Analog Computation in the Digital Age

By James M. Bryant

The Science Museum in South Kensington, London, was founded during Queen Victoria's reign, and although it is imaginatively run and well worth a visit, it is scarcely renowned for its humor. But while I was in their Computer Science Area recently, I laughed so much, and so loudly, that I barely escaped arrest for disorderly behavior. The cause of my uncontrolled mirth was a glass case labeled "Obsolete Analog Computing Technology," containing, among other things, an AD534 analog multiplier. ADI has been making this device for more than 30 years, and it continues to generate substantial sales revenue. In fact, there are a number of operations where analog processing has clear advantages over digital, and this article will discuss some of them.

Analog Multipliers

A multiplier is a device having two input ports and an output port. The signal at the output is the product of the two input signals. If both input and output signals are voltages, the transfer characteristic is the product of the two voltages divided by a scaling factor, K, which has the dimension of voltage (see Figure 1).

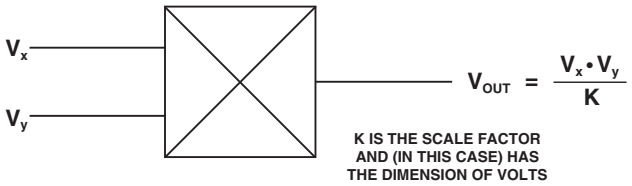


Figure 1. The basic multiplier function.

From a mathematical point of view, multiplication is a "four-quadrant" operation, which is to say that both inputs may be either positive or negative, as may be the output. Some of the circuits used to produce electronic multipliers, however, are limited to signals of one polarity. If both signals must be unipolar, we have a "single-quadrant" multiplier, and the output will also be unipolar. If one of the signals is unipolar, but the other may have either polarity, the multiplier is a "two-quadrant" multiplier, and the output may have either polarity (and is "bipolar").

Type	V_x	V_y	V_{OUT}
One Quadrant	Unipolar	Unipolar	Unipolar
Two Quadrant	Bipolar	Unipolar	Bipolar
Four Quadrant	Bipolar	Bipolar	Bipolar

Figure 2. One-, two-, and four-quadrant multipliers.

The circuitry used to produce one- and two-quadrant multipliers may be simpler than that required for four-quadrant multipliers, and since there are many applications where full four-quadrant multiplication is not required, it is not uncommon to find accurate devices that work only in one or two quadrants. An example is the AD539, a wideband, dual, two-quadrant multiplier that has a single unipolar V_y input with a relatively limited bandwidth of 5 MHz, and two bipolar V_x inputs, one per multiplier, with bandwidths of 60 MHz. A block diagram of the AD539 is shown in Figure 3.

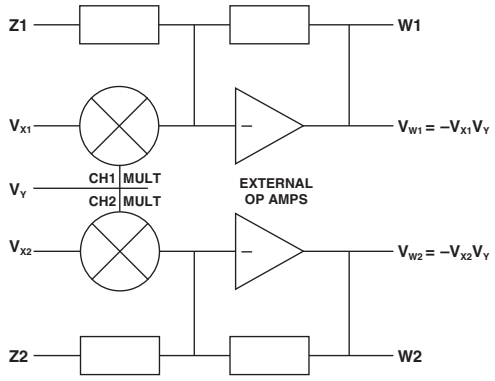


Figure 3. AD539 functional block diagram.

The simplest electronic multipliers use logarithmic amplifiers. The computation relies on the fact that the antilog of the sum of the logs of two numbers is the product of those numbers (see Figure 4).

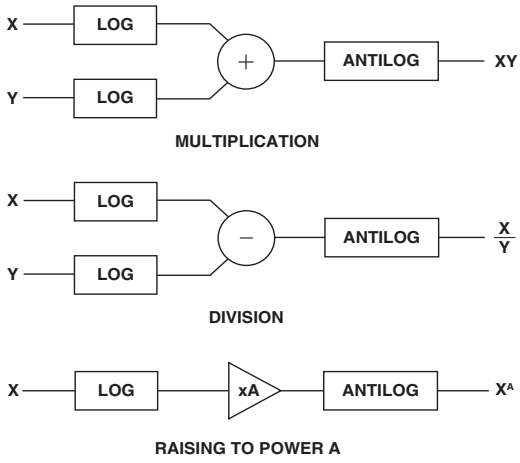


Figure 4. Computation with log and antilog circuits.

The disadvantages of this type of multiplication are the very limited bandwidth and single-quadrant operation. A far better type of multiplier uses the "Gilbert Cell." This structure was invented by Barrie Gilbert in the late 1960s (see References 1 and 2).

There is a linear relationship between the collector current of a silicon junction transistor and its transconductance (gain), which is given by

$$dI_c/dV_{be} = qI_c/kT$$

where:

I_c = the collector current

V_{be} = the base-emitter voltage

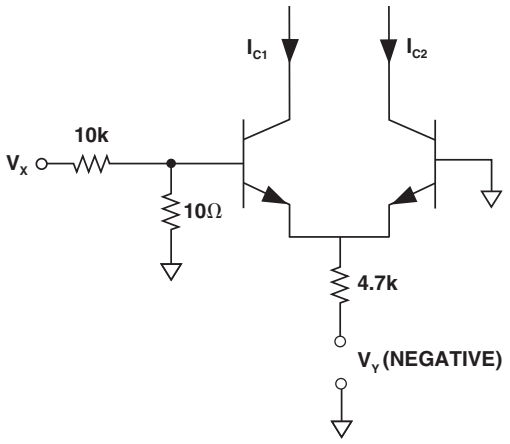
q = the electron charge (1.60219×10^{-19})

k = Boltzmann's constant (1.38062×10^{-23})

T = the absolute temperature

This relationship may be exploited to construct a multiplier with a long-tailed pair of silicon transistors, as shown in Figure 5.

This is a rather poor multiplier because (1) the Y input is offset by the V_{be}, which changes nonlinearly with V_y; (2) the X input is nonlinear as a result of the exponential relationship between I_c and V_{be}; and (3) the scale factor varies with temperature.



$$I_{C1} - I_{C2} = \Delta I_C = \frac{q}{kT}\left(\frac{V_Y + V_{be}}{4.7 \times 10^3}\right)\left(\frac{10}{10{,}010}\right)V_X$$

$$= 8.3 \times 10^{-6}\,(V_Y + 0.6)\,V_X @ 25°C$$

Figure 5.

Gilbert realized that this circuit could be linearized and made temperature stable by working with currents rather than voltages and by exploiting the logarithmic I_c/V_{be} properties of transistors (see Figure 6). The X input to the Gilbert Cell takes the form of a differential current, and the Y input is a unipolar current. The differential X currents flow in two diode-connected transistors, and the logarithmic voltages compensate for the exponential V_{be}/I_c relationship. Furthermore, the q/kT scale factors cancel. This gives the Gilbert Cell the linear transfer function

$$\Delta I_c = \frac{\Delta I_x I_y}{I_x}$$

$$\Delta I_c = I_{c1} - I_{c2}$$

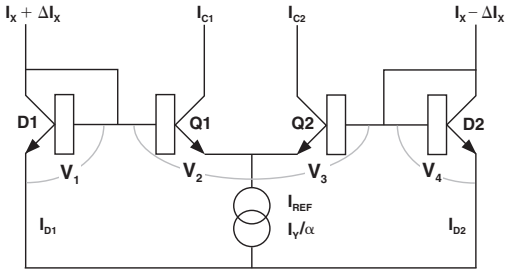


Figure 6. The Gilbert Cell: a linear two-quadrant multiplier.

As it stands, the Gilbert Cell has three inconvenient features: (1) its X input is a differential current; (2) its output is a differential current; and (3) its Y input is a unipolar current—so the cell is only a two-quadrant multiplier.

By cross-coupling two such cells and using two voltage-to-current converters (as shown in Figure 7), we can convert the basic architecture to a four-quadrant device with voltage inputs, such as the AD534. At low and medium frequencies, a subtractor amplifier may be used to convert the differential current at the output to a voltage. Because of its voltage output architecture, the bandwidth of the AD534 is only about 1 MHz, although the AD835, a later version using a much faster dielectrically isolated complementary bipolar process (XFCB), has a bandwidth of 250 MHz.

