necessary for constant output amplitude is

$$g(V) = \frac{V_o}{V} \tag{3.13}$$

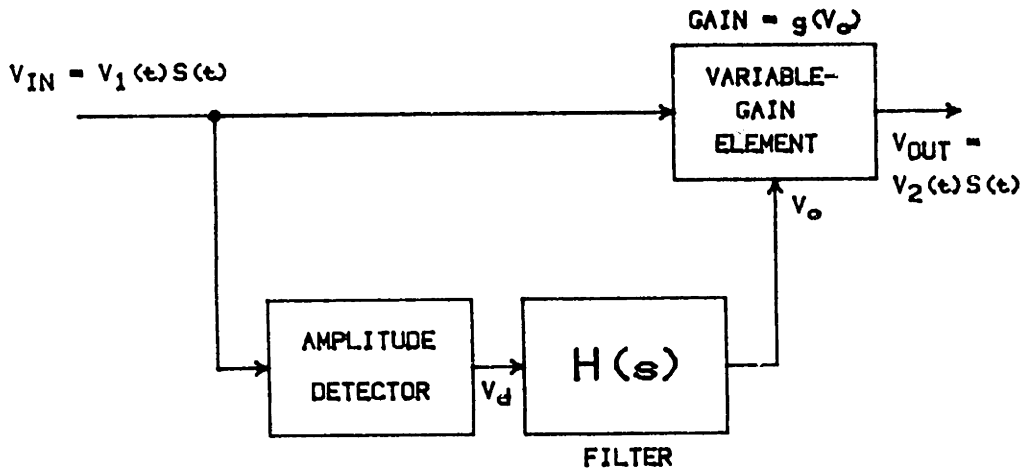


Figure 3.2 -- Feedforward-type AGC System

where V_o is now the desired output amplitude. For Figure 3.2, a small-signal analysis using (3.1) and (3.13) yields

$$\frac{v_2}{v_1} = q(V_{1q})[1-H(s)] \triangleq G(s) \tag{3.14}$$

and in the case of a single-pole lowpass filter,

$$H(s) = \frac{1}{1 + s/\omega_o} \tag{3.15}$$

this becomes

$$G(s) = G_o \left(\frac{s}{s + \omega_o}\right) \tag{3.16}$$

with $G_o = g(V_{1q})$--exactly the same frequency response as the feedback-type AGC with an integrator as the loop filter.

In principle, the feedforward AGC topology allows for a sharp-cutoff averaging filter, for accurate discrimination between envelope and signal frequencies. However, as with any feedforward system, there is a simultaneity issue: both input and control must arrive at the VGE in synchronism. Any "skew" will cause distortion in the output. A high-order averaging filter will exhibit, at best, substantial phase delay; worse, the different frequency components will tend to disperse if the phase response is not linear. This application suggests the use of maximally-flat-delay (so-called Besel or Thomson) filter characteristics.

The existence of bipolar-transistor large-signal VGEs with highly predictable gain-control laws has made possible high-quality audio AGC systems using feedforward topologies.*

*The products of dbx, Incorporated make extensive use of this technique.

In some cases, the output-versus-input amplitude relation is continuously adjustable, from zero output change, through equal output and input, to an output excursion greater than that of the input (dynamic range expansion). Such a capability is extremely difficult in a feedback-type AGC.

3.3 Oscillator Amplitude Stabilization

The classical sinewave oscillator of the resonant circuit type synthesizes a pair of conjugate imaginary poles and, by perturbing the positions of the pair, regulates the amplitude of oscillation. Clarke and Hess [3.12] identified at least four different ways of accomplishing this regulation; these may be divided into two basic groups, nonlinear and linear time-variant. The nonlinear method of amplitude regulation, by far the most common, employes the soft-limiting action of an (essentially memoryless) nonlinearity in the signal path. Although not strictly amenable to s-plane analysis, the nonlinear method, roughly speaking, achieves the pole-shifting by the dependence of incremental gain on amplitude in the nonlinear element. This is obtained at the expense of some waveform distortion and a shift of oscillation frequency away from the ideal value of a truly linear system. The shift is quantified by Groszkowski's Theorem [3.12].

The alternative approach, linear time-variant amplitude regulation, uses only linear elements in the signal path, but one or more is made to vary its gain in response to a control signal. The oscillation amplitude is sensed and used to drive the control signal, varying the circuit's pole locations (which migrate along the root locus) and hence the oscillation amplitude.

This has the advantage of very low waveform distortion and good frequency stability even with low-"Q" resonant elements. It is generally preferred for low (e.g., audio) frequency and low-distortion oscillators that must be tunable.*

The linear time-variant approach is an application of variable-gain elements similar to automatic gain control. As with the nonlinear approach, a pole-zero analysis is not strictly valid (the network parameters are varying with time). However, it is an excellent approximation since the rate of variation is normally much slower than the oscillation frequency.

Figure 3.3 shows a basic amplitude-stabilized oscillator in which the VGE follows the output. This is a range-expanding application, since the VGE's input amplitude is fixed and its output amplitude must rise or fall to compensate for the variations in the frequency-determining network. The alternative is to place the VGE just before the oscillator output, which would impose a range-compressing role. However, the version shown in Figure 3.3 has the advantage that any distortion in the VGE's signal path is filtered by F(s), which is typically a lowpass or bandpass function.

The dynamics of the amplitude-stabilization loop depend on F(s), H(s), and the control law g(·). In particular, there is usually an implicit integration in the small-signal transfer function between V_c and V_2. A small increment or decrement to

---

*Lest the nonlinear approach be slighted, it should be noted that based on that approach, some years ago, Messrs Hewlett and Packard founded an enterprise that was later to prove successful.

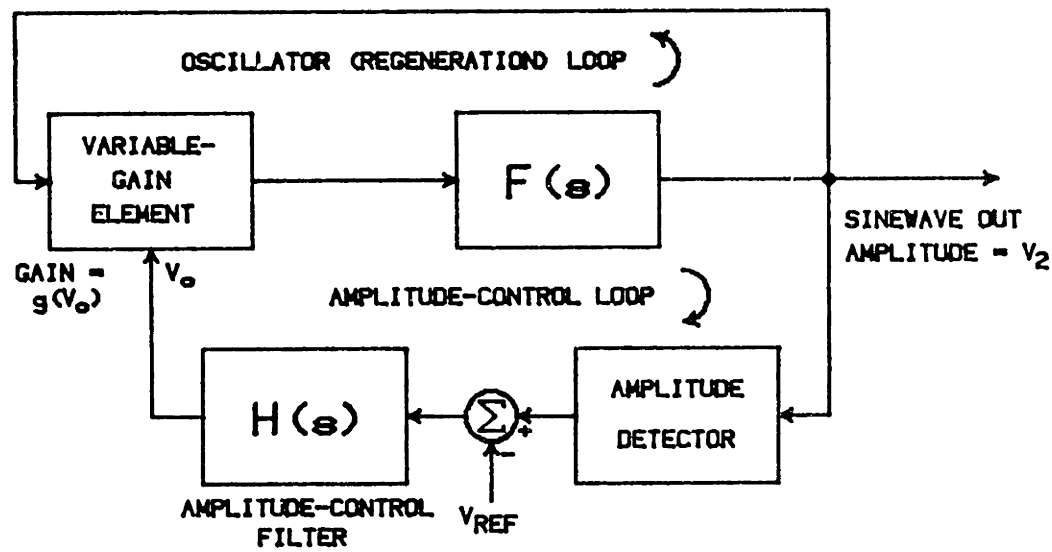


Figure 3.3 -- Amplitude-stabilized Sinewave Oscillator

Amplitude-control loop forces the output amplitude V_2 to equal V_{REF}. F(s) is the resonant or phase-shift network that determines oscillation frequency.

V_c causes the poles of the regeneration loop to move slightly into the left or right half plane, and therefore causes V_2 to begin ramping up or down. The regeneration loop may be analyzed in this manner, and represented by a certain transfer function from V_c to V_2. The H(s) response may then be chosen for the desired amplitude-stabilization properties. An example of this was treated by Roberge [3.13] in the case of a quadrature oscillator with a FET VGE.

Provided that H(s) contains an integration, it should be possible to use a wide range of VGE control laws in such an application.* Provided that the VGE has a very linear signal path, this technique can yield an extremely low distortion level in the output sinusoid. For this reason it is used in high-quality low-frequency laboratory oscillators.

3.4 Amplitude Modulation and Remote Amplitude Control

The use of a VGE in an amplitude-modulator role is a rather obvious possibility, and demands a linear control law (i.e., the VGE is an analog multiplier). This is a good example of a "range-expanding" application: the input amplitude is constant and the output amplitude changes according to the control (modulation) input. For such applications the appropriate VGE will have an input overload level that is independent of gain.

Remote amplitude control is, in effect, an electronic substitute for a potentiometer. This may arise in a

---

*According to David Hodsdon, MIT Lincoln Laboratory, considerations similar to those in section 3.2.4 lead to an "optimal" control law of the form g(V) = K/V.

computer-controlled data acquisition system, an automatic audio mixing console, or a programmable signal generator, for example. It will be either a range-compressing or range-expanding application depending on whether the output level or input level is fixed (respectively).

Typically, a substantial gain range is required from a VGE in such a situation. This tends to rule out elements whose attenuation is derived from the cancellation of two opposing signal paths. Although sometimes adequate for amplitude modulation, such structures (typically found in four-quadrant analog multipliers) exhibit unstable nulling and output noise that tends not to drop with gain. Moreover, the cancellation will often degrade rapidly with frequency, so that the useful gain range may be as low as one decade.

More successful are the fundamentally two-quadrant approaches based on two pairs of bipolar transistors operated at different current densities. A number of these are discussed in chapter 6.

3.5 Electronic Parameter Adjustment

The technical literature has included a plethora of circuits in which variable-gain elements impart electronic programmability by a straightforward replacement of fixed gain paths. Only a few representative cases will be considered here.

Afuso and Ishikawa [3.14] obtained an electronically-variable synthetic grounded inductance by means of a variable gyrator with a capacitive load. The gyrator was rendered variable by inserting a large-signal bipolar-transistor current steering

network in the signal path. The result was a current-controlled gyrator, which Afuso and Ishikawa operated in a voltage-controlled mode through series resistors.

Sparkes and Sedra [3.15] inserted off-the-shelf analog multipliers into second-order state-variable networks to obtain active filters with voltage-controlled parameters. Because the multipliers used had substantial internal phase shift (in an internal I-to-V conversion), some modification of the networks was required. This filter concept is a straightforward modification of a standard state-variable structure and has appeared many times. Recently, Fukahori [3.16] described a monolithic programmable filter with a similar architecture. Fukahori's device, however, merges the multipliers and integrators, yielding, in effect, a state-variable filter network whose integrators have programmable time constants. This avoids much of the phase shift introduced in earlier circuits by repeated voltage-current-voltage conversion using op amps.

From programmable linear filters it is but a small step to programmable sinewave oscillators (a step which, no doubt, many have taken inadvertently). Ryan [3.17] described an early application of a bipolar-transistor four-quadrant multiplier to a sinusoidal oscillator that could be tuned electronically over a fraction of its center frequency. Bhattacharyya and Viswanathan [3.18] produced an electronically variable inductor from a fixed inductor and a variable current gain cell (Figure 3.4). This was then used in a standard LC oscillator configuration. The technique

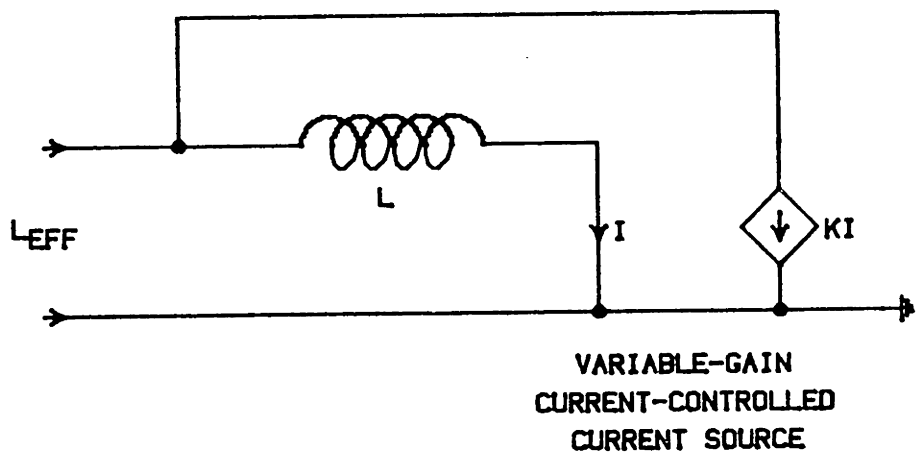


Figure 3.4 --

Electronically Variable Inductor

The current gain cell gives $L_{EFF} = L/(1+K)$.

After Bhattacharyya and Viswanathan [3.18]

could be used with other impedance elements as well. By selecting the proper control law, a linear frequency-to-control input relation can in principle be realized this way.

CHAPTER 4

MISCELLANEOUS GAIN CONTROL TECHNIQUES

Many different devices and circuits have been used for electronically variable gain, with varying degrees of success. Korn and Korn [4.1] give an extensive bibliography emphasizing analog multiplication, although it is somewhat dated (1972). Some of the approaches were considered and abandoned as analog multipliers, but still may be useful for other variable-gain functions or specialized applications. The important ones are included in this chapter.

Desirable attributes of a VGE include signal-input linearity, flat frequency response, low control and signal feedthrough, repeatability in manufacture, and compatibility with standard solid-state device fabrication. The techniques in this chapter are all deficient in one or more of these areas, especially the last. Most large-signal applications today are handled economically with the bipolar-transistor VGEs of chapter 6 or, less often, the FET techniques of chapter 5. However, some of the "miscellaneous" methods have particular strengths and maintain their importance despite this competition.

4.1 Small-Signal Methods

Although peripheral to the present study, the theory of small-signal gain control will be briefly reviewed because of its practical importance. It is encountered most often in narrowband (e.g., receiver) automatic gain control and in low-cost applications.

The theory has been treated extensively in the communications literature; the book by Clarke and Hess [4.2] summarizes typical cases.

Small-signal gain control works by assuming that the signal input is practically infinitesimal; then any device it is applied to will appear incrementally linear. The driving-point impedance or gain, for a small-signal input, is controlled by varying the large-signal operating point. Every conceivable device with a monotonic large-signal static nonlinearity has at one time or another been used this way: pn and Schottky diodes, bipolar transistors, field-effect transistors, vacuum tubes, and others. Sometimes variations on standard structures have superior properties for this application: remote-cutoff vacuum tubes and tetrode BJTs, for example. Other cases are discussed sporadically in subsequent chapters.

For a concrete example, consider a field-effect transistor in saturated operation. The relationship of drain current I_D to gate-source voltage V_{GS} in such a device can be approximated by

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_T}\right)^2 \tag{4.1}$$

where I_{DSS} and V_{GS} are parameters. The range of validity of (4.1) is for V_T between zero and V_T, typically (in a junction FET), so the characteristic is monotonic. If we now take $V_{GS} = V_1 + v$, where v is the incremental signal input, we obtain

$$I_D = I_1 + i,$$

where

$$I_1 = I_{DSS}\left(1 - \frac{V_1}{V_T}\right)^2$$

and

$$i = v\frac{dI_D}{dV_{GS}} = -2\frac{I_{DSS}}{V_T}\left(1 - \frac{V_1}{V_T}\right)v \qquad (4.2)$$

as long as v is "sufficiently small".

The last qualification is present in all such systems. Because small-signal gain control exploits the linear term in a Taylor series -- hence the derivative in (4.2) -- distortion will increase steadily as the signal input v increases. In wideband applications this constrains dynamic range, since the signal input amplitude is hemmed in by noise at one extreme and distortion at the other. In narrowband situations this is not so; much of the noise power will be filtered out, as will distortion components, although in-band intermodulation products may still limit maximum signal amplitude.

The above example includes control feedthrough (the I_1 component in I_D), although, as explained in chapter 2, this is often acceptable for a small-signal VGE. It can be eliminated with a balanced circuit configuration. Many small-signal VGE circuits using bipolar transistors employ a differential (emitter-coupled) pair for this and other reasons. Examples are given in chapter 6; also, an analysis by Yen [4.3] gives convenient closed-form distortion results.

4.2 Piecewise-Linear Approximation

An arbitrary nonlinear function of one variable can be approximated by a series of straight line segments. This is the classical technique for generating nonlinearities in analog computers, and it has been used to determine the waveform in variable-frequency oscillators, converting a triangle wave to an arbitrary shape (e.g., the Intersil 8038 voltage-controlled sinewave oscillator). The segment slopes and breakpoints are set with a series of resistor values, while clipping elements (e.g., diodes) provide the switching from one slope to the next.

This method provided accurate, temperature-insensitive squaring operators for the "quarter-square" multipliers used in analog computers; Korn and Korn [4.1] treat this topic extensively. These could be considered a variable-gain element. More directly, Wilkinson [4.4] and others used diode-resistor networks to approximate functions of two or more arguments, specifically multiplication.

These techniques are capable of good static accuracy when driven with known input amplitudes. Like other fixed-scale multipliers, they tend to have serious fractional distortion under wide gain-range operation. In particular, because of the piecewise approximation, they exhibit nondifferentiable (cuspate) rather than smooth signal distortion. Moreover, the array of slopes and breakpoints is clumsy to implement in low-cost (monolithic) form.

4.3 PIN Diodes

Semiconductor diodes with a p-intrinsic-n (PIN) doping profile provide a large-signal variable resistance to signals at

frequencies up to several Gigahertz. This behavior, however, depends on a large separation between control and signal frequencies. Signal components below some critical frequency limit, typically a few Megahertz, begin to modulate the resistance, and signal distortion results. Such a constraint is inevitable in a one-port VGE where the same port receives both signal and control, distinguished only by their frequency ranges.

The high-frequency resistance depends on low-frequency control current with a roughly 1/I characteristic. PIN diodes can accommodate high voltages (hundreds of volts) and substantial power levels (tens of watts, more if pulsed) when properly designed. They are most important in RF and microwave applications [4.5-4.7] where few devices can perform a variable-gain function with large signal excursions. Used as the resistive elements in a pi-type attenuator, they provide constant input and output resistance (when suitably driven) and an attenuation range of 40dB or more [4.8, 4.9].

Different device geometries allow various tradeoffs between minimum signal frequency and maximum control frequency. PIN diodes are confined to RF applications (Megahertz and above), so they are in a different class from most of the variable-gain elements discussed here. The others are capable of low-frequency and even DC operation.

4.4 Photoelectric Systems

One of the oldest forms of large-signal gain control consists of a light-sensitive resistor illuminated by a controllable light source. Early forms used incandescent bulbs, which suffered from

aging problems, poorly-controlled gain functions and slow control response. Thermal lag in the lamp generally imposed turn-on times in the tens of milliseconds. Sophisticated versions used tricks to improve this figure. Aiken and Swisher [4.10] achieved response times in the tens of microseconds, suitable for high-performance audio gain control, by adding a derivative component to the control current applied to the lamp.

More recently, light-emitting diodes have been used in place of lamps, with attendant improvements in dynamics and reliability. Edwards [4.11] incorporated an LED-photoresistor structure in a high-performance AGC circuit with very low signal distortion.

The compound structure of these devices requires expensive assembly and the components must be preselected to give predictable characteristics. Edwards circumvented the predictability problem, and the poorly-controlled gain law, with a feedback stabilizing technique. Since his signal input occupied a narrow frequency range, he was able to inject a DC component in the photoresistor and use it to sense the resistance, which was then forced to the desired value. This technique would, of course, be impractical if the signal and control bandwidths were to overlap, as is the case in more general applications.

4.5 Thermoelectric Systems

Heat-sensitive resistors are another natural device for variable-resistance gain control. These have been constructed from "discrete components", i.e. resistance wire and heating filament wound together. They suffer from predictability and time-constant problems

just as photoelectric versions do, with response times in the tens of milliseconds. Székely [4.12] described a more sophisticated version that was completely monolithic.

The use of a self-heating thermistor for audio AGC is one of the few cases where the thermal time constant works to the user's advantage. A thermistor with negative temperature coefficient, employed as a load resistor in an amplifier, tends to lower its resistance (and hence the gain) in response to increasing average signal power. The resulting regulation and dynamics are far from ideal, but the simplicity of a one-component AGC is remarkable. Kornev [4.13] reported accurate amplitude regulation and constant phase response over a 100:1 range of input amplitudes at 1KHz.

4.6 Modulated-Carrier Systems

There is a broad category of systems in which a high-frequency waveform or carrier is used to facilitate multiplication (or a related function) of two much-lower-frequency signals. Pulseheight/width or "time-division" multipliers, discussed at length by Korn and Korn [4.1], are the best-known of this class. With static accuracies as high as 0.01 percent, time-division multipliers are commercially important in the highest-accuracy low-speed applications. Holt [4.14] has described a low-cost monolithic implementation.

Many other carrier-type topologies have been developed. Brown's multiplier [4.15] derived multiplication and division from the exponential time waveform in a periodically-triggered RC lag circuit. Davis [4.16] used a pair of tuned amplitude-modulation circuits in a feedback-linearizing system (section 2.5.3). Each

modulator had an output amplitude that was the product of an amplitude-limiting voltage and some function of a gain-control input. Racoveanu [4.17] obtained a multiplier of limited accuracy and dubious stability by multiplying a (voltage-controlled) squarewave frequency with a (voltage-controlled) varactor capacitance. In another variable-frequency scheme, Kraus [4.18] used a fashionable switched-capacitor architecture for a multiplier. Packets of charge, proportional to one input voltage, were transferred to a summing node at a frequency proportional to the other input voltage. This gave an average output current proportional to the product of the two inputs.

These modulated-carrier systems share a drawback common to thermoelectric and photoelectric approaches: limited speed. Moreover, the limited frequency response usually applies to the signal input as well as the control input, rendering them inadequate for any but the slowest variable-gain applications. The heavy output filtering that is required to remove carrier components generally limits the signal frequency to about 1/100 of the carrier frequency [4.1], hence at most a few kilohertz.

CHAPTER 5

LARGE-SIGNAL GAIN CONTROL USING FIELD-EFFECT DEVICES

5.1 Introduction

Junction field-effect transistors (JFETs), metal-oxide-silicon field-effect transistors (MOSFETs), and related devices have a number of properties that make them useful in variable-gain functions, especially analog multiplication. They are subordinate only to bipolar transistors as solid-state devices for large-signal, wideband gain control.

The different ways in which FETs can be used reflect the distinct operating regions of the devices. In the nonsaturation region, the FET channel resistance can be modulated electronically, and this variable resistance--although inherently nonlinear--is a natural starting point for various signal-path modulation schemes. This is, at present, the major mode for FET-based variable-gain elements.

In the saturation region of operation, the FET loses its variable-resistor property, but its square-law transfer characteristic can be used indirectly to yield large-signal multiplication. The MOSFET has yet another useful mode, the subthreshold regime where its channel is weakly inverted. An exponential current-voltage characteristic results, which, while not as accurate as the corresponding relation in bipolar transistors, potentially can be exploited in the same way.

Finally, various field-effect devices other than conventional transistors have been explored as gain control elements. The last section of the chapter discusses some alternative structures that have been used or proposed.

Much of the work in this area occurred prior to the emergence, in the early 1970s, of bipolar-transistor "translinear" circuits as the dominant technique for analog multiplication. Some of the more recent work has been directed at the exploitation of MOS IC technology for analog circuits, especially in the context of analog charge-coupled devices (CCDs). The many FET variable-gain circuits in the literature are really variations on a few key ideas, and the major purpose of this chapter is to explore these ideas systematically. The development proceeds from the static current-voltage characteristics of the basic devices.

Throughout this chapter, n-channel devices are assumed for convenience, and the voltage and current polarities are chosen accordingly. For analytical flexibility, basic device equations are expressed both in terms of individual terminal voltages (V_D, V_S, etc., all referred to an arbitrary zero-volt reference node) and in terms of the more conventional interterminal voltage drops (V_{DS}, V_{GS}, etc.).

5.2 FETS in the Nonsaturation Region

In the nonsaturation region of operation (also called the ohmic, linear, prepinchoff, or triode region), a FET presents a relatively low resistance from drain to source. The drain-to-source voltage, or V_{DS}, is small enough that the FET channel is not fully depleted or "pinched off" along any part of its length; the drain

current then depends strongly on both V_{DS} and gate bias. Interaction between these two dependences forms the basis for multiplication or gain control in the nonsaturated mode. This interaction includes a rather accurate multiplicative term in the case of MOSFETs, while in JFETs there is a more complex characteristic that can approximate the same effect. Various ingenious schemes have been developed to cancel both undesired terms in the drain-gate interaction and process-dependent FET parameters that are unpredictable and temperature-sensitive.

Before examining actual circuits, it is worthwhile to review basic device behavior. The discussion will emphasize those aspects most important in electronically-variable gain applications

5.2.1 JFET Characteristics in Nonsaturation

The junction FET in nonsaturation is basically a voltage-controlled resistor, in which a conducting channel is bounded by depletion regions. The depletion regions are established by a reverse bias on the gate-to-channel pn junction; changing the bias modulates the width of the depletion regions and hence of the channel. Because the terminal properties of the JFET derive from the behavior of these depletion regions, it is necessary to specify the doping profile in the channel. From elementary pn junction theory, the voltage-dependent shape of the regions is a function of doping profile.

For a three-terminal n-channel JFET with a rectangular channel region, a constant channel dopant density N_d, a one-sided step junction between gate and channel regions, and symmetrical source and drain, the classical expression for drain current in nonsaturation is [5.1, 5.2]:

$$I_D = \frac{W}{L} q\mu_n N_d t \left\{ V_D - V_S - \frac{2}{3}\left(\frac{2\varepsilon_s}{qN_d t^2}\right)^{1/2} \left[(\phi_i - V_G + V_D)^{3/2} - (\phi_i - V_G + V_S)^{3/2} \right] \right\} \tag{5.1a}$$

in terms of individual terminal voltages V_D, V_S, V_G (with respect to an arbitrary reference), or

$$I_D = \frac{W}{L} q\mu_n N_d t \left\{ V_{DS} - \frac{2}{3}\left(\frac{2\varepsilon_s}{qN_d t^2}\right)^{1/2} \left[(\phi_i - V_{GS} + V_{DS})^{3/2} - (\phi_i - V_{GS})^{3/2} \right] \right\} \tag{5.1b}$$

in terms of V_{DS} and V_{GS}, where

W is the width of the channel, L the length

μ_n is electron mobility in the channel

t is metallurgical channel depth

$\phi_i = \frac{kI}{q} \ln\left(\frac{N_d}{n_i}\right)$ is the built-in voltage for the "one-sided" gate-channel junction

In addition, the "threshold voltage" $V_T = \phi_i - \frac{qN_d t^2}{2\varepsilon_s}$ can be defined: when V_{GS} reaches this (negative) value, the channel is entirely pinched off (depleted). The nonsaturation region, where the above equations apply, is delimited by

$$V_{DS} < V_{GS} - V_T$$

for positive V_{DS}, with a similar constraint (but roles of drain and source reversed) for negative V_{DS}. A further restriction is that the gate-channel junction not become sufficiently forward-biased that current flows across the junction.

Equations (5.1) follow directly from the original development of Shockley [5.3] some thirty years ago. The derivation makes use of the "gradual channel" approximation (in the depletion regions, electric fields are essentially perpendicular to the channel) and neglects series resistances in the drain and source paths due to the bulk regions of the semiconductor. Implicit in the derivation of (5.1) are two important assumptions: that the gate-channel junction has an abrupt doping profile, with uniform doping in the channel; and that the electron mobility μ_n in the channel is constant.

The 3/2-power behavior of the terms containing V_G in (5.1) is a consequence of the assumed doping profile. Different dopings in the channel and gate yield modified versions (5.1) with different functional forms. While it has been demonstrated that JFET characteristics can be rather independent of doping profile in saturation [5.4-5.7], this is less true in the nonsaturation region. Bockemuehl [5.8] considered the effects of certain limiting doping profile cases. Cobbold and Trofimenkoff [5.9] treated various combinations of abrupt and linearly-graded profiles, leading for example to a 5/3-power dependence in a simple linearly-graded case. For dopings more elaborate than a simple linear grading, the theoretical analysis generally becomes intractable and numerical techniques must be used. In modern JFETs with epitaxially-grown channel regions and (often) ion-implanted gates [5.2], the agreement with (5.1) can be good,

at least within the limits imposed by the constant-mobility approximation.

The approximation of constant mobility assumes that electron (or in p-channel FETs, hole) velocity is directly proportional to electric field. In practice, it has long been known that this velocity tends to saturate at high values of field, so that the effective carrier mobility falls off. The effect in JFETs is to produce a lower I_D than would otherwise be predicted, under large values of V_{DS}. Moreover, both the low-field mobility and the shape of the mobility-versus-field characteristic depend on the total impurity level (N_d+N_a) in a given region.

Trofimenkoff [5.10] proposed a practical, empirical model for field-dependent mobility, and derived corresponding JFET current-voltage characteristics. A more thorough derivation was given subsequently by Cobbold in his book [5.11]. Trofimenkoff's basic expression for mobility in the channel is

$$\mu = \frac{\mu_0}{1 + (E/E_0)} \tag{5.2}$$

where μ_0 is the low-field mobility, E is the local field magnitude along the channel, and E_0 is an empirically determined critical field value (typically in the range 10-20 KV/cm for silicon). In contrast to other formulations, (5.2) has proven both analytically convenient and well suited to the modelling of measured behavior in silicon.

Under certain assumptions, the resulting drain current expression, as derived by Cobbold [5.11], takes the form

$$I_D = \frac{I_D\Big|_{\mu_n=\mu_{n0}}}{1 + (V_{DS}/LE_0)} \tag{5.3}$$

where the numerator is the I_D of (5.1) and L is the channel length. This permits field-dependent mobility to be incorporated into (5.1) with slight modification, and also illustrates that low values of V_{DS} and long channels tend to minimize mobility-degradation effects.

Thus, both the doping profile and non-constant mobility influence the validity of (5.1). Additional refinements are possible; for example, a four-terminal JFET formulation in which the gate regions on either side of the channel are electrically independent. This has been treated in detail by Cobbold [5.11]. Such a modification may be necessary in monolithic JFETs whose channel region is bounded on one side by a common substrate,* although the effect will be minor if the substrate is lightly doped in comparison with the gate region. Another complication is the resistive bulk regions which inevitably separate the drain and source terminals from the active channel region, causing ohmic voltage drops.

With these limitations in mind, (5.1) will now be used to explore the basic suitability of JFETs as variable-gain elements. Although (5.1) is usually derived on the assumption that V_{DS} is positive, it applies also for V_{DS} negative, as a simple substitution will show. This assumes, of course, that the JFET remains out of saturation and the gate-channel junction is not forward-biased.

---

* This applies, for example, to FETs whose channel is formed in an epitaxial layer. Other structures, such as the ion-implanted shallow surface channel FET of certain compatible JFET-bipolar processes, may not have this complication.

We can therefore regard V_{GS} as a control signal and examine the relationship of V_{DS} to I_D at a given V_{GS}. A series expansion of (5.1) around $V_{DS} = 0$ gives

$$I_D = \frac{W}{L} G_x \left\{ \left[1 - \left(\frac{\phi_i - V_{GS}}{\phi_i - V_T} \right)^{1/2} \right] V_{DS} - \frac{1}{4} \frac{V_{DS}^2}{(\phi_i - V_{GS})^{1/2} (\phi_i - V_T)^{1/2}} \right.$$

$$\left. + \frac{1}{24} \frac{V_{DS}^3}{(\phi_i - V_{GS})^{3/2} (\phi_i - V_T)^{1/2}} - \ldots \right\} \qquad (5.4)$$

where $(\phi_i - V_T) = \frac{qN_d t^2}{2\epsilon_s}$ and $G_x = q\mu_n N_d t$.

In the limit as V_{DS} approaches zero, we can approximate this with the linear term

$$ID \cong \frac{W}{L} G_x \left[1 - \left(\frac{\phi_i - V_{GS}}{\phi_i - V_T} \right)^{1/2} \right] V_{DS} \qquad (5.5)$$

which is indeed a voltage-controlled resistance, but a small-signal one with a nonlinear control law.

In order to admit some nonlinearity in V_{DS}, a quadratic approximation can be used, and this is often done in practice. The above expansion yields

$$ID \cong \frac{W}{L} G_x \left\{ \left[1 - \left(\frac{\phi_i - V_{GS}}{\phi_i - V_T} \right)^{1/2} \right] V_{DS} - \frac{1}{4} \frac{V_{DS}^2}{(\phi_i - V_{GS})^{1/2} (\phi_i - V_T)^{1/2}} \right\} \tag{5.6}$$

The nonlinearity in the "parameter" V_{GS}, which may appear formidable, is in fact rather weak, except when V_{GS} approaches V_T. The detailed V_{GS} dependence in (5.6) is of limited value in practice since it follows from the idealized doping distribution assumed in (5.1). A much more popular approach is a semi-empirical expression, quadratic in V_{DS}, which suppresses <u>all</u> nonlinearity with respect to V_{GS}. It has the form

$$I_D \cong K \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \tag{5.7}$$

for some transconductance constant K.

Middlebrook [5.5] derived this using a simple charge-control analysis, without considering the device doping profile explicitly. Von Ow [5.12] obtained the same expression by fitting a quadratic function of V_{DS} to the presumed behavior of I_D at the origin ($V_{DS} = I_D = 0$) and at the onset of saturation ($V_{DS} = V_{GS} - V_T$, $I_D = I_{DSS}(1 - V_{GS}/V_T)^2$--see section 5.3). Von Ow compared this approximation graphically to the measured characteristics of a practical FET; according to his data, the agreement was within a few percent.

The approximation of (5.7) has propagated through the literature as a basis for analog multiplication and gain control; the basic idea is to exploit the $V_{GS}V_{DS}$ product term in (5.7), discarding the other terms by suitable circuit design. It must be remembered, however, that (5.7) neglects higher-order nonlinearities appearing in a more fundamental (and specific) result such as (5.1). The nonlinear V_{GS} dependence that appears in (5.1) but is neglected in (5.7) is an important facet in the comparison of JFETs and MOSFETs, to be considered in section 5.2.3.

5.2.2 MOSFET Characteristics in Nonsaturation

A Metal-Oxide-Semiconductor FET is not unlike a JFET (whose "gate" is the bulk or "body" of the MOSFET*) with an additional oxide-insulated gate terminal. As long as this gate is biased above threshold, a surface inversion layer exists directly beneath the gate oxide, forming a conducting channel of opposite polarity from the bulk region (an n-type channel, if the bulk is p-type). If the gate is biased substantially below threshold, the channel disappears and so does the analogy to a junction FET.

In both the JFET and the MOSFET, the shape of the conducting channel is influenced by an adjacent depletion region. However, in the MOSFET, conductance of the channel is controlled primarily by varying the amount of charge in the inversion layer rather than

---

* The term "substrate" is avoided here in reference to the bulk region of a MOSFET, both because the subscript "B" denotes this region or terminal and because the region is not necessarily the actual substrated of the silicon chip.

by directly modulating the depletion region. Such modulation can of course be accomplished through bias of the MOSFET bulk region, but this lightly-doped region forms a poor control gate and is rarely used for this purpose. Whereas in the JFET the channel is defined by explicit doping, the MOSFET channel is electrically induced. The dependence of I_D on V_{GS} in nonsaturation is a function not of an a-priori doping profile but of the charge-to-voltage relationship in the oxide; the dependence is therefore inherently linear.

Corresponding to (5.1) for JFETs is the n-channel MOSFET nonsaturation-region drain current expression from a distributed analysis [5.13, 5.14]:

$$I_D = \frac{W}{L}\mu_n \left\{ C'_{ox}\left[V_G - V_{FB} - 2\phi_{Fp} - \left(\frac{V_D}{2}\right) - \left(\frac{V_S}{2}\right)\right](V_D - V_S) - \frac{2}{3}(2\varepsilon_s q N_a)^{1/2}\left[(2\phi_{Fp}+V_D-V_B)^{3/2} - (2\phi_{Fp}+V_S-V_B)^{3/2}\right]\right\} \quad (5.8a)$$

in terms of the voltages V_D, V_S, V_G, V_B (with respect to an arbitrary reference), or

$$I_D = \frac{W}{L}\mu_n \left\{ C'_{ox}\left[V_{GS} - V_{FB} - 2\phi_{Fp} - \frac{V_{DS}}{2}\right]V_{DS} - \frac{2}{3}(2\varepsilon_s q N_a)^{1/2}\left[(2\phi_{Fp}+V_{DS}+V_{SB})^{3/2} - (2\phi_{Fp}+V_{SB})^{3/2}\right]\right\} \quad (5.8b)$$

in terms of node-pair voltages V_{DS}, V_{GS}, V_{SB}, where

C'_{ox} is gate oxide capacitance per unit area

V_{FB} is the flatband voltage, a constant determined by fabrication details

$\phi_{Fp} = \frac{kT}{q} \ln\left(\frac{N_a}{n_i}\right)$ is the Fermi potential of the p-type bulk silicon, i.e., the contact potential between intrinsic silicon and the bulk material

N_a is the bulk dopant concentration

and the subscript "B" refers to the bulk or body terminal. The flatband voltage V_{FB}[5.15, 5.16] allows for nonzero oxide and fixed surface charges and for a contact potential between the gate electrode and the bulk silicon. The other parameters are as in (5.1); note, however, that the correct value for μ_n in the present case is the surface mobility, which is lower than the bulk mobility by about a factor of two.

Equations (5.8) again neglect both high-field mobility degradation and extrinsic source and drain resistances. The corresponding threshold voltage, which depends on bulk bias, is

$$V_T = V_{FB} + 2\phi_{Fp} + \frac{1}{C'_{ox}} [2\varepsilon_s q N_a (2\phi_{Fp} + V_{SB})]^{1/2} \qquad (5.9)$$

The condition for nonsaturation is more complicated for a MOSFET than for a JFET. Instead of a limit on V_{DS}, there are constraints on both drain and source voltages with respect to the bulk:

$$V_{SB} < V'$$

$$V_{DB} < V'$$

where

$$V' = V_{GB} - V_{FB} - 2\phi_{Fp} - \frac{\varepsilon_s q N_a}{C'^2_{ox}} \left\{ \left[1 + \frac{2C'^2_{ox}}{\varepsilon_s q N_a} (V_{GB} - V_{FB}) \right]^{1/2} - 1 \right\} \tag{5.10}$$

Equations (5.8) and (5.10) provide a summary of device behavior in nonsaturation and are valid for both positive and negative V_{DS}.

Note that (5.8) is linear in V_{GS}, with the constant of proportionality $(W/L)\mu_n C'_{ox}$. There is a difference of 3/2-power terms, as in the JFET case, but these arise from the effect of bulk bias and do not depend on V_{GS}.

As with junction FETs, carrier mobility in the channel is reduced in the presence of high electric fields, and this must be taken into account if a more accurate expression than (5.8) is desired. The situation is somewhat different for MOSFETs since there is a substantial, usually dominant, electric field through the oxide, perpendicular to the channel. Because of this, mobility degradation sets in at high gate voltages even if $V_{DS} \cong 0$ (an example is plotted in Grove [5.17]). With JFETs, in contrast, the channel is bounded by depletion regions and the electric field in the channel is directed essentially between source and drain. Moreover, the MOSFET channel is located at or near an oxide-silicon interface,

and the presence of this interface greatly influences the mobility [5.13].

The effective surface mobility has been measured as a function of electric field through the surface, or, equivalently, as a function of induced surface charge. The results have been well summarized by Sze [5.13], Grove [5.18], and Cobbold [5.19]. The upshot is that effective mobility appears independent of surface field up to about 150 KV/cm for typical doping levels, and above this value the effective mobility decreases with increasing field. As was pointed out by Cobbold, the value of 150 KV/cm corresponds approximately to the onset of degenerate conditions in the inversion layer, and represents a very high level of channel conductivity. Unfortunately there is no exact theoretical description for this effect, and empirical methods must be used as with JFET mobility effects.

In one of the few multiplier studies to actually consider mobility degradation, Bosshart [5.20] has obtained the empirical approximation

$$\frac{\text{Effective Mobility}}{\text{Low-Field Mobility}} \cong \frac{1}{(1 + (V_{GS}/20)^2)^{1/2}}$$

for the data shown in Sze [5.13]. Bosshart used this to estimate nonlinearity in the I_D-versus-V_{GS} relation.

Along with this surface-field dependence, attributed to scattering effects at the oxide-silicon interface, the mobility is also degraded by high fields along the channel due to carrier velocity saturation. As with JFETs, large values of V_{DS} or short channels cause an additional mobility reduction. Klaassen [5.21, 5.22]

has demonstrated that both surface-normal field (V_{GS}-dependent) and longitudinal-field (V_{DS}-dependent) mobility degradation can be accurately modelled by an empirical expression of the form (compare (5.3))

$$\frac{\text{Effective Mobility}}{\text{Low-Field Mobility}} = \frac{1}{1 + \theta(V_{GS}-V_T) + (V_{DS}/LE_0)} \tag{5.11}$$

This may be used to predict distortion in a device where (5.8) is exploited for a gain-control function; it summarizes major deviations from the linear dependence of I_D on V_{GS}.

