In practice, the operation of a wireless communication system deviates from this ideal setting—hence the need to speak of an "estimate" of the complex impulse response $\tilde{h}(t)$ at the output of the matched filter. Hereafter, we refer to this estimate as $\tilde{h}_{\text{est}}(t)$. Ignoring the effect of channel noise, we now note from Eqs. (4.45) and (4.46) that

$$\tilde{h}_{\text{est}}(t) = \rho(t) \otimes \tilde{h}(t + T_c) \tag{4.47}$$

where $\tilde{h}(t)$ is the actual value of the complex baseband impulse response of the channel. The effect of ignoring $w(t)$ is justified only when the signal-to-noise ratio is high.

4.15.2 Viterbi Equalization

As mentioned previously, a primary objective of the baseband processor is to undo the convolution performed on the transmitted signal by the channel, which is indeed a task well suited for the Viterbi algorithm functioning as an equalizer. This is yet another novel application of the algorithm, building on what we said earlier in Section 4.9 on convolutional decoding: the Viterbi algorithm is a maximum-likelihood sequence estimator.

Consider, then, a channel with *memory* l, requiring the use of a *Viterbi equalizer* with a window of length l. (For the application at hand, a value of 4 is considered typical for l.) Correspondingly, the equalizer has 2^l possible states, with each state consisting of l symbols. As with convolutional decoding, we need a metric for the design of the equalizer, which, in turn, requires the generation of two kinds of waveforms:

1. *Estimated received waveforms.* This set of waveforms is generated by cycling a local modulator and channel model through all the possible l-bit sequences for every bit period. The combination of local modulator and channel model based on the channel estimate $\tilde{h}_{\text{est}}(t)$ is termed the *estimated waveform generator*, with its output denoted by the complex waveform $\tilde{\xi}_{\text{est}}(t)$.
2. *Compensated received waveform.* From Eq. (4.45), we note that, except for a delay, the complex impulse response estimate $\tilde{h}_{\text{est}}(t)$ equals the actual complex impulse response $\tilde{h}(t)$, convolved with the real autocorrelation function $\rho(t)$. Since the estimated waveform generator embodies the channel model, it follows that the actual received signal $\tilde{x}_{\text{data}}(t)$ should also be convolved with $\rho(t)$. Then the compensated version of $\tilde{x}_{\text{data}}(t)$, namely,

$$\tilde{\xi}_{\text{data}}(t) = \rho(t) \otimes \tilde{h}(t) \tag{4.48}$$

would be on par with the estimated received waveform $\tilde{\xi}_{\text{est}}(t)$.

Note that $\tilde{\xi}_{\text{data}}(t)$ and $\tilde{\xi}_{\text{est}}(t)$ are both continuous-time signals.

In order to generate fairly accurate digital representations of $\tilde{\xi}_{\text{data}}(t)$ and $\tilde{\xi}_{\text{est}}(t)$, it is necessary that they be sampled at a rate of n times the incoming bit rate, so as to prevent aliasing. That is, each bit of data (actual as well as estimated) is represented by n samples, where n is an integer equal to or greater than two.

The *squared Euclidean distance* between the ith sample of the kth bit in the actual received waveform $\tilde{\xi}_{\text{data}}(t)$ and the corresponding ith sample of the estimated

received waveform $\tilde{\xi}_{\text{est}}(t)$ pertaining to a possible state v of the equalizer may be expressed as the sum of two squared terms, one due to the in-phase components of these two waveforms and the other due to their quadrature components:

$$\mu_{k,v}^2(i) = (\tilde{\xi}_{\text{data},I}(k,i) - \tilde{\xi}_{\text{est},I}(v,i))^2 + (\tilde{\xi}_{\text{data},Q}(k,i) - \tilde{\xi}_{\text{est}Q}(v,i))^2$$

With η samples per bit, the sample index i ranges from 0 to $\eta - 1$. Hence, we may define the transition metric for bit k of the actual signal and possible state v of the equalizer as

$$\mu_{k,v} = \sum_{i=0}^{\eta-1} \mu_{k,v}^2(i) \tag{4.49}$$

where $v = 0,1,...,2^{l-1}$

Putting the ideas discussed here together, we may now formulate the block diagram of Fig. 4.22 for the baseband processor for channel estimation and equalization, leading to data recovery. The processor consists of three subsystems: the *estimated waveform generator*, *transition metric computer*, and *Viterbi equalizer*.

Building on the idea of the Viterbi algorithm as a maximum-likelihood sequence estimator, we may now describe the way in which the Viterbi equalizer performs its computations. The basic difference between the Viterbi equalizer and the Viterbi decoder (discussed in Section 4.9) lies in what is used for the transition metric. Specifically, we use $\mu_{k,v}$ (defined in Eq. (4.49)) for equalization and the Hamming distance for hard-decision-based convolutional decoding. Accordingly, the steps involved in the Viterbi equalization are as follows:

1. Compute the transition metric $\mu_{k,v}$ for bit k of the actual received signal and state v of the equalizer, where $v = 0,1,...,2^{l-1}$ and l stands for the window length of the equalizer.
2. Compute the accumulated transition metric for every possible path in the trellis representing the equalizer. The metric for a particular path is defined as the squared Euclidean distance between the estimated received waveform represented by that path and the actual received waveform. For each node in the trellis

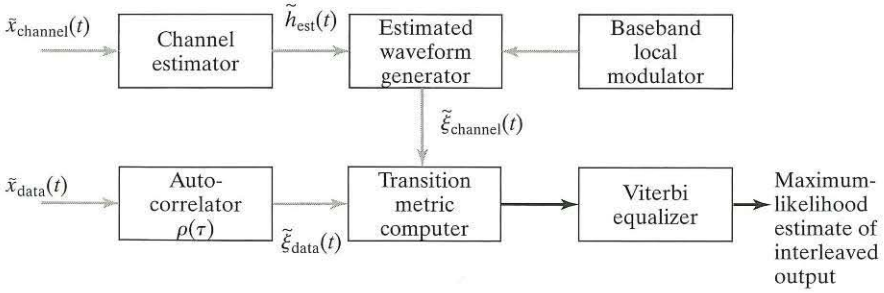


FIGURE 4.22 Block diagram of baseband processor for channel estimation and equalization. The lighter arrows indicate complex signals.

(i.e., each state of the equalizer), the Viterbi algorithm compares the two paths entering that node. The path with the lower metric is retained, and the other path is discarded.

3. Repeat the computation for every bit of the received signal.
4. The survivor, or active, path discovered by the algorithm defines the l-bit sequence applied to the local modulator in Fig. 4.22 for which the estimated received waveform $\tilde{\xi}_{\text{est}}(t)$ is the closest to the actual received waveform $\tilde{\xi}_{\text{data}}(t)$ in Euclidean distance. With this sequence at hand, the tasks of channel estimation and equalization are completed.

One last comment is in order: The window length l assigned to the equalizer depends not only on the memory of the channel, but also on whether the modulator used in the transmitter has memory of its own or not (i.e., partial-response modulation). Let l_{mem} denote the memory of the modulator and l_{channel} denote the memory of the channel. Then we may express the window length of the equalizer as

$$l = l_{\text{mem}} + l_{\text{channel}} \tag{4.50}$$

For example, $l_{\text{mem}} \approx 2$ for GMSK. With $l_{\text{channel}} = 4$, for example, we thus have $l \approx 6$.

4.16 TIME-DIVISION MULTIPLE ACCESS

The discussion thus far has focused on specific functional blocks (i.e., speech coding, channel shaping, coding, and modulation) that are basic to the design of a digital communication system. With this material at hand, we are now ready to discuss TDMA, a widely used form of multiple access for wireless communications.

The purpose of a *time-division multiple-access (TDMA) system* is to permit a number of users, say, N, to access a wireless communication channel of bandwidth B on a time-shared basis. The immediately apparent features that distinguish TDMA from FDMA are twofold:

1. Each user has access to the full bandwidth B of the channel, whereas in FDMA each user is assigned a fraction of the channel bandwidth, namely, B/N.
2. Each user accesses the channel for only a fraction of the time that it is in use and on a periodic and orderly basis, with the transmission rate being N times the user's required rate. By contrast, in FDMA, each user accesses the channel on a continuous-time basis.

Both of these features have significant implications for the operation of a TDMA wireless communication system. Access to the full bandwidth of the wireless channel means that we may now be dealing with wideband data transmission, which makes the TDMA system vulnerable to frequency-selective fading. In contrast, FDMA deals with narrowband transmission, which means that the fading channels are typically frequency flat. To combat the frequency-selective fading problem requires the use of sophisticated signal-processing techniques. Access to the channel on a time-shared basis has implications of its own. In particular, the transmission of information-bearing data over the channel takes place in the form of *bursts*, which, in turn, further complicates the requirement of synchronizing the receiver to the transmitter.

In the context of implementation, unfortunately, there is no TDMA structure applicable to all TDMA wireless systems in operation. Nevertheless, they do share a common feature: Each *frame* of the TDMA structure contains N time *slots* of equal duration. It is in the detailed structure of each slot and in the way in which the transmitting and receiving slots are assigned in time that TDMA systems differ from one another.

Typically, the bits constituting each slot of a TDMA frame are divided into two functional groups:

- *Traffic data bits*, which represent digitized speech or other forms of information-bearing data.
- *Overhead bits*, whose function is to assist the receiver in performing some auxiliary functions that are essential for satisfactory TDMA operation.

The auxiliary functions include synchronization and channel estimation. Specifically, the *synchronization* bits in a slot enable the receiver to recover sinusoidal carrier and bit-timing information, which are needed for coherent demodulation. The *framing* bits are used to estimate the unknown impulse response of the channel, which is needed for estimating the transmitted signal. As already mentioned, in TDMA systems, the transmission of information-bearing signals pertaining to any user is not continuous in time. Rather, it is *discontinuous*, requiring the use of a *buffer-and-burst* strategy. The burst form of data transmission over the channel results in an increase in the synchronization overhead, as each receiver is required to piece the transmitted signal (e.g., speech) together as it is received over a succession of frames.

Up to now, the discussion of the TDMA framing structure has been of a generic nature. Theme Example 1, presented in Section 4.17, describes the frame structure of a specific system.

4.16.1 Advantages of TDMA over FDMA

The following are some of the advantages TDMA has over FDMA:

1. With TDMA, the use of a diplexer can be avoided at the mobile terminal. A diplexer is a complicated and expensive arrangement of filters that allows the mobile terminal to transmit and receive data at the same time without jamming its own information-bearing signal. In TDMA, the terminal need not transmit and receive at the same time; hence, a diplexer is not needed.
2. With TDMA, only one RF carrier at a time is present in the channel. If the channel includes a nonlinearity, then the effects of the nonlinearity are much reduced on a single carrier than if multiple carriers were present, as in FDMA. Examples of such nonlinearities are the power amplifiers employed in base stations or those in satellite transponders.
3. With voice, a significant portion of the call consists of quiet time, when neither party is speaking. With a TDMA strategy, special processing techniques can be employed to fill the quiet times with data or other voice calls to improve the channel's efficiency.

4. With FDMA, the base station must have a channel unit (transmitter/receiver pair) for each active session. With TDMA, the same channel unit is shared between multiple sessions; thus, the base station hardware can be significantly simplified.

To achieve some of these advantages requires the use of complicated signal-processing techniques, which, in turn, necessitates a reliance on digital signal-processing technology in the form of silicon chips for cost-effectiveness. Moreover, this same enabling technology has made it possible to implement other functional needs of TDMA systems efficiently:

- sophisticated timing and fast acquisition operations for synchronizing the receiver to the transmitter, and
- source coding and channel coding techniques for the efficient and reliable transmission of information over the channel.

Putting all these operational advantages and practical realities together has made TDMA preferable to FDMA. However, a major disadvantage of TDMA is that its deployment requires an increased rate of data transmission across the wireless channel, which, in turn, may result in increased intersymbol interference (ISI), making channel equalization in the receiver a necessary practical requirement in TDMA systems.