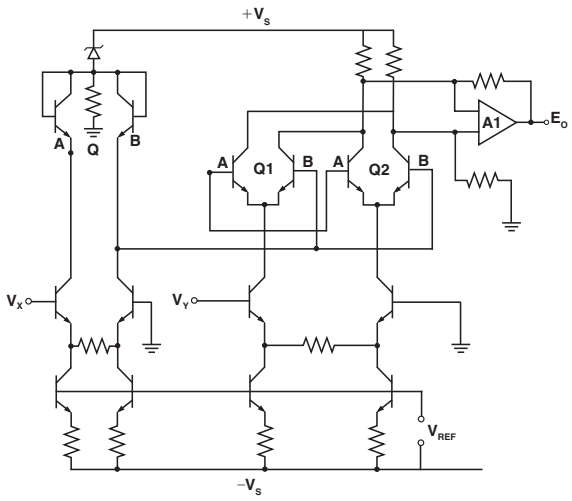


Figure 7. AD534.

In Figure 7, Q1A and Q1B and Q2A and Q2B form the two core long-tailed pairs of the two Gilbert Cells, while Q3A and Q3B are the linearizing transistors for both cells. In Figure 7, there is an operational amplifier acting as a differential current to single-ended voltage converter, but for higher speed applications, the cross-coupled collectors of Q1 and Q2 form a differential open collector current output (as in the AD834 500 MHz multiplier).

The translinear multiplier relies on the matching of a number of transistors and currents. This is easily accomplished on a monolithic chip. Even the best IC processes have some residual errors, however, and these show up as four dc error terms in such multipliers (see Figure 8). In early Gilbert Cell multipliers, these errors had to be trimmed by means of resistors and potentiometers external to the chip, which was somewhat inconvenient. With modern analog processes, which permit the laser trimming of SiCr thin film resistors on

the chip itself, it is possible to trim these errors during manufacture so that the final device has very high accuracy. Internal trimming has the additional advantage that it does not reduce the high frequency performance, as may be the case with external trim potentiometers.

Because the internal structure of the translinear multiplier is necessarily differential, the inputs are usually differential as well (after all, if a single-ended input is required, it is not hard to ground one of the inputs). This is convenient not only in allowing common-mode signals to be rejected, but also in permitting more complex computations to be performed.

Multipliers can be placed in the feedback loop of op amps to form several useful functions. Figure 10 illustrates the basic principle of analog computation that a function generator in a negative feedback loop computes the inverse function (provided, of course, that the function is monotonic over the range of operations).

Figure 11 shows a multiplier and an op amp configured as a divider in both inverting and noninverting mode.

TRIMMABLE ERRORS IN MULTIPLIERS

- **X-Input Offset Voltage: Y Feedthrough**
- **Y-Input Offest Voltage: X Feedthrough**
- **Z-Input (Output Amplifier) Voltage Offset: DC Output Offset Voltage**
- **Resistor Mismatch: Gain Error**

Figure 8.

KEY FEATURES OF THE TRANSLINEAR MULTIPLIER

- **High Accuracy: Better than 0.1% Possible**
- **Wide Bandwidth:**
  - **Over 250MHz Voltage Output**
  - **Over 500MHz Current Output**
- **Simplicity, Low Cost, and Ease of Use**

Figure 9.

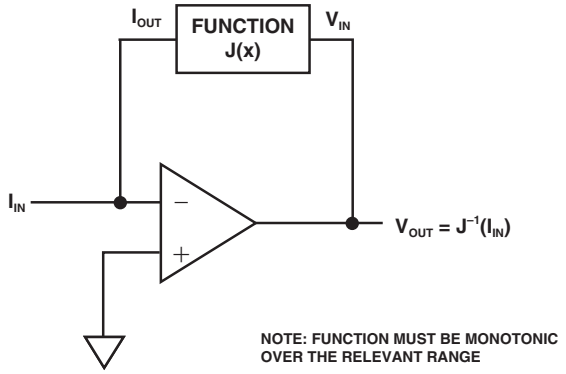


Figure 10. Inverse function generated by function generator in feedback.

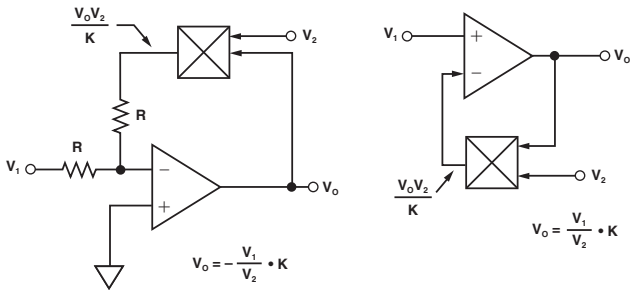


Figure 11. Multipliers and op amps used to divide.

Modulators and Mixers

A modulator (also called a mixer when it is used as a frequency changer) is closely related to a multiplier. The output of a multiplier is the instantaneous product of its inputs. The output of a modulator is the instantaneous product of a signal on one of its inputs (known as the signal input) and the sign of the signal on the other input (known as the carrier input). A modulator may be modeled as an amplifier whose gain is switched positive and negative by the output of a comparator on its carrier input (as in the case of the AD630 balanced modulator)—or as a multiplier with a high gain limiting amplifier between the carrier output and one of its ports (see Figure 12). Both architectures have been used to produce modulators, but the switched amplifier version, although potentially very accurate, tends to be rather slow. Most high speed integrated circuit modulators consist of the translinear multiplier (based on the Gilbert Cell) with a limiting amplifier in the carrier path.

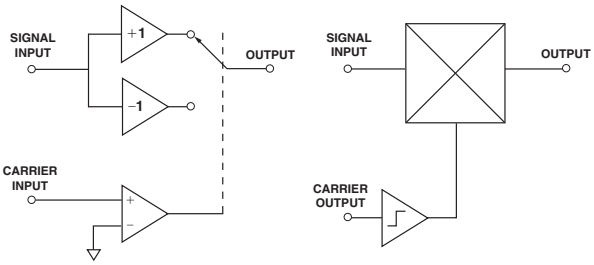


Figure 12. The modulator function (two ways of looking at it).

If two periodic waveforms, $A_m\cos(\omega_m t)$ and $A_c\cos(\omega_c t)$, are applied to the inputs of a multiplier (with a scale factor of 1 V for simplicity of analysis), then the output will be given by

$$V_o(t) = 1/2A_mA_c[\cos(\omega_m + \omega_c)t + \cos(\omega_m - \omega_c)t]$$

This signal contains signals at the sum and difference frequencies, but not at the original frequencies. Some RF engineers also call these the upper and lower sidebands. There is a 6 dB loss in an ideal modulator. Note that using the cosine formula rather than the sine formula makes the equations easier to manipulate because cos(a) = cos(–a) (which makes sign unimportant during simplification), and because cos(0) = 1, so that for dc signals (when $a\omega t = 0$), cos($a\omega t$) is equal to unity.

When we say that the original frequencies are not present in the output of a modulator, we make the assumption that the modulator is perfectly balanced—i.e., that neither its signal port nor its carrier port has any offset. In practice, both ports will have some offset, and so there will be some signal and carrier leakage. Trimming offset on the inputs of a modulator will reduce the leakages, but there will always be untrimmable residual leakages, which are due to coupling by stray capacitance and to nonlinearities in the core, rather than to offsets.

This “sum and difference mixer” is the function that we expect of modulators. However, if we use a linear multiplier as a modulator, we find that any noise or modulation on the carrier input appears in the

CAUSES OF SIGNAL AND CARRIER LEAKAGE IN MODULATORS

- **Offset on the signal port causes carrier leakage.**
- **Offset on the carrier port causes signal leakage.**
- **Even when all offsets have been trimmed out, there is residual signal and carrier leakage caused by stray capacitance and core nonlinearities.**

Figure 13.

output signal. If we replace the simple multiplier with a modulator, any amplitude variation on the carrier input disappears.

Like an analog multiplier, a modulator multiplies two signals. But unlike analog multipliers, the multiplication is not linear. Instead, the signal input is "chopped" by the local oscillator carrier signal that alternates between +1 and –1 in sign (i.e, a square wave). This is equivalent to passing the carrier signal $A_c\cos(\omega_c t)$ through a comparator, or limiting amplifier. The square wave with a frequency of $\omega_c t$ has the form represented by the Fourier series of odd harmonics:

$$K[\cos(\omega_c t) - 1/3\cos(3\omega_c t) + 1/5\cos(5\omega_c t) - 1/7\cos(7\omega_c t) + ...]$$

The sum of the series: [+1, –1/3, +1/5, –1/7 +...] is $\pi/4$. Therefore, the value of K is $4/\pi$, such that a balanced modulator acts as a unity gain amplifier when a positive dc signal is applied to its carrier input.

Therefore, if a modulator is driven by a signal $A_m\cos(\omega_m t)$ and a carrier $\cos(\omega ct)$ (the carrier amplitude is unimportant provided it is great enough to drive the limiting amplifier), then the output will be the product of the signal and the squared carrier above.

