CHAPTER 2

Signals and Systems

This text assumes a basic background in the representation of linear, time-invariant systems and the associated continuous-time and discrete-time signals, through convolution, Fourier analysis, Laplace transforms and $\mathcal{Z}$-transforms. In this chapter we briefly summarize and review this assumed background, in part to establish notation that we will be using throughout the text, and also as a convenient reference for the topics in the later chapters. We follow closely the notation, style and presentation in *Signals and Systems*, Oppenheim and Willsky with Nawab, 2nd Edition, Prentice Hall, 1997.

2.1 SIGNALS, SYSTEMS, MODELS, PROPERTIES

Throughout this text we will be considering various classes of signals and systems, developing models for them and studying their properties.

Signals for us will generally be real or complex functions of some independent variables (almost always time and/or a variable denoting the outcome of a probabilistic experiment, for the situations we shall be studying). Signals can be:

- 1-dimensional or multi-dimensional
- continuous-time (CT) or discrete-time (DT)
- deterministic or stochastic (random, probabilistic)

Thus, a DT deterministic time-signal may be denoted by a function $x[n]$ of the integer time (or clock or counting) variable n.

Systems are collections of software or hardware elements, components, subsystems. A system can be viewed as mapping a set of input signals to a set of output or response signals. A more general view is that a system is an entity imposing constraints on a designated set of signals, where the signals are not necessarily labeled as inputs or outputs. Any specific set of signals that satisfies the constraints is termed a behavior of the system.

Models are (usually approximate) mathematical or software or hardware or linguistic or other representations of the constraints imposed on a designated set of signals by a system. A model is itself a system, because it imposes constraints on the set of signals represented in the model, so we often use the words "system" and "model" interchangeably, although it can sometimes be important to preserve the distinction between something truly physical and our representations of it mathematically or in a computer simulation. We can thus talk of the behavior of a model.

A mapping model of a system comprises the following: a set of input signals $\{x_i(t)\}$, each of which can vary within some specified range of possibilities; similarly, a set of output signals $\{y_j(t)\}$, each of which can vary; and a description of the mapping that uniquely defines the output signals as a function of the input signals. As an example, consider the following single-input, single-output system:

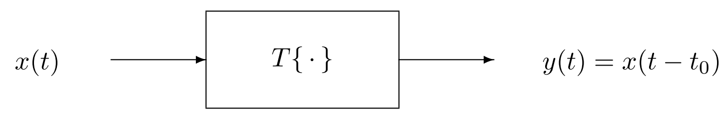

FIGURE 2.1 Name-Mapping Model

Given the input $x(t)$ and the mapping $T\{\cdot\}$, the output $y(t)$ is unique, and in this example equals the input delayed by t_0.

A behavioral model for a set of signals $\{w_i(t)\}$ comprises a listing of the constraints that the $w_i(t)$ must satisfy. The constraints on the voltages across and currents through the components in an electrical circuit, for example, are specified by Kirchhoff's laws, and the defining equations of the components. There can be infinitely many combinations of voltages and currents that will satisfy these constraints.

2.1.1 System/Model Properties

For a system or model specified as a mapping, we have the following definitions of various properties, all of which we assume are familiar. They are stated here for the DT case but easily modified for the CT case. (We also assume a single input signal and a single output signal in our mathematical representation of the definitions below, for notational convenience.)

- **Memoryless** or **Algebraic** or **Non-Dynamic**: The outputs at any instant do not depend on values of the inputs at any other instant: $y[n_0] = T\{x[n_0]\}$ for all n_0.
- **Linear**: The response to an arbitrary linear combination (or "superposition") of inputs signals is always the same linear combination of the individual responses to these signals: $T\{ax_A[n] + bx_B[n]\} = aT\{x_A[n]\} + bT\{x_B[n]\}$, for all x_A, x_B, a and b.

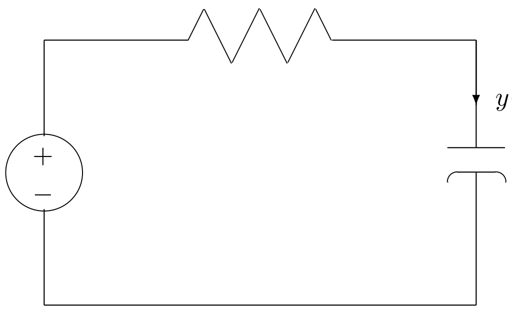

FIGURE 2.2 RLC Circuit

- **Time-Invariant**: The response to an arbitrarily translated set of inputs is always the response to the original set, but translated by the same amount: If $x[n] \to y[n]$ then $x[n - n_0] \to y[n - n_0]$ for all x and n_0.
- **Linear and Time-Invariant (LTI)**: The system, model or mapping is both linear and time-invariant.
- **Causal**: The output at any instant does not depend on future inputs: for all n_0, $y[n_0]$ does not depend on $x[n]$ for $n > n_0$. Said another way, if $\widehat{x}[n], \widehat{y}[n]$ denotes another input-output pair of the system, with $\widehat{x}[n] = x[n]$ for $n \le n_0$, then it must be also true that $\widehat{y}[n] = y[n]$ for $n \le n_0$. (Here n_0 is arbitrary but fixed.)
- **BIBO Stable**: The response to a bounded input is always bounded: $|x[n]| \le M_x < \infty$ for all n implies that $|y[n]| \le M_y < \infty$ for all n.

EXAMPLE 2.1 System Properties

Consider the system with input $x[n]$ and output $y[n]$ defined by the relationship

$$y[n] = x[4n + 1] \tag{2.1}$$

We would like to determine whether or not the system has each of the following properties: memoryless, linear, time-invariant, causal, and BIBO stable.

memoryless: a simple counter example suffices. For example, $y[0] = x[1]$, i.e. the output at $n = 0$ depends on input values at times other than at $n = 0$. Therefore it is not memoryless.

linear: To check for linearity, we consider two different inputs, $x_A[n]$ and $x_B[n]$, and compare the output of their linear combination to the linear combination of their outputs.

$$\begin{aligned} x_A[n] &\rightarrow x_A[4n+1] = y_A[n] \\ x_B[n] &\rightarrow x_B[4n+1] = y_B[n] \\ x_C[n] = (ax_A[n] + bx_B[n]) &\rightarrow (ax_A[4n+1] + bx_B[4n+1]) = y_C[n] \end{aligned}$$

If $y_C[n] = ay_A[n] + by_B[n]$, then the system is linear. This clearly happens in this case.

time-invariant: To check for time-invariance, we need to compare the output due to a time-shifted version of $x[n]$ to the time-shifted version of the output due to $x[n]$.

$$\begin{aligned} x[n] &\rightarrow x[4n+1] = y[n] \\ x_B[n] = x[n+n_0] &\rightarrow x[4n+n_0+1] = y_B[n] \end{aligned}$$

We now need to compare $y[n]$ time-shifted by n_0 (i.e. $y[n+n_0]$) to $y_B[n]$. If they're not equal, then the system is not time-invariant.

$$\begin{aligned} y[n+n_0] &= x[4n+4n_0+1] \\ \text{but} \qquad y_B[n] &= x[4n+n_0+1] \end{aligned}$$

Consequently, the system is not time-invariant. To illustrate with a specific counter-example, suppose that $x[n]$ is an impulse, $\delta[n]$, at $n = 0$. In this case, the output, $y_\delta[n]$, would be $\delta[4n+1]$, which is zero for all values of n, and $y[n+n_0]$ would likewise always be zero. However, if we consider $x[n+n_0] = \delta[n+n_0]$, the output will be $\delta[4n+1+n_0]$, which for $n_0 = 3$ will be one at $n = -4$ and zero otherwise.

causal: Since the output at $n = 0$ is the input value at $n = 1$, the system is not causal.

BIBO stable: Since $|y[n]| = |x[4n+1]|$ and the maximum value for all n of $x[n]$ and $x[4n+1]$ is the same, the system is BIBO stable.

2.2 LINEAR, TIME-INVARIANT SYSTEMS

2.2.1 Impulse-Response Representation of LTI Systems

Linear, time-invariant (LTI) systems form the basis for engineering design in many situations. They have the advantage that there is a rich and well-established theory for analysis and design of this class of systems. Furthermore, in many systems that are nonlinear, small deviations from some nominal steady operation are approximately governed by LTI models, so the tools of LTI system analysis and design can be applied incrementally around a nominal operating condition.

