1

Signals and Systems

1.1 Chapter Objectives

On completion of this chapter, the reader should:

1) Be able to apply mathematical principles to waveforms.
2) Be conversant with some important terms and definitions used in telecommunications, such as root-mean-square for voltage measurements and decibels for power.
3) Understand the relationship between the time- and frequency-domain descriptions of a signal and have a basic understanding of the operation of frequency-selective filters.
4) Be able to name several common building blocks for creating more complex systems.
5) Understand the reasons why impedances need to be matched, to maximize power transfer.
6) Understand the significance of noise in telecommunication system design and be able to calculate the effect of noise on a system.

1.2 Introduction

A signal is essentially just a time-varying quantity. It is often an electrical voltage, but it could be some other quantity, which can be changed or *modulated* easily, such as radio-frequency power or optical (light) power. It is used to carry information from one end of a communications channel (the sender or transmitter) to the receiving end. Various operations can be performed on a signal, and in designing a telecommunications transmitter or receiver, many basic operations are employed in order to achieve the desired, more complex operation. For example, modulating a voice signal so that it may be transmitted through free space or encoding data bits on a wire all entail some sort of processing of the signal.

Communication Systems Principles Using MATLAB®, First Edition. John W. Leis.

Companion website: www.wiley.com/go/Leis/communications-principles-using-matlab

A voltage that changes in some known fashion over time is termed a *waveform*, and that waveform carries information as a function of time. In the following sections, several operations on waveforms are introduced.

1.3 Signals and Phase Shift

In many communication systems, it is necessary to delay a signal by a certain amount. If this delay is relative to the frequency of the signal, it is a constant proportion of the total cycle time of the signal. In that case, it is convenient to write the delay not as time, but as a phase angle relative to 360° or 2π rad (radians). As with delay, it is useful to be able to advance a signal, so that it occurs earlier with respect to a reference waveform. This may run a little counter to intuition, since after all, it is not possible to know the value of a signal at some point in the future. However, considering that a signal repetitive goes on forever (or at least, for as long as we wish to observe it), then an advance of say one-quarter of a cycle or 90° is equivalent to a delay of $90 - 360 = -270°$.

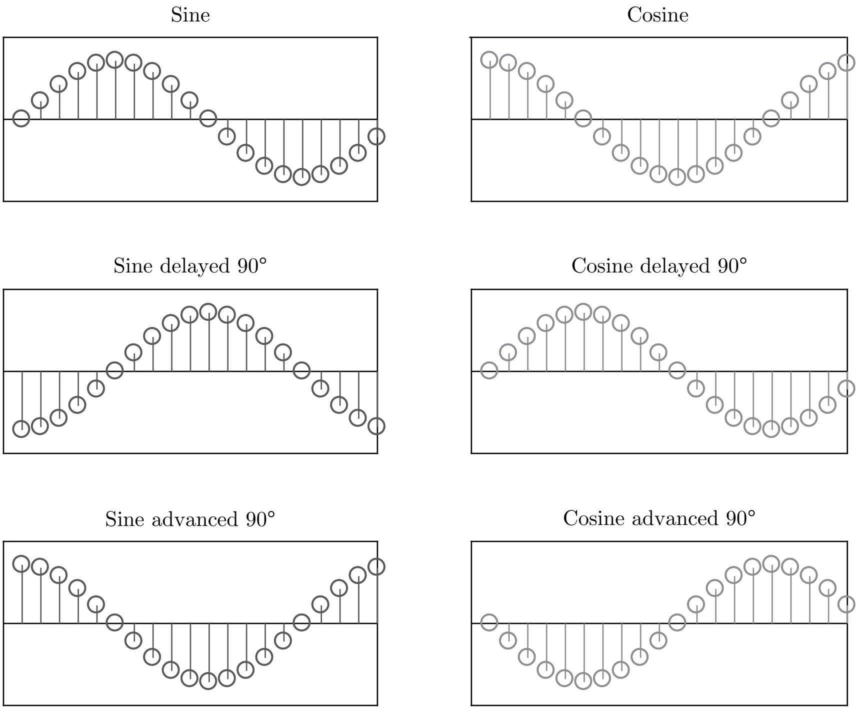

Figure 1.1 Sine and cosine, phase advance, and phase retard. Each plot shows amplitude $x(t)$ versus time t.

To see the effect of phase advance and phase delay, consider Figure 1.1, which shows these operations on both sine and cosine signals. The left panels show a sine wave, a delayed signal (moved later in time), and an advanced signal (moved earlier). The corresponding equations are

$$x(t) = \sin \omega t$$
$$x(t) = \sin\left(\omega t - \frac{\pi}{2}\right)$$
$$x(t) = \sin\left(\omega t + \frac{\pi}{2}\right)$$

Starting with a cosine signal, Figure 1.1 shows on the right the original, delayed (or retarded), and advanced signals, respectively, with equations

$$x(t) = \cos \omega t$$
$$x(t) = \cos\left(\omega t - \frac{\pi}{2}\right)$$
$$x(t) = \cos\left(\omega t + \frac{\pi}{2}\right)$$

1.4 System Building Blocks

Telecommunication systems can be understood and analyzed in terms of some basic building blocks. More complicated systems may be "built up" from simpler blocks. Each of the simpler blocks performs a specific function. This section looks initially at some simple system blocks and then at some more complex arrangements.

1.4.1 Basic Building Blocks

There are many types of blocks that can be specified according to need, but some common ones to start with are shown in Figure 1.2. The generic *input/output block* shows an input $x(t)$ and an output $y(t)$, with the input signal waveform being altered in some way on passing through. The alteration of the signal may be simple, such as multiplying the waveform by a constant A to give $y(t) = Ax(t)$. Alternatively, the operation may be more complex, such as introducing a phase delay. The *signal source* is used to show the source of a waveform – in this case, a sinusoidal wave of a certain frequency ω_o. The addition (or subtraction) block acts on two input signals to produce a single output signal, so that $y(t) = x_1(t) \pm x_2(t)$ for each time instant t. Similarly, a multiplier block produces at its output the product $y(t) = x_1(t) \times x_2(t)$.

These basic blocks are used to encapsulate common functions and may be combined to build up more complicated systems. Figure 1.3 shows two system blocks in cascade. Suppose each block is a simple multiplier – that is, the output

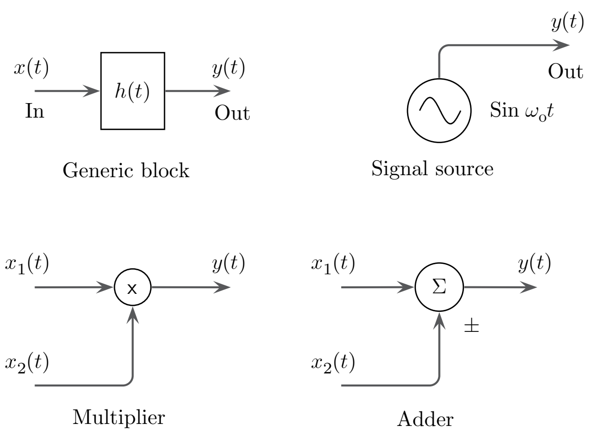

Figure 1.2 Basic building blocks: generic input/output, signal source, adder, and multiplier.

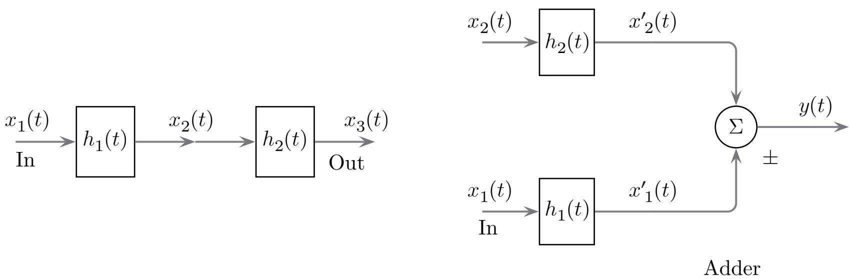

Figure 1.3 Cascading blocks in series (left) and adding them in parallel (right).

is simply the input multiplied by a gain factor. Let the gain of the $h_1(t)$ block be G_1 and that of the $h_2(t)$ block be G_2. Then, the overall gain from input to output would be just $G = G_1G_2$.

To see how it might be possible to build up a more complicated system from the basic blocks, consider the system shown on the right in Figure 1.3. In this case, the boxes are simply gain multipliers such that $h_2(t) = G_1$ and $h_2(t) = G_2$, and so the overall output is $y(t) = G_1x_1(t) + G_2x_2(t)$.

1.4.2 Phase Shifting Blocks

In Section 1.3, the concept of phase shift of a waveform was discussed. It is possible to develop circuits or design algorithms to alter the phase of a waveform, and it is very useful in telecommunication systems to be able to do this. Consequently, the use of a phase-shifting block is very convenient. Most commonly, a phase shift of $\pm 90°$ is required. Of course, $\pi/2$ radians in the phase angle is equivalent to $90°$. As illustrated in the block diagrams of

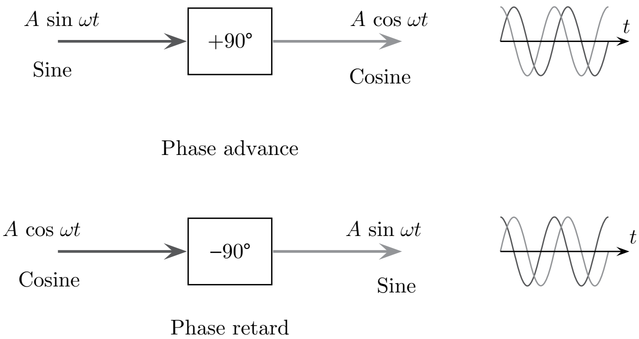

Figure 1.4 Phase shifting blocks. Note the input and output equations.

Figure 1.4, we use +90° to mean a *phase advance* of 90° and, similarly, −90° to mean a *phase delay* of 90°.

1.4.3 Linear and Nonlinear Blocks

Let us examine more closely what happens when a signal is passed through a system. Suppose for the moment that it is just a simple DC voltage. Figure 1.5 shows a transfer characteristic, which maps the input voltage to a corresponding output voltage. Two input values separated by δx, with corresponding outputs separated by δy, allow determination of the *change* in output as a function of the *change* in input. This is referred to as the *gain* of the system.