The final output is given by

$$\begin{aligned} 2A_m/\pi[&\cos(\omega_m + \omega_c)t + \cos(\omega_m - \omega_c)t \\ &-1/3\{\cos(\omega_m + 3\omega_c)t + \cos(\omega_m - 3\omega_c)t\} \\ &+1/5\{\cos(\omega_m + 5\omega_c)t + \cos(\omega_m - 5\omega_c)t\} \\ &-1/7\{\cos(\omega_m + 7\omega_c)t + \cos(\omega_m - 7\omega_c)t\} + ...] \end{aligned}$$

This output contains sum and difference frequencies of the signal and carrier, and of the signal and each of the odd harmonics of the carrier (in the ideal, perfectly balanced modulator, products of even harmonics are not present—in real modulators, which have residual offsets on their carrier ports, low level, even harmonic products are also present; just how low their level depends on the size of the offset). In most applications, a filter is used to remove the products of the higher harmonics so that, effectively, the modulator does behave like a multiplier. (In analyzing the above expressions, we must remember that $\cos(A) = \cos(-A)$, so that $\cos(\omega_m - N\omega_c)t = \cos(N\omega_c - \omega_m)t$, so we do not have to worry about "negative frequencies.") After filtering, the modulator output is given by

$$2A_m/\pi[\cos(\omega_m + \omega_c)t + \cos(\omega_m - \omega_c)t]$$

Because of the $2/\pi$ term, a modulator has a minimum 3.92 dB insertion loss, in the absence of any gain. (The AD831 has a gain of 3.92 dB to provide unity gain from RF to IF.)

The most obvious application of a modulator is a mixer or frequency changer. If we apply an input signal at F1 and a carrier at F2 to a modulator, we find that the output contains signals at the sum and difference frequencies as shown in Figure 14. This applies even if the signal is a modulated signal containing a number of frequency components.

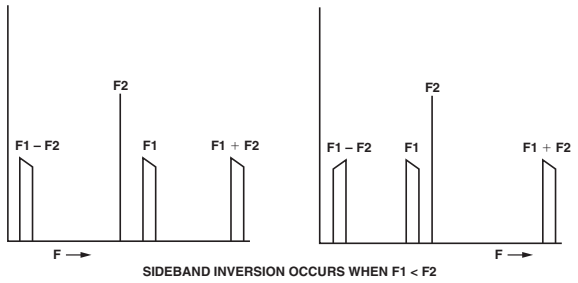


Figure 14. The modulator as a mixer (frequency changer).

As we have mentioned above, we cannot have "negative frequencies," and so if F1 – F2 is negative, what we actually see is a frequency of F2 – F1. If F1 is a complex signal containing a number of components, however, we find that if the carrier frequency, F2, is less than F1, then the sidebands are inverted in both the sum and difference products. However, if F2 is greater than F1, then in the difference product, the sidebands are inverted, as shown in the diagram.

The *mixer,* or frequency changer, is a key component in most radio receivers. While it is inappropriate to go into a detailed discussion of receiver design in this section, it is perhaps useful to point out two important features of modulators for use in receivers. These are *noise* and *strong signal performance* (see References 3 and 4).

Suppose that we have a mixer with a noisy carrier channel that causes the carrier frequency, F2, to spread out on either side of its center as shown in Figure 15. If we are receiving a small wanted signal, F1, then we shall see a small IF output from the mixer at F2 – F1. If, however, there is a strong unwanted signal at F3, then the product, F4 – F3, of F3 and that part of the broadband carrier noise indicated by F4 on the diagram, will also fall at the IF and, being larger than the wanted IF component, will swamp it. This phenomenon is known as *reciprocal mixing* and can cause severe limitation on the dynamic range of a receiver. While it is probably more commonly caused by noise or spurious synthesizer sidebands in the oscillator driving the modulator carrier port, it is common for it to be caused by noise in the modulator itself, and it should certainly be considered when choosing a mixer for a radio receiver (see Reference 5).

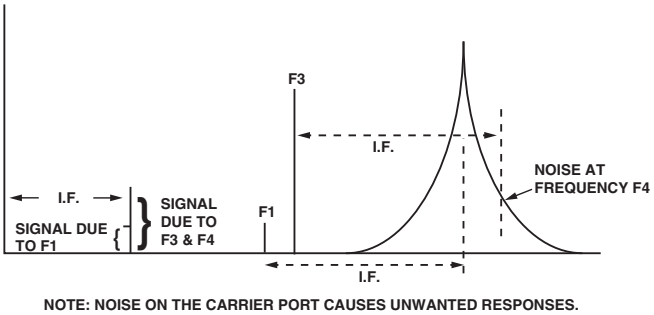


Figure 15. Reciprocal mixing.

In the past, the sensitivity of a radio receiver has been one of its most important features. Today, while sensitivity is still important, the behavior of the receiver in the presence of strong signals is equally important. The characteristic chosen as a measure of a

mixer's performance in this respect is its *third-order intermodulation* performance. The key specification is the *third-order intercept point.*

Consider a nonlinear amplifier with two large input signals, at F1 and F2, as shown in Figure 16. The nonlinearity gives rise to additional output components at F1 + F2 and F1 – F2: these are known as *second-order intermodulation products.* These second-order products mix with the original signals and produce *third-order intermodulation products* at frequencies 2F1 – F2 and 2F2 – F1.

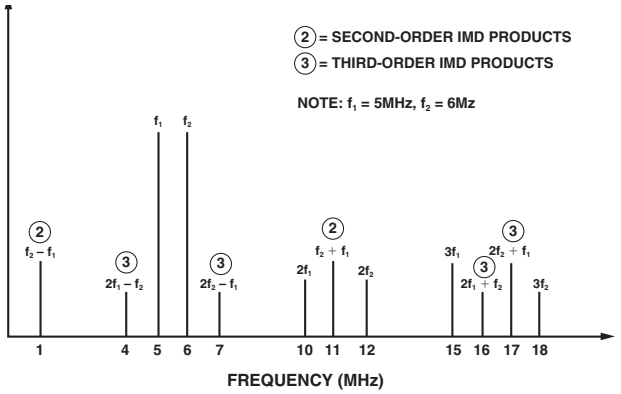


Figure 16. Second- and third-order intermodulation products for f_1 = 5 MHz, f_2 = 6 MHz.

The third-order intermodulation products are a major nuisance in radio reception, especially in channelized systems, because they fall close to the signals causing them. As an example, consider a receiver monitoring a frequency of 145.5000 MHz. In Europe, this frequency is the calling frequency of the 2-meter amateur band. Working channels in this band are separated by 12.5 kHz. Suppose that there are two transmitters working at 145.475 MHz and 145.4875 MHz, respectively. The third-order IMD products of these two frequencies fall at 145.4625 MHz and 145.500 MHz. If the receiver is liable to third-order IMD, it will respond to the third-order products—which it itself produces—and appear to be receiving a signal at 145.500 MHz.

It is impossible to design an amplifier or mixer that is unaffected by third-order intermodulation. All that can be done is to minimize the problem. The third-order intercept point mentioned above is the parameter that measures how susceptible a device is to third-order IMD.

If we plot the input vs. the output amplitudes of an amplifier on a log/log (dB/dB) scale as shown in Figure 17, we obtain a straight line slope of unity. At a certain input level, the device saturates, and the output ceases to rise. A measure of this saturation point is known as the *1 dB compression point.* If we plot the level of the second-order IMD products in the output against the level of a two-tone input on the same axes, we obtain a straight line with a slope of 2. This line also ceases to rise when it reaches some limit. If, however, we extend the two straight lines past their limiting values, they will eventually cross. The value of the power in one of the two-tone inputs at this "intercept point" is the *second-order intercept point* of the device or system being measured. The level of the third-order IMD products can also be plotted as a function of input, and the slope of the straight line is 3. The intersection of this line with the extension of the unity-slope line is known as the *third-order intercept point.*

In mixers for receivers, values of the third-order intercept point can vary from –15 dBm to over +45 dBm. For a radio working with an external antenna of reasonable size, any value below 0 dBm is generally considered poor, and good performance requires values of

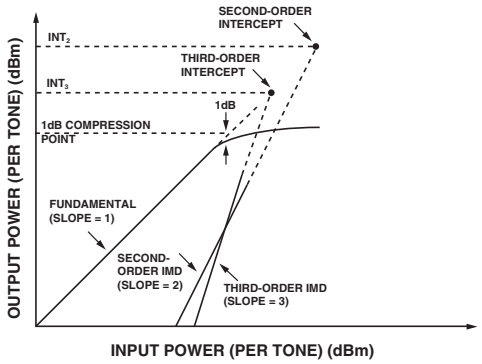


Figure 17. Intercept points, gain compression, and IMD.

at least +15 dBm and, preferably, more—radios with small internal antennas are less demanding. The second-order intercept point may also be specified, but it is generally of less concern.

