Chapter 1

INTRODUCTION

1.1 Motivation

Wireless communications is one of the most active areas of technology development of our time. This development is being driven primarily by the transformation of what has been largely a medium for supporting voice telephony into a medium for supporting other services, such as the transmission of video, images, text, and data. Thus, similar to the developments in wireline capacity in the 1990s, the demand for new wireless capacity is growing at a very rapid pace. Although there are, of course, still a great many technical problems to be solved in wireline communications, demands for additional wireline capacity can be fulfilled largely with the addition of new private infrastructure, such as additional optical fiber, routers, switches, and so on. On the other hand, the traditional resources that have been used to add capacity to wireless systems are radio bandwidth and transmitter power. Unfortunately, these two resources are among the most severely limited in the deployment of modern wireless networks: radio bandwidth because of the very tight situation with regard to useful radio spectrum, and transmitter power because mobile and other portable services require the use of battery power, which is limited. These two resources are simply not growing or improving at rates that can support anticipated demands for wireless capacity. On the other hand, one resource that is growing at a very rapid rate is that of processing power. Moore's Law, which asserts a doubling of processor capabilities every 18 months, has been quite accurate over the past 20 years, and its accuracy promises to continue for years to come. Given these circumstances, there has been considerable research effort in recent years aimed at developing new wireless capacity through the deployment of greater intelligence in wireless networks (see, e.g., [145,146,270,376,391] for reviews of some of this work). A key aspect of this movement has been the development of novel signal transmission techniques and advanced receiver signal processing methods that allow for significant increases in wireless capacity without attendant increases in bandwidth or power requirements. The purpose of this book is to present some of the most recent of these receiver signal processing methods in a single place and in a unified framework.

Wireless communications today covers a very wide array of applications. The telecommunications industry is one of the largest industries worldwide, with more than $1 trillion in annual revenues for services and equipment. (To put this in per-

spective, this number is comparable to the gross domestic product of many of the world's richest countries, including France, Italy, and the United Kingdom.) The largest and most noticeable part of the telecommunications business is telephony. The principal wireless component of telephony is mobile (i.e., cellular) telephony. The worldwide growth rate in cellular telephony is very aggressive, and analysts report that the number of cellular telephony subscriptions worldwide has now surpassed the number of wireline (i.e., fixed) telephony subscriptions. Moreover, at the time of this writing in 2003, the number of cellular telephony subscriptions worldwide is reportedly on the order of 1.2 billion. These numbers make cellular telephony a very important driver of wireless technology development, and in recent years the push to develop new mobile data services, which go collectively under the name *third-generation* (3G) *cellular*, has played a key role in motivating research in new signal processing techniques for wireless. However, cellular telephony is only one of a very wide array of wireless technologies that are being developed very rapidly at the present time. Among other technologies are wireless piconetworking (as exemplified by the Bluetooth radio-on-a-chip) and other personal area network (PAN) systems (e.g., the IEEE 802.15 family of standards), wireless local area network (LAN) systems (exemplified by the IEEE 802.11 and HiperLAN families of standards, called WiFi systems), wireless metropolitan area network (MAN) systems (exemplified by the IEEE 802.16 family of standards, called WiMax systems), other wireless local loop (WLL) systems, and a variety of satellite systems. These additional wireless technologies provide a basis for a very rich array of applications, including local telephony service, broadband Internet access, and distribution of high-rate entertainment content such as high-definition video and high-quality audio to the home, within the home, to automobiles, and so on (see, e.g., [9, 41, 42, 132, 159, 161, 164, 166, 344, 361, 362, 365, 393–395, 429, 437, 449, 457, 508, 558, 559] for further discussion of these and related applications). Like 3G, these technologies have spurred considerable research in signal processing for wireless.

These technologies are supported by a number of transmission and channel-assignment techniques, including time-division multiple access (TDMA), code-division multiple access (CDMA), and other spread-spectrum systems, orthogonal frequency-division multiplexing (OFDM) and other multicarrier systems, and high-rate single-carrier systems. These techniques are chosen primarily to address the physical properties of wireless channels, among the most prominent of which are multipath fading, dispersion, and interference. In addition to these temporal transmission techniques, there are spatial techniques, notably beamforming and space-time coding, that can be applied at the transmitter to exploit the spatial and angular diversity of wireless channels. To obtain maximal benefit from these transmission techniques, to exploit the diversity opportunities of the wireless channel, and to mitigate the impairments of the wireless channel, advanced receiver signal processing techniques are of interest. These include channel equalization to combat dispersion, RAKE combining to exploit resolvable multipath, multiuser detection to mitigate multiple-access interference, suppression methods for co-channel interference, beamforming to exploit spatial diversity, and space-time processing to

jointly exploit temporal and spatial properties of the signaling environment. These techniques are all described in the ensuing chapters.

1.2   Wireless Signaling Environment

1.2.1   Single-User Modulation Techniques

To discuss advanced receiver signal processing methods for wireless, it is useful first to specify a general model for the signal received by a wireless receiver. To do so, we can first think of a single transmitter, transmitting a sequence or *frame* $\{b[0], b[1], \ldots, b[M-1]\}$ of channel symbols over a wireless channel. These symbols can be binary (e.g., ± 1), or they may take on more general values from a finite alphabet of complex numbers. In this treatment, we consider only *linear* modulation systems, in which the symbols are transmitted into the channel by being modulated linearly onto a signaling waveform to produce a transmitted signal of this form:

$$x(t) \; = \; \sum_{i=0}^{M-1} b[i] w_i(t), \tag{1.1}$$

where $w_i(\cdot)$ is the modulation waveform associated with the ith symbol. In this expression, the waveforms can be quite general. For example, a single-carrier modulation system with carrier frequency ω_c, baseband pulse shape $p(\cdot)$, and symbol rate $1/T$ is obtained by choosing

$$w_i(t) \; = \; A\, p(t - iT)\, e^{\jmath(\omega_c t + \phi)}, \tag{1.2}$$

where $A > 0$ and $\phi \in (-\pi, \pi)$ denote carrier amplitude and phase offset, respectively. The baseband pulse shape may, for example, be a simple unit-energy rectangular pulse of duration T:

$$p(t) \; = \; p_T(t) \; \triangleq \; \begin{cases} \dfrac{1}{\sqrt{T}}, & 0 \le t < T, \\ 0, & \text{otherwise,} \end{cases} \tag{1.3}$$

or it could be a raised-cosine pulse, a bandlimited pulse, and so on. Similarly, a direct-sequence spread-spectrum system is produced by choosing the waveforms as in (1.2) but with the baseband pulse shape chosen to be a spreading waveform:

$$p(t) \; = \; \sum_{j=0}^{N-1} c_j \psi(t - j\, T_c), \tag{1.4}$$

where N is the spreading gain, $c_0, c_1, \ldots, c_{N-1}$, is a pseudorandom spreading code (typically, $c_j \in \{+1, -1\}$), $\psi(\cdot)$ is the chip waveform, and $T_c \triangleq T/N$ is the chip interval. The chip waveform may, for example, be a unit-energy rectangular pulse of duration T_c:

$$\psi(t) \; = \; p_{T_c}(t). \tag{1.5}$$

Other choices of the chip waveform can also be made to lower the chip bandwidth. The spreading waveform of (1.4) is periodic when used in (1.2), since the same spreading code is repeated in every symbol interval. Some systems (e.g., CDMA systems for cellular telephony) operate with *long spreading codes*, for which the periodicity is much longer than a single symbol interval. This situation can be modeled by (1.1) by replacing $p(t)$ in (1.2) by a variant of (1.4) in which the spreading code varies from symbol to symbol; that is,

$$p_i(t) \;=\; \sum_{j=0}^{N-1} c_j^{(i)} \psi(t - j\,T_c). \tag{1.6}$$

Spread-spectrum modulation can also take the form of frequency hopping, in which the carrier frequency in (1.2) is changed over time according to a pseudorandom pattern. Typically, the carrier frequency changes at a rate much slower than the symbol rate, a situation known as *slow frequency hopping*; however, *fast hopping*, in which the carrier changes within a symbol interval, is also possible. Single-carrier systems, including both types of spread spectrum, are widely used in cellular standards, in wireless LANs, Bluetooth, and others (see, e.g., [42, 131, 150, 163, 178, 247, 338, 361, 362, 392, 394, 407, 408, 449, 523, 589]).