Suppose such a linear transfer characteristic with zero offset (that is, it passes through $x = 0, y = 0$) is subjected to a sinusoidal input. The output $y(t)$ is a linear function of input $x(t)$, which we denote as a constant α. Then,

$$y(t) = \alpha\, x(t) \tag{1.1}$$

With input $x(t) = A \sin \omega t$, the output will be

$$y(t) = \alpha\, A \sin \omega t \tag{1.2}$$

Thus the change in output is simply in proportion to the input, as expected.

This linear case is somewhat idealistic. Usually, toward the maximum and minimum range of voltages which an electronic system can handle, a characteristic that is not purely linear is found. Typically, the output has a limiting or saturation characteristic – as the input increases, the output does not increase directly in proportion at higher amplitudes. This simple type of nonlinear behavior is illustrated in Figure 1.6. In this case, the relationship between the input and output is not a simple constant of proportionality – though note that if the input is kept within a defined range, the characteristic may well be approximately linear.

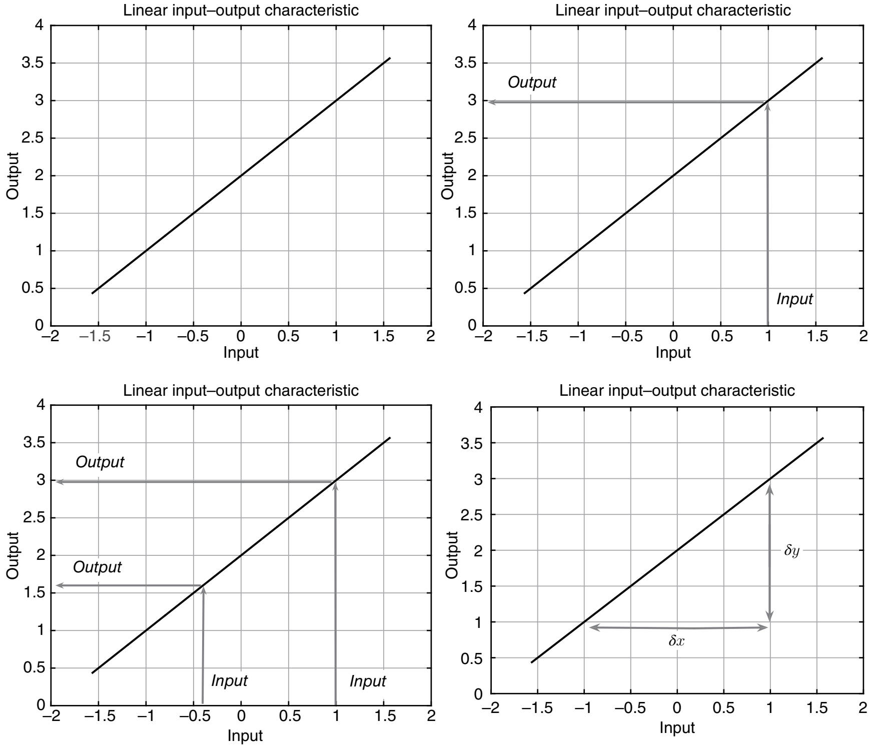

Figure 1.5 The process of mapping an input (horizontal axis) to an output (vertical), when the block has a linear characteristic. The constant or DC offset may be zero, or nonzero as illustrated.

To fix ideas more concretely, suppose the characteristic may be represented by a quadratic form, with both a linear constant multiplier α and a small amount of signal introduced that is proportional to the square of the input, via constant β. If the input $x(t)$ is again a sinusoidal function, the output may then be written as

$$\begin{aligned} y(t) &= \alpha\ x(t) + \beta x^2(t) \\ &= \alpha\ A \sin \omega t + \beta\ A^2 \sin^2 \omega t \end{aligned} \tag{1.3}$$

This is straightforward, but what does the sinusoidal squared term represent? Using the trigonometric identities

$$\cos(a + b) = \cos a \cos b - \sin a \sin b \tag{1.4}$$

$$\cos(a - b) = \cos a \cos b + \sin a \sin b \tag{1.5}$$

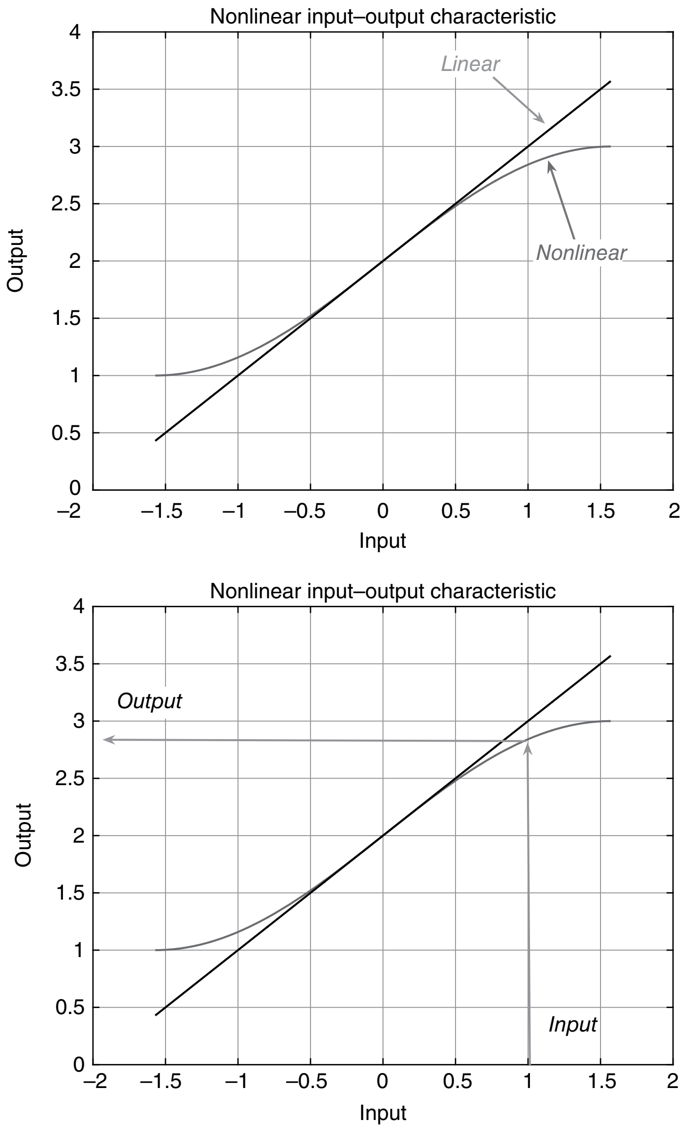

Figure 1.6 Example of mapping an input (horizontal axis) to an output (vertical), when the block has a nonlinear characteristic. Other types of nonlinearity are possible, of course.

we have by subtracting the first from the second, and then putting $b = a$,

$$\begin{aligned} \sin a \sin b &= \frac{1}{2}[\cos(a-b) - \cos(a+b)] \\ \therefore \qquad \sin^2 a &= \frac{1}{2}[\cos(a-a) - \cos(a+a)] \\ &= \frac{1}{2}(1 - \cos 2a) \end{aligned} \tag{1.6}$$

After application of this relation, and simplification, the output may be written as

$$y(t) = \alpha\, A \sin \omega t + \frac{1}{2}\beta\, A^2(1 - \cos 2\omega t) \tag{1.7}$$

This can be broken down into a constant or DC term, a term at the input frequency, and a term at twice the input frequency:

$$y(t) = \overbrace{\alpha\, A \sin \omega t}^{\text{Linear term}} + \overbrace{\frac{1}{2}\beta\, A^2}^{\text{Constant term}} - \overbrace{\frac{1}{2}\beta\, A^2 \cos 2\omega t}^{\text{Double-frequency term}} \tag{1.8}$$

This is an important conclusion: the introduction of nonlinearity to a system may affect the frequency components present at the output. A linear system always has frequency components at the output of the exact same frequency as the input. A nonlinear system, as we have demonstrated, may produce harmonically related components at other frequencies.

1.4.4 Filtering Blocks

A more complicated building block is the frequency-selective filter, almost always just called a *filter*. Typically, a number of filters are used in a telecommunication system for various purposes. The common types are shown in Figure 1.7. The sine waves (with and without cross-outs) shown in the middle of each box are used to denote the operation of the filter in terms of frequency selectivity. For example, the lowpass filter shows two sine waves, with the lower one in the vertical stack indicating the lower frequency. The higher frequency is crossed out, thus leaving only lower frequency components. Representative input and output waveforms are shown for each filter type. Consider, for example, the bandpass filter. Lower frequencies are attenuated (reduced in amplitude) when going from input to output. Intermediate frequencies are passed through with the same amplitude, while high frequencies are attenuated. Thus, the term *bandpass filter* is used. Filters defining highpass and bandstop operation may be designated in a similar fashion, and their operation is also indicated in the figure.

When it comes to more precisely defining the operation of a filter, one or more cutoff frequencies have to be specified. For a lowpass filter, it is not sufficient to say merely that "lower" frequencies are passed through unaltered. It

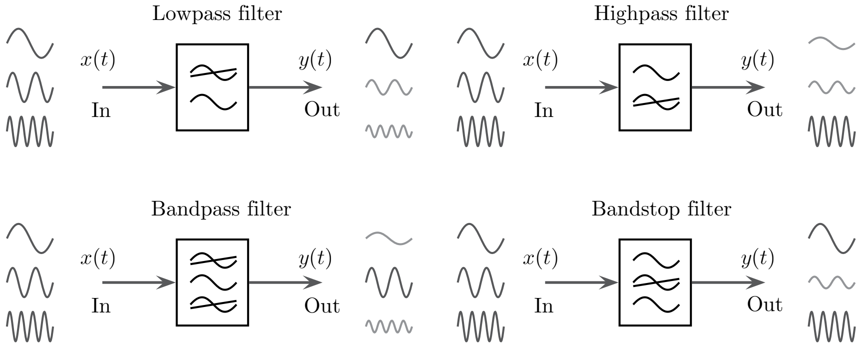

Figure 1.7 Some important filter blocks and indicative time responses. The waveforms and crossed-out waveforms in the boxes, arranged high to low in order, represent high to low frequencies. Input/ouput waveform pairs represent low, medium, and high frequencies, and the amplitude of each waveform at the output is shown accordingly.

is necessary to specify a boundary or cutoff frequency ω_c. Input waveforms whose frequency is below ω_c are passed through, but (in the ideal case) frequencies above ω_c are removed completely. In mathematical terms, the lower frequencies are passed through with a gain of one, whereas higher frequencies are multiplied by a gain of zero.

