II Function Fitting

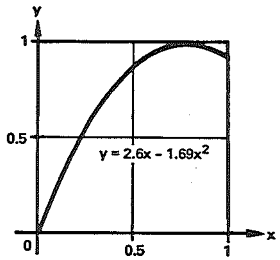
Chapter 1

Function fitting, in general, is the translation of a mathematical or empirical relationship (between a *dependent* variable –output– and one or more *independent* variables –inputs) from one medium, such as a table, a mathematical formula, or a set of curves, to another medium, usually a physically-realizable device or system having an output and one or more inputs. A function may be fit by an "exact" relationship, or it may be approximated.

There are two basic steps in function fitting (Figure 1). The first is the establishment of a close-enough approximation in terms of ideal building blocks, that is, a conceptual model. The second is the successful employment of actual devices to embody the function within an acceptable set of constraints, such as range, scale factor, accuracy, drift, response time, complexity, cost, etc.

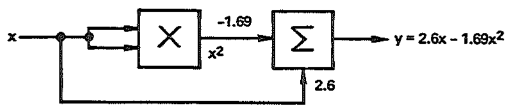
In this chapter, we shall consider functions that can be realized by circuits which embody "instantaneous dc" relationships between sets of voltages having a limited (not infinite) dynamic range of variation. For the most part, we consider single-valued functions, i.e., each set of input values creates a unique value of output, independent of history. (Through switching, or hysteresis, though, single-valued functions can become multi-valued.) We omit, to a great extent, the fitting of dynamic response relationships, such as linear transfer functions, or filter characteristics, whether in the frequency domain or the time domain, since the circuit theorists and filter designers have explored that area with great enthusiasm and have produced a profuse output of published material.

We also assume that the functions to be fit by analog techniques are statistically satisfactory. That is, we do not consider the class of problem in which the data points characterizing a relationship are scattered. Any necessary statistical smoothing has already been accomplished. However, it is possible to use random noise multiplicatively to simulate operators having appreciable stochastic fluctuation.

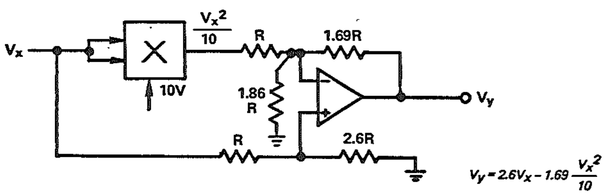


x	y
0	0
0.1	0.243
0.2	0.452
0.3	0.628
0.4	0.77
0.5	0.878
0.6	0.952
0.7	0.992
0.8	0.998
0.9	0.971
1.0	0.91

a. Curve table, and equation



b. Block diagram



c. Circuit, scaled to 10V full scale, $V_y = 10y$, $V_x = 10x$

Figure 1. Analog function fitting

The limitations of space and time do not permit as thorough a coverage of the subject as it deserves, but we do hope to leave the reader with the outlines of analog function-fitting techniques (both as a bag of tricks and a guide to further thinking), some pointers for successful application, and a few examples.

WHY ANALOG FUNCTION FITTING?

We indicated in the introductory chapter that measurements of phenomena that one might wish to obtain linearly often come out nonlinearly. For example, a thermocouple is a cheap and simple (but low-level) means of measuring temperature differences. It can consist of as little as three segments of wire, with two thermal junctions at different temperatures. But its output voltage departs significantly from a linear function of temperature, depending on the materials it is made of, and the temperature range under consideration. Other examples might include the Wheatstone bridge, a simple way of measuring resistance deviation, but nonlinear for large deviations, and the deflection of an oscilloscope's spot, a nonlinear function that depends on the electron gun, the shape of the tube face, and the voltages applied to the X and Y axes.

It is possible to linearize these measurements (i.e., obtain an output reading that linearly represents the desired indication) by using nonlinearity to either compensate for or obtain the inverse of a nonlinear transducer function. The use of function-fitting techniques for such applications is discussed in Chapters 2-3 and 2-5.

Other applications of analog function fitting include calibration, simulation of nonlinear relationships in analog computers and computer-based instruments, translating indirect measurements into useful form economically, and generating time functions of arbitrary shape.

Nonlinear function fitting can also be performed digitally by read-only memories (ROM's) –often in conjunction with A/D and D/A converters– and by combinations of hardware and software. Decreasing costs and wide availability of digital hardware, plus an ever-increasing library of algorithms would appear to make this approach seem increasingly tempting, despite its inflexibility and complexity. But the cost of analog IC's and modules (both op amps and functional operators) has also decreased dramatically; analog approaches are still a "best buy" for simple relationships, and lower cost (for improved accuracy and increased circuit sophistication) makes them more competitive for increasingly-complex relationships.

RATIONAL FUNCTIONS

The simplest functions to fit, in concept, are those that can be expressed "exactly" by an equation involving basic operations: squaring, rooting, multiplication, addition, logarithms, and arbitrary power and roots. The basic operators mentioned in the introductory chapter can of course be used to fit functional relationships that are identical to the operations they perform.

They can also be combined to fit a wide range of explicit functions of 1 or more variables, such as

$$u = 1 + 0.3w^2 \quad (1)$$

$$u = v\,(\mathrm{r} - w) \quad (2)$$

$$u = v^{\mathrm{m}} + w^{\mathrm{n}} \quad (3)$$

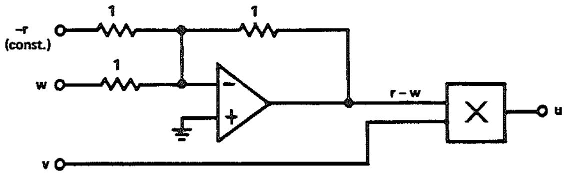
$$u = \frac{v \cdot w}{1 + w^{\mathrm{k}}} \quad (4)$$

$$u = (v + \mathrm{m})\,(w + \mathrm{n})^{\mathrm{r}} \quad (5)$$

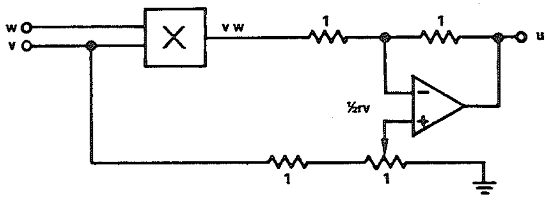
$$u = \frac{v}{1 + \frac{v}{2}} \quad (6)$$

where u, v, and w may be variables, and the other terms fixed or adjustable constants. The summation operations are often performed with op amps.

In some cases, equations can be rewritten for embodiment in several ways. For example, equation (2) can also be written $u = v \cdot \mathrm{r} - v \cdot w$ (Figure 2). If r is a constant, either equation can be embodied with a single multiplier and a single subtractor. In general, one tends to choose the equation that will give the best compromise of error vs. cost. In this instance, if w is always small in magnitude compared to r, the configuration of Figure 2b is probably a better choice, because the more-important term can be handled with linear circuit elements, and errors of the $v \cdot w$ term will be of lower order. On the other hand, if w can be comparable to r in magnitude, it is probably better to take the difference of two large numbers *before* performing a nonlinear operation; therefore one would use the configuration of Figure 2a. The underlying assumption here (nearly always justified) is that accurate linear operations are cheaper than nonlinear operations of comparable accuracy.



a. u = v (r – w), r = constant, preferred when vr and vw are of comparable magnitude



b. u = vr – vw, r = constant, preferred when vr >> vw, as when fitting a function that is essentially linear, with a small deviation.

Figure 2. The way an equation is written affects both circuit configuration and performance

For more-complicated functions, range and error analyses of all the available alternatives should always be performed, either analytically or empirically (if one happens to have "worst-case" components on hand).

SCALE FACTORS

Having optimized the circuit and chosen a set of devices likely to implement it economically, one is confronted by the *scaling* problem, i.e., determining the exact relationship between the electrical circuit and the function it fits, including constants (gains and biases) and the ranges of all voltages.

Every accessible voltage or current in the analog circuit corresponds to a variable in the original functional operation. If it is an important term, requiring good accuracy, its range should be close to full scale of the device producing (or accepting) it; but it should not appreciably exceed the full-scale range, for any combination of inputs or outputs (unless it happens to be a Bounds function). Input and output ranges are usually predetermined by other elements of the overall system, but ranges at intermediate locations are somewhat flexible and can be tailored for optimum dynamic range.

Since the ranges are determined by the configuration, scale factors (i.e., electrical coefficients or "gains.") should be chosen after the configuration has been adopted. It should be noted, though, that availability of appropriate scale factors may be a factor in the choice of configuration.

The experienced instrumenter can usually derive scale factors directly from the equation to be implemented, based partly on intuition, partly on common sense, and partly on a set of well-learned but perhaps unverbalized rules. Others may benefit by observing the following principles:

1. The original equation should be dimensionally correct.

2. If not already in dimensionless form, it can be normalized by multiplying and dividing each variable by a multiple of its range, usually 100%. Consider this example: for

$$y = \mathrm{A}x \sin\theta + \mathrm{K} \quad \text{(Figure 3a)} \tag{7}$$

$$\mathrm{Y_m}\frac{y}{\mathrm{Y_m}} = \mathrm{A\,X_m}\frac{x}{\mathrm{X_m}} \sin\theta + \mathrm{K} \tag{8}$$

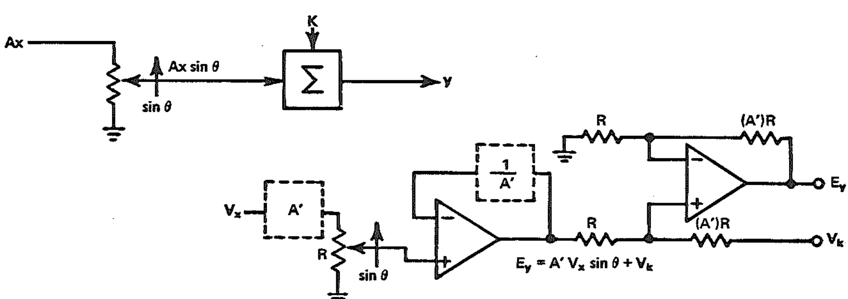


Figure 3a. Block diagram and electrical schematic for linear scaling example

where Y_m, X_m, are the full-range values of x and y. Defining the ratio-to-its-range as the dimensionless variable, and dividing both sides of the equation by Y_m, the equation becomes, in terms of dimensionless variables, y' and x'

$$y' = A\frac{X_m}{Y_m} x' \sin\theta + \frac{K}{Y_m} \tag{9}$$

3. Write the equation of the analogous electrical circuit (Figure 3a), using (unknown) coefficients A′ and K′ to relate the various voltages, and including any known electrical scale factors (such as those inherent in multipliers and log devices).

$$E_y = A'V_x \sin\theta + K'V_{Km} \tag{10}$$

4. To determine the unknown constants, multiply and divide by the expected maximum values of voltage, to normalize the electrical equation:

$$\begin{aligned} E_{ym}\frac{E_y}{E_{ym}} &= A'V_{xm}\frac{V_x}{V_{xm}}\sin\theta + K'V_{Km} \\ \frac{E_y}{E_{ym}} &= A'\frac{V_{xm}}{E_{ym}}\frac{V_x}{V_{xm}}\sin\theta + K'\frac{V_{Km}}{E_{ym}} \end{aligned} \tag{11}$$

The normalized equations, 9 and 11, must be identical, therefore

$$A'\frac{V_{xm}}{E_{ym}} = A\frac{X_m}{Y_m}, \text{ therefore } A' = A\frac{X_m}{Y_m}\frac{V_{xm}}{E_{ym}}$$

and

$$K'\frac{V_{Km}}{E_{ym}} = \frac{K}{Y_m}, \text{ therefore } K' = \frac{K}{Y_m}\frac{E_{ym}}{V_{Km}}$$

5. The electrical constants are now substituted in the electrical system equation (10). The process is by no means as formidable

as it may appear at first glance, because usually, $V_{xm} = E_{ym} = V_{Km} = 10V$; so

$$A' = A\frac{X_m}{Y_m} \text{ and } K' = \frac{K}{Y_m}$$

The scaling, as described in this example, has (so far) applied to an essentially linear circuit, corresponding to the case in which $\sin\theta$ is a dimensionless gain (e.g., a potentiometer setting). The process differs somewhat if the sine function is the output voltage of a sine operator (Figure 3b), and must be multiplied by V_x

$$V_s = V_{sm} \sin\frac{\theta_m}{V_{\theta m}}V_\theta \tag{12}$$

The electrical equation becomes

$$E_y = A'\frac{V_x}{V_r} V_{sm} \sin\frac{\theta_m}{V_{\theta m}}V_\theta + K'V_{Km} \tag{13}$$

where V_r is the multiplier's scale constant. The equation normalizes to

$$\frac{E_y}{E_{ym}} = A'\frac{V_{xm}}{E_{ym}}\frac{V_x}{V_{xm}}\frac{V_{sm}}{V_r}\sin\frac{\theta_m}{V_{\theta m}} V_\theta + K'\frac{V_{Km}}{E_{ym}} \tag{14}$$

Again, recognizing that equations 9 and 14 are identical,

$$A'\frac{V_{xm}}{E_{ym}}\frac{V_{sm}}{V_r} = A\frac{X_m}{Y_m}, \text{ and } A' = A\frac{X_m}{Y_m}\frac{E_{ym}}{V_{xm}}\frac{V_r}{V_{sm}}$$

If $V_{xm} = E_{ym} = V_r = V_{sm} = 10V$,

$$A' = A\frac{X_m}{Y_m} \text{ and } K' = \frac{K}{Y_m}$$

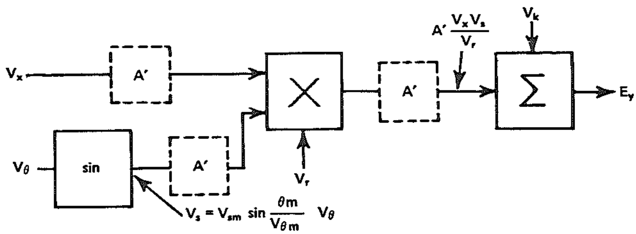


Figure 3b. Electrical block diagram of nonlinear version, showing possible locations for scale factor

It is important to note that the scale factor A′, when associated with a multiplication, can be applied at any one of the three terminals, or distributed among two or more of them, if necessary to optimize the dynamic range for both inputs and the output. For example, if V_{sm} = 5V and V_r = V_{xm} = E_{ym} = 10V, A′ will be doubled. Most likely, a factor of 2 should be applied between the output of the sine operator and its input to the multiplier, if it is desired to make full use of the multiplier's input range.

After the scale factors have been computed, they should be checked, by considering various extremes of input and output signals; any indicated modifications should be made. While the approach suggested here works, it is no better than the assumptions. Awkward assumptions will lead to awkward dynamic ranges.

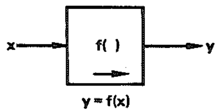
5. Note that the assignment of a "maximum" value E_{ym} to E_y does not automatically *guarantee* that V_y will not exceed full scale, unless the set of normalizing voltages is fully consistent. With practice, one will develop a near-intuitive feeling for proper scale factors and will find much of the above procedure unnecessary to plow through in detail. Incidentally, time-dependent devices may also be scaled in this manner. Where time appears in an equation, it is multiplied and divided by a nominal "unit value," usually 1 second, but often the characteristic time of the slowest integration, in high-speed analog computing devices or systems. ($T \frac{t}{T} = T\,t'$)

In chapter 2–3, it is shown how one might develop and scale a thermocouple-compensation circuit.

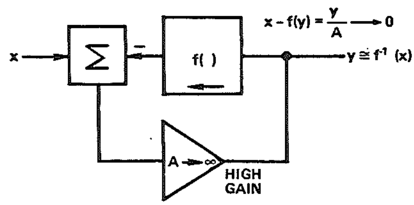
INVERSE FUNCTIONS

If $u = f(v)$, the inverse function, $v = f^{-1}(u)$, may be obtained (in concept) by the use of $f(v)$ in a high-gain negative feedback circuit (Figures 4 and 5a). This is already widely exploited in:

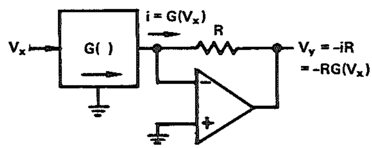
(a) the generation of logarithmic operations. The exponential I-V relationship of a diode in the feedback path of an operational amplifier matches the input current, enforcing a logarithmic output voltage (see Figure 4c and 4d),



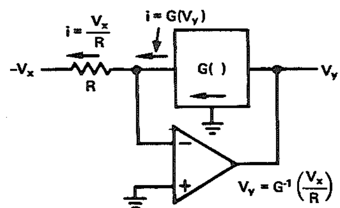
a. Function. Arrow inside block indicates causality



b. Inverse function



c. Direct function, transconductance and op amp



d. Inverse function, with op amp

Figure 4. Direct and inverse functional operations

(b) the use of multipliers for division. The product of one input and the output is made to equal the second input, therefore the output is proportional to their ratio,

(c) the use of squarer-connected multipliers for square-rooting. The product of the output multiplied by itself is made to equal the input, hence the output is the square-root of the input (Figure 6a).

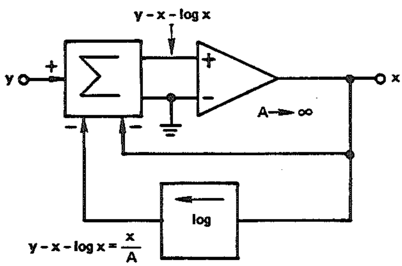
Such schemes can be applied to combined functional operations for generating operators that are more easily obtainable in the inverse form. For example, if $y = x + \log x$, there is no closed-form solution to this transcendental equation if one desires x. One configuration for obtaining x, given y, is a high-gain feedback loop around $x + \log x$, as shown in Figure 5a.

There are a number of evident restrictions to the use of this technique:

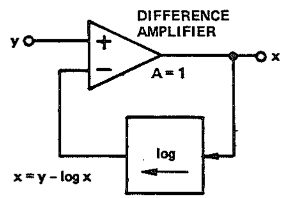
1. The net incremental feedback must be negative over the range of interest.

2. Instabilities resulting in oscillation or "latchup" should be ruled out. Stabilizing and range-limiting circuitry may be necessary, with possible restriction of range, bandwidth, or accuracy. Loop gain and phase shift must be examined under all conditions. Adjustably-offset random noise may be employed as an input to detect sensitive frequency and amplitude bands.

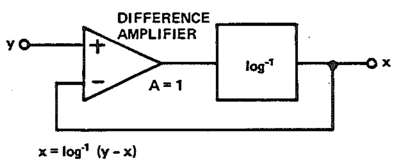
3. In general the functions should be single-valued and monotonic in the range of interest. For example, $\sin^{-1}(x)$ should be limited to within a range of ±90°.



a. Inverse solution of y = x + log x



b. Implicit solution of y = x + log x



c. Another implicit solution of y = x + log x

Figure 5. Inverse and implicit solutions. Note that (a) can be implemented identically to (b), in this case, but (b) is conceptually simpler.

IMPLICIT SOLUTIONS ($x = f\{x, y, ...\}$)

A powerful feedback technique for solving for any variable that can be made to appear twice in an equation (by non-redundant summation, factoring, or other trickery) is the *implicit* use of the variable in solving for itself, without necessitating the explicit use of high gain to enforce the feedback constraints. Figure 5b shows an implicit solution as an alternative to the inverse for obtaining x in $y = x + \log x$. Here are a few additional examples of this use of algebraic analog computing.

1. Square rooting (Figure 6b) Any divider may be used as a square-rooter; non-feedback types tend to be the most successful. If the input, x, is divided by the output, y, to obtain y,

$$y = \frac{x}{y}, \text{ or } y^2 = x \tag{15}$$

and

$$y = \sqrt{x} \tag{16}$$

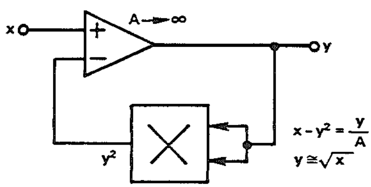


Figure 6a. Square root as an inverse function, using a squarer in a high-gain feedback loop

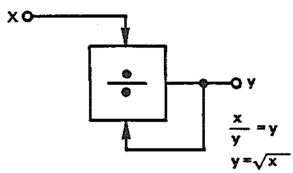


Figure 6b. Square root as an implicit function. If the divider uses a multiplier fed back with high gain, the configuration is identical to 6a. But a device specifically designed for division will retain low error over a much wider dynamic range.

2. Root mean-square (see page 17) A multiplier-divider (uv/w) may be used, followed by an averaging filter, to compute the average:

$$y = \text{ave. } (x^2/y) \tag{17}$$

For stationary waveforms, and using a filter having a sufficiently long time constant, y will be constant, and

$$y = \sqrt{\text{ave. } (x^2)} \tag{18}$$

3. Vector sum and difference (see page 21) If $w = \sqrt{u^2 + v^2}$, one can compute w as follows, using a multiplier-divider

$$w^2 - u^2 = v^2 = (w + u)\,(w - u) \tag{19}$$

Dividing by $(w + u)$

$$\frac{v^2}{w + u} = w - u \tag{20}$$

and

$$w = u + \frac{v^2}{w + u} \tag{21}$$

Additional variables may be embraced simply by adding terms; for example, to compute $w = \sqrt{u^2 + v^2 + x^2 + y^2}$,

$$w = u + \frac{v^2}{w + u} + \frac{x^2}{w + u} + \frac{y^2}{w + u} \tag{22}$$

Given w and u, the vector *difference* $v = \sqrt{w^2 - u^2}$ may be computed by dividing equation 19 by v, whence (Figure 7a)

$$v = \frac{(w + u)\,(w - u)}{v} \tag{23}$$

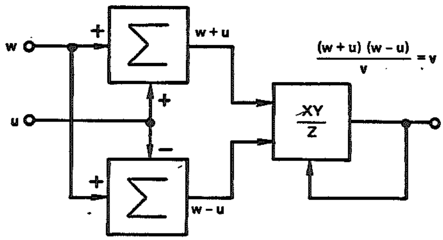
4. Bridge linearization The output of a Wheatstone bridge configuration with one leg variable is of the form

$$y = \frac{x}{1 + x} \tag{24}$$

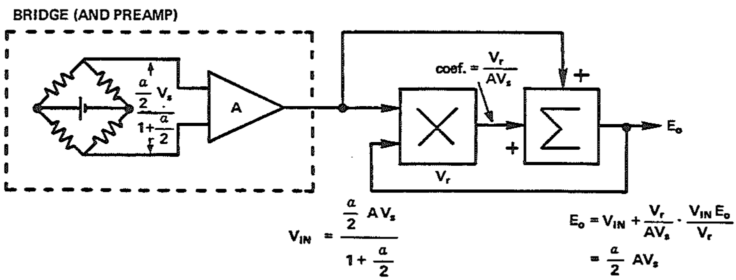
This response is linear only for small values of the deviation, x. It can be linearized by solving implicitly for x (Figure 7b)

$$x = y + x \cdot y \tag{25}$$

Since the deviations are usually small, a multiplier with very modest specifications may be used, provided that the signals are scaled to use near-full-scale capability, and that drift of the second (i.e., correction) term is low.



a. Vector difference $v = \sqrt{w^2 - u^2}$



b. Linearizing a Wheatstone-bridge output

Figure 7. Applications of implicit feedback

There are three principal reasons for using implicit solutions, when they are appropriate:

a. To simplify the block diagram. In the case of vector summation, a $u \cdot v/w$ multiplier and two operational amplifiers replace two squarers, a square-rooter, and a summing amplifier.

b. To avoid expansion of dynamic range. Squaring a signal with 100:1 dynamic range results in a signal with 10,000:1 dynamic range. Noise, drifts, and reduced bandwidth can impair overall accuracy. The $u \cdot v/w$ operation, on the other hand, remains net first order.

c. To provide an improved fit with few additional components.

Figure 14 and the appendix to this chapter show the great improvement in fitting sin x due to using modified equations involving an additional feedback term.

Like inverse functions, implicit functions must be single-valued. Unlike inverse functions, they need not always be monotonic. For example, sin x can be approximated from $-\pi$ to $+\pi$ with greatly improved accuracy, using feedback.

FITTING ARBITRARY FUNCTIONS

For the purpose of this section, "arbitrary functions" are defined to include all functions that cannot be fit "exactly" by a conceptual closed-form equation, whether explicit or implicit. In other words, there is almost always a residual theoretical error, which must be considered along with the device errors to determine the overall closeness of fit.

Besides such obviously arbitrary functions as empirically-determined circuit and system nonlinearities requiring calibration curves, "arbitrary functions," by the above definition, include such analytic but non-rational functions as sin θ, $\tan^{-1} x$, and a whole host of functions characterizable by infinite series.

In general, to be fittable by more-or-less simple analog circuits, a functional operation must be bounded (i.e., defined within a finite range of all variables involved), single-valued in terms of inputs, and free from singularities (except where they can be satisfactorily fit by diode breakpoints, switching, or comparator "jump" functions). To be practical for analog circuit elements, there is the additional constraint that circuit complexity (consequently the cost) must be competitive with digital function generation (ROM's alone, or ROM's plus digital processing, plus at least one step of conversion).

Relationships are smoothly fit using logarithmic, exponential, or power-law elements, or they may be fit more-or-less directly with a set of straight-line segments produced by diode breakpoints (Figure 8). The former technique requires more-sophisticated mathematics and error analysis, but the output is differentiable, and the error function is satisfyingly smooth. The latter technique is well-suited to quick, empirical fitting of functions of one variable, but the error consists of a series of cusps that can be troublesome if differentiation is used or if the function is employed within a feedback loop. Also, arbitrary functions of 2 or more variables (which are difficult to fit, in any event) run into structural limitations, due to the sheer number of "piecewise-planar" elements; the influence of each faceted element also poses tricky visualization problems. Smooth functions, in linear, or nonlinear combinations —on the other hand— pose no interpolation problems. Any point is readily calculable, though not necessarily an accurate fit.

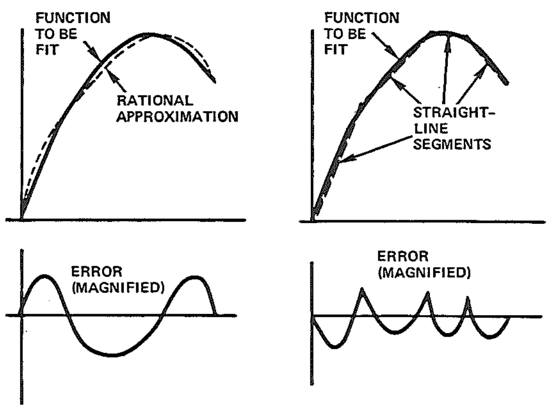


Figure 8. Smooth fit vs. piecewise-linear fit

SMOOTH APPROXIMATIONS

Although these approaches require more mathematics than do piecewise-linear approximations, the manipulations are of a kind that is not difficult if one is armed with a mechanized calculating device, such as an HP-35 pocket calculator. For mechanized optimization of the fit, a programmable engineering calculator, or access to computers, is helpful (but not essential unless one has the job of fitting many similar-but-different functions, or it is necessary to solve a large number of simultaneous equations to obtain a high-accuracy, high-order fit).

The ready availability of large amounts of calculating power, combined with the ready availability (at low cost) of today's multipliers, dividers, power-and-root devices, and logarithmic elements, makes smooth analog approximations (with errors typically varying from 0.1% to 1%) far more practical now than they have ever been.

The approach to fitting a function $y = f(x, A, B, C, \ldots)$, where x and y are variables, and A, B, C,... are constants, to the desired prototype shape, involves the following steps.

1. It is helpful to start with the data in normalized form.

2. Postulate a function that is likely to have the "right shape."

3. Write the equation for as many specific points as there are constants to be solved-for in the approximating function. The fit will be exact at those points.

4. Solve the set of simultaneous equations for the constants. Plug them into the equation, and check, by substituting the specific values chosen for "exact fit" into the equation.

5. Try out the equation at other intermediate values of x. Solve each for y and subtract the expected value of y to obtain the error. It may be helpful to plot an error curve.

6. If the errors are of reasonable magnitude, but are greater between one pair of calculated points than another, new intermediate points may be selected, and the equations written, solved, and tried-out for the new points. This process may be repeated as often as necessary to give (for example) equal maximum errors in all ranges. (Interpolation formulas may be used to shorten the process.)

If the errors are obviously too large, a different function may be tried. The reader will recognize that both experience and creativity will be of great help in proposing a function that has small inherent errors for a given shape. Here are some suggestions that those unfamiliar with the process may try as a starter:

- Try to find a "natural law" (e.g., logarithmic response)
- Try to fit deviations from linearity or from simple functional relationships, having a somewhat similar shape to the curve in question, such as $\log x$, $1/x$, ϵ^x, x^m, etc.

- Try truncated power series ($A + Bx + Cx^2 + Dx^3$, etc.)
- Try series involving non-integral exponents, e.g., $A + Bx + Cx^m$
- Try implicit functions, e.g., $y = (A - y)\,x^m = Ax^m/(1 + Ax^m)$. While the two expressions are identical, the first uses fewer elements.
- Try the "Hoerl equation" $y = Ax^B \epsilon^{Cx} = A\ln^{-1}(B \ln x + Cx)$
- Try a more-easily-fit complementary function, such as $\cos x = \sin(\pi/2 - x)$

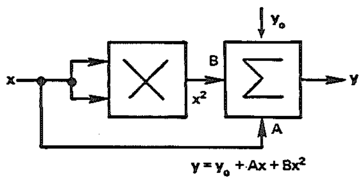


Figure 9. 2nd-degree polynomial using single multiplier. Op-amp configuration depends on polarity of constants in this figure and those that follow.

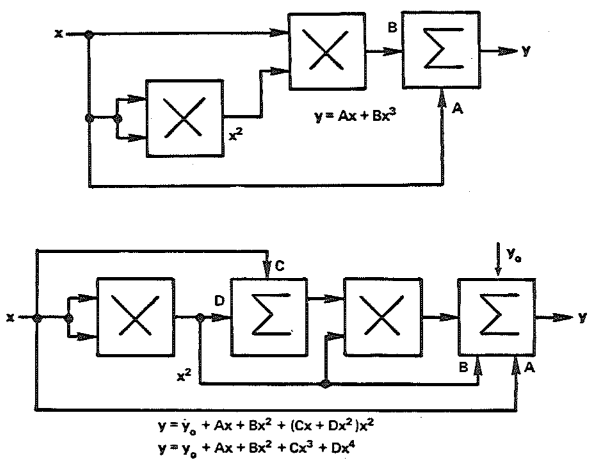


Figure 10. Odd-function 3rd-degree polynomial and generalized 4th-degree polynomial, using 2 multipliers. For complete generality, the origin may be offset along the X-axis by an amount h by adding a bias to the input. The x's then become x' = x - h.

POLYNOMIALS AND POWER SERIES

Polynomials can be modeled with multipliers and operational amplifiers. The minimum number of multipliers required to fit truncated power series of various degrees are:

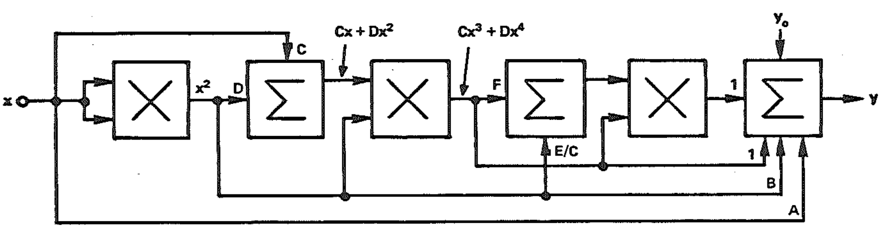
2nd degree (involves x^2) . . . 1 (Figure 9)

4th degree (involves x^4 and lesser powers) . . . . 2 (Figure 10)

8th degree (involves x^8 and lesser powers) . . . 3 (Figure 11)

However, if implicit feedback is used, any of these truncated series may be converted into an *infinite* series, convergent over a limited (but adequate) range (Figures 12 and 13). The resulting enrichment can greatly improve the theoretical fit.

For example, a cubic ($y = Ax + Cx^3$) can fit sin x to within ±0.6% of full scale, from $\pi/2$ to $-\pi/2$, or within ±13.2%, from π to $-\pi$. But with the simple addition of a feedback term ($y = Ax + Cx^3 + Ex^2y$), the theoretical error becomes less than ±0.01% ($\pi/2 > x > -\pi/2$); and over the wider range of angle (π to $-\pi$), the error is still less than ±1.2%. The following example shows how dimensionless coefficients are derived, and the appendix to this chapter provides comparative details of a variety of sine-function-fitting schemes.

