Lock in the Devil, Educe Him, or Take Him for the Last Ride in a Boxcar,

L C Fisher P.A.R. Research Note, Vol. 6, No. 1, 1960.

Signal Processing

Box car + Lock-in

NO. 1

...E HIM, OR TAKE HIM FOR ...ST RIDE IN A BOXCAR?

In sp... ...pearances, this is *not* a message from S.P.E.C.T.R.E. to James Bond. It is an allegorical expression of the riddle facing the modern scientist who wishes to measure ever smaller changes in the parameters of the physical world, only to find that Nature has conspired to thwart him by veiling her secrets in a miasma of noise. In our allegory, we shall personify all the noise demons of the universe in the form of Old Devil Decibel, the Mephistopheles of Measurement, beside whose machinations the schemes of Goldfinger seem like mere fund-raising for the Community Chest.

The protagonist of our allegory is the Knight of Knowledge, who represents all the unsung legions of scientific researchers, toiling in their laboratories to unravel the mysteries of the universe. Our Knight has not the gleaming broadsword of Sir Lancelot, nor the six-gun of Matt Dillon. The monotony of his struggle is seldom relieved by the amorous intrusions of the shapely young creatures who distract 007 while he tools his Aston-Martin down Alpine roads. Our Knight is far from helpless, however, for his arsenal includes such exotica as memory oscilloscopes, spectrophotometers, and magnetic-resonance spectrometers, not to mention numerous other complexitrons. In fact, the Armory of Science bristles with so many sophisticated weapons that our Knight of Knowledge may find his choice consumes so much time and energy that he is exhausted before the battle begins. One could scarcely blame our hero for yearning now and then for the simple days of old when Galileo spent his time dropping cannon balls from the leaning tower of Pisa and denying charges of heresy by the Inquisition.

But he cannot turn back the clock. He must choose his weapons and go on with the struggle. The purpose of this issue of our homely journal is to set forth some guidelines which may assist our hero in making his choice. To this end, it behooves us to take a closer look at old Beelzebub.

THE FACE OF THE DEVIL

With apologies to those readers who recall that we examined this matter in an earlier issue (HOW TO HEAR A HUMMINGBIRD IN A HURRICANE), we proceed now to define certain fundamental concepts. The experimental scientist, whatever his specialty, has one or the other of two objectives: he seeks to compile a catalog of empirical facts about the physical world from which he can construct theories to explain these facts, or he seeks to observe facts with which he can verify or refute existing theories. In either case, he must make quantitative measurements of physical parameters such as temperature, pressure, current, frequency, etc., by means of various devices we call *instruments*. The process of physical measurement is essentially one of detecting *signals* in Nature, and we shall define these as follows: *A signal is any variation with time in the value of a physical parameter, which conveys information about one or more members of the whole group of physical phenomena which influence the value of that parameter.*

It should be noted that there must be a *change with time* in the value of the parameter in order for information to be conveyed to the observer. If the parameter remains constant, the observer knows only that the related phenomena have not changed. While this may be construed as *information* in the popular usage

*Technical Director, American Dynamics Corporation.

inescapable, for tha.

From this starting poin... *in the value of a physical parameter the observer wishes to study out of the w...* be noted that we have defined noise as *unwan...* which may offend the purists of special disciplines. However, we w... discussion as closely as possible to the intuitive notions which our readers will from their everyday contacts with noise. As we observed in HUMMINGBIRD, noise, invades every channel of communication between the human brain and the outside world, including those extensions of the senses which are the scientist's instruments. Although his brain evolves adaptive responses to the noise of everyday living, the scientist's instruments have no such automatic compensations for the noises which they introduce into the measurement system. Every piece of his measurement hardware must be constructed of matter, and matter, because of its discrete nature and its thermal energy at all temperatures above absolute zero, is noisy stuff. At high signal levels, this intrinsic noise is seldom troublesome, but as the *resolution* (the minimum detectable signal increment) of his instruments increases, and the signal levels decrease, noise becomes the final limitation on his ability to make measurements.

Most of the physical measurements made by scientists today involve the use of electricity at some point in the measurement system, either because the signal is electrical to start with, or because of the highly developed technology of electrical devices and the availability of a large variety of *transducers* to convert other signals into electrical variations. Since all electronic instruments must be built with wires, resistors, diodes, transformers, tubes, and transistors, among numerous other things, and since these elements possess intrinsic noise of one sort or another, Old Devil Decibel casts his spell even upon the weapons our Knight forges to fight him.

Consider an ordinary resistor at room temperature. When externally short-circuited upon itself, with no sources of electrical energy included in the circuit, it carries no *average* current in either direction, but *instantaneously* there are very perceptible currents in one direction or the other, because of the thermal energy of the free electrons. Just as a sociologist is doomed to failure when he seeks to discern any lasting social significance in the fluctuating length of women's skirts, an observer watching these tiny electronic currents on some indicator like an oscilloscope would never discern any pattern of repetition, because the time variation is truly random. But, like the irreversible flow of dollars into the garment makers' coffers associated with the chaos of ladies' fashions, there are certain predictable phenomena connected with thermal electron noise. If the resistor at room temperature be connected in parallel with another of comparable resistance value, but held at a temperature close to absolute zero, through wires whose thermal conductance to the second resistor can be made insignificant, it will be observed that there is a unidirectional flow of energy from the warm resistor to the cold one, *transferred electrically*, and that over long periods the time-average rate of energy transfer will be a function only of the difference between the absolute temperatures of the two resistors.

Thus, although we cannot predict the exact value of the random electronic noise current at any given moment, we can accurately predict the *long-term time-average of the square of its magnitude*, upon which the energy transfer depends. By the methods of statistics, we can also draw certain non-idiotic conclusions about the instantaneous magnitudes of these so-called *Johnson-noise* currents. In arriving at these conclusions, we abandon our hope for *certainties* and content ourselves instead with *probabilities*.

The foregoing conclusions can be derived from observations made over finite intervals of time, which learned analysts like to call *studies in the time domain*. Thanks to a brilliant French mathematician, Jean Baptiste Joseph Fourier, whose devotion to science was not shaken by the rolling heads of the guillotined nobility nor the exploding cannonballs of Napoleon, we have another view of the problem, known as *studies in the frequency domain*. Fourier demonstrated that every repetitive variation in a time-dependent parameter, which repeats the same cycle of values over and over with a fixed period of repetition, and which satisfies the so-called *Dirichlet Conditions**, can be resolved into an exactly equivalent sum of an infinite series of simple harmonic (sinusoidal) variations whose frequencies are integer multiples of the fundamental (lowest) frequency, which is the reciprocal of the period of repetition. The amplitudes of these harmonic variations are determined from certain integrals of the time-varying phenomenon over the repetition period. Thus, nearly every periodic function of time can be characterized by a *spectrum*, or graphical plot whose ordinates are the amplitudes of the Fourier harmonics, and whose abscissae are the frequencies. A periodic time function possesses a *line spectrum* (one in which only *discrete* frequencies occur). By extension of the basic concepts, we can represent non-periodic time functions, such as the Johnson-noise current in a resistor, by a *Fourier integral*, in which the period of integration becomes infinite, and the sinusoidal constituents have all possible frequencies from zero to infinity. Random noise thus has a *continuous* spectrum rather than a *line* spectrum.

Because of the continuous nature of this spectrum, it is practically meaningless to think of an amplitude associated with each particular value of frequency. All that the Fourier integral (and any empirical measurement of noise) can give us is an *amplitude-density* function which is the first derivative of amplitude with respect to frequency. Over a finite interval of frequencies, called a *bandwidth*, we can then calculate or measure an effective r.m.s. (root-mean-square) noise amplitude. The amplitude-density function of a continuous spectrum, however, is not often of great utility in studying the effects of noise on practical measurement systems, because it lacks any information about *phase*, one of the three variables which characterize sinusoidal functions of time: amplitude, frequency, and phase.

But the concept of *spectral power density*, which is dependent upon the *square* of the amplitude density function, has great practical utility, for it is independent of phase, and can always be measured for noise spectra in the real world, even though the Fourier integral itself cannot, in some cases, be evaluated analytically. In physical terms, each noise source can be characterized by a curve whose ordinates are the *available* power per unit bandwidth and abscissae the frequencies about which the infinitesimal bandwidths are measured in taking the derivative of power with respect to frequency. The *available* power is that which the source could deliver to a *matched* load within the specified bandwidth, and is the maximum power per unit bandwidth which can be delivered by the source to any load.

At this point, it should not be necessary to take a trip on an LSD-powered flying mattress to perceive that Fourier has made it possible to analyze much of the time-domain chaos of noise by the steady-state techniques of the frequency domain which have been so highly developed in the theory of alternating-current circuits. The responses of various linear circuits and networks to specific noise-power-density spectra can be evaluated to show us how measurements of signals can still be made in the presence of noise. Our analysis, though analytically correct as far as it goes, will aim to be illustrative rather than exhaustive, lest too much rigor in the text may produce rigor mortis in the reader. We shall assume only a knowledge of calculus, complex algebra, and the fundamentals of a.c. circuit theory.

Consider Figure 1, which depicts the resistor we have spoken of, together with a sketch of the time variation of the electronic noise current which flows when the resistor is shorted on itself through an essentially noise-free current detector of zero resistance. We shall assume for now that the resistor has negli-

* Within its period of repetition, the function must be single-valued, must have a finite number of maxima and minima, and must have a finite number of finite discontinuities (jumps) in amplitude. All real-world periodic phenomena are of this type, and hence Fourier-analyzable.

gible reactance effects, and shows a constant resistance for all frequencies. The r.m.s. (square root of the long-time average of the square) current flowing due to thermal electronic motions within a bandwidth of frequencies is

$$I_{NO} = \sqrt{4kT\cdot\Delta f/R} \tag{1}$$

where k = Boltzmann's Constant = 1.38×10^{-23} JOULE/°K;
T = the absolute temperature of the resistor in °K;
Δf = the bandwidth in HERTZ;
R = the resistance in OHMS.

This is the *short-circuit* current due to thermal noise. According to Thevenin's Theorem, if we opened the switch of Figure 1, we would expect to see an open-circuit e.m.f. of r.m.s. value

$$E_{NO} = I_{NO}R = \sqrt{4kT\cdot\Delta f\cdot R}. \tag{2}$$

To put some cardinal points on the scale of our noisemeter, Equation (2) says that the r.m.s. noise voltage across a 1,000-ohm resistor at room temperature (300°K) is 4.07×10^{-8} volt in a 100-hertz bandwidth. Although this may seem as quiet as a caterpillar in a caramel custard, remember that there are modern instruments capable of resolving one nanovolt (1×10^{-9} volt)!

If the switch is now closed on the left side, connecting the warm, noisy resistor to a cold (near $T = 0°K$), quiet resistor of equal value R, then the r.m.s. current which flows is given by

$$I_N = E_{NO}/2R = \sqrt{kT\cdot\Delta f/R},$$

and the average noise power transferred to the cold resistor from the hot one is

$$\Delta W_N = I_N^2 R = kT\cdot\Delta f. \tag{3}$$

Note that this is the *maximum* power which can flow from the noisy resistor to the quiet one, because the external resistance of the load equals the internal resistance of the source, and we have neglected reactance effects. We see now that the *spectral power density,* or the *available power per unit bandwidth,* is given by

$$w_N(f) = \frac{dW_N}{df} = \frac{\Delta W_N}{\Delta f} = kT. \tag{4}$$

Because $w_N(f)$ is constant, independent of frequency, we say that Johnson noise is *white.* However, the analogy with white light is inexact, because white light has a constant available power per unit *wavelength,* giving a power per unit bandwidth which rises as frequency decreases.

The simple relation of Equation (4) becomes inexact as T approaches zero, or as f rises, and must be replaced by the more precise formulation

$$w_N(f) = \frac{hf}{\exp(hf/kT) - 1}, \tag{5}$$

where h = Planck's Constant = 6.625×10^{-34} JOULE-SECOND.

This relation should be used when hf is comparable in magnitude to kT. If hf is less than one-tenth kT, then the exponential may be replaced by the first two terms of its infinite-series expansion, or $(1 + hf/kT)$, and (5) simplifies to (4). If $hf \leq 0.1kT$, then we have

$$\frac{f}{T} \leq 0.1\frac{k}{h} = \frac{0.1 \times 1.38 \times 10^{-23}}{6.625 \times 10^{-34}} \simeq 2 \times 10^{9} \text{ HERTZ/°}K.$$

Not only does (5) give more accurate values for low T and high f, but it also avoids the mental trauma induced in mathematical minds by the realization that (4) implies *infinite total power* if *all* frequencies are considered.

Another type of noise commonly encountered in measurement systems is the so-called $1/f$-noise, whose name derives from the fact that the available power per unit bandwidth varies as the reciprocal of some power of the frequency:

$$w_N(f) = K/f^n, \tag{6}$$

where K = the power-density constant, WATTS/ HERTZ $^{1-n}$;
n = a numeric, $0 \leq n \leq 2$.

If $n = 0$, we have white noise. Strong line spectra of noise also can occur, as in the case of electric-power-line frequency (60 hertz) and its harmonics. In general, there is a wide variety of noise-power spectra which can infest the measurement systems of scientists. One can readily perceive that Beelzebub has a thousand faces, simply by doing what Einstein called a gedanken-prufung, or thought-experiment. Imagine a young biophysicist trying to get his well stacked lab assistant in the mood to visit his pad for an evening of discussion on polypeptide chains (with occasional breaks for less scientific activities). He decides to take her to dinner at a cool bistro, followed by a visit to the philharmonic. While attempting to whisper sweet nothings into her left sonar transducer, he discovers that the right one is being driven to saturation by a strong noise-power spectrum known as Chopin's Etude in C Minor. With apologies to Bach, Grieg, Korsakov, and Satchmo Armstrong, every musical composition is a noise spectrum quite distinct from every other, and the merest acquaintance with the theory of permutations and combinations tells us that a veritable infinity of spectra can be constructed from musical notes.

But, whatever mask Old Devil Decibel may wear, he can always be characterized by a specific spectrum of available noise power per unit bandwidth. Figure 2 shows plots of such spectra for white noise, $1/f$-noise and an assumed spectrum having peaks and valleys. In each case the power which the noise source can deliver to a *matched* load is given by

$$W_N = \int_{f_1}^{f_2} w_N(f)\, df. \tag{7}$$

This can be evaluated graphically as the area under the curve of $w_N(f)$ in each case, between the limits f_1 and f_2. As our sharp-eyed subscriber in Bozrah, Connecticut has no doubt already noticed, if the bandwidth $\Delta f = f_2 - f_1$ is narrow and lies in a valley in the spectrum, the noise power dumped into his measurement system will be minimized.

This frequency-domain characterization of noise, though very useful in our subsequent study of methods for separation of signals from noise, is somewhat like a description of Miss America in terms of 3 circumferences: fascinating, but not exhaustive. A complete description must also include some discussion of the statistical properties of noise as a function of time, but for economy of space we shall defer this to a future issue of TEK-TALK.

PUTTING THE SLUG ON BEELZEBUB

We note that Equation (7) gives the power delivered by a noise source to a *matched* load: a pure, constant resistance equal to the internal resistance of a non-reactive source, or a complex impedance equal to the conjugate of the complex internal impedance of a reactive source. Harkening back to the ancient Celtic discovery that a funnel transmits beer in one direction but only foam in the other, we wonder whether our measurement systems can be made less receptive to noise while preserving their ability to transmit signals. Fortunately, such is the case, for the response of an electric network to input stimuli is, in general, dependent upon frequency. Let us now imagine a black* box having 2 input terminals and 2 output terminals, as shown in Figure 4, whose input terminals are excited by an alternating potential difference

$$v_i = \hat{V}_i \sin 2\pi f t, \text{ VOLTS,} \tag{8}$$

where $\hat{V}_i$ = the peak value of the p.d. in VOLTS;
t = time in SECONDS.

The contents of the box need not concern us, whether they be resistors and capacitors, or machines operated by trained ants, provided that the output voltage is always related to the input voltage by the same unique function, which we shall call the *transfer function* of the box. This output voltage may be written

$$v_o = \hat{V}_o \sin (2\pi f t + \theta), \text{ VOLTS.} \tag{9}$$

Confining our attention to sinusoidal variations at a single frequency f, we note that a sinusoidal variation with time may be regarded as the Y-axis projection of a radial line whose length is proportional to the peak value, rotating in the X-Y plane at a constant cyclic frequency f about the origin of coordinates. The angle ϕ between the X-axis and the radial line is the time-angle of the wave, $2\pi f t$, where f is the frequency in HERTZ(CYCLES/SECOND), and t is the elapsed time in SECONDS from the start of the process. Figure 3(a) shows such a rotating radius, which we shall call a *phasor*. When we deal with more than one sinewave at the same frequency, we are usually concerned more with relative peak or r.m.s. magnitudes and relative phases than with instantaneous values, so we let our X and Y-axes rotate with the phasors, as shown in Figure 3(b), and then determine the phasors representing sums and differences of two or more sinusoids by combining their individual phasors according to the geometric rules for the addition of *static vectors*, or by the more potent methods of complex algebra as applied to those points in the X-Y plane which are the outer tips of the phasors. Phasors of different frequency cannot be combined on a single plane except to represent conditions at one instant of time.

Hence we may express the input and output voltages in the familiar complex-algebra notation for phasor quantities (sometimes loosely called *vectors*):

$$\bar{V}_i = V_i e^{j0} = V_i + j0 = V_i, \text{ VOLTS,} \tag{10}$$

and

$$\bar{V}_o = V_o e^{j\theta} = V_o (\cos\theta + j \sin\theta), \text{ VOLTS,} \tag{11}$$

where e = the base of the natural logarithms = 2.71828;
$j = \sqrt{-1}$.

The ratio of output voltage to input voltage is the *transfer function*, or

$$\bar{g} \equiv \bar{V}_o/\bar{V}_i = g e^{j\theta} \tag{12}$$

Therefore,

$$\bar{V}_o = \bar{V}_i g e^{j\theta}, \tag{13}$$

and the magnitudes (or absolute values) of the input and output are related by

$$V_o = |\bar{V}_o| = |\bar{g}| \cdot |\bar{V}_i| = g V_i. \tag{14}$$

The net effect of the black box is to alter the magnitude and shift the time-phase of the input voltage, in a manner which is dependent only on the frequency if the constituents of the box are constants independent of the magnitude of the excitation and independent of other parameters such as temperature, etc. Since $\bar{g}$ is a function of frequency, we shall express it in the more general functional notation

$$\bar{g} = \bar{g}(f).$$

* A chartreuse box would do equally well, but black is the time-honored color of the boxes used by circuit analysts.

If this black box is now subjected to an input excitation consisting of a whole *spectrum* of sinusoidal voltages, the output voltage may be determined as the sum of the separate responses to each of the distinct, single-frequency components of the exciting voltage. If the input spectrum is a line spectrum of the Fourier-series type, then the output spectrum will be of the same type, and will contain one harmonic term corresponding to each of the harmonic terms of the input spectrum:

$$v_o(t) = \bar{g}(kf) \sum_{k=1,2,3}^{\infty} \hat{V}_{ik} \sin(2\pi kft + \phi_k) = \sum_{k=1,2,3}^{\infty} g(kf)\hat{V}_{ik} \sin(2\pi kft + \phi_k + \theta_k) \tag{15}$$

where k = the order of the Fourier harmonic = 1, 2, 3, 4, 5, etc.;
$\hat{V}_{ik}$ = the peak value of the kth harmonic wave;
ϕ_k = the time-phase angle of the kth harmonic wave;
θ_k = the time-phase shift due to the black box for the kth harmonic of the fundamental frequency f.

If the input spectrum is a continuous one, of the Fourier-integral type, then the r.m.s. voltage-density functions at the input and output are related by the equation

$$v_o(f) = |\bar{g}(f)| v_i(f), \text{ VOLTS/HERTZ}, \tag{16}$$

and the r.m.s. voltage over any bandwidth is given by

$$V_o = \int_{f_1}^{f_2} v_o(f)\, df = \int_{f_1}^{f_2} |\bar{g}(f)| v_i(f)\, df, \text{ VOLTS}. \tag{17}$$

It should be pointed out, for the record, that a transfer function is not always a dimensionless (numeric) quantity. It may relate two parameters of different physical dimensions. For example, the impedance of a simple series circuit containing resistance and inductance is the transfer function relating voltage and current:

$$\bar{Z} = R + j2\pi fL = \bar{V}/\bar{I}, \text{ OHMS}, \tag{18}$$

where $\bar{Z}$ is the complex impedance in OHMS;
L is the inductance in HENRIES.

If, as in Figure 4, we interpose a 4-terminal black box with transfer function $\bar{g}(f)$ between a noise source whose available power per unit bandwidth is $w_N(f)$, and a load which would otherwise be matched to this source, then the available power per unit bandwidth at the output of the black box is given by the expression

$$w_{No}(f) = w_{Ni}(f)|\bar{g}(f)|^2. \tag{19}$$

Clearly we can reduce the available noise power density if we can contrive a black box whose transfer function gives nearly unity transmission within the range of frequencies where we expect our *signals* to occur, but has nearly zero transmission for all other frequencies. This is a time-honored device in signal-processing technology, and is known as *narrow-banding* the information system. In one form or another, this is the essential scheme used in all analog systems which separate signals from noise.

Consider Figure 5, wherein we depict a hypothetical transfer function, showing the square of its magnitude characteristic as a function of frequency. We shall define an *equivalent noise bandwidth* for this transfer function as the frequency-axis width of a rectangle whose height is equal to the peak value of the actual squared-amplitude curve, g_m^2, and whose area is equal to the actual area under this curve, for all frequencies from zero to infinity. Analytically this can be expressed as

$$\text{Equivalent Noise Bandwidth} \equiv \beta_N = \frac{1}{g_m^2}\int_0^{\infty} |\bar{g}(f)|^2\, df, \text{ HERTZ}. \tag{20}$$

Since the total noise power available to the load of Figure 4 is

$$W_N = \int_0^\infty w_{N\,o}(f)\,df = \int_0^\infty w_{N\,i}(f)|\bar{g}(f)|^2\,df, \tag{21}$$

we see that a white-noise source having a spectral density of w_NWATTS/HERTZ could deliver to this load, via the intermediate black box, a total average power of

$$W_N = \int_0^\infty w_N|\bar{g}(f)|^2\,df = w_N\int_0^\infty |\bar{g}(f)|^2\,df = w_N g_m^2 \beta_N,\ \text{WATTS}. \tag{22}$$

For lack of a generally accepted name, we shall call the product $w_N g_m^2 \beta_N$ the *white-noise power transmittance* of the black box.

Now let us have a closer look at the trained ants in a few typical black boxes which we frequently use in the real world. Figure 6 shows an idealized series circuit containing an inductance, capacitance, and resistance, connected to a voltage generator producing a constant sinusoidal output at a single frequency f, which may be varied. A wideband voltage detector of high input impedance connected across the output terminals would measure a sinusoidal signal whose amplitude and phase relative to the generator are functions of frequency. We have assumed lumped constants and pure elements, which, at high frequencies, are as hard to find as billboards in a cemetery. However, close approximations to this idealized filter can be constructed using active amplifiers as well as passive elements, and can be made usable over rather wide frequency ranges. Such practical details as suitable input and output impedances, buffering between reiterated stages, etc. can be adequately provided by the rather vast body of electronic technology now at our disposal. The examples we are about to study are only a few out of numerous frequency-selective black boxes, some of which are highly ingenious, but they illustrate the principles and possibilities of this approach to noise reduction.

In Figure 6, the ratio of output voltage to input voltage is given by

$$\bar{g}(f) = \frac{\bar{V}_o}{\bar{V}_i} = \frac{R}{R + j2\pi fL - j/2\pi fC}\,. \tag{23}$$

The maximum value of $|\bar{g}(f)|$ occurs where

$$f = f_o = 1/2\pi\sqrt{LC}, \tag{24}$$

(C is the capacitance in FARADS), and f_o is the *resonant frequency*, or *center frequency* of this circuit. The maximum amplitude of the transfer function is given by the expression

$$g_m \equiv \left|\frac{\bar{V}_o}{\bar{V}_i}\right|_{f=f_o} = \frac{R}{R + j\sqrt{L/C} - j\sqrt{L/C}} = 1. \tag{25}$$

The sharpness of the frequency selectivity of this circuit is measured by a parameter Q, which is defined here as

$$Q \equiv 2\pi f_o L/R,\ \text{NUMERIC} \tag{26}$$

In practical *electrical* networks Q will lie between zero and 100. In *mechanical* networks it is easy to attain values of Q of 1,000 or more.

When the parameters f_o and Q are introduced into Equation (23), it simplifies to the form

$$\bar{g}(f) = \frac{1}{1 + jQ(f/f_o - f_o/f)}\,. \tag{27}$$

Figure 7 shows the amplitude and phase characteristics of this transfer function for 3 values of Q. We note that, for $f/f_o \ll 0.10$ or $f/f_o \gg 10$, the amplitude characteristic on this log-log plot is a straight line whose slope is such that the amplitude changes by 20 decibels for every decade of change in f/f_o, amplitude rising with increasing frequency to the left of resonance ($f/f_o = 1$), and falling to the right. In each case, the phase shift goes from $+90°$ at very low frequencies through $0°$ at resonance to $-90°$ at very high frequencies. As Q is increased, the steepness of the phase change through resonance is accentuated.

For those of our readers who are not electronics sharks, *decibel* notation is used to express ratios whose values may range through many orders of magnitude (factors of 10). To find the decibel value of an amplitude or intensity ratio, multiply the common logarithm of the ratio by 20. For a power ratio (proportional to the *square* of the corresponding amplitude ratio), multiply the common logarithm of the ratio by 10. When the ratio, either for amplitude or for power, is unity, the decibel value is zero. For ratios greater than unity, the decibel value is positive, and for ratios less than unity, it is negative. To express the decibel value of the reciprocal of a ratio, merely multiply the decibel value of the ratio by -1. When ratios are multiplied (as in the case of several black boxes in concatenation), their decibel values are added algebraically. When one ratio is divided by another, the decibel value of the quotient is found by subtracting the decibel value of the denominator ratio from that of the numerator ratio. For example, if $V_o/V_i = 10$, the ratio is equivalent to 20 decibels. For the power ratio $W_o/W_i = (V_o/V_i)^2 = 100$, we have $10 \times 2 = 20$ decibels. $V_i/V_o = 0.10$ or -20 decibels. The abbreviation for decibel is *db*.