Multicarrier systems can also be modeled in the framework of (1.1) by choosing the signaling waveforms $\{w_i(\cdot)\}$ to be sinusoidal signals with different frequencies. In particular, (1.2) can be replaced by

$$w_i(t) \;=\; A\,p(t)\,e^{j(\omega_i t + \phi_i)}, \tag{1.7}$$

where now the frequency and phase depend on the symbol number i but all symbols are transmitted simultaneously in time with baseband pulse shape $p(\cdot)$. We can see that (1.2) is the counterpart of this situation with time and frequency reversed: All symbols are transmitted at the same frequency but at different times. (Of course, in practice, multiple symbols are sent in time sequence over each of the multiple carriers in multicarrier systems.) The individual carriers can also be direct-spread, and the baseband pulse shape used can depend on the symbol number i. (For example, the latter situation is used in *multicarrier CDMA*, in which a spreading code is used across the carrier frequencies.) A particular case of (1.7) is OFDM, in which the baseband pulse shape is a unit pulse p_T, the intercarrier spacing is $1/T$ cycles per second, and the phases are chosen so that the carriers are orthogonal at this spacing. (This is the minimal spacing for which such orthogonality can be maintained.) OFDM is widely believed to be among the most effective techniques for wireless broadband applications and is the basis for the IEEE 802.11a high-speed wireless LAN standard (see, e.g., [354] for a discussion of multicarrier systems).

An emerging type of wireless modulation scheme is ultra-wideband (UWB) modulation, in which data are transmitted with no carrier through the modulation of extremely short pulses. Either the timing or amplitude of these pulses can be used to carry the information symbols. Typical UWB systems involve the transmission of many repetitions of the same symbol, possibly with the use of a direct-sequence

type of spreading code from transmission to transmission (see, e.g., [569] for a basic description of UWB systems).

Further details on the modulation waveforms above and their properties will be introduced as needed throughout this treatment.
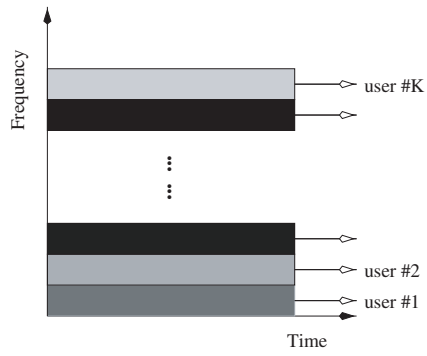
1.2.2   Multiple-Access Techniques

In Section 1.2.1 we discussed ways in which a symbol stream associated with a single user can be transmitted. Many wireless channels, particularly in emerging systems, operate as multiple-access systems, in which multiple users share the same radio resources.
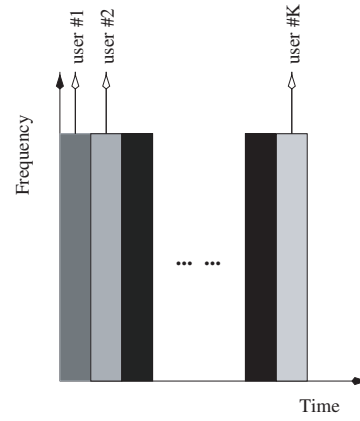
There are several ways in which radio resources can be shared among multiple users. These can be viewed as ways of allocating regions in frequency, space, and time to different users, as shown in Fig. 1.1. For example, a classic multiple-access technique is *frequency-division multiple access* (FDMA), in which the frequency band available for a given service is divided into subbands that are allocated to individual users who wish to use the service. Users are given exclusive use of their subband during their communication session, but they are not allowed to transmit signals within other subbands. FDMA is the principal multiplexing method used in radio and television broadcast and in first-generation (analog voice) cellular telephony systems, such as the Advanced Mobile Phone System (AMPS) and Nordic Mobile Telephone (NMT), developed primarily in the 1970s and 1980s (cf. [458]). FDMA is also used in some form in all other current cellular systems, in tandem with other multiple-access techniques that are used to further allocate the subbands to multiple users.

Similarly, users can share the channel on the basis of *time-division multiple access* (TDMA), in which time is divided into equal-length intervals, which are further divided into equal-length subintervals, or time slots. Each user is allowed to transmit throughout the entire allocated frequency band during a given slot in each interval but is not allowed to transmit during other time slots when other users are transmitting. So, whereas FDMA allows each user to use part of the spectrum all of the time, TDMA allows each user to use all of the spectrum part of the time. This method of channel sharing is widely used in wireless applications, notably in a number of second-generation cellular (i.e., digital voice) sytems, including the widely used Global System for Mobile (GSM) system [178, 407, 408] and in the IEEE 802.16 wireless MAN standards. A form of TDMA is also used in Bluetooth networks, in which one of the Bluetooth devices in the network acts as a network controller to poll the other devices in time sequence.
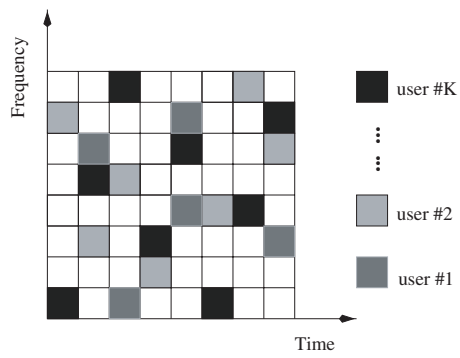
FDMA and TDMA systems are intended to assign orthogonal channels to all active users by giving each, for their exclusive use, a slice of the available frequency band or transmission time. These channels are said to be *orthogonal* because interference between users does not, in principle, arise in such assignments (although, in practice, there is often such interference, as discussed further below). *Code-division multiple access* (CDMA) assigns channels in a way that allows all users to use all of the available time and frequency resources simultaneously, through the assignment of a pattern or code to each user that specifies the way in which these resources

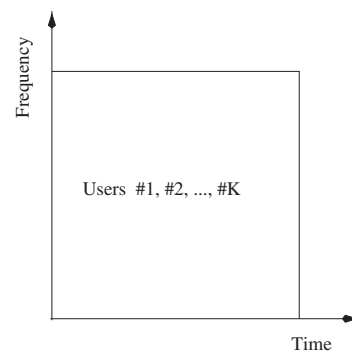**Frequency-Division Multiple-Access (FDMA)**

Time-Division Multiple-Access (TDMA)

Frequency-Hopping Code-Division
Multiple-Access (FH-CDMA)

Direct-Sequence Code-Division
Multiple-Access (DS-CDMA)

Figure 1.1. Multiple-access schemes.

will be used by that user. Typically, CDMA is implemented via spread-spectrum modulation, in which the pattern is the pseudorandom code that determines the spreading sequence in the case of direct sequence, or the hopping pattern in the case of frequency hopping. In such systems, a channel is defined by a particular pseudorandom code, so each user is assigned a channel by being assigned a pseudorandom code. CDMA is used, notably, in the second-generation cellular standard IS-95 (Interim Standard 95), which makes use of direct-sequence CDMA to allocate subchannels of larger-bandwidth (1.25 MHz) subchannels of the entire cellular band. It is also used, in the form of frequency hopping, in GSM to provide isolation among users in adjacent cells. The spectrum spreading used in wireless LAN systems is also a form of CDMA in that it allows a number of such systems to operate in the same lightly regulated part of the radio spectrum. CDMA is also the basis for the principal standards being developed and deployed for 3G cellular telephony (e.g., [130, 361, 362, 407]).

Any of the multiple-access techniques discussed here can be modeled analytically by considering multiple transmitted signals of the form (1.1). In particular, for a system of K users, we can write a transmitted signal for each user as

$$x_k(t) \;=\; \sum_{i=0}^{M-1} b_k[i]\, w_{i,k}(t), \qquad k = 1, 2, \ldots, K, \tag{1.8}$$

where $x_k(\cdot)$, $\{b_k[0], b_k[1], \ldots, b_k[M-1]\}$, and $w_{i,k}(\cdot)$ represent the transmitted signal, symbol stream, and ith modulation waveform, respectively, of user k. That is, each user in a multiple-access system can be modeled in the same way as in a single-user system, but with (usually) differing modulation waveforms (and symbol streams, of course). If the waveforms $\{w_{i,k}(\cdot)\}$ are of the form (1.2) but with different carrier frequencies $\{\omega_k\}$, say, this is FDMA. If they are of the form (1.2) but with time-slotted amplitude pulses $\{p_k(\cdot)\}$, say, this is TDMA. Finally, if they are spread-spectrum signals of this form but with different pseudorandom spreading codes or hopping patterns, this is CDMA. Details of these multiple-access models will be discussed in the sequel as needed.