A very general way of representing an LTI mapping from an input signal x to an output signal y is through convolution of the input with the system impulse response. In CT the relationship is

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \tag{2.2}$$

where $h(t)$ is the unit impulse response of the system. In DT, we have

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]\, h[n-k] \tag{2.3}$$

where $h[n]$ is the unit sample (or unit "impulse") response of the system.

A common notation for the convolution integral in (2.2) or the convolution sum in (2.3) is as

$$y(t) = x(t) * h(t) \tag{2.4}$$

$$y[n] = x[n] * h[n] \tag{2.5}$$

While this notation can be convenient, it can also easily lead to misinterpretation if not well understood.

The characterization of LTI systems through the convolution is obtained by representing the input signal as a superposition of weighted impulses. In the DT case, suppose we are given an LTI mapping whose impulse response is $h[n]$, i.e., when its input is the unit sample or unit "impulse" function $\delta[n]$, its output is $h[n]$. Now a general input $x[n]$ can be assembled as a sum of scaled and shifted impulses, as follows:

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\, \delta[n-k] \tag{2.6}$$

The response $y[n]$ to this input, by linearity and time-invariance, is the sum of the similarly scaled and shifted impulse responses, and is therefore given by (2.3). What linearity and time-invariance have allowed us to do is write the response to a general input in terms of the response to a special input. A similar derivation holds for the CT case.

It may seem that the preceding derivation shows all LTI mappings from an input signal to an output signal can be represented via a convolution relationship. However, the use of infinite integrals or sums like those in (2.2), (2.3) and (2.6) actually involves some assumptions about the corresponding mapping. We make no attempt here to elaborate on these assumptions. Nevertheless, it is not hard to find "pathological" examples of LTI mappings — not significant for us in this course, or indeed in most engineering models — where the convolution relationship does not hold because these assumptions are violated.

It follows from (2.2) and (2.3) that a necessary and sufficient condition for an LTI system to be BIBO stable is that the impulse response be absolutely integrable (CT) or absolutely summable (DT), i.e.,

$$\text{BIBO stable (CT)} \iff \int_{-\infty}^{\infty} |h(t)|dt < \infty$$

$$\text{BIBO stable (DT)} \iff \sum_{n=-\infty}^{\infty} |h[n]| < \infty$$

It also follows from (2.2) and (2.3) that a necessary and sufficient condition for an LTI system to be causal is that the impulse response be zero for $t < 0$ (CT) or for $n < 0$ (DT)

2.2.2 Eigenfunction and Transform Representation of LTI Systems

Exponentials are eigenfunctions of LTI mappings, i.e., when the input is an exponential for all time, which we refer to as an "everlasting" exponential, the output is simply a scaled version of the input, so computing the response to an exponential reduces to just multiplying by the appropriate scale factor. Specifically, in the CT case, suppose

$$x(t) = e^{s_0 t} \tag{2.7}$$

for some possibly complex value s_0 (termed the complex frequency). Then from (2.2)

$$\begin{aligned} y(t) &= h(t) * x(t) \\ &= \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau \\ &= \int_{-\infty}^{\infty} h(\tau)e^{s_0(t-\tau)}d\tau \\ &= H(s_0)e^{s_0 t} \end{aligned} \tag{2.8}$$

where

$$H(s) = \int_{-\infty}^{\infty} h(\tau)e^{-s\tau}d\tau \tag{2.9}$$

provided the above integral has a finite value for $s = s_0$ (otherwise the response to the exponential is not well defined). Note that this integral is precisely the bilateral Laplace transform of the impulse response, or the transfer function of the system, and the (interior of the) set of values of s for which the above integral takes a finite value constitutes the region of convergence (ROC) of the transform.

From the preceding discussion, one can recognize what special property of the everlasting exponential causes it to be an eigenfunction of an LTI system: it is the fact that time-shifting an everlasting exponential produces the same result as scaling it by a constant factor. In contrast, the one-sided exponential $e^{s_0 t}u(t)$ — where $u(t)$ denotes the unit step — is in general not an eigenfunction of an LTI mapping: time-shifting a one-sided exponential does not produce the same result as scaling this exponential.

When $x(t) = e^{j\omega t}$, corresponding to having s_0 take the purely imaginary value $j\omega$ in (2.7), the input is bounded for all positive and negative time, and the corresponding output is

$$y(t) = H(j\omega)e^{j\omega t} \tag{2.10}$$

where

$$H(j\omega) = \int_{-\infty}^{\infty} h(t)e^{-j\omega t}\, dt \tag{2.11}$$

EXAMPLE 2.2 Eigenfunctions of LTI Systems

While as demonstrated above, the everlasting complex exponential, $e^{j\omega t}$, is an eigenfunction of any stable LTI system, it is important to recognize that $e^{j\omega t}u(t)$ is not. Consider, as a simple example, a time delay, i.e.

$$y(t) = x(t - t_0) \tag{2.12}$$

The output due to the input $e^{j\omega t}u(t)$ is

$$e^{-j\omega t_0}e^{+j\omega t}u(t - t_0)$$

This is not a simple scaling of the input, so $e^{j\omega t}u(t)$ is not in general an eigenfunction of LTI systems.

The function $H(j\omega)$ in (2.10) is the system frequency response, and is also the continuous-time Fourier transform (CTFT) of the impulse response. The integral that defines the CTFT has a finite value (and can be shown to be a continuous function of ω) if $h(t)$ is absolutely integrable, i.e. provided

$$\int_{-\infty}^{+\infty} |h(t)|\, dt < \infty$$

We have noted that this condition is equivalent to the system being bounded-input, bounded-output (BIBO) stable. The CTFT can also be defined for signals that are not absolutely integrable, e.g., for $h(t) = (\sin t)/t$ whose CTFT is a rectangle in the frequency domain, but we defer examination of conditions for existence of the CTFT.

We can similarly examine the eigenfunction property in the DT case. A DT everlasting "exponential" is a geometric sequence or signal of the form

$$x[n] = z_0^n \tag{2.13}$$

for some possibly complex z_0 (termed the complex frequency). With this DT exponential input, the output of a convolution mapping is (by a simple computation that is analogous to what we showed above for the CT case)

$$y[n] = h[n] * x[n] = H(z_0)z_0^n \tag{2.14}$$

where

$$H(z) = \sum_{k=-\infty}^{\infty} h[k]z^{-k} \tag{2.15}$$

provided the above sum has a finite value when $z = z_0$. Note that this sum is precisely the bilateral $\mathcal{Z}$-transform of the impulse response, and the (interior of the) set of values of z for which the sum takes a finite value constitutes the ROC of the $\mathcal{Z}$-transform. As in the CT case, the one-sided exponential $z_0^n u[n]$ is not in general an eigenfunction.

Again, an important case is when $x[n] = (e^{j\Omega})^n = e^{j\Omega n}$, corresponding to z_0 in (2.13) having unit magnitude and taking the value $e^{j\Omega}$, where Ω — the (real) "frequency" — denotes the angular position (in radians) around the unit circle in the z-plane. Such an $x[n]$ is bounded for all positive and negative time. Although we use a different symbol, Ω, for frequency in the DT case, to distinguish it from the frequency ω in the CT case, it is not unusual in the literature to find ω used in both CT and DT cases for notational convenience. The corresponding output is

$$y[n] = H(e^{j\Omega})e^{j\Omega n} \tag{2.16}$$

where

$$H(e^{j\Omega}) = \sum_{n=-\infty}^{\infty} h[n]e^{-j\Omega n} \tag{2.17}$$

The function $H(e^{j\Omega})$ in (2.17) is the frequency response of the DT system, and is also the discrete-time Fourier transform (DTFT) of the impulse response. The sum that defines the DTFT has a finite value (and can be shown to be a continuous function of Ω) if $h[n]$ is absolutely summable, i.e., provided

$$\sum_{n=-\infty}^{\infty} |\, h[n] \,| < \infty \tag{2.18}$$

We noted that this condition is equivalent to the system being BIBO stable. As with the CTFT, the DTFT can be defined for signals that are not absolutely summable; we will elaborate on this later.

