

2

Signals and noise

Types of signals

The nature of signals, and their relationship to noise and interfering signals, determines appropriate design all the way from the system level down to the component selection level. In this chapter we will take a look at signals and noise, and how each affects the design of amplification and other RF circuits.

Signals can be categorized several ways, but one of the most fundamental is according to *time domain* behaviour (the other major category is *frequency domain*). We will therefore consider signals of the form $v = f(t)$ or $i = f(t)$. The time domain classes of signals include: *static*, *quasistatic*, *periodic*, *repetitive*, *transient*, *random*, and *chaotic*. Each of these categories has certain properties that can profoundly influence appropriate design decisions.

Static and quasistatic signals

A *static* signal (Fig. 2.1A) is, by definition, unchanging over a very long period of time (T_{long} in Fig. 2.1A). Such a signal is essentially a DC level, so must be processed in low drift DC amplifier circuits. This type of signal does not occur at radio frequencies because it is DC, but some RF circuits may produce a DC level, e.g. a continuous wave, constant amplitude RF signal applied to an envelope detector.

The term *quasistatic* means ‘nearly unchanging’, so a quasistatic signal (Fig. 2.1B) refers to a signal that changes so slowly over long times that it possesses characteristics more like static signals than dynamic (i.e. rapidly changing) signals.

Periodic signals

A *periodic* signal (Fig. 2.1C) is one that exactly repeats itself on a regular basis. Examples of periodic signals include sine waves, square waves, sawtooth waves, triangle waves, and so forth. The nature of the periodic waveform is such that each waveform is identical at like points along the time line. In other words, if you advance along the time line by exactly one period (T), then the voltage, polarity and direction of change of the waveform will be repeated. That is, for a voltage waveform, $V(t) = V(t + T)$.

Repetitive signals

A *repetitive* signal (Fig. 2.1D) is quasiperiodic in nature, so bears some similarity to the periodic waveform. The principal difference between repetitive and periodic signals is