The equivalent noise bandwidth of this circuit is

$$\beta_{NQ} \equiv \frac{1}{g_m^2} \int_0^\infty |\bar{g}(f)|^2 \, df = \int_0^\infty \frac{df}{1 + Q^2(f/f_o - f_o/f)^2} . \tag{28}$$

Emulating the learned practitioners of higher mathematics, we neatly sidestep the tunneling-with-teaspoons drudgery of straightforward evaluation of this integral, and call upon the ingenious legerdemain of H. W. Bode, whose famous resistance-integral theorem makes it possible to derive the result easily. For those who may feel that this is as unorthodox as digging beach clams with a plumber's helper, we suggest consultation of Bode's book, NETWORK ANALYSIS AND FEEDBACK AMPLIFIER DESIGN, published by D. Van Nostrand in 1945. If we express the transfer function as $\bar{g}(f) = g_r(f) + jg_i(f)$, then Bode says that

$$\int_0^\infty g_r(f) \, df = -\frac{\pi}{2} \left[\lim_{f \to \infty} (f g_i(f)) \right] . \tag{29}$$

Since the integrand in (28) is precisely equal to $g_r(f)\, df$, (29) is applicable here and leads to the final result that

$$\beta_{NQ} = \frac{\pi f_o}{2Q} . \tag{30}$$

Consider now another useful black box, shown in Figure 8, which is commonly referred to as a *low-pass filter*. The transfer function of this device is

$$\bar{g}_1(f) \equiv \frac{\bar{V}_o}{\bar{V}_i} = \frac{1}{1 + j2\pi f RC} . \tag{31}$$

The maximum transmission for this circuit occurs when $f = 0$, and is

$$g_{m1} = |\bar{g}_1(f)|_{f=o} = 1 . \tag{32}$$

Hence we have the expression for the equivalent noise bandwidth,

$$\beta_{N1} \equiv \frac{1}{g_{m1}^2} \int_0^\infty |\bar{g}_1(f)|^2 \, df = \int_0^\infty \frac{df}{1 + 4\pi^2 f^2 R^2 C^2},$$

$$\beta_{N1} = \left[\frac{1}{2\pi RC} \tan^{-1} (2\pi fRC)\right]_0^\infty = \frac{1}{2\pi RC}\left[\frac{\pi}{2} - 0\right],$$

$$\beta_{N1} = \frac{1}{4RC}. \tag{33}$$

Figure 9 shows the amplitude and phase characteristics of this circuit. Note that when $2\pi fRC \geq 10$, or $f \geq 5/\pi RC$, then the low-pass filter behaves like an *ideal integrator*, which is characterized in the frequency domain by a phase lag of 90° and a magnitude attenuation of 20 decibels per decade of increase in $2\pi fT$, where T is the time-constant of the integrator ($T = RC$ in this case). For those who are musically inclined, the attenuation rate of an ideal integrator is approximately 6 decibels per octave of frequency change.

Returning for the moment to the transfer function in Equation (27), we see that this also is the characteristic of an ideal integrator if $f/f_o \gg 10$, because the phase shift is $-90°$ and the amplitude characteristic under these conditions approaches a slope of -20 db/decade:

$$\bar{g}(f) = -jf_o/fQ. \tag{27a}$$

When $f/f_o \ll 0.10$, Equation (27) approaches the characteristic of an *ideal differentiator*, because the phase shift is $+90°$ and the amplitude characteristic has a slope of $+20$ *db*/decade:

$$\bar{g}(f) = jf/f_oQ. \tag{27b}$$

Figure 10 shows 2 identical low-pass filters in concatenation, with a unity-gain impedance-buffering amplifier between them for the purpose of preventing the second stage from loading the first. In this case, the overall transfer function is

$$\bar{g}_2(f) = \left[\frac{1}{1 + j2\pi fRC}\right]^2 = \frac{1}{(1 - 4\pi^2 f^2 R^2 C^2) + j4\pi fRC}. \tag{34}$$

When $f = 0$, $g_{m2} = 1$. Hence the equivalent noise bandwidth of this device is

$$\beta_{N2} = \int_0^\infty \frac{df}{(1 + 4\pi^2 f^2 R^2 C^2)^2} = \left[\frac{f}{2(1 + 4\pi^2 f^2 R^2 C^2)}\right]_0^\infty + \frac{1}{2}\left[\frac{1}{2\pi RC} \tan^{-1} 2\pi fRC\right]_0^\infty,$$

$$\beta_{N2} = \frac{1}{8RC}. \tag{35}$$

Outside its pass band, $0 \leq f \leq 1/2\pi RC$, this 2-section low-pass filter attenuates at the rate of 40 decibels per decade.

Now let us examine an ideal integrator as a noise-bandwidth-limiting device. Its transfer function is, for *all* frequencies,

$$\bar{g}_i(f) = -j/2\pi f = \frac{1}{2\pi f} e^{-j90°} \tag{36}$$

From Equation (20), we see that the equivalent noise bandwidth of the ideal integrator is

$$\beta_{Ni} = \lim_{f_0 \to 0} \left[4\pi^2 f_o^2 \int_0^\infty \frac{df}{4\pi^2 f^2}\right] = \lim_{f_0 \to 0} \left[\frac{-f_o^2}{f}\right]_{f_o}^\infty = \lim_{f_0 \to 0} f_o = 0. \tag{37}$$

This seems to be as valuable as a machine for making gold bars out of garbage cans, but there is a small difficulty. In real integrators, the integrating time T_i must be finite, and hence the lowest frequency f_o to which it behaves like an ideal integrator is greater than zero. Thus, in practice, we may say that

$$\beta_{Ni} = f_o = 1/T_i. \tag{38}$$

To get around this difficulty, we could employ a concatenation of real integrators, each separated from its adjacent members in the sequence by buffer amplifiers having unity gain, zero phase shift, very high input impedance, and very low output impedance. If the integrating times are all equal to T_i, then the equivalent noise bandwidth of the sequence works out to be

$$[\beta_{Ni}]_n = \frac{1}{(2n - 1)T_i}, \tag{39}$$

where n is the number of integrators used.

Theoretically, we could make $[\beta_{Ni}]_n$ as small as we wish by adding more integrators to the chain, but there is still a scorpion in the salad. The *signal* we wish to transmit will be hopelessly distorted in waveshape by either single or multiple frequency-domain integration, unless the highest significant frequency in the Fourier spectrum for the signal is less than $f_o = 1/T_i$. Therefore, the maximum *usable* value of T_i is less than the *physically attainable* value of integration time, and hence there is a lower limit set on the usable value of equivalent noise bandwidth.

What this means in practice is that in any case where the waveshape of the signal must be preserved, and we cannot work merely with a peak value or some integrated by-product of the signal wave, an *ideal* frequency-domain integrator (i.e. a *continuous* integrator running *forever* in the time domain) cannot be used. However, it is feasible to use a real integrator having an integrating time T_i much less than the shortest time in which a significant change of signal intensity can occur, and let this integrate the signal over adjacent time increments approximately equal to T_i, transmitting one output level for each such interval and then resetting itself for integration of the signal in the next interval. We shall later discuss this technique in some detail.

So we see that in a resonant circuit we can make the equivalent noise bandwidth arbitrarily narrow by making Q sufficiently large, and in a low-pass filter or chain of such filters, by making RC large enough. Thus we can reduce to any desired degree the amount of noise power transmitted to a measurement system by means of interposing a bandwidth-limiting black box in the signal-processing chain between signal source and final detector. Can we finally nail Old Devil Decibel in the barrel and send him over the falls? Remembering the ancient Arabian discovery that amputation of the head has other results besides the relief of sinus pains, we decide to proceed with caution.

DOES ANYBODY HERE SPEAK THE LANGUAGE?

Like the prudent tourist from Brooklyn who takes a fast course at Berlitz in local dialects before visiting Millinocket, Maine, we perceive that some prior knowledge of the signals we hope to hear might be helpful in separating them from the accompanying noise. Assume once more that our prime signal can be converted into an electrical variation by some appropriate transducer, and consider the simple signal shown in Figure 11, wherein a cosine wave of constant frequency f_s is amplitude-modulated at another single frequency f_M. Let us also assume that $f_s \gg f_M$. Analytically this signal may be expressed as

$$v_s(t) = \hat{V}_s(1 + m \cos 2\pi f_M t) \cos 2\pi f_s t. \tag{40}$$

Using the familiar trigonometric identity

$$\cos\theta \cos\phi = \tfrac{1}{2} \cos(\theta + \phi) + \tfrac{1}{2} \cos(\theta - \phi), \tag{41}$$

we expand (40) into the form

$$v_s(t) = \hat{V}_s\left[\cos 2\pi f_s t + \frac{m}{2}\cos 2\pi(f_s + f_M)t + \frac{m}{2}\cos 2\pi(f_s - f_M)t\right]. \tag{42}$$

The first term in the brackets in (42) is called the *carrier* and the second and third terms are called the *upper* and *lower sidebands*, respectively. The Fourier spectrum of this wave therefore contains only 3 lines: one for the carrier frequency f_s and one for each of the sideband frequencies, $f_s + f_M$ and $f_s - f_M$, as shown in Figure 12. Thinking back to our discussion of information, we see that the information here is transmitted by the sidebands, and that the carrier itself merely serves as a vehicle on which the information rides, because the per-unit modulation m appears as a factor only in the sideband terms.

Now let us suppose that a more complicated waveform of modulation occurs, with the same carrier frequency, as shown in Figure 13. In this case, the modulation envelope can be resolved into a Fourier series of harmonics of the lowest modulation frequency, f_M, and the expression for the wave becomes

$$v_s(t) = \hat{V}_s\left[1 + \sum_{k=1,2,3}^{\infty} m_k \cos\,(2\pi k f_M t + \theta_k)\right]\cos 2\pi f_s t. \tag{43}$$

The expansion of (43) by (41) leads to

$$v_s(t) = \hat{V}_s\left[\cos 2\pi f_s t + \sum_{k=1,2,3}^{\infty} \frac{m_k}{2}\cos\,(2\pi[f_s \pm k f_M]t + \theta_k)\right]. \tag{44}$$

In this case we have a whole spectrum of sidebands, whose frequencies are found by forming all the possible sums and differences of the carrier and modulation frequencies. Since the harmonic orders extend to infinity, the difference frequency $(f_s - kf_M)$ must eventually become negative. This need not cause us any analytical schizophrenia, however, because we have based our analysis only upon positive frequencies, and we simply exclude from consideration all negative frequency values as having no physical meaning here.

Negative frequencies correspond to backward angular rotation of the phasors which generate simple harmonic functions of time. We could have based our analysis upon frequencies extending from minus infinity to plus infinity, by postulating that each sinewave is generated as the instantaneous sum of the simultaneous Y-axis projections of two phasors, each having *half* the peak value as its radial length, one phasor rotating forward and the other backward with equal angular speeds, from zero-time starting directions 180° apart and parallel to the X-axis. A cosine wave could be similarly generated by the sum of the X-axis projections of two identical phasors, each of half-peak-value length, rotating in opposite directions at the same angular speed, but starting from congruent positions parallel to the postive X-axis at time zero. Every sinusoid would then take the analytical form

$$v(t) = \tfrac{1}{2}\hat{V}\sin 2\pi f t + \tfrac{1}{2}\hat{V}\sin\,(-2\pi f t - 180°).$$

Since $\sin\phi = -\sin\,(-\phi)$, and $\sin\phi = -\sin\,(\phi + 180°)$, the foregoing expression is equivalent to $v(t) = 2 \times \frac{1}{2}\hat{V}\sin 2\pi f t$, which is a sinewave generated by a single phasor having a positive cyclic frequency f. If we wish to include negative frequencies in our analysis, half the intensity and half the power in each circuit must be assigned to the frequencies from minus infinity to zero, and the other half to those from zero to plus infinity. Except in those cases where the complex-algebra expressions for the Fourier series are used, for the sake of their compact mathematical form, there is no analytical advantage to the inclusion of negative frequencies. Since the *real* form of the Fourier series must be used in the numerical evaluation of harmonic amplitudes of real-world periodic functions anyway, consideration of negative frequencies only complicates our analysis. Therefore we postulate that all the power in our circuits is carried by simple harmonic functions having frequencies which are either zero or positive numbers, and that negative frequencies have no physical significance for us.

Although theoretically we must take account of all harmonic frequencies from zero to infinity in order to reproduce the actual waveform exactly, in practice we can usually attain a sufficiently good approximation by considering frequencies only out to some definite harmonic order k, at which we judge the sum of the Fourier harmonics differs from the true wave by a tolerable error.

If we decide to apply the voltage wave of Equation (44) to the series-resonant circuit of Figure 6, so as to reduce the amount of accompanying noise which can get through to the final detector, then we must be sure that two conditions are fulfilled: (1) *the resonant frequency f_o of the circuit must be equal to the carrier frequency*; and (2) *the highest significant modulation frequency kf_M must be small enough relative to f_s that the corresponding sideband frequencies, $(f_s + kf_M)$ and $(f_s - kf_M)$, lie well within the passband of the tuned circuit.* Otherwise, some of the sideband waves will suffer appreciable magnitude attenuation and phase shift, which will distort the total wave corresponding to the Fourier summation at the output of the tuned circuit. The term *appreciable* in the foregoing sentence is the key word, and it must be defined in each individual situation, but a few guideposts may be helpful.

In this, and in all our subsequent analysis, unless we explicitly make a statement to the contrary, *it will be assumed that our signal-processing devices must preserve the waveshape of the input signal.* Mere peak-value or average-value information is not, in general, sufficient.

Let us consider a specific example. The modulation might consist of a train of rectangular pulses, like that shown in Figure 14, where the repetition period of the pulses is T, their time-width is τ, and their height A_{pp}. Such a modulation would correspond to turning the carrier wave on and off abruptly in short bursts of length τ. By choosing our zero of time exactly at the center of one of the pulses, we can simplify the evaluation of the Fourier coefficients for this wave. Monsieur Fourier's description of this wave has the general form

$$f(t) = A_o + \sum_{k=1,2,3}^{\infty} \left(\hat{A}_k \cos \frac{2\pi kt}{T} + \hat{B}_k \sin \frac{2\pi kt}{T} \right), \tag{45}$$

where A_o is the average value of the wave over any whole number of repetition periods (or the average over any time interval very much larger than one repetition period), and is given by

$$A_o = \frac{1}{T} \int_{-T/2}^{T/2} f(t)\, dt. \tag{46}$$

The amplitudes of the cosine and sine terms are given by

$$\hat{A}_k = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos \frac{2\pi kt}{T}\, dt, \tag{47}$$

and

$$\hat{B}_k = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin \frac{2\pi kt}{T}\, dt. \tag{48}$$

Equation (46) gives us

$$A_o = \frac{1}{T} \int_{-\tau/2}^{\tau/2} A_{pp}\, dt = \frac{A_{pp}\tau}{T} = A_{pp}\delta, \tag{49}$$

where $\delta = \tau/T$, the per-unit duty cycle of the pulse train. The harmonic amplitudes evaluate as

$$\hat{A}_k = \frac{2}{T} \int_{-\tau/2}^{\tau/2} A_{pp} \cos \frac{2\pi kt}{T}\, dt = \frac{2A_{pp}}{\pi k} \sin \pi k\delta, \tag{50}$$

and

$$\hat{B}_k = \frac{2}{T} \int_{-\tau/2}^{\tau/2} A_{pp} \sin \frac{2\pi kt}{T}\, dt = 0. \tag{51}$$

We have blithely skipped over the algebraic steps in evaluation of the integrals here, much as a politician neglects to explain how he plans to carry out his campaign promises, but, in brief, the sine-term amplitudes are zero because we chose the zero of time relative to the pulse train so that the wave is an *even* function of time or one for which $f(t) = f(-t)$. Since the sum of a series of even functions is itself even, and the sum of an even function and an odd function is neither even nor odd, only cosine terms (which are even functions) remain in the final result, which is

$$f_p(t) = A_{pp}\left[\delta + \sum_{k=1,2,3}^{\infty} \frac{2}{\pi k} \sin \pi k\delta \cdot \cos \frac{2\pi kt}{T}\right]. \tag{52}$$

A special case of considerable practical interest is that for which $\delta = \frac{1}{2}$. This is the familiar *square* wave. For a square wave which is zero-based (i.e. one having no reversals of sign or polarity), the Fourier series is found directly from (52) by setting $\delta = \frac{1}{2}$ and adding a $\pm$ sign to the average term to account for waves of either polarity:

$$f_{so}(t) = A_{pp}\left[\pm\frac{1}{2} + \sum_{k=1,2,3}^{\infty} \frac{2}{\pi k} \sin \frac{\pi k}{2} \cdot \cos \frac{2\pi kt}{T}\right]. \tag{53}$$

We note that when $k = 2, 4, 6, 8, \cdots$, $\sin \pi k/2 = 0$. When $k = 1, 3, 5, 7, 9, \cdots$, then $\sin \pi k/2 = \pm 1$. Hence (53) simplifies to

$$f_{so}(t) = A_{pp}\left[\pm\frac{1}{2} + \frac{2}{\pi}\left(\frac{1}{1} \cos \frac{2\pi t}{T} - \frac{1}{3} \cos \frac{6\pi t}{T} + \frac{1}{5} \cos \frac{10\pi t}{T} \cdots\right)\right]. \tag{54}$$

For a symmetrical square wave having equal positive and negative amplitudes, and hence no average value, the constant term of $\frac{1}{2}$ in the brackets of Equation (54) disappears, and we have the expression

$$f_{ss}(t) = \frac{2A_{pp}}{\pi}\left(\frac{1}{1} \cos \frac{2\pi t}{T} - \frac{1}{3} \cos \frac{6\pi t}{T} + \frac{1}{5} \cos \frac{10\pi t}{T} \cdots\right). \tag{55}$$

Now we return to the burning question of what minimum bandwidth we must have in a tuned-circuit black box of the type in Figure 6 in order to transmit a reasonable replica of the input square wave to the output terminals. To answer this question, we must first find a reply to another tantalizing question, namely "How square is square?" We realize that this is akin to such riddles as "How high is up?" and "When is it too late?", but we have not flipped our wig. We merely make an arbitrary judgment, based upon experience, that the transmitted wave must approximate the input wave to within a specified error.

We recall that the essence of Fourier's method is that the addition of an infinite series of sinusoidal variations to one another, each with a prescribed amplitude and phase relative to all the others, yields the repetitive waveform we seek. Clearly, any frequency-selective network will cause *some* distortion in any wave other than a pure sinewave at the center frequency of the network, because such a network alters the *relative* phases and amplitudes of the Fourier harmonics which constitute the wave. In Equations (52) through (55), a little thought reveals that the zero crossings of the harmonics are repetitively coincident: every second zero crossing of the second harmonic coincides with one of the fundamental; etc. Any shift of the zero crossings of any harmonic relative to the fundamental is certain to cause an alteration in the resultant total waveform. If there is a simultaneous *amplitude* change for each harmonic, the distortion of the resultant is even more drastic. However, it is intuitively clear that if the correct phases and amplitudes are preserved up to a particular harmonic order, then the maximum distortion caused by *attenuation* and phase shift of the higher harmonics from this order onward cannot exceed the sum of these remaining harmonics, and that this limiting value of total distortion must ultimately decrease toward zero as the order of the highest *undistorted* harmonic is increased.

For networks which give frequency-selective *amplification* at frequencies other than the fundamental, the distortion criterion is more difficult to specify. However, such networks are seldom used where reduction of total noise power is the objective.

In the case of the square wave described by (53), let us assume that a sufficient approximation can be obtained if the black box passes the 13th harmonic with less than 10% attenuation and no more than 10° of phase shift. Further, let us assume that the frequency of our carrier, which is being modulated by this zero-based square wave, is a specific multiple of the fundamental frequency of the square wave:

$$f_s = 100/T.$$

Finally, let us assume that the resonant frequency of the series R-L-C circuit is set equal to the carrier frequency: $f_o = f_s$, or $f_s = 1/2\pi\sqrt{LC}$. Now our original question is no longer just psychedelic gibberish, but devolves into the simpler one of determining the maximum permissible value of Q. Either the attenuation criterion or the phase-shift criterion may be the determining one, but not both. Let us first examine phase shift.

From Equation (27), we derive the attenuation and phase characteristics of the tuned circuit:

$$|\bar{g}(f)| = \frac{1}{\sqrt{1 + Q^2(f/f_o - f_o/f)^2}}, \tag{56}$$

and

$$\theta = -\tan^{-1} Q(f/f_o - f_o/f). \tag{57}$$

Here our highest significant modulation frequency is 13 times the fundamental of the square wave, and the fundamental of the square wave is 0.01 times the carrier frequency, which is equal to the resonant frequency of the tuned circuit. Looking back at Equation (42), we see that the process of amplitude modulation produces sums and differences of the carrier frequency and the harmonics in the modulation spectrum. Hence the *highest* value of f we must consider in Equations (52) and (53) is

$$f_{\max} = f_s + 13/T = 100/T + 13/T = 113/T.$$

But $f_o = f_s = 100/T$, so we have $f_{\max}/f_o = 113/100 = 1.13$. Looking first at the phase criterion for this upper sideband, we have

$$\theta = 10^\circ = \tan^{-1} Q(1.13 - 1/1.13),$$

$$0.24504Q = \tan 10^\circ = 0.17633,$$

$$Q = 0.17633/0.24504 = 0.7196.$$

The *lowest* value of f we must consider is the difference between the carrier and the 13th harmonic of the modulation, or

$$f_{\min} = f_s - 13/T = 100/T - 13/T = 87/T.$$

Hence, for the phase criterion at this lower sideband, we have

$$Q(0.87 - 1/0.87) = \tan(-10^\circ) = -0.17633,$$

$$Q = 0.17633/0.2794 = 0.6311.$$

Since this is the lower value of Q it is the one we must choose. Substitution of this value of Q into (52), together with the *lower* sideband value of f/f_o (which gives the more severe attenuation) yields

$$|\bar{g}(f)| = \frac{1}{\sqrt{1 + 0.6311^2 \times 0.2794^2}} = \frac{1}{\sqrt{1.0233}},$$

$$|\bar{g}(f)| = 0.9885.$$

Since our criterion is that $|\bar{g}(f)| \geq 0.900$, this is acceptable.

It should be evident that we could have phrased the question differently: we could have started with a specified value of Q, and then determined the *minimum* ratio of carrier frequency to modulation fundamental in order that the 13th harmonic would produce upper and lower sidebands that would be shifted in phase by no more than 10°, or attenuated by no more than 10%.

But, like the little old grandmother who prefers live tadpoles to olives in her evening martinis, each experimenter must choose his own criteria of acceptability of filter performance.

There are many other interesting types of frequency-selective networks and circuits, such as the famous twin-T (or parallel-T), whose inventor, Mr. Augustadt, was granted U. S. Patent No. 2,106,785 on this ingenious network. All frequency-selective networks can be categorized into general classes, according to the shapes of their transfer-function loci when plotted on a complex plane. The length of a phasor with one end fixed in the origin of coordinates represents the magnitude, $|\bar{g}(f)|$, and the angle between this phasor and the positive X-axis represents the time-phase angle, $\theta(f)$. The sequence of points occupied by the outer tip of this phasor for all values of f from zero to infinity is the complex-plane locus of the transfer function $\bar{g}(f)$. Figure 15 shows these loci for several common black boxes. We note that any transfer function which plots as a semicircle or full circle will have an attenuation rate which approaches ± 20 *db* per decade as frequency goes to very low or very high values, and a phase characteristic which approaches $\pm 90°$ as a limiting value. For transfer functions plotting as complete circles, we may always define a parameter Q, and the value of Q merely affects the manner in which frequency varies with position along the complex-plane locus. Reiteration of these simple networks with wideband, impedance-buffering amplifiers between successive stages leads to more elaborate geometries for the complex plots, but a simple rule still holds. For n buffered identical black boxes of the circular-locus type in concatenation, the attenuation characteristic approaches $\pm 20n$ decibels/decade, and the phase characteristic $\pm 90n°$ as frequency becomes very low or very high.

One final commentary is in order here, regarding the meaning of the parameter Q for circuits or networks whose transfer loci are circles, but which are not simple R-L-C circuits like that of Figure 6. We note from Figure 15 that when $f = f_o$ the typical lead-lag network shows a transfer amplitude of $1 + j0$. Below f_o there will be a frequency f_1 which gives a phase shift of $+45°$, and above f_o a frequency f_2 giving a phase shift of $-45°$. At the frequencies f_1 and f_2 the transfer amplitude is $1/\sqrt{2} = 0.707$; hence at these frequencies the network will transmit only half the available input power, which is proportional to the square of the transmitted intensity or amplitude. From Equation (57), with $f = f_1$ we have

$$\tan \theta = \tan 45° = 1 = Q(f_o/f_1 - f_1/f_o),$$

or

$$Q = (f_o/f_1 - f_1/f_o)^{-1}. \tag{58}$$

By choosing the other half-power frequency f_2 in Equation (57), we obtain

$$Q = (f_2/f_o - f_o/f_2)^{-1}. \tag{59}$$

Since the same value of Q must exist in both cases, we combine (58) and (59) to give us the equality

$$(f_o/f_1 - f_1/f_o) = (f_2/f_o - f_o/f_2). \tag{60}$$

Equation (60) will be satisfied only if we have

$$f_o = \sqrt{f_1 f_2}, \tag{61}$$

which is to say that the resonant frequency is the geometric mean of the half-power frequencies for any black box whose transfer locus is a circle having its center on the X-axis and passing through the origin for $f = 0$ and through $1 + j0$ for $f = f_o$. Equation (61) is equivalent to

$$f_o/f_1 = f_2/f_o. \tag{62}$$

Solving (60) for the per-unit difference in the half-power frequencies, we obtain

$$f_2/f_o - f_1/f_o \equiv \Delta f_{21}/f_o = f_o/f_1 - 2f_1/f_o + f_o/f_2.$$

But substitution from (62) reduces this to the form

$$\Delta f_{21}/f_o = f_o/f_1 - f_1/f_o = Q^{-1},$$

or

$$Q = f_o/\Delta f_{21}. \tag{63}$$

Therefore, we may take this as a definition of the tuning constant Q: Q is the ratio of center frequency f_o to the difference between the half-power frequencies, $f_2 - f_1 = \Delta f_{21}$.

HOW SMALL CAN WE MAKE THE DEVIL'S CAGE?