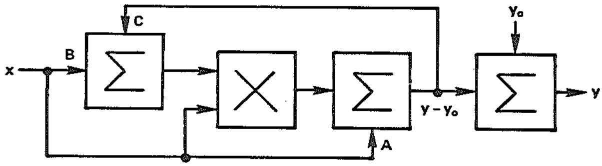


$$y = y_o + Ax + Bx^2 + Cx^3 + Dx^4 + (\frac{E}{C}x^2 + F\,Cx^3 + F\,Dx^4)\,(Cx^3 + Dx^4)$$

$$= y_o + Ax + Bx^2 + Cx^3 + Dx^4 + Ex^5 + (F\,C^2 + \frac{E}{C}D)x^6 + 2F\,C\,Dx^7 + F\,D^2\,x^8$$

$$= y_o + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 + a_6x^6 + a_7x^7 + a_8x^8$$

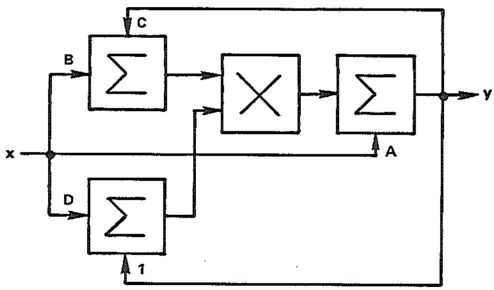
Figure 11. Generalized 8th-degree polynomial, using 3 multipliers. This configuration obtains its relative simplicity at the cost of 2 degrees of freedom (a_7 and a_8 are functions of a_3, a_4, a_5, a_6, and a_6 is not independent of a_3, a_4, a_5).



$$y - y_o = Ax + x\,(Bx + C\,[y - y_o]\,) = \frac{Ax + Bx^2}{1 - Cx}$$

$$y = y_o + Ax + (B + AC)x^2 + C(B + AC)x^3 + C^2\,(B + AC)\,x^4 + \ldots$$

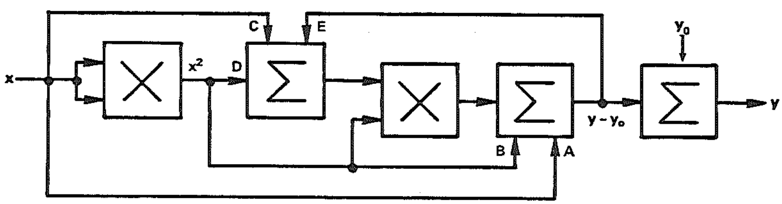
a. Second-degree polynomial with implicit feedback produces infinite series, convergent for Cx <1, has three degrees of freedom.



$$y = Ax + (Bx + Cy)\,(Dx + y) = \frac{Ax + BDx^2}{1 - (B + CD)x - Cy}$$

b. Second-degree polynomial in both x and y has four degrees of freedom, but coefficients are derived with greatly-increased difficulty. If y(0) ≠ 0, y_o is added outside the loop, as in 12a.

Figure 12. Implicit approximations with a single multiplier.



$$y - y_o = Ax + Bx^2 + x^2\,(Cx + Dx^2 + E\,[y - y_o]\,)$$

$$= \frac{Ax + Bx^2 + Cx^3 + Dx^4}{1 - Ex^2}$$

$$y = y_o + Ax + Bx^2 + (C + AE)x^3 + (D + BE)x^4 + E(C + AE)x^5 + E(D + BE)x^6 + \ldots$$

Figure 13. Fourth-degree polynomial using 2 multipliers with implicit feedback produces infinite series, convergent for Ex^2 <1, has up to five degrees of freedom. For odd function, B = D = 0; for even function, A = C = 0.

Earlier in this chapter, we have mentioned the practical limitations to the degree of fit; device cost and performance, and circuit complexity. To these must be added the difficulty (even in simple, low-cost configurations) of coping with the many degrees of freedom as the number of coefficients is increased. Three coefficients is a reasonable maximum for an engineer with a hand calculator, unless he is of a mathematical bent and enjoys solving this kind of problem. If mechanized stored-program calculators and computers are at hand, the device cost and performance limitations become more significant.

AN EXAMPLE: $y = f(x) \simeq \sin x \quad \left(0 \leqq x \leqq \frac{\pi}{2}\right)$

The appendix to this chapter, as mentioned, shows a number of equations and configurations that provide theoretical fits (of $\sin x$) to varying degrees of accuracy and suitability. We will derive here, as an example of the function-fitting process, the simplest of the approximations, a quadratic polynomial, using a single multiplier,

$$y = \mathrm{A}x + \mathrm{B}x^2 \qquad [y(0) = 0] \tag{26}$$

and compare it with a more-accurate version, still using a single multiplier, but adding an implicit feedback

$$y = \frac{\mathrm{A}x + \mathrm{B}x^2}{1 - \mathrm{C}x} = \mathrm{A}x + x(\mathrm{B}x + \mathrm{C}y) \tag{27}$$

To obtain a trial set of coefficients in (26), substitute y and x (in radians) at two points. Let us use the end point, $x = \pi/2$, and an experimental intermediate point, $x = 1$ rad $= 57.296°$:

$$\sin \pi/2 = 1 \quad = \mathrm{A}\,(\pi/2) + \mathrm{B}\,(\pi/2)^2 \tag{28}$$

$$\sin 1 = 0.8415 = \mathrm{A} + \mathrm{B} \tag{29}$$

Solving simultaneously for A and B, we find that B = −0.3589 and A = 1.2004; hence, $y = 1.2004x - 0.3589x^2$.

In testing this approximation over the range of angles 0 to $\pi/2$,* maximum error (y - sin x) appears at 21.6° (error < 3.4% F.S.) and at 74.5° (error = - 0.96% F.S.) The error is zero at 0, 1 radian, and $\pi/2$ radians.

By choosing a different value of angle for intermediate zero error in (29) and solving for new coefficients, testing them, repeating, etc., it is possible to arrive at a "best" fit, with symmetrical maximum errors of about ±2.1%. This approximation is

$$y = 1.155x - 0.33x^2 \tag{30}$$

Intermediate zero-error occurs at about 42.2°. The maximum errors occur at 17.4° and 68.6°. Error plots appear in Figure 14.

The block diagram of a configuration that would produce this approximation is shown in Figure 9. By adding an implicit feed-

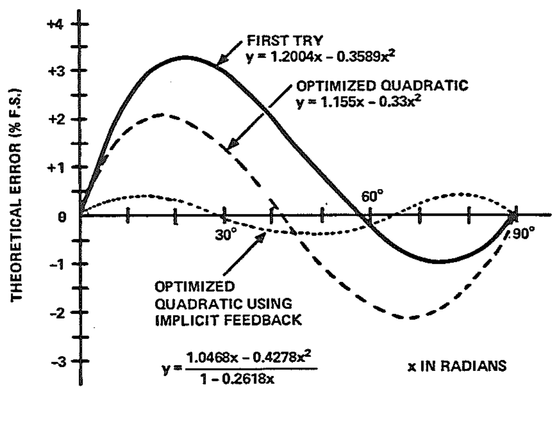


Figure 14. Errors of quadratic approximation to y = sin x ($0 < x < \frac{\pi}{2}$). Percent error = 100 (f(x) - sin x).

*Maximum error can be determined by an error plot, or by differentiating the error equation ($f(x)$ − sin x) and solving for the values of x at which the derivative is zero.

back, the configuration of Figure 12a, characterized by equation 27, is achieved, resulting in a reduction of the errors to less than ±0.5% (Figure 14).

The coefficients of equation (27) can be calculated by a similar process to that outlined above. However, the additional degree of freedom provided by C requires that two intermediate zero-error angles be determined, and we can expect *three* error maxima. The solution of three simultaneous equations makes the process somewhat more arduous, but still manageable, and the results are rewarding, since a fourfold reduction of error is obtained, at the cost of one additional op amp. The new equation, with optimized coefficients, is

$$y = \frac{1.0468x - 0.4278x^2}{1 - 0.2618x} \quad 0 < x < \pi/2$$

$$= 1.0468x - x(0.4278x + 0.2618y) \qquad (31)$$

Maximum theoretical errors occur at 11.5° (0.42% F.S.), 47.1° (−0.44% F.S.), and 80.4° (0.44% F.S.), with zero error at 0°, 28°, 65.5°, and 90°.

It should be noted here that some reduction of maximum error could be obtained by allowing non-zero error at the end points.

In theory, this is tenable, though it necessitates an additional constant (and hence an additional simultaneous equation), but in practice it can be disastrous, since it makes calibration more difficult, and magnifies sensitivity to variations in device tolerances.

It should also be noted that errors discussed here are expressed in percentage of full-scale, rather than percent of the ideal value of $\sin x$. To convert the plotted errors to the latter form, they should be divided by $\sin x$. The ratio errors will be found to be larger, the error maxima will occur at different values of angle, and they will no longer be equal in magnitude. To test the approximations with the aim of minimizing ratio-to-ideal-value errors, the error function is $f(x)/\sin x - 1$.

A much better theoretical fit can be obtained, using two multipliers, with the additional bonus that it can be made to work in two quadrants (i.e., $-\pi/2 < x < \pi/2$). The simple cubic (Figure 10) gives ±0.6% maximum error, but with implicit feedback, the theoretical error can be reduced to less than ±0.01%, which probably represents greater accuracy than might be expected from any of the devices currently available to implement the approximation at reasonable cost. That is, the error is limited by the devices, rather than the approximation. For single-quadrant fitting, a single $U{\cdot}V^{m}$ device (such as Model 433), with m set at 2.0, can replace the two multipliers.

Finally, using a $u{\cdot}v^{m}$ device in the same configuration, but with m set at a non-integral value, the maximum theoretical error can be reduced to ±0.15% F.S. open-loop, and ±0.004% with implicit feedback. The above approximations are all included in the appendix to this chapter.

PIECEWISE-LINEAR FUNCTION FITTING
(A Brief Introduction)

As Figure 15 shows, a nonlinear relationship is fit by summing gain segments (S_1, ΔS_2, ΔS_3, etc.) that have zero contribution until a threshold is crossed. Beyond the threshold, the output of a given segment contributes linearly. The nature of the ideal contribution (and the errors, too) is determined by both the location of the thresholds and the incremental gains attributed to the segments, as well as the means of implementation.

The simplest "diode function generators," or DFG's, as such devices are commonly termed, use segments that provide either zero or linear response (from the threshold to full-range input), as shown. Since the contributions accumulate, sharp reversals require a large amount of gain to overcome the accumulated gain of earlier segments. Circuits have been built using truncated segments to avoid the accumulation of gain, but they tend to lead to an unwieldy amount of circuitry; in addition, they require careful matching of break points to avoid "glitches" where one segment leaves off and another starts.

The conceptually-simplest segment is obtained with a biased diode and a precision resistor, but its temperature sensitivity leaves much to be desired. Chapter 3-5 discusses several more-practical approaches to individual segments. "Ideal-diode" op-amp circuits are an obvious possibility, because of their stable thresholds, sharp corners, and precise gains, as well as the low cost of operational amplifiers.

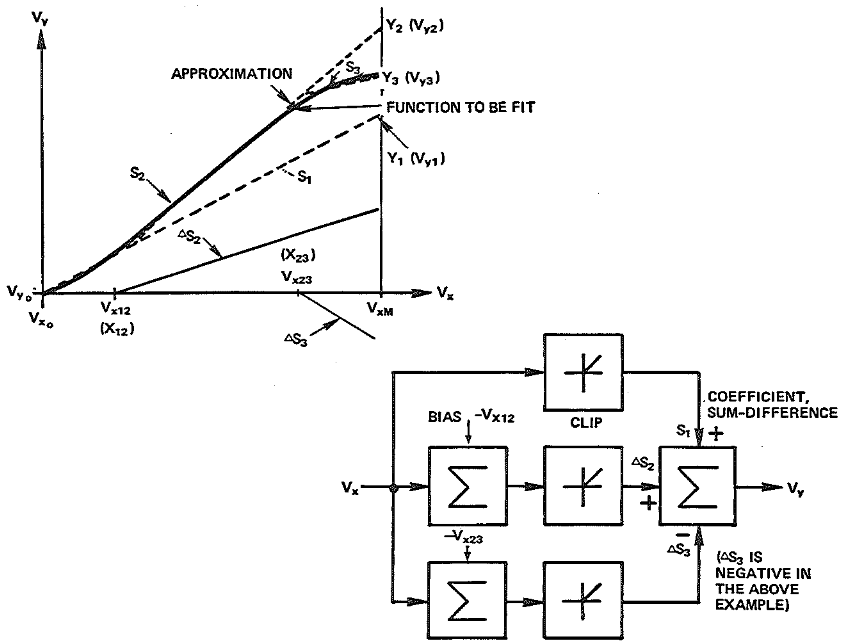


Figure 15. Basic 3-segment piecewise-linear function fitter.

Positive or negative contributions from individual segments are obtained by the use of a subtractive output circuit, usually consisting of an inverting output amplifier and an intermediate current inverter. For special-purpose function fitting (which comprises the great majority of applications), gains and thresholds may be computed, and fixed resistance values —with minor "tweaks"— are used. For general-purpose function fitting, potentiometers typically are used for setting each threshold (bias) and gain. To obtain the gamut of positive-to-negative gains, potentiometers straddle the positive and negative summing buses (Figure 16).

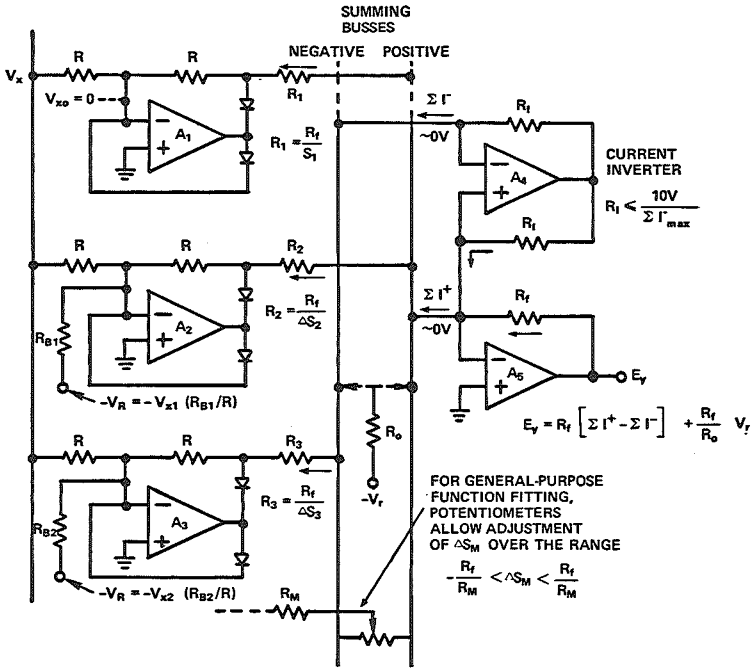


Figure 16. Piecewise-linear function fitter. For adjustable thresholds, typically the R_B's = R and $V_R = V_{xj}$, adjusted by individual potentiometers connected to a stable reference source.

Fitting functions using the piecewise-linear approach is, to begin with, a paper exercise. It can be done graphically or numerically. The graphic approach simply involves a large normalized plot of the function to be simulated. Draw the minimum number of connected straight lines to fit the curve to the required accuracy. (Figure 15). This fixes the break points, X_{12}, X_{23}, X_{34},... The straight lines are all extended to the ordinate corresponding to the maximum value of x, X_m. This permits the incremental gains to be computed accurately, even for short segments. The incremental gains, S_1, ΔS_2, ΔS_3, for the example shown are

$$\frac{Y_1}{X_m}, \frac{Y_2 - Y_1}{X_m - X_{12}}, \frac{Y_3 - Y_2}{X_m - X_{23}} \qquad (32)$$

When the circuit has been assembled, final setting of the coefficients of the function*

$$V_y = V_{Y0} + S_1(V_x - V_{x0}) + \Delta S_2(V_x - V_{x_{12}}) + \Delta S_3(V_x - V_{x_{23}}) \quad (33)$$

can be simply done as follows:

1. Set the thresholds, $V_{x_{12}}$, $V_{x_{23}}$, etc.

2. With $V_x = V_{x0}$, and all gains at zero, set the output bias $V_y = V_{y0}$.

3. With $V_x = V_{xm}$, adjust S_1 for $V_y = V_{y_1}$, ΔS_2 for $V_y = V_{y_2}$, etc., in that order, keeping all gains at zero until the previous gains have been set. Use overall output attenuation (temporarily-reduced R_f), if necessary to keep V_y within reasonable limits.

Because all the adjustments are made with $V_x = V_{xm}$, there is a tendency for cumulative gain errors to be reduced. The function can now be checked at the intermediate points. If the breakpoints are not sharp, this factor should be taken into account on the paper plot before establishing the values of V_{y_1}, V_{y_2}, etc. The fit can be refined, if necessary, by minor adjustments to the thresholds, and repeating step 3.

A WORD ABOUT SUMMING-AMPLIFIER CONFIGURATIONS

Sum-and-difference amplifiers are well-known, having been discussed in just about every textbook and tutorial article on the basic applications of differential op amps.

In function-fitting applications, there is usually an amplifier that bears the brunt of summing a number of arbitrary inputs with a variety of gains of either polarity. The choice is usually between a differential amplifier and two inverting op amps in a subtracting configuration.

FOR SMOOTH APPROXIMATIONS, the inputs are usually taken from either op amps or nonlinear modules (or IC's), which have low-impedance operational-amplifier outputs. For these applica-

*$\Delta S_j = 0$ for $V_x - V_{xij} = 0$

tions, either the differential subtractor or the inverting subtractor of Figure 16 may be used. Because the inverting subtractor operates at ground level, it is more suitable for applications where gains must be adjustable, and the impedance changes associated with a specific gain adjustment must not disturb the other gains. However, if the gains are fixed, the differential subtractor is somewhat less costly; resistance ratios are easy to compute if the basic rule associated with Figure 17b is observed. See also *Electronics*, June 12, 1975, pp 125–126.

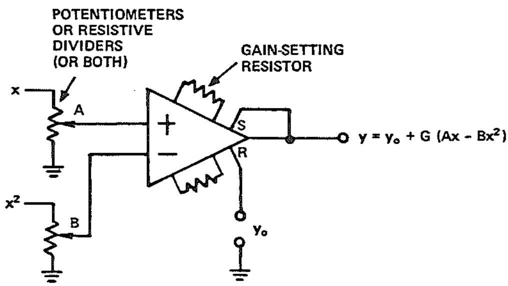


Figure 17a. Use of fixed-gain differential amplifier for 2-variable system with coefficients of opposing polarities (see Figure 9).

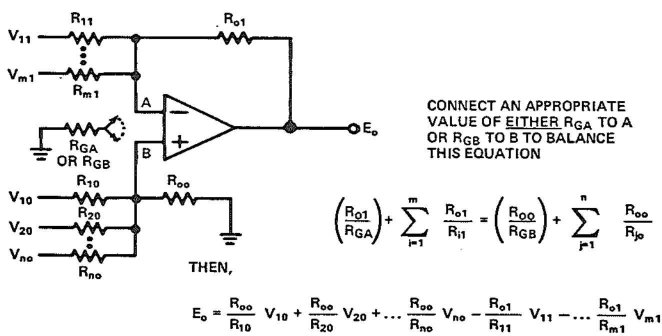


Figure 17b. Use of differential op amp for summing and differencing an arbitrary number of inputs with arbitrary fixed gains.

FOR PIECEWISE-LINEAR APPROXIMATIONS, the source impedance of the additive terms is usually nonlinear, being low in the conducting state and high in the open state. Thus, the isolation afforded by the summing-point of an inverting amplifier is not only desirable, but necessary, to avoid interaction.

A close look at Figure 16 will disclose the interesting fact that the positive input of A4, instead of being grounded, is connected to the summing-point of A5. In this connection, A4 serves as a *current inverter or reflector*, rather than as a voltage inverter. The purpose becomes clear if one considers that the summing point of A5 is loaded by the high output impedance of a current source rather than the usually-low resistance R_I, thus minimizing the closed-loop gain of A5, increasing bandwidth, and reducing the amplification of drift and noise.

PRACTICAL MATTERS

We have dealt with an "ideal building-block" approach to function-fitting, while appearing to ignore the practical characteristics of the building blocks that are to be used. The purpose was to avoid interjecting issues that, while highly appropriate, would tend to serve as digressions and dilute the main course of the argument. Also, each case must be analyzed in terms of the specific functional operations, their configuration, and the allowable ranges of input and output. Since the variety of permutations and combinations is broad, it is virtually impossible even to begin to cover them all in the detail they deserve in the available space. Nevertheless, this chapter would be incomplete if it didn't provide some guidance toward practical implementation of the ideas.

Practical considerations include scaling, component choice, errors (and their sensitivity to parameter variation and drift), response speed, and (for feedback configurations) stability.

If the reader is mathematically gifted, he will have little difficulty determining the scaling, the sensitivity to parameter tolerance (within the limits specified for the real devices and passive elements), or computing the approximate speed of response. While stability may be investigated theoretically, it is perhaps better to explore it experimentally.

More typically, the reader will have sufficient mathematical facility to compute the constants and perform the scaling (an example is given in the appendix), but may have difficulty with the mathematical formulations involved in error and stability analysis. In that case, as regards error analysis, he should perform a series of "brute force" calculations involving changes in the constants to find out which are the most tolerant and those that are the most sensitive. One approach is to make (say) a 0.1% change in a given constant, and determine the effect (magnitude and direction) on the maximum error. Another is to make (say) a 0.1% change in an input variable, and determine its effect on the output error.

It may be useful, in this day of calculators, to perform the computations of theoretical constants to many significant digits, then to round off, one digit at a time, until a significant effect on the error is seen. The theoretical examples given in this chapter and its appendix have all been worked out to an excessive number of places for the accuracy involved.

In any case, the reader should study the chapters in Part 3 that pertain to the *devices* to be used, and those in Part 4 that pertain to their application in the specific *operations* to be used (e.g., multiplication, division, logs, etc.). Naturally, familiarity with the data sheets for the devices actually to-be-chosen is essential, to be sure that they are physically and electrically compatible with the rest of the system and that there are no unpleasant surprises in the list of specifications.

Performance should always be checked on a "breadboard" that includes a facility for investigating response, stability, and the effects of parameter variations in those portions of the circuit that analysis (or intuition) suggests are most sensitive.

The resistors can be chosen at the next-lower (for example) standard values, with appropriate tolerances, temperature-sensitivity, and cost; the effects of parameter variations can be studied experimentally by "tweaking" incremental resistances connected in series.

Dynamic responses and stability are best studied experimentally, using large and small sine and square waves (and perhaps noise) biased at various levels. Response can often be improved, especially where subtraction is involved, by seeking to match the approximate responses of branches being summed by delaying the faster using an R-C lag circuit.

X-Y plots on the oscilloscope screen (input horizontal, output vertical), using sine- or triangular waves, can be quite helpful in observing the shape of the curve(s), determining that there have been no gross errors of fit, finding amplitude-sensitive instabilities, and (by frequency adjustment) determining simultaneously both amplitude and "phase" response. Not only the output behavior can be observed; one can also observe behavior at intermediate stages.

If the function involves a deviation from linearity, errors can be more-sensitively explored by subtracting the output from a signal proportional to the input, and observing just the deviation, plotted against the input.

Errors can also be determined point-by-point, using voltage sources and precise digital voltmeters, or by comparing X-Y plots on a chart recorder with hand-plotted curves. Where large numbers of identical functions are to be monitored, or trimmed, computer-test techniques can be brought into play in various ways, for example, by programming the input, and comparing the output with the stored "correct" values, either digitally (go-no), or with an analog readout established by computer graphics.

CONCLUSION

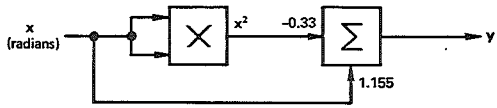
This chapter has sought to introduce the reader to the basic ideas and techniques relating to analog function fitting, and to encourage the increased application of low-cost nonlinear analog devices in calibration, compensation, and measurement. Some of these ideas will reappear, perhaps in amplified form, in subsequent chapters. The concentration in this chapter has been on the development of conceptual models. The following chapters will utilize some of these ideas in the context of their applications.

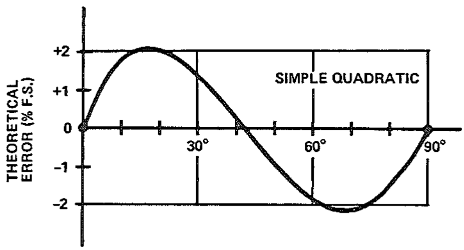
APPENDIX TO CHAPTER 2-1

Analog Approximations for sin x with Ideal Devices

1. Quadratic, one-quadrant, single-multiplier

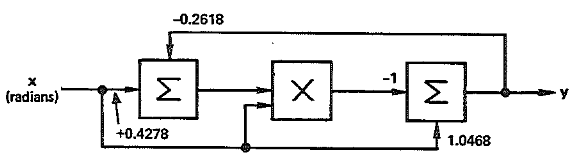
A. Explicit function: $y = 1.155\ x - 0.33\ x^2$

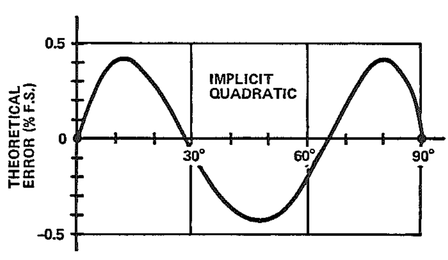




B. Implicit function: $y = 1.0468\ x - x\ (0.4278\ x - 0.2618\ y)$

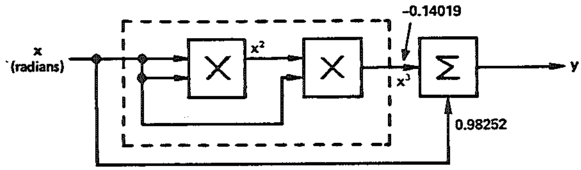
$$= \frac{1.0468\ x - 0.4278\ x^2}{1 - 0.2618\ x}$$

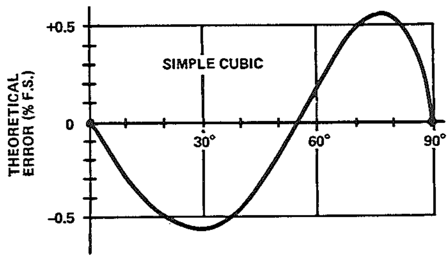




2. Cubic, two-quadrant, 2-multiplier, or one-quadrant UV^2

A. Explicit function: $y = 0.98252\,x - 0.14019\,x^3$

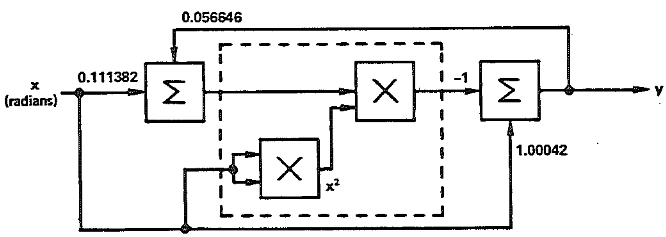


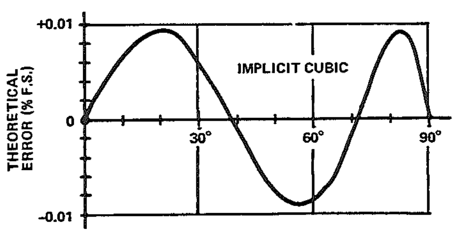


B. Implicit function:

$$y = 1.00042\,x - x^2\,(0.111382\,x + 0.056646\,y)$$

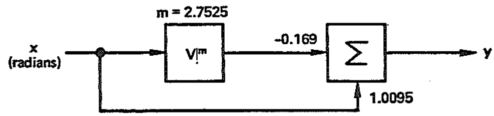
$$= \frac{1.00042\,x - 0.111382\,x^3}{1 + 0.056646\,x^2}$$

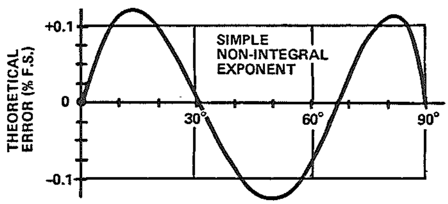




3. Non-integral exponent, one-quadrant, single UV^m

A. Explicit function: $y = 1.0095\ x - 0.169\ x^{2.7525}$

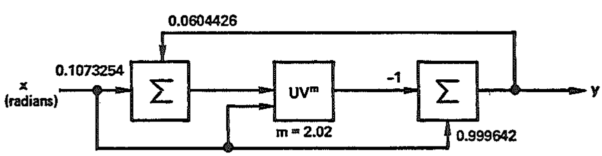


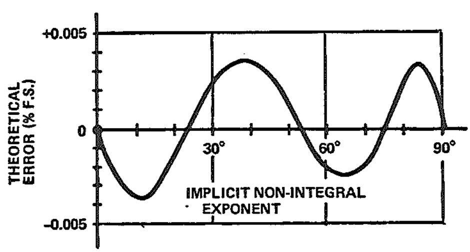


B. Implicit function:

$$y = 0.999642\ x - x^{2.02}\ (0.1073254\ x + 0.0604426\ y)$$

$$= \frac{0.999642\ x - 0.1073254\ x^{3.02}}{1 + 0.0604426\ x^{2.02}}$$



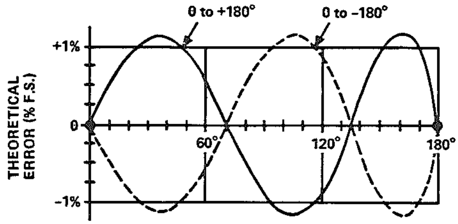


4. Extended angular range

A. Implicit cubic, 4-quadrant $-\pi \leqslant x \leqslant +\pi$, 2 multiplications

$$y = \frac{1.0287\,x - 0.10423\,x^3}{1 + 0.0904\,x^2}\text{, Circuit similar to 2(B)}$$

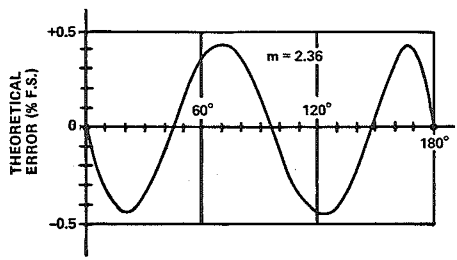
$$= 1.0287\,x - x^2\,(0.10423\,x + 0.0904\,y)$$



B. Implicit non-integral exponent, 2 quadrant, $0 \leqslant x \leqslant \pi$, single UV^m

$$y = \frac{0.9790\,x - 0.0657\,x^{3.36}}{1 + 0.0814\,x^{2.36}}\text{, Circuit similar to 3(B)}$$

$$= 0.9790\,x - x^{2.36}\,(0.0657\,x + 0.0814\,y)$$



SCALING EXAMPLE

To demonstrate the application of the scaling principles discussed in the chapter, electrical coefficients for example 2B will be derived, using the following assumptions:

1. 10V full-scale input corresponds to $\pi/2$ radians.
2. 10V full-scale output corresponds to $\sin \pi/2$.
3. Multiplier transfer functions are $V_1 V_2/10 = E_{out}$

$$y = \frac{A\,x - B\,x^3}{1 + C\,x^2} = A\,x - x^2\,(B\,x + C\,y) \cong \sin x$$

where

$$A = 1.00042$$
$$B = 0.111382$$
$$C = 0.056646$$

Since the maximum value of y is 1, y is already normalized. However, though x is dimensionless (radians), it is not normalized. To normalize x to its maximum value, $\pi/2$, multiply and divide by $\pi/2$, wherever x appears.

$$y = A\,\frac{\pi}{2}\,\frac{x}{\pi/2} - \left(\frac{\pi}{2}\right)^2 \left(\frac{x}{\pi/2}\right)^2 \left[B\,\frac{\pi}{2}\,\frac{x}{\pi/2} + C\,y\right]$$

If we let A′, B′, and C′ be the (unknown) coefficients of the electrical equation, the following equation describes the ideal performance of the electrical equivalent, taking into account the multiplier transfer functions.

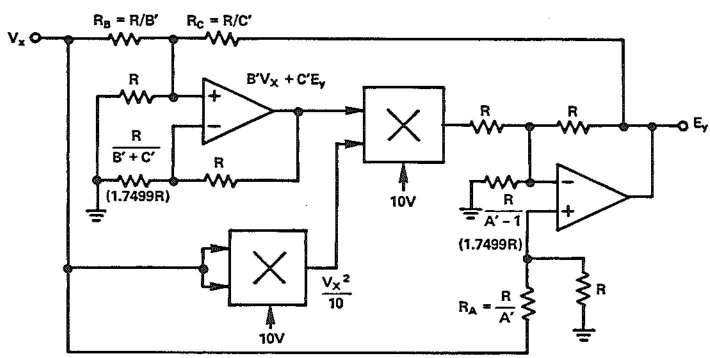
$$E_y = A'V_x - \frac{{V_x}^2}{10} \cdot \frac{B'V_x + C'E_y}{10}$$

Normalizing,
$$\frac{E_y}{10} = A'\,\frac{V_x}{10} - \left(\frac{V_x}{10}\right)^2 \left[B'\,\frac{V_x}{10} + C'\,\frac{E_y}{10}\right]$$

Because the normalized equations must be identical,

$$A' = A\,(\pi/2) = 1.571456$$
$$B' = B\,(\pi/2)^3 = 0.431693$$
$$C' = C\,(\pi/2)^2 = 0.139768$$

A circuit that embodies these coefficients, using ideal multipliers, op amps, and resistors, is:



$R_B = 2.31646R$, $R_C = 7.15471R$, $R_A = 0.636353R$

II Time—Function Generation

Chapter 2

Two products that revolutionized electronic instrumentation in the '30's and '40's were the oscilloscope and the sine-wave generator. The latter applied stimuli to systems or devices under test; the former permitted observation and time-domain measurement of the response. Since then, time-function generators as instruments have become greatly sophisticated; today, digitally-programmed sine- and square-wave, pulse, triangular, and even ROM-determined arbitrary function generators are available, in speeds from mHz to MHz.