1.2.3   Wireless Channel

From a technical point of view, the greatest distinction between wireless and wireline communications lies in the physical properties of wireless channels. These physical properties can be described in terms of several distinct phenomena, including ambient noise, propagation losses, multipath, interference, and properties arising from the use of multiple antennas. Here we review these phenomena only briefly. Further discussion and details can be found, for example, in [38,46,148,216,405,450,458,465].

Like all practical communications channels, wireless channels are corrupted by ambient noise. This noise comes from thermal motion of electrons on the antenna and in the receiver electronics and from background radiation sources. This noise is well modeled as having a very wide bandwidth (much wider than the bandwidth of any useful signals in the channel) and no particular deterministic structure (structured noise can be treated separately as interference). A very common and useful

model for such noise is additive white Gaussian noise (AWGN), which as the name implies, means that it is additive to the other signals in the receiver, has a flat power spectral density, and induces a Gaussian probability distribution at the output of any linear filter to which it is input. Impulsive noise also occurs in some wireless channels. Such noise is similarly wideband but induces a non-Gaussian amplitude distribution at the output of linear filters. Specific models for such impulsive noise are discussed in Chapter 4.

Propagation losses are also an issue in wireless channels. These are of two basic types: diffusive losses and shadow fading. *Diffusive losses* arise because of the open nature of wireless channels. For example, the energy radiated by a simple point source in free space will spread over an ever-expanding spherical surface as the energy propagates away from the source. This means that an antenna with a given aperture size will collect an amount of energy that decreases with the square of the distance between the antenna and the source. In most terrestrial wireless channels, the diffusion losses are actually greater than this, due to the effects of ground-wave propagation, foliage, and so on. For example, in cellular telephony, the diffusion loss is inverse square with distance within line of sight of the cell tower, and it falls off with a higher power (typically, 3 or 4) at greater distances. As its name implies, *shadow fading* results from the presence of objects (buildings, walls, etc.) between the transmitter and receiver. Shadow fading is typically modeled by an attenuation (i.e., a multiplicative factor) in signal amplitude that follows a log-normal distribution. The variation in this fading is specified by the standard deviation of the logarithm of this attenuation.

Multipath refers to the phenomenon by which multiple copies of a transmitted signal are received at the receiver, due to the presence of multiple radio paths between the transmitter and receiver. These multiple paths arise due to reflections from objects in the radio channel. Multipath is manifested in several ways in communications receivers, depending on the degree of path difference relative to the wavelength of propagation, the degree of path difference relative to the signaling rate, and the relative motion between the transmitter and receiver. Multipath from scatterers that are spaced very close together will cause a random change in the amplitude of the received signal. Due to central-limit effects, the resulting received amplitude is often modeled as being a complex Gaussian random variable. This results in a random amplitude whose envelope has a Rayleigh distribution, and this phenomenon is thus termed *Rayleigh fading*. Other fading distributions also arise, depending on the physical configuration (see, e.g., [396]). When the scatterers are spaced so that the differences in their corresponding path lengths are significant relative to a wavelength of the carrier, the signals arriving at the receiver along different paths can add constructively or destructively. This gives rise to fading that depends on the wavelength (or, equivalently, the frequency) of radiation, which is thus called *frequency-selective fading*. When there is relative motion between the transmitter and receiver, this type of fading also depends on time, since the path length is a function of the radio geometry. This results in *time-selective fading*. (Such motion also causes signal distortion due to Doppler effects.) A related phenomenon arises when the difference in path lengths is such that the time delay of

arrival along different paths is significant relative to a symbol interval. This results in dispersion of the transmitted signal, and causes *intersymbol interference* (ISI); that is, contributions from multiple symbols arrive at the receiver at the same time.

Many of the advanced signal transmission and processing methods that have been developed for wireless systems are designed to contravene the effects of multipath. For example, wideband signaling techniques such as spread spectrum are often used as a countermeasure to frequency-selective fading. This both minimizes the effects of deep frequency-localized fades and facilitates the resolvability and subsequent coherent combining of multiple copies of the same signal. Similarly, by dividing a high-rate signal into many parallel lower-rate signals, OFDM mitigates the effects of channel dispersion on high-rate signals. Alternatively, high-data-rate single-carrier systems make use of channel equalization at the receiver to counteract this dispersion. Some of these issues are discussed further in Section 1.3.

Interference, also a significant issue in many wireless channels, is typically one of two types: multiple-access interference and co-channel interference. *Multiple-access interference* (MAI) refers to interference arising from other signals in the same network as the signal of interest. For example, in cellular telephony systems, MAI can arise at the base station when the signals from multiple mobile transmitters are not orthogonal to one another. This happens by design in CDMA systems, and it happens in FDMA or TDMA systems due to channel properties such as multipath or to nonideal system characteristics such as imperfect channelization filters. *Co-channel interference* (CCI) refers to interference from signals from different networks, but operating in the same frequency band as the signal of interest. An example is the interference from adjacent cells in a cellular telephony system. This problem is a chief limitation of using FDMA in cellular systems and was a major factor in moving away from FDMA in second-generation systems. Another example is the interference from other devices operating in the same part of the unregulated spectrum as the signal of interest, such as interference from Bluetooth devices operating in the same 2.4-GHz ISM band as IEEE 802.11 wireless LANs. Interference mitigation is also a major factor in the design of transmission techniques (e.g., the above-noted movement away from FDMA in cellular systems) as well as in the design of advanced signal processing systems for wireless, as we shall see in the sequel.

The phenomena we have discussed above can be incorporated into a general analytical model for a wireless multiple-access channel. In particular, the signal model in a wireless system is illustrated in Fig. 1.2. We can write the signal received at a given receiver in the following form:

$$r(t) = \sum_{k=1}^{K} \sum_{i=0}^{M-1} b_k[i] \int_{-\infty}^{\infty} g_k(t,u) w_{i,k}(u) \; \mathrm{d}u + i(t) + n(t), \qquad -\infty < t < \infty, \quad (1.9)$$

where $g_k(t,u)$ denotes the impulse response of a linear filter representing the channel between the kth transmitter and the receiver, $i(\cdot)$ represents co-channel interference, and $n(\cdot)$ represents ambient noise. The modeling of the wireless channel as a linear system seems to agree well with the observed behavior of such channels. All of the quantities $g_k(\cdot,\cdot)$, $i(\cdot)$, and $n(\cdot)$ are, in general, random processes. As noted
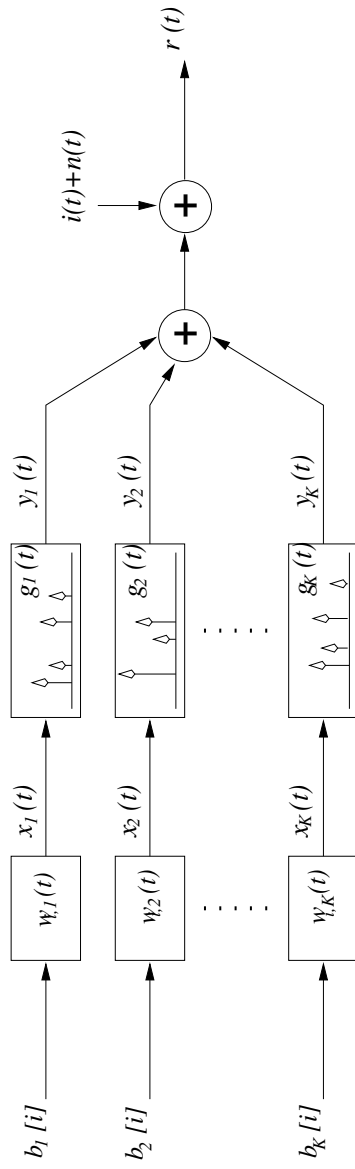
Figure 1.2. Signal model in a wireless system.

above, the ambient noise is typically represented as a white process with very little additional structure. However, the co-channel interference and channel impulse responses are typically structured processes that can be parameterized.

An important special case is that of a pure multipath channel, in which the channel impulse responses can be represented in the form

$$g_k(t, u) = \sum_{\ell=1}^{L_k} \alpha_{\ell,k} \delta(t - u - \tau_{\ell,k}), \tag{1.10}$$

where L_k is the number of paths between user k and the receiver, $\alpha_{\ell,k}$ and $\tau_{\ell,k}$ are the gain and delay, respectively, associated with the ℓth path of the kth user, and $\delta(\cdot)$ denotes the Dirac delta function. Note that this is the situation illustrated in Fig. 1.2, in which we have written the time-invariant impulse response as $g_k(t) \equiv g_k(t, 0)$. This model is an idealization of the actual behavior of a multipath channel, which would not have such a sharply defined impulse response. However, it serves as a useful model for signal processor design and analysis. Note that this model gives rise to frequency-selective fading, since the relative delays will cause constructive and destructive interference at the receiver, depending on the wavelength of propagation. Often, the delays $\{\tau_{\ell,k}\}$ are assumed to be known to the receiver or are spaced uniformly at the inverse of the bulk bandwidth of the signaling waveforms. A typical model for the path gains $\{\alpha_{\ell,k}\}$ is that they are independent complex Gaussian random variables, giving rise to Rayleigh fading.