By now we perceive wistfully that we cannot exterminate Old Devil Decibel, because an equivalent noise bandwidth of zero means a signal bandwidth of zero, and the total loss of ability to transmit information. So, mindful of the advice contained in the Lemma of Ling Hsiung Chin*, an obscure Mongolian philosopher of the 8th century B.C., we inquire how much improvement can realistically be made in our signal/noise ratio, which we shall abbreviate as S/N, and take to be the *amplitude* ratio of r.m.s. signal to r.m.s. noise. The exact answer depends upon the specific noise-power spectrum of our measurement system, but for now we shall assume the noise to be white: $w_N(f) = K$. Also, since every *practical* measurement system has a *finite* power-transmission bandwidth, let us postulate that our measurement system has an equivalent noise bandwidth of β_{NO} before we do anything to improve S/N. Subject to the limitations discussed in the preceding section, we now interpose a frequency-selective black box into the measurement chain at some suitable point between original signal source and final readout device. If this box has an equivalent noise bandwidth of β_N, which is narrower than β_{NO}, then the improvement in signal/noise *power* ratio is

$$p = \beta_{NO}/\beta_N, \tag{64}$$

and in *amplitude* ratio

$$a = \sqrt{\beta_{NO}/\beta_N}. \tag{65}$$

If the black box contains a lead-lag network of circular transfer locus, with a tuning sharpness of Q and a center frequency of f_o, then the factor by which S/N is improved is, from (30),

$$a_Q = \sqrt{2Q/\pi f_o}\sqrt{\beta_{NO}}. \tag{66}$$

For a single-section low-pass filter, we have, from (33),

$$a_1 = 2\sqrt{RC}\sqrt{\beta_{NO}}. \tag{67}$$

To scribe a few marks on our yardstick, assume a lead-lag network which gives unity transmission and zero phase shift at $f_o = 10$ hertz, and has a Q of 25. Such values are easily attainable in practical networks. For this black box, Equation (66) gives $a_Q = 1.261\sqrt{\beta_{NO}}$. Limiting values for realizable electrical black boxes of this type would be in the vicinity of $Q = 100$ and $f_o = 1$ hertz. For these values we have $a_Q = 7.979\sqrt{\beta_{NO}}$. By contrast, consider a single-section low-pass filter having $RC = 100$ seconds, which is readily attainable with available components. From (67), we have $a_1 = 20\sqrt{\beta_{NO}}$. If a 2-section low-pass filter with inter-stage buffering and $RC = 100$ seconds in each section is used, we have, from (35), $a_2 = 2\sqrt{2RC}\sqrt{\beta_{NO}} = 28.28\sqrt{\beta_{NO}}$.

In general, the improvement in S/N for white noise is proportional to the inverse square root of the equivalent noise bandwidth of the device we insert into the measurement loop, provided that this device

* It is better to discover a little cockroach in a dish full of meatballs than a little meat in a dish full of cockroaches.

causes negligible attenuation of the signal carrier and its sidebands. Although it should scarcely need mention, the buffering amplifiers should be designed to introduce as little additional noise of their own as possible. Fortunately, this is possible in practice, and the degradation of S/N due to active amplifiers can be held to 5 *db* or less, even for amplifiers having amplitude gain factors of 80 *db* or more.

Later on we shall consider the question of S/N improvement attainable with noise-power spectra other than white.

Stated in simple terms, our technique of fighting old Noisy Nick is like forcing a fat man to reduce by putting him in a room with no windows and a 12-inch-wide door. However, it involves more than just this. There are numerous ingenious tricks we can employ in the process of incorporating the narrow-door technique into actual instruments. In our subsequent discussion, we shall examine three basic instruments which extract signals from noise, but which differ significantly in design and operation, and which are intended for the recovery of different types of signal.

THE LOCK-IN AMPLIFIER

Suppose that we wish to measure a signal of the type described by Equations (40) and (42) and shown in Figure 11. In most practical cases, such a simple-minded signal would not be encountered, but it serves to illustrate the operation of the lock-in amplifier (sometimes referred to as a *synchronous amplifier* or *phase-sensitive detector*). We assume that the basic physical parameter, such as pressure, temperature, light intensity, etc. has been converted into an electric potential difference by an appropriate transducer. One of the joys of the analyst is that he can make such sweeping assumptions, secure in the knowledge that he, unlike the poor slob who is a synthesist, will never be called upon to see that his assumptions are fulfilled in practice. Our postulate, however, is not quite so grandiose as that in the opening instruction of the home chemistry set on the preparation of magnesium, which begins by telling the user to fill a container with one cubic mile of seawater, because there are numerous transducers available today for the conversion of almost any physical parameter into an electrical one.

We next make an assumption which is as easy to fulfill as finding the teeth on an alligator: the signal is mixed with noise. This noise will, in general, contain two distinct components: (1) noises which are coherent with the signal carrier, and (2) those which are non-coherent. By a *coherent* noise we mean any simple harmonic variation whose frequency f_N is always a rational multiple of the signal-carrier frequency, f_s:

$$f_N = \frac{p}{q} f_s, \tag{68}$$

where p and q are *integers*. If (68) is fulfilled *at all instants of time*, then the noise wave whose frequency is f_N is *phase-locked* to the signal-carrier wave of frequency f_s. In other words, there will be a definite pattern of repetition of the zero crossings of one wave relative to those of the other, with p crossings of the N-wave for every q of the S-wave. Under this condition, a *stationary* lissajous figure could be obtained on the screen of an oscilloscope by applying one wave to the vertical-axis deflection plates and the other to the horizontal.

Examples of coherent, or phase-locked, noises are the Fourier harmonics of any real periodic waveform in which we wish to look at only one particular harmonic, such as the fundamental. In this case, p/q takes on all integer values: 1, 2, 3, 4, 5, 6, 7, etc.

Non-coherent noises, which constitute the much broader category in the real world, are all those for which Equation (68) is not true. Such noise waves will exhibit either a regular or a random phase drift with respect to the S-wave. All random noise in the physical world is non-coherent with any signal wave of *constant* frequency and definite time origin.

To eliminate coherent noises, we can do an excellent job by employing a tuned bandpass circuit of the type whose transfer function is given by Equation (27), having a Q lying between 5 and 25. This will not eliminate frequencies for which p/q lies close to unity, but these are uncommon among coherent noises.

Since this tuned circuit must have a signal bandwidth Δf_{21} greater than zero in order to pass the signal carrier and its sidebands with tolerable attenuation, its equivalent noise bandwidth β_{NQ} must also be greater than zero, and so Old Devil Decibel shoves his hand through the bars. Non-coherent noises generally include *all* frequencies in the spectrum from zero to infinity, and no matter where we locate the passband on the frequency scale, some of this noise comes through.

Consider Figure 16 (a), which shows the essentials of a synchronous detector. The input-buffer amplifier is a tuned amplifier of the type just discussed, with a Q high enough to reduce the coherent noise and to limit the accompanying non-coherent noise to some tolerable level which will not saturate the following stages. The input buffer is followed by a double-pole, double-throw reversing switch, driven by a motor or actuator which is synchronized with the carrier wave in such a way that for every polarity reversal of the carrier wave, there is one of the switch. The waveform of switching will be assumed to be a perfect square wave of zero average value (as shown in Figure 16(b), crossing from (+) to (−) polarity in zero time, and having a dwell period on either polarity which is just half the repetition period, T_s. The phase difference, ϕ, between the zero crossings of the carrier wave and those of the switch will be assumed to be under our control in such a way that it can be varied slowly from zero to 360°. Following the synchronous switch, there is an output-buffer amplifier whose transfer function, $\bar{g}_b(f)$, is a constant over a wide range of frequencies from zero upward: $\bar{g}_b(f) = G_b + j0$. Finally, there is a low-pass filter.

The electrologists (electric-circuit sharks) in our audience will readily perceive that we have here a *synchronous rectifier.* Its output, after amplification by the wideband buffer, $\bar{g}_b(f)$, has the waveform shown in Figure 16(c), whose time-average value is given by

$$(v_r)_{AV} = \frac{2}{\pi} G_a G_b \hat{V}_s \cos \phi, \tag{69}$$

where G_a is the numerical value of the gain of the input buffer at the resonant frequency, $f_o = f_s$. In Equation (69) we are tacitly assuming that the average is being taken over a long enough period of time to include any integer multiple of one full period of the modulation, $T_M = 1/f_M$, or else a very large number of periods T_M.

It is clear that $\cos \phi$ is a constant for any wave coherent with the signal carrier, but that for non-coherent waves, $\cos \phi$ will fluctuate between +1 and −1 in a manner which causes the *average* value of $\cos \phi$ to approach zero as the averaging time is increased:

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \cos \phi_N(t)\, dt = 0. \tag{70}$$

For any non-coherent wave of *constant* frequency $f_N \neq \frac{p}{q} f_s$, the time-average of $\cos \phi(t)$ will be zero over any whole number of time intervals each equal to one period of the *drift* frequency, $\Delta t = 1/|f_s - f_N|$. But for non-coherent waves whose phases drift *randomly* with time, such as Johnson-noise voltages, there is no cyclic pattern of repetition of the changes in $\phi_N(t)$, and hence the time average approaches zero only in the limit when the averaging interval becomes infinitely large.

Of the coherent components which may get through the tuned input buffer, the waves whose frequencies are odd-integer multiples of the switch or carrier frequency are the only ones which contribute anything to the long-time average value of the output from the synchronous rectifier. This can be demonstrated mathematically, but we shall leapfrog over the symbol-shifting toil of a routine analytical proof by making use of one of those maneuvers which learned mathematicians fondly refer to as *ingenious devices.*

We observe that our synchronous switch is really doing something startlingly simple to the input wave-train: *it is multiplying the input wave by a unit square wave symmetrical about the zero axis.* While our electrological readers smile smugly like the Maine farmer watching the big-city tourists picking poison ivy, we also note from Equations (40), (41), and (42) that multiplication of one wavetrain by another leads to a new wavetrain whose spectral frequencies are completely determined by forming all the possible sums and differences of the frequencies in the spectrum of one wave with those in the spectrum of the other. From Equation (55), we see that the square wave of switch polarity contains only frequencies which are odd-integer multiples of the fundamental or signal-carrier frequency:

$$f = f_s,\ 3f_s,\ 5f_s,\ \cdots\ (2k - 1)f_s,\ \cdots,$$

where $k = 1, 2, 3, 4, 5, \cdots$. The input-signal wave contains only 3 frequencies: $f = f_s$, $(f_s + f_M)$, and $(f_s - f_M)$. The output spectrum, therefore, is expressible as

$$f_r = [f_s \pm (2k - 1)f_s],\ [(f_s + f_M) \pm (2k - 1)f_s],\ [(f_s - f_M) \pm (2k - 1)f_s] \tag{71}$$

Since we have assigned no physical significance to negative frequencies, we shall disregard difference terms corresponding to negative values in the expansion of (71) through all integer values of k. In order for the wavetrain whose spectrum of frequencies is described by (71) to have any average (d.c.) value greater than zero, this spectrum must contain at least one frequency equal to zero, because a *steady* value corresponds to alternation at a zero rate. Clearly the *sum*-frequency terms cannot equal zero if we consider only positive values of frequency, so we examine the *difference* terms in (71). Setting the first bracket equal to zero, and choosing the $(-)$ sign, we obtain

$$f_s - (2k - 1)f_s = 0, \quad \text{or} \quad k = 1. \tag{72}$$

This shows that the average value given by Equation (69) is at least partly due to the interaction of the time fundamental of the switch wave with the carrier of the signal wave. Do any other components of different frequencies contribute? Setting either the second or the third bracket of (71) equal to zero, we obtain the result

$$(f_s \pm f_M) - (2k - 1)f_s = 0,$$

or

$$(f_s \pm f_M)/f_s = (2k - 1). \tag{73}$$

Equation (73) says that the only sidebands of the signal carrier which will contribute to the average value of the output of the synchronous switch are those whose frequencies are odd-integer multiples of the fundamental frequency of the switch.

Because it produces sums and differences of the signal and switch frequencies, the synchronous rectifier is often called a *balanced mixer*. To the non-electrologist, however, this name conveys about as much meaning as the words of a beatnik love song.

Anticipating that immortal question of the philosophers, "So what else is new?", we take the next step. We see that what we need now is an averaging device which passes the average value of the switch output undiminished, but eliminates the fluctuations due to the non-coherent noises accompanying the input signal. The low-pass filter following the output-buffer amplifier in Figure 16(a) approximates a true time-averager to a degree which improves as the product RC is made larger, because it passes only d.c. ($f = 0$) with no attenuation, and it attenuates fluctuations at a rate which ultimately approaches 20 decibels per decade when the factor $2\pi fRC$ exceeds unity. But again we are partially trapped by Krausemeyer's Theorem*.

Referring back to Equation (71), we see that setting $k = 1$ in the second bracket yields the result

$$[(f_s + f_M) \pm (2k - 1)f_s]_{k=1} = \begin{cases} 2f_s + f_M, \\ f_M \end{cases} \tag{74}$$

* (with apologies to those who read HUMMINGBIRD): If the liverwurst *can* fall out of the sandwich, it *will* fall out.

The first result is a frequency high enough so that its counterpart in the output spectrum will be heavily attenuated by the low-pass filter. The second result is precisely the modulation frequency of the input signal, and corresponds to a component of the rectifier-output spectrum whose amplitude is proportional to the modulation amplitude of the original signal:

$$(v_r)_M = \frac{m}{2} G_a G_t \frac{2}{\pi} \hat{V}_s \cos \phi \cdot \cos 2\pi f_M t = \frac{m}{\pi} G_a G_t \hat{V}_s \cos \phi \cdot \cos 2\pi f_M t. \tag{75}$$

This is really what we are interested in recovering. It appears as a modulation on the d.c. output level of the buffer amplifier $\bar{g}_b(f)$, and it can be recovered all by itself in the final output voltage $v_o(t)$ if the product RC is chosen properly.

Looking back for a moment, we can see that if the input tuned amplifier has a Q of 10, and if we wish the phase shift of the upper sideband of the signal to be less than 10° (remembering that the center frequency f_o is set equal to f_s), then from Equation (57) we must have

$$\begin{gathered} 10\,[(f_s + f_M)/f_s - f_s/(f_s + f_M)] \leq \tan 10° = 0.1763, \\ (f_s + f_M)/f_s \leq 1.00885, \quad \text{or} \quad f_M \leq 0.00885 f_s. \end{gathered} \tag{76}$$

As a rough rule of thumb, then, let us say that f_M must not be greater than 1% of the carrier frequency if the phase shift of the upper sideband by the tuned amplifier $\bar{g}_o(f)$ is not to exceed 10° when $Q = 10$. Under this condition, we may expand (71), arranging all the positive frequencies (we neglect the negative values) in the rectified output spectrum in ascending order:

$$\begin{aligned} f_r &= 0, f_M, (2f_s - f_M), 2f_s, (2f_s + f_M), (3f_s - f_M), 3f_s, (3f_s + f_M), \cdots \\ f_r &= 0,\ 0.01f_s,\ 1.99f_s,\ 2.00f_s,\ 2.01f_s,\ 2.99f_s,\ 3.00f_s,\ 3.01f_s, \text{ etc.} \end{aligned} \tag{77}$$

The term $0.01f_s$ is the frequency of the modulation envelope we seek to recover from the rectified output spectrum. There will be a negligible attenuation and phase shift of this modulation wave by the low-pass filter of Figure 16(*a*) if we set $2\pi \times 0.01 f_s RC \leq 0.1$, or

$$RC \leq 1.5916/f_s. \tag{78}$$

For this value of filter time-constant the amplitude transmission of the wave corresponding to the third term in Equation (77) is found from (31) by setting $f = 1.99f_s$:

$$\left|\frac{v_o}{v_r}\right| = \frac{1}{\sqrt{1 + 4\pi^2 f^2 R^2 C^2}} = \frac{1}{\sqrt{1 + 396.0}} = 0.05019. \tag{79}$$

Equation (79) tells us that the third member and all higher-frequency terms in the rectified-output spectrum will be reduced in amplitude to 1/20 or less of their values at the input terminals of the filter. The attenuation of the $(2f_s - f_M)$-term is 26 decibels, and for every doubling in frequency thereafter another 6 decibels of attenuation will be added.

The final output then consists only of a steady d.c. level plus a sinusoidal modulation at the frequency f_M, plus random fluctuations due to non-coherent noises whose frequencies lie chiefly within the range $0 \leq f_N \leq 0.1f_s$. In this range the attenuation of the low-pass filter will not exceed 3 decibels, which corresponds to a voltage ratio of 0.7071.

Remembering our previous discussion of the required bandwidth for preservation of a complicated amplitude modulation, such as a square wave, we realize that we cannot make RC indefinitely large to remove *all* the noise, because then our system would not transmit the waveshape of the modulation, which contains our desired information. If the modulation is, in fact, a whole *spectrum* rather than just a simple cosine wave of frequency f_M, our foregoing analysis is still valid, provided that in Equations (76) through (79) *we regard f_M as the highest significant frequency in the Fourier spectrum of the modulation.*

Without further discussion, we shall point out that the final result could be improved by reiteration of the buffer amplifier and low-pass filter in the output stage of the synchronous detector, so that two or more stages are in sequence. This has the effect of increasing the attenuation rate for high frequencies by the factor n, where n is the number of identical stages in sequence, so that the rate becomes $20n$ db/decade. The phase shift, likewise, is multiplied by n, and ultimately approaches $-90n°$. This has the effect of attenuating harmonic variations whose frequencies are *outside* the passband more rapidly than those whose frequencies are *inside* it, although the passband itself is narrowed. Consequently we can use a lower value of RC to achieve the desired equivalent noise bandwidth. From Equations (33) and (35), we may write the general result for n identical low-pass filters in concatenation, with unity-gain, wideband, impedance-buffering amplifiers between successive stages:

$$\beta_{Nn} = 1/4nRC, \text{ HERTZ.} \tag{80}$$

In practice, however, 2 or 3 stages is the maximum number which can be utilized without unwarranted complication of the circuitry.

A filter scheme which is of theoretical interest is the so-called *comb filter*. Let us suppose that we know in advance that the modulation envelope of our signal contains Fourier harmonics whose frequencies are given by kf_o, where k takes on all integer values and f_o is the reciprocal of the longest repetition period discernible in the modulation. Since we wish to preserve these harmonics in the output of the lock-in amplifier, but remove the rectified carrier and all the fluctuations due to the non-coherent noise, let us take an array of Q-tuned filters, each one tuned to one of the harmonic frequencies of interest, with their input terminals in parallel across the output terminals of the synchronous rectifier, and their outputs connected for *summing* of their output voltages. Let us assume that only k such filters are used, and that the Q of each is high enough so that the passband of the 10th filter has negligible overlap with that of the 9th, etc. If the tuning of the filters is done carefully, the harmonics of the signal modulation for orders 1, 2, 3, 4, 5, $\cdots$, up to k. will come through to the final output terminals unchanged in magnitude and phase. Hence the fidelity of the output replica to the input modulation wave depends only on the highest order k to which the summation is taken. The equivalent noise bandwidth of this "comb" is, from (30), never greater than

$$\beta_{NC} = \sum_{1,2,3}^{k} \frac{\pi k f_o}{2Q}. \tag{81}$$

Because k takes on all integer values, this is the sum of an arithmetic progression, and we have

$$\beta_{NC} = \frac{k(k+1)}{2} \cdot \frac{\pi f_o}{2Q} = \frac{k(k+1)\pi f_o}{4Q}. \tag{82}$$

If $k = 10$ and $Q = 20$, we have $\beta_{NC} = 1.375\pi f_o$. By contrast, a single-section low-pass filter having $RC = 0.1/2\pi \times 10f_o$, which would pass the 10th harmonic with negligible attenuation and phase shift, has an equivalent noise bandwidth of $\beta_{N1} = 1/4RC = 2\pi \times 100f_o/4 = 50\pi f_o$. The comb filter is thus a vast improvement over the low-pass filter in *performance*, but its cost, complexity, and cumbersome tuning procedure make its use impractical in most cases.

Just what have we accomplished up to here by synchronous rectification of our signal and noise? Why not just use a *passive* rectifier (i.e. one which is not phase-sensitive) and a low-pass filter? Well, let us take a look.

If a sinewave of constant amplitude and frequency f is rectified by a full-wave, phase-insensitive rectifier, the output wavetrain contains a d.c. level plus a Fourier spectrum of even-integer multiples of f:

$$|\hat{A} \sin 2\pi ft| = \frac{2}{\pi}\hat{A}\left[1 - \sum_{k=2,4,6,8}^{\infty} \frac{2}{(k^2-1)} \cos 2\pi kft\right]. \tag{83}$$

From this we see that every simple harmonic variation in the input wave contributes to the average value of the output. Furthermore, any noise waves which are present make a contribution to the average output value. If any noise having a Gaussian magnitude distribution and zero d.c. value is passively full-wave rectified by an ideal rectifier, the long-term average value of the output from the rectifier is given by

$$AV|v_N(t)| = \sqrt{2/\pi}\ V_N = 0.7979V_N, \tag{84}$$

where V_N represents the r.m.s. value of the noise voltage over whatever bandwidth we are considering. Johnson noise and shot noise, which are white in terms of spectral power density, both have Gaussian distributions of magnitudes. In general, any noise which results from the simultaneous action of many independent random factors will have a Gaussian distribution, even though its spectral power density may not be white.

In particular, if the modulated wave of our previous discussion, Equation (42), is applied to such a passive full-wave rectifier, the contribution made by the upper and lower sidebands to the average output value is totally indistinguishable from that due to the carrier, and is further masked by the short-term random fluctuations in the *total* average value which are caused by the noises accompanying the signal. These fluctuations can be removed to any arbitrary degree by low-pass filtering, but the *waveshape* of the signal modulation is lost to us because each Fourier harmonic thereof appears in the output only as an increment to the d.c. level, plus an infinite train of even-order multiples of the original frequency. The vital amplitude-and-phase-vs.-frequency character of the original Fourier series has been obscured like a glass eye in a candyball jar.

Suppose the modulation of Equation (40) could be made to appear by itself, without any carrier wave. Is the carrier really helping us, or is it more like a chain-link parachute? From (40), the modulation alone is

$$v_M(t) = \hat{V}_s(1 + m \cos 2\pi f_M t). \tag{85}$$

What happens if we merely apply this to a low-pass filter? Who needs a lock-in amplifier? While it is true that jumping off the top of the Empire State Building will get us to the street faster than taking the elevator, the end results are not always identical. Let us look further.

Suppose that our noise-power spectral density is given by the expression

$$w_N(f) = W_N/f, \text{ WATTS/HERTZ}, \tag{86}$$

where W_N is a constant, in WATTS. Theoretically, if we integrate from zero to infinite frequency, we would get infinite power available, a result which is impossible in the real world. In actuality, we can seldom perceive any noise power due to frequencies at the very low end of the scale, because we do not let our measurement system run long enough to sense these slow changes, which manifest themselves as zero-level, or baseline, drift. In practice we provide our measurement system with pre-filtering which cuts off somewhere near $f_M/10$ at the low end and $10f_M$ at the high end, because a bandwidth extending higher or lower simply increases the transmitted noise without any improvement in signal. Therefore, in practical situations, the total input noise power available is finite, and we may take it to be that found by integrating (86) between the limits $0.10f_M$ and $10f_M$. Let us further assume that we set the half-power frequency of our low-pass filter equal to f_M. If the modulation is a complicated wave instead of the simple one in (85), we may regard f_M as the highest significant frequency, and reduce our lower perceptible frequency limit accordingly. In the simple case at hand, we have an available signal bandwidth of $\Delta f = 10f_M - 0.10f_M = 9.9f_M \simeq 10f_M$, before filtering.

The total noise power available from the $1/f$-source between the limits f_1 and f_2 is

$$W_{N12} = \int_{f_1}^{f_2} w_N(f)\, df = W_N \int_{f_1}^{f_2} \frac{df}{f} = W_N \log_e \frac{f_2}{f_1}. \tag{87}$$

If we interpose a low-pass filter between the noise source and the rest of the measurement system, the available noise power at the output terminals of the filter is

$$W_{f12} = \int_{f_1}^{f_2} \frac{W_N\, df}{f(1 + 4\pi^2 f^2 R^2 C^2)} = \frac{W_N}{2}\left[\log_e\left(\frac{f^2}{1 + 4\pi^2 f^2 R^2 C^2}\right)\right]_{f_1}^{f_2},$$

$$W_{f12} = W_N \log_e\left[\frac{f_2}{f_1}\sqrt{\frac{1 + 4\pi^2 f_1^2 R^2 C^2}{1 + 4\pi^2 f_2^2 R^2 C^2}}\right] \tag{88}$$

In this case we have $f_2/f_1 = 100$, and we have set $2\pi f_M RC = 1$, to pass signals at frequency f_M with half power. Since $f_1 = f_M/10$ and $f_2 = 10 f_M$, we have $2\pi f_1 RC = 0.1$ and $2\pi f_2 RC = 10$. For these values, Equations (87) and (88) give us $W_{N12} = W_N \log_e 100 = 4.605 W_N$, and $W_{f12} = W_N \log_e\left[100\sqrt{\frac{1.01}{101}}\right]$, or $W_{f12} = 2.3025 W_N$. Hence the improvement in S/N by low-pass filtering alone is $a_f = \sqrt{W_{N12}/W_{f12}} = \sqrt{2} = 1.4141$. If the noise were white, the improvement under otherwise identical conditions would be

$$a_w = \sqrt{9.9 f_M \times 4RC} = \sqrt{9.9 \times 2/\pi} = 2.510.$$

Thus we see that low-pass filtering alone is more effective in this case with white noise than with $1/f$-noise. This result is not always true, however, and depends in general upon f_2/f_1 and the time constant, RC, of the filter.

Now let us make $v_M(t)$ appear as an amplitude modulation on a carrier wave of frequency f_s, and since we are at liberty to choose f_s arbitrarily (within wide limits), let us set $f_s = 100 f_M$. If we still preserve an input signal bandwidth of $10 f_M$, approximately centered about f_s, the upper cutoff lies at $f_s + 5 f_M$, or $105 f_M$, and the lower at $95 f_M$. By using a lock-in amplifier to detect the amplitude-modulated carrier, with its low-pass filter again set to transmit half power at f_M, we have an available noise power at the input to the synchronous rectifier of $W_{N12} = W_N \log_e (105 f_M/95 f_M) = 0.09985 W_N$. (Note that we have omitted the gain factor of the input-buffer amplifier, because we are interested in evaluating only the effects of synchronous detection.) Comparison of this value to the previous available input noise power shows that the upscale frequency translation effected by the process of causing the signal to appear as an amplitude modulation on a higher-frequency carrier wave, has reduced the available noise power by a factor of 46.12 at the input to the low-pass filter.