The AD831 is a low distortion, wide dynamic range, monolithic mixer for use in such applications as RF to IF downconversion in HF and VHF receivers, the second mixer in digital mobile radio base stations, direct-to-baseband conversion, quadrature modulation and demodulation, and doppler-frequency shift detection in ultrasound imaging applications. The mixer includes a local oscillator driver and a low noise output amplifier. The AD831 provides a +24 dBm third-order intercept point for –10 dBm local oscillator power, thus improving system performance and reducing system cost, compared to passive mixers, by eliminating the need for a high power local oscillator driver and its associated shielding and isolation problems. A simplified block diagram of the AD831 is shown in Figure 18, and key specifications in Figure 19. (Since this article was originally

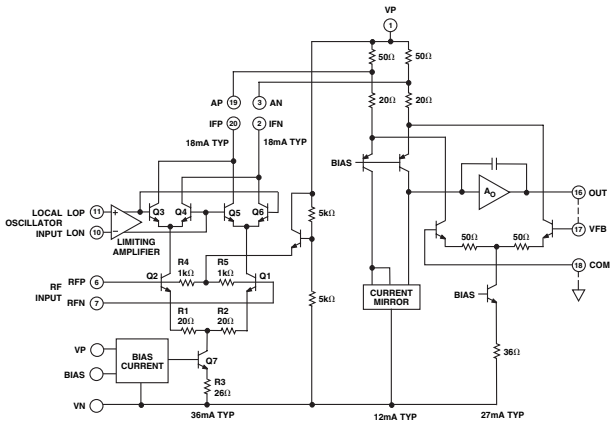


Figure 18. The AD831 low distortion mixer.

AD831 MIXER KEY FEATURES

- **Double-balanced mixer**
- **Low Distortion:**
  - **+24dBm Third-Order Intercept**
  - **+10dBm Compression Point**
- **Low LO Drive Required: –10dBm**
- **Bandwidth**
  - **500MHz RF and Lo Input Bandwidths**
  - **250MHz Differential Circuit IF Output**
  - **DC to >200MHz Single-Ended Voltage IF Output**

Figure 19.

written, a number of more advanced mixers have been developed at Analog Devices. Their details may be found on our website at *www.analog.com*.)

The basic mixing property of modulators is also used for many operations where dynamic range is far less important. These include frequency synthesis by mixing, frequency changing with fixed level signals, and sideband generation.

One application worth considering is using a modulator as a precision rectifier. If an ac signal is applied to both inputs of a modulator as shown in Figure 20, the instantaneous output will be equal to the input, if the input is positive, and to the inverse of the input (and therefore still positive), if the input is negative. This arrangement, therefore, behaves as a precision rectifier.

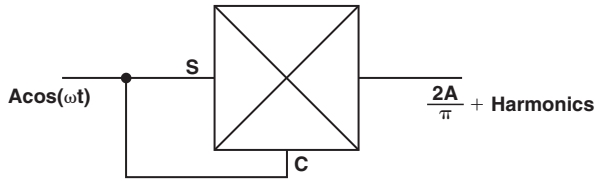


Figure 20. Modulator used as a precision rectifier.

If, instead of applying a signal to both ports of a modulator, a signal is applied to the signal port and a reference signal at the same frequency (but not necessarily the same phase) to the carrier port, then the output will be proportional both to the amplitude of the signal input and the cosine of their phase difference. In this mode, a modulator acts as a phase-sensitive rectifier (see Figure 21).

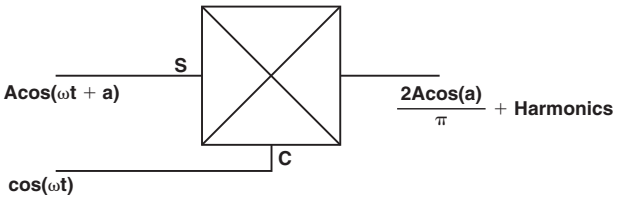


Figure 21. Modulator used as a phase-sensitive rectifier.

RMS-to-DC Converters

The root mean square (rms) is a fundamental measurement of the magnitude of an ac signal. Defined practically, the rms value assigned to the ac signal is the amount of dc required to produce an equivalent amount of heat in the same load. Defined mathematically, the rms value of a voltage is defined as the value obtained by squaring the signal, taking the average, and then taking the square root. The averaging time must be sufficiently long to allow filtering at the lowest frequencies of operation desired. A complete (if now slightly dated) discussion of rms-to-dc converters can be found in Reference 6, but we will show a few examples of how efficiently analog circuits can perform this function.

The first method, called the *explicit* method, is shown in Figure 22. The input signal is first squared (in the diagram by a multiplier with both inputs driven by the signal). The average value is then taken by using an appropriate filter, and the square root is taken using an op amp with a second squarer in the feedback loop. This circuit has relatively limited dynamic range because the stages following the squarer must handle a signal that varies enormously in amplitude—especially if the signal has a high crest factor. Modern IC rms-to-dc converters of this type have dynamic ranges of up to 60 dB—and will work at frequencies up to several GHz.

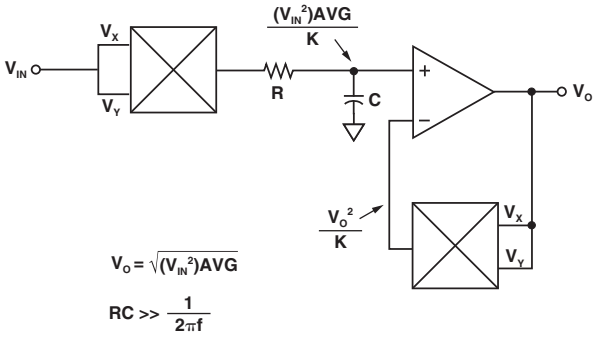


Figure 22. Explicit rms computation.

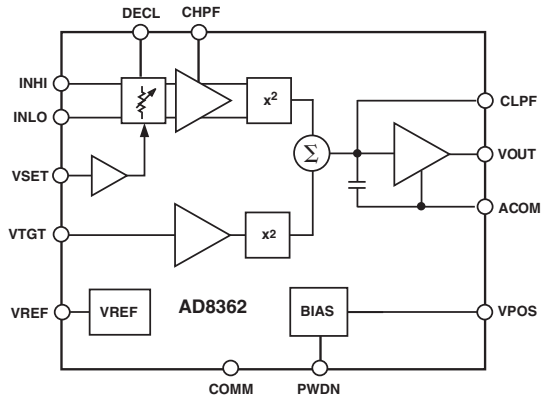


Figure 23. The AD8362 rms-to-dc converter works from VLF (<<50 Hz) to 2.7 GHz and has a dynamic range of 60 dB.

Figure 24 shows the circuit for computing the rms value of a signal using the implicit method. Here, the output is fed back to the direct-divide input of a multiplier such as the AD734. In this circuit, the output of the multiplier varies linearly (instead of as the square) with the rms value of the input. This considerably increases the dynamic range of the implicit circuit as compared to the explicit circuit. The disadvantage of this approach is that it generally has less bandwidth than the explicit computation.

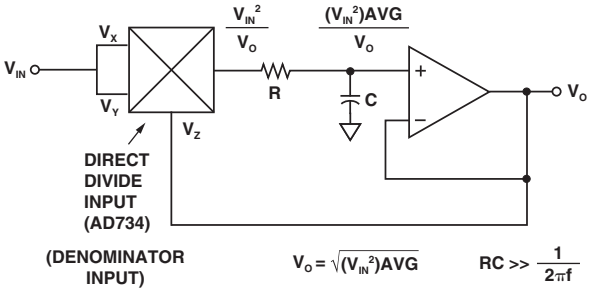


Figure 24. Implicit rms computation.

While it is possible to construct such an rms circuit from an AD734 multiplier/divider, it is far simpler to use a dedicated rms circuit. On a chip, the V_{IN}^2/V_Z circuit may be current driven and need only be one quadrant if the input first passes through an absolute value circuit.