Note from (2.17) that the frequency response for DT systems is always periodic, with period 2π. The "high-frequency" response is found in the vicinity of $\Omega = \pm\pi$, which is consistent with the fact that the input signal $e^{\pm j\pi n} = (-1)^n$ is the most rapidly varying DT signal that one can have.

When the input of an LTI system can be expressed as a linear combination of bounded eigenfunctions, for instance (in the CT case),

$$x(t) = \sum_{\ell} a_\ell e^{j\omega_\ell t} \tag{2.19}$$

then, by linearity, the output is the same linear combination of the responses to the individual exponentials. By the eigenfunction property of exponentials in LTI systems, the response to each exponential involves only scaling by the system's frequency response. Thus

$$y(t) = \sum_{\ell} a_\ell H(j\omega_\ell)e^{j\omega_\ell t} \tag{2.20}$$

Similar expressions can be written for the DT case.

2.2.3 Fourier Transforms

A broad class of input signals can be represented as linear combinations of bounded exponentials, through the Fourier transform. The synthesis/analysis formulas for the CTFT are

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\omega)\, e^{j\omega t} d\omega \quad \text{(synthesis)} \tag{2.21}$$

$$X(j\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t} dt \quad \text{(analysis)} \tag{2.22}$$

Note that (2.21) expresses $x(t)$ as a linear combination of exponentials — but this weighted combination involves a continuum of exponentials, rather than a finite or countable number. If this signal $x(t)$ is the input to an LTI system with frequency response $H(j\omega)$, then by linearity and the eigenfunction property of exponentials the output is the same weighted combination of the responses to these exponentials, so

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(j\omega)X(j\omega)\, e^{j\omega t} d\omega \tag{2.23}$$

By viewing this equation as a CTFT synthesis equation, it follows that the CTFT of $y(t)$ is

$$Y(j\omega) = H(j\omega)X(j\omega) \tag{2.24}$$

Correspondingly, the convolution relationship (2.2) in the time domain becomes multiplication in the transform domain. Thus, to find the response Y at a particular frequency point, we only need to know the input X at that single frequency, and the frequency response of the system at that frequency. This simple fact serves, in large measure, to explain why the frequency domain is virtually indispensable in the analysis of LTI systems.

The corresponding DTFT synthesis/analysis pair is defined by

$$x[n] = \frac{1}{2\pi}\int_{<2\pi>} X(e^{j\Omega})\, e^{j\Omega n} d\Omega \quad \text{(synthesis)} \tag{2.25}$$

$$X(e^{j\Omega}) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-j\Omega n} \quad \text{(analysis)} \tag{2.26}$$

where the notation $< 2\pi >$ on the integral in the synthesis formula denotes integration over any contiguous interval of length 2π, since the DTFT is always periodic in Ω with period 2π, a simple consequence of the fact that $e^{j\Omega}$ is periodic with period 2π. Note that (2.25) expresses $x[n]$ as a weighted combination of a continuum of exponentials.

As in the CT case, it is straightforward to show that if $x[n]$ is the input to an LTI mapping, then the output $y[n]$ has DTFT

$$Y(e^{j\Omega}) = H(e^{j\Omega})X(e^{j\Omega}) \tag{2.27}$$

2.3 DETERMINISTIC SIGNALS AND THEIR FOURIER TRANSFORMS

In this section we review the DTFT of deterministic DT signals in more detail, and highlight the classes of signals that can be guaranteed to have well-defined DTFTs. We shall also devote some attention to the energy density spectrum of signals that have DTFTs. The section will bring out aspects of the DTFT that may not have been emphasized in your earlier signals and systems course. A similar development can be carried out for CTFTs.

2.3.1 Signal Classes and their Fourier Transforms

The DTFT synthesis and analysis pair in (2.25) and (2.26) hold for at least the three large classes of DT signals described below.

Finite-Action Signals. Finite-action signals, which are also called absolutely summable signals or ℓ_1 ("ell-one") signals, are defined by the condition

$$\sum_{k=-\infty}^{\infty} \Big| x[k] \Big| < \infty \tag{2.28}$$

The sum on the left is called the 'action' of the signal. For these ℓ_1 signals, the infinite sum that defines the DTFT is well behaved and the DTFT can be shown to be a continuous function for all Ω (so, in particular, the values at $\Omega = +\pi$ and $\Omega = -\pi$ are well-defined and equal to each other — which is often not the case when signals are not ℓ_1).

Finite-Energy Signals. Finite-energy signals, which are also called square summable or ℓ_2 ("ell-two") signals, are defined by the condition

$$\sum_{k=-\infty}^{\infty} \Big| x[k] \Big|^2 < \infty \tag{2.29}$$

The sum on the left is called the 'energy' of the signal.

In discrete-time, an absolutely summable (i.e., ℓ_1) signal is always square summable (i.e., ℓ_2). (In continuous-time, the story is more complicated: an absolutely integrable signal need not be square integrable, e.g., consider $x(t) = 1/\sqrt{t}$ for $0 < t \leq 1$ and $x(t) = 0$ elsewhere; the source of the problem here is that the signal is not bounded.) However, the reverse is not true. For example, consider the signal $(\sin \Omega_c n)/\pi n$ for $0 < \Omega_c < \pi$, with the value at $n = 0$ taken to be Ω_c/π, or consider the signal $(1/n)u[n-1]$, both of which are ℓ_2 but not ℓ_1. If $x[n]$ is such a signal, its DTFT $X(e^{j\Omega})$ can be thought of as the limit for $N \to \infty$ of the quantity

$$X_N(e^{j\Omega}) = \sum_{k=-N}^{N} x[k] e^{-j\Omega k} \tag{2.30}$$

and the resulting limit will typically have discontinuities at some values of Ω. For instance, the transform of $(\sin \Omega_c n)/\pi n$ has discontinuities at $\Omega = \pm\Omega_c$.

Signals of Slow Growth. Signals of 'slow' growth are signals whose magnitude grows no faster than polynomially with the time index, e.g., $x[n] = n$ for all n. In this case $X_N(e^{j\Omega})$ in (2.30) does not converge in the usual sense, but the DTFT still exists as a generalized (or singularity) function; e.g., if $x[n] = 1$ for all n, then $X(e^{j\Omega}) = 2\pi\delta(\Omega)$ for $|\Omega| \leq \pi$.

Within the class of signals of slow growth, those of most interest to us are bounded (or ℓ_∞) signals:

$$\Big|x[k]\Big| \leq M < \infty \tag{2.31}$$

i.e., signals whose amplitude has a fixed and finite bound for all time. Bounded everlasting exponentials of the form $e^{j\Omega_0 n}$, for instance, play a key role in Fourier transform theory. Such signals need not have finite energy, but will have finite average power over any time interval, where average power is defined as total energy over total time.

Similar classes of signals are defined in continuous-time. Specifically, finite-action (or L_1) signals comprise those that are absolutely integrable, i.e.,

$$\int_{-\infty}^{\infty} \Big|x(t)\Big| dt < \infty \tag{2.32}$$

Finite-energy (or L_2) signals comprise those that are square summable, i.e.,

$$\int_{-\infty}^{\infty} \Big|x(t)\Big|^2 < \infty \tag{2.33}$$

And signals of slow growth are ones for which the magnitude grows no faster than polynomially with time. Bounded (or L_∞) continuous-time signals are those for which the magnitude never exceeds a finite bound M (so these are slow-growth signals as well). These may again not have finite energy, but will have finite average power over any time interval.

