LOGARITHMIC AMPLIFIERS

Logarithmic amplifiers are specialized nonlinear signal-processing elements used wherever a signal of large dynamic range must be represented by an output of substantially smaller range, and where equal *ratios* in the input domain are usefully transformed to equal *increments* in the output domain. In communications and instrumentation applications, the logarithmic transformation has the additional value of providing a measure of the input expressed in decibel form.

Nonlinear signal conversion invariably has consequences that can be puzzling if the fundamental nature of the transformation is not kept clearly in mind. This is especially true of logarithmic conversion. For example, an attenuator inserted between the source V_{IN} and a conventional linear amplifier simply changes the *slope* $\partial V_{OUT}/\partial V_{IN}$ at the output, which is simply the gain. But this modification would have no effect on the slope (defined in the same way) at the output of a logarithmic amplifier. Similarly, a dc offset voltage at the output of a linear amplifier has no relevance to the processing of an ac signal, whereas an offset introduced at the output of a demodulating logarithmic amplifier alters the apparent ac magnitude of its input.

These important distinctions might be clearer if a term such as "decibel-to-linear converter" were used for such elements, but the description "logarithmic amplifier" or simply *logamp,* the term used throughout this article, has become so widespread that convention demands the continued use of this somewhat misleading name. It should be noted, however, that the logarithmic function can be realized without the use of amplification in the ordinary sense, though that operation is often involved.

There are several different types of logamps, having a similar transfer characteristic between their input signal and the output. Figure 1 shows an idealized input–output response of a generalized logamp, which will later be formalized mathematically. The input might be a signal burst in a cellular phone; the instantaneous value of a unipolar baseband pulse in an airborne, marine, or automotive radar system; the slowly varying carrier envelope in a spectrum analyzer; the current output of a photodiode, or some other measured variable in a high-dynamic-range analytical instrument; and so on. The form of the input signal will be quite different in each case, and the time domain over which events occur ranges from a few nanoseconds in a high-resolution radar system to many seconds in chemical analysis equipment.

For the moment we do not need to be concerned with these distinctions, and it is convenient for now to suppose that the input is a dc voltage, $\boldsymbol{V}_X$. (Boldface symbols indicate input and output signals. Scaling parameters and other incidental variables will be in lightface). This simplification will be particularly useful later, when we undertake a formal analysis of the behavior of logamps based on piecewise linear techniques.

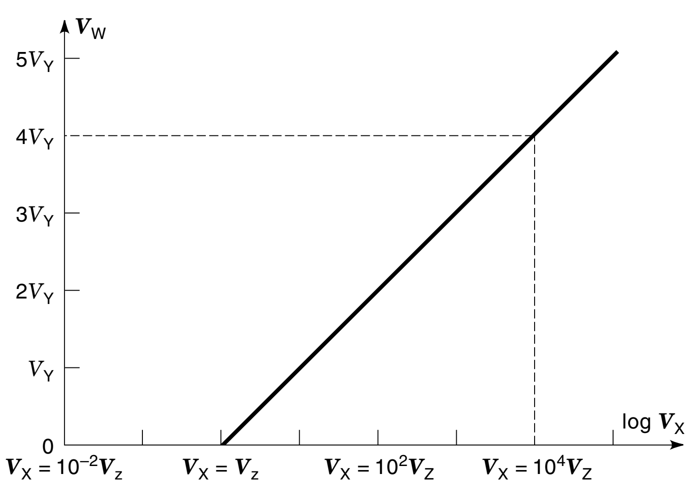


Figure 1. Response of an idealized logarithmic amplifier.

Over a range of several decades, each ratio (say, an octave or decade) of change in $\boldsymbol{V}_X$ causes a fixed unit of change in the output $\boldsymbol{V}_W$. The parameter defining this scaling attribute, V_Y, will here be called the *logarithmic slope,* usually expressed in millivolts per decibel. However, dimensional consistency in equations requires that formally $\boldsymbol{V}_Y$ be identified as the *slope voltage.*

At a certain value of input signal, the output or, more commonly, the *extrapolated output,* will pass through zero, which is here called the *intercept voltage,* and assigned the variable V_Z. If the logamp were perfect, this intercept would actually occur at the unique input $\boldsymbol{V}_X = V_Z$. The need to use an extrapolated value arises because at low input levels, internal amplifier noise or residual offset voltages at the input of a practical circuit will cause significant errors in $\boldsymbol{V}_W$, which will read higher than the correct value, and the output may never actually pass through zero, as depicted in the figure. The intercept is an important *scaling parameter,* since only through knowing both the slope V_Y and the intercept V_Z can the input level be accurately determined. At high input levels the limited signal-handling capability of the circuit cells, either at the input or at the output, will eventually impose some limit on the upper extent of the dynamic range.

Note that the conversion characteristic does not need to be logarithmic to achieve useful compression. An amplifier having a square-root transfer characteristic would halve the decibel-specified dynamic range; a cube-root response would reduce it by a factor of three. *Compandors* performs such power-law operations on the *envelope* amplitude of an ac signal. The logarithmic function is especially valuable because it uniquely provides an output that changes by the same amount over any given ratio of input amplitudes, rendering the output particularly easy to interpret. For example, the output of a logamp with a slope of 1 V/decade changes by 1 V for any tenfold change in the magnitude of the input within its dynamic range. Since a factor-of-ten change in input level corresponds to 20 dB, a logarithmic response is also useful in representing decibel levels; a slope of 1 V/decade corresponds to 50 mV/dB.

Specifying logarithmic circuit performance requires care in defining terms. The literature abounds with incomplete explanations of critical fundamental issues. In calling them *amplifiers,* their strongly nonlinear nature is in danger of being obscured. In some cases (for example, progressive compression logamps) they actually do provide the needed amplification, and in these cases, the logarithmic output, invariably called the *received signal strength indication* (RSSI) may be only an incidental function of the part.

The focus of this article will be on practical circuits that closely approximate the logarithmic function for a wide variety of signal types. We will exercise constant vigilance in matters of scaling, that is, in ensuring *formal traceability* of the constants V_Y and V_Z back to one or more reference voltages, to ensure the highest possible accuracy in the transformation in an actual implementation. This challenge in precision nonlinear design has received inadequate attention in the literature.

CLASSIFICATION

Logamps may be placed in three broad groups, according to the technique used, with one subclassification:

- Direct translinear
- Exponential loop
- Progressive compression:
  - Baseband
  - Demodulating

Direct translinear logamps invoke the highly predictable log-exponential properties of the bipolar transistor. (See TRANSLINEAR CIRCUITS.) Practical translinear logamps can be designed to provide a dynamic range of well over 120 dB for current-mode inputs (e.g., 1 nA to 1 mA). They are most useful where the signal is essentially static or only slowly varying. The design challenge here is to achieve excellent dc accuracy, usually with little emphasis on dynamic behavior.

Figure 2 shows a rudimentary translinear logamp having a current input. It is clearly very incomplete, but it is immediately apparent that this circuit bears no resemblance to any sort of familiar amplifier; in fact, it is a special kind of *transresistance* element, that is, an element generating a voltage output in response to a current input. The practical translinear logamps use operational amplifiers to force the collector-current input signal and further process the voltage-mode output so as to eliminate temperature and device dependence.

Exponential-loop logamps are essentially high-precision *automatic gain control* (AGC) systems. A crucial prerequisite for accurate implementation is the availability of a *variable-gain amplifier* (VGA) having an exact exponential relationship between the control variable (usually a voltage) and the gain. In linear IF strips, the RSSI function is usually derived from the AGC voltage. The Analog Devices X-AMP technique for precision low-noise VGA implementation (1) allows the realization of very accurate exponential-loop logamps; practical examples will be presented.

Figure 3 shows the basic form. The technique is particularly attractive in ac applications, where the *envelope* amplitude of the signal is changing relatively slowly, such as spectrum analysis. It has the further valuable property that the calibration metric may now be peak, average, or rms, depending on the choice of detector type: a representative square-law cell is shown.

The principle is simple: the control loop through the integrator, which seeks to null the difference $V_{DC} - V_R$, adjusts the gain control voltage, which is also the logarithmic output voltage $\boldsymbol{V}_W$, so that the output of the detector cell, V_{DC}, will be equal to the fixed reference voltage V_R. To do that, the gain must have a specific value, depending on the input amplitude $\boldsymbol{V}_X$; using an exponential-controlled VGA, this causes $\boldsymbol{V}_W$ to assume a logarithmic relationship to $\boldsymbol{V}_X$.

Progressive compression is a key concept for high-speed envelope and pulse applications, and most logamps for wideband IF and RF applications are based on this technique, which utilizes a series-connected chain of amplifier cells, each having a simple nonlinear transfer characteristic. In one case, the incremental gain of these cells is A up to a certain instantaneous input amplitude (called the *knee* voltage E_K) falling to 1 above this input. We will refer to these as $A/1$ cells. Using this technique, the logarithmic output is developed at the output of the *last* cell in the chain. $A/1$-based logamps were once widely used for baseband applications.

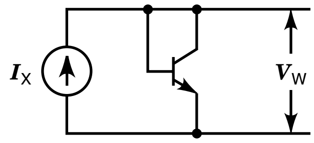


Figure 2. A rudimentary translinear logamp.

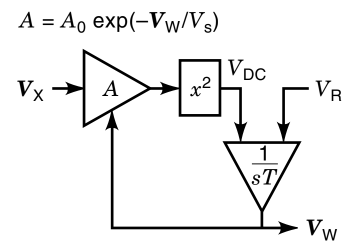


Figure 3. A demodulating logamp based on an exponential VGA.

In a second type, the gain of the cell drops to zero above the knee voltage; this will be called an $A/0$ or *limiter* cell. The logarithmic output is now developed by summing the output of all the cells, either linearly, for baseband applications, or via a half-wave or full-wave rectifier (called a *detector*), in demodulating applications. The ease of implementing the $A/0$ limiter cell in monolithic form, its flexibility (a basic design can readily be adapted for use in either baseband or demodulating operation), and its high degree of accuracy all make it an excellent choice.

In both $A/0$ and $A/1$ types, the N-stage amplifier chain has very high incremental gain A^N for small signals (which is a fundamental requirement of any logamp), and this gain progressively declines as the input amplitude increases. The logarithmic response over a given dynamic range can be approximated to arbitrary accuracy by choice of A and N, with some obvious practical limitations (including noise and offset voltages at the low end, and cell overload at the high end). Because the gain is distributed over many stages, the signal bandwidth can be very high.

Commonplace bipolar technologies readily provide cell bandwidths of over 500 MHz, and operation to well over 10 GHz is possible using advanced *heterojunction bipolar transistor* (HBT) technologies. The AD8313 logamp uses a 25 GHz process to achieve a 3.5 GHz cell bandwidth and provides accurate operation at signal frequencies of 2.5 GHz. These techniques can also be realized using complementary metal-oxide-semiconductor (CMOS) cells in a submicron process, although scaling accuracy is somewhat harder to achieve in CMOS designs.

Baseband logamps are also known as "video logamps," although they are rarely used in video (that is, display-related) applications. They respond to the *instantaneous* value of a rapidly changing input signal. Many board-level and hybrid baseband logamps accept inputs having only one polarity, and they are usually dc-coupled. They are used to compress pulse signals in which the baseline must be accurately preserved. When used *after* a microwave detector (typically a backward diode), the combination is referred to as a *detector video logarithmic amplifier,* or DVLA.

Figure 4 shows a typical application. The numbers are for illustrative purpose only; in practice, the smallest input may be only a few tens of microvolts, calling for unusually low input-offset voltages, sometimes achieved by using an autonulling or dc restoration technique. The dynamic range of a video logamp typically ranges between 40 dB and 80 dB. A DVLA

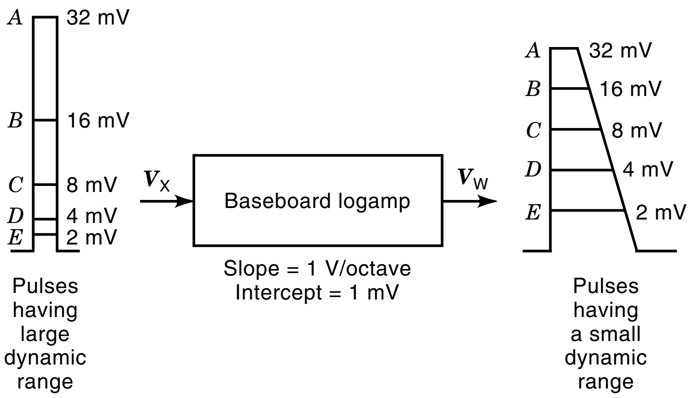


Figure 4. Context of a baseband logamp.

will often incorporate some deliberate deviation from an exact logarithmic response, to first-order compensate the nonlinearities of the preceding microwave detector diode at the extreme upper and lower ends of the signal range.

A baseband logamp that can accept inputs of either polarity and generate an output whose sign follows that of the input is sometimes called a "true logamp," although this is a misnomer, since a true log response would require the output to have a singularity of $\pm\infty$ as the input passes through zero, and anyway the log function has no simple meaning for negative arguments. Instead, the output of a practical logamp of this type passes through *zero* when the input does, just as for any amplifier having a bipolar response. We will later show that the formal response function for this type of logamp is the inverse hyperbolic sine ($\sinh^{-1}$), also called the "ac log" function.

Demodulating logamps rectify the ac signals applied to their input and those that appear at every cell along the amplifier chain. Detection at each stage is followed by summation and low-pass filtering to extract the running average. The logarithmic output is then a baseband signal, essentially a varying dc level, corresponding to the *modulation,* or *envelope,* amplitude of the RF input, rather than its instantaneous value. This structure is called a *successive-detection logamp* (SDLA). Practical demodulating logamps provide dynamic ranges of from 40 dB to over 120 dB. Signal frequencies can extend from near-dc to several gigahertz. The low-frequency capability is invariably determined by the use of a high-pass signal path, needed to suppress the accumulation of small offset voltages. This high-pass corner is typically between a few tens of kilohertz and several megahertz and can be lowered to subaudio frequencies in some general-purpose SDLAs, such as the AD8307. Figure 5 shows the general applications context; the input is an RF carrier, the output is a quasi-dc voltage.

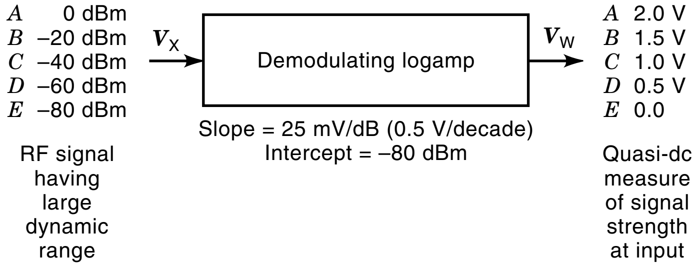


Figure 5. Context of a successive-detection logamp.