Also paralleling the JFET case, (5.8) can be approximated by an expression quadratic in V_{DS}. A simplified form that neglects the variation of potential along the length of the channel (i.e., assumes $V_D \cong V_S$) is

$$I_D \cong \frac{W}{L} \mu_n C'_{ox} \left[(V_{GS}-V_T)V_{DS} - \frac{V_{DS}^2}{2} \right] \tag{5.12}$$

which may be obtained directly from a charge-control analysis or from (5.8) in the limit $V_{DS} \cong 0$. The approximation becomes more accurate as the background concentration N_a is reduced or the oxide capacitance C'_{ox} is increased. Extensions of (5.12), replacing the leading factor $\frac{W}{L} \mu_n C'_{ox}$ by an empirical constant and adding an empirical coefficient in the V_{DS}^2 term, can achieve fair agreement with measurements for nonzero V_{DS} as well.

5.2.3 JFETs versus MOSFETs

Both JFETs and MOSFETs have been used in the nonsaturation mode for large-signal variable-gain applications, in circuits so similar that one might conclude the devices are completely interchangeable. The broad properties of MOSFETs and JFETs are indeed similar (compare (5.7) and (5.12), for example), but there are important and revealing differences in second-order behavior.

In comparing the two types, the objective here is not so much to recommend one over the other as to elucidate the key attributes of each. Often, in fact, the designer does not have a choice, so that a recommendation is not particularly useful.

Both devices exhibit a controllable, nonlinear channel resistance in which the dependence of current on voltage is approximately quadratic. Charge-control arguments show that this quadraticity is the natural tendency in a field-effect device regardless of construction details. Thus the same first-order expression is widely applied to both:

$$I_D \cong K\left[(V_{GS}-V_T)V_{DS} - \frac{V_{DS}^2}{2}\right] \tag{5.13}$$

The major differences from a circuit viewpoint are in the gate properties rather than the channel characteristics. This reflects the difference in control mechanisms. The JFET channel is built in during manufacture and selectively depleted of carriers by a reverse-biased pn junction, while the MOSFET channel is created and controlled by the field through an oxide layer.

As was cited in section 5.2.2, the MOSFET drain current contains a "clean" $V_{GS}V_{DS}$ product. This is because the conductivity in its channel region depends linearly on field through the oxide, hence on gate voltage. In the JFET, the relation between channel conductivity and gate voltage is determined by the electrostatic properties of the gate-to-channel depletion region, yielding a nonlinear function that varies with doping profile. The MOSFET, then, has an inherent superiority as a low-distortion multiplying element, at least until mobility degradation sets in.

Against this distortion advantage stand a number of practical problems in the MOSFET. Because it is influenced (through V_{FB}) by charges in the oxide and at the oxide-silicon interface, the MOSFET threshold voltage is difficult to control accurately and may even fluctuate with age. The device has a well-earned reputation for DC instability. Even today, after many fabrication problems have been overcome, it is rare to see (for example) MOSFETs used at the input of a discrete-component or hybrid operational amplifier, chiefly because of DC offset and instability problems.

Moreover, MOSFETs are four-terminal devices, and the disposition of the bulk electrode is an important issue. Bias on this electrode can affect drain-source characteristics ("body effect"), especially when the FET is being used as a floating "variable resistor." This is addressed in subsequent sections. JFETs, by contrast, may or may not be four-terminal elements. Although the classical JFET has one gate junction bounding each side of the channel, discrete versions and some monolithic types avoid having

separate gate terminals either by tying them together internally or by an asymmetric structure with control applied to only one side of the channel.

The following sections detail practical gain-control schemes that exploit the device behavior outlined above. The starting point in each case is the first-order model of (5.13). Many of these techniques are useful for both JFETs and MOSFETs; peculiarities of each are considered where necessary.

5.2.4 The FET as a Simple "Voltage-Controlled Resistor"

The basic expression (5.13) consists of two terms:

- A simple multiplicative term, I_D proportional to $(V_{GS}-V_T)V_{DS}$; and
- A V_{DS}^2 term which accounts for the curvature of the non-saturated I_D-V_{DS} characteristics.

The second term is very small at small values of V_{DS}, specifically at magnitudes of V_{DS} well below $(V_{GS}-V_T)$. Many simple gain-modulation schemes and early multipliers got around the V_{DS}^2 nonlinearity by simply neglecting it, and keeping $|V_{DS}|$ small.

In the simplest FET variable-gain scheme, then, we assume

$$I_D \cong K(V_{GS}-V_T)V_{DS} \tag{5.14}$$

whereupon the FET channel presents a conductance

$$\frac{I_D}{V_{DS}} \cong g_{ds} = K(V_{GS}-V_T) \tag{5.15}$$

Thus, a single FET may be regarded as an electronically variable resistance, although the signal range for linear operation is restricted. The quadratic nonlinearity is fairly weak, so the "voltage-controlled resistor" viewpoint, although a small-signal approximation, is more useful in this case than with many other devices. In pn junction diodes operated as variable resistors, for example, distortion sets in much more quickly as a function of signal amplitude.

A variable resistor is not a complete variable gain element, but it can easily form the basis for one. To effect controllable gain/loss, FET "resistors" of this kind have been used variously as one arm of a resistive divider [5.12, 5.23], as input and/or feedback resistors in op amp gain circuits [5.24], as the emitter-degeneration resistor in differential transistor pairs [5.25, 5.26], and as general replacements for resistors in a variety of "cookbook" applications [5.27, 5.28].

Limitations that may be more serious than the quadratic distortion are the dependence on the parameters K and V_T, hence on fabrication and temperature, and the control-dependent dynamic range. If we wish to limit the V_{DS}^2 term to some fraction f of the amplitude of the multiplicative term, we obtain the constraint

$$|V_{DS}| < 2f(V_{GS}-V_T) \tag{5.16}$$

or

$$|V_{DS}| < (2f/K)g_{DS} \tag{5.17}$$

A signal voltage limit that is directly proportional to conductance is the wrong kind of constraint for most circuit applications. As the resistance goes up the permissible signal level goes down--the opposite of the desired behavior if, for example, the FET is being used as the bottom leg of a voltage divider in an amplitude-modulation task. In order to minimize distortion from the FET while maximizing dynamic range, the circuit environment must be such that maximum V_{DS} is applied when g_{DS} is greatest. If the FET is substituted for a gain-setting resistor in an amplifier or attenuator circuit, this criterion is usually not satisfied.

Leighton [5.29] considered this problem for signal-range-compressing (specifically AGC) applications. He noted that bridge-type circuits--where signal gain depends on the imbalance of two resistances--can have the desired property, and he showed a practical version using a differential op amp. Other authors have generally neglected this consideration, or have simply recommended fixed limits on $|V_{DS}|$, which gives a variable distortion level.

More sophisticated schemes using FETs in nonsaturation improve on this "voltage-controlled resistor" idea in one or both of the following ways:

- Linearization of the I_D vs. V_{DS} characteristic by removing the V_{DS}^2 term to give a true large-signal controllable resistance
- Eliminating the nonlinearity and/or dependence on the parameters K and V_T by using multiple-FET configurations.

These refinements are the subjects of the next two sections.

5.2.5 Linearization of Drain-Source Characteristics

The major limitation of the simple FET "voltage-controlled resistor" is signal-path nonlinearity, due (at least) to the V_{DS}^2 term in (5.13). One successful solution to this is the cancellation of the V_{DS}^2 term through a modification of the way in which the gate is driven. Instead of a fixed V_{GS} being imposed, the gate is biased with respect to a voltage halfway between drain and source potentials. For a theoretical FET that obeys (5.13), the result is a truly linear, large-signal, controllable resistance between drain and source. More accurate models for the JFET and MOSFET reveal some residual nonlinearity, but it is greatly reduced, and always symmetrical for both polarities of V_{DS}.

Consider the simple model of (5.13) expressed in terms of individual electrode voltages (all referred to an arbitrary zero-volt reference):

$$I_D = K\left[(V_G - V_T) - \frac{(V_D+V_S)}{2}\right](V_D - V_S) \tag{5.18}$$

Now, if we shift the gate voltage by adding the average of the drain and source voltages, so that

$$V_G = V_x + \frac{(V_D+V_S)}{2} \tag{5.19}$$

for some V_x, the result is

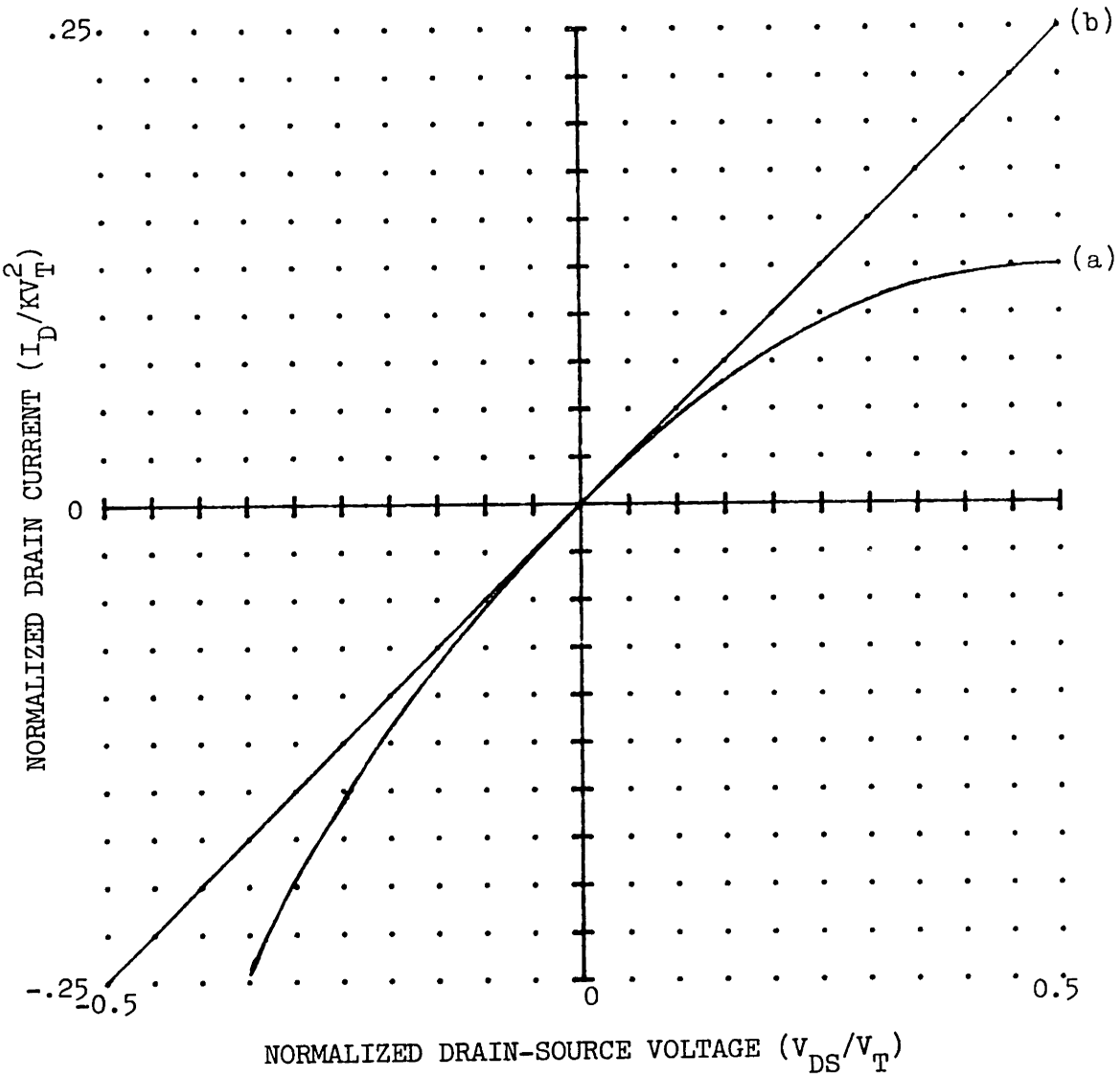


Figure 5.1 -- Comparison of theoretical drain characteristics for FETs in the nonsaturation region.

(a) Simple FET with $(V_{GS}-V_T) = V_T/2$. From equation (5.13).

(b) Linearized FET of section 5.2.5, with $(V_X-V_T) = V_T/2$. From equation (5.20).

$$I_D = K(V_x - V_T)V_{DS} \tag{5.20}$$

(see Figure 5.1). In effect, the small-signal behavior of the FET channel around $V_{DS}=0$, as in (5.14), has been extended to include the entire nonsaturation region.

This technique is usually presented in a special case where the FET source terminal is grounded. It is explained as "adding half the drain voltage to the gate" (note that $V_S=0$ in this case). The more general case is described succinctly by (5.19). Another way to view this is that the drain and source experience equal and opposite swings with respect to a reference voltage, and the gate bias is referred to this voltage as well. The situation is illustrated in Figure 5.2. It is easy to see how, if the FET is constructed symmetrically (drain and source interchangeable), such an arrangement treats drain and source nodes identically and results in symmetry in the I_D-V_{DS} plane. Linearity follows, in theory, from the form of (5.13), and more fundamentally because gate-source and drain-source biases have equal and opposite influence on drain conductance in a charge-control analysis.

In the usual implementation of this technique, the average of the drain and source voltages is obtained with a pair of equal resistors, as shown in Figure 5.3. A version with a convenient control (V_x) input, which does however attenuate the control voltage by a factor of two, is illustrated in Figure 5.4. Often the unity-gain voltage buffer is omitted; however, this does add the resistors and control signal to the controlled resistance port. An asymmetrical FET construction may be accommodated by a resistor

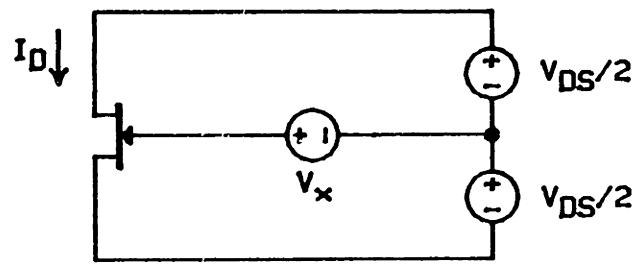


Figure 5.2 -- Linearization of Channel Characteristics

Offsetting the gate bias by a potential halfway between drain and source yields a nearly-linear relation between I_D and V_{DS}. V_x controls the scale factor (i.e., resistance).

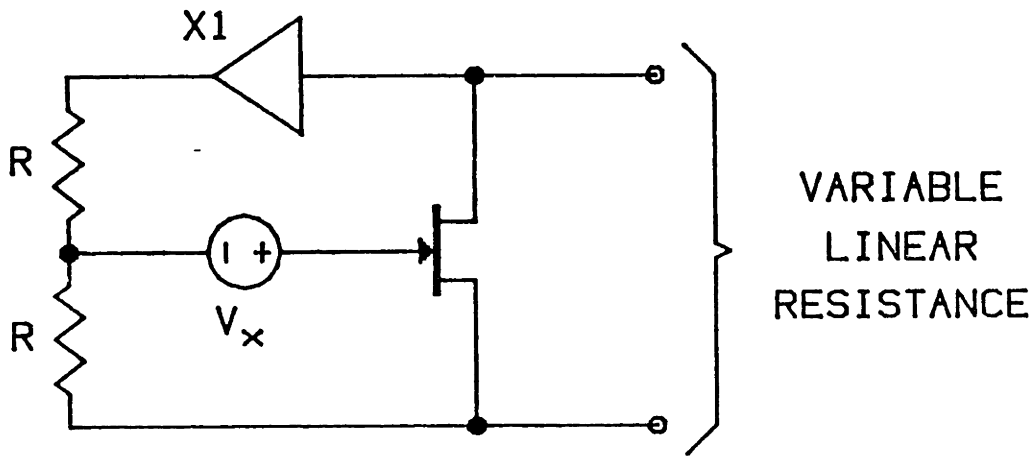


Figure 5.3

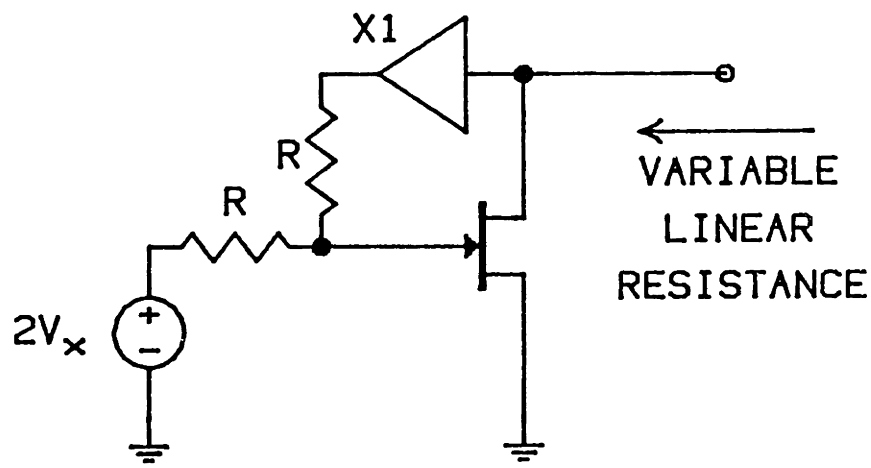


Figure 5.4

Practical realizations of Figure 5.2. The unity-gain voltage buffers are optional.

ratio other than 1:1. This has been done [5.12], although it is not clear whether such finesse is called for in the simple applications to which the technique is usually put. Departures from the idealized model of (5.13) place an ultimate limit on the available linearity, and the chief merit of this approach is its great simplicity.

The technique, attributed to Martin [5.30] originally, has been discussed by several others assuming junction FETs. Elliott [5.31] presented a one-page qualitative explanation; Todd [5.32], in a primer on applications of JFETs in nonsaturation, showed experimental results (which show the imperfect linearity achieved with practical JFETs). Leighton [5.29] and von Ow [5.12, 5.33] emphasized voltage-controlled attenuator applications, Leighton considering in detail a configuration to maximize signal dynamic range and von Ow using the FET as the lower [5.12] or upper [5.33] arm in a resistive voltage divider. Both Leighton and von Ow employed variations on Figure 5.3 in which the gate was coupled to the other terminals of the FET through capacitors. Von Ow claimed [5.12] a reduction in harmonic distortion by a factor of 50 over a simple FET; he allowed one of the resistors in Figure 5.3 to be trimmed.

A more detailed JFET expression reveals that the technique will leave some residual odd-order nonlinearity. Using (5.1) instead of (5.13), and substituting (5.19), gives

$$I_D = \frac{W}{L} q\mu_n N_d t \left\{ V_{DS} - \frac{2}{3} (\phi_i - V_T)^{-1/2} \left[(\phi_i - V_x + \frac{1}{2} V_{DS})^{3/2} - (\phi_i - V_x - \frac{1}{2} V_{DS})^{3/2} \right] \right\} \qquad (5.21)$$

which is an odd function of V_{DS}, at a given V_x.

Thus the drain-source characteristic is again symmetrical about the origin in the I_D-V_{DS} plane; it is also nonlinear, although weakly so. A similar result is obtained if the 3/2-power terms in (5.21) are replaced with other power-law dependences, such as those obtained with a nonuniform doping profile in the channel region.

In applying the linearization to a MOSFET rather than a JFET, a complication immediately arises. The MOSFET is inherently a four-terminal device; what do we do with the bulk terminal?* Bilotti [5.34] considered the MOSFET version, using the usual simplified expression (5.13), and recommended that the bulk be left floating. This worked well for Bilotti, with a discrete FET in an isolated test circuit, but it is a questionable practice because of the possibility of leakage or capacitive coupling of charge into the sensitive floating node. Moreover, in monolithic MOSFET circuits the FET bulk is often committed to a common bias voltage and is therefore not available to the designer. Exceptions are the isolation-well transistors of CMOS processes--the n-channel devices in traditional CMOS, and the p-channel FETs in the now-ascendant n-well CMOS. Bilotti further suggested, in passing, that the bulk terminal could also be tied to the source terminal of the FET.

The formulation of (5.13) implicitly deals with bulk bias through the V_T term, as in (5.9). To more rigorously analyze the linearized MOSFET, including bulk bias, consider the detailed model

---

* The same question arises in a JFET with two gates, and similar conclusions apply.

of (5.8a). Using (5.19) for the gate voltage, with control voltage V_x,

$$I_D = \frac{W}{L}\mu_n \left\{ C'_{ox}[V_x - V_{FB} - 2\phi_{Fp}]V_{DS} - \frac{2}{3}(2\varepsilon_s q N_a)^{1/2}[(2\phi_{Fp}+V_D-V_B)^{3/2} - (2\phi_{Fp}+V_S-V_B)^{3/2}] \right\} \tag{5.22}$$

Note that, unlike the JFET case, the predicted dependence on the control voltage V_x is exactly linear. Equation (5.22) reveals that the linearized MOSFET behaves like the ideal variable resistance of (5.20)--the part of (5.22) before the $\frac{2}{3}$ -- in parallel with a nonlinear resistance, independent of V_x, due to the 3/2-power terms (the "body effect"). The relative magnitudes of these two resistive components depend on C'_{ox} and N_a; the shape of the body-effect nonlinearity depends on how the bulk terminal is driven.

The bulk potential V_B enters through the 3/2-power terms. Because of the symmetry of this pair of terms, their difference will be an odd function of V_{DS} if the drain and source experience equal and opposite excursions with respect to V_B. This suggests that, if the bulk terminal is at the disposal of the designer, it should be biased in a manner similar to the gate; see Figure 5.5. A practical implementation could follow the approach of Figure 5.4 with a second voltage divider to drive the bulk terminal. Figure 5.5 could even be applied to MOSFETs whose bulk regions are connected in a common substrate, as long as balanced voltage swings can be maintained on drain and source.

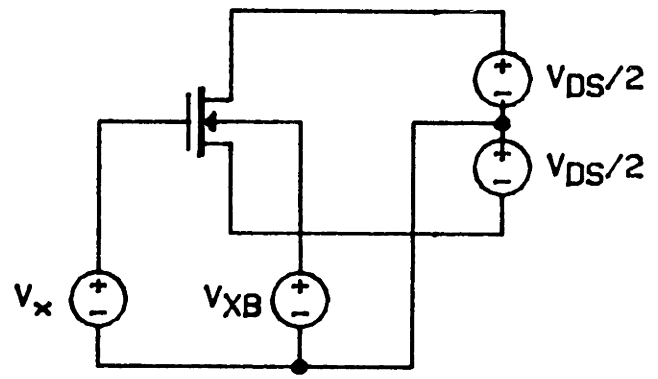


Figure 5.5 -- Bulk-terminal Bias Scheme for a Linearized MOSFET

THE BIAS VOLTAGE V_{XB} NORMALLY WOULD BE NEGATIVE FOR AN N-CHANNEL FET.

If the bulk and source are connected together, as recommended by Bilotti, we find

$$I_D = \frac{W}{L}\mu_n \left\{ C'_{ox}[V_x - V_{FB} - 2\phi_{Fp}]V_{DS} - \frac{2}{3}(2\varepsilon_s q N_a)^{1/2}[(2\phi_{Fp}+V_{DS})^{3/2} - (2\phi_{Fp})^{3/2}] \right\} \tag{5.23}$$

so that I_D is no longer an odd function of V_{DS}. There is also a tendency for the drain-bulk junction to become forward-biased when V_{DS} swings negative; this will limit the negative voltage excursion across the voltage-controlled "resistor" to the order of one-half volt. Such a limit can be avoided in the circuit of Figure 5.5 since a substantial reverse bias can be applied to the bulk terminal.

Regardless of the bulk connection, linearity of the channel resistance is greatly improved over a simple MOSFET. The crucial modification of (5.19) cancels the V_{DS}^2 term present in (5.8b), and hence the majority of the curvature in the nonsaturation drain characteristics.

This discussion has established that most of the channel-resistance nonlinearity in both JFETs and MOSFETs can be eliminated if the gate (and bulk, if any) is properly driven. It is indeed possible, then, to construct a large-signal electronically variable resistor from a FET; however, as a gain control component, it has practical shortcomings. The characteristics of the variable resistor depend strongly on device parameters such as C'_{ox} and N_a; the control input V_x must be offset by a FET threshold voltage. In a monolithic design, obtaining either the balanced drain-source drive of Figure 5.2

or the voltage divider of Figures 5.3 and 5.4 may be a major obstacle. The linearized FET is probably most useful as a very simple variable-resistance element with modest performance or as a component in multiple-FET topologies that cancel the remaining parameter dependences.

5.2.6 Multiple-FET Configurations

As with bipolar transistors, both fabrication-dependent parameters and signal-path nonlinearity can be circumvented by using multiple matched FETs. A judicious choice of topology will cancel unwanted behavior which is common to different devices, while preserving the device characteristics that yield the desired gain modulation.

As before, a first-order model common to both junction and MOS FETs will be assumed initially. A further assumption is that device parameters are well matched between FETs. Analysis of mismatch effects is straightforward, and is not included here.

Since the essential property of the FET in nonsaturation is a voltage-controlled channel resistance, it is natural to substitute FETs (with or without the linearization of section 5.2.5) for gain-setting resistors in various gain/loss networks. The preceding sections described single FETs used in resistive voltage dividers and similar networks. They have also been used in pairs in an effort to cancel common device parameters.

Shore [5.35] used a pair of MOSFETs with identical gate-source voltages. One FET served as a gain-setting resistor in an op amp signal-path circuit and the other was a reference device, its drain current set by a control input. Shore's basic approach was to assume the FETs were linear voltage-controlled resistors (that is, he neglected the V_{DS}^2 term of (5.13)) and then to force

the two V_{GS}s to be equal. The result was that the parameters K and V_T cancelled, leaving the basic relation

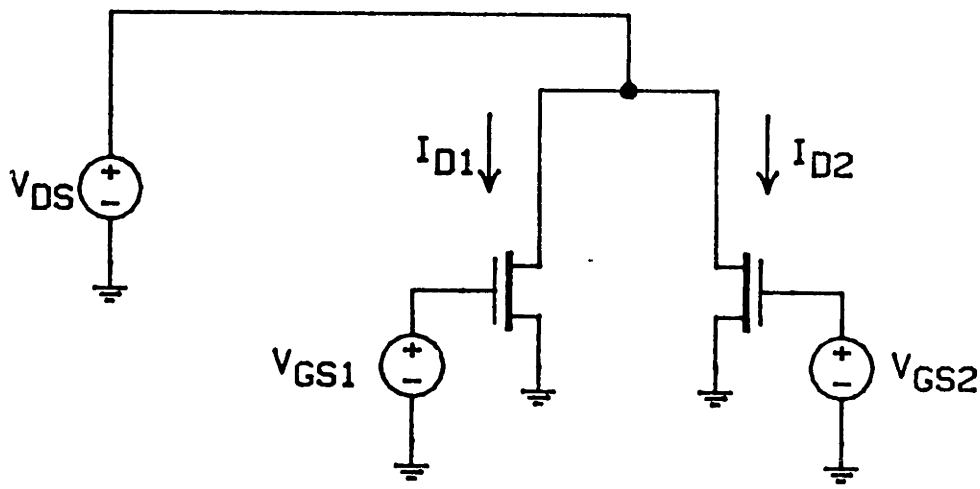
$$\frac{I_{D1}}{V_{DS1}} = \frac{I_{D2}}{V_{DS2}} \tag{5.24}$$

which could then yield, say, a multiply/divide function with one of these four variables taken as the output and the other three as inputs. The neglected quadratic term in (5.13) limits this approach to signal excursions well below the bounds of the nonsaturation regime.

Miller [5.36] employed a bridge-like scheme where the two input resistors to a differentially connected op amp were replaced by linearized JFETs. His circuit compensated for the parameter K in (5.13) but not for V_T. Trofimenkoff and Smallwood [5.37] used a similar pair of linearized JFETs as shunts at the inputs of a differential amplifier. The result was sensitive to the FET parameters and required a thermistor to maintain temperature stability.

A different approach involving a pair of FETs has been used very successfully to cancel both the nonlinearity and the threshold-voltage dependence. It ranks with the single-FET linearization of section 5.2.5 as one of the most elegant and useful FET subcircuits for the class of gain-control applications. The basic configuration, as shown in Figure 5.6, is two FETs with:

- Identical V_{DS} (which forms one input variable)
- Differential V_{GS} (the difference $V_{GS1}-V_{GS2}$ is the second input)
- Drain currents which are subtracted to form the output.



$$I_{OUT} = I_{D1} - I_{D2}$$

Figure 5.6 -- FET-pair Multiplier Cell

OUTPUT I_{OUT} IS ACCURATELY PROPORTIONAL TO THE PRODUCT OF V_{DS} AND THE GATE VOLTAGE DIFFERENCE

From the model of (5.13), this gives

$$I_{D1} = K\left[(V_{GS1}-V_T)V_{DS} - \frac{V_{DS}^2}{2}\right]$$

$$I_{D2} = K\left[(V_{GS2}-V_T)V_{DS} - \frac{V_{DS}^2}{2}\right]$$

$$I_{OUT} = I_{D1} - I_{D2} = KV_{DS}(V_{GS1}-V_{GS2}) \tag{5.25}$$

Both V_T and the V_{DS}^2 term have cancelled in the output. Also, both V_{DS} and $(V_{GS1}-V_{GS2})$ can have either polarity, so this is inherently a four-quadrant multiplication.

This technique appears to have been first disclosed by Highleyman and Jacob in 1962 [5.38], although, like many such artifices, it is reinvented regularly. Østerfjells [5.39], Bosshart [5.20, 5.40],* Mavor et al. [5.41], and McCaughan et al. [5.42], for example, have presented it in various forms. The methods of imposing the inputs and extracting the output vary considerably. Highleyman and Jacob drove JFETs through transformers; McCaughan et al. actually time-shared the input stage of a CCD to act as two MOSFET devices. The latter authors [5.20, 5.40-5.42] were addressing the problem of building compatible (i.e., all-MOS) analog multipliers on analog CCD chips.

According to (5.25), the technique yields exact cancellation of nonlinear channel conductance and hence an accurate multiplication.

---

*The FET equations in Bosshart's ISSCC paper [5.40] contain typographical errors, but the result is correct.

In fact, this result holds for MOSFETs even when more accurate drain current expressions are used. In (5.8), the generalized formula for MOSFETs, there are additional 3/2-power terms arising from the bulk-to-channel interaction. As Bosshart has pointed out [5.20], these terms cancel in the configuration of Figure 5.6 since they are independent of gate-source voltage and are thus identical for both FETs. The bulk bias or "body effect" does not, therefore, enter directly in the output (although it does affect the saturation threshold, and therefore the multiplier input range). The result of the more detailed MOSFET model is basically the same as (5.25), subject to higher-order effects such as high-field mobility degradation (section 5.2.2). Measurements [5.20, 5.40] confirm that accuracies better than 1% are attainable with MOSFETs.

For junction FETs with uniform channel doping the conclusion is not quite so clean, since exact cancellation of all nonlinearities is not predicted. However, an approximate analysis reveals the same general behavior. If we assume balanced gate drive so that

$$V_{GS1} = V_{GS0} + \Delta V$$

$$V_{GS2} = V_{GS0} - \Delta V$$

then for the configuration of Figure 5.6, (5.1) gives

$$I_{D1} - I_{D2} = \frac{2}{3} A[(V_R+V_{DS}+\Delta V)^{3/2} + (V_R-\Delta V)^{3/2} - (V_R+V_{DS}-\Delta V)^{3/2} - (V_R+\Delta V)^{3/2}] \qquad (5.26)$$

where $A = \frac{W}{L} \frac{q\mu_n N_d t}{(\phi_i - V_T)^{1/2}}$ and $V_R = \phi_i - V_{GSO}$. The important thing to notice is the group of 3/2-power terms, in which the input quantities V_{DS} and ΔV appear. Using the series expansion

$$(V+v)^{3/2} = V^{3/2} + \frac{3}{2} V^{1/2} v + \frac{3}{8} V^{-1/2} v^2 - \ldots$$

and discarding terms beyond v^2, (5.26) collapses to

$$I_{D1} - I_{D2} \cong \frac{A}{V_R^{1/2}} V_{DS} \Delta V$$

$$\cong \frac{A}{2V_R^{1/2}} V_{DS} (V_{GS1} - V_{GS2}) \qquad (5.27)$$

which, at least for sufficiently small signals, confirms (5.25) for JFETs that satisfy (5.1).

5.2.7 Linearity and Parameter-Independence: The Nonsaturation-Mode Analogs of Translinear Circuits

For FETs in nonsaturation, as for certain other active devices, it is possible to systematically combine multiple devices so as to realize "pure" algebraic input-output relations. In such cases the output depends entirely on controllable variables (inputs, references, dimensionless scale factors) and not on unpredictable, temperature-sensitive FET parameters--at least in principle. Equation (5.24) is an example of such a relation; however, (5.24) is

a small-signal result. More accurate and general functions may be obtained by replicating either the FET-pair multiplier of section 5.2.6 or the linearized FET of section 5.2.5.

The scheme of Abu-Zeid and Groendijk [5.43] uses two of the FET pairs of Figure 5.6, with a feedback loop driving the second pair in a "slave" arrangement as shown in Figure 5.7. The effect of the feedback loop is to force the drain current difference in the second pair to equal the drain current difference in the first pair. Under this condition, (5.25) gives (allowing for possibly different Ks in the two different pairs)

$$K_1 V_{DS1} \Delta V_{GS1} = K_2 V_{DS2} \Delta V_{GS2} \tag{5.28}$$

$$\text{or} \qquad \Delta V_{GS2} = \frac{K_1}{K_2} \frac{V_{DS1} \Delta V_{GS1}}{V_{DS2}} \tag{5.29}$$

Equation (5.29) is a flexible multiply/divide function in which each voltage on the righthand side may be either an input or a reference (constant). Unlike the circuits discussed previously, the relation is independent of all FET parameters except the ratio K_1/K_2. This ratio will be essentially independent of temperature for isothermal FETs of similar construction, such as those on a common monolithic substrate; it may not be so for FETs only pairwise matched, although Abu-Zeid and Groendijk obtained good results even then. The ratio will be unity if all FETs are identical; it may be intentionally made other than unity (by design of FET

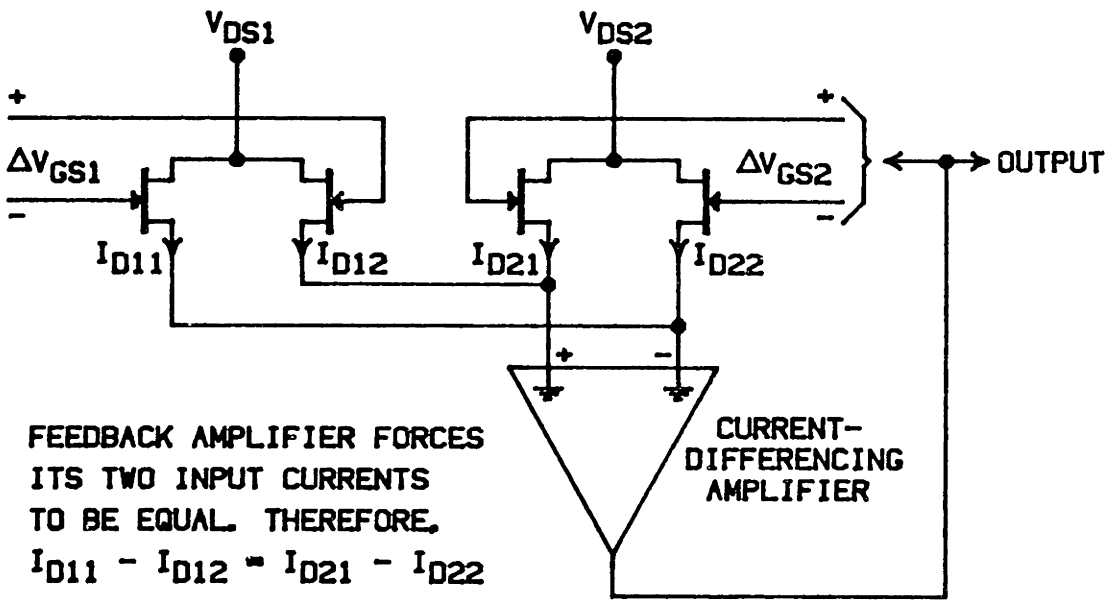


Figure 5.7 -- Multiplier Using Two FET Pairs and a Servo Loop

AFTER ABU-ZEID AND GROENDIJK [5.43]

geometry or by paralleling identical FETs) to give a fixed scale factor.

Although (5.29) and Figure 5.7 imply that V_{GS2} is taken as an output, the fundamental relation of (5.28) may be exploited in other ways as well. For example, a modification of the feedback loop in Figure 5.7 could force the drain current differences to be equal by driving the drains of the second pair rather than the gates. This drain voltage could then be taken as the output, so that the roles of V_{DS2} and ΔV_{GS2} in (5.29) would be reversed. Further practical considerations are the choice of common-mode voltage on the gates, which affects the standing drain currents, and the need for a fast feedback amplifier, which (as in any such servo system) largely determines the bandwidth.

Abu-Zeid and Groendijk [5.43] described a special case of this topology using two monolithic JFET pairs. They claimed a maximum error of ±1% as an analog multiplier with an input voltage range of ±0.3V (the FET thresholds were on the order of -2V). Since (5.25) is particularly accurate when MOSFETs are used, they should provide comparable or better results.

A different approach was disclosed earlier by Abu-Zeid et al. [5.44] and, in a more specialized form, earlier still by Hutcheon and Puddefoot [5.45]. It follows a similar principle: equating internal, parameter-dependent variables to obtain a parameter-independent relation analogous to (5.28). In this case, however, the basic subcircuit is the linearized FET of section 5.2.5. Recall that the linearization of Figure 5.3 or Figure 5.4 removed

the major nonlinearity in I_D vs. V_{DS} and gave a relation of the form

$$I_D = K(V_x - V_T)V_{DS} \tag{5.30}$$

in which the device parameters K and V_T are still present. V_x is the gate bias with respect to a voltage halfway between V_D and V_S; V_{DS} may be of either polarity. Shore's topology [5.35] led, as cited earlier, to (5.24), but only in the limit of very small signal excursions. The linearized FET has the capability of extending Shore's result to large signals as well.

Rearrangement of (5.30) gives

$$V_x = \frac{I_D}{KV_{DS}} + V_T \tag{5.31}$$

If we now arrange two such circuits with identical values of V_x, we find from (5.31) that K and V_T cancel out. The result of equating the two V_xs, allowing again for different values of K, is

$$\frac{I_{D1}}{K_1 V_{DS1}} = \frac{I_{D2}}{K_2 V_{DS2}} \tag{5.32}$$

or
$$I_{D1}V_{DS2} = \frac{K_1}{K_2} I_{D2}V_{DS1} \tag{5.33}$$

As with (5.28), this is a flexible relation that may be used in a

variety of ways. One possible implementation is shown in Figure 5.8, where V_{DS1} is a constant, I_{D1} and V_{DS2} are inputs, and the product is obtained by measuring I_{D2}. This was the configuration of Abu-Zeid et al., and uses op amps to force the desired node conditions. Hutcheon and Puddefoot actually used a single linearized FET and time-shared it to serve as both devices.

Abu-Zeid et al. employed discrete JFETs and reported an accuracy of about 1% for an input range of ±0.3V. As Abu-Zeid and Groendijk later pointed out [5.43], the linearization technique does require greater complexity than the approach of Figure 5.7. Moreover, the requirement of accurately ratioed resistors may be difficult to meet if, for example, an all-MOS version is needed.

There is an interesting sidelight to these two schemes. Equations (5.28) and (5.32) both result from forcing certain internal signals to be equal, and in both cases the parameters V_T and K (except for a ratio) are absent in the final equation. The topologies that led to (5.28) and (5.32) may be generalized: instead of using a pair of subcircuits and equating the corresponding internal variables (ΔI in the case of Figure 5.7, V_x in the case of Figure 5.8) in the pair, we set up an arbitrary number of similar subcircuits and impose a linear constraint on these internal variables.

For example, the basic linearized-FET subcircuit of section 5.2.5 obeys (5.31). Suppose that we have four identical linearized FETs arranged so that

$$V_{x1} + V_{x2} = V_{x3} + V_{x4} \tag{5.34}$$

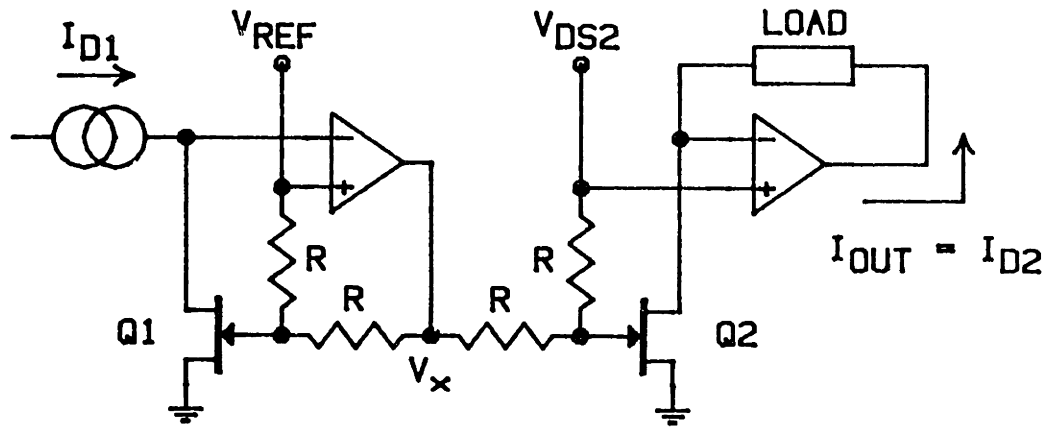


Figure 5.8 -- Multiplier Using Two "Linearized" FET Subcircuits

Then substitution of (5.31) gives

$$\frac{I_{D1}}{V_{DS1}} + \frac{I_{D2}}{V_{DS2}} = \frac{I_{D3}}{V_{DS3}} + \frac{I_{D4}}{V_{DS4}} \tag{5.35}$$

which could be exploited in a number of ways, depending on which of the variables were considered inputs, constants and output. If V_{DS1} were the output, the other V_{DS} variables all inputs and the currents all set to the same reference value, for example, (5.35) would give

$$V_{DS1} = \frac{1}{\frac{1}{V_{DS3}} + \frac{1}{V_{DS4}} - \frac{1}{V_{DS2}}} \tag{5.36}$$

This is not a function that one frequently needs; however, it illustrates a nontrivial algebraic function that is easily accommodated with such a circuit.