Figure 25 shows a simplified diagram of a typical monolithic rms-to-dc converter, the AD536A. It is subdivided into four major sections: absolute value circuit (active rectifier), squarer/divider, current mirror, and buffer amplifier. The input voltage, V_{IN}, which can be ac or dc, is converted to a unipolar current, I_1, by the absolute value circuit A_1, A_2. I_1 drives one input of the one-quadrant squarer/divider,

which has the transfer function: $I_4 = I_1^2/I_3$. The output current, I_4, of the squarer/divider drives the current mirror through a low-pass filter formed by R1 and an externally connected capacitor, C_{AV}. If the $R1C_{AV}$ time constant is much greater than the longest period of the input signal, then I_4 is effectively averaged. The current mirror returns a current, I_3, which equals $AVG[I_4]$, back to the squarer/divider to complete the implicit rms computation. Thus,

$$I_4 = AVG\,[I_1^2/I_4] = I_1\,\text{rms}$$

The current mirror also produces the output current, IOUT, which equals $2I_4$. IOUT can be used directly or converted to a voltage with R2 and buffered by A4 to provide a low impedance voltage output. The transfer function becomes

$$V_{OUT} = 2R2 \bullet I\ rms = V_{IN}\ rms$$

The dB output is derived from the emitter of Q3, since the voltage at this point is proportional to $-\log V_{IN}$. Emitter follower, Q5, buffers and level shifts this voltage, so that the dB output voltage is zero when the externally supplied emitter current (I_{REF}) to Q5 approximates I_3. However, the gain of the dB circuit has a TC of approximately 3300 ppm/°C and must be temperature-compensated.

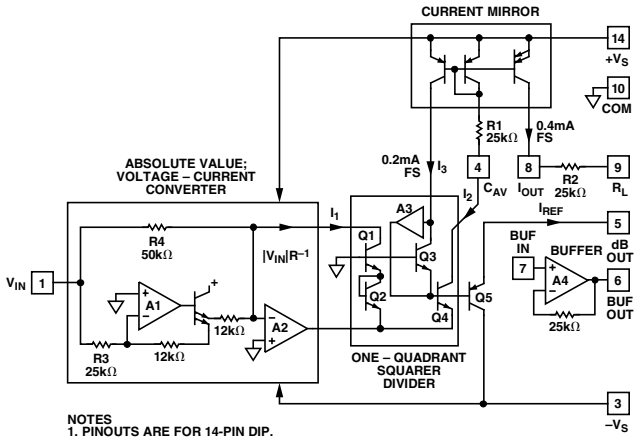


Figure 25. The AD536A monolithic rms-to-dc converter.

Dynamic Range Compression and Logarithmic Amplifiers

A wide dynamic range is often an essential aspect of a signal, something to be preserved at all costs. This is true, for example, in the high quality reproduction of music. However, it is often necessary to compress the signal to a smaller range without significant loss of information. Compression is often used in magnetic recording, where the upper end of the dynamic range is limited by tape saturation and the lower end by the granularity of the medium. In professional noise-reduction systems, compression is "undone" by precisely matched nonlinear expansion during reproduction. Similar techniques are often used in conveying speech over noisy channels, where the performance is more likely to be measured in terms of word intelligibility than audio fidelity. The reciprocal processes of compressing and expanding are implemented using "compandors," and many schemes have been devised to achieve this function.

Logarithmic amps find wide applications where signals having wide dynamic ranges (perhaps greater than 100 dB) must be handled by elements such as ADCs, which may have a more limited dynamic range. Log amps have maximum incremental gain for small signals; the gain decreases in inverse proportion to the magnitude of the input. This permits the amplifier to accept signals with a wide input dynamic range and compress them substantially.

Log amps provide nonlinear dynamic range compression and are used in applications where low harmonic distortion is not a requirement. All types of log amps produce a low dynamic range output without the need to first acquire some measure of the signal amplitude for use in controlling gain.

There is another class of *linear* dynamic range compression systems where the gain of the amplifiers in the signal processing chain is independent of the instantaneous amplitude of the signal, but is controlled by a closed-loop system in such a way as to render the output (that is, the peak, or rms value) essentially constant. The harmonic distortion is relatively low. These systems use what are often called *variable gain amplifiers*. While correct, this lacks precision, because *nonlinear* amplifiers (such as log amps) also exhibit variable gain, but in direct response to the signal magnitude. The term *voltage controlled amplifier* (VCA) is preferred in this context; it clearly describes the way in which the gain control is implemented, while allowing latitude in regard to the actual circuit means used to achieve the function. The gain may be controlled by a *current* within the circuit, but usually a voltage. Analog multipliers may be used as VCAs, but other topologies are usually better. Although such dynamic range compression is quite obviously an example of analog computation, it is, perhaps, less relevant to the main topic of this article and will not be discussed further.

The term "logarithmic amplifier" (generally abbreviated to "log amp") is something of a misnomer, and "logarithmic converter" would be a better description. The conversion of a signal to its equivalent logarithmic value involves a nonlinear operation, the consequences of which can be confusing if not fully understood. It is important to realize that many of the familiar concepts of linear circuits are irrelevant to log amps. For example, the incremental gain of an ideal log amp approaches infinity as the input tends to zero, and a change of offset at the output of a log amp is equivalent to a change of amplitude at its input—not a change of input offset.

For the purposes of simplicity in our initial discussions, we shall assume that both the input and the output of a log amp are voltages, although there is no particular reason why logarithmic current, transimpedance, or transconductance amplifiers should not also be designed.

If we consider the equation $y = \log(x)$, we find that every time x is multiplied by a constant, A, y increases by another constant, $A1$. Thus, if $\log(K) = K1$, then $\log(AK) = K1 + A1$, $\log(A2K) = K1 + 2A1$, and $\log(K/A) = K1 - A1$. This gives a graph as shown in Figure 26, where y is zero when x is unity, where y approaches minus infinity as x approaches zero, and which has no values for x for which y is negative.

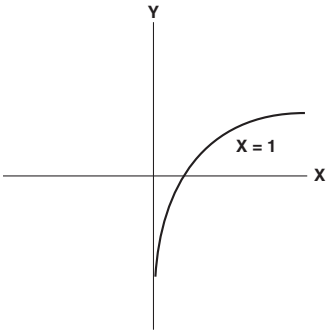


Figure 26. Graph of y = log(x).

On the whole, log amps do not behave in this way. Apart from the difficulties of arranging infinite negative output voltages, such a

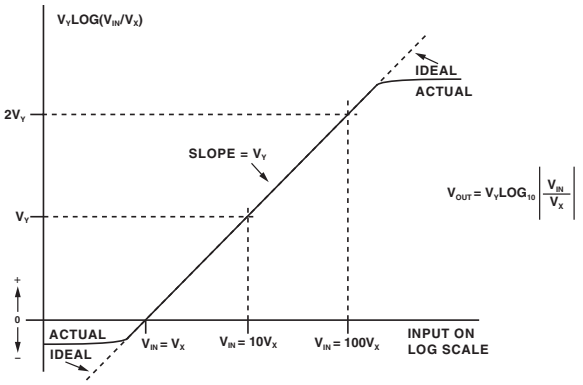


Figure 27. Log amp transfer function.

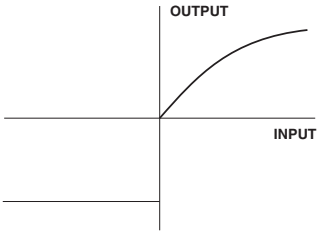


Figure 28. Basic log amp (saturates with negative input).

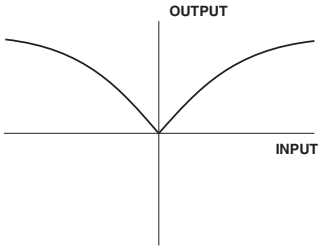


Figure 29. Detecting log amp (output polarity independent of input polarity).

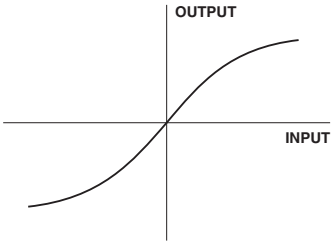


Figure 30. Log video or "true log amp" (symmetrical response to positive or negative signals).

device would not, in fact, be very useful. A log amp must satisfy a transfer function of the form

$$V_{OUT} = V_y \log(V_{IN}/V_x)$$

over some range of input values that may vary from 100:1 (40 dB) to well over 1,000,000:1 (120 dB).

With inputs very close to zero, log amps cease to behave logarithmically, and most then have a linear V_{IN}/V_{out} law. This behavior is often lost in device noise. Noise often limits the dynamic range of a log amp. The constant, V_y, has the dimensions of voltage, because the output is a voltage. The input, V_{IN}, is divided by a voltage, V_x, because the argument of a logarithm must be a simple dimensionless ratio.

A graph of the transfer characteristic of a log amp is shown in Figure 3.2. The scale of the horizontal axis (the input) is logarithmic, and the ideal transfer characteristic is a straight line. When $V_{IN} = V_x$, the logarithm is zero (log 1 = 0). V_x is therefore known as the intercept voltage of the log amp because the graph crosses the horizontal axis at this value of V_{IN}.