In all high-frequency logamps, the management of noise poses a major challenge. Noise is particularly troublesome in demodulating logamps, because once it has been converted to a baseband signal, it is indistinguishable (at the output) from a constant low-level input, thus limiting the attainable dynamic range. Bandpass filters are sometimes inserted between stages to lower the noise bandwidth. A bandpass response may also be desirable as part of the overall system function, for example, in the IF amplifier of a cordless or cellular phone, or in a spectrum analyzer. However, the design of *bandpass logamps* needs great care, since the scaling parameters, which define the logarithmic response, are now inherently frequency-dependent.

Demodulating logamps do *not* respond directly to input power, even though their input is often specified in dBm. Rather, it is the signal *voltage* that determines the output. A root mean square (rms) voltage of 223.6 mV represents a power of 1 mW in a 50 Ω load for a sinusoidal signal; this is written 0 dBm, meaning 0 dB relative to 1 mW. At this input level a logamp would respond with a certain output, say 1 V. Now, if, in the design, we merely alter the *impedance* of the input to 100 Ω (without changing the voltage), the input power halves, but the logamp response is unchanged. On the other hand, logamps are sensitive to signal waveform: thus an input of the same rms value but with a square or triangular waveform would result in a different output magnitude from that generated by a sinusoidal excitation; specifically, the logarithmic intercept is altered. The topic of waveform dependence will be addressed at length.

SCALING OF LOGAMPS

Generating the logarithm of a signal represents a significant transformation. Close attention to the matter of *scaling* is essential. By regarding the logamp as a precision nonlinear element rather than a special kind of amplifier, the designer is forced to think carefully about the source of these scaling parameters. If they cannot be defined with adequate precision, it is likely that the circuit will not be stable with respect to variations in supply voltage and temperature.

Thus, logamp design should begin with a clear formulation of the basic function to be synthesized. For all voltage-input, voltage-output logarithmic converters, of whatever type, this must have the form

$$\boldsymbol{V}_W = V_Y \log(\boldsymbol{V}_X/V_Z) \tag{1}$$

where $\boldsymbol{V}_W$ is the output voltage, $\boldsymbol{V}_X$ is the input voltage, V_Y is the *slope* voltage, and V_Z is the intercept voltage. From the outset, we are careful to use variables of the correct dimensions (all are voltages, in this case). Signals $\boldsymbol{V}_X$ and $\boldsymbol{V}_W$ are uniformly shown in ***bold*** to differentiate them from constants and internal voltages; they stand for the *instantaneous* values of the input and output, at this juncture.

Normally, V_Y and V_Z are fixed scaling voltages, but they could be scaling control inputs. For example, it may be useful to arrange for the slope to track a supply voltage when this is also used as the scaling reference for an analog-to-digital converter (ADC), which subsequently processes the logamp's output. Equation (1) is completely general and dimensionally consistent; this is important in developing a theory of logamps, and in designing them, since it maintains a strong focus on the detailed sources of the function's scaling. The practice of using factors of unclear dimension such as $e_{OUT} = K_1 \log K_2 e_{IN}$ is discouraged (2).

The choice of logarithmic base is arbitrary. To preserve generality, it is not defined in Eq. (1). A change of base merely results in a change in V_Y. We will generally adopt base-10 logarithms, identified by the symbol lgt, in keeping with the decibel-oriented context. However, in order to evaluate certain integrals when discussing the effect of waveform on intercept in demodulating logamps, we will occasionally switch to ln, that is, natural (base-*e*) logarithms.

It is apparent from Eq. (1) that $\boldsymbol{V}_W$ increases by an amount V_Y for each unit increase in the quantity $\log(\boldsymbol{V}_X/V_Z)$. When the logarithm is to base ten, that statement reads: for each decade increase in $\boldsymbol{V}_X$. In that particular case V_Y has the meaning of "volts per decade." Figure 1 showed this function for $V_Y = 1$ V and $V_Z = 1$ μV. The logarithmic output $\boldsymbol{V}_W$ would ideally cross zero when $\boldsymbol{V}_X = V_Z$. In other words, V_Z represents the *intercept* of the transfer function on the horizontal axis. This may not actually occur; V_Z will often be the *extrapolated* intercept, and for reasons of design its value may be so small that the lower boundary on $\boldsymbol{V}_W$ will be first limited by noise or input offset voltage.

It is very easy to arrange for the intercept of a logamp to have any desired value. This can be readily appreciated from the following expansion:

$$\begin{aligned}\boldsymbol{V}_W &= V_Y \log(\boldsymbol{V}_X V_N / V_Z V_N) \\ &= V_Y \log(\boldsymbol{V}_X / V_N) + V_A\end{aligned} \tag{2}$$

where V_N is the new value of the intercept achieved by adding some constant V_A to the output of the log converter, having the value

$$V_A = V_Y \log(V_N/V_Z) \tag{2a}$$

Clearly, V_A, and therefore V_N, can have whatever value we wish, and this voltage need not be physically sensible; it could be as low as, say, 1 nV. Usually, it will be a few microvolts. The intercept may be repositioned by adding or subtracting a voltage to the output of a logamp, corresponding to V_A (see Fig. 6), often included as part of a temperature-compensation technique, though internally the offsetting quantity is conveniently in current-mode form. The introduction of attenuation at the *input* of a logamp only changes the effective intercept, and does not affect the logarithmic slope:

$$\begin{aligned}\boldsymbol{V}_W &= V_Y \log(K\boldsymbol{V}_X/V_Z) \\ &= V_Y \log(\boldsymbol{V}_X/\boldsymbol{V}_N), \qquad \text{where} \quad V_N = V_Z/K\end{aligned} \tag{3}$$

These transformations will later prove useful when we need to compensate basic temperature effects in the gain cells typically used in logamps.

A well-designed logamp has at least one high-accuracy dc reference source, from which both V_Y and V_Z are derived. The Analog Devices AD640 provides an early example of a monolithic logamp designed with close attention to the matter of calibration accuracy. It uses two laser-trimmed reference generators, of which one, a bandgap circuit, sets V_Y and the other, a cell providing a bias voltage that is *proportional to absolute temperature* (PTAT), accurately determines the cell gains, which affect both V_Y and V_Z. In a logamp, these voltage references play a role as important as those in an ADC or analog multiplier. In the case of the AD606 and AD608, the reference is the supply voltage. This is a deliberate simplification: in their intended application, the ADC that processes the logarithmic (RSSI) output uses the same supply voltage for its scaling reference.

The choice of *voltage* for inputs and outputs is mainly to provide a suitable frame of reference for formal analysis. We could have just as easily cast Eq. (1) in terms of a current-input and voltage-output device:

$$\boldsymbol{V}_W = V_Y \log(\boldsymbol{I}_X/\boldsymbol{I}_Z) \tag{1a}$$

where V_Y, $\boldsymbol{I}_X$, and I_Z have equivalent specifications. This is the function of the rudimentary translinear logamp of Fig. 2, elaborated in the next section.

Alternatively, *all* signals may be in current form:

$$\boldsymbol{I}_W = I_Y \log(\boldsymbol{I}_X/I_Z) \tag{1b}$$

This is less common, but certainly quite practical. Finally, the function could be in the form of voltage input and current output:

$$\boldsymbol{I}_W = I_Y \log(\boldsymbol{V}_X/V_Z) \tag{1c}$$

This is the form found *internally* in RF logamps that use transconductance cells for demodulation; in these cases, the intermediate output current $\boldsymbol{I}_W$ is later converted back to a voltage $\boldsymbol{V}_W$, using a transresistance cell.

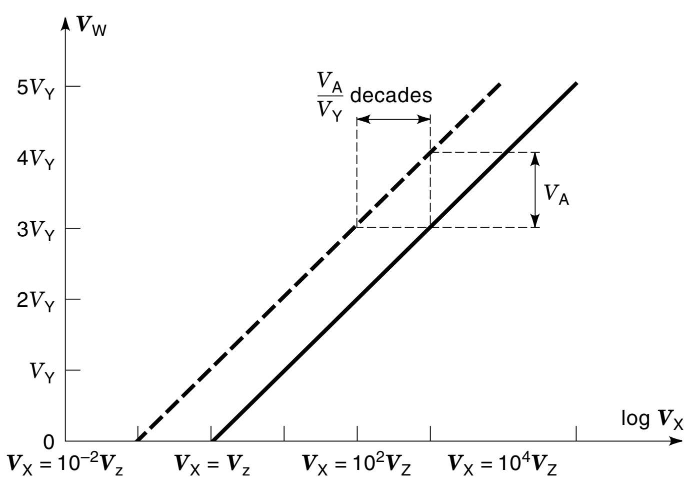


Figure 6. Repositioning the intercept.

Region Near Zero

As the input $\boldsymbol{V}_X$ tends toward zero from positive values, the output $\boldsymbol{V}_W$ will ideally approach $-\infty$. Differentiating Eq. (1), using base-*e* logarithms, and ignoring at this point any resultant scale changes in V_Y, we can see that the *incremental gain* of a logamp approaches $+\infty$ as $\boldsymbol{V}_X$ approaches zero:

$$\begin{aligned}\frac{\partial \boldsymbol{V}_W}{\partial \boldsymbol{V}_X} &= \frac{\partial}{\partial \boldsymbol{V}_X} V_Y (\ln \boldsymbol{V}_X - \ln V_Z) \\ &= \frac{V_Y}{\boldsymbol{V}_X}\end{aligned} \tag{4}$$

The elimination of V_Z in the expression for incremental gain is consistent with the fact, already pointed out, that we can arbitrarily alter V_Z after logarithmic conversion by the addition or subtraction of a dc term at the output.

Since the overall amplifier gain must ideally tend to infinity for near-zero-amplitude inputs, it follows that the low-level accuracy of a practical logamp will be limited at the outset by its maximum small-signal gain (determined by the gain of the amplifier stages and the number of stages, for the progressive compression logamps described later), and ultimately by the noise level of its first stage (in the case of demodulating logamps) or by the input-referred dc offset (in the case of baseband logamps).

The incremental gain—a familiar metric for linear amplifiers, where it ought to be independent of the signal level—will be found to vary radically for a logamp, from a maximum value of perhaps 100 dB to below 0 dB at high inputs: Eq. (4) shows that it is unity when the input voltage $\boldsymbol{V}_X$ is equal to the scaling voltage V_Y. In fact, the incremental gain of a logamp is never of great importance in the design process, but its tremendous variation demonstrates the inadvisability of lapsing into the use of small-signal analysis and simulation studies when dealing with logamps.

So far, we have not mentioned the *polarity* of the signal $\boldsymbol{V}_X$. For a baseband converter $\boldsymbol{V}_X$ might be a positive dc voltage or pulse input, so Eq. (1) can be used without further consideration. But what happens when $\boldsymbol{V}_X$ becomes negative? There is no simple meaning to the log function when its argument is negative. Fortunately, we do not have to consider the mathematical consequences of this, because practical baseband logamps can be designed to handle inputs of *either* polarity, or, using appropriate techniques, inputs of *both* polarities. If mathematical rigor is needed, we can adapt Eq. (1) to handle this situation by assuming that the circuit is arranged in some way to respond only to the magnitude of $\boldsymbol{V}_X$ and then restore its sign at the output:

$$\boldsymbol{V}_W = \mathrm{sgn}(\boldsymbol{V}_X)\, V_Y \ln(|\boldsymbol{V}_X|/V_Z) \tag{5}$$

The bisymmetric function described by this equation refers to what is sometimes called the "ac logarithm." This function is still not practical, however, because it requires that the output undergo a transition from $-\infty$ to $+\infty$ as $\boldsymbol{V}_X$ passes through zero, whereas in practical amplifiers intended to handle bipolar inputs $\boldsymbol{V}_W$ will pass through *zero* when $\boldsymbol{V}_X = 0$, because of the finite gain of its component sections. The situation described by Eq. (5) and its practical limitations can be handled by replacing the logarithmic function by the inverse hyperbolic sine function:

$$\sinh^{-1} u = \ln(u + \sqrt{u^2 + 1}) \qquad \text{for} \quad u > 1 \tag{6}$$

$$\sim \ln 2u \qquad \text{for} \quad u \gg 1 \tag{7}$$

Note that

$$\sinh^{-1}(-u) = -\sinh^{-1} u$$

Figure 7 compares the ideal bisymmetric logarithmic function in Eq. (5) with the inverse hyperbolic sine in the region

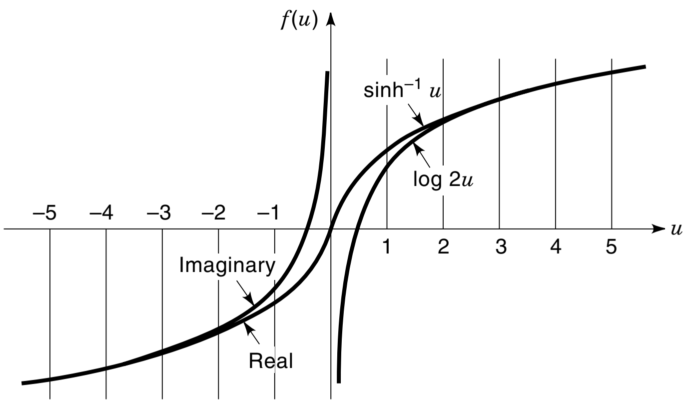


Figure 7. Log and $\sinh^{-1}$ functions for small arguments.

near $u = 0$. The "ac log" function may thus be very closely approximated by

$$\boldsymbol{V}_W = V_Y \sinh^{-1}(\boldsymbol{V}_X/2V_Z) \tag{8}$$

As a practical matter, the region of operation corresponding to extremely small values of $\boldsymbol{V}_X$ will invariably be dominated by noise, which appears to occupy an inordinate amount of the output range. The use of a nonlinear *low-pass filter* (LPF), whose corner frequency depends on the instantaneous output of the logamp, is helpful. For outputs near zero, this filter "idles" with a low bandwidth of, say, 1 kHz; a rapid increase in the input to this adaptive filter immediately raises the bandwidth, and the step response remains fast. A dead-zone response near zero can be used to obscure low-level noise. In marine radar applications, this is called an *anti-clutter* filter.

Effect of Waveform on Intercept

We have seen that a demodulating logamp operates from an ac input signal and internally has detector cells that convert the alternating signals along the amplifier chain into quasi-dc signals, which become the logamp output after low-pass filtering. Now, we need to consider not just the amplitude of $\boldsymbol{V}_X$, but also its waveform, since this can have significant practical consequences. In the performance specifications for a RF logamp, the signal is invariably assumed to be sinusoidal, and the intercept, usually specified as a power in dBm, also assumes this. For other waveforms, such as those arising for a complex modulation mode, as in *code-division multiple-access* (CDMA) systems, the effective value of the intercept will be different.

If the input is an amplitude-symmetric square wave, the rectification inherent in this type of logamp results in an intercept that would be identical to that for a constant dc level, assuming the logamp is dc-coupled and uses full-wave detectors. For a sinusoidal input, where $\boldsymbol{V}_X$ is specified as the amplitude (*not* the rms value), it will be exactly double this dc value. For an amplitude-symmetric triangle wave, the intercept will appear to be increased by a factor of $e \approx 2.718$. For a noise input with some prescribed probability density function (PDF) it will have a value dependent on the PDF: when this is Gaussian, the effective intercept is increased by a factor of 1.887. While it is unusual for the behavior of a logamp to be quantified for waveforms other than sinusoidal, it is valuable to establish these foundations before proceeding with practical designs (3).