The technique may be further generalized by allowing an arbitrary number of V_xs to be summed on each side of (5.34), and by permitting ratios of Ks, which gives added design freedom.

This is reminiscent of the translinear principle for bipolar transistors (see chapter 6), although it yields a different class of algebraic relation--an equation in weighted sums of quotients, such as (5.35). Among the drawbacks are the need to "stack" values of V_x (analogous to V_{BE} in translinear circuits), and limited ultimated accuracy due to imperfect FET linearization as discussed

in section 5.2.5. The general structure is a "loop" of V_xs obeying Kirchhoff's Voltage Law, which may be cast in the form

$$\sum_{CW} V_{xn} = \sum_{CCW} V_{xn} \tag{5.37}$$

where the notations cw, ccw refer to those n where the sense of V_{xn} is clockwise or counterclockwise around the voltage loop. From (5.31), any set of linearized-FET subcircuits whose V_x values satisfy (5.37) will obey

$$\sum_{CW} \frac{I_{Dn}}{K_n V_{DSn}} = \sum_{CCW} \frac{I_{Dn}}{K_n V_{DSn}} \tag{5.38}$$

A similar line of reasoning leads to a generalization of the other parameter-independent multiplier/divider discussed above (and epitomized in Figure 5.7). If the two-FET subcircuit of Figure 5.6 is replicated four times, for example, and appropriate circuitry imposes the condition

$$I_{OUT1} + I_{OUT2} = I_{OUT3} + I_{OUT4} \tag{5.39}$$

(note that each I_{OUT} variable represents a drain current <u>difference</u>), then (5.28) generalizes to

$$K_1 V_{DS1} \Delta V_{GS1} + K_2 V_{DS2} \Delta V_{GS2} = K_3 V_{DS3} \Delta V_{GS3} + K_4 V_{DS4} \Delta V_{GS4} \tag{5.40}$$

in section 5.2.5. The general structure is a "loop" of V_xs obeying Kirchhoff's Voltage Law, which may be cast in the form

$$\sum_{CW} V_{xn} = \sum_{CCW} V_{xn} \tag{5.37}$$

where the notations cw, ccw refer to those n where the sense of V_{xn} is clockwise or counterclockwise around the voltage loop. From (5.31), any set of linearized-FET subcircuits whose V_x values satisfy (5.37) will obey

$$\sum_{CW} \frac{I_{Dn}}{K_n V_{DSn}} = \sum_{CCW} \frac{I_{Dn}}{K_n V_{DSn}} \tag{5.38}$$

A similar line of reasoning leads to a generalization of the other parameter-independent multiplier/divider discussed above (and epitomized in Figure 5.7). If the two-FET subcircuit of Figure 5.6 is replicated four times, for example, and appropriate circuitry imposes the condition

$$I_{OUT1} + I_{OUT2} = I_{OUT3} + I_{OUT4} \tag{5.39}$$

(note that each I_{OUT} variable represents a drain current <u>difference</u>), then (5.28) generalizes to

$$K_1 V_{DS1} \Delta V_{GS1} + K_2 V_{DS2} \Delta V_{GS2} = K_3 V_{DS3} \Delta V_{GS3} + K_4 V_{DS4} \Delta V_{GS4} \tag{5.40}$$

In practice, as with Figure 5.7, some of the signal variables in this equation will be employed to force the condition of (5.39). Corresponding to the constraint of (5.37) is the more general condition

$$\sum_n s_n I_{OUTn} = 0 \tag{5.41}$$

where s_n is +1 or -1 depending on the polarity with which the current difference I_{OUTn} is summed. Note that it is not necessary to use an even number of elementary subcircuits in this case, even though the above example uses four. The preceding technique did require an equal number of terms on each side of (5.37), hence an even number of linearized FETs, in order that the V_T components of (5.31) would properly cancel.

Applying (5.25) to the condition of (5.41) yields the general result of this second technique: a relation of the form

$$\sum_n s_n K_n V_{DSn} \Delta V_{GSn} = 0 \tag{5.42}$$

which is different from the form of (5.38). It is also more accurate; the underlying relation, in (5.25), is (at least for MOSFETs) subject to fewer aberrations than that of (5.30) and (5.31). Note that (5.41), in describing the way that the subcircuits are linked, is an expression of Kirchhoff's Current Law while (5.37) corresponds to the Voltage Law.

The ability to obtain theoretically accurate, temperature-independent relations like (5.35), (5.36) and (5.40) from field-effect

transistors is a novel extension of the multiplier/divider techniques that began this section. Although only a cursory description has been given, without circuit examples, it should be evident that many different configurations can be developed from the general theory presented here. The novelty of the idea is tempered somewhat by the restricted types of nonlinear relations that can be realized and by the various second-order effects that constrain the available accuracy.

5.2.8 Summary: Strategies for Exploiting the Nonsaturation Region

In view of the length of section 5.2, it seems appropriate to review the main topics and conclusions. From the basic idea of a variable channel resistance, this discussion has developed the theory of nonsaturation-region FET variable-gain elements. Sections 5.2.4 through 5.2.7 have presented a sequence of increasingly refined approaches in which various device limitations are dealt with.

With few exceptions, the published papers on nonlinear analog circuits using FETs have developed theoretical results from elementary FET models without explicitly considering the hypotheses of the models. An attempt has been made to consider them here, and to cast the diverse circuit techniques in a consistent large-signal formulation.

FETs are not precision nonlinear elements. Although their qualitative behavior is pleasingly simple, they lack the strikingly predictable static characteristics of bipolar junction devices. Their peculiarities become manifest when accuracies beyond a few percent of full scale are sought. At the same time they are very

simple to make, and are increasingly important as analog devices in monolithic form. Unlike the bipolar transistor, the FET in nonsaturation presents a useful relationship between three signal variables (I_D, V_{DS}, V_{GS}); consequently some useful gain-control functions can be obtained with remarkably simple circuits.

In the simplest "voltage-controlled resistors" of section 5.2.4, neither the inherent nonlinearity of the FET channel resistance nor the dependence on device parameters is compensated for, yielding a controllable two-terminal resistance that is fabrication-sensitive, temperature-sensitive and nonlinear. In the MOSFET, the excellent linearity of drain current as a function of gate voltage suggests that the gate would have advantages as a signal input, although the versatility of a controlled resistor would be lost. In section 5.2.5, a technique is introduced whereby the linearity of the controlled resistance may be greatly improved, without addressing the parameter dependence. If a MOSFET is used, residual nonlinearity is due primarily to body-bias effect. The nonlinearity may be made symmetrical to eliminate even-order distortion provided the bulk terminal is available. Low bulk-region doping and thin gate oxide are desirable characteristics for such a MOSFET.

In section 5.2.6 a variety of multiple-FET topologies are cited. The use of FET pairs can give a simple four-quadrant analog multiplication function that is independent of threshold voltage and has good linearity with respect to either input. This is especially true with MOSFETs, and the need for low-body-effect devices is removed.

Finally, section 5.2.7 discusses two classes of circuits that approach the ideal: input-output characteristics insensitive to device parameters and free of signal distortion. One class uses two of the FET-pair multipliers of the previous section, cancelling the remaining parameter dependence, and gives a versatile equation relating the product of two voltage variables to the product of another two. Since four FETs must be well matched and at the same temperature, monolithic construction is almost mandatory. The other class employs a pair of the linearized FETs of section 5.2.5, and leads to an equality-of-ratios equation in which two of the four variables are voltages and the others are currents. Generalizations of these two classes can in theory produce more elaborate nonlinear functions in a manner similar to the translinear principle for bipolar junction devices.

Little has been said here about signal excursion limits in these circuits. Always present in the background of the discussion is the inescapable nonsaturation constraint. There are interdependent bounds on the terminal voltages in each FET, as detailed in sections 5.2.1 and 5.2.2. Consequently, a device like a multiplier may be restricted to a subset of the XY input plane in order to keep the key transistors out of saturation. Further constraints occur, especially with JFETs, because of the possibility of forward-biasing the gate-channel (or bulk-channel) junction. Such considerations are intimately tied to the specific circuit involved; however, they form a common shortcoming of nonsaturation-mode large-signal gain control elements.

5.3 FETs in Other Modes of Operation

In the nonsaturation region, the drain current is sensitive to both gate bias and channel voltage drop (V_{DS}); a plethora of multiplication schemes make use of this dual sensitivity. In contrast, a FET operating in the saturation region (also called the pinchoff or pentode region) has a drain current that depends in a useful way only on gate bias.

The definition of the saturation threshold for JFETs is different from that for MOSFETs with significant body effect; see sections 5.2.1 and 5.2.2. Above this threshold (i.e., for a sufficiently large drain voltage), the FET channel is completely depleted beyond a "pinchoff" point near the drain end; between this point and the drain diffusion is a high-field depletion region which absorbs further increases in drain voltage. Consequently the voltage along the nondepleted part of the channel is effectively fixed, accounting for the insensitivity of I_D to V_{DS}. Any large-signal multiplicative function in this mode must therefore exploit the I_D-to-V_{GS} relation.

In the MOSFET, operation with V_{GS} below the threshold voltage does not correspond to a completely "pinched off" channel, as in a JFET, but to a weakly inverted silicon surface. Relatively small drain currents can flow in this condition, and the dependence of I_D on both V_{GS} and V_{SB} is approximately exponential over a few decades of current. The exponential property can be exploited as in bipolar junction dervices, providing yet another useful mode for MOSFETs.

5.3.1 JFETs and MOSFETs in Saturation

Although saturation is the most common region of operation for FETs in nondigital applications (in amplifiers, for example), it has been little used for large-signal gain control. Ikeda [5.46] operated a MOSFET with a large drain-to-source bias and coupled the drain voltage to the gate through a capacitor. This gave an AC resistance from drain to source that could be modulated by varying DC drain current (hence transconductance), but it was in fact a small-signal AC resistance, with quadratic distortion properties similar to the circuits of section 5.2.4.*

Házman [5.47] used a more elaborate circuit with two differentially connected pairs of MOSFETs, whose source currents were supplied by a third pair, whose sources were in turn grounded. The result was an approximate multiplication with an approximate cancellation of some parameter dependence, but a severe residual dependence on temperature and bias. Neither of these two approaches is very attractive for large-signal applications.

5.3.2 Device Characteristics

Expressions for I_D in saturation can be obtained from the nonsaturation results of sections 5.2.1 and 5.2.2, evaluated at the onset of saturation. As the drain voltage increases beyond the saturation threshold, I_D will increase slightly as the pinchoff point

---

* Ikeda emphasized that a FET in saturation could stand off a DC drain-source voltage without changing its characteristics, unlike a FET in nonsaturation. Of course, the same thing could be accomplished by adding a capacitor to a FET in nonsaturation. Ikeda evidently confused control-input linearity with signal-path linearity in claiming that his circuit was at least as "linear" as that of Bilotti [5.34].

moves further into the active channel, effectively decreasing the channel length. This effect has been treated extensively in the literature and can be incorporated into the analysis, but it is a distraction in the present context so it will be neglected. The discussion will assume instead that the FETs are operating at the onset of saturation, with the drain current determined entirely by V_{GS} (and, in the case of MOSFETs, V_{SB}).

Consider first the result of evaluating (5.1) at the onset of saturation, which corresponds to $V_{DS} = V_{GS} - V_T$. Equation (5.1) gives the drain current for a JFET with uniform doping throughout the channel region. Taken at the onset of saturation, (5.1b) yields

$$I_D = \frac{W}{L} q\mu_n N_d t \left[V_{GS} - \phi_i + \frac{1}{3} \frac{qN_d t^2}{2\epsilon_s} + \frac{2}{3} \left(\frac{2\epsilon_s}{qN_d t^2} \right)^{1/2} (\phi_i - V_{GS})^{3/2} \right]$$

$$= I_p \left[1 - 3 \left(\frac{\phi_i - V_{GS}}{\phi_i - V_T} \right) + 2 \left(\frac{\phi_i - V_{GS}}{\phi_i - V_T} \right)^{3/2} \right] \qquad (5.43)$$

where $I_p = \frac{1}{3} \frac{W}{L} \frac{q^2 \mu_n N_d^2 t^3}{2\epsilon_s}$ and the other terms are defined in section 5.2.1. Now, as Sze [5.1] has pointed out, the corresponding result for a JFET whose channel doping is a thin, dense sheet rather than uniform (virtually the opposite extreme from the assumptions in (5.1)) is of the form

$$I_D = I_p \left[1 - \left(\frac{\phi_i - V_{GS}}{\phi_i - V_T} \right) \right]^2 \qquad (5.44)$$

Equations (5.43) and (5.44), relating I_D to V_{GS} for two very different theoretical JFETs, do not at first look at all similar. Yet when plotted together they very nearly coincide. This was illustrated by Sze. Indeed, the close numerical similarity of FET characteristics in saturation is typical. In a series of widely quoted papers [5.4-5.7], Richer and Middlebrook established that JFET saturation-region static characteristics fall into a narrow range, both in theory and for many measured devices. They showed [5.7] that under certain weak restrictions, the theoretical result has the form

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_T}\right)^{n} \tag{5.45}$$

where n is always near 2 and is bounded by $2 \leq n \leq 2.25$. Here, I_{DSS} is the drain current for zero V_{GS} (following the usual industry notation) and is different from the I_p of the preceding equations.

Equation (5.45), with n=2, is exactly the form predicted by the first-order nonsaturation model of (5.13) at the onset of saturation. The first-order model

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_T}\right)^{2} \tag{5.46}$$

(modified perhaps by the inclusion of ohmic source resistance) is often used in practice.

Deviations due to high-field mobility degradation in the JFET are more likely in saturation than in nonsaturation, since the electric field along the channel is greatest in saturation. Cobbold and Trofimenkoff [5.9] showed that such effects could account for values of n as low as 1.5 in (5.45), in contrast to the lower limit of 2 predicted by the simpler analysis of Richer and Middlebrook. Cobbold and Trofimenkoff also measured values of n as low as 1.57 and as high as 2.7 in some JFETs, although most showed n close to 2.

Miller and Mever [5.48] demonstrated vividly that two FET characteristics whose plots are nearly coincident may yield vastly different results in some applications. They found that classical theoretical results such as (5.43), augmented by allowing for a series resistance in the source lead, agreed well graphically with measured JFET behavior. Yet, for predicting high-order nonlinearities in the JFET transfer characteristic, the theoretical curve was inadequate. Nonlinearities above the quadratic depended on exactly those differences that appeared so slight when (5.43) and (5.44) were plotted together. Miller and Meyer found it necessary to work from measured FET data where such nonlinearities were important.*

For the MOSFET, the drain current at the onset of saturation may be found by evaluating (5.8b) with a drain-to-bulk voltage given

---

* This seems tc be the case, in both nonsaturation and saturation, whenever the exact static characteristics of JFETs are critical in circuit design. Published work on FET circuits makes heavy use of loose expressions like (5.46) or, less often, theoretical abstractions like (5.43). To accurately predict high-order behavior in JFETs, it will probably always be necessary to forsake these models and work directly from measured characteristics.

by (5.10). The result is*

$$I_D = \frac{1}{2}\left(\frac{W}{L}\mu_n C'_{ox}\right)\left\{(V_{GS}-V_{FB}-2\phi_{Fp})^2 - \frac{\gamma^4}{4}\left[2 + \frac{4}{\gamma^2}(V_{GS}+V_{SB}-V_{FB}) - 2\sqrt{1 + \frac{4}{\gamma^2}(V_{GS}+V_{SB}-V_{FB})}\right]\right\} - \frac{2}{3}\gamma\left(\frac{W}{L}\mu_n C'_{ox}\right)\left\{\left[V_{GS}+V_{SB}-V_{FB} - \frac{\gamma^2}{2}\left(\sqrt{1 + \frac{4}{\gamma^2}(V_{GS}+V_{SB}-V_{FB})} - 1\right)\right]^{3/2} - (2\phi_{Fp}+V_{SB})^{3/2}\right\} \tag{5.47}$$

where $\gamma = \frac{(2\varepsilon_s q N_a)^{1/2}}{C'_{ox}}$ and the other terms are defined in section 5.2.2.

The factor γ, sometimes called "body effect coefficient," is a measure of the sensitivity of the FET to bulk (body) bias compared with gate bias. If the body effect is negligible, γ (in volts to the 1/2 power) approaches zero and (5.47) simplifies (considerably) to

$$I_D = \frac{1}{2}\frac{W}{L}\mu_n C'_{ox}(V_{GS}-V_T)^2 \tag{5.48}$$

with $V_T = V_{FB} + 2\phi_{Fp}$.

This has the form of (5.46), although slightly rearranged, and supports the frequent assumption that MOSFETs fundamentally are better square-law devices in saturation than are JFETs. Equation (5.47)

---

* The standard books on device theory (Sze, Grove, etc.) do not give this expression explicitly. I understood why after deriving it.

is subject also to mobility-degradation effects as discussed in section 5.2.2 [5.13].

In summary, both JFETs and MOSFETs operating in saturation may be described as imperfect square-law transconductors. The near-square-law behavior is widely exploited in large-signal high-frequency circuits for communications electronics [5.49-5.51]. To date it does not appear to have been used much for nonlinear analog functions of the kind central to this thesis. One way in which it might be used is the classical quarter-square approach to analog multiplication.

5.3.3 Quarter-Square Architectures

If accurate square-law devices are available, they can synthesize an analog multiplication by exploiting the identity

$$XY = \frac{1}{4}\left[(X+Y)^2 - (X-Y)^2\right] \tag{5.49}$$

This was once the dominant approach to accurate, high-speed multiplication, using piecewise-linear diode squaring operators. A direct implementation of (5.49) requires a pair of squaring circuits, which must accommodate bipolar arguments (inputs) if the resulting multiplier is to operate in at least two quadrants of the X-Y plane.

To the extent that they obey (5.46), FETs can in theory act as the squaring elements in a multiplier similar to this "quarter-square" structure. The formulation will be slightly different from (5.49) to allow for the idiosyncrasies of the FET "square law": a one-quadrant "input" (V_{GS}) and an offest term inside the squaring operator. To see the basic idea, suppose that X and Y are normalized

input variables such that $0 \leq X \leq 1$ and $-1 \leq Y \leq 1$. One modification of the quarter-square identity in (5.49) is

$$\frac{1}{4}\left[(1+X+Y)^2 - (1+X-Y)^2\right] = XY + Y \tag{5.50}$$

Notice that the quantities within parentheses are now unipolar (nonnegative) rather than bipolar, and that the inputs span two quadrants of the X-Y plane (unipolar X, bipolar Y). The added Y term on the righthand side of (5.50) may be regarded as "Y feedthrough" or, equivalently, "X offset," and could in principle be subtracted off to yield a pure multiplication. With suitable scaling of the input variables, (5.50) may be applied to FETs in saturation, since the two squaring operations in (5.50) have forms similar to the ideal FET square law, (5.46).

To provide a concrete example, consider two identical FETs satisfying (5.46) and having gate-source voltages given by:

$$V'_{GS} = \frac{V_T}{2} + V_X + V_Y \tag{5.51a}$$

$$V''_{GS} = \frac{V_T}{2} + V_X - V_Y \tag{5.51b}$$

where V_X and V_Y are the multiplier input variables. V_X will be unipolar, between zero and $\frac{V_T}{4}$; V_Y will be bipolar, between $\frac{-V_T}{4}$ and $\frac{V_T}{4}$ (note that V_T is negative for JFETs and depletion-mode MOSFETs). For variable-gain applications V_X would be the gain

control input, V_Y the signal input. The output is then the drain current difference, given by

$$\Delta I_D = I_D' - I_D'' = \frac{4I_{DSS}}{V_T^2} \left(V_X - \frac{V_T}{2}\right) V_Y \tag{5.52}$$

This is very similar in form to (5.20), which applies to a "linearized" nonsaturation-region FET, even though the structure is completely different. The desired $V_X V_Y$ product is accompanied by a device-dependent scale factor and an offset on the V_X input. The input voltage limits, furthermore, are proportional to V_T. However, it seems likely that these aberrations could be overcome by compound configurations similar to those discussed in section 5.2. For example, if a second such multiplier were connected in a "slave" arrangement such that its output (drain current difference) were forced equal to the output of the first multiplier, the result would be to impose the condition

$$\left(V_{X1} - \frac{V_T}{2}\right) V_{Y1} = \left(V_{X2} - \frac{V_T}{2}\right) V_{Y2} \tag{5.53}$$

which could be exploited for a multiplication independent of the scale factor $4I_{DSS}/V_T^2$. Another approach might force the drain current difference of one multiplier to a certain (input) value, drive a second multiplier such that the two multipliers had the same V_X, and obtain the result

$$\Delta I_{D2} = \frac{\Delta I_{D1} V_{Y2}}{V_{Y1}} \tag{5.54}$$

--a parameter-independent, bipolar-input multiplier/divider developed entirely from FETs in saturation.

This discussion has introduced some conceptual approaches to analog multiplication using saturation-mode FETs. Many variations on the "quarter-square" theme are possible, and this area should be investigated further. The accuracy of any such circuit hinges, of course, on the applicability of (5.46) and the matching of the FETs, neither of which can be taken for granted.

Other nonlinear analog functions can be obtained from (5.46) as well. Seriki and Newcomb [5.52] described a practical squaring circuit based on a pair of MOSFETs. The gate-to-source voltages are driven with signal increments of opposite polarity and the drain currents are summed. This cancels any odd-order components in the V_{GS}-to-I_D relation, enhances the accuracy of the fundamental squaring behavior in the FETs, and accommodates a bipolar input. Seriki and Newcomb also suggested a quarter-square multiplier as an application of the squaring circuit.

5.3.4 MOSFETs in Weak Inversion

The threshold voltage of a MOSFET is usually defined as the value of gate-to-source voltage at which the strong inversion condition occurs at the silicon surface under the oxide. When the gate-to-source voltage is reduced below this value, the transistor enters a region of operation where drain current flows by diffusion rather than drift, is relatively insensitive to drain-to-source voltage,

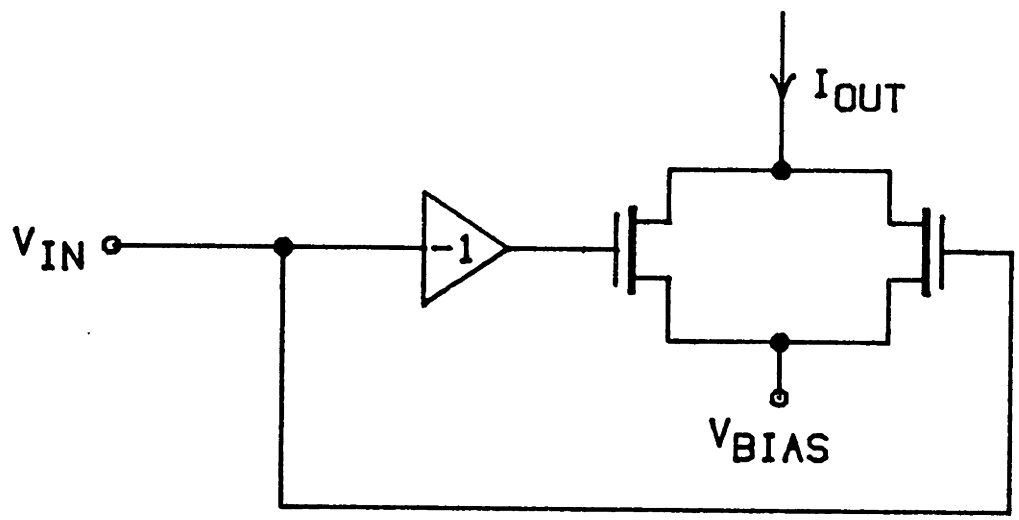


Figure 5.9 -- Simple Squaring Cell Using FETs in Saturation

SYMMETRY OF THE CIRCUIT ENHANCES THE FET SQUARE-LAW TENDENCY BY CANCELLING ODD-ORDER DISTORTION. THIS IS THE BASIS OF THE CIRCUIT OF SERIKI AND NEWCOMB [5.52].

and depends more or less exponentially on V_{GS}. This effect, considered a nuisance in digital circuits, is potentially useful for large-signal analog functions.

Vittoz and Fellrath [5.53] have reviewed the work in this area and concisely summarized the theory. In terms of V_{GS}, V_{DS} and V_{SB}, their expression for drain current is

$$I_D = \frac{W}{L} I_{D0} \, e^{\left(\frac{1-n}{n}\right)\frac{qV_{SB}}{kT}} \, e^{\frac{qV_{GS}}{nKT}} \left[1 - e^{\frac{-qV_{DS}}{kT}}\right] \qquad (5.55)$$

in which I_{D0} is a characteristic current that varies with processing and n is a capacitance-ratio parameter related to the surface depletion-region capacitance C'_d and the oxide capacitance C'_{ox} (both per unit area) by

$$n = 1 + \frac{C'_d}{C'_{ox}} \qquad (5.56)$$

The other parameters follow the usual MOSFET definitions. The value of n can be controlled somewhat through the process design; it ranges from 1 to about 3, tending to 1 for a lightly doped bulk region.

The factor in brackets in (5.55) will be effectively unity under normal operation (substantial V_{DS}). The parameters n and I_{D0} are both weak functions of V_{SB}; moreover, I_D depends significantly on V_{SB} through the first exponential in (5.55). For $n > 2$ the dependence is greater on V_{SB} than on V_{GS}.

Equation (5.55) is applicable over some three or four decades of drain current typically, limited at the lower extreme by leakage currents and at the upper by departure from weak inversion conditions. The weakly inverted surface does not necessarily correspond to extremely low drain currents, although it does for MOSFETs of geometries used normally in the other operating modes, with (W/L) factors clustered around unity. In order to use (5.55) at currents above the nanoampere region, the (W/L) factor must be much larger than unity. Vittoz and Fellrath derived the upper limit on drain current

$$I_D \leq \frac{n-1}{e^2} \frac{W}{L} \mu_n C'_{ox} \left(\frac{kT}{q}\right)^2 \qquad (5.57)$$

for weak inversion operation.

The exponential nature of (5.55) immediately suggests the use of some powerful design techniques developed for bipolar transistors. Exact large-signal transfer characteristics could in principle be created following the "translinear" approach (see chapter 6). Although Vittoz and Fellrath, as well as others, have investigated some useful analog circuits (amplifiers, oscillators, current sources, etc.), these primarily have been other than translinear configurations. Implementation of multipliers and other variable-gain elements is a natural and potentially fruitful area that needs research.

This discussion will not include any practical variable-gain configurations using weak inversion operation, since they

have not yet been investigated in depth. However, the obvious potential for translinear circuits in this mode prompts some conclusions regarding such circuits and the important differences between (5.55) and the analogous relation in bipolar transistors.

First, weak inversion is a low-current regime, or more accurately a mode of low current per unit channel width. Because the oxide and reverse-biased junction capacitances are not correspondingly small, there are severe limits to the operating frequencies and large-singal slewing rates in these circuits. While it may be possible to cover the audio range and perhaps somewhat beyond, it is unreasonable to expect analog functions extending into megahertz bandwidths.

The weak-inversion MOSFET is inescapably a four-terminal device. Because drain current depends exponentially on bulk-to-source bias, it is very advantageous to have independent bulk electrodes. Processes such as standard NMOS, with the bulk regions occupying the common substrate, will require significantly different circuit topologies than their bipolar counterparts in order to realize translinear functions. The source potentials will be tightly constrained; circuits depending on the matching of V_{GS} will need identical source voltages, and transistors will be unable to operate at source voltages much above the substrate voltage without being effectively shut off.

These considerations derive from the exponential behavior in (5.55). Additional problems are caused by the sensitivities of n and I_{DO} to V_{SB}, sensitivities which are not shown explicitly in (5.55). Typically, the accuracy of translinear circuits depends

critically on the fidelity of the exponential voltage-to-current relationship, as discussed in chapters 7 and 8. As pointed out by Vittoz and Fellrath, such circuits can only be used with transistors whose V_{SB}s are within a few $\frac{kT}{q}$ of each other. Thus, the "stacking" of V_{GS}s, analogous to the emitter-to-base coupling of multiple bipolar transistors, is ruled out whenever the bulk voltages are predetermined.

The FETs are not without advantages in this mode, especially if the bulk regions are isolated, so that, for example, the bulk of each can be tied to the source. Current flow into the gate is effectively zero, thus removing one of the common problems encountered in bipolar-transistor translinear circuits. Also, the FETs can be designed with a wide range of aspect ratios (W/L). Translinear circuits with weak-inversion MOSFETs are a provocative concept, worthy of further study.

5.4 Special-Purpose Field-Effect Devices

The discussion thus far in this chapter has assumed the use of standard junction and MOS FETs, with emphasis on the planar structures compatible with monolithic circuits. Other FETs and even other transistor types have been used and proposed for variable-gain applications. The emphasis generally has been on improving the controllable small-signal characteristics (channel resistance or transconductance), which falls outside the main thrust of this thesis. However, the topic is significant in the broader context of variable gain, and it illuminates some limitations of conventional FET structures.

5.4.1 Devices with Extended Nonsaturation Region

Gosling, in his study of voltage-controlled attenuators [5.23], was concerned with nonsaturation-region JFETs operated in a simple "voltage-controlled resistor" mode (see section 5.2.4). He noted that the classical uniform-channel JFET shows a rapid increase in small-signal drain-to-source resistance as the gate-to-source bias approaches the pinchoff (threshold) value. The importance of this depends on the application; in feedback-type automatic level control circuits, for example, such a characteristic can give undesirable loop dynamics.

Gosling advocated the use of an etched-alloyed JFET, an annular structure that exhibits a relatively "soft" saturation with respect to V_{DS} as well as a more gradual sensitivity to V_{GS}. This was attractive at the time (1965), planar transistors being relatively novel and monolithic analog FET circuits being nonexistent. Today, however, it is less readily available and it is quite incompatible with monolithic fabrication. Similar results might be possible with a special monolithic structure, if indeed the benefits justify it, which is uncertain.

Strachan and Townsend [5.54] suggested a monolithic MOSFET structure to achieve a similar aim. In the MOSFET, as they pointed out, the small-signal drain-to-source conductance is essentially a linear function of V_{GS} in nonsaturation. If several MOSFETs with different threshold voltages and channel aspect ratios are connected in parallel, the resulting device exhibits a very wide range of conductance as successive MOSFETs turn on at different values of gate voltage. The concept is similar to the method of obtaining

wide transconductance range in the remote-cutoff pentode vacuum tube. Strachan and Townsend proposed a planar MOSFET structure in which both channel length and local threshold voltage would vary across the device, giving a continuous equivalent of the multiple-FET configuration.

Nishizawa et al. [5.55] have developed an entirely different transistor structure, the static induction transistor (SIT). Under gate-to-source bias this device has a resistive drain-to-source characteristic that does not saturate as a FET does. However, the drain-to-source behavior becomes nonlinear at low currents; the device does not exhibit the ohmic property found in a FET around $V_{DS} = 0$.

5.4.2 Tetrode FETs for Variable Gain

Yang [5.56] investigated the use of a two-gate JFET for amplitude modulation. The device considered was a standard rectangular FET with electrically independent gates on either side of the channel. In saturation, such a FET can provide a small-signal transconductance from one gate to the drain that is nearly a linear function of the D.C. bias on the other gate. However, the large-signal transconductance is significantly nonlinear, introducing signal-path distortion, and Yang found it necessary to use a narrow-band configuration with a tuned load circuit in order to mitigate the distortion.

CHAPTER 6

LARGE-SIGNAL GAIN CONTROL USING BIPOLAR JUNCTION DEVICES

6.1 Introduction

This chapter presents a brief survey of the various techniques for creating large-signal variable-gain elements from bipolar junction transistors (BJT) and related devices. The number of basic configurations is surprisingly small, and the objective here is to examine the many published examples in a unified manner.

Because this topic has received a far more systematic treatment in the literature than the subject of the previous chapter, the discussion will defer to the references on many details. In particular, simplified transistor models will be used for clarity. Base currents often will be neglected. Chapter 7 goes into details of the logarithmic current-voltage characteristics on which much of this material is based.

6.2 Underlying Principles

6.2.1 Current-Voltage Characteristics

Bipolar transistors in forward-active bias are one of several device types obeying relations of the form

$$V_{BE} = \frac{kT}{q} \ln\left(\frac{I_C}{I_S}\right) \tag{6.1a}$$

$$I_C = I_S \exp\left(\frac{qV_{BE}}{kT}\right) \tag{6.1b}$$

Such characteristics hold, in modified form, for diodes and various other semiconductor devices. Details and limitations in equations (6.1) are treated at length in section 7.2.

A single device obeying (6.1) can be readily employed to give a controllable small-signal conductance or transconductance. We have

$$g_m = \frac{dI_C}{dV_{BE}}$$

$$= \frac{qI_S}{kT} \exp\left(\frac{qV_{BE}}{kT}\right) \tag{6.2a}$$

$$= I_C / \left(\frac{kT}{q}\right) \tag{6.2b}$$

Circuits are easily (and frequently) constructed to employ this property. The transconductance yields a signal gain, or alternatively a driving-point conductance, that follows a linear control law if I_C is the controlling variable or an exponential one if V_{BE} is the control. Such a small-signal application is available with any nonlinear device. The forms of control law with the bipolar transistor are rather convenient, but the basic result is temperature-sensitive, and signals causing V_{BE} excursions of the order of kT/q or greater violate the "small-signal" condition, causing excessive distortion.

Because of the logarithmic/exponential functional relation, a rich set of large-signal variable-gain techniques is also available from the BJT. These techniques depend on using two or more transistors to produce a fundamentally linear signal-path characteristic. At the

same time, undesirable dependence on device parameters (T, I_S, etc.) can be removed. These circuits have enjoyed great industrial success, due largely to the accuracy and predictability of (6.1).

6.2.2 Addition of Logarithms

One rather obvious way to exploit (6.1a) is to take the logarithms of two currents, sum them, and then obtain the antilogarithm of the sum, using (6.1b). This is the basis for some successful analog multipliers. As stated, however, it is a one-quadrant multiplication of currents. The requirement for a bipolar signal input in variable-gain applications requires a more indirect approach. Moreover, many variable-gain applications do not use a linear control law (see chapter 3).

In practice, differential transistor pairs are commonly used in variable-gain elements. These conveniently accommodate bipolar signals; moreover, they have properties that make them, perhaps, more fundamental than the transistor in such circuits.

6.2.3 Two Ways to Use Differential Pairs

Much of the utility of bipolar transistors in large-signal variable-gain circuits follows from a key characteristic of differential (long-tailed) pairs. For a given total collector current, the fraction of this current that flows in each transistor depends on the base-emitter voltage difference but not on the total current (Figure 6.1). This is not true of field-effect transistors; it follows directly from the logarithmic/exponential characteristic of (6.1). Indeed, requiring that a differential pair of three-terminal active

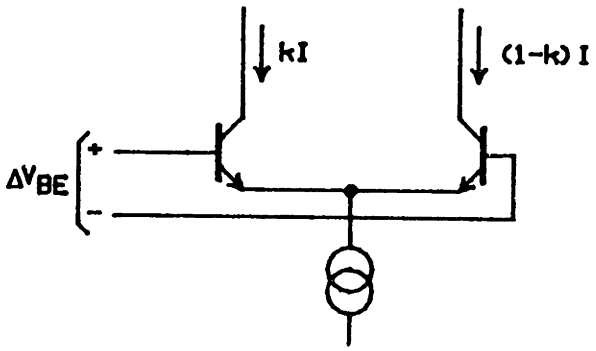


Figure 6.1—Differential BJT pair.

k IS THE FRACTION OF TOTAL CURRENT THAT FLOWS IN THE LEFTHAND COLLECTOR. IT DOES NOT DEPEND ON THE TOTAL CURRENT I, BUT ONLY ON ΔV_{BE} (EXCEPT FOR SECOND-ORDER EFFECTS).

devices have this property leads to a differential equation whose solution is the exponential of (6.1b) for each device.

In practice this remarkable current-steering property is exploited in two distinct ways. Most successful large-signal variable-gain circuits using bipolar transistors employ one or both of these techniques.

Class 1:

Differential collector currents are imposed on a pair of transistors in response to the signal input. The ΔV_{BE} from this pair is conveyed to the bases of a second pair operating at a different total collector current. The output is the differential collector current in the second pair; control is accomplished by modulating the total currents or the absolute V_{BE} values.

Class 2:

A unipolar current, containing a signal component, is imposed as the total emitter (or collector) current of a transistor pair. An imbalance of V_{BE} steers part of this signal current into an output circuit while the remainder is discarded or subtracted out. The V_{BE} imbalance is the control input.

Class-1 circuits appear in various forms; a typical case is shown in Figure 6.2(a). This is known as a "linearized base-driven pair" or "Gilbert's quad." Various manifestations of class-2 circuits have been called, for example, "emitter-driven pair" and (in a

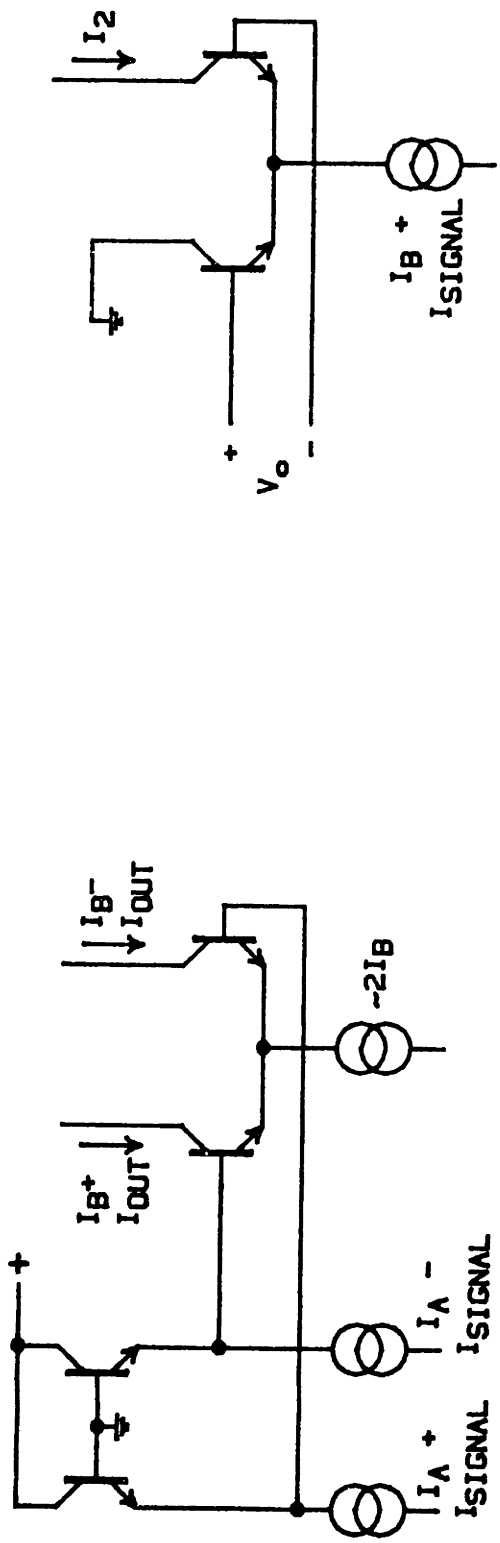


Figure 6.2—Examples of the two standard configurations for differential-pair VGEs

different form) "controlled-cascode multiplier." Figure 6.2(b) shows the basic concept.

6.2.4 Translinear Circuits and Current Variables

Class 1, defined above, is in fact a special case of a rich set of bipolar-transistor analog circuits. These combine multiple transistors to obtain current-input current-output transfer functions, linear and nonlinear, that are fundamentally accurate, independent of temperature and limited only by second-order effects. They are based directly on the exponential law of (6.1b), which implies a trans-conductance linear in collector current; hence, translinear.

The general principle was first articulated [6.1] and has been extensively explored by Gilbert.* A slight generalization was noted by Hart [6.2], although it was actually implicit in Gilbert's early papers [6.3, 6.4]. The theoretical basis is reviewed in section 7.3.3 of this report, in preparation for treating certain second-order effects. This chapter will emphasize practical examples.

A crucial message of the translinear formalism is that any large-signal gain circuit (fixed or variable) built from BJTs, in order to be exactly linear, must use currents as input and output variables. In practice, of course, many voltage-mode amplifiers achieve approximate linearity through the use of linearizing artifices (feedback or feedforward). But since the variable-gain mechanism in BJT circuits mandates open-loop operation, all high-performance VGE configurations require a current-mode signal input and output.

---

*Although named more recently [6.1], the principle appeared in an appendix to Gilbert's classic "amplifier" paper [6.3].