4.16.2 TDMA Overlaid on FDMA

From the discussion presented thus far, it may appear that TDMA is implemented in a rigorous, pure form. In reality, however, TDMA is implemented in an overlaid fashion on FDMA, for practical reasons. To appreciate this point, it is important that we first recognize that the usable radio spectrum extends from tens of hertz to tens of gigahertz, which represents over nine orders of magnitude. By international agreement, this spectrum is shared by allotting certain portions of it to certain applications. For example, in North America, the band from 118 to 130 MHz is dedicated to aeronautical safety communications, and the bands from 824–849 to 869–894 MHz are dedicated to public telephony. In a very high level sense, this is a form of FDMA: *sharing the spectrum on the basis of frequency.*

Hence, every wireless communication system has an FDMA baseline, and multiple-access schemes such as TDMA are overlaid on this baseline. One of the issues is the *granularity of the underlying FDMA structure.* In this sense, TDMA comes in three basic forms:

1. *Wideband TDMA.* In this form of TDMA, there is only one or a small number of frequency channels, typically several megahertz wide. Wideband TDMA has been used in satellite communication systems in which the TDMA service occupies the full bandwidth of the satellite transponder.
2. *Medium-band TDMA.* In this form of TDMA, there is a significant number of frequency (FDMA) channels, but the bandwidth of each channel is still large enough (100 to 500 kHz) that frequency-selective fading can be expected. The GSM system discussed as a theme example in Section 4.17 is an example of medium-band TDMA, with sufficient FDMA channels being available to assign different-frequency channels to different cells and to perform the necessary task of interference management.

3. *Narrowband TDMA*. This last form of TDMA is a simple step up from a pure FDMA system. The number of users time-sharing a single channel is small, and the number of frequency channels is typically large. The bandwidth of a channel in narrowband TDMA is relatively small (usually less than 50 kHz), and, as a consequence, we can usually assume the multipath phenomenon to be flat fading. The North American IS-54 digital telephone system is an example of a narrowband TDMA system. (See Note 2 of Chapter 3.)

The appropriate choice of granularity for the underlying FDMA systems depends upon several factors:

- In a cellular system, the granularity has to be sufficient to allow different frequency assignments in a neighboring cell and perform flexible interference management.
- System complexity increases with the channel bandwidth and data-transmission rate, with the increase in complexity occurring in both the synchronization and processing aspects of the system. Lower bandwidths tend to imply lower cost solutions and lower power requirements.
- Propagation conditions may favor higher bandwidth systems, but only if appropriate measures are implemented to use this advantage. Frequency-selective fading that occurs in medium and wideband TDMA systems can provide a diversity advantage, but only if the receiver includes an effective equalizer.

One final comment is in order: TDMA is the not the only choice of multiple access for overlaying on an FDMA baseline. In Chapter 5, we will present code-division multiple access (CDMA), which can be considered a wideband system, but, in reality, is still overlaid on an FDMA baseline. Moreover, wireless communication is not limited to a single overlay. For example, from the discussion to be presented in Section 4.17, we will see that the GSM system is not simply TDMA overlaid on FDMA; rather, it also includes a third multiple-access strategy known as frequency hopping (FH); that is, GSM is, in reality, an FDMA/TDMA/FH system. Frequency hopping is also discussed in Chapter 5.

4.17 THEME EXAMPLE 1: GSM[13]

The *Global System for Mobile (GSM) communications* is a digital wireless communication system that is used all over the world. Figure 4.23 displays the basic TDMA frame structure of GSM. The structure is composed of eight 577-μs slots, which makes the total frame duration equal to 4.616 ms. The 1-bit *flag* adjacent to each data burst of 57 bits is used to identify whether the data bits are digitized speech or some other information-bearing signal. The 3 *tail* bits, all logical zeros, are used in convolutional decoding of the channel-encoded data bits. The 26-bit training sequence in the middle of the time slot is used for channel equalization. Finally, the *guard time*, occupying 8.25 bits, is included at the end of each slot to prevent data bursts received at the base station from mobile users from overlapping with each other; this is achieved by transmitting no signal at all during the guard time.

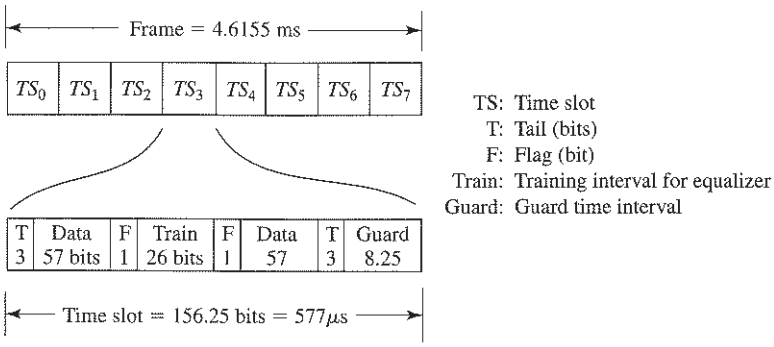


FIGURE 4.23 Frame structure of GSM communications.

The *frame efficiency* of a TDMA system is defined as the number of bits representing information-bearing signals (e.g., digitized speech), expressed as a percentage of the total number of bits (including the overhead) that are transmitted in a frame. With each slot consisting of 156.25 bits, of which 40.25 bits are overhead (ignoring the 2 flag bits), the frame efficiency of GSM is

$$\left(1 - \frac{40.25}{156.25}\right) \times 100 = 74.24\%$$

It is important to note, however, that (as remarked in Section 4.16.2) GSM is *not* a pure TDMA system. Rather, it combines TDMA with frequency hopping. Accordingly, a physical channel is *partitioned in both time and frequency.* The channel is partitioned in time because, with eight slots in a TDMA frame, each carrier frequency supports eight physical channels mapped onto the eight slots. A time slot assigned to a particular physical channel is naturally used in every TDMA frame for as long as that channel is engaged by a mobile user. Consequently, partitioning of the channel in frequency arises because the carrier assigned to such a slot changes its frequency from one frame to the next in accordance with a frequency-hopping algorithm.

In Section 4.10, we introduced the idea of interleaving as a way of combatting the Rayleigh fading problem. Frequency hopping combined with interleaving enables a TDMA system to combat the fading problem even more effectively. In the context of a TDMA system, the *principle of frequency hopping* embodies the following two considerations:

1. The carrier used to modulate a TDMA frame changes its frequency from one frame to the next.
2. If a particular TDMA frame happens to be in a deep fade, then it is highly unlikely that the next TDMA frame will also be in a deep fade, provided that the change in carrier frequency applied by the frequency-hopping algorithm from the particular frame in question to the next one is sufficiently large.

For uplink transmission, in Europe, GSM uses the frequency band 890 to 915 MHz, and for downlink transmission, it uses the frequency band 935 to 960 MHz. In either case, the maximum frequency change from one frame to the next is 25 MHz. Expressed as a

percentage of the mean carrier frequency, the *maximum frequency hopping* for the downlink is approximately

$$\frac{25}{900} \times 100 = 2.8\%$$

With this percentage of maximum frequency hopping, it turns out that the time spent by a rapidly moving mobile user in a deep fade is reduced to about 4.6 ms, which is essentially the frame duration. In the case of slowly moving mobile users (e.g., pedestrians), the frequency-hopping algorithm built into the design of GSM produces substantial gains against fades.

GSM employs a moderately complicated, 13-kilobits/s regular pulse-excited speech codec (coder/decoder) with a long-term predictor. To provide error protection for the speech-encoded bits, concatenated convolutional codes and multilayer interleaving are employed. An overall speech delay of 57.5 ms occurs in the system.

Turning next to the type of digital modulation used in GSM, we find that the method of choice is Gaussian minimum-shift keying (GMSK), which was discussed in Sections 3.7 and 4.14. For GSM, the time–bandwidth product WT of GMSK is standardized at 0.3, which provides the best compromise between increased bandwidth occupancy and resistance to cochannel interference. Ninety-nine percent of the radio frequency (RF) power of GMSK signals so specified is confined to a bandwidth of 250 kHz, which means that, for all practical purposes, the sidelobes of the GMSK signal are insignificant, for all practical purposes, outside this frequency band.

The available spectrum is divided into 200-kHz-wide subchannels, each of which is assigned to a GSM system transmitting data at 271 kb/s. Figure 4.24 depicts the power spectrum of a channel in relation to its two adjacent channels; this plot is the

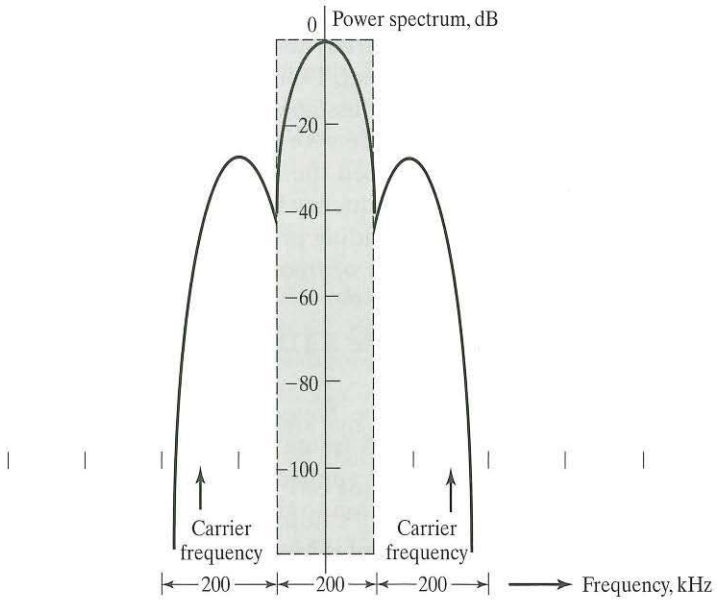


FIGURE 4.24 Power spectrum of GMSK signal for GSM communications.

passband version of the baseband power spectrum of Fig. 3.21 corresponding to $WT_b = 0.3$. From Fig. 4.24, we may make the following important observation: The RF power spectrum of the subchannel shown shaded is down by an amount close to 40 dB at the carrier frequencies of both adjacent subchannels, which means that the effect of cochannel interference in GSM is small.

4.18 THEME EXAMPLE 2: JOINT EQUALIZATION AND DECODING[14]

The material presented in Section 4.12 has taught us an important principle in digital communication theory, hereafter referred to as the *turbo coding principle.* The principle may be stated as follows:

> *The performance of the receiver of a digital communication system, embodying the plot of bit error rate (BER) versus transmitted signal energy per bit-to-noise spectral density ratio, E_b/N_0, may be significantly improved by using*
>
> ***(i)*** *a concatenated encoding strategy at the transmitter and*
>
> ***(ii)*** *an iterative receiver, with all of its components operating in soft-input, soft-output (i.e., analog) form.*

The iterative receiver is the hallmark of the turbo coding principle.

In Fig. 4.15, the concatenated encoding strategy is implemented in *parallel* form, so called because encoder 1 and encoder 2 operate in parallel on their respective inputs. Moreover, the two encoder inputs are essentially statistically independent by virtue of the turbo interleaver that separates them.

Alternatively, we can implement the concatenated encoding strategy in *serial* form, as illustrated in Fig. 4.25(a). Although, at first sight, this structure looks familiar

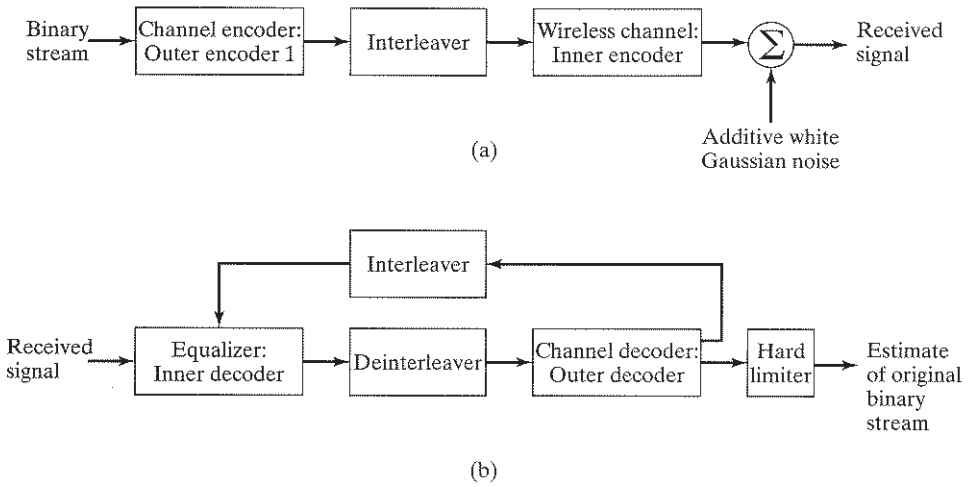


FIGURE 4.25 Joint equalization-and-decoding problem.
(a) Turbo encoder of the serial form, with the channel viewed as the inner encoder.
(b) Iterative Turbo decoder, highlighting the application of feedback around the two decoding stages.

in the context of a fast-fading wireless communication system, the viewpoint embodied in its description as a *two-stage encoder* can be justified along the following lines:

- The channel encoder, introduced into the transmitter chain to improve the reliability of communication, is viewed as the *outer encoder*.
- The wireless channel, essential for the communication process, is viewed as the *inner encoder*.
- The *channel interleaver*, introduced to disperse the burst of errors produced by the possible presence of the fast-fading phenomenon in the wireless channel, separates the two encoders in accordance with the transmit part of the turbo-coding principle.