In both continuous-time and discrete-time there are many important Fourier transform pairs and Fourier transform properties developed and tabulated in basic texts on signals and systems (see, for example, Chapters 4 and 5 of Oppenheim and Willsky). For convenience, we include here a brief table of DTFT pairs. Other pairs are easily derived from these by applying various DTFT properties. (Note that the δ's in the left column denote unit samples, while those in the right column are unit impulses!)

$$\textbf{DT Signal} \longleftrightarrow \textbf{DTFT for } -\pi < \Omega \leq \pi$$

$$\delta[n] \longleftrightarrow 1$$

$$\delta[n-n_0] \longleftrightarrow e^{-j\Omega n_0}$$

$$1 \text{ (for all } n) \longleftrightarrow 2\pi\delta(\Omega)$$

$$e^{j\Omega_0 n} \quad (-\pi < \Omega_0 \leq \pi) \longleftrightarrow 2\pi\delta(\Omega-\Omega_0)$$

$$a^n u[n]\ ,\ |a|<1 \longleftrightarrow \frac{1}{1-ae^{-j\Omega}}$$

$$u[n] \longleftrightarrow \frac{1}{1-e^{-j\Omega}} + \pi\delta(\Omega)$$

$$\frac{\sin\Omega_c n}{\pi n} \longleftrightarrow \begin{cases} 1, & -\Omega_c < \Omega < \Omega_c \\ 0, & \text{otherwise} \end{cases}$$

$$\left.\begin{matrix} 1, & -M \leq n \leq M \\ 0, & \text{otherwise} \end{matrix}\right\} \longleftrightarrow \frac{\sin[\Omega(2M+1)/2]}{\sin(\Omega/2)}$$

In general it is important and useful to be fluent in deriving and utilizing the main transform pairs and properties. In the following subsection we discuss a particular property, Parseval's identity, which is of particular significance in our later discussion.

There are, of course, other classes of signals that are of interest to us in applications, for instance growing one-sided exponentials. To deal with such signals, we utilize $\mathcal{Z}$-transforms in discrete-time and Laplace transforms in continuous-time.

2.3.2 Parseval's Identity, Energy Spectral Density, Deterministic Autocorrelation

An important property of the Fourier transform is Parseval's identity for ℓ_2 signals. For discrete time, this identity takes the general form

$$\sum_{n=-\infty}^{\infty} x[n]y^*[n] \;=\; \frac{1}{2\pi}\int_{<2\pi>} X(e^{j\Omega})Y^*(e^{j\Omega})\,d\Omega \tag{2.34}$$

and for continuous time,

$$\int_{-\infty}^{\infty} x(t)y^*(t)dt \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\omega)Y^*(j\omega)\,d\omega \tag{2.35}$$

where the * denotes the complex conjugate. Specializing to the case where $y[n] = x[n]$ or $y(t) = x(t)$, we obtain

$$\sum_{n=-\infty}^{\infty} |x[n]|^2 \;=\; \frac{1}{2\pi}\int_{<2\pi>} |X(e^{j\Omega})|^2\,d\Omega \tag{2.36}$$

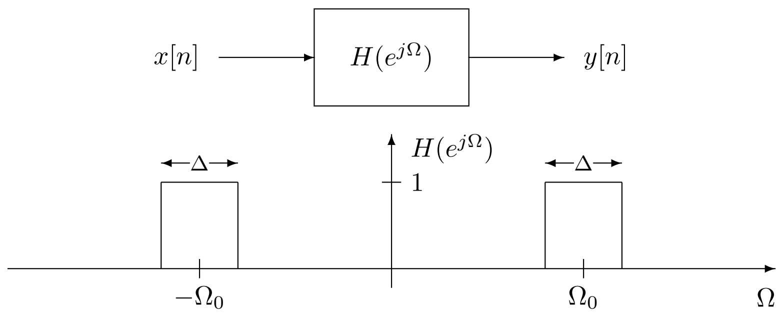

FIGURE 2.3 Ideal bandpass filter.

$$\int_{-\infty}^{\infty} |x(t)|^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} |X(j\omega)|^2 \, d\omega \tag{2.37}$$

Parseval's identity allows us to evaluate the energy of a signal by integrating the squared magnitude of its transform. What the identity tells us, in effect, is that the energy of a signal equals the energy of its transform (scaled by $1/2\pi$).

The real, even, nonnegative function of Ω defined by

$$\overline{S}_{xx}(e^{j\Omega}) = |X(e^{j\Omega})|^2 \tag{2.38}$$

or

$$\overline{S}_{xx}(j\omega) = |X(j\omega)|^2 \tag{2.39}$$

is referred to as the energy spectral density (ESD), because it describes how the energy of the signal is distributed over frequency. To appreciate this claim more concretely, for discrete-time, consider applying $x[n]$ to the input of an ideal bandpass filter of frequency response $H(e^{j\Omega})$ that has narrow passbands of unit gain and width Δ centered at $\pm\Omega_0$ as indicated in Figure 2.3. The energy of the output signal must then be the energy of $x[n]$ that is contained in the passbands of the filter. To calculate the energy of the output signal, note that this output $y[n]$ has the transform

$$Y(e^{j\Omega}) = H(e^{j\Omega})X(e^{j\Omega}) \tag{2.40}$$

Consequently the output energy, by Parseval's identity, is given by

$$\begin{aligned}\sum_{n=-\infty}^{\infty} |y[n]|^2 &= \frac{1}{2\pi}\int_{<2\pi>} |Y(e^{j\Omega})|^2 \, d\Omega \\ &= \frac{1}{2\pi}\int_{\text{passband}} \overline{S}_{xx}(e^{j\Omega}) \, d\Omega \end{aligned}\tag{2.41}$$

Thus the energy of $x[n]$ in any frequency band is given by integrating $\overline{S}_{xx}(e^{j\Omega})$ over that band (and scaling by $1/2\pi$). In other words, the energy density of $x[n]$ as a function of Ω is $\overline{S}_{xx}(\Omega)/(2\pi)$ per radian. An exactly analogous discussion can be carried out for continuous-time signals.

Since the ESD $\overline{S}_{xx}(e^{j\Omega})$ is a real function of Ω, an alternate notation for it could perhaps be $\mathcal{E}_{xx}(\Omega)$, for instance. However, we use the notation $\overline{S}_{xx}(e^{j\Omega})$ in order to make explicit that it is the squared magnitude of $X(e^{j\Omega})$ and also the fact that the ESD for a DT signal is periodic with period 2π.

Given the role of the magnitude squared of the Fourier transform in Parseval's identity, it is interesting to consider what signal it is the Fourier transform of. The answer for DT follows on recognizing that with $x[n]$ real-valued

$$|X(e^{j\Omega})|^2 = X(e^{j\Omega})X(e^{-j\Omega}) \tag{2.42}$$

and that $X(e^{-j\Omega})$ is the transform of the time-reversed signal, $x[-k]$. Thus, since multiplication of transforms in the frequency domain corresponds to convolution of signals in the time domain, we have

$$\overline{S}_{xx}(e^{j\Omega}) = |X(e^{j\Omega})|^2 \iff x[k] * x[-k] = \sum_{n=-\infty}^{\infty} x[n+k]x[n] = \overline{R}_{xx}[k] \tag{2.43}$$

The function $\overline{R}_{xx}[k] = x[k]*x[-k]$ is referred to as the deterministic autocorrelation function of the signal $x[n]$, and we have just established that the transform of the deterministic autocorrelation function is the energy spectral density $\overline{S}_{xx}(e^{j\Omega})$. A basic Fourier transform property tells us that $\overline{R}_{xx}[0]$ — which is the signal energy $\sum_{n=-\infty}^{\infty} x^2[n]$ — is the area under the Fourier transform of $R_{xx}[k]$, scaled by $1/(2\pi)$, namely the scaled area under $\overline{S}_{xx}(e^{j\Omega}) = |X(e^{j\Omega})|^2$; this is just Parseval's identity, of course.

The deterministic autocorrelation function measures how alike a signal and its time-shifted version are, in a total-squared-error sense. More specifically, in discrete-time the total squared error between the signal and its time-shifted version is given by

$$\begin{aligned} \sum_{n=-\infty}^{\infty} (x[n+k] - x[n])^2 &= \sum_{n=-\infty}^{\infty} |x[n+k]|^2 \\ &\quad + \sum_{n=-\infty}^{\infty} |x[n]|^2 - 2 \sum_{n=-\infty}^{\infty} x[n+k]x[n] \\ &= 2(\overline{R}_{xx}[0] - \overline{R}_{xx}[k]) \end{aligned} \tag{2.44}$$

Since the total squared error is always nonnegative, it follows that $\overline{R}_{xx}[k] \le \overline{R}_{xx}[0]$, and that the larger the deterministic autocorrelation $\overline{R}_{xx}[k]$ is, the closer the signal $x[n]$ and its time-shifted version $x[n+k]$ are.