This introduces a voltage-to-current conversion issue if the circuit is to process voltage signals, as is often true. That issue is not treated here.* Note that the current requirement applies to the signal input and not the control input of the VGE. Thus the common emitter node of a BJT differential pair is a natural signal (current) input (class 2, Figure 6.2(b)) while additional transistors are required to convert the bases to a signal input (class 1, Figure 6.2(a)). Of course, linearizing drive is also required on the bases in Figure 6.2(b) if a linear control law is sought. The circuit then becomes similar to Figure 6.2(a) except for the definitions of signal and control inputs.

6.3 Early BJT Variable-Gain Circuits

Prior to the evolution of systematic large-signal methods using bipolar transistors, numerous specialized circuits were explored for multiplication and gain control. Hurtig's circuit [6.5], one of the earliest (1955), was actually a crude form of the "controlled-cascode" multiplier, which recently has been used for high-performance variable gain [6.30]. A current-mode input signal was steered between a common-base (germanium) transistor and an alternate path through a diode to ground. The steering was controlled by a current input.

Other early VGEs were based mostly on small-signal approximations, and were therefore limited in dynamic range. Kahn [6.6] and Romano [6.7] described applications of forward-biased diodes as current-controlled variable attenuator elements. Although not a

---

*See, for example, the study of bipolar-transistor voltage-current converter circuits by Mack [6.15].

viable large-signal technique, this was relatively successful for signal-normalizing applications like automatic gain control (see sections 3.1 and 3.2). In such applications, the peak voltage excursion across the diodes, and hence the signal-path nonlinearity, is approximately independent of gain setting. Put another way, the VGE has minimum gain when the input signal is maximum. Figure 6.3 shows a simple example; more sophisticated versions, such as Romano's [6.7], used a pair or bridge of diodes for symmetric saturation behavior and reduced control feedthrough.

Other small-signal approaches used the current-controlled transconductance of bipolar transistors. Deb and Sen [6.8] described a rudimentary multiplier using AC-coupled common-base transistors. Frater [6.9] and Morton [6.10] used differential transistor pairs in "variable transconductance" schemes; these were basically forerunners of translinear multipliers, without the linearizing current-input circuits. Morton's circuit did use a linear voltage-current converter to supply the emitter currents of balanced differential pairs; thus it had a linear large-signal response on one input (the class-2 technique).

The limitations of a small-signal philosophy were recognized early. In 1963, Kundu [6.11] pointed out the narrow dynamic range of such circuits by using a large-signal analysis. He subsequently published [6.12] a large-signal multiplier design using logarithmic/exponential techniques. It was, however, an unwieldy AC-coupled configuration.

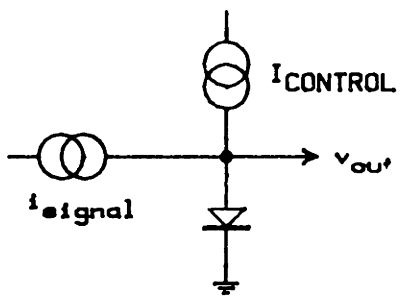


Figure 6.3 - A diode as a simple variable-gain element.

DISTORTION CAN BE SMALL IN APPLICATIONS LIKE AUTOMATIC GAIN CONTROL, WHERE GAIN OF CIRCUIT FALLS AS INPUT AMPLITUDE RISES.

Novel use was made of saturated bipolar transistors in large-signal, albeit device-sensitive, VGE circuits. Davidson and Leighton [6.13] showed how a pair of parallel, saturated BJTs, one inverted, gave a linear variable resistance to ground. Grasselli and Stefanelli [6.14] described in 1966 what amounts to a translinear two-quadrant multiplier circuit, with current inputs and output. This circuit used a pair of saturated transistors, connected in inverse-parallel, driving the bases of a differential pair.

6.4 High-Performance BJT Variable-Gain Circuits

The "contemporary" circuit techniques that follow are distinguished by good signal-path linearity and highly predictable characteristics compared to the early methods. Although distortion sources remain (the subject of chapters 7 and 8), they arise from deviations from (6.1) and device mismatch, rather than equations (6.1) themselves.

6.4.1 VGE Configurations Using Linearized Base-Driven Differential Pairs (Class 1)

Several references [6.3, 6.4, 6.16-6.24] deal with variations on Figure 6.2(a), which is perhaps the quintessential translinear circuit. This circuit is also used to investigate the effects of log-conformance error in chapter 7.

The "input pair" in Figure 6.2(a) operates at a quiescent current of I_A per device, with a differential signal current of I_{SIGNAL}. Assuming that beta is large, the "output pair" operates at a quiescent collector current of I_B per device, and the output current, the signal-dependent component appearing at the collectors, is

$$I_{OUT} = (I_B/I_A) I_{SIGNAL} \tag{6.3}$$

It is assumed here that all four devices are matched. The basic result is treated in more detail in section 7.4. The saturation limits, where both input and output reach a hard bound, are given by

$$\left|I_{SIGNAL}\right| \leq I_A \tag{6.4a}$$

$$\left|I_{OUT}\right| \leq I_B \tag{6.4b}$$

A useful feature of the class-1 variable-gain circuits is their ability to operate in either a range-compressing or a range-expanding mode (section 3.1). Either of the currents I_A and I_B, or both at once, may be used as gain-control input. The choice determines how the linear ranges of signal input and output, as well as the noise level, will vary with gain setting.

For example, if G is the desired current gain, we may take

$$I_A = I_{REF} \ , \ I_B = GI_{REF}$$

or

$$I_A = I_{REF}/G \ , \ I_B = I_{REF}$$

or

$$I_A = \frac{I_{REF}}{\sqrt{G}}\ ,\ I_B = I_{REF}\sqrt{G}$$

(where I_{REF} is some reference current).

All three of these control strategies give the same current gain G, but different relationships between gain and linear range. In the first example, the input overload point remains constant while the output excursion expands and contracts with gain. In the second case the reverse is true. In the third, both input and output limits vary, but over a smaller range than the current gain.

Moreover, the noise behavior is different. Noise studies by Sansen and Meyer [6.24] and Bahnas et al. [6.22] have demonstrated that transistor base resistance is the dominant noise source at typical operating currents, and that the noise magnitude depends significantly on the output-pair current I_B. Bahnas et al. derived an expression for white noise in the circuit of Figure 6.2(a), from which we can obtain

$$\overline{(I_{OUT})^2}/\Delta f \cong \frac{4qr_b}{V_t} I_B^2 + (G+2)qI_B \qquad (6.5)$$

The expression gives power spectral density (mean squared noise current per Hertz) in I_{OUT} when the signal input is zero. q is electron charge, V_T is kT/q, r_b is the ohmic base resistance, and G is the nominal current gain, I_B/I_A. This assumes four transistors of identical construction.

Thus, if I_A is fixed and I_B varies, as in the first example above, then the noise level in the output will be roughly proportional to gain (taking R.M.S. noise current, which is the square root of the last expression). On the other hand, if I_B is held fixed and I_A varies to set the gain, the situation is different. The R.M.S. output noise will tend to remain constant if the first term on the righthand side of (6.5) dominates, or to vary with gain if the second term is larger. This depends on the particular values of r_b, I_B and G. This analysis neglects the noise in the current sources, which is sometimes very important as well.

Thus, the signal limits and noise behavior, as functions of gain, may be tailored to best fit the application. In a range-expanding role, such as amplitude modulation or audio dynamic range expansion, the input-pair current I_A should be fixed while the output-pair current I_B controls the gain. This causes the output noise level to fall as the output signal is reduced.

On the other hand, a range-compressing application, like automatic gain control, has different requirements. The output swing of the variable-gain circuit will be approximately constant, so it is inappropriate to modulate gain by varying the output-pair current. Instead I_B should be held constant and I_A should be varied. If r_b dominates among noise sources, as is usually the case, output signal-to-noise ratio will be maximized this way. This method yields a reciprocal-type control law (the gain varies as $1/I_A$), but the control law can be changed at will, in principle, by adding a translinear nonlinearity to drive the input-pair currents.

An alternative approach for range-compressing applications is to vary I_B and use the resulting variable-gain element in the feedback path of an amplifier. Bandwidth of the resulting circuit is generally determined by the amplifier, and will tend to be much lower than for the open-loop VGE. This method was used by Todd [6.19]; the same VGE served for both range expansion and (in a feedback path) range compression in an audio compander.

Nonideal effects that impair the performance of class-1 circuits include odd-order distortion due to log-conformance error (e.g., ohmic resistances), treated in chapter 7, and transistor mismatch. Unequal I_S values in the four transistors, due to processing variations or mismatched discrete devices, yield even-order distortion on the order of one percent per millivolt net V_{BE} mismatch (at 300 degrees Kelvin). This was examined in detail by Gilbert [6.4]. Trimming may be necessary to zero the base-emitter voltage offset if the signal-path linearity is critical.

Because it is an open-loop circuit with current-mode input and output and low-impedance internal nodes, Figure 6.2(a) potentially has very wide bandwidth. Gilbert showed that the dynamics are dominated by the r_e of the input pair (i.e., $1/g_m$) forming a pole with the C_π of the output pair [6.3]. This gives a pole frequency of about f_T/G, where f_T is the transition frequency for the transistors.

However, the collector-to-base capacitances (C_μ) of the output pair provide an out-of-phase feedforward path that limits the available attenuation at high frequencies. As in other situations where a capacitance feeds around an inverting transconductance, a right-half-plane transmission zero is created at a frequency

$$\omega_z = \frac{g_m}{C_\mu}$$

This also distorts the phase response of the circuit. Sansen and Meyer [6.23], Gilbert and Holloway [6.30], and others have noted this problem, which is one of the main recommendations for the alternative (class-2) VGE structure in wideband applications.

The original and definitive development of these circuits is due to Gilbert, in a pair of papers [6.3, 6.4] and their precursor, an ISSCC report [6.16]. Subsequently, Sansen and Meyer obtained fundamental results on distortion [6.23] and noise [6.24] in these and related (class-2 and six-pack) variable-gain circuits. Schlotzhauer and Viswanathan [6.17] presented a practical version, with attention to base-current effects, for use as a multiplier. Jung [6.18] discussed practical versions, of moderate performance, that could be constructed from available general-purpose integrated circuits. Todd, as mentioned previously, used the technique in a commercial monolithic audio compander, the Signetics NE570 [6.19].

Two other versions that have appeared are different from Figure 6.2(a) but retain the essential class-1 structure, defined in section 6.2.3. Curtis's voltage-controlled amplifier [6.20] is an example of gain control accomplished not by varying pair currents but by directly changing the absolute V_{BE} level of one of the pairs (independent of V_{BE} <u>difference</u>, which is signal-dependent and identical in both pairs). Figure 6.4 shows the basic circuit. The differential current gain of the circuit is unity when V_e is zero; nonzero V_e results in gain ($V_e<0$) or loss ($V_e>0$). In general,

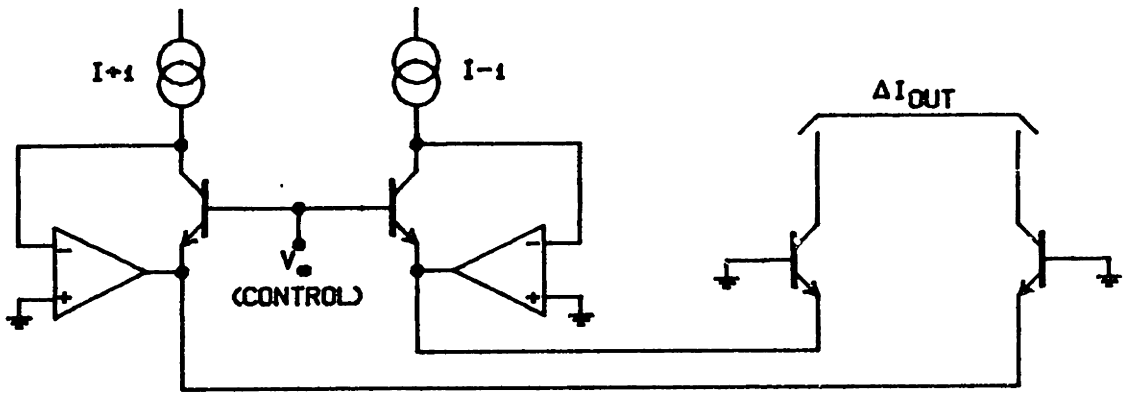


Figure 6.4 - Curtis's VCA circuit.

THIS IS A CLASS-1 CIRCUIT WITH CONTROL APPLIED TO V_{BE} RATHER THAN I_C

$$\text{CURRENT GAIN} = \exp\left(\frac{-qV_e}{kT}\right) \tag{6.6}$$

The exponential form and temperature sensitivity are to be expected whenever a V_{BE} (voltage) difference forms the control input.

The current-gain cell of Hamilton and Finch [6.21] is an ingenious variant that links the input and output pairs of a class-1 circuit in a tight positive feedback loop. The result is a circuit with single-ended current (signal) input and output, with the input and output at the same DC level. It realizes the full bandwidth of the basic Gilbert cell of Figure 6.2(a), unlike other single-ended versions that employ high-gain feedback loops to avoid a differential input. The circuit also exhibits a symmetrical reverse transmission from output to input, which may be a nuisance for some applications.

6.4.2 VGE Configurations Using Emitter-Driven Differential Pairs (Class 2)

References [6.23-6.30] consider variable-gain circuits based on the second method of exploiting differential transistor pairs. For the circuit of Figure 6.2(b), the large-signal transfer characteristic (assuming large beta and neglecting resistive parasitics) is

$$I_2 = \frac{I_B + I_{SIGNAL}}{1 + \exp(qV_c/kT)} \tag{6.7}$$

The inherently linear dependence of I_2 on I_{SIGNAL} motivates this approach.

In practice, balanced configurations are usually preferred over the single pair of Figure 6.2(b). Often two such pairs are driven with opposite polarities of I_{SIGNAL} and the two outputs subtracted. This results in cancellation of both control feedthrough (due to the I_B term in the numerator of (6.7)) and the even-order distortion that arises from nonzero resistive parasitics in the transistors [6.23, 6.28, 6.29].

Class-2 variable-gain circuits differ from class-1 circuits in several notable respects:

- Gain is determined by the voltage input V_C and not by the bias current I_B. Therefore, class-2 circuits operate naturally in a range-expanding mode. This contrasts with class-1 circuits, which can be designed for either range-expanding or range-compressing operation.
- Class-2 circuits are capable of controlled attenuation over a very wide bandwidth, since the differential pairs are operating common-base. Signal feedthrough via device capacitances is still a problem, but less so than in the class-1 case. Feedthrough is in-phase.
- Distortion due to resistive parasitics occurs, in the same manner as class-1; but the effect is different: even as well as odd-order distortion in the signal path. The magnitude of distortion, for a given operating current, is lower for class-2 [6.23].

- Class-2 provides current gains less than unity,

  whereas class-1 has no such constraint.

Class-2 circuits achieved early acceptance as high-frequency variable-gain elements, as in the IF amplifiers of Davis and Solomon [6.25] and McCalla [6.26]. Sansen and Meyer, in their work on noise and distortion, identified a balanced class-2 form as superior in several respects to alternative circuits. They developed a high-dynamic range monolithic VGE based on it [6.24]. Both Sansen and Meyer [6.23] and Yen [6.28, 6.29] derived low- and high-frequency distortion results for the basic circuit of Figure 6.2(b).

Faulkner and Grimbleby [6.27] obtained range-compressing operation from class-2 circuits by placing them in the feedback path of a high-gain operational amplifier. They employed a balanced configuration of two differential pairs, one NPN, one PNP. More recently, Gilbert and Holloway [6.30] described a class of "controlled-cascode" multipliers consisting of balanced class-2 circuits with both voltage- and current-mode control inputs.

6.4.3 A VGE Based on a Pair of One-Quadrant Cells

One of the few successful variable-gain circuits not based on class 1 or class 2 is shown in Figure 6.5. This configuration is used in the 160 series of audio amplitude compressors manufactured by dbx, Incorporated. It illustrates yet another approach: two precision one-quadrant variable-gain cells, each processing one polarity of the signal.

When I_{SIGNAL} is positive, the NPN pair is active and the PNP pair is cut off, with reverse-biased base-emiter junctions. The NPN pair has a collector-current ratio

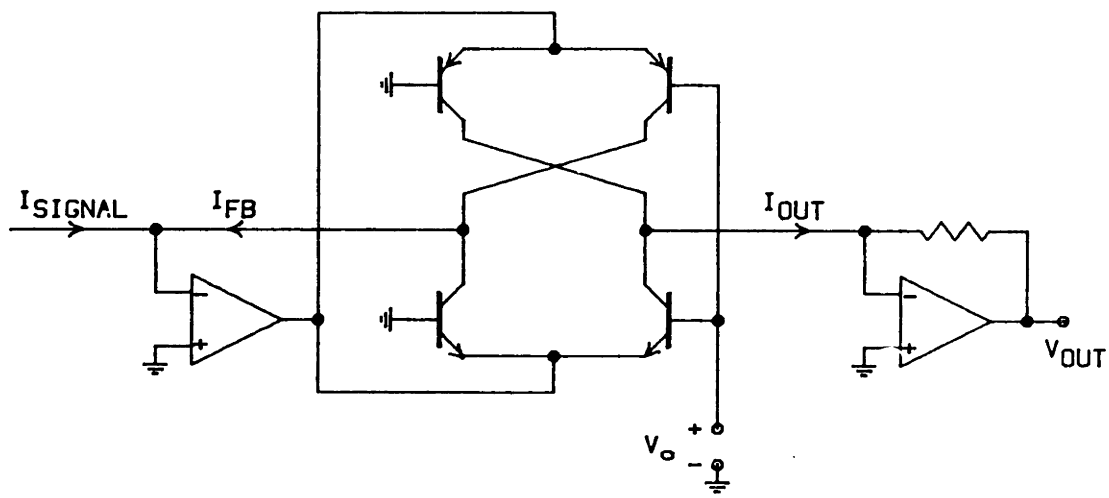


Figure 6.5 — The dbx VCA circuit.

BASED ON TWO UNIPOLAR CURRENT SCALERS, ONE FOR EACH POLARITY OF THE INPUT.

$$\frac{I_{OUT}}{I_{FB}} = \exp\left(\frac{qV_c}{kT}\right) \tag{6.8}$$

(compare with (6.6)). The lefthand op-amp forces $I_{FB} = -I_{SIGNAL}$, so the signal current gain is also given by (6.8). The righthand op-amp provides a virtual-ground summing node and a voltage output.

When I_{SIGNAL} is negative, the NPN pair cuts off and the PNP pair becomes active; thus the two pairs act as precision current-ratio circuits, providing the same ratio for their two corresponding signal polarities. A fast op-amp is necessary in such a system to avoid a momentary dead zone at the zero-crossing.

6.4.4 "Six-Pack" Multipliers

One of the best-known translinear circuits is the four-quadrant linearized transconductance multiplier or "six-pack" [6.4, 6.16, 6.22-6.24, 6.31-6.34], shown in essence in Figure 6.6. This may be regarded as the nexus of class 1 and class 2 circuits, with inputs applied to both the bases and emitters of the righthand quad.

This configuration is discussed extensively in Gilbert's early papers [6.4, 6.16]. High-performance versions where feedback compensates for imperfections in the voltage-to-current converters have been described also [6.31, 6.32]. At the same time, it has been applied to the remote programming of wideband communications channels [6.34].

The six-pack is primarily a multiplier rather than a variable-gain element. Although it has been used occasionally in VGE applications, as in the expander circuit of Erratico and Caprio [6.33], it

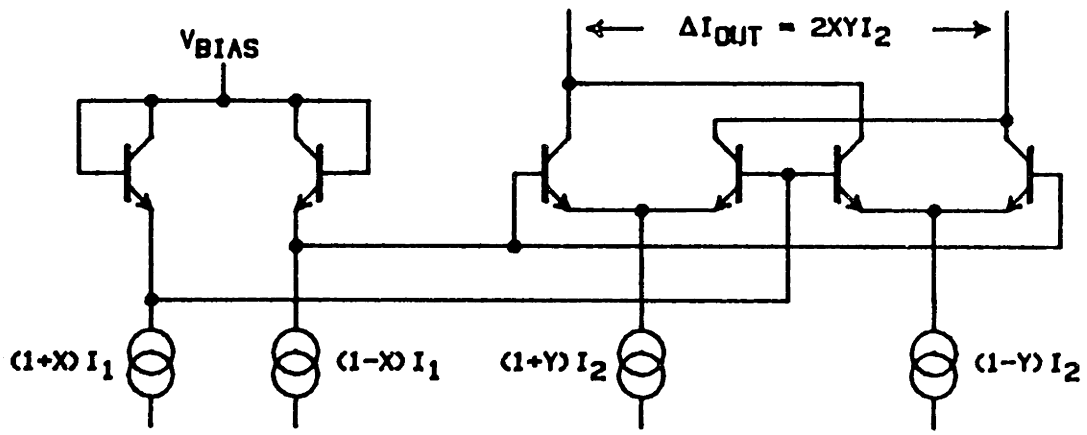


Figure 6.6 —
The basic "six-pack" multiplier

has major drawbacks. Because it depends on cancellation of two out-of-phase signal paths, the zero-output setting is very sensitive to mismatch. Fractional distortion and signal-to-noise ratio degrade rapidly at high values of attenuation, as noted by Sansen and Meyer [6.23, 6.24].

The major feature of the six-pack that distinguishes it from related circuits is an immunity to resistive parasitics. Log-conformance error due to ohmic resistances cancels out, provided the transistors are properly matched. The signal distortion that plagues class-1 and class-2 circuits at high currents and large gain ranges is thus avoided.

6.4.5 Brüggemann's Multiplier

Another four-quadrant configuration, due to Brüggemann [6.35, 6.36] is a variation on the "six-pack" scheme. Although not as attractive as the six-pack for monolithic work, it uses a distinct and novel principle.

The major innovation in going from small-signal "transconductance" multipliers [6.10] to the "linearized transconductance" six-pack (Figure 6.6) was the addition of the leftmost two transistors. This provided an input whose effect on the current steering in the righthand four devices was linear. Brüggemann developed an alternative approach, using a feedback loop (Figure 6.7). While one combination of collector currents gives the desired multiplier output, as in Figure 6.6, a different combination, as it happens, may be used to sense the degree of modulation in the two output pairs.

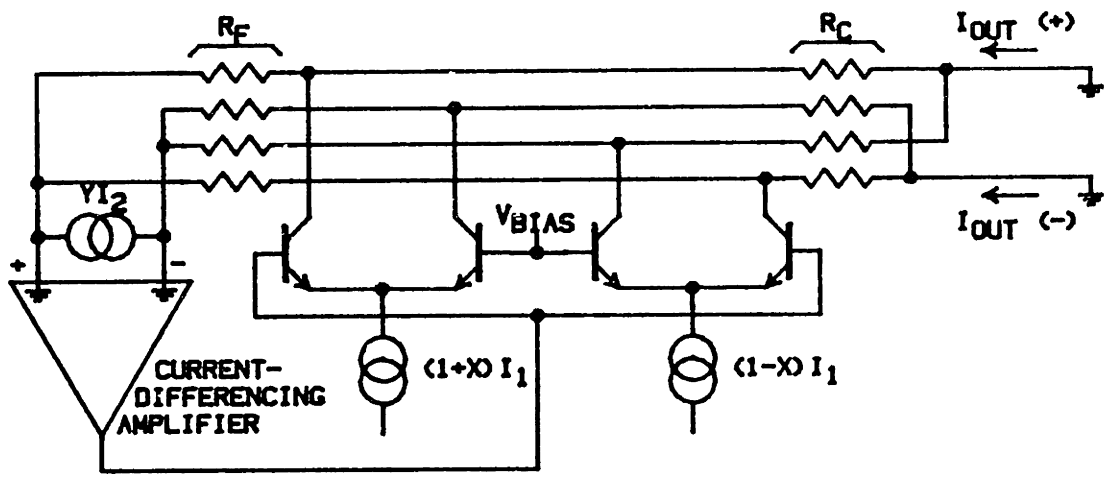


Figure 6.7 — Brüggemann's multiplier.

EACH COLLECTOR CURRENT IS SPLIT BETWEEN A LEFTHAND AND RIGHTHAND RESISTOR

Brüggemann's circuit splits each collector current and uses it for two purposes. The lefthand portions in Figure 6.7 are used to sense current steering, which is then forced equal to a desired input (yI_2) by a feedback loop. The righthand portions of collector current are combined to give the customary six-pack output signal (which is differential).

The disadvantages of this technique are the (slow) feedback operation and, most important, the need for a large number of matched resistors.

6.4.6 Multipliers Included for Completeness

To be noted in passing are a few other configurations that have appeared in the literature. Faulkner and Grimbleby [6.37] described a one-quadrant translinear multiplier based on stacked (NPN and PNP) square-law subcircuits. The multiplier obtained a product by the roundabout scheme

$$I_z = \sqrt{(I_x^2)(I_y^2)}$$

Viswanathan and Deep [6.38] also used an NPN/PNP transliner circuit for a one-quadrant multiplier; a version with offsets yielded four-quadrant operation. Scratchely [6.39] showed a four-quadrant multiplier with single-ended input and output, using an exotic translinear configuration.

These three circuits were rather complex and have little to recommend them in view of subsequent (and previous) designs that performed at least as well.

6.5 Special-Purpose Bipolar Devices

As in the case of FETs, some successful use has been made of non-standard semiconductor devices in variable-gain elements. Voulgaris and Yang [6.40] demonstrated linear applications of a PNPN four-layer device designed as a silicon controlled switch. Their application as a variable-gain amplifier and analog multiplier was basically limited to small-signal conditions, however.

More significant was the carrier-domain multiplier developed by Gilbert [6.41]. This was one of a family of special devices, constructed in a standard bipolar process, whose structure directly performed a useful operation. Multipliers were successfully fabricated [6.42, 6.43] and investigated as an alternative to "six-pack" cells. To date, performance has not been sufficient to give them a clear-cut advantage, but carrier-domain devices retain some promise in this area.

CHAPTER 7

LOG-CONFORMANCE ERROR EFFECTS IN TRANSLINEAR VARIABLE-GAIN CIRCUITS

7.1 Introduction

Bipolar junction devices (transistors and diodes) exhibit a logarithmic I-to-V relation that is undoubtedly the most useful large-signal nonlinearity in the realm of analog circuits. This relation is at the heart of all large-signal variable-gain circuits, and many other classes of circuits, using such devices. Circuits of this kind are often very sensitive to the fidelity of the logarithmic relation. It is therefore profitable to examine deviations from the ideal logarithmic law, and the effect these have on the input-output properties of translinear variable-gain configurations.

This chapter first examines aspects of device operation that give rise to log-conformance error (LCE). The basic consequences of such error for translinear circuits in general, and variable-gain circuits in particular, are then considered. The last section presents experimental results from a prototype variable-gain configuration. This chapter sets the stage for the discussion of LCE correction techniques in chapter 8.

In the bipolar trasnsitor, the logarithmic I_C-V_{BE} relation is useful over perhaps eight decades of collector current. While

this may at first seem adequate for most circuit applications, other considerations (such as bandwidth and gain) often constrain the designer to operate at the upper end of the current range. Practical interest, therefore, focuses on high- rather than low-current aberrations in the logarithmic behavior. This and the following chapter emphasize static nonlinearities (in a controlled signal path) that result from such aberrations. The work does not address other, potentially important, error sources such as mismatch errors or localized thermal effects, and treats finite-beta and high-frequency considerations only briefly.

7.2 Origins of Log-Conformance Error

7.2.1 Basic Log Characteristic

Bipolar junction transistors, pn diodes Schottky barrier diodes, MOS field-effect transistors with weakly inverted channels, and various other semiconductor devices all exhibit static current-voltage relationships of the general form

$$V = \frac{nkT}{q} \ln\left(\frac{I}{I_S}\right) \tag{7.1a}$$

or equivalently

$$I = I_S \exp\left(\frac{qV}{nkT}\right) \tag{7.1b}$$

over some range of currents. Here, n is a dimensionless constant on the order of unity, I_S is a device current parameter, and the

other constants have the usual meanings. I and V may both refer to the same pair of terminals (in two-terminal diodes) or they may refer to different terminal pairs.

The behavior of (7.1) is fundamental; it arises when a controllable energy barrier affects the thermal flux of carriers asymmetrically. The different devices that obey (7.1) are characterized by different values of I_S and n, and by different current ranges over which (7.1) is valid.

These equations are sometimes presented in a slightly different form, arising from a theoretical derivation. For example, a variant of (7.1b), often called the "ideal diode law," is

$$I = I_S \left[\exp \left(\frac{qV}{kT} \right) - 1 \right] \tag{7.2}$$

The -1 term arises from theoretical arguments and allows conveniently for I = 0 at V = 0. This -1 term is then cast aside as quickly as possible in discussions of active bias, since the argument of the exponential is then well above unity.

The approach preferred here is to do away with the "-1" term altogether, both because substantial forward current flow is indeed assumed and because the model of (7.2) is rarely if ever valid for values of I near I_S, contrary to the frequent implication. For the device types mentioned above, second-order effects generally determine I at the low-current extreme.

Of the various devices that satisfy versions of (7.1), bipolar transistors offer a quite remarkable current range and a value of n very close to unity when V is taken to be base-emitter voltage and I is collector current. Data obtained by Sah [7.1], Gibbons and Horn [7.2], and others for silicon transistors with zero collector-to-base voltage indicate adherence to (7.1) over some eight or nine decades of collector current (at room temperature). In these cases the measured value of n was within about one percent of unity over the current range. The scaling of current extremes depends on transistor construction, but as a rule of thumb, for a typical discrete transistor of the type measured, the useful range is from 10 picoamps to 10 milliamps.

Bipolar transistors are unique both in the wide range of currents and in the value of n so close to unity. Both pn junction diodes (the other major "log" elements) and the BJT emitter current behavior exhibit a much narrower useful range, as will be discussed shortly. For this reason, practical application of (7.1) to translinear and other circuits has emphasized an I_C-V_{BE} embodiment.

The value of n, sometimes called the emission coefficient (the letter m is also used) is unimportant in translinear circuits, to the extent that it is truly constant. In fact, any current-dependent shift in n is a form of log-conformance error (LCE) and will cause errors in such circuits. This occurs under high-level injection conditions in BJTs. Under low-level injection conditions, the value of n is often assumed to be exactly unity in BJTs, although slight discrepancies have appeared in the

published measurements. Recently Hart [7.3], building on the earlier work of Gummel and Poon (summarized by Getreu [7.4]), demonstrated how slight deviations from n = 1 could be predicted from fluctuations of base majority carrier charge. Hart also identified a tradeoff between achieving high current gain (beta) and n near unity through the design of the transistor. Typical values of n were within 0.5% of unity in Hart's examples, consistent with earlier measurements.

The collector-base voltage in a forward-active transistor influences both collector current and current gain through base-width modulation. Moreover, nonzero collector-base bias introduces a small leakage current across the terminals, although it is often negligible. Generally the circuits treated here operate transistors with fixed, possibly zero, V_{CB}, so that in principle there is no need to modify the form of (7.1). In translinear circuits, if all V_{CB} values are equal, their effects will tend to cancel; any significant collector leakage current becomes an equivalent input or output offset.

Unfortunately, practical transistors--especially low-power monolithic transistors--have appreciable series collector resistance. This causes two complications at high collector currents. First, significant signal-dependent voltage appears across the internal collector resistance, so true V_{CB} is no longer constant. This variation can be modelled in terms of its effect on V_{BE} through base-width modulation. The second phenomenon is a possible saturation of the transistor if the DC drop across

the collector resistance is too large. This will effectively limit the maximum collector current at low or zero V_{CB}.

With these limitations in mind, a basic large-signal static relationship useful at low-to-moderate currents in bipolar transistors is

$$V_{BE} = \frac{nkT}{q} \ln\left(\frac{I_C}{I_S}\right) \tag{7.3}$$

with n very close to unity (often taken equal to unity as a good approximation). However, as cited in the introduction, considerations such as bandwidth often preclude operating solely at "low-to-moderate" currents. If they did not, the remainder of this chapter would be unnecessary.

7.2.2 Limitations in Diodes

It is appropriate to briefly mention why the log conformance in pn diodes is greatly inferior to that of bipolar transistors, a distinction that has long been recognized [7.5].

A number of mechanisms are responsible for current flow in a forward-biased pn junction. In a diode, unlike a forward-active transistor, all of the current flows through the two terminals where the "log" voltage appears. At moderate current levels, the diode current is due mainly to the injection of minority carriers across the junction. This current follows the dependence of (7.1b) on applied voltage, with $n \cong 1$. At lower current levels, in silicon diodes, the dominant current is due to recombination within the depletion region; this current component follows (7.1b)

with $1 < n < 2$ (and a different value of I_S) [7.6, 7.7]. Additional currents may arise because of surface effects (the different properties of the depletion region, and the possibility of a channel, at the surface). Sah's classic and unified treatment of this topic [7.1] identified values of n between 1 and 4, and occasionally larger, for the different components.

The result of this heterogeneous current makeup is a current-voltage characteristic that behaves like (7.1) over short current intervals but undergoes one or more changes of the coefficient n over several decades of current. Additional deviations arise at high currents because of bulk resistances and high-level injection, but these phenomena also occur in transistors. The useful current range for a silicon diode, where n is substantially constant, will typically span two to four decades. Faulkner and Buckingham [7.8] observed a range of four decades with some diodes, with a value of n between 1 and 2.

Thus, although potentially useful in translinear configurations, the pn junction diode is less accurate than the bipolar transistor, as well as less convenient (lacking the separation between I and V ports). Similar accuracy considerations also yield an important distinction between collector and emitter currents in a transistor.

7.2.3 Complications in Base and Emitter Currents

A forward-biased base-emitter junction, like a forward-biased pn diode, carries various voltage-dependent current components. Only one of them--minority carriers in the base injected

from the emitter--contributes significantly to collector current. This accounts for the excellent fidelity of the logarithmic law when I_C is related to V_{BE}.

The remaining components flow in the base and emitter leads. At moderate collector currents, the base current is due essentially to back injection into the emitter and (less so in modern monolithic transistors) recombination of minority carriers in the base. Base current tends to be a constant (and small) fraction of collector current, so collector and emitter currents may be used almost interchangeably in (7.3). This situation prevails over an "optimal" range (perhaps one decade wide) of collector current. At high currents, corresponding to the onset of high-level effects that invalidate (7.3), back injection from the base to the emitter increases and the current gain falls.

At low collector currents, the other components of base current (chiefly that of recombination in the base-emitter space-charge region) assume greater importance [7.9]. As I_C is decreased, base current begins to fall less quickly than collector current (i.e., beta drops off). At a sufficiently low collector current (on the order of a nanoamp, typically) the base current will exceed the collector current. Yet (7.3) continues to hold down to currents so low that I_B is <u>substantially</u> greater than I_C.

Thus, as with a junction diode, the I_E-V_{BE} relation has a much smaller range of utility than the I_C-V_{BE} version. Emitter currents can be and are used as variables in translinear circuits, but over a restricted and well-chosen current range.

The discussion of diodes and emitter currents has established that significant log-conformance error problems apply to both, especially at low currents. The main thrust of this chapter, however, is the area of log-conformance error at moderate-to-high currents. Such error arises from device bulk resistances and high-level injection, and causes significant deviations from (7.3), the I_C-V_{BE} relation, at current levels that are otherwise attractive for analog circuits.

7.2.4 Base Resistance

The bulk material separating the base contact from the base-emitter junction is probably the single most troublesome source of log-conformance error in the I_C-V_{BE} relation. Its resistance introduces an inevitable voltage drop in series with the base lead, and hence an additive error in the righthand side of (7.3). At higher operating currents this voltage drop is a nonlinear large-signal function of collector current.

Both current crowding and high-level injection contribute to the nonlinearity, making the effective resistance of the base bulk region dependent on base current. This is complicated further by the nonlinear large-signal relation between base and collector currents (beta nonlinearity). The base resistance is important in a number of other contexts, whose diversity is apparent in the different ways of measuring and modelling the resistance. For small-signal applications where noise or frequency response is the major issue, a linear resistance r_b is used as a model. The measured value of this resistance then depends on operating point.

This small-signal formulation is not very useful for predicting log-conformance error in general, but a useful variant does arise in certain gain-control circuits, as will be shown.

The basic concept of current crowding was explained graphically by Gray et al. in the SEEC series [7.10]. Their approach was to model the base resistance as a truly linear, "extrinsic" resistance (outside of the active junction area) in series with a distributed, nonlinear "intrinsic" resistance that decreases with increasing current. At about the same time, Hauser (no relation to the author) developed a quantitative two-dimensional analysis [7.11]. There have been additional refinements in the theoretical modelling of nonlinear base resistance, some of them reviewed by Getreu [7.4]. Rather than predicting it theoretically, the present work is concerned with measuring the actual effect and developing a correction strategy that does not require a-priori knowledge of the base resistance or its current dependence.

7.2.5 Other Resistive Parasitics

Inevitable series resistances appearing in the emitters and collectors of practical transistors, as well as incidental resistance in the interconnections, will cause additional voltage drops. Because the emitter is much more heavily doped than the base in modern (high-beta) transistors, the resistivity of the bulk emitter region tends to be very small, and the series resistance in the emitter lead is due primarily to the contact interface [7.4].

This is usually on the order of an ohm, and substantially independent of current (compared with the base resistance).

The collector is a high-resistivity region, and the series resistance separating the collector contact from the active collector-base junction will depend on transistor construction, especially the presence or absence of a low-resistance diffusion ("collector plug") shorting the epitaxial collector material. Moreover, the collector resistance varies with current. However, this resistance is far less influential in normal operation than is the base resistance, because voltage drops in series with the collector have a very small effect on base-emitter voltage. The constant of proportionality is the base-width modulation factor η (typically between 10^{-3} and 10^{-4}), as long as the transistor remains out of saturation. If the collector current remains constant, an incremental change ΔV in collector-to-base voltage causes a change $-\eta\Delta V$ in base-emitter voltage. The negative sign arises because an increase in V_{CB} requires a reduction in V_{BE} to maintain a constant collector current.

The voltage drop resulting from series resistance in the collector may be represented by a modified series emitter resistance. If an otherwise ideal transistor has common-base current gain α, series emitter resistance r_e' and series collector resistance r_c (both assumed linear for the moment), and the far side of r_c is connected to a voltage source (see Figure 7.1), then to first order the combined effect of r_c and r_e' is equivalent to a resistor r_{ec} in the emitter alone. The value is

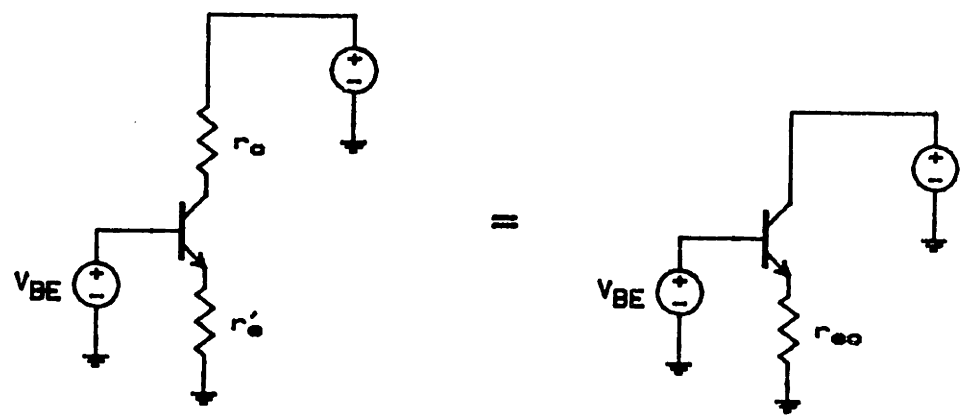


Figure 7.1 -- Consolidating parasitic emitter and collector resistances

$$r_{ec} = r'_e + \alpha\eta r_c \tag{7.4}$$

With typical r_c values in the hundreds of ohms for small monolithic transistors and η less than 10^{-3}, the effect of internal collector resistance should be small in terms of resistive voltage drops in V_{BE}. Potentially more important is the tendency for the collector-base junction to become forward-biased (saturation) if a substantial DC drop appears across this resistance.

Further resistive effects will occur because of resistance in the contact interfaces at the surface of the transistor structure, and because of the finite conductivity of the metallization that interconnects the different parts of the circuit. Some of these resistive parasitics, specifically those other than interconnect resistance, will naturally scale as multiple devices are connected in parallel; the effect of such paralleling on interconnect resistance is a function of detailed layout topology.

All of these elements introduce excess voltage drops in series with the ideal (logarithmic) V_{BE}. Some of the drops are linear (ohmic) with current, and some are very small. However, they all contribute to distortion in circuits that depend on the accuracy of (7.3). The net effect of the various resistive parasitics, including base resistance, may be lumped into a net log-conformance error voltage appearing in series with the V_{BE} of (7.3). Significant aspects of this error voltage are its dependence on device geometry (e.g., on junction area scaling) and its departure from linear (ohmic) behavior, especially at high currents.

7.2.6 High-Level Injection

When the density of injected carriers in a region of a semiconductor becomes comparable to the dopant concentration, qualitative changes occur in the operation of the device [7.12]. Conductivity becomes current-dependent in the high-injection region, so that the current-to-voltage relation for a section of bulk material is no longer linear (ohmic). This accounts in part for the current dependence of the parasitic base and (especially) collector resistances. The classic work of Hauser [7.11] on base current crowding did not allow for high-level injection, although others (such as Roulston et al. [7.13]) have since incorporated it. High-level injection causes an additional reduction in effective base resistance at high collector currents.