6 *RF Components and Circuits*



Figure 2.1 Various types of signal: (A) static

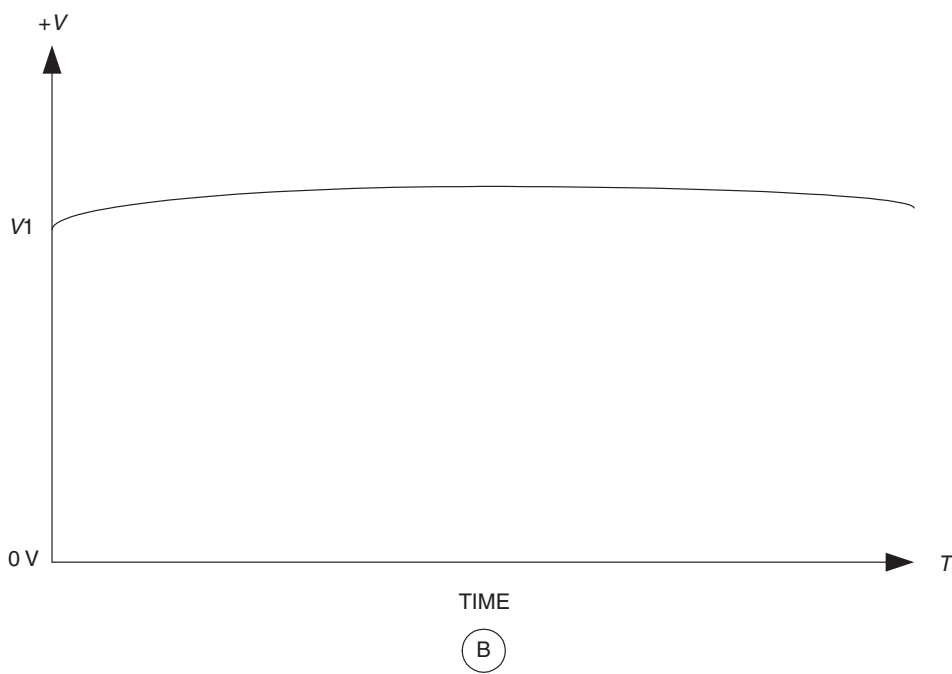


Figure 2.1 (B) quasi-static

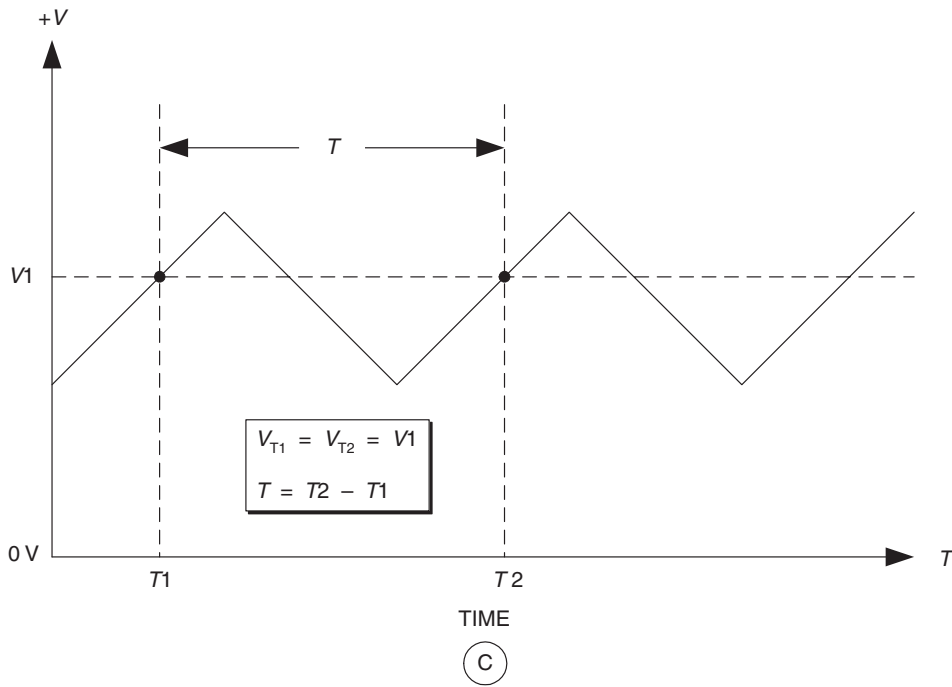


Figure 2.1 (C) periodic

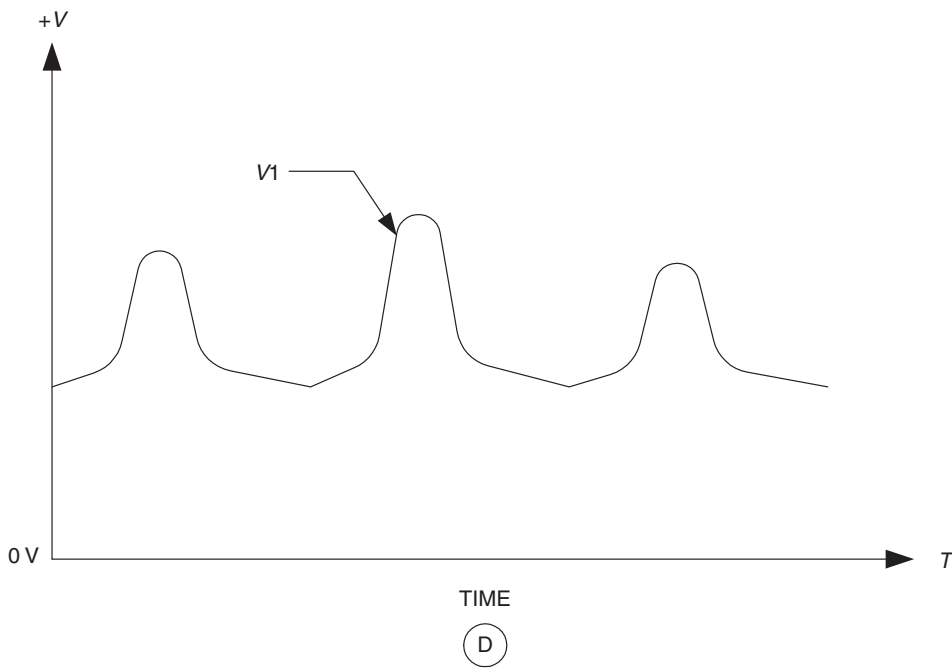


Figure 2.1 (D) quasi-periodic

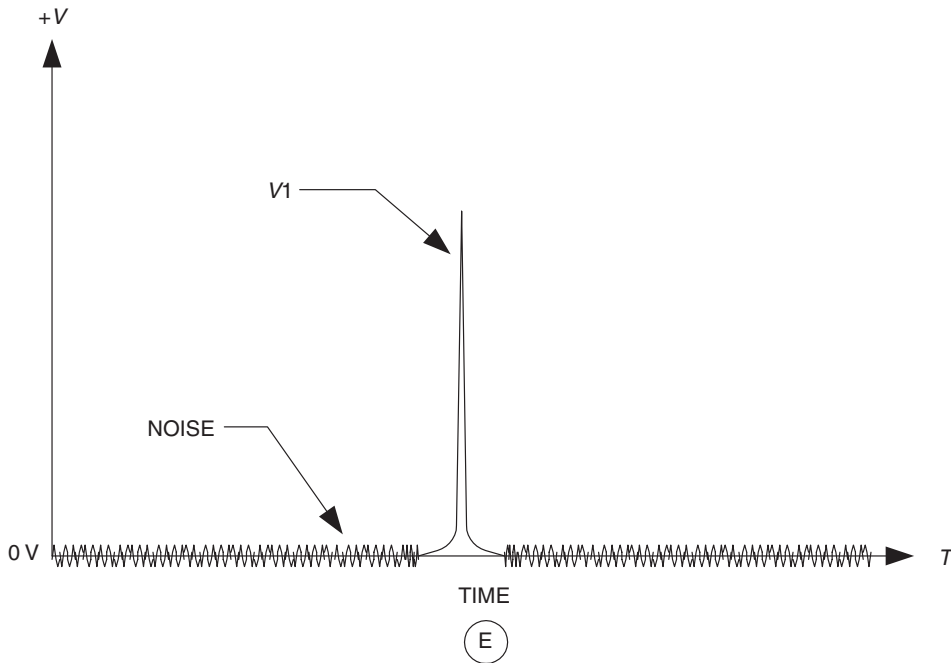


Figure 2.1 (E) spectrum of single frequency

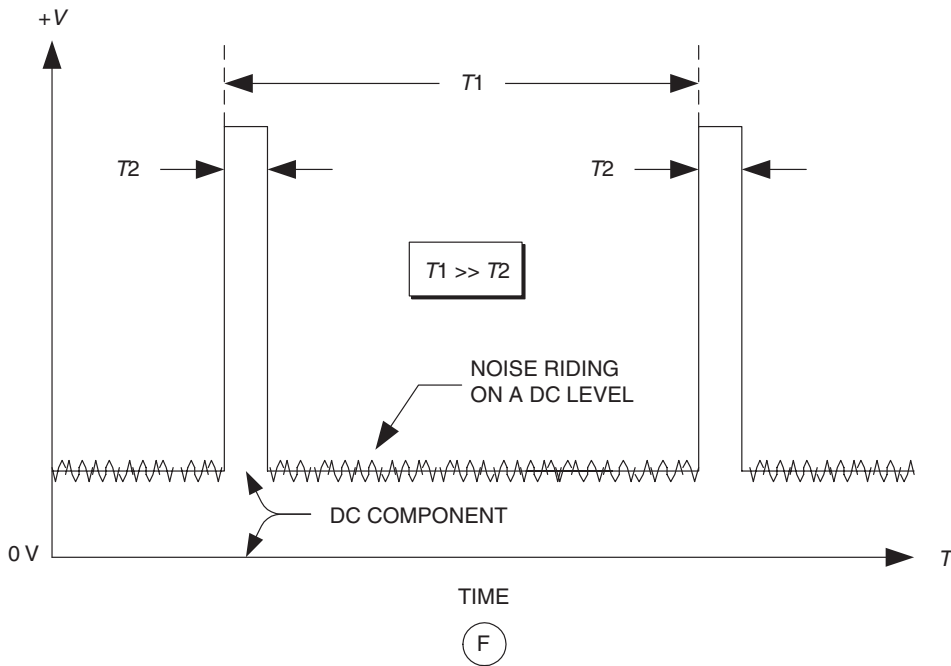


Figure 2.1 (F) two pulses.

seen by comparing the signal at $f(t)$ and $f(t + T)$, where T is the period of the signal. Unlike periodic signals, in repetitive signals these points might not be identical although they will usually be similar. The general waveshape is nearly the same. The repetitive signal might contain either transient or stable features that vary from period to period.

Transient signals and pulse signals

A *transient* signal (Fig. 2.1E) is either a one-time event, or a periodic event in which the event duration is very short compared with the period of the waveform (Fig. 2.1F). In terms of Fig. 2.1F, the latter definition means that $t_1 \ll t_2$. These signals can be treated as if they are transients. In RF circuits these signals might be intentionally generated as pulses (radar pulses resemble Fig. 2.1F), or a noise transient (Fig. 2.1E).

Fourier series

All continuous periodic signals can be represented by a fundamental frequency sine wave, and a collection of sine or cosine harmonics of that fundamental sine wave, that are

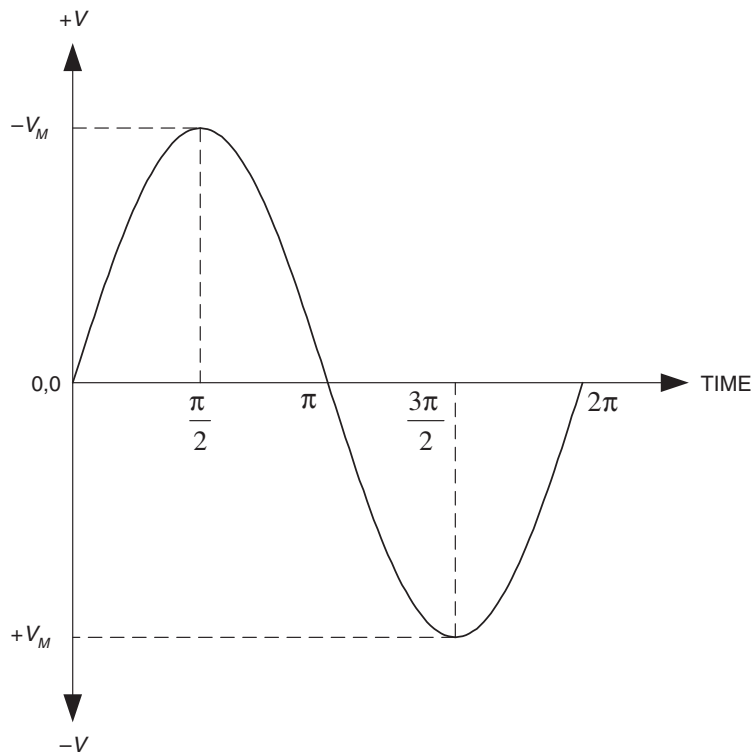


Figure 2.2 Sine wave.

10 RF Components and Circuits

summed together linearly. These frequencies comprise the *Fourier series* of the waveform. The elementary sine wave (Fig. 2.2) is described by:

$$v = V_m \sin(\omega t) \quad (2.1)$$

Where:

v is the instantaneous amplitude of the sine wave

V_m is the peak amplitude of the sine wave

ω is the angular frequency ($2\pi F$) of the sine wave

t is the time in seconds

The *period* of the sine wave is the time between repetition of identical events, or $T = 2\pi/\omega = 1/F$ (where F is the frequency in cycles per second).

The *Fourier series* that makes up a waveform can be found if a given waveform is decomposed into its constituent frequencies either by a bank of frequency selective filters, or a digital signal processing algorithm called the *fast Fourier transform* (FFT). The Fourier series can also be used to construct a waveform from the ground up. Figure 2.3 shows square wave (Fig. 2.3A), sawtooth wave (Fig. 2.3B) and peaked wave (Fig. 2.3C) signals constructed from fundamental sine waves and their harmonic sine and cosine functions.

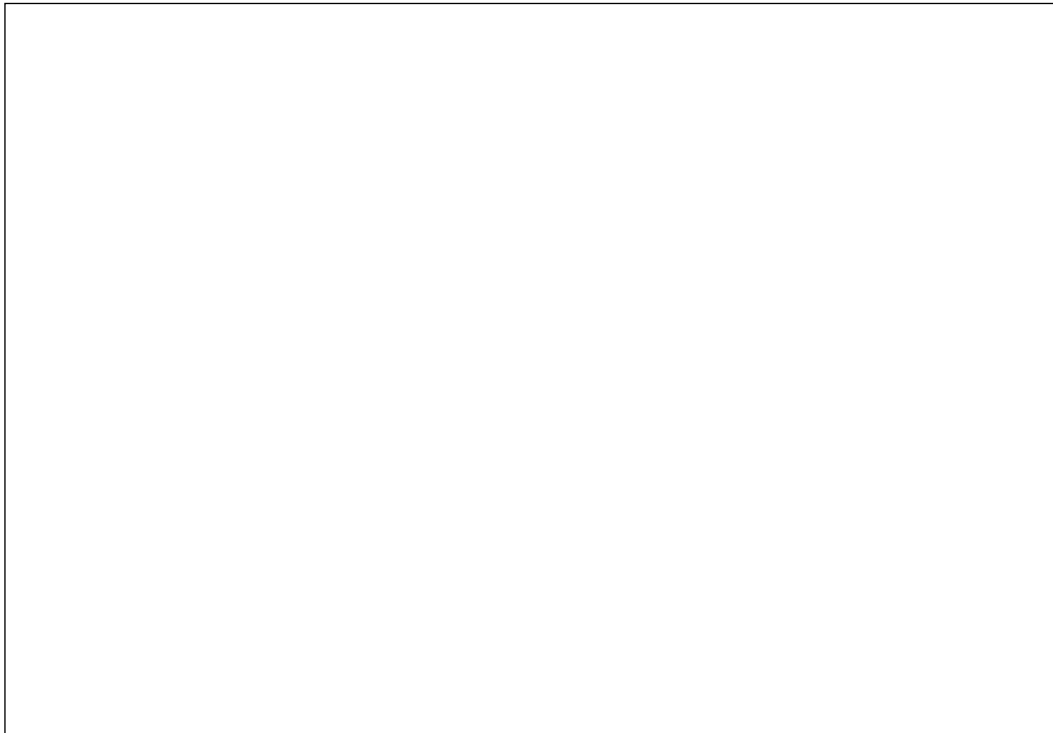


Figure 2.3

The Fourier series for any waveform can be expressed in the form:

$$f(t) = \frac{a_o}{2} + \sum_{n=1}^{\infty} [a_n \cos(n\omega t) + b_n \sin(n\omega t)] \quad (2.2)$$

Where:

a_n and b_n are the amplitudes of the components (see below)

n is an integer ($n = 1$ is the fundamental)

Other terms are as previously defined

The amplitude coefficients (a_n and b_n) are expressed by:

$$a_n = \frac{2}{T} \int_0^T f(t) \cos(n\omega t) dt \quad (2.3)$$

and,

$$b_n = \frac{2}{T} \int_0^T f(t) \sin(n\omega t) dt \quad (2.4)$$

Because only certain frequencies, fundamental plus harmonics determined by integer n , are present the spectrum of the periodic signal is said to be *discrete*.

The term $a_o/2$ in the Fourier series expression (Eq. (2.2)) is the average value of $f(t)$ over one complete cycle (one period) of the waveform. In practical terms, it is also the *DC component* of the waveform. When the waveform possesses *half-wave symmetry* (i.e. the peak amplitude above zero is equal to the peak amplitude below zero at every point in t , or $+V_m = |-V_m|$), there is no DC component, so $a_o = 0$.