The operation of common filters may be depicted in the frequency domain as shown in the diagrams of Figure 1.8. First, consider the lowpass filter. This type of filter would, ideally, pass all frequencies from zero (DC) up to a specified cutoff frequency. Ideally, the gain in the *passband* would be unity, and the gain in the *stopband* would be zero. In reality, several types of imperfections mean that this situation is not always realized. The *order* of the filter determines how rapidly the response changes from one gain level to another. The order of a filter determines the number of components required for electronic filters or the number of computations required for a digitallyprocessed filter.

A low-order filter, as shown on the left, has a slower transition than a high-order filter (right). In any given design, a tradeoff must be made between a lower-cost, low-order filter (giving less rapid passband-to-stopband transitions) and a more expensive high-order filter.

Lowpass filters are often used to remove noise components from a signal. Of course, if the noise exists across a large frequency band, a filter can only remove or attenuate those components in its stopband. If the frequency range of the signal of interest also contains noise, then a simple filter cannot differentiate the desired signal from the undesired one.

In a similar fashion, a highpass filter may be depicted as also shown in Figure 1.8. As we would expect, this type of filter passes frequencies that are higher than some desired cutoff. A hybrid characteristic leads to a bandpass filter or bandstop filter. These types of filters are used in telecommunication

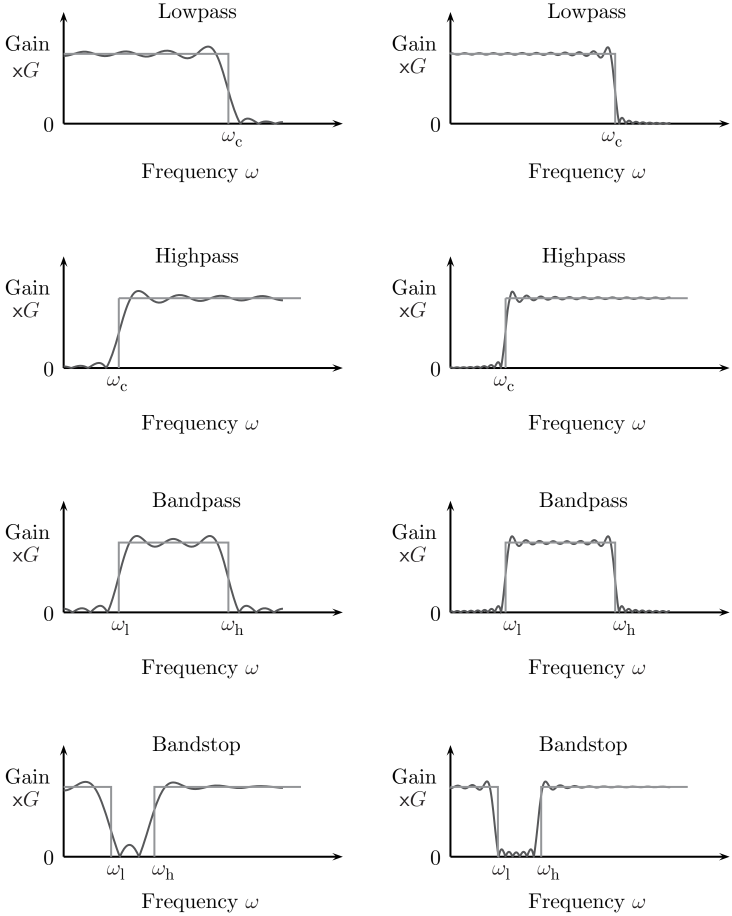

Figure 1.8 Primary filter types: lowpass, highpass, bandpass, and bandstop, with a low-order filter shown on the left and higher-order on the right. Ideally, the passband has a fixed and finite signal gain, whereas the stopband has zero gain.

systems for special purposes. For example, the bandstop filter may be used to remove interference at a particular frequency, and a bandpass filter may be used to pass only a particular defined range of frequencies (a channel or set of channels, for example).

1.5 Integration and Differentiation of a Waveform

This section details two signal operations that are related to fundamental mathematical operations. First, there is *integration*, which in terms of signals means

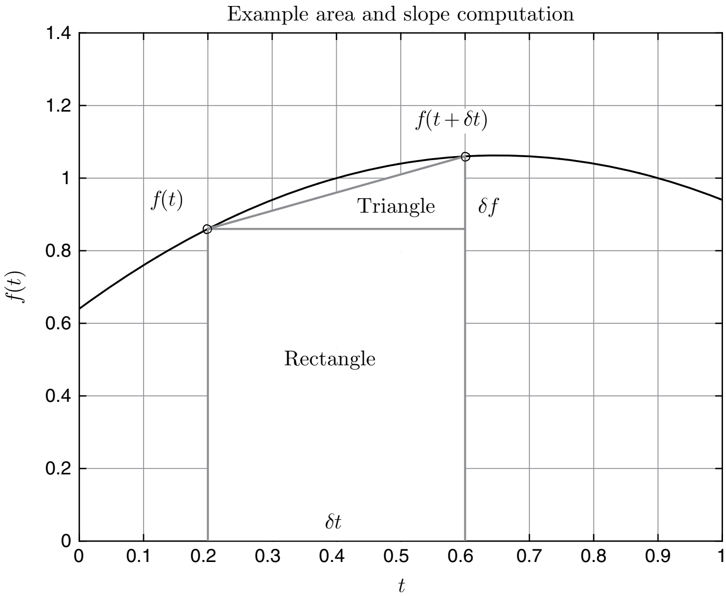

Figure 1.9 Calculating the area over a small time increment δt using a rectangle and the slope of the curve using a triangle.

the cumulative or sum total of a waveform over time. The opposite operation, *differentiation*, essentially means the rate of change of the voltage waveform over time. These are really just the two fundamental operations of calculus: Integration and differentiation. These are the inverse of each other, as will be explained. This intuition is useful in understanding the signal processing for communication systems presented in later chapters. The functions are presented in terms of time t, as this is the most useful formulation when dealing with time-varying signals.

Figure 1.9 shows the calculation of the area (integral) and slope (derivative) for two adjacent points. At a specific time t, the function value is $f(t)$, and at a small time increment δt later, the function value is $f(t + \delta t)$. The area (or actually, a small increment of area) may be approximated by the area of the rectangle of width δt and height $f(t)$. This small increment of area δA is

$$\delta A \approx f(t)\ \delta t \tag{1.9}$$

It could be argued that this approximation would be more accurate if the area of the small triangle as indicated were taken into account. This additional area would be the area of the triangle or $(1/2)(\delta t\ \delta f)$, which would diminish rapidly as the time increment gets smaller ($\delta t \to 0$). This is because it is not one small quantity δt, but the product of two small quantities $\delta t\ \delta f$.

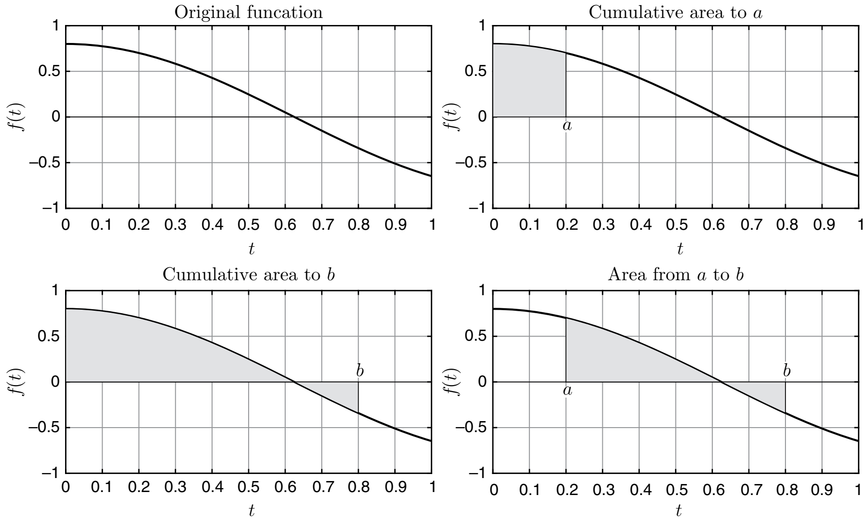

Figure 1.10 A function $f(t)$, calculating its cumulative area to a and b, and the area between $t = a$ and $t = b$. Note the negative portions of the "area" below the $f(t) = 0$ line.

Similarly, the slope at point $(t, f(t))$ is $\delta f/\delta t$. This is the instantaneous slope or derivative, which of course varies with t, since $f(t)$ varies. This slope may be approximated as the slope of the triangle, which changes from $f(t)$ to $f(t + \delta t)$ over a range δt. So the slope is

$$\frac{\delta f}{\delta t} \approx \frac{f(t + \delta t) - f(t)}{\delta t} \tag{1.10}$$

The calculation of the derivative or slope of a tangent to a curve is a point-by-point operation, since the slope will change with $f(t)$ and hence the t value (the exception being a constant rate of change of value over time, which has a constant slope). The integral or area, though, depends on the range of t values over which we calculate the area. Since the integral is a continuous function, it extends from the left from as far back as we wish to the right as far as we decide. Figure 1.10 shows a function and its integral from the origin to some point $t = a$ (note that we have started this curve at $t = 0$, but that does not have to be the case). In the lower-left panel, we extend the area to some point $t = b$. This is essentially the same concept, except that the area below the horizontal $f(t) = 0$ line is in fact negative. While the concept of "negative area" might not be found in reality, it is a useful concept. In this case, the negative area simply subtracts from the positive area to form the net area. Finally, the lower-right panel illustrates the fact that the area from $t = a$ to $t = b$ is simply

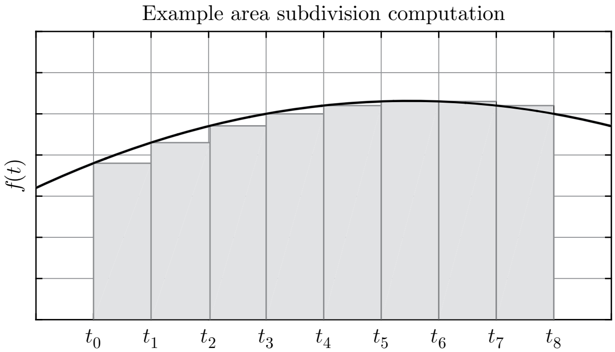

Figure 1.11 Calculating area using a succession of small strips of width δt.

the area to $t = b$, less the area to $t = a$. Mathematically, this is written as

$$\int_a^b f(t)\,\mathrm{d}t = F(b) - F(a) \tag{1.11}$$

where $F(\cdot)$ represents the cumulative area to that point. This is called the definite integral – an integration or area calculation with definite or known start and end boundaries.[1]