Significant other effects arise under high-level injection. The electrostatics of pn-junction depletion regions are modified [7.14], for example. In the bipolar transistor, high-level injection in the base region brings about a reduction in current gain and, most importantly, a fundamental change in the I_C-V_{BE} relation. This asymptotically approaches the form [7.12]:

$$V_{BE} \sim \frac{2kT}{q} \ln\left(\frac{I_C}{I'_S}\right) \tag{7.5}$$

This may be difficult to observe in practice because of the series voltage drop in the base resistance.

High-level, high-frequency signals can introduce still further compications. At frequencies above about f_T/β, the AC

base current rises with frequency; high-level injection and crowding, with attendant effects on DC behavior, amy be introduced by the presence of AC signals [7.10, 7.15].

7.2.7 Special Transistor Designs

The phenomena detailed in sections 7.2.4 through 7.2.6 all relate to deviations from the ideal I_C-V_{BE} relation of (7.3) at high collector currents, or more properly at high current densities (since the critical current levels can always be raised by paralleling devices). For small-geometry transistors in a typical analog IC process, log-conformance errors large enough to cause problems are observed at collector currents on the order of 100 μA. Yet this is not a fundamental limit; the upper range of collector current can be extended considerably further. The fact is that neither the processing nor (especially) the layout of "standard" analog transistors is optimized to accommodate high current densities with good log conformance.

Chapter 8 of this report treats circuit approaches to circumventing log-conformance error in variable-gain configurations. Not to be overlooked is the possibility of reducing the error through judicious design of the key transistors. Parasitic resistances may of course be reduced at will by effectively paralleling multiple transistors; a more meaningful objective is to minimize the resistances for a given total device area. Simple modifications (such as extra base contacts) to a conventional transistor geometry can be helpful and cost very little in terms of area and layout anxiety.

Some designers have gone far to reduce resistive parasitics, especially in the context of low-noise amplification of low-resistance signal sources. Willemsen and Bel [7.16], for example, described a transistor structure that resulted from an all-out effort to minimize series base and interconnect resistances. This and similar devices that have been reported are large-area transistors intended for high total collector current; they are not designed specifically for logarithmic conformance, but for low thermal noise.

7.3 Influence of Log-Conformance Error in a General Translinear Network

Having discussed some practical sources of deviation from (7.3), the ideal logarithmic law, we now consider some implications for translinear circuits that exploit (7.3). This section is, of necessity, rather broad and basic. Detailed analysis of the input-output error depends on the particular translinear configuration. Section 7.4 deals with the particular case of a four-V_{BE} translinear variable-gain circuit.

7.3.1 Currents and Current Densities

For a given transistor process, the errors in V_{BE} associated with resistive parasitics and high-level injection are all functions of transistor geometry and terminal currents. Actually, the desired logarithmic voltage and (to a large extent) the error terms are dependent on current density in the vicinity of the base-emitter junction. If the effective device area is doubled by paralleling two identical transistors, the effect is identical to

halving the currents in a single device. The I_S parameter in (7.3) tends to scale with base-emitter junction area, as do the current sensitivities of the various LCE effects, even in the regime where their voltage-current relation is nonlinear.

In actuality this tendency is not exact for arbitrary scalings of junction area, because of three-dimensional effects in the transistor. It becomes exact when "area scaling" is taken to mean integer ratios of base-emitter junction area A_E, obtained by paralleling multiple identical transistors or identical base-emitter junctions in a single transistor structure (the typical monolithic procedure). This will generally be assumed from now on.

To allow for the equivalence of scaling junction areas and scaling currents, it is convenient to work in terms of current densities rather than currents. With the equivalence J=I/A, it is understood that a ratio of two current densities (J) may be realized by ratioing the currents or junction areas or both. Moreover, the true base-emitter junction current density, besides being spatially dependent over the extent of the junction, is not very useful as it incorporates current components that do not appear at the collector (see 7.2.3). For a precise treatment of logarithmic V_{BE}-versus-I_C and log-conformance effects, this report uses a formalized "current density" variable defined by

$$J_C = I_C/A_E \tag{7.6}$$

where I_C is collector current in the corresponding transistor and A_E is its base-emitter junction area. This "J_C" is used

mainly in ratios. It does not correspond strictly to any physical current density, but embodies the desired dependence on the particular current and area of interest.

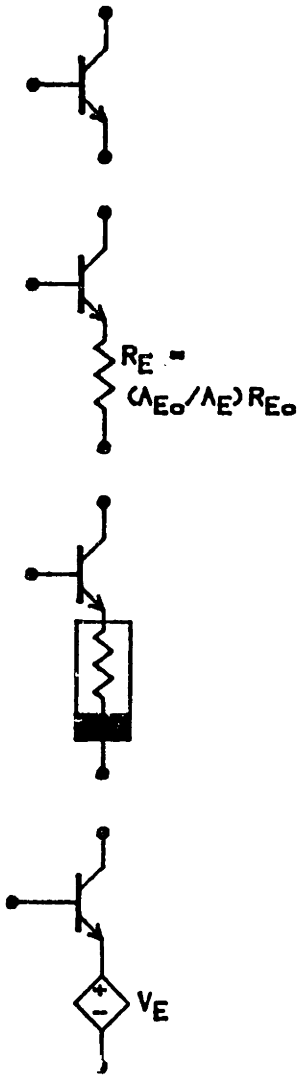
Cast in this more general "current density" form, (7.3) becomes

$$V_{BE} = \frac{kT}{q} \ln\left(\frac{J_C}{J_S}\right) \tag{7.7}$$

where J_S is now a saturation current density, no longer dependent on the effective base-emitter junction area. The emission coefficient "n" is taken to be unity for simplicity. This is the idealized relation, lacking log-conformance error terms. It must be emphasized that J_C is not itself a fundamental signal variable, but rather a shorthand for the ratio of I_C, a signal variable, with A_E, a design parameter. Usually it is not the absolute magnitude of A_E that is significant, but the ratio of A_E between different transistors in a circuit.

7.3.2 Models of LCE in a Single Transistor

Figure 7.2 illustrates successive levels of generality in modelling the I_C-V_{BE} relation of a forward-active transistor. In Figure 7.2(a), for reference, is an idealized device with perfect log conformance. In Figure 7.2(b) a linear (ohmic) resistance is introduced in the emitter. To the extent that beta is independent of collector current, this may accurately represent the net effect of linear resistances present in the emitter, base and even (see section 7.2.5) collector leads.



(a) IDEAL BJT. $V_{BE} = \frac{kT}{q}\ln\left\{\frac{I_C}{\Lambda_E J_S}\right\}$

(b) IDEAL BJT PLUS LINEAR RESISTANCE IN EMITTER. CONSTANT BETA.

$$V_{BE} = \frac{kT}{q}\ln\left\{\frac{I_C}{\Lambda_E J_S}\right\} + (\Lambda_{Eo}/\Lambda_E) I_E R_{Eo}$$

(c) IDEAL BJT PLUS NONLINEAR RESISTANCE IN EMITTER. CONSTANT BETA.

$$V_{BE} = \frac{kT}{q}\ln\left\{\frac{I_C}{\Lambda_E J_S}\right\} + V_E\left\{\frac{I_E}{\Lambda_E}\right\}$$

(d) ARBITRARY I_C-DEPENDENT LOG-CONFORMANCE ERROR VOLTAGE ADDED TO V_{BE}.

$$V_{BE} = \frac{kT}{q}\ln\left\{\frac{I_C}{\Lambda_E J_S}\right\} + V_E\left\{\frac{I_C}{\Lambda_E}\right\}$$

THE QUANTITY (I_C/Λ_E) CAN BE EXPRESSED AS A CURRENT DENSITY, J_C

Figure 7.2 --
Models for LCE in a transistor.

The model of Figure 7.2(b) has been used extensively in past analyses. The resistance R_E was taken to represent the effect of any (presumed linear) resistance in the base circuit, divided by $(\beta+1)$, along with any additional series resistance in the emitter lead. To distinguish it from the "dynamic emitter resistance" (α/g_m) of the transistor, R_E has usually been called "ohmic" emitter resistance. The usage implies that R_E is truly linear (obeys Ohm's Law). This is a legitimate approximation at low collector currents.

Over a sufficiently wide range of collector current, inevitable nonlinearity develops in both the I_C-I_B relation and the resistance of the base region. This is provided for in the model of Figure 7.2(c), through an equivalent emitter resistor with arbitrary current-controlled nonlinearity. High-level injection in the base may also be accounted for in this way. The implied dependence of log-conformance error voltage on emitter current, although in principle quite general, may be inconvenient.

A completely arbitrary additive error voltage in V_{BE} may be accommodated by simply adding a controlled voltage source, as shown in Figure 7.2(d). The voltage source V_E is shown dependent on collector current with an arbitrary functional relationship. In principle this format is interchangeable with the model of Figure 7.2(c), in that arbitrary additive current-dependent error in V_{BE} can be represented either way. However, Figure 7.2(d) uses the same independent variable (collector current) in both the desired (logarithmic) and error components of V_{BE}; it avoids

the implicit sensitivity to beta that accompanies the use of emitter current. For this reason it may be attractive to use Figure 7.2(d) even if the LCE is predominantly ohmic.

Because of the complex of process-, current-, and temperature-dependent nonlinearities that contribute to log-conformance error, it is dangerous to make conclusive a priori statements about its functional form. The approach taken here is to analyze the consequences of linear LCE and then to examine the departure from linearity in some practical transistors. There is no reason why explicit LCE could not be measured (other than experimental difficulty) and fitted, say, to a power-series form; such a fit could then be applied to the general results of this chapter and the next, yielding numerical values. This is a potential topic for further study.

At sufficiently low currents, below the levels where current crowding and high-level injection are significant, a linear (ohmic) LCE is typically observed. In this case, Figure 7.2(d) applies with a linear $V_e(\cdot)$ function:

$$V_E(J_C) = A_o R_o J_C \tag{7.8}$$

where A_o is a normalizing area and R_o is an equivalent resistance that would exhibit the same voltage at a current $A_o J_C$. The form of (7.8) is useful even if V_E is not a linear function of J_C; in that case, R_o may be regarded as a current-density-dependent parameter.

The general expression for V_{BE}, assuming arbitrary deviation from (7.7), is

$$V_{BE} = \frac{kT}{q} \ln\left(\frac{J_C}{J_S}\right) + V_E(J_C) \tag{7.9}$$

This expression tacitly assumes that the error term V_E depends not on I_C and A_E separately but on their ratio J_C. In order for the assumption to be valid, all sources of excess voltage drop in V_{BE} must be duplicated in any junction-area replication that accomplishes A_E-scaling. This may not be true in the case of interconnection resistance; deliberate attention to layout is necessary, otherwise the V_E term will contain a resistive component that cannot be assumed to scale with area.

7.3.3 The Basic Translinear Network

An elementary translinear circuit is a loop of an even number of forward-biased pn junctions. There is no constraint on the orientation (polarity) of each junction, except that half of them must be oriented in the opposite direction as the other half (clockwise versus counterclockwise voltage drops). The junctions are at the same temperature. Associated with the forward voltage drop of each junction is a current density; these are denoted by V_n and J_n respectively, where n is the number of the junction. Junction voltage drops and current densities are, ideally, related by (7.7) for each junction, with a possible prefactor "n" if the emission coefficient is not unity. Saturation current density "J_S" in (7.7) is assumed the same for each junction.

Figure 7.3 shows a typical case, where the junctions are the base-emitter junctions of bipolar transistors and the associated current densities are as defined in (7.6). Although all of the "clockwise" junctions are shown on one side of the loop and those "counterclockwise" on the other, this is by no means a necessary orientation; the two polarities are often interspersed in practice. The associated circuitry that forward-biases the junctions, drives the currents, etc., is not shown; Figure 7.3 is only the essential core of the circuit. The topology imposes, through Kirchhoff's Voltage Law,

$$\sum_{cw} V_n - \sum_{ccw} V_n = 0 \tag{7.10}$$

If (7.7) is satisfied by each transistor, the sums of logarithms of J values convert to products of J values, leaving the classic translinear result

$$\frac{\prod_{cw} J_n}{\prod_{ccw} J_n} = 1 \tag{7.11}$$

With various J values, or linear combinations of them, used as inputs or outputs,* (7.11) gives rise to a large and striking variety of linear and nonlinear function circuits. These are described in more detail in chapter 6. The important result is

*Actually the corresponding currents serve as inputs or outputs; current densities figure in the circuit equation.

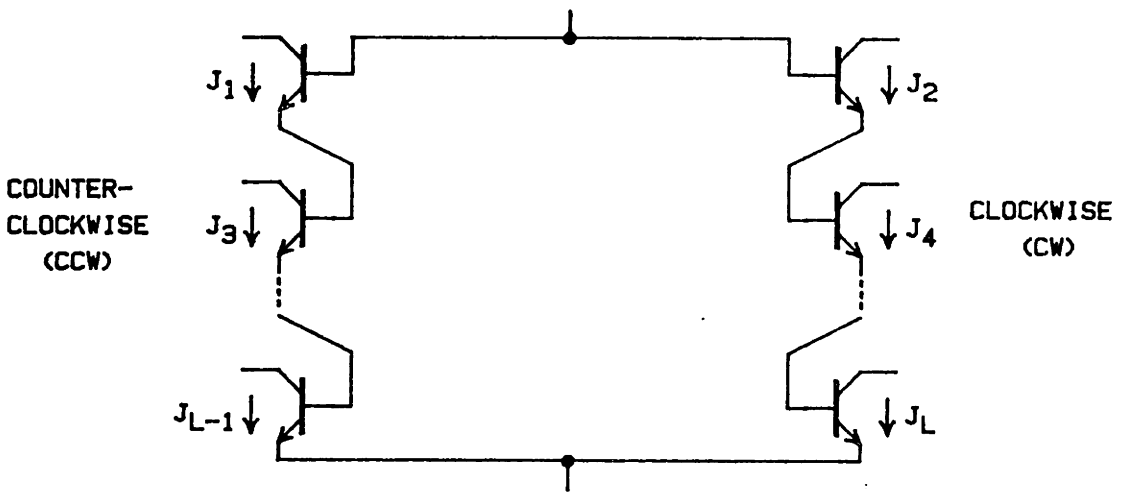


Figure 7.3 -- Translinear network made up of bipolar transistors.

BIAS CIRCUITRY, INPUT/OUTPUT CONNECTIONS, ETC, NOT SHOWN

(7.11), which is fundamentally temperature-insensitive and provides precise large-signal algebraic relations between the different current densities. All of this depends on exact conformance to the ideal relation of (7.7).

7.3.4 Translinear Networks in the Presence of Log-Conformance Error

Arbitrary current-density-dependent log-conformance error yields a relation like (7.9) for each junction in Figure 7.3. Schematically, the situation is then equivalent to a loop of ideal junctions in series with an excess loop voltage, V_X. Voltage V_X represents the algebraic sum of the V_E terms in all of the junctions; by Kirchhoff's Voltage Law these V_E terms can be grouped together for the purpose of analysis.

The situation is shown in Figure 7.4, which now satisfies

$$\frac{\prod_{cw} J_n}{\prod_{ccw} J_n} = \exp\left(\frac{qV_X}{kT}\right) \tag{7.12}$$

where

$$V_X = \sum_{ccw} V_E(J_n) - \sum_{cw} V_E(J_n) \tag{7.13}$$

Equation (7.12), of course, reduces to (7.11) in the absence of log-conformance error (V_X=0).

Assessing the impact of log-conformance error on the circuit requires specifying the detailed dependence of V_X on

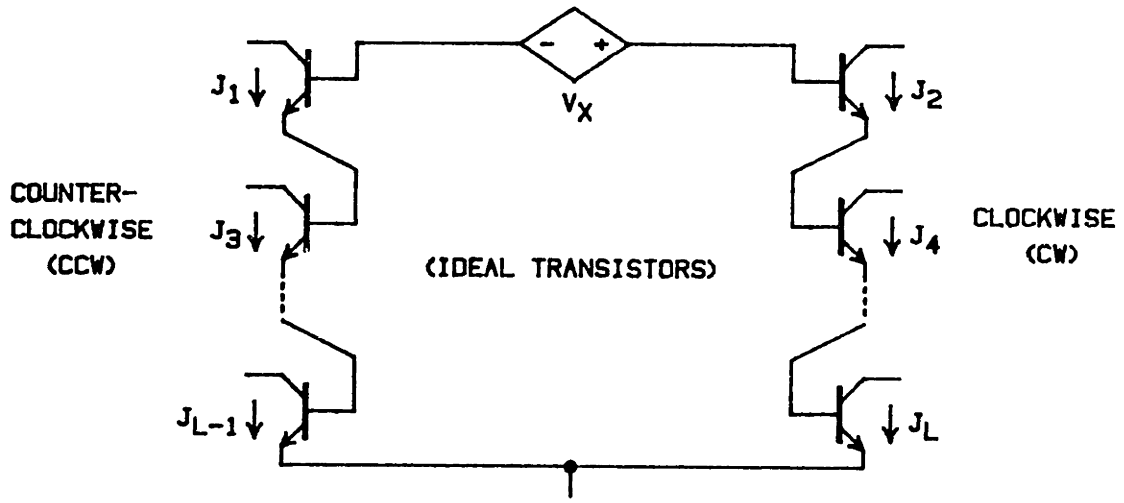


Figure 7.4 -- Translinear network with a net excess loop voltage V_X.

the signal variables (J values). In general this will yield an implicit, transcendental relation among the J values. For example, if the error voltage is a linear function of J in each device, so that (7.8) applies, then (7.12) and (7.13) yield

$$\frac{J_n}{J_n} = \exp \frac{qA_o R_o}{kT} \quad J_n - \quad J_n \tag{7.14}$$

Note that an excess loop voltage V_X that is proportional to absolute temperature (PTAT) will cancel the temperature dependence in the righthand side of (7.12). This yields a relation similar to (7.11) but with a dimensionless, temperature-stable factor on the right-hand side.

Solving an expression such as (7.14) for one of the J values in terms of the others is not usually possible by analytical means. Numerical solution is available, however; the result of log-conformance error is generally a smooth deviation from the simpler solution of (7.11). Again, it is necessary to be more specific regarding the particular translinear configuration. The discussion therefore focuses on a class of circuits useful in gain-control applications.

7.4 Influence of Log-Conformance Error in a Four-Junction Variable-Gain Circuit

Figure 7.5 shows the translinear configuration to be studied. This is the classic two-quadrant multiplier of chapter 6, and was essentially the first translinear circuit to be published [7.17]. Again, only the central core of the circuit is shown. This is a special case of Figure 7.3 -- four base-emitter junctions in a loop. The analysis will use the collector currents of all four transistors as variables; this departs slightly from Gilbert's original formulation, in which the emitter currents of the lefthand pair served as inputs. The use of collector currents throughout permits adherence to the translinear formalism without introducing alpha factors. This convention is followed also in the measurements of section 7.5; the experimental setup forces the desired collector currents in the input pair.

7.4.1 Ideal Circuit Behavior

Associated with the four collector currents I_1 - I_4 in Figure 7.5 are four current densities J_1 - J_4, where each is the ratio of collector current to base-emitter junction area, as in (7.6). It is assumed that each pair of currents is balanced; i.e., additional circuitry not shown imposes

$$I_1 + I_2 = 2I_A \tag{7.15a}$$

$$I_3 + I_4 = 2I_B \tag{7.15b}$$

for some fixed values I_A, I_B.

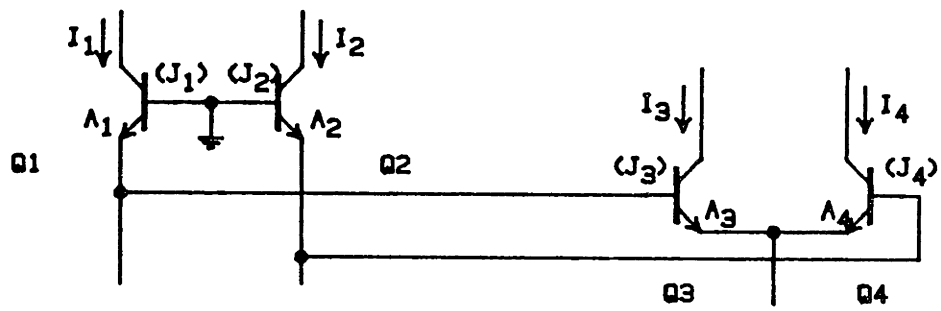


Figure 7.5 -- Basic test circuit

ADDITIONAL CIRCUITRY IMPOSES BALANCED COLLECTOR CURRENTS:

$I_1+I_2=2I_A$. $I_3+I_4=2I_B$

If the transistors have perfect log conformance, then from (7.11) the four current densities will satisfy the translinear-circuit constraint

$$\frac{J_2 J_4}{J_1 J_3} = 1 \tag{7.16}$$

If the areas are pairwise matched, so that $A_1 = A_2 = A_A$ and $A_3 = A_4 = A_B$, then (7.16) is equivalent to

$$\frac{I_2 I_4}{I_1 I_3} = 1 \tag{7.17}$$

This, in combination with the balanced-drive condition of (7.15), yields the basic variable-gain result

$$I_4 = \left(\frac{I_B}{I_A}\right) I_1 \tag{7.18}$$

or, in differential form,

$$(I_4 - I_3) = \left(\frac{I_B}{I_A}\right)(I_1 - I_2) \tag{7.19}$$

The gain from the input $(I_1 - I_2)$ to the output $(I_4 - I_3)$ is set by the ratio of the two fixed currents I_B and I_A. Now consider the consequences of a linear log-conformance error.

7.4.2 Effect of Linear LCE

If the V_{BE} of each transistor contains an error voltage proportional to current density, as in the simple model of resistive parasitics, then (7.16) becomes

$$\frac{J_2 J_4}{J_1 J_3} = \exp\left[\frac{qA_o R_o}{kT}\,(J_1+J_3-J_2-J_4)\right] \tag{7.20}$$

--a special case of (7.14). This again does not exactly correspond to the emitter-resistor model of Figure 7.2(b), since the error depends on collector current. Now assuming $A_1 = A_2 = A_A$ and $A_3 = A_4 = A_B$, and using the balanced-current conditions of (7.15), the result in terms of currents is

$$\frac{(I_1-2I_A)I_4}{I_1(I_4-2I_B)} = \exp\left\{\frac{2qR_o}{kT}\left[\frac{A_o}{A_A}\,(I_1-I_A) - \frac{A_o}{A_B}\,(I_4-I_B)\right]\right\} \tag{7.21}$$

In the particular case where the pair currents are ratioed the same as the pair areas, so that $I_B/I_A = A_B/A_A$, equation (7.21) has the pleasing solution

$$I_4 = \left(\frac{I_B}{I_A}\right) I_1$$

so that the linear log-conformance error has no effect. Any other ratio (I_B/I_A) causes distortion. This result is well known.

The form and magnitude of the distortion were studied in the classic works of Gilbert [7.17, 7.18] and Sansen and Meyer [7.19].

The transcendental nature of (7.21) precludes a general closed-form solution for I_4 in terms of I_1. Instead, specialized approaches have been taken. Gilbert obtained small-signal (gain vs. operating point) [7.17] and approximate large-signal (assuming low distortion) [7.18] expressions. The basic form of the error is a symmetric, odd-order (S-shaped) distortion of the I_1-to-I_4 transfer curve. The distortion is approximately cubic, and its polarity depends on the relative sizes of I_B/I_A and A_B/A_A (when they are equal of course, the distortion vanishes).

Gilbert's approximate large-signal result, recast according to the nomenclature of this chapter, was

$$\left(\frac{I_4}{2I_B}\right) = \left(\frac{I_1}{2I_A}\right) + D_R \tag{7.22}$$

with

$$D_R = \frac{qA_oR_o}{2kT}\left(\frac{I_B}{A_B} - \frac{I_A}{A_A}\right)(X^3-X) \tag{7.23}$$

where $X = I_1/I_A - 1$. Note that the quantities in parentheses in (7.22) are normalized input and output variables that range from 0 to 1; similarly X is a normalized input variable between -1 and 1. The factor (X^3-X) has peak values of ± 0.385 at $X = \mp 0.577$ and vanishes at $X = 0$ and $X = \pm 1$. These results are taken directly from Gilbert [7.18].

Sansen and Meyer investigated the distortion in terms of third-order intermodulation. For low-frequency distortion, they

obtained numerical point-by-point solutions to (7.21) with a sinusoidal input current, and expanded the resulting waveform in a Fourier series. For the high-frequency regime (above about f_T/β), a Volterra series expansion was used, leading to a frequency-dependent distortion estimate.

A different, but related, problem is the effect of linear LCE in emitter-driven differential transistor pairs, the basic element of the "controlled cascode" multiplier [7.20]. Yen has concisely derived expressions for low-frequency [7.21] and high-frequency [7.22] distortion effects in this case.

Because Figure 7.5 suffers from distortion at all values of current gain but one, its utility as a variable-gain element suffers. The severity of the distortion depends on the average current densities I_A/A_A and I_B/A_B in the two pairs; the larger of these will account for most of the fractional error, as implied by (7.23). This sets a fundamental limit on both the absolute current levels and the gain range in Figure 7.5. Although not treated in detail here, other translinear circuits are subject to the same kind of errors and signal limits. An exception is the six-transistor cell commonly used for four-quadrant analog multiplication, but that configuration is poorly suited for variable-gain applications, as discussed in chapter 6.

7.4.3 Nonlinear LCE

Nonlinearities in the log-conformance error (e.g., current-dependent base resistance) will of course require a modification of (7.20). However, if the nonlinearity in $V_E(J)$ for each transistor is

sufficiently weak, a simplified model may be adequate. This exploits the fact that J_1 and J_2 will usually be of the same order, as will J_3 and J_4, even if the two pairs have widely different magnitudes of current density. Based on this assumption, the log-conformance error function $V_E(J)$ is regarded as locally linear (i.e., between each pair), but with different constants of proportionality for the two pairs. In effect, the equivalent resistance R_o of (7.8) and (7.20) is considered weakly current-density-dependent and thus possibly different for the two transistor pairs.

Equation (7.20) then becomes

$$\frac{J_2 J_4}{J_1 J_3} = \exp\left\{\frac{qA_o}{kT}\left[R_A(J_1-J_2) - R_B(J_4-J_3)\right]\right\} \tag{7.24}$$

This approximation is in fact borne out by the experimental results, with the restriction that the pair imbalance (e.g., J_1-J_2) not approach full scale. This is reasonable, as it insures that J_1 and J_2 will remain of the same order.

7.5 Measured Results

A series of measurements was performed to evaluate LCE-induced distortion from real transistors. The translinear variable-current-gain configuration of Figure 7.5 was set up using monolithic transistor arrays from an analog-IC fabrication line. The static transfer characteristics were then measured under various operating conditions. These characteristics illustrate quantitatively the consequences in a translinear circuit of log-conformance error, and

the degree to which the linear-error approximation of section 7.4.2 is valid.

The same monolithic transistors were used subsequently to test an LCE-correction strategy. The results are reported in chapter 8.

7.5.1 Experimental Arrangement

Figure 7.6 shows, in detail, the test circuit used for the measurements. Transistors Q_1 - Q_4 correspond to the devices in Figure 7.5, the basic test configuration. Associated circuitry in Figure 7.6 established bias conditions, imposed input and monitored output.

The base-emitter voltage offset in the Q_1 - Q_4 quad was adjusted to zero, and periodically monitored, in order to remove its influence from the circuit. Failure to do so yields a characteristic even-order distortion [7.17] that could interfere with the LCE measurements. Precision operational amplifiers maintained matched collector voltages in each of the two pairs (within 2 mV). The Q_1 - Q_2 pair was operated at a prescribed collector-base voltage, as described in the measurements, while the Q_3 - Q_4 pair operated at a collector-base voltage equal to the average V_{BE} of the first pair.

The quad was configured as a variable-gain current-normalizing amplifier. That is, the input pair operated at an adjustable total current ($2I_A$) while the output pair (Q_3,Q_4) operated at a fixed total current ($2I_B$) of 20.00 microamps. Input and output were monitored with series DC digital microammeters. The input was a differential current produced by splitting a single ($2I_A$) current source between

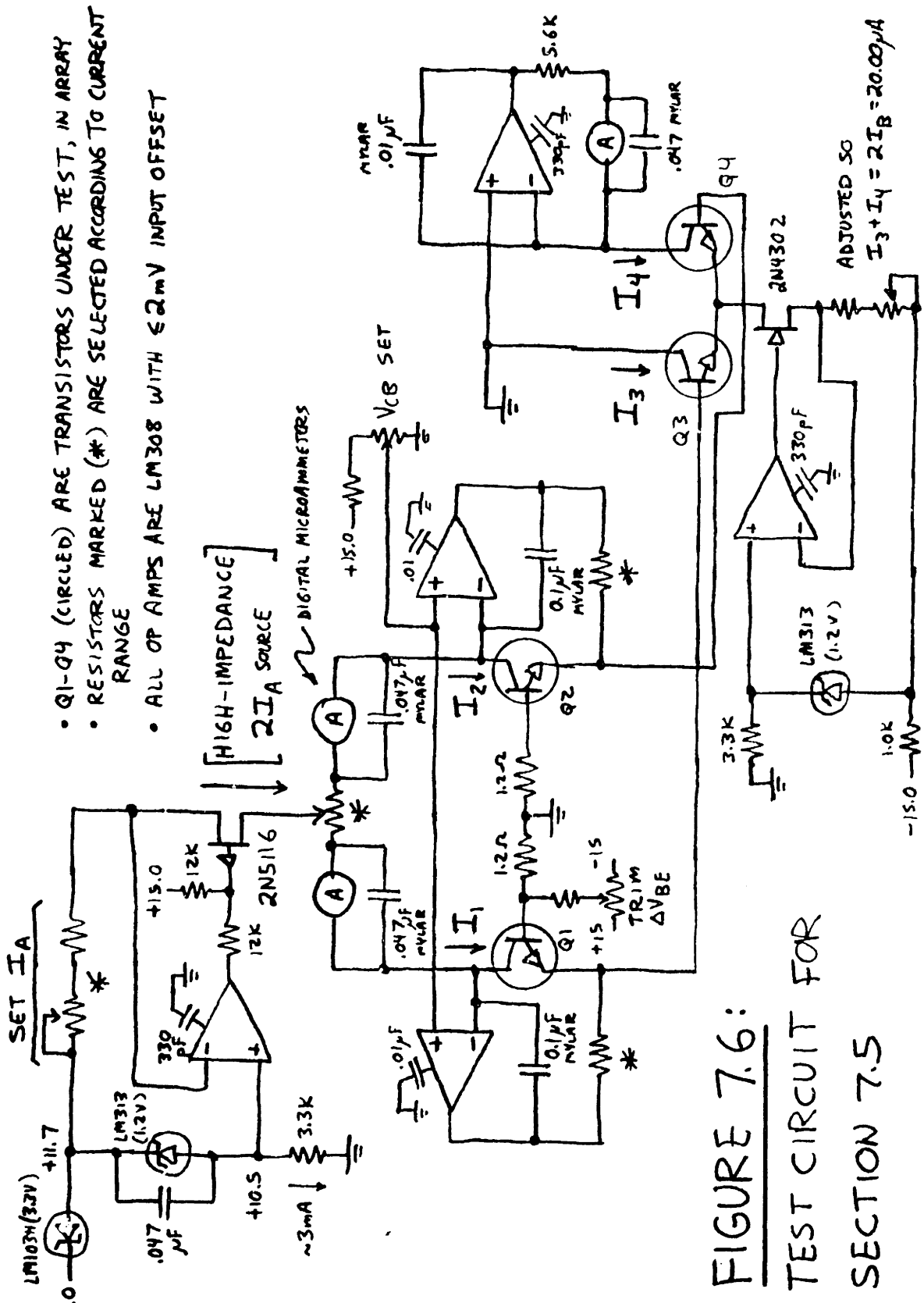


FIGURE 7.6: TEST CIRCUIT FOR SECTION 7.5

the collectors of Q_1 and Q_2, while the output was Q_4's collector current. Care was taken to regulate and measure collector currents at all times so that transistor current gain would not figure in the nominal transfer function.

Tests were performed using eight packaged monolithic transistor arrays provided by Analog Devices Semiconductor. The transistors in these arrays were typical small-area NPN devices used for contemporary analog circuits. They had 0.8-mil diameter circular emitters with a nominal emitter saturation current (I_S) of 1.2×10^{-15} amperes. Typical early voltage was 45 volts, and BV(CEO) breakdown was 35 volts. Typical DC beta was about 160 at 10 microamps collector current, 230 at 100 microamps, and 300 at 1000 microamps. V_{BE} match was better than 0.5 mV.

7.5.2 Signal-Path Linearity

The test circuit ideally performs a variable current gain function, with perfect input-output linearity, as described by (7.19). It is convenient to define normalized input and output variables:

$$X_i = \frac{I_1 - I_2}{2I_A} \tag{7.25a}$$

$$X_o = \frac{I_4 - I_3}{2I_B} \tag{7.25b}$$

These take on values in the range -1 to +1 for given values of the gain-programming currents I_A and I_B.

The ideal input-output relation, from (7.19), becomes

$$X_0 = X_i \tag{7.26}$$

for all values of I_A and I_B (in the measurements, I_B is fixed and I_A varies). The use of normalized variables allows easy comparison of input-output linearity at different current gain settings.

Figure 7.7 shows output versus normalized input for I_A = 10 μA, 250 μA, and 2000 μA. As in all of these measurements, I_B is 10 μA. Accuracy is ±0.1% or better. This introductory plot is from chip #5 and is typical of all devices tested.

For I_A = 10 μA, the input-output curve cannot be distinguished from a straight line. Since the four transistors have the same area, the ohmic LCE is expected to cancel in this case (I_A=I_B), as discussed in section 7.4.2. Indeed, no error was observed, to the accuracy of the test setup. For I_A = 250 μA, the peak error of 2.1% is barely visible. For I_A = 2000 μA, the peak error was 12.7%, now quite obvious. These percentages correspond to an output scale of ±100%. That is,

$$\% \text{ error} = (X_o - X_i) \times 100\% \tag{7.27}$$

Subsequent plots differ from Figure 7.7 in two aspects of presentation. First, to better illustrate <u>deviations</u> from the ideal straight-line transfer function, percentage <u>error</u> is plotted rather than output. Second, the symmetry between first and third quadrants

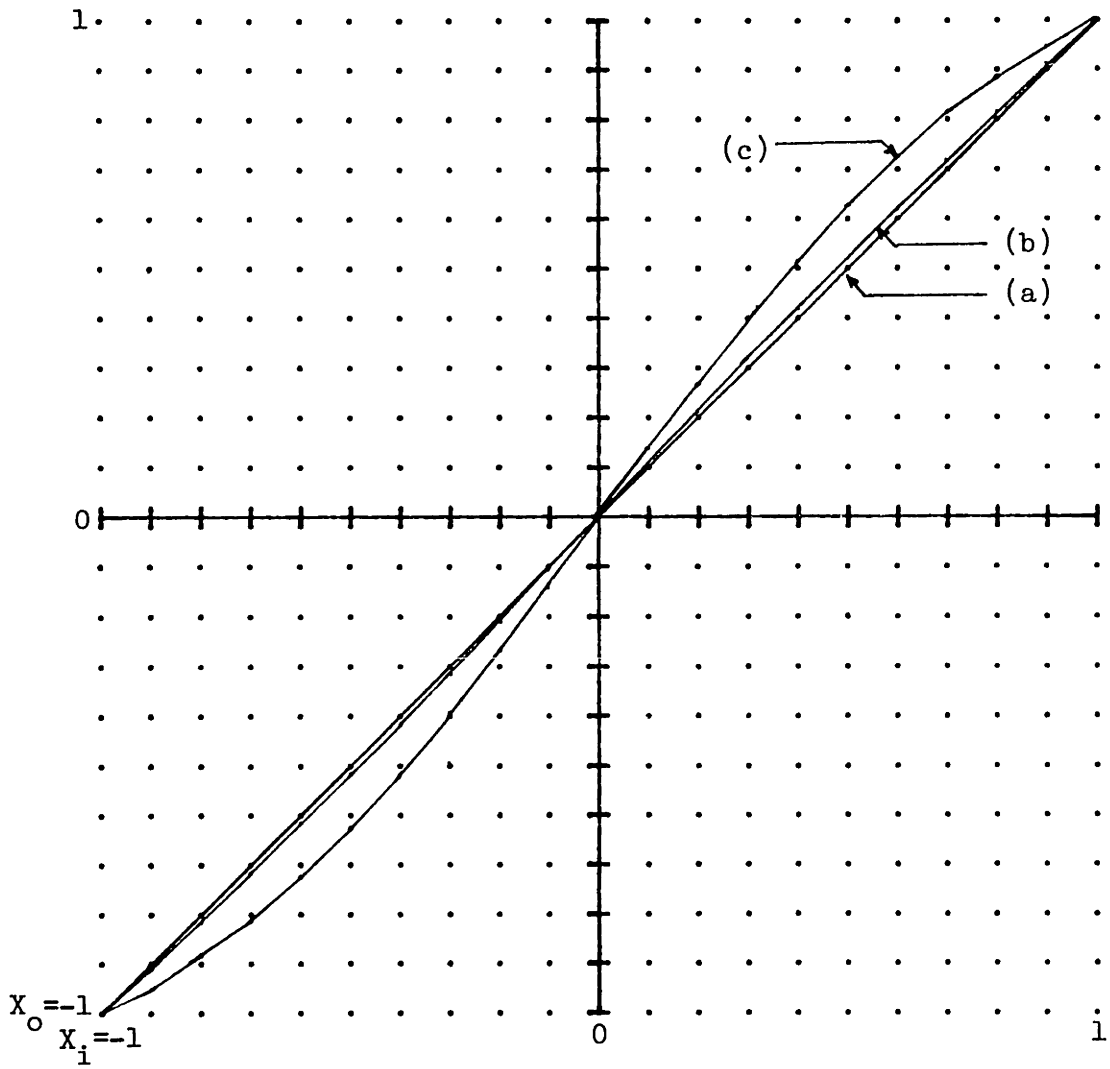


Figure 7.7 -- Output vs. normalized input for circuit of Figure 7.5.

Input-pair currents (I_A) are (a) 10uA, (b) 250uA, (c) 2000uA.
Output-pair current (I_B) is 10uA.

in Figure 7.7 is essentially perfect, and this was typical of all measurements. Therefore, only the first quadrant is presented hereafter.

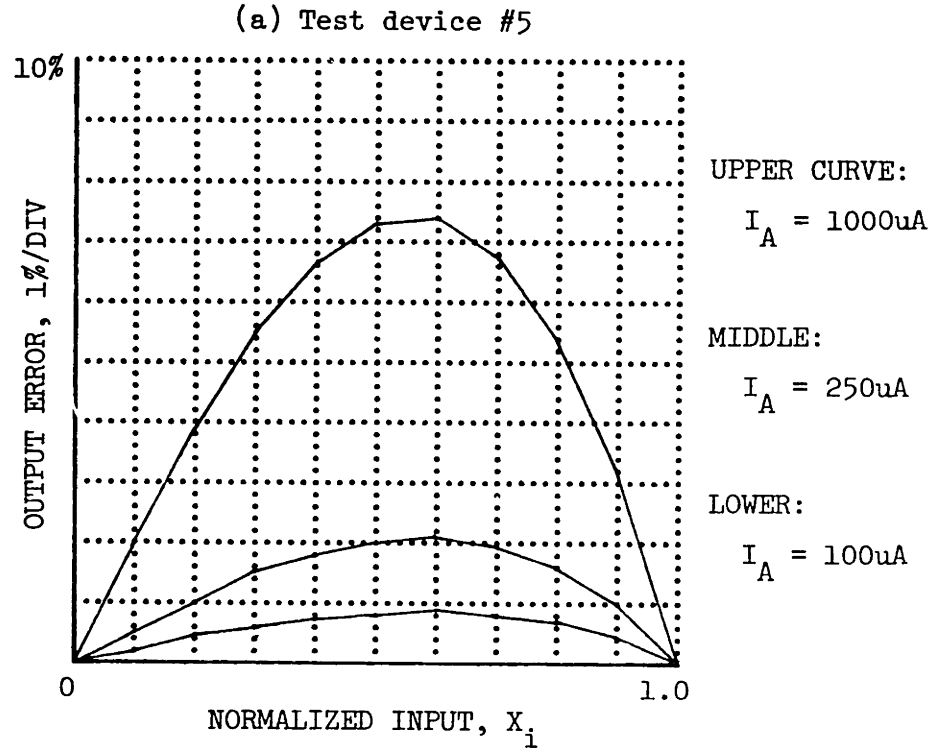
Figure 7.8 presents output error as a function of normalized input for two different test chips at I_A = 100 μA, 250 μA, and 1000 μA. Corresponding peak error is 0.9%, 2.1%, and 7.5% for the device in Figure 7.8(a); in Figure 7.8(b) it is 0.8%, 1.9%, and 6.5%. Collector-to-base voltage for the input pair was zero in these measurements.

The influence of collector-base bias is shown in Figure 7.9. The first plot, Figure 7.9(a), was obtained at a "moderate" I_A value of 250 μA. For the transistors used, this current would be considered close to the upper limit for reliable log conformance. The approximate ohmic resistance to which this error corresponds is treated in section 7.5.3.