The slope of the line is proportional to V_y. When setting scales, logarithms to the base 10 are most often used because this simplifies the relationship to decibel values: when $V_{IN} = 10V_x$, the logarithm has the value of 1, so the output voltage is V_y. When $V_{IN} = 100V_x$, the output is $2V_y$, and so forth. V_y can therefore be viewed either as the "slope voltage" or as the "volts per decade factor."

The logarithm function is indeterminate for negative values of x. Log amps can respond to negative inputs in three different ways: (1) they can give a full-scale negative output as shown in Figure 28; (2) they can give an output that is proportional to the log of the absolute value of the input and that disregards its sign as shown in Figure 29 (this type of log amp can be considered to be a full-wave detector with a logarithmic characteristic, and is often referred to as a *detecting* log amp); or (3) they can give an output that is proportional to the log of the absolute value of the input and has the same sign as the input as shown in Figure 30. This type of log amp can be considered to be a video amp with a logarithmic characteristic, and may be known as a *logarithmic video* (*log video*) amplifier or, sometimes (and misleadingly), a *true log amp*.

There are three basic architectures that may be used to produce log amps: the *basic diode log amp*, the *successive detection log amp*, and the *true log amp*, which is based on cascaded semi-limiting amplifiers.

The voltage across a silicon diode is proportional to the logarithm of the current through it. If a diode is placed in the feedback path of an inverting op amp, the output voltage will be proportional to the log of the input current as shown in Figure 31. In practice, the dynamic range of this configuration is limited to 40 dB to 60 dB because of nonideal diode characteristic, but if the diode is replaced with a diode-connected transistor as shown in Figure 32, the dynamic range can be extended to 120 dB or more. This type of log amp has three disadvantages: (1) both the slope and intercept are temperature dependent; (2) it will only handle unipolar signals; and (3) its bandwidth is both limited and dependent on signal amplitude. Where several such log amps are used on a single chip to produce an analog computer that performs both log and antilog operations, the temperature variation in the log operations is unimportant, since it is compensated by a similar variation in the antilogging. This makes possible the AD538, a monolithic analog computer that can multiply, divide, and raise to powers (see Figure 33A). Where actual logging is required, however, the AD538 and similar circuits require temperature compensation (see Reference 7).

Recently the need for accurate attenuation measurements in fiber-optic systems has driven the development of log amps of this type with integral temperature compensation. The ADL5310, shown in Figure 33B, is an example—it has a current input, allowing a dynamic range of 120 dB at high accuracy (the log conformance is 0.4 dB typical (0.6 dB worst case) over 120 dB from 3 nA to 3 mA). The AD8304 has an even wider dynamic range of 160 dB, but its log conformance is slightly worse over the full range.

The major disadvantage of this type of log amp for high frequency applications, though, is its limited frequency response—which cannot be overcome. However carefully the amplifier is designed, there will always be a residual feedback capacitance, C_c (often known as Miller capacitance), from output to input, which limits the high frequency response (see Figure 32).

What makes this Miller capacitance particularly troublesome is that the impedance of the emitter-base junction is inversely

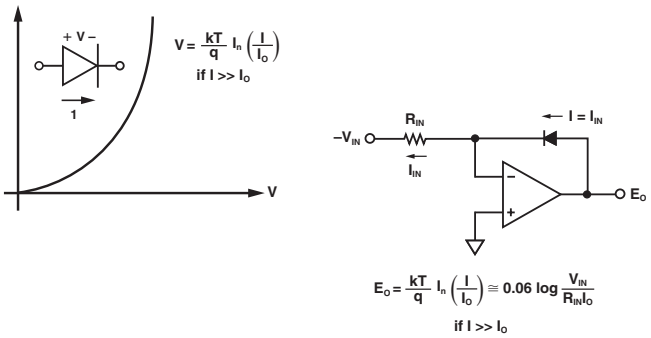


Figure 31. The diode/op amp log amp.

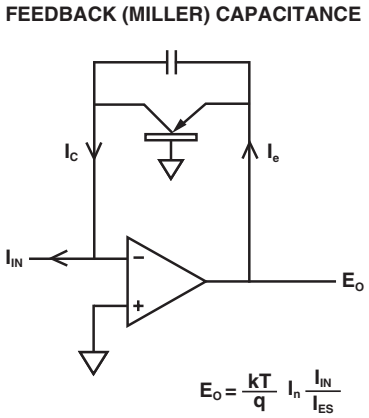


Figure 32. Transistor/op amp.

proportional to the current flowing in it—so that if the log amp has a dynamic range of 1,000,000:1, then its bandwidth will also vary by 1,000,000:1. In practice, the variation is less because other considerations limit the large signal bandwidth, but it is very difficult to make a log amp of this type with a small-signal bandwidth greater than a few hundred kHz.

For high frequency applications, therefore, detecting and "true log" architectures are used. Although these differ in detail, the general principle behind their design is common to both: Instead of one amplifier having a logarithmic characteristic, these designs use a number of similar cascaded linear stages having well-defined large signal behavior.

Consider N cascaded limiting amplifiers, the output of each driving a summing circuit as well as the next stage (Figure 34). If each amplifier has a gain of A dB, the small signal gain of the strip is NA dB. If the input signal is small enough for the last stage not to limit, the output of the summing amplifier will be dominated by the output of the last stage.

As the input signal increases, the last stage will limit. It will now make a fixed contribution to the output of the summing amplifier, but the incremental gain to the summing amplifier will drop to (N – 1)A dB. As the input continues to increase, this stage in turn will limit and make a fixed contribution to the output, and the incremental gain will drop to (N – 2)A dB, and so forth—until the first stage limits, and the output ceases to change with increasing signal input.

The response curve is thus a set of straight lines as shown in Figure 35. The total of these lines, though, is a very good approximation to a logarithmic curve, and in practical cases, is an even better one, because few limiting amplifiers, especially high frequency ones, limit quite as abruptly as this model assumes.

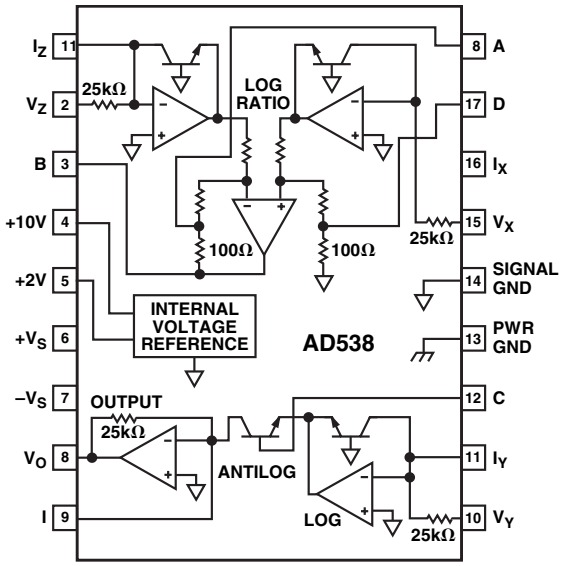


Figure 33A. AD538 log amp simplified diagram.

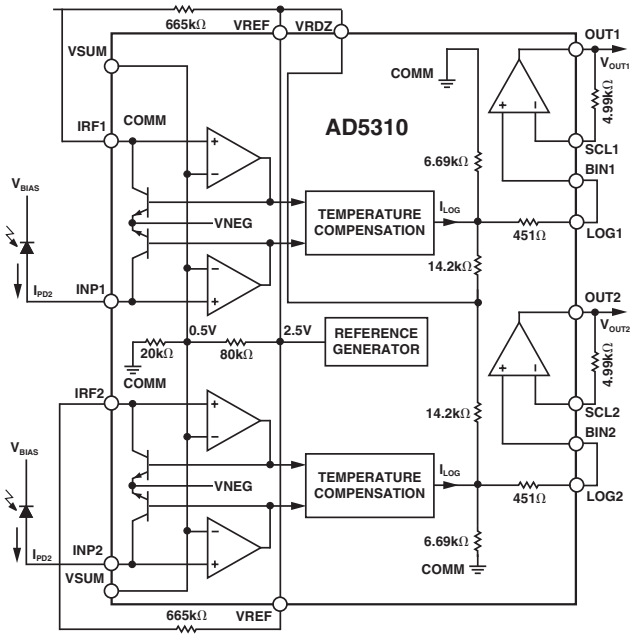


Figure 33B. AD5310 log amp simplified diagram.

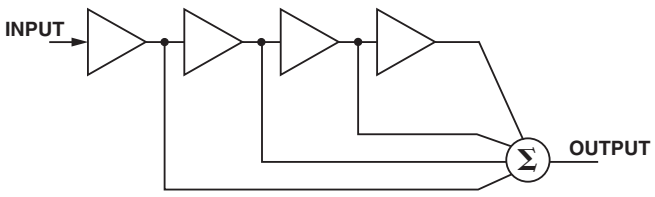


Figure 34. Basic multistage log amp architecture.