Correspondingly, the *two-stage decoder* is configured as an *iterative receiver*, as shown in Fig. 4.25(b), in accordance with the receiver part of the turbo-coding principle. Herein lies the basis of a novel receiver structure made up of the following constituents:

- A soft-input, soft-output *channel equalizer*, designed to mitigate the effect of intersymbol interference (ISI) produced by the transmission of the encoded-interleaved signal across the channel; the equalizer acts as the *inner decoder*.
- A soft-input, soft-output *channel decoder*, designed to improve the estimates of encoded data symbols; the channel decoder acts as the *outer decoder*.
- A *channel deinterleaver*, designed to undo the permutation that is present in soft outputs produced by the equalizer, so as to facilitate proper channel decoding.
- An *interleaver*, designed to repermute the soft outputs produced by in the channel decoder, so that the feedback signal applied to the equalizer assumes a form consistent with the received signal.

Now, if we were to open the feedback loop in Fig. 4.25(b) by removing the interleaver in the *feedback path*, we would be left with a conventional receiver defined by the *forward path* made up of the channel-equalizer, channel-deinterleaver, channel-decoder chain. The practical advantage of the iterative receiver is that it performs *joint equalization and decoding*, thereby offering the potential for improving the performance of the receiver by virtue of the feedback around the two stages of processing: channel equalization and channel decoding.

In particular, the reduction in bit error rate through joint equalization and decoding performed iteratively can be explained by observing that each component in the receiver, namely, the equalizer and the decoder, helps to bootstrap the performance of the other. The *bootstrap* action manifests itself as follows:

- The equalizer uses *frequency diversity* in the channel to improve the decoder performance through *ISI reduction*.
- The decoder uses *time diversity* in the code to improve the equalizer performance through *improved estimates of uncoded data symbols.*

The net result of this bootstrap action is that, in the course of three to five iterations, a significant reduction in the bit error rate is accomplished, as is illustrated in a simple, yet insightful, computer experiment described next.

4.18.1 Computer Experiment

Consider the serial concatenated encoder of Fig. 4.25(a), with the following specifications:

1. Channel encoder (outer encoder): convolutional encoder
   Code rate = 1/2
   Constraint length, $K = 3$
   Generator polynomials:
   $g^{(1)}(D) = 1 + D^2$
   $g^{(2)}(D) = 1 + D + D^2$
2. Interleaver:
   Type: pseudorandom interleaver
   Block size: 1000 bits
3. Wireless channel: Tapped-delay-line model with the following tap weights (see Fig. 4.26, where T denotes the symbol duration):

$$w_0 = 0.93$$
$$w_1 = -0.17$$
$$w_2 = 0.35$$

   Euclidean norm of the tap-weight vector $\mathbf{w}$:

$$\begin{aligned} \|\mathbf{w}\| &= (w_0^2 + w_1^2 + w_2^2)^{1/2} \\ &= ((0.93)^2 + (-0.17)^2 + (0.35)^2)^{1/2} \\ &\approx 1 \end{aligned}$$

4. Modulation (not shown in Fig. 4.25(a)): Binary phase-shift keying (BPSK). With this simple method of modulation, the baseband model of the system assumes a real-valued form throughout the system.

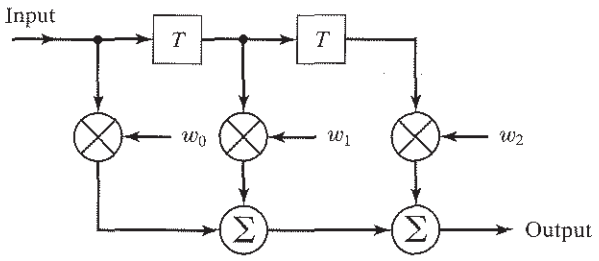


FIGURE 4.26 Tapped-delay-line model of wireless channel with three tap-weights; the blocks labeled T act like unit-delay operations.

The iterative two-stage receiver of Fig. 4.25(b) was implemented as follows:

1. Equalizer (inner decoder).
   - The channel impulse response, assumed to be known.
   - The decoding trellis, formed on the basis of the channel impulse response (i.e., tap weights of the tapped-delay-line model) and BPSK.
2. Deinterleaver, designed to deinterleave the soft outputs produced by the equalizer.
3. Channel decoder (outer decoder).
   - The decoding trellis, formed on the basis of the convolutional encoder's generator polynomials $g^{(1)}(D)$ and $g^{(2)}(D)$
   - Construction of the decoding trellis, discussed in Section 4.7
4. Interleaver, designed to interleave the soft outputs produced by the channel decoder.
5. Decoding algorithm for both the equalizer and channel decoder: The logarithmic form of the maximum a posterior probability (MAP) algorithm, discussed in Section 4.12.

Using computer simulations of the encoder/decoder system of Fig. 4.25, we plot the receiver performance, in terms of BER versus E_b/N_0, in Fig. 4.27, on the basis of which we may make the following observations:

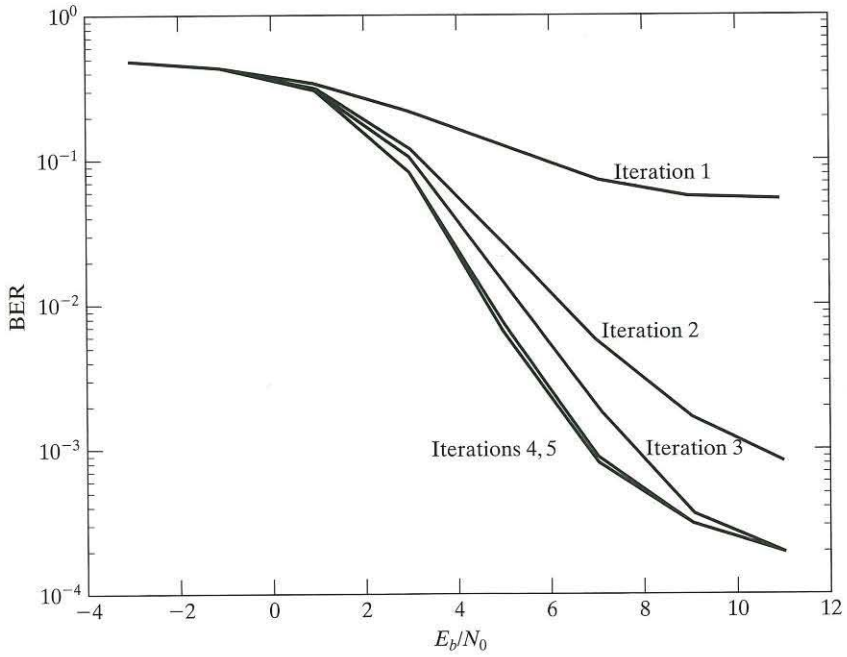


FIGURE 4.27 Performance receiver curves for the iterative joint equalization-and-decoding experiment.

1. Iterative detection, performed in accordance with the turbo coding principle, provides a significant improvement in receiver performance measured with respect to the first iteration; in effect, iteration 1 represents what is achievable with a noniterative (i.e., conventional) receiver.
2. The receiver converges in about five iterations.
3. Little change in receiver performance occurs in going from iteration 4 to iteration 5.

Problem 4.5 The baseband model used in the computer experiment on joint equalization and decoding is real valued, which is justified for BPSK modulation. To improve spectral efficiency, QPSK modulation is commonly used. Discuss the modifications that would have to be made to the baseband model in order for it to handle QPSK modulation. ■

The computer experiment just presented assumes that the receiver has perfect knowledge of the *channel state information* (CSI). In practice, we have to deal with a wireless channel that is typically nonstationary, in which case the equalizer structure has to be expanded to include a CSI estimator. (See Problem 4.24.)

4.19 THEME EXAMPLE 3: RANDOM-ACCESS TECHNIQUES

There are many instances in multiaccess communications in which a user terminal is required to send a packet of information to the base station at a random instant in time. Such instances occur, for example, when the terminal wishes to log onto the system or when the user wishes to make a telephone call. The system must provide a means by which these random requests can be serviced. This could be done in a number of ways:

1. The system could permanently assign one channel to each user.
2. The system could *poll* each user at regular intervals to see if he or she had anything to transmit.
3. The system could provide a *random-access* channel that the users could access at any time.

Since a typical user has a low duty cycle, the first approach is wasteful of spectrum. The second approach could result in long delays if there is a large number of users, and if the users are mobile, the polling process can become complicated. In this section, we will consider the third approach of assigning a random-access channel.

4.19.1 Pure Aloha[15]

Consider the following model of the random-access channel: Let us assume that there is a large population of user terminals that operate independently of each other and that each terminal has no knowledge of when the other terminals will transmit. Each terminal transmits random packets of length P, and the average transmission rate is λ. That is, there are, on average, λ packets transmitted per second by the entire user population.

This situation is commonly modeled as a *Poisson process*. From Appendix C, with Poisson arrivals, the probability that there are k arrivals in the period $[0,t]$ is given by

$$\text{Prob}(X(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{4.51}$$

Poisson processes have a *memoryless property*; that is, the probability that there are k arrivals in the period $[0,t]$ is the same as the probability that there are k arrivals in the period $[s,t+s]$ for some arbitrary s.

If two packets collide (i.e., they overlap in time), it is assumed that the information in both packets is lost. It is of interest to determine the throughput S of this channel; *throughput* is the number of packets that can be transmitted per slot, on average. If there was only one user terminal, then, clearly, the maximum throughput would be unity. In the case of a large number of terminals, however, we must consider the probability that two or more packets will collide. In Fig. 4.28, we show several types of collisions. From the figure, it should be clear that, for a packet transmitted at time t_0, any packet transmitted in the interval $[t_0 - T, t_0 + T]$, where T is the packet duration, will cause a collision.

With a Poisson model for packet arrival times, if a packet is being transmitted, then the probability that no additional packets arrive during the period $[t_0 - T, t_0 + T]$ is given by Eq. (4.51) with $k = 0$, or

$$\text{Prob}(\text{No additional packets in time } 2T) = e^{-2\lambda T} \tag{4.52}$$

Consequently, the throughput of an Aloha system is given by the product of the packet arrival rate and the probability that a packet is successfully received; that is,

$$S = \lambda e^{-2\lambda T} \tag{4.53}$$

If we normalize this equation to packets received per packet time T, then the *normalized throughput* is given by

$$\begin{aligned} S_0 &= \lambda T e^{-2\lambda T} \\ &= G e^{-2G} \end{aligned} \tag{4.54}$$

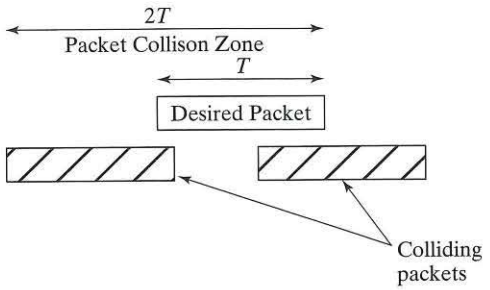


FIGURE 4.28 Illustration of packet collision zone.