In the process of synchronous rectification, the carrier wave is demodulated, and the signal spectrum is translated back *downscale* again by the amount of the carrier frequency (the spectrum is also shifted *upscale* by the amount f_s in the train of *sum* frequencies, out of the synchronous rectifier, but these components are all heavily attenuated by the low-pass filter, and we may neglect them). Negative frequencies, as before, are disregarded in our analysis. The low-pass filter, then, is exposed to a signal spectrum like that given by Equation (77). The noise-power spectrum to which it is exposed is found by imagining the actual spectrum $w_N(f)$ to be cut in two at $f = f_s$, and the portion for $f > f_s$ to be translated back downscale to the origin, as in Figure 17, so that at $f = 0$ the available noise power per unit bandwidth is only W_N/f_s. The new curve, from $f = 0$ to infinity, can be described analytically by setting $f = f_s + f$ in Equation (86):

$$w_N(f + f_s) = W_N/(f + f_s). \tag{89}$$

Since this is of exactly the same mathematical form as (86), integration to find W_{f12} leads to the result

$$W_{f12} = W_N \log_e\left[\frac{f_2 + f_s}{f_1 + f_s}\sqrt{\frac{1 + 4\pi^2 (f_1 + f_s)^2 R^2 C^2}{1 + 4\pi^2 (f_2 + f_s)^2 R^2 C^2}}\right]. \tag{90}$$

Now we have $2\pi RC = 1/f_M$ and $f_2 = 5f_M$ and $f_1 = 0$. Hence we see that

$$W_{f12} = W_N \log_e \left[\frac{105f_M}{100f_M} \sqrt{\frac{1 + (100)^2}{1 + (105)^2}} \right] \simeq 0. \tag{91}$$

The available noise power at the output of the low-pass filter is not exactly zero, but very small. The reason for this is that the lock-in has cut off all that portion of the $w_N(f)$ spectrum below $f = f_s$, and the low-pass filter has cut off the upper end of the noise spectrum so that it is no longer infinite in area included under the curve. Furthermore, we must remember that our *signal* has been cut in half, because only half the signal is carried by the upper sideband of the carrier, so that the result in (91) is not quite as magical as it looks. However, the net effect is a very powerful improvement in S/N, due almost entirely to the frequency-translation capability of the lock-in amplifier. For white noise, the lock-in amplifier offers no essential advantage over straight low-pass filtering, but for any noise of the form K/f^n, the lock-in amplifier will do better by a factor which increases as n becomes larger. It is also superior to the simple filter in any case where the noise-power spectrum has peaks and valleys, because f_s can be chosen to lie in a region of low $w_N(f)$.

For the hecklers in the balcony, we must answer one more important question: If we begin with a signal like that of Equation (85), or something more complicated, how do we make it appear as an amplitude modulation on a carrier wave? The essence of the answer is the same in all cases: we turn the signal on and off with a switch, chopper, or commutating device driven at a constant repetition frequency, f_s, in such a way that this chopper can be synchronized with the driver of the rectifier in the lock-in amplifier. This rectifier-driver circuit is usually called the *reference channel*, and customarily it can be driven by an external synchronizing signal from the chopper, or else the chopper can be driven from an internal oscillator in the reference channel which simultaneously drives the rectifier.

Consider a hypothetical signal, consisting of a unidirectional parameter which is varying with time in a cosinusoidal fashion above and below its long-term average value. If this signal is turned on and off by a square-wave chopper having a duty factor of δ, as shown in Figure 18, the resulting signal at the output terminals of the chopper can be described analytically by an equation of the form of (52), but with the amplitude constant A_{pp} replaced by $a_{pp} = A_{pp}(1 + m \cos 2\pi f_M t)$:

$$f_c(t) = A_{pp}(1 + m \cos 2\pi f_M t) \left[\delta + \sum_{k=1,2,3\cdots}^{\infty} \frac{2}{\pi k} \sin \pi \delta k \cdot \cos \frac{2\pi k t}{T} \right],$$

$$f_c(t) = \delta A_{pp} + \delta A_{pp} m \cos 2\pi f_M t + A_{pp}(1 + m \cos 2\pi f_M t) \sum_{k=1,2,3}^{\infty} \frac{2}{\pi k} \sin \pi k \delta \cdot \cos \frac{2\pi k t}{T} \tag{52a}$$

Capacitive or transformer coupling of the input stage of the lock-in amplifier effectively blocks out the first term of Equation (52a). The tuned amplifier stage rejects the second term, whose frequency is f_M, and all those members of the third-term summation whose harmonic orders k are different from unity, provided that Q is high enough and that the center frequency f_o is set equal to $1/T$, where T is the period of the chopper. It must also be understood that $T \ll 1/f_M$, or else Equation (52a) is not valid, because otherwise the factor A_{pp} in the integrals of (49), (50), and (51), for the Fourier coefficients, will not be a constant over the integration period T.

The end result is that the signal applied to the synchronous rectifier of the lock-in (set to switch at a frequency $f = 1/T$) is of the form

$$f_c'(t) = \frac{2G_o A_{pp}}{\pi} \sin \pi\delta (1 + m \cos 2\pi f_M t) \cos \frac{2\pi t}{T}. \tag{52b}$$

We note from Equation (52b) that if δ is zero or unity, the signal amplitude seen by the synchronous rectifier is zero, and that for $\delta = \frac{1}{2}$ it will have a maximum value. When (52b) is expanded to show the carrier and its sidebands separately, the amplitude of the upper sideband for $\delta = \frac{1}{2}$ is

$$A_u = G_a m A_{pp}/\pi. \tag{52c}$$

Our previous comments about the effect of Q on the attenuation and phase shift of the signal sidebands are in order here also.

If the time-variation of the original unidirectional parameter had been a more complicated function of time, like the sequence of peaks and valleys obtained in taking the optical transmission of a filter as a function of wavelength, where the wavelength is a linear or monotonically changing function of time, we would have a whole Fourier series for the modulation, as in Equations (43) and (44). Our tuned amplifier and low-pass filter would have to be adjusted to pass this spectrum up to some limiting order k, as before.

If we wish to use a direct-coupled input to the lock-in (which is almost never done, however, because of $1/f$-noise), we could eliminate the d.c. component and the modulation-frequency term in Equation (52a) by externally chopping the unidirectional parameter with a synchronous double-pole, double-throw reversing switch just like the one used in the lock-in amplifier. This process is called *inversion*, and it produces a square-wave carrier of zero average value bearing an *envelope* of amplitude modulation which has the shape of the original, but is symmetrical about the zero axis.

Practical interface hardware is available for the commutating function, such as a motor-driven slotted disc for chopping a beam of light in a photometric experiment, which simultaneously generates a synchronized reference signal. In optical spectrometry the light may either be chopped or be commutated by means of rotating or oscillating mirrors. In Hall-effect studies, the carrier is provided by using a sinusoidally varying magnetic field or exciting current, or both. Application details would fill many volumes. One very important precaution must be observed, however: *the commutating device must be located in such a position in the measurement system that it switches on and off only the signal, but not the noise.* Otherwise, the noise appears as part of the modulation spectrum on the carrier, and the lock-in offers no significant advantage over a plain low-pass filter.

With a few refinements, such as a calibrated gain control, variable Q, tunable amplifiers in both signal and reference channels, and a selection of values for RC, with a choice of single- or double-section filtering, the simple system of Figure 16 becomes a practical lock-in amplifier.

THE 2-PHASE LOCK-IN AMPLIFIER

Let us now imagine two identical lock-in amplifiers, with the same signal connected to the input terminals of both instruments, and both reference channels driven from the same synchronizing source, which is phase-locked to the signal carrier. All adjustments of gain, Q, RC, etc. are the same in both instruments, except that the switching phase of the first unit is ϕ_1, and that of the second unit is $\phi_1 - 90°$. From Equation (69), we see that the average values of the final outputs of these two units will be

$$AV(v_{o1}) = \frac{2}{\pi} G_a G_b \hat{V}_s \cos \phi, \tag{92}$$

and

$$AV(v_{o2}) = \frac{2}{\pi} G_a G_b \hat{V}_s \cos (\phi - 90°),$$

$$AV(v_{o2}) = \frac{2}{\pi} G_a G_b \hat{V}_s \sin \phi. \tag{93}$$

We now harken back to the timely words of that ancient Greek swinger, Pythagoras, whose thoughts about squares gave us the famous Pythagorean Theorem. The trigonometric expression of this theorem is

$$\cos^2 \phi + \sin^2 \phi = 1. \tag{94}$$

Squaring both sides of (92) and (93) and adding the resulting equations together, we obtain

$$AV^2(v_{o1}) + AV^2(v_{o2}) = \left(\frac{2}{\pi} G_a G_b \hat{V}_s\right)^2 (\cos^2 \phi + \sin^2 \phi),$$

or

$$AV^2(v_o) = \left(\frac{2}{\pi} G_a G_b \hat{V}_s\right)^2,$$

and hence

$$AV(v_o) = \frac{2}{\pi} G_a G_b \hat{V}_s. \tag{95}$$

So we see that we can make a phase-insensitive system by taking the square root of the sum of the squares of the final outputs of two phase-sensitive detectors whose mixers (rectifiers) are always in time-phase quadrature! This is sometimes as handy as polystyrene pants in the electric chair, for there are situations where the phase of the signal carrier drifts in an unpredictable way relative to the reference. Sonar tracking of moving targets under water produces signals whose phases relative to the outgoing "ping" are highly variable.

Before our subscriber in Deposit, New York writes us to ask how to take the square root of the sum of the squares, we hasten to say it is as simple as incurring a tax liability. Merely connect the cosine output to the X-axis, and the sine output to the Y-axis, of an X-Y recorder whose X and Y gain factors are set equal, and whose pen is positioned at the center of the chart when both inputs are zero. The radial distance of the pen from the zero point, when both signals are applied, is proportional to $AV(v_o)$, given by (95), and the angle between this radius vector and the X-axis is equal to ϕ. If the value of $\hat{V}_s$ is constant while ϕ varies from 0° to 360°, the pen will trace a complete circle about the zero point as a center. If ϕ is constant and $\hat{V}_s$ varies, the pen will travel back and forth along a fixed radial line centered in the zero point, at an angle ϕ with the positive X-axis.

For those who do not have an X-Y recorder, the same result can be achieved by using 2 square-wave, low-noise, electromechanical choppers, with their ON periods overlapping slightly but not completely. One chops the cos ϕ output and applies it to the X-axis of a cathode-ray oscilloscope via an a.c.-coupled amplifier, and the other chops the sin ϕ output and applies it to the Y-axis via an a.c.-coupled amplifier of equal gain. The resulting display on the screen is a rectangle with 4 bright corners, whose diagonal is a measure of $AV(v_o)$, and the angle between this diagonal and the horizontal is ϕ.

The use of two entirely separate lock-ins is not usually necessary here, unless we wish to resolve an inphase component of the signal carrier which is much smaller or much larger than the quadrature. This occurs when ϕ is near zero or 90°. The difficulty with using a single signal channel feeding two rectifiers in quadrature is that a gain setting which gives a measurable output for the small component may cause saturation of the circuitry by the large component. This leads us to make the general observation that saturation of the circuitry in a lock-in amplifier is as undesirable as drinking sarsaparilla in the Pepsi generation, because it leads to cross-modulation of the signal by the noise. Hence all well built lock-ins are provided with overload indicators which show the onset of non-linearity in any stage of the internal circuitry in the direct signal-processing chain up to and including the synchronous rectifier.

When the quadrature and inphase components of the signal carrier are comparable in magnitude, then the whole system can be combined into one box, using a single signal-tuned amplifier and two mixers in quadrature, each followed by its own separate buffer amplifier and low-pass-filter system. Such a device is known as a 2-phase lock-in amplifier.

THE BOXCAR INTEGRATOR

We have seen that a lock-in amplifier is an extremely useful device for recovering *continuous* signals from noise, but what happens if the signal consists of a train of short pulses, separated by relatively long intervals of zero infomation? Such a pulse train is shown in Figure 19(a). If the repetition period of the pulses is constant, this pulse train can be resolved into a Fourier spectrum whose fundamental frequency is the reciprocal of the repetition period. (If the period is not constant, the Fourier-series expansion can be used to describe one pulse over one such period, but it will not define the pulses outside the period chosen for evaluating the harmonic-coefficient integrals.) From Equation (52), for a train of zero-based rectangular pulses, we get a qualitative idea of the variation of harmonic amplitude with harmonic order. We see that as the pulse duty factor, δ, approaches zero, the amplitudes of the fundamental and all the harmonics get smaller and more nearly equal to one another. Since the lock-in amplifier must be tuned to a *single* carrier frequency, its use in this case is made difficult by the fact that any given harmonic has a relatively small amplitude, and hence the gain of the amplifier must be made large in order to recover any usable signal. High gain (up to 10^8) can be achieved in real lock-ins, but it leads to problems such as increase of internal amplifier noise, zero stability, etc.

So we are led to inquire whether there are better ways to recover pulses from noise. It so happens that there are. We shall take a lesson from the deaf man watching television, who turns down his hearing aid during the commercials: we shall simply connect our lock-in to the signal source just before each pulse occurs, and disconnect it shortly afterward. In this way, we can greatly improve S/N by "looking" at the signal only when it is present, and closing our electronic eyes when there is nothing but noise.

This process of switching the input to the amplifier on and off is called *gating*. In a way, this is an unfortunate term, because a closed switch is an open gate, and vice-versa; however, we shall try to make our discussion unambiguous. A variety of highly efficient black boxes has been developed for gating functions in electronic instruments. For now, we shall describe a *gate* only by saying that for our purposes it is an ideal switch whose contacts close when a proper stimulus, or trigger, is applied to its *driving* terminals (isolated from its *gating* terminals), with virtually zero time required to effect the transition from ON to OFF. In the ON position, the switch has zero impedance through its gating terminals, and in the OFF position infinite impedance.

Before rushing out to the local electronic supermarket to buy some gating circuits for our lock-in, think the problem through a little further, and come to the interesting conclusion that if we are going to use a gate before the lock-in, there is really no great need for the *mixer* (synchronous rectifier), which is really just a polarity-reversing gate of duty factor $\delta = \frac{1}{2}$. So we bypass the mixer stage and go directly to the rectifier-buffer amplifier and low-pass filter. But wait a moment. What about the input tuned amplifier? We decide to retain its impedance-buffering properties, but to make its Q very low, so that its gain is constant at G_a from zero (d.c.) to very high frequencies. This means that $\bar{g}_a(f) = G_a + j0$, and no phase-shift through the amplifier will occur at any frequency. Next, because electronic gates are inherently noisy devices, we decide to locate the gate between the input-buffer amplifier and the output buffer, where it will operate on signals amplified to a relatively high level, well above the intrinsic noise level of the gate, and where the not-quite-ideal ON/OFF impedances of the gate will have less significance when the signal source has very high internal impedance. We also decide to shift the output-buffer amplifier, $\bar{g}_b(f)$, to the output side of the low-pass filter, because the latter is now well buffered by the broadbanded input stage on its input side, and needs only output-side impedance buffering to prevent the readout circuitry from loading the capacitor. All that remains now is to add appropriate trigger circuitry, and to provide a means by which the turn-on of the gate can be delayed for an adjustable time interval after the occurrence of the trigger, so that we can allow for triggers which do not coincide with the pulses which they herald. It is clear that we must have one trigger for each pulse, and that the time interval between trigger and pulse must be constant, or else our gated amplifier will suffer from an electronic form of paranoia. Finally, we see that it might be helpful to provide an independent adjustment of the time interval between the triggered turn-on of the gate and the automatic turn-off which follows.

After all this surgery, we stand triumphantly amid a pile of excised parts and contemplate our new creation, shown in block form in Figure 20, which is customarily called a *boxcar integrator*. The name is an exceedingly unhappy choice, for it suggests some monstrous device used in a railroad switchyard. The *boxcar* terminology was first suggested by the fact that each ON/OFF cycle of the gate feeds a rectangular time-pulse of information into the low-pass filter, which behaves as an ideal integrator for frequencies greater than $1/2\pi RC$, thus accounting for the *integrator* nomenclature. There is a feeble connection with boxcars, in that boxcars have rectangular silhouettes. Hopefully, this trend will not spread, for it could lead to such absurdities as calling the classified phone directory the *banana book*, because it has yellow covers. However, the nomenclature is now widely accepted, and we are stuck with it.

Let us now suppose that we can somehow derive a train of trigger pulses having the same frequency of repetition as the pulse train of Figure 19(a), and exactly synchronized with it, as shown in Figure 19(b). We sometimes can use the signal pulses themselves as triggers, but if they are below the noise level, then triggering of the gate will be haphazard and controlled by the noise. We adjust the gate delay, D, and duration, τ, so that each signal pulse lies completely within one gate, ON occurring at the start of the pulse, and OFF exactly at the finish. Clearly this does not alter the time function which describes the pulse, because the gate multiplies the pulse by unity when the pulse function is different from zero, and by zero when the pulse function is zero anyway. Hence the pulse train is transmitted undistorted to the low-pass filter. From Equation (45), the time-function describing the pulse train is

$$v_s(t) = A_o + \sum_{k=1,2,3}^{\infty} \left(\hat{A}_k \cos \frac{2\pi kt}{T_s} + \hat{B}_k \sin \frac{2\pi kt}{T_s} \right).$$

We may re-write this in the more compact form

$$v_s(t) = V_o + \sum_{k=1,2,3}^{\infty} \hat{V}_k \cos \left(\frac{2\pi kt}{T_s} - \theta_k \right), \tag{96}$$

where

$$\hat{V}_k = \sqrt{\hat{A}_k^2 + \hat{B}_k^2}, \tag{97}$$

and

$$\theta_k = \tan^{-1} (\hat{B}_k / \hat{A}_k). \tag{98}$$

Equation (96) describes a spectrum of cosinusoidal voltage waves plus a constant which is the average value of one pulse over the repetition period:

$$V_o = A_o = \frac{1}{T_s} \int_0^{T_s} v_s(t)\, dt = \frac{1}{T_s} \int_0^{\tau} v_s(t)\, dt. \tag{99}$$

The Fourier series of (96) is amplified by the broadband input-buffer stage, which merely puts a constant factor G_a in front of the whole expression. It is then applied to the gate and low-pass filter. Although the *signal* is not altered by the action of the gate, the *impedance of the signal source*, as seen by the low-pass filter, is altered in a periodic fashion from virtually zero during the ON period to infinity during the OFF period.

In order to analyze the effects of this periodic change of apparent source impedance, we shall first note that the charging current to the capacitor must be zero during the OFF periods, and since the input impedance of the output-buffer amplifier is assumed to be infinite, there can be *no change* in the potential difference across the filter capacitor during these OFF intervals. Thus the low-pass filter is held in a state of suspended activity except during the ON periods of the gate.

We shall make certain assumptions to simplify our analysis, which do not alter the general validity of our conclusions about the effects of the gate on the output of the filter. First, let us assume that $\theta_k = 0$ in Equation (96), which is tantamount to saying that the pulse train is an even function of time (i.e. we choose $t = 0$ at the center of any one pulse, and assume that each pulse is symmetrical about its center instant).

Second, we shall assume that the time-constant of the filter, RC, is much larger than the fundamental period, T_s, of the pulse train. This second assumption means that

$$RC/T_s \gg 1. \tag{100}$$

Hence

$$RC/T_k = kRC/T_s \gg 1. \tag{101}$$

Finally, we recall that the duty factor of the gate lies between zero and unity:

$$0 \leq \delta \equiv \tau/T_s \leq 1. \tag{102}$$

Expressions (100) and (102) together imply that

$$\tau/RC \ll 1. \tag{103}$$

Inequality (100) is a reasonable assumption for the normal modes of operation of a boxcar integrator.

Expression (101) means that, if δ were unity, the capacitive reactance of the low-pass filter would be negligible by comparison with the series resistance:

$$|R/X_c| = 2\pi kRC/T_s \gg 1. \tag{104}$$

Under this condition, the charging current to the capacitor is determined only by $v_g(t)$ and the resistance R, and for every signal-voltage harmonic order k, the wave of charging current is in time-phase with the voltage wave, $v_{gk}(t)$. When δ is less than unity, this is still true, but the *net* voltage acting to charge the capacitor is the instantaneous difference between $v_{gk}(t)$ and the voltage $v_{fk}(t)$ (see Figure 20) retained on the capacitor from its previous exposures to the gated harmonic voltage wave $v_{gk}(t)$. By examining the final (steady-state) contribution of each gated harmonic of the signal pulse train to the total capacitor voltage, and summing all these contributions up to the highest significant order k, we can determine the net effect of the gating upon the action of the low-pass filter on the pulse train. We shall consider the effects of noise later.

Figure 19(d) shows the cosine wave of kth order in the Fourier series of the pulse train, if $\theta_k = 0$ in (96) so that a series of only cosine terms results. Between points 1 and 2 of this voltage wave the gate is ON. The first ON period of the gate, with no initial charge on the capacitor, produces a charging current

$$i_{k1}(t) = v_{gk}(t)/R,$$

or

$$i_{k1}(t) = \frac{G_a \hat{V}_k}{R} \cos \frac{2\pi kt}{T_s}. \tag{105}$$

The voltage increment added to the capacitor during the first ON period is given by

$$\Delta V_{fk1} = \frac{1}{C}\int_{-\tau/2}^{\tau/2} i_{k1}(t)\,dt = \frac{G_a \hat{V}_k}{RC}\int_{-\tau/2}^{\tau/2} \cos \frac{2\pi kt}{T_s}\,dt = \frac{G_a \hat{V}_k T_s}{\pi kRC} \sin \pi k\delta. \tag{106}$$

When the gate turns off, this voltage remains on the capacitor until the next instant of turn-on. At this instant of the second turn-on, the capacitor is suddenly connected to a source whose terminal voltage has exactly the same set of values during this ON period as it had during the first ON interval, and whose internal impedance is virtually zero. Two processes now take place: (1) the charge on C accumulated during the first ON period now partially leaks off back through the source, while (2) the source adds more charge to the capacitor. The *net* charging current to the capacitor is the algebraic sum of the charging and discharging currents during this period. In view of Equation (103), we may say that the voltage increment ΔV_{fk1} acquired in the first gate ON interval is negligibly diminished by the discharge current in the second ON interval. (This is valid if $\tau/RC \leq \frac{1}{5}$; at this value less than 0.68% of the previous charge is lost during the next interval.) Therefore, the net charging current during the second ON period is given by

$$i_{k2}(t) = \frac{v_{k1}(t) - \Delta V_{fk1}}{R}. \tag{107}$$

Integration of (107) and division by the capacitance C gives the net voltage increment accumulated during the second ON period:

$$\Delta V_{fk2} = \frac{1}{RC}\int_{-\tau/2}^{\tau/2} v_{k1}(t)\,dt - \frac{1}{RC}\int_{-\tau/2}^{\tau/2} \Delta V_{fk1}\,dt,$$

$$\Delta V_{fk2} = \Delta V_{fk1} - \frac{\tau}{RC}\Delta V_{fk1} = \Delta V_{fk1}\left(1 - \frac{\tau}{RC}\right). \tag{108}$$

Reiteration of this analytical process gives us the expression for the net voltage increment acquired during the nth ON period of the gate:

$$\Delta V_{fkn} = \Delta V_{fk1}(1 - \tau/RC)^{n-1}. \tag{109}$$

The *total* voltage accumulated on the capacitor due to the kth voltage harmonic, after n ON periods is

$$v_{fkn} = \Delta V_{fk1}\sum_{p=1}^{n}(1 - \tau/RC)^{p-1}. \tag{110}$$

The mathematicians in our audience will recognize this as a geometric progression summed through n terms, where the constant is ΔV_{fk1} and the ratio is $(1 - \tau/RC)$. From the well known formula for the sum of n terms of such a progression, we have

$$v_{fkn} = \frac{RC}{\tau}\Delta V_{fk1}[1 - (1 - \tau/RC)^n]. \tag{111}$$

Because $(1 - \tau/RC)$ is less than unity, the progression converges to a finite sum, and the final value of voltage on the capacitor (after a very large number of gate cycles) is

$$V_{fk} = \lim_{n\to\infty} v_{fkn} = \frac{RC}{\tau}\Delta V_{fk1} = G_a\hat{V}_k\frac{\sin \pi k\delta}{\pi k\delta}. \tag{112}$$

If we now evaluate the average value over the ON interval of the kth harmonic of the voltage wave across the terminals 3–5 in Figure 20, we have

$$A V]_\tau v_{ak}(t) = \frac{1}{\tau}\int_{-\tau/2}^{\tau/2} G_a\hat{V}_k\cos\frac{2\pi kt}{T_s}\,dt = \frac{G_a\hat{V}_kT_s}{\pi k\tau}\sin \pi k\frac{\tau}{T_s} = G_a\hat{V}_k\frac{\sin\pi k\delta}{\pi k\delta} = V_{fk}. \tag{113}$$

Thus we see that the long-term (steady-state) contribution of the kth harmonic of the signal-pulse train to the total voltage accumulated on the capacitor of the gated low-pass filter is precisely the *average* value of this harmonic *within any ON interval* of the gate. It should be noted that this result is valid even if $\theta_k \neq 0$ in Equation (96), because the *form* of the integrals in Equations (106) and (113) is not altered by the particular value chosen for θ_k. Our assumption of $\theta_k = 0$ merely made the final expressions more compact. The conclusion expressed in (113) is dependent only on the assumption that the actual filter time-constant RC is much greater than the pulse-repetition period, T_s.

Now we must examine the response of the gated low-pass filter to the constant term in Equation (96). V_o represents a constant voltage existing at all instants of time, and the gating operation sequentially connects it to, and disconnects it from, the filter. If the gate were turned on continuously ($\delta = 1$), then the variation of voltage across the initially uncharged filter capacitor from the first ON instant would be described by the equation

$$v_f(t) = G_aV_o(1 - \epsilon^{-t/RC}). \tag{114}$$

The time required by the capacitor to charge to 99% of the final voltage level is $t_{99} = RC\log_\epsilon(1 - 0.99)^{-1} = RC\log_\epsilon 100 = 4.6052RC$. Thus, after a lapse of 5 time-constants following the first turn-on of the gate, when the gate is ON continuously, the filter capacitor will have reached its final voltage within a deviation of less than 1%.