Corresponding results hold in continuous time, and in particular

$$\overline{S}_{xx}(j\omega) = |X(j\omega)|^2 \iff x(\tau) * x(-\tau) = \int_{-\infty}^{\infty} x(t+\tau)x(t)dt = \overline{R}_{xx}(\tau) \tag{2.45}$$

where $\overline{R}_{xx}(t)$ is the deterministic autocorrelation function of $x(t)$.

2.4 THE BILATERAL LAPLACE AND $\mathcal{Z}$-TRANSFORMS

The Laplace and $\mathcal{Z}$-transforms can be thought of as extensions of Fourier transforms and are useful for a variety of reasons. They permit a transform treatment of certain classes of signals for which the Fourier transform does not converge. They also augment our understanding of Fourier transforms by moving us into the complex plane, where the theory of complex functions can be applied. We begin in Section 2.4.1 with a detailed review of the bilateral $\mathcal{Z}$-transform. In Section 2.4.3 we give a briefer review of the bilateral Laplace transform, paralleling the discussion in Section 2.4.1.

2.4.1 The Bilateral $\mathcal{Z}$-Transform

The bilateral $\mathcal{Z}$-transform is defined as:

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n=-\infty}^{\infty} x[n]z^{-n} \tag{2.46}$$

Here z is a complex variable, which we can also represent in polar form as

$$z = re^{j\Omega}\,, \quad r \geq 0\,, \;\; -\pi < \Omega \leq \pi \tag{2.47}$$

so

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]r^{-n}e^{-j\Omega n} \tag{2.48}$$

The DTFT corresponds to fixing $r = 1$, in which case z takes values on the unit circle. However there are many useful signals for which the infinite sum does not converge (even in the sense of generalized functions) for z confined to the unit circle. The term z^{-n} in the definition of the $\mathcal{Z}$-transform introduces a factor r^{-n} into the infinite sum, which permits the sum to converge (provided r is appropriately restricted) for interesting classes of signals, many of which do not have discrete-time Fourier transforms.

More specifically, note from (2.48) that $X(z)$ can be viewed as the DTFT of $x[n]r^{-n}$. If $r > 1$, then r^{-n} decays geometrically for positive n and grows geometrically for negative n. For $0 < r < 1$, the opposite happens. Consequently, there are many sequences for which $x[n]$ is not absolutely summable but $x[n]r^{-n}$ is, for some range of values of r.

For example, consider $x_1[n] = a^n u[n]$. If $|a| > 1$, this sequence does not have a DTFT. However, for any a, $x[n]r^{-n}$ is absolutely summable provided $r > |a|$. In particular, for example,

$$X_1(z) = 1 + az^{-1} + a^2z^{-2} + \cdots \tag{2.49}$$

$$= \frac{1}{1 - az^{-1}}\,, \quad |z| = r > |a| \tag{2.50}$$

As a second example, consider $x_2[n] = -a^n u[-n-1]$. This signal does not have a DTFT if $|a| < 1$. However, provided $r < |a|$,

$$X_2(z) = -a^{-1}z - a^{-2}z^2 - \cdots \tag{2.51}$$

$$= \frac{-a^{-1}z}{1 - a^{-1}z}\,, \quad |z| = r < |a| \tag{2.52}$$

$$= \frac{1}{1 - az^{-1}}\,, \quad |z| = r < |a| \tag{2.53}$$

The $\mathcal{Z}$-transforms of the two distinct signals $x_1[n]$ and $x_2[n]$ above get condensed to the same rational expressions, but for different regions of convergence. Hence the ROC is a critical part of the specification of the transform.

When $x[n]$ is a sum of left-sided and/or right-sided DT exponentials, with each term of the form illustrated in the examples above, then $X(z)$ will be rational in z (or equivalently, in z^{-1}):

$$X(z) = \frac{Q(z)}{P(z)} \tag{2.54}$$

with $Q(z)$ and $P(z)$ being polynomials in z.

Rational $\mathcal{Z}$-transforms are typically depicted by a pole-zero plot in the z-plane, with the ROC appropriately indicated. This information uniquely specifies the signal, apart from a constant amplitude scaling. Note that there can be no poles in the ROC, since the transform is required to be finite in the ROC. $\mathcal{Z}$-transforms are often written as ratios of polynomials in z^{-1}. However, the pole-zero plot in the z-plane refers to the polynomials in z. Also note that if poles or zeros at $z = \infty$ are counted, then any ratio of polynomials always has exactly the same number of poles as zeros.

Region of Convergence. To understand the complex-function properties of the $\mathcal{Z}$-transform, we split the infinite sum that defines it into non-negative-time and negative-time portions: The non-negative-time or one-sided $\mathcal{Z}$-transform is defined by

$$\sum_{n=0}^{\infty} x[n]z^{-n} \tag{2.55}$$

and is a power series in z^{-1}. The convergence of the finite sum $\sum_{n=0}^{N} x[n]z^{-n}$ as $N \to \infty$ is governed by the radius of convergence $R_1 \geq 0$, of the power series, i.e. the series converges for each z such that $|z| > R_1$. The resulting function of z is an analytic function in this region, i.e., has a well-defined derivative with respect to the complex variable z at each point in this region, which is what gives the function its nice properties. The infinite sum diverges for $|z| < R_1$. The behavior of the sum on the circle $|z| = R_1$ requires closer examination, and depends on the particular series; the series may converge (but may not converge absolutely) at all points, some points, or no points on this circle. The region $|z| > R_1$ is referred to as the region of convergence (ROC) of the power series.

Next consider the negative-time part:

$$\sum_{n=-\infty}^{-1} x[n]z^{-n} = \sum_{m=1}^{\infty} x[-m]z^{m} \tag{2.56}$$

which is a power series in z, and has a radius of convergence R_2. The series converges (absolutely) for $|z| < R_2$, which constitutes its ROC; the series is an analytic function in this region. The sum diverges for $|z| > R_2$; the behavior for the circle $|z| = R_2$ takes closer examination, and depends on the particular series; the series may converge (but may not converge absolutely) at all points, some points, or no points on this circle. If $R_1 < R_2$ then the $\mathcal{Z}$-transform converges (absolutely) for $R_1 < |z| < R_2$; this annular region is its ROC, and is denoted by $\mathcal{R}_X$. The transform is analytic in this region. The sum that defines the transform diverges for $|z| < R_1$ and $|z| > R_2$. If $R_1 > R_2$, then the $\mathcal{Z}$-transform does not exist (e.g., for $x[n] = 0.5^n u[-n-1] + 2^n u[n]$). If $R_1 = R_2$, then the transform may exist in a technical sense, but is not useful as a $\mathcal{Z}$-transform because it has no ROC. However, if $R_1 = R_2 = 1$, then we may still be able to compute and use a DTFT (e.g., for $x[n] = 3$, all n; or for $x[n] = (\sin\omega_0 n)/(\pi n)$).

Relating the ROC to Signal Properties. For an absolutely summable signal (such as the impulse response of a BIBO-stable system), i.e., an ℓ_1-signal, the unit circle must lie in the ROC or must be a boundary of the ROC. Conversely, we can conclude that a signal is ℓ_1 if the ROC contains the unit circle because the transform converges absolutely in its ROC. If the unit circle constitutes a boundary of the ROC, then further analysis is generally needed to determine if the signal is ℓ_1. Rational transforms always have a pole on the boundary of the ROC, as elaborated on below, so if the unit circle is on the boundary of the ROC of a rational transform, then there is a pole on the unit circle, and the signal cannot be ℓ_1.

For a right-sided signal it is the case that $R_2 = \infty$, i.e., the ROC extends everywhere in the complex plane outside the circle of radius R_1, up to (and perhaps including) ∞. The ROC includes ∞ if the signal is 0 for negative time.

We can state a converse result if, for example, we know the signal comprises only sums of one-sided exponentials, of the form obtained when inverse transforming a rational transform. In this case, if $R_2 = \infty$, then the signal must be right-sided; if the ROC includes ∞, then the signal must be causal, i.e., zero for $n < 0$.

For a left-sided signal, one has $R_1 = 0$, i.e., the ROC extends inwards from the circle of radius R_2, up to (and perhaps including) 0. The ROC includes 0 if the signal is 0 for positive time.

In the case of signals that are sums of one-sided exponentials, we have a converse: if $R_1 = 0$, then the signal must be left-sided; if the ROC includes 0, then the signal must be anti-causal, i.e., zero for $n > 0$.