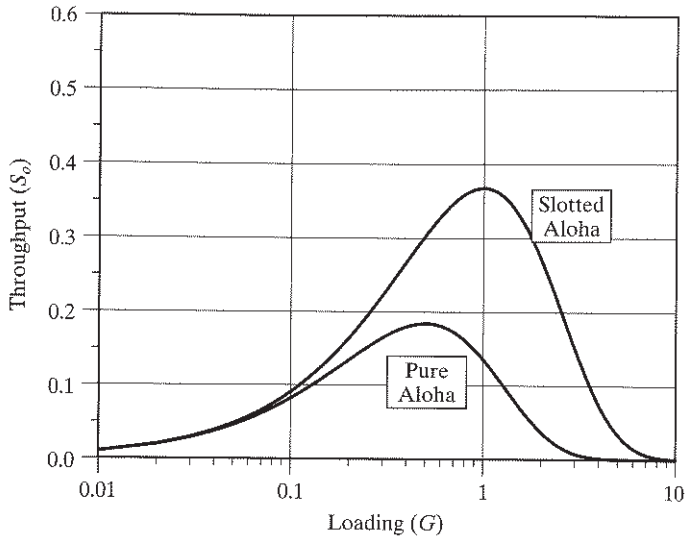


FIGURE 4.29 Normalized throughput for unslotted and slotted Aloha.

where $G = \lambda T$ is the *normalized loading per packet period*—that is, the average number of packets per slot time T. This throughput is plotted as a function of the offered load in Fig. 4.29. The peak throughput occurs at $G = 1/2$, yielding $S_0 = (2e)^{-1} \approx 0.184$ packets per packet time. That is, with an Aloha random-access channel, the maximum throughput is less than 19% of the full channel capacity. In practice, the throughput is maintained at a much smaller value so as to ensure stability of the approach.

4.19.2 Slotted Aloha

The performance of an Aloha system can be improved if a framing structure is provided. This framing structure includes fixed slot times, and user terminals are required to synchronize their transmissions with the slot times. Often, the timing of the Aloha frame is based on the timing of a forward-link broadcast channel. This form of random access with framing is referred to as *slotted Aloha*.

With slotted Aloha, a collision occurs only if the two user terminals transmit during the same T second slot. In a manner analogous to the development for the unslotted Aloha case, the normalized throughput of the slotted Aloha is given by

$$S_0 = Ge^{-G} \tag{4.55}$$

The normalized throughput of unslotted Aloha is also plotted in Fig. 4.29. The peak throughput with slotted Aloha, $S_{\max} = 1/e \approx 0.36$ packet per slot, is double that of unslotted Aloha.

4.19.3 Carrier-Sense Multiple Access[16]

The Aloha protocol was first applied in satellite networks in which the user terminals were dispersed over a wide geographic area. The transmissions were received by the

satellite and then re-broadcast over the whole area. There are two consequences of this approach. The first is that user terminals can hear the broadcast by the satellite and thus determine immediately whether a collision occurred and there is no need to transmit an acknowledgment. Second, these systems use geostationary satellites that have an altitude of approximately 36,000 kilometers, resulting in a significant transmission delay. Consequently, the user terminal can "hear" a collision, but it is too late to avoid it.

In terrestrial networks, it is often the case that each user terminal can hear the transmissions of all the other user terminals. This process of listening to the channel is referred to as *sensing*. Such a situation led to the development of another random-access protocol known as *carrier-sense multiple access* (CSMA). In its simplest form, this protocol has the following three steps:

1. If the channel is sensed as *idle*, the user terminal transmits the packet.
2. If the channel is sensed as *busy*, the transmission is scheduled for a later time according to a specified random distribution.
3. At the new point in time, the user terminal senses the channel and repeats the algorithm.

If transmission were instantaneous, then collisions would occur in the CSMA protocol only if two terminals transmitted at exactly the same time; this should be a rare occurrence. Although the transmission delay is smaller in terrestrial networks than it is in satellite networks, it is not negligible. Let τ be the maximum transmission delay between any pair of user terminals. Then collisions can occur between packets that have the timing shown in Fig. 4.30.

To analyze the throughput of CSMA, we use the following model: As in the Aloha case, we assume that packet arrivals have a Poisson distribution with average rate λ and packet duration T. Analogously to transmission in the Aloha case, a packet is transmitted successfully if it is the only packet to be transmitted in time τ. Consequently, the probability that a packet is transmitted successfully is given by

$$\text{Prob(No additional packets in time } \tau) = e^{-\lambda\tau} \tag{4.56}$$

Because of the sensing strategy, the average throughput calculation is more complex than in the Aloha case. Since a packet arriving at a terminal does not mean that it will be transmitted immediately, we have to calculate the average transmission rate; this involves calculating the average busy time per packet and the average idle time per packet. The sum of the two gives the average time between packet transmissions.

The average busy time of a channel is the packet duration, plus the propagation delay, plus the relative delay of the last colliding packet, as illustrated in Fig. 4.30. The relative delay of the last colliding packet is a random variable denoted by Y. To determine the distribution of this random variable, we note that, due to the memoryless property of the Poisson process,

$$\text{Prob (No additional packets in the interval } (t, t+s)) = e^{-\lambda s} \tag{4.57}$$

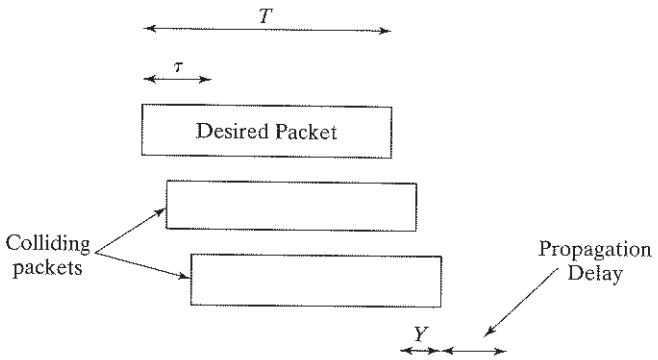


FIGURE 4.30 Illustration of collision conditions for CSMA.

So the probability that the last packet is transmitted at time y or before is equivalent to the probability that no packets are transmitted in the interval $(y,\tau]$, or

$$\text{Prob}(Y \le y) = e^{-\lambda(\tau - y)} \qquad \text{for } 0 < y < \tau \tag{4.58}$$

By determining the probability density function of random variable Y from Eq. (4.58), we may compute the average relative delay of the last colliding packet (see Problem 4.33):

$$\mathbf{E}[Y] = \tau - \frac{1 - e^{-\lambda\tau}}{\lambda} \tag{4.59}$$

Combining all of the components, we find that the average busy time of a channel per transmission is

$$\begin{aligned} T_{\text{busy}} &= T + \tau + \mathbf{E}[Y] \\ &= T + 2\tau - \frac{1 - e^{-\lambda\tau}}{\lambda} \end{aligned} \tag{4.60}$$

For a Poisson distribution, the average idle time is given by

$$T_{\text{idle}} = \frac{1}{\lambda} \tag{4.61}$$

The throughput of the CSMA channel is therefore

$$\begin{aligned} S &= \frac{\text{Prob(successful transmission)}}{T_{\text{busy}} + T_{\text{idle}}} \\ &= \frac{e^{-\lambda\tau}}{T + 2\tau - \dfrac{1 - e^{-\lambda\tau}}{\lambda} + \dfrac{1}{\lambda}} \\ &= \frac{\lambda e^{-\lambda\tau}}{\lambda(T + 2\tau) + e^{-\lambda\tau}} \end{aligned} \tag{4.62}$$

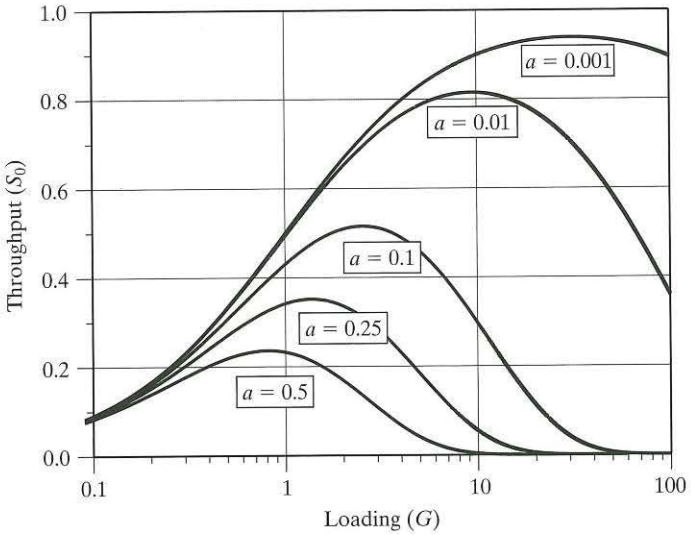


FIGURE 4.31 Throughput of the CSMA channel for varying parameter a.

If we normalize the transmission delay by setting $a = \tau/T$ and $G = \lambda T$, then the normalized throughput, in packets per packet time, is given by

$$S_0 = \frac{Ge^{-aG}}{G(1+2a)+e^{-aG}} = \frac{G}{1+(1+2a)Ge^{aG}} \tag{4.63}$$

We plot this normalized throughput for various delay parameters a in Fig. 4.31 as a function of the load offered. Intuitively, the smaller a is, the larger is the expected throughput of CSMA, which is evident in the figure.

4.19.4 Other Considerations with Random-Access Protocols

The analysis to this point applies to a datagram type of service: packets are sent and forgotten. In many systems, if there is a collision, the packet is retransmitted. If the number of retransmissions is small, then the previous results hold. But if the collision rate is high, retransmissions can add appreciably to the total traffic presented to the network at any particular time.

With random-access protocols, collisions cause delays in delivering the message. There is a trade-off between throughput and delay. With high loading, the throughput approaches zero, and the average delay per packet approaches infinity. The delay depends upon the retransmission strategy and the propagation delay. Consequently, it is difficult to compare the delays of different random-access strategies. Retransmission

cannot occur as soon as a collision is detected, as that will most certainly result in another collision. Instead, there has to be a random back-off such that the user terminal waits a random period of time before retransmitting.

Note also that the results presented herein have assumed that there are no packet errors due to noise.

4.20 SUMMARY AND DISCUSSION

In this chapter, we discussed digital wireless communications built around time-division multiple access (TDMA). Both the functional blocks that constitute the transmitter of a TDMA wireless communication system and the corresponding ones in the receiver involve features that follow from some powerful theorems in Shannon's information theory. Among these features are the following:

- *Source coding*, which is used to remove redundant bits inherent in an information-bearing signal in accordance with the source-coding theorem and rate distortion theory, thereby improving the spectral efficiency of the system.
- *Channel coding*, which involves the purposeful addition of redundant bits to the transmitted signal in a controlled manner in accordance with the channel-coding theorem, thereby providing protection against channel noise.
- *Interleaving*, which involves the pseudorandomization of the bits in a TDMA frame so as to combat the fading problem.

From the preceding features, it is apparent that the deployment of a TDMA wireless communication system offers a practical means of improving receiver performance at the expense of a significant increase in system complexity. However, system complexity is not an issue of concern, because electronics are inexpensive, thanks to the ever-increasing computing power and cost-effectiveness of silicon chips and computer software.

The other major issue discussed in the chapter is that of channel estimation and equalization. Since TDMA channels are typically wider than FDMA channels, they are more likely to be frequency selective, hence posing a more difficult channel-estimation problem. The need for channel estimation arises because the impulse response of the channel is unknown. This matter is taken care of by transmitting a known sequence over the channel. The issue of equalization relates to the need for undoing the convolution performed on the transmitted signal by the channel. The Viterbi equalizer, based on the Viterbi algorithm acting as a maximum-likelihood sequence estimator, provides a powerful method for solving this second problem.

GSM was discussed in the chapter as the first of three theme examples. In GSM, TDMA is combined with frequency hopping so as to combat the problem of deep fades in a more effective manner than would be possible otherwise. The combined use of time slots in TDMA frames and changes in the carrier frequency from one frame to the next in accordance with a frequency-hopping algorithm results in partitioning of the physical channel in both time and frequency. For the transmission of TDMA frames over the channel, GSM uses Gaussian minimum-shift keying (GMSK), which is

a spectrally efficient form of continuous-phase frequency-shift keying. A combination of regular pulse-excited speech codecs (with long-term prediction), concatenated convolutional codecs, and multilayer interleavers/deinterleavers is used for the processing of individual speech channels.

The second theme example addressed a novel application of the turbo coding principle to solve the joint equalization-and-decoding problem in an iterative manner. The procedure described therein is different from the traditional approach, in that the channel equalizer and channel decoder operate inside a closed feedback loop, helping to bootstrap the performance of each other. The net result is a receiver that achieves, for a prescribed E_b/N_0, a bit error rate significantly lower than that attainable by traditional means (typically, in three to five iterations).