However, if the gate is switched regularly through the ON/OFF sequence depicted in Figure 19(b), the charge-up of the capacitor will be greatly extended in time as measured by a clock running *continuously* from $t = 0$ at the first moment of turn-on of the gate. If the capacitor is initially uncharged, in the first ON period it will acquire a voltage increment

$$\Delta V_{f1} = G_a V_o(1 - e^{-\tau/RC}). \tag{115}$$

When the gate turns off, the capacitor merely remains charged to this level, because we have postulated an infinite input impedance for the output-buffer amplifier in Figure 20. At the next turn-on of the gate, another voltage increment is added (equal to the first one), because of a positive charging current from the source of $G_a V_o$, and some of the first voltage increment is lost because of a discharging current from the capacitor back through the source. The result is that the *net* voltage increment added during the second turn-on is less than the first. The total voltage on the capacitor at the end of the second ON period is

$$v_{f2} = \Delta V_{f1} + \Delta V_{f2} = G_a V_o(1 - e^{-\tau/RC}) - \Delta V_{f1} e^{-\tau/RC},$$

or

$$\Delta V_{f2} = G_a V_o(1 - e^{-\tau/RC}) - \Delta V_{f1}(1 - e^{-\tau/RC}) = (G_a V_o - \Delta V_{f1})(1 - e^{-\tau/RC}) \tag{116}$$

Equation (116) tells us that the net voltage increment during the second ON period is computed as if the capacitor were *un*-charged and connected to an equivalent source whose voltage is the actual source voltage minus the voltage existing on the capacitor at the end of the previous charging period. Substitution of the value of ΔV_{f1} from (115) into (116) gives us

$$\Delta V_{f2} = [G_a V_o - G_a V_o(1 - e^{-\tau/RC})](1 - e^{-\tau/RC}),$$

or

$$\Delta V_{f2} = G_a V_o e^{-\tau/RC}(1 - e^{-\tau/RC}) = \Delta V_{f1} e^{-\tau/RC}. \tag{117}$$

For the third ON period, we have a net voltage increment, according to (116), of

$$\Delta V_{f3} = (G_a V_o - \Delta V_{f1} - \Delta V_{f2})(1 - e^{-\tau/RC}). \tag{118}$$

Substitution of the values for ΔV_{f1} and ΔV_{f2} from (115) and (116) into (118) gives us

$$\Delta V_{f3} = G_a V_o[1 - (1 - e^{-\tau/RC}) - (1 - e^{-\tau/RC})e^{-\tau/RC}](1 - e^{-\tau/RC}),$$

$$\Delta V_{f3} = G_a V_o(1 - e^{-\tau/RC})[e^{-2\tau/RC}] = \Delta V_{f1}(e^{-\tau/RC})^2. \tag{119}$$

Reiteration of this analysis shows that the net voltage increment during the nth ON period is

$$\Delta V_{fn} = \Delta V_{f1}(e^{-\tau/RC})^{n-1}. \tag{120}$$

The total voltage remaining on the filter capacitor at the end of the nth ON period is

$$v_{fn} = \sum_{p=1}^{n} \Delta V_{fp} = \Delta V_{f1} \sum_{p=1}^{n} (e^{-\tau/RC})^{p-1}. \tag{121}$$

As in Equation (110), this is the sum of n terms of a convergent geometric progression:

$$v_{fn} = \Delta V_{f1} \frac{(1 - e^{-n\tau/RC})}{(1 - e^{-\tau/RC})}.$$

Substitution of ΔV_{f1} from (115) gives us

$$v_{fn} = G_a V_o(1 - e^{-n\tau/RC}). \tag{122}$$

This result takes on the same mathematical form as (111), if we again assume that $\tau/RC \ll 1$, and replace the exponential term by its infinite-series expansion through the first two terms:

$$e^q = 1 + q + q^2/2! + q^3/3! + q^4/4! + \cdots \tag{123}$$

For $\tau/RC \leq 0.2$, we have the exponential adequately described by the sum of the first two terms of (123):

$$(e^{-\tau/RC})^n = (1 - \tau/RC)^n \tag{124}$$

If this is substituted into (122), the result is identical in *form* to (111).

The voltage across the capacitor after a very large number of gate cycles approaches the value

$$V_f = \lim_{n \to \infty} v_{fn} = G_a V_o. \tag{125}$$

Once again, the capacitor charges to the *average value* of this Fourier-series component *over the gate interval*. This is also the *long-term* average of the constant in Equation (96), since it is, in fact, a constant. This result was intuitively evident at the outset, but the derivation was useful, nevertheless, because Equation (122) reveals one very interesting property of the gated low-pass filter, which we shall discuss later.

To summarize what we have learned, the gated low-pass filter acts on every member of the Fourier series describing the pulse train of Figure 19(b) in such a way as to accumulate on the filter capacitor, after a long time lapse, a voltage component which is the average value of that series member over the gate interval. Since the average of the sum of any number of independent functions of time over a given period is equal to sum of their individual averages over this period, *the final net voltage across the filter capacitor* (which is the sum of the contributions from all the members of the Fourier series) *is equal to the average value of the signal pulse over the ON interval of the gate.*

How long does it take for the gated low-pass filter to attain this average? Returning now to Equation (122), we make the substitution $\tau = \delta T_s$:

$$v_{fn} = G_a V_o (1 - e^{-n\delta T_s/RC}). \tag{126}$$

But we see that the elapsed time since the first turn-on, as measured by a continuous clock, is $t = nT_s$. Substitution of this value into (126) gives us

$$v_{fn} = G_a V_o \left(1 - e^{-\frac{t}{RC/\delta}}\right). \tag{127}$$

Comparison of (127) with (114) tells us that an observer measuring only the potential differences across the terminal pairs 3–5 and 6–7 in Figure 20, and ignorant of the presence of a synchronized gate within the black box having 3–5 and 6–7 as its only input and output terminals, would say that the box contained a simple low-pass filter having an *effective* time-constant

$$T_{eff} = RC/\delta. \tag{128}$$

The same conclusion emerges from (111), which describes the response of the gated filter to the kth harmonic of the pulse train, if we make the substitutions

$$\tau/RC = \delta T_s/RC, \tag{129}$$

and

$$(1 - \delta T_s/RC)^n \simeq 1 - (n\delta T_s/RC),\ \delta T_s/RC \ll 1. \tag{130}$$

Equation (111) then becomes

$$v_{fkn} = \frac{RC}{\delta T_s} \Delta V_{fk1} (1 - n\delta T_s/RC). \tag{131}$$

Substitution of ΔV_{fk1} from (106) gives us the final result:

$$v_{fkn} = G_a \Gamma_k \frac{\sin \pi k\delta}{\pi k\delta} \left(1 - \frac{t}{RC/\delta}\right). \tag{132}$$

In summary, we see that, if $T_s/RC \gg 1$, the boxcar integrator is a time-averager of the signal variations over the ON interval of its gate, and that its effective time-constant is the actual time-constant divided by the duty factor of the gate. If our assumption in Equation (100) is not fulfilled, the steady-state output voltage of the gated low-pass filter will contain residual fluctuations due to imperfect filtering of the instantaneous voltage changes occurring along the cosine waves of the pulse harmonics within the ON periods of the gate. Therefore, as $RC/T_s \to 0$, the output voltage of the gated filter becomes a progressively poorer approximation to the true time-average of the signal fluctuations over the ON period.

To gain some insight into how much the actual steady-state capacitor voltage differs from the true time-average of the signal over the gate interval, we make a very fundamental observation about the action of the gate. *Insofar as the variations of capacitor voltage in the gated low-pass filter during the ON periods are concerned, the filter behaves as if it were ungated ($\delta = 1$) and subjected to a recurrent signal waveform whose shape over one cycle is identical to the actual signal within the ON period, and whose frequency of recurrence is $1/\tau$.* This must be so, because the net effect of the gate is simply to disconnect the filter from the signal source during the OFF periods, so that no change of internal conditions can occur until the next ON period, and therefore the initial conditions of any given ON period are identical to the final conditions of the preceding one. Since each ON interval subjects the gated filter to the same time function of signal voltage, and the initial conditions of any ON period are exactly the final conditions of the previous one, the filter output in each ON interval must take on the same time-sequence of values as if the OFF intervals were shrunk to zero and that portion of the signal wave lying within the ON periods were repeatedly applied to the filter with no intervening dead periods.

On this artificial scale of *active* filter time, which we shall designate as t_f, time passes more slowly than on the real time scale:

$$t_f = \delta t. \tag{133}$$

If we now imagine the pulses of Figure 19(a) and 19(b) compressed together along the time axis so that each pulse begins exactly where the previous one ends, the Fourier series of this pulse train on the filter-time scale has the same form as Equation (96), except that the amplitude constants $\hat{V}_k$ will each be $1/\delta$ times as large as those of the real-time waveform, and the harmonic frequencies will each be multiplied by $1/\delta$. This is so because each real-time amplitude is derived from an integral of the form

$$\hat{A}_k = \frac{a}{T_s}\int_{-T_s/2}^{T_s/2} v_s(t)\cos\frac{2\pi kt}{T_s}\,dt. \begin{cases} a = 1 \text{ or } 2 \\ k = 0, 1, 2, 3, \cdots, \end{cases} \tag{134}$$

and when this integral is converted to the filter-time scale, the interval of integration (the repetition period) becomes $\tau = \delta T_s$.

Hence the Fourier series for the filter-time scale becomes, from (96),

$$v_s(t_f) = \frac{V_o G_a}{\delta} + \sum_{k=1,2,3}^{\infty} \frac{G_a \hat{V}_k}{\delta}\cos\left(\frac{2\pi k t_f}{\tau} - \theta_k\right). \tag{135}$$

On this time scale, the amplitude ratio of the transfer function of the filter becomes

$$\left|\frac{v_{fk}(t_f)}{v_{ak}(t_f)}\right| = \frac{1}{\sqrt{1 + \dfrac{4\pi^2 k^2 R^2 C^2}{\tau^2}}} = \frac{\tau}{2\pi k RC}, \tag{136}$$

if $RC \geq \tau$. When the wavetrain described by Equation (135) is applied to the low-pass filter, the output wave is given by

$$v_f(t_f) = \frac{G_a V_o}{\delta} + \sum_{k=1,2,3}^{\infty} \frac{G_a \hat{V}_k \tau}{2\pi k \delta RC}\cos\left(\frac{2\pi k t_f}{\tau} - \theta_k - 90^\circ\right). \tag{137}$$

But $G_a V_o/\delta$ is precisely the average value of the gated signal over the gate ON interval, and $\tau/\delta = T_s$. Hence the filter output is

$$v_f(t_f) = AV]_\tau v_g(t) + \sum_{k=1,2,3}^{\infty} \frac{G_a \hat{V}_k T_s}{2\pi k RC} \cos\left(\frac{2\pi k t_f}{\tau} - \theta_k - 90°\right). \tag{138}$$

The root-mean-square value of all the harmonic fluctuations together is the square root of the sum of the squares of the individual r.m.s. values:

$$V_f = \sqrt{\sum_{k=1,2,3}^{\infty} \left[\frac{1}{\sqrt{2}} \cdot \frac{G_a \hat{V}_k T_s}{2\pi k RC}\right]^2} = \frac{G_a T_s}{2\sqrt{2}\pi RC} \sqrt{\sum_{k=1,2,3}^{\infty} \left(\frac{\hat{V}_k}{k}\right)^2}. \tag{139}$$

Equation (139) shows us that the net r.m.s. voltage across the gated filter capacitor is directly proportional to T_s/RC for any given pulse shape within the ON interval of the gate. Hence, as we assumed in Equation (100), the *real* filter time-constant must be very large relative to the repetition period to suppress residual fluctuations in the steady output voltage of the gated low-pass filter.

But now we must return to the hot potato which we cleverly dropped on the floor at the start of this discussion. What about the noise accompanying the pulse train? The two viewpoints thus far acquired of our low-pass filter are both valid here: (1) the R-C circuit is an *approximate time-domain averager* which averages the noise progressively closer to zero as the averaging time increases (this time is proportional to RC); (2) the R-C circuit is a *frequency-domain filter* having an equivalent noise bandwidth which becomes steadily smaller as RC is increased, passing only zero-frequency (d.c.) signals in the limit at infinite time-constant. From Equations (33) and (128), we have the expression for the equivalent noise bandwidth of the gated low-pass filter:

$$\beta_{Ng} = \frac{\delta}{4RC}. \tag{140}$$

Hence the white-noise-power transmittance of the gated filter is, from (22),

$$W_{Ng} = w_N g_m^2 \delta/4RC = w_N \delta/4RC, \tag{141}$$

since $g_m = 1$. This is simply δ times the transmittance of an ungated filter of the same time-constant:

$$W_{Ng} = \delta W_N. \tag{142}$$

We can readily see that this must be so, because the effect of the gate is to change the available noise power at terminal pair 4–5 from W_N during the ON period to zero during the OFF period, so that the average noise power available over any very large whole number of gate cycles is

$$W_{Ng} = \frac{1}{T_s}[W_N \tau + 0 \times (T_s - \tau)] = \delta W_N. \tag{143}$$

It appears that we have finally found a solid shillelagh to dent the Devil's dome, but Krausemeyer still has his sticky fingers in the sandwich. Unfortunately, the boxcar integrator perceives *changes in the signal* appearing within the ON periods of the gate according to the effective time-constant RC/δ, and so once again the noise bandwidth has been reduced at the expense of the signal bandwidth, so that the *ratio* of signal bandwidth to equivalent noise bandwidth is the same as in the ungated low-pass filter. As always in the battle with noise, we ultimately improve S/N only by being patient and contriving our experiments so that the significant physical changes are spread out over longer periods of time, thus reducing the highest important signal frequency and requiring a smaller signal bandwidth.

So far we have considered only zero-based pulses of one polarity and relatively short duration compared to their repetition period. The boxcar integrator is useful with many other waveforms, however. Consider the one shown in Figure 19(c), and assume the gate width is made very short, while D is such as to place each gate ON period somewhere in the midportion of the signal cycle. Now what happens?

First let us disregard the noise, because we have already studied the effect of the boxcar on the noise. Here we have a *continuous* signal wave with polarity reversals and with no long intervals of zero value. The action of the gate is such that from D to $D + \tau$, in each cycle of the wave, the signal transmitted to the low-pass filter is the actual signal multiplied by G_a and by unity, and for all other instants of time it is multiplied by zero. Therefore, an observer looking only at terminals 4–5 in Figure 20 would see only the narrow pulse train shown shaded in Figure 19(c), and would be unaware that the signal wave had any other shape. As before, the output voltage of the filter across terminals 6–7 will approximate the average value of any one shaded pulse over the duration of this pulse, τ, provided that $RC \gg T_s$. If we denote this average value by $V_A(D, \delta)$, then

$$\lim_{RC/T_s \to \infty} v_f(t) = G_a V_A(D, \delta). \tag{144}$$

In general, this average value will approximate the instantaneous value of the signal at a time t_g midway through the ON period of the gate, and the approximation improves as the ON duration is shortened.

$$\lim_{\delta \to 0} G_a V_A(D, \delta) = v_a(t_g). \tag{145}$$

Flexing our mental muscles a bit, let us now imagine that we set $D = 0$, and $\delta \ll 1$ and wait approximately $5RC/\delta$ for the gated filter capacitor to charge almost to the signal value at this setting which is approximately equal to $V_A(0, \delta)$. Working quickly, we record the value of the output from the boxcar on a d.c. voltmeter, then push a button which directly short-circuits the capacitor while we re-set $D = \tau$. The gate is now retarded in time-phase by exactly one gate width, τ, and it will now let the filter capacitor charge up to the new value $V_A(\tau, \delta)$ after another time lapse of $5RC/\delta$. We repeat this process until we have reached $D = T_s$. Our set of values derived from the output meter can then be plotted vs. the corresponding values of D to give us a replica of the waveform $v_s(t)$, differing by a scale factor G_aG_b, and the small residual error caused by the finite (greater than zero) gate width. If RC is sufficiently large, most of the accompanying noise will have been removed in the process.

If we have nimble fingers, well coordinated with our eyes, we could carry out this process in a total elapsed time just a little greater than

$$t = \frac{5RC}{\delta} \times \frac{T_s}{\tau} = \frac{5RC}{\delta^2}, \tag{146}$$

which is simply the charging time at each delay setting, multiplied by the number of settings required to cover one full signal period with no overlaps or gaps between successive gate positions on the time axis.

Letting our cranial computer run a while longer, we sense that it would be much easier to provide a circuit which automatically *sweeps* the delay D at a constant, adjustable rate through all values from zero to T_s, and to apply the output of the boxcar to the Y-axis of an X-Y recorder whose X-axis is driven by an auxiliary voltage output from the boxcar proportional to the gate delay. The recorder will then plot a replica of the signal waveform which becomes more exact as δ and/or the scan rate, dD/dt, is reduced. If the total scan period is T_D, then (146) suggests that we should set

$$T_D \geq 5RC/\delta^2. \tag{147}$$

We can empirically judge the correctness of our scan rate by increasing T_D on successive scans until the resulting *change* in the plotted waveform becomes negligible for any given value of δ. A procedure which allows the decision to be made more readily is to sweep D from zero to T_s, and then from T_s back to zero. If the two resulting plots on the X-Y recorder coincide with negligible *hysteresis*, then the scan rate is slow enough.

For our mathematical readers, we may describe this delay-scan process as an analog cross-correlation between the periodic signal plus random noise and a train of narrow pulses of unit height and the same period as the signal. As the pulse width approaches zero, and the integration period becomes longer, the mathematical limit of the cross-correlation is precisely the periodic time function, without the accompanying noise. We shall explore the uses of auto- and cross-correlation more fully in a future issue of TEK TALK.

Many uses can be imagined for the boxcar integrator, and we lack the space to list them all. One very interesting application is the detection of the precise instant of a peak or a zero-crossing in a repetitive waveform which exhibits "phase jitter" due either to random noise or to cyclic irregularities in the signal generator. The boxcar's gate is set for a low value of δ, and the delay D is manually varied in the presumed vicinity of the peak or zero until the output meter indicates a relatively steady maximum or zero value with minor residual excursions. In this case, the boxcar acts as a time averager, taking the average of a number of signal repetitions roughly equal to

$$\Lambda = hRC/\delta T_s, \text{ NUMERIC}, \tag{148}$$

where the value of h lies somewhere between 2 and 5.

In summary, we see that the boxcar integrator is really a gated low-pass filter, which takes the approximate gate-time average of whatever passes through its gate, according to an effective time-constant which is the actual RC product divided by the duty factor of the gate. It can be used to measure instantaneous values of periodic waveforms, trace out replicas of these waveforms, or read the gate-time average of a pulse train, all in the presence of rather strong random noise. Unlike the lock-in amplifier, it has the ability to detect any *recurrent* signal, even though its manner of recurrence may be irregular, provided that one trigger pulse precedes each recurrence by the same time advance. It is also useful in many applications where noise is not a problem, but where the unique characteristics of the boxcar allow particular types of measurement to be made.

THE WAVEFORM EDUCTOR[T.M.*]

Like the Neanderthal mechanic who perceived that square wheels on his stone cart would be a vast improvement over the old reliable triangular ones, we suddenly have an inspiration: why waste the time needed for the boxcar integrator to trace a repetitive waveform by a *delay scan* of a *single* narrow gate, when we could use many identical boxcars, each connected to the same signal source and activated by the same trigger train, and each one having the same gate width τ, such that the signal period T_s is just equal to the gate width multiplied by the number of boxcars: $T_s = N_b\tau$. All we have to do is set the delay D of each gate so that the several gates turn on in a regular sequence, the first at $D = 0$, the second at $D = \tau$, the third at $D = 2\tau$, etc. Thus there will be a sample taken *simultaneously* at each of N_b points of the wave on each repetition, and the total elapsed time needed to produce a replica of the waveform is reduced to just $5RC/\delta$. But halfway through the construction of the 58th boxcar, we see that there is a fair amount of redundancy in this scheme, because the input and output circuitry of each boxcar are being utilized only for the fraction δ of the total elapsed time. Why not use just one input buffer, one output buffer, one trigger circuit, one series resistor, and N_b low-pass-filter capacitors, each with its own gate and separate delay adjustment? So we proceed to construct the device shown in block form in Figure 21, having N_b gates and N_b capacitors. Each gate has an ON duration of $\tau = T_s/N_b$ and a delay D_p, such that $D_{p+1} = D_p + \tau$, or $D_{p-1} = D_p - \tau$. Because it is convenient in practice, we provide for a continuous adjustment of T_s, and of the initial delay D_1 between the trigger and the turn-on of the first gate in the sequence.

Now we have created what is called a *WAVEFORM EDUCTOR*, because it has the ability to lead waveforms out of noise. If we apply this device to the waveform of Figure 22(*a*), and set the scan duration (sometimes called the *sweep* duration) just equal to T_s, then a cathode-ray oscilloscope connected to the output terminals in Figure 21 will display a stepwise approximation to the signal wave, as shown in Figure 22(*b*), if

*WAVEFORM EDUCTOR is a registered trademark of Princeton Applied Research Corporation.

the sweep of the oscilloscope is synchronized with the trigger train driving the waveform eductor. From the instant of first turn-on until a steady state output has been attained, the step display on the screen of the oscilloscope will grow in size from zero to a final value proportional to the gate-time average of the signal at each gate position along the waveform. The elapsed time necessary to attain this steady state is $5RC/\delta$, as in the boxcar integrator, and during the transient buildup period the displayed waveform will be a replica whose ordinates are all proportional to their final values by the same factor, $(1-e^{-\delta t/RC})$, provided that the value of capacitance is exactly the same for each channel (one gate plus one capacitor is one channel). To prevent transient distortion of the waveform during the buildup period, it is necessary to provide a real EDUCTOR with trimmer adjustments on the capacitors to compensate for the distributed capacitance due to the wiring in the circuitry, so that each value of C is exactly equal to all the others. Variations in τ could also cause transient distortion, but the binary-clock circuitry used to drive can readily be designed so that τ is a constant for all channels.

The stepwise nature of the output replica is the result of the finite number of individual gates and "memory" capacitors, each of which "sees" the signal for a time greater than zero and takes an approximate average of the range of values covered by the signal during this ON time of the gate. If the number of channels were infinite, then a smooth output curve would result. However, in practice, the number of channels must always be finite, so the stepped nature of the final result cannot be entirely eliminated. But it is possible to smooth out the steps to some extent without employing a very large number of channels. The stepped output replica can be regarded as the sum of the actual signal wave and a sawtoothed wave symmetrical about the zero axis (time axis), whose fundamental period is T_s/N_b, and whose peak-to-peak amplitude is proportional to the slope of the signal wave. If the slope of the signal wave is not constant, the zero-crossings of this sawtooth wave are not exactly uniform in their time spacing, but if the number of channels is large enough, this zero-crossing shift is insignificant. The sawtooth can then be regarded as the sum of a Fourier series having a zero average value and a fundamental frequency of

$$f_{st} = N_b/T_s = 1/\tau, \tag{149}$$

with modulation sidebands symmetrically spaced about each harmonic by an interval of $\pm f_{stm}$, where f_{stm} is the highest significant frequency in the Fourier series representing the *amplitude modulation* of the sawtooth due to the changing slope of the signal wave. If the number of channels is large enough so that the step-replica is reasonably similar to the actual signal wave, then $f_{stm} \ll f_{st}$, and for practical purposes the lowest significant frequency in the sawtooth spectrum is just slightly lower than f_{st}. Under these conditions we can effectively remove the steps from the output wave by inserting a low-pass filter after the readout-buffer amplifier in Figure 21, with a time-constant of

$$T_f = 10/2\pi f_{st} = 10\tau/2\pi = 1.592\tau. \tag{150}$$

This smoothing filter will not seriously distort the fine-structure of the true signal curve except in those channels where the signal exhibits both changes of slope from positive to negative, and changes of amplitude comparable to the average value over the interval τ.

Since τ is always a fixed fraction of the sweep duration, the control which varies T_s can be ganged to that which changes T_f, in such a way that Equation (150) is always fulfilled. If the stepped output presentation is desired, the smoothing filter can always be switched out of the circuit.

Since the first-gate delay, D_1, and the sweep duration, T_s, can be independently adjusted, we can *compress* all the channels of the WAVEFORM EDUCTOR into a time interval $N_b\tau$ which is less than the trigger repetition period T_t and which occurs anywhere between one trigger and the next. Or, we can *expand* the sweep duration to values larger than T_t. This gives the WAVEFORM EDUCTOR a time-resolution capability analogous to the focusing properties of a "zoom" microscope, so that we can examine small portions of the signal wave for details, or take a wide view for coarse features. In practice, the range of attainable values is

$$\tau = 1 \times 10^{-7} \text{ to } 1.0 \text{ second;}$$

$$T_s = 1 \times 10^{-5} \text{ to } 100 \text{ seconds.}$$

(We have here assumed $N_b = 100$ channels.)

Because the sweep duration, T_s, can be made much shorter than the trigger-repetition period, T_t, it is evident that δ, the real-time duty factor ($\delta = \tau/T_t$), can be made very much less than unity, so that the effective time-constant of each channel, $T_{eff} = RC/\delta$, can become very large. As in the case of the boxcar integrator, this has a very powerful effect in reducing the noise which gets through to the output terminals, but it has the same effect on the ability of the EDUCTOR to perceive changes in the shape or amplitude of the signal wave. Practical limits on the operating parameters T_s, δ, and τ are set also by the realizable impedances of the gates and buffer amplifiers, and the internal leakage resistance of the memory capacitors. The readout of signal can be done simultaneously with analysis of the signal plus noise, or it can be done separately with all the channels isolated from the input circuitry and connected only to the output terminals as the gates close and open in their normal sequence. In this "interrogation" of the memory capacitors without any signal replenishment, some of the stored information (electric charge) will be lost each time the readout circuitry scans sequentially across the channels. However, the amnesia per scan can be reduced by good design to a few parts per hundred million. Furthermore, a real WAVEFORM EDUCTOR allows the interrogation to be carried out at a much slower rate than the original analysis, down to the rates required for oscilloscope displays at very low writing speeds, or even to a one-shot readout on an X-Y recorder. Hence, in practice, very long memory durations can be achieved. Auxiliary outputs, such as internally generated triggers and scan ramps of voltage must be provided, of course, for the readout-only mode. Finally, the ultimate in operational convenience is provided by a switch for fast oblivion, so that everything stored in the memories can be erased to allow study of new signals with a minimum of transitional delay.

What improvement in S/N can we expect from this demon-grinder? This question cannot be answered exactly until we have specified both the spectral power density, $w_N(f)$, and the effective input bandwidth, β_{No}, of the noise, as well as the waveform distortion we can tolerate in cases where the signal changes with time. Let us assume white noise, and a stationary signal waveform. From Equation (65), we recall that the improvement in S/N is

$$a = \sqrt{\frac{\beta_{NO}}{\beta_{Ne}}}. \tag{65}$$

Because the WAVEFORM EDUCTOR is essentially a multi-section, sequentially gated low-pass filter, its equivalent bandwidth for real-time, continuous white noise is $\beta_{Ne} = \delta/4RC$. We can estimate β_{No}, the input noise bandwidth, by noting that the WAVEFORM EDUCTOR achieves some pre-reduction of the total noise power transmitted to the memories, by means of a low-pass filter ahead of the broadbanded input-buffer amplifier. The input buffer must have a flat frequency response up to some specified harmonic of the highest trigger-repetition frequency to which the EDUCTOR can respond. However, when the sweep duration is set for longer times than the minimum capability of the instrument, it is senseless to utilize the full bandwidth of the input buffer. Therefore, it is advantageous to gang the time-constant control of the input filter to the sweep-duration control in such a way that the half-power frequency of the pre-filter lies at some value r/T_s, $r \geq 1$, where r is large enough to preserve harmonics of the signal fundamental up to some order k which allows relatively faithful transmission of complex waveshapes. Hence the equivalent noise bandwidth at the input terminals of the input-buffer amplifier is

$$\beta_{NO} = 1/4T_i = \pi r/2T_s. \tag{151}$$

Making the foregoing substitutions for β_{Ne} and β_{NO} in (65), we obtain the following expression for the S/N improvement with white noise:

$$a = \sqrt{\frac{4RC}{\delta} \times \frac{\pi r}{2T_s}} = \sqrt{\frac{2\pi r RC}{\delta T_s}}. \tag{152}$$

In practice, the trigger-repetition period T_t is usually approximately equal to the sweep duration T_s, and, by definition. $\delta T_t = \tau$. Hence

$$a = \sqrt{\frac{2\pi r RC}{\tau}}. \tag{153}$$

Since the charge-up of each memory capacitor is an exponential process, the capacitor reaches full charge for a steady signal only after a theoretically infinite time, but after a total charging time of $5RC$, the level reaches 99.33% of the final value. The number of exposures of the capacitor to the signal necessary to accumulate final charge level is, then, approximately

$$\Lambda = 5RC/\tau. \tag{154}$$

Since each capacitor gets one exposure per sweep, this is also the number of *sweeps* needed to reach substantially final charge on all memories.