Note that the error magnitude falls slightly as V_{CB} is increased. A likely mechanism for this is the increase in DC beta (due to basewidth modulation) that accompanies an increase in V_{CB}. Resistive parasitics in the base region will have less effect as beta rises.

A more pathological phenomenon occurs at higher collector currents. Not only are resistive errors in base and emitter larger, but the drop across the internal collector resistance may actually drive the transistors into saturation. In this case the external V_{CB} can have great influence on the circuit's behavior.

Figure 7.9(b), with a different vertical scale, illustrates this. With I_A = 2000 μA, the same test device as in Figure 7.9a



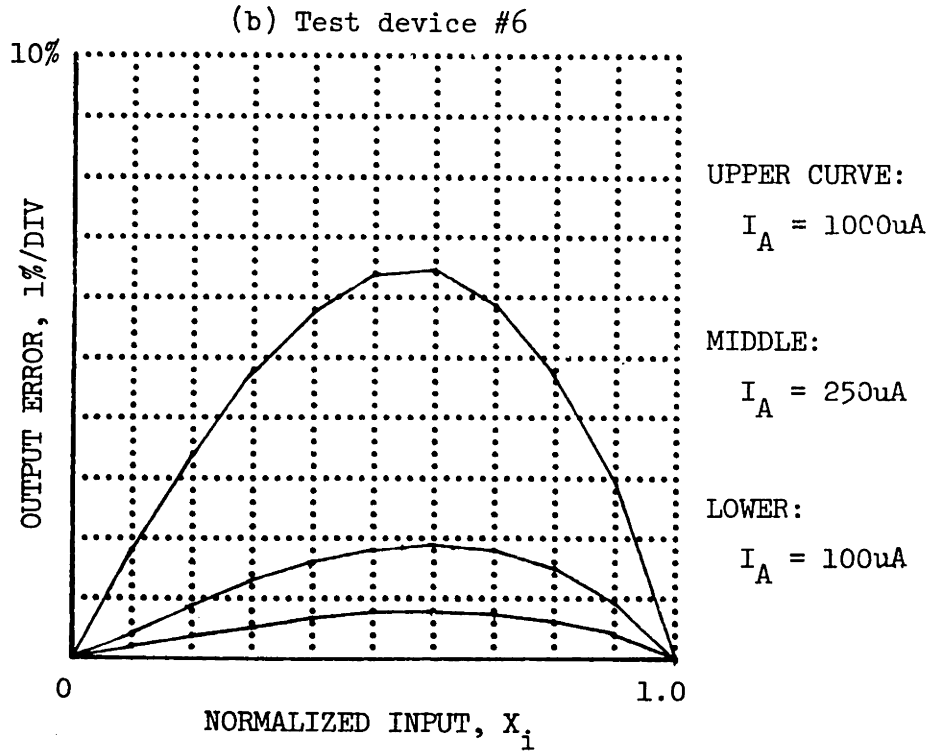
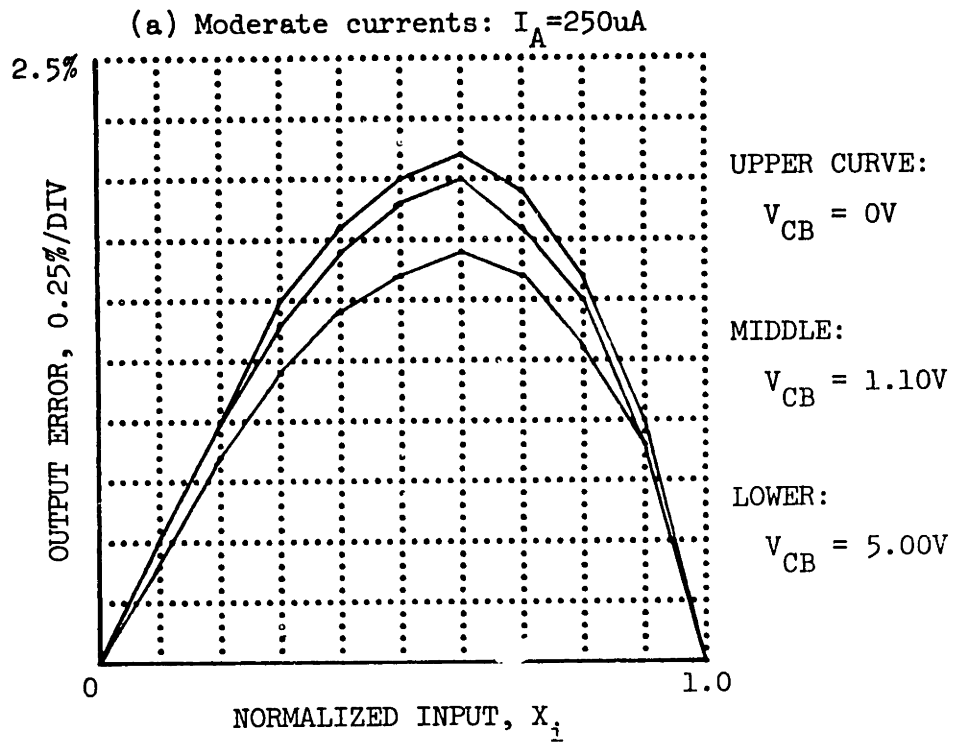


Figure 7.8 -- Measured input-output error for two transistor arrays.



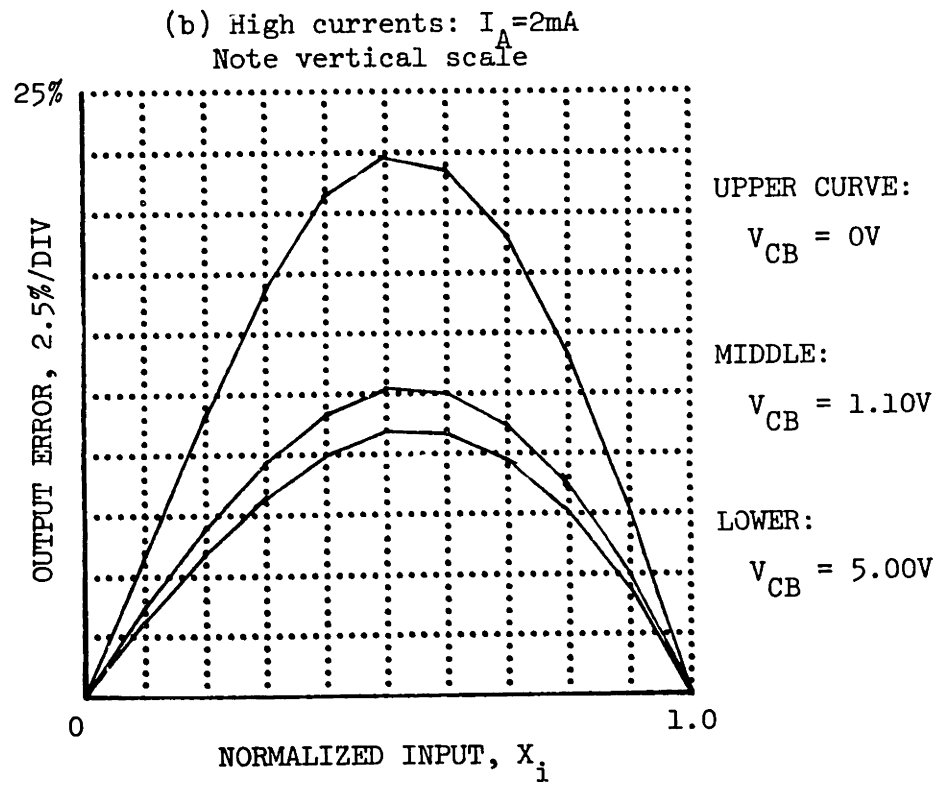


Figure 7.9 -- Measured input-output error versus input-pair V_{CB}.

shows a peak error of 22.4% when the input pair has $V_{CB} = 0$. The error falls to 12.7% for $V_{CB} = 1.10V$ and 10.9% for $V_{CB} = 5.00V$; the minor change between 1.10V and 5.00V suggests that the devices are out of saturation. Also, the value of X_i at the error peak has shifted slightly, from $X_i \cong 0.6$ in Figure 7.9a to $X_i \cong 0.5$. The shift implies a breakdown of the small-error estimate, equation (7.23). Moreover, the amplitude of the error peak increases less rapidly with current than predicted by the linear-LCE model.

7.5.3 Loop Error Voltage and Equivalent Resistance

We would like to extract an equivalent resistance, possibly current-dependent, that relates LCE voltage in the test transistors to collector current. Although the measurement of resistive parasitics in the base-emitter circuit has been treated extensively in the literature, the emphasis usually has been different. A dominant interest in noise, high-frequency or switching implications of base and emitter resistances has guided past work (for example, [7.23-7.27]). As a result, the measurements lack the large-signal LCE formulation sought here.

Rather than attempting to measure the transistors directly, an indirect approach will be taken here. Log-conformance error in the individual transistors will be inferred from the input-output error in the translinear current-gain circuit. First, the observed transfer functions will be related to an excess loop voltage V_X, as in section 7.3.4. An equivalent resistance then can be derived.

In the ideal case where the V_{BE} of each transistor is an exact logarithmic function of collector current, Figure 7.5 obeys

the ideal translinear result of (7.16). As discussed in section 7.3.4, nonzero log-conformance error terms in the four base-emitter voltages may be lumped into an error voltage V_X. Assuming matched areas, or the electrical equivalent (trimmed V_{BE} match, as in Figure 7.6), Figure 7.5 satisfies, from (7.12),

$$\frac{I_2 I_4}{I_1 I_3} = \exp\left(\frac{qV_X}{kT}\right) \tag{7.28}$$

In this implicit relation, V_X is also a function of I_1 through I_4; the dependence is not shown in (7.28). From (7.13), the exact form of V_X will depend on the LCE function $V_E(\cdot)$ of the transistors used. However, (7.28) permits us to calculate V_X from the current-transfer data, since all four I variables are known simultaneously. Thus,

$$\frac{qV_X}{kT} = \ln\left[\frac{I_2 I_4}{I_1 I_3}\right] \tag{7.29}$$

Note that the dimensionless variable qV_X/kT is actually calculated, rather than V_X, which would require specifying junction temperature. This derived result is shown as a function of the normalized current input X_i in Figure 7.10 and, in more detail, Figure 7.11.

These figures show the excess loop voltage to be approximately a linear function of input current at each value of I_A. This

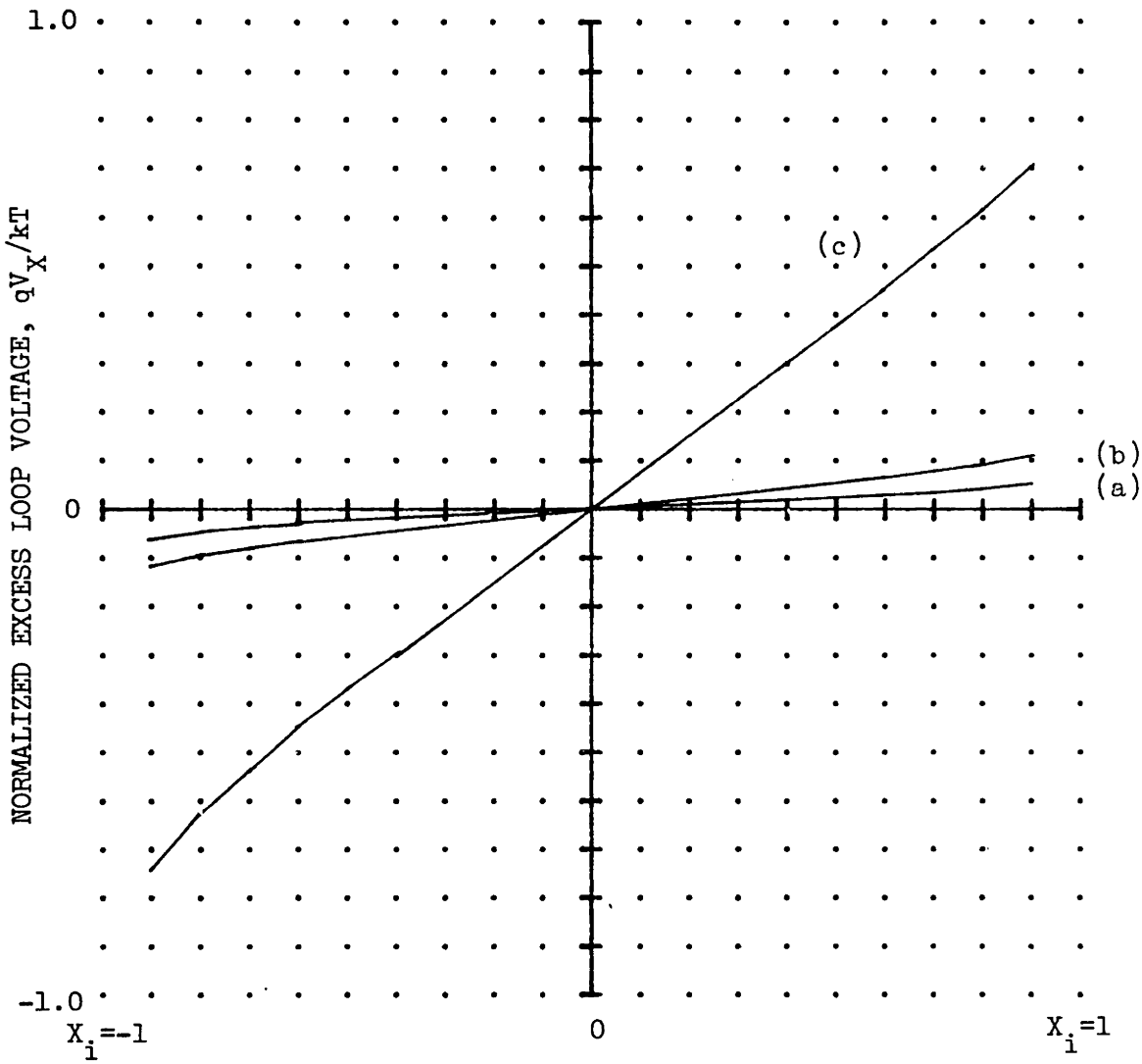


Figure 7.10 -- Derived excess loop voltage vs. normalized input.

Input-pair currents (I_A) are (a) 100uA, (b) 250uA, (c) 2000 uA.

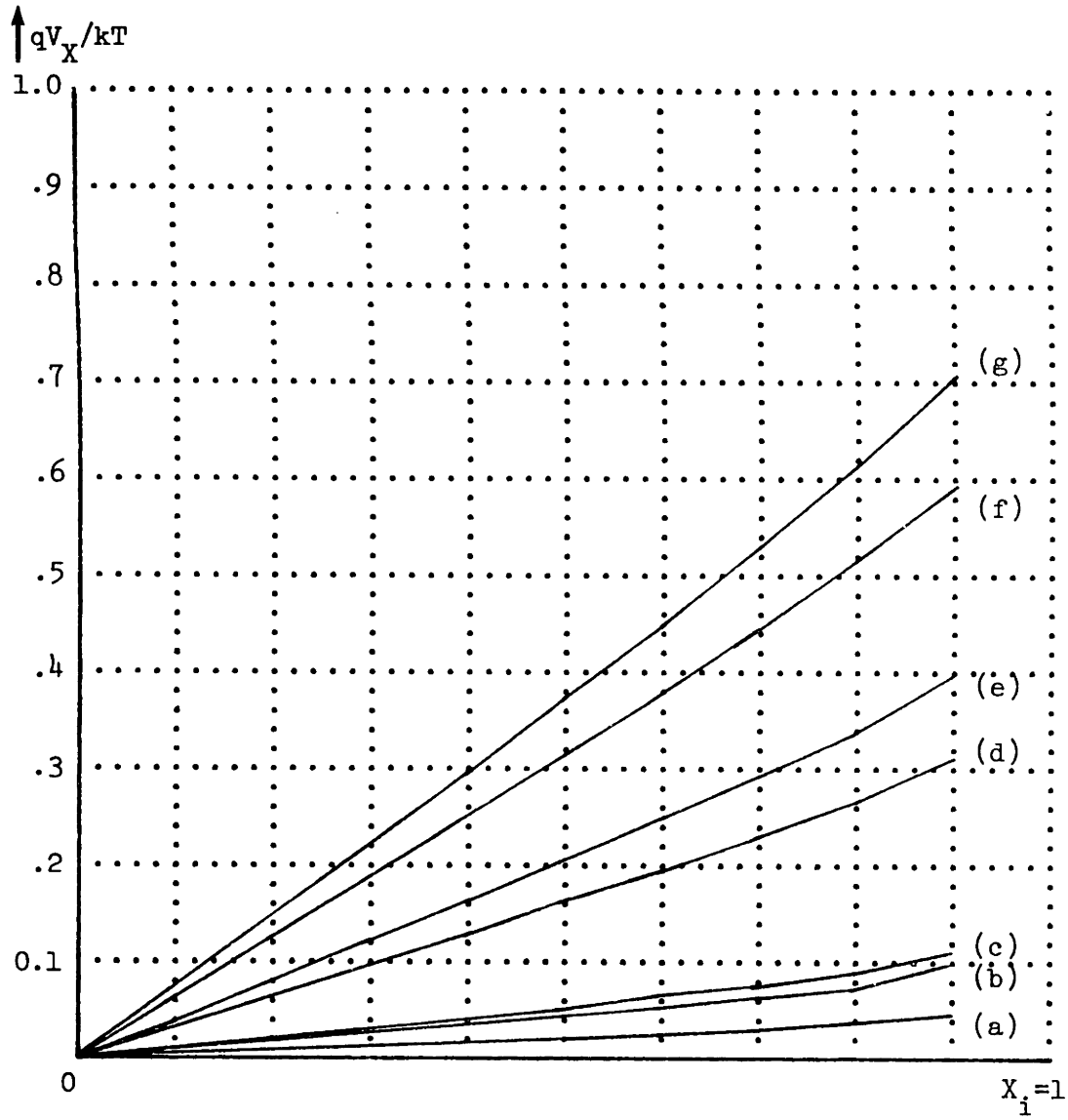


Figure 7.11 -- Excess loop voltage for positive input, with different values of I_A and input-pair V_{CB}.

(a)	I_A = 100uA,	V_{CB} = 0V
(b)	250uA,	5.00V
(c)	250uA,	0V
(d)	1000uA,	5.00V
(e)	1000uA,	0V
(f)	2000uA,	5.00V
(g)	2000uA,	1.10V

is to be expected, since V_X is the algebraic sum of "non-logarithmic" voltages in the base-emitter drops, and these voltages are roughly ohmic. If the log-conformance error were exactly ohmic, then (7.20) would apply, giving the form

$$V_X = R_o(I_1+I_3-I_2-I_4)$$

$$= 2R_o(I_A X_i - I_B X_o) \qquad (7.30)$$

For most of the data plotted, I_A is much greater than I_B, so a linear V_X-versus-X_i relation would follow.

This linear-LCE case is hypothetical. Rather than assuming it is true, let us turn this around and define an equivalent resistance.

$$R_{EQ} = V_X/2(I_A X_i - I_B X_o) \qquad (7.31)$$

R_{EQ} is not a physical resistance, but rather a voltage-current ratio that depends, in general, on signal and bias levels. It may be regarded as the equivalent (possibly current-dependent) resistance R_o that would fit the transistors measured to the model of (7.8). Similarly, within a factor of alpha, it represents an equivalent emitter resistance for a single transistor. The degree to which R_{EQ} is constant reflects the validity of the "ohmic" LCE model.

Figure 7.12 shows computed values of R_{EQ} from the data of Figures 7.10 and 7.11. In obtaining this result, it was necessary

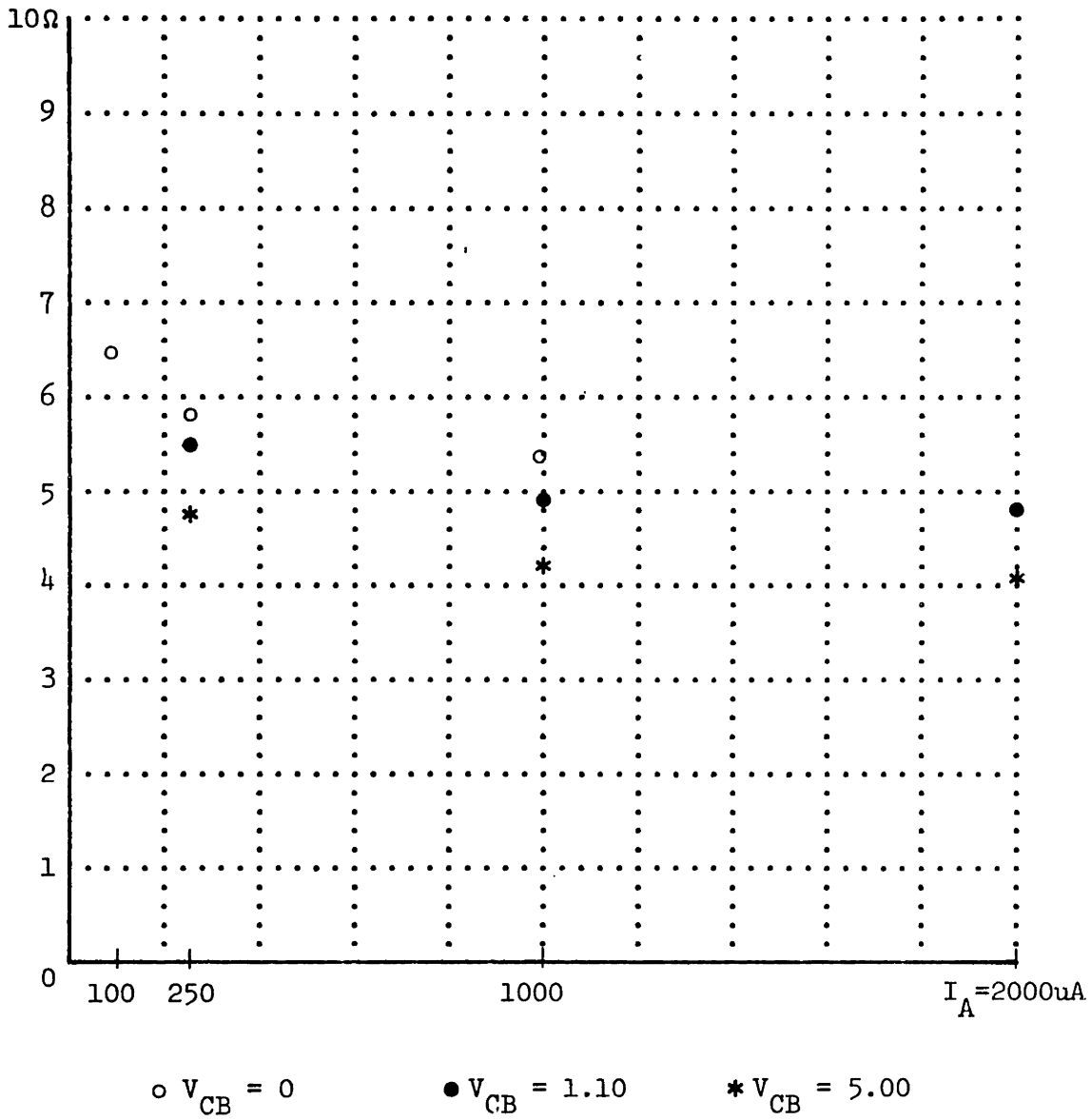


Figure 7.12 -- Dependence of the hypothetical resistance R_{EQ} on I_A and V_{CB}, for one particular monolithic transistor quad.

R_{EQ} values estimated by averaging equation (7.31) for the data sets pictured in Figure 7.11, and others.

to assign a nominal value for junction temperature. Since this parameter was not controlled in the measurements, a nominal 300 degrees Kelvin was assumed; the results are therefore approximate.* However, ambient temperature was constant to a few degrees during the measurements, and the currents used were low enough that self-heating was minimal. The packaging of the chips yielded a thermal resistance corresponding to about 0.2% increase in kT/q per milliamp of I_A.

The calculation yielding R_{EQ} is, unfortunately, somewhat ill-conditioned; small errors in the measured data yield significant errors in R_{EQ}. However, trends are apparent in Figure 7.12. We observe an equivalent resistance that falls somewhat at large values of I_A and V_{CB}. Part of this dependence is undoubtedly due to changes in beta, which increases with both V_{CB} and collector current. Within each curve, R_{EQ} is only a weak function of the input X_i except at the endpoints. This supports the model of section 7.4.3: weakly current-dependent resistance relating $V_E(J)$ and J. The resistance changes with I_A but is approximately constant over the input range for a given I_A.

---

*"A little inaccuracy sometimes saves tons of explanation."
--Hector Hugh Munro.

CHAPTER 8

CIRCUIT TECHNIQUES FOR CORRECTING LOG-CONFORMANCE ERROR

8.1 Introduction

The previous chapter demonstrated the magnitudes of signal-path distortion arising from imperfect log conformance in practical translinear circuits. This traditionally has been a limitation in electronic gain control using such circuits. Designers have been forced to limit current levels in the crucial transistors, which in turn restricts signal swing and gain range.

This chapter examines a new circuit technique for mitigating log-conformance error (LCE) effects in bipolar-transistor circuits, including translinear variable-gain elements.* The technique has various practical implementations, and it may be incorporated into existing circuits. In variable gain circuits it affords a significant improvement in signal-path linearity over a wide gain range, as demonstrated by experimental results.

Notable advantages over past correction techniques are the lack of need for a-priori knowledge of resistive parasitics, and the ability to correct automatically for nonlinear as well as ohmic LCE.

---

*The original ideas that led to the subject of this chapter were developed by the author while at Tektronix, Incorporated, in the summer of 1978. Tektronix patent claims apply to some of the material in this chapter, and their permission to disclose is appreciated.

8.2 Other Approaches

The issue of log conformance in bipolar-transistor analog circuits is not new. A number of authors have demonstrated circuit techniques that correct for ohmic errors in the V_{BE}-to-I_C relation, extending the accuracy or current range of a particular analog function. In general, these work by assuming an ohmic resistance in series with the base or emitter of an otherwise ideal transistor. A compensating voltage is then developed across an external resistor by passing through it a replica of the critical base or emitter current. This compensating voltage, properly inserted, cancels the undesired ohmic term in the critical V_{BE}.

Morgan [8.1] used this technique in a single-ended logarithmic-converter cell. McGinn [8.2] used it indirectly to reduce distortion in a translinear four-quadrant multiplier of unusual configuration. Both of these were basically discrete-component circuits.

More recently, Schmoock [8.3] demonstrated a monolithic translinear circuit incorporating LCE compensation. Schmoock used monolithic resistors designed to approximate the ohmic resistance, referred to the emitter, of the critical transistors. Huijsing and van Steenwijk, in a sophisticated monolithic exponential converter [8.4], followed a similar approach, formulated in terms of base resistance. A modified Caprio circuit [8.5], essentially a negative impedance converter, synthesized a linear negative resistance whose magnitude approximated the ohmic resistance of a transistor (referred to the base circuit). Inserted in series with the bases of critical transistors, this cancelled the linear LCE.

All of these circuits depended on approximating a transistor's ohmic resistances with separate resistors. It is inherently difficult to do this accurately, since resistive parasitics in a transistor are subject to a complex of processing factors, making them difficult to predict and match. Moreover, such techniques correct only for linear LCE, and not for current crowding or other nonlinear effects.

It would be more satisfactory to use the actual parasitics of similar transistors as a source of correction signal. The problem in doing so is to expose these second-order effects without being swamped by the (much larger) logarithmic part of V_{BE}. That is the subject of this chapter; section 8.3 lays out, in detail, a practical approach.

In a recent and apparently independent development, Blauschild [8.6] has demonstrated an LCE correction circuit using the principle of this chapter. Blauschild's correction cell, also incorporated in a translinear variable-gain element, is a special case of the correction quad of Figure 8.3.

8.3 A General Correction Strategy

Rather than obtaining compensating voltage drops in explicit resistors, as past correction circuits have, the method advocated here derives a correction signal from transistors similar to those bearing the signal of interest. The method is general because it applies to a wide range of circuits, and because it does not depend on an a-priori knowledge of ohmic resistance values in the devices being corrected. Indeed, it can in principle correct automatically for all log-conformance error effects, nonlinear as well as ohmic.

This new strategy is based on configuring a number (at most four) of transistors to extract only the non-logarithmic part of the V_{BE}-to-I_C relation, suppressing the logarithmic component -- the opposite of the usual aim. The resulting correction cell produces exactly the voltage drop needed to cancel log-conformance error in an existing circuit, such as a variable-gain amplifier.

8.3.1 The Basic Idea

Consider the hypothetical circuit of Figure 8.1. We assume large beta so that base currents can be neglected. This circuit has a superficial resemblance to those of Caprio [8.5], who employed cross-connected transistor quads in voltage-to-current converter circuits. However, unlike Caprio's circuits, it contains no resistors and the upper pair of devices have a different base-emitter junction area from the lower pair.

The circuit would in fact be remarkably useless if all four transistors had the same area. In that case the voltage V would always be zero, regardless of the currents I_1 and I_2. Because of the cross connection, the two bottom emitters would be at the same potential, which would fluctuate with changes in I_1 and I_2.

Consider now what happens when the upper pair have some nominal junction area A_1 while the lower pair have a very large junction area A_2. Since the two pairs operate at the same current levels, the lower pair will therefore have a very small current density. Recalling that log-conformance error increases with current density, the lower pair can then be taken to have negligible LCE, so that each device obeys the idealized relation (7.7). The

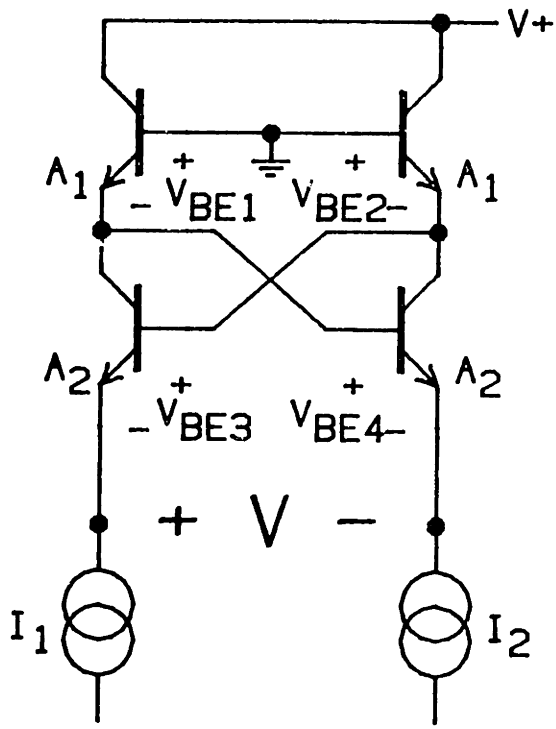


FIGURE 8.1 -- CONCEPTUAL CIRCUIT
FOR ISOLATING LOG-CONFORMANCE ERROR

upper pair, however, exhibits log-conformance error, expressed generally by (7.9).

The voltage V is then

$$V = V_{BE1} + V_{BE4} - V_{BE2} - V_{BE3}$$

$$= \left[\frac{kT}{q} \ln\left(\frac{I_1}{A_1 J_S}\right) + V_E\left(\frac{I_1}{A_1}\right)\right] + \frac{kT}{q} \ln\left(\frac{I_2}{A_2 J_S}\right)$$

$$- \left[\frac{kT}{q} \ln\left(\frac{I_2}{A_1 J_S}\right) + V_E\left(\frac{I_2}{A_1}\right)\right] - \frac{kT}{q} \ln\left(\frac{I_1}{A_2 J_S}\right)$$

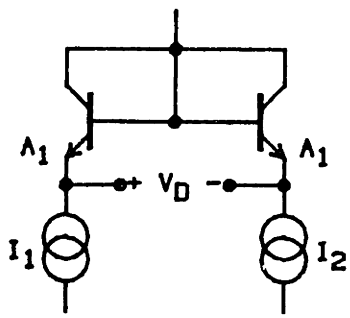
$$= V_E\left(\frac{I_1}{A_1}\right) - V_E\left(\frac{I_2}{A_1}\right) \qquad (8.1)$$

The logarithmic components have cancelled, and V contains only the log-conformance error terms from the upper pair. This voltage, dependent on the currents I_1 and I_2, is available as a correction signal to be inserted at a suitable point in a prototype circuit. A very simple example is given in Figure 8.2. The pair of diode-connected transistors (again, we are neglecting base currents) yields

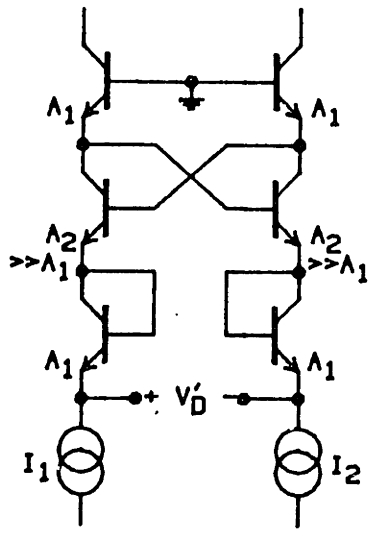
$$V_D = \frac{kT}{q} \ln\left(\frac{I_2}{I_1}\right) + V_E\left(\frac{I_2}{A_1}\right) - V_E\left(\frac{I_1}{A_1}\right) \qquad (8.2)$$

The correction circuit adds in a signal-dependent LCE term of opposite polarity, giving

$$V_D' = \frac{kT}{q} \ln\left(\frac{I_2}{I_1}\right) \qquad (8.3)$$



(a) PROTOTYPE CIRCUIT



(b) PROTOTYPE CIRCUIT WITH CORRECTION QUAD ADDED

Figure 8.2

Unfortunately, this conceptual correction circuit requires a transistor pair of very large junction area. If such a pair were available, why not use it for the prototype circuit of Figure 8.2(a)? The utility of the new correction technique rests on the following facts:

- Finite-area transistors may be used in the two pairs.
- This correction technique is useful with prototype circuits of far more complexity than that of Figure 8.2(a).

The key is to use two pairs of transistors, with the current densities in one pair proportional to the current densities in the other pair, but of different magnitude. Then when the base-emitter junctions of all four devices are properly connected in series, the logarithmic parts of the V_{BE}s will all cancel but any non-logarithmic terms will not. For example, if the lower pair in Figure 8.1 had any finite area other than A_1, the resulting voltage V would still be due only to the LCE effects in the four transistors. The exact magnitude and current dependence of V would of course depend on the areas A_1 and A_2 (i.e., on the current density ratio between upper and lower pairs).

Obtaining a useful correction circuit with finite-area transistors requires the proper choice of current density ratio. Moreover, there are many alternative ways to construct a correction quad and to insert it into a prototype circuit (translinear or otherwise). We first consider a generalization of the above examples.

8.3.2 Generalized Correction Quad

Figure 8.3 shows a general correction quad (the upper four transistors) imbedded in a general prototype circuit. The prototype circuit contains a series connection of some number L of base-emitter voltages, an arbitrary series network (block "C"), and incidental circuitry (blocks "B") that provides some or all of the currents in the L transistors. The essential feature of the prototype circuit is a series connection of semiconductor junction voltages, each of which is subject to log-conformance error. The prototype circuit may contain many other features not shown, and may be part of a larger circuit containing other groups of series junction voltages.

This prototype circuit may represent any of a vast number of analog circuits that exploit the detailed junction behavior. The translinear circuits of chapters 6 and 7 are a special case, without the block "C"; this block allows for non-V_{BE} terms in the loop equation. The "loop of voltage drops" in Figure 8.3 refers to the series of combination of the correction quad, the L base-emitter voltages in the prototype circuit, and block "C". The dotted line indicates where the correction quad has been added into the prototype circuit by breaking a connection.

Note the cross-proportionality of current densities in the correction quad, which insures that the logarithmic V_{BE} components from the quad will sum to zero in V_{CORR}. In the special case of Figure 8.1, the required cross-proportionality was accomplished by exchanging the upper pair of devices and connecting upper emitters to lower collectors. The two pairs then operated at the same current magnitudes and area scaling gave the factor c.

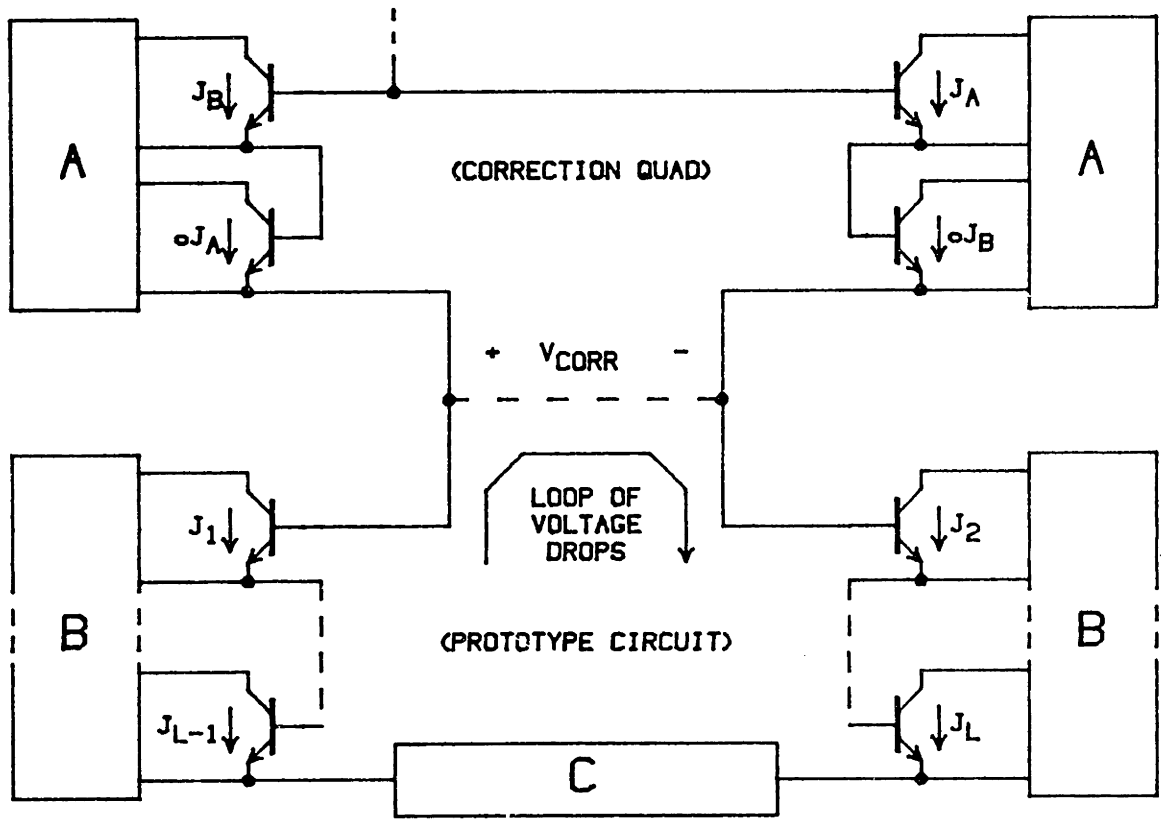


Figure 8.3 -- Correction quad added to an arbitrary prototype circuit that contains a loop of base-emitter junction voltage drops. The total number of such drops in the prototype circuit is L.

Blocks "A" denote current-establishing circuitry for the correction quad.

Blocks "B" denote current-establishing circuitry for the prototype circuit.

Block "C" denotes elements in the voltage loop, other than junction voltage drops, that are part of the prototype circuit.

There are other ways to impose the factor c. For example, all four devices in the quad could have the same area while the associated circuitry (blocks "A" in Figure 8.3) supplied ratioed currents to the four devices. This issue is taken up in section 8.4.4, and illustrated in Figure 8.8.

The two independent current densities (J_A and J_B) in the correction quad must be related to those in the prototype circuit (J_1 through J_L) in such a way that the net log-conformance error of the correction quad compensates for the net LCE in the prototype circuit. This is treated quantitatively in section 8.4.1.

The purpose of the correction quad is to introduce a small, signal-dependent correction voltage into the overall loop. By Kirchoff's Voltage Law, the proper correction voltage inserted in series with the L series-connected junctions has the effect of correcting all of the individual LCE voltages of these junctions. Similarly, although the correction quad is added to the top of the prototype circuit in Figure 8.3, the correction devices actually could be placed anywhere in the loop, so long as their orientation was preserved. They could even be interspersed with the various junctions in the prototype circuit to be corrected.

In Figure 8.3, the net voltage added to the loop by the correction quad is

$$V_{CORR} = V_E(J_A) + V_E(cJ_B) - V_E(cJ_A) - V_E(J_B) \qquad (8.4)$$

8.3.3 Non-Translinear Applications

The log-conformance correction technique sketched so far applies to any prototype circuit where fidelity of the V_{BE}-to-I_C relation is critical. Although the emphasis in this thesis is translinear applications, and variable-gain circuits in particular, it is not necessary that the prototype circuit operate with current-mode inputs and outputs, or that the number of clockwise base-emitter drops equal the number of counterclockwise drops in Figure 8.3.

Following the analysis in this chapter, the same correction strategy should be useful for the two other classes of circuits (besides translinear) where the V_{BE}-to-I_C relation is used directly. In the first class, a difference of V_{BE} drops is used as an input or an output. This, for example, is implicit in Figure 8.2, where a pair of diodes approximates (8.3). This yields the logarithm of a current ratio; or, if the current ratio is fixed, a voltage proportional to absolute temperature.