The third theme example addressed random-access techniques in which each user terminal is required to send a packet of information to the base station at a random instant in time. In this third example, we discussed the following configurations with increasing levels of performance:

- Pure Aloha, involving a large population of user terminals that operate independently of each other and that have no knowledge of when the other terminals will transmit.
- Slotted Aloha, using fixed slot times and requiring the user terminals to synchronize their transmissions with the slot times.
- Carrier-sense multiple access, following three steps:
  1. Transmit the packet if the channel is idle
  2. Schedule the transmission for a later time if the channel is busy
  3. Sense the channel at a new point in time, and repeat the algorithm.

According to the approach taken to describe the channel coding process in this chapter, channel encoding is performed separately from modulation; likewise for demodulation and decoding in the receiver. Moreover, the provision for error correction is made by transmitting additional redundant bits (i.e., parity-check bits) in the code, which has the effect of lowering the spectral efficiency in bits per second per hertz. That is, bandwidth utilization is traded for increased power efficiency. To attain a more efficient utilization of the two communication resources, namely, channel bandwidth and transmit power, the processes of channel coding and modulation would have to be treated as a single entity rather than two separate ones. This treatment is precisely what is done in trellis-coded modulation.[17] When this method of modulation is applied to wireless communications, the usual procedure is to to insert an interleaver between the encoder and signal-space mapper so as to overcome the multipath fading problem.

The interleaving process may be viewed as a form of time diversity introduced into the transmitted signal. (A detailed treatment of diversity, with emphasis on spatial diversity, is presented in Chapter 6.) In a variant of coded modulation known as the bit-interleaved coded modulation (BICM), the interleaving process is applied at the level of encoded bits rather the encoded symbols.[18] The motivation behind BICM is to

produce a coded system with a high degree of time diversity, whereby the bit error rate of the receiver is governed by some product of d_{free} terms, where d_{free} refers to the free binary Hamming distance of the code. It turns out that for the same E_b/N_0, BICM outperforms the baseline approach for coded modulation by producing a smaller BER.

NOTES AND REFERENCES

[1] The celebrated classic paper of Shannon (1948) laid down the foundations of information theory, and with it, a principled approach to the design of digital communication systems. For a detailed treatment of Shannon's theory, see Cover and Thomas (1991). An introductory treatment of the subject is presented in Chapter 9 of Haykin (2001).

[2] Chapter 3 of Steele and Hanzo (1999) presents a detailed discussion of speech-coding techniques, with an emphasis on their relevance to wireless communications.

[3] For detailed treatments of traditional error-control coding techniques, see Clark and Cain (1981), Lin and Costello (1983) and Michelson and Levesque (1985). These techniques are also discussed in Chapter 4 of Steele and Hanzo (1999), which emphasizes their relevance to wireless communications. Convolutional codes were first described by Elias (1955).

[4] The free distance of convolutional codes is discussed in Viterbi and Omura (1979), Benedetto and Biglieri (1999), and Proakis (1995).

[5] The classic paper on the Viterbi algorithm is due to Viterbi (1967). For a tutorial paper on this algorithm, see Forney (1973).

[6] Interleaving, of both the block and convolutional types, is discussed in some detail in Clark and Cain (1981) and in lesser detail in Sklar (2001).

[7] Discussions of pseudorandom interleaving in the context of turbo coding are presented in Vucetic and Yuan (2000) and Heegard and Wicker (1999). We may also mention the so-called *S-constraint interleaver*, devised by Divsalar and Pollara (1995). This new interleaver is based on the generation of N uniformly distributed integers, subject to an S-constraint defined by the minimum interleaving distance. Specifically, in the construction of a turbo code, S is chosen to correspond to the maximum input error pattern length to be broken by the interleaver.

[8] Turbo codes were invented by Berrou *et al.* (1993); see also Berrou and Glavieux (1996, 1998). Heegard and Wicker (1999), Vucetic and Yuan (2000), and Hanzo *et al.* (2002) discuss turbo codes as well. For a detailed treatment of iterative decoding, see Chugg *et al.* (2001). Hanzo *et al.* discusses the performance of turbo codes over fading channels.

The discovery of turbo processing has emerged as a revolution, not only affecting the development of good codes for reliable communications and channel equalization as discussed in this chapter, but also affecting other design aspects of digital communication systems, as summarized here:

- Turbo-like codes for source coding as well as joint source-channel coding
- Iterative timing recovery and phase estimation for synchronizing the receiver to the transmitter
- Turbo-like MIMO (multiple-input, multiple output) wireless communications

Turbo-BLAST, an example of turbo-MIMO wireless communications, is discussed in Chapter 6.

[9] The BCJR algorithm is named in honor of its four originators: Bahl, Cocke, Jelinek, and Raviv, who coauthored the first forward–backward method for implementing maximum a posterior probability decoding in 1974.

[10] For the original exposition of the EXIT chart, see ten Brink (1999).

[11] For matters relating to issues 1 and 2 discussed under Subsection 4.13.6 on joint equalization and decoding, see Benedetto and Montrosi (1996) and Berrou (2003), respectively.

[12] The material presented in this section follows Chapter 6 of Steele and Hanzo (1999).

[13] The abbreviation GSM originally stood for the French name *Groupe de travail pour les Services Mobiles*. In recognition of the widespread use of GSM communications all over the globe, it was later renamed *Global System of Mobile* communications.

[14] The issue of an iterative receiver for dealing with joint equalization and decoding is discussed in Chugg *et al.* (2001), pp. 105–110.

[15] The term "Aloha" is Hawaiian, and it literally means "love," but it is commonly used as a greeting or farewell. The original ALOHA system was developed at the University of Hawaii by Abramson (1970) as a protocol for a terrestrial radio system that broadcast packets for computer communications. It was later adapted to satellite communications. The collection of papers edited by Abramson (1993) has numerous contributions describing the ALOHA protocol variants and their performance.

[16] In carrier-sense multiple-access, if every terminal cannot hear every other one, then there can be a degradation in performance. In Tobagi and Kleinrock (1975), it is shown that if users can be divided into two groups A and B such that members within a group can hear each other, but cannot hear all the members in the other group, then the performance of CSMA rapidly degrades such that it is worse than slotted Aloha. There is further degradation with three or more groups.

[17] Trellis-coded modulation was discovered by Ungerboeck; see his seminal paper published in 1982; see also the two-part tutorial paper (Ungerboeck, 1987).

[18] Bit-interleaved coded modulation (BICM), using a rate-2/3, 8-PSK modulator, was first described in Zehavi (1992). For a detailed information-theoretic treatment of BICM, see Caire *et al.* (1998). These two papers discuss two different ways of implementing BICM, depending on how the bit-interleaving process is itself implemented:

- The encoded bits are interleaved separately.
- A common bit interleaver is used.

The latter implementation makes it possible to take a more general approach to the treatment of BICM.

ADDITIONAL PROBLEMS

Convolutional coding

Note: For Problem 4.6 through 4.10. the same message sequence, 10111..., *is used so that we may compare the outputs of different encoders for the same input.*

Problem 4.6 Consider the convolutional encoder of Fig. 4.32 with rate $r = 1/2$ and constraint length $K = 2$. Find the encoder output produced by the message sequence 10111....

Problem 4.7 Figure 4.33 shows a convolutional encoder with rate $r = 1/2$ and constraint length $K = 4$. Determine the encoder output produced by the message sequence 10111....

Problem 4.8 Consider the convolutional encoder of Fig. 4.34 with rate $r = 2/3$ and constraint length $K = 2$. Determine the code sequence produced by the message sequence 10111....

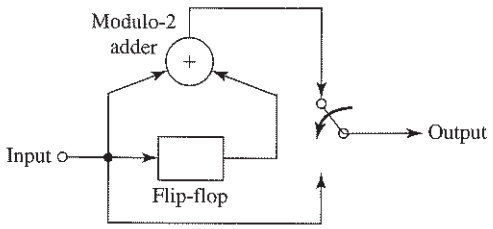


FIGURE 4.32 Diagram for problem 4.6.

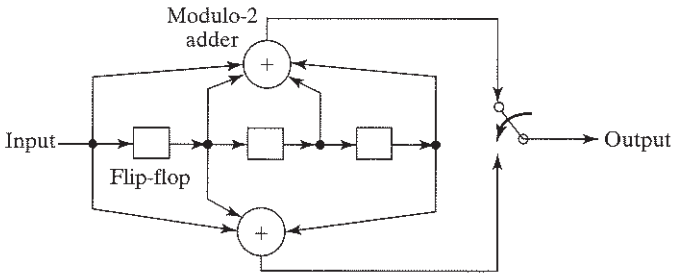


FIGURE 4.33 Diagram for problem 4.7.

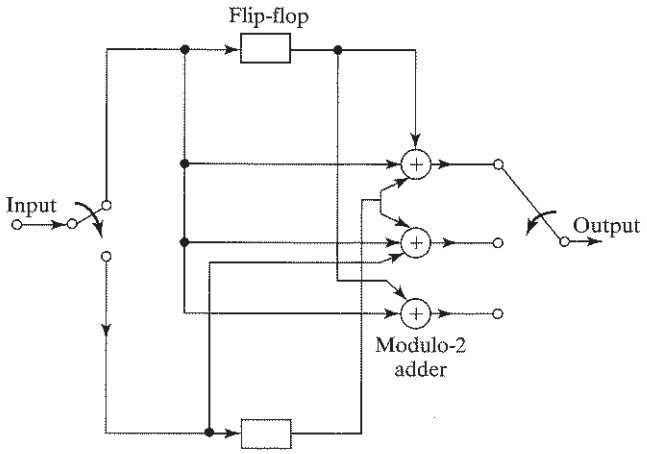


FIGURE 4.34 Diagram for problem 4.8.

Problem 4.9 Construct the trellis diagram for the encoder of Fig. 4.33, assuming a message sequence of length 5. Trace the path through the trellis corresponding to the message sequence 10111.... Compare the resulting encoder output with that found in Problem 4.7.

Problem 4.10 Construct the state diagram for the encoder of Fig. 4.33. Starting with the all-zero state, trace the path that corresponds to the message sequence 10111..., and compare the resulting code sequence with that found in Problem 4.7.

Problem 4.11 The code rate of the convolutional codes discussed in Section 4.7 is $1/n$, where n is the number of modulo-2 adders used in the encoder. A new convolutional code with code rate k/n is required, where k is an integer. How would you generate such a convolutional code? Justify your answer.

Problem 4.12 A convolutional code has the constraint length $K = 5$.

(a) Assuming that the code is of a nonsystematic nature and using the Viterbi algorithm, how many errors can be corrected by the code for a "bursty" wireless channel?

(b) Repeat the problem for a systematic code.

Problem 4.13 Consider the nonsystematic convolutional codes listed in Table 4.3 for constraint length $K = 6$ and 7.

(a) Construct the generator polynomials for these two codes.

(b) Find the free distance for each of the codes.

(c) Identify which, if any, of the two codes can deal with a fading channel that produces a burst of five transmission errors.

Interleaving

Problem 4.14 The expected length of an error burst produced by a fast-fading channel is inversely proportional to the speed of a mobile unit and directly proportional to the bit rate transmitted by the unit. Justify the validity of this statement.

Problem 4.15

(a) Continuing with the classical block interleaver/deinterleaver discussed in Subsection 4.10.1, show that the end-to-end delay produced by the interleaving/deinterleaving action is $2LN$ bits and that the memory requirement is LN.

(b) Show that the corresponding results for the convolutional interleaver discussed in Subsection 4.10.2 are LN and $LN/2$, respectively.

Problem 4.16 The classical block interleaver can be operated in four different permuted ways, depending on the order in which the incoming symbols are written into the rows of the matrix of memory elements and the order in which they are read out along the columns:

(a) Left-to-right write in/top-to-bottom read out (LR/RB)

(b) Left-to-right write in/bottom-to-top read out (LR/BT)

(c) Right-to-left write in/top-to-bottom read out (RL/TB)

(d) Right-to-left write in/bottom-to-top read out (RL/BT)

Consider an interleaver in which $N = 4$ rows and $L = 4$ columns.

(a) Given the input sequence {0,1,2,3,4,...,14,15,16,...}, construct the output sequence for each of the preceding four permutations.