These issues only became of more than academic interest with the advent of *fully calibrated* logamps. Prior to that time, the intercept had to be adjusted by the user, and demodulating RF logamps were calibrated using sinusoidal inputs. Of course, the waveform dependence of the intercept does not arise in the case of baseband ("video") logamps, where there is a direct mapping between the instantaneous value of the input and the output. It is entirely a consequence of the signal rectification of the detectors and the averaging behavior of the post-detection low-pass filter, neither of which is present in a baseband logamp.

We begin with the sine case and use base-ten logarithms, denoted by lgt. We can write Eq. (1) in the form

$$\boldsymbol{V}_{\mathrm{W}} = V_{\mathrm{Y}} \,\mathrm{lgt}\, \frac{\boldsymbol{E}_{\mathrm{A}} \sin\theta}{V_{\mathrm{Z}}} \tag{9}$$

$\boldsymbol{V}_{\mathrm{W}}$ describes the *instantaneous* value of the output; however, for a demodulating logamp, we will be concerned with the *average* value of $\boldsymbol{V}_{\mathrm{W}}$, that is, the output of some postdemodulation low-pass filter. In this equation, $\boldsymbol{E}_{\mathrm{A}}$ is the amplitude of the sine input and θ is its angle, more usually written as the time-domain function ωt. The mathematical inconvenience of negative logarithmic arguments can be avoided by considering the behavior of Eq. (9) over the range for which $\sin\theta$ is positive. In fact, we need only concern ourselves with the principal range $0 \le \theta \le \pi/2$, since the average over a full period will be simply four times the average over this range, assuming the use of full-wave rectification in the detectors. The demodulated and filtered output is

$$\begin{aligned}
\mathrm{Ave}(\boldsymbol{V}_{\mathrm{W}}) &= \frac{2}{\pi}\int_0^{\pi/2} V_{\mathrm{Y}} \,\mathrm{lgt}\, \frac{\boldsymbol{E}_{\mathrm{A}}\sin\theta}{V_{\mathrm{Z}}}\, d\theta \\
&= \frac{2V_{\mathrm{Y}}}{\pi}\int_0^{\pi/2} \left(\mathrm{lgt}\sin\theta + \mathrm{lgt}\,\frac{\boldsymbol{E}_{\mathrm{A}}}{V_{\mathrm{Z}}}\right) d\theta \\
&= \frac{2V_{\mathrm{Y}}}{\pi \ln 10}\int_0^{\pi/2} \left(\ln\sin\theta + \ln\frac{\boldsymbol{E}_{\mathrm{A}}}{V_{\mathrm{Z}}}\right) d\theta
\end{aligned} \tag{10}$$

The definite integral of $\ln\sin\theta$ over the range of interest is $-(\pi/2)\ln 2$ and the complete integral yields

$$\begin{aligned}
\mathrm{Ave}(\boldsymbol{V}_{\mathrm{W}}) &= \frac{2V_{\mathrm{Y}}}{\pi\ln 10}\left(-\frac{\pi}{2}\ln 2 + \frac{\pi}{2}\ln\frac{\boldsymbol{E}_{\mathrm{A}}}{V_{\mathrm{Z}}}\right) \\
&= \frac{V_{\mathrm{Y}}}{\ln 10}\left(\ln\frac{\boldsymbol{E}_{\mathrm{A}}}{V_{\mathrm{Z}}} - \ln 2\right)
\end{aligned} \tag{11}$$

$$= V_{\mathrm{Y}}\,\mathrm{lgt}(\boldsymbol{E}_{\mathrm{A}}/2V_{\mathrm{Z}}) \tag{12}$$

Simply stated, the response to an input having a sinusoidal waveform and an *amplitude* $\boldsymbol{E}_{\mathrm{A}}$ will be the same as for a *constant* dc input having a magnitude of $\boldsymbol{E}_{\mathrm{A}}/2$. The logarithmic transfer function is shifted to the right by 6.02 dB for the case of sine excitation, relative to the basic dc response.

The functional form of Eq. (11) deserves further attention. Inside the parentheses we have the difference between a logarithmic term with the normalized argument $\boldsymbol{E}_{\mathrm{A}}/V_{\mathrm{Z}}$ and a second term, ln 2, which is a function of the waveform. This term can be viewed as a *waveform signature.*

Using a similar approach for the triangular-wave input, we can write

$$\boldsymbol{V}_{\mathrm{W}} = V_{\mathrm{Y}}\,\mathrm{lgt}\left(\frac{\boldsymbol{E}_{\mathrm{A}}}{V_{\mathrm{Z}}}\frac{4t}{T}\right) \tag{13}$$

to describe the instantaneous output, where $\boldsymbol{E}_{\mathrm{A}}$ is now the amplitude of a triwave of period T. The demodulated and filtered output is then

$$\begin{aligned}
\mathrm{Ave}(\boldsymbol{V}_{\mathrm{W}}) &= \frac{4}{T}\int_0^{T/4} V_{\mathrm{Y}}\,\mathrm{lgt}\left(\frac{\boldsymbol{E}_{\mathrm{A}}}{V_{\mathrm{Z}}}\frac{4t}{T}\right) dt \\
&= \frac{4V_{\mathrm{Y}}}{T}\int_0^{T/4}\left(\mathrm{lgt}\,t + \mathrm{lgt}\frac{4\boldsymbol{E}_{\mathrm{A}}}{V_{\mathrm{Z}}T}\right) dt \\
&= \frac{4V_{\mathrm{Y}}}{T\ln 10}\int_0^{T/4}\left(\ln t + \ln\frac{4\boldsymbol{E}_{\mathrm{A}}}{V_{\mathrm{Z}}T}\right) dt
\end{aligned} \tag{14}$$

The integral of $\ln t$ is simply $t(\ln t - 1)$, yielding

$$\begin{aligned}
\mathrm{Ave}(\boldsymbol{V}_{\mathrm{W}}) &= \frac{4V_{\mathrm{Y}}}{T\ln 10}\left(\frac{T}{4}\ln\frac{T}{4} - \frac{T}{4} + \frac{T}{4}\ln\frac{4\boldsymbol{E}_{\mathrm{A}}}{V_{\mathrm{Z}}T}\right) \\
&= \frac{V_{\mathrm{Y}}}{\ln 10}\left(\ln\frac{\boldsymbol{E}_{\mathrm{A}}}{V_{\mathrm{Z}}} - 1\right)
\end{aligned} \tag{15}$$

In this case, the waveform signature is just 1; however, since this may be written as $\ln e$, the output becomes

$$\mathrm{Ave}(\boldsymbol{V}_{\mathrm{W}}) = V_{\mathrm{Y}}\,\mathrm{lgt}(\boldsymbol{E}_{\mathrm{A}}/eV_{\mathrm{Z}}) \tag{16}$$

Thus, a triangle-wave input will effectively cause the intercept to shift to the right by a factor of e, or 8.69 dB. Intuitively, this is not unreasonable: for any given amplitude the triwave spends less time at its higher values than a sinusoidal waveform does, and consequently its average contribution to the filtered output is reduced.

For a noise input having a Gaussian PDF with an rms value of $\boldsymbol{E}_{\sigma}$, the effective intercept is most easily calculated by first reducing the formulation to a generalized form. The average value μ of some variable x having a unit standard deviation, which has been subjected to a logarithmic transformation, can be expressed as

$$\mu = \frac{\int_0^\infty e^{-x^2/2}\ln x\, dx}{\int_0^\infty e^{-x^2/2}\, dx} \tag{17}$$

Note that the variable x represents the *instantaneous* value of the input noise voltage [so it is actually $x(t)$, but the time argument is an unnecessary complication for this calculation]. The numerator and denominator are both standard forms (4):

$$\int_0^\infty e^{-\alpha x^2}\ln x\, dx = \frac{(\gamma + \ln 4\alpha)\sqrt{\pi}}{4\alpha}$$

where γ is Euler's constant, and

$$\int_0^\infty e^{-\alpha x^2}\, dx = \frac{\sqrt{\pi}}{4\alpha}$$

Hence

$$\mu = \frac{\gamma + \ln 2}{4}$$

which evaluates to ln(1/1.887). In other words, the average value of the logarithmic output in response to a Gaussian input of unit rms value is equivalent to a dc input of 1/1.887. For a general input $\boldsymbol{E}_\sigma$,

$$\mathrm{Ave}(\boldsymbol{V}_{\mathrm{W}}) = V_{\mathrm{S}}\,\mathrm{lgt}(\boldsymbol{E}_\sigma/1.887V_{\mathrm{Z}}) \tag{18}$$

where V_S is a scaling parameter. This corresponds to an intercept shift of 5.52 dB. It is interesting to note that this is only 0.5 dB different from the rms calibration for a sine-wave input and might easily be attributed to measurement error in the evaluation of the noise response of practical logamps.

THE TRANSLINEAR LOGAMP

Logamps intended for use at dc or moderate frequencies traditionally invoke a translinear technique, though that term has not generally been used in a logamp context. The word *translinear* (5,6) refers to the remarkably exact logarithmic relationship between the base–emitter voltage V_{BE} and the collector current I_C in a *bipolar junction transistor* (BJT), of pervasive significance in the design of analog bipolar circuits. In particular, it results in the *trans*conductance being *linear* in I_C. For the present purposes, this relationship can be written as

$$V_{\mathrm{BE}} = V_{\mathrm{T}} \ln\left(\frac{I_{\mathrm{C}}}{I_{\mathrm{S}}+1}\right) \tag{19}$$

where I_S is a basic scaling parameter for the BJT, called the *saturation current,* an extremely strong function of temperature, and V_T is the thermal voltage kT/q. Thus, there might at first seem little prospect of taming this temperature variability. In fact, translinear logamps can be developed to a high degree of refinement. We will first convert Eq. (19) to base-10 logarithms to bring it into line with the decibel-world logamp perspective, slightly rearrange things, and again show the key signal variables in boldface:

$$\boldsymbol{V}_{\mathrm{BE}} = V_{\mathrm{Y}}\,\mathrm{lgt}\left(\frac{\boldsymbol{I}_{\mathrm{C}}+I_{\mathrm{S}}}{I_{\mathrm{S}}}\right) \tag{20}$$

where

$$V_{\mathrm{Y}} = V_{\mathrm{T}} \ln 10 \tag{21}$$

The logarithmic slope V_Y is PTAT, and evaluates to 59.52 mV per decade at $T = 300$ K. The logarithmic intercept is simply the saturation current I_S, typically between 10^{-18} A and 10^{-15} A at this temperature. In Eq. (20) the signal input $\boldsymbol{I}_C$ is augmented by this tiny current; we later address the consequences of this slight anomaly in the otherwise straightforward logamp form of the equation.

Figure 8 shows a scheme often used to force the collector current I_C to equal $\boldsymbol{I}_X$, the signal current applied to the logamp. (Compare with Fig. 2). This is sometimes called a "transdiode connection" or "Paterson diode" (7). The usual

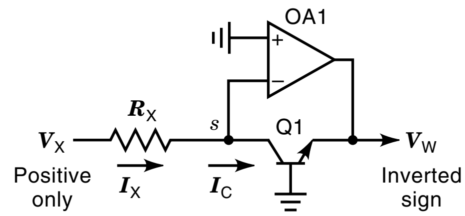


Figure 8. Translinear logamp using an opamp to force I_C.

npn form is shown, requiring $V_X \geq 0$, but it is obvious that use of a *pnp* transistor would simply reverse the required polarity of input current and the resulting polarity of output voltage.

The transistor can be replaced by a diode or diode-connected transistor, with certain advantages, one of which is that a bipolar response can now be achieved by using two parallel opposed diodes. Another benefit is that the loop gain around the opamp becomes essentially independent of signal current, simplifying high-frequency (HF) compensation and potentially raising the bandwidth. However, this technique requires the opamp to have very low offset voltage V_{OS}. Also, since V_{BE} bears a highly accurate, multidecade relationship only to the *collector* current I_C, logamps built using diode-connected transistors, in which it is I_E that is forced, will exhibit inherently lower accuracy.

The opamp OA1 forces the collector current of the transistor Q1 to equal the input current $\boldsymbol{I}_X$ while maintaining its collector–base voltage V_{CB} very close to zero. The condition $V_{CB} = 0$ is not essential: for most purposes little harm will result if the collector junction is reverse biased (this effectively increases I_S), or even becomes slightly forward biased. It can be shown that there is actually an advantage to using a very specific value of the reverse collector bias ($V_{CB} \approx 50$ mV) in certain applications.

The logarithmic output is taken from the emitter node; the opamp allows this to be loaded while preserving accuracy. In most cases, I_S will be very much less than $\boldsymbol{I}_X$ and, replacing I_C by $\boldsymbol{I}_X$, we can simplify Eq. (20) to

$$\boldsymbol{V}_{\mathrm{W}} = -V_{\mathrm{Y}}\,\mathrm{lgt}(\boldsymbol{I}_{\mathrm{X}}/I_{\mathrm{S}}) \tag{22}$$

Figure 9 shows an illustrative simulation result for this rudimentary circuit using an idealized *npn,* having $I_S = 10^{-16}$ A at $T = 300$ K, operating at temperatures of -50°C, $+50$°C, and $+150$°C. (The sign of the output has been flipped to maintain a uniform presentation.) The temperature dependences of the slope and intercept are apparent. The output converges on the bandgap voltage $E_{G0} \approx 1.2$ V at very high currents.

For high-temperature operation at very low currents, I_S becomes comparable with the input current $\boldsymbol{I}_X$. The departure from logarithmic behavior in this region can be corrected by using a particular values of V_{CB}, which is useful in logamps that must operate accurately down to low-picoampere inputs. The details lie beyond this treatment but are included in the practical design shown in Fig. 10, which also includes means (C_C and R_E) to ensure HF stability of the first loop around OA1.