The area may be approximated by creating successive small strips of width δt as before, and joining enough of them together to make the desired area. This is illustrated in Figure 1.11, for just a few subdivisions. Using the idea of $F(t)$ as the cumulative area function under the curve $f(t)$, consider the area under the curve from t to $t + \delta t$, where δt is some small step of time. The *change* in area over that increment is

$$\delta A = F(t + \delta t) - F(t) \tag{1.12}$$

Also, the change in area is *approximated* by the rectangle of height $f(t)$ and width δt, so

$$\delta A = f(t)\,\delta t \tag{1.13}$$

Equating this change of area δA,

$$f(t)\,\delta t = F(t + \delta t) - F(t) \tag{1.14}$$

$$f(t) = \frac{F(t + \delta t) - F(t)}{\delta t} \tag{1.15}$$

1 The $\int$ symbol comes from the "long s" of the 1700s, so you can see the connection with the idea of "summation."

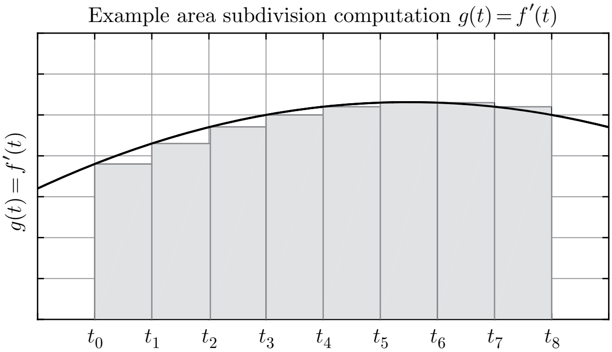

Figure 1.12 The area under a curve $g(t)$, but the curve happens to be the derivative of $f(t)$.

This is the same form of equation we had earlier for the definition of slope. Now, it is showing that the slope of some function $F(t)$, which happens to be the integral or area under $f(t)$, is actually equal to $f(t)$. That is, **the derivative of the integral equals the original function**. This is our first important conclusion.

Next, consider how to calculate the cumulative area by subdividing a curve $f(t)$ into successive small strips. However, instead of the plain function $f(t)$, suppose we plot its derivative, $f'(t)$ instead. This is illustrated in Figure 1.12, for just a few strips of area from t_0 at the start to an ending value t_8.

The *cumulative* area (call it $A(t)$) under this curve $f'(t)$ – which we defined to be the derivative of $f(t)$ – is the summation of all the individual rectangles, which is

$$A(t) = \delta t\, f'(t_0) + \delta t\, f'(t_1) + \cdots + \delta t\, f'(t_{n-1}) \tag{1.16}$$

$$= \delta t\ [f'(t_0) + f'(t_1) + \cdots + f'(t_{n-1})] \tag{1.17}$$

Now we can use the same concept for slope as developed before, where we had the approximation to the derivative

$$f'(t) = \frac{f(t+\delta t) - f(t)}{\delta t} \tag{1.18}$$

Substituting this for all the derivative terms, we have

$$A(t) = \delta t\ \left\{ \left[\frac{f(t_0+\delta t) - f(t_0)}{\delta t}\right] + \left[\frac{f(t_1+\delta t) - f(t_1)}{\delta t}\right] + \cdots + \left[\frac{f(t_{n-1}+\delta t) - f(t_{n-1})}{\delta t}\right] \right\} \tag{1.19}$$

Canceling the δt and using the fact that each $t_k + \delta t$ is actually the next point t_{k+1} (for example, $t_1 = t_0 + \delta t$, $t_2 = t_1 + \delta t$), we can simplify things to

$$A(t) = \{[f(t_1) - f(t_0)] + [f(t_2) - f(t_1)] + \cdots + [f(t_n) - f(t_{n-1})]\} \qquad (1.20)$$

Looking carefully, we can see terms that will cancel, such as $f(t_1)$ in the first square brackets, minus the same term in the second square brackets. All these will cancel, except for the very first $-f(t_0)$ and the very last $f(t_n)$ to leave us with

$$A(t) = f(t_n) - f(t_0) \qquad (1.21)$$

So this time, we have found that the area under some curve $f'(t)$ (which happens to be the derivative or slope of $f(t)$) is actually equal to the original $f(t)$. That is, the *area under the slope curve* equals *the original function evaluated at the end* (right-hand side), less any start area. The subtraction of the start area seems reasonable, since it is "cumulative area to b less cumulative area to a," as we had previously. Thus, our second important result is that **the integral of a derivative equals the original function**.

We can see the relationship between differentiation and integration at a glance in the following figures. Figure 1.13 shows taking a function (top) and integrating it (lower); if we then take this integrated (area) function as shown

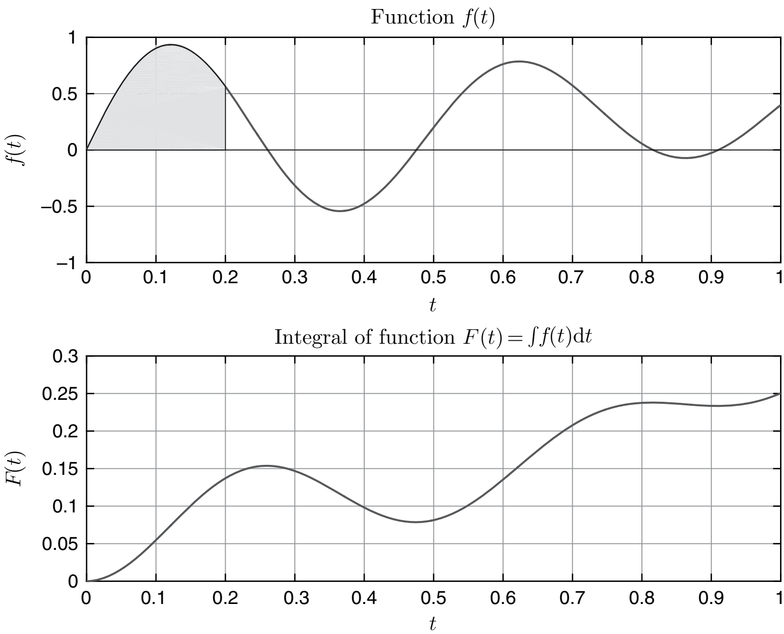

Figure 1.13 The cumulative area under $f(t)$. Each point on $F(t)$ represents the area up to the right-hand side of the shaded portion at some value of t (here $t = 0.2$ for the shaded portion). Note that when $f(t)$ becomes negative, the area reduces.

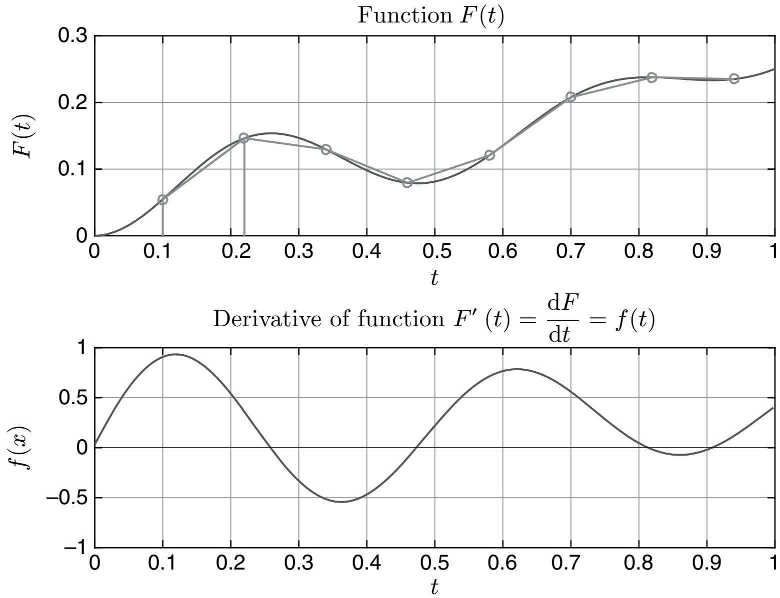

Figure 1.14 The derivative of $f(t)$ as a function. It may be approximated by the slopes of the lines as indicated, though the spacing is exaggerated for the purpose of illustration.

in Figure 1.14 (top) and then take the derivative of that (Figure 1.14, lower), *we end up with the original function* that we started with. And the process is invertible: Take the derivative of the top function in Figure 1.14 to obtain the lower plot of Figure 1.14. Transferring this to the top of Figure 1.13, and then integrating it, we again end up where we started: the original function. So it is reasonable to say that integration and differentiation are the inverse operations of each other. We just have to be careful with the integration, since it is cumulative area, and that may or may not have started from zero at the leftmost starting point.

1.6 Generating Signals

Communication systems invariably need some type of waveform generation in their operation. There are numerous methods of generating sinusoids, which have been devised over many years, and each has advantages and disadvantages. The ability to generate not just one, but several possible frequencies (that is, to tune the frequency), is a desirable attribute. So too is the spectral purity of the waveform: How close it is to an ideal sine function. One method, which is relatively simple, has a tunable frequency, and can generate a wide range of

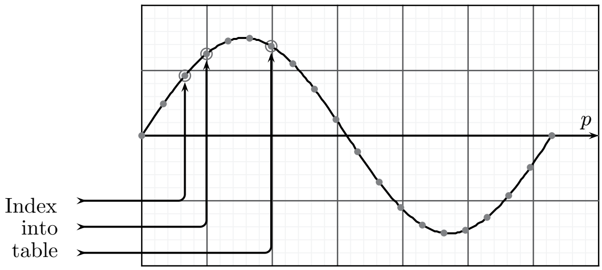

Figure 1.15 Generating a sinusoid using an index p into a table. The value at each index specifies the required amplitude at that instant.

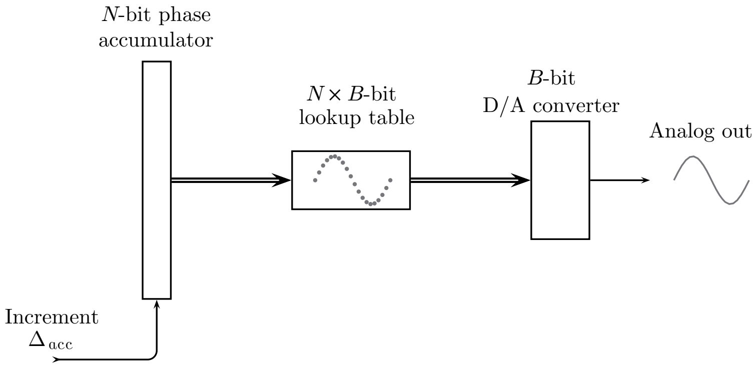

Figure 1.16 Using a lookup table to generate a waveform. Successive digital (binary-valued) steps are used to index the table. The digital-to-analog (D/A) converter transforms the sample value into a voltage.

possible frequencies, is the Direct Digital Synthesizer (DDS), whose working principle was originally introduced in Tierney et al. (1971).