The other important class of circuits exploits the full V_{BE} dependence, rather than a V_{BE} difference. The common example is "bandgap" reference circuits. Again, the fidelity of V_{BE}-to-I_C is critical. In this case, however, the current range is relatively narrow and other factors usually influence accuracy to a greater extent than does the log conformance.

It is also possible to use a "one-sided" correction circuit, rather than a correction quad. In this case, however, there is a residual temperature-proportional offset voltage of the form (8.3). This is potentially useful only where the offset may be tolerated, as for example in an explicit temperature sensor.

8.3.4 Non-BJT Applications

As was pointed out in section 7.2.1, the logarithmic current-voltage characteristic of (7.1) is not unique to bipolar junction transistors (BJT). Although the area has not been explored much to date, it is entirely possible that useful translinear circuits may be constructed from weakly-inverted (subthreshold) MOSFETs*, permeable-base transistors [8.7], or other devices that exhibit this behavior. Because the log conformance in these devices is generally poorer than in the BJT, correction for LCE may have even greater importance.

It should be evident from the preceding sections that the proposed technique applies to any semiconductor devices that approximate (7.1). The strategy is the same in each case: a correction quad of devices that cancels its internal logarithmic terms, presenting an LCE correction signal that is inserted into a prototype circuit.

8.4 Design Considerations

8.4.1 Specification of the Correction Quad

Given the arrangement of Figure 8.3, the designer must select J_A, J_B and c so that the log-conformance error in the correction quad substantially or completely cancels the log-conformance error in the rest of the circuit. In practice J_A and J_B must vary along with J_1 through J_L of the prototype circuit, so J_A and J_B must be some realizeable (i.e., simple) function of J_1 through J_L.

In the notation of Figure 7.4, the various log-conformance errors in the prototype circuit can be represented by a net error voltage V_X connected in series with a loop of ideal junction voltages.

---

*See section 5.3.4.

Comparing Figure 7.4 with Figure 8.3, perfect correction will clearly be obtained if

$$V_{CORR} = V_X \tag{8.5}$$

or, substituting in (7.13) and (8.4),

$$V_E(J_A) + V_E(cJ_B) - V_E(cJ_A) - V_E(J_B) = \sum_{ccw} V_E(J_n) - \sum_{cw} V_E(J_n) \tag{8.6}$$

where the summations refer to J values in transistors oriented counter-clockwise or clockwise in the prototype circuit of Figure 8.3.

To obtain this perfect correction we must know the exact shape of the LCE function $V_E(\cdot)$. Recognizing the difficulty of this, we use instead a pragmatic correction criterion: we assume that $V_E(J)$ is proportional to J; in other words, that the log-conformance error is ohmic. This choice of correction criterion reduces (8.6) to a convenient form; it yields exact correction of ohmic LCE, and partial correction of nonlinear effects due to current crowding, etc. (see section 8.5.1).

More sophisticated correction criteria could be applied, given additional knowledge of $V_E(\cdot)$. For example, the log-conformance error function $V_E(J)$ could be measured for a particular type of transistor. The cross-proportional current densities in the correction quad could then be driven so that the correction quad minimized some error measure, such as mean squared residual loop error voltage. The linear method derived below, however, works well in practice

and has the advantage of applying to all pn junction devices regardless of the exact behavior of LCE at high currents.

Thus, assuming temporarily that $V_E(J)$ is proportional to J, and substituting such a form into (8.6), we obtain

$$(1-c)(J_A-J_B) = \sum_{CCW} J_n - \sum_{CW} J_n \tag{8.7}$$

(the proportionality constant in $V_E(J)$ is not important since it cancels out).

If the current densities in Figure 8.3 satisfy (8.7), then the correction quad will eliminate the effects of linear LCE in the prototype circuit, resulting in improved performance of the prototype circuit's function. In particular, equation (8.7) is satisfied by

$$J_A = \frac{1}{(1-c)} \sum_{CCW} J_n \tag{8.8a}$$

$$J_B = \frac{1}{(1-c)} \sum_{CW} J_n \tag{8.8b}$$

These are easy to satisfy in practice; the currents that must be established in the correction quad will be linear combinations of the currents in the original L transistors.

The parameter c, ranging from zero to unity, is set by the designer and allows a degree of design freedom. As c approaches zero, the required current density ratio in the quad grows large, but the correction becomes exact for all log-conformance errors, both linear and nonlinear. Values of c greater than 1 imply a correction signal

of the opposite polarity, which would enhance rather than cancel the errors.

8.4.2 The Area-Ratioed Quad, and Applications

Although by no means the only embodiment, configurations similar to Figure 8.1 are convenient correction quads since they permit the required current density ratioing to be achieved with junction area ratios. Figure 8.4 shows a typical case. This circuit has differential input and output, and provides some buffering because of the transistor current gain. The currents I_A and I_B also may be recovered at the top collectors. In terms of the preceding discussion,

$$V_{CORR} = V_{OUT} - V_{IN}$$

$$J_A = I_A/A_1$$

$$J_B = I_B/A_1$$

$$c = A_1/A_2 \qquad (8.9)$$

(neglecting base currents).

To use this in conjunction with a prototype circuit, I_A and I_B, as well as A_1 and A_2, must have the proper relation to the current densities in the prototype circuit. Equations (8.8), along with (8.9), give

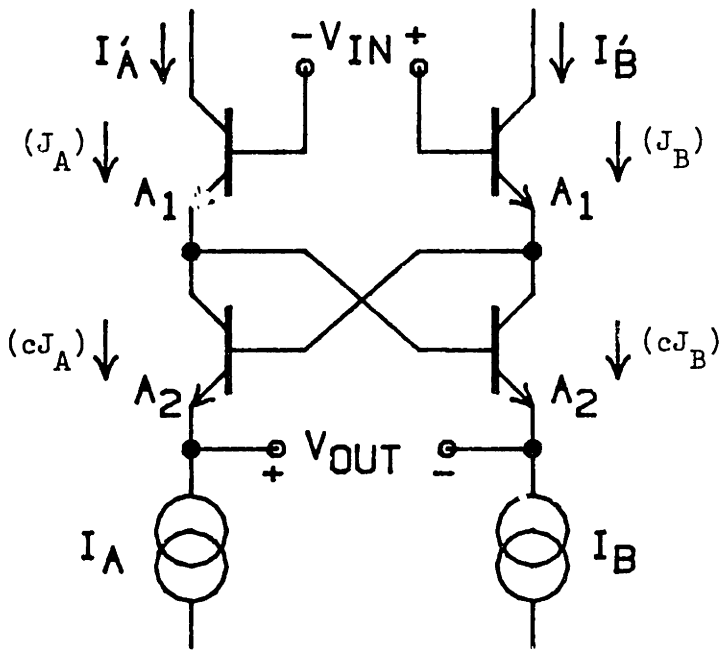


FIGURE 8.4 -- A USEFUL FORM OF CORRECTION QUAD.

$$I_A = \frac{A_1}{(1-c)} \sum_{ccw} \left(\frac{I_n}{A_n}\right) \tag{8.10a}$$

$$I_B = \frac{A_1}{(1-c)} \sum_{cw} \left(\frac{I_n}{A_n}\right) \tag{8.10b}$$

where now explicit currents and areas have been inserted for the L prototype junctions.

In the simple (and common) case where L = 2, there are only two critical devices in the prototype circuit, both of junction area A. Equations (8.10) become

$$I_A = \frac{1}{(1-c)} \frac{A_1}{A} I_1 \tag{8.11a}$$

$$I_B = \frac{1}{(1-c)} \frac{A_1}{A} I_2 \tag{8.11b}$$

In practice, it is convenient for the currents I_1 and I_2 also to form the currents in the correction quad, I_A and I_B. Figures 8.5 and 8.6 illustrate two such cases. This added condition allows complete specification of device areas in the correction quad. From (8.11a), with $I_A = I_1$,

$$A_1 = (1-c)\ A \tag{8.12a}$$

Equation (8.9) then gives

$$A_2 = \frac{(1-c)}{c} A \tag{8.12b}$$

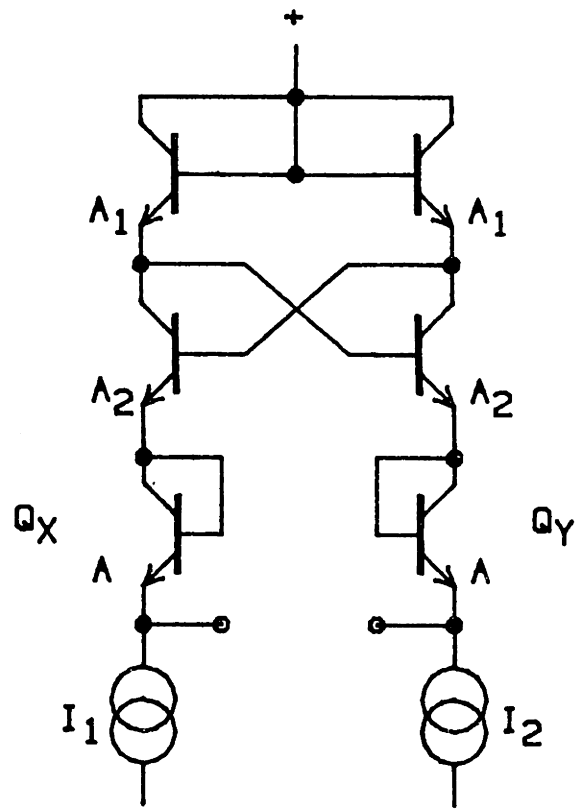


Figure 8.5 -- The area-ratioed quad of Figure 8.4 used to cancel log-conformance error in a critical diode pair (Q_X, Q_Y).

Using the zero-ohmic-error criterion, the base-emitter junction areas in the quad are given in terms of those of the diode pair and the ratio parameter "c" by

$$A_1 = (1-c)A, \quad A_2 = \frac{(1-c)}{c}A$$

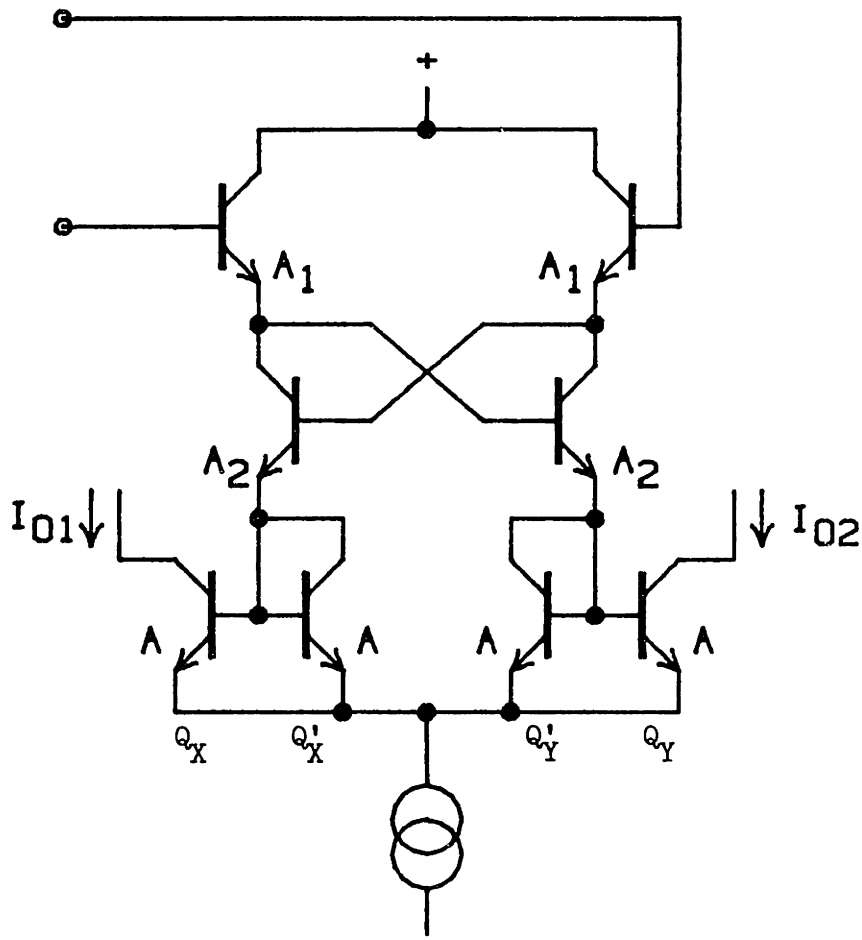


Figure 8.6 -- Differential transistor pair incorporating log-conformance error correction.

These are design equations for A_1 and A_2 in terms of prototype device area A and correction parameter c.

Examples showing two-transistor prototype circuits are important in translinear variable-gain elements. Distortion in such elements often can be attributed to a single, critical, differentially-connected pair of junctions, as in section 8.6.

Figure 8.5 is a corrected diode pair. This is similar to Figure 8.2(b), except that now all devices have small junction areas, and the difference in the areas accomplishes the correction. Q_X and Q_Y are connected as diodes in Figure 8.5; they could in principle be placed at the top of the circuit, above the correction quad. However, the configuration shown has superior AC characteristics, as discussed in section 8.6.

Figure 8.6 is a differential transistor pair (Q_X and Q_Y) with correction circuit added. The principle is identical to the previous example. Note that the upper bases of the correction quad now serve as input terminals. The extra transistors Q_X' and Q_Y' form current mirrors with Q_X and Q_Y, respectively, so that the proper signal currents flow in the correction quad. This makes it necessary for the tail current source to have twice the value that it normally would have.

Figure 8.7 gives a concrete example involving a prototype circuit with <u>four</u> critical base-emitter junctions. Transistors Q_W, Q_X, Q_Y and Q_Z form one of the (less common) translinear four-quadrant multipliers [8.8]. The inputs are the current sources at the bottom (a and b are normalized variables in the range -1 to +1). The additional four transistors in that part of the circuit sense the

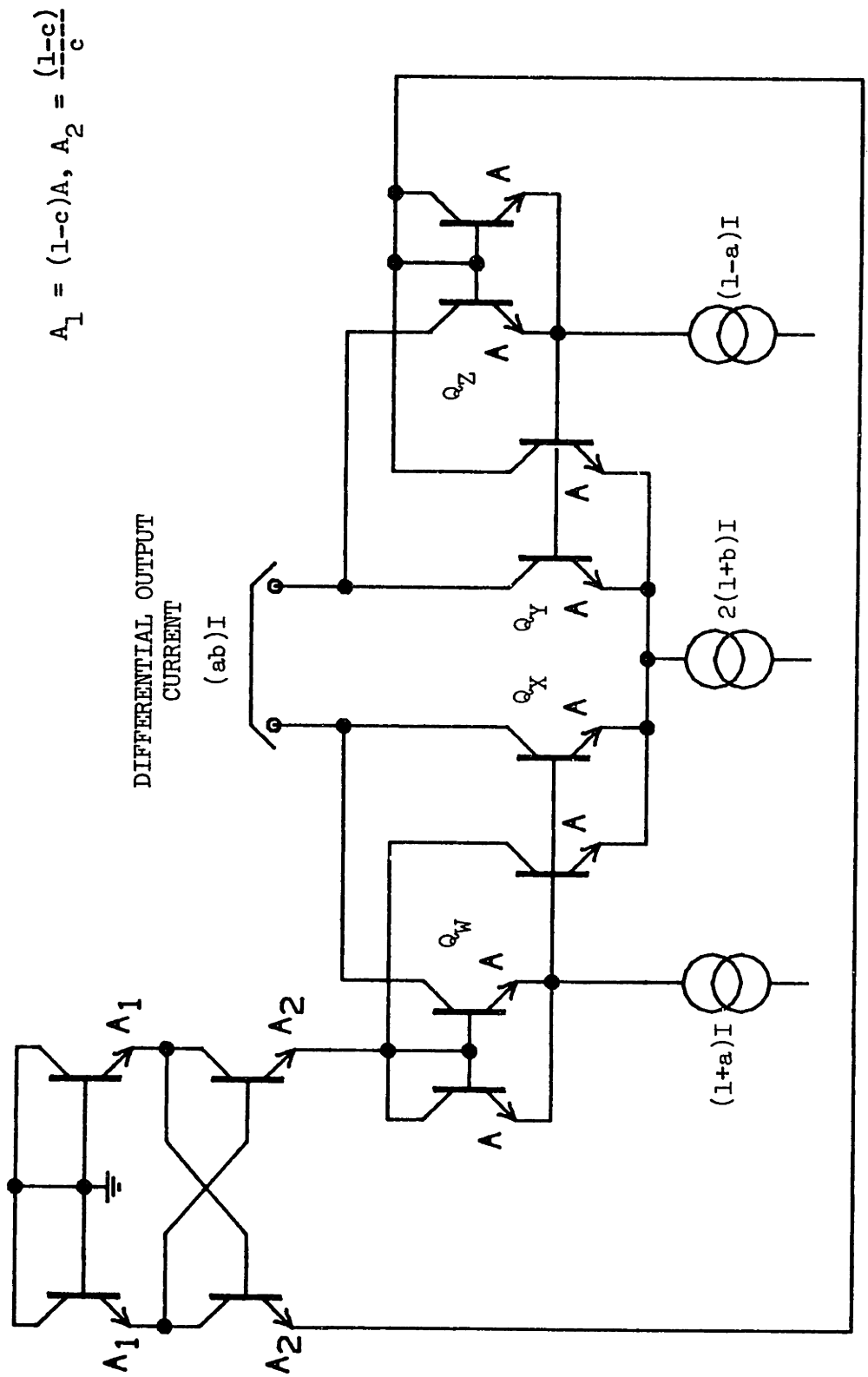


Figure 8.7 -- Four-quadrant analog multiplier incorporating log-conformance error correction. Errors in the four critical devices Q_W, Q_X, Q_Y, Q_Z are corrected.

currents in Q_W, Q_X, Q_Y, Q_Z and feed them to the correction quad, in a manner similar to Q_X' and Q_Y' of Figure 8.6.

8.4.3 Available Area Ratios using Multiple Emitters

One standard technique for obtaining scaled emitter areas is to place multiple identical emitter diffusions in the same monolithic transistor. An equivalent procedure with discrete transistors is to connect some number of matched devices in parallel. The result gives more accurate scaling of current characteristics than, for example, the use of rectangular emitter geometries of variable length. However, it imposes the requirement that all areas be integer multiples of some standard value.

To use the area-ratioed correction quad of Figure 8.4 with multiple-emitter transistors, the junction areas in the quad and the prototype circuit must be chosen from a discrete set that satisfies both the integer-multiple constraint and the correction equations, (8.10). The results will be presented for the case where the prototype circuit has a single pair of transistors, as in Figures 8.5 and 8.6, or multiple pairs with identical area, as in Figure 8.7.

As before, we denote the area in the prototype pair (or pairs) by A, and those in the correction quad of Figure 8.4 by A_1 and A_2. Let all three areas be integer multiples of a standard (perhaps minimum-size) area A_{min}:

$$A = pA_{min} \quad \text{(prototype pair)}$$

$$A_1 = mA_{min} \quad \text{(small-area pair in quad)}$$

$$A_2 = nA_{min} \quad \text{(large-area pair in quad)}$$

Then in order to satisfy the design equations of (8.12), we require that the integers (p,m,n) satisfy

$$\frac{p}{m} - \frac{p}{n} = 1 \tag{8.13}$$

Such a set yields a value

$$c = m/n \tag{8.14}$$

Recall that c is, in effect, an inverse figure of merit; as c approaches zero, the correction circuit compensates for all non-logarithmic components in V_{BE}. Equation (8.13) defines sets of area factors (p,m,n) giving exact correction of linear log-conformance error, but not all such sets have a minimum total area for the corresponding value of c. Considering only minimum-area correction quads, it can be shown that

$$m = p - 1$$

$$n = p(p-1)$$

$$c = 1/p \tag{8.15}$$

The parameters of the quad thus follow from the prototype area factor p, which directly determines the value of c. Table 8.1 shows the first five multiple-emitter area sets, along with resulting

MULTIPLES OF MINIMUM EMITTER AREA			RESULTING	TOTAL NUMBER OF EMITTERS
A	A_1	A_2	c	IN QUAD
2	1	2	0.500	6
3	2	6	0.333	16
4	3	12	0.250	30
5	4	20	0.200	48
6	5	30	0.167	70

Table 8.1 -- Compatible sets of areas for the prototype-circuit (A) and correction-quad (A_1, A_2) emitters when multiple-emitter area ratioing is used.

values of c and the total number of emitters in the quad. The total number of minimum-size emitters grows as the square of p, due mostly to the large-area (A_2) transistor pair. Thus, it is difficult to realize very small values of c in an area-ratioed quad using multiple emitters. It may be possible to do so using non-integer area factors that satisfy (8.13) or electronic current ratioing (discussed in the next section).

As a practical matter, the first set of areas in Table 8.1 (p=2, m=1, n=2) is attractive since it is the smallest area-ratioed combination possible using multiple emitters. This set of areas was used for the experimental work in section 8.6.

8.4.4 Area Ratioing versus Electronic Current Ratioing

Although the examples shown so far have employed ratioed emitter areas, this is not fundamentally necessary. Equations (8.8) prescribe values for the quad current densities J_A and J_B in terms of other variables. These current densities are related to those in the prototype circuit through the factor (1-c); moreover, there is an additional ratio of c within the quad, as shown generally in Figure 8.3.

Such current density ratios, which are at the heart of this correction technique, may be realized using junction area ratios or current ratios or both. With electronic current ratioing, all transistors can have the same size; some form of fixed current gain or loss circuit achieves the current density ratios.

This is illustrated in Figure 8.8, where electronic ratioing of emitter currents provides the factor of c within the quad. Were

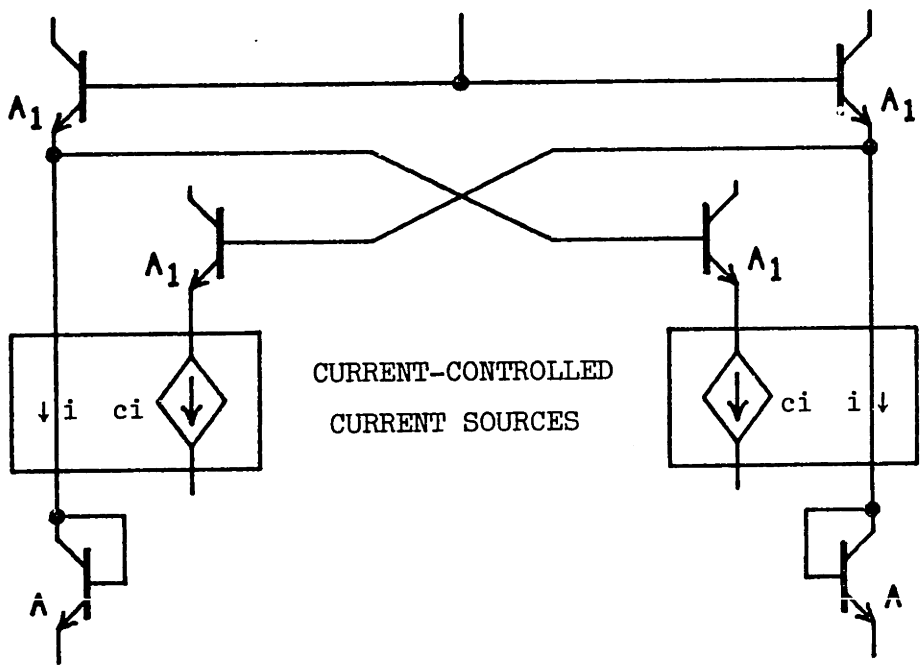


Figure 8.8 -- An example of electronic current ratioing. The current-loss elements take the place of junction-area ratioing within the correction quad. Compare with Figure 8.5.

the same technique used to generate currents for both pairs of the quad, then all six transistors would have the same area (A).

This technique has not been explored in depth. It has the potential, however, of freeing the correction circuit from the large area ratios required for a small value of "c" when multiple-emitter transistors are used. The complication is the requirement for a designable, stable current-ratioing or splitting element. In the past, such an element has been realized often with mutliple-emitter transistors, which clearly defeats the purpose here.

8.5 Second-Order Effects

We now relax some of the simplyfing assumptions used in the previous analysis. We assume again that the correction circuit is being incorporated into a prototype circuit with a critical pair of transistors operating at (signal-dependent) current densities J_1 and J_2. For a larger number of junctions in the prototype circuit, these single current densities are to be repaced by, respectively, sums of "counterclockwise" and "clockwise" current densities, following Figure 8.3.

The following equations summarize the function of the general correction quad shown in Figure 8.3, with independent current densities J_A and J_B derived as in section 8.4.1:

$$J_A = J_1/(1-c) \tag{8.16a}$$

$$J_B = J_2/(1-c) \tag{8.16b}$$

$$V_{CORR} = V_E(J_A) + V_E(cJ_B) - V_E(cJ_A) - V_E(J_B) \tag{8.16c}$$

where the function $V_E(J)$ is the log-conformance error voltage in a single V_{BE}. The voltage V_{CORR} is used to correct the V_{BE} difference in the prototype pair:

$$\Delta V'_{BE} = V_{BE1} - V_{BE2} - V_{CORR} \tag{8.16d}$$

8.5.1 Non-Ohmic Log-Conformance Error

The correction criterion of section 8.4.1 yielded a family of correction circuits that exactly cancel ohmic error terms in V_{BE}; i.e., additive error proportional to collector current (or, in a more general logarithmic device, the appropriate analog to collector current). At the same time a partial correction of non-ohmic LCE (nonlinear with respect to collector current) was alluded to. This will be examined now.

Suppose that an arbitrary log-conformance error voltage is represented in power-series form:

$$V_{BE}(J) = \frac{kT}{q} \ln\left(\frac{J}{J_S}\right) + \sum_{n=1}^{\infty} a_n \left(\frac{J}{J_o}\right)^n \tag{8.17}$$

Here, J_o is some normalizing value included so that the coefficients a_n will be dimensionless. The n = 1 term is the familiar ohmic error; in the notation of (7.8), we have

$$a_1 = J_o A_o R_o \tag{8.18}$$

Applying (8.17) to the correction-circuit equations (8.16) yields a residual-error expression showing the effect of the correction on higher-order terms (n=2, 3, etc.). The residual error is defined as

$$\Delta V_E' = \Delta V_{BE}' - \frac{kT}{q} \ln\left(\frac{J_1}{J_2}\right) \tag{8.19}$$

The result is

$$\begin{aligned}\Delta V_E' &= V_E(J_1) - V_E(J_2) - V_E\left(\frac{J_1}{1-c}\right) - V_E\left(\frac{cJ_2}{1-c}\right) \\ &\quad + V_E\left(\frac{cJ_1}{1-c}\right) + V_E\left(\frac{J_2}{1-c}\right) \\ &= \sum_{n=1}^{\infty} \frac{a_n}{J_o^{\,n}} \left[1 - \left(\frac{1}{1-c}\right)^n + \left(\frac{c}{1-c}\right)^n\right] (J_1^{\,n} - J_2^{\,n})\end{aligned} \tag{8.20}$$

The original (uncorrected) error was

$$\begin{aligned}\Delta V_E &= V_E(J_1) - V_E(J_2) \\ &= \sum_{n=1}^{\infty} \frac{a_n}{J_o^{\,n}} J(J_1^{\,n} - J_2^{\,n})\end{aligned}$$

Thus, in (8.20), each loop error voltage term of order n has been multiplied by the factor

$$f(c,n) = \left[1 - \left(\frac{1}{1-c}\right)^n + \left(\frac{c}{1-c}\right)^n\right] \qquad (8.21)$$

For n = 1 this is zero; the ohmic error cancels, as expected. For c very small (that is, for a large current-density ratio in the correction quad), the last term in (8.21) vanishes, leaving

$$f(c,n) \tilde{=} -cn, \quad n \geq 2.$$

The function f(c,n) is plotted in Figure 8.9. It is difficult to gain insight into the shape of the residual nonlinear error without knowing an explicit series expansion. However, Figure 8.9 provides a guide in the case of isolated higher-order terms, such as a quadratic or cubic nonlinearity in $V_E(J)$. For a given order n, there is a maximum value of c, beyond which the correction circuit will enhance rather than reduce distortion of that order.

Treating polynomial distortion terms separately provides a worst-case measure of the circuit's effect, and may yield an exaggerated estimate of residual error for practical log-conformance characteristics. However, Figure 8.9 clearly indicates that distortion of all orders is made worse when c = 1/2. Therein lies a drawback of the smallest-area version of the area-ratioed quad. Smaller values of c afford correction of higher-order LCE components but at the cost of increased total area, unless electronic current ratioing (section 8.4.4) is used.

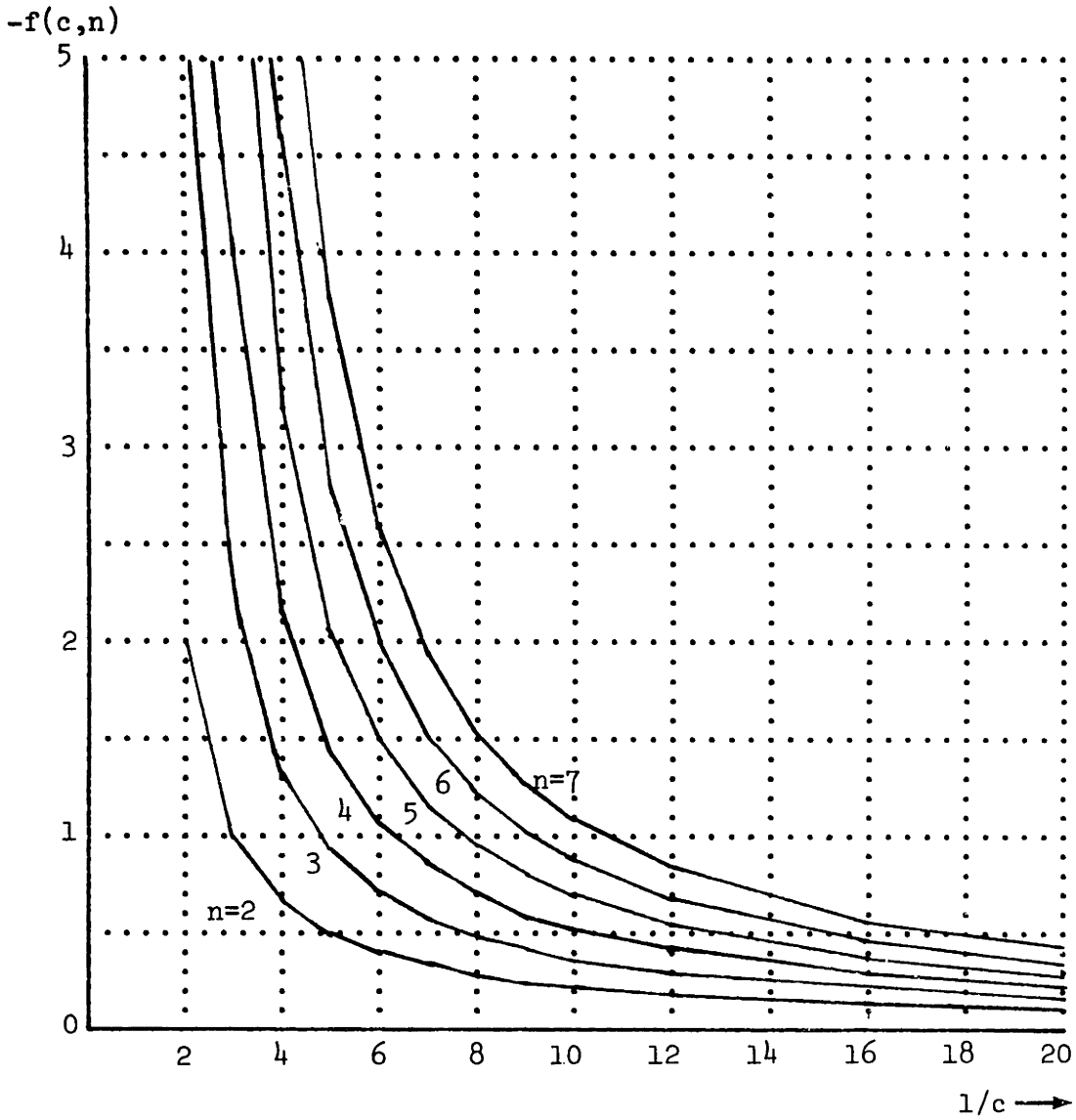


Figure 8.9 -- Plot of the function f(c,n) from equation (8.21).

Each curve gives the factor by which the correction circuit scales the nonlinear log-conformance error term of degree n. The quantity f(c,n) is negative, so its magnitude is shown.

8.5.2 Current Density Mismatches

Even with purely linear log-conformance error, distortion may occur due to imperfect establishment of current densities in the correction quad. Values other than those indicated in Figure 8.3 may arise through two mechanisms:

- Mismatch of junction-area values in the quad.
  Also, differences in J_S or LCE amplitude due to processing gradients (in a monolithic circuit) may be cast in this form.
- Mismatch of currents applied to the quad.

Assume that instead of the nominal values J_A, cJ_B, cJ_A, J_B, the four devices have, respectively, the current densities k_1J_A, k_2cJ_B, k_3cJ_A, and k_4J_B, with k values close to unity. The values J_A and J_B are defined as before by (8.16a) and (8.16b); base currents will be neglected, and linear LCE will be assumed, so that $V_E(J) = JA_oR_o$.

Because of the k factors, the logarithmic terms in the four base-emitter voltages no longer sum to zero, and (8.16c) must be replaced with the more general expression (including k factors)

$$V_{CORR} = V_{BE}(k_1J_A) + V_{BE}(k_2cJ_B) - V_{BE}(k_3cJ_A) - V_{BE}(k_4J_B) \qquad (8.22)$$

Such a correction voltage leaves a residual error, defined in (8.19), of

$$\Delta V_E' = A_o R_o (J_1 - J_2) - V_{CORR}$$

$$= \frac{kT}{q} \ln\left[\frac{k_3 k_4}{k_1 k_2}\right] + A_o R_o \left[J_1\left(1 - \frac{k_1 - k_3 c}{1 - c}\right) - J_2\left(1 - \frac{k_4 - k_2 c}{1 - c}\right)\right]$$

(8.23)

The first term in (8.23) is an offset voltage, proportional to absolute temperature but independent of signal. Such a voltage arises in translinear circuits whenever there is unintentional current-density mismatch; it may be trimmed out with suitable circuitry, although this is not always convenient. The remainder of (8.23) is a residual ohmic error. Note that this error is <u>not</u> a symmetric function of $(J_1 - J_2)$, in general, unless the current-density mismatch factors obey

$$k_1 - k_3 c = k_4 - k_2 c$$

A mismatch situation may arise in a "benign" manner when the correction quad is intended to cancel ohmic effects in a prototype pair with unbalanced resistances. In that case the residual loop error voltage $\Delta V_E'$ will be zero again, except for the temperature-proportional offset component, which must be removed by other means.

8.5.3 <u>Base Currents in the Area-Ratioed Quad</u>

Consider now the effect of finite current gain in the transistors of Figure 8.4. The common-base current gain α will be assumed identical for all four devices, and independent of current density.

With the definitions of (8.9) retained, the current densities in the quad are different from those indicated in Figure 8.4 due to base currents. The resulting value of V_{CORR} is

$$V_{CORR} = V_{BE}(\alpha^2 J_A + \alpha[1-\alpha]J_B) + V_{BE}(\alpha c J_B)$$

$$- V_{BE}(\alpha^2 J_B + \alpha[1-\alpha]J_A) - V_{BE}(\alpha c J_A)$$

$$= \frac{kT}{q} \ln\left[\frac{\alpha J_A J_B + (1-\alpha)J_B^2}{\alpha J_A J_B + (1-\alpha)J_A^2}\right] + \alpha(2\alpha-1-c)A_o R_o (J_A - J_B) \qquad (8.24)$$

With the substitution of (8.16a) and (8.16b), to obtain a result in terms of current densities J_1 and J_2 in the prototype pair, and the approximation $\ln(1+x) \cong x$ for small x, this gives

$$V_{CORR} \cong \frac{1}{\beta}\frac{kT}{q}\left[\frac{J_2}{J_1} - \frac{J_1}{J_2}\right] - (1-\alpha)\left[\frac{1 + 2\alpha - c}{1 - c}\right]A_o R_o (J_1 - J_2)$$

$$+ A_o R_o (J_1 - J_2) \qquad (8.25)$$

with the condition that the spread between J_1 and J_2 be much less than β (usually true in practice). The last term in (8.25) is the <u>desired</u> value of V_{CORR} and the others are error components.

The residual error is therefore

$$\Delta V_E' = A_o R_o (J_1 - J_2) - V_{CORR}$$

$$\cong -\frac{1}{\beta}\frac{kT}{q}\left[\frac{J_2}{J_1} - \frac{J_1}{J_2}\right] + \frac{1}{\beta}\left[\frac{3 - c}{1 - c}\right] A_o R_o (J_1 - J_2) \qquad (8.26)$$

In order to interpret this, consider the special (but common) case where J_1 and J_2 are driven differentially. That is,

$$J_1 = J + j, \; J_2 = J - j$$

Then (8.26) becomes

$$\Delta V_E' \cong \frac{4}{\beta}\frac{kT}{q}\left[\frac{J}{J^2 - j^2}\right] j + \frac{2}{\beta}\left[\frac{3 - c}{1 - c}\right] A_o R_o j \qquad (8.27)$$

For deviations (signals) about $j = 0$, this has the following significance. The first term is a new component, similar to linear LCE, introduced by the base currents. It corresponds to an ohmic emitter resistance about $2/\beta$ as large as the intrinsic r_e of the prototype transistors (and of negative polarity). Thus there will be a small β-dependent error in the corrected circuit's transfer function even at low currents, where LCE is normally not a problem. The second term in (8.27) is a residual ohmic error, reduced in magnitude from the original value by the factor $(3-c)/\beta(1-c)$.

8.6 Measured Results

A breadboard version of a variable-gain current amplifier was constructed to test the basic correction strategy. This current amplifier incorporated an area-ratioed quad with c = 1/2, which optionally could be removed from the circuit in order to measure the uncorrected performance of the amplifier.

Figure 8.10 shows the test configuration. The circuit was constructed from three transistor-array packages obtained from the same wafer; the characteristics of the transistors were discussed in section 7.5.1. The area ratioing in the input-stage quad was obtained by parallel connection of identical transistors, as suggested in section 8.4.3. Because of the large number of minimum-area transistors required, the input stage was assembled from two five-transistor arrays; the disposition of individual transistors was arranged so that each half of the quad was matched to a corresponding half of the prototype pair. Close thermal coupling between the two array packages was obtained by mounting them to a common aluminum bar with the aid of thermal paste.

The basic circuit is similar to the test configuration of chapter 7, except for the emitter followers Q33-Q34 that decouple the input and output sections. Also, signal current is presented to the emitters of the prototype pair rather than the collectors; thus an alpha factor is to be expected. As before, a trim adjustment allowed zeroing of any loop offset voltage; this was readjusted as necessary to accommodate slight temperature drifts. The theoretical behavior of the complete circuit is very similar to that of Figure 7.5, with a variable input-pair current and fixed output-pair current.

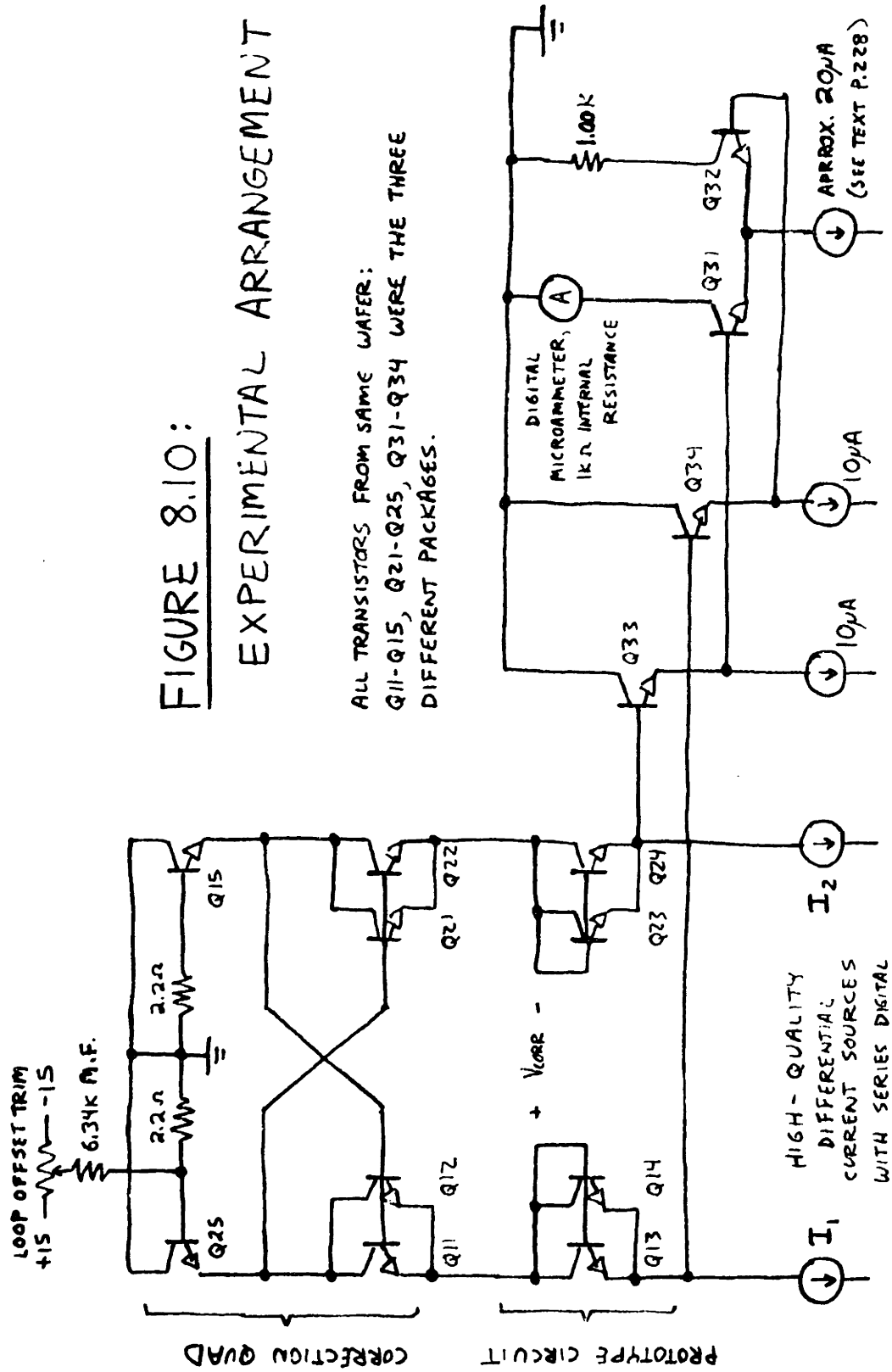


FIGURE 8.10:

EXPERIMENTAL ARRANGEMENT

ALL TRANSISTORS FROM SAME WAFER:
Q11-Q15, Q21-Q25, Q31-Q34 WERE THE THREE DIFFERENT PACKAGES.