An alternative Fourier series expression replaces the $a_n \cos(n\omega t) + b_n \sin(n\omega t)$ with an equivalent expression of another form:

$$f(t) = \frac{2}{T} \sum_{n=1}^{\infty} C_n \sin(n\omega t - \phi_n) \quad (2.5)$$

Where:

$$C_n = \sqrt{a_n^2 + b_n^2}$$

$$\phi_n = \arctan\left(\frac{a_n}{b_n}\right)$$

All other terms are as previously defined

Waveform symmetry

One can infer certain things about the Fourier spectrum of a waveform by examination of its *symmetries*. One would conclude from the above equations that the harmonics extend to infinity on all waveforms. Clearly, in practical systems a much less than infinite

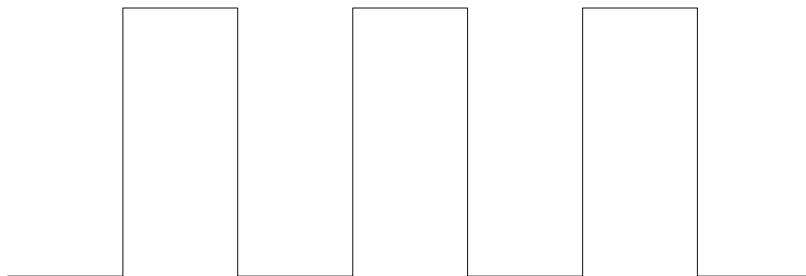


Figure 2.4 Square wave.

bandwidth is found, so some of those harmonics will be removed by the normal action of the electronic circuits. Also, it is sometimes found that higher harmonics might not be truly significant, so can be ignored. As n becomes larger, the amplitude coefficients a_n and b_n tend to become smaller. At some point, the amplitude coefficients are reduced sufficiently that their contribution to the shape of the wave is either negligible for the practical purpose at hand, or are totally unobservable in practical terms. The value of n at which this occurs depends partially on the *rise time* of the waveform. Rise time is usually defined as the time required for the RF pulse waveform to rise from 10 per cent to 90 per cent of its final amplitude.

Figure 2.4 shows an RF pulse waveform based on a square impulse. The square wave represents a special case because it has an extremely fast rise time. Theoretically, the square wave contains an infinite number of harmonics, but not all of the possible harmonics are present. For example, in the case of the square wave only the odd harmonics are found (e.g. 3, 5, 7). According to some standards, accurately reproducing the square wave requires 100 harmonics, while others claim that 1000 harmonics are needed. Which standard to use may depend on the specifics of the application.

Another factor that determines the profile of the Fourier series of a specific waveform is whether the function is *odd* or *even*. Figure 2.5A shows an odd-function square wave, and Fig. 2.5B shows an even-function square wave. The even function is one in which $f(t) = f(-t)$, while for the odd function $-f(t) = f(-t)$. In the even function only cosine harmonics are present, so the sine amplitude coefficients b_n are zero. Similarly, in the odd function only sine harmonics are present, so the cosine amplitude coefficients a_n are zero.

Both *symmetry* and *asymmetry* can occur in several ways in a waveform (Fig. 2.6), and those factors can affect the nature of the Fourier series of the waveform. In Fig. 2.6A we see the case of a waveform with a DC component. Or, in terms of the Fourier series equation, the term a_0 is non-zero. The DC component represents a case of asymmetry in a signal. This offset can seriously affect instrumentation electronic circuits that are DC-coupled.

Two different forms of symmetry are shown in Fig. 2.6B. *Zero-axis symmetry* occurs when, on a point-for-point basis, the waveshape and amplitude above the zero baseline is equal to the amplitude below the baseline (or $|+V_m| = |-V_m|$). When a waveform possesses zero-axis symmetry it will usually not contain even harmonics, only odd harmonics are present; this situation is found in square waves, for example (Fig. 2.7A). Zero-axis symmetry is not found only in sine and square waves, however, as the sawtooth waveform in Fig. 2.6C demonstrates.

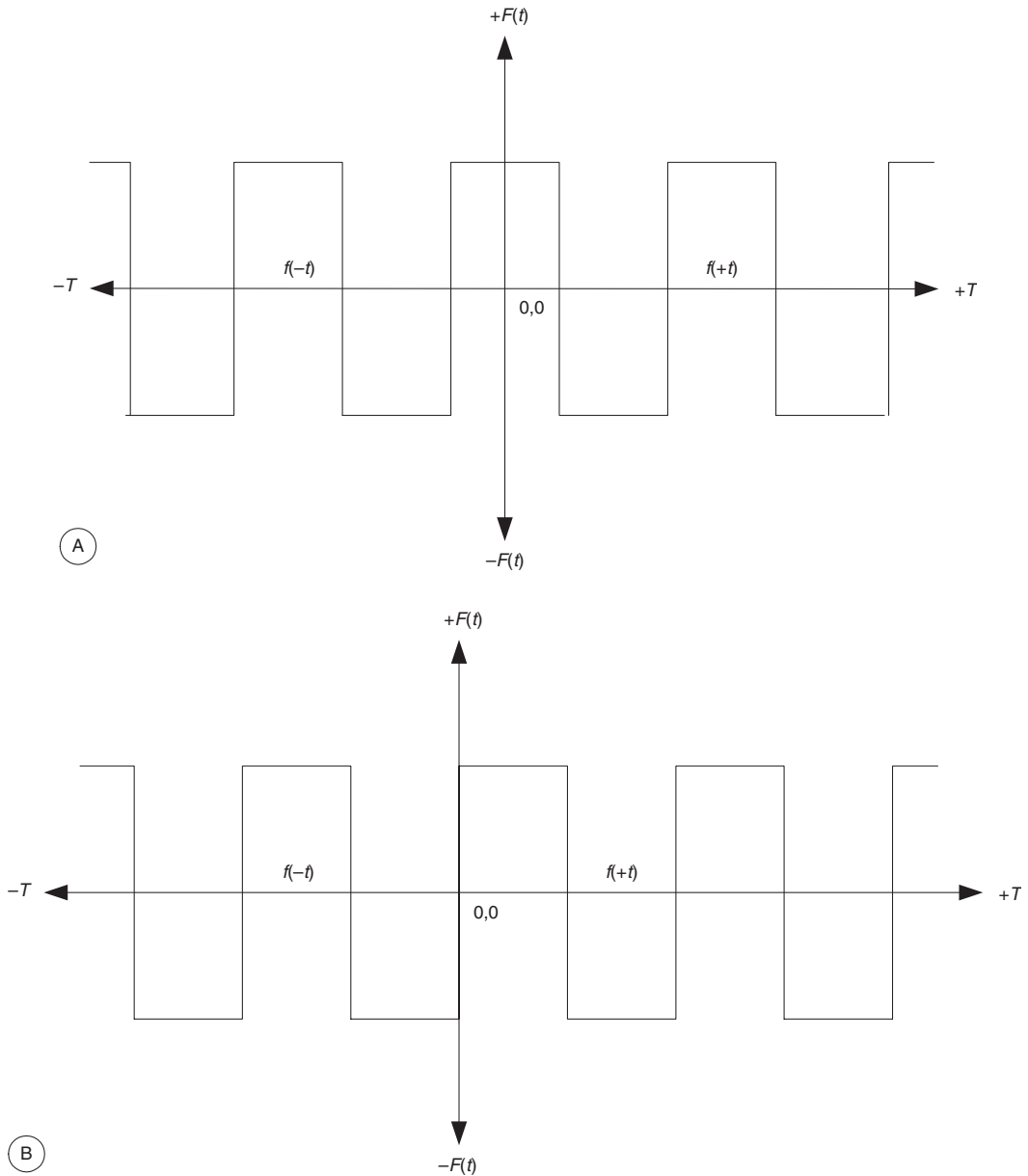


Figure 2.5 Types of symmetry.

An exception to the 'no even harmonics' general rule is that there will be even harmonics present in the zero-axis symmetrical waveform (Fig. 2.7B) if the *even harmonics are in-phase with the fundamental sine wave*. This condition will neither produce a DC component, nor disturb the zero-axis symmetry.

Also shown in Fig. 2.6B is *half wave symmetry*. In this type of symmetry the *shape* of the wave above the zero baseline is a mirror image of the shape of the waveform below the