(b) With $N = L$, the LR/TB and RL/BT permutations are self-inverse; that is,

$$\mathbf{I}_{\text{LR/TB}}^{-1} = \mathbf{I}_{\text{LR/TB}} \qquad \text{and} \qquad \mathbf{I}_{\text{RL/TB}}^{-1} = \mathbf{I}_{\text{RL/TB}}$$

where $\mathbf{I}$ is the square matrix describing the interleaver and $\mathbf{I}^{-1}$ is the inverse matrix describing the deinterleaver. Demonstrate the self-inverse property for the example specified in part (a).

Problem 4.17 The classical block (N,L) interleaver has an important property: *A burst of less than L contiguous transmission errors results in isolated errors at the deinterleaver output that are separated from each other by at least N symbols.* Demonstrate this property for each of the following block interleavers:

(a) $N = 4, L = 4$

(b) $N = 4, L = 5$

(c) $N = 4, L = 6$

Problem 4.18 Compare the advantages and disadvantages of block interleavers with those of convolutional interleavers in the context of four issues:

(a) End-to-end delay

(b) Memory requirement

(c) Synchronization of commutators in the receiver with those in the transmitter

(d) Complexity of implementation

Baseband processing for channel estimation and equalization

Problem 4.19 Estimation of the impulse response of a wireless communication channel requires the transmission of a sounding sequence, preferably with the following desirable features:

1. The autocorrelation function of the sequence closely approximates a delta function.
2. The sequence can be locally generated in the receiver in synchrony with the transmitter.

Describe a sounding sequence that satisfies these two requirements, and justify the sequence.

Problem 4.20 Equation (4.50) states that the window length of the Viterbi equalizer equals the sum of two memories, one due to the shaping filter used in the partial-response modulation and the other due to the wireless channel. Justify the validity of that equation.

Turbo decoding

Problem 4.21 Construct a table comparing the features that distinguish the Viterbi algorithm and the BCJR algorithm, discussed in Sections 4.9 and 4.12, respectively.

Problem 4.22 Suppose that a turbo encoder of the parallel form is used as the channel encoder (i.e., outer encoder) in the serial two-stage encoder of Fig. 4.25(a).

(a) Develop the structure of the new iterative receiver.
(b) In what ways does this receiver differ from that of Fig. 4.25(b)?
(c) What can be improved by the new receiver? What about limitations?

Problem 4.23 The iterative joint equalization-and-decoding receiver of Fig. 4.25 is an example of a closed-loop feedback system. In such a system, the receiver may become *unstable* (i.e., diverge). In qualitative terms, explain how such a phenomenon can arise in practice.

Problem 4.24 In the computer experiment presented in Section 4.18, it was assumed that the receiver "knows" about the state of the channel. In TDMA wireless communication systems, this requirement is usually satisfied by including a training sequence in each packet transmitted. For example, in GSM, the training sequence, which is "known" to the receiver, occupies 20 percent of each packet.

With this background, the requirement is to postulate a procedure for estimating the channel impulse response.

(a) Adapting Eq. (4.9) to the problem at hand, formulate a procedure for performing this estimation.
(b) Given the joint equalization-and-decoding strategy described in Fig. 4.25, discuss how the use of bootstrapping (made possible by the strategy) can be used to improve the estimate of the channel impulse response.

Problem 4.25 In this problem, we revisit the turbo principle applied to the joint equalization-and-decoding problem discussed as Theme Example 2 in Section 4.18. Specifically, we look at a more difficult multipath channel represented by the tapped-delay-line model of Fig. 4.26 with the following tap weights:

$$w_0 = 0.5$$
$$w_1 = \sqrt{0.5}$$
$$w_2 = 0.5$$

One way to tackle this problem is to consider the following system specifications:

1. Channel code: rate-1/2, 16-state convolutional code with the following two generator polynomials:

$$g^{(1)}(D) = 1 + D^3 + D^4$$
$$g^{(2)}(D) = 1 + D^2 + D^4$$

2. Block interleaver of size 57×30
3. QPSK modulation with Gray coding
4. Packet size: 57 symbols

Contrast the issues involved in the equalization-and-decoding problem embodied in this problem versus those discussed in the computer experiment in Section 4.18.

Time-division multiple access

Problem 4.26 The composition of a TDMA frame permits the use of a single-carrier frequency for both forward and reverse transmissions of information-bearing signals. Describe how such a two-way communication can be accomplished.

Problem 4.27 A TDMA frame uses a preamble of n_P bits, N time slots, and n_{T_1} trail bits. Each time slot contains n_{T_2} trail bits, n_G guard bits, n_S synchronization and channel estimation bits, and n_I information-bearing bits. Derive formulas for (a) the framing efficiency η of the system and (b) the size of a TDMA frame.

Problem 4.28 As mentioned in Section 3.8, FDMA relies on the use of highly selective bandpass filters for its operation. A bandpass filter is said to be highly selective if its *quality factor* or, simply *Q-factor*, is high compared with unity; the Q-factor is defined as the ratio of the midband frequency of the filter to its bandwidth.

In this context, TDMA enjoys an advantage over FDMA in that it relaxes this requirement. Discuss how TDMA attains its advantage.

Random-access techniques

Problem 4.29 Suppose a system with a user population of 100 terminals plans to use short packets for making reservations on a longer demand-assigned TDMA channel. There are two options for making the reservations: a polling method or slotted Aloha. Assume that the packet length is equal to T in either case. Which method would be the most efficient and under what conditions?

Problem 4.30

(a) Prove that the peak throughput of a pure Aloha system is $1/2e$, where e is the base of the natural algorithm.

(b) Suppose a system has two different packet lengths: T and $2T$. Find the throughput of a slotted Aloha system in this case.

Problem 4.31 Suppose that, in a slotted Aloha system, there is a 10% packet error rate due to noise, in addition to the error rate associated with those packet errors due to collisions.

(a) Discuss how the system throughput will be affected.

(b) Repeat the discussion for a carrier-sensitive multiple-access (CSMA) system.

Problem 4.32 Assume that, in a local area network using carrier-sensitive multiple access (CSMA), the average number of packets offered per second is 1500, with a packet size of 50 microseconds. If the maximum diameter of a local area network is 200 meters, determine the expected throughput of the system in packets per second.

Problem 4.33 Following the discussion on page 247, show that, in a CSMA system, the average relative delay of the last colliding packet is

$$\mathbf{E}[Y] = \tau - \frac{1 - e^{-\lambda\tau}}{\lambda}$$

CHAPTER 6

Diversity, Capacity and Space-Division Multiple Access

6.1 INTRODUCTION

Up to now, we have emphasized the multipath fading phenomenon as an inherent characteristic of the wireless channel. Given this physical reality, how do we make the communication process across the wireless channel into a *reliable* operation? The answer to this fundamental question lies in the use of *diversity*, which may be viewed as a form of redundancy. In particular, if several replicas of the information-bearing signal can be transmitted simultaneously over independently fading channels, then there is a good likelihood that at least one of the received signals will not be severely degraded by channel fading. There are several methods for making such a provision. In the context of the material covered in this book, we may identify three approaches to diversity:

1. Frequency diversity
2. Time (signal-repetition) diversity
3. Space diversity

In *frequency diversity*, the information-bearing signal is transmitted by means of several carriers that are spaced sufficiently apart from each other to provide independently fading versions of the signal. This may be accomplished by choosing a frequency spacing equal to or larger than the coherence bandwidth of the channel. The frequency-hopping form of spread-spectrum modulation, discussed in Chapter 5, is an example of frequency diversity.

In *time diversity*, the same information-bearing signal is transmitted in different time slots, with the interval between successive time slots being equal to or greater than the coherence time of the channel. If the interval is less than the coherence time of the channel, we can still get some diversity, but at the expense of performance. In any event, time diversity may be likened to the use of a repetition code for error-control coding. In a more general setting, we may view channel coding with interleaving, discussed in Chapter 4, as a form of time diversity.

In *space diversity*, multiple transmit or receive antennas, or both, are used, with the spacing between adjacent antennas being chosen so as to ensure the independence of possible fading events occurring in the channel. In practice, however, we find that

antenna spacings which result in correlations as high as 0.7 may incur a performance loss of at most half a decibel, compared with the ideal case of independent channels. Of the three kinds of diversity, space diversity is the subject of interest in this chapter. Depending on which end of the wireless link is equipped with multiple antennas, we may identify three different forms of space diversity:

1. *Receive diversity*, which involves the use of a single transmit antenna and multiple receive antennas.
2. *Transmit diversity*, which involves the use of multiple transmit antennas and a single receive antenna.
3. *Diversity on both transmit and receive*, which combines the use of multiple antennas at both the transmitter and receiver. Clearly, this third form of space diversity includes transmit diversity and receive diversity as special cases.

In the literature, a wireless channel using multiple antennas at both ends is commonly referred to as a *multiple-input, multiple-output (MIMO) channel*. Technology built around MIMO channels resolves the fundamental issue of having to deal with two practical realities of wireless communications:

- a user terminal of limited battery power, and
- a channel of limited RF bandwidth.

Given fixed values of transmit power and channel bandwidth, this new technology offers a sophisticated approach to exchanging increased system complexity for *boosting* the channel capacity (i.e., the spectral efficiency of the channel, measured in bits per second per hertz) up to a value significantly higher than that attainable by any known method based on a single-input, single-output channel. More specifically, when the wireless communication environment is endowed with rich Rayleigh scattering, the MIMO channel capacity is roughly proportional to the number of transmit or receive antennas, whichever is smaller. That is to say, we have a spectacular increase in spectral efficiency, with the channel capacity being roughly doubled by doubling the number of antennas at both ends of the link.

Another approach to increasing the spectral efficiency of wireless communications is to use highly *directional antennas*, whereby user terminals are separated in space by virtue of their angular directions. This approach is the basis of *space-division multiple access* (SDMA), discussed in the latter part of the chapter.

The chapter is organized as follows: Section 6.2 discusses the notion of space diversity on receive, using four techniques for its implementation, namely, selection combining, maximal-ratio combining, equal-gain combining, and square-law combining. Section 6.3 describes a mathematical model of MIMO wireless communications. This discussion is followed by Section 6.4 on the channel capacity of MIMO systems, assuming that the receiver "knows" the state of the channel. Section 6.5 presents another viewpoint of the input–output relation of the MIMO channel by applying a transformation known as singular-value decomposition to the channel matrix; the resulting decomposition is insightful in the context of fading correlation. Section 6.6 discusses space–time block codes for the joint coding of multiple transmit antennas in MIMO wireless communications. Section 6.7 examines differential space-time block

codes, used to simplify the receiver design; such an approach eliminates the need for the receiver to "know" the state of the channel.

Space-division multiple access and the related use of smart antennas are discussed in Section 6.8. The next three sections, 6.9 through 6.11, present three theme examples respectively dealing with (1) a type of coherent MIMO wireless communication system popularized as a BLAST architecture, (2) the practical merits of different antenna diversity techniques and the spectral efficiency of space–time block codes and BLAST systems, and (3) keyhole channels that arise when the channel matrix of a MIMO wireless link is rank deficient.

6.2 "SPACE DIVERSITY ON RECEIVE" TECHNIQUES

In "space diversity on receive," multiple receiving antennas are used, with the spacing between adjacent antennas chosen so that their respective outputs are essentially independent of each other. This requirement may be satisfied by spacing the adjacent receiving antennas by as much as 10 to 20 radio wavelengths or less apart from each other. Typically, an elemental spacing of several radio wavelengths is deemed adequate for space diversity on receive. The much larger spacing is needed for elevated base stations, for which the angle spread of the incoming radio waves is small; note that the spatial coherence distance is inversely proportional to the angle spread. Through the use of diversity on receive as described here, we create a corresponding set of fading channels that are essentially independent. The issue then becomes that of combining the outputs of these statistically independent fading channels in accordance with a criterion that will to provide improved receiver performance. In what follows, we describe four diversity-combining techniques: selection combining, maximal-ratio combining, equal-gain combining, and square-law combining; the first three involve the use of linear receivers, and the fourth utilizes a nonlinear receiver.

6.2.1 Selection Combining

The block diagram of Fig. 6.1 depicts a diversity-combining structure that consists of two functional blocks: N_r linear receivers and a logic circuit. This diversity system is said to be of a *selection-combining* kind, in that, given the N_r receiver outputs produced by a common transmitted signal, the logic circuit *selects* the particular receiver output with the *largest signal-to-noise ratio* as the received signal. In conceptual terms, selection combining is the simplest form of "space diversity on receive" techniques.