Note that, in general, the receiver will see the following *composite* modulation waveform associated with the symbol $b_k[i]$:

$$f_{i,k}(t) = \int_{-\infty}^{\infty} g_k(t, u) w_{i,k}(u) \, \mathrm{d}u. \tag{1.11}$$

If these waveforms are not orthogonal for different values of i, ISI will result. Consider, for example, the pure multipath channel of (1.10) with signaling waveforms of the form

$$w_{i,k}(t) = A_k s_k(t - iT), \tag{1.12}$$

where $s_k(\cdot)$ is a normalized signaling waveform $\left[\int |s_k(t)|^2 \, \mathrm{d}t = 1\right]$, A_k is a complex amplitude, and T is the inverse of the single-user symbol rate. In this case, the composite modulation waveforms are given by

$$f_{i,k}(t) = f_k(t - iT), \tag{1.13}$$

with

$$f_k(t) = A_k \sum_{\ell=1}^{L_k} \alpha_{\ell,k} s_k\left(t - \tau_{\ell,k}\right). \tag{1.14}$$

If the delay spread (i.e., the maximum of the differences of the delays $\{\tau_{\ell,k}\}$ for different values of ℓ) is significant relative to T, ISI may be a factor. Note that

for a fixed channel, the delay spread is a function of the physical geometry of the channel, whereas the symbol rate depends on the data rate of the transmitted source. Thus, higher-rate transmissions are more likely to encounter ISI than are lower-rate transmissions. Similarly, if the composite waveforms for different values of k are not orthogonal, MAI will result. This can happen, for example, in CDMA channels when the pseudorandom code sequences used by different users are not orthogonal. It can also happen in CDMA and TDMA channels, due to the effects of multipath or asynchronous transmission. These issues are discussed further in the sequel as the need arises.

This model can be further generalized to account for multiple antennas at the receiver. In particular, we can modify (1.9) as follows:

$$\boldsymbol{r}(t) = \sum_{k=1}^{K} b_k[i] \int_{-\infty}^{\infty} \boldsymbol{g}_k(t, u) w_{i,k}(u) \, \mathrm{d}u \; + \; \boldsymbol{i}(t) \; + \; \boldsymbol{n}(t), \quad -\infty < t < \infty, \quad (1.15)$$

where the boldface quantities denote (column) vectors with dimensions equal to the number of antennas at the received array. For example, the pth component of $\boldsymbol{g}_k(t, u)$ is the impulse response of the channel between user k and the pth element of the receiving array. A useful such model is to combine the pure multipath model of (1.10) with a model in which the spatial aspects of the array can be separated from its temporal properties. This yields channel impulse responses of the form

$$\boldsymbol{g}_k(t, u) \; = \; \sum_{\ell=1}^{L_k} \alpha_{\ell,k} \boldsymbol{a}_{\ell,k} \delta(t - u - \tau_{\ell,k}), \qquad (1.16)$$

where the complex vector $\boldsymbol{a}_{\ell,k}$ describes the response of the array to the ℓth path of user k. The simplest such situation is the case of a *uniform linear array* (ULA), in which the array elements are uniformly spaced along a line, receiving a single-carrier signal arriving along a planar wavefront and satisfying the *narrowband array assumption*. The essence of this assumption is that the signaling waveforms are sinusoidal carriers carrying narrowband modulation and that all of the variation in the received signal across the array at any given instant in time is due to the carrier (i.e., the modulating waveform is changing slowly enough to be assumed constant across the array). In this case, the array response depends only on the angle $\phi_{\ell,k}$ at which the corresponding path's signal is incident on the array. In particular, the response of a P-element array is given in this case by

$$\boldsymbol{a}_{\ell,k} \; = \; \begin{bmatrix} 1 \\ e^{-\jmath \gamma \sin \phi_{\ell,k}} \\ e^{-\jmath 2\gamma \sin \phi_{\ell,k}} \\ \vdots \\ e^{-\jmath (P-1)\gamma \sin \phi_{\ell,k}} \end{bmatrix}, \qquad (1.17)$$

where $\jmath$ denotes the imaginary unit and where $\gamma \triangleq 2\pi d/\lambda$, with λ the carrier wavelength and d the interelement spacing (see [126, 266, 269, 404, 445, 450, 510] for further discussion of systems involving multiple receiver antennas).

It is also of interest to model systems in which there are multiple antennas at both the transmitter and receiver, called *multiple-input/multiple-output* (MIMO) *systems*. In this case the channel transfer functions are matrices, with the number of rows equal to the number of receiving antennas and the number of columns equal to the number of transmitting antennas at each source. There are several ways of handling the signaling in such configurations, depending on the desired effects and the channel conditions. For example, transmitter beamforming can be implemented by transmitting the same symbol simultaneously from multiple antenna elements on appropriately phased versions of the same signaling waveform. Space-time coding can be implemented by transmitting frames of related symbols over multiple antennas. Other configurations are of interest as well. Issues concerning multiple-antenna systems are discussed further in the sequel as they arise.

1.3 Basic Receiver Signal Processing for Wireless

This book is concerned with the design of advanced signal processing methods for wireless receivers, based largely on the models discussed in preceding sections. Before moving to these methods, however, it is of interest to review briefly some basic elements of signal processing for these models. This is not intended to be a comprehensive treatment, and the reader is referred to [145, 146, 270, 376, 381, 385, 391, 396, 510, 520, 523] for further details.

1.3.1 Matched Filter/RAKE Receiver

We consider first the particular case of the model of (1.9), in which there is only a single user (i.e., $K = 1$), the channel impulse $g_1(\cdot, \cdot)$ is known to the receiver, there is no CCI [i.e., $i(\cdot) \equiv 0$], and the ambient noise is AWGN with spectral height σ^2. That is, we have the following model for the received signal:

$$r(t) \;=\; \sum_{i=0}^{M-1} b_1[i] f_{i,1}(t) + n(t), \qquad -\infty < t < \infty, \tag{1.18}$$

where $f_{i,1}(\cdot)$ denotes the composite waveform of (1.11), given by

$$f_{i,1}(t) \;=\; \int_{-\infty}^{\infty} g_1(t, u) w_{i,1}(u) \; \mathrm{d}u. \tag{1.19}$$

Let us further restrict attention, for the moment, to the case in which there is only a single symbol to be transmitted (i.e., $M = 1$), in which case we have the received waveform

$$r(t) \;=\; b_1[0] f_{0,1}(t) + n(t), \qquad -\infty < t < \infty. \tag{1.20}$$

Optimal inferences about the symbol $b_1[0]$ in (1.20) can be made on the basis of the likelihood function of the observations, conditioned on the symbol $b_1[0]$, which is given in this case by [377]

$$
\begin{aligned}
\mathcal{L}\Big(r(\cdot)|b_1[0]\Big) = \\
\exp\left\{\frac{1}{\sigma^2}\left[2\Re\left\{b_1^*[0]\int_{-\infty}^{\infty}f_{0,1}^*(t)r(t)\ \mathrm{d}t\right\} - |b_1[0]|^2\int_{-\infty}^{\infty}|f_{0,1}(t)|^2\ \mathrm{d}t\right]\right\}, \quad (1.21)
\end{aligned}
$$

where the superscript asterisk denotes complex conjugation and $\Re\{\cdot\}$ denotes the real part of its argument.

Optimal inferences about the symbol $b_1[0]$ can be made, for example, by choosing *maximum-likelihood* (ML) or *maximum a posteriori probability* (MAP) values for the symbol. The ML symbol decision is given simply by the argument that maximizes $\mathcal{L}(\ r(\cdot)\ |\ b_1[0]\ )$ over the symbol alphabet, $\mathcal{A}$:

$$
\begin{aligned}
\hat{b}_1[0] &= \arg\left\{\max_{b\in\mathcal{A}}\mathcal{L}\Big(\ r(\cdot)\ |\ b_1[0] = b\ \Big)\right\} \\
&= \arg\left\{\max_{b\in\mathcal{A}}\left[2\Re\left\{b^*\int_{-\infty}^{\infty}f_{0,1}^*(t)r(t)\ \mathrm{d}t\right\} - |b|^2\int_{-\infty}^{\infty}|f_{0,1}(t)|^2\ \mathrm{d}t\right]\right\}. \quad (1.22)
\end{aligned}
$$

It is easy to see that the corresponding symbol estimate is the solution to the problem

$$
\min_{b\in\mathcal{A}}\ |b - z|^2, \quad (1.23)
$$

where

$$
z \triangleq \frac{\int_{-\infty}^{\infty}f_{0,1}^*(t)r(t)\ \mathrm{d}t}{\int_{-\infty}^{\infty}|f_{0,1}(t)|^2\ \mathrm{d}t}. \quad (1.24)
$$

Thus, the ML symbol estimate is the closest point in the symbol alphabet to the observable z.