As the uses of general-purpose function generators spread, the possibilities for low-cost, compact, in-house-designed *special-purpose* function generators for use in specialized equipment became apparent. The availability of operational amplifiers at low cost enabled some of these possibilities to become realities, and now the collateral availability of low-cost circuit elements with controlled, predictable nonlinearity should stimulate the greatly-increased use of function generators in OEM equipment.

A few examples of applications for function generation include establishing "profiles" (temperature, flow, velocity) in control systems, adjusting programmed parameters in test and instrumentation systems, and providing time bases of special form (e.g., logarithmic) in chart-recorder and oscilloscopic readout devices. Other applications for nonlinearity include variable-frequency polyphase oscillators, voltage-controlled filters, and low-cost signal generation with precise control of amplitude, frequency, and/or phase. These last include, of course, the classical sine-, square-, and triangular-wave generators, variable duty-cycle pulse generators, and one-shots, as well as the famous phase-locked loop. Also, one should not forget random-noise generators.

In this chapter, we discuss some of the principles of function generation and suggest ways of accomplishing a few basic functions with available standard building blocks. There exists, of course, a voluminous body of publications describing a gamut from simple circuits—involving transistors and passive elements—to the catalogues of manufacturers devoted to test instrumentation. Our aim is to neither replace nor surpass these efforts. Rather, it is to provide the designer with a modest indication of the range of rôles that controlled nonlinearity can play in function generation, with the thought that it will form a respectable complement to his bag of design tools, tricks, and ideas.

FUNCTION GENERATORS ARE MULTI-FACETED

It is possible to conceive of an extremely-wide range of function generators, classified in many different ways. What is common to them all is the use of nonlinearity: it is quite difficult to imagine a means of independently generating time functions, starting with a dc power source, without in some purposeful way involving nonlinear devices.

While this chapter deals with only a few specific examples of function generation, the following inclusive inventory of function-generator properties may be helpful to the reader who is seeking insight into remote (as well as better-known) aspects of this all-embracing field.

1. PERIODICITY: Aperiodic (single-shot), Stationary, Modulated, Random. Although the familiar connotation of "single-shot" is a pulse generator that delivers a single pulse in response to a stimulus, the term also should suggest such possibilities as a single half-sine, or a damped exponential train, or an arbitrary velocity or torque profile for a dynamometer test. *Stationary* waveforms are those having statistical properties that do not change with time. In practice, if a determinate waveform's amplitude, frequency, phase, or shape, or a random waveform's mean, variance, amplitude, distribution, and frequency spectrum are constant over a lengthy period

before, during, and after a measurement it is involved in, it may be considered stationary. *Modulated* refers to the variation of some property of the waveform during any observation interval (or from interval to interval) in response to a signal, for example, amplitude, phase, frequency, pulse-width, pulse position, presence or absence. It specifically includes voltage-to-frequency conversion, which can be viewed either as generation of a signal having a voltage-determined frequency, or modulating a signal about a fixed frequency.

2. SPEED: Very Low (fractions of 1Hz), Low, Audio, High, Video (> 1MHz). These distinctions are largely qualitative, but they are important insofar as they affect the choice of components or approach, the criticality of design, the limitations of accurate behavior, and the difficulty of use and measurement. The easiest portions of the spectrum to design for are in the middle, from about 1Hz to 30kHz. Suitable passive elements are small and cheap, and active devices have low drift and noise, as well as reasonable bandwidths. Thermal effects, that plague the low end, and stray capacitance and inductance, that complicate life at the high end, are rather manageable in the middle. Since this is not a complete text on the design of function generators, most of the specific circuits suggested have their best performance in the low-to-audio range of speeds and frequencies.

3. SHAPING: Simple vs. Complex. Simple functions are those that Nature allows to be achieved (in concept) with a minimum of basic hardware. They include sine-waves, as produced by resonant elements, square waves (produced by switching), triangular waves (often a by-product of square waves, or vice versa), exponential waves (also a by-product of square waves), and pulse trains. *Complex* functions involve operations on simpler functions, including modulation, filtering, and nonlinear function fitting (analog or digital) applied to simple waveforms. Analog function fitting involves ramps and function fitters; digital involves pulse trains, read-only memories (ROM's), and D/A converters with appropriately-filtered output; both can be combined to advantage (see Figure 11, this chapter). Most of today's commercial sine-square-triangle generators obtain the sine in a *complex* fashion: a triangular wave is applied to a fitted sine operator (Figure 1).

Random noise can be generated simply (for example, by amplifying resistor or junction noise) or in complex fashion (by generating a pseudo-random waveform having sufficiently-low autocorrelation, using a pulse train, tapped shift-register, exclusive-or'd feedback, D/A conversion, and filtering).

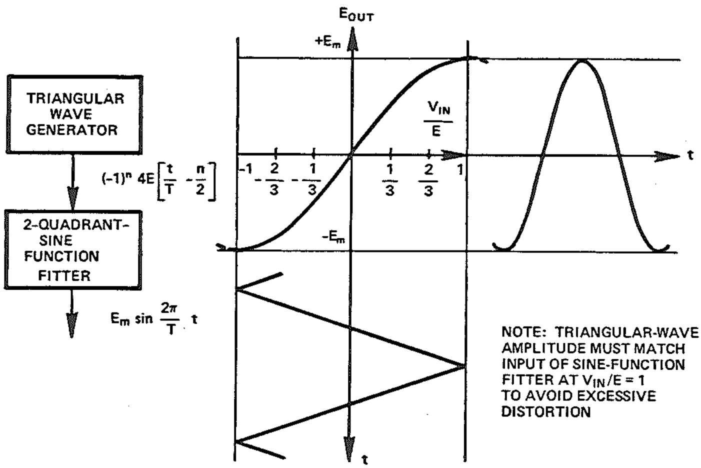


Figure 1. Generating a sine wave by function fitting

4. CONTROLLED PARAMETERS: Amplitude, Frequency, Phase, Mark/Space Ratio, Ranges, Shape, Measure. These parameters are affected by the manner of function generation; their fidelity to the desired behavior comprises the basic performance specifications of the function generator. Mark-space ratio takes on a broader meaning than just pulse on-off time: as a measure of symmetry, it also refers to a ratio between up-going and down-going intervals of ramps and sweeps. Departure from specified *shape* may be specified as "distortion." *Measure* indicates a form of average measurement that may be used instead of, or in addition to, *amplitude*, to characterize the waveform; for example, RMS or mean absolute-value. Crest factor is the ratio of peak amplitude to RMS. (With random noise, it is easier to measure RMS repeatably, than to observe peak amplitudes; the probabilities of various crest factors are a function of the noise distribution.)

5. FORM OF PARAMETER VARIATION: Fixed, Manually Adjustable (Continuously, or continuously, in ranges), Discretely, Digitally-Programmed, Auto-Ranged, Continuously Variable (modulated). This is entirely determined by the application, but it has profound effect on the design. Depending on this choice, for example, a parameter may be set by a resistor, by a pot (and switched fixed-resistors), by resistance-decade switches, by a D/A converter, or by an analog multiplier.

6. PARAMETRIC ACCURACY CLASS: 0.01%, 0.1%, 1%, Externally-Calibrated. This category is a catchall that includes such terms as "absolute accuracy," relative accuracy, precision, repeatability, stability. The above numbers represent orders of magnitude of *error* and each of these desirable characteristics is generally specified by a small number that represents the *deviation* from perfection. A function generator designed for a given application may have parameters that differ widely in error magnitudes; for example, frequency may be held to within parts-per-million, but amplitude variations and shape distortion may be of the order of 10%. The widespread availability of low-cost multipliers and D/A converters has made it possible for test-systems to be built that depend, not on costly fixed calibration of every generator used, but rather on a single programmed reference against which all generators are automatically calibrated and computer-adjusted to the desired settings before each measurement for which each is used.

7. INDEPENDENCE: Free-Running, Synchronized, Slaved. A free-running function generator depends for its accuracy and timing entirely upon its own internal reference sources, and to some extent (usually minimized) on the supply voltage. A synchronized device is allowed to free-run most of the time, but is from time-to-time brought "up-to-speed." A slaved device follows its speed reference, cycle by cycle.

8. FREQUENCY-DETERMINING ELEMENT: Resonant, Level-Controlled, External. Function generators that use internal crystal oscillators, Wien bridge, phase-shift or integrator-loop oscillators are *resonant*. Those that switch phase when a threshold has been

crossed are *level-controlled*. Though level-controlled types (multivibrators, one-shots, etc.) can be low in cost, their timing usually depends on an RC time constant, a reference supply, and a comparator; resonant types depend only on linear parameters, such as RC time constant. All types must of course take into account amplifier phase shifts and parasitic reactances. The amplitude-control arrangements for resonant types affect the damping, and may thereby marginally affect the frequency.

BASIC TRIANGULAR/SQUARE-WAVE GENERATOR

Figure 2 shows the configuration common to many varieties of level-controlled oscillators. It consists of a hysteretic comparator and an integrator. The output of the hysteresis element has two stable states, E_{o+} and E_{o-}; it switches to E_{o+} when the input exceeds V_{1+}, and it remains in that state until the input is less than V_{1-}, whereupon it switches to E_{o-}. It remains in that state until the input once again exceeds V_{1+}.

Suppose that the output has just switched to E_{o+}; it is applied to the integrator input. The integrator's output, starting from V_{1+}, decreases linearly with time at a rate E_{o+}/RC. At the end of the interval

$$\Delta t_1 = RC\frac{V_{1+} - V_{1-}}{E_{o+}} \qquad E_{o+} > 0 \qquad (1)$$

the output of the integrator is V_{1-}, and the output of the hysteretic comparator switches to E_{o-}. The integrator's output now *increases* linearly with time at the rate $-E_{o-}/RC$, until the output of the integrator is once again V_{1+}, which occurs at the end of the interval

$$\Delta t_2 = RC\frac{V_{1+} - V_{1-}}{-E_{o-}} \qquad -E_{o-} > 0 \qquad (2)$$

The period is

$$T = \Delta t_1 + \Delta t_2 = RC\left(\frac{V_{1+} - V_{1-}}{E_{o+}}\right)\left(1 - \frac{E_{o+}}{E_{o-}}\right) \qquad (3)$$

The frequency is

$$f=\frac{1}{T}=\frac{E_{o+}}{\left(1-\frac{E_{o+}}{E_{o-}}\right)\left(V_{1+}-V_{1-}\right)RC} \tag{4}$$

The mark-space ratio of the square-wave is

$$M/S = \Delta t_1/\Delta t_2 = -E_{o-}/E_{o+} \tag{5}$$

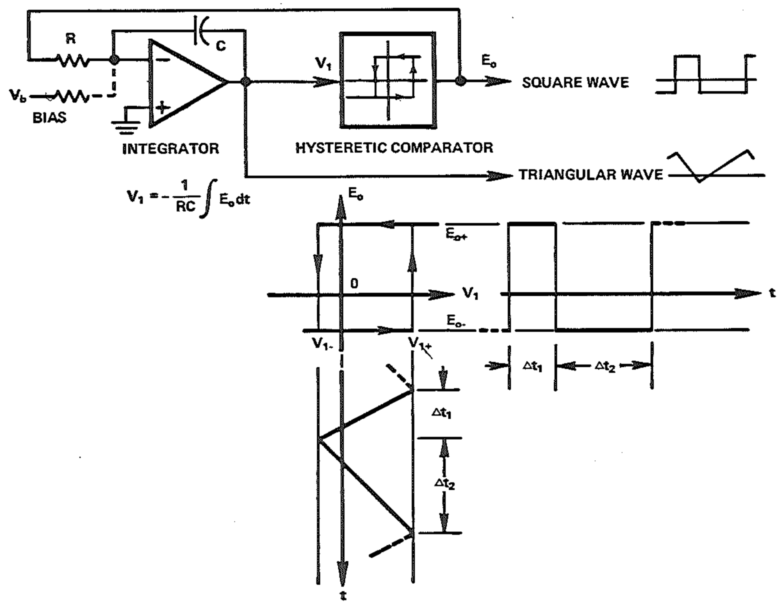


Figure 2. Basic triangular/square-wave generator

The peak-to-peak amplitudes of the triangular wave and the square wave are $(V_{1+} - V_{1-})$ and $(E_{o+} - E_{o-})$, respectively. For amplitude symmetry of the triangular wave, $V_{1+} = - V_{1-}$; For amplitude symmetry of the square wave, $E_{o+} = - E_{o-}$.

If a symmetrical square wave is desired with a mark/space ratio other than unity, a suitable bias V_b may be added to the integrator input. This bias is added to both E_{o+} and E_{o-} for computing the periods and mark/space ratios. However, the output levels of the hysteretic comparator are unaffected. For example, if $E_{o+} = -E_{o-} = +10V$, and a mark/space ratio of 2:1 is desired,

$$2 = \frac{-(E_{o-} + V_b)}{(E_{o+} + V_b)} = \frac{10 - V_b}{10 + V_b} \quad (6)$$

Solving, $V_b = -10/3$ volts, or $-E_{o+}/3$. In general,

$$\frac{V_b}{E_{o+}} = \frac{-E_{o-}/E_{o+} - M/S}{1 + M/S} \quad (7)$$

For symmetrical square waves,

$$\frac{V_b}{E_o} = \frac{1 - M/S}{1 + M/S} \quad (8)$$

Another commonly-used expression, related to mark/space ratio, is *duty cycle*, η,

$$\eta = \frac{M/S}{1 + M/S} \quad (9)$$

Equation (8), rewritten in terms of duty cycle, is

$$\frac{V_b}{E_o} = 1 - 2\eta, \text{ or } V_b = E_o - 2\eta E_o \quad (10)$$

If the bias voltage added at the integrator input is a constant, E_o, less a variable, $V_m = 2\eta E_o$, the duty-cycle will be a linear function of V_m (linear pulse-width modulation). Unfortunately, the frequency will not remain constant; it will be a function of V_m.

An additive bias at the integrator input may also be used to obtain time symmetry (M/S = 1) if the comparator has asymmetrical output levels. An additive bias is essential to meet the constraints of (1) and (2) if the comparator has unipolar output, e.g., if its outputs are in the TTL logic range (say, 5V & 0.5V).

If the output is symmetrical, the frequency can be linearly controlled by introducing a multiplication between the comparator output and the integrator input (Figure 3). For manual control, the "multiplier" can be a potentiometer; for voltage-control, it can be a multiplier; and for digital control, it can be a multiplying D/A converter. If a multiplier with a 10V scale constant is used, the frequency is

$$f = \frac{V_f E_o}{20\ RC} \cdot \frac{1}{V_{1+} - V_{1-}} \quad (11)$$

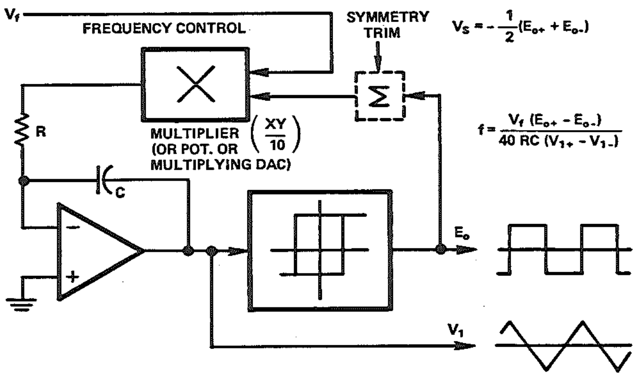


Figure 3. Controlling the oscillator frequency

OPERATIONAL AMPLIFIER AS HYSTERETIC COMPARATOR

Figure 4 shows a simple operational amplifier circuit, using positive feedback to develop hysteresis. Amplifiers that limit "hard", within a volt or so of the power supply, are especially useful for these circuits. For greater stability, a temperature-compensated zener diode regulator circuit could be used. This stabilizes, not only the amplitude of E_o, but also the frequency and mark/space ratio, and the triangular-wave amplitude, all of which depend on E_o.

To illustrate how it works, consider that the output has just switched to E_{o+}, as V_1 reached the threshold V_{1+}, V_1 then decreases linearly and will continue to do so until the voltage at the amplifier's positive input terminal goes negative. That occurs when V_1 reaches V_{1-}

$$-V_{1-}\left(\frac{R_2}{R_1+R_2}\right)=E_{o+}\left(\frac{R_1}{R_1+R_2}\right) \qquad (12)$$

The output switches to E_{o-}, the integrator's output starts back up, and continues to climb until the amplifier's input terminal goes positive (when V_1 reaches V_{1+})

$$V_{1+}\left(\frac{R_2}{R_1+R_2}\right)=-E_{o-}\left(\frac{R_1}{R_1+R_2}\right) \qquad (13)$$

Thus, the output switches at V_{1+} and V_{1-}, when

$$V_1 \geqq -\frac{R_1}{R_2}E_{o-} \text{ and when } V_1 \leqq -\frac{R_1}{R_2}E_{o+} \qquad (14)$$

The theoretical frequency of the oscillator of Figure 2, using the hysteretic comparator of Figure 4 is

$$f=\frac{R_2}{R_1}\cdot\frac{1}{RC}\cdot\frac{-\frac{E_{o+}}{E_{o-}}}{\left(1-\frac{E_{o+}}{E_{o-}}\right)^2} \qquad (15)$$

The triangular-wave amplitude is

$$\left(V_{1+}-V_{1-}\right)=\frac{R_1}{R_2}\left(E_{o+}-E_{o-}\right) \qquad (16)$$

The frequency may be controlled independently of the triangular of square-wave amplitudes, by adjusting RC, or by placing a gain adjustment in the feedback path to the integrator input. Symmetry of the triangular-wave amplitude may be controlled by introducing a bias current at the hysteresis summing point, via resistor R_0, connected to a voltage source of appropriate polarity. For fine trim, if the comparator output is nearly symmetrical, the adjustment may be connected between the supplies, with R_0 fairly large. If, on the other hand, a large offset must be dealt with (as when the comparator output swings between 0.5 and 5V), the adjustment may be a variable resistance in series with a fixed resistance.

For that case, if V_s is a negative voltage, at symmetry,

$$-\frac{V_s}{R_o} = \frac{1}{2}\left(\frac{E_{o+} + E_{o-}}{R_2}\right) \tag{17}$$

The biasing of the triangular wave doesn't affect its amplitude, frequency, or mark/space ratio.

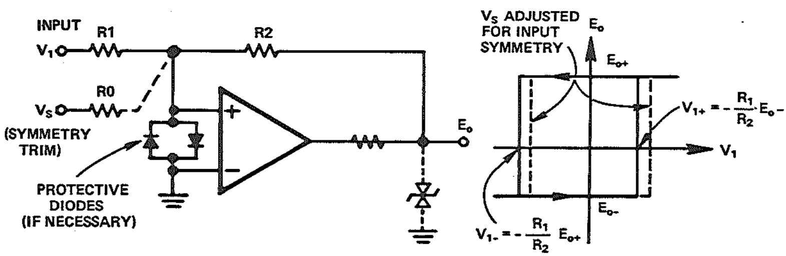


Figure 4. Operational amplifier as hysteretic comparator

A PRACTICAL OSCILLATOR CIRCUIT[1] (Figure 5)

This circuit, using low-cost components, provides square waves of about ±14V, with near-unity mark/space ratio, and triangular waves of about ±10V, with reasonable symmetry, at about 100Hz, for the values given. Frequency, triangular-wave amplitude, symmetry, and mark/space ratio may all be adjusted, by the means discussed above. Because A1 is a FET-input amplifier, frequencies as low as 0.1Hz and less are feasible, using large values for C and R

[1]"Triangular and square-wave generator has wide range," by R.S. Burwen, *EDN Magazine*, December 1, 1972.

(10MΩ for 0.1Hz) and/or an attenuator ahead of R. Bias current and offset voltage in A1 act in the same way as an external bias of $V_{os} + I_bR$, producing a slight modification of the mark/space ratio. Square-wave rise time is about 1.5μs, and fall time is about 0.5μs.

The frequency is affected by the saturation voltages of A2, and by the power-supply voltages. However, as equation (15) can show, sensitivity to symmetrical power-supply variations is quite small, and even individual variations as large as 20% cause no more than a couple-of-percent change. If stable passive components are used, a frequency stability of ±0.02%/°C is attainable. Capacitor C is preferably a polycarbonate type for stability, and also to ensure linearity of the triangular wave.

Although frequency stability is excellent, amplitude stability depends on the power supplies, the output-transistor saturation voltages, and the load (and their variations with temperature). For most applications, however, the outputs would be followed by adjustable-gain circuits. When amplitude stability is of critical importance, E_{o+} and E_{o-} should be determined by temperature-compensated zener diodes with fixed load, or —for variability— by a precision bound circuit (see Part 1).

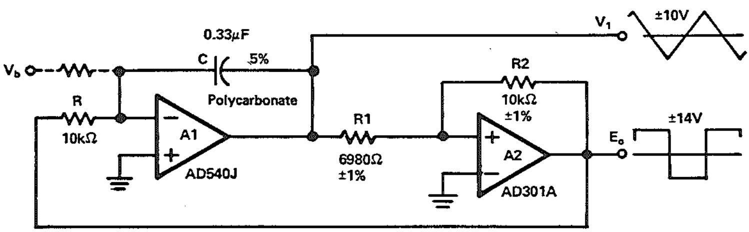


Figure 5. Practical oscillator circuit

Besides the inherent square-wave and triangular wave, and the variety of pulse widths, other functions, including sine waves, may be generated by feeding the output of the triangular-wave generator into one of the many varieties of function fitter described in Chapter 2-1. Trapezoidal waves may be generated by feeding the triangular-wave output into a set of bounds. Triangular pulses may be produced by feeding the triangular wave into a dead zone (Figure 6). Exponential responses are obtained with simple high- or low-pass filters.

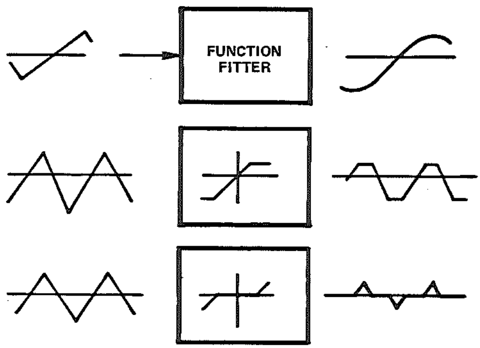


Figure 6. Shaped triangular waves: arbitrary, trapezoid, triangular pulses

ONE-SHOT (Figure 7)

The circuit of Figure 5 may be slightly modified to obtain a one-shot. A diode across the feedback capacitor prevents the output of A1 from ever becoming sufficiently negative to switch the output of A2 to $E_{o}-$. A capacitively-coupled positive-going logic pulse, applied to the negative amplifier input, starts the cycle by switching the output to $E_{o}-$. The integrator output ramps upward until it reaches V_{1+}, switching the output of A2 to E_{o+}. The integrator then ramps back through zero and stops at one negative diode-drop. The circuit is then ready for the next *start* pulse. The driving rate should be slow enough to allow each cycle to be completed, and the pulses should be narrow, compared to the cycle time.

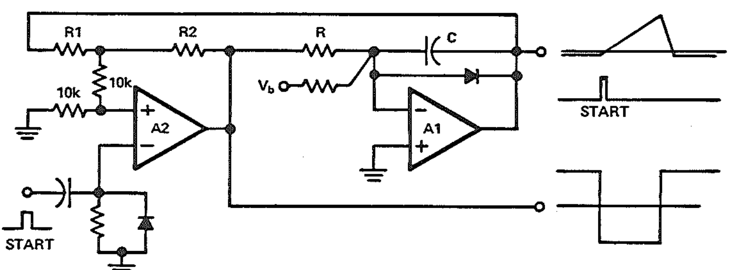


Figure 7. One-shot function generator

The ramp output may be used as an input to a functional operator to generate arbitrary functions of time that occur but once. The diode drop can be biased out in the input stages of the associated circuitry. If the descending ramp is not desired, the change-of-state of A2 can operate a switch to disconnect the output of A1. (The change back, with the *start* pulse, can reconnect it.)

SINE-WAVE OSCILLATOR

We have already noted that a triangular wave can be shaped to a low-distortion sinusoid, using function-fitting techniques. Some of the smooth-function techniques of Chapter 2-1 may yield considerably more-faithful sinusoids than the conventional piecewise-linear diode shaping networks. It has also been noted that the frequency (and symmetry) of such oscillators is dependent on voltage thresholds, as well as RC time constants.

For some purposes (for example, if the waveform is to be differentiable, with low distortion), an oscillator that relies solely on passive components for frequency control may be more desirable. The class of oscillator that uses RC networks includes the Wien bridge, phase-shift oscillators, twin-T oscillators, and *state-variable* oscillators.

This last type is an analog-computer equivalent to an L-C circuit. It consists of two integrators in a negative-feed-back loop, with damping appropriate to maintain amplitude control (Figure 8). It has some interesting features: first, since integrators have a fixed 90° phase shift, with unity gain at the frequency of oscillation, it is inherently a two-phase oscillator, producing both sin ωt and cos ωt; second, two analog multipliers or dividers will allow a voltage to set (or modulate) the coefficients that determine frequency or period (respectively); third, the system can either free-run or be started at any arbitrary point in the cycle, determined by preset initial conditions; fourth, the damping can be set to produce exponentially-decreasing or increasing waveforms.

For the free-running case (stationary amplitude), a slight amount of regenerative damping ensures that the oscillation will build up.

When one of the outputs reaches a level established by comparison with an amplitude reference, degenerative damping is applied at the peaks, reducing the last increment of buildup, and maintaining successive peaks at the same amplitude. While this introduces some distortion, it is integrated (smoothed) before appearing at one of the outputs and is integrated again before appearing at the other output.

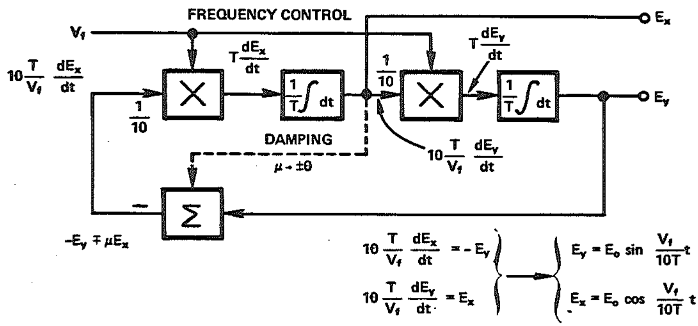


Figure 8. Block diagram of variable-frequency 2-phase sinusoidal oscillator. For fixed frequency, replace multipliers by coefficients. If not free running, apply initial conditions to integrators in SET. If driven from summing-point as 2nd-order filter, $\mu > 0$, E_y is low-pass output, E_x is band-pass, and output of Σ is high-pass.

A PRACTICAL 2-PHASE SINE-WAVE OSCILLATOR

Low-cost, high-performance complete-on-a-single-chip IC multipliers, such as the AD533, make it feasible to build oscillators having two-phase sine-wave output, with frequency controllable by a voltage. The frequency may be varied over a wide range, depending on the dynamic range of the multiplier, for frequency-sweep applications, or it may be centred about a fixed frequency for highly-linear frequency modulation. While an IC multiplier is used for the example of Figure 9 because of its low cost, there is no inherent barrier to using a wideband multiplier, such as the 429, for increased bandwidth, or a high-accuracy multiplier for increased low-frequency accuracy and resolution, or even multiplying D/A converters, for digital control of frequency.

The oscillator shown in Figure 9[2] delivers a 2-phase sine-wave output tuneable over a 10:1 frequency range by means of the DC control voltage. The output amplitude is stabilized by zener reference diodes at about 7Vrms and maintained constant within 1dB over the range of frequencies.

The oscillator system consists of two integrators, A1 and A2, and a unity-gain inverter, A3, forming a negative feedback loop. The effective time constants ($T = a$ RC) of the integrators are varied by a pair of multipliers, M1 and M2, which serve to (in effect) increase the conductance of R1 and R2 as the control voltage is increased, thus decreasing the time constant and increasing the natural frequency. Viewed in terms of gain and phase, at frequency $f_n = 1/(2\pi a RC)$, with $a = 1$, ($V_f = 10V$, $a = V_f/10V = 10/10$) both integrators have 90° phase lag and unity gain, the multipliers also have unity gain, and there are three sign inversions, all of which looks like a loop gain of $1\underline{/0°}$ at f_n (and only at f_n).

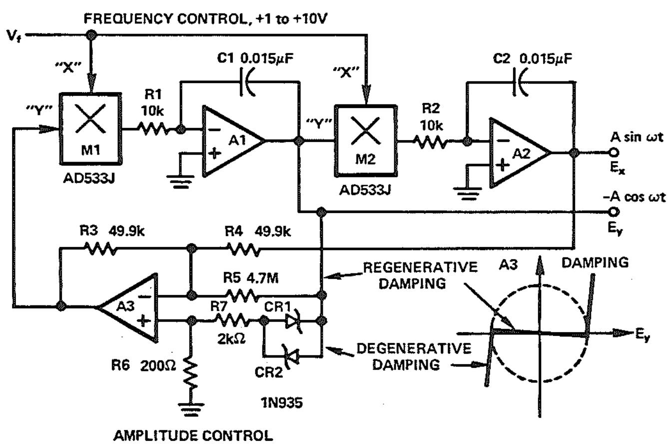


Figure 9. Practical version of configuration shown in Figure 8

To ensure sufficient regeneration to start and maintain the oscillation, a small amount of positive feedback is fed from the output of A1 through R5 to the input of A3. This causes the oscillation to build up until one or both of the zener diodes CR1, CR2, begin to

[2]"Frequency Modulator" by R.S. Burwen, *Analog Dialogue*, Volume 5, No. 5.

conduct at the tips of the waveform and produce increased negative feedback via the positive input of A3. The positive feedback must be kept small enough to provide buildup at a reasonable rate without requiring a large amount of negative feedback to keep the amplitude under control, since the zener diodes introduce some distortion. (Fortunately, this small distortion is integrated once in A1 and again in A2, so that the output of A2 is quite clean, and that of A1 is "oscilloscope-clean.")

With the values shown, the oscillator can be tuned from 100Hz to 1kHz. Distortion at the cosine output was measured at 0.74% at 100Hz and 0.46% at 1kHz. At the sine output, distortion was 0.64% at 100Hz and 0.18% at 1kHz. Distortion, especially at the lower end of the tuning range, is somewhat affected by the nonlinear feedthrough in the multipliers.* Multiplier nonlinearity and drift (using low-cost IC's) placed a limit on the useful tuning range.

It is easy to modify this design to operate with frequency modulation about a fixed frequency. For example, to operate at 1kHz, with ±10% frequency variation linearly controlled by V_f (±10V range), change R1 and R2 to 100kΩ, and add 10kΩ resistors between the output of A3 and the input of A1, and between the output of A1 and the input of A2.

SWEEP CIRCUITS

Linear sweeps, like the output of a triangular-wave generator, are usually produced by an integrator within a feedback loop; but instead of a linear retrace, a fast return is obtained by "dumping" the capacitor charge through a switch. The retrace is blanked (oscilloscope) or the pen lifted (recorder) during the retrace interval.

Nonlinear sweeps are desirable for some purposes. For example, in swept-frequency measurements, either the sweep may be logarithmic, or the frequency may be varied exponentially by applying an exponential input to control a variable-frequency oscillator.

*This distortion can be reduced by use of the "cross-feeding" technique for improving multiplier linearity, as discussed in Chapter 3-2.

(An ordinary linear display sweep may be used, since equal increments of time will represent equal ratios of frequency.) Starting at the high-frequency end, such a sweep can be obtained by passing a step through a simple RC coupling element (Figure 10). The output is $V_{in}\epsilon^{-t/RC}$. If it is used to control a frequency, the frequency will decrease by equal ratios in equal intervals of time.

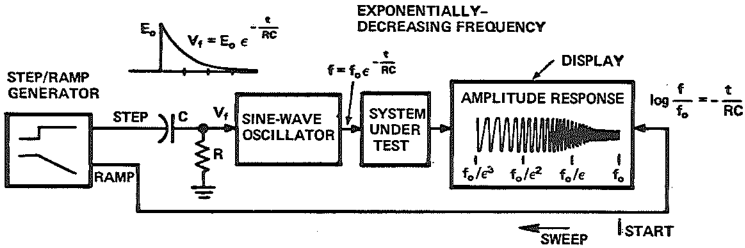


Figure 10. Use of logarithmic sweep for frequency-response measurements.