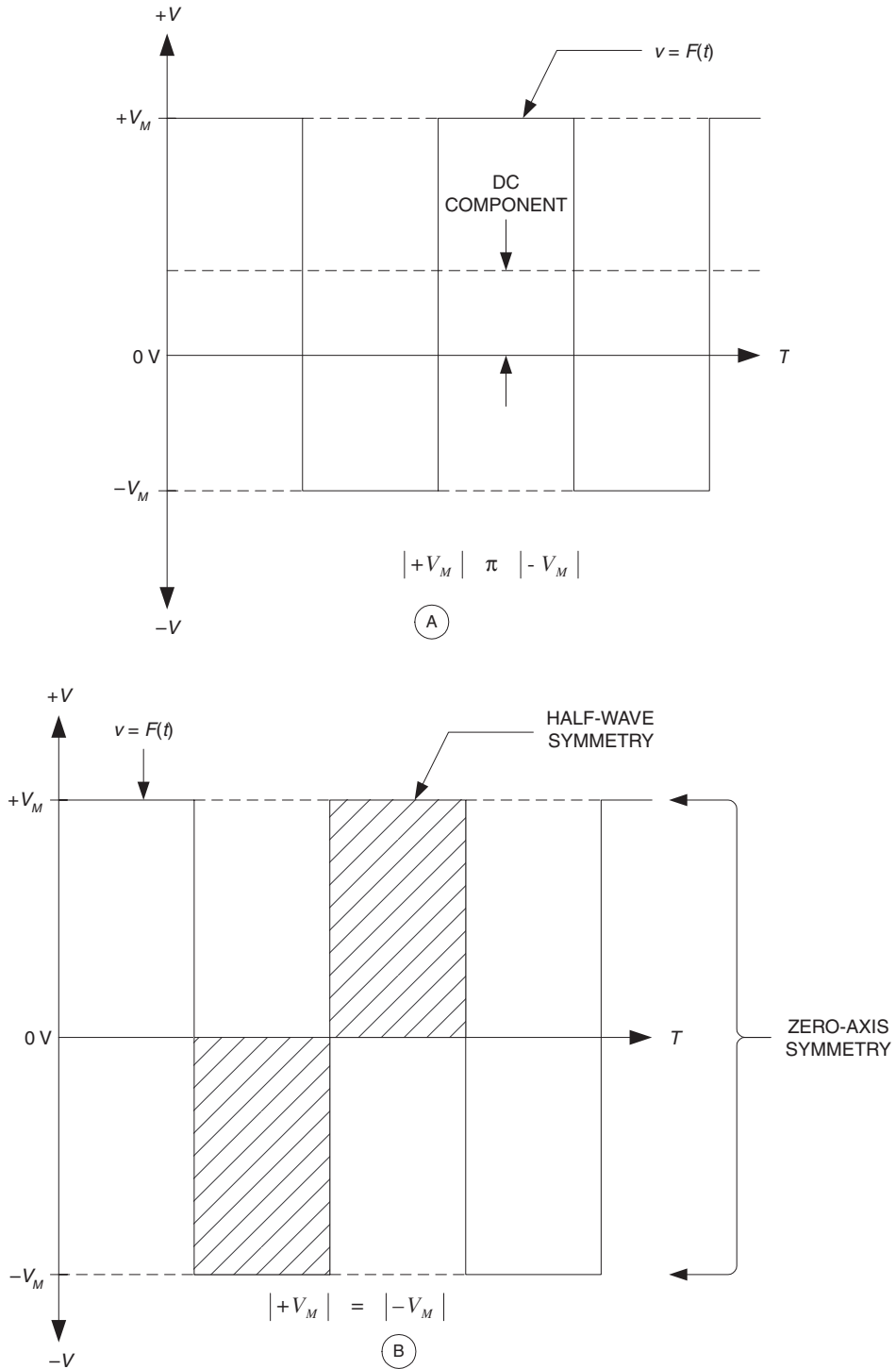


Figure 2.6 (A) Waveform with DC component; (B) half-wave and zero-axis symmetry

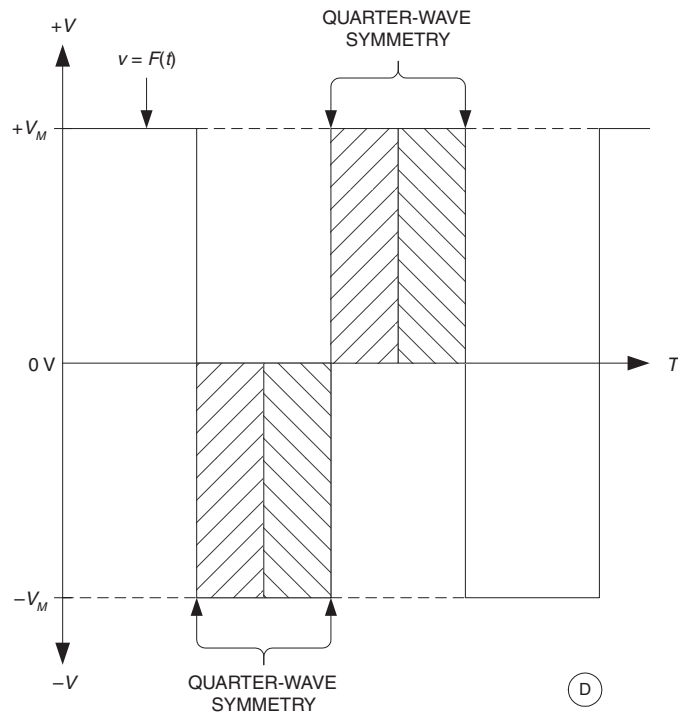
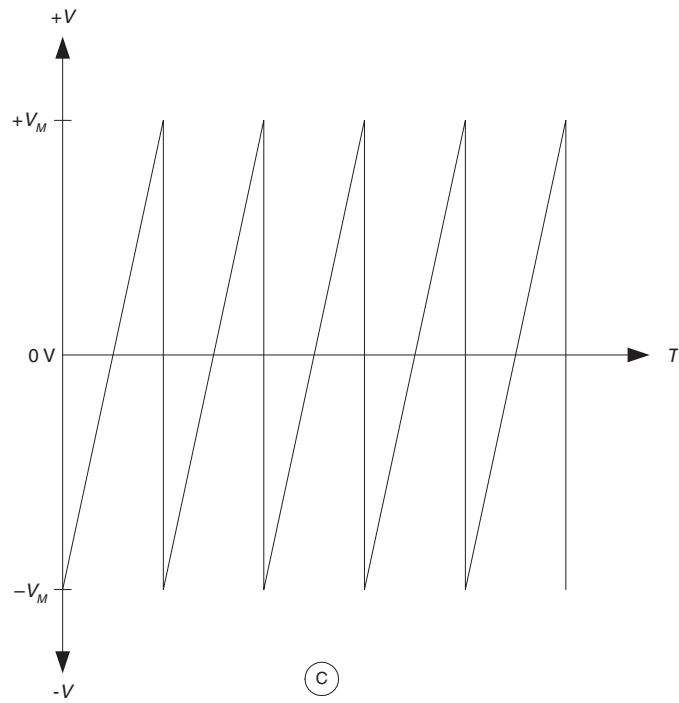


Figure 2.6 (C) triangle waveform; (D) quarter-wave symmetry.

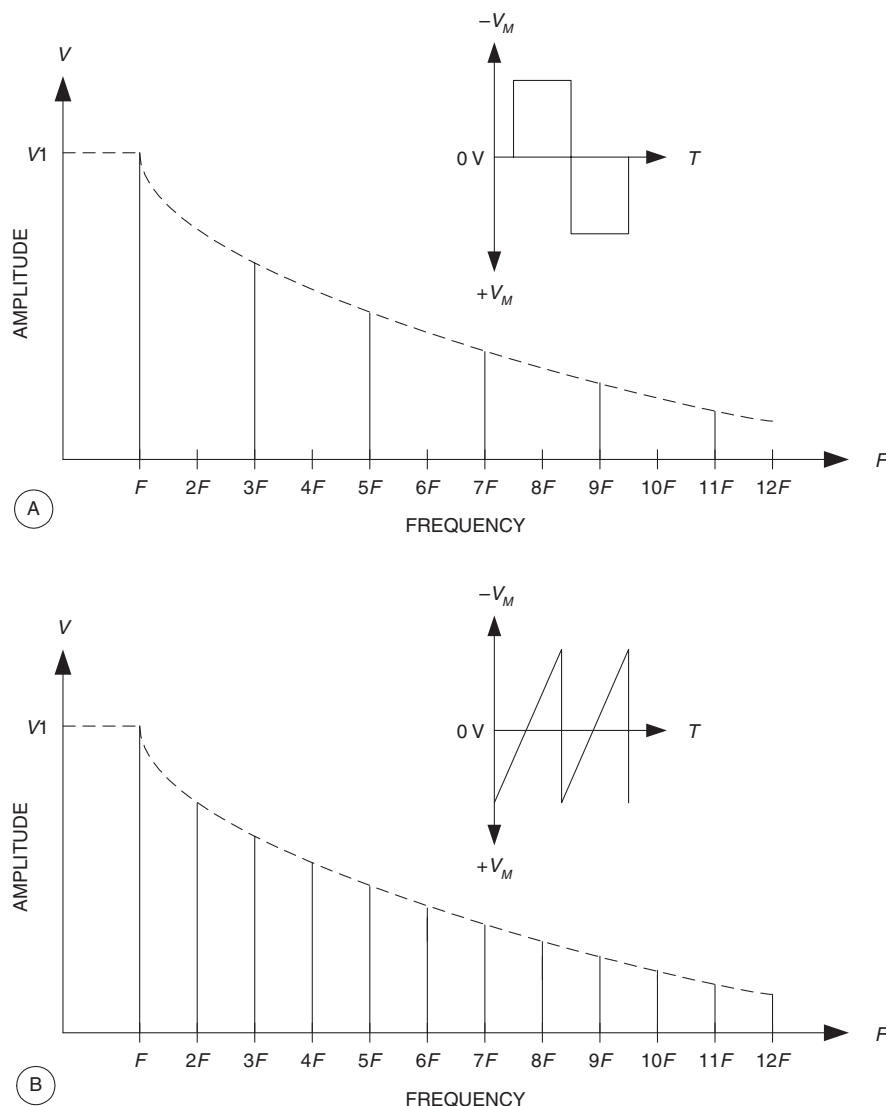


Figure 2.7 Spectrum of two waveforms.

baseline (shaded region in Fig. 2.6B). Half-wave symmetry also implies a lack of even harmonics.

Quarter-wave symmetry (Fig. 2.6D) exists when the left half and right half sides of the waveforms are mirror images of each other on the same side of the zero-axis. Note in Fig. 2.6D, that above the zero-axis the waveform is like a square wave, and indeed the left- and right-hand sides are mirror images of each other. Similarly, below the zero-axis the rounded waveform has a mirror image relationship between left and right sides. In this case, there is a full set of even harmonics, and any odd harmonics that are present are in-phase with the fundamental sine wave.

Transient signals

A transient signal is an event that occurs either once only, or occurs randomly over a long period of time, or is periodic but has a very short duration compared with its period (i.e. it is a very short duty cycle event). Many pulse signals fit the latter criterion even though mathematically they are actually periodic.

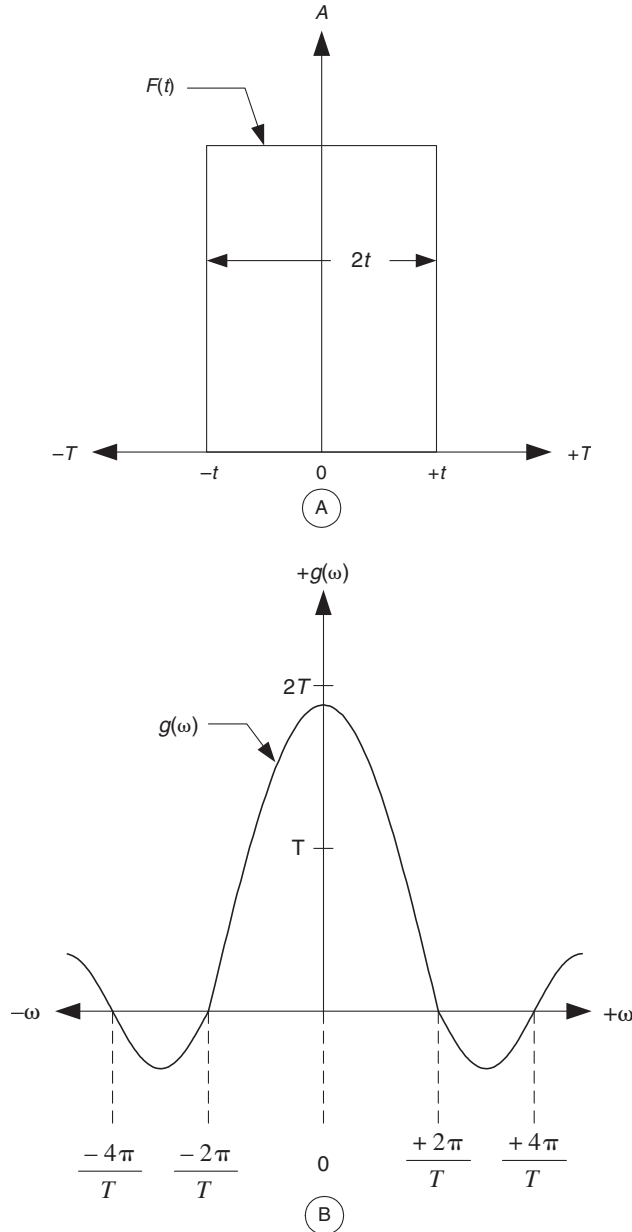


Figure 2.8

Transient signals are not represented properly by the Fourier series, but can nonetheless be represented by sine waves in a spectrum. The difference is that the spectrum of the transient signal is *continuous* rather than discrete. Consider a transient signal of period $2T$, such as Fig. 2.8A. The *spectral density*, $g(\omega)$, is:

$$g(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t} dt \quad (2.6)$$

Given a spectral density the original waveform can be reconstructed from:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} g(\omega)e^{j\omega t} d\omega \quad (2.7)$$

The shape of the spectral density is shown in Fig. 2.8B. Note that the negative frequencies are a product of the mathematics, and do not have physical reality. The shape of Fig. 2.8B is expressed by:

$$g(\omega) = \frac{\sin \omega t}{\omega t} \quad (2.8)$$

The general form $\sin x/x$ is used also for repetitive pulse signals as well as the transient form shown in Fig. 2.8B.