The basic scheme shown in Fig. 8 would exhibit large temperature variations, due to the dependence of I_S in the underlying BJT equations, which directly determines the log inter-

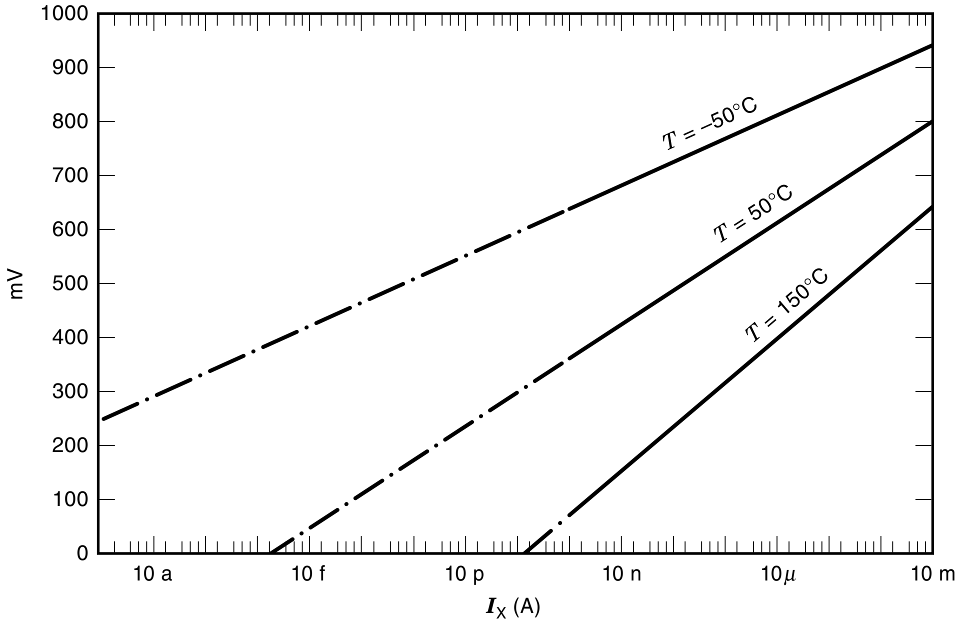


Figure 9. Output of the basic translinear logamp.

cept. The practical design includes a means for canceling the temperature dependency of I_S, using a second transistor Q2, presumed here to be identical to Q1, and a second operational amplifier OA2. Now we have

$$\begin{aligned} \boldsymbol{V}_W &= -V_T \ln \frac{\boldsymbol{I}_X}{I_S(T)} + V_T \ln \frac{\boldsymbol{I}_Z}{I_S(T)} \\ &= -V_T \ln(\boldsymbol{I}_X/I_Z) \end{aligned} \tag{23}$$

Thus, the intercept has been altered from a very uncertain value (I_S) to one of arbitrarily high accuracy (I_Z) provided from an external source.

Equation (23) still has a temperature-dependent slope voltage, $V_T = kT/q$. Also, the fairly small and awkward scaling factor (≈59.52 mV/decade at 300 K) will usually need to be raised to a larger and more useful value. This is achieved in Fig. 10 using a temperature-corrected feedback network. R_{PT} is a resistor with a large positive temperature coefficient (TC), while R_F is a zero-TC component. If the ratio R_F/R_{PT} were very high, R_{PT} would need to be exactly PTAT ($+3300 \times 10^{-6}$/°C at $T = 30$°C), but for any finite ratio this resistor must have a higher TC. Such resistors are readily available. Both the slope and the intercept are now substantially free of temperature effects. Figure 11 shows a typical result and the improvement that can be achieved at low input currents by applying a small PTAT bias to the base of Q1.

While some wide-dynamic-range transducers (photodiodes and photomultiplier tubes) do generate current-mode signals, the input signal will often be in the form of a voltage, $\boldsymbol{V}_X$. It is a simple matter to adapt the logamp shown in Fig. 10 to voltage-mode signals, using a resistor between $\boldsymbol{V}_X$ and the summing node. When the opamp uses a bipolar input stage, and therefore exhibits considerable input bias current, the inclusion of an equal resistor in series with the noninverting input of the opamp will serve to cancel its effect.

EXPONENTIAL AGC LOGAMPS

The logarithm function is the inverse of the exponential function. In classical analog computing tradition, function in-

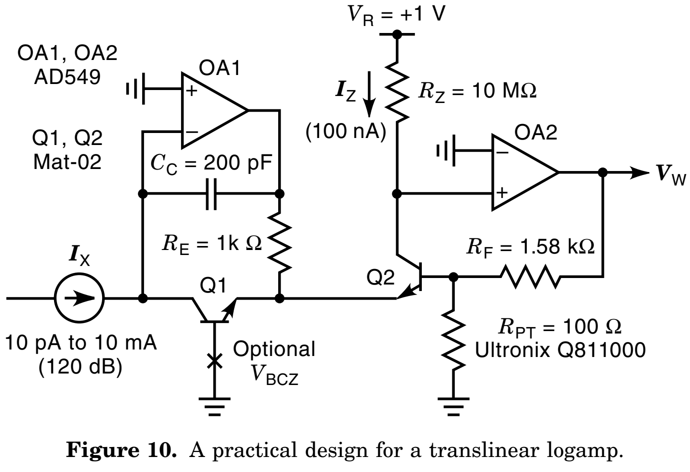


Figure 10. A practical design for a translinear logamp.

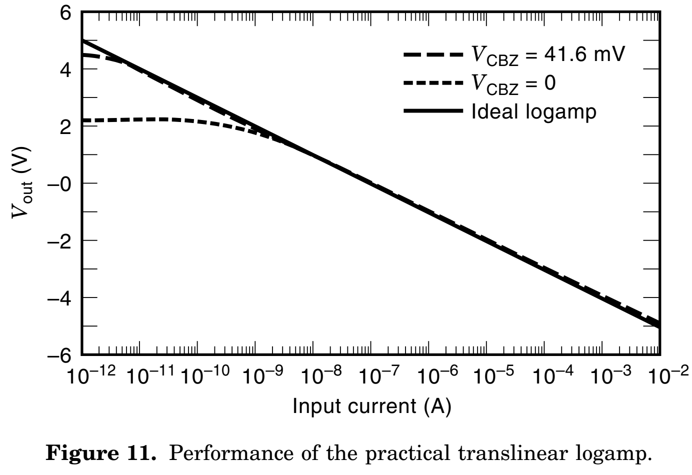


Figure 11. Performance of the practical translinear logamp.

verses are generated by enclosing the function in a feedback path around an opamp, which forces the output of the function to be equal to some input, at which time, the output of the opamp (that is, the input to the function block) is the desired inverse. This is precisely what happens in the case of the translinear logamp, where the *forward* direction through the function block—in that case a transistor—is exponential.

However, there are severe bandwidth limitations in attempting to cover a wide dynamic range in a single stage. A special type of VGA, having precisely exponential control of gain, can be used in place of the transistor, as shown in Fig. 12. Here, the gain of the amplifier cell *decreases* with increasing value of its control input, to which is applied the output voltage $\boldsymbol{V}_W$, with a scaling voltage of V_Y. The output is thus

$$V_A = \boldsymbol{V}_X A_0 \exp(-\boldsymbol{V}_W/V_Y) \tag{24}$$

The second active block, essentially an error integrator, forces $\boldsymbol{V}_W$ to the value that results in the VGA output being maintained at the reference voltage V_R applied to the inverting input of the error amplifier. When that condition is satisfied, we have

$$\boldsymbol{V}_X A_0 \exp(-\boldsymbol{V}_W/V_Y) = V_R \tag{25}$$

Thus

$$\boldsymbol{V}_W = V_Y \ln(\boldsymbol{V}_X/V_Z) \tag{26}$$

where

$$V_Z = V_R/A_0 \tag{27}$$

The use of an exponential VGA response and an integrator results in a simple single-pole low-pass response, for small perturbations, independent of the magnitude of $\boldsymbol{V}_X$, over many decades. Thus, we have realized a baseband logamp having a constant small-signal bandwidth. Good operation can be achieved even using a single VGA cell, which might use translinear principles to realize the exponential gain law. In practice, however, several lower-gain VGA cells will often be used to realize the main amplifier, which will also be ac-coupled in many applications. For N cells the gain is

$$A = [A_0 \exp(-\boldsymbol{V}_W/V_Y]^N = A_0^N \exp(-N\boldsymbol{V}_W/V_Y) \tag{28}$$

To cover an 80 dB range of inputs, we might use four cells, each of which provides a gain variation of one decade. A final detector cell must be added, to convert the ac output to a quasi-dc value, as shown in Fig. 13. Assuming for now that the detector cell has an effective gain of unity, the use of N stages simply alters the intercept to $V_Z' = V_R/A_0^N$ and the slope to $V_Y' = V_Y/N$.

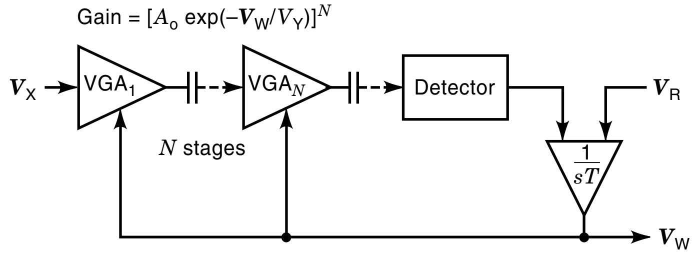


Figure 13. Logamp based on cascaded exponential VGA cells.

A variety of demodulator styles is possible. The simplest is a half-wave detector, based on a single junction; this provides an averaging response. It has a mean output that is $0.318E_A$ for a sine-wave signal $E_A \sin \omega t$, is $E_A/2$ for a square wave signal of amplitude E_A, is $E_A/4$ for a triwave signal, and so on. A full-wave rectifier would simply double these numbers. Alternatively, we might use a peak detector, changing the dynamics of the loop considerably.

With a two-quadrant square-law detector, that is, one responding equally to signals of either polarity, followed by filtering to extract the mean square, the resulting loop implements a *root-mean-square* (rms) measurement system without having to use a square-rooting circuit (Fig. 14). Here, the loop integrator seeks to null its input by forcing the mean squared value of the detector output to the fixed reference V_{R1}. There is obviously no need to include the rooting function before making this comparison; however, a more careful analysis of the scaling issues will show that a square-law detector has its own scaling voltage:

$$V_{SQR} = V_{OUT}(t)^2/V_{R2} \tag{29}$$

and the low-pass filtered output is thus

$$\boldsymbol{V}_W = \text{Ave}(V_{OUT}(t)^2/V_{R2} \tag{30}$$

where V_{R2} is the squaring cell's scaling voltage. From Eq. (28),

$$V_{OUT} = V_{IN} A_0^N \exp(-N\boldsymbol{V}_W/V_Y) \tag{31}$$

and the loop forces $\boldsymbol{V}_W$ to equal V_R, so we have

$$\frac{\text{Ave}([V_{IN} A_0^N \exp(-N\boldsymbol{V}_W/V_Y)]^2)}{V_{R2}} = V_{R1} \tag{32}$$

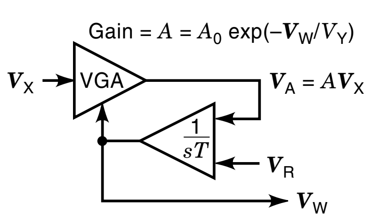


Figure 12. A logamp based on a VGA with exponential gain control.

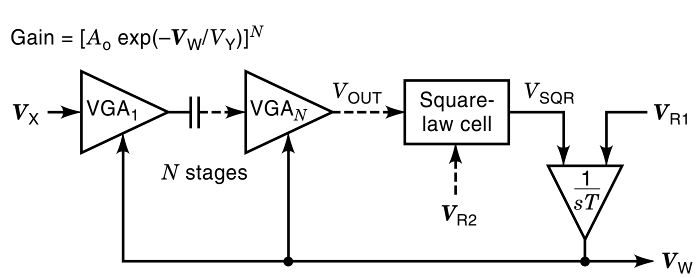


Figure 14. Exponential AGC logamp providing rms metric.

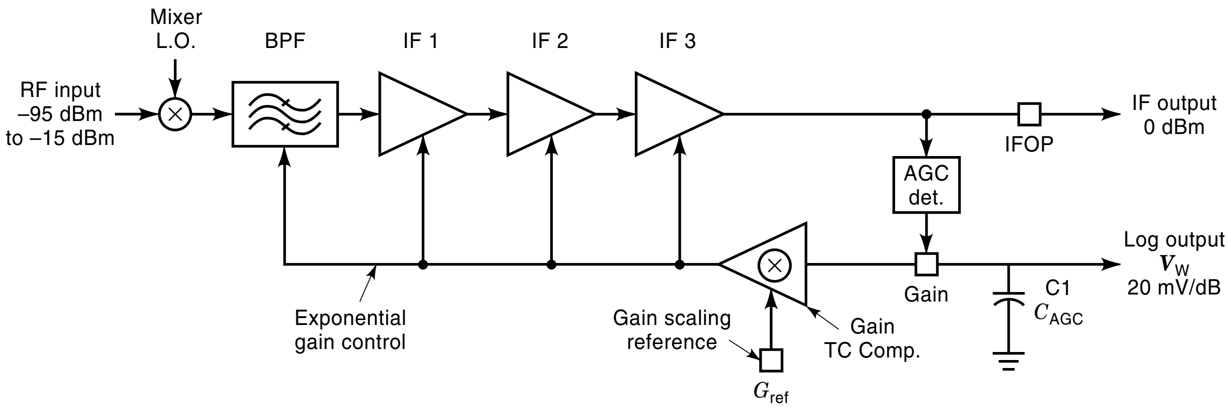


Figure 15. Exponential AGC logamp using AD607.

After some further manipulation, noting that the rms value of the input signal can be equated to V_{IN} through $V_{rms}^2 = \text{Ave}(V_{IN}^2)$, we find

$$\exp\left(-\frac{N\boldsymbol{V}_W}{V_Y}\right) = \frac{\sqrt{V_{R1}V_{R2}}}{V_{rms}A_0^N} \tag{33}$$

$$\boldsymbol{V}_W = \frac{V_Y}{N}\ln\frac{V_{rms}}{V_Z} \tag{34}$$

where the effective intercept voltage is now

$$V_Z = \frac{\sqrt{V_{R1}V_{R2}}}{A_0^N} \tag{35}$$

The possibility of measuring true rms is of great value in power measurements over the entire frequency span from subaudio to microwave. This is a unique capability of this type of logamp, which thus combines the rms feature with the multidecade range of the logarithmic function. As noted earlier, the effect of waveform on logamp behavior can be quite complex, and the progressive-compression logamps to be described next do not respond to the rms input (the true measure of signal power) but in waveform-dependent ways to the input.

Note one further important advantage of this method. The squaring circuit is forced to operate at constant output (V_{R1}). Therefore, it does not need to cope with a large dynamic range, and can be very simple, provided that it exhibits an accurate square-law response on peaks of signals of high crest factor. Note that the amplifiers cells must also have sufficient dynamic headroom for high-crest-factor operation. A monolithic realization of an exponential AGC logamp using a *mean-responding* detector is to be found in the Analog Devices AD607, a single-chip receiver capable of operation from inputs over at least an 80 dB dynamic range, from −95 dBm to −15 dBm (5.6 μV to 56 mV amplitude for sine inputs), at frequencies up to 500 MHz via its mixer (Fig. 15). The mixer and three IF stages are each variable-gain elements, each with a gain range of 25 dB, for a total of 100 dB, providing a generous 10 dB of overrange at both the top and the bottom of the signal range. The gain is controlled by the voltage, which is accurately scaled to 20 mV/dB, and, due to the use of special circuit techniques, is temperature-stable (8).

Figure 16 shows an AGC-style logamp based on a special amplifier topology called an X-AMP, in this case, the AD600. As the name suggests, these provide an exponential gain control function that is very exactly "linear in dB" but does not depend on the translinear properties of the BJT to generate the exponential function (1). The signal input is applied to a passive resistive attenuator of N sections each having an attenuation factor A. Thus, the overall attenuation is A^N. In the AD600, $A = 0.5$ (that is, an R–$2R$ ladder is used) and $N = 7$, so the total attenuation is 42.14 dB. By means that we need not discuss here, the voltage along the continuous "top surface" of the ladder network can be sensed and amplified, and the position of the "slider" can be linearly controlled by a gain control voltage V_G.