In this example, and throughout the chapter, it is tacitly considered that the rate of variation of "frequency" is so slow compared to variations *at* the frequency being controlled that there is little difficulty with the assumption that the waveform is stationary. Since this chapter deals with techniques rather than analysis, it must be assumed that for clearly interactive situations, in which frequency must be defined incrementally, the reader has an understanding of the mathematical implications and can deal with their consequences. The circuits, little caring about the complexity of the mathematics that describes their behavior, will perform nevertheless.

MARRYING ANALOG AND DIGITAL CIRCUITS

It is possible to generate linear sweeps of precisely-maintained amplitude and frequency, with arbitrary resolution, independent of the properties of capacitors and analog comparators, by driving a D/A converter with a counter that is itself driven by a train of pulses from a clock generator. The clock may be crystal-controlled, with frequency adjusted by counting down or a binary-rate multiplier, or it may be a simpler circuit.

Frequency depends only on the ratio of the clock rate to the total number of counts used, and amplitude can be scaled at the output of the converter. If the converter is a multiplying type, the sweep amplitude can be scaled by a voltage. The upper limit on speed is determined by the maximum clock rate, resolution, and settling-time of the converter. The converter circuitry should be "glitch-free," that is, there should be no large spikes at major-carry transition points (e.g., from 0 1 1 1 1 to 1 0 0 0 0).

A digitally-generated sweep, of appropriate resolution, with (or without) filtering may be applied to an analog function-fitter circuit (Chapter 2-1) to generate waveforms of any shape, in the same way that a purely-analog sweep might be applied (Figure 11). This is often a good deal less costly and more versatile than using a read-only memory (ROM) for shaping. Yet, like a ROM, it has the added possible benefit of being completely under the time control of the system. Not only is it slaved to the clock frequency— it can be started, stopped, *held* indefinitely, and reset, with simple logic circuitry. This would appear to be a happy combination of the best of analog and digital technology, characterized by simplicity, low cost, and versatility.

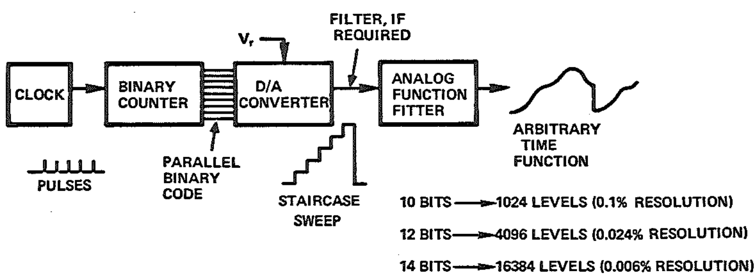


Figure 11. Arbitrary analog waveforms synchronized to digital clock

VOLTAGE-TO-FREQUENCY CONVERSION

The circuits of Figure 3 and Figure 5 are, in a sense, voltage-to-frequency converter circuits, but they have several limitations. Perhaps the most serious is that the range of continuous variation is limited, at best, to about 100:1. Also, they cannot be easily synchronized without some means of "dumping" capacitor charge.

Figure 12 shows a more-sophisticated circuit that is capable of 1:10,000 resolution and nonlinearity, gain stability (with external reference) to within 10ppm/°C, and practically negligible sensitivity to the dc power supplies. It is operated by a 100kHz clock, to which the output is synchronized. For an input variation of 0 to -10V, the output frequency varies proportionally from 0 to 50kHz. A synchronized 50kHz pulse train is also available, as a frequency reference.

$$f = 5 \times 10^4 \left(\frac{V_{in}}{-V_r}\right) \tag{18}$$

The AD301A amplifier operates in the linear mode; that is, the negative input terminal tracks the voltage at the positive input. Therefore, the current through R, equal to V_{in}/R, flows toward the capacitor. Q1A is a switching transistor that either has zero collector current, or a current equal to $V_r/½R$, flowing *away* from the capacitor. When Q1A turns on, the capacitor is charged by the net current $(2V_r + V_{in})/R$; When Q2 turns off, the capacitor discharges at $^-V_{in}/R$. Thus, to maintain equilibrium, for each time Q1 charges, the number of equal intervals spent discharging must be $(2V_r + V_{in})/-V_{in} = -2V_r/V_{in} - 1$. If each interval is 10μs, the total time per charge-discharge cycle is 10μs $(1 - 2V_r/V_{in} - 1) = -20V_r/V_{in}$μs. If, now, each charge-discharge cycle produces a pulse, the number of pulses per second will be $5 \times 10^4(-V_{in}/V_r)$, as noted in (18).

When output Q of the flip-flop is low, the emitter voltage of Q2 is less than the base voltage, and it is turned off. Since the bases of Q1A and Q1B are driven together, and the emitter circuitry is identical, their collector currents should track rather precisely. Thus, the collector current of Q1A should be equal to $2V_r/R$. When output Q of the flip-flop is high, Q2 is able to conduct; it furnishes enough current through the emitter resistor to raise the emitter voltage of Q1 above the base line, turning off the collector current.

Whenever the output of A1 is slightly below the threshold of the D input of the flip-flop, the next pulse causes $\overline{Q}$ (the output of the circuit) to go high. It also causes current to flow through Q1A, and a large increment of charge to raise the output of A1 by $\Delta V_1 = I_1 \Delta t/C$ (where $I_1 = (2V_r + V_{in})/R$). The next clock pulse finds the output of A1 high, $\overline{Q}$ goes low, and Q goes high, cutting off the flow of current through Q1A. The decrease of charge during this interval is $\Delta V_2 = I_2 \Delta t/C$, (where $I_2 = -V_{in}/R$). At the next clock pulse, unless $V_{in} = -10V$, the output of A1 is still high, $\overline{Q}$ remains low, and Q remains high, allowing a further decrease of charge. This process is repeated until the output of A1 is again slightly below the threshold of the D input, a cycle has been completed, and a new cycle begins.

When $V_{in} = -10V$, the charge and discharge periods are equal in number, and the output is at a 50kHz rate. The second half of the flip-flop counts down by 2, so that the reference pulse train is also at 50kHz.

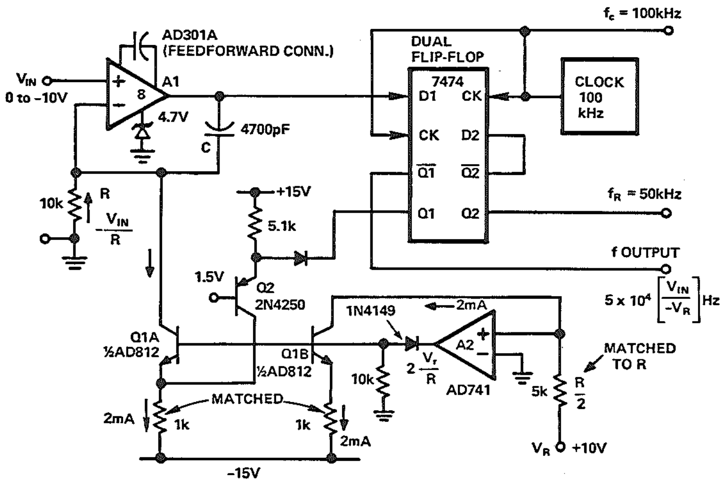


Figure 12. High-accuracy synchronized voltage-to-frequency converter

CONCLUSION

This chapter has sought to give an overview of function generators in general, and to provide details of a few useful circuits in particular. The objective is to arouse interest in special-purpose function generation, with particular emphasis on the cooperative rôles of linear and nonlinear analog devices and the possibilities of their fruitful collaboration with digital circuits.

II Instruments & Data Acquisition

Chapter 3

The design of instruments and "front-end" circuitry for data-acquisition systems is perhaps the area of greatest prospective payoff to users of nonlinear computational devices.

The circuits discussed in this chapter produce analog information that may be either directly read out by a human operator, or digitized and transmitted from a remote location to a control center, without requiring further interpretation. The analog data-reduction circuits covered here are simple and more-or-less universally applicable. The closely-related treatment of measurement and control circuits in Chapter 2-5 complements (and to some extent overlaps) the material presented here, and is somewhat similar in basic form; but it tends to include more-ramified analog computation, applied to situations that are more specialized.

ANALOG DATA REDUCTION

The primary goal of the configurations discussed here is to *reduce* data by analog techniques. To *reduce* data, as used here, means to extract significant information from one or more analog inputs, and transmit it –either to the human eye or to an interface– as meaningful, compact, well-paced data.

For a single variable, data-reduction can consist of extracting the peak, average, RMS, mean-square, or some other measure that is consistent in the presence of large numbers of individual data points. If the process is *stationary*, it may involve an average; if one-shot, it may call for a peak, integral, or final-value. If the measurement is nonlinear, it may call for *linearization*; if wide-ranging, it may call for *compression*.

If the measured data comprises many variables, further combination may be in order: summing and differencing (linear, vector, or root-square), ratios or products (linear, log, or otherwise), multiplexing.

The reduced data may be read out via analog or digital panel meters. Of it may be digitized (perhaps by a digital panel meter, that also provides a readout) and transmitted in digital form, to a remote control station (for further processing or remote printout on a teletypewriter or CRT terminal) via some compatible system, such as SERDEX*.

Figure 1 shows a single-channel data-acquisition subsystem, typical of those encountered in the *Analog-Digital Conversion Handbook.*[1] Whereas much space is given, in that volume, to preamplification, grounding, conversion, sample-hold, and analog multiplexing, this chapter (and related chapters) will be concerned with the blocks in which analog data is transformed into more-useful (but still analog) forms to meet specific needs.

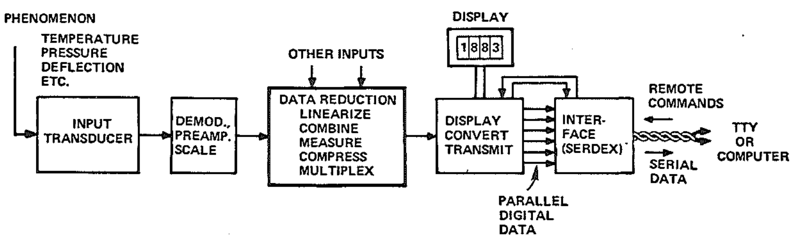


Figure 1. Typical data-acquisition channel

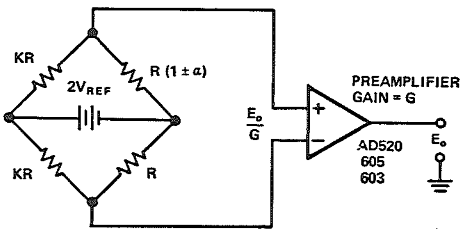
*SERDEX: SERial Data EXchange (Analog Devices trade-name), a means of simply controlling conversion processes, and transmitting data and commands in serial ASCII format under control of a teletype keyboard (or a computer programmed in a high-level language, such as BASIC) via an isolated current loop employing a simple twisted-pair of wires. While not strictly within the scope of this volume, it is nevertheless of great potential usefulness to the hardware-oriented analog-digital system designer. Complete data and applications information is available from Analog Devices, Inc.

[1] *Analog-Digital Conversion Handbook*, edited by D.H. Sheingold, Analog Devices, Inc., 1972, 402pp.

LINEARIZING

The system designer must strike an economic balance between convenience of measurement and convenience of dealing with the measured information. The simplest and most convenient transducers frequently have a nonlinear relationship between the variable being measured and the electrical output. Linear transducers, if available, often turn out to be less sensitive, more costly, or difficult to implement. Linearization can make it possible to obtain greater sensitivity by using nonlinear regions that are usually shunned.

For example, the simple Wheatstone bridge, a 4-terminal device used in a wide variety of pressure, force, strain, and electrical measurements, has an inherent nonlinearity (Figure 2a), which increases with sensitivity (e.g., it is 50% at K = 1, α = −1). By opening one leg, and using a readout operational amplifier to drive a portion of the bridge, one can obtain linear response (Figure 2b).



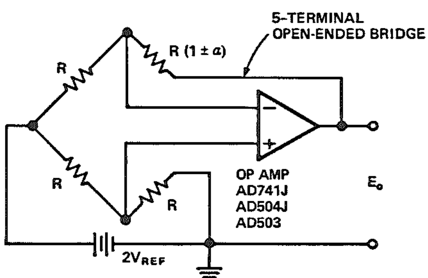
$$\frac{E_o}{G} = 2V_{REF}\left[\frac{R(1+a)}{(1+K)R + aR} - \frac{R}{(1+K)R}\right]$$

$$= 2V_{REF} \cdot \frac{K}{1+K} \quad \frac{\frac{a}{1+K}}{1+\frac{a}{1+K}}$$

IF K = 1,

$$\frac{E_o}{G} = V_{REF}\frac{a/2}{1+a/2} = V_{REF}\frac{X}{1+X}$$

a. Nonlinear response of Wheatstone bridge



$$V_{REF} = \frac{R(1+a)}{R(1+a/2)} V_{REF} + \frac{R}{2R(1+a/2)} E_o$$

$$E_o = -aV_{REF}$$

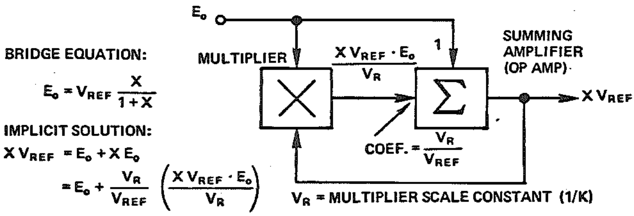
b. Linear version of bridge using operational amplifier

Figure 2. Nonlinear and linear bridge circuits

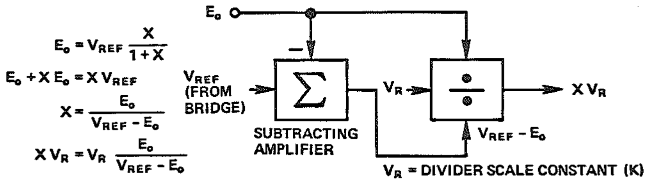
However, there are a number of significant costs: First, an amplifier *must* be used (whereas a Wheatstone bridge can be read out with a passive analog meter); also, five terminals are necessary, and the cable connecting the amplifier with the transducer affects loop stability; in addition, if gain is needed, an extra amplifier is needed; finally, 4-terminal bridges are cheap, widely available, and standard in many transducers —which leaves the designer with no alternative at the transducer level.

Fortunately, bridge nonlinearity is described by a simple mathematical relationship, and it can be compensated for completely by the use of a multiplier and an operational amplifier, as we have indicated in Chapter 2-1. The simplest approach is to use the configuration of Figure 3a, where implicit feedback is used to obtain the inverse of the bridge nonlinearity function. It has the benefit of summing a purely-linear term with a correction term. It is also possible to compute the inverse directly, using division (Figure 3b). Although this approach makes good use of a divider (the maximum dynamic range of the denominator is only 3:1), it relies on the inherent linearity of the divider over the whole range of variation. Since, at full scale ($\alpha = 1$, $K = 1$), the correction term is 50% of the output (Figure 3c), multiplier nonlinearities in the circuit of Figure 3a are in effect attenuated by 50%, while the divider nonlinearities are not attenuated. On the other hand, if V_{REF} (in the denominator) is the actual bridge-reference voltage, the divider circuit will also compensate for reference-voltage variations.

The correction terms should be scaled to represent the portion of the range of resistance variation represented by α. Usually, a transducer is chosen to operate over the most linear portion of the bridge's range (small α and large K or large α and small K) to avoid the need for linearization. But this means throwing away sensitivity and signal-to-noise ratio for the sake of linearity, since the output is in either event a small fraction of the supply voltage. *A major advantage of linearization* is the prospect of using a more sensitive (albeit grossly nonlinear) bridge, in which the variable arm can conceivably go from zero to more than 200% of the fixed resistance, to deliver outputs comparable in magnitude to the

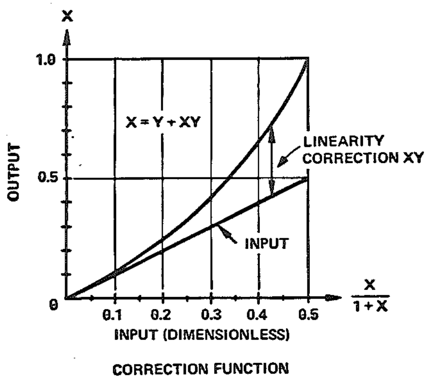


a. Bridge linearizer using implicit solution



b. Bridge Linearizer using divider. Note that gain of this ratiometric circuit can be made independent of V_{REF}

X	$Y = \frac{X}{1+X}$	CORRECTION
0.	0.	0.
0.0101	0.01	0.0001
0.0204	0.02	0.0004
0.0526	0.05	0.0026
0.1111	0.10	0.0111
0.25	0.20	0.05
0.429	0.30	0.129
0.667	0.40	0.267
1.00	0.50	0.500



c. Tabulation and plot of bridge linearization function

Figure 3. Bridge linearization circuits

bridge-supply voltage.

It is evident that, for a high-level signal, a preamplifier is unneeded. If the bridge supply is floating, the multiplier and summing amplifier can be single-ended. If the bridge supply is returned to system ground, one can use amplifiers and multipliers that have differential inputs.

But bridge linearization alone may not be enough. The tacit assumption has been that the resistance variation, αR, is proportional to the primary variable that causes the resistance to vary. But what if the resistance variation, α, is itself a nonlinear function of the primary variable? The designer has two choices: to linearize the bridge and resistance functions separately, or to linearize the overall response (Figure 4). The former has advantages of using standard circuitry and eliminating immediately a predictable source of nonlinearity (usually the major one); the latter has the advantage of possibly simpler and less-costly circuitry (but perhaps involves greater setup cost). With either approach, the designer can use a function fitter (Chapter 2-1) that employs either a smooth or a piecewise-linear approximation to the inverse of the function to be linearized.

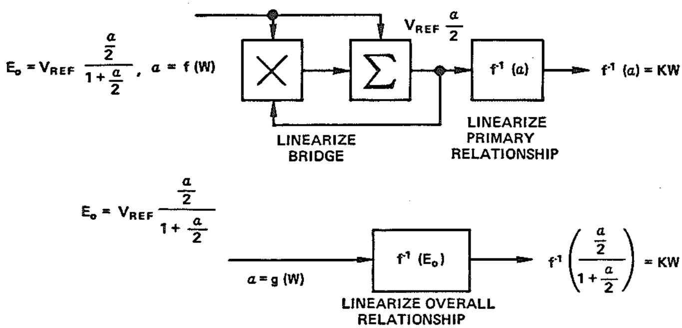
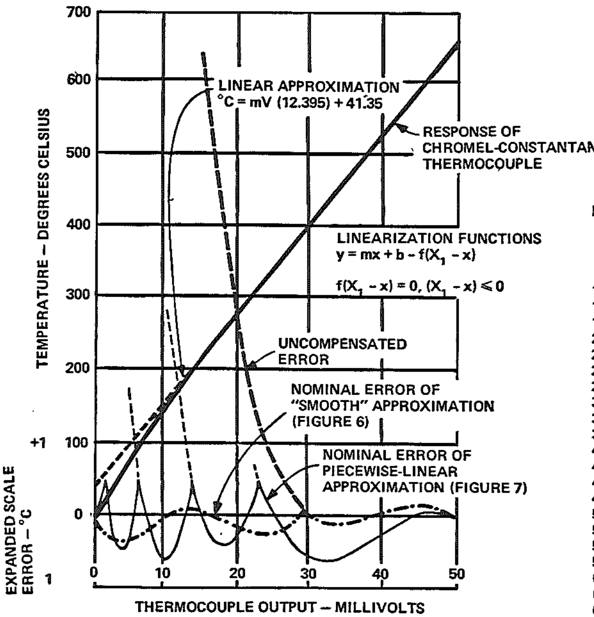


Figure 4. Two ways of linearizing a bridge-transducer measurement when the deviation is a nonlinear function of the primary variable (W)

LINEARIZATION EXAMPLE: THERMOCOUPLE

Figure 5 includes a tabulation of the relationship[2] between temperature and output voltage of a nickel-chromium X copper-nickel (Chromel-Constantan) thermocouple, with 0°C reference junction, over the range from 0° to 661.1°C (0 to 50mV). From the plot, it can be seen that the output is linear within ±1°C from about 340°C to beyond 650°C. The deviation from linearity increases at lower temperatures to about 40°C at zero.

[2]The figures in the table are based on a tabulation in *The Omega Temperature Measurement Handbook* (1973), page A-9, published by Omega Engineering, Inc., Stamford, Connecticut 06907, based on 1971 figures from the National Bureau of Standards.



Degrees C	Millivolts
0	0
50	3.047
80.26	5.000
100	6.317
150	9.787
153	10.
200	13.419
221.21	15
250	17.178
286.71	20
300	21.033
350	24.961
350.49	25
400	28.943
413.2	30
450	32.96
475.27	35
500	36.999
537.09	40
550	41.045
598.95	45
600	45.085
650	49.109
661.1	50

Figure 5. Nonlinear thermocouple response and theoretical residual errors, using two different linearizing functions

If, for any reason, it is necessary to obtain linear temperature measurement within ±1°C over the indicated temperature range, using this device, a linearizing circuit must be used to obtain an output voltage proportional to temperature, given the millivolt output (mV) of the thermocouple. Following preamplification, two approaches may be used to compensate for the nonlinearity at the lower end: a smooth fit, or a piecewise-linear fit. Examples of the errors experienced with a smooth cubic fit, and with a 5-segment piecewise-linear approximation, are shown in Figure 5.

In both cases, the desired response is fit by an operation of the form

$$°C = \underbrace{(\text{slope})(mV) + (\text{intercept})}_{\text{linear portion}} - \underbrace{f(\text{mV}_0 - mV)}_{\text{correction}} \qquad (1)$$

For values of mV greater than the threshold, mV_0, the correction term is zero. In both cases, a "breakpoint" enforces this condition. For the specific case considered here, the correction functions providing the theoretical error plots in Figure 5 are:

$$f = 0.2391(28.943 - mV) + [0.09464(28.943 - mV)]^{3.512} \quad (2)$$

and

$$f = 0.6356(23 - mV) + 1.021(13.419 - mV) + 1.473(6.317 - mV) + 1.17(1.495 - mV) \quad (3)$$

In both equations, the coefficients of the bracketed terms are positive if the bracketed terms are positive, and zero if they are negative. A circuit corresponding to (2) is shown in Figure 6, and one corresponding to (3) is shown in Figure 7. The fitting

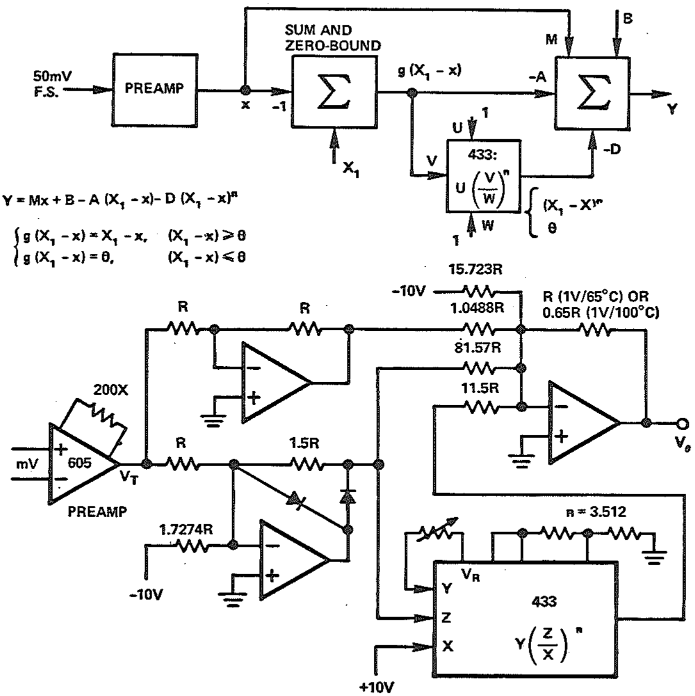


Figure 6. Block diagram and circuit for linearizing, using smooth approximation

process is aided by using slightly different coefficients for the linear portion: (12.395) (mV) + 41.35 for the exponential case, to reduce error, and (12.424) (mV) + 39.88 for the piece-wise-linear case, to reduce the number of breakpoints.

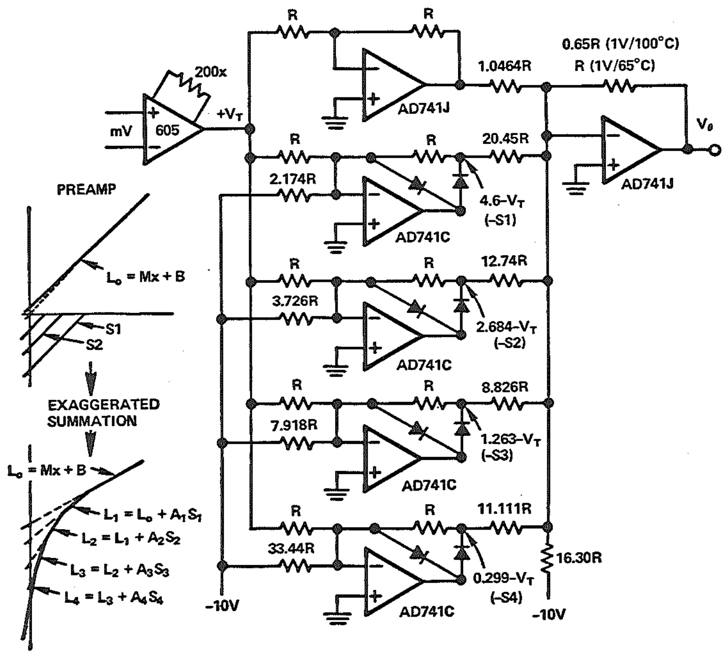


Figure 7. Circuit for linearizing using piecewise-linear approximation

The values calculated for the plots were carried out to a sufficient number of places to make the computational errors negligible for the idealized configuration. However, it should be evident that, for a practical circuit, the tolerances can in most cases be considerably looser than the numbers in (2) and (3) imply. The functions were fitted using the techniques discussed in Chapter 2-1. Then, as noted there, the next steps are to derive electrical scaling, nominal circuit values, and allowable device tolerances.

Equations (4) and (5) are electrically-scaled equations for the two cases, assuming that a gain-of-200 preamplifier is used, providing 1V/5mV at the input of the linearizer. The output scale factor is

1V/65°C. *The resistance values in Figures 6 and 7 are the nominal values required to provide the needed gain relationships. The table following equation (4) lists the tolerances necessary to embody that equation with less than 0.13°C (i.e., 2mV) error contribution by each term. V_θ is the scaled output voltage, and V_T is the scaled thermocouple voltage.

$$V_\theta = 0.9535 V_T + 0.6362 - 0.01226(8.6835 - 1.5 V_T) - (0.087)(10)\left[\frac{8.6835 - 1.5 V_T}{10}\right]^{3.512} \qquad (4)^\dagger$$

The tolerances of the terms in (4), determined by differentiation, ($\partial V_\theta / \partial A_i = S_i$), setting ΔV_θ = 2mV, solving for $\Delta A_i / A_i = 2/A_i S_{i_{max}}$, and rounding down, are:

0.9535	0.02%
0.6362	0.3%
0.01226	1.5%
8.6835	0.1%
1.5	0.3%
0.87	0.35%
10 (denom)	0.1%
3.512	0.8%

The "ideal-diode" limiting circuit in Figure 6 ensures that the bracketed terms have no contribution when negative.

Equation 5 is the scaled equation for the piecewise-linear case. Though the error is "lumpier" than that of the exponential approximation, and there are more circuit details to attend to, the circuitry is repetitive, and the tolerances are somewhat looser.

*This scaling was chosen to obtain the benefits of using the full output range. Though not making use of the full-scale range, a scale-factor of 100°/1V would permit direct readout of temperature on a 3 or 4-digit panel meter. It can be obtained without further modification by appropriately attenuating the output.

†This equation is derived from equation (2) by normalizing it, then setting the normalized equation equal to a normalized electrical equation, thus arriving at the constant voltages and coefficients. A further step was to recognize that it would be advantageous to use the major portion of the full-scale range of nonlinear devices. To do so, the difference terms were multiplied and divided by 1.5, resulting in larger input voltages and smaller overall coefficients. The constants in the exponent term were manipulated to provide the 10V denominator and 10V input multiplier desirable for a $Y(Z/X)^m$ device, in an application where Z is the only active input.

$$V_\theta = 0.9557V_T + 0.6135 - 0.0489(4.6 - V_T) - 0.0785(2.684 - V_T)$$
$$- 0.1133(1.263 - V_T) - 0.09(0.299 - V_T) \qquad (5)$$

The tolerances of the terms in equation (5) are:

0.9557	±0.02%
0.6135	±0.3%
0.0489	±0.75%
4.6	±0.8%
0.0785	±0.9%
2.684	±0.9%
0.1133	±1.2%
1.263	±1.2%
0.09	±5%
0.299	±5%

for less than ±2mV error from any term, or 6.3mV (0.41°C) root-sum-of-squares error (allowing for ±0.6°C of theoretical error). Tolerances, as applied to circuit elements, should take into account resistance-ratio mismatch, and the drift variations of amplifiers, resistances, and references, with time and ambient temperature, as well as scale factor, drift errors, and shape errors of the exponentiating device (e.g., the Model 433, if applied as the $Y(Z/X)^m$ in Figure 6).

It is interesting to note, as an exercise in function fitting, the value of plotting a curve. While it would appear natural to fit a function by seeking the best numerical fit, starting with a linear slope from the origin, this case proves the contrary. From the plot, it is immediately obvious that the departure from linearity is *greatest* at the origin, and that the most rewarding approach is to offset and reverse the "origin of nonlinearity."

AMPLITUDE COMPRESSION

If the result of an analog measurement, having a modest frequency content and a wide range of variation, must be made available at some distance, with an intervening noisy medium that is likely to result in pickup and loss of amplitude information, the designer has a number of possible options. Popular ones include:

1. Transmission as a frequency-modulated signal

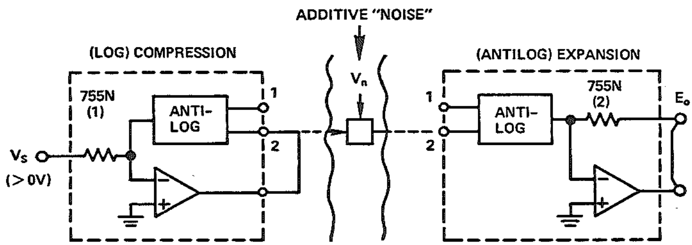
2. Conversion to digital form and transmission either serially or in parallel
3. Logarithmic compression and analog transmission
4. Logarithmic compression and digital transmission

Some general comments can be made about these options:

1. Frequency modulation calls for wide bandwidth, depending on the dynamic range and frequency content of the signal, and highly-linear modulation and demodulation. If the DC level is important, a precise phase or frequency reference must be made available.

2. A/D conversion requires adequate resolution (16 bits for less than 30% error for the smallest signal in a 10,000:1 dynamic range). Adequate sampling rate and bandwidth, and a stable clock are necessary for serial (2+ wires) transmission; many wires are required for parallel transmission. All alternatives are costly, but SERDEX (see Page 88) is more convenient than most, if its bit-rate is adequate.

3. Logarithmic compression can be implemented at low cost (Figure 8). The signal-to-noise ratio of a compressed signal depends only on the noise level and the choice of log scaling; it is essentially independent of the signal level over a wide dynamic range. Band-width requirements are those of the analog signal, in its compressed



$$E_o = V_{r2} \log^{-1}\left[-\frac{1}{K_2}\left(V_n - K_1 \log \frac{V_S}{V_{r1}}\right)\right]$$

IF $V_{r2} = V_{r1}$ and $K_2 = K_1 = K = 2$

$$\frac{E_o}{V_S} = 10^{-V_n/K} \quad \text{(independent of } V_S\text{)}$$

$$\left(\text{FOR THE CASE OF UNPROCESSED TRANSMISSION, } E_o = V_S + V_n,\ \frac{E_o}{V_S} = 1 + \frac{V_n}{V_S}\right)$$

Figure 8. Log compression used for improving dynamic range of transmitted signal

form. Though suitable for transmitting small or large signals impartially, the compression process is inherently insensitive to small signal components riding on larger signals. That is, the signal-to-noise ratio, even when mediocre, is independent of amplitude.

4. Logarithmic compression, combined with digital transmission (Figure 9), results in greatly-increased signal-to-noise, as long as the induced noise is below the logic thresholds. A further advantage of compression is the reduction of the required digital resolution: a 10,000:1 dynamic range can be comfortably resolved to within 1% of the actual value at any level using a 12-bit converter (cf. 2 above). Besides the obvious cost savings, there is also a slight reduction of the number of wires (parallel transmission) or an improvement in speed (serial transmission).