If we think of the low-pass filter consisting of one memory capacitor and the common series resistor as an approximate time-averager, we see that whatever averaging is done is essentially finished after Λ sweeps of a signal to which the system is newly exposed, starting from a zero condition in all memories. In this sense, Λ may be regarded as the *number of sweeps averaged* by the WAVEFORM EDUCTOR. Substitution of (154) into (153) gives us

$$a = \sqrt{\frac{2}{5}\pi r \Lambda}. \tag{155}$$

If $r = 10$, which means that the pre-filter is set to pass the 10th signal harmonic with 0.707 amplitude ratio and $-45°$ phase shift, we have

$$a = 3.545\sqrt{\Lambda}.$$

Although Equation (155) expresses a useful point of view, it suggests an inexact description of the WAVEFORM EDUCTOR, which is not a *true* time-averager unless $RC/T_s \gg \infty$. Mathematically, the process of time averaging is one of integration of some function of time, followed by division of the integral by the time of integration. If the average of the same stationary periodic function of time is repeatedly taken over the same time interval, and the results of successive averages are stored in a memory device, each repetition of the process contributes the *same* increment to the total as do all the others. Thus there is an increase in the accumulated information in the memory which is a *linear* function of time if we view this accumulation process on a time-scale long enough to make the discrete increments of individual integrations insignificant by comparison with the total. If the number of repetitions of the averaging process is allowed to increase without limit, then the memory accumulation also increases without limit. However, when a gated low-pass filter is repeatedly exposed to the same periodic stationary waveform, the voltage accumulation on the capacitor is an *exponential* function of time, and the memory increment due to the *nth* exposure is *not* equal to any of the others. From Equation (120), we see that each successive increment is *smaller* than the preceding one by a factor $e^{-\tau/RC}$, and that the final memory accumulation, therefore, *asymptotically* approaches a *finite limit.*

By means of *digital* devices (as opposed to *analog* instruments), it is possible to perform the averaging process in exact mathematical fashion. If we imagine a large number of gate switches, each with an ON duration of τ, being closed sequentially by a master clock circuit so that the signal-plus-noise seen within each gate interval is applied to an analog-digital converter, then we see that digital-computer techniques can be employed to perform true integration of the variations within each gate interval. One of the common techniques is to use a voltage-to-frequency converter so that the instantaneous input voltage is proportional to the instantaneous value of the frequency of an output sinewave having constant amplitude. If the minimum output frequency is high by comparison with $1/\tau$, then an ordinary electronic cycle counter can be used to integrate the output of the voltage-frequency converter, because

$$F_\tau = \int_t^{t+\tau} f dt = h \int_t^{t+\tau} [v_s(t) + v_N(t)]\, dt, \tag{156}$$

where F_τ is the number of cycles passed through by a sinusoidal variation whose instantaneous frequency, f, is h times an instantaneous voltage consisting of signal plus noise.

If each channel of this multi-channel analyzer has one memory device such as a count register, then each memory will receive an integrated increment F_τ for each time the trigger occurs and causes the analysis sequence to be traversed. After Λ sweeps have been made, the averaging process is completed by normalization, or division by Λ, to obtain the time-average for one sweep.

Since the *signal* waveform is stationary, the signal increment in the memory of any given channel on each sweep is always the same:

$$F_{s\tau} = h\tau V_s(D, \tau), \tag{157}$$

where $V_s D, \tau)$ is the time-average value of the signal over the ON interval τ commencing a time D after each trigger. After Λ sweeps, the net signal accumulation in this memory is

$$F_s(\Lambda) = \Lambda F_{s\tau} = \Lambda h\tau V_s(D, \tau). \tag{158}$$

Coherent noise will cause a memory accumulation of the same form as that given by (158), but we shall assume that only random noise is present.

We shall further assume that our random noise has a long-term average value of zero. If it does not, it is a simple matter to block out the average value by a.c. coupling of the input circuitry to the analog-digital converter. We shall assume, finally, that the noise voltage, $v_N(t)$, is *stationary* and *ergodic*. The meaning of these esoteric terms from statistical theory is simply that the averaged results of a very large number of observations made on *this one* noise source over a long period of time are always equal to each other, and to the averaged results of *simultaneous* observations of a very large number of *separate* but *identical* noise sources just like this one.

The gated integrator takes repeated short samples of this random noise over sequential time intervals each of length τ, and for any one memory there will be Λ samples, separated in time from each other by the trigger-repetition period, T_t. For the memory whose gate delay is D, let us denote the instantaneous noise voltage over the *pth* ON period as $v_{Np}(t.)$ Whether or not the noise voltage has a long-term average different from zero, we may regard this sample of noise voltage as being the sum of a steady unidirectional voltage which is the time-average of $v_N(t)$ over this *pth* interval of duration τ, and an alternating voltage whose time-average over this particular interval is zero:

$$v_{Np}(t) = V_{Ndp} + v_{Nap}(t). \tag{159}$$

The contribution of this noise voltage sample to the memory of this channel over the *p*th sampling period is

$$F_{Np} = h \int_{t_p}^{t_p+\tau} v_{Np}(t)\, dt = h\tau V_{Ndp}. \tag{160}$$

After Λ sweeps, the net accumulation of noise in the memory is

$$F_N(\Lambda) = \sum_{p=1}^{\Lambda} F_{Np} = h\tau \sum_{p=1}^{\Lambda} V_{Ndp}. \tag{161}$$

But we see that

$$\sum_{p=1}^{\Lambda} V_{Ndp} = \Lambda V_{Nd}(\Lambda),$$

where $V_{Nd}(\Lambda)$ is the arithmetic average of the Λ separate values of V_{Ndp}. Therefore we have

$$F_N(\Lambda) = h\tau\Lambda V_{Nd}(\Lambda). \tag{162}$$

From the theory of probability and statistics, we now borrow a theorem about the distribution of samples taken from a continuous random variable:

> If $v_N(t)$ is a random variable with an expectation (infinite-time-average value) of V_{Nd} and a variance (infinite-time-average of $v_N^2(t)$ minus the square of the infinite-time-average of $v_N(t)$) of σ_N^2, from which a random set of Λ instantaneous samples is taken, then the expectation of the arithmetic mean of the set of samples is equal to the expectation of $v_N(t)$, and the variance of the arithmetic mean of the sample set is equal to $1/\Lambda$ times the variance of $v_N(t)$.

This theorem says that in our present case,

$$\lim_{\Lambda\to\infty}\left[\frac{1}{\Lambda}\sum_{p=1}^{\Lambda} V_{Nd}(p)\right] = V_{Nd}(\infty) = V_{Nd} = 0, \tag{163}$$

and

$$\lim_{\Lambda\to\infty}\left[\frac{1}{\Lambda}\sum_{p=1}^{\Lambda} V_{Nd}^2(p) - \left\{\frac{1}{\Lambda}\sum_{p=1}^{\Lambda} V_{Nd}(p)\right\}^2\right] = \frac{\sigma_N^2}{\Lambda}. \tag{164}$$

But, by definition, the variance of $v_N(t)$ is

$$\sigma_N^2 = V_N^2 - V_{Nd}^2 = V_N^2, \tag{165}$$

and from (163) we see that the second term in the brackets of (164) is equal to zero. Hence (164) simplifies to

$$\lim_{\Lambda\to\infty}\left[\frac{1}{\Lambda}\sum_{p=1}^{\Lambda} V_{Nd}^2(p)\right] = \frac{V_N^2}{\Lambda}, \tag{166}$$

where V_N is the root-mean-square (r.m.s.) value of the noise voltage over a very long period of time.

From (162) and (163), we see that the noise accumulation in the memory of any one channel after a very large number of sweeps is proportional to the long-term average of the noise itself, which is zero in this case:

$$\lim_{\Lambda\to\infty} F_N(\Lambda) = h\tau\Lambda V_{Nd} = 0. \tag{167}$$

Equation (167) says that if the sampling process is continued indefinitely, the ratio S/N becomes infinitely large. However, in a real signal-sampling device there are practical limitations on the maximum number of samples which can be taken, and during the course of obtaining any finite number of samples there will be short-term fluctuations of $F_N(\Lambda)$ above and below the long-term average of zero. Hence it is more meaningful for us to define S/N in this case as the *average* signal accumulation divided by the *root-mean-square* value of the noise accumulation, after Λ sweeps:

$$\frac{S}{N} \equiv \frac{F_s(\Lambda)/\Lambda}{R.M.S]_\Lambda F_N(\Lambda)} = \frac{h\tau V_s(D,\tau)}{h\tau\sqrt{\dfrac{1}{\Lambda}\sum_{p=1}^{\Lambda} V_{Nd}^2(p)}} = \frac{V_s(D,\tau)}{\sqrt{\dfrac{1}{\Lambda}\sum_{p=1}^{\Lambda} V_{Nd}^2(p)}}. \tag{168}$$

From (166) and (168) we see that as Λ increases

$$\lim_{\Lambda \to \infty} \frac{S}{N} = \frac{V_s(D, \tau)}{V_N} \sqrt{\Lambda}. \tag{169}$$

The theorem from which we derived the result in Equation (169) is based upon the postulate that each of the Λ samples of noise is *stochastically* (randomly) *independent* of all the others. For noise whose bandwidth is limited to β_N, this will be true if the time-lapse between successive samples is at least equal to $1/2\beta_N$. For random noise of any bandwidth, limited or not, irrespective of the interval between samples, Equation (169) is correct, because we have taken the limit as Λ becomes infinite. As an ever-increasing number of samples is accumulated, the later samples must ultimately become randomly independent of the earlier ones. If, however, a finite number of samples is taken, and these are all stochastically independent of one another, then S/N may approach the limit shown in (169) after a small number of samples.

The problem of determining the noise conditions under which this $\sqrt{\Lambda}$ -law holds for small numbers of samples has been investigated by R. R. Ernst in THE REVIEW OF SCIENTIFIC INSTRUMENTS, Volume 36, Number 12, December 1965, p. 1689. He concludes that S/N is proportional to $\sqrt{\Lambda}$, for all values of Λ, only if the spectral power density of the noise is of the type

$$w_N(f) = K/f^n, \; 0 \leq n \leq 1. \tag{170}$$

This includes white noise ($n = 0$) and $1/f$-noise ($n = 1$). For noises of other power spectra the $\sqrt{\Lambda}$-law is a reasonable approximation which becomes better as Λ increases. He also concludes that if the spectral power density increases toward the lower frequencies ($n \neq 0$), a higher final value of S/N will result from using the total available analysis time to make many short averaging sweeps rather than one long averaging sweep. However, if $n = 0$, the noise is white and the final S/N ratio is proportional to the square-root of the *total averaging time*, irrespective of how many sweeps are made. Ernst's conclusions are based upon an analysis of the correlation function of the noise, a topic which we shall examine in future issues of TEK TALK.

The digital multi-channel analyzer is widely known as a *computer of average transients*, from which popular usage has derived the acronym CAT*. The CAT has a voracious appetite for Old Devil Decibel, but it is an extremely complex device in terms of circuitry, and its cost is roughly twice that of its analog counterpart, the WAVEFORM EDUCTOR. Furthermore, it requires counting of the number of sweeps integrated, so that the final signal sequence stored in the memories can be normalized (divided by Λ). Otherwise, the scale factor of the output is unknown, and always greater than unity if more than one sweep has been made. Although, in theory, the CAT continues to improve S/N forever, in practice it is limited by the long-term stability of the gain of its input amplifiers, the linearity of its analog-digital converters, and the zero stability of its output circuitry. It does offer the advantages of truly permanent memories and the availability of binary-coded output data for further digital-computer processing. It is limited, however, to a minimum time per channel of about 30 microseconds, whereas the waveform eductor is capable of operation with τ at 1 microsecond or lower. Furthermore, the total time required to sweep through all channels is completely used for analysis in the WAVEFORM EDUCTOR, while the CAT uses only a fraction of this time for actual integration, because of digitizing and other internal operations which must be done separately in each channel. Within the range where their operating parameters overlap, they give comparable results in terms of S/N improvement.

CHOOSE YOUR WEAPON

And so, like the swimmer holding the crocodile's mouth shut with both hands, we come to the inevitable moment of decision. Which device is best for our purposes? Though partisans may argue this question the

*CAT is a registered trademark of the Technical Measurements Corporation.

way Civil-War buffs debate the effect of Grant's favorite whiskey on the battle of Shiloh, we can offer a few simple guidelines. The primary criterion is the period T_M, required for the signal to pass through one complete set of values. A secondary criterion is the waveshape of the signal, and we shall assume in the following discussion that we must preserve *shape* information in the time-wave of the signal, as well as mere *amplitude* information. Other criteria, such as the noise-power spectrum and the time stability of the signal source, may be important. We offer the following general suggestions:

FIRST If the signal gamut is naturally non-recurrent (i.e. the experimenter must excite the system once for each sweep or once for each point by causing some independent variable to pass through a range of values, such as the wavelength of a monochromator used in optical transmission studies), and the total time required for the signal to cover the desired range of values has an *irreducible minimum value* of 100 seconds or longer, then chop the signal and use a lock-in amplifier. Choose the chopping frequency f_s so that the highest significant modulation-harmonic frequency in the Fourier series describing the signal over the period T_M is very much lower than f_s. If the 10th Fourier harmonic must be preserved. then Expression (76) suggests that for an input Q of 10, the signal frequency should be 100 times the 10th harmonic of the fundamental modulation frequency, or $f_s \geq 100 \times 10/T_M = 1000/T_M$. For example, assume that $T_M = 100$ seconds, and that the input-noise bandwidth is 100 hertz on either side of f_s. Then we must have $f_s \geq 10$ hertz. If we use two buffered, concatenated, identical low-pass filters each having a time-constant RC then we must have $RC \leq 0.1T_M/20\pi$ in order to pass the 10th harmonic of the modulation with negligible attenuation and phase shift, or $RC = 0.1592$ second. The output noise bandwidth will be $\beta_{N2} = 1/8RC = 0.7852$ hertz at this value of RC, and for white noise, the S/N improvement is therefore, $a = \sqrt{100/0.7852} = \sqrt{127.36} = 11.28$. In order to achieve this improvement by true repetitive time-averaging in a multi-channel analyzer, the *minimum total elapsed time* would be 128 × 100 seconds, or 213 minutes! If the sweep time T_M were increased to 213 minutes, and the filter-time-constant of the lock-in amplifier increased accordingly, we could set $RC = 204$ seconds, giving $\beta_{N2} = 6.127 \times 10^{-4}$ hertz, and obtain an improvement of $a = \sqrt{100/6.127 \times 10^{-4}} = 404.0$!

If the noise is non-white, and the spectral power density increases with declining frequency, an even greater improvement results because of the upscale-frequency-translation-and-demodulation characteristics of the lock-in amplifier, as we have previously demonstrated. *This upscale noise-power reduction with $1/f^n$-type spectra more than offsets the disadvantage stated by Ernst for one long averaging period vs. many short ones of the same total duration, since the carrier frequency can nearly always be made high enough to reduce the available noise power by one or more orders of magnitude (factors of 10).*

SECOND If the signal gamut is either automatically recurrent or can be externally triggered to recur, with a cycle duration of T_M which has a *maximum possible value* of 0.1 second or less, then the lock-in amplifier is a poor choice. For measurement of isolated peak values or zero crossings in the signal gamut, use the boxcar integrator with a narrow gate width and an RC product adequate for the existing noise. In this case the gamut period becomes equal to the time-base width: $T_M = T_s = 1/f_s$. If the *complete waveshape* must be recovered, use the WAVEFORM EDUCTOR and set $T_M = T_s$, the sweep duration. If one trigger occurs for each repetition of the signal gamut, then $T_t = T_s$, and the gate-duty factor is $\delta = \tau/T_s$ in either case. Do not attempt the use of extremely large RC products with low values of δ, however, because the effective time-constant of each memory is RC/δ, and it requires a lapse of $5RC/\delta$ to attain a steady signal level in each memory. If this required time lapse becomes very large (in the order of *days*), the internal and external leakage decrements of the memory capacitors during the OFF intervals may become comparable to the information increments during the ON periods.

Suppose that $T_M = T_s = 0.1$ second, $\delta = 0.01$, $RC = 1$ second, and the noise is white and externally band-limited to an upper cutoff frequency of 200 hertz (which would let the 10th harmonic of the signal-repetition frequency pass without serious distortion). The equivalent noise bandwidth at the output of the gated low-pass filter is $\beta_{N_g} = \delta/4RC$, or 0.0025 hertz. The S/N improvement is, therefore, $a = \sqrt{200/0.0025}$, or 282.8. The lock-in amplifier under similar conditions would require a minimum carrier (signal) frequency of $f_s = 100/T_M = 1000$ hertz, and the filter time-constant would have to be such that $RC = 0.1T_M/2\pi \times 10 = 0.0001906$ second, giving an equivalent noise bandwidth for a 2-section filter of $\beta_{N2} = 1/8RC = 655.8$ hertz. Since we have already assumed the noise was pre-limited to 100 herz either side of f_s (200 hertz total bandwidth), the lock-in gives no further improvement in S/N. By increasing RC to 0.001906 second, and accepting some phase shift at the 10th modulation harmonic, we could reduce β_{N2} to 65.58 hertz and achieve a gain in S/N of $a = \sqrt{100/65.58} = 1.235$. This is scarcely a startling improvement.

If the noise has a $1/f$ type spectrum of power density, the boxcar or WAVEFORM EDUCTOR could be preceded by a band-limiting black box passing noise only within the range of, let us say, $0.1/T_M$ to $10/T_M$. From (88), the available input noise power would then be $W_{N12} = W_N \log_e 100$, or $W_{N12} = 4.605\ W_N$. The available output noise power from the gated low-pass filter is, then,

$$W_{f12} = W_N \log_e \left[100 \sqrt{\frac{1 + 4\pi^2 \times 1^2 \times 100^2}{1 + 4\pi^2 \times 100^2 \times 100^2}} \right] \simeq 0.$$

The improvement here is more spectacular than it was for white noise, because RC/δ is very large by comparison with the reciprocal of the low-frequency limit of the noise band. This is feasible in this case because we are not concerned with *changes* of signal waveshape, and hence are free to use a relatively large effective time-constant in our gated filter(s).

If the signal modulation were a pure sinewave of constant amplitude and frequency $f_M = 1/T_M$, we could use the lock-in amplifier to somewhat better advantage than with a complicated waveshape, because the filter time-constant could be made larger. We would still have to make the carrier frequency much higher than f_M, as previously pointed out, in order to preserve the sinusoidal modulation on the d.c. output level. At the maximum permissible filter time-constant which still preserves the modulation at the output terminals, the lock-in would have a significantly poorer S/N ratio than the boxcar or WAVEFORM EDUCTOR.

THIRD If T_M lies between 0.1 second and 100 seconds, with either driven or naturally recurrent sweeps of the signal through its gamut of values, all three instruments are feasible to consider, and the choice must be based upon criteria other than sweep time alone. For example, in magnetic-resonance spectroscopy, a lock-in amplifier inherently gives an approximation to the *dS/dB derivative* of the absorption-mode resonance peaks when the air-gap magnetic field at the sample is audio-frequency-modulated in the customary manner about the linearly swept average value of flux density, B. To obtain the absorption-mode spectrum, the output of the lock-in must be integrated electronically or graphically, which may be inconvenient for the experimenter. The WAVEFORM EDUCTOR, on the other hand, with a suitable r.f. detection stage, can avoid the need for audio modulation of the magnetic field and recover the resonance-signal variations directly if B is a monotonically increasing or decreasing function of time. Furthermore, in this case, it may not be necessary to wait for the full $5RC/\delta$ needed for attainment of a steady memory level if the experimenter is content with a crude but early recognition of the important peaks in the spectrum, especially with regard to the values of B at which these peaks occur.

We realize all too well that there is no simple rule for making your choice of weapons to bash Old Devil Decibel, and that measurement techniques are as varied as the number of ways an Italian chef can serve spaghetti with sauce. But fortunately there is a handy, free accessory for each of these instruments, which greatly simplifies the decision process. We call it a technical-applications engineer. Our technical personnel have a wide background of experience in literally hundreds of different problems requiring extraction of signals from noise. This experience, as well as a free, in-your-lab evaluation of a lock-in, boxcar integrator, or WAVEFORM EDUCTOR, is always at your disposal on short notice. So why delay? Put your finger into that analog-digital converter known as the dial on your telephone, and send a few decibels to our end of the line. That's one kind of noise we *like* to hear!

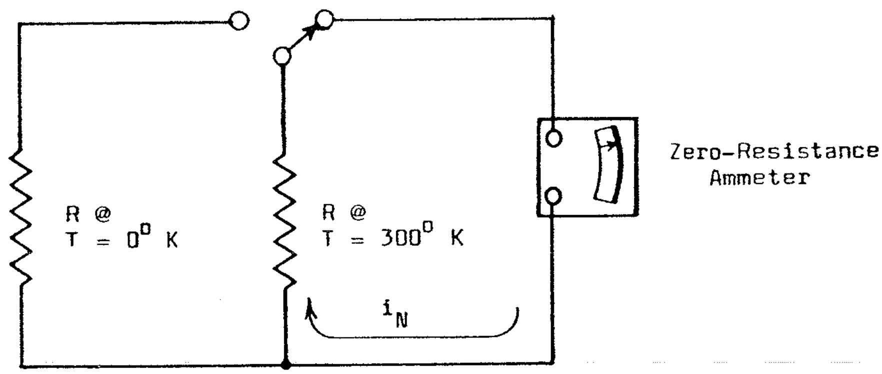


FIGURE 1(a) - Cold Resistor and Identical Warm Resistor in a Circuit.

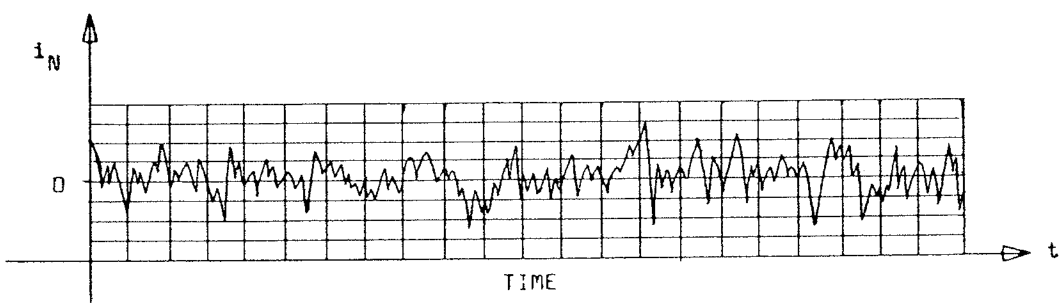


FIGURE 1(b) - Time-Variation of Instantaneous Noise Current in Short-Circuited Resistor at T = 300° Kelvin.

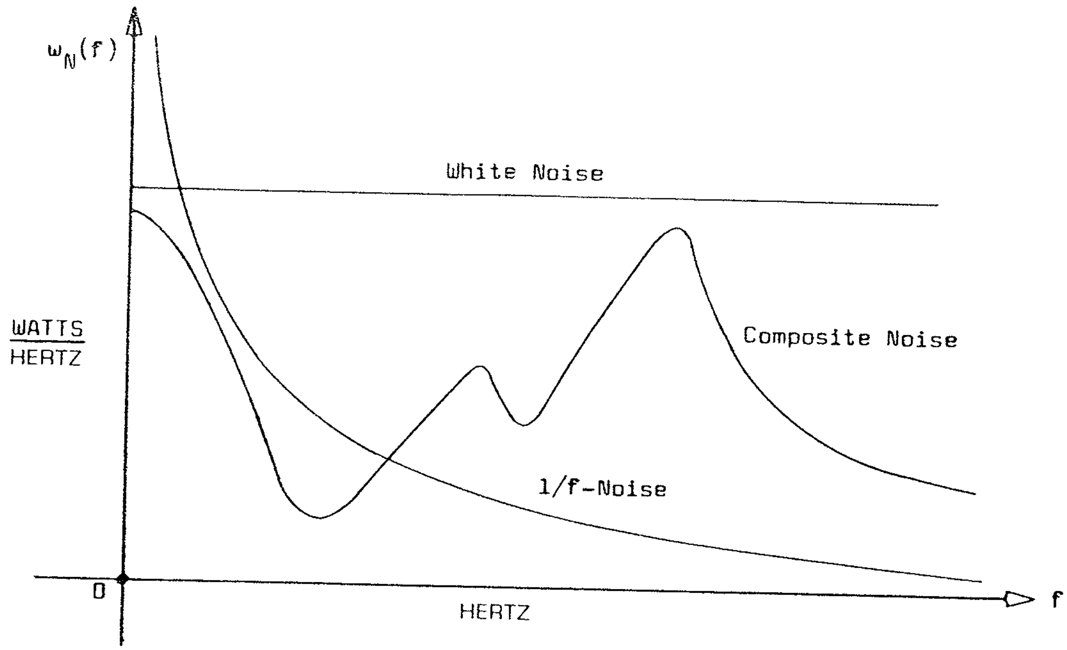
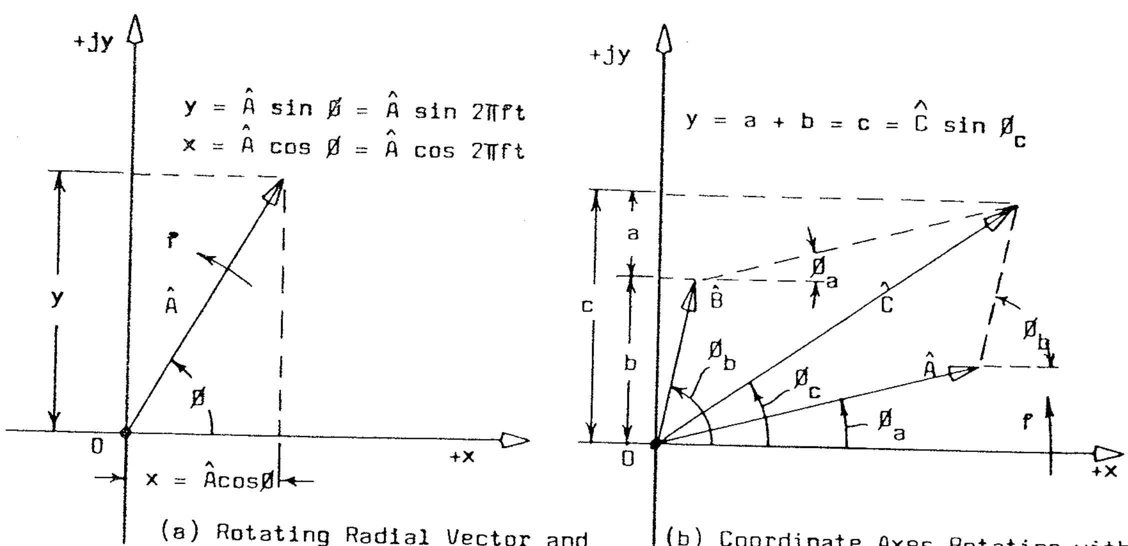


FIGURE 2 - Power-Density Spectra of Various Noises.