It is also important to note that the ROC cannot contain poles of the $\mathcal{Z}$-transform, because poles are values of z where the transform has infinite magnitude, while the ROC comprises values of z where the transform converges. For signals with *rational* transforms, one can use the fact that such signals are sums of one-sided exponentials to show that the possible boundaries of the ROC are in fact precisely determined by the locations of the poles. Specifically:

(a) the outer bounding circle of the ROC in the rational case contains a pole and/or has radius ∞. If the outer bounding circle is at infinity, then (as we have already noted) the signal is right-sided, and is in fact causal if there is no pole at ∞;

(b) the inner bounding circle of the ROC in the rational case contains a pole and/or has radius 0. If the inner bounding circle reduces to the point 0, then (as we have already noted) the signal is left-sided, and is in fact anti-causal if there is no pole at 0.

2.4.2 The Inverse $\mathcal{Z}$-Transform

One way to invert a rational $\mathcal{Z}$-transform is through the use of a partial fraction expansion, then either directly "recognizeing" the inverse transform of each term in the partial fraction representation, or expanding the term in a power series that converges for z in the specified ROC. For example, a term of the form

$$\frac{1}{1 - az^{-1}} \tag{2.57}$$

can be expanded in a power series in az^{-1} if $|a| < |z|$ for z in the ROC, and expanded in a power series in $a^{-1}z$ if $|a| > |z|$ for z in the ROC. Carrying out this procedure for each term in a partial fraction expansion, we find that the signal $x[n]$ is a sum of left-sided and/or right-sided exponentials. For non-rational transforms, where there may not be a partial fraction expansion to simplify the process, it is still reasonable to attempt the inverse transformation by expansion into a power series consistent with the given ROC.

Although we will generally use partial fraction or power series methods to invert $\mathcal{Z}$-transforms, there is an explicit formula that is similar to that of the inverse DTFT, specifically,

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(z) z^n \, d\omega \Big|_{z=\bar{r}e^{j\omega}} \tag{2.58}$$

where the constant $\bar{r}$ is chosen to place z in the ROC, $\mathcal{R}_X$. This is not the most general inversion formula, but is sufficient for us, and shows that $x[n]$ is expressed as a weighted combination of discrete-time exponentials.

As is the case for Fourier transforms, there are many useful $\mathcal{Z}$-transform pairs and properties developed and tabulated in basic texts on signals and systems. Appropriate use of transform pairs and properties is often the basis for obtaining the $\mathcal{Z}$-transform or the inverse $\mathcal{Z}$-transform of many other signals.

2.4.3 The Bilateral Laplace Transform

As with the $\mathcal{Z}$-transform, the Laplace transform is introduced in part to handle important classes of signals that don't have CTFT's, but also enhances our understanding of the CTFT. The definition of the Laplace transform is

$$X(s) = \int_{-\infty}^{\infty} x(t)\, e^{-st}\, dt \tag{2.59}$$

where s is a complex variable, $s = \sigma + j\omega$. The Laplace transform can thus be thought of as the CTFT of $x(t)\, e^{-\sigma t}$. With σ appropriately chosen, the integral (2.59) can exist even for signals that have no CTFT.

The development of the Laplace transform parallels closely that of the $\mathcal{Z}$-transform in the preceding section, but with e^{σ} playing the role that r did in Section 2.4.1. The (interior of the) set of values of s for which the defining integral converges, as the limits on the integral approach $\pm\infty$, comprises the region of convergence (ROC) for the transform $X(s)$. The ROC is now determined by the minimum and maximum allowable values of σ, say σ_1 and σ_2 respectively. We refer to σ_1, σ_2 as the abscissa of convergence. The corresponding ROC is a vertical strip between σ_1 and σ_2 in the complex plane, $\sigma_1 < Re(s) < \sigma_2$. Equation (2.59) converges absolutely within the ROC; convergence at the left and right bounding vertical lines of the strip has to be separately examined. Furthermore, the transform is analytic (i.e., differentiable as a complex function) throughout the ROC. The strip may extend to $\sigma_1 = -\infty$ on the left, and to $\sigma_2 = +\infty$ on the right. If the strip collapses to a line (so that the ROC vanishes), then the Laplace transform is not useful (except if the line happens to be the $j\omega$ axis, in which case a CTFT analysis may perhaps be recovered).

For example, consider $x_1(t) = e^{at}u(t)$; the integral in (2.59) evaluates to $X_1(s) = 1/(s-a)$ provided $\mathrm{Re}\{s\} > a$. On the other hand, for $x_2(t) = -e^{at}u(-t)$, the integral in (2.59) evaluates to $X_2(s) = 1/(s-a)$ provided $\mathrm{Re}\{s\} < a$. As with the $\mathcal{Z}$-transform, note that the expressions for the transforms above are identical; they are distinguished by their distinct regions of convergence.

The ROC may be related to properties of the signal. For example, for absolutely integrable signals, also referred to as L_1 signals, the integrand in the definition of the Laplace transform is absolutely integrable on the $j\omega$ axis, so the $j\omega$ axis is in the ROC or on its boundary. In the other direction, if the $j\omega$ axis is strictly in the ROC, then the signal is L_1, because the integral converges absolutely in the ROC. Recall that a system has an L_1 impulse response if and only if the system is BIBO stable, so the result here is relevant to discussions of stability: if the $j\omega$ axis is strictly in the ROC of the system function, then the system is BIBO stable.

For right-sided signals, the ROC is some right-half-plane (i.e. all s such that $\mathrm{Re}\{s\} > \sigma_1$). Thus the system function of a causal system will have an ROC that is some right-half-plane. For left-sided signals, the ROC is some left-half-plane. For signals with rational transforms, the ROC contains no poles, and the boundaries of the ROC will have poles. Since the location of the ROC of a transfer function relative to the imaginary axis relates to BIBO stability, and since the poles identify the boundaries of the ROC, the poles relate to stability. In particular, a system with a right-sided impulse response (e.g., a causal system) will be stable if and only if all its poles are in the left-half-plane, because this is precisely the condition that allows the ROC to contain the imaginary axis. Also note that a signal with a rational transform is causal if and only if it is right-sided.

A further property worth recalling is connected to the fact that exponentials are eigenfunctions of LTI systems. If we denote the Laplace transform of the impulse response $h(t)$ of an LTI system by $H(s)$, referred to as the system function or transfer function, then $e^{s_0 t}$ at the input of the system yields $H(s_0)\,e^{s_0 t}$ at the output, provided s_0 is in the ROC of the transfer function.

2.5 DISCRETE-TIME PROCESSING OF CONTINUOUS-TIME SIGNALS

Many modern systems for applications such as communication, entertainment, navigation and control are a combination of continuous-time and discrete-time subsystems, exploiting the inherent properties and advantages of each. In particular, the discrete-time processing of continuous-time signals is common in such applications, and we describe the essential ideas behind such processing here. As with the earlier sections, we assume that this discussion is primarily a review of familiar material, included here to establish notation and for convenient reference from later chapters in this text. In this section, and throughout this text, we will often be relating the CTFT of a continuous-time signal and the DTFT of a discrete-time signal obtained from samples of the continuous-time signal. We will use the subscripts c and d when necessary to help keep clear which signals are CT and which are DT.

2.5.1 Basic Structure for DT Processing of CT Signals

The basic structure is shown in Figure 2.4. As indicated, the processing involves continuous-to-discrete or C/D conversion to obtain a sequence of samples of the CT signal, then DT filtering to produce a sequence of samples of the desired CT output, then discrete-to-continuous or D/C conversion to reconstruct this desired CT signal from the sequence of samples. We will often restrict ourselves to conditions such that the overall system in Figure 2.4 is equivalent to an LTI continuous-time system. The necessary conditions for this typically include restricting the DT filtering to be LTI processing by a system with frequency response $H_d(e^{j\Omega})$, and also requiring that the input $x_c(t)$ be appropriately bandlimited. To satisfy the latter requirement, it is typical to precede the structure in the figure by a filter whose purpose is to ensure that $x_c(t)$ is essentially bandlimited. While this filter is often referred to as an anti-aliasing filter, we can often allow some aliasing in the C/D conversion if the discrete-time system removes the aliased components; the overall system can then still be a CT LTI system.