Note that the two simplest and most common choices of symbol alphabet are M-ary phase-shift keying (MPSK) and quadrature amplitude modulation (QAM). In MPSK, the symbol alphabet is

$$
\mathcal{A} = \left\{e^{j2\pi m/M}\ |\ m \in \{0, 1, \dots, M-1\}\right\}, \quad (1.25)
$$

or some rotation of this set around the unit circle. (M as used in this paragraph should not be confused with the framelength M.) For QAM, a symbol alphabet containing $M \times N$ values is

$$
\mathcal{A} = \left\{b_R + jb_I\ |\ b_R \in \mathcal{A}_R\ \text{and}\ b_I \in \mathcal{A}_I\right\}, \quad (1.26)
$$

where $\mathcal{A}_R$ and $\mathcal{A}_I$ are discrete sets of amplitudes containing M and N points, respectively; for example, for $M = N$ even, a common choice is

$$\mathcal{A}_R \;=\; \mathcal{A}_I \;=\; \left\{ \pm\frac{1}{2}, \pm\frac{3}{2}, \dots, \pm\frac{M}{4} \right\} \tag{1.27}$$

or a scaled version of this choice. A special case of both of these is that of binary phase-shift keying (BPSK), in which $\mathcal{A} = \{-1, +1\}$. The latter case is the one we consider most often in this treatment, primarily for the sake of simplicity. However, most of the results discussed herein extend straightforwardly to these more general signaling alphabets.

ML symbol estimation [i.e., the solution to (1.23)] is very simple for MPSK and QAM. In particular, since the MPSK symbols correspond to phasors at evenly spaced angles around the unit circle, the ML symbol choice is that whose angle is closest to the angle of the complex number z of (1.24). For QAM, the choices of the real and imaginary parts of the ML symbol estimate are decoupled, with $\Re\{b\}$ being chosen to be the closest element of $\mathcal{A}_R$ to $\Re\{z\}$, and similarly for $\Im\{b\}$. For BPSK, the ML symbol estimate is

$$\hat{b}_i[0] \;=\; \text{sign}\left\{ \Re\{z\} \right\} \;\triangleq\; \text{sign}\left\{ \Re\left\{ \int_{-\infty}^{\infty} f_{0,1}^*(t) r(t) \; dt \right\} \right\}, \tag{1.28}$$

where $\text{sign}\{\cdot\}$ denotes the signum function:

$$\text{sign}\{x\} \;=\; \left\{ \begin{array}{ll} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ +1 & \text{if } x > 0. \end{array} \right. \tag{1.29}$$

MAP symbol detection in (1.20) is also based on the likelihood function of (1.21), after suitable transformation. In particular, if the symbol $b_1[0]$ is a random variable, taking values in $\mathcal{A}$ with known probabilities, the *a posteriori* probability distribution of the symbol conditioned on $r(\cdot)$ is given via Bayes' formula as

$$P\Big(b_1[0] = b \mid r(\cdot) \Big) \;=\; \frac{\mathcal{L}\left(\, r(\cdot) \mid b_1[0] = b \, \right) P(b_1[0] = b)}{\sum_{a \in \mathcal{A}} \mathcal{L}\left(\, r(\cdot) \mid b_1[0] = a \, \right) P(b_1[0] = a)}, \qquad b \in \mathcal{A}. \tag{1.30}$$

The MAP criterion specifies a symbol decision given by

$$\begin{aligned} \hat{b}_1[0] \;&=\; \arg\left\{ \max_{b \in \mathcal{A}} P\left(b_1[0] = b \mid r(\cdot) \right) \right\} \\ &=\; \arg\left\{ \max_{b \in \mathcal{A}} \left[\mathcal{L}\left(\, r(\cdot) \mid b_1[0] = b \, \right) P(b_1[0] = b) \right] \right\}. \end{aligned} \tag{1.31}$$

Note that in this single-symbol case, if the symbol values are equiprobable, the ML and MAP decisions are the same.

The structure of the ML and MAP decision rules above shows that the main receiver signal processing task in this single-user, single-symbol, known-channel case is the computation of the term

$$y_1[0] \triangleq \int_{-\infty}^{\infty} f_{0,1}^*(t)r(t) \ \mathrm{d}t. \tag{1.32}$$

This structure is called a *correlator* because it correlates the received signal $r(\cdot)$ with the known composite signaling waveform $f_{1,0}(\cdot)$. This structure can also be implemented by sampling the output of a time-invariant linear filter:

$$\int_{-\infty}^{\infty} f_{0,1}^*(t)r(t) \ \mathrm{d}t \ = \ (h \star r)(0), \tag{1.33}$$

where $\star$ denotes convolution and h is the impulse response of the time-invariant linear filter given by

$$h(t) = f_{0,1}^*(-t). \tag{1.34}$$

This structure is called a *matched filter*, since its impulse response is matched to the composite waveform on which the symbol is received. When the composite signaling waveform has a finite duration so that $h(t) = 0$ for $t < -D \leq 0$, the matched-filter receiver can be implemented by sampling at time D the output of the causal filter with the following impulse response:

$$h_D(t) \ = \ \begin{cases} f_{0,1}^*(D-t) & \text{if } t \geq 0, \\ 0 & \text{if } t < 0. \end{cases} \tag{1.35}$$

For example, if the signaling waveform $s_{0,1}(t)$ has duration $[0,T]$ and the channel has delay spread τ_d, the composite signaling waveform will have this property with $D = T + \tau_d$.

A special case of the correlator (1.32) arises for a pure multipath channel in which the channel impulse response is given by (1.10). The composite waveform (1.11) in this case is

$$f_{0,1}(t) \ = \ \sum_{\ell=1}^{L_1} \alpha_{\ell,1} s_{0,1}(t - \tau_{\ell,1}), \tag{1.36}$$

and the correlator output (1.32) becomes

$$y_1[0] \ \triangleq \ \sum_{\ell=1}^{L_1} \alpha_{\ell,1}^* \int_{-\infty}^{\infty} s_{0,1}^*(t - \tau_{\ell,1})r(t) \ \mathrm{d}t, \tag{1.37}$$

a configuration known as a RAKE receiver. Further details on this basic receiver structure can be found, for example, in [396].

1.3.2   Equalization

We now turn to the situation in which there is more than one symbol in the frame of interest (i.e., when $M > 1$). In this case we would like to consider the likelihood function of the observations $r(\cdot)$ conditioned on the entire frame of symbols, $b_1[0], b_1[1], \ldots, b_1[M-1]$, which is given by

$$\mathcal{L}\Big(r(\cdot)|b_1[0], b_1[1], \ldots, b_1[M-1]\Big) = \exp\left\{\frac{1}{\sigma^2}\left[2\Re\left\{\boldsymbol{b}_1^H \boldsymbol{y}_1\right\} - \boldsymbol{b}_1^H \boldsymbol{H}_1 \boldsymbol{b}_1\right]\right\}, \quad (1.38)$$

where the superscript H denotes the conjugate transpose (i.e., the Hermitian transpose), $\boldsymbol{b}_1$ denotes a column vector whose ith component is $b_1[i], i = 0, 1, \ldots, M-1$, $\boldsymbol{y}_1$ denotes a column vector whose ith component is given by

$$y_1[i] \triangleq \int_{-\infty}^{\infty} f_{i,1}^*(t) r(t) \, \mathrm{d}t, \qquad i = 0, 1, \ldots, M-1, \qquad (1.39)$$

and $\boldsymbol{H}_1$ is an $M \times M$ Hermitian matrix, whose (i, j)th element is the cross-correlation between $f_{i,1}(t)$ and $f_{j,1}(t)$:

$$\boldsymbol{H}_1[i, j] = \int_{-\infty}^{\infty} f_{i,1}^*(t) f_{j,1}(t) \, \mathrm{d}t. \qquad (1.40)$$

Since the likelihood function depends on $r(\cdot)$ only through the vector $\boldsymbol{y}_1$ of correlator outputs, this vector is a sufficient statistic for making inferences about the vector $\boldsymbol{b}_1$ of symbols [377].