FIGURE 3 - Generation of Sine & Cosine Functions by Rotating Radial Vectors (Phasors) in a Plane.



(a) Rotating Radial Vector and Stationary Coordinate Axes, Generating Simple-Harmonic Functions of Time.

(b) Coordinate Axes Rotating with Radial Vectors, for Addition of Phasors to Find Phasor Generating the Sum of Two Sine Functions.

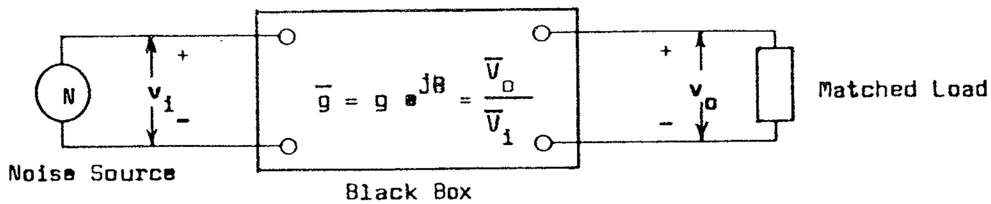


FIGURE 4 - Frequency-Selective Network Interposed Between a Noise Source and a Matched Load to Limit the Noise Power Transmitted to the Load.

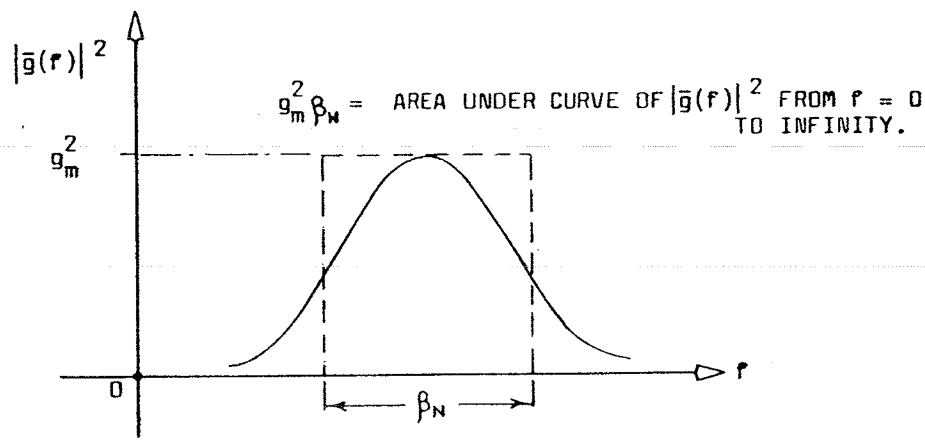


FIGURE 5 - Equivalent Noise Bandwidth of Network in Black Box

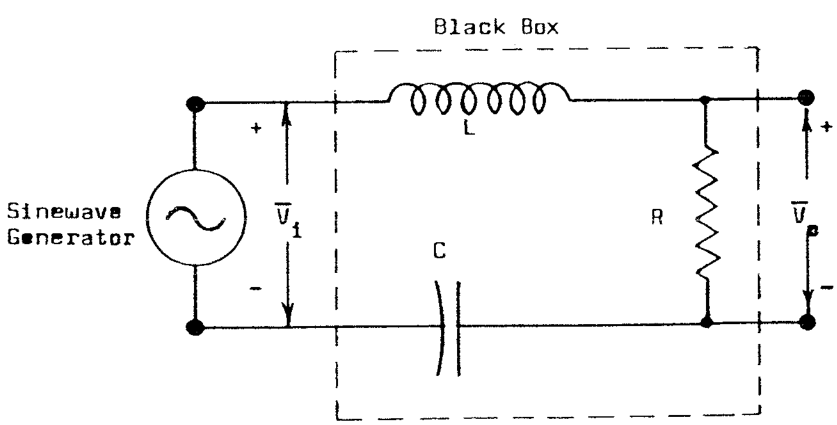


FIGURE 6 - Series R-L-C Circuit in Black Box, With Sinusoidal Excitation.

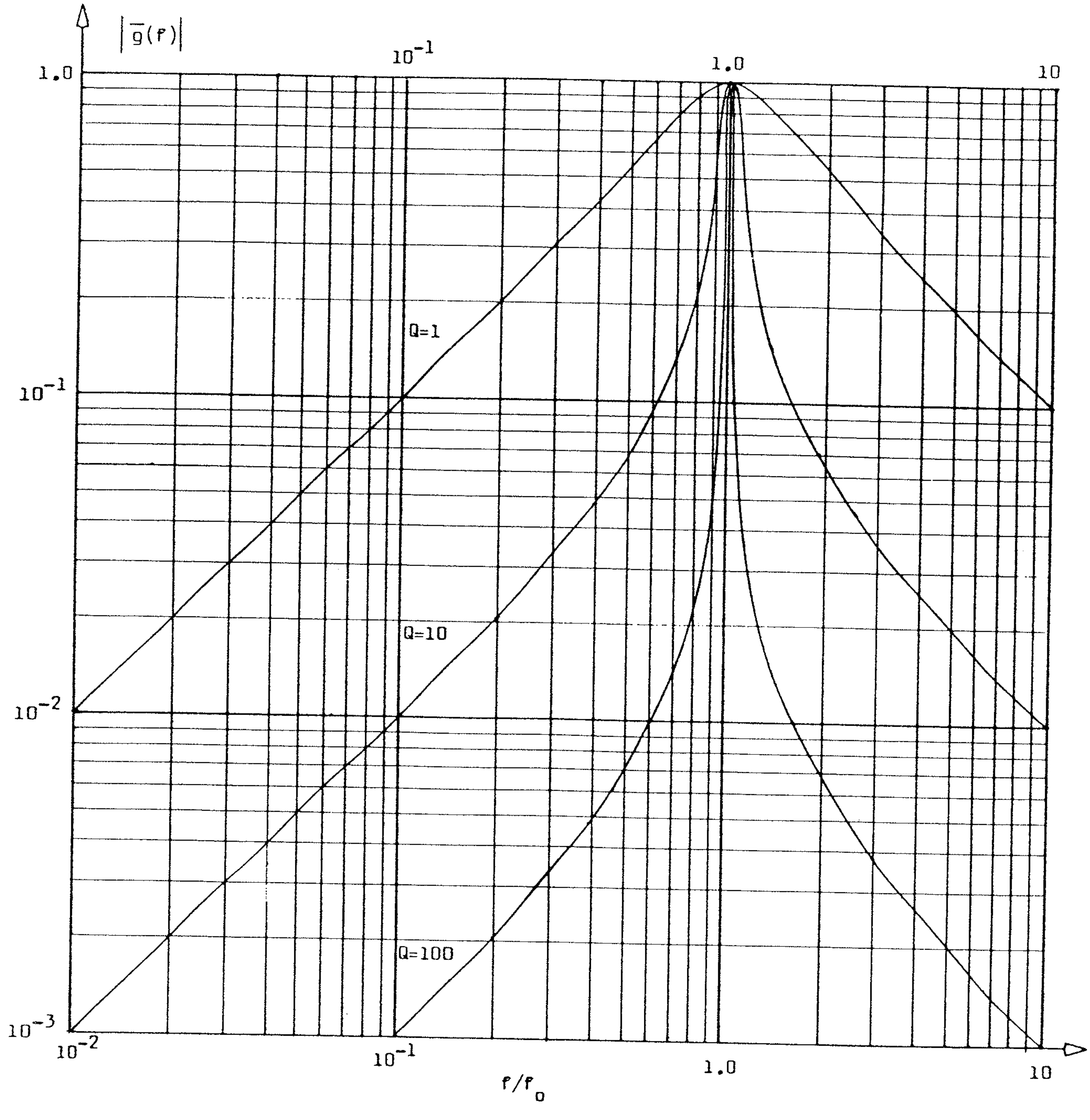
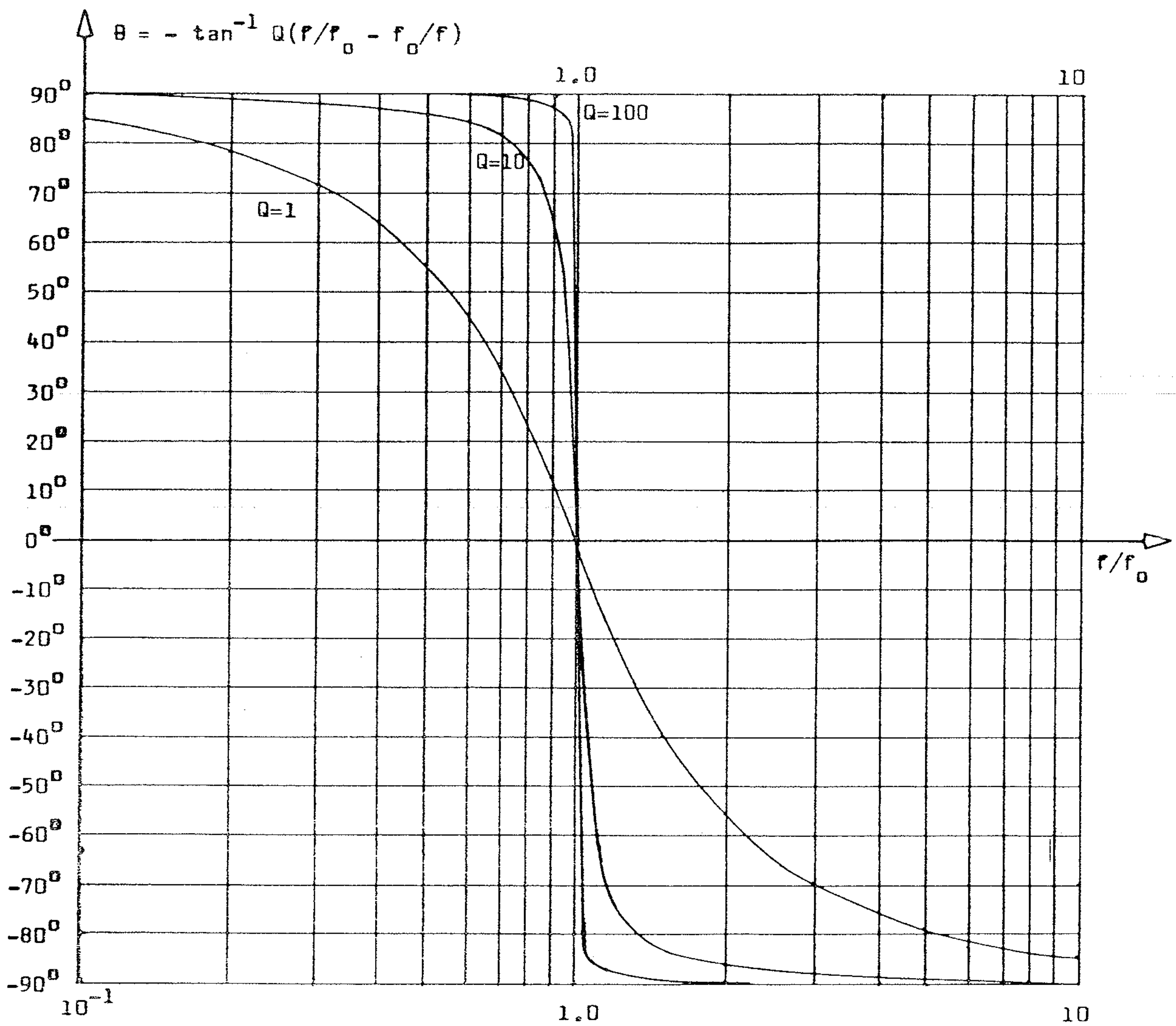


FIGURE 7(a) - Amplitude Characteristic of Transfer Function of Series R-L-C Circuit in Figure 6.



<u>FIGURE 7(b)</u> – Phase Characteristic of Series R-L-C Circuit.

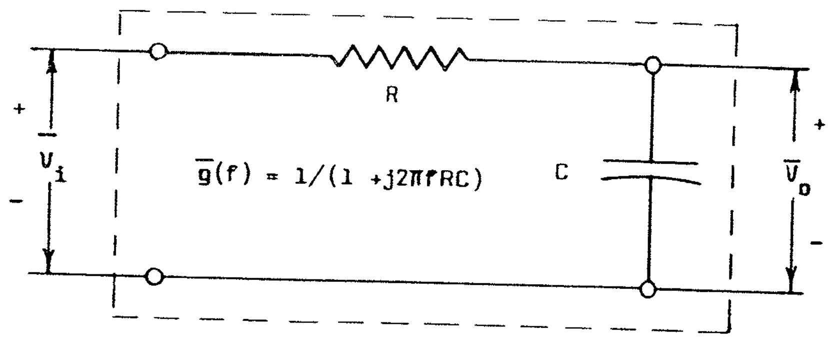


FIGURE 8 - Low-Pass Filter.

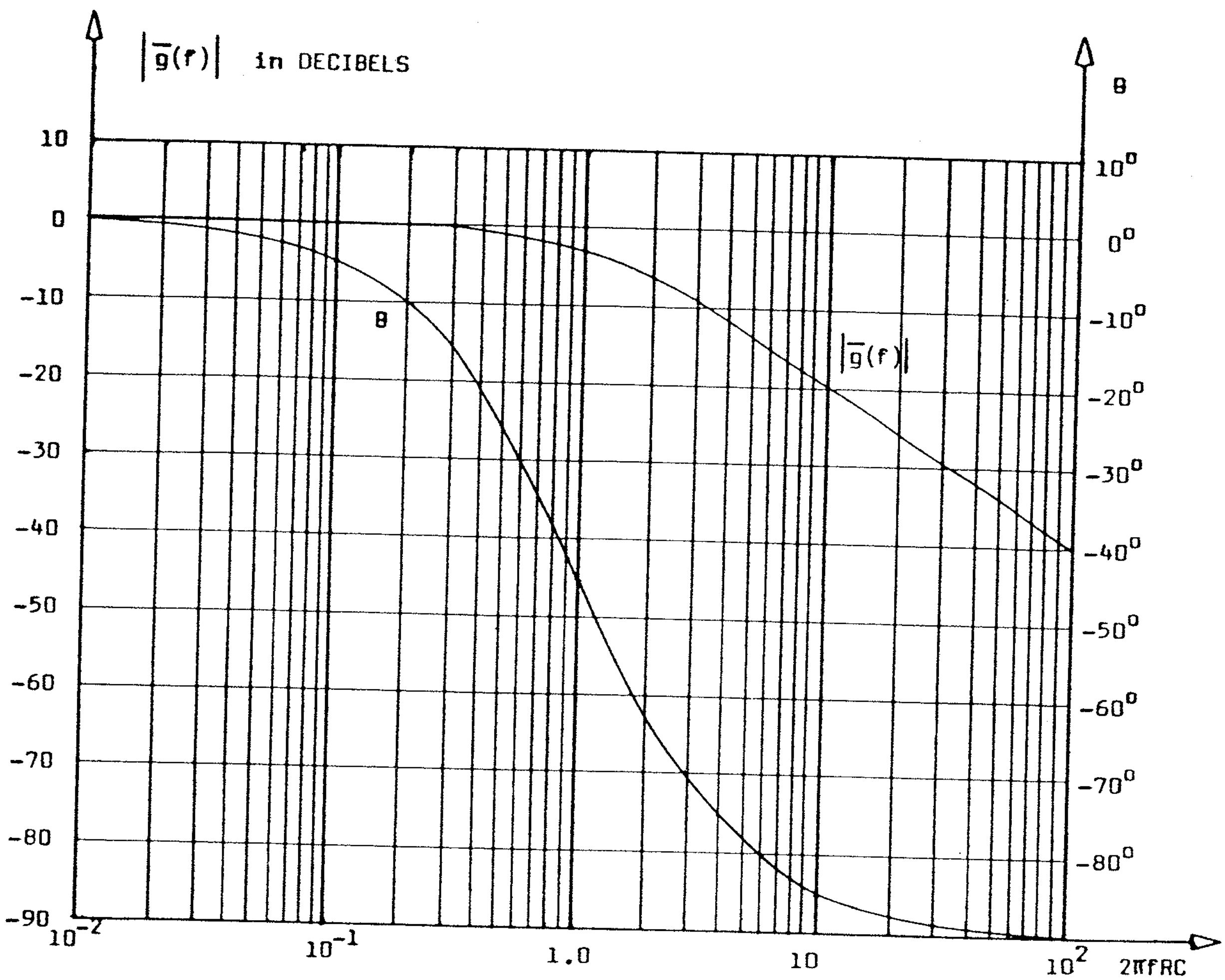


FIGURE 9 - Amplitude & Phase Characteristics of Transfer Function of Low-Pass Filter.

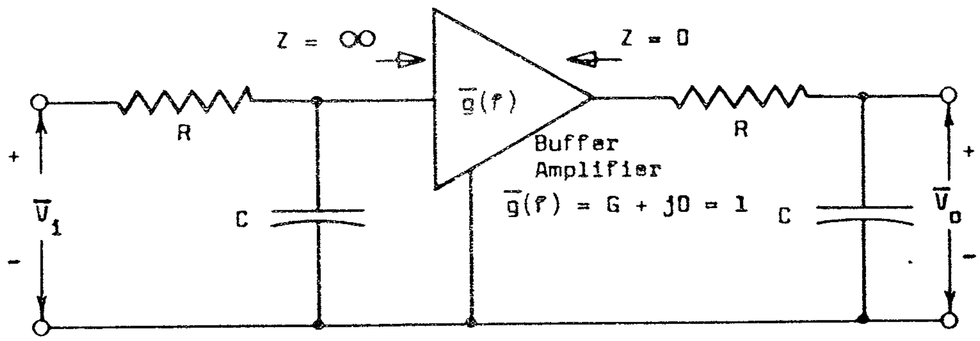


FIGURE 10 - Concatenated Identical Low-Pass Filters with Intermediate Impedance-Buffering Amplifier.

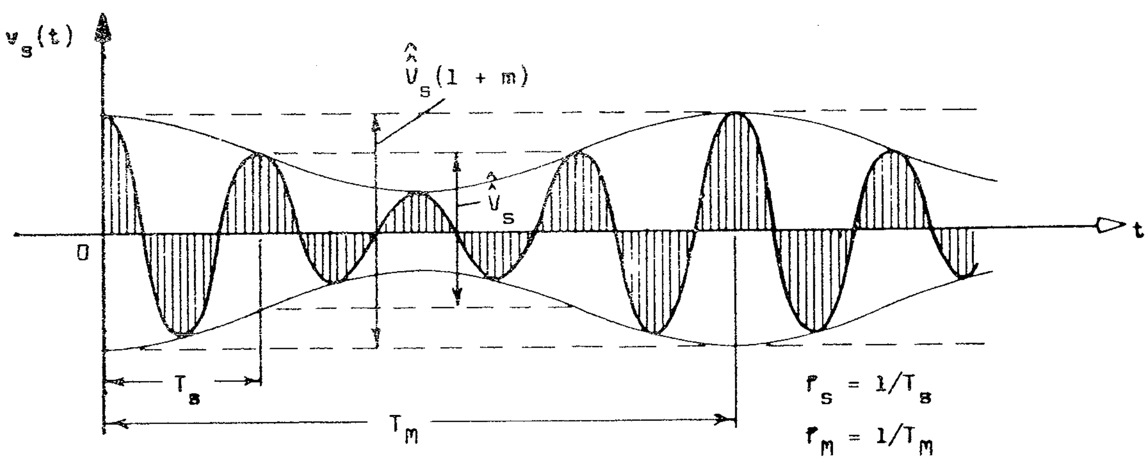


FIGURE 11 - Cosine Wave of Voltage Amplitude-Modulated by a Cosine Wave of Lower Frequency.

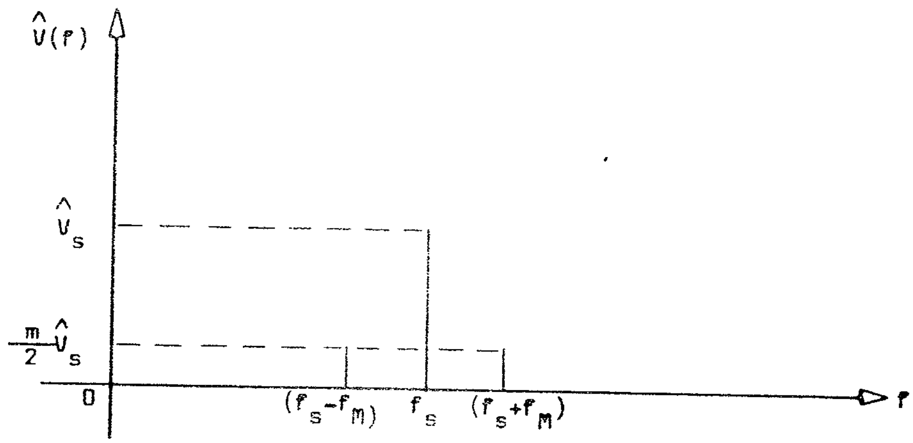


FIGURE 12 - Spectrum of Amplitude-Modulated Wave in Figure 11.

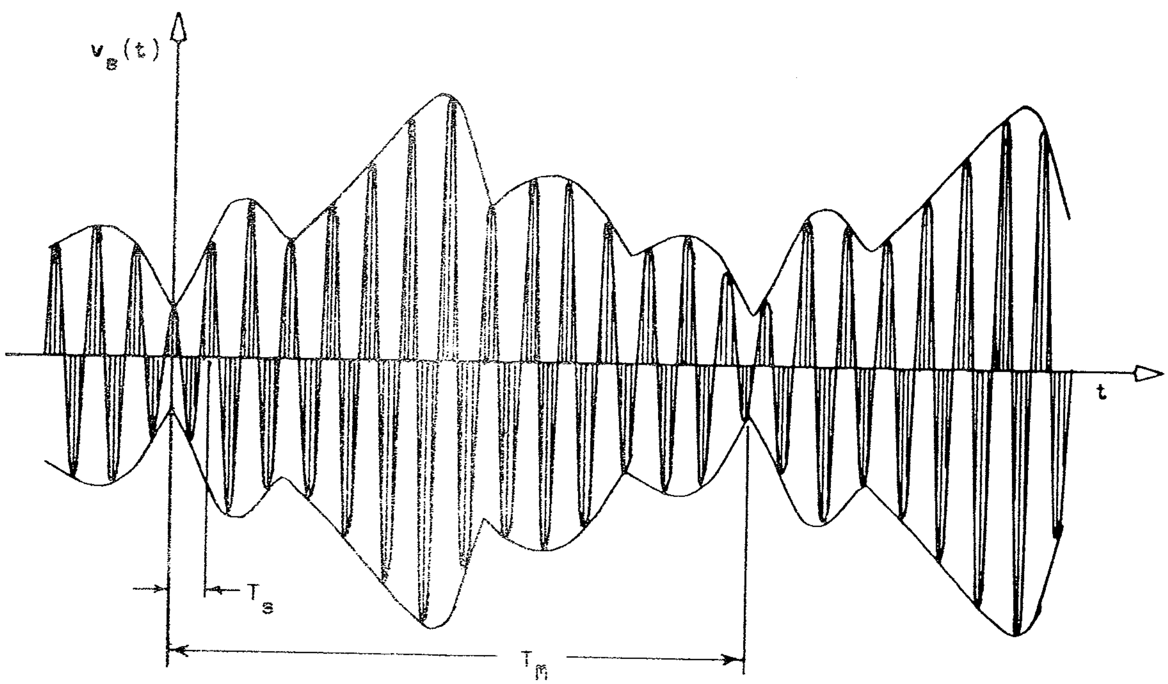


FIGURE 13 - Cosine Wave of Voltage with Complicated Amplitude Modulation.