Computing the actual samples of a sine function is often not feasible in real time for high frequencies. However, precomputing the values and storing in a table – a Lookup Table or LUT – is possible. Stored-table sampling with indexing is illustrated in Figure 1.15. Effectively, the index of each point in the table is the phase value, and each point's value represents the amplitude at that particular phase. All that is then required is to step through the table with a digital counter as shown in Figure 1.16.

The number of points on the waveform determines the accuracy and also the resolution of frequency tuning steps. This resolution is the clock frequency f_{clk} divided by the number of points 2^N, where N is the number of bits in the address counter. However, this also requires a table of size 2^N. It follows that for finer frequency tuning steps, N should be as large as possible.

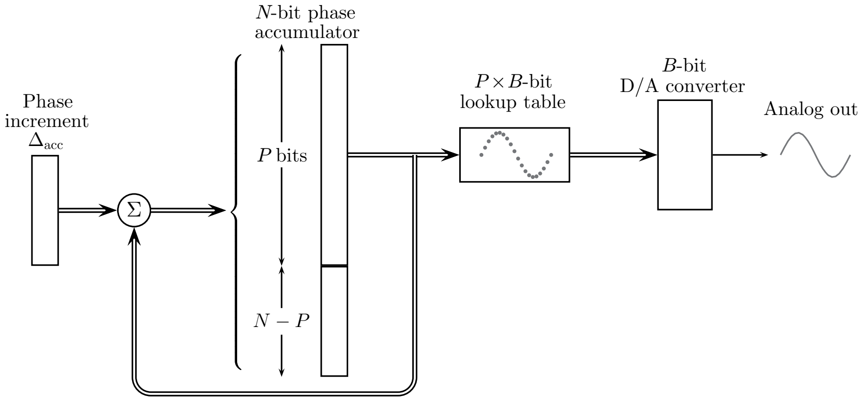

Figure 1.17 A Direct Digital Synthesizer (DDS) using a reduced lookup table. Samples are produced at a rate of f_s and for each new sample produced, a phase step of Δ_{acc} is added to the current index p to locate the next sample value in the Lookup Table (LUT).

In order to reduce the size of the lookup table, a compromise is to employ a smaller table, which is indexed by only the upper P bits of the phase address counter. This is shown in Figure 1.17. In order to compute the next point on the waveform, a phase increment Δ_{acc} is added for each point in the generated waveform. A smaller Δ_{acc} means that the table is stepped through more slowly, hence resulting in a lower frequency waveform. Conversely, a larger Δ_{acc} means the table is stepped through more rapidly, resulting in a higher frequency waveform. The tradeoff in using a smaller LUT means that the preciseness of the waveform is reduced, which is shown in Figure 1.18 for a small table size.

An interesting problem then arises. If the phase accumulator step Δ_{acc} is a power of 2, then at the end of the LUT, the counter will wrap back to the same relative starting position. The only problem with the output frequency spectrum will be the harmonics generated by stepping through at a faster rate, and these harmonics will not vary over time. However, if the step is such that, upon reaching the end of the table, the addition of the step takes the pointer back to a different start position, the next cycle of the waveform will be generated from a slightly different starting point. This means that there will be some jitter in the output waveform, and the frequency spectrum will contain additional phase noise components, as shown in Figure 1.19.

The DDS structure is able to generate multiple waveforms by using multiple index pointers. For example, sine and cosine may be generated by offsetting one pointer by the equivalent of a quarter of a cycle in the table. The phase and frequency are also easily changed by changing the relative position of the index pointer, and this is useful for generating modulated signals (discussed in Chapter 3).

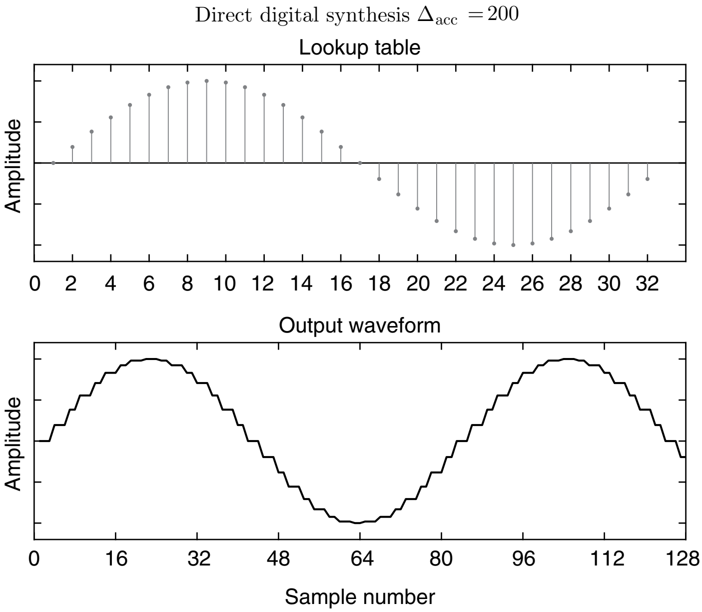

Figure 1.18 A lookup table (top) with $2^P = 32$ entries, requiring $P = 5$ bits. One possible waveform generated by stepping through at increments of $\Delta_{acc} = 200$ is shown below, when the total phase accumulator has $N = 14$ bits.

1.7 Measuring and Transferring Power

This section discusses the concept of the power transferred by a signal and the related concept of impedance of a circuit. The notion of power is important in telecommunications, since how much power is used to send a signal is clearly important, how far can a signal travel and how much power is enough are relevant questions. The impedance of a circuit appears a great deal in discussions about power and information transfer. It basically describes how much a current flow is "impeded" along its way.

1.7.1 Root Mean Square

Sinusoidal signals have their amplitude determined directly by the factor A in the equation of a sinusoid, $x(t) = A\sin(\omega t + \varphi)$. However, not all signals are pure sinusoids. It is useful to have a definition of power, which is not dependent on the wave shape of the underlying signal.

One of the most commonly used is the RMS, or Root Mean Square. This means that first, we square the signal and then take the mean or average of that result. This is necessary so as to measure power over a normalized time interval.

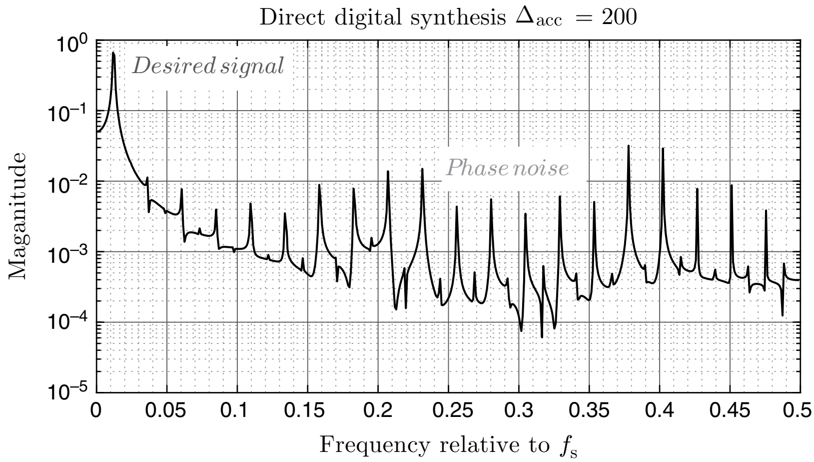

Figure 1.19 The frequency spectrum of the waveform, showing the magnitude of each signal component. Ideally, only one component should be present, but the stepping approach means that other unwanted components with smaller magnitudes are also produced. Note that the vertical amplitude scale is logarithmic, not linear.

Finally, to "undo" the squaring operation, we take the square root. Graphically, Figure 1.20 illustrates this operation for a sine wave. The first step is to square the waveform, which means that negative values are converted into positive, since squaring a negative value results in a positive result.

The second step after squaring the waveform is to add up all the squared values, as illustrated in Figure 1.21. This diagram shows individual bars or samples of the waveform in order to illustrate the point – in reality, the signal has no discontinuities. Next, we divide by the time we have averaged over. In the illustration, this is exactly one cycle of the wave. If need not we do say 2 or 100 cycles, then the summation would be correspondingly larger, and dividing by the number of samples (in the discrete-bar case) or the total time (for the continuous wave) would normalize things out. Finally, we take the square root of this quantity, and we have the RMS value.

We can calculate this mathematically for known signals. A simple and commonly used case is the pure sine wave, and to work this out let the period be

$$\tau = \frac{2\pi}{\omega_o} \tag{1.22}$$

where ω_o is the radian frequency (rad s^{-1}). To convert from Hertz frequency f to radian frequency ω, the formula $\omega = 2\pi f$ is used, where f is in Hertz, or cycles per second, and ω is in radians per second. The equation of the sine wave is

$$x(t) = A \sin \omega t \tag{1.23}$$

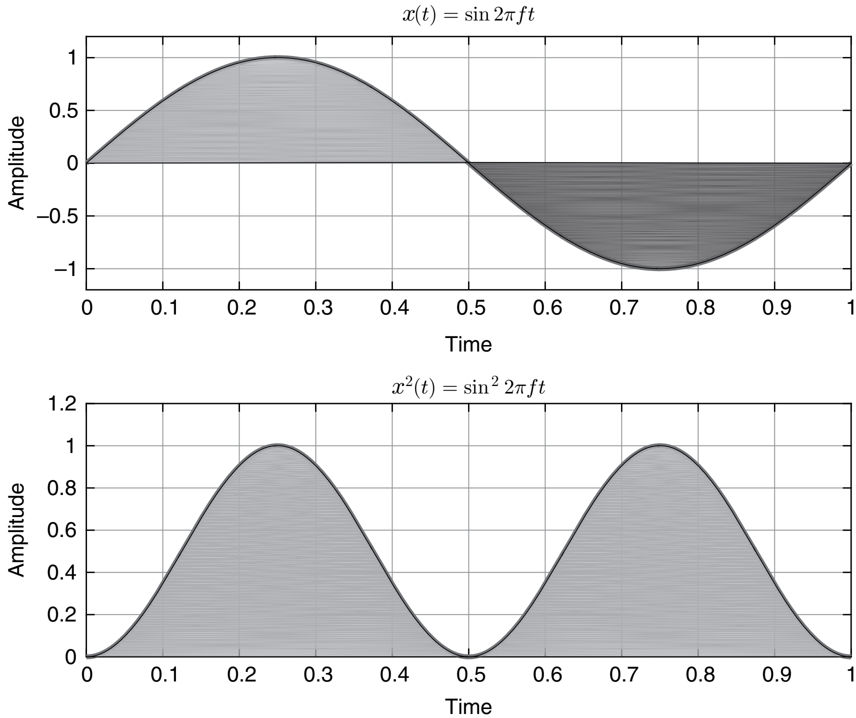

Figure 1.20 Graphical illustration of the calculation of RMS value. Squaring the waveform at the top results in the lower waveform.