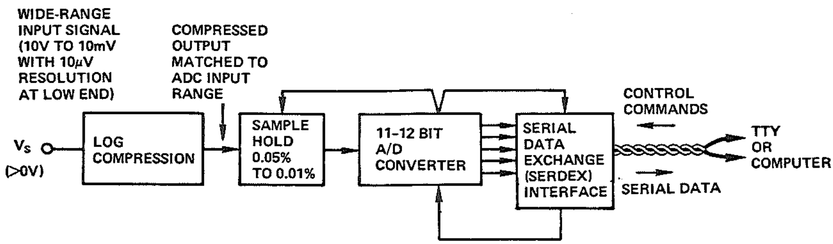


Figure 9. Log compression allows signal having wide dynamic range to be converted to digital at moderate resolution, and transmitted digitally via standard twisted-pair 20mA current loop with high noise immunity

The logarithmic compression process involves a logarithmic operator, such as the Model 755N, which computes $-K\log_{10}(V_s/V_r)$, where K may be 1V or 2V (per decade), and $V_r = 0.1V$. If the signal is transmitted in this logarithmic form (K = 2), a span of 10,000:1 of V_s is translated to a span of 8V at the compression output. An input swing of 1-10V will produce a 2V output change; so will an input swing of 1mV to 10mV. Thus, high-level signals are attenuated (average gain = 2/9 = 0.22) and low-level signals are amplified (average gain = 2/0.009 = 222). Noise picked up or induced in transmission will add to the logarithmic version of the signal. That signal-to-noise ratio can be greatly improved for small signals should be evident.

At the receiving end, the Model 755N is used as an antilog operator (a difference "instrumentation" amplifier may be used to reject common-mode errors, if appropriate), producing the inverse operation: $V_r\,(10)^{V_{in}/-K}$. Table 1 shows what happens to an instantaneous voltage V_s —assuming ideal (or matched) log conformance, that the V_r's are matched, and that the K's are adjusted for net unity gain—in the presence of a spurious instantaneous "noise" voltage, V_n. The "signal-to-noise" ratio with logarithmic compression is compared to what it would be without compression.

It can be easily seen that the signal-to-noise ratio depends only on the noise level, and that 10mV of noise is rejected in the same ratio, whether the signal is 1mV or 10V. While linear transmission does a much better job at high levels, it is virtually useless at low levels. Table 2, which is extracted (and interpolated) from Table 1, shows the comparable dynamic range available at different choices of signal-to-noise level.

TABLE 1. RESPONSE IN THE PRESENCE OF NOISE

V_s sig.	$-K\log(V_s/V_r)$ V_T	V_n	V_n+V_T	E_o	$\lvert E_o-V_s\rvert$ error	S/N log	S/N lin.
1mV	+4V	− 1mV	+3.999V	1.001mV	1.15μV	868	1
		− 10mV	+3.99V	1.012mV	11.6 μV	86	0.1
		+100mV	+4.1V	0.891mV	0.11mV	9.2	—
		−100mV	+3.9V	1.12 mV	0.122mV	8.2	—
		+ 1V	+5V	0.32mV	0.7mV	1.5	—
		− 1V	+3V	3.16mV	2.2mV	0.5	—
10mV	+2V	− 1mV	+1.999V	10.01mV	11.5μV	868	10
		− 10mV	+1.99V	10.12mV	116 μV	86	1
		−100mV	+1.9V	11.22mV	1.22mV	8.2	0.1
		− 1V	+1V	31.6mV	22mV	0.5	—
100mV	0V	− 1mV	-0.001V	0.1001V	115 μV	868	100
		− 10mV	-0.01V	0.101V	1.16mV	86	10
		−100mV	-0.1V	0.112V	12.2mV	8.2	1
		− 1V	-1V	0.316V	0.22V	0.5	—
1V	-2V	− 1mV	-2.001V	1.001V	1.15mV	868	10^3
		− 10mV	-2.01V	1.012V	11.6mV	86	10^2
		−100mV	-2.1V	1.122V	122mV	8.2	10
		− 1V	-3V	3.16V	2.2V	0.5	1
10V	-4V	− 1mV	-4.001V	10.01V	11.5mV	868	10^4
		− 10mV	-4.01V	10.12V	116mV	86	10^3
		−100mV	-4.1V	11.22V	1.22V	8.2	10^2
		− 1V	-5V	31.6V!!	21.6V	0.5	10

TABLE 2. DYNAMIC RANGE VS. SIGNAL-TO-NOISE RATIO

	|Noise|	Dynamic Range Log Channel	Dynamic Range Linear Channel
A. S/N >865	1mV	10V:1mV	10V:865mV
	10mV	–	10V:8.65V
B. S/N >85	1mV	10V:1mV	10V:85mV
	10mV	10V:1mV	10V:850mV
	100mV	–	10V:8.5V
C. S/N >8.5	1mV	10V:1mV	10V:8.5mV
	10mV	10V:1mV	10V:85mV
	100mV	10V:1mV	10V:850mV
	1V	–	10V:8.5V

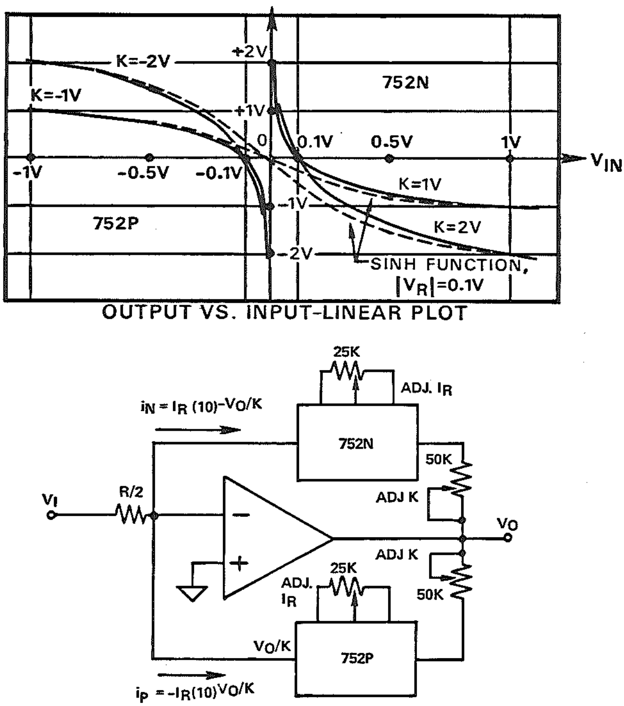
To determine the effects of errors in the log devices (especially variations of the coefficients with temperature), the complete relationship may be used:

$$E_o = V_{r_2} \cdot 10^{\left[\frac{K_1}{K_2} \log \frac{V_s}{V_{r_1}} - \frac{V_n}{K_2}\right]}$$

$$= \left[\frac{V_{r_2}}{V_{r_1}{}^{K_1/K_2}}\right] \cdot V_s{}^{K_1/K_2} \cdot 10^{-V_n/K_2} \qquad (6)$$

If K_1 and K_2 are equal and track one another, and if V_{r_1} and V_{r_2} are equal and tracking, $E_o = V_s \cdot 10^{-V_n/K}$, giving the results in column 5 of Table 1. If they differ, equation (6) provides a means of exploring the errors. Errors of log conformance can be treated as additive values of V_n.

Since a logarithmic function is inherently unipolar (the logarithm is real only for positive values of the argument –positive signals require a 755N, negative signals a 755P), it is far from ideal for signals that are inherently zero-centered. While it may be useful to bias some types of input signals into a single polarity, functions that demand symmetrical treatment may be badly distorted by the wide variation, in both resolution and speed, between zero and full-scale input. Such functions would profit by a type of precise compression that is symmetrical about zero. An example of an easily-obtained form is a $\sinh^{-1}$ function (Figure 10), which involves two complementary antilog transconductors (752P and 752N) in the feedback path of an operational amplifier. The resulting function is logarithmic for larger values of input, but it passes through zero essentially linearly.



$$2\frac{V_I}{R} = i_N + i_P = I_R(10)^{-V_O/K} - I_R(10)^{+V_O/K}$$

$$-\frac{V_I}{I_R R} = \frac{\epsilon^{2.3V_O/K} - \epsilon^{-2.3V_O/K}}{2} = \sinh\left(2.3\frac{V_O}{K}\right)$$

$$-V_O = \frac{K}{2.303}\sinh^{-1}\left[\frac{V_I}{I_R R}\right]$$

Figure 10. Bipolar signal compression using complementary logarithmic transconductors to synthesize $\sinh^{-1}$ function

Used in the forward path, the pair provides an inverse function, proportional to the hyperbolic sine. Assuming appropriate symmetry, matching, and tracking, the overall response, in the presence of an added noise voltage, is (instant by instant)

$$E_o = V_r \sinh\left[\sinh^{-1} \frac{V_s}{V_r} - \frac{\ln(10)}{K} V_n\right] \qquad (7)$$

If the magnitudes of V_r and K are, respectively, 0.1V and 2V,

$$E_o = 0.1 \sinh(\sinh^{-1} 10V_s - 1.1513V_n) \qquad (8)$$

A table similar to Table 1 may be derived to compare signal-to-noise and dynamic ranges

TABLE 3. IDEAL RESPONSE OF BIPOLAR COMPRESSION/EXPANSION IN THE PRESENCE OF A NOISE VOLTAGE

V_s signal magnitude	$\sinh^{-1} 10V_s$	V_n "noise"	E_o output	$E_o - V_s$ error	S/N nonlin.	S/N linear
+10V	5.2983	−0.001V	10.0121V	0.0121V	827	10,000
		−0.01	10.1164	0.116	86	1,000
		−0.1	11.221	1.221	8.2	100
+1V	2.9982	−0.001V	1.00123V	0.00123V	810	1,000
		−0.01	1.0117	0.0117	85	100
		−0.1	1.123	0.123	8.2	10
+0.1V	0.8814	−0.001V	0.10017V	0.00017V	600	100
		−0.01	0.10164	0.0016	61	10
		−0.1	0.117	0.017	5.9	1
+0.01V	0.09983	−0.001V	0.01012	0.00017	85	10
		−0.01	0.01116	0.0012	8.6	1
		−0.1	0.0217	0.012	0.9	—
+0.001V	0.010	−0.001V	0.00112	0.00012	8.7	1
		−0.01V	0.0022	0.0012	0.9	—
		−0.1V	0.0125	0.012	—	—

While not quite as impressive as the logarithmic function, because of limited gain through the origin, the hyperbolic compression/expansion can be improved by extending the logarithmic portion of the range. If, for example, V_r is reduced to 0.01V, the signal-to-noise ratio for $V_s = V_n = 0.01V$, is increased to 60. The penalty that is paid is in the maximum speed through zero, since the effective feedback resistance and gain in the vicinity of zero are multiplied tenfold.

A final word: The $\sinh^{-1}$ function can also be used as a high-accuracy tapered null meter with calibrated (approximately-equal) intervals off-null for equal ratios of change. It has great sensitivity at null, wide dynamic range, and continuous indication of direction of approach to the null.

EXTRACTING A MEASURE

We have seen how nonlinear analog-computing circuits may be used to compensate for transducer nonlinearity and to reduce the effects of noise in transmitting data. A historically important factor in data reduction (especially in the days before the use of oscilloscopes), still in widespread use, and likely to remain so forever, is the use of the meter (i.e., *measure*) to "boil down" information to a simple reading or trend of readings that can be interpreted by the human eye instantly, and serve as the basis for a decision. Metering is also used in computer-control and measuring systems, where the computer seeks to reduce a large number of measurements to a few significant indications of the present status and trend of an element or a process, as ingredients of a series of decisions.

There is a large universe of computational meters, and it would be presumptuous in these pages to seek to accomplish more than to skim the surface lightly, picking out those operations that have great usefulness, universal validity, and specific appropriateness to the devices under discussion here. These include, to begin with,

Mean
Mean Absolute Value
Root Mean-Square
Peak (or Valley)

While one ordinarily thinks of free-running devices, with fixed averaging times, it might be profitable to at least consider, in addition, single-shot measurements and variable-period measurements.

In addition, there are a number of measurements that involve two or more variables. A few interesting and popular ones include

Power (instantaneous, peak, and average)

Energy and energy-per-cycle

Power factor (and phase angle)

Vector Sum (Root-sum-of squares)

Ratio and log ratio (dB)

The circuitry that produces these functions will usually operate on the input after it has been preamplified and scaled, and perhaps linearized, but quite often it can be implemented with devices having sufficient stability, input impedance, common-mode rejection, or what-have-you, to operate directly on the transducer output.

MEAN AND MEAN ABSOLUTE-VALUE

Figure 11a shows the usual circuit employed for a running average, a simple unit-lag. In Figure 11b, an inverting version is shown; the averaging time-constant can be increased, without resorting to high-value resistors, by employing "T" networks (Figure 11c) at the cost of increased voltage drift because of the attenuation-and gain. This circuit is adequate for determining averages of signals having high-frequency fluctuations and relatively-slow mean variations. The settling time for a step change is 4.61RC to 1%, 6.91RC to 0.1%, and 9.21RC to 0.01%. High-frequency attenuation is modest: 3db at $f_0 = 1/2\pi RC$, 20dB at $10f_0$, 40dB at $100f_0$, etc.

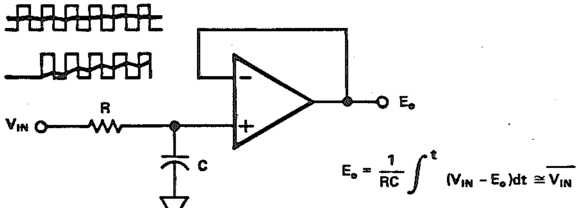
If the average must respond more quickly to changes of non-stationary functions, one needs a filter having a response continuously approximating the ideal average response over a period τ,

$$E_0 = \frac{1}{\tau}\int_{t-\tau}^{t} v(t)dt\text{, or operationally } \frac{1-\epsilon^{-\tau p}}{\tau p} \qquad (9)$$

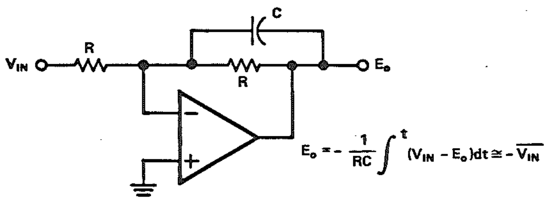
A reasonable approximation is the transfer function

$$\frac{e_o}{v_i} = \frac{1 + \overline{RCp}^2}{(1 + 1.2RCp + 1.6\overline{RCp}^2)\ (1 + 2RCp)} \tag{10}$$

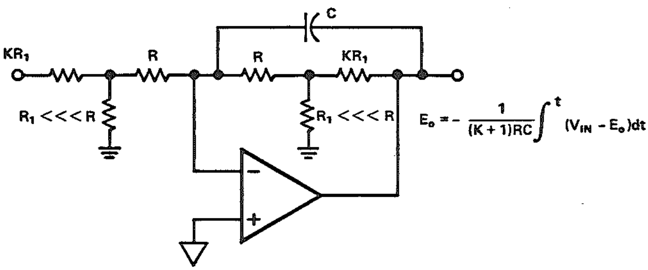
(where p is the Heaviside derivative operator). It can be stably embodied with a unit-lag and a second-order state-variable (integrator-loop) band-reject filter or, less stably but more compactly, with a single operational amplifier, 8 precision resistors and 6 precision capacitors.[3]



a. Non-inverting unit lag



b. Inverting unit lag

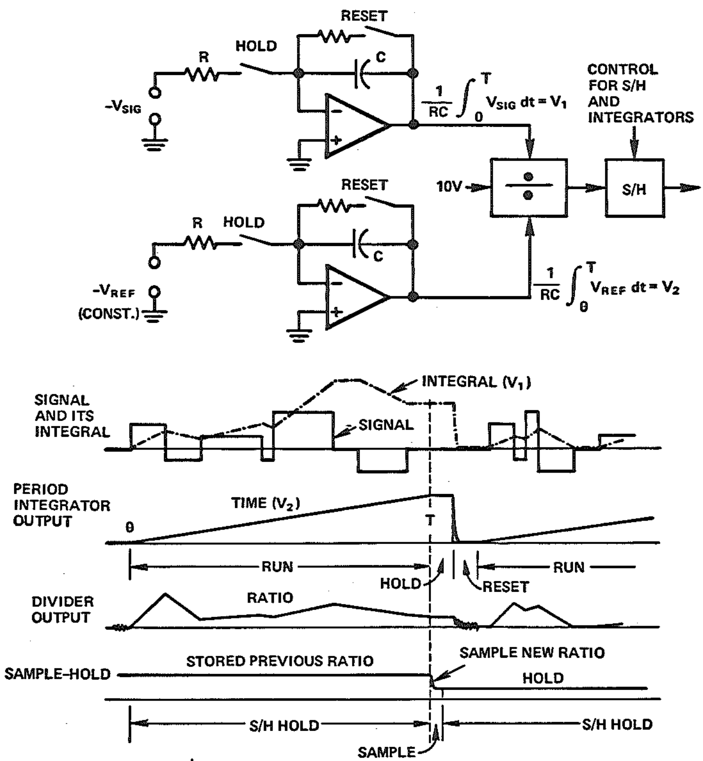


c. Long-period unit-lag, using T-network to avoid extremely-large R and C values

Figure 11. Classical unit-lag running-average circuits.

[3]"The Lightning Empiricist," Vol. 13, 1965.

Another kind of "running average" is the average over each cycle of a train of events having differing periods, for example, the blood pressure averaged over each heartbeat, or the volume of CO_2 averaged over each breath. This can be accomplished with two integrators, a divider, and a sample-hold (Figure 12). At the beginning of each cycle (determined independently), the two integrators (*signal* and *period*) are gated to *run*. At the conclusion of the cycle, the integrators are momentarily placed in *hold*, the divider output (accumulated signal divided by period) is *sampled*, and the integrators are then dumped in preparation for the next cycle, which may start immediately (or after a wait of arbitrary duration). Meanwhile, the sample-hold retains the last average reading



$$10\frac{V_1}{V_2} = 10\frac{\frac{T}{RC}\left.\overline{V_{SIG}}\right|_T}{\frac{T}{RC}V_{REF}} = \frac{10}{V_{REF}}\bullet\left.\overline{V_{SIG}}\right|_T \quad \text{(INDEPENDENT OF T)}$$

Figure 12. Averaging signals over variable periods

and holds it until updated. The divider-and-sample-hold could be a fast variable-reference A/D converter, especially desirable if long intervals are required between integrations.

Although averaging usually is considered to be a linear, or time-varying process (except where ratios are involved, as above), it has a place in the discussion of applying nonlinear devices in data reduction and measurement, because averaging is one of the most universally-used forms of data reduction.

The average, or mean, is not always directly appropriate, as a measure of a signal. For example, AC measurements seek to ignore the DC level (or *mean*), and instead concern themselves with the mean absolute deviation (from the mean). This is done by first establishing a "zero" level, usually the mean, then rectifying and averaging. Though there are many ways of accomplishing this, a widely used op-amp approach, involving "ideal diodes", is shown in Figure 13.

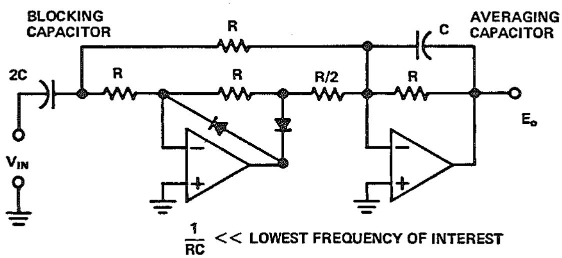


Figure 13. Circuit for computing mean absolute deviation at low frequencies. For wider bandwidth, external input follower and output averager will permit higher currents through diodes and greater RC's using reasonable capacitance. If input average is zero or can be zeroed, coupling capacitor is unnecessary.

ROOT MEAN-SQUARE

For many applications, particularly where voltage or current measurements provide information about the average energy generated, transmitted, or dissipated, the root-mean-square (rms) is a more useful measure. Straightforwardly, it involves squaring an input, taking the average, and obtaining the square root.

$$E_{rms} = \sqrt{\frac{1}{T}\int_0^T (V_{in})^2\, dt} \tag{11}$$

Classically, it has been measured by meters sensitive to the heating effect of an rms level. Electronically, because of the difficulty of tailoring a general-purpose instrument to fit the characteristics of a given thermal device, most "rms" meters for many years didn't

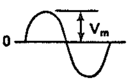
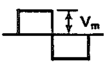
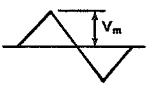
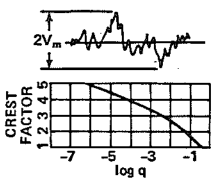
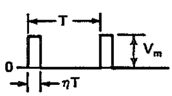
WAVEFORM		RMS	MAD	$\frac{RMS}{MAD}$	CREST FACTOR
	SINE WAVE	$\frac{V_m}{\sqrt{2}}$ 0.707 V_m	$\frac{2}{\pi}V_m$ 0.637 V_m	$\frac{\pi}{2\sqrt{2}} = 1.111$	$\sqrt{2} = 1.414$
	SYMMETRICAL SQUARE WAVE OR DC	V_m	V_m	1	1
	TRIANGULAR WAVE OR SAWTOOTH	$\frac{V_m}{\sqrt{3}}$	$\frac{V_m}{2}$	$\frac{2}{\sqrt{3}} = 1.155$	$\sqrt{3} = 1.732$
	GAUSSIAN NOISE CREST FACTOR IS THEORETICALLY UNLIMITED. q IS THE FRACTION OF TIME DURING WHICH GREATER PEAKS CAN BE EXPECTED TO OCCUR	RMS	$\sqrt{\frac{2}{\pi}}$ RMS = 0.798 RMS	$\sqrt{\frac{\pi}{2}}$ 1.253	C.F. / q 1 / 32% 2 / 4.6% 3 / 0.37% 3.3 / 0.1% 3.9 / 0.01% 4 / 63ppm 4.4 / 10ppm 4.9 / 1ppm 6 / 2×10^{-9}
 η: "DUTY CYCLE"	PULSE TRAIN η / MARK/SPACE 1 / ∞ 0.25 / 0.3333 0.0625 / 0.0667 0.0156 / 0.0159 0.01 / 0.0101	$V_m\sqrt{\eta}$ V_m $0.5V_m$ $0.25V_m$ $0.125V_m$ $0.1V_m$	$V_m\eta$ V_m $0.25V_m$ $0.0625V_m$ $0.0156V_m$ $0.01V_m$	$\frac{1}{\sqrt{\eta}}$ 1 2 4 8 10	$\frac{1}{\sqrt{\eta}}$ 1 2 4 8 10

Figure 14. RMS, MAD, and crest factor of some common waveforms. See also Table 1, Chapter 3-7, for additional waveforms.

measure rms at all. They measured the mean absolute deviation ("ac average") but indicated it on an "rms" scale calibrated to the ratio of rms-to-mean for sine waves. This was all right as long as sine waves were being measured (if they weren't badly distorted). It was even acceptable if signals having a more-or-less constant ratio of rms/mad, such as symmetrical square waves, Gaussian noise, or symmetrical triangular waves, were being measured, so long as a calibration was provided (see Figure 14). But, for unpredictable waveforms, variable-width pulse trains, and SCR'd sine waves, average-measuring rms meters were useless.

Accurate, wide-range "true-rms" circuits are made possible at reasonable cost by the availability of transconductance multiplier/dividers (XY/Z), such as the AD531, and by stable log-antilog circuits, such as the 433, and (more recently) by the 440 rms module. Examples of practical rms circuits can be seen in Chapter 3–7. The basic scheme is shown in Figure 15. It employs squaring, averaging, and *implicit* square-rooting. The crucial dynamic-range characteristic of rms devices is *crest factor*, the ratio of peak input to the rms value of the waveform. For example, an input signal having a dynamic range of rms of 20:1 and a crest factor of 5, calls for a device having substantially greater-than-100:1 dynamic range.

Straightforward open-loop schemes –square, average, root– call for excessive dynamic range internally; for example, a 100:1 ratio of maximum to minimum input, when squared, becomes 10,000:1, placing near-impossible demands on an open-loop square-rooter. When a transconductance-type multiplier-divider is connected for *implicit* square-rooting, the squarer's gain is controlled by the output, reducing the dynamic range, in the steady state, to first-order. Types employing logarithmic circuitry are even more effective, if slower, because they can reduce wide dynamic ranges to equal per-decade internal voltage swings.

For simplicity, filters are almost always first-order unit-lags. However, it is not unfeasible to specify filters having a more-nearly ideal "running-average" response.

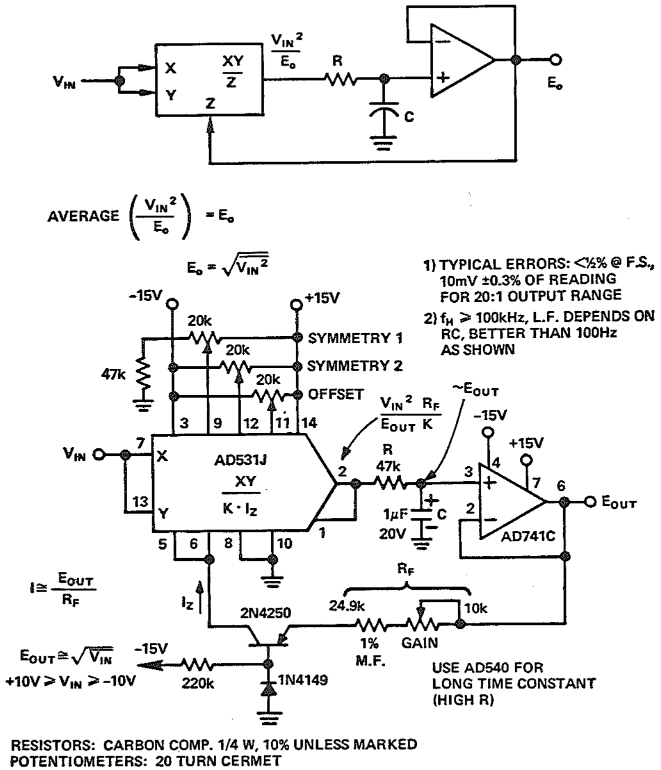


Figure 15. Block diagram of rms circuit and a practical circuit employing an I.C. multiplier-divider. See also Chapter 3-7.

PEAK AND VALLEY

For some purposes, averaging-type measurements do not provide adequate information about a waveform. Examples include periodic signals with rapidly-changing amplitudes, signals with variable crest factors, and amplitude-modulated waveforms. There are also applications in which it is necessary to determine the largest peak (or valley, or p-p spread) of a waveform over a given time interval.

Peak-detection-and-measurement circuits are numerous, and the choice (and cost) depends on the characteristics to be optimized. Such characteristics include accuracy, speed, leakage rate, sensitivity, and complexity. For free-running applications, a built-in leak is necessary; for one-shot applications, very long *hold* time (with little or no degradation) plus a *reset* circuit may be required. For noisy signals, the very definition of a "peak" may be in question, requiring either preliminary filtering, a hysteretic response (that ignores small fluctuations), or slow response (ignoring fast "blips").

The basic peak-measurement circuit consists of a comparator and a switched storage element. Figure 16 shows a simple circuit embodying the function. Operational amplifier A1 serves as the comparator. When the input voltage exceeds the charge stored on the capacitor, the amplifier acts as a unity-gain follower, causing the charge (supplied via the diode) to follow the input. When the input voltage drops back from the peak, the feedback loop is opened, and the capacitor retains its charge. Amplifier A2 unloads the capacitor and makes its voltage available at low output impedance. If low leakage is necessary, A1 and A2 must both have low-leakage inputs (e.g., FET's). A1 must be capable of fast recovery from the open-loop condition, it must have large phase margin, the ability to drive a capacitive load stably, and high input impedance in the open-loop condition (i.e., internal "protection" should be unnecessary).

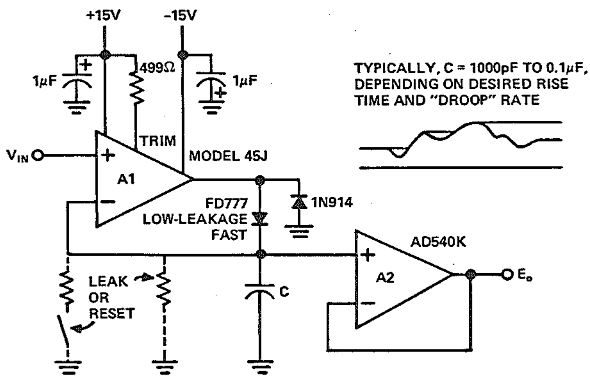


Figure 16. Peak-follower circuit

The capacitor determines both the charging rate and the "droop:" $dE_O/dt = I/C$. If 10mA are available for charging the capacitor (1000pF), the slewing rate is $10^{-2}/10^{-9} = 10V/\mu s$. If the total leakage current is 100pA, the droop rate will be $10^{-10}/10^{-9} = 0.1$V/s. If the capacitance is increased to 0.1μF, the droop rate will be reduced to 1mV/s, and the maximum charging rate will be 0.1V/μs. If the circuit is to free-run, a leak must be provided to allow downward variations of the peak level to be followed. A resistor will provide exponential decay (proportional to the last peak), and a current sink will provide linear decay at a fixed rate. The output of the peak follower may be averaged to determine the average variation of the peaks. If, on the other hand, the circuit is to provide a one-shot measurement of the highest of a series of peaks, a *reset* switch must be provided to discharge the capacitor before the next series of readings.

The negative-going edge at the output of A1 can be used to indicate that a peak has just occurred. If fast following and long *hold* are necessary to "catch" a single fast peak, two of these circuits may be cascaded, the first using a small capacitor (paralleled by a reverse-biased diode to the negative supply to ensure a small downward leak), and the second using a large capacitor for leisurely acquisition and long *hold.*

The "valley" follower is essentially the same circuit, but the diodes are reversed. It will track negative-going voltages that are below the stored level, and *hold* the lowest level experienced. For peak-to-peak measurements, the outputs of the followers can feed a simple subtractor-connected op amp. Alternatively, the Al-diode-capacitor portions of a peak and a valley circuit may be used, with the capacitor voltages applied to the inputs of a differential instrumentation amplifier, such as the AD520, the 603, or the 605. If the peak-to-peak circuit requires a leakage path to enable it to follow an envelope, the capacitors can feed directly a subtractor-connected FET input op amp, with resistors of appropriate magnitude for the desired leakage rate. As an added bonus, capacitors may be connected across the feedback resistors to filter out the cyclical swings of the peak measurements.

Usually, peaks are above ground, and valleys are below ground. However, if it is desired to measure peaks or valleys of widely-ranging signals anywhere in the range, this can be done by connecting the capacitor and the leak resistor (or the *reset* switch) to a voltage lower than the lowest peak (peaks) or higher than the highest valley (valleys), instead of to ground; typically, the negative and positive supply voltages serve the purpose. The *reset* switch should, of course, always have resistance in series for protection.

As indicated earlier, there are many circuits for peak-following. They include single op amps with diode-capacitor inputs (outside the feedback loop), multiple-op-amp loops, sample-holds with comparators (input is compared with the S/H output, and the comparator operates the S/H control logic, often in synchronism with a clock to avoid oscillation), and A/D converters.

Converters used for peak-following are typically the counter-DAC-comparator type. A D/A converter continuously provides an output voltage proportional to the state of a digital counter. The converter output is compared with the signal input. If the input is the lesser, the comparator continuously inhibits the count. If the input is the greater, the counter accumulates clock pulses until the comparator threshold is crossed, and the count is again inhibited with the next clock pulse. Though slow, the A/D converter types have the advantage of essentially "infinite" *hold* times, since retention of data does not depend on the charge stored in a capacitor. The A/D converter is an ideal second stage of a two-stage peak follower (note that the DAC output, corresponding to the digital count, is an analog quantity). It is also obvious that if the reduced data must be converted to digital form, this is an ideal way to "kill two birds with one stone", if peak information is a suitable measure of the input.

POWER MEASUREMENT (Figure 17)

Analog multipliers are well-suited to the accurate measurement of *instantaneous* power ($e \cdot i$). Their outputs can be averaged to obtain *average* power, applied to peak-detectors to obtain *peak* power, and integrated to obtain *energy*. Furthermore, the energy output can be computed and divided by the period, in the manner indicated in Figure 12, to obtain *energy per cycle*.