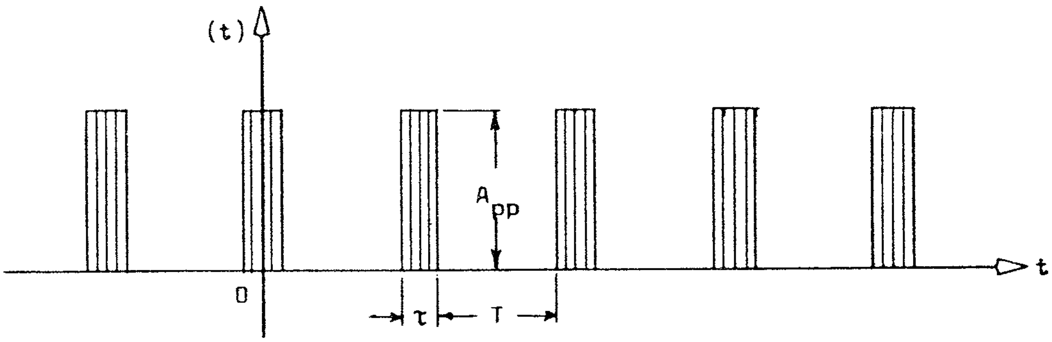
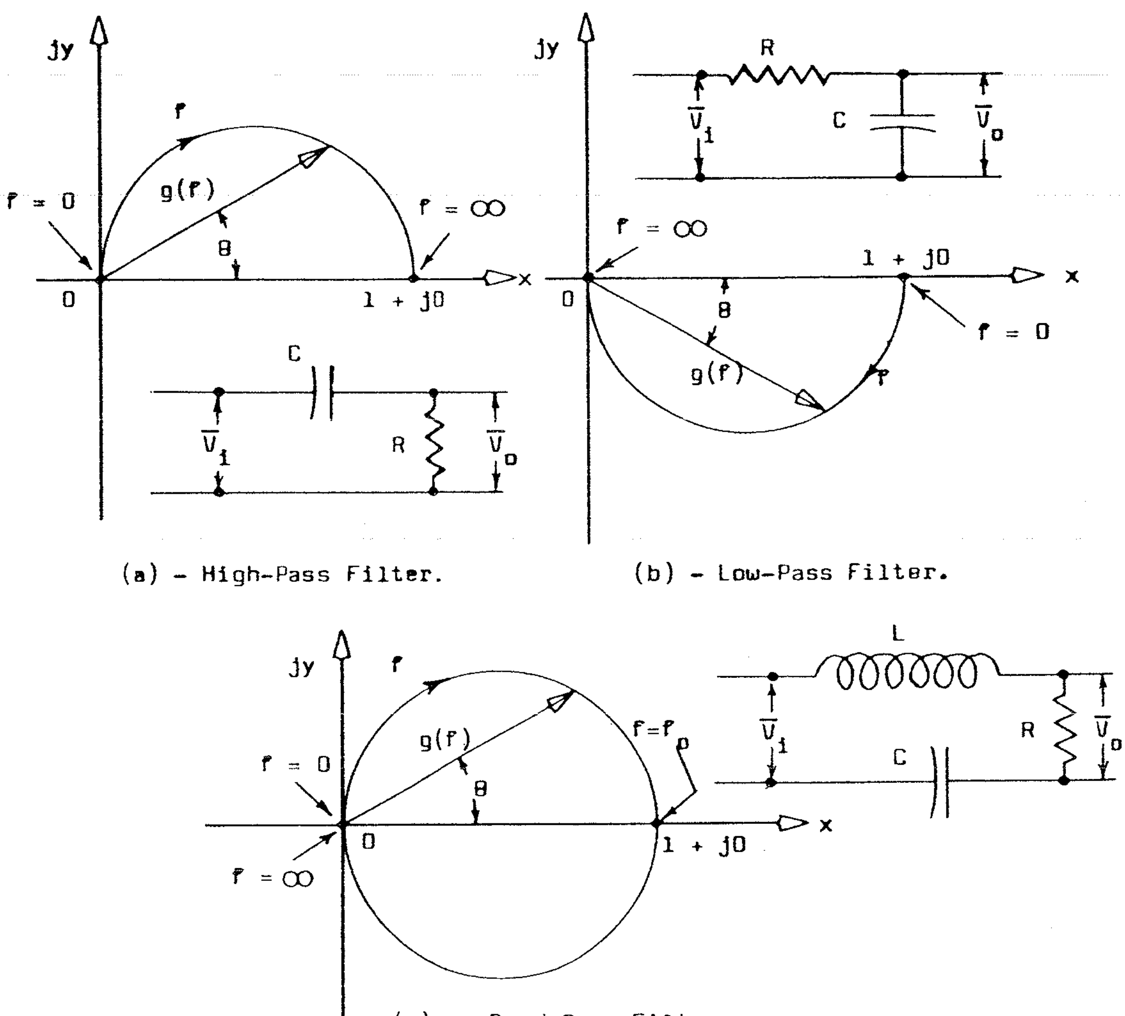


FIGURE 14 - Zero-Based Train of Rectangular Pulses.

FIGURE 15 - Transfer-Function Loci of Common Filter Circuits.



(c) - Band-Pass Filter.

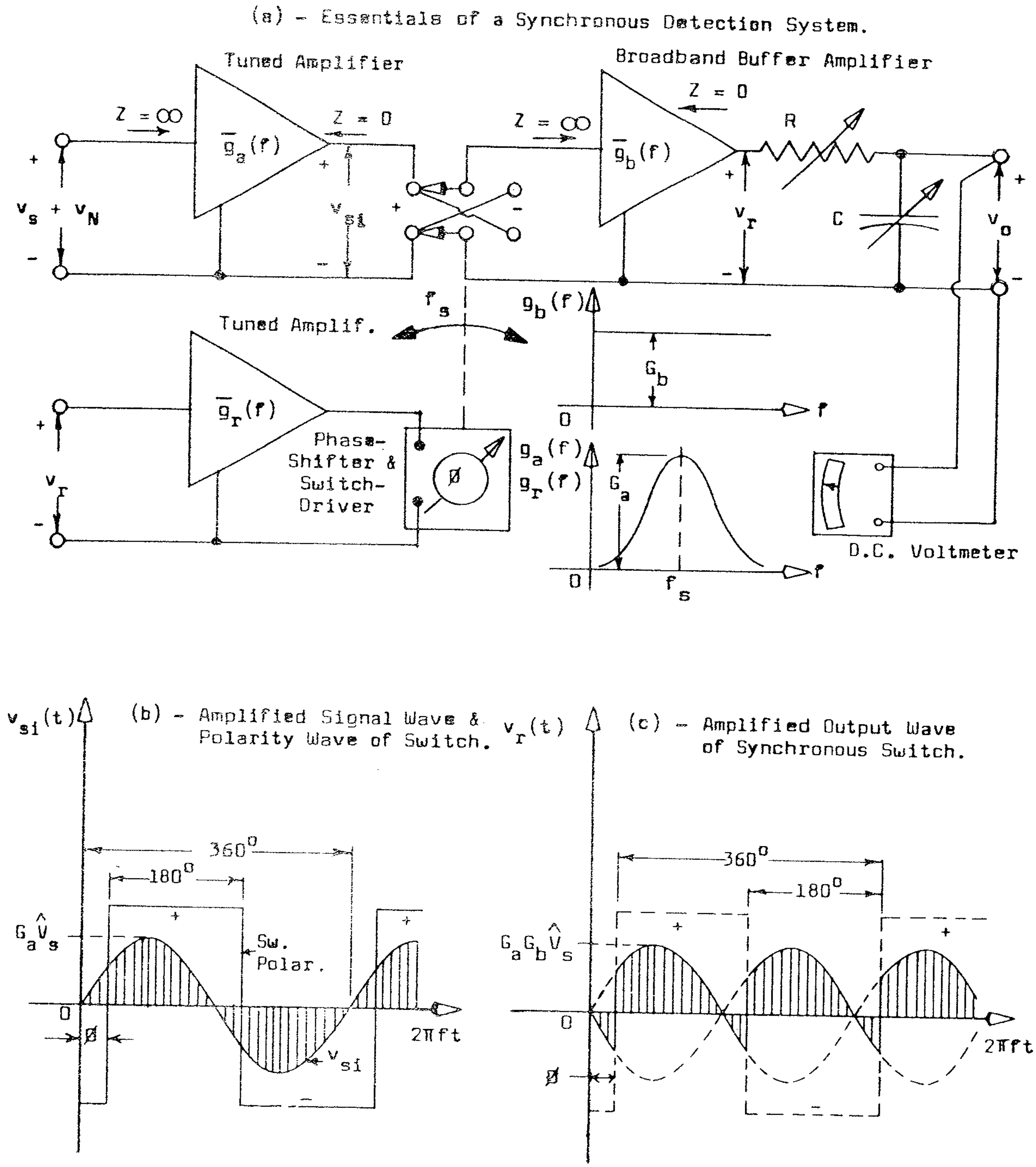


FIGURE 16 - Block Diagram & Waveforms of a Synchronous Detector.

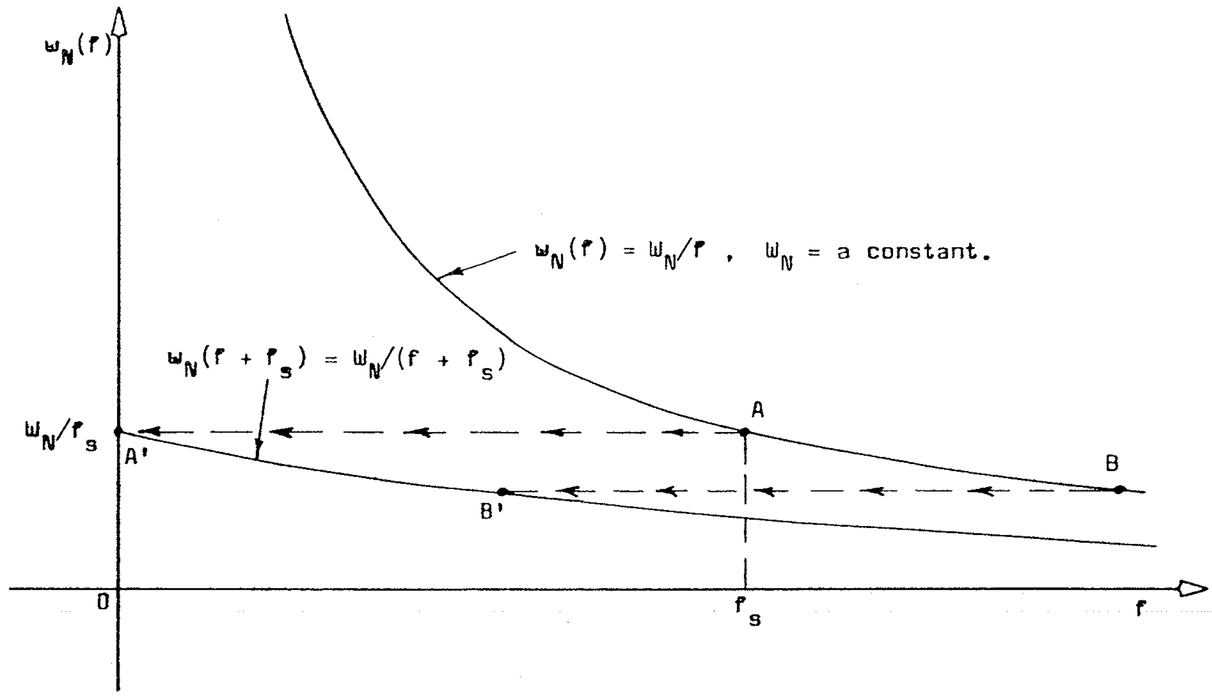


FIGURE 17 - Effect of Frequency-Translation Capability of Lock-in Amplifier on Available Noise-Power Density of 1/f Source.

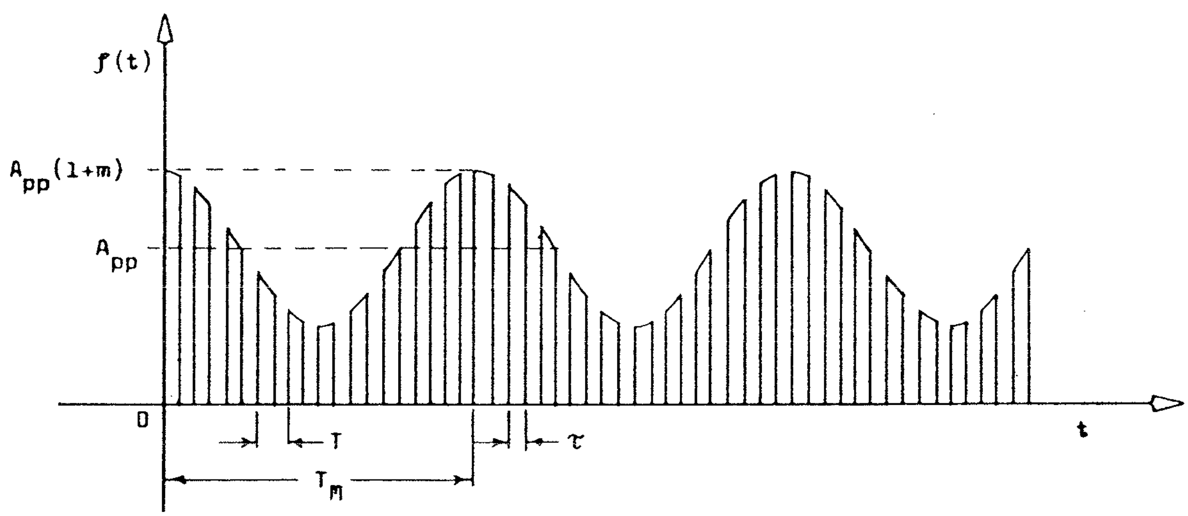
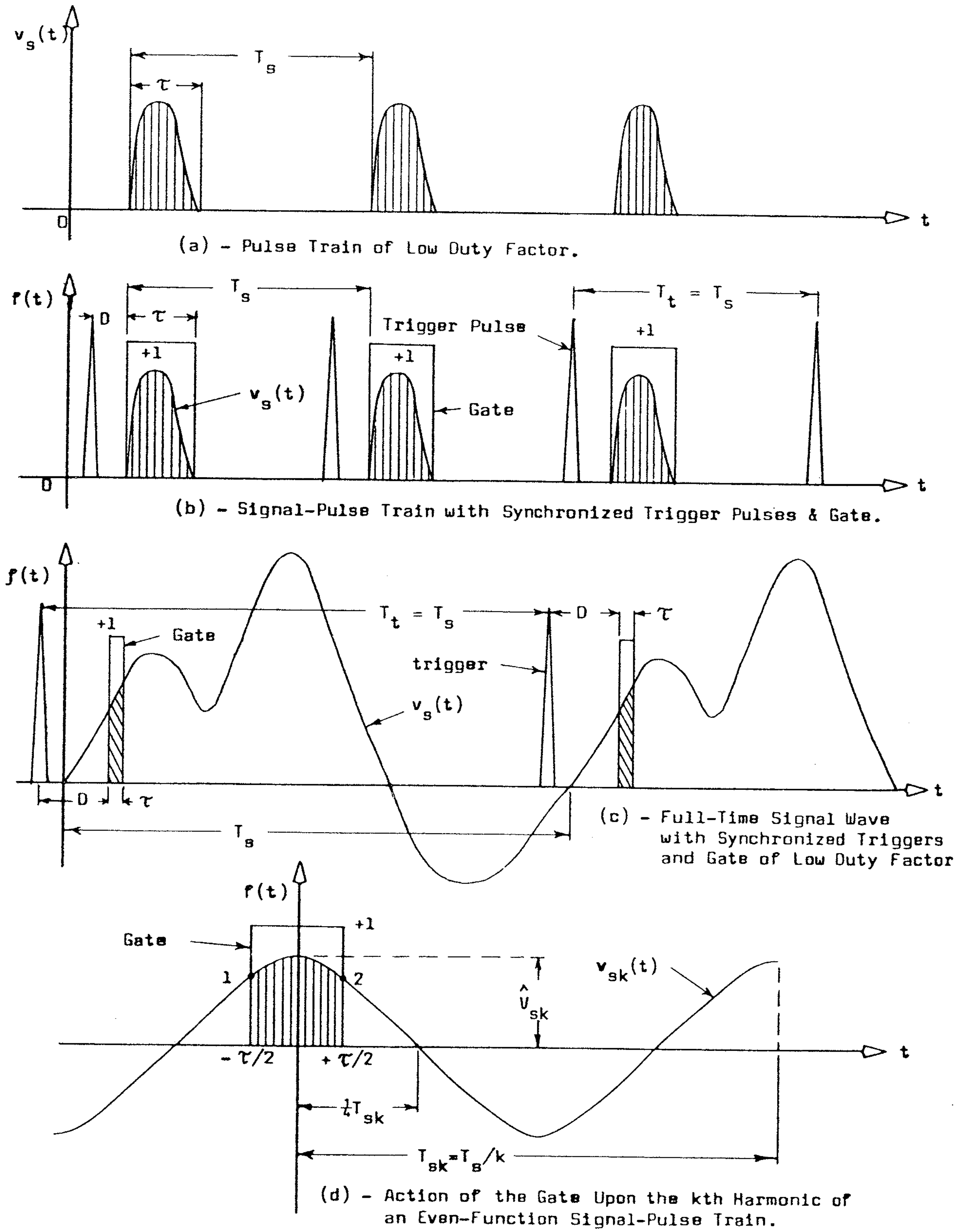


FIGURE 18 - Use of Square-Wave On-Off Chopper to Make a Unidirectional Cosine-Varying Signal Appear as an Amplitude Modulation on a Carrier Wave at an Arbitrary Frequency $f = 1/T$.



$v_s(t)$
T_s
τ
0
t
(a) - Pulse Train of Low Duty Factor.
f(t)
D
+1
Trigger Pulse
$T_t = T_s$
$v_s(t)$
Gate
(b) - Signal-Pulse Train with Synchronized Trigger Pulses & Gate.
f(t)
Gate
trigger
$v_s(t)$
(c) - Full-Time Signal Wave with Synchronized Triggers and Gate of Low Duty Factor
1
2
$\hat{V}_{sk}$
$v_{sk}(t)$
$-\tau/2$
$+\tau/2$
$\frac{1}{4}T_{sk}$
$T_{sk}=T_s/k$
(d) - Action of the Gate Upon the kth Harmonic of an Even-Function Signal-Pulse Train.

FIGURE 19

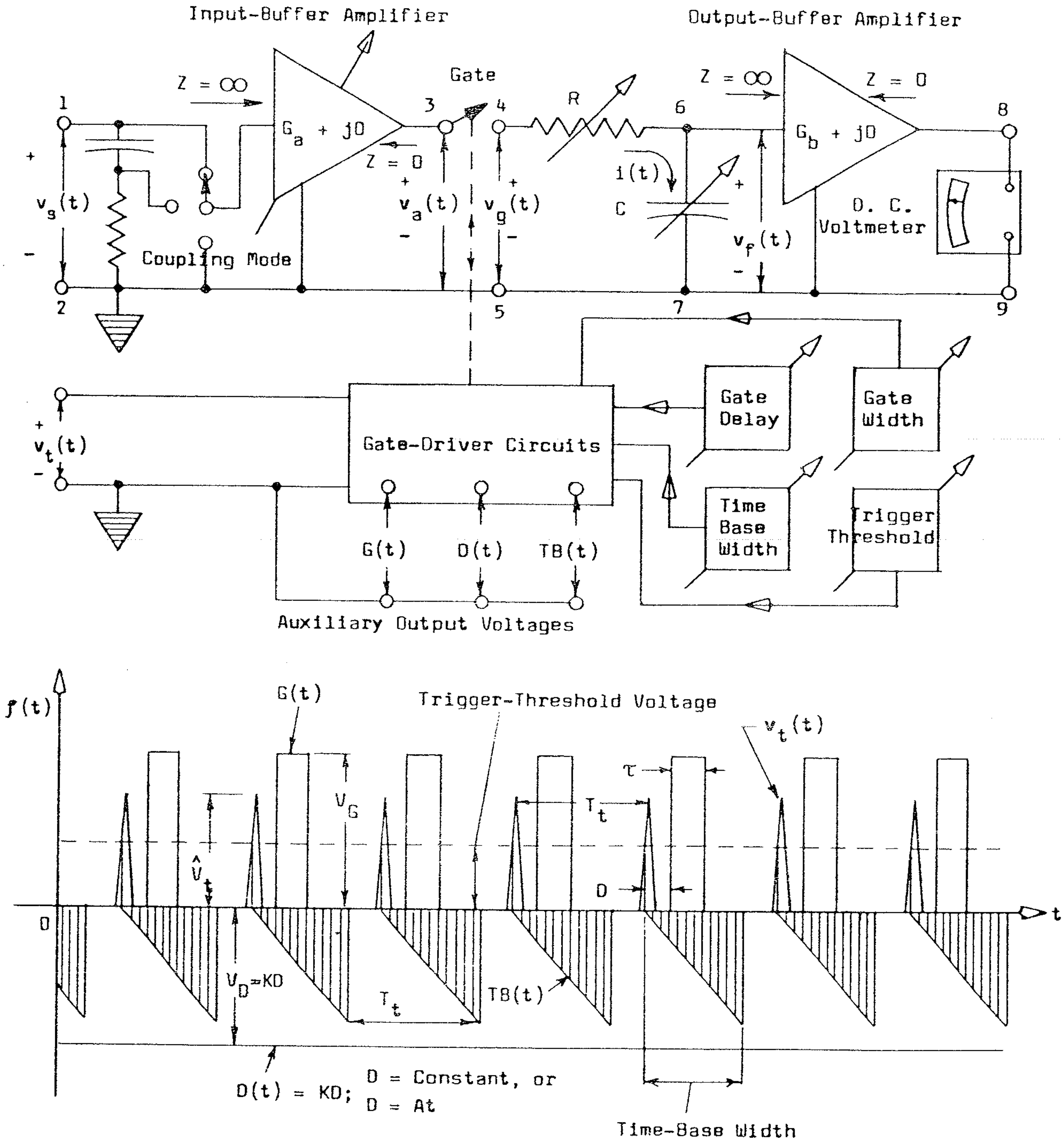


FIGURE 20 - Block Diagram of a Boxcar Integrator, and Waveforms of Auxiliary Output Voltages with an External Recurrent Trigger.

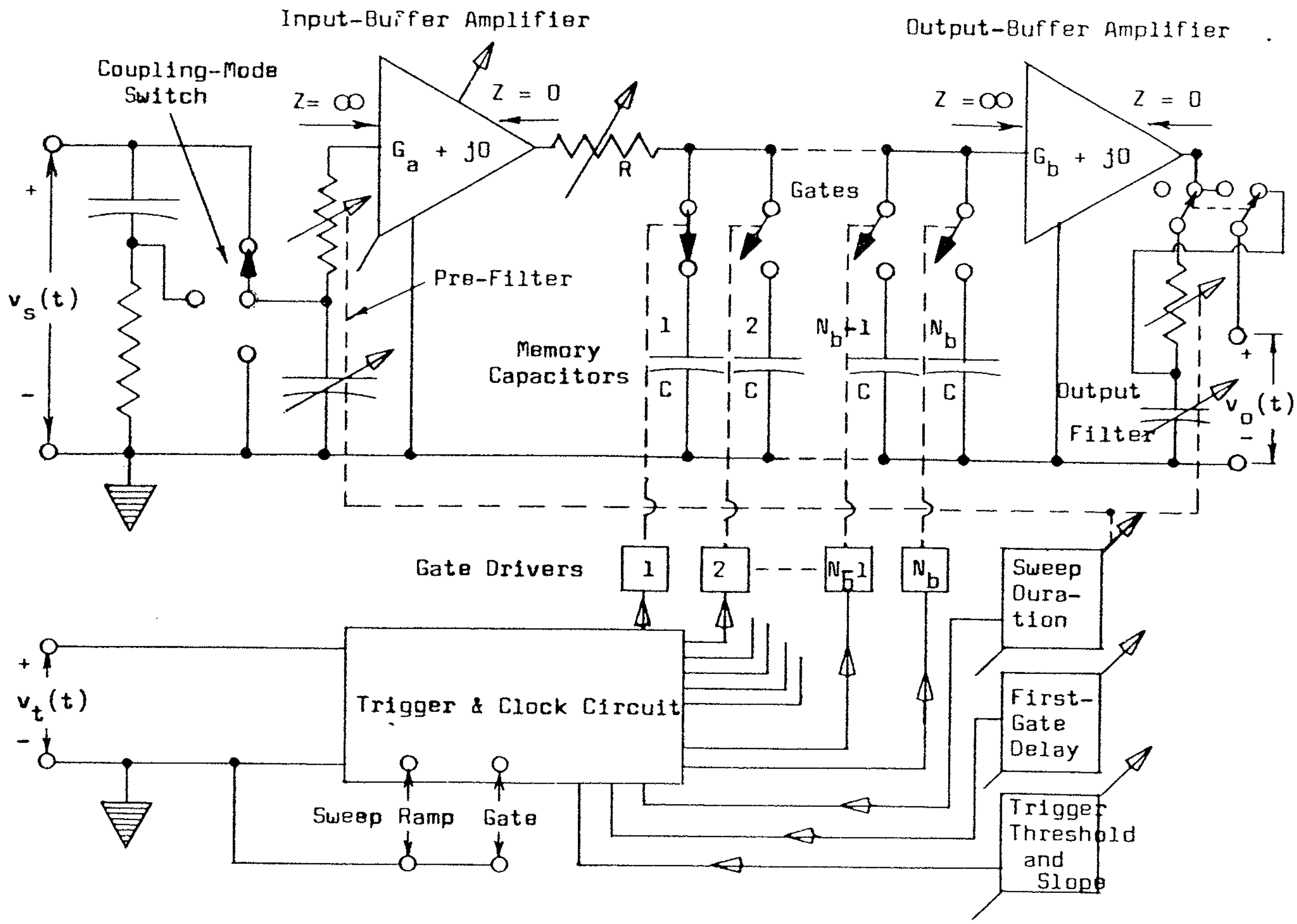


FIGURE 21 - Essentials of a WAVEFORM EDUCTOR.

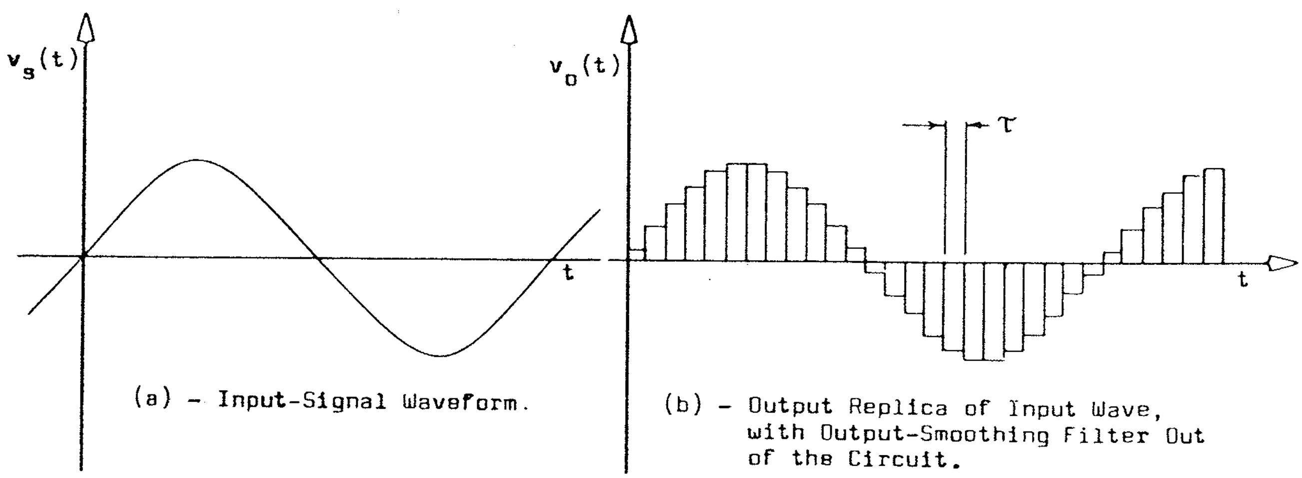


FIGURE 22 - Input and Output Waves of WAVEFORM EDUCTOR for a Sinusoidal Input Signal Without Noise.

[illegible] EARCH
6 DE [illegible]NGELO [illegible] BEDFORD, [illegible] 01730