Squaring gives

$$x^2(t) = A^2 \sin^2 \omega t \tag{1.24}$$

In order to calculate the mean square over one period τ, we need to integrate the squared waveform

$$\overline{x^2} = \frac{1}{\tau} \int_0^\tau A^2 \sin^2 \omega t \, \mathrm{d}t \tag{1.25}$$

Evaluating this integral, we find that the mean-square value of a sine wave is

$$\overline{x^2} = \frac{A^2}{2} \tag{1.26}$$

The RMS is just the square root of this, or

$$\text{RMS } \{x(t)\} = \frac{A}{\sqrt{2}} \tag{1.27}$$

This is a very common result. It tells us that the RMS value of a sine wave is the peak divided by $\sqrt{2}$, or approximately 1.4. Equivalently, the peak is multiplied

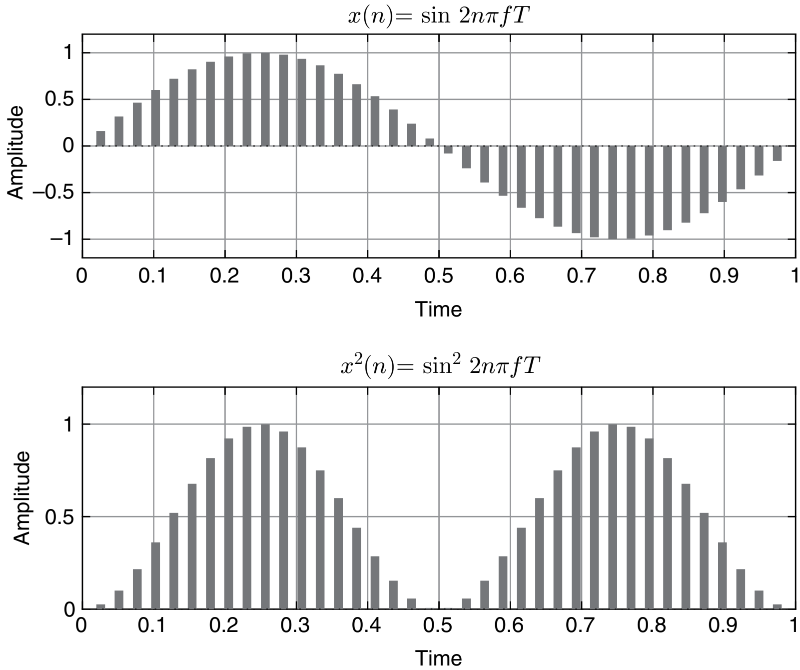

Figure 1.21 Imagining RMS calculation as a series of bars, with each bar equal to the height of the waveform at that point. The period between samples is T, with sample index n. The substitution required is then $t = nT$.

by $1/\sqrt{2} \approx 0.7$ to obtain the RMS value. Alternatively, if we know the RMS value, we multiply it by $\sqrt{2} \approx 1.4$ to find the peak value. The following MATLAB code shows how to generate a sine wave and calculate the RMS value from the peak.

```
% waveform parameters
dt = 0.01;
tmax = 2;
t = 0:dt:tmax;
f = 2;

% generate the signal
x = 1*sin(2*pi*f*t);
plot(t,x);

% calculate the signal's RMS value
sqrt((sum(x.*x)*dt)/tmax)
ans =
```

```
    0.7071

% it is a known factor
1/sqrt(2)
ans =
    0.7071
```

So, what use is the RMS value? Even though we calculated a mathematical expression relating the amplitude of a sine wave to its RMS value, the concept is applicable to *any* waveform. It gives us a measure of the power that the signal can deliver. If, for example, we simply averaged the waveform, then a sine wave would yield a figure of zero (since it is symmetrical about the time axis). This is not a very useful result. In the next section, it is demonstrated that the RMS value may be related to another quantity, termed the *decibel*, which is commonly used in telecommunication systems.

1.7.2 The Decibel

Another quantity that is frequently encountered in telecommunications is the *decibel* (dB). It is used in different contexts. One is to show the power of a signal, and in that way it might be regarded as similar to the RMS value mentioned above. Another context in which the decibel is used is to measure the gain or loss of a communication processing block, such as an amplifier.

The first use is to denote power, or more precisely, power relative to some reference value. For a power P, the relative *power in decibels* is calculated as

$$P_{\mathrm{dB}} = 10\log_{10}\left(\frac{P}{P_{\mathrm{ref}}}\right) \tag{1.28}$$

where P_{ref} is the reference power. There are several important points to note about this formula. First, it does not measure absolute power as such, but rather power relative to a defined reference power level. Secondly, we use the logarithm to base 10 in the computation of the decibel. The relative power is usually a standard amount, in which case standard symbols are used to denote this. For example, dBW is used when the reference power P_{ref} is 1W (Watt) and dBm when the reference power P_{ref} is 1mW (milliwatt), or 1×10^{-3} W.

The concept of power is meaningless in a practical sense unless it is applied to a load. The load must have a certain impedance. Suppose we had a purely resistive load of 50Ω. Power is $P = IV$ and Ohm's law is $V = IR$, and so power is V^2/R. Thus for a power of 1mW, we have

$$\begin{aligned} P &= \frac{V^2}{R} \\ \therefore V &= \sqrt{P \times R} \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{1\text{mW} \times 50\Omega} \\
&= \sqrt{0.05} \\
&= 0.2236\text{V} \approx 223\text{mV} \qquad (1.29)
\end{aligned}
$$

This is the voltage needed across the load resistance to develop the given amount of power. Note that the voltage is RMS, not peak amplitude.

The second common use of the decibel is in measuring the gain of a system. That is to say, given an input power P_{in}, and a corresponding output power P_{out}, the *power gain* is defined as

$$G_{\text{dB}} = 10\log_{10}\left(\frac{P_{\text{out}}}{P_{\text{in}}}\right) \qquad (1.30)$$

The basic formula is similar, taking the logarithm of a power ratio and then multiplying by 10 (the "deci" part). What was the reference power in the previous example has now become the input power. This is not unreasonable, since the "reference" is at the input to the system we are considering.

Suppose a system has a power gain of 2. That is, the output power is twice the input power. The power gain in dB is

$$
\begin{aligned}
G_{\text{dB}} &= 10\log_{10}2 \\
&\approx 3 \text{ dB} \qquad (1.31)
\end{aligned}
$$

Now suppose another system has a power gain of $1/2$. In that case, the power gain in dB is

$$
\begin{aligned}
G_{\text{dB}} &= 10\log_{10}\frac{1}{2} \\
&\approx -3 \text{ dB} \qquad (1.32)
\end{aligned}
$$

Notice how these are the same values, but negated. This gives us a clue as to one of the useful properties of decibels: increasing the power is a positive dB figure, whereas decreasing is a negative dB figure. So what about the same power for input and output, when $P_{\text{out}} = P_{\text{in}}$? It is not hard to show that this gives a figure of 0 dB.

A common use of the decibel is to state the *voltage* gain of a circuit or system in decibels rather than the power gain. Suppose we have two power flows P_{out} and P_{in} as above and that they each drive a load resistance of R. We can determine the voltage at the input and output using $P = V^2/R$ as before, and noting that $\log x^a = a \log x$, the decibel ratio becomes

$$
\begin{aligned}
G_{\text{dB}} &= 10\log_{10}\left(\frac{P_{\text{out}}}{P_{\text{in}}}\right) \\
&= 20\log_{10}\left(\frac{V_{\text{out}}}{V_{\text{in}}}\right) \qquad (1.33)
\end{aligned}
$$

So, now we have a multiplier factor of 20× rather than 10×.

It is useful to keep some common decibel figures in mind. The most commonly encountered one is a doubling of power, and 3 dB corresponds approximately to a double ratio

$$3\ \text{dB} \approx 10\log_{10}2$$

The exact figure is 3.0103, but 3 is close enough for most practical use. Similarly

$$\begin{aligned}2\ \text{dB} &\approx 10\log_{10}1.6\\ 4\ \text{dB} &\approx 10\log_{10}2.5\end{aligned}$$

From these values, it is possible to derive many other dB figures fairly easily. For example, 6 dB is

$$\begin{aligned}6\ \text{dB} &= 3\ \text{dB} + 3\ \text{dB}\\ \therefore 6\ \text{dB} &\rightarrow 2 \times 2\\ &= 4\times\end{aligned}$$

and so the ratio is 4. Since adding logarithms corresponds to multiplication, it follows that subtracting corresponds to division. So, for example,

$$\begin{aligned}1\ \text{dB} &= 4\ \text{dB} - 3\ \text{dB}\\ \therefore 1\ \text{dB} &\rightarrow \frac{2.5}{2}\\ &= 1.25\times\end{aligned}$$

Finally, note that the dB when used as a difference represents a ratio, and not a normalized power. So, for example, using two power values referenced to 1 mW,

$$4\ \text{dBm} - 3\ \text{dBm} = 1\ \text{dB}$$

We have two power figures (in dBm) but the difference is a ratio and is expressed in dB. Remember, because of the logarithmic function, a seemingly small number – such as a power loss of 20 dB – in fact represents a 99% power loss.

1.7.3 Maximum Power Transfer

When a signal is received by an antenna, that signal is likely to be exceedingly small. It follows that we do not want to waste any of that signal in the transmission from the antenna to the receiver. Similarly, if a transmitter is connected to an antenna, ideally the maximum amount of power would be transferred, implying no loss along the connecting wires. How can this be achieved?