It will be apparent that the logarithmic law in this case is built into the attenuator. The advantage of the X-AMP topology is that a fixed-gain feedback amplifier of constant bandwidth, optimized for ultralow noise, can be used. This is directly connected to the signal at maximum gain, while at high gains the signal is attenuated in the passive network, maintaining full bandwidth and linearity. Each section of the AD600 provides a nominal 40 dB gain range (42 dB max), to achieve a >80 dB logarithmic range. The gain-control interface is differential and at high impedance (50 MΩ). The basic gain scaling is 37.5 mV/dB for each section, but this is altered in the Fig. 16 example to provide an overall logarithmic scaling factor of 100 mV/dB.

Further advances in logarithmic amplifiers based on the use of exponential AGC loops are expected. In particular, the use of new monolithic variable-gain amplifier cell topologies combined with wideband square-law detectors has been shown to provide 60 dB of true-power measurement range at frequencies up to 2.5 GHz, placing this technique on an equal footing with the more usual progressive-compression logamps for microwave applications.

PROGRESSIVE-COMPRESSION LOGAMPS

It was shown in Eq. (4) that a logamp must have high gain for small signals. Translinear logamps are of little utility in high-frequency applications, mainly because all the gain is provided by a single opamp having a very limited gain–bandwidth product. Exponential AGC logamps are valuable

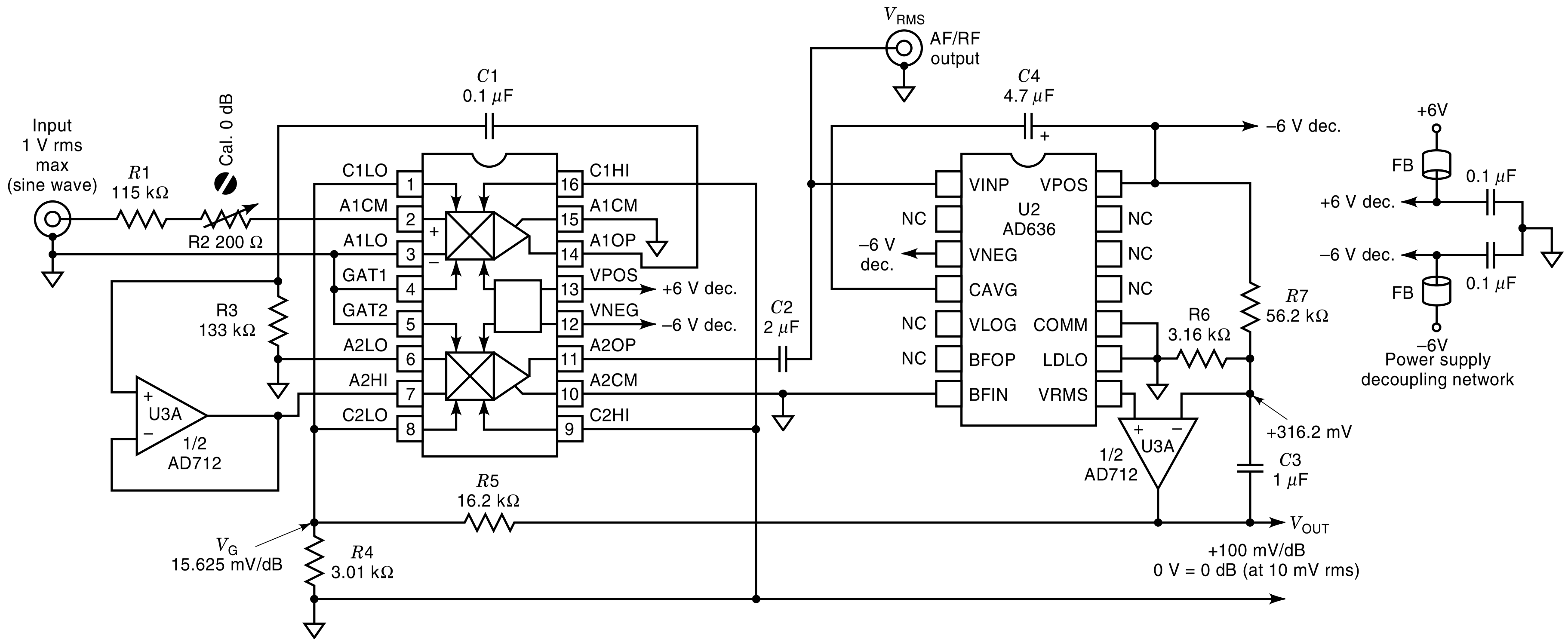


Figure 16. An 80 dB demodulating logamp with rms detector.

in high-frequency applications, where the high gain is provided by several variable-gain stages operating in cascade. But these provide a relatively low loop bandwidth, since signal averaging is needed after the single detector stage. They are therefore useful in determining the *envelope* level of a signal whose power is varying at a moderate rate (from hertz to megahertz).

Baseband and demodulating logamps based on progressive techniques achieve their high internal gain over a large number of cascaded cells and do not involve any kind of feedback (2). Very high gain–bandwidth products (of over 20,000 GHz in practical monolithic products) can thus be achieved. They do not depend on the nonlinearity of a semiconductor device to achieve the desired logarithmic conversion. Rather, they approximate a logarithmic law, in a deliberate and formally correct fashion, through a type of piecewise linear approximation, over a wide dynamic range limited mainly by fundamental noise considerations. Demodulating types provide a logarithmic output that is a measure of signal strength (the RSSI function), and these may also provide a hard-limited output for use in applications where the signal modulation is encoded in FM or PM form. Baseband types provide an output that bears a point-by-point mapping between input and output.

The internal structure of the two types is similar, and cell design techniques can often be shared. We will begin with a development of the underlying theory for a baseband logamp based on a particular type of amplifier cell, then move to cell design that is easier to implement in monolithic form, and finally show how the demodulation function is introduced. The mathematical theory of progressive-compression logamps is poorly developed in the literature, particularly with regard to the essential matter of scaling, that is, the comprehensive consideration of the fundamentals on which the accuracy of this nonlinear function depends. In developing a theory from first principles, we will be paying close attention to this topic.

A baseband logamp operates on the instantaneous value of its input voltage, $\boldsymbol{V}_X$, to generate

$$\boldsymbol{V}_W = V_Y \log(\boldsymbol{V}_X/V_Z) \tag{1}$$

where $\boldsymbol{V}_W$ is the output voltage, $\boldsymbol{V}_X$ is the input voltage, V_Y is the *slope* voltage, and V_Z is the *intercept* voltage. We start from these formal foundations, because we wish to develop a sound theory of progressive-compression logamps, on which the design of robust, manufacturable products can be based, rather than simply discuss logamp behavior in general terms. Our objective will be to find the scaling parameters V_Y and V_Z for specific circuits, of increasing complexity, starting with a baseband logamp built from a chain of simple amplifier cells, each with very simple scaling attributes.

Consider first an amplifier stage having the dc transfer function shown in Fig. 17. For the time being, we will be concerned only with its response to positive inputs, but the theory is completely applicable to bipolar inputs. Furthermore, throughout the development of the theory, we will not be concerned with the frequency-dependent aspects of the amplifier.

The gain for small inputs is A, a well-defined quantity moderately greater than one (typically between 2 and 5), and remains so up to an input (*knee*) voltage of E_K, at which point the gain abruptly drops to unity. We will call this a *dual-gain* amplifier, or $A/1$ amplifier. Thus

$$V_{OUT} = AV_{IN} \qquad \text{for} \quad V_{IN} \le E_K \tag{36}$$

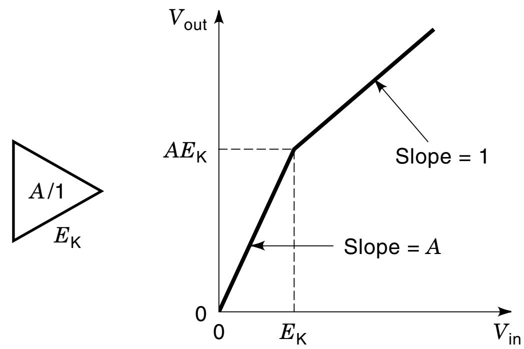


Figure 17. The dc transfer function of an $A/1$ amplifier cell.

and

$$\begin{aligned} V_{OUT} &= AE_K + (V_{IN} - E_K) \\ &= (A-1)E_K + V_{IN} \qquad \text{for} \quad V_{IN} > E_K \end{aligned} \tag{37}$$

We can immediately reach some conclusions about the behavior of a logamp built from a series-connected set of N such amplifier sections. First, because the amplifier behavior just defined is piecewise linear, it follows that the overall function, while more complicated, can never be anything but a piecewise linear approximation. It is also clear that when more stages, of lower gain, are used to cover a given dynamic range, the closer this approximation can be. That is, we can expect the approximation error to be some increasing function of A.

Second, we can be quite certain that the logarithmic slope V_Y and the intercept voltage V_Z in the target function are both directly proportional to the knee voltage E_K, that is, we can expect them to have the general form

$$V_Y = yE_K \quad \text{and} \quad V_Z = zE_K$$

where y is some function of A alone and z is a function of A and N alone. We can predict this simple proportionality with total assurance, because if some polynomial in E_K were needed, there would need to be other parameters with the dimension of voltage within the system, in order to restore dimensional consistency.

Our immediate challenge is to find the functional form of y and z for the cascade of N dual-gain amplifier sections shown in Fig. 18. This will provide a firm foundation for understanding all classes of logamps using progressive compression techniques. The overall input is labeled $\boldsymbol{V}_X$ and the output $\boldsymbol{V}_W$ in observance of the nomenclature already used. For very small inputs, the overall gain is simply A^N. At some critical value $\boldsymbol{V}_X = \boldsymbol{V}_{X1}$ the input to the last (that is, Nth) stage

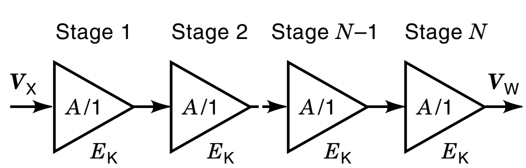


Figure 18. A baseband logamp comprising a cascade of $A/1$ amplifier cells.

reaches its knee voltage E_K. Since the gain of the preceding $N - 1$ stages is A^{N-1}, this must occur at a voltage

$$V_{X1} = E_K/A^{N-1} \tag{38}$$

This is called the *lin–log* transition, because for smaller inputs the cascade is simply a linear amplifier, while for larger values of $\boldsymbol{V}_X$ it enters a region of pseudologarithmic behavior. Above this point, the overall incremental gain falls to A^{N-1}. As the input is raised further, a second critical point is reached, at which the input to the $(N - 1)$th section reaches *its* knee. Then

$$V_{X2} = E_K/A^{N-2} \tag{39}$$

which is simply A times larger than the first critical voltage. We can call this the first *midlog* transition. Above this point, the incremental gain falls by a further factor of A, to A^{N-2}, and so on. It will be apparent that the cascade is characterized by a total of N transitions, the last occurring at $\boldsymbol{V}_{XN} = E_K$. Figure 19 shows the voltages in an illustrative four-stage system at its four transition points, which occur at input voltages separated by a constant ratio, equal to the gain A of each amplifier section. This already looks promising, since if $\boldsymbol{V}_X$ is represented on a logarithmic axis, these transitions occur at equally spaced increments on that axis, corresponding to a ratio of A, while the output changes by equal increments of $(A - 1)E_K$ over this ratio.

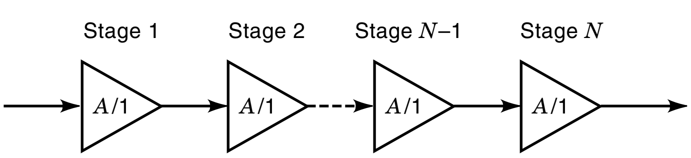
The next step is to find the corresponding values of $\boldsymbol{V}_W$ for all intervals above the lin–log transition and up to $\boldsymbol{V}_X = E_K$. From Eq. (37),

$$V_W = (A-1)E_K + V_{Ni} \tag{40}$$

where V_{Ni} is the input to the Nth stage. But at the lin–log transition, $V_{Ni} = E_K$, and therefore $\boldsymbol{V}_W = AE_K$. Further, because the first $N - 1$ stages of the cascade are still in a linear mode up to the second transition, V_N in this interval is just $\boldsymbol{V}_X A^{N-1}$. Thus,

$$V_W = (A-1)E_K + V_X A^{N-1} \tag{41}$$

We could use this starting point to find an expression for $\boldsymbol{V}_W$ for all values of $\boldsymbol{V}_X$. However, we do not need to delineate all possible values of $\boldsymbol{V}_W$ to determine the effective slope and intercept of the overall piecewise linear function. At the first midlog transition, the output of the $(N - 1)$th stage is simply



E_K/A^3	E_K/A^2	E_K/A	E_K	AE_K
E_K/A^2	E_K/A	E_K	AE_K	$(2A–1)E_K$
E_K/A	E_K	AE_K	$(2A–1)E_K$	$(3A–2)E_K$
E_K	AE_K	$(2A–1)E_K$	$(3A–2)E_K$	$(4A–3)E_K$

Figure 19. Voltages along four cells at the transition points.

AE_K, so the output of the next stage, which is also the final output, is

$$\begin{aligned} V_W &= (A-1)E_K + AE_K \\ &= (2A-1)E_K \end{aligned} \tag{42}$$

So the output increased from AE_K to $(2A - 1)E_K$, an amount $(A - 1)E_K$, for a *ratio* change of A in $\boldsymbol{V}_W$. Continuing this line of reasoning, we can demonstrate that at the next transition $\boldsymbol{V}_W = (3A - 2)E_K$, and so on: *the change in* $\boldsymbol{V}_W$ *is always by the fixed amount* $(A - 1)E_K$ *as* $\boldsymbol{V}_X$ *increases by each factor of* A. Now, a factor of A can be stated as some fractional part of a decade, which is just lgt A, where lgt denotes a logarithm to base 10. For example, a ratio of 4 is slightly over six-tenths of a decade, since lgt 4 = 0.602. We can therefore state that the slope of the output function, corresponding to a line drawn through all the transition points, is

$$V_Y = \frac{\text{absolute voltage change in } V_W}{\text{ratio change in } V_X} = \frac{(A-1)E_K}{\operatorname{lgt} A} \tag{43}$$

As expected, V_Y is proportional to E_K, while the slope is unaffected by the number of stages, N. Since we are here using base-10 logarithms, V_Y can be read as volts per decade. The slope can be approximated by $V_Y = [2.4 + 0.85(A - 1)]E_K$ to within ±2.5% between A = 1.2 and 5.5. To determine the intercept, we insert one point into the target equation and use the resulting value of the slope. We can conveniently choose the lin–log transition, at which point $\boldsymbol{V}_X = E_K/A^{N-1}$ and $\boldsymbol{V}_W = AE_K$. Thus

$$AE_K = \frac{(A-1)E}{\operatorname{lgt} A} \operatorname{lgt} \frac{E_K}{V_Z A^{N-1}} \tag{44}$$

which solves to

$$V_Z = \frac{E_K}{A^{N+1/(A-1)}} \tag{45}$$

Suppose A = 4, N = 8, and E_k = 50 mV. The choice of a gain of 4 for each section is consistent with high accuracy and wide bandwidth in a simple amplifier cell; using eight stages, the dynamic range will be slightly over 4^8, which corresponds to 96 dB; the choice E_K = 50 mV will become apparent later, when it will be shown to arise from $2kT/q$ (= 51.7 mV at T = 300 K). With these figures, the slope evaluates to 0.25 V/decade and the intercept is positioned at about 0.5 μV; the response is shown in Fig. 20. In a practical amplifier handling several decades and operating within the constraints of a 2.7 V supply, a somewhat higher value of E_K could be used; values between 15 mV/dB and 30 mV/dB are common. As noted, the slope and intercept can be readily altered by peripheral modifications.