The input section was driven differentially, with an average current $I_A = (I_1+I_2)/2$. The emitter followers operated with fixed emitter currents of 10 microamps each, while the output pair (Q31, Q32) were fed from an emitter current source adjusted to give a total collector current of 20 microamps at zero signal. In the tests, the current I_A varied from 10 to 1000 microamps. The correction quad could be electrically removed by grounding the collectors of the input transistor pair (Q13-Q14 and Q23-Q24) and applying the offset trim voltage to the corresponding bases.

Figure 8.11 shows the error in the input-output transfer curve at I_A = 1000 microamps, with and without the correction quad in place. As in section 7.5.2, this error is defined as

$$\% \text{ error} = (X_o - X_i) \times 100\%$$

$$= \left(\frac{I_C \text{ of Q31} - 10\mu A}{10\mu A} - \frac{I_1 - I_2}{2I_A} \right) \times 100\% \qquad (8.28)$$

A practical variable-gain amplifier would be operated over some fraction of its input range to avoid limiting effects. A figure of 70% of full scale was chosen to derive typical distortion results. Some of the observed error over this region may be attributed to gain inaccuracy; indeed, this is to be expected because of base current effects in the input section.

It is therefore convenient to separate linear and nonlinear error terms. Over the range of 70% of full scale, the observed error curves were fitted to a straight-line approximation through

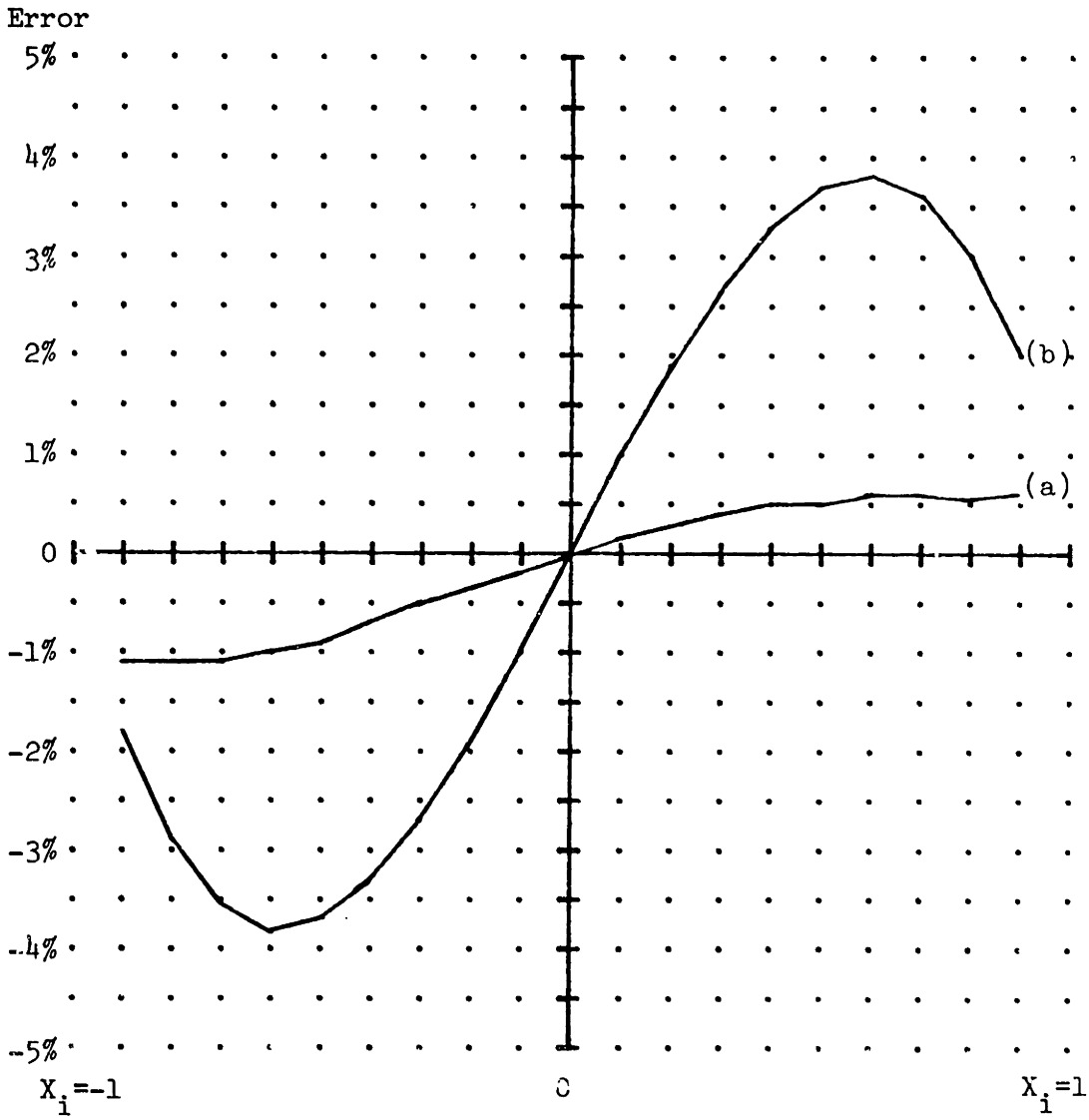


Figure 8.11 -- Measured input-output error in a breadboard variable-gain amplifier with input-pair average current (I_A) of one milliamp.

(a) Correction quad in place; (b) correction quad removed

the origin, using a least-square-error fit. This linear estimate was then subtracted off, leaving a measure of input-output nonlinearity as a function of input.

Table 8.2 summarizes the results. "RMS error" is the raw figure, from both linear and nonlinear components. The effect of the correction quad is most noticeable in the high-current case, which is to be expected. At low currents the quad actually degrades the linearity of the amplifier, although the difference is on the order of the measurement accuracy (±0.1% error).

Worst-case error was reduced by about a factor of five through the use of a correction quad. In view of the discrete-component breadboard limitations, a better result would be expected in an all-monolithic realization.

It is also possible to place the prototype transistor pair above the correction quad, or between the upper and lower pairs of the quad. However, both of these modifications, when applied to the amplifier of Figure 8.10, leave the emitters of a cross-coupled transistor pair in the lowermost position. the load impedance seen by the collectors of this pair is maximized, and this has a destabilizing effect. The resulting circuit may be compared to an emitter-coupled multivibrator [8.9]. Indeed, experiments with such configurations demonstrated oscillations at up to 100 Megahertz when a few picofarads were placed directly across the emitters. Moreover, it is desirable to keep the smallest-area transistor pair at the top of the circuit. This permits the collector voltage to be raised, if necessary, to guarantee reverse bias of the collector-base junctions

	I_A=1000uA	I_A=100uA	I_A=10uA
RMS ERROR:			
Uncorrected Amplifier	2.91%	.22%	.40%
With Correction Quad	.60%	.17%	.30%
RMS NONLINEARITY:			
Uncorrected Amplifier	.60%	.07%	.05%
With Correction Quad	.15%	.17%	.18%

Table 8.2 -- Summary of errors measured in the circuit of Figure 8.10.

"RMS Error" includes both nonlinearity and gain error effects.

in this pair. Saturation problems are more likely in this pair than in the others, since it has the highest current density.

8.7 Conclusions

A method for correcting for troublesome log-conformance error effects was proposed, and some practical implications examined, in this chapter. It is unlikely that such a method will remove all limits to precision in bipolar-transistor circuits that exploit logarithmic V_{BE} behavior, since a host of problems remains, including thermal effects (localized and inter-transistor heating), mismatches, series collector resistance and base currents. Indeed, an LCE correction circuit whose complexity is too great may introduce more DC errors than it removes, to say nothing of noise and dynamic aberrations. The proposed technique simply provides another little subcircuit that the designer can invoke when appropriate.

Notable features of the technique are:

- Applicability to any "logarithmic" semiconductor device.
- The ability to compensate for log-conformance error terms that do not depend linearly on current, if a large enough current-density ratio is used.
- Design flexibility; the correction circuit may use area-ratioed active devices, or it may use identical devices and electronic current ratioing.
- Applicability to circuits other than translinear networks, such as log converters, or even unbalanced

circuits like bandgap references -- any group of V_{BE} drops.

A simple correction quad is limited to improving the current-mode accuracy of a prototype circuit by at most about $3/\beta$ if no care is taken to compensate for base-current effects in the quad. Experimental results, using a fairly pessimistic breadboard configuration, demonstrated the basic viability of the technique. Issues of noise and dynamics need further investigation, as this treatment has been concerned mainly with static error sources and their mitigation. Correction circuits with current-density ratios greater than two (i.e. c factors less than 0.5) are another area that warrants practical experimentation.

A Brief Editorial*

The conventional practice in writing a thesis is to look through the literature of the field, or past theses, and to copy a convenient reference format. This is an excellent way to propagate bad habits. Certain peculiarities, like abbreviating the titles of even the obscure journals, come to be mimicked in theses even though their original justifications do not apply there (if indeed they ever applied). Moreover, some scholars in engineering mercilessly cite things like "private communications" (most of which belong in text or footnotes), and works they know to be inaccessible (which don't belong at all). A reference list is supposed to be a guide to sources.

Having been exposed to this many times in reading the literature, I resolved not to do the same to my readers. In the list that follows, bibliographic numerics have the format VOLUME(NUMBER): PAGINATION, DATE. Presentation of author and title information follows the style that is standard, or becoming standard, in many fields, with electrical engineering a notable exception.

Actually, the purpose of this harangue is not so much to gripe as to recommend a particualarly useful guide. A Handbook for Scholars, by Mary-Claire van Leunen (Alfred A. Knopf paperback, 1979), covers a motley collection of topics relevant to technical writing. I have learned a great deal from this little volume.

---

*"The soapbox phenomenon. Given any slim excuse, 99.624 percent of all persons will sound off. Given no excuse at all, 99.608 percent of them will do so.

-- van Leunen

Chapter 3

[3.1] B. A. Blesser. Audio dynamic range compression for minimum perceived distortion. IEEE Transactions on Audio and Electroacoustics AU-17(1):22-32, March 1969.

[3.2] Barry Blesser and Allan R. Kent. Analysis of a feedback-controlled limiter using a logarithmic measuring scale. IEEE Transactions on Audio and Electroacoustics AU-16(4): 481-485, December 1968.

[3.3] B. Blesser and J. M. Kates. Digital processing in audio signals. In Alan V. Oppenheim, editor, Applications of Digital Signal Processing, pages 29-116, Prentice-Hall, 1978.

[3.4] F. Doorenbosch and Y. Goinga. Integrable, wideband, automatic volume control (A.V.C.) using Pythagoras' law for amplitude detection. Electronics Letters 12(16): 418-420, 5 August 1976.

[3.5] B. M. Oliver. Automatic volume control as a feedback problem. Proceedings of the IRE 36(4):466-473, April 1948.

[3.6] W. K. Victor and M. H. Brockman. The application of linear servo theory to the design of AGC loops. Proceedings of the IRE 48(2):234-238, February 1960.

[3.7] R. McFee and S. Dick. Reducing nonlinear distortion due to automatic gain-control circuits. Electronics Letters 1(4):102-104, June 1965.

[3.8] V. G. Vol'pyan. Synthesis of the characteristics of an automatic control amplifier. Telecommunications and Radio Engineering (part 2) 1968(9):148-150, September 1968. English translation of Radiotekhnika 23(9):103-105, September 1968.

[3.9] Richard Smith Hughes. Design automatic gain control loops the easy way. EDN 123-128, 5 October 1978.

[3.10] Jack Porter. AGC loop design using control system theory. RF Design 3(6):27-32, June 1980.

[3.11] M. L. Zelenz. On AGC transient-response behavior in cascaded amplifiers. Proceedings of the IEEE 58(7): 1130-1131, July 1970.

[3.12] Kenneth K. Clarke and Donald T. Hess. Communication Circuits: Analysis and Design, chapter 6. Addison-Wesley, 1971.

[3.13] James K. Roberge. Operational Amplifiers: Theory and Practice, chapter 12. Wiley, 1975.

[3.14] C. Afuso and S. Ishikawa. A temperature-insensitive voltage-controlled variable inductor. IEEE Journal of Solid-State Circuits SC-10(6):540-542, December 1975.

[3.15] Robert G. Sparkes and Adel S. Sedra. Programmable active filters. IEEE Journal of Solid-State Circuits SC-8(1): 93-95, February 1973.

[3.16] Kiyoshi Fukahori. Monolithic tunable notch filter. In 1981 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pages 130-131, 19 February 1981.

[3.17] C. R. Ryan. Applications of a four-quadrant multiplier. IEEE Journal of Solid-State Circuits SC-5(1):45-48, February 1970.

[3.18] A. K. Bhattacharyya and T. R. Viswanathan. Sinusoidal current-controlled oscillator. Electronics Letters 12(23):614-615, 11 November 1976.

Chapter 4

[4.1] Granino A. Korn and Theresa M. Korn. Electronic Analog and Hybrid Computers, chapter 6. McGraw-Hill, second edition 1972.

[4.2] Kenneth K. Clarke and Donald T. Hess. Communications Circuits: Analysis and Design, chapter 7. Addison-Wesley, 1971.

[4.3] Chu-Sun Yen. Distortion analysis of base-driven differential pairs. In Record of the Fourteenth Asilomar Conference on Circuits, Systems and Computers, pages 303-306. Pacific Grove, California, 17-19 November 1980. Published by the IEEE Computer Society.

[4.4] R. H. Wilkinson. A method of generating functions of several variables using analog diode logic. IEEE Transactions on Electronic Computers EC-12(2):112-129, April 1963.

[4.5] Alpha Industries, Inc. PIN diode basics. Application note 80200, July 1976.

[4.6] Hewlett-Packard Co. The PIN diode. Application note 904, 15 February 1966.

[4.7] Hewlett-Packard Co. Applications of PIN diodes. Application note 922, 1971.

[4.8] B. M. Potts. PIN diode weight circuits. Technical note #1979-26, MIT Lincoln Laboratory, 28 December 1979.

[4.9] Hewlett-Packard Co. High-performance PIN attenuator for low cost AGC applications. Application note 936, October 1971.

[4.10] William Ross Aiken and Charles F. Swisher. An improved automatic level control device. Journal of the Audio Engineering Society 16(4):447-449, October 1968.

[4.11] Allen P. Edwards. Precise, convenient analysis of modulated signals. Hewlett-Packard Journal 30(11):3-18, November 1979. Page 14 in particular.

[4.12] V. Székely. New type of thermal-function I.C.: the four-quadrant multiplier. Electronics Letters 12(15):372-373, 22 July 1976.

[4.13] V. N. Kornev. Automatic gain control by thermistors in transistor amplifiers. Soviet Journal of Instrumentation and Control 1967(2):64-65, February 1967. English translation of Pribori i Systemy Upravleniya 1967(2):55, February 1967.

[4.14] James G. Holt. A two-quadrant analog multiplier integrated circuit. IEEE Journal of Solid-State Circuits SC-8(6):434-439, December 1973.

[4.15] B. H. Brown. Simple analogue divider and multiplier. Electronics Letters 1(7):206, September 1965.

[4.16] Q. V. Davis. Analogue divider. Electronics Letters 2(11): 425-426, November 1966.

[4.17] N. Racoveanu. Analogue multiplier using p-n junction capacitance. Electronics Letters 3(5):217-218, May 1967.

[4.18] Kamil Kraus. Switched V-F converter linearizes analog multiplier. Electronics 53(15):133, 3 July 1980.

Chapter 5

[5.1] S. M. Sze (also references [5.13], [5.15]). Physics of Semiconductor Devices, chapter 7. Wiley, 1969.

[5.2] Richard S. Muller and Theodore I. Kamins (also references [5.14], [5.16]). Device Electronics for Integrated Circuits, chapter 3. Wiley, 1977.

[5.3] W. Shockley. A unipolar "field-effect" transistor. Proceedings of the IRE 40(11):1365-1376, November 1952.

[5.4] I. Richer and R. D. Middlebrook. Power-law nature of field-effect transistor experimental characteristics. Proceedings of the IEEE 51(8):1145-1146, August 1963.

[5.5] R. D. Middlebrook. A simple derivation of field-effect transistor characteristics. Proceedings of the IEEE 51(8): 1146-1147, August 1963.

[5.6] I. Richer. Basic limits on the properties of field-effect transistors. Solid-State Electronics 6(5):539-542, September 1963.

[5.7] R. D. Middlebrook and I. Richer. Limits on the power-law exponent for field-effect transistor transfer characteristics. Solid-State Electronics 6(5):542-544, September 1963.

[5.8] R. R. Bockemuehl. Analysis of field effect transistors with arbitrary charge distribution. IEEE Transactions on Electron Devices ED-10(1):31-34, January 1963.

[5.9] R. S. C. Cobbold and F. N. Trofimenkoff. Theory and application of the field-effect transistor, part 1: Theory and d. c. characteristics. Proceedings of the IEE (UK) 111(12):1981-1992, December 1964.

[5.10] F. N. Trofimenkoff. Field-dependent mobility analysis of the field-effect transistor. Proceedings of the IEEE 53(11):1765-1766, November 1965.

[5.11] Richard S. C. Cobbold (also reference [5.19]). Theory and Applications of Field-Effect Tranasistors, chapter 3. Wiley, 1970.

[5.12] H. P. von Ow. Reducing distortion in controlled attenuators using FET. Proceedings of the IEEE 56(10):1718-1719, October 1968.

[5.13] Sze [5.1], chapter 10.

[5.14] Muller and Kamins [5.2], chapter 8.

[5.15] Sze [5.1], chapter 9.

[5.16] Muller and Kamins [5.2], chapter 7.

[5.17] A. S. Grove (also reference [5.18]). Physics and Technology of Semiconductor Devices, page 325. Wiley, 1967.

[5.18] Grove, chapter 12.

[5.19] Cobbold [5.11], chapter 6.

[5.20] Patrick William Bosshart. A monolithic analog correlator using charge-coupled devices, chapter 4. Master's thesis, Massachusetts Institute of Technology, February 1976.

[5.21] F. M. Klaassen. A MOS model for computer-aided design. Philips Research Reports 31(1):71-83, 1976.

[5.22] F. M. Klaassen. Review of physical models for MOS transistors. In Proceedings of the NATO Advanced Study Institute on Process and Device Modelling for Integrated Circuit Design, pages 541-571, Louvain-la-Neuve, Belgium, 19-29 July 1977. Published by Noordhoff International, Leyden, The Netherlands and Reading, Massachusetts.

[5.23] William Gosling. Voltage controlled attenuators using field effect transistors. IEEE Transactions on Audio AU-13(5): 112-120, September/October 1965.

[5.24] Erich Hoelken. FET feedback configurations provide accurate logic functions. Electronic Design 12(26):52-54, 21 December 1964.

[5.25] Gerald W. Beene. An analog multiplier. Proceedings of the IEEE 55(7):1206, July 1967.

[5.26] Alan B. Grebene. Analog Integrated Circuit Design, section 8.8. Van Nostrand Reinhold, 1972. Reprinted by Robert E. Krieger Co., Huntington, New York, 1978.

[5.27] James S. Sherwin. FETs as voltage-controlled resistors. Solid State Design 6(8):12-14, August 1965.

[5.28] Siliconix, Inc. FETs as voltage-controlled resistors. Application note AN73-1, 1973.

[5.29] Howard N. Leighton. An optimized gain-control configuration using the field-effect transistor. IEEE Journal of Solid-State Circuits SC-3(4):441-447, December 1968.

[5.30] T. B. Martin. Circuit applications of the field-effect transistor. Semiconductor Products 5:30-38, March 1962. Cited in Bilotti [5.34].

[5.31] W. Y. Elliott, Jr. Field effect transistor as a linear variable resistance. IBM Technical Disclosure Bulletin 7(1):111, June 1964.

[5.32] Carl David Todd. FETs as voltage-variable resistors. Electronic Design 13(19):66-69, 13 September 1965.

[5.33] H. P. von Ow. A simple circuit using FETs for fading audio signals in and out exponentially in time. Proceedings of the IEEE 57(2):222-224, February 1969.

[5.34] Alberto Bilotti. Operation of a MOS transistor as a variable resistor. Proceedings of the IEEE 54(8):1093-1094, August 1966.

[5.35] Benjamin Shore. Pocket-size analog computer divides and multiplies. Electronics 40(26):66-67, 25 December 1967.

[5.36] A. K. H. Miller. Temperature-compensated analogue multiplier. Electronics Letters 7(18):539-540, 9 September 1971.

[5.37] F. N. Trofimenkoff and R. E. Smallwood. Four-quadrant JFET multipliers. IEEE Journal of Solid-State Circuits SC-12(3): 316-319, June 1977.

[5.38] W. H. Highleyman and E. S. Jacob. An analog multiplier using two field effect transistors. IRE Transactions on Communications Systems CS-10(3):311-317, September 1962.

[5.39] S. Østerfjells. Analog multiplier with field effect transistors. Proceedings of the IEEE 53(5):521, May 1965.

[5.40] Patrick Bosshart. An integrated analog correlator using charge-coupled devices. In 1976 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pages 198-199, 20 February 1976.

[5.41] J. Mavor, J. W. Arthur, and P. B. Denyer. Analogue C.C.D. correlator using monolithic M.O.S.T. multipliers. Electronics Letters 13(12):373-374, 9 June 1977. Note figures 3a, 3b exchanged.

[5.42] D. V. McCaughan, J. C. White, and J. R. Hill. A novel C.C.D. multiplying input technique. Electronics Letters 14(6):165-166, 16 March 1978.

[5.43] M. M. Abu-Zeid and H. Groendijk. Field-effect-transistor-bridge multiplier-divider. Electronics Letters 8(24): 591-592, 30 November 1972.

[5.44] M. M. Abu-Zeid, H. Groendijk, and A. Willemse. Temperature-compensated F.E.T. multiplier. Electronics Letters 4(16): 324-325, 9 August 1968.

[5.45] I. C. Hutcheon and D. J. Puddefoot. New solid-state electronic multiplier-divider. Proceedings of the IEE (UK) 112(8): 1523-1531, August 1965.

[5.46] Hiroaki Ikeda. MOST variable resistor and its application to an AGC amplifier. IEEE Journal of Solid-State Circuits SC-5(1):43-45, February 1970.

[5.47] I. Házman. Four-quadrant multiplier using M.O.S.F.E.T. differential amplifiers. Electronics Letters 8(3):63-65, 10 February 1972.

[5.48] David M. Miller and Robert G. Meyer. Nonlinearity and cross modulation in field-effect transistors. IEEE Journal of Solid-State Circuits SC-6(4):244-250, August 1971.

[5.49] Jakob S. Vogel. Nonlinear distortion and mixing processes in field-effect transistors. Proceedings of the IEEE 55(12):2109-2116, December 1967.

[5.50] Kenneth K. Clarke and Donald T. Hess. Communication Circuits: Analysis and Design, chapter 7. Addison-Wesley, 1971.

[5.51] Ed Oxner. FETs in balanced mixers. Siliconix, Inc., application note AN72-1, 1972.

[5.52] O. A. Seriki and R. W. Newcomb. Direct-coupled MOS squaring circuit. IEEE Journal of Solid-State Circuits SC-14(4): 766-768, August 1979.

[5.53] Eric Vittoz and Jean Fellrath. CMOS analog integrated circuits based on weak inversion operation. IEEE Journal of Solid-State Circuits SC-12(3):224-231, June 1977.

[5.54] A. J. Strachan and W. Godfrey Townsend. A proposed voltage variable resistor MOST. Proceedings of the IEEE 56(11): 2096-2097, November 1968.

[5.55] Jun-Ichi Nishizawa, Takeshi Terasaki, and Jiro Shibata. Field-effect transistor versus analog transistor (static induction transistor). IEEE Transactions on Electron Devices ED-22(4):185-197, April 1975.

[5.56] E. S. Yang. Four-terminal field-effect transistors for amplitude modulation. IEEE Journal of Solid-State Circuits SC-4(2):84-86, April 1969.

Chapter 6

[6.1] B[arrie] Gilbert. Translinear circuits: a proposed classification. Electronics Letters 11(1):14-16, 9 January 1975.

[6.2] B. L. Hart. Translinear circuit principle: a reformulation. Electronics Letters 15(24):801-803, 22 November 1979.

[6.3] Barrie Gilbert. A new wide-band amplifier technique. IEEE Journal of Solid-State Circuits SC-3(4):353-365, December 1968.

[6.4] Barrie Gilbert. A precise four-quadrant multiplier with subnanosecond response. IEEE Journal of Solid-State Circuits SC-3(4):365-373, December 1968.

[6.5] C. R. Hurtig. Constant-resistance AGC attenuator for transistor amplifiers. IRE Transactions on Circuit Theory CT-2(2):191-196, June 1955.

[6.6] Herbert L. Kahn. Multiplication and division using silicon diodes. The Review of Scientific Instruments 33(2):235-238, February 1962. See also comments, 34(5):592-593, May 1963.

[6.7] S. A. Romano, Jr. Swamp out distortion in wide-range AGC. Electronic Design 14(8):72-74, 12 April 1966.

[6.8] S. Deb and J. K. Sen. Transistorized electronic analog multiplier. The Review of Scientific Instruments 32(2): 189-192, February 1961.

[6.9] R. H. Frater. Accurate wideband multiplier-square-law detector. The Review of Scientific Instruments 35(7): 810-813, July 1964.

[6.10] R. R. A. Morton. A simple D.C. to 10 Mc/s analogue multiplier. Journal of Scientific Instruments 43:165-168, March 1966.

[6.11] P. Kundu. Transistorized electronic analog multiplier. The Review of Scientific Instruments 34(5):593-594, May 1963.

[6.12] P. Kundu and S. Banerji. Transitorized multiplier and divider and its applications. IEEE Transactions on Electronic Computers EC-13(3):288-295, June 1964.

[6.13] F. S. Davidson and H. N. Leighton. Linearity improvement for bipolar transistors used as electrically variable resistors. IBM Technical Disclosure Bulletin 9(7):915, December 1966.

[6.14] A. Grasselli and R. Stefanelli. New analogue multiplier-divider. Electronics Letters 2(1):2-3, January 1966. Corrections appear in 2(5):188.

[6.15] William Mack. Wideband transconductance amplifiers. Master's thesis, University of California at Berkeley, June 1979.

[6.16] B[arrie] Gilbert. A DC-500 MHz amplifier/multiplier principle. In 1968 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pages 114-115, 16 February 1968. A considerably expanded version appears in references [6.3] and [6.4].

[6.17] K. G. Schlotzhauer and T. R. Viswanathan. New bipolar analogue multiplier. Electronics Letters 8(16):425-427, 10 August 1972.

[6.18] Walter G. Jung. Get gain control of 80 to 100 dB. Electronic Design 22(13):94-99, 21 June 1974.

[6.19] Craig C. Todd. A monolithic analog compandor. IEEE Journal of Solid-State Circuits SC-11(6):754-762, December 1976.

[6.20] Douglas R. Curtis. A monolithic voltage-controlled amplifier employing log-antilog techniques. Journal of the Audio Engineering Society 24(2):93-102, March 1976.

[6.21] D. J. Hamilton and K. B. Finch. A single-ended current gain cell with AGC, low offset voltage, and large dynamic range. IEEE Journal of Solid-State Circuits SC-12(3):322-323, June 1977.

[6.22] Y. Z. Bahnas, G. G. Bloodworth, A. Brunnschweiler. The noise properties of the linearized transconductance multiplier. IEEE Journal of Solid-State Circuits SC-12(5):580-584, October 1977.

[6.23] Willy M. C. Sansen and Robert G. Meyer. Distortion in bipolar transistor variable-gain amplifiers IEEE Journal of Solid-State Circuits SC-8(4):275-282, August 1973.

[6.24] Willy M. C. Sansen and Robert G. Meyer. An integrated wide band variable-gain amplifier with maximum dynamic range. IEEE Journal of Solid-State Circuits SC-9(4):159-166, August 1974.

[6.25] W. Richard Davis and James E. Solomon. A high-performance monolithic IF amplifier incorporating electronic gain control. IEEE Journal of Solid-State Circuits SC-3(4):408-416, December 1968.

[6.26] William J. McCalla. An integrated IF amplifier. IEEE Journal of Solid-State Circuits SC-8(6):440-447, December 1973.

[6.27] E. A. Faulkner and J. B. Grimbleby. New type of analogue multiplier with wide dynamic range. Electronics Letters 8(6):145-146, 23 March 1972.

[6.28] Chu-Sun Yen. Distortion in emitter-driven variable-gain pairs. IEEE Journal of Solid-State Circuits SC-14(4): 771-773, August 1979.

[6.29] Chu-Sun Yen. High-frequency distortion in emitter-driven variable-gain pairs. IEEE Journal of Solid-State Circuits SC-15(3):375-377, June 1980.

[6.30] Barrie Gilbert and Peter Holloway. A wideband two-quadrant multiplier. In 1980 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pages 200-201, 15 February 1980.

[6.31] Barrie Gilbert. A high-accuracy analog multiplier. In 1974 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pages 128-129, 14 February 1974. Expanded version appears in reference [6.32].

[6.32] Barrie Gilbert. A high-performance monolithic multiplier using active feedback. IEEE Journal of Solid-State Circuits SC-9(6):364-373, December 1974.

[6.33] Pietro G. Erratico and Raimondo Caprio. An integrated expandor circuit. IEEE Journal of Solid-State Circuits SC-11(6):762-772, December 1976.

[6.34] J. N. Wright. Transconductance multipliers for weight circuits in an adaptive nulling system. Technical note #1979-11, MIT Lincoln Laboratory, 1 June 1979.

[6.35] H[arro] Brüggemann. New feedback-stabilised analogue multiplier. Electronics Letters 5(5):86-88, 6 March 1969. Corrections appear in 5(25):670.

[6.36] Harro Brüggemann. Feedback-stabilized four-quadrant analog multiplier. IEEE Journal of Solid-State Circuits SC-5(4): 150-159, August 1970.

[6.37] E. A. Faulkner and J. B. Grimbleby. High-performance analogue multiplier based on the characteristics of bipolar transistors. Electronics Letters 6(12):379-380, 11 June 1970.

[6.38] T. R. Viswanathan and G. S. Deep. Simple 4-quadrant analogue multiplier. Electronics Letters 6(23):725-726, 12 November 1970.

[6.39] Edward W. Scratchely. Single-ended-input single-ended-output four-quadrant analog multiplier. IEEE Journal of Solid-State Circuits SC-6(6):394-395, December 1971. See also comments by Gilbert, SC-7(5):434.

[6.40] Nicholas C. Voulgaris and Edward S. Yang. Linear applications of a p-n-p-n tetrode. IEEE Journal of Solid-State Circuits SC-5(4):146-150, August 1970.

[6.41] B[arrie] Gilbert. New planar distributed devices based on a domain principle. In 1971 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pages 166-167, 19 February 1971.

[6.42] Barrie Gilbert. A new technique for analog multiplication. IEEE Journal of Solid-State Circuits SC-10(6):437-447, December 1975.

[6.43] Jim Smith. A second-generation carrier domain four-quadrant multiplier. IEEE Journal of Solid-State Circuits SC-10(6):448-457, December 1974.

Chapter 7

[7.1] Chih-Tang Sah. Effect of surface recombination and channel on p-n junction and transistor characteristics. IRE Transactions on Electron Devices ED-9(1):94-108, January 1962.

[7.2] J. F. Gibbons and H. S. Horn. A circuit with logarithmic transfer response over 9 decades. In 1963 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pages 38-39, 20 February 1963.

[7.3] B. L. Hart. Remarks on the emission coefficient of a bipolar transistor. Proceedings of the IEEE 69(5):665-666, May 1981.

[7.4] Ian Getreu. Modeling the Bipolar Transistor, section 2. Tektronix, Inc., Beaverton, Oregon, 1976.

[7.5] William L. Paterson. Multiplication and logarithmic conversion by operational amplifier-transistor circuits. The Review of Scientific Instruments 34(12):1311-1316, December 1963.

[7.6] S. M. Sze. Physics of Semiconductor Devices, chapter 3. Wiley, 1969.

[7.7] A. S. Grove (also reference [7.12]). Physics and Technology of Semiconductor Devices, chapter 6. Wiley, 1967.

[7.8] E. A. Faulkner and M. J. Buckingham. Modified theory of the current/voltage relation in silicon p-n junctions. Electronics Letters 4(17):359-360, 23 August 1968.

[7.9] Charles A. Bittmann, Garth H. Wilson, Ronald J. Whittier, and Robert K. Waits. Technology for the design of low-power circuits. IEEE Journal of Solid-State Circuits SC-5(1):29-37, February 1970.

[7.10] Paul E. Gray, David DeWitt, A. R. Boothroyd, and James F. Gibbons. Physical Electronics and Circuit Models of Transistors, chapter 8. Wiley, 1964. Volume 2 of the SEEC series.

[7.11] J. R. Hauser. The effects of distributed base potential on emitter-current injection density and effective base resistance for stripe transistor geometries. IEEE Transactions on Electron Devices ED-11(5):238-242, May 1964.

[7.12] Grove [7.7], chapter 7.

[7.13] D. J. Roulston, S. G. Chamberlain, and J. Sehgal. High-level asymptotic variation of transistor base resistance and current gain. Electronics Letters 7(15):438-440, 29 July 1971.

[7.14] J. Cornu. Analytical model for p-n junctions under high-injection conditions. Electronics Letters 7(18):509-510, 9 September 1971.

[7.15] M. J. Gay. Base resistance in silicon planar transistors. Electronics Letters 1(7):212-213, September 1965. Corrections appear in 1(9):271.

[7.16] Albert Willemsen and Niko Bel. Low base resistance integrated circuit transistor. IEEE Journal of Solid-State Circuits SC-15(2):245-246, April 1980.

[7.17] Barrie Gilbert. A new wide-band amplifier technique. IEEE Journal of Solid-State Circuits SC-3(4):353-365, December 1968.

[7.18] Barrie Gilbert. A precise four-quadrant multiplier with subnanosecond response. IEEE Journal of Solid-State Circuits SC-3(4):365-373, December 1968.

[7.19] Willy M. C. Sansen and Robert G. Meyer. Distortion in bipolar transistor variable-gain amplifiers. IEEE Journal of Solid-State Circuits SC-8(4):275-282, August 1973.

[7.20] Barrie Gilbert and Peter Holloway. A wideband two-quadrant analog multiplier. In 1980 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pages 200-201, 15 February 1980.

[7.21] Chu-Sun Yen. Distortion in emitter-driven variable-gain pairs. IEEE Journal of Solid-State Circuits SC-14(4): 771-773, August 1979.

[7.22] Chu-Sun Yen. High-frequency distortion in emitter-driven variable-gain pairs. IEEE Journal of Solid-State Circuits SC-15(3):375-377, June 1980.

[7.23] G. Rey and J. P. Babary. Sur deux méthodes de mesure de la résistance de base. Electronics Letters 2(12):462-464, December 1966. In French.

[7.24] Willy M. C. Sansen and Robert G. Meyer. Characterization and measurement of the base and emitter resistances of bipolar transistors. IEEE Journal of Solid-State Circuits SC-7(6):492-498, December 1972.

[7.25] John Choma, Jr. Error minimization in the measurement of bipolar collector and emitter resistances. IEEE Journal of Solid-State Circuits SC-11(2):318-322, April 1976.

[7.26] J. S. T. Huang. Rapid determination of emitter- and collector-bulk resistances. IEEE Journal of Solid-State Circuits SC-11(2):343-344, April 1976.

[7.27] A. Zafer Incecik. Computer-aided determination of emitter and collector resistances of integrated bipolar transistors. IEEE Journal of Solid-State Circuits SC-14(6):1108-1111, December 1979.

Chapter 8

[8.1] Dennis R. Morgan. Get the most out of log amplifiers by understanding the error sources. EDN 19(2):52-55, 20 January 1973.

[8.2] M. McGinn. All n-p-n 4-quadrant analogue multiplier. Electronics Letters 9(20):472-473, 4 October 1973.

[8.3] James C. Schmoock. Monolithic multiplier/divider. In 1979 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pages 230-231, 16 February 1979.

[8.4] Johan H. Huijsing and J. Anne van Steenwijk. A monolithic analog exponential converter. IEEE Journal of Solid-State Circuits SC-15(2):162-168, April 1980.

[8.5] Raimondo Caprio. Precision differential voltage-current convertor. Electronics Letters 9(6):147-148, 22 March 1973.

[8.6] Robert A. Blauschild. An open-loop programmable amplifier with a maximum gain-bandwidth product of 1GHz. In 1981 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pages 190-191, 19 February 1981.

[8.7] Carl O. Bozler and Gary D. Alley. Fabrication and numerical simulation of the permeable base transistor. IEEE Transactions on Electron Devices ED-27(6):1128-1141, June 1980.

[8.8] B[arrie] Gilbert. Translinear circuits: a proposed classification. Electronics Letters 11(1):14-16, 9 January 1975. See Figure 2A.

[8.9] Paul R. Gray and Robert G. Meyer. Analysis and Design of Analog Integrated Circuits, chapter 10. Wiley, 1977.

ACKNOWLEDGEMENTS

My thanks to the Fannie and John Hertz Foundation, whose financial support made this whole project possible. Thanks also to the RF Techniques and Communications Processors Groups at the MIT Lincoln Laboratory, for giving me access to laboratory facilities for the experimental part of this work. To the technical staff of Analog Devices Semiconductor, Wilmington, Massachusetts, and especially Lew Counts and Paul Brokaw, for providing experimental components, and for advice and encouragement.

Professor James Roberge, my thesis advisor, supervised the project and was the source of considerable wisdom along the way. He also convinced me, though he may not believe it, to restrict the scope of the work. I am grateful to the members of the Digital Signal Processing Group at MIT, for their hospitality in spite of my incongruous research topic, and to Professor Alan Oppenheim, for his active interest in the progress of my thesis. Bob McCharles and Jim Eaton, of Hewlett-Packard Laboratories, helped out in the later portion of the work by making available, with a wink and a nod, some nifty computer resources.

I am particularly in debt to Barrie Gilbert, formerly with Tektronix and now a major constituent of Analog Devices Northwest Labs. Barrie fostered my interest in this subject matter when I worked under him at Tektronix in the summer of 1978. During the early planning of this thesis he steered me toward some important

issues and shared his knowledge of the translinear gain-control literature. His enthusiasm and scholarly style have been inspiring.

The typing of this report was a transcontinental effort. The expert services of Joanne Klys (Cambridge) and Paula Conant (Berkeley) converted my dingy draft to a magnificent monograph.

AUTHOR'S BIOGRAPHY

Max W. Hauser was born in Berkeley, California, in 1956. He received the B.S. degree in electrical engineering and computer science from the University of California, Berkeley, in 1979.

In the summer of 1972 he worked for the University of California, Department of Physiology/Anatomy, designing biomedical electronic equipment. From 1975 to 1977 he was with the Electronics Research Laboratory of the University, designing laboratory instrumentation for plasma physics. In the summer of 1978 he was at Tektronix, Incorporated, Beaverton, Oregon, working on analog IC design. In 1979 he was with the MIT Lincoln Laboratory, Lexington, Massachusetts working on technology for spread-spectrum communications. In the summer of 1981 he worked at the Hewlett-Packard Laboratories, Palo Alto, California, on MOS analog IC design. He has served as a consultant to other firms.

He has contributed articles to the trade press on analog circuits and other topics. Since 1979 he has been a Hertz Foundation Fellow, first at the Massachusetts Institute of Technology and subsequently at the University of California, Berkeley, where he is now studying for a doctoral degree. In 1979 he was named outstanding electrical engineering student in the United States by the Eta Kappa Nu Association. He is a member of Eta Kappa Nu, Tau Beta Pi, and Phi Beta Kappa.