Maximum-likelihood detection in this situation is given by

$$\hat{\boldsymbol{b}}_1 = \arg\left\{\max_{\boldsymbol{b} \in \mathcal{A}^M} \left[2\Re\left\{\boldsymbol{b}^H \boldsymbol{y}_1\right\} - \boldsymbol{b}^H \boldsymbol{H}_1 \boldsymbol{b}\right]\right\}. \qquad (1.41)$$

Note that if $\boldsymbol{H}_1$ is a diagonal matrix (i.e., all of its off-diagonal elements are zero), (1.41) decouples into a set of M independent problems of the single-symbol type (1.22). The solution in this case is correspondingly given by

$$\hat{b}_1[i] = \arg\min_{b \in \mathcal{A}} |b - z_1[i]|^2, \qquad (1.42)$$

where

$$z_1[i] \triangleq \frac{y_i[i]}{\int_{-\infty}^{\infty} |f_{i,1}(t)|^2 \, \mathrm{d}t}. \qquad (1.43)$$

However, in the more general case in which there is intersymbol interference, (1.41) will not decouple and the optimization must take place over the entire frame, a problem known as *sequence detection.*

The problem of (1.41) is an integer quadratic program which is known to be an NP-complete combinatorial optimization problem [380]. This implies that the complexity of (1.41) is potentially quite high: exponential in the frame length M,

which is essentially the complexity order of exhausting over the sequence alphabet $\mathcal{A}^M$. This is a prohibitive degree of complexity for most applications, since a typical frame length might be hundreds or even thousands of symbols. Fortunately, this complexity can be mitigated substantially for practical ISI channels. In particular, if the composite signaling waveforms have finite duration D, the matrix $\boldsymbol{H}_1$ is a banded matrix with nonzero elements only on those diagonals that are no more than $\Delta = \lceil D/T \rceil$ diagonals away from the main diagonal (here $\lceil \cdot \rceil$ denotes the smallest integer not less than its argument); that is,

$$|\boldsymbol{H}_1[i,j]| \;=\; 0, \quad \forall\, |i-j| > \Delta. \tag{1.44}$$

This structure of the matrix permits solution of (1.41) with a dynamic program of complexity order $\mathcal{O}\left(|\mathcal{A}|^\Delta\right)$, as opposed to the $\mathcal{O}\left(|\mathcal{A}|^M\right)$ complexity of direct search. In most situations $\Delta \ll M$, which implies an enormous savings in complexity (see, e.g., [380]). This dynamic programming solution, which can be structured in various ways, is known as a *maximum-likelihood sequence detector* (MLSD).

MAP detection in this model is also potentially of very high complexity. The *a posteriori* probability distribution of a particular symbol, say $b_1[i]$, is given by

$$P\Big(b_1[i] = b|r(\cdot)\Big) = \frac{\sum_{\{\boldsymbol{a} \in \mathcal{A}^M | a_i = b\}} \mathcal{L}\left(\, r(\cdot)|\boldsymbol{b}_1 = \boldsymbol{a}\right) P(\boldsymbol{b}_1 = \boldsymbol{a})}{\sum_{\{\boldsymbol{a} \in \mathcal{A}^M\}} \mathcal{L}\left(\, r(\cdot)|\boldsymbol{b}_1 = \boldsymbol{a}\right) P(\boldsymbol{b}_1 = \boldsymbol{a})}, \qquad b \in \mathcal{A}. \tag{1.45}$$

Note that these summations have $\mathcal{O}\left(|\mathcal{A}|^M\right)$ terms and thus are of complexity similar to those of the maximization in (1.41) in general. Fortunately, like (1.41), when $\boldsymbol{H}_1$ is banded these summations can be computed much more efficiently using a generalized dynamic programming technique that results in $\mathcal{O}\left(|\mathcal{A}|^\Delta\right)$ complexity (see, e.g., [380]).

The dynamic programs that facilitate (1.41) and (1.45) are of much lower complexity than brute-force computations. However, even this lower complexity is too high for many applications. A number of lower-complexity algorithms have been devised to deal with such situations. These techniques can be discussed easily by examining the sufficient statistic vector $\boldsymbol{y}_1$ of (1.39), which can be written as

$$\boldsymbol{y}_1 \;=\; \boldsymbol{H}_1 \boldsymbol{b}_1 \;+\; \boldsymbol{n}_1, \tag{1.46}$$

where $\boldsymbol{n}_1$ is a complex Gaussian random vector with independent real and imaginary parts having identical $\mathcal{N}(\boldsymbol{0}, \frac{\sigma^2}{2}\boldsymbol{H}_1)$ distributions. Equation (1.46) describes a linear model, and the goal of equalization is thus to fit this model with the data vector $\boldsymbol{b}_1$. The ML and MAP detectors are two ways of doing this fitting, each of which has exponential complexity with exponent equal to the bandwidth of $\boldsymbol{H}_1$. The essential difficulty of this problem arises from the fact that the vector $\boldsymbol{b}_1$ takes on values from a discrete set. One way of easing this difficulty is first to fit the linear model without constraining $\boldsymbol{b}_1$ to be discrete, and then to quantize the resulting (continuous) estimate of $\boldsymbol{b}_1$ into symbol estimates. In particular, we can use a linear fit, $\boldsymbol{M}\boldsymbol{y}_1$, as a continuous estimate of $\boldsymbol{b}_1$, where $\boldsymbol{M}$ is an $M \times M$ matrix. In this way, the ith symbol decision is

$$\hat{b}_1[i] \;=\; q\left([\boldsymbol{M}\boldsymbol{y}_1]_i\right), \tag{1.47}$$

where $[\boldsymbol{M y}_1]_i$ denotes the ith component of $\boldsymbol{M y}_1$ and where $q(\cdot)$ denotes a quantizer mapping the complex numbers to the symbol alphabet $\mathcal{A}$. Various choices of the matrix $\boldsymbol{M}$ lead to different linear equalizers. For example, if we choose $\boldsymbol{M} = \boldsymbol{I}_M$, the $M \times M$ identity matrix, the resulting linear detector is the common matched filter, which is optimal in the absence of ISI. A difficulty with the matched filter is that it ignores the ISI. Alternatively, if $\boldsymbol{H}_1$ is invertible, the choice $\boldsymbol{M} = \boldsymbol{H}_1^{-1}$ forces the ISI to zero,

$$\boldsymbol{H}_1^{-1}\boldsymbol{y}_1 \;=\; \boldsymbol{b}_1 \;+\; \boldsymbol{H}_1^{-1}\boldsymbol{n}_1, \tag{1.48}$$

and is thus known as the *zero-forcing equalizer* (ZFE). Note that this would be optimal (i.e., it would give perfect decisions) in the absence of AWGN. A difficulty with the ZFE is that it can significantly enhance the effects of AWGN by placing high gains on some directions in the set of M-dimensional complex vectors. A trade-off between these extremes is effected by the *minimum-mean-square-error* (MMSE) *linear equalizer*, which chooses $\boldsymbol{M}$ to give an MMSE fit of the model (1.46). Assuming that the symbols are independent of the noise, this results in the choice

$$\boldsymbol{M} \;=\; (\boldsymbol{H}_1 \;+\; \sigma^2\boldsymbol{\Sigma}_b^{-1})^{-1}, \tag{1.49}$$

where $\boldsymbol{\Sigma}_b$ denotes the covariance matrix of the symbol vector $\boldsymbol{b}_1$. (Typically, this covariance matrix will be in the form of a constant times $\boldsymbol{I}_M$.) A number of other techniques for fitting the model (1.46) have been developed, including iterative methods with and without quantization of intermediate results [decision-feedback equalizers (DFEs)], and so on. For a more detailed treatment of equalization methods, see [396].