To describe the benefit of selection combining in statistical terms, we assume that the wireless communication channel is described by a *frequency-flat, slowly fading Rayleigh channel*. The implications of this assumption are threefold:

1. The frequency-flat assumption means that all the frequency components constituting the transmitted signal are characterized by the same random attenuation and phase shift.
2. The slow-fading assumption means that fading remains essentially unchanged during the transmission of each symbol.
3. The fading phenomenon is described by the Rayleigh distribution.

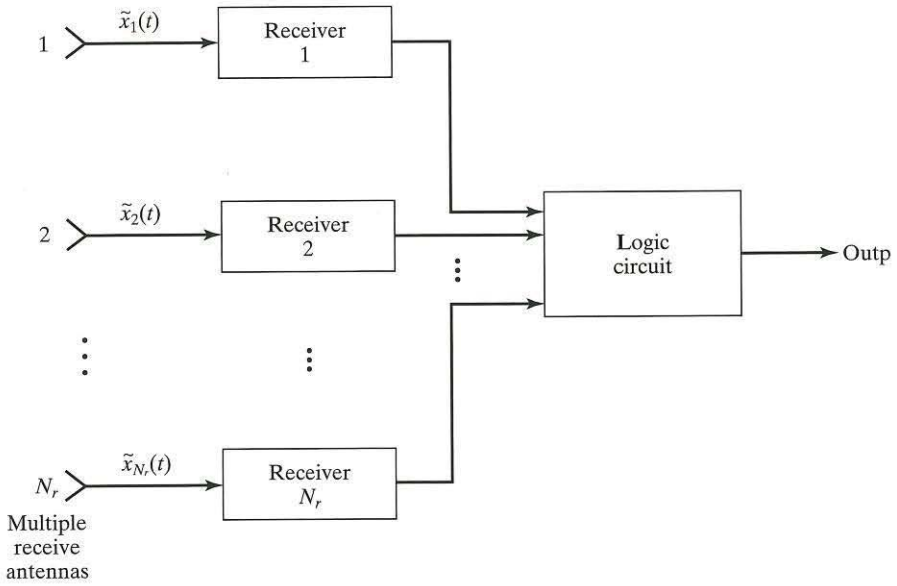


FIGURE 6.1 Block diagram of selection combiner, using N_r receive antennas.

Let $\tilde{s}(t)$ denote the complex envelope of the modulated signal transmitted during the symbol interval $0 \le t \le T$. Then, in light of the assumed channel, the complex envelope of the received signal of the kth diversity branch is defined by

$$\tilde{x}_k(t) = \alpha_k e^{j\theta_k}\tilde{s}(t) + \tilde{w}_k(t) \qquad \begin{matrix} 0 \le t \le T \\ k = 1, 2, \ldots, N_r \end{matrix} \tag{6.1}$$

where, for the kth diversity branch, the fading is represented by the multiplicative term $\alpha_k e^{j\theta_k}$ and the additive channel noise is denoted by $\tilde{w}_k(t)$. With the fading assumed to be slowly varying relative to the symbol duration T, we should be able to estimate and then remove the unknown phase shift θ_k at each diversity branch with sufficient accuracy, in which case Eq. (6.1) simplifies to

$$\tilde{x}_k(t) \approx \alpha_k\tilde{s}(t) + \tilde{w}_k(t) \qquad \begin{matrix} 0 \le t \le T \\ k = 1, 2, \ldots, N_r \end{matrix} \tag{6.2}$$

The signal component of $\tilde{x}_k(t)$ is $\alpha_k\tilde{s}(t)$ and the noise component is $\tilde{w}_k(t)$. The average signal-to-noise ratio at the output of the kth receiver is therefore

$$(\text{SNR})_k = \left(\frac{\mathbf{E}[|\alpha_k\tilde{s}(t)|^2]}{\mathbf{E}[|\tilde{w}_k(t)|^2]}\right) = \frac{\mathbf{E}|\tilde{s}(t)|^2}{\mathbf{E}|\tilde{w}_k(t)|^2}E[\alpha_k^2] \qquad k = 1, 2, \ldots, N_r \tag{6.3}$$

Ordinarily, the mean-square value of $\tilde{w}_k(t)$ is the same for all k. Accordingly, we have

$$(\text{SNR})_k = \frac{E}{N_0}\mathbf{E}[\alpha_k^2] \qquad k = 1, 2, \ldots, N_r \tag{6.4}$$

where E is the symbol energy and N_0 is the one-sided noise spectral density. For binary data, E equals the transmitted signal energy per bit, namely, E_b.

Let γ_k denote the *instantaneous* signal-to-noise ratio measured at the output of the kth receiver during the transmission of a given symbol. Then replacing the mean-square value $\mathbf{E}[|\alpha_k|^2]$ by the instantaneous value $|\alpha_k|^2$ in Eq. (6.4), we may write

$$\gamma_k = \frac{E}{N_0}\alpha_k^2 \qquad k = 1, 2, \ldots, N_r \tag{6.5}$$

Under the assumption that the random amplitude α_k is Rayleigh distributed, the squared amplitude α_k^2 will be *exponentially distributed* (i.e., chi-squared with two degrees of freedom; see Section 2.6.1). If we further assume that the average signal-to-noise ratio $(\text{SNR})_k$ over the short-term fading is the same, namely, Γ_{av}, for all the N_r diversity branches, then we may express the probability density functions of the random variables Γ_k pertaining to the individual branches as

$$f_{\Gamma_k}(\gamma_k) = \frac{1}{\gamma_{\text{av}}}\exp\left(-\frac{\gamma_k}{\gamma_{\text{av}}}\right) \qquad \begin{matrix}\gamma_k \geq 0 \\ k = 1, 2, \ldots, N_r\end{matrix} \tag{6.6}$$

Problem 6.1 Following the material presented in Section 2.6.1, derive Eq. (6.6). ■

For some signal-to-noise ratio γ, the associated cumulative distributions of the individual branches are

$$\begin{aligned}\text{Prob}(\gamma_k \leq \gamma) &= \int_{-\infty}^{\gamma} f_{\Gamma_k}(\gamma_k)d\gamma_k \\ &= 1 - \exp\left(-\frac{\gamma}{\gamma_{\text{av}}}\right) \qquad \gamma \geq 0\end{aligned} \tag{6.7}$$

Since, by design, the N_r diversity branches are essentially statistically independent, the probability that all the diversity branches have a signal-to-noise ratio less than the threshold γ is the product of the individual probabilities that $\gamma_k < \gamma$ for all k; that is,

$$\begin{aligned}\text{Prob}(\gamma_k < \gamma \text{ for } k = 1, 2, \ldots, N_r) &= \prod_{k=1}^{N_r} \text{Prob}(\gamma_k < \gamma) \\ &= \prod_{k=1}^{N_r}\left[1 - \exp\left(-\frac{\gamma}{\gamma_{\text{av}}}\right)\right] \\ &= \left[1 - \exp\left(-\frac{\gamma}{\gamma_{\text{av}}}\right)\right]^{N_r} \qquad \gamma \geq 0\end{aligned} \tag{6.8}$$

which decreases in numerical value with increasing N_r.

The cumulative distribution function of Eq. (6.8) is the same as the cumulative distribution function of the random variable Γ_{sc} described by the value

$$\gamma_{sc} = \max\{\gamma_1, \gamma_2, \ldots, \gamma_{N_r}\} \tag{6.9}$$

which is less than the threshold γ if, and only if, the individual signal-to-noise ratios $\gamma_1, \gamma_2, \ldots, \gamma_{N_r}$ are all less than γ. Indeed, the cumulative distribution function of the selection combiner (i.e., all of the N_r diversity branches that have a signal-to-noise ratio less than γ) is given by

$$F_\Gamma(\gamma_{sc}) = \left[1 - \exp\left(-\frac{\gamma_{sc}}{\gamma_{av}}\right)\right]^{N_r} \qquad \gamma_{sc} \geq 0 \tag{6.10}$$

By definition, the probability density function $f_\Gamma(\gamma_{sc})$ is the derivative of the cumulative distribution function $F_\Gamma(\gamma_{sc})$ with respect to the argument γ_{sc}. Hence, differentiating Eq. (6.10) with respect to γ_{sc} yields

$$\begin{aligned} f_\Gamma(\gamma_{sc}) &= \frac{d}{d\gamma_{sc}} F_\Gamma(\gamma_{sc}) \\ &= \frac{N_r}{\gamma_{av}} \exp\left(-\frac{\gamma_{sc}}{\gamma_{av}}\right)\left[1 - \exp\left(-\frac{\gamma_{sc}}{\gamma_{av}}\right)\right]^{N_r - 1} \qquad \gamma_{sc} \geq 0 \end{aligned} \tag{6.11}$$

For convenience of graphical presentation, we use the scaled probability density function

$$f_X(x) = \gamma_{av} f_{\Gamma_{sc}}(\gamma_{sc})$$

where the normalized variable x is defined by

$$x = \gamma_{sc}/\gamma_{av}$$

Figure 6.2 plots $f_X(x)$ versus x for a varying number of receive-diversity branches, N_r, under the assumption that the short-term signal-to-noise ratios for all the N_r branches share the common value γ_{av}. From the figure, we can make the following observations:

1. As the number of diversity branches, N_r, is increased, the probability density function $f_X(x)$ of the normalized random variable $X = \Gamma_{sc}/\gamma_{av}$ moves progressively to the right.
2. The probability density function $f_X(x)$ becomes more and more symmetrical, and Gaussian, as N_r is increased.

Stated another way, a frequency-flat, slowly fading Rayleigh channel is modified through the use of selection combining into a Gaussian channel, provided that the number N_r of diversity channels is sufficiently large. Realizing that a Gaussian channel is a "digital communication theorist's dream," we can now see the practical benefit of using selection combining.

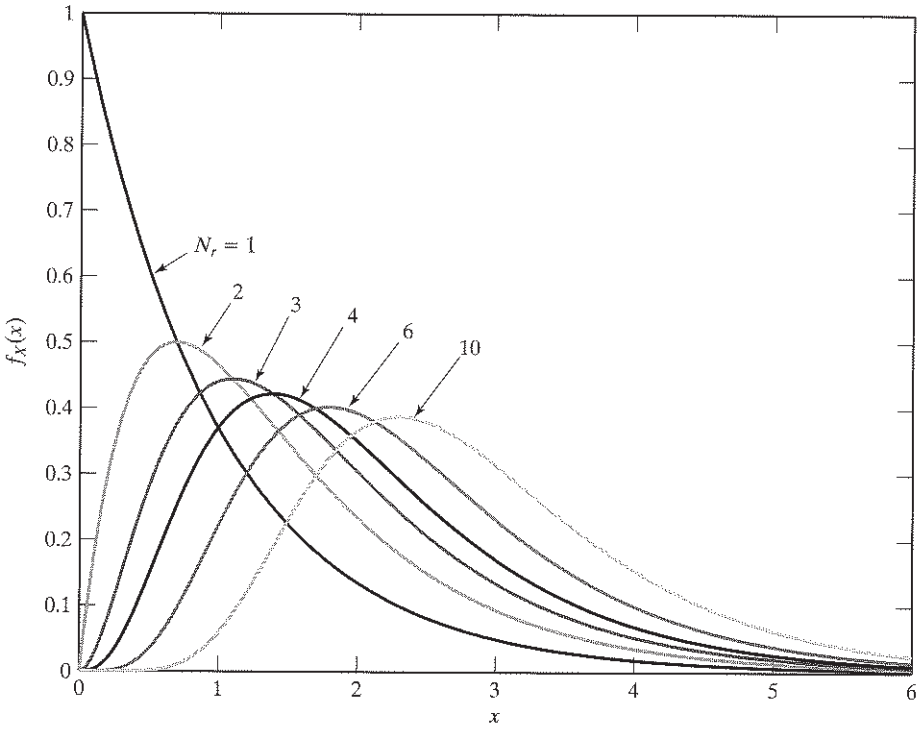


FIGURE 6.2 Normalized probability density function $f_X(x) = N_r \exp(-x)(1 - \exp(-x))^{N_r-1}$ for a varying number N_r of receive antennas.