The ideal C/D converter in Figure 2.4 has as its output a sequence of samples of $x_c(t)$ with a specified sampling interval T_1, so that the DT signal is $x_d[n] = x_c(nT_1)$. Conceptually, therefore, the ideal C/D converter is straightforward. A practical analog-to-digital (or A/D) converter also quantizes the signal to one of a finite set of output levels. However, in this text we do not consider the additional effects of quantization.

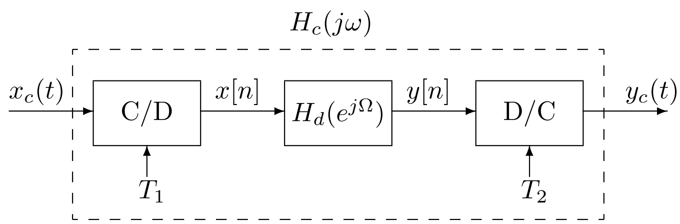

FIGURE 2.4 DT processing of CT signals.

In the frequency domain, the CTFT of $x_c(t)$ and the DTFT of $x_d[n]$ are related by

$$X_d\left(e^{j\Omega}\right)\Big|_{\Omega=\omega T_1} = \frac{1}{T_1}\sum_k X_c\left(j\omega - jk\frac{2\pi}{T_1}\right) . \tag{2.60}$$

When $x_c(t)$ is sufficiently bandlimited so that

$$X_c(j\omega) = 0 \; , \qquad |\omega| \geq \frac{\pi}{T_1} \tag{2.61}$$

then (2.60) can be rewritten as

$$X_d\left(e^{j\Omega}\right)\Big|_{\Omega=\omega T_1} = \frac{1}{T_1}X_c(j\omega) \qquad |\omega| < \pi/T_1 \tag{2.62a}$$

or equivalently

$$X_d\left(e^{j\Omega}\right) = \frac{1}{T_1}X_c\left(j\frac{\Omega}{T_1}\right) \qquad |\Omega| < \pi \; . \tag{2.62b}$$

Note that $X_d(e^{j\Omega})$ is extended periodically outside the interval $|\Omega| < \pi$. The fact that the above equalities hold under the condition (2.61) is the content of the sampling theorem.

The ideal D/C converter in Figure 2.4 is defined through the interpolation relation

$$y_c(t) = \sum_n y_d[n]\frac{\sin\left(\pi\left(t - nT_2\right)/T_2\right)}{\pi(t - nT_2)/T_2} \tag{2.63}$$

which shows that $y_c(nT_2) = y_d[n]$. Since each term in the above sum is bandlimited to $|\omega| < \pi/T_2$, the CT signal $y_c(t)$ is also bandlimited to this frequency range, so this D/C converter is more completely referred to as the ideal bandlimited interpolating converter. (The C/D converter in Figure 2.4, under the assumption (2.61), is similarly characterized by the fact that the CT signal $x_c(t)$ is the ideal bandlimited interpolation of the DT sequence $x_d[n]$.)

Because $y_c(t)$ is bandlimited and $y_c(nT_2) = y_d[n]$, analogous relations to (2.62) hold between the DTFT of $y_d[n]$ and the CTFT of $y_c(t)$:

$$Y_d\left(e^{j\Omega}\right)\Big|_{\Omega=\omega T_2} = \frac{1}{T_2} Y_c(j\omega) \qquad |\omega| < \pi/T_2 \tag{2.64a}$$

or equivalently

$$Y_d\left(e^{j\Omega}\right) = \frac{1}{T_2} Y_c\left(j\frac{\Omega}{T_2}\right) \qquad |\Omega| < \pi \tag{2.64b}$$

One conceptual representation of the ideal D/C converter is given in Figure 2.5. This figure interprets (2.63) to be the result of evenly spacing a sequence of impulses at intervals of T_2 — the reconstruction interval — with impulse strengths given by the $y_d[n]$, then filtering the result by an ideal low-pass filter $L(j\omega)$ of amplitude T_2 in the passband $|\omega| < \pi/T_2$. This operation produces the bandlimited continuous-time signal $y_c(t)$ that interpolates the specified sequence values $y_d[n]$ at the instants $t = nT_2$, i.e., $y_c(nT_2) = y_d[n]$.

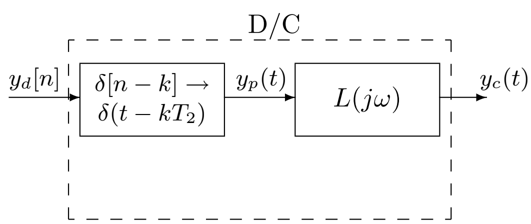

FIGURE 2.5 Conceptual representation of processes that yield ideal D/C conversion, interpolating a DT sequence into a bandlimited CT signal using reconstruction interval T_2.

2.5.2 DT Filtering, and Overall CT Response

Suppose from now on, unless stated otherwise, that $T_1 = T_2 = T$. If in Figure 2.4 the bandlimiting constraint of (2.61) is satisfied, and if we set $y_d[n] = x_d[n]$, then $y_c(t) = x_c(t)$. More generally, when the DT system in Figure 2.4 is an LTI DT filter with frequency response $H_d\left(e^{j\Omega}\right)$, so

$$Y_d(e^{j\Omega}) = H_d(e^{j\Omega})X_d(e^{j\Omega}) \tag{2.65}$$

and provided any aliased components of $x_c(t)$ are eliminated by $H_d(e^{j\Omega})$, then assembling (2.62), (2.64) and (2.65) yields:

$$Y_c(j\omega) = H_d\left(e^{j\Omega}\right)\Big|_{\Omega=\omega T} X_c(j\omega) \qquad |\omega| < \pi/T \tag{2.66}$$

The action of the overall system is thus equivalent to that of a CT filter whose frequency response is

$$H_c(j\omega) = H_d\left(e^{j\Omega}\right)\Bigg|_{\Omega=\omega T} \qquad |\omega| < \pi/T. \tag{2.67}$$

In other words, under the bandlimiting and sampling rate constraints mentioned above, the overall system behaves as an LTI CT filter, and the response of this filter is related to that of the embedded DT filter through a simple frequency scaling. The sampling rate can be lower than the Nyquist rate, provided that the DT filter eliminates any aliased components.

If we wish to use the system in Figure 2.4 to implement a CT LTI filter with frequency response $H_c(j\omega)$, we choose $H_d\left(e^{j\Omega}\right)$ according to (2.67), provided that $x_c(t)$ is appropriately bandlimited.

If $H_c(j\omega) = 0$ for $|\omega| \geq \pi/T$, then (2.67) also corresponds to the following relation between the DT and CT impulse responses:

$$h_d[n] = T\,h_c(nT) \tag{2.68}$$

The DT filter is therefore termed an impulse-invariant version of the CT filter. When $x_c(t)$ and $H_d(e^{j\Omega})$ are not sufficiently bandlimited to avoid aliased components in $y_d[n]$, then the overall system in Figure 2.4 is no longer time invariant. It is, however, still linear since it is a cascade of linear subsystems.

The following two important examples illustrate the use of (2.67) as well as Figure 2.4, both for DT processing of CT signals and for interpretation of an important DT system, whether or not this system is explicitly used in the context of processing CT signals.

EXAMPLE 2.3 Digital Differentiator

In this example we wish to implement a CT differentiator using a DT system in the configuration of Figure 2.4 . We need to choose $H_d\left(e^{j\Omega}\right)$ so that $y_c(t) = \frac{dx_c(t)}{dt}$, assuming that $x_c(t)$ is bandlimited to π/T. The desired overall CT frequency response is therefore

$$H_c(j\omega) = \frac{Y_c(j\omega)}{X_c(j\omega)} = j\omega \tag{2.69}$$

Consequently, using (2.67) we choose $H_d(e^{j\Omega})$ such that

$$H_d\left(e^{j\Omega}\right)\Bigg|_{\Omega=\omega T} = j\omega \qquad |\omega| < \frac{\pi}{T} \tag{2.70a}$$

or equivalently

$$H_d\left(e^{j\Omega}\right) = j\Omega/T \qquad |\Omega| < \pi \tag{2.70b}$$

A discrete-time system with the frequency response in (2.70b) is commonly referred to as a digital differentiator. To understand the relation between the input $x_d[n]$

and output $y_d[n]$ of the digital differentiator, note that $y_c(t)$ — which is the bandlimited interpolation of $y_d[n]$ — is the derivative of $x_c(t)$, and $x_c(t)$ in turn is the bandlimited interpolation of $x_d[n]$. It follows that $y_d[n]$ can, in effect, be thought of as the result of sampling the derivative of the bandlimited interpolation of $x_d[n]$.