To motivate the development, consider a simple circuit as shown in Figure 1.22. The question may be framed for this case as: What value of load resistance R_L will give the maximum amount of power transferred to that load? The assumption is that the source has a certain resistance R_S, and in practice

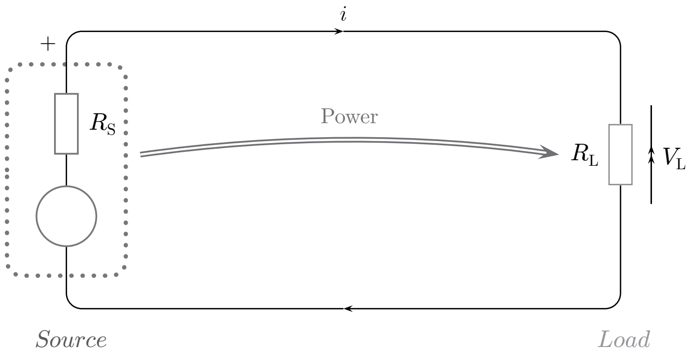

Figure 1.22 Transferring power from a source to a load. The source resistance R_S is generally quite small, and is inherent in the power source itself. We can adjust the load resistance R_L to maximize the power transferred.

this may be composed of the voltage source's own internal resistance or the equivalent resistance of the driving circuit.

For a simple circuit that has purely resistive impedances, we may write some basic equations governing the operation. The equivalent series resistance is

$$R_{eq} = R_S + R_L \tag{1.34}$$

Ohm's law applied to the circuit gives

$$\begin{aligned} V_S &= i\, R_{eq} \\ &= i\,(R_S + R_L) \\ \therefore i &= \frac{V_S}{R_S + R_L} \end{aligned}$$

and so the load power and current are

$$\begin{aligned} V_L &= i\, R_L \\ \therefore i &= \frac{V_L}{R_L} \end{aligned} \tag{1.35}$$

The power dissipated in the load, which is our main interest, is

$$\begin{aligned} P_L &= i\, V_L \\ &= i^2 R_L \\ &= \frac{V_S^2}{(R_S + R_L)^2}\, R_L \end{aligned} \tag{1.36}$$

A simulation of this scenario, using only the basic equations for voltage, current, and power, helps to confirm the theory. Using the MATLAB code below, the power as the load resistance varies is calculated, with the result shown in Figure 1.23.

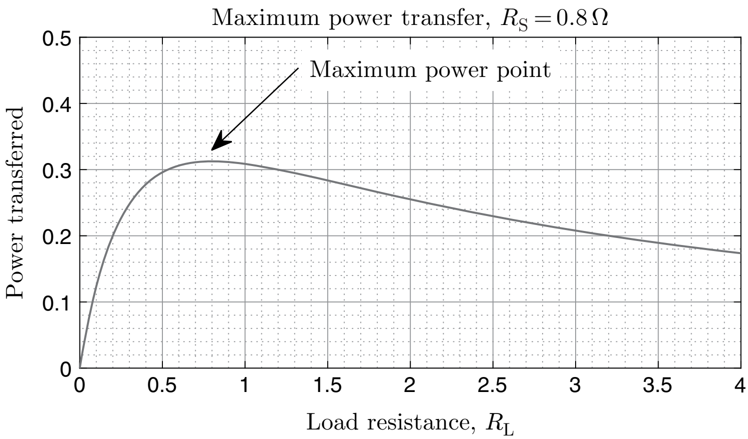

Figure 1.23 The power transferred to a load as the load resistance is varied. There is a point where the maximum amount of power is transferred, and this occurs when the load resistance exactly matches the source resistance.

```
% parameters of the simulation
Vsrc = 1;
Rsrc = 0.8;

% load resistance range
Rload = linspace(0, 4, 1000);

% equations
Req = Rsrc + Rload;
i = Vsrc./Req;
Pload = (i.*i).*Rload;

% plotting
plot(Rload, Pload);
xlabel('Load Resistance R_{load}')
ylabel('Power Transferred P_{load}')
```

From the figure, we can see that there is a point where the amount of power transferred is a maximum. Why does this occur? If the load resistance is very high, the current flowing through it will be low, and the voltage drop across it will be high. If the load resistance is low, the current flowing through it will be higher, but the voltage drop across it will be lower. Since the power dissipated in the load depends on both voltage and current, there is obviously an interplay between these factors.

How can we verify this analytically? We need to find the maximum P_L as a function of R_L. The governing equation was derived as

$$\begin{aligned} P_L &= \frac{V_S^2}{(R_S + R_L)^2} R_L \\ &= \frac{V_S^2}{R_S^2 + 2R_S R_L + R_L^2} R_L \\ &= \frac{V_S^2}{R_S^2/R_L + 2R_S + R_L} \end{aligned} \tag{1.37}$$

Reasoning that this power is a maximum when the denominator is a minimum, we define an auxiliary function and try to minimize that

$$\begin{aligned} f(R_L) &= R_S^2/R_L + 2R_S + R_L \\ \frac{df}{dR_L} &= -\frac{R_S^2}{R_L^2} + 0 + 1 \end{aligned} \tag{1.38}$$

Setting this to zero, we have that $R_L = R_S$, and so the conclusion is that **the maximum amount of power is dissipated in the load if the load resistance equals the source resistance**. Equivalently, the maximum power is dissipated (transferred) when the source resistance equals the load resistance.

In a communication system, we might have an antenna (load) fed by a source and transmission line. Thus, the line resistance (actually, impedance, which is resistance at certain frequency) must match the source and load resistance.

Note that *maximum power transfer* does not equal *maximum efficiency*. Defining efficiency η as the power delivered to the load over the total power dissipated,

$$\begin{aligned} \eta &= \frac{iR_L}{iR_S + iR_L} \\ &= \frac{R_L}{R_S + R_L} \\ &= \frac{1}{1 + \frac{R_S}{R_L}} \end{aligned} \tag{1.39}$$

If the source resistance were zero ($R_S = 0$), which is not really a practical scenario, the efficiency would be 100%. However, for some other resistance, if we arranged that $R_S = R_L$, then the efficiency would be 50%.

1.8 System Noise

Any real system is subject to the effects of extraneous noise. This may come from devices that deliberately radiate energy, such as radio or wireless transmissions, or nearby electronics such as computers and switch-mode power supplies, which radiate interference as an unintended but inevitable consequence of their operation. There is also noise present naturally – as a result of cosmic background radiation and from the thermal agitation of electrons in conductors. In this section, we briefly summarize some important concepts encountered when dealing with noise in a system.

A key result found in the early development of radio and electronics was that noise is present in any resistance that is at a temperature above absolute zero. Johnson (1928) is generally deemed to be the first to have experimentally assessed this phenomenon, which was further explained by Nyquist (1928). As a result, thermal noise is often termed Johnson Noise or Johnson–Nyquist Noise. The key result was that current was proportional to the square root of the temperature, and as a result the noise power $\mathcal{N}$ dissipated in a load is

$$\mathcal{N} = kTB \tag{1.40}$$

where T is the absolute temperature (in Kelvin), B is the bandwidth of the system being measured, and k is a constant due to Planck, but usually termed Boltzmann's constant, which has an approximate value of $k \approx 1.38 \times 10^{-23}$ J K^{-1}. Importantly, this result shows that noise power is dependent on temperature, but not on resistance. Furthermore, since the bandwidth employed in a particular application may not be known in advance, the noise power is often expressed as a power per unit bandwidth, or dBm/Hz. Following on from this, the noise voltage is then V/$\sqrt{\text{Hz}}$.

The amount of noise present in a system is not usually considered in isolation, but rather with respect to the size of the desired signal that carries information. Thus, the signal-to-noise ratio (SNR) is defined as the signal power divided by noise power and is usually expressed in decibels:

$$\frac{S}{N} = \frac{P_{\text{signal}}}{P_{\text{noise}}} \tag{1.41}$$

It is usually expressed as a dB figure:

$$\text{SNR}_{\text{dB}} = 10\log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \text{ dB} \tag{1.42}$$

Telecommunication systems are composed of numerous building blocks, such as amplifiers, filters, and modulators. An excessive amount of noise results in audible distortion for analog audio systems, and an increase in the bit error rate (BER) for digital systems. In extreme cases, digital systems may not function if the BER is over a maximum tolerable threshold. It is therefore

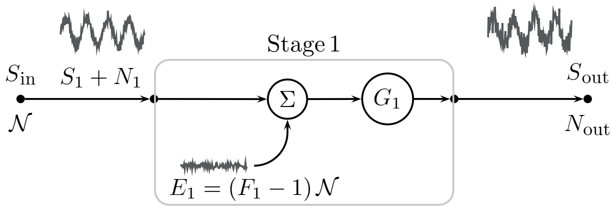

Figure 1.24 Modeling the noise transfer of a system. The noise at the input of the first block is $\mathcal{N} = N_1$, and this is used as a "noise reference" when subsequent blocks are added after the first. The quantity E is the excess noise added by the stage.

useful to know the effect of thermal noise on one particular block in isolation, and also the net result of cascading several blocks. This is done with the *noise factor* or *noise ratio*. The noise factor (or ratio) is defined as the SNR at the input terminals of a device, divided by the SNR at the output terminals. On the assumption that the block incorporates amplification, and the bandwidth of the block is not a limiting factor, then a noise ratio of unity would mean that no noise is added by the block, whereas a noise ratio of greater than one would imply that there is more noise at the output than the input (or, equivalently, the particular element reduces the SNR).

Often, the *noise figure* is expressed in dB, which is derived from the noise ratio as

$$F_{\mathrm{dB}} = 10\log_{10} F \tag{1.43}$$

Using the noise figure concept, an important step in analyzing block-level design is Friis's noise equation, first devised in Friis (1944) and covered in many textbooks in detail (for example, Haykin and Moher, 2009). To illustrate the basic idea, consider a block within a system as shown in Figure 1.24, which performs amplification of a signal by a factor of G_1. The input on the left may be an antenna, or some other receiver such as an optical sensor. Since any system block will add some noise to the overall system design, it is good to be able to quantify just how much noise is added.

Referring to Figure 1.24, we have an input signal S_{in} and thermal noise $\mathcal{N}$. These are assumed to be additive, with a resulting signal input $S_1 + N_1$ seen by the input of the block. It is assumed that the gain is greater than one and that the bandwidth is sufficient to pass the signal.