The output is seen to deviate from the ideal line, with a periodic ripple at intervals of A along the horizontal axis. An analysis of the ripple amplitude, expressed in decibel form, shows that it is dependent only on A:

$$\text{error}_{\text{pk dB}} = 10\,\frac{(A + 1 - 2\sqrt{A}) \operatorname{lgt} A}{A - 1} \tag{46}$$

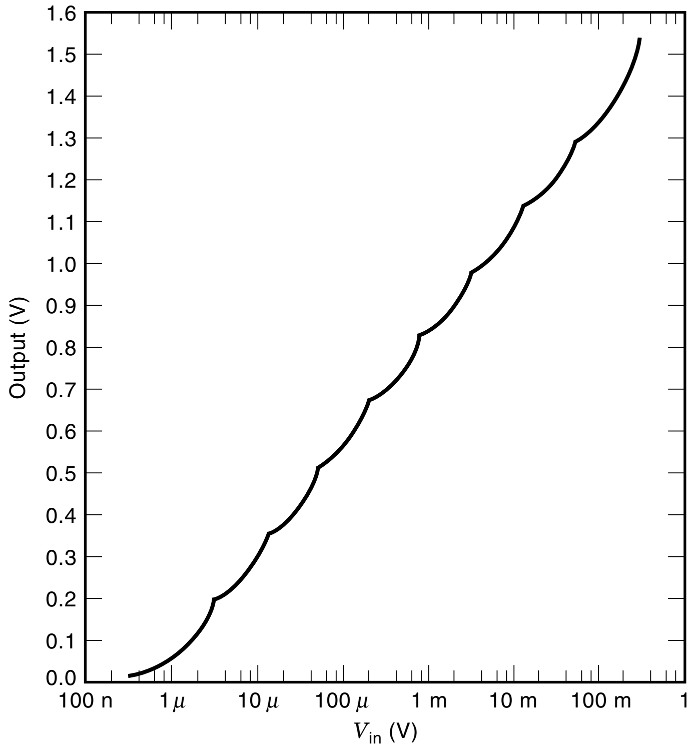


Figure 20. Output of an eight-stage system using $E_K = 51.7$ mV, $A = 4$.

For $A = 4$, this evaluates to 2.01 dB; some other values are 0.52 dB for $A = 2$, 1 dB for $A = 2.65$, 1.40 dB for $A = \sqrt{10}$ (10 dB gain), and 2.67 dB for $A = 5$. However, using practical amplifier cells, which usually have an incremental gain that is a continuously varying function of input voltage, the ripple is much lower in amplitude and roughly sinusoidal, rather than a series of parabolic sections as in the idealized case.

Numerous circuit arrangements are possible to implement a dual-gain stage at low frequencies. An easily realized practical form, providing operation from dc to a few tens of kilohertz, using off-the-shelf components, is shown in Fig. 21. The gain A of each opamp section switches from $\times 3.2$ ($\approx$10 dB) to unity at an effective E_K of 0.436 V, determined by the two-terminal bandgap reference and resistor ratios. The last three cells are slightly modified to improve the low-end accuracy. The ± 1 dB dynamic range is 95 dB (220 μV to 12 V), the intercept is at 100 μV, and the slope is 100 mV/dB (2 V/decade), making decibel reading on a DVM straightforward. The output for a 10 V input is thus 2 lgt($10/10^{-4}$) = 10 V. Figure 22 shows a typical result.

Use of Limiting Cells

A simpler cell topology, more suited to monolithic integration, can achieve the same function at very high frequencies (over 3 GHz in a practical embodiment such as the AD8313), and with better accuracy than $A/1$ cells. In this nonlinear amplifier cell, the incremental gain is A for small signals, but drops to zero for inputs above the knee voltage E_K. This will be called an amplifier–limiter stage, and is denoted by the symbol $A/0$. Figure 23 shows the transfer function of this cell, now for bipolar inputs. The basic equations are

$$V_{OUT} = -AE_K \quad \text{for} \quad V_{IN} < -E_K \tag{47a}$$

$$V_{OUT} = AV_{IN} \quad \text{for} \quad -E_K \le V_{IN} \le E_K \tag{47b}$$

$$V_{OUT} = AE_K \quad \text{for} \quad V_{IN} > E_K \tag{47c}$$

Figure 24 shows the structure of a baseband logamp made up of N such $A/0$ stages. It will be immediately apparent that we can no longer use just the output of the final stage, since as soon as this stage goes into limiting, when $\boldsymbol{V}_X = E_K/A^{N-1}$, the output will simply limit at AE_K and will not respond to further increases in $\boldsymbol{V}_X$. To generate the logarithmic response, the outputs of all stages must be summed. The milestones along the log-input axis are at exactly the same values of $\boldsymbol{V}_X$ as for the $A/1$ case; so the challenge is to find the corresponding values of the output $\boldsymbol{V}_W$ for all values of input $\boldsymbol{V}_X$ up to, and slightly beyond, E_K.

For small inputs, below the lin–log transition, and for equal weighting of the individual cell outputs,

$$\begin{aligned}\boldsymbol{V}_W &= A^N\boldsymbol{V}_X + A^{N-1}\boldsymbol{V}_X + \cdots + \boldsymbol{V}_X \\ &= (A^N + A^{N-1} + \cdots + 1)\boldsymbol{V}_X\end{aligned} \tag{48}$$

At the lin–log transition,

$$\begin{aligned}\boldsymbol{V}_W &= (A^N + A^{N-1} + \cdots + 1)E_K/A^{N-1} \\ &= \left(A + 1 + \cdots + \frac{1}{A^{N-1}}\right)E_K\end{aligned} \tag{49}$$

At the first midlog transition,

$$\boldsymbol{V}_W = \left(A + A + 1 + \cdots + \frac{1}{A^{N-2}}\right)E_K \tag{50}$$

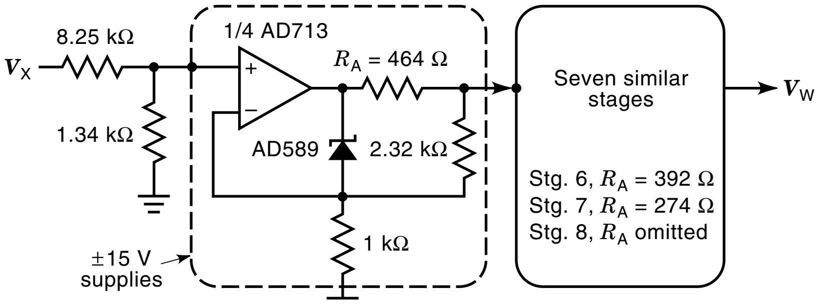


Figure 21. A practical low-frequency logamp using $A/1$ cells.

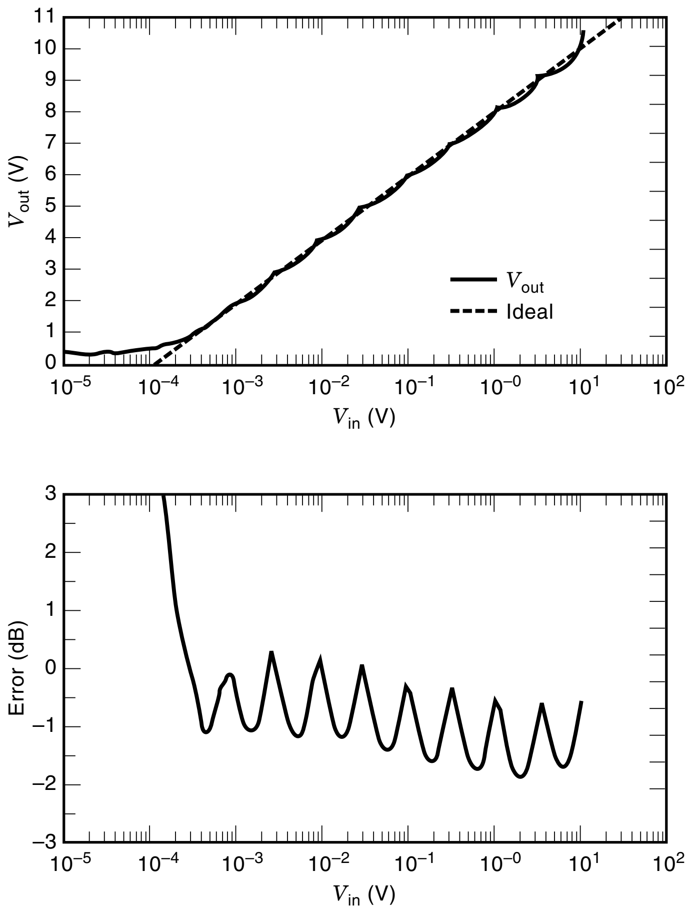


Figure 22. Measured output and absolute error (dB).

Between these two inputs, the output has changed by

$$\Delta \boldsymbol{V}_{\mathrm{W}} = \left(A - \frac{1}{A^{N-2}}\right) E_{\mathrm{K}}$$

while the input increased by a factor A, or lgt A decades. Thus, the slope for this first interval, measured on the transition coordinates, is

$$V_{\mathrm{Y1}} = \frac{(A - 1/A^{N-2})E_{\mathrm{K}}}{\mathrm{lgt}\, A} \tag{51}$$

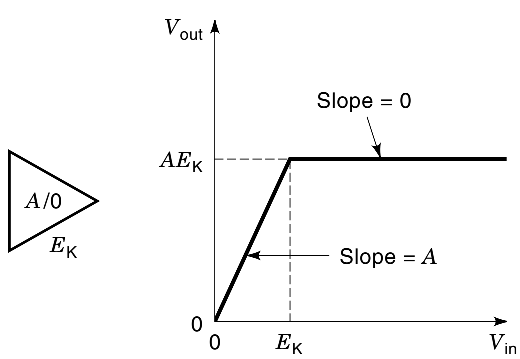


Figure 23. The dc transfer function of an $A/0$ cell.

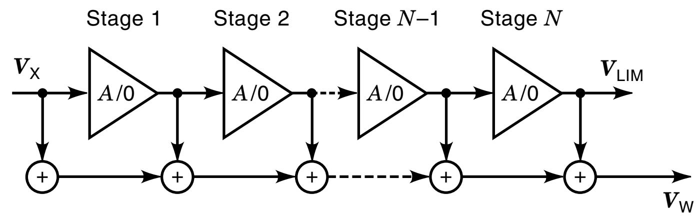


Figure 24. A baseband logamp using $A/0$ stages.

For typical values of A and N, this is very close to AE_{K}/lgt A. At the second midlog transition,

$$\boldsymbol{V}_{\mathrm{W}} = \left(A + A + A + 1 + \cdots + \frac{1}{A^{N-3}}\right) E_{\mathrm{K}} \tag{52}$$

Thus, over the second interval the slope is

$$V_{\mathrm{Y2}} = \frac{(A - 1/A^{N-3})E_{\mathrm{K}}}{\mathrm{lgt} A} \tag{53}$$

Again, for typical values of A and N, this remains close to AE_{K}/lgt A. For example, if $A = 4$ and $N = 8$, the exact value of V_{Y2} is $6.642E_{\mathrm{K}}$, while the approximate value is $6.644E_{\mathrm{K}}$. It is therefore reasonable to use the expression

$$V_{\mathrm{Y}} = \frac{AE_{\mathrm{K}}}{\mathrm{lgt}\, A} \tag{54}$$

for the slope of this logamp over the entire lower portion of its dynamic range. It can be shown that there is a slight reduction in the slope over the last few transition intervals. This artifact can be corrected, and the top-end logarithmic conformance in a high-accuracy design can be improved, by simply using a higher summation weighting on the output from the first stage, as will be shown.

To determine the intercept, we follow the procedure used earlier: insert one input–output point into the target equation and use the known value of the slope. The solution is almost identical to that derived for the system using $A/1$ cells, given in Eq. (45). The ripple calculation also follows the same approach as used above and yields essentially identical results.

The top end of the dynamic range gradually deviates from the slope established at lower levels when the $A/0$ system is used. This can be corrected by a technique first used in the AD606. The analysis lies beyond the scope of this review; the result is that the weighting of just the voltage at the input must be altered by the factor $(A - 1)/A$; when this is done, the slope for *all* intervals is now exactly as given in Eq. (54). It is of interest to note that the slope has a minimum sensitivity to the actual value of the basic gain when this is set to $A = e$ (whatever the base of the logarithm). Thus, the logamp scaling can be rendered less dependent on lot-to-lot variations in gain (for example, due to ac beta variations) by using a gain close to this optimum. Note also that the slope function for the $A/1$-style logamp, namely $V_{\mathrm{Y}} = (A - 1)E_{\mathrm{K}}$/lgt A, does not behave so helpfully: it merely increases with A, and can be approximated by $V_{\mathrm{Y}} = [2.4 + 0.85(A - 1)]E_{\mathrm{K}}$ to within ±2.5% between $A = 1.2$ and 5.5.

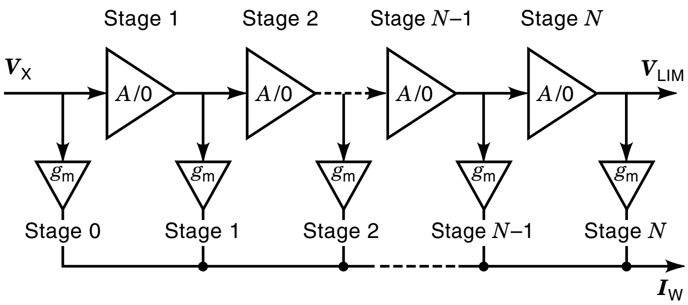


Figure 25. An $A/0$ baseband logamp using g_m cells for summation.

Signal Summation

Figure 24 was unclear about the way in which the signals along the amplifier chain are summed. In a monolithic circuit, this will be effected in the current domain, using a transconductance (g_m) stage at each cell output. This approach is appealing for three reasons. First, current-mode signals can be summed by simply connecting the outputs of all stages together: the conversion back to voltage form can then be accomplished using either a simple resistive load or a transresistance stage. Second, they provide *unilateral* transmission to the output nodes, minimizing the likelihood of unwanted reverse coupling to sensitive nodes early in the signal path.