Sampled signals

The digital computer is incapable of accepting analogue input signals, but rather requires a digitized representation of that signal. The *analogue-to-digital* (A/D) converter will convert an input voltage (or current) to a representative binary word. If the A/D converter is either clocked or allowed to run asynchronously according to its own clock, then it will take a continuous string of samples of the signal as a function of time. When combined, these signals represent the original analogue signal in binary form.

But the sampled signal is not exactly the same as the original signal, and some effort must be expended to ensure that the representation is as good as possible. Consider Fig. 2.9. The waveform in Fig. 2.9A is a continuous voltage function of time, $V(t)$; in this case a triangle waveform is seen. If the signal is sampled by another signal, $p(t)$, with frequency F_s and sampling period $T = 1/F_s$, as shown in Fig. 2.9B, and then later reconstructed, the waveform may look something like Fig. 2.9C. While this may be sufficiently representative of the waveform for many purposes, it would be reconstructed with greater fidelity if the sampling frequency (F_s) is increased.

Figure 2.10 shows another case in which a sine wave, $V(t)$ in Fig. 2.10A, is sampled by a pulse signal, $p(t)$ in Fig. 2.10B. The sampling signal, $p(t)$, consists of a train of equally spaced narrow pulses spaced in time by T . The sampling frequency F_s equals $1/T$. The resultant is shown in Fig. 2.10C, and is another pulsed signal in which the amplitudes of the pulses represent a sampled version of the original sine wave signal.

The sampling rate, F_s , must by *Nyquist's theorem* be twice the maximum frequency (F_m) in the Fourier spectrum of the applied analogue signal, $V(t)$. In order to reconstruct the original signal after sampling, it is necessary to pass the sampled waveform through a

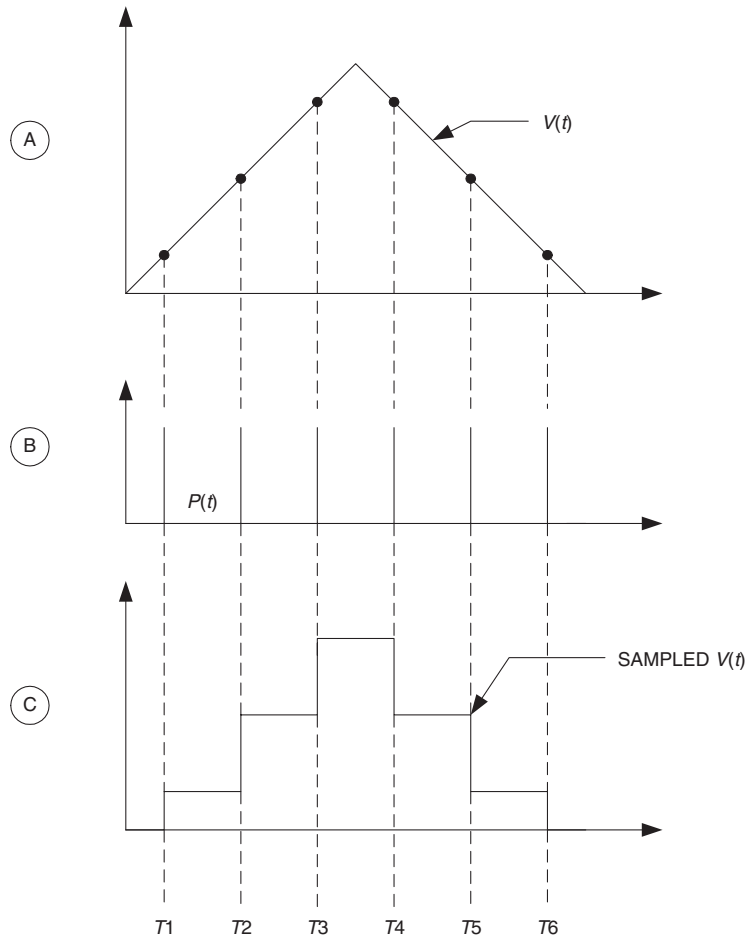


Figure 2.9 Sampled waveform and its reconstruction.

low-pass filter that limits the pass band to F_s . In practical RF systems, you will find that many engineers determine that the minimum Nyquist rate is insufficient for good fidelity reproductions of the sampled waveform, so will specify a faster rate. Also, some oversampling methods are used to dramatically reduce noise.

The sampling process is analogous to a form of *amplitude modulation* (AM), in which $V(t)$ is the modulating signal, with spectrum from DC to F_m , and $p(t)$ is the carrier frequency. The resultant spectrum is shown partially in Fig. 2.11, and resembles the double sideband with carrier AM spectrum. The spectrum of the modulating signal appears as 'sidebands' around the 'carrier' frequency, shown here as F_o . The actual spectrum is a bit more complex, as shown in Fig. 2.12. Like an unfiltered AM radio transmitter, the same spectral information appears not only around the fundamental frequency (F_s) of the carrier (shown at zero in Fig. 2.12), but also at the harmonics spaced at intervals of F_s up and down the spectrum.

Providing that the sampling frequency $F_s \geq 2F_m$, the original signal is recoverable from the sampled version by passing it through a low-pass filter with a cut-off frequency F_c , set

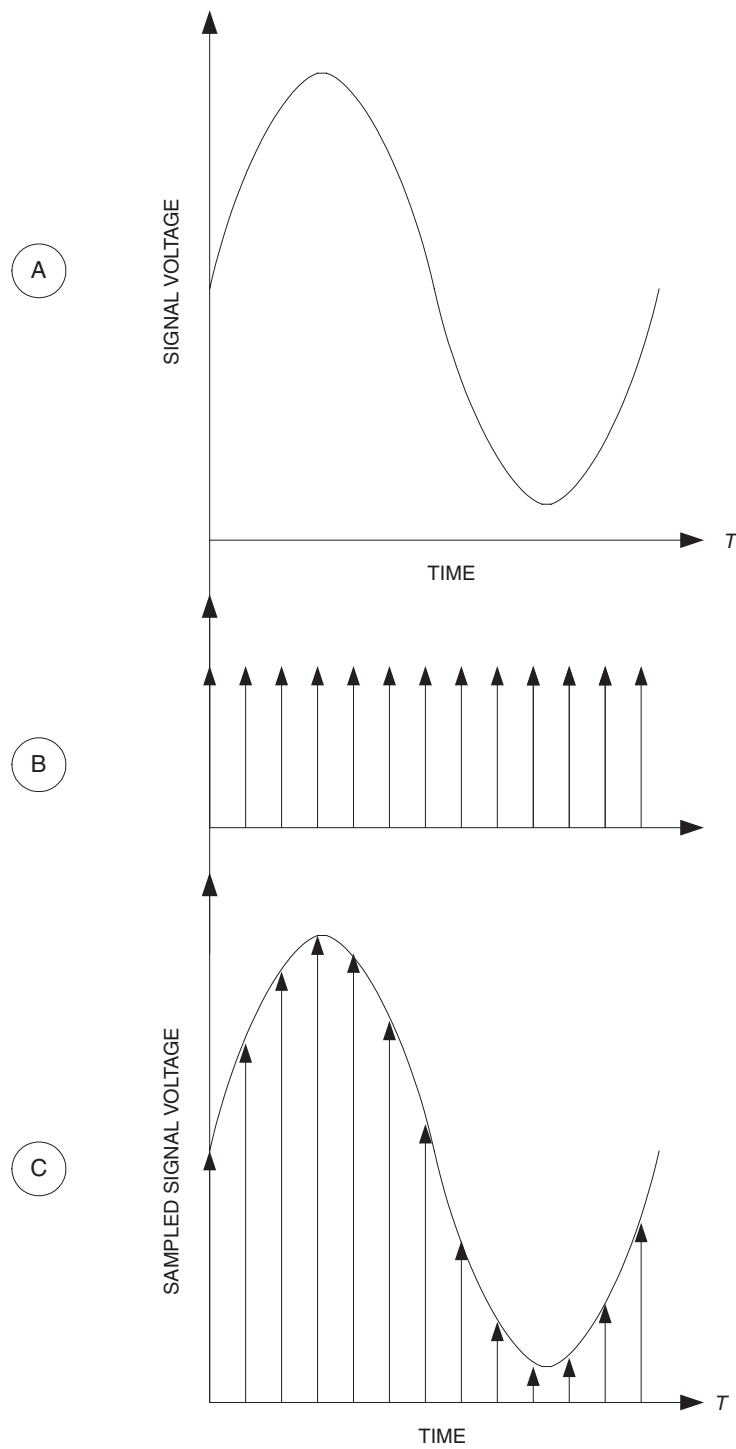


Figure 2.10 Sampled sine wave and its reconstruction.