In the present context, we would like to know how much a given system degrades the SNR overall. To that end, we define a noise factor F, which pertains to how much noise is added when a signal passes through a system block. It is the SNR at the input, divided by the SNR at the output:

$$F = \frac{S_{\text{in}}/N_{\text{in}}}{S_{\text{out}}/N_{\text{out}}} \tag{1.44}$$

Referring to Figure 1.24, the signal output is simply the signal input multiplied by the gain of the block, so mathematically $S_{\text{out}} = G_1 S_1$. Assuming that noise is added to the signal, the input noise is also multiplied by the gain of the block. However, the block may also add its own noise, so we may write

$$N_2 = F_1 G_1 N_1 \tag{1.45}$$

where F_1 is a multiplicative factor greater than one. If $F_1 = 1$, it would imply a perfect block, which adds no additional noise. So, we could calculate the noise ratio as

$$\begin{aligned} R &= \frac{S_{\text{in}}/N_{\text{in}}}{S_{\text{out}}/N_{\text{out}}} \\ &= \left(\frac{S_1}{N_1}\right)\left(\frac{F_1 G_1 N_1}{G_1 S_1}\right) \\ &= F_1 \end{aligned} \tag{1.46}$$

So, the noise figure F is actually the noise ratio, defined as SNR at the input divided by SNR at the output.

It is useful in a practical sense to refer the output noise of cascaded blocks back to the noise appearing at the input. To follow the path of this noise, we write it as $\mathcal{N}$, where $\mathcal{N} = N_1$ is the noise at the input. Referring to Figure 1.24, we may rewrite noise at the output as

$$N_2 = G_1[\overbrace{(F_1 - 1)\mathcal{N}}^{\text{“excess noise”}} + \mathcal{N}] \tag{1.47}$$

This turns out to be useful in analyzing a cascade of two systems, as shown in Figure 1.25. The noise at the output of the second stage will be the input noise, multiplied by the gain factor, plus any additional noise from the system itself. This gives

$$\begin{aligned} N_3 &= F_1 G_1 G_2 \mathcal{N} + G_2(F_2 - 1)\mathcal{N} \\ &= G_2[F_1 G_1 \mathcal{N} + (F_2 - 1)\mathcal{N}] \end{aligned} \tag{1.48}$$

As a result, the overall noise figure (or noise ratio) is

$$\begin{aligned} F_{12} &= \left(\frac{S_1}{N_1}\right)\left(\frac{N_3}{S_3}\right) \\ &= \left(\frac{S_1}{\mathcal{N}}\right)\left\{\frac{G_2[F_1 G_1 \mathcal{N} + (F_2 - 1)\mathcal{N}]}{G_1 G_2 S_1}\right\} \\ &= F_1 + \frac{F_2 - 1}{G_1} \end{aligned} \tag{1.49}$$

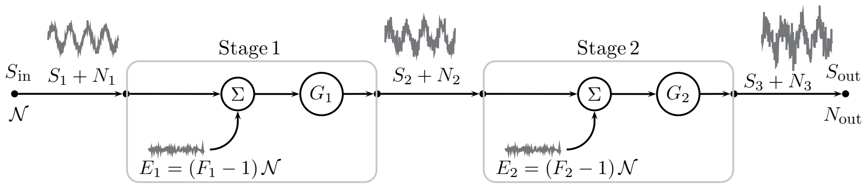

Figure 1.25 Analysis of two systems in cascade. The *E* values refer to the hypothetical noise added if referred back to the input of the first stage, whose noise is $\mathcal{N}$.

The significance of this is that **the first stage in a multistage system dominates the noise figure overall**. Subsequent stages contribute an amount lessened by the gain; in this case, the contribution of stage 2, which is $(F_2 - 1)$, is reduced by a factor equal to the gain of the previous stage G_1.

This could be extrapolated to any number of stages, for which the Friis equation for overall noise figure becomes

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots + \frac{F_n - 1}{G_1 G_2 \ldots G_n} \tag{1.50}$$

Thus, it makes sense to maximize efforts to reduce the noise in the very first stage. Additionally, a high gain is helpful in the first stage, to reduce the effects of subsequent stages.

1.9 Chapter Summary

The following are the key elements covered in this chapter:

- The description of a waveform as a time-evolving quantity.
- The description of signal as comprising various frequency components, and how these components may be affected by filtering.
- Operations such as averaging, multiplication, and phase shifting, which may be applied to a waveform.
- One method of variable-frequency waveform generation: the DDS.
- The significance of power transfer, impedance matching, and noise in telecommunication system design.
- Thermal noise, and how noise may be characterized in a cascade of system blocks.

Problems

1.1 The decibel requires the calculation $10\log_{10}(P_{\text{out}}/P_{\text{in}})$. Using $P = V^2/R$ and assuming V_{out} is the voltage at the output, V_{in} the voltage at the

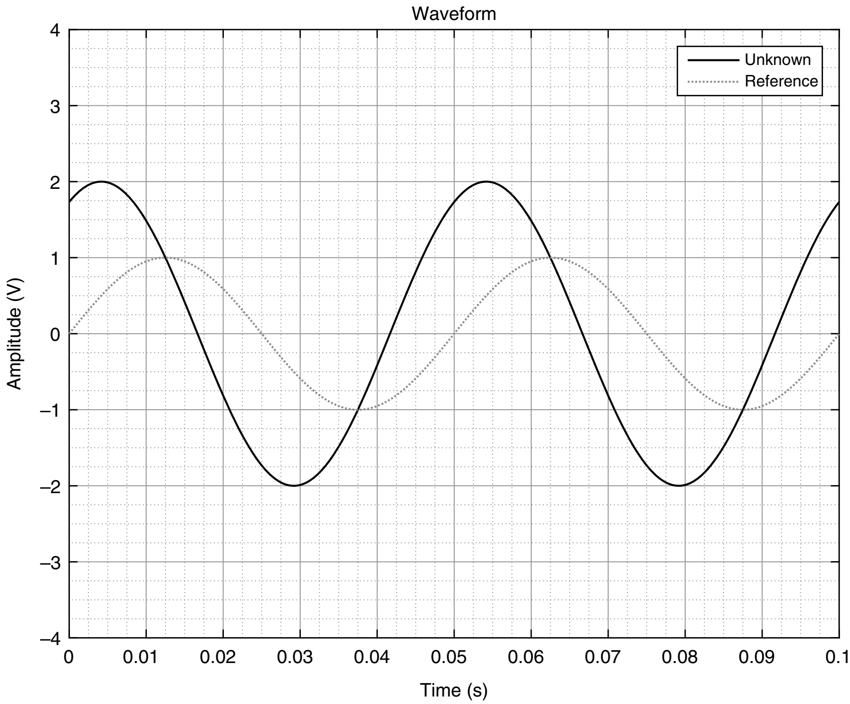

Figure 1.26 Waveform parameter problem.

input, and that the impedances of both are $R\ \Omega$, show that an equivalent calculation is $20\log_{10}(V_{\text{out}}/V_{\text{in}})$.

1.2 The input to a Radio Frequency (RF) spectrum analyzer states that the input impedance is 50Ω, and that the maximum input power is +10dBm. What would be the maximum safe voltage in that case?

1.3 A copper communications line has a noise level of 1 mV RMS when a signal of 1 V RMS is observed. What is the SNR?

1.4 Determine the parameters (amplitude, phase, and frequency) of the waveform shown in Figure 1.26.

1.5 Given the mathematical description of a signal $x(t) = A\sin\omega t$, show that over one period $\tau = 2\pi/\omega$ the mean-square value is $\overline{x^2}(t) = A^2/2$. Hence show that the RMS value is $A/\sqrt{2}$. *Hint: Remember that the* *arithmetic mean is really an average, so you integrate the square value over one period. You may need the trigonometric identity* $\sin^2\theta = (1/2)(1 - \cos 2\theta)$.

1.6 Given a signal equation and the system transfer function, we can work out the output for both linear and nonlinear systems.

a) Given the system transfer function $y(t) = \alpha\, x(t)$, show that for an input $x(t) = A \sin \omega t$, the output is $y(t) = \alpha\, A \sin \omega t$. Is this system linear?

b) Given the system described by $y(t) = \alpha\, x(t) + \beta x^2(t)$, show that for an input $x(t) = A \sin \omega t$, the output can be simplified to the summation of a constant (or DC) term, a term at the same frequency as the input, and a term at twice the frequency of the input. Is this system linear? *Hint: You may need the trigonometric identity* $\sin^2\theta = (1/2)(1 - \cos 2\theta)$.

c) From the above results, can you infer what might happen if you had cubic-form transfer function, such as $y(t) = \gamma x^3(t)$?

1.7 Systems may be defined in terms of basic building blocks.

a) Given two series blocks as depicted on the left of Figure 1.3, what is the overall gain if each block's gain is given in decibels?

b) Would the same rule apply if the blocks were added in parallel? Why not?

1.8 The correspondence between dB and ratio is approximately

$$2 \text{ dB} \approx 1.6\times$$
$$3 \text{ dB} \approx 2\times$$
$$4 \text{ dB} \approx 2.5\times$$

a) Explain why the dB figure goes up in equal increments of one, but the ratio figure goes up in differing increments (0.4 then 0.5).

b) Plot a graph of ratio r versus $10\log_{10} r$ for $r = 0.1$ to $r = 10$ in steps of 0.1, and explain the shape.

c) Plot a graph of ratio r versus $10\log_{10} r$ for $r = 10$ to $r = 100$ in steps of 1, and explain the shape. Compare the two graphs and explain their shapes as well as the values on the vertical axis.

1.9 Many concepts in telecommunications deal with very large or very small signals or cover a very wide range of values. In these cases, a *logarithmic* scale is useful rather than the usual linear scale. A good example is the decibel for measuring power. Suppose the frequency response of a certain system is defined by a function $g(f) = 1/(f + 1)$.

a) Explain what is deficient in the following approach, and suggest a better way.

```
f = 0.01:1:100;
g = 1./(f + 1);
plot(f, g, 's-');
set(gca, 'xscale', 'log');
grid('on');
```

b) Noting that the exponents of 10 on the frequency axis go from −2 to +2, change the code to

```
r = -2:0.04:2;
f = 10.^r;
g = 1./(f + 1);
plot(f, g, 's-');
set(gca, 'xscale', 'log');
grid('on');
```

Why does this give a proportional spacing of the data points, and hence a better plot?

c) Investigate the difference between the MATLAB functions `linspace()` and `logspace()` and briefly comment on why they are useful.

1.10 An amplifier has an SNR of 50 dB and Noise Figure of 3 dB. Determine the output SNR.

1.11 This question investigates the extension to two-stage systems as shown in Figure 1.25, in order to find an expression for the cascaded noise figure.

a) Draw a block diagram for this system, labeling all the "useful" signals and the unwanted noise signals.

b) Show mathematically that the noise factor for a three-stage system is

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2}$$