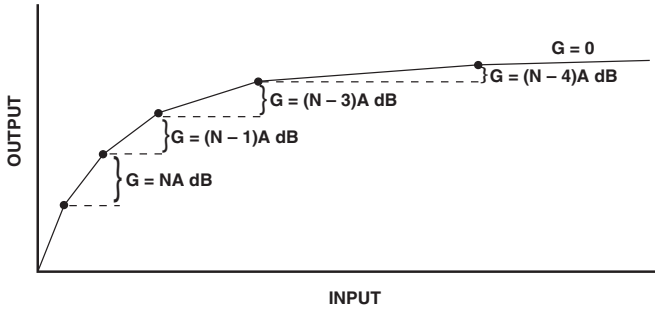


Figure 35. Basic multistage log amp response (unipolar case).

The choice of gain, A, will also affect the log linearity. If the gain is too high, the log approximation will be poor. If it is too low, too many stages will be required to achieve the desired dynamic range. Generally, gains of 10 dB to 12 dB (3× to 4×) are chosen.

This is, of course, an ideal and very general model—it demonstrates the principle, but its practical implementation at very high frequencies is difficult. Assume that there is a delay in each limiting amplifier of t nanoseconds (this delay may also change when the amplifier limits, but let's consider first-order effects!). The signal that passes through all N stages will undergo delay of Nt nanoseconds, while the signal that only passes one stage will be delayed only t nanoseconds. This means that a small signal is delayed by Nt nanoseconds, while a large one is "smeared" and arrives spread over Nt nanoseconds. A nanosecond equals a foot at the speed of light, so such an effect represents a spread in position of Nt feet in the resolution of a radar system—which may be unacceptable in some systems (for most log amp applications, this is not a problem).

A solution is to insert delays in the signal paths to the summing amplifier, but this can become complex. Another solution is to alter the architecture slightly so that instead of limiting gain stages, we have stages with small signal gain of A and large signal (incremental) gain of unity (0 dB). We can model such stages as two parallel amplifiers, a limiting one with gain, and a unity gain buffer, which together feed a summing amplifier as shown in Figure 36.

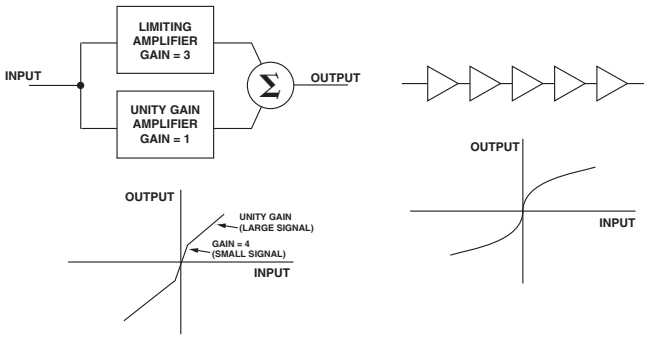


Figure 36. Structure and performace of "true" log amp element and of a log amp formed by several such elements.

Figure 36 shows that such stages, cascaded, form a log amp without the necessity of summing from individual stages. Both the multistage architectures described above are *video* log amplifiers, or *true log* amplifiers, but the most common type of high frequency log amplifier is the *successive detection* log amp architecture shown in Figure 37.

The *successive detection* log amp consists of cascaded limiting stages as described above, but instead of summing their outputs directly, these outputs are applied to detectors, and the detector

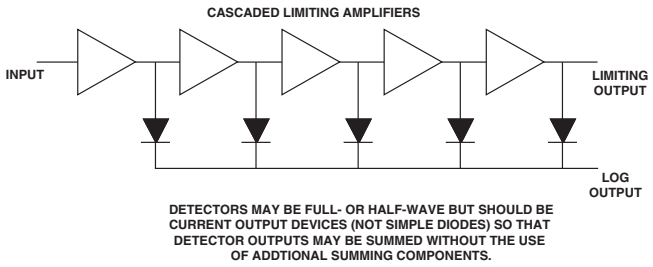


Figure 37. Successive detection log amp.

outputs are summed as shown in Figure 37. If the detectors have current outputs, the summing process may involve no more than connecting all the detector outputs together.

Log amps using this architecture have two outputs: the log output and a limiting output. In many applications, the limiting output is not used, but in some (FM receivers with "S"-meters, for example), both are necessary.

The log output of a successive detection log amplifier generally contains amplitude information, and the phase and frequency information is lost. This is not necessarily the case, however, if a half-wave detector is used and attention is paid to equalizing the delays from the successive detectors—but the design of such log amps is demanding.

The specifications of log amps will include *noise*, *dynamic range*, *frequency response* (some of the amplifiers used as successive detection log amp stages have low frequency as well as high frequency cutoff), the *slope of the transfer characteristic* (which is expressed as V/dB or mA/dB depending on whether we are considering a voltage- or current-output device), the *intercept point* (the input level at which the output voltage or current is zero), and the *log linearity* (see Figures 38 and 39).

KEY PARAMETERS OF LOG AMPS

NOISE: The noise referred to the input (RTI) of the log amp. It may be expressed as a noise figure or as a noise spectral density (voltage, current, or both) or as a noise voltage, a noise current, or both.

DYNAMIC RANGE: Range of signal over which the amplifier behaves in a logarithmic manner (expressed in dB).

FREQUENCY RESPONSE: Range of frequencies over which the log amp functions correctly.

SLOPE: Gradient of transfer characteristic in V/dB or mA/dB.

INTERCEPT POINT: Value of input signal at which output is zero.

LOG LINEARITY: Deviation of transfer characteristic (plotted on log/lin axes) from a straight line (expressed in dB).

Figure 38.

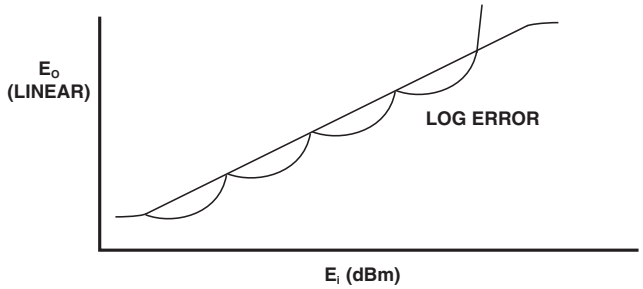


Figure 39. Log linearity.

In the past, it has been necessary to construct high performance, high frequency successive detection log amps (called log strips) using a number of individual monolithic limiting amplifiers such as the one shown in Figure 40—which is over 40 years old! (See Reference 8.) Notice that such components are not, themselves, log amps, but are components from which log amps may be made. The circuit contains a limiting amplifier that drives both the output and an internal current-output half-wave rectifier.

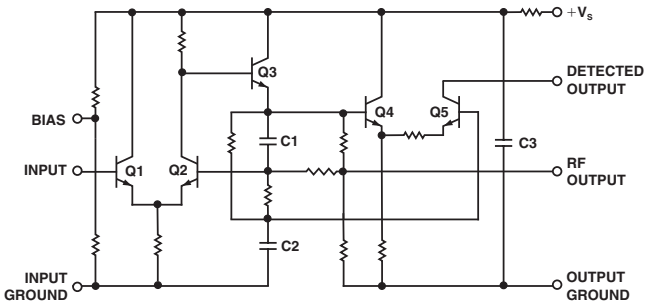


Figure 40. Simplified schematic of monolithic limiting amplifier with half-wave detector.

Figure 41 shows the overall transfer function of a log amp made with four such stages. The detected current output of each stage is plotted against input, as is the sum of all the outputs. It is clear that the sum of these currents approximates a straight line for inputs between 300 μV and 100 mV—about 48 dB.

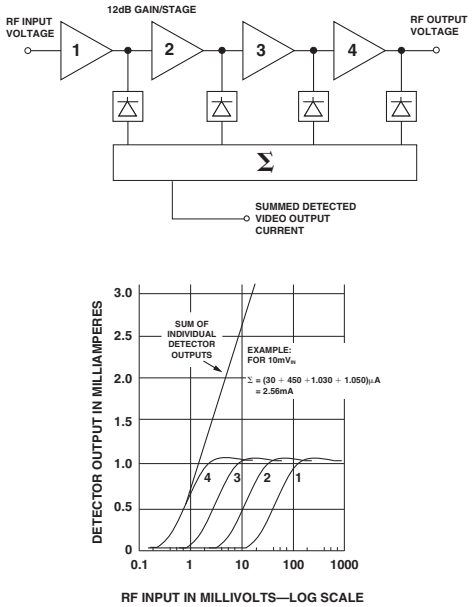


Figure 41. Four-stage successive detection log strip using SL521 "log amps."