1.3.3   Multiuser Detection

To finish this section we turn finally to the full multiple-access model of (1.9), within which data detection is referred to as *multiuser detection.* This situation is very similar to the ISI channel described above. In particular, we now consider the likelihood function of the observations $r(\cdot)$ conditioned on all symbols of all users. Sorting these symbols first by symbol number and then by user number, we can collect them in a column vector $\boldsymbol{b}$ given as

$$\boldsymbol{b} \;=\; \begin{bmatrix} b_1[0] \\ b_2[0] \\ \vdots \\ b_K[0] \\ \vdots \\ b_1[M-1] \\ b_2[M-1] \\ \vdots \\ b_K[M-1] \end{bmatrix}, \tag{1.50}$$

so that the nth element of $\boldsymbol{b}$ is given by

$$[\boldsymbol{b}]_n = b_k[i] \ \ \text{with} \ \ k \triangleq [n-1]_K \ \ \text{and} \ \ i \triangleq \left\lfloor \frac{n-1}{K} \right\rfloor, \qquad n = 1, 2, \dots, KM, \quad (1.51)$$

where $[\cdot]_K$ denotes reduction of the argument modulo K and $\lfloor \cdot \rfloor$ denotes the integer part of the argument. Analogously with (1.38) we can write the corresponding likelihood function as

$$\mathcal{L}\left(\, r(\cdot) \mid \boldsymbol{b} \, \right) \ = \ \exp\left\{ \frac{1}{\sigma^2} \left[2\Re\left\{ \boldsymbol{b}^H \boldsymbol{y} \right\} \ - \ \boldsymbol{b}^H \boldsymbol{H} \boldsymbol{b} \right] \right\}, \qquad (1.52)$$

where $\boldsymbol{y}$ is a column vector that collects the set of observables

$$y_k[i] \ \triangleq \ \int_{-\infty}^{\infty} f_{i,k}^*(t) r(t) \ \mathrm{d}t, \qquad i = 0, 1, \dots, M-1, \quad k = 1, 2, \dots, K, \quad (1.53)$$

indexed conformally with $\boldsymbol{b}$, and where $\boldsymbol{H}$ denotes the $KM \times KM$ Hermitian cross-correlation matrix of the composite waveforms associated with the symbols in $\boldsymbol{b}$, again with conformal indexing:

$$\boldsymbol{H}[n, m] \ = \ \int_{-\infty}^{\infty} f_{i,k}^*(t) f_{j,\ell}(t) \ \mathrm{d}t, \qquad (1.54)$$

with

$$k \triangleq [n-1]_K, \quad i \triangleq \left\lfloor \frac{n-1}{K} \right\rfloor, \quad \ell \triangleq [m-1]_K, \quad \text{and} \quad j \triangleq \left\lfloor \frac{m-1}{K} \right\rfloor. \qquad (1.55)$$

Comparing (1.52), (1.53), and (1.54) with their single-user counterparts (1.38), (1.39), and (1.40), we see that $\boldsymbol{y}$ is a sufficient statistic for making inferences about $\boldsymbol{b}$, and moreover that such inferences can be made in a manner very similar to that for the single-user ISI channel. The principal difference is one of dimensionality: Decisions in the single-user ISI channel involve simultaneous sequence detection with M symbols, whereas decisions in the multiple-access channel involve simultaneous sequence detection with KM symbols. This, of course, can increase the complexity considerably. For example, the complexity of exhaustive search in ML detection, or exhaustive summation in MAP detection, is now on the order of $|\mathcal{A}|^{MK}$. However, as in the single-user case, this complexity can be mitigated considerably if the delay spread of the channel is small. In particular, if the duration of the composite signaling waveforms is D, the matrix $\boldsymbol{H}$ will be a banded matrix with

$$\boldsymbol{H}[m, n] \ = \ 0, \qquad \forall \ |n - m| > K\Delta, \qquad (1.56)$$

where, as before, $\Delta = \lceil D/T \rceil$. This bandedness allows the complexity of both ML and MAP detection to be reduced to the order of $|\mathcal{A}|^{K\Delta}$ via dynamic programming.

Although further complexity reduction can be obtained in this problem within additional structural constraints on $\boldsymbol{H}$ (see, e.g., [380]), the $\mathcal{O}\left(|\mathcal{A}|^{K\Delta} \right)$ complexity of ML and MAP multiuser detection is not generally reducible. Consequently, as

with the equalization of single-user channels, a number of lower-complexity sub-optimal multiuser detectors have been developed. For example, analogously with (1.47), linear multiuser detectors can be written in the form

$$\hat{b}_k[i] \;=\; q\left([\boldsymbol{M}\boldsymbol{y}]_n\right), \quad \text{with} \;\; k \triangleq [n-1]_K \quad \text{and} \quad i \triangleq \left\lfloor \frac{n-1}{K} \right\rfloor, \qquad (1.57)$$

where $\boldsymbol{M}$ is a $KM \times KM$ matrix, $[\boldsymbol{M}\boldsymbol{y}]_n$ denotes the nth component of $\boldsymbol{M}\boldsymbol{y}$, and where, as before, $q(\cdot)$ denotes a quantizer mapping the complex numbers to the symbol alphabet $\mathcal{A}$. The choice $\boldsymbol{M} = \boldsymbol{H}^{-1}$ forces both MAI and ISI to zero and is known as the decorrelating detector, or *decorrelator*. Similarly, the choice

$$\boldsymbol{M} \;=\; (\boldsymbol{H} \;+\; \sigma^2 \boldsymbol{\Sigma}_b^{-1})^{-1}, \qquad (1.58)$$

where $\boldsymbol{\Sigma}_b$ denotes the covariance matrix of the symbol vector $\boldsymbol{b}$, is known as the *linear MMSE multiuser detector*. Linear and nonlinear iterative versions of these detectors have also been developed, both to avoid the complexity of inverting $KM \times KM$ matrices and to exploit the finite-alphabet property of the symbols (see, e.g., [520]).

   As a final issue here we note that all of the discussion above has involved direct processing of continuous-time observations to obtain a sufficient statistic (in practice, this corresponds to hardware front-end processing), followed by algorithmic processing to obtain symbol decisions. Increasingly, an intermediate step is of interest. In particular, it is often of interest to project continuous-time observations onto a large but finite set of orthonormal functions to obtain a set of observables. These observables can then be processed further using digital signal processing (DSP) to determine symbol decisions (perhaps with intermediate calculation of the sufficient statistic), which is the principal advantage of this approach. A tacit assumption in this process is that the orthonormal set spans all of the composite signaling waveforms of interest, although this will often be only an approximation. A prime example of this kind of processing arises in direct-sequence spread-spectrum systems [see (1.6)], in which the received signal can be passed through a filter matched to the chip waveform and then sampled at the chip rate to produce N samples per symbol interval. These N samples can then be combined in various ways (usually, linearly) for data detection. In this way, for example, the linear equalizer and multiuser detectors discussed above are particularly simple to implement. A significant advantage of this approach is that this combining can often be done adaptively when some aspects of the signaling waveforms are unknown. For example, the channel impulse response may be unknown to the receiver, as may the waveforms of some interfering signals. This kind of processing is a basic element of many of the results discussed in this book and will be revisited in more detail in Chapter 2.

1.4   Outline of the Book

In Section 1.3 we described the basic principles of signal reception for wireless systems. The purpose of this book is to delve into advanced methods for this problem

in the contexts of the signaling environments that are of most interest in emerging wireless applications. The scope of the treatment includes advanced receiver techniques for key signaling environments, including multiple-access, MIMO, and OFDM systems, as well as methods that address unique physical issues arising in many wireless channels, including fading, impulsive noise, co-channel interference, and other channel impairments. This material is organized into nine chapters beyond the current chapter. The first five of these deal explicitly with multiuser detection (i.e., with the mitigation of multiple-access interference) combined with other channel features or impairments. The remaining four chapters deal with the treatment of systems involving narrowband co-channel interference, time-selective fading, or multiple carriers, and with a general technique for receiver signal processing based on Monte Carlo Bayesian techniques. These contributions are outlined briefly in the paragraphs below.

Chapter 2 is concerned with the basic problem of adaptive multiuser detection in channels whose principal impairments (aside from multiple-access interference) are additive white Gaussian noise and multipath distortion. Adaptivity is a critical issue in wireless systems because of the dynamic nature of wireless channels. Such dynamism arises from several sources, notably from mobility of the transmitter or receiver and from the fact that the user population of the channel changes due to the entrance and exit of users and interferers from the channels and due to the bursty nature of many information sources. This chapter deals primarily with *blind multiuser detection*, in which the receiver is faced with the problem of demodulating a particular user in a multiple-access system, using knowledge only of the signaling waveform (either the composite receiver waveform or the transmitted waveform) of that user. The "blind" qualifier means that the receiver algorithms to be described are to be adapted without knowledge of the transmitted symbol stream. In this chapter we introduce the basic methods for blind adaptation of the linear multiuser detectors discussed in Section 1.3 via traditional adaptation methods, including least-mean-squares (LMS), recursive least-squares (RLS), and subspace tracking. The combination of multiuser detection with estimation of the channel intervening the desired transmitter and receiver is also treated in this context, as is the issue of correlated noise.