EXAMPLE 2.4 Half-Sample Delay

It often arises in designing discrete-time systems that a phase factor of the form $e^{-j\alpha\Omega}$, $|\Omega| < \pi$, is included or required. When α is an integer, this has a straightforward interpretation, since it corresponds simply to an integer shift by α of the time sequence.

When α is not an integer, the interpretation is not as straightforward, since a DT sequence can only be directly shifted by integer amounts. In this example we consider the case of $\alpha = 1/2$, referred to as a half-sample delay. To provide an interpretation, we consider the implications of choosing the DT system in Figure 2.4 to have frequency response

$$H_d(e^{j\Omega}) = e^{-j\Omega/2} \qquad |\Omega| < \pi \tag{2.71}$$

Whether or not $x_d[n]$ explicitly arose by sampling a CT signal, we can associate with $x_d[n]$ its bandlimited interpolation $x_c(t)$ for any specified sampling or reconstruction interval T. Similarly, we can associate with $y_d[n]$ its bandlimited interpolation $y_c(t)$ using the reconstruction interval T. With $H_d\left(e^{j\Omega}\right)$ given by (2.71), the equivalent CT frequency response relating $y_c(t)$ to $x_c(t)$ is

$$H_c(j\omega) = e^{-j\omega T/2} \tag{2.72}$$

representing a time delay of $T/2$, which is half the sample spacing; consequently, $y_c(t) = x_c(t - T/2)$. We therefore conclude that for a DT system with frequency response given by (2.71), the DT output $y_d[n]$ corresponds to samples of the half-sample delay of the bandlimited interpolation of the input sequence $x_d[n]$. Note that in this interpretation the choice for the value of T is immaterial. (Even if $x_d[n]$ had been the result of regular sampling of a CT signal, that specific sampling period is not required in the interpretation above.)

The preceding interpretation allows us to find the unit-sample (or impulse) response of the half-sample delay system through a simple argument. If $x_d[n] = \delta[n]$, then $x_c(t)$ must be the bandlimited interpolation of this (with some T that we could have specified to take any particular value), so

$$x_c(t) = \frac{\sin(\pi t/T)}{\pi t/T} \tag{2.73}$$

and therefore

$$y_c(t) = \frac{\sin\Big(\pi(t - (T/2))/T\Big)}{\pi(t - (T/2))/T} \tag{2.74}$$

which shows that the desired unit-sample response is

$$y_d[n] = h_d[n] = \frac{\sin\Big(\pi(n-(1/2))\Big)}{\pi(n-(1/2))} \tag{2.75}$$

This discussion of a half-sample delay also generalizes in a straightforward way to any integer or non-integer choice for the value of α.

2.5.3 Non-Ideal D/C converters

In Section 2.5.1 we defined the ideal D/C converter through the bandlimited interpolation formula (2.63); see also Figure 2.5, which corresponds to processing a train of impulses with strengths equal to the sequence values $y_d[n]$ through an ideal low-pass filter. A more general class of D/C converters, which includes the ideal converter as a particular case, creates a CT signal $y_c(t)$ from a DT signal $y_d[n]$ according to the following:

$$y_c(t) = \sum_{n=-\infty}^{\infty} y_d[n]\, p(t-nT) \tag{2.76}$$

where $p(t)$ is some selected basic pulse shape and T is the reconstruction interval or pulse repetition interval. This too can be seen as the result of processing an impulse train of sequence values through a filter, but a filter that has impulse response $p(t)$ rather than that of the ideal low-pass filter. The CT signal $y_c(t)$ is thus constructed by adding together shifted and scaled versions of the basic pulse shape; the number $y_d[n]$ scales $p(t-nT)$, which is the basic pulse delayed by nT. Note that the ideal bandlimited interpolating converter of (2.63) is obtained by choosing

$$p(t) = \frac{\sin(\pi t/T)}{(\pi t/T)} \tag{2.77}$$

We shall be talking in more detail in Chapter 12 about the interpretation of (2.76) as pulse amplitude modulation (PAM) for communicating DT information over a CT channel.

The relationship (2.76) can also be described quite simply in the frequency domain. Taking the CTFT of both sides, denoting the CTFT of $p(t)$ by $P(j\omega)$, and using the fact that delaying a signal by t_0 in the time domain corresponds to multiplication by $e^{-j\omega t_0}$ in the frequency domain, we get

$$\begin{aligned} Y_c(j\omega) &= \Big(\sum_{n=-\infty}^{\infty} y_d[n]\, e^{-jn\omega T}\Big) P(j\omega) \\ &= Y_d(e^{j\Omega})\Big|_{\Omega=\omega T} P(j\omega) \end{aligned} \tag{2.78}$$

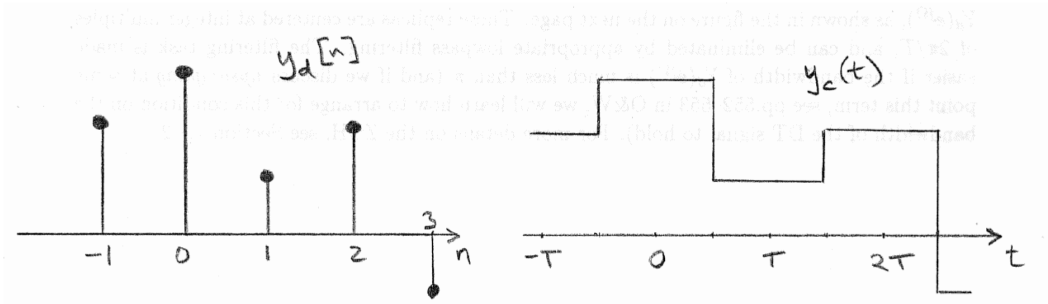

FIGURE 2.6 A centered zero-order hold (ZOH)

In the particular case where $p(t)$ is the sinc pulse in (2.77), with transform $P(j\omega)$ corresponding to an ideal low-pass filter of amplitude T for $|\omega| < \pi/T$ and 0 outside this band, we recover the relation (2.64).

In practice an ideal low-pass filter can only be approximated, with the accuracy of the approximation closely related to cost of implementation. A commonly used simple approximation is the (centered) zero-order hold (ZOH), specified by the choice

$$p(t) = \begin{cases} 1 & \text{for } |t| < (T/2) \\ 0 & \text{elsewhere} \end{cases} \tag{2.79}$$

This D/C converter holds the value of the DT signal at time n, namely the value $y_d[n]$, for an interval of length T centered at nT in the CT domain, as illustrated in Figure 2.6. Such ZOH converters are very commonly used. Another common choice is a centered first-order hold (FOH), for which $p(t)$ is triangular as shown in Figure 2.7. Use of the FOH represents linear interpolation between the sequence values.

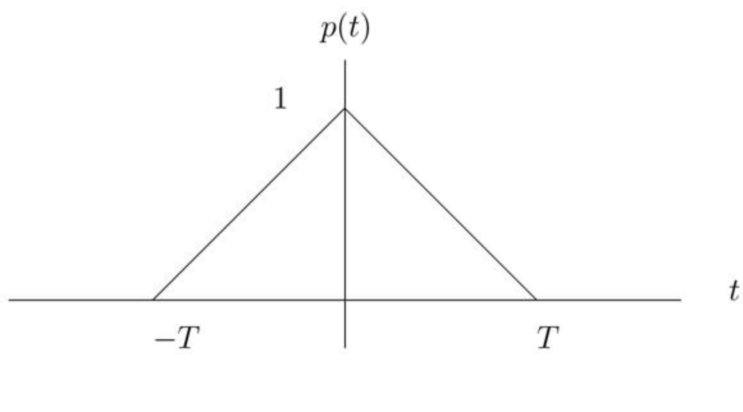

FIGURE 2.7 A centered first order hold (FOH)