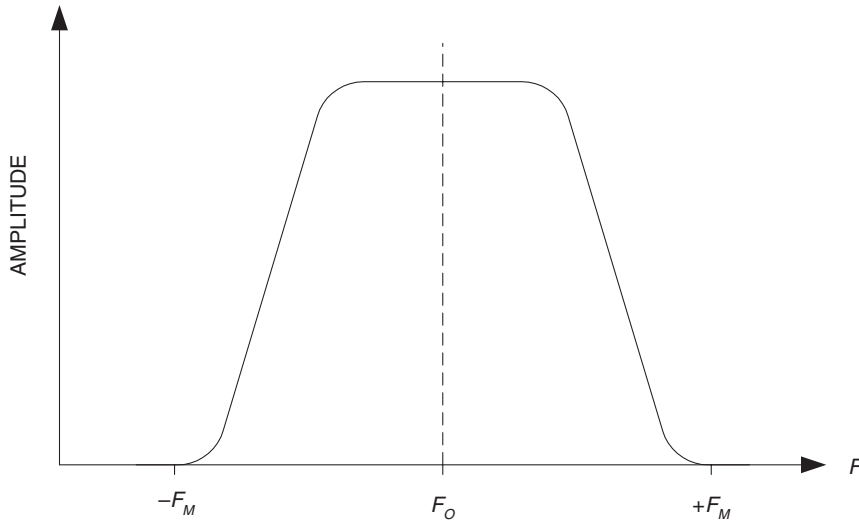


Figure 2.11 Spectrum.

to pass only the spectrum of the analog signal – but not the sampling frequency. This phenomenon is shown with the dotted line in Fig. 2.12.

When the sampling frequency $F_s < 2F_m$, then a problem occurs (see Fig. 2.13). The spectrum of the sampled signal looks similar to before, but the regions around each harmonic overlap such that the value of $-F_m$ for one spectral region is less than $+F_m$ for the next lower frequency region. This overlap results in a phenomenon called *aliasing*. That is, when the sampled signal is recovered by low-pass filtering it will produce not the original sine wave frequency F_o but a lower frequency equal to $(F_s - F_o)$... and the information carried in the waveform is thus lost or distorted.

The solution, for accurate sampling of the analogue waveform for input to a computer, is to:

- 1 Bandwidth limit the signal at the input of the sampler or A/D converter with a low-pass filter with a cut-off frequency F_c selected to pass only the maximum frequency in the waveform (F_m) and not the sampling frequency (F_s).
- 2 Set the sampling frequency F_s at least twice the maximum frequency in the applied waveform's Fourier spectrum, i.e. $F_s \geq 2F_m$.

Noise

An ideal electronic circuit produces no noise of its own, so the output signal from the ideal circuit contains only the noise that was in the original signal. But real electronic circuits and components do produce a certain level of inherent noise. Even a simple fixed value resistor is noisy. Figure 2.14A shows the equivalent circuit for an ideal, noise-free resistor. The inherent noise is represented in Fig. 2.14B by a noise voltage source, V_n , in series with the ideal, noise-free resistance, R_i . At any temperature above *absolute zero* (0 K or about

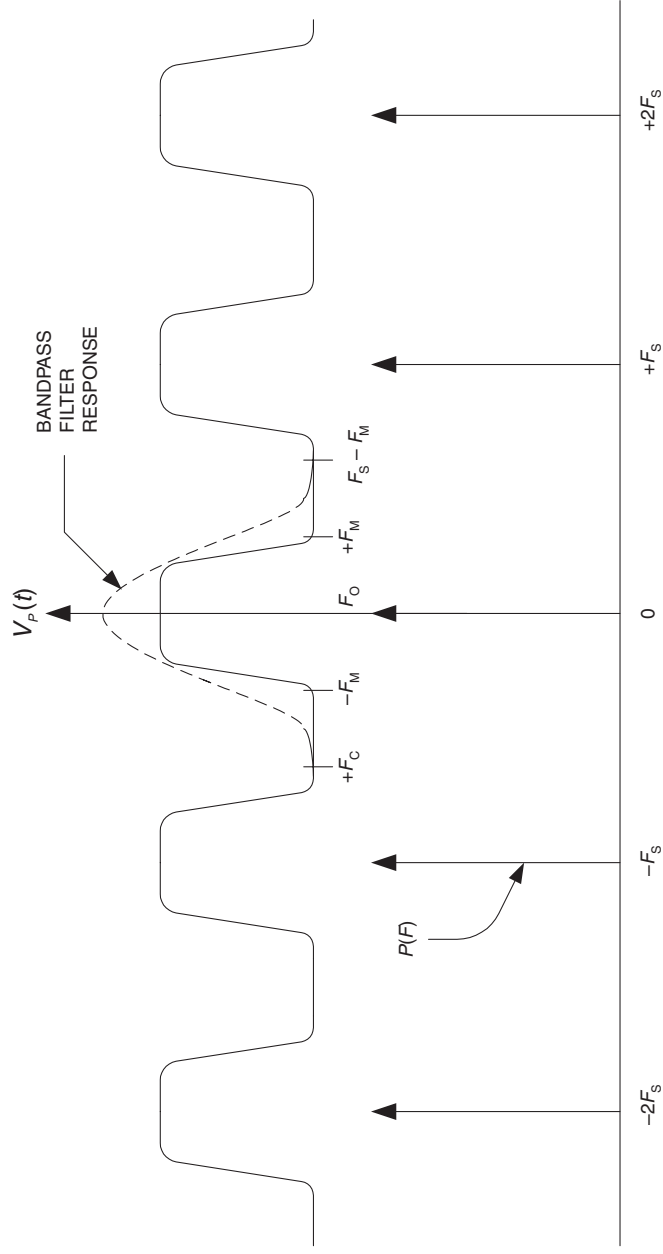


Figure 2.12 Spectrum of sampled signal.

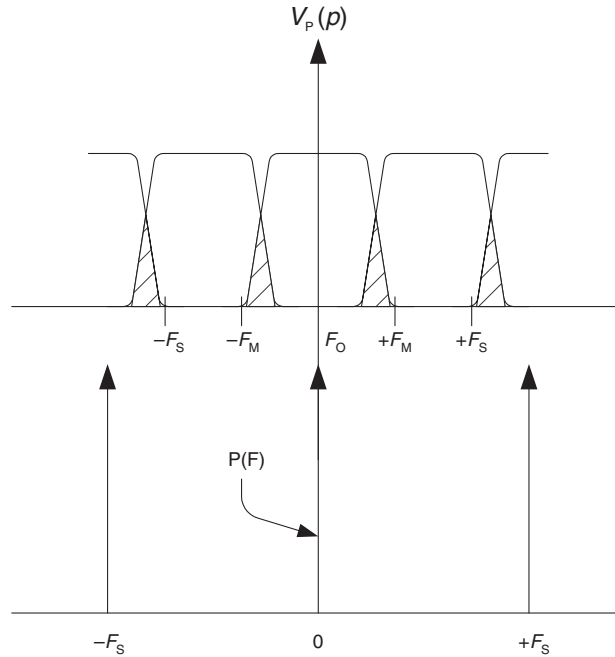


Figure 2.13 Aliasing.

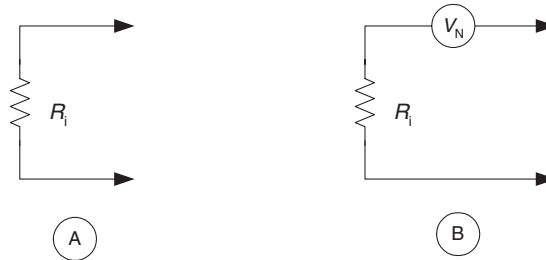


Figure 2.14 (A) Ideal resistor; (B) real practical resistor.

-273°C) electrons in any material are in constant random motion. Because of the inherent randomness of that motion, however, there is no detectable current in any one direction. In other words, electron drift in any single direction is cancelled over short time periods by equal drift in the opposite direction. Electron motions are therefore statistically decorrelated. There is, however, a continuous series of random current *pulses* generated in the material, and those pulses are seen by the outside world as a noise signal. This signal is called by several names: *thermal agitation noise*, *thermal noise*, or *Johnson noise*.

Johnson noise is a so-called 'white noise' because it has a very broadband (nearly gaussian) spectral density. The thermal noise spectrum is essentially flat. The term 'white

noise' is a metaphor developed from white light, which is composed of all visible colour frequencies. The expression for Johnson noise is:

$$V_N = \sqrt{4KTRB} \text{ V} \quad (2.9)$$

Where:

V_n is the noise voltage (V)

K is Boltzmann's constant (1.38×10^{-23} J/K)

T is the temperature in kelvin (K)

R is the resistance in ohms (Ω)

B is the bandwidth in hertz (Hz)

At normal room temperature, with the constants collected, and the resistance normalized to 1 kohms, Eq. (2.9) reduces to:

$$V_N = 4 \sqrt{\frac{R}{1 \text{ k}\Omega}} \frac{\text{n V}}{\sqrt{\text{Hz}}} \quad (2.10)$$

The evaluated solution of Eq. (2.10) is normally read *nanovolts (nV) per square root hertz*. In this equation, a 1 megohm resistor will have a thermal noise of 126 nV/ $\sqrt{\text{Hz}}$.

Several other forms of noise are present in linear ICs and other semiconductor amplifiers to one extent or another. For example, because current flow at the quantum level is not smooth and predictable, an intermittent burst phenomenon is sometimes seen. This noise is called *popcorn noise*, and consists of pulses of many milliseconds duration. Another form of noise is *shot noise* (also called *Schottky noise*). The name 'shot' is derived from the fact that the noise sounds like a handful of B-B shot thrown against a metal surface. Shot noise is a consequence of DC current flowing in any conductor, and is found from:

$$I_n = \sqrt{2qIB} \frac{\text{A}}{\sqrt{\text{Hz}}} \quad (2.11)$$

Where:

I_n is the noise current in amperes (A)

q is the elementary electric charge (1.6×10^{-19} coulombs)

I is the current in amperes (A)

B is the bandwidth in hertz (Hz)

Finally, there is *flicker noise*, also called *pink noise* or *1/f noise*. The latter name applies because flicker noise is predominantly a low frequency (<1000 Hz) phenomenon. This type of noise is found in all conductors, and becomes important in IC devices because of manufacturing defects.