The methods of Chapter 2 are of particular interest in downlink situations (e.g., base to mobile), in which the receiver is interested in the demodulation of only a single user in the system. Another scenario is that the receiver has knowledge of the signaling waveforms used by a group of transmitters and wishes to demodulate this entire group while suppressing the effects of other interfering transmitters. An example of a situation in which this type of problem occurs is the reverse, or mobile-to-base, link in a CDMA cellular telephony system, in which a given base station wishes to demodulate the users in its cell while suppressing interference from users in adjacent cells. Chapter 3 continues with the issue of blind multiuser detection, but in this more general context of group detection. Here, both linear and nonlinear methods are considered, and again the issues of multipath and correlated noise are examined.

Channels in which the ambient noise is assumed to be Gaussian are considered in Chapters 2 and 3. Of course, this assumption of Gaussian noise is a very common one in the design and analysis of communication systems, and there are often good reasons for this assumption, including tractability and a degree of physical reality stemming from phenomena such as thermal noise. However, many practical channels involve noise that is decidedly not Gaussian. This is particularly true in urban and indoor environments, in which there is considerable impulsive noise due to human-made ambient phenomena. Also, in underwater acoustic channels (which are not specifically addressed in this book but which are used for tetherless communications) the ambient noise tends to be non-Gaussian. In systems limited by multiple-access interference, the assumption of Gaussian noise is a reasonable one, since it allows the focus to be placed on the main source of error—multiple-access interference. However, as we shall see in Chapters 2 and 3, the use of multiuser detection can return such channels to channels limited by ambient noise. Thus, the structure of ambient noise is again important, particularly since the performance and design of receiver algorithms can be affected considerably by the shape of the noise distribution even when the noise energy is held constant. In Chapter 4 we consider the problem of adaptive multiuser detection in channels with non-Gaussian ambient noise. This problem is a particularly challenging one because traditional methods for mitigating non-Gaussian noise involve nonlinear front-end processing, whereas methods for mitigating MAI tend to rely on the linear separating properties of the signaling multiplex. Thus, the challenge for non-Gaussian multiple-access channels is to combine these two methodologies without destroying the advantages of either. A powerful approach to this problem based on nonlinear regression is described in Chapter 4. In addition to the design and analysis of basic algorithms for known signaling environments, blind and group-blind methods are also discussed. It is seen that these methods lead to methods for multiuser detection in non-Gaussian environments that perform much better than linear methods in terms of both absolute performance and robustness.

In Chapter 5 we introduce the issue of multiple antennas into the receiver design problem. In particular, we consider the design of optimal and adaptive multiuser detectors for MIMO systems. Here, for known channel and antenna characteristics, the basic sufficient statistic [analogous to (1.53)] is a space-time matched-filter bank, which forms a generic front end for a variety of space-time multiuser detection methods. For adaptive systems, a significant issue that arises beyond those in the single-antenna situation is lack of knowledge of the response of the receiving antenna array. This can be handled through a novel adaptive MMSE multiuser detector described in this chapter. Again, as in the scalar case, the issues of multipath and blind channel identification are considered as well.

In Chapter 6 we treat the problem of signal reception in channel-coded multiple-access systems. In particular, the problem of joint channel decoding and multiuser detection is considered. A turbo-style iterative technique is presented that mitigates the high complexity of optimal processing in this situation. The essential idea of this turbo multiuser detector is to consider the combination of channel coding followed by a multiple-access channel as a concatenated code, which can

be decoded by iterating between the constituent decoders—the multiuser detector for the multiple-access channel and a conventional channel decoder for the channel codes—exchanging soft information between each iteration. The constituent algorithms must be soft-input/soft-output (SISO) algorithms, which implies MAP multiuser detection and decoding. In the case of convolutional channel codes, the MAP decoder can be implemented using the well-known Bahl, Cocke, Jelinek, and Raviv (BCJR) algorithm. However, the MAP multiuser detector is quite complex, and thus a SISO MMSE detector is developed to lessen this complexity. A number of issues are treated in this context, including a group-blind implementation to suppress interferers, multipath, and space-time coded systems.

In Chapter 7 we turn to the issue of narrowband interference suppression in spread-spectrum systems. This problem arises for many reasons. For example, in multimedia transmission, signals with different data rates make use of the same radio resources, giving rise to signals of different bandwidths in the same spectrum. Also, some emerging services are being placed in parts of the radio spectrum which are already occupied by existing narrowband legacy systems. Many other systems operate in license-free parts of the spectrum, where signals of all types can share the same spectrum. Similarly, in tactical military systems, jamming gives rise to narrowband interference. The use of spread-spectrum modulation in these types of situations creates a degree of natural immunity to narrowband interference. However, active methods for interference suppression can yield significant performance improvements over systems that rely simply on this natural immunity. This problem is an old one, dating to the 1970s. Here we review the development of this field, which has progressed from methods that exploit only the bandwidth discrepancies between spread and narrowband signals, to more powerful "code-aided" techniques that make use of ideas similar to those used in multiuser detection. We consider several types of narrowband interference, including tonal signals and narrowband digital communication signals, and in all cases it is seen that active methods can offer significant performance gains with relatively small increases in complexity.

Chapter 8 is concerned with the problem of Monte Carlo Bayesian signal processing and its applications in developing adaptive receiver algorithms for tasks such as multiuser detection, equalization, and related tasks. Monte Carlo Bayesian methods have emerged in statistics over the past few years. When adapted to signal processing tasks, they give rise to powerful low-complexity adaptive algorithms whose performance approaches theoretical optima for fast and reliable communications in the dynamic environments in which wireless systems must operate. The chapter begins with a review of the large body of methodology in this area that has been developed over the past decade. It then continues to develop these ideas as signal processing tools, both for batch processing using Markov chain Monte Carlo (MCMC) methods and for online processing using sequential Monte Carlo (SMC) methods. These methods are particularly well suited to problems involving unknown channel conditions, and the power of these techniques is illustrated in the contexts of blind multiuser detection in unknown channels and blind equalization of MIMO channels.

Although most of the methodology discussed in the preceding paragraphs can deal with fading channels, the focus of those methods has been on quasi-static channels in which the fading characteristics of the channel can be assumed to be constant over an entire processing window, such as a data frame. This allows representation of the fading with a set of parameters that can be well estimated by the receiver. An alternative situation arises when the channel fading is fast enough that it can change at a rate comparable to the signaling rate. For such channels, new techniques must be developed in order to mitigate the fast fading, either by tracking it simultaneously with data demodulation or by using modulation techniques that are impervious to fast fading. Chapter 9 is concerned with problems of this type. In particular, after an overview of the physical and mathematical modeling of fading processes, several basic methods for dealing with fast-fading channels are considered. In particular, these methods include application of the expectation-maximization (EM) algorithm and its sequential counterpart, decision-feedback differential detectors for scalar and space-time-coded systems, and sequential Monte Carlo methods for both coded and uncoded systems.

Finally, in Chapter 10, we turn to problems of advanced receiver signal processing for coded OFDM systems. As noted previously, OFDM is becoming the technique of choice for many high-data-rate wireless applications. Recall that OFDM systems are multicarrier systems in which the carriers are spaced as closely as possible while maintaining orthogonality, thereby efficiently using available spectrum. This technique is very useful in frequency-selective channels, since it allows a single high-rate data stream to be converted into a group of many low-rate data streams, each of which can be transmitted without intersymbol interference. The chapter begins with a review of OFDM systems and then considers receiver design for OFDM signaling through unknown frequency-selective channels. In particular, the treatment focuses on turbo receivers in several types of OFDM systems, including systems with frequency offset, a space-time block coded OFDM system, and a space-time coded OFDM system using low-density parity-check (LDPC) codes.

Taken together, the techniques described in these chapters provide a unified methodology for the design of advanced receiver algorithms to deal with the impairments and diversity opportunities associated with wireless channels. Although most of these algorithms represent very recent research contributions, they have generally been developed with an eye toward low complexity and ease of implementation. Thus, it is anticipated that they can be applied readily in the development of practical systems. Moreover, the methodology described herein is sufficiently general that it can be adapted as needed to other problems of receiver signal processing. This is particularly true of the Monte Carlo Bayesian methods described in Chapter 8, which provide a very general toolbox for designing low-complexity yet sophisticated adaptive signal processing algorithms.

Note to the Reader Each chapter of this book describes a number of advanced receiver algorithms. For convenience, the introduction to each chapter contains a list of the algorithms developed in that chapter. Also, the references cited for all chapters are listed near the end of the book. This set of references comprises an extensive, although not exhaustive, bibliography of the literature in this field.