Finally, they provide the means to decouple the slope calibration from the parameters that control the behavior of the main amplifier. This last benefit is central to the scaling of monolithic logamps using differential bipolar pairs as the gain cells, since the voltage E_K, which controls the slope in all the structures considered so far, is proportional to absolute temperature. If totally linear, the interposition of these g_m cells would make no difference to the analysis, but because in practice they are also in the nature of analog multipliers (being constructed of bipolar differential pairs, whose transconductance is proportional to the bias current), we have full and independent control of the temperature behavior of scaling.

A voltage input V_j (the signal at any node j) generates a current output GV_j. The maximum output from a $G/0$ stage (for $V_j \geq E_K$) is GE_K, fully analogous to AE_K for the voltage-gain stage. The dimensional change inherent in the g_m stage means that this peak output is a current, which will here be called I_D. The subscript D refers to "detector," anticipating the function provided by this cell in demodulating logamps, though we are here still considering baseband operation. The currents I_D, which will be provided by a precision biasing means, control the logarithmic slope. The summed outputs are converted back to the voltage domain using a simple load resistance R_D, or a transresistance stage of the same effective scaling. We will define a parameter

$$V_D = I_D R_D \tag{55}$$

Figure 25 shows the revised scheme. Consistent with summing all the voltages at the amplifier nodes, we have added another g_m stage at the front, and labeled this the 0th cell. The current I_D for this cells is altered in according with the above theory to improve the logarithmic-law conformance at the top end of the dynamic range. With the modified weighting D_0 on just the 0th $G/0$ stage, it can be shown that the change in $\boldsymbol{V}_W$ between *any* adjacent pair of transitions is exactly V_D. Thus

$$V_Y = \frac{V_S}{\lg t\, A} \tag{56}$$

that is, the voltage AE_K has been replaced by a stable, independently controllable parameter. The intercept, however, remains proportional to E_K, which will be PTAT when in a typical monolithic implementation. This will be addressed later.

Fully differential topologies are generally used in monolithic logamps, since they have a high degree of immunity to noise on supply lines and can provide good dc balance. All signals, including the summation signals from the $G/0$ stages, have completely defined current circulation paths, keeping unwanted signals away from common and supply lines. At the very high gains and bandwidths typical of multistage logamps, only a very small amount of feedback from a downstream stage to the input may cause oscillation. For example, an 80 dB amplifier chain having an overall bandwidth of 2 GHz has a gain–bandwidth product of 20,000 GHz.

WIDEBAND MONOLITHIC LOGAMPS

Monolithic logamps of the progressive compression type, utilizing bipolar technologies, have been developed to a high level of refinement during the past decade. A "workhorse" gain cell, providing both the $A/0$ and $G/0$ functions, based on the bipolar differential pair, allows the easy implementation of both baseband and demodulating logamps, which can operate on a single 2.7 V supply voltage or lower.

We will first discuss baseband logamps. The design of demodulating logamps, which is mainly a matter of adding suitable detector (rectifying) cells to a structure that is otherwise very much like a baseband logamp, is presented later.

Figure 26 shows the ubiquitous bipolar differential pair with resistive loads, a simple but versatile amplifier–limiter cell, which, with special biasing techniques, can have accurate gain even when the transistors have finite beta, ohmic resistances, and other imperfections. Using ideal transistors, the small-signal gain is

$$A = \frac{\partial V_{\mathrm{OUT}}}{\partial V_{\mathrm{IN}}} = \frac{R_C}{r_E} = \frac{R_C I_T}{2V_T} \tag{57}$$

Thus, the tail current I_T should be PTAT if the gain is to be temperature-stable. It is important to note that *the gain is a*

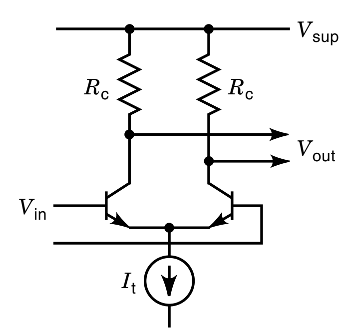


Figure 26. Differential amplifier–limiter–multiplier cell.

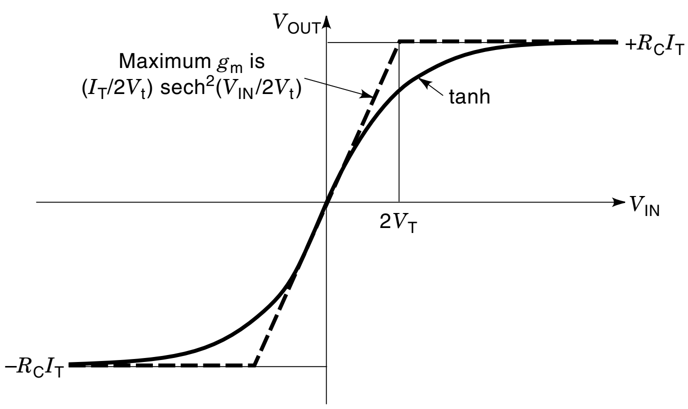


Figure 27. Fitting $A/0$ to the differential amplifier's tanh function.

linear function of I_T, in other words, this is also a *multiplier cell,* an important asset in the development of the $G/0$ and detector cells.

The peak differential output is

$$\pm V_{\mathrm{OUT\,max}} = \pm R_C I_T = \pm 2V_T A \tag{58}$$

Thus, a 10 dB amplifier ($A = \sqrt{10}$) has a peak output of 51.7 mV $\times$ 3.162 = 163.5 mVP. (The suffix P indicates a PTAT quantity, referenced to T = 300 K.) Without further consideration of the precise nonlinearity of this stage, we can already fit this behavior to that of the ideal $A/0$ cell, noting that, in general, an amplifier with a gain of A that limits at an output of $2V_T A$ implies a knee voltage of

$$E_k = 2V_T \tag{59}$$

The full form of the transfer function is

$$V_{OUT} = R_C I_T \tanh(V_{IN}/2V_T) \tag{60}$$

Figure 27 shows how this fits the $A/0$ approximation. Because the transition from a gain of A to a gain of zero is smooth, we can expect the ripple in the log conformance of an amplifier constructed from such cells to be lower than that using ideal $A/0$ stages with abrupt gain transitions, and such is the case. In fact, the tanh function is highly desirable in this application.

The input-referred noise spectral density of this cell evaluates to

$$e_n = \frac{0.9255\,\mathrm{nV/Hz}^{1/2}}{\sqrt{I_T}} \tag{61}$$

when I_T is expressed in milliamperes. The attainment of low noise is very important for the first one or two stages of a logamp. To conserve overall current consumption, a *tapered* biasing scheme is useful in a fully monolithic multistage design: the first stage will be scaled to use a higher tail current, with a corresponding reduction in R_C and a proportional increase in the size of the transistors. This is done in the AD608, where the first stage operates at $4I_T$, the second at $2I_T$, and all further stages at I_T; though not completely optimal, these are convenient in that the transistor and resistor sizes bear simple binary ratios. Similar methods are used in the highly calibrated laser-trimmed AD8306, the low-cost AD8307, the general-purpose 500 MHz AD8309 with limiter output, and the 0.1 GHz to 2.5 GHz power-controlling AD8313.

A valuable property of this gain cell is that, for moderate gains, it can be dc-coupled and cascaded indefinitely without level-shifting or other intermediate components, such as emitter followers. Under zero-signal conditions, all transistors in these cells operate at zero collector–base bias V_{CB}. Using a product $R_C I_T = 8V_T$ = 206.8 mVP, a gain of $\times 4$ (12.04 dB) can be maintained over the temperature range -55°C to $+125$°C using a supply voltage of only 1.2 V. Since even lower gains may be used in a wideband logamp, it will be apparent that single-cell operation is possible. A high-f_t process is important in minimizing the ac loading of each cell by the next.

However, most practical designs achieve higher versatility through the use of a 2.7 V minimum supply, and these may also include the use of emitter followers between stages, whose use increases the bandwidth and improves robustness, through the power gain afforded and consequent reduction in cell loading, and also because this increases the V_{CB} from zero to V_{BE}, or about 800 mV. Furthermore, the overload behavior of the gain cells is improved, by avoiding saturation when in the limiting condition. However, for a 12 dB gain, now a supply voltage of at least 2.2 V is required at -55°C, and the power consumption is roughly doubled.

The basic (unloaded) cell already has some gain error due to finite dc beta (β_0), since the collector current in each transistor is lowered by the factor $\alpha_0 = \beta_0/(1 + \beta_0)$, and the tail current I_T is also reduced by this factor. Fortunately, the multiplier aspect of this gain cell allows us to address this problem very simply, by raising the basic bias current according to suitable corrective algorithms, built into the bias generator; the correction can be precise without knowing the value of beta a priori.

Likewise, real transistors have ohmic resistances, $r_{bb'}$ and $r_{ee'}$, associated with the base–emitter junction, which lower the gain because they increase the effective value of the incremental emitter resistance r_e. Once again invoking the multiplier nature of the gain cell, and noting that by increasing the bias current we can lower the basic r_e and restore the gain,

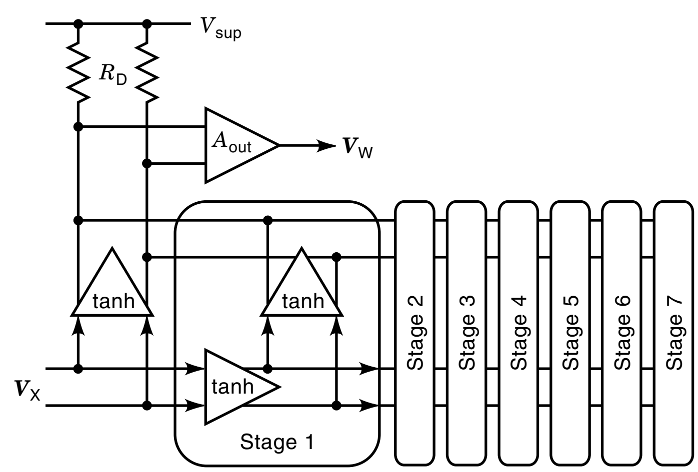


Figure 28. A baseband logamp using $A/0$ cells.

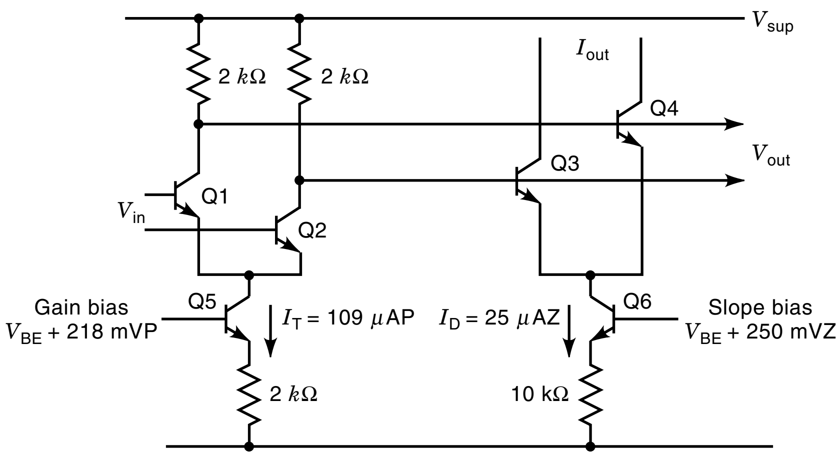


Figure 29. One stage of the monolithic baseband logamp.

correction can be built into the cointegrated bias cell (8,9). Indeed, the utilization of synergistic biasing means is an essential aspect of contemporary monolithic logamp design.

A fairly complete baseband logamp, for simulation studies, is shown in Figs. 28 and 29. It uses R_C = 2 kΩ and I_T = 109 μAP to provide a gain of 12.5 dB (A = 4.217). Seven such cells are used in this example, sufficient to demonstrate typical behavior, plus an eighth g_m cell at the input, to extend the dynamic range upward. The current outputs of all $G/0$ cells are summed by direct connection, converted to a voltage output by the load resistors R_D, and buffered by a difference amplifier, whose gain, A_{OUT}, is chosen to set up a convenient overall logarithmic slope, V_Y. Since the overall gain of the eight cells is 87.5 dB, we can expect to cover a dynamic range of slightly more than this with good accuracy, aided by the extra top-end cell.

A temperature-stable I_D of 25 μA is used. With load resistors of R_D = 2 kΩ, the voltage change over each 12.5 dB interval at the input is 50 mV (that is, 25 μA × 2 kΩ), or 4 mV/dB; using A_{OUT} = 5, the slope voltage is thus 20 mV/dB. The input g_m cell is operated at $I_d(A - 1)/A$, that is, at a current about 30% higher, to improve the top-end law conformance. This is a "true logamp" or "ac logamp," since it can handle either positive or negative inputs. Figure 30 shows the output for small inputs (±10 mV), and the difference between this and the ideal $\sinh^{-1}$ response, as formulated in Eq. (8), exactly scaled by 0.4 V/ln 10; the peak error of ±1.5 mV amounts to ±0.075 dB.

Driving the input over a much larger input voltage range, using an exponential dc sweep, we obtain the output shown in Fig. 31; the intercept occurs at 0.5 μV. The middle panel shows that the dynamic range for a ±1 dB error extends from 1 μV to 60 mV, that is, 95.6 dB. The lower panel shows that the log ripple (the deviation from an ideal logarithmic response) is ±0.06 dB. Note that with I_T = 109 μA and I_D = 25 μA, we have used only 963 μA, including the top-end correction, or 2.6 mW from a 2.7 V supply. Results of this sort demonstrate that amazingly accurate performance is possible using very simple, low-power cells; these simulation-based predictions have been amply proven in numerous commercial products. Using low-inertia IC processes, several-hundred-megahertz bandwidths can be achieved at milliwatt power levels.

In practice, the accuracy of a baseband logamp at the lower end of the input range will be degraded by the input offset voltage V_{OS} and by noise. However, once good basic adherence to the logarithmic function has been achieved, V_{OS} can be eliminated in critical applications by a corrective loop, while noise (in this *nondemodulating* logamp) can be filtered even after the final amplifier stage. Finally, a fully robust design requires close attention to many biasing details. The cells generating I_D and I_T will be specially designed to essentially eliminate the sensitivity of all scaling parameters to temperature, supply voltage, lot-to-lot variations in beta and ohmic resistance, device mismatches, and so on. While most of these go beyond the scope of this review, we need to briefly discuss how the intercept may be stabilized.