The noise spectrum in any given instrumentation system will contain elements of several kinds of noise, although in some systems one form or another may dominate the others. It is common to characterize noise from a single source using the *root mean square* (rms) value of the voltage amplitudes:

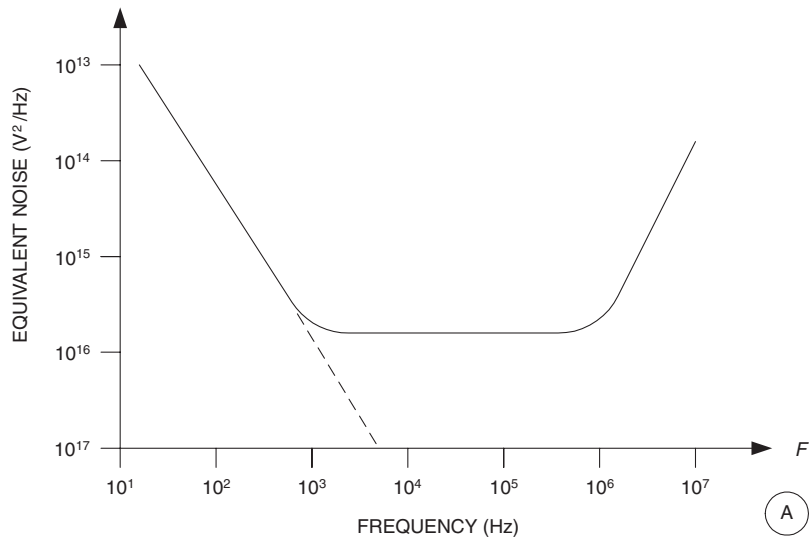


Figure 2.15 (A) Equivalent noise temperature at frequencies to 10 MHz

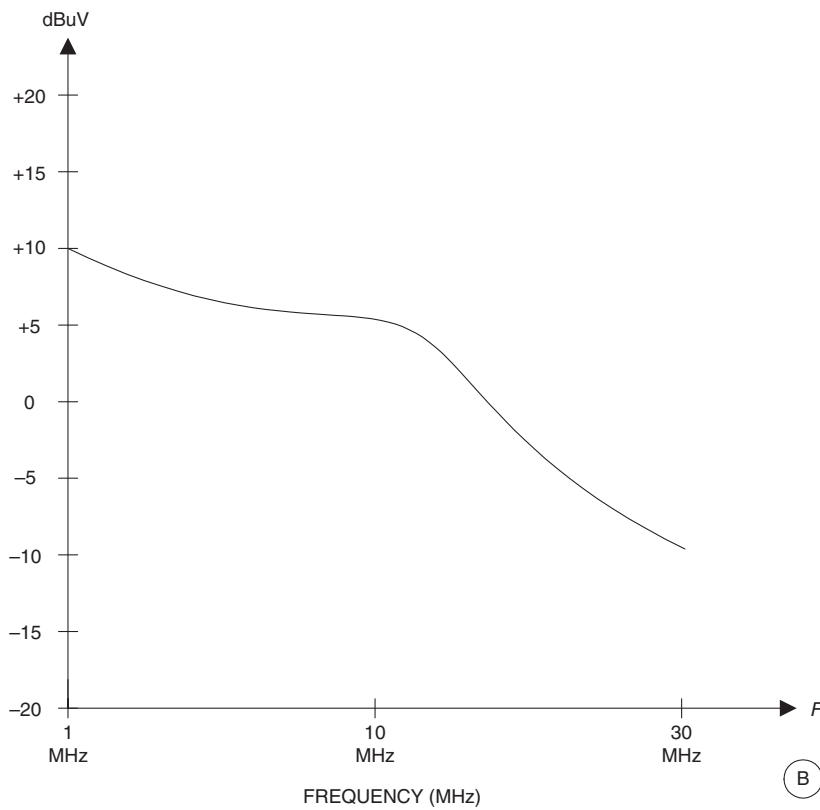


Figure 2.15 (B) universal noise sources

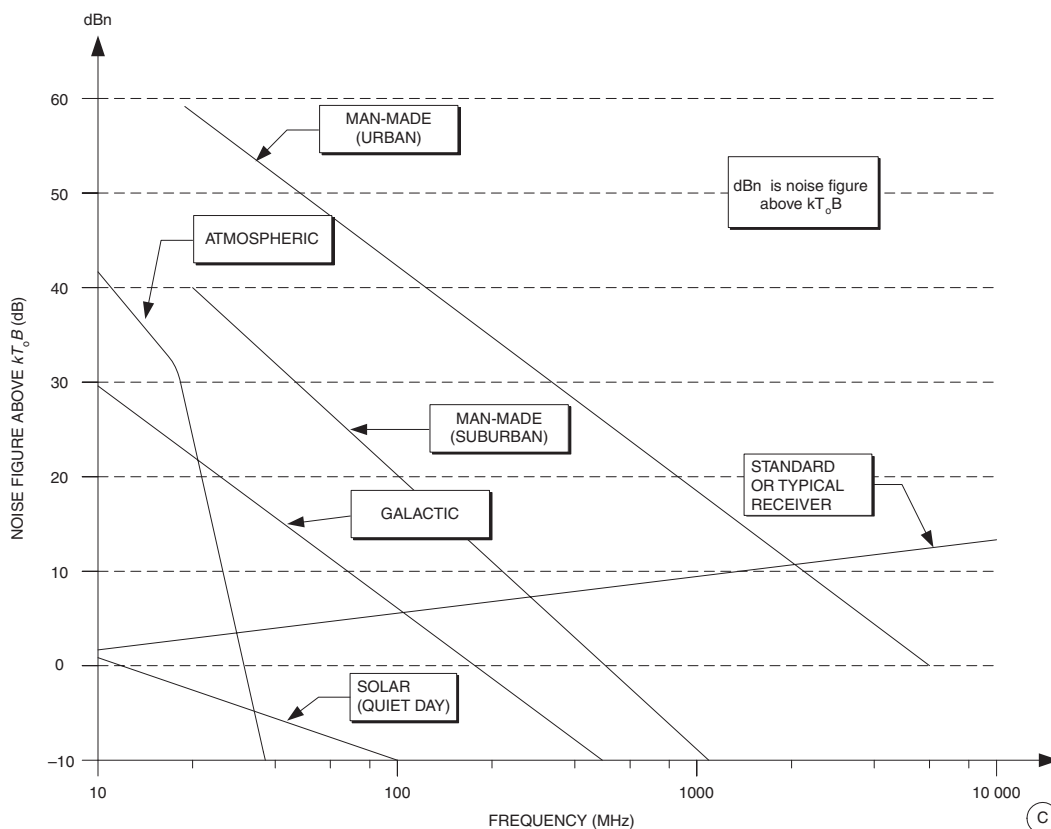


Figure 2.15 (C).

$$V_{n(fns)} = \sqrt{\frac{1}{T} \int_0^T [F(t)^2] dt} \quad (2.12)$$

Figure 2.15A shows the noise spectrum profile for a typical system that contains $1/F$ noise, thermal or ‘white’ noise, and some high frequency noise. Noise comes in a number of different guises, but in the case of radio reception we can divide noise into two classes: *sources external to the receiver* and *sources internal to the receiver*. There is little one can do about the external noise sources, for they consist of natural and man-made electromagnetic signals that fall within the passband of the receiver. Figure 2.15A shows an approximation of the external noise situation from the middle of the AM broadcast band to the low end of the VHF region. A somewhat different view, which captures the severe noise situation seen by receivers, is shown in Fig. 2.15B. One must select a receiver that can cope with external noise sources, especially if the noise sources are strong.

Some natural external noise sources are extraterrestrial. It is these signals that form the basis of radio astronomy. For example, if you aim a beam antenna at the eastern horizon prior to sunrise, a distinct rise of noise level occurs as the Sun slips above the horizon, especially in the VHF region (the 150–152 MHz band is used to measure solar flux). The

reverse occurs in the west at sunset, but is less dramatic, probably because atmospheric ionization decays much slower than it is generated. During World War II, it was reported that British radar operators noted an increase in received noise level any time the Milky Way was above the horizon, decreasing the range at which they could detect in-bound German bombers. There is also some well-known, easily observed noise from the planet Jupiter in the 18 to 30 MHz range.

Signal-to-noise ratio (SNR or S_n)

Amplifiers can be evaluated on the basis of *signal-to-noise ratio* (S/N or 'SNR'), denoted S_n . The goal of the circuit or instrument designer is to enhance the SNR as much as possible. Ultimately, the minimum signal level detectable at the output of an amplifier is that level which appears above the noise floor level. Therefore, the lower the system noise floor, the smaller the *minimum allowable signal*. Although often thought of as a radio receiver parameter, SNR is applicable in other amplifiers where signal levels are low.

Noise resulting from thermal agitation of electrons is measured in terms of *noise power* (P_n), and carries the units of power (watts or its sub-units). Noise power is found from:

$$P_n = KTB \quad (2.13)$$

Where:

P_n is the noise power in watts (W)

K is Boltzmann's constant (1.38×10^{-23} J/K)

B is the bandwidth in hertz (Hz)

Notice in Eq. (2.13) that there is no centre frequency term, only the bandwidth (B). Thus, in bandwidth limited systems, such as a practical amplifier or network, the total noise power is related to temperature and bandwidth. We can conclude that a 3000 Hz bandwidth centred on 1 MHz produces the same thermal noise level as a 3000 Hz bandwidth centred on 60 MHz or any other frequency.

Noise sources can be categorized as either *internal* or *external*. The internal noise sources are due to thermal currents in the semiconductor material resistances. It is the noise component contributed by the amplifier under consideration. If noise, or S/N ratio, is measured at both input and output of an amplifier, the output noise is greater. The

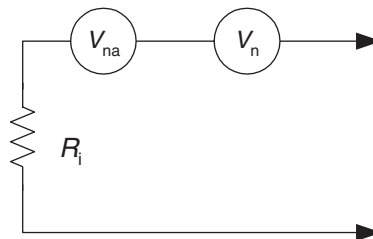


Figure 2.16 Noise and signal voltages.