If we add stages, the dynamic range increases by 12 dB with each stage until the strip limits on the noise of its own input stage. This occurs with six stages if they are simply connected together broadband. If the noise figure is 5 dB at 450 Ω, this gives about 70 μV broadband noise (assuming 220 MHz bandwidth). The limiting amplifier limits with 100 mV drive, so there must be a gain of less than 1428 (63 dB) to the input of the last stage. At 12 dB/stage, this requires five stages, so, with the output stage, we cannot have more than six stages without limiting on noise. This gives a dynamic range of less than 70 dB.

We can increase the dynamic range by placing an interstage filter between the third and fourth stages of the strip to limit the bandwidth as shown in Figure 42. If we reduce the bandwidth to 10 MHz, the noise is reduced by the square root of 22 (13.5 dB), so we are still limited to seven stages. The interstage filter used does not affect the accuracy of the log response, *provided* that it has a voltage gain that is precisely unity throughout the pass band.

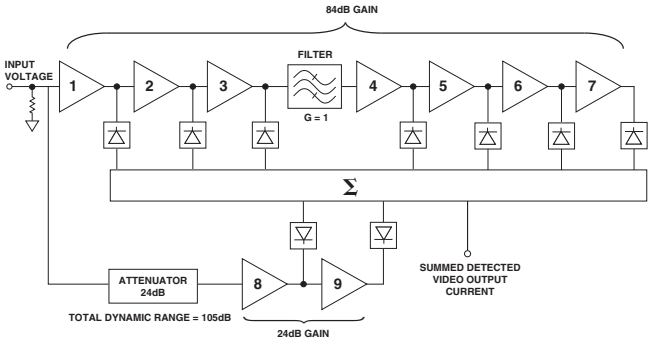


Figure 42. 105 dB log amp using an auxilary strip and an interstage filter.

If we allow for the effects of noise, seven stages will only give some 80 dB dynamic range. If further dynamic range is required, we must use an auxiliary strip. This makes use of the fact that although the output of the limiting amplifier saturates with 100 mV input, the device operates without problems with inputs up to 1.9 V.

If, therefore, we add another two-stage strip of limiting amplifiers, with a 24 dB attenuator at its input, in parallel with the existing strip, and summing its outputs to those of the existing strip, we can add another 24 dB of detector range before the input to the main seven-stage strip is overloaded. This gives us nine stages, for a theoretical dynamic range of 108 dB—in practice, it is possible to achieve about 103 dB to 105 dB.

When constructing a log strip such as the one described above, there are various considerations of coupling, decoupling, filter design, and feedback via the detector pins that must be addressed in any successful design. The single-stage limiting amplifier building block has a low frequency cutoff of about 10 MHz, which makes it impossible to use in many lower frequency applications.

Recent advances in IC processes have allowed the complete log strip function to be integrated into a single chip, thereby eliminating the need for complex and costly hybrid log strips. It is possible, using dielectrically isolated high frequency complementary bipolar processes such as XFCB, to achieve bandwidths of 10 GHz and dynamic ranges up to 100 dB (but not [yet] both at once!).

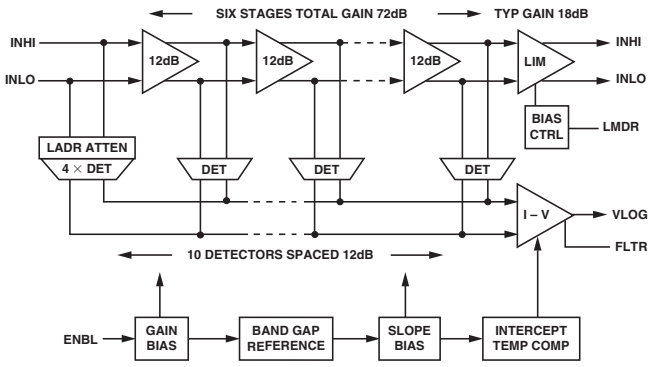


Figure 43. The AD8309 log amp.

As an example, the AD8309 works from LF to 400 MHz and uses a similar configuration to that shown in Figure 42 (but, because the process is lower noise, there is no need for an interstage filter).

Because of their high accuracy, the actual waveform driving log amps of this type must be considered when calculating responses. When a waveform passes through a log function generator, the mean value of the resultant waveform changes. This does not affect the slope of the response, but the apparent intercept is modified according to Figure 44.

The waveform also affects the ripple or nonlinearity of the log response. This ripple is greatest for dc or square wave inputs because every value of the input voltage maps to a single location on the transfer function, and thus traces out the full nonlinearities of the log response. By contrast, a general time-varying signal has a continuum of values within each cycle of its waveform. The averaged output is thereby "smoothed" because the periodic deviations away from the ideal response, as the waveform "sweeps over" the transfer function, tend to cancel. As is clear in Figure 45, this smoothing effect is greatest for a triwave.

Input Waveform	Peak or RMS	Intercept Factor	Error (Relative to a DC Input)
Square Wave	Either	1	0.00 dB
Sine Wave	Peak	2	–6.02 dB
Sine Wave	RMS	1.414 (√2)	–3.01 dB
Triwave	Peak	2.718 (e)	–8.68 dB
Triwave	RMS	1.569 (e/√3)	–3.91 dB
Gaussian Noise	RMS	1.887	–5.52 dB

Figure 44. The effect of waveform on intercept point.

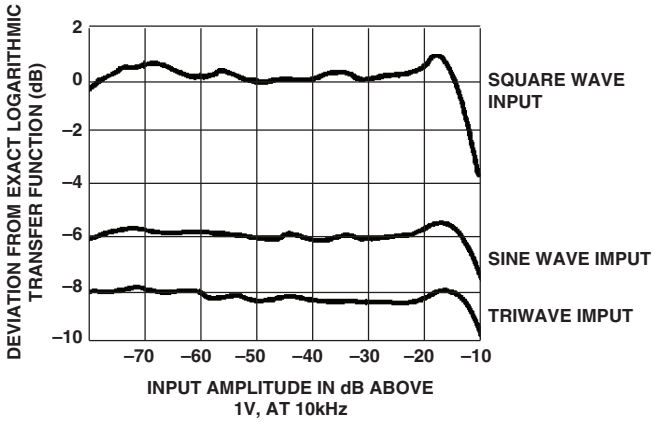


Figure 45. The effect of waveform on log amp log linearity and intercept point.

Conclusion

This paper has shown that although digital computing is a cheap and very powerful tool, there still remain a number of areas where analog computing can compete with it very successfully and offer considerable cost and performance advantages.

References

1. Gilbert, Barrie, *ISSCC Digest of Technical Papers 1968*, February 16, 1968, pp. 114-115.

2. Gilbert, Barrie, *Journal of Solid State Circuits*, Vol. SC-3, December 1968, pp. 353-372.

3. Bryant, James M., *Mixers for High Performance Radio, Wescon 1981: Session 24* (Published by Electronic Conventions, Inc., Sepulveda Blvd., El Segundo, CA).

4. Chadwick, P.E., *High Performance IC Mixers*, IERE Conference on Radio Receivers and Associated Systems, Leeds, 1981, IERE Conference Publication No. 50.

5. Chadwick, P.E., *Phase Noise, Intermodulation, and Dynamic Range*, RF Expo, Anaheim, CA, January 1986.

6. Kitchin, Charles and Lew Counts, *RMS-to-DC Conversion Application Guide, Second Edition*, Analog Devices, Inc., 1986.

7. Sheingold, Daniel H., Editor, *Nonlinear Circuits Handbook*, Analog Devices, Inc., l974.

8. Gay, M. S., *SL521 Application Note*, Plessey Semiconductors, 1966.

Analog Devices, Inc. Worldwide Headquarters
Analog Devices, Inc.
One Technology Way
P.O. Box 9106
Norwood, MA 02062-9106
U.S.A.
Tel: 781.329.4700
(800.262.5643,
U.S.A. only)
Fax: 781.461.3113

Analog Devices, Inc. Europe Headquarters
Analog Devices, Inc.
Wilhelm-Wagenfeld-Str. 6
80807 Munich
Germany
Tel: 49.89.76903.0
Fax: 49.89.76903.157

Analog Devices, Inc. Japan Headquarters
Analog Devices, KK
New Pier Takeshiba
South Tower Building
1-16-1 Kaigan, Minato-ku,
Tokyo, 105-6891
Japan
Tel: 813.5402.8200
Fax: 813.5402.1064

Analog Devices, Inc. Southeast Asia Headquarters
Analog Devices
22/F One Corporate Avenue
222 Hu Bin Road
Shanghai, 200021
China
Tel: 86.21.5150.3000
Fax: 86.21.5150.3222