When power is measured, voltage can be picked off, differentially,

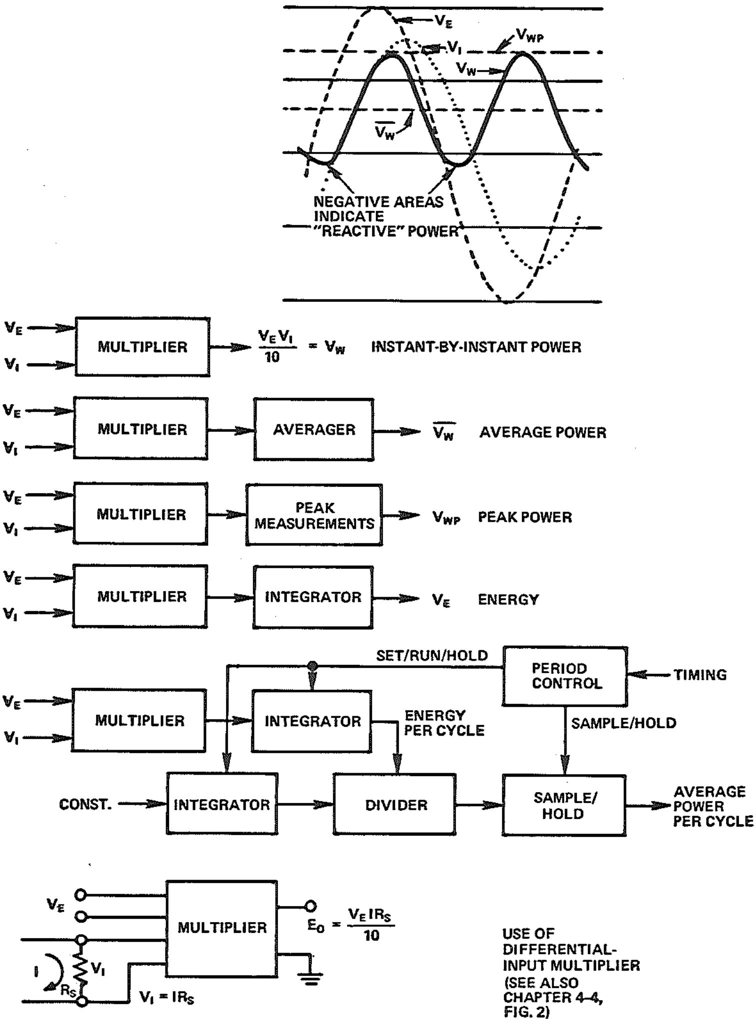


Figure 17. Power and energy measurement

if necessary, by a differential or isolation amplifier, and scaled to the multiplier input. Current can be measured by a differential pickoff across a shunt. If the passively-scaled voltage and current happen to fall within the common-mode and amplitude limitations of available analog multipliers, it is worthwhile to consider the use of *differential-input* multipliers, such as the low-cost monolithic AD532.

Commercially-available complete multipliers are available with bandwidths as high as 10MHz. For the measurement of average power, it is well to consider the input and output response of the multiplier separately. At high frequencies, the phase relationships of instantaneous power start to deteriorate significantly at frequencies as low as 1/50 of the "−3dB frequency," principally because of lags in the output stages. However, the *average* power, which depends critically on the *input* phase relationships, can be computed accurately at frequencies up to 1/10 of the −3dB frequency in transconductance multipliers. (The output undergoes averaging in any case.)

"Power factor," the ratio of average power to average volt-amperes, equal to the cosine of the phase angle (for sinusoids), can be determined by fairly simple analog circuitry. Figure 18 shows a scheme for performing such measurements. By phase-shifting one of the inputs by 90°, the *sine* of the phase angle may be computed; for small angles, it is approximately equal to the angle. For larger angles ($< \pi/2$), an arc-sine function fitter may be used, if a direct

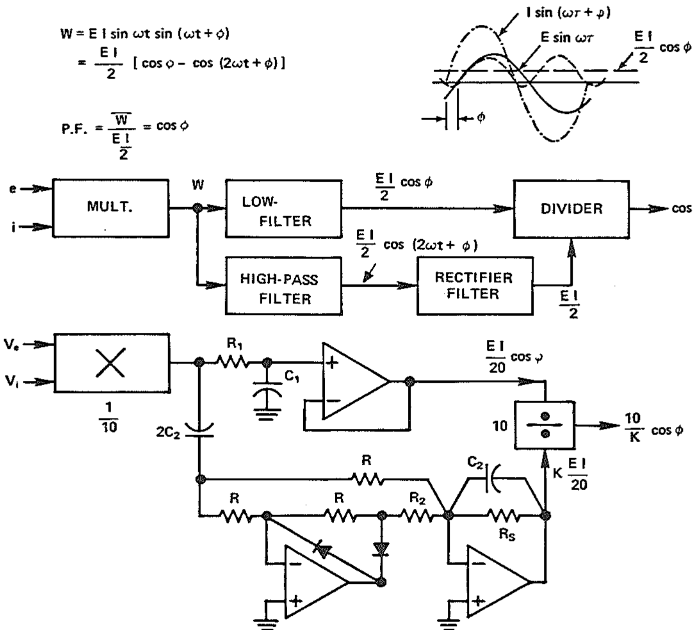


Figure 18. Power-factor measurement

measurement of the angle is required.

Impedance (magnitude) may be measured by computing the scaled ratio of the average or rms voltage to the average or rms current, using a divider (Figure 19). It is important not to fall into the trap of seeking to take the ratio of two ac quantities by an instantaneous measurement. Conceptually, the measurement will not be finite for zero denominator unless the two signals are $n\pi$ apart in phase ($n = 0$ or any integer). As a practical matter, analog dividers call for unipolar denominators; with the added complication of polarity-switching, bipolar denominators may be handled, but the vicinity of zero is ordinarily forbidden.

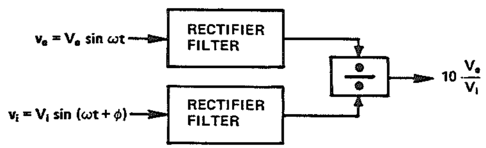


Figure 19. Impedance-magnitude measurement

VECTOR SUM

The vector sum of any number of mutually orthogonal voltages may be obtained by a circuit that solves the equation

$$E_o = \sqrt{V_1{}^2 + V_2{}^2 + V_3{}^2 + \ldots V_n{}^2} \qquad (12)$$

As noted in the introductory chapter, and confirmed elsewhere,[4] the straightforward approach (squaring each input, summing, then taking the square-root of the sum) can be expensive, and may lead to poor results over wide dynamic ranges of E_o, because of the expansion of dynamic range inherent in the squaring operation.

An implicit approach, using a ZY/X device, such as the 433, solving the equation

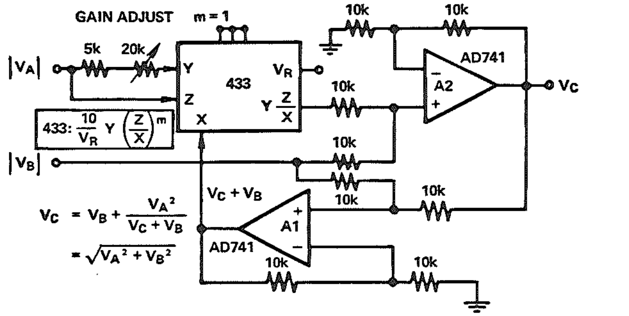
$$E_o = V_2 + \frac{V_1{}^2}{E_o + V_2} \qquad (13)$$

[4]See *Analog Dialogue,* Vol. 6, No. 3, page 3.

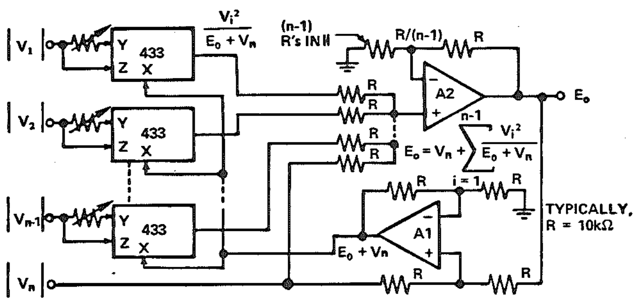
for two variables, and, in general

$$E_o = V_n + \frac{V_1{}^2}{E_o + V_n} + \frac{V_2{}^2}{E_o + V_n} + \frac{V_3{}^2}{E_o + V_n} + \ldots \qquad (14)$$

is far more satisfactory, because each nonlinear term, $V_i^2/(E_o + V_n)$, is net first order, with no external manifestation of square-law dynamic range (Figure 20).



a. Preferred circuit to compute $\sqrt{V_A{}^2 + V_B{}^2}$



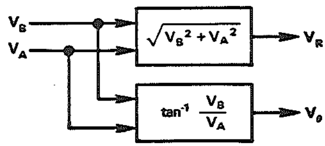
b. Extension of the technique to n input signals

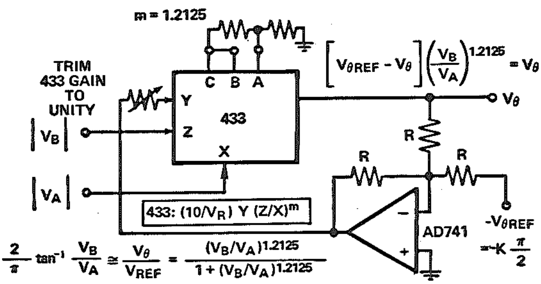
Figure 20. Square-root of sum of squares

Performance, adjustment, and choice of components are straightforward. An important consideration, not immediately obvious, is the need to scale down inputs and outputs to avoid overdriving amplifier A1 or the 433's. If all inputs can have the same maximum value simultaneously, for the maximum output of A1 to be less than an arbitrary level, E_{max}, the maximum input value, $V_i = E_{max}/(1 + \sqrt{n})$. For E_{max} = 10V, the corresponding values of n and V_{max} are:

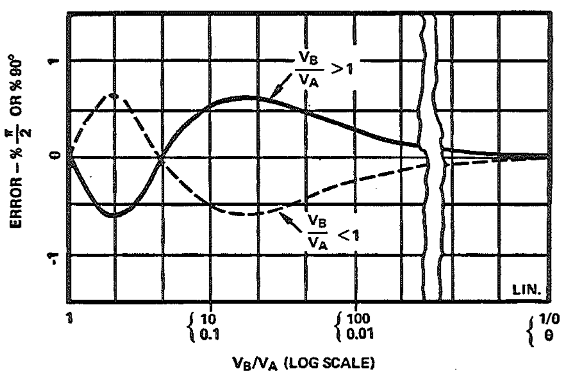
n	V_{max}
2	4.14V
3	3.66V
4	3.33V
5	3.09V

If n = 2, and A1 can swing to 12.1V, V_{max} is 5V.





Basic Circuit for Approximating the Arctangent of a Ratio.



Theoretical Errors of Arctan Approximation

Figure 21. Arctangent circuit, with error plot

Magnitude is one aspect of vector composition, but not the whole story. In addition to magnitude, phase-angle is often desired. If the phase angle, θ, is equal to the arctangent of the ratio V_B/V_A, it can be approximated by function fitting. An excellent first-quadrant fit ($V_B, V_A \geqslant 0$) can be obtained simply, to within 0.75% (theoretically), using a single 433 and an operational amplifier, in an implicit feedback circuit. It maintains its accuracy over an extremely wide range of ratios, because the ratio never appears explicitly – only as a difference of logarithms within the 433. The circuit of Figure 21, which embodies the approximation, solves the normalized equation:

$$\theta = \left[\frac{\pi}{2} - \theta\right] \left[\frac{V_B}{V_A}\right]^{1.2125} = \frac{\pi}{2} \cdot \frac{\left[\frac{V_B}{V_A}\right]^{1.2125}}{1 + \left[\frac{V_B}{V_A}\right]^{1.2125}} \cong \tan^{-1} \frac{V_B}{V_A} \qquad (15)$$

with a maximum theoretical error less than 0.75% $\frac{\pi}{2}$ (or 0.68°). If V_B is negative (IVth quadrant), its absolute value is applied as the input to the $\tan^{-1}$ circuit. Its polarity, determined by a comparator, operates a sign/magnitude circuit, to furnish the proper polarity of the angle. With suitable logic (to add or subtract $\pi/2$), ranges of angle up to $\pm\pi$ can be made available.

RATIO AND LOG RATIO, dB (Figure 22)

Dividers can be used for direct readout of such ratios as efficiencies, losses or gains, % distortion, impedance magnitudes, elasticity (stress/strain). Ratios may be taken of instantaneous, average, rms, or peak quantities. Furthermore, in conjunction with sample/hold devices, ratios may be taken of any of these measurements at different instants of time.

Ratiometric measurements are by no means new, but the low (and still-decreasing) cost of analog dividers (and of variable-reference A/D converters) should serve as an encouragement to designers to consider employment of the technique as a realistic alternative (or adjunct) to tightly-regulated reference supplies for measurements, ultra-stable light sources, etc.

To eliminate the effects of a common parameter, whether physical or electrical, many measurements can profitably involve the use of ratio techniques. For example, in bridge measurements, variations of the power supply directly affect the scale factor. But if the output is divided by the bridge-supply voltage, the scale-factor stability depends only on the stability of the divider. This scheme can be combined with linearization, as shown in Figure 3b. Naturally, the divider should be at least as stable as the bridge-reference voltage if the ratiometric compensation is to be useful.

Compensation for reference-voltage variations is an example of reducing the effects of a common *electrical* parameter. However, ratios can also be used to eliminate the effects of a common *physical* parameter. For example, in light-transmission measurements, it is common to compensate for variations in light intensity by transmitting two beams, one through a reference medium, the other through the medium being measured, and to take the ratio of the two measurements.

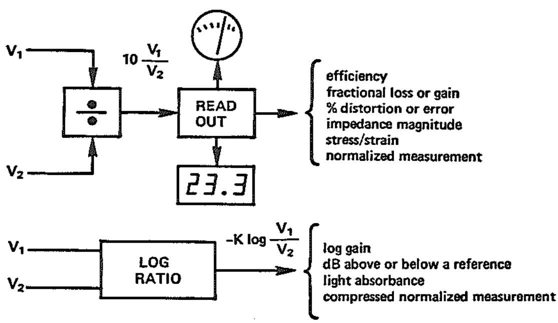


Figure 22. Ratio and log-ratio measurements

Often, logarithmic ratios are more useful than linear ratios. There are two broad categories of such measurements. The first is the measurement of phenomena covering a wide dynamic range, with reference to a normalized level, with log-compression, and either the display of the results on a limited-range meter scale, or the transmission of the measurement through a noisy medium. The second category consists of those measurements that are performed linearly but are normally characterized (or thought about) in terms of logarithmic ratios. One example is light transmission measurements. For another, electrical gain or attenuation may be measured as a ratio of output to input; by the use of a log ratio device, such as Model 756, a direct measurement of the log ratio, or "dB,"* may be performed.

CONCLUSION

In this chapter, we have suggested a number of uses of analog nonlinearities in the reduction of data for display or transmission. It is not unlikely that the thoughtful reader's experience and needs will suggest many more.

*The decibel, one-tenth of a bel (B), is the logarithm of an electrical power ratio of 1.259. The number of dB corresponding to a power ratio is $10 \log_{10} (P_2/P_1)$. If resistance is constant, the number of dB also is equal to $20 \log_{10}(V_2/V_1)$ or $20 \log_{10}(I_2/I_1)$, since power is proportional to the square of voltage or current. The term has been widely corrupted to express log ratios of any two quantities (even engineers' salaries), by the definition $dB = 20 \log_{10}(Q_2/Q_1)$. Though confusing (some would say deplorable), it is almost universally understood.

II Communications & Signal Processing

Chapter 4

Nonlinear devices have always been used in audio signal communications to stabilize or modulate oscillator amplitude and frequency, achieve automatic gain control, and demodulate the received signal. Classically, diode, transistor, and thermionic-device characteristics have been used. Effectiveness and stability of such operations have generally depended on the designer's skill and ingenuity in circuit design and on the availability of components having suitable stability, parameter match, "linearity" (i.e., *parametric conformance*), and low cost.

Now with the availability of operational amplifiers, multiplier-dividers, and logarithmic elements, with their tightly-specified (guaranteed) parameters, convenient (modular or black-box) packages, and low (or decreasing) cost, the designer has a set of new options to make his job easier and more fruitful. In addition to standard signal-processing circuits, he can now consider new approaches to waveform synthesis and control, and the design and uses of such tools as voltage-controlled amplifiers, filters, and oscillators (VCA, VCF, VCO) with uniform behavior and predictable characteristics. Combining these operations with some of the "hybrid" techniques described in Chapters I-4 and I-5B of the *Analog-Digital Conversion Handbook*, one becomes aware of a formidable arsenal of signal-manipulating possibilities, virtually at one's fingertips.

Just a few additional examples of the applications of nonlinear analog devices in signal-handling include compression and expansion, phase-sensitive detectors, phasemeters, phase-locked loops, low-noise recording systems, correlators, spectrum analyzers, speech and music synthesizers, and so on.

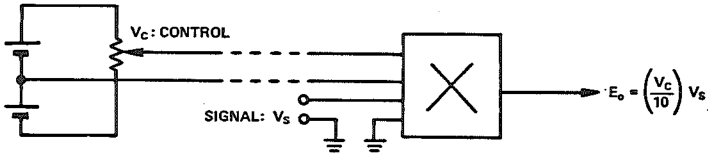
AUTOMATIC GAIN CONTROL

An analog multiplier or divider is inherently a gain controller (Figure 1a) since the signal applied to one of its inputs can be considered a dependent variable, either multiplied or divided by a second input that controls its gain (or attenuation). Since the gain-setting voltage can be derived from any source, there is a wide range of possible applications. For example, a DC voltage applied from a remote manually-adjustable source can cause the multiplier to act as a potentiometer with a "long shaft." The control voltage can be derived as a measure of one or more other voltages in a system and used to control the gain in response to their variation. A useful special case is *automatic gain control* (Figure 1b).

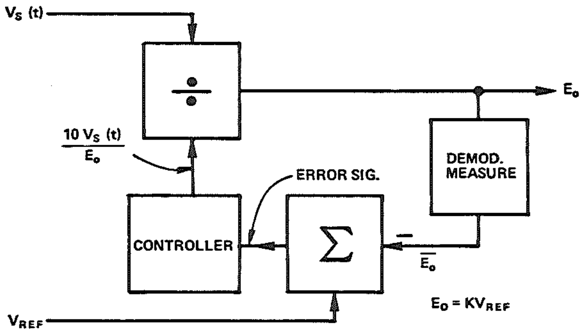
The circuit of Figure 1c is a practical example illustrating the application of the low-cost AD531 I.C. multiplier-divider in an AGC application. It maintains 3V peak-to-peak output for inputs ranging from 0.1Vp-p to more than 12Vp-p, with better than 2% regulation from 0.4Vp-p to 6Vp-p, and distortion well below 1%. Input frequency can range from 30Hz to 400kHz (−3dB). The set point is adjustable either manually or by an external DC reference voltage. The input signal can be either single-ended or differential.

The feedback circuit works in a straightforward manner: if the input signal increases, the output will tend to increase. Its negative peaks, as recognized by the diode and stored on the 1μF capacitor, tend to increase, causing the output of the inverting integrator to increase. This, in turn, causes the denominator to increase, reducing the gain of the AD531 multiplier-divider (an XY/I device), and tending to keep the output level constant.

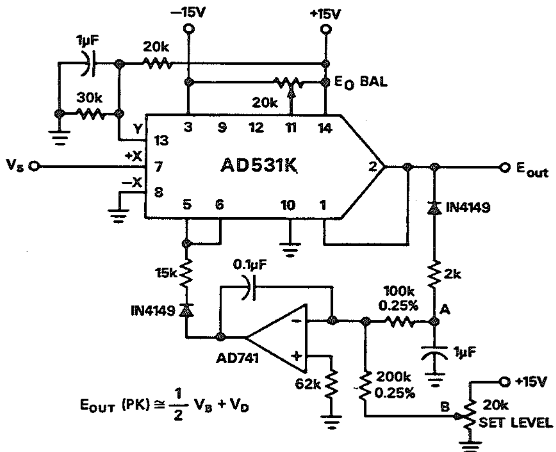
In the steady state, the average voltage at point A must be ideally equal to $\frac{1}{2}V_B$, but of opposite polarity, making the net input to the integrator equal to zero, and holding the output of the integrator at whatever constant level is necessary to keep the loop in balance. In that state, the negative peak value of E_{out} is approximately one diode-drop below V_A, so $|E_{out}\text{ (peak)}| \cong \frac{1}{2}V_B$ + diode drop



a. Analog multiplier as gain controller



b. Typical AGC loop using divider



c. Practical AGC loop using IC multiplier-divider

Figure 1. Gain control with multipliers and dividers

In practice, the *set level* potentiometer would be adjusted empirically to calibrate the output at the desired level.

In the simple practical example given here, to illustrate the principle, an unembellished half-wave diode-and-capacitor circuit reads the peak level of the waveform. Naturally, other measures of the waveform, such as mean absolute-value or RMS, might be used; in addition, somewhat more-sophisticated temperature-compensated rectification circuitry might be used, depending on the needs of the application.

The control voltage (V_c) at the output of the amplifier ranges from about −2V (lowest AD531 gain) to the amplifier's lower limit, −13.5V (to handle the smallest input signals). Linearity of V_c is not important, since it is a manipulated variable inside the loop.

COMPRESSION AND EXPANSION

In Chapter 2-3, the possibilities of logarithmic compression and expansion in transmitting small voltages safely through a noisy medium were touched upon. Though it operates instant-by-instant, a drawback of the scheme is that the logarithmic gains must be matched to ensure linear overall response. Another approach, that can be applied to quasi-stationary waveforms, is to divide the signal by a voltage that is a measure of some property, such as squared-peak, transmit the modified signal through the medium, then multiply the received signal by its squared peak-value (Figure 2a). Since the control voltage varies more slowly than the signal (essentially DC), it does not affect the signal's shape, only its amplitude. The high gain for small signals and low gain for large signals produces a predictable compression function. At the receiving end, the inverse function is applied, and the output amplitude variation is recovered. Mismatches affect only the overall gain, without introducing distortion.

An example of a typical application of this technique is in high-fidelity tape recording systems. The Burwen Laboratories Model 2000, outlined in Figure 2b,[1] has a 110dB dynamic range when used with a 15ips tape-recorder.

[1] "Design of a Noise Eliminator System," by R.S. Burwen, Audio Engineering Society Preprint No. 838(B-8), October, 1971.

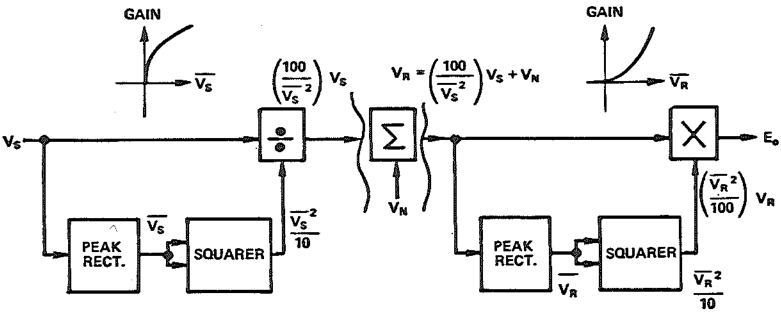
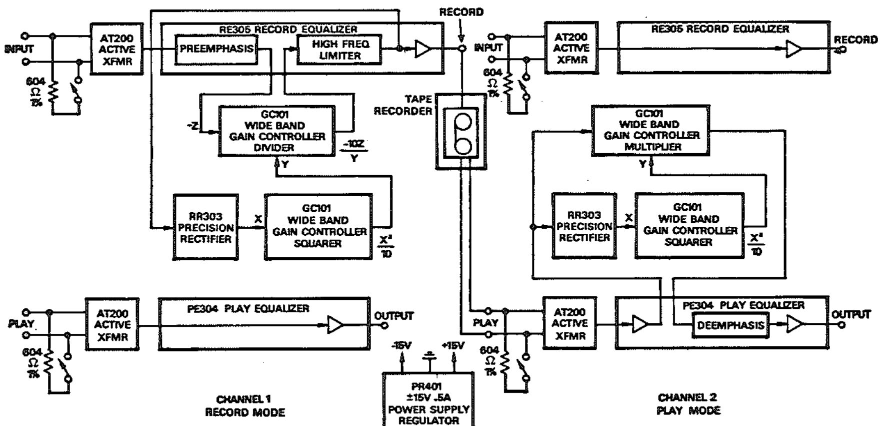


Figure 2a. Gain compression-expansion. Gain function is nonlinear, but signal is transmitted throughout essentially without distortion. Small signals are greatly amplified before transmission. Noise is either suppressed by squaring or masked by high signal levels.



Courtesy of Burwen Laboratories, Inc., Burlington, Mass.

Figure 2b. A commercial wide-range record-playback system having 110dB dynamic range.

SIGNAL GENERATION

A number of schemes for signal generation are discussed in Chapter 2-2, including a variable-frequency two-phase oscillator. As noted there, nonlinear elements can be used to control frequency, phase, amplitude, etc. As a further example, Figure 3 is a schematic diagram of a very low distortion (0.01%) fixed-frequency (1kHz) single-phase phase-shift sine-wave oscillator. Its amplitude (about 7Vrms) is controlled by an AGC loop that applies linear damping in greater or lesser degree without affecting the waveform.

Amplifier A1 is connected as a non-inverting amplifier with a gain of +3. The band-pass filter R1, C1, R5, C2, tuned to 1kHz, provides frequency-selective positive feedback, causing the circuit to oscillate at $f_o = (2\pi RC)^{-1}$.

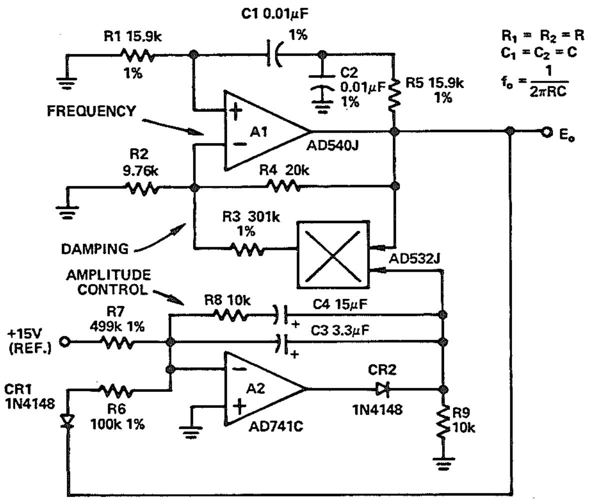


Figure 3. Low-distortion oscillator

The output amplitude is measured via diode CR1 and compared with a reference current through R7. The error is accumulated by the integrator (A2) and, applied to one of the multiplier inputs, increases or decreases the negative feedback around A1, appropriately affecting its gain and the damping of the oscillator. In the steady state, the net input to the integrator is zero, its output is constant, and R4 is in effect paralleled by a large trim resistance of exactly the right magnitude to keep the oscillation stable at a constant amplitude.

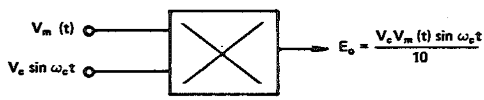
Since the multiplier output is essentially linear and is attenuated to provide a "vernier" gain adjustment on the oscillator amplifier, its distortion has a negligible effect on the output. The distortion is affected primarily by the nonlinearity of operational amplifier

A1 at the frequency of oscillation. The AD540J FET-input op amp provides distortion in the neighborhood of 0.01%. If distortion of 0.04% is tolerable, an AD741C may be used.

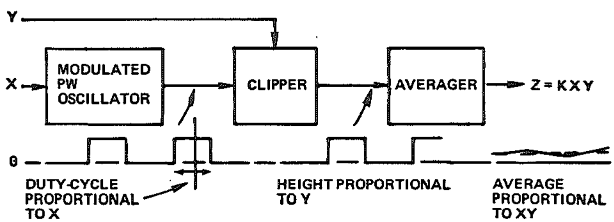
Capacitors C1 and C2 may be changed to obtain other frequencies of oscillation. The amplitude reference (the +15V supply in Figure 3) can be provided by a zener reference diode (for a 9V diode, reduce R7 to 301kΩ).

MODULATION

The terms "multiplier" and "modulator" are closely related. The modulation process almost invariably either uses or creates a multiplication operation. To illustrate this, Figure 4a shows that the "balanced modulator" is simply an analog multiplier; Figure 4b shows the block diagram of a "pulse-height-pulse-width" multiplier—one variable modulates the amplitude, the other modulates the duty cycle, and the area (measured by an averager) is proportional to the product of the two inputs. Historically, modulation was used in the design of multipliers far more frequently than multipliers were used for modulation. But now, with the coming of low-cost IC transconductance multipliers, the pendulum is swinging the other way. Analog multipliers are considered for a variety of modulation applications, from amplitude modulators (Figure 5) to frequency-modulated triangular, square, and sine waves (Chapter 2-2, Figures 3, 8, and 9).



a. Multiplier as balanced modulator



b. Pulse-height, pulse-width-modulation multiplier, first quadrant

Figure 4. Modulation and multiplication

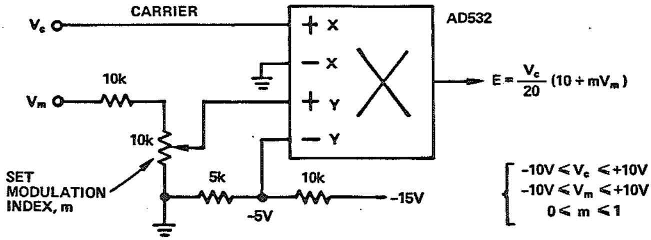


Figure 5. Multiplier as amplitude modulator

Voltage-symmetrical (but not necessarily time-symmetrical) triangular waves may be used to produce duty-cycle-modulated square pulse trains by biasing the triangular waves with the modulating waveform and detecting zero crossings with a precision comparator (Figure 6).

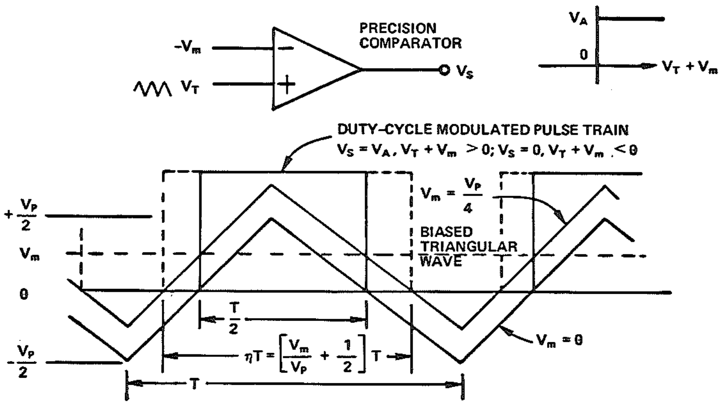
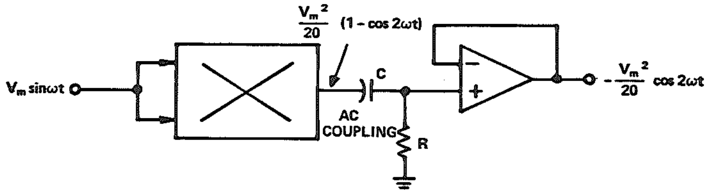


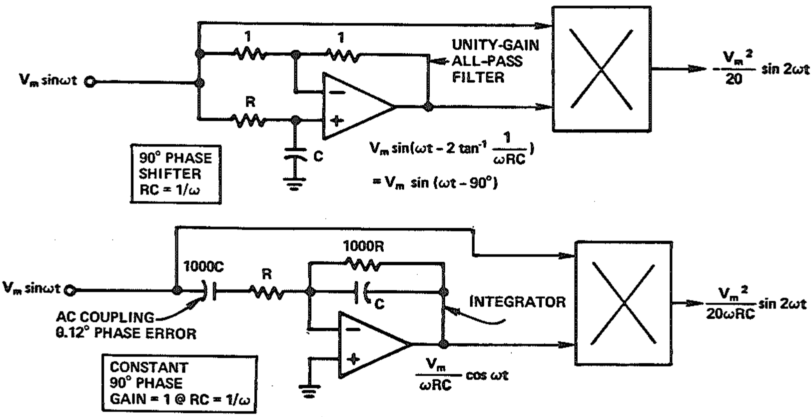
Figure 6. Duty-cycle-modulated triangular wave

FREQUENCY DOUBLING AND n-TUPLING

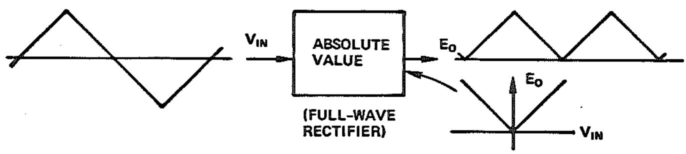
A multiplier, connected as a squarer, can be used to obtain low-distortion sine waves of twice the frequency of an input sine wave. The DC component of the output can be removed with a high-pass filter (Figure 7a). Alternatively, one of the inputs can be phase-shifted by 90°, using either an integrator or an all-pass filter (Figure 7b). This alternative has the advantage that amplitude variations do not result in large transient "bounces" at the output; however, its performance is somewhat sensitive to frequency,



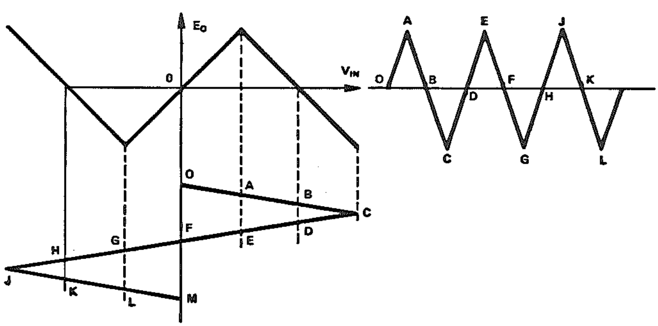
a. Multiplier as frequency doubler



b. Multiplier as "low-bounce" frequency doubler – two approaches



c. Absolute-value as triangular-wave frequency doubler



d. Triangular-wave frequency tripler using 3 piecewise-linear segments

Figure 7. Frequency multiplication (see chapters 2-1, 2-3, 3-5)

whereas that of Figure 7a is wideband for frequencies well above the filter's crossover. Typically, phase error of the double-frequency signal becomes significant at 1/100 of the multiplier's −3dB frequency, and the envelope amplitude loses accuracy above 1/10 of the −3dB frequency.