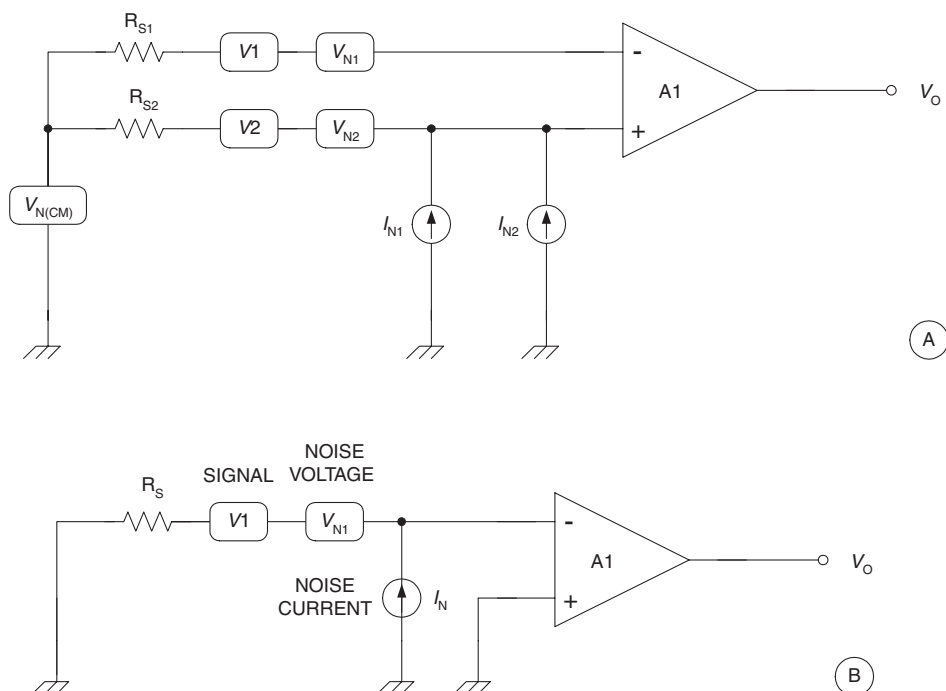


Figure 2.17 Various noise sources in amplifier circuit.

internal noise of the device is the difference between output noise level and input noise level.

External noise is the noise produced by the signal source, so is often called *source noise*. Part of this noise signal is due to thermal agitation currents in the signal source. In fact, the simple terminated noise level might be higher than V_n because of component construction. For example, the noise signal produced by a carbon composition resistor has an additional noise source modelled as V_{na} in Fig. 2.16. This noise generator is a function of resistor construction and manufacturing defects.

Figure 2.17A shows a circuit model showing that several voltage and current noise sources exist in an op-amp. The relative strengths of these noise sources, hence their overall contribution, varies with op-amp type. In an FET-input op-amp, for example, the current noise sources are tiny, but voltage noise sources are significant. For bipolar op-amps the opposite situation applies.

All of the noise sources in Fig. 2.17A are uncorrelated with respect to each other, so one cannot simply add noise voltages; only noise power can be added. To characterize noise voltages and currents they must be added in the *root sum squares* (RSS) manner.

Models such as Fig. 2.17A are too complex for most situations, so it is standard practice to lump all of the voltage noise sources into one source, and all of the current noise sources into another source. The composite sources have a value equal to the RSS voltage (or current) of the individual sources. Figure 2.17B is such a model in which only a single current source and a single voltage source are used. The *equivalent AC noise* in Fig. 2.17B

is the overall noise, given a specified value of source resistance, R_s , and is found from the RSS value of V_n and I_n :

$$V_{nt} = \sqrt{V_n^2 + (I_n R_s)^2} \quad (2.14)$$

Noise factor, noise figure and noise temperature

The noise of a system or network can be defined in three different but related ways: *noise factor* (F_n) – a simple ratio; *noise figure* (NF) – a decibel ratio; and *equivalent noise temperature* (T_e).

Noise factor (F_n)

For components such as resistors, the noise factor is the ratio of the noise produced by a real resistor to the simple thermal noise of a perfect resistor. The noise factor of a system is the ratio of output noise power (P_{no}) to equivalent input noise power (P_{ni}):

$$F_N = \frac{P_{no}}{P_{ni}} | T = 290 \text{ K} \quad (2.15)$$

In order to make comparisons easier the noise factor is always measured at the standard temperature (T_o) 290 K (standardized room temperature).

The equivalent input noise power P_{ni} is defined as the product of the source thermal noise at standard temperature (T_o) and the amplifier gain (G):

$$P_{ni} = GKBT_o \quad (2.16)$$

It is also possible to define noise factor F_n in terms of output and input S/N ratio:

$$F_n = \frac{S_{ni}}{S_{no}} \quad (2.17)$$

which is also:

$$F_n = \frac{P_{no}}{KT_o BG} \quad (2.18)$$

Where:

S_{ni} is the input signal-to-noise ratio

S_{no} is the output signal-to-noise ratio

P_{no} is the output noise power

K is Boltzmann's constant (1.38×10^{-23} J/K)

T_o is 290 kelvin

B is the network bandwidth in hertz

G is the amplifier gain

The noise factor can be evaluated in a model that considers the amplifier ideal, and therefore only amplifies through gain G the noise produced by the 'input' noise source:

$$F_n = \frac{KT_b BG + \Delta N}{KT_o BG} \quad (2.19)$$

or,

$$F_n = 1 + \frac{\Delta N}{KT_o BG} \quad (2.20)$$

Where:

ΔN is the noise added by the network or amplifier

All other terms are as defined above

Noise figure (NF)

The noise figure is a frequently used measure of an amplifier's 'goodness', or its departure from *goodness*. Thus, it is a *figure of merit*. The noise figure is the noise factor converted to decibel notation:

$$NF = 10 \log_{10} F_n \quad (2.21)$$

Where:

NF is the noise figure in decibels (dB)

F_n is the noise factor

Noise temperature (T_e)

The noise temperature is a means for specifying noise in terms of an equivalent temperature. The idea is to pretend that the noise added by the amplifier comes instead from thermal noise from the source. The noise temperature is how hot a perfect input resistor would have to be to give rise to the output noise. It is a particularly useful concept in situations where the source noise does not come from thermal noise at room temperature, e.g. a UHF or microwave antenna.

Note that the equivalent noise temperature T_e is *not* the physical temperature of the amplifier. The noise temperature is related to the noise factor by:

$$T_e = (F_n - 1)T_o \quad (2.22)$$

and to the noise figure by:

$$T_e = [10^{NF/10} - 1] \times T_o \quad (2.23)$$

Now that we have noise temperature T_e , we can also define noise factor and noise figure in terms of noise temperature:

$$F_n = \frac{T_e}{T_o} + 1 \quad (2.24)$$

and,

$$NF = 10 \log \left[\frac{T_e}{T_o} + 1 \right] \quad (2.25)$$

The total noise in any amplifier or network is the sum of internally generated and externally generated noise. In terms of noise temperature:

$$P_{n(\text{total})} = GKB(T_o + T_e) \quad (2.26)$$

Where:

$P_{n(\text{total})}$ is the total noise power

All other terms are as previously defined

Noise in cascade amplifiers

A noise signal is treated by any following amplifier as a valid input signal. Thus, in a cascade amplifier the final stage sees an input signal that consists of the original signal and noise amplified by each successive stage. Each stage in the cascade chain amplifies signals and noise from previous stages, and also contributes some noise of its own. The overall noise factor for a cascade amplifier can be calculated from *Friis' noise equation*:

$$NF = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \dots + \frac{F_N - 1}{G_1 G_2 \dots G_{N-1}} \quad (2.27)$$

Where:

NF is the overall noise factor of N stages in cascade

F_N is the noise factor of the N th stage

G_N is the gain of the N th stage

As you can see from Eq. (2.27), the noise factor of the entire cascade chain is dominated by the noise contribution of the first stage or two. Later stages are less important where noise is concerned, provided that the input stages have sufficient gain.

Noise reduction strategies

Although noise is a serious problem for the designer, especially where low signal levels are experienced, there are a number of common sense approaches to minimize the effects of noise on a system. In this section we will examine several of these methods. For example:

- 1 Keep the source resistance and the amplifier input resistance as low as possible. Using high value resistances will increase thermal noise voltage.

- 2 Total thermal noise is also a function of the bandwidth of the circuit. Therefore, reducing the bandwidth of the circuit to a minimum will also minimize noise. But this job must be done mindfully because signals have a Fourier spectrum that must be preserved for faithful reproduction or accurate measurement. The solution is to match the bandwidth to the frequency response required for the input signal.
- 3 Prevent external noise from affecting the performance of the system by appropriate use of grounding, shielding and filtering.
- 4 Use a low noise amplifier (LNA) in the input stage of the system.
- 5 For some semiconductor circuits, use the lowest DC power supply potentials that will do the job.

Noise reduction by signal averaging

If a signal is either periodic or repetitive, or can be made so, then it is possible to enhance signal-to-noise ratio (S_n) by signal averaging. The basis for this simple signal processing technique is the assumption that noise meets the definition of either random or chaotic processes. If so, then noise tends to integrate to zero or near-zero over time. If time-averaging integration is performed in a coherent manner, then a repetitive signal tends to build in value, while noise levels (being decorrelated) decrease. If we assume that the signal-to-noise ratio is:

$$S_n = 20 \log \left(\frac{V_{in}}{V_n} \right) \quad (2.28)$$

Then, for systems where $V_i < V_n$, the noise reduction by time averaging is:

$$\bar{S}_n = 20 \log \left(\frac{V_{in}}{V_n/\sqrt{N}} \right) \quad (2.29)$$

Where:

S_n is the time-averaged SNR

V_n is the unprocessed SNR

N is the number of repetitions of the signal

Example

An RF amplifier processes a $5 \mu\text{V}$ signal in the presence of a $500 \mu\text{V}$ random noise level. Calculate the unprocessed SNR, the processed SNR for 1000 repetitions of the signal, and the processing gain.

Solution:

A. *Unprocessed SNR:*

$$\begin{aligned} S_{no} &= 20 \log(V_i/V_n) \\ &= 20 \log(5 \mu\text{V}/500 \mu\text{V}) = -40 \text{ dB} \end{aligned}$$

B. *Processed SNR:*

$$\begin{aligned}\bar{S}_n &= S_n/[M]^{1/2} \\ &= (-40 \text{ dB})/[1000]^{1/2} = -1.3 \text{ dB}\end{aligned}$$

C. *Processing gain:*

$$\begin{aligned}G_p &= S_n - \bar{S}_{no} \\ &= (-1.3 \text{ dB}) - (-40 \text{ dB}) = +38 \text{ dB}\end{aligned}$$

The effect of time averaging is to increase the time required to collect data, so (by $F = 1/T$) time averaging is effectively a means of decreasing the bandwidth of the system.

Coherency is maintained in a system by ensuring that repetitive data points are processed in a consistent time relationship with respect to each other. The averaging will be triggered by a repetitive event, and that action starts the process. Data points are always matched to other data points taken at the same elapsed time after the trigger for previous iterations. For example, the i th datum point following a current sweep is paired with all other i th points from previous sweeps, and none other.

An example of signal averaging used to extract weak signals from larger noise signals is found in radar systems where multiple reflected pulses are integrated (i.e. time averaged) to improve SNR.