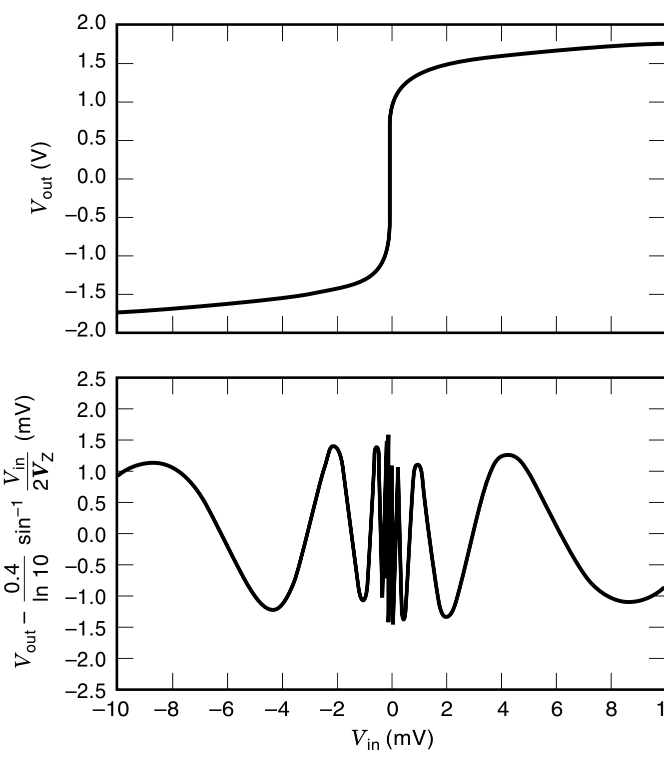


Figure 30. Output of the baseband logamp and deviation from the $\sinh^{-1}$ function.

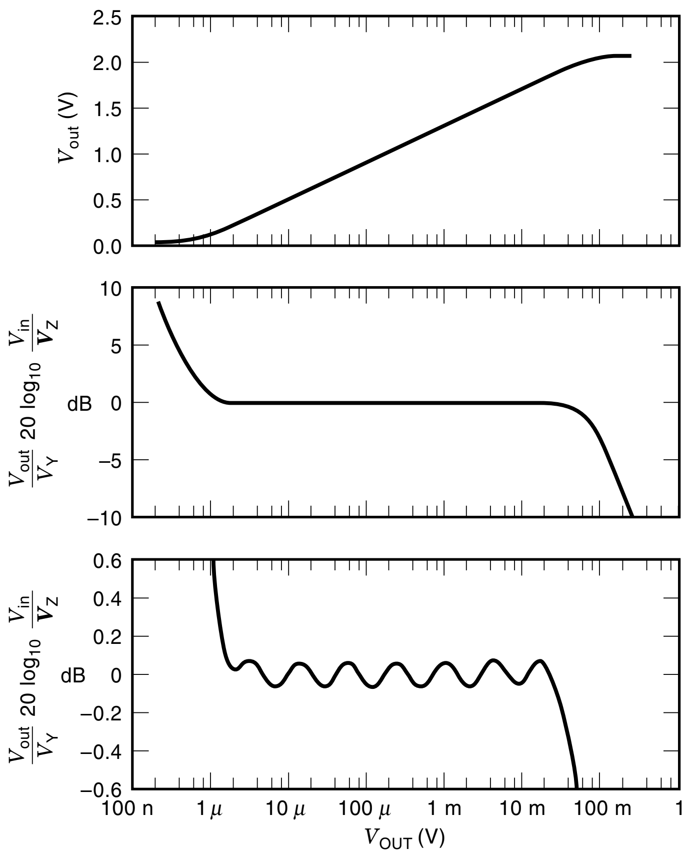


Figure 31. Output and deviation from ideal log function (expanded in lower panel).

Temperature Stabilization of Intercept

The use of the $G/0$ stages eliminates the PTAT form of E_K from the slope calibration, but we have not yet addressed this direct temperature dependence of the intercept. It can be expressed in decibels per degree Celsius in the following way. The fundamental Eq. (1) can be written

$$\boldsymbol{V}_W = V_Y \log\left(\frac{\boldsymbol{V}_X}{V_{Z0}}\frac{T_0}{T}\right) \tag{62}$$

for an input $\boldsymbol{V}_X$ that is temperature-stable and an intercept V_Z that is PTAT, having a value of V_{Z0} at T_0. The decibel variation in output for a one-degree temperature change in the vicinity of $T_0 = 300$ K is

$$\Delta_{dB} = 20\,\mathrm{lgt}(300/301) = -0.029\,\mathrm{dB}/^\circ\mathrm{C} \tag{63}$$

For a −55°C to +125°C operating range the total change in intercept is over 5 dB.

There are several solutions to this problem. Two methods, both of which are provided in the AD640, will be described. First, note that if we could somehow multiply $\boldsymbol{V}_X$ by a factor that is PTAT, we would completely cancel the reciprocal factor in Eq. (62). It makes no sense to consider doing this using an analog multiplier based on, say, translinear techniques (though in principle that is possible): it would be noisy, its dynamic range would likely be much less than that provided by the logamp; it would have dc offsets, it would limit the attainable bandwidth, and so on.

However, a passive attenuator with PTAT loss has none of these problems. We have already used such an attenuator in the translinear logamp described earlier. In a monolithic implementation, the PTAT resistor can be made using, in part, the aluminum interconnect metalization, which may be quite low in value. If the input impedance is 50 Ω, about 3 Ω of aluminum is needed, providing a nominal attenuation of about 24 dB, with a corresponding increase in the intercept. In the AD640 (where the input resistance is 500 Ω), laser trimming is used to eliminate the large ratio uncertainty in the two resistor layers. The use of an attenuator has the added advantage, in certain cases, of raising the upper end of the dynamic range, from about ±60 mVP for the basic structure described here, to ±1 V. The attenuator has no effect on the slope voltage V_Y. This method provides an essentially perfect fix, without any artifacts. The intercept remains within about 0.2 dB over the full military temperature range, and the limits of the dynamic range are temperature-invariant.

The second approach illustrates another use of Eq. (2), which showed that the intercept can be moved by simply adding or subtracting a suitable offset to the output. In this case, the offset must vary with temperature. For a current-summing system, this can be achieved most simply by adding a PTAT current directly to one of the log-summing nodes. For a demodulating logamp, the output is unipolar, and the correction current is easily added. In the case of the baseband logamp, it may be achieved just as readily, by using a correctly proportioned PTAT bias current for the last $G/0$ cell in place of the stable current I_D.

Figure 32 shows the result of using this simple modification on the baseband logamp shown in Fig. 28. This result shows that the left–right shift in the basic response remains; that is, unlike the use of a PTAT attenuator, this technique slightly erodes the dynamic range. It can be seen that the net ±1 dB range now extends from 2.5 μV (at 125°C) to 50 mV (at −55°C), which is 86 dB. (In practice, the reduction will be less, because the lower limit of the dynamic range is determined by the noise floor.) On the other hand, the intercept now varies by only +0.4 dB at either temperature extreme.

Looking at Eq. (62) more carefully, it is apparent that the desired temperature compensation shape is not quite PTAT, even though the intercept is. The reason is simply that there is a logarithmic transformation between input and output axes. A more exact function, of the form $V_{FIX} = V_Y \log(T/T_0)$, can be readily generated in a monolithic circuit, and is used, for example, in the AD640.

Range Extension Using an Auxiliary Logamp

The top end of the dynamic range of this BJT logamp is limited by the signal capacity of the first $G/0$ cell. However, by using two logamps operating in a parallel manner, the range can be considerably extended (Fig. 33). The L amplifier, handling the lower part of the range, is driven directly by the signal; it would be optimized for ultralow noise, and given special attention with regard to thermal behavior in overload. The U amplifier is driven via the attenuator and handles the upper end of the dynamic range.

In this way, a very sensitive low-end response can be combined with the ability to handle input amplitudes that, with

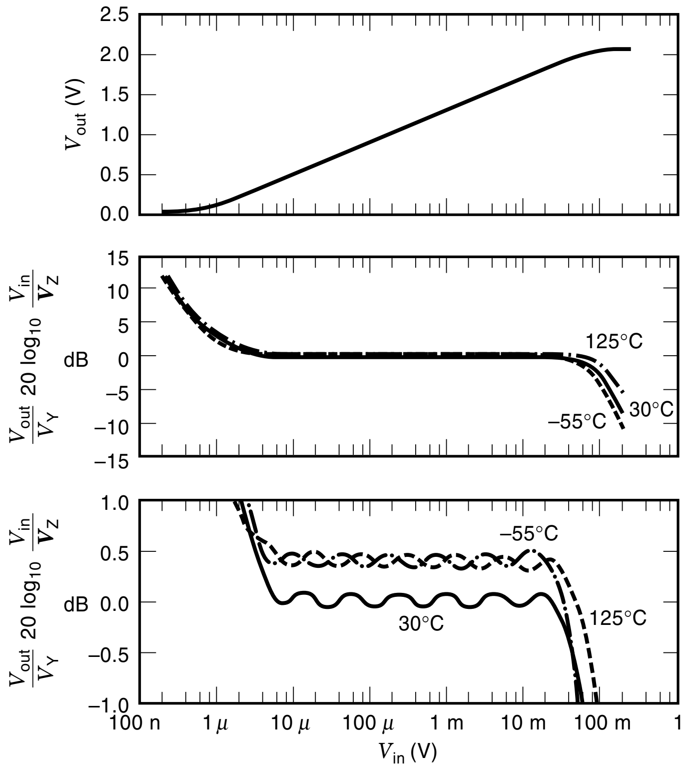


Figure 32. Output (top panel) and error (lower panels) of baseband logamp after intercept compensation, at $T = -55°C$ (solid curves), 30°C (dashed curves), and 125°C (dot–dash curves).

appropriate design techniques, can be as large as the supplies (for example, ±5 V), provided that the emitter–base breakdown voltage of the input transistors in the L amplifier is not exceeded. (The U amplifier never sees large inputs.) Both baseband and demodulating types can benefit from this treatment.

The choice of the attenuation ratio depends on several considerations, but it must have one of the values A, A^2, A^3, ... Thus, using $A = 4$, the choices would be 4, 16, 64, . . ., extending the 1 dB upper input from ±62.5 mV to ±0.25 V, ±1 V, ±4 V, ... A somewhat different approach is used in the logamp section of the AD608, the AD8306, and the AD8309. The upper end of the dynamic range in these cases is extended using independent attenuator sections, each followed by a detector ($G/0$) cell.

The demodulation function can be introduced by a modification of the basic structure. The literature contains descriptions of many different practical ways to go about this, and other methods are found in commercial products such as the

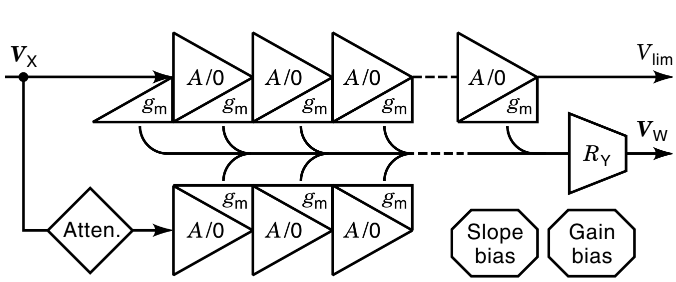


Figure 33. Wide-range logamp uses two parallel-driven sections.

Motorola MC3356. A large amount of the knowledge about logamps relates to discrete designs, and must, in a time of 30 GHz monolithic technologies, be regarded as all but obsolete.

The main features of logamps intended for the rapid determination of the envelope amplitude of an RF input are similar to those delineated for a baseband logamp:

- The necessary high gain is distributed over many low-gain, high-bandwidth stages of the amplifier–limiter, or $A/0$, type, which subject the input to a process of progressive compression, and the logarithmic function approximation is analyzed using essentially the same mathematics as for the baseband logamp.
- The output of all the amplifiers stages, plus the direct input, is summed through the mediation a type of transconductance ($G/0$) cell; similar small adjustments to the weighting of these cells can be used to improve the accuracy of the law conformance.
- Differential circuit topologies are used to achieve a high degree of HF balance and to minimize common-mode effects, such as the intrusion of spurious signals and noise from the power-supply. L and U sections are used to extend the dynamic range; the attenuator sections are also built in differential form.
- The stabilization of the E_K proportional logarithmic intercept over temperature can be achieved using either signal multiplication by a PTAT attenuator at the input or the addition of a temperature-dependent correction at the output.

The chief *differences* are that:

- The logamp now incorporates the demodulation function. Its input is an ac signal (for example, a modulated sinusoidal carrier of from 100 kHz to several gigahertz, or possibly an audio signal), and its output is a single-polarity (usually positive) voltage proportional to the logarithm of the amplitude of this input.
- This output is generated by rectifying the signals along the amplifier chain and then averaging the resulting fluctuating output over some finite time in a low-pass filter, usually integrated into the output amplifier. Either full-wave or half-wave rectification can be used. The former is preferable, since it doubles the carrier frequency and thus reduces the residual carrier feedthrough at the output of the low-pass filter. These rectifier cells (usually called *detectors* in a logamp context) operate in a transconductance mode.
- The cyclical ripple in the error curve is lower in a demodulating logamp, for similar values of A, than for baseband operation, because the instantaneous value of a sinusoidal input voltage is sweeping over a wide range of values during each cycle of the RF carrier. It is roughly halved for sine excitation.

The design of monolithic demodulating logamps is a specialist topic, the details of which have yet to be fully described in a comprehensive text. The material provided here has provided the essential framework and emphasized the importance of approaching all aspects of logamp synthesis with the fundamental issue of scaling firmly in mind.

BIBLIOGRAPHY

1. B. Gilbert, A low-noise wideband variable-gain amplifier using an interpolated ladder attenuator, *IEEE ISSCC Tech. Digest,* 1991, pp. 280–281, 330.
2. R. S. Hughes, *Logarithmic Video Amplifiers,* Norwood, MA: Artech House, 1971.
3. Analog Devices, *High-Speed Amplifier Applications,* Boston: Seminar Notebook, 1992.
4. I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products,* New York: Academic Press, 1980, Art 3.321.3, p. 307; Art. 4.333, p. 574.
5. B. Gilbert, Translinear circuits: A proposed classification, *Electron. Lett.,* **11** (1), 14–16; Errata, **11** (6), 136.
6. B. Gilbert, Translinear circuits: An historical overview, in *Analog Integrated Circuits and Signal Processing,* Norwell, MA: Kluwer, 1996, Vol. 9, pp. 95–118.
7. W. L. Paterson, Multiplication and logarithmic conversion by operational amplifier–transistor circuits, *Rev. Sci. Instrum.,* **34**, 1311–1316, 1963.
8. B. Gilbert, Advances in BJT techniques for high-performance transceiver, *Eur. Solid-State Circuits Conf. Rec.,* Sept. 1997, pp. 31–38.
9. B. Gilbert, IF amplifiers for monolithic bipolar communications systems, presented at *EPFL Electronics Laboratories Advanced Engineering Course on RF Design for Wireless Communications Systems,* Lausanne, 1996.

Reading List

B. Gilbert, *Temperature compensated logarithmic circuit,* US patent No. 4,604,532, August 1986.

B. Gilbert, Aspects of translinear amplifier design, in W. Sansen et al. (eds.), *Analog Circuit Design,* Norwell, MA: Kluwer, 1996, Part III, pp. 257–290.

BARRIE GILBERT
Analog Devices Inc.

LOGIC. See FORMAL LOGIC; LOGIC DESIGN.