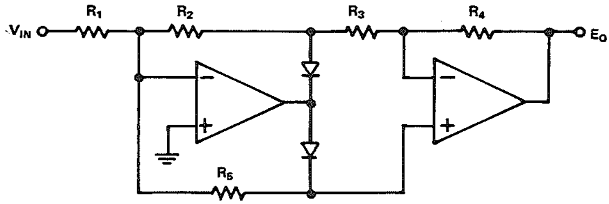
Frequencies of triangular waves can be doubled by the use of an absolute-value circuit (Figure 7c). If amplitude is constant, the dc level can be biased out. Otherwise, ac coupling can be used, with a cutoff frequency well below the fundamental (phase shift does not affect the shape of a sine wave, but it does distort triangular waves).

Square-wave and triangular waves can be tripled in frequency, or in general multiplied by any whole number, using a piecewise-linear voltage sawtooth operator (Figure 7d). Factors much larger than 3 tend to become impractical because of sensitivity to break-point drift and incremental gain settings. It is worth noting that the output of a tripled triangular wave can be shaped into sinusoidal form, if desired, using a function fitter (Chapter 2-1).

DEMODULATION

We have touched on peak, average, and RMS measurements in Chapter 2-3. Similar techniques are used for demodulating amplitude-modulated signals. Figure 8 shows two "ideal diode" high-accuracy full-wave rectifier circuits that are perhaps less well-known than Figure 13 of Chapter 2-3. The circuit of Figure 8a uses 5 equal resistors to obtain unity gain and has only a single path from the input source. The circuit of Figure 8b has high input impedance, an especially useful feature if the signal source must be unloaded or if ac coupling with long time constants is necessary. A possible disadvantage in both cases is that the output-averaging filtering must be performed in a separate stage. These circuits may be followed by a peak-reading circuit, if desired.

If the waveform contains polarity information, *synchronous-detection* may be useful. In the scheme shown in Figure 9, a square-wave reference signal multiplies the alternate half-cycles by positive and negative constant voltages. If signal and reference are in

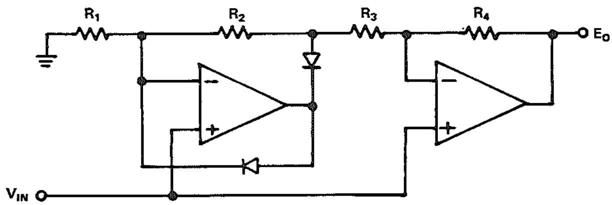


GAIN CONSTRAINT: $\frac{E_O}{V_{IN}} = \frac{R_4}{R_3}\,\frac{R_2}{R_1}$, $V_{IN} \geq 0$

SYMMETRY CONSTRAINT: $\frac{E_O}{V_{IN}} = \frac{R_5 \parallel (R_2 + R_3)}{R_1}\left(1 + \frac{R_4}{R_2 + R_3}\right)$, $V_{IN} \leq 0$

$\begin{cases} \text{FOR } R_1 = R_2 = R_3 = R_4 = R_5,\ \text{GAIN} = 1 \\ \text{FOR } R_1 = R_2 = R_3 = R,\ R_4 = R_5 = 2R,\ \text{GAIN} = 2 \end{cases}$

a. Full-wave rectifier circuit



$$\begin{cases} \frac{E_O}{V_{IN}} = 1, \quad V_{IN} \geq 0 \\ \frac{E_O}{V_{IN}} = 1 + \frac{R_4}{R_3} - \frac{R_4}{R_3}\left(1 + \frac{R_2}{R_1}\right) = 1 - \frac{R_4}{R_3}\,\frac{R_2}{R_1}, \quad V_{IN} \leq 0 \end{cases}$$

CONSTRAINT: $\frac{R_4}{R_3}\,\frac{R_2}{R_1} = 2$, e.g., $R_1 = R_2 = R_3 = R$, $R_4 = 2R$

b. High-input-impedance full-wave rectifier circuit

Figure 8. Absolute-value (full-wave rectifier) circuits

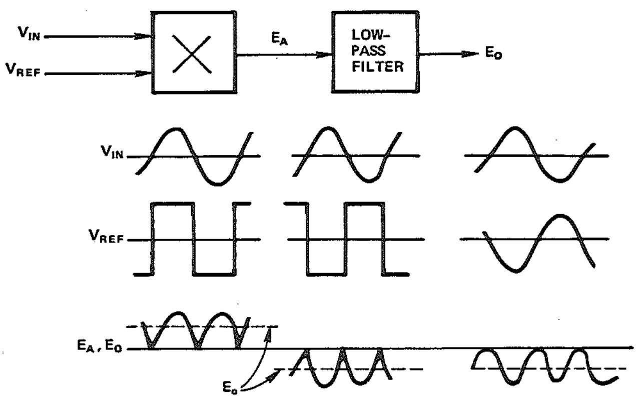


Figure 9. Synchronous (phase-sensitive) detection

phase, the full-wave-rectified output is positive; if they are in opposite phase, the output is negative. If the signal and reference are sinusoidal, the average value of the output will be equal to $(V_{rm} V_{sm}/20) \cos\theta$, where θ is the phase angle and V_{rm} and V_{sm} are the reference and signal amplitudes.[2] Small phase shifts do not greatly affect detection accuracy; for example, 0.8° gives 0.01% error, 2.56° gives 0.1%, 8° gives 1%, and 18° gives 5%. If the signal and reference are 180° out of phase, the average output will be negative, with the same ideal phase tolerances.

If, on the other hand, it is desired to measure small *phase* deviations, one of the inputs can be shifted 90°; the average output will then be proportional to the sine of the phase angle. The following brief table outlines the theoretical error inherent in the assumption that $\sin\theta = \theta$.

Angle		Sine	Fractional Error (% 1rad)
θ_{rad}	θ°	$\sin\theta$	$\lvert\sin\theta - \theta_{rad}\rvert$
0.084	4.813	0.0839	<0.01%
0.180	10.31	0.1790	<0.1%
0.390	22.35	0.380	<1.%
0.490	28.1	0.471	<2.%
0.670	38.4	0.621	<5.%

Function-fitting techniques can be used to reduce the error if the range of angle is too large for the desired accuracy.

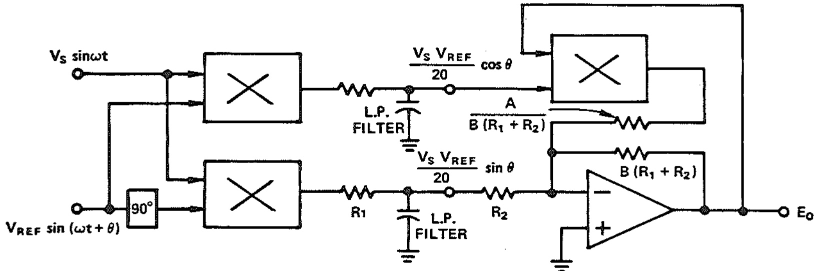
Greatly-improved linearity can be obtained by combining sine and cosine demodulation with an implicit feedback loop to obtain "tan-lock" demodulation[3]. A tan-lock demodulator solves the equation

$$-E_O = \frac{B \sin\theta}{1 + A\cos\theta} = B \sin\theta + AE_o \cos\theta \cong K\theta \qquad (1)$$

[2]See Figure 18, Chapter 2-3.

[3]"Use this Tan-Lock Demodulator," by R.P. Hennick, *Electronic Design* No. 25, December 6, 1970, pp74-75.

as shown in Figure 10. In addition to the improved linearity over a wide range of angle, as shown in the error plot, with its attendant reduction of distortion, one might expect to realize improvements in noise threshold, hold-in range, and pull-out frequency (see discussion below).



$$-E_O = A\,\frac{V_S V_{REF} E_O}{200}\cos\theta + B\,\frac{V_S V_{REF}}{20}\sin\theta$$

$$-\frac{E_O}{10} = \frac{\frac{B}{2}\,\frac{V_S}{10}\,\frac{V_{REF}}{10}\sin\theta}{1+\frac{A}{2}\,\frac{V_S}{10}\,\frac{V_{REF}}{10}\cos\theta} \simeq \frac{\theta^\circ}{100^\circ} = \frac{9}{5}\,\frac{\theta_R}{\pi}$$

IF $V_S = V_{REF} = 10$

$$-\frac{E_O}{10} = \frac{\frac{B}{2}\sin\theta}{1+\frac{A}{2}\cos\theta} \simeq \frac{9}{5}\,\frac{\theta}{\pi}\;\begin{cases} B = 1.8 \\ A = 1.1716 \end{cases}$$

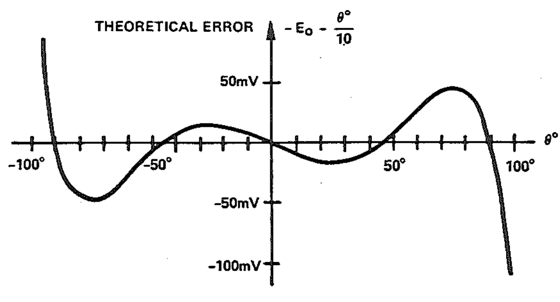


Figure 10. "Tan-lock" demodulator circuit, scaling, and theoretical error

Phase demodulators are often preceded by AGC or limiting circuits to ensure constant ac input amplitude and avoid amplitude modulation of the output. Wideband multipliers, such as the 429, have less than 1° of differential phase shift at 1MHz. The output double-frequency phase shift of about 24° @1MHz is unimportant, since only the dc component of the output is used; the dc level depends critically only on the input frequency characteristics.

PHASE-LOCKED LOOPS

A phase detector may be used as the "summing point" of a feedback loop that generates a frequency that is compared with the average input frequency and "locked in" to that frequency with a fixed (e.g., 90°) phase relationship (in the steady state, ideally, for sine waves, cos θ = 0). Phase error is usually in the form of a dc voltage that drives the local frequency generator (a voltage-controlled oscillator) through a high-gain amplifier. Thus, the loop, if stable, seeks to maintain the phase error at zero (Figure 11).

To anyone familiar with the principles of feedback (and today, that includes anyone who uses op amps creatively and successfully), the phase-locked loop would appear analogous to an operational amplifier, except that phase is the input variable and frequency (rate-of-change of phase) is fed back. The "loop gain" of a phase-locked loop is expressed in terms of %Δf/radian.

The basic elements of a phase-locked loop, as mentioned above, are the phase detector, a filter-amplifier (to remove ac components from the dc voltage that represents the phase error and to amplify the error signal), and a voltage-controlled oscillator (VCO).

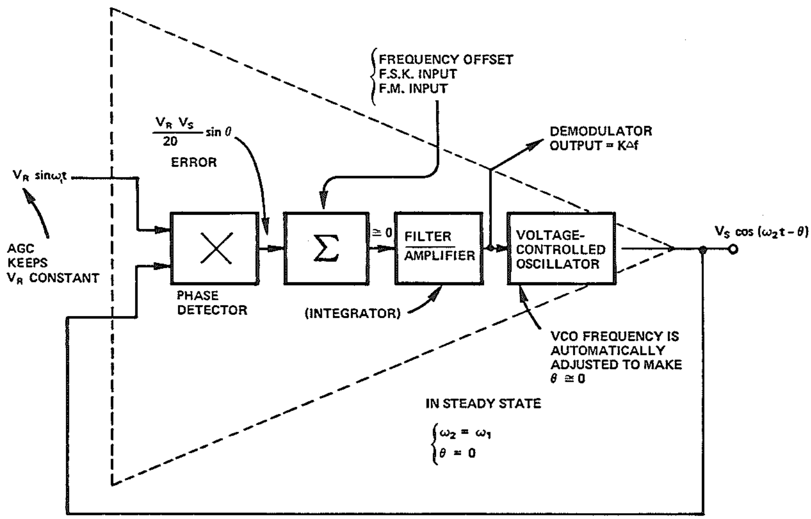


Figure 11. Phase-locked loop, or phase follower, with response to sinusoidal signals.

Applications of phase-locked loops are in two basic classes: frequency re-creation and multiplication, and narrow-band filtering, based on the ability to respond to an input frequency; and frequency modulation and demodulation, based on the ability of a well-designed phase detector and VCO to respond accurately, stably, and linearly to a dc voltage.

In the first class, a received signal may be noisy, distorted, and actually jittering in frequency. The job of the phase-locked loop is to generate a clean waveform of appropriate shape that is locked-in to the average signal frequency. The filter prevents the output frequency from responding to rapid fluctuations of phase; it tends to null out the average phase error. The VCO need not be very linear (the same may be said for the phase detector), but the range of frequencies, or phase, within which the loop is captured (i.e., under control) must be sufficiently wide to embrace the expected fluctuations, either at the signal frequency or a harmonic (if the job of the loop is frequency multiplication).

In the second class, the VCO and/or the phase detector should have a linear relationship over a dynamic range corresponding to the maximum deviation present in the modulated signal. The filter should be slow enough to filter out carrier, but fast enough to follow the modulation. If the loop is acting as a demodulator for a frequency-modulated signal, the output of the VCO will track the modulation, and the "dc" input of the VCO will be the demodulated output voltage. If the loop is acting as a modulator, the modulating signal is added at a voltage summing point after the phase detector. The output frequency will change to the degree necessary to create a phase-error voltage that will continuously balance out the disturbance caused by the modulating signal, while remaining locked-in to the input frequency.

An intermediate class is "frequency-shift keying" (FSK), in which the modulating signal is a step change of voltage (such as a change of binary logic levels), which produces a step-change of frequency. On the receiving end, the loop responds to a step-change of frequency with a step-change of voltage. This form of operation is often referred to as "modem" (modulate-demodulate). A linear relationship is not required, but the phase-voltage-frequency band should be adequate.

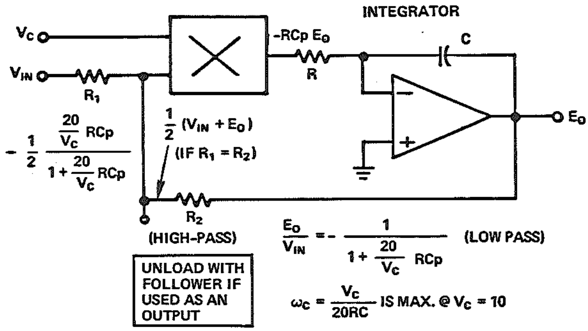
Signals and VCO outputs may have any waveshape, as long as one can ensure that the loop will not lock in on a harmonic (unless so desired). Since a phase-locked loop is an active closed-loop system, it must be designed not to be dynamically unstable (i.e., to run away or oscillate, rather than behaving as desired). Two ranges of frequency are usually important in governing lock-in: *capture* (pull-in) *range*, the band of frequencies within which a lock-in condition can be acquired, and *lock* (dropout) *range* (tracking or holding), the band of frequencies within which lock-in can be maintained, always wider than the capture range.

Some ideas about VCO design may be found in Chapter 2-2 (and in many other places in the literature). The low cost of modular and IC multipliers makes high-performance medium-frequency phase-locked loops quite practical today. For the future, one may expect to find integrated-circuit phase-locked loops that transcend in performance today's elementary IC devices, requiring considerably fewer external components, and more suitable for high-precision applications.

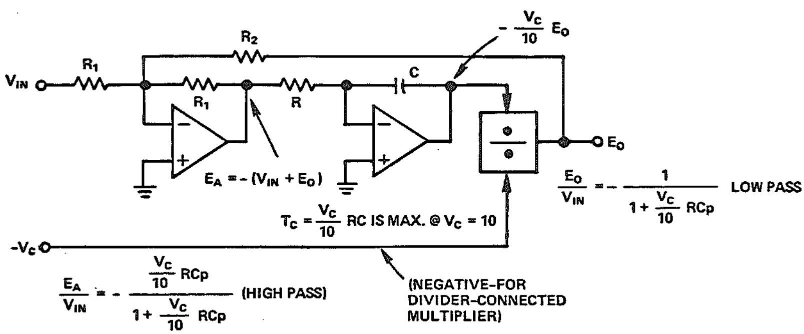
VOLTAGE-CONTROLLED FILTERS

A "state-variable" active filter is one in which an analog-computing feedback loop (or loops), involving one or more integrators, is used to simulate the desired transfer function. Though less compact than the usual active-filter circuit, in terms of the number of op amps required to obtain a transfer function characterized by an *n*-degree polynomial, it has an important advantage: If an integrator is preceded (or followed) by an analog multiplier (or divider), the overall characteristic frequency ω_0 (or characteristic time T_0) will be directly proportional to the multiplying or dividing input voltage. This makes it possible to build filters in which the capacitors are, in effect, adjustable, either directly or inversely, by a control voltage. For a filter that does not involve inductors, is thus possible ideally to manipulate the frequency scale of a filter by means of a single voltage, without affecting any other parameters. The cost is one multiplier per capacitor, formerly impractical, but now quite feasible, because of the low price of multiplier/dividers.

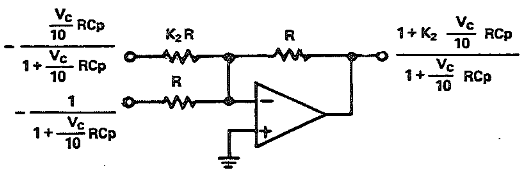
Figure 12 shows how a multiplier (a) or a divider (b) can be used to adjust the "break" frequency or time constant of a first-order lead or lag (for a lead-lag (c), the outputs are summed in an external adder-subtractor with appropriate polarities and coefficients). Note that the multiplier is used ahead of the integrator, with passive summation at its input, while the divider follows the



a. Multiplier adjusts break frequency of unit-lag circuit.



b. Divider adjusts time-constant of unit-lag circuit



c. Adder circuit combines outputs of (b) to form lead-lag response (normalized)

Figure 12. First-order variable filters using multipliers or dividers

integrator. The reason for this can be seen if one considers the consequences of, for example, using the multiplier following the integrator. If the multiplier's output is 10V, then for *any* value of V_c less than 10V, the integrator output must be greater than 10V. Placing the multiplier ahead of the integrator solves the out-of-range problem, because the integrator output, in the closed loop, can never be greater than the input to the circuit (multiplied by R_2/R_1), and the multiplier output can never be greater than 10V because of its inherent $V_1V_2/10$ scaling. In like fashion, out-of-range problems are avoided for this case if the divider *follows* the integrator.

If the divider is a conventional multiplier in a feedback configuration, requiring a negative denominator voltage, the configuration of Figure 12b makes available both the output (low-pass) and its derivative (high-pass). If the divider has *positive* gain, the inverting summing amplifier may be omitted (low-pass only) or replaced by a *non*-inverting summer (high pass and lead-lag).

Multiplier-integrator elements can be combined to form higher-order state-variable filters. For example, Figure 13 shows a second-order filter (note the similarity to the oscillator circuit of Fig. 9, Chapter 2-2). Depending on which output or combination of outputs is used, it can serve as a high-pass, low-pass, band-pass, band-reject, all-pass, etc. Again, it is important to note that, for ideal circuit elements, the control voltage V_c affects only the frequency scale; damping and coefficient weightings, normalized frequency-response characteristics, and normalized time response are all unaffected.

This feature is especially useful when first- and second-order filter responses are cascaded to obtain *n*th order Butterworth, Chebyshev, Bessel, or other response characteristics. Variation of V_c to adjust cutoff frequency does not affect the coefficient weightings, once the relative-frequency-and-damping relationships have been set.

Where digital control is desired, the multiplier blocks could be embodied by multiplying D/A converters, such as the IC AD7520.

Variable time-constant integrators, and the tunable filter networks that they make possible, can be usefully and profitably employed

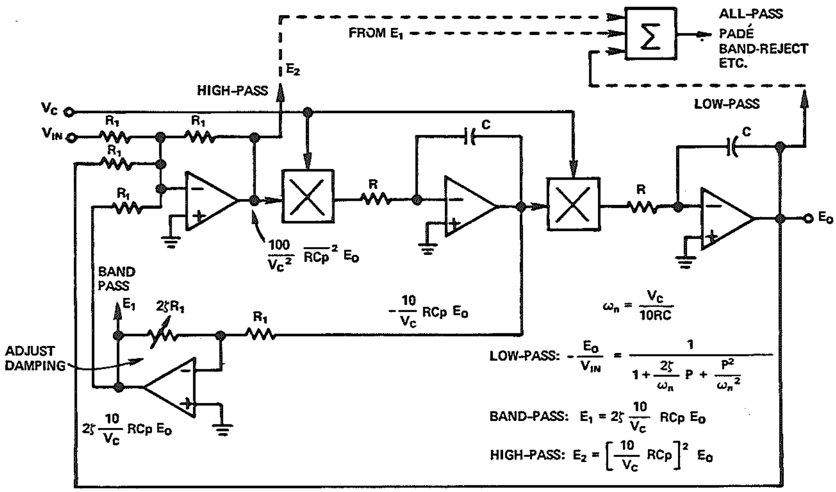
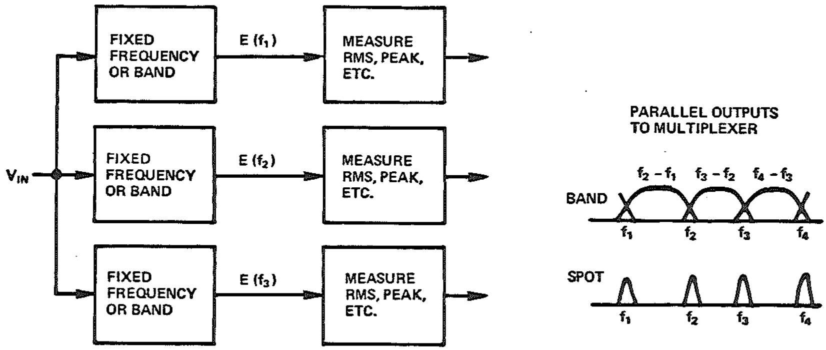


Figure 13. 2nd Order filter with variable natural frequency, using two multipliers. If desired, damping could also be controlled via a multiplier

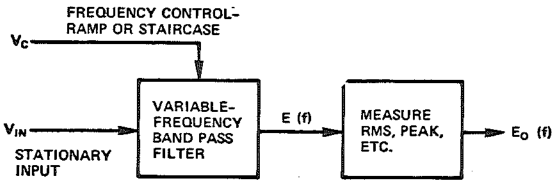
in a number of ways. Examples include adaptive control (adjustment of time constants to achieve an automatically-minimized control-loop error function), spectrum analysis (variable-frequency sweeps to obtain amplitude spectra of stationary waveforms), variable analog delay lines, variable-bandwidth systems for audio "hiss" and "rumble" noise reduction, tunable-carrier transmitters and receivers for narrow-band signals, programmable filters, etc. If the filter is controlling an oscillator frequency, it may be used in FM detection with a phase-locked loop, where the input to the filter is the phase-detector output voltage required for tracking, proportional to the modulating signal.

SPECTRUM ANALYZERS

There are many ways of analyzing and plotting a signal spectrur A few that are relevant to the techniques and devices described in these pages are "spot" measurements (or frequency "combs"), band measurements, and swept measurements. The first two types can be achieved either with fixed-frequency (narrow-band or bandpass) filters in parallel or with a single stepped (narrow-band or bandpass) filter (Figure 14). If many frequencies or bands are to be measured, the serially-stepped-filter approach is considerably



a. Filters in parallel



b. Single variable-frequency filter

Figure 14. Spectrum analyzers

more economical in terms of equipment, but consumes more time for the measurement. The swept filter may provide a continuous "spot" measurement; however, the sweep must be slow enough to not introduce substantial errors as a result of its rate of variation. The dc measure can be obtained in terms of rms, peak, average, or "one-shot-per-step" integral measurement.

MUSIC SYNTHESIZERS

These versatile instruments serve as a tonal palette embracing an extremely wide range of audio waveforms and sounds for the ministrations of the musical composer, performing artist, and special-effects creator. They permit a wide range of pitches, tones, attack-decay-sustain-release sequences, amplitudes, and combinations, both linear, and nonlinear. They tend to use the whole gamut of waveform processing trickery, including voltage-controlled amplifiers, voltage-controlled oscillators, voltage-controlled filters, modu-

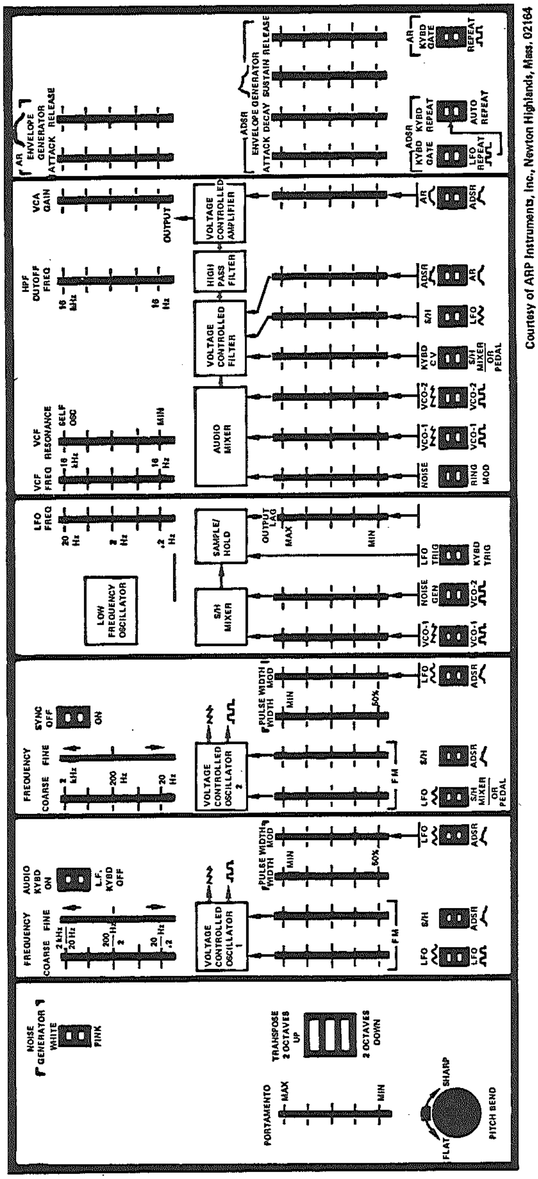


Figure 15. Typical music-synthesizer functions

lators and demodulators, phase-locked loops, sample-holds, noise generators, and pressure-sensitive transducers. Both analog and digital (ROM) functions are used. Figure 15 shows the control panel of a typical commercially-produced moderately-priced keyboard instrument, the ARP Odyssey.

CONCLUSION

This chapter has sought to touch briefly and suggestively on a number of techniques used in audio communications and signal-processing, and on the possible contributions of today's low-cost, compact, comprehensively-specified modular and IC nonlinear devices. It is hoped that the reader will consider omissions and elisions (due to the pressures of space and time) as a challenge to creativity.

II

Computing & Control

Chapter 5

In this final chapter of the Applications section, we discuss a few ways that nonlinear analog computing techniques are used in industry, suggest a few additional ones, review further applications of ideas suggested earlier, and, in effect, present a modest list of topics that, possibly landing in fertile ground, may be fruitful in terms of the ideas that are inspired in thoughtful readers. It is always important to bear in mind that improvements in device performance, reductions in cost and size, and ready availability from multiple sources, have brought many of these ideas from the status of "merely interesting" to feasibility as everyday tools of the designer.

FILTER-FREE THREE-PHASE POWER MEASUREMENT

Figure 1 shows the block diagram of a simple scheme for

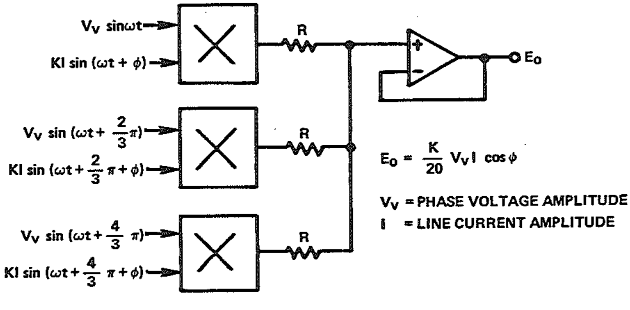


Figure 1. Three-phase average-power measurement without filters

computing the average power in a three-phase system.[1] Voltages proportional to the phase voltages and the corresponding line currents are multiplied individually in three multipliers, and the outputs are summed. The output is ripple-free for a balanced system, with the three average levels averaged and the double-frequency ac components cancelled.

Since low-pass filtering, with its essential delays, is not needed, rapid measurement or detection of the power level is made possible. If one input of each multiplier is phase-shifted by 90°, the output will be a continuous measurement of *reactive* power, $KV_V I \sin\phi$. The 90° phase shift can be obtained by measuring the *line-to-line* voltages, and absorbing the stray $\sqrt{3}$ factor in the analog circuitry.

The improved speed of response makes possible faster-responding, more-stable control loops, and clean, easily-metered monitoring signals. An interesting application is in the excitation control of synchronous motors.

Besides real and reactive power, other useful output signals may be obtained by combining the real and reactive power measurements in various ways. For example, the square-root of the sum of the squares (Figure 20, chapter 2-3) may be used to compute total volt-amperes. The ratio of the power to volt-amperes is the *power factor* ($\cos\phi$), while the ratio of reactive power to volt-amperes is a good approximation to the phase angle for small angles. The ratio of real power to reactive power, $\tan\phi$, a nonlinear function of the power factor, is particularly useful as a control signal because of its high sensitivity.

Finally, it is often useful to control the excitation of a synchronous motor so that it is overexcited at low loads, with reduced excitation as load increases, to avoid exceeding the normal current limitations of the motor at full load. The control criteria for this operation can be established by setting a simple linear combination of the reactive and real power equal to a constant.

[1] "Detection and Measurement of Three-Phase Power, Reactive Power, and Power Factor, with Minimum Time Delay," by I. R. Smith and L. A. Snyder, *Proc. IEEE*, November, 1970, p. 1866.

RATIOMETRIC MEASUREMENTS – LIGHT TRANSMISSION

Figure 2 shows a scheme commonly employed to measure transmittance or absorbance of light by an unknown medium, independently of variations of the light-source level. The light is transmitted through a reference medium (which might be air or vacuum) and through an unknown. Both samples are transduced to current by a pair of matched photosensitive detectors. The 756 log-ratio module converts the output currents (at essentially zero input impedance) to an output voltage proportional to the logarithm of their ratio.

Since the measurement is ratiometric, it is independent of the source intensity. Since it is logarithmic, it can deal accurately with a wide range (4 decades) of unknowns, and furthermore it can be read out directly in logarithmic absorbance or transmittance units.

Always useful, ratiometric measurements (with analog dividers) in general, and log-ratio measurements in particular, are becoming increasingly feasible and accessible for an ever-wider variety of applications as cost decreases and availability and performance increase.

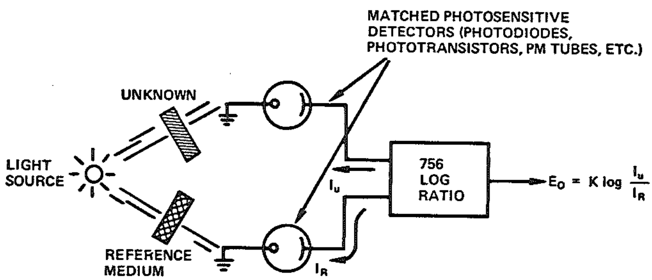


Figure 2. Measuring light transmission independently of light-source variations. Log-ratio gives direct reading of relative transmittance or absorbance

EXPONENTIAL DECAY TIME-CONSTANT

Figure 3 shows a circuit that can be used to rapidly measure, compute, and display continuously the time constant of an exponential decay. For example, a 10-minute time constant can be measured within seconds.

The operating principle is simple: the time-derivative of $\epsilon^{-t/\tau}$ is equal to $-(1/\tau)\epsilon^{-t/\tau}$. Therefore, if we divide the argument by the negative of its time-derivative, the result is the time constant τ, available immediately after the startup transient has died away.

In order that the differentiator be stable and not have excessive noise at high frequencies, it has a second-order rolloff (R_cC and RC_c). Naturally, these time constants should be short compared to the shortest time constants being measured, but they should be no shorter than is necessary on that account.

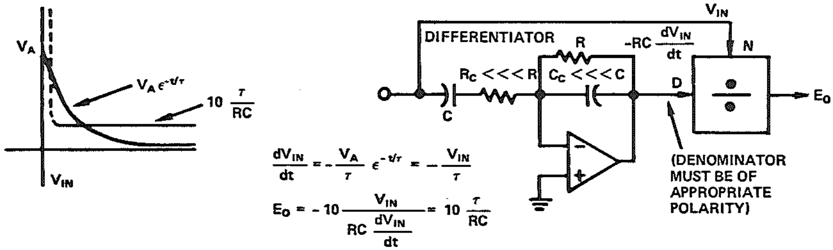


Figure 3. Circuit for determining time constant of exponential decay

The divider should be capable of dealing with signals having a wide dynamic range, if the range of time constants to-be-measured is substantial. It may be a log-ratio device if log τ is acceptable.

Applications include calibration and capacitor measurements. It is especially suitable for obtaining rapid measurements of slowly-varying phenomena, such as battery discharge and capacitor-dielectric "soakage." By recording the measurement continuously, or sampling the waveform from time-to-time, the "quality" of the time constant can be investigated (e.g., is the response truly exponential?). The differentiator must not introduce substantial errors; therefore, though the initial tolerance is not critical, the capacitor should be the highest grade available (polystyrene, teflon, polycarbonate, etc.), and the amplifier should have low leakage current and low noise.

MASS GAS FLOW COMPUTATION

This is an example of the use of low-cost nonlinear analog circuit elements in the conditioning of transducer outputs to obtain an essentially direct measurement of a quantity that depends on a number of variables.

The measurement of gas flow through a resistive element, such as a nozzle, venturi, or an orifice, requires that we know the absolute pressure, the absolute temperature, and the pressure-drop. An equation typically used to relate the gas flow to these variables is*

$$F = K_1 \left(1 - K_2 \frac{\Delta P}{P}\right) \sqrt{\frac{P \Delta P}{T}} \qquad (1)$$

If ΔP is small compared to P, this expression simplifies to the frequently-used

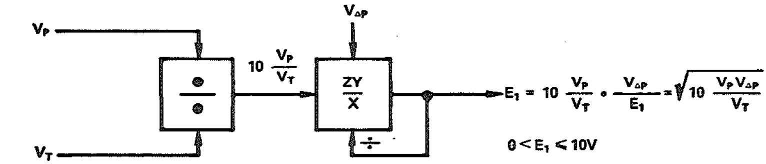
$$F = K \sqrt{\frac{P \Delta P}{T}} \qquad (1a)$$

Figure 4a shows how a divider and a multiplier-divider can be used to compute equation (1a). For a fixed value of K, the electrical inputs are scaled in the preceding preamplifiers so as to utilize the full output range (E_1), and as much of the input ranges as is consistent with the various combinations that produce full output. If the input divider is a multiplier-divider, the third input can be a constant voltage with the effect of adjusting K^2.

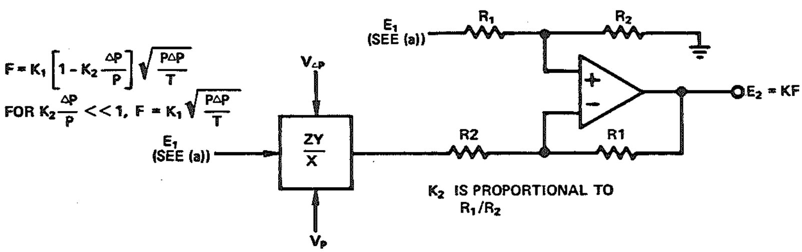
Figure 4c shows how logarithmic circuits can be used to embody equation (1a). If all three variables can vary widely, the logarithmic approach is the more useful, because it allows the scaling to be flexible, without fear of overranging, as long as the output is properly scaled.

If equation (1) is used, it can be embodied by feeding the output of equation (1a) into the circuit of Figure 4b. It amounts to subtracting from E_1 a correction term proportional to $E_1 V_{\Delta P}/V_P$,

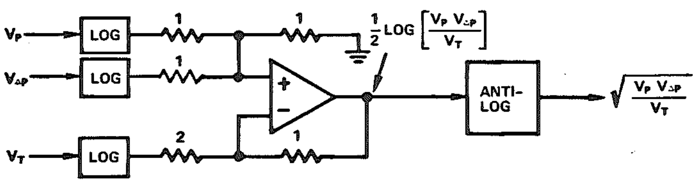
NASA Tech Brief 71-10407, Lewis Research Center, J. Watson, D. Noga, J. Dolce, and J. Gaby, Jr., "Low-Cost Logarithmic Mass Flow Computer" with a coefficient K_2, determined by the resistor ratio R_1/R_2. The output of the flow circuit can be integrated to determine the total mass transferred over a period of time, or it can be averaged by a unit-lag or other averaging filter.



a. *Approximate mass gas flow,* $F = K\sqrt{\frac{P\Delta P}{T}}$



b. *Mass gas flow* $= K_1\left[\sqrt{\frac{P\Delta P}{T}} - K_2\frac{\Delta P}{P}\sqrt{\frac{P\Delta P}{T}}\right]$



c. *Logarithmic circuit*

Figure 4. Mass gas flow configurations

*The subtractor configuration of Figure 4b may be found useful for implementing many of the circuits in this book (and elsewhere) that call for differences of the form (x - Ay). The gain or attentuation, -A, is determined by R_1/R_2, and the gain at the plus input will be unity if the resistor ratio is matched as shown.

OXYGEN CONCENTRATION WITH ANTILOGS

Electrical measurements of ion concentrations are logarithmic. For solutions at 29°C, a relative concentration change of one decade (i.e., X10 or X0.1) produces a 60mV change, of appropriate polarity, at the measuring electrode.

If it is desired to measure the actual (relative) concentration, the output of the detector is preamplified and applied to the input of an antilog circuit. A circuit for performing this job is shown in Figure 5.[2]

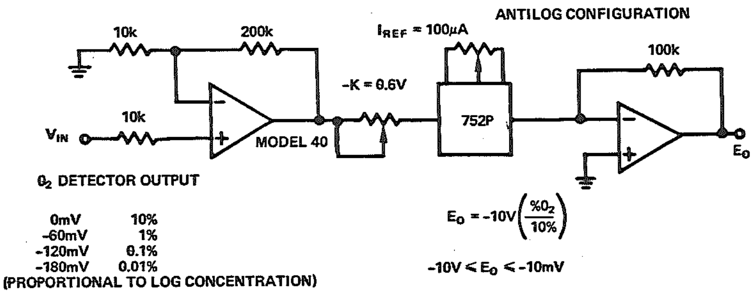


Figure 5. Linearizer for oxygen detector (see Chapter 4–3)

TRANSIENT-FREE RANGING PICOAMMETER

The conventional electrometer circuit using an inverting operational amplifier requires large-value feedback resistors to convert the input current to an output voltage. If the input covers a wide range, either manual or automatic range-switching may be needed, involving several large-value resistors.

There are a number of discomforting factors to consider when designing such circuits. First of all, resistances in the 10kMΩ region, and greater, are difficult to obtain with tight tolerances and good stability vs. time and temperature. Stray capacitance tends to make the response of these circuits quite slow. Range switches tend to have leakage and capacitance; besides the inherent steady-state errors, the settling time after switching can

[2] See also Chapter 4–3.

be of the order of many seconds, certainly inappropriate for autoranging. In addition, there are all the inherent problems of low-level current measurement by *any* means: cable problems, stray capacitance and leakage, and amplifier input-circuit problems.[3,4]

A workable answer to this problem (Figure 6) involves the use of log-antilog circuitry, similar to that employed in the Model 434 multiplier-divider, but with one of the inputs designed specifically for electrometer-level current-handling. The input current is "logged" in the feedback circuit of A1, and referred to an input I_{REF}. The ratio I_{IN}/I_{REF} is multiplied by adding the log of a voltage reference, V_R, and antilogged in the circuitry associated with A4. As equation (2) shows,

$$E_o = \frac{R_2}{R_1} \cdot \frac{I_{IN}}{I_{REF}} V_R \tag{2}$$

the output scale factor may be adjusted in a number of ways, separately or concurrently: by the resistor ratio, R_2/R_1, by a reference voltage, V_R, or by a reference current, I_{REF}. I_{REF}, in turn, may be determined by a stable low-current source.

Since the scale factor is proportional to V_R, the gain may be set directly by a voltage, without the need for switch circuitry in automatic ranging. The saturation current of the log transistors is quite low, typically well below 10^{-13}A at 25°C. Since the saturation currents of the two transistors in each pair are monolithically matched, temperature affects the ratio negligibly. Because the monolithic dual transistors are essentially at the same temperature (in close proximity), the kT/q terms cancel, and the performance of the circuit is essentially independent of temperature. When ranges are switched, the circuit recovers quickly (from milliseconds to microseconds), because the switching is remote from the picoampere-level circuitry.

[3] See "The World of fA—Op Amps as Electrometers," *Analog Dialogue*, Volume 5, No. 2.

[4] "High-Performance Flame-Ionization Detector System for Gas Chromatography," *Hewlett-Packard Journal*, Volume 24, No. 7.

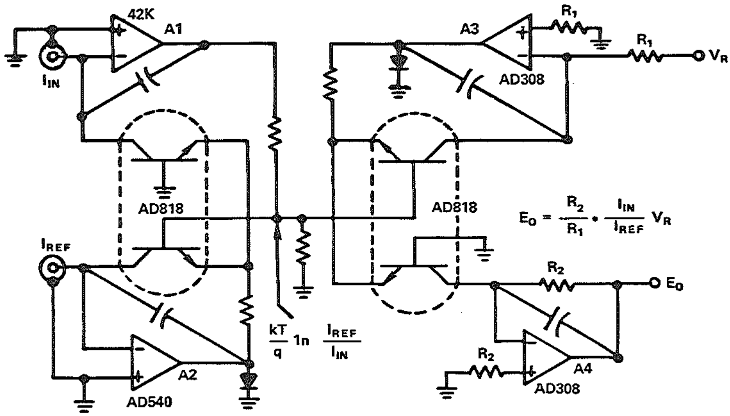


Figure 6. Temperature-compensated wide-range picoammeter with normal resistance values, non-interactive range scaling and voltage-adjustable scale factor

CORRELATION AND CONVOLUTION

These topics involve equations of the typical form

$$F(\tau) = \int_0^{\tau} f(t) \cdot g(\tau - t)dt \qquad (3)$$

While there is simply not enough space available in the present volume even to touch (however inadequately) on these topics, they must nevertheless be mentioned, because the high speed and low cost of multipliers and multiplying D/A converters, and their small space requirement, makes analog or partially-analog approaches more competitive with digital techniques than has been the case in the past. A typical circuit that embodies (3) is shown in Figure 7.

Correlation is used as a means of recovering information in the presence of noise or unrelated signals. If the information is sinusoidal, and the "noise" is an out-of-phase component at the same frequency, equation (3) can be recognized as a phase-sensitive detector, where τ is the delay corresponding to the phase-shift ϕ. For signal waveshapes having less-predictable properties, τ is the delay of an adjustable delay line. The integration is performed a number of times (depending on the desired resolution), for various values of τ up to the full period, and each integration reconstructs one point on the correlation function. Adjustable delay lines are available, for short delays, in analog form (e.g., "bucket-brigade" types), and for arbitrary delays, in digital form.[5] The most-popular forms of correlation are auto-correlation and cross-correlation, determined by the relationship between the functions f( ) and g( ).

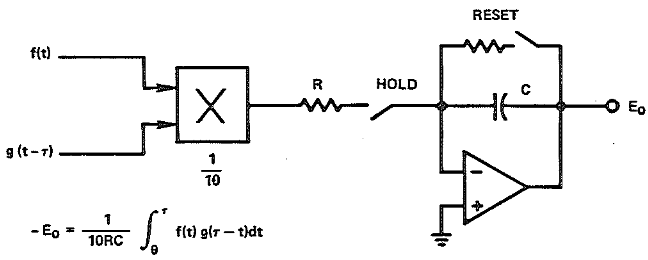


Figure 7. Basic analog correlation circuit

Convolution of time functions corresponds to multiplication of their transforms in the *s* or jω domains. Multiplying the transform of an analog signal by the transform of the indicial (step, pulse, etc.) response of a linear circuit (i.e., its complex transfer function) provides the transform of the time response of the circuit to the analog signal. It is therefore possible to *model* the time response of a circuit to a time waveform without actually building the circuit by a series of convolutions of the input waveform with an independently-generated waveform that has been fitted* to conform to the desired indicial time response. This is an especially useful technique if the desired indicial response requires an unreasonable or not-physically-attainable transfer function.

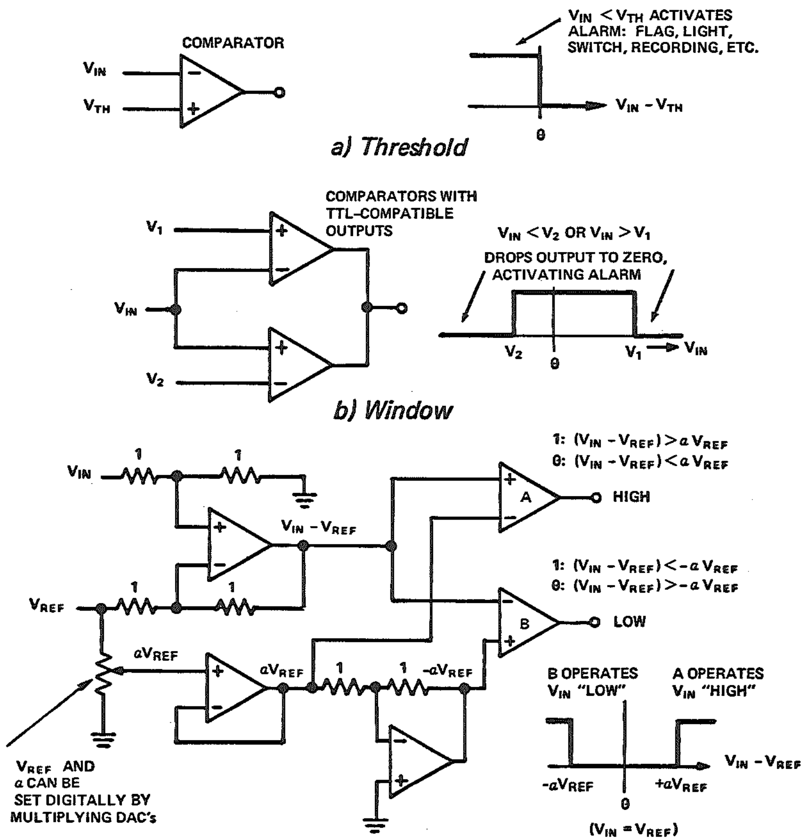
[5]*Analog-Digital Conversion Handbook*, Analog Devices, Inc., 1972

*See chapters 2-1 and 2-2 on function fitting and function generation.

ALARM CIRCUITS

In a system, there are usually a number of variables the magnitudes of which are unimportant from the standpoint of their contribution to the equations of the performance or efficiency of the on-going process, but must nevertheless be maintained within given tolerances. While it is possible to convert and record or observe these variables, it is usually cheaper and simpler to take note of them only when they have deviated beyond one or more sets of thresholds.

Figure 8 shows three circuits that can be used to activate alarms if



c) Fractional deviation ($\pm\alpha V_{REF}$) from preset reference (V_{REF})

Figure 8. Alarm circuits

the input exceeds or falls below a preset threshold (a), if the input departs from a prescribed range of operation (b), or if the input departs from a reference value by more than a prescribed percentage (c).

These are the simplest kinds of deviations requiring alarm. Most others can be constructed using them as basic elements. Using the Serdex system (see page 88), it is possible to obtain remote information on the state of both alarm-only and measured variables at the same time via the same twisted-pair, in the form of a coded printout.

Naturally, these same alarm functions can be recognized as key elements of tolerance control in automated production operations of all sorts: machining parts, adjusting precision resistances, keeping machine speed within limits, etc.

CLASSIFICATION

For processes that must measure certain properties of objects (size, current-gain, resistance, brightness) and identify those units falling into specific classifications, the circuit of Figure 9 may be found useful.

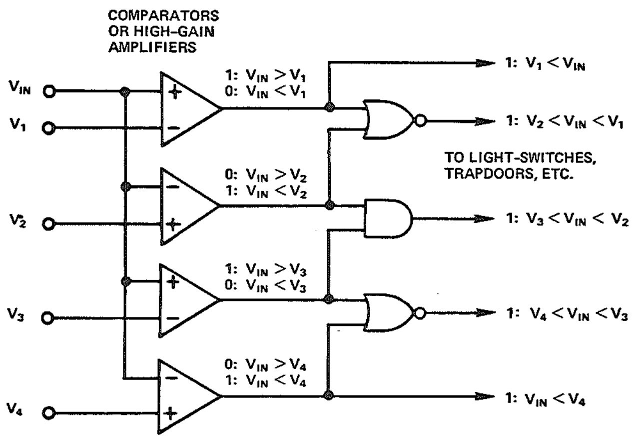


Figure 9. Classification circuit. Latching-type comparators and/or hysteresis may be used to reduce ambiguities and "cycling" due to noise

The input voltage, corresponding to the measurement of the property of interest, is compared with a series of graduated reference levels. The outputs of the comparators are processed by simple logical operations, the outputs of which indicate uniquely into which grade the object whose property is being measured should be placed. These outputs activate the appropriate trapdoor, indicator light, etc.

MEDIAN CIRCUIT

An analog signal may be transmitted along several redundant paths to improve reliability or reduce noise. While simple summation of the outputs will reduce uncorrelated noise, the output may be greatly in error if one path has failed in a saturation mode. One means of combination that secures both noise reduction and

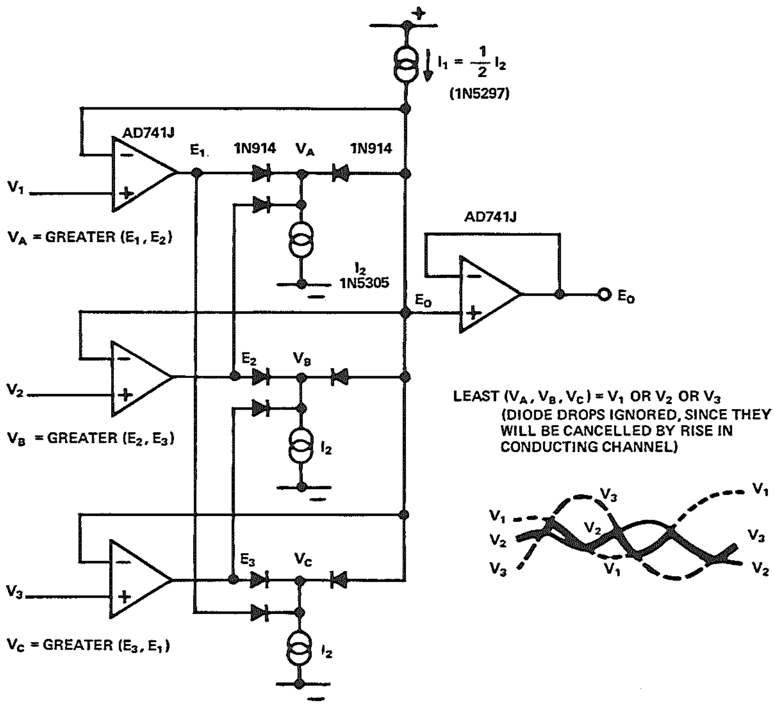


Figure 10. Median circuit continuously selects middle value among three inputs for noise reduction or improved reliability through redundant circuitry

protects against failure of one path involves computing the median signal. That is, the output is always the signal the value of which is between the other two signals.

Figure 10 shows one form of circuit that can compute the median of three input signals V_1, V_2, and V_3. It computes the greater of each pair (V_1, V_2), (V_2, V_3), (V_3, V_1), and then follows the least of the three "greaters." It will therefore follow continuously the signal which is neither the greatest nor the least, irrespective of which one it happens to be at a given instant.

It has been claimed possible, using similar circuitry, to design a circuit that will follow the *m*th in magnitude among *n* input signals.[6]

TRIGONOMETRIC FUNCTIONS AND COMBINATIONS

Rectangular-to-polar conversion involves computations of the form (vector composition)

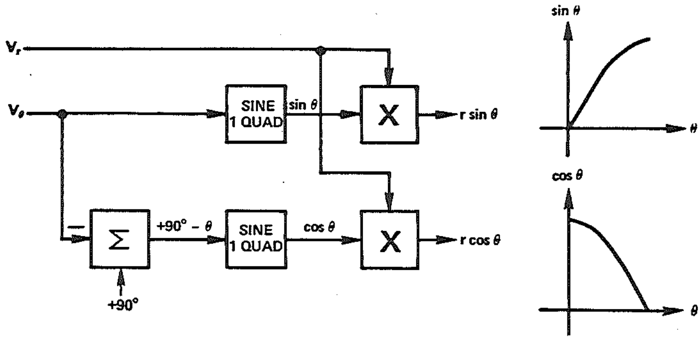
$$\left.\begin{aligned} r &= \sqrt{x^2 + y^2} \\ \theta &= \tan^{-1}(y/x) \end{aligned}\right\} \qquad (4)$$

and polar-to-rectangular (vector resolution) involves the inverse operation,

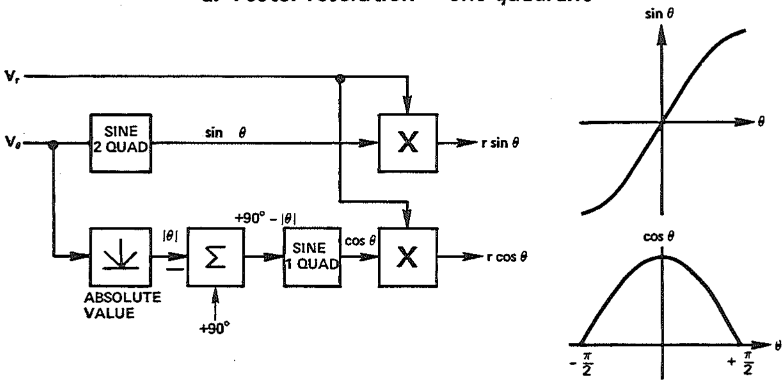
$$\left.\begin{aligned} y &= r\sin\theta \\ x &= r\cos\theta \end{aligned}\right\} \qquad (5)$$

Chapter 2-1 has discussed in great detail the manner of fitting $\sin\theta$. Since the cosine of an angle θ is the sine of $(90^\circ - \theta)$, circuits that fit $\sin\theta$ can also be used to fit $\cos\theta$ (Figure 11). Usually, two sine-function-fitters are needed per angle, but if the signal varies slowly, multiplexing may be used to allow one function fitter to share sine and cosine (of a number of different angles, if necessary).

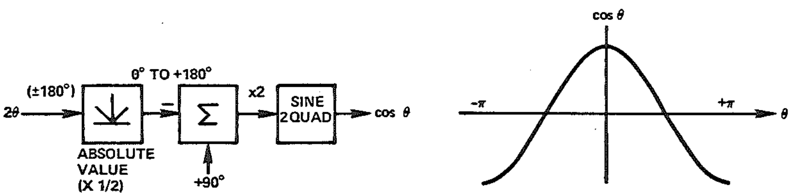
[6]"Analog Sorting Network Ranks Inputs by Amplitude and Allows Selection," *Electronic Design* 2, January 18, 1973, and sequel, *Electronic Design* 17, August 16, 1973, p. 7.



a. Vector resolution — one quadrant



b. Vector resolution — two-quadrant

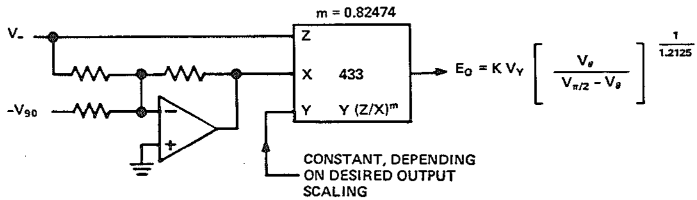


c. Cosine approximation — 4 quadrants

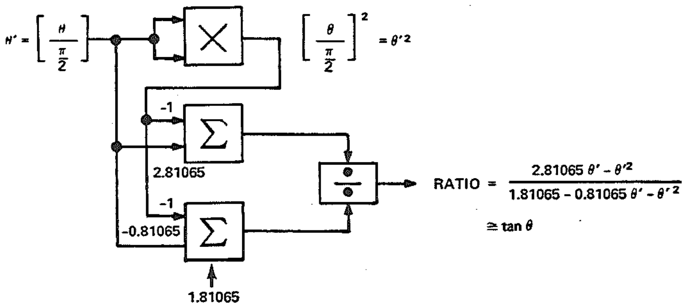
Figure 11. Cosine approximations using sine function fitters

Vector composition (4) is discussed in the text that accompanies Figures 20 and 21 in Chapter 2-3.

If it is necessary to compute the tangent of an angle, the two schemes outlined in Figure 12 may be of interest. The first, based on the arctangent scheme, is the simpler but somewhat less accurate (within 1.4% of tan 45° up to 50°, within 2.5% of ideal value up to 80°, within 2% of 10 (i.e., tan 84.3°) up to 86°. A more-accurate scheme, involving squaring and division, is shown in 12b.[7]



a. Tangent approximation based on arctan circuit of Figure 21, Chapter 2-3



b. Tangent approximation with multiplication, division, and summing. Theoretical absolute error less than 0.0001 to 63°, relative error less than 0.1% to 72°, less than 0.5% to 81°, less than 1.2% for all values of θ < 90°.

Figure 12. Approximations to the tangent

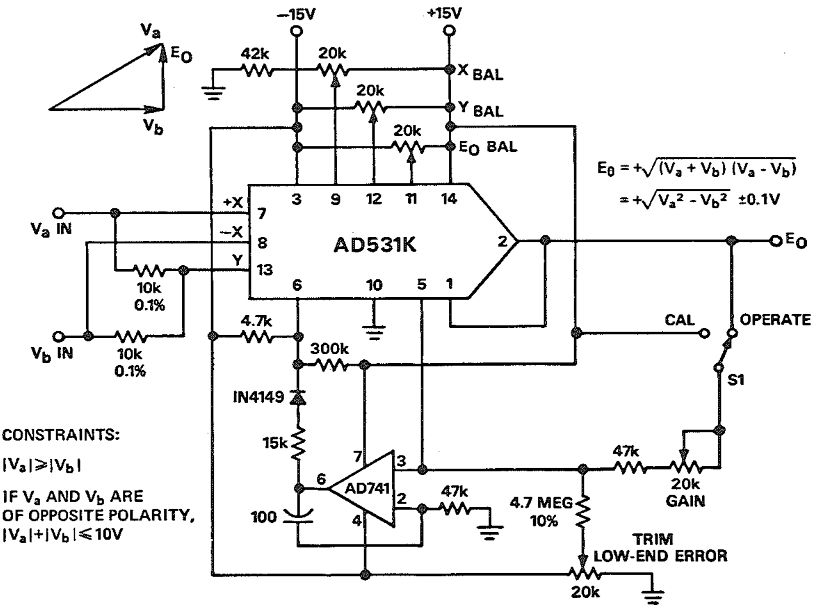
Figure 13 is a circuit for computing the *difference* between two orthogonal quantities, using an XY/Z device (e.g., the AD531 multiplier-divider) for implicit square-rooting:

$$E_o = \sqrt{V_a^2 - V_b^2} = (V_a + V_b)(V_a - V_b)/E_o \qquad (6)$$

To minimize the use of external amplifiers, the sum-term is obtained passively and the difference is inherently available at the

[7]The equation that this scheme embodies is similar to that used (in a different guise) in a patented digital-to-synchro converter designed by F. H. Fish, of the U.S. Naval Avionics Facility, Indianapolis, Indiana.

differential "x" inputs. The AD741J feedback amplifier converts the output voltage to a current. When properly calibrated, and adjusted for less than 100mV error at full-scale, the output of this circuit will differ from the theoretical value by less than ±100mV for any pair of input voltages over an output dynamic range between 10V:0.3V and 10V/0.1V. Bandwidth is dc to 100kHz for best accuracy and 600kHz for -3dB error.



CALIBRATING THE CIRCUIT

Step	Condition	Adjust	For
1.	CAL, $V_a = V_b = 0V$	E_O BAL	$E_O = 0V$
2.	CAL, V_a = 20Vp-p, 10Hz, $V_b = 0V$ Pin 13 (AD531) grounded Scope sensitivity 50mV/cm(V), 100ms/cm(H)	Y BAL	Min. E_O swing
3.	CAL, $V_a = V_b = 0$, 20Vp-p to pin 13 Scope sensitivity as in (2)	X BAL	Min. E_O swing
4.	OPERATE, V_a = 10.00V, V_b = 0V	GAIN	E_O = 10.00V
5.	OPERATE, V_a = 1.00V, V_b = 0V	"Low end"	E_O = 1.00V
6.	OPERATE, $V_a = V_b = 0V$	E_O BAL	E_O = 0V

Figure 13. Vector-difference circuit using AD531

ADAPTIVE CONTROL (Figure 14)

It has been noted, in Chapter 2-4, that the multiplier is essentially a remotely-operated gain control. Because it is free from the reliability and speed problems of servoed potentiometers, and is several orders-of-magnitude lower in cost, it is a natural choice for variable gains in adaptive control systems. If the gain criteria are determined by analog computation, analog multipliers (or dividers) are used; if they are determined digitally, multiplying D/A converters, such as the monolithic AD7520, are used.

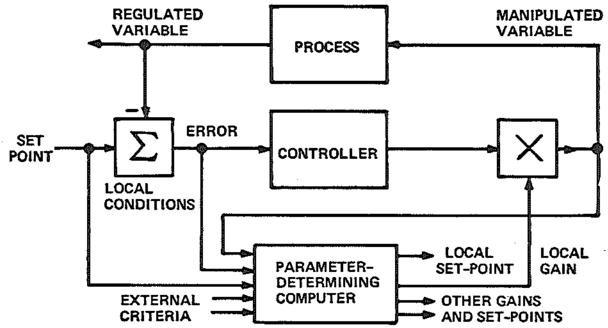


Figure 14. Adaptive control loop: multiplier as loop-gain adjuster

LINEARIZATION

Besides transducer linearization (discussed at length in Chapter 2-3), another interesting applications area for nonlinear devices, especially high-speed multipliers, lies in linearizing inherently nonlinear displays. For example, aside from other sources of nonlinearity and error, the spot-position on the face of a non-spherical cathode-ray tube is subject to "pincushion" distortion and defocusing as a consequence of the inherent geometrical relationships.

As a result of such distortion, the x-coordinate, the y-coordinate, and the spot width are multiplied by a factor of the form

$$\sqrt{(a_1 x)^2 + (a_2 y)^2 + 1} \qquad (7)$$

where x and y are the deflection voltages.

In order to correct for this distortion, the deflection voltages must, in effect, be divided by this term. Although it is feasible, speed and accuracy limitations make it preferable to use an approach in which a correction term (usually small) is *added* to the deflection voltages at the deflection amplifier. In this way, introduction of additional nonlinearity is minimized, and additional delay through the correction circuit applies only to a small correction rather than the entire deflection signal.

The choice of a suitable additive function is not a matter agreed upon by all designers. Examples of functions that have been employed are:

$$kX = AV_x + BV_x(CV_x + DV_x^2 + EV_xV_y) \qquad (8)^8$$

$$kX = AV_x + BV_x(V_x^2 + V_y^2) \qquad (9)^9$$

$$kX = AV_x + BV_x + CV_x\sqrt{V_x^2 + V_y^2 + D} \qquad (10)^{10}$$

The X-correction is shown in the above examples, but the Y-correction is similar in form, as is the focus correction.

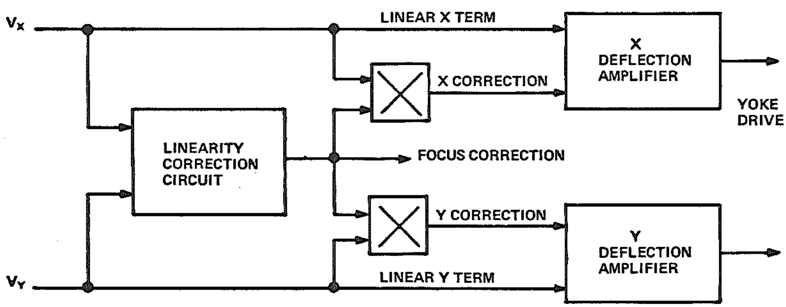


Figure 15. Linearity correction for cathode-ray tubes

[8] "Linearize Your CRT Displays," *Electronic Design* 17, August 16, 1970.

[9] "IC Op Amps Straighten Out CRT Graphic Displays," *Electronics*, January 4, 1971.

[10] *Distortion Correction in Precision Cathode-Ray Tube Display Systems,* Intronics, Inc., 1970.

CONCLUSION

We have shown in these chapters many of the ways that nonlinearity can be, is being, and will be used by system designers to do jobs in a practical, economical, and (often) uncomplicated manner. We have considered function fitting, function generation, instrumentation and measurement, signal-processing, and sundry other applications, as examples of the near-universality of the analog approach.

In the next section, we shall inspect the devices that are used for these applications more closely, with an eye to learning more about how they are designed and understanding their important properties, characteristics, and specifications.