According to the theory described herein, the selection-combining procedure requires that we monitor the receiver outputs in a continuous manner and, at each instant of time, select the receiver with the strongest signal (i.e., the largest instantaneous signal-to-noise ratio). From a practical perspective, such a selective procedure is rather cumbersome. We may overcome this practical difficulty by adopting a *scanning* version of the selection-combining procedure as follows:

- Start the procedure by selecting the receiver with the strongest output signal.
- Maintain the procedure by using the output of this particular receiver as the combiner's output, so long as its instantaneous signal-to-noise ratio does not drop below a prescribed threshold.
- As soon as the instantaneous signal-to-noise ratio of the combiner falls below the threshold, select a new receiver that offers the strongest output signal, and continue the procedure.

This technique has a performance similar to that of the nonscanning version of selective diversity.

EXAMPLE 6.1 Outage Probability of Selection Combiner

The *outage probability* of a diversity combiner is defined as *the percentage of time the instantaneous output signal-to-noise ratio of the combiner is below some prescribed level for a specified number of branches.* Using the cumulative distribution function of Eq. (6.10), Fig. 6.3 plots the outage curves for the selection combiner with N_r as the running parameter. The horizontal axis of the figure represents the instantaneous output signal-to-noise ratio of the combiner relative to 0 dB (i.e., the 50-percentile point for $N_r = 1$), and the vertical axis represents the outage probability, expressed as a percentage. From the figure, we observe that the fading depth introduced through the use of space diversity on receive diminishes rapidly with the increase in the number of diversity branches. ■

6.2.2 Maximal-Ratio Combining[2]

The selection-combining technique just described is relatively straightforward to implement. However, from a performance point of view, it is not optimum, in that it ignores the information available from all the diversity branches except for the particular branch that produces the largest instantaneous power of its own demodulated signal.

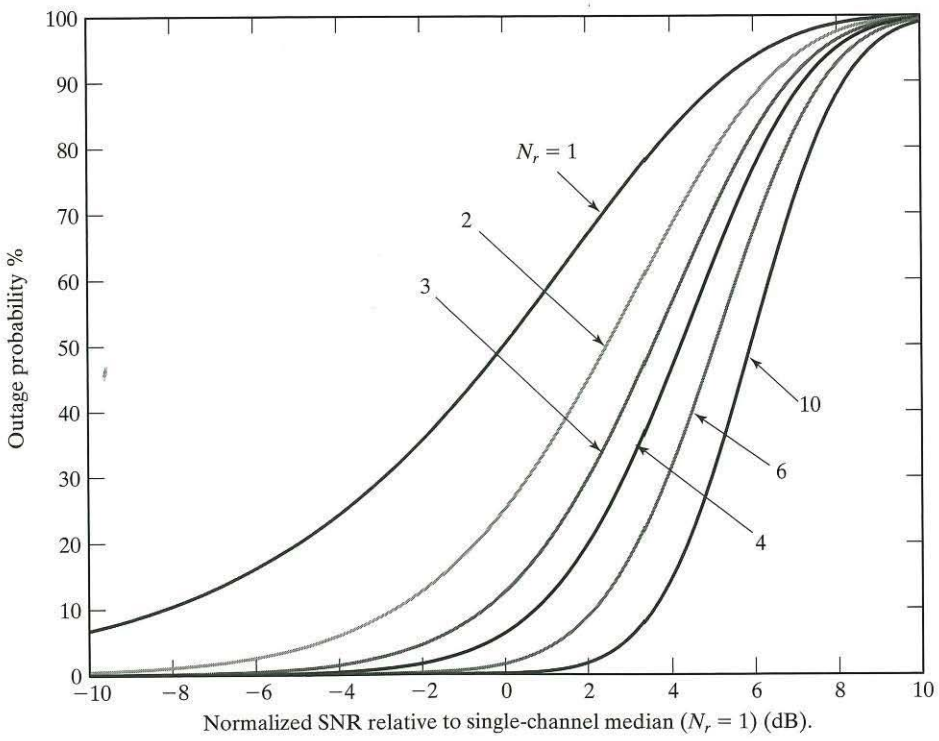


FIGURE 6.3 Outage probability for selector combining for a varying number N_r of receive antennas.

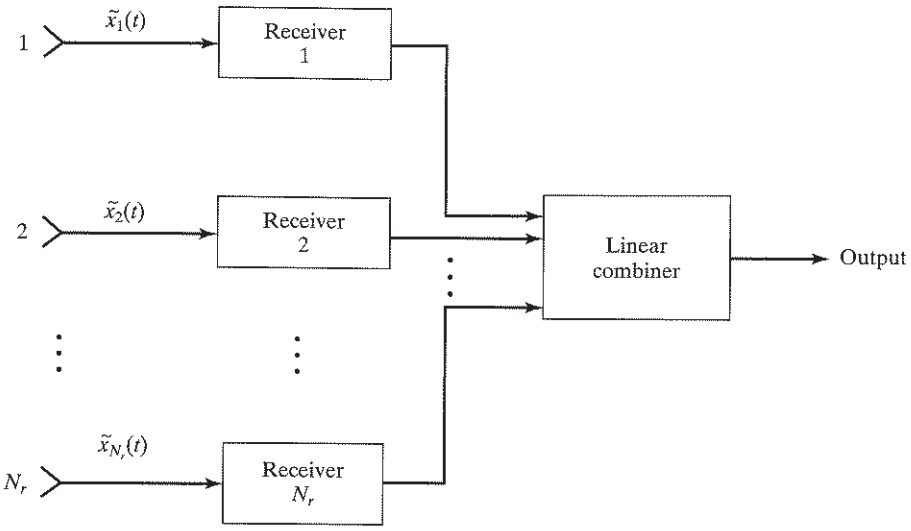


FIGURE 6.4 Block diagram of maximal-ratio combiner using N_r receive antennas.

This limitation of the selection combiner is mitigated by the *maximal-ratio combiner*, the composition of which is described by the block diagram of Fig. 6.4. The maximal-ratio combiner consists of N_r linear receivers, followed by a linear combiner. Using the complex envelope of the received signal at the kth diversity branch given in Eq. (6.1), we find that the corresponding complex envelope of the linear combiner output is defined by

$$\begin{aligned}\tilde{y}(t) &= \sum_{k=1}^{N_r} a_k \tilde{x}_k(t) \\ &= \sum_{k=1}^{N_r} a_k [\alpha_k e^{j\theta_k} \tilde{s}(t) + \tilde{w}_k(t)] \\ &= \tilde{s}(t) \sum_{k=1}^{N_r} a_k \alpha_k e^{j\theta_k} + \sum_{k=1}^{N_r} a_k \tilde{w}_k(t)\end{aligned} \tag{6.12}$$

where the a_k are *complex weighting parameters* that characterize the linear combiner. These parameters are changed from instant to instant in accordance with variations in signals in the N_r diversity branches over the short-term fading process. The requirement is to design the linear combiner so as to maximize its output signal-to-noise ratio at each instant of time. From Eq. (6.12), we note the following two points:

1. The complex envelope of the output signal equals $\tilde{s}(t) \sum_{k=1}^{N_r} a_k \alpha_k e^{j\theta_k}$.

2. The complex envelope of the output noise equals $\sum_{k=1}^{N_r} a_k \tilde{w}_k(t)$.

Assuming that the $\tilde{w}_k(t)$ are mutually independent for $k = 1,2,...,N_r$, the output signal-to-noise ratio of the linear combiner is therefore

$$
\begin{aligned}
(\mathrm{SNR})_c &= \frac{\mathbf{E}\left[\left|\tilde{s}(t)\sum_{k=1}^{N_r} a_k \alpha_k e^{j\theta_k}\right|^2\right]}{\mathbf{E}\left[\left|\sum_{k=1}^{N_r} a_k \tilde{w}_k(t)\right|^2\right]} \\
&= \frac{\mathbf{E}[|\tilde{s}(t)|^2]}{\mathbf{E}[|\tilde{w}_k(t)|^2]} \cdot \frac{\mathbf{E}\left[\left|\sum_{k=1}^{N_r} a_k \alpha_k e^{j\theta_k}\right|^2\right]}{\mathbf{E}\left[\sum_{k=1}^{N_r} |a_k|^2\right]} \\
&= \left(\frac{E}{N_0}\right) \frac{\mathbf{E}\left[\left|\sum_{k=1}^{N_r} a_k \alpha_k e^{j\theta_k}\right|^2\right]}{\mathbf{E}\left[\sum_{k=1}^{N_r} |a_k|^2\right]}
\end{aligned}
\tag{6.13}
$$

where E/N_0 is the *symbol energy-to-noise spectral density ratio.*

Let γ_c denote the *instantaneous output signal-to-noise ratio* of the linear combiner. Then, using

$$
\left|\sum_{k=1}^{N_r} a_k \alpha_k e^{j\theta_k}\right|^2 \text{ and } \sum_{k=1}^{N_r} |a_k|^2
$$

as the instantaneous values of the expectations in the numerator and denominator of Eq. (6.13), respectively, we may write

$$
\gamma_c = \left(\frac{E}{N_0}\right) \frac{\left|\sum_{k=1}^{N_r} a_k \alpha_k e^{j\theta_k}\right|^2}{\sum_{k=1}^{N_r} |a_k|^2} \tag{6.14}
$$

The requirement is to maximize γ_c with respect to the a_k. This maximization can be carried out by following the standard differentiation procedure, recognizing that the

weighting parameters a_k are complex. However, we choose to follow a simpler procedure based on the Cauchy–Schwarz inequality, as described next.

Let a_k and b_k denote any two complex numbers for $k = 1,2,...,N_r$. According to the *Cauchy–Schwarz inequality* for complex numbers, we have

$$\left|\sum_{k=1}^{N_r} a_k b_k\right|^2 \le \sum_{k=1}^{N_r} |a_k|^2 \sum_{k=1}^{N_r} |b_k|^2 \tag{6.15}$$

which holds with equality for $a_k = cb_k^*$, where c is some arbitrary complex constant and the asterisk denotes complex conjugation.

Thus, applying the Cauchy–Schwarz inequality to the instantaneous output signal-to-noise ratio of Eq. (6.14), with a_k left intact and b_k set equal to $\alpha_k e^{j\theta_k}$, we obtain

$$\gamma_c \le \left(\frac{E}{N_0}\right) \frac{\sum_{k=1}^{N_r} |a_k|^2 \sum_{k=1}^{N_r} \left|\alpha_k e^{j\theta_k}\right|^2}{\sum_{k=1}^{N_r} |a_k|^2} \tag{6.16}$$

Cancelling common terms in Eq. (6.16) readily yields

$$\gamma_c \le \left(\frac{E}{N_0}\right) \sum_{k=1}^{N_r} \alpha_k^2 \tag{6.17}$$

Equation (6.17) proves that, in general, γ_c cannot exceed $\sum_k \gamma_k$, where γ_k is as defined in Eq. (6.5). The equality in Eq. (6.17) holds for

$$\begin{aligned} a_k &= c(\alpha_k e^{j\theta_k})^* \\ &= c\alpha_k e^{-j\theta_k} \qquad k = 1, 2, \ldots, N_r \end{aligned} \tag{6.18}$$

where c is some arbitrary complex constant. Equation (6.18) defines the complex weighting parameters of the maximal-ratio combiner. On the basis of this equation, we may state that the optimal weighting factor a_k for the kth diversity branch has a magnitude proportional to the amplitude α_k of the signal and a phase that cancels the signal's phase θ_k to within some value that is identical for all the N_r diversity branches. The phase alignment just described has an important implication: It permits the *fully coherent addition* of the N_r receiver outputs by the linear combiner.

Equation (6.17) with the equality sign defines the instantaneous output signal-to-noise ratio of the maximal-ratio combiner, which is written as

$$\gamma_{\text{mrc}} = \left(\frac{E}{N_0}\right) \sum_{k=1}^{N_r} \alpha_k^2 \tag{6.19}$$

According to Eq. (6.5), however, $(E/N_0)\,\alpha_k^2$ is the instantaneous output signal-to-noise ratio of the kth diversity branch. Hence, the maximal-ratio combiner produces an instantaneous output signal-to-noise ratio that is the sum of the instantaneous signal-to-noise ratios of the individual branches; that is,

$$\gamma_{\text{mrc}} = \sum_{k=1}^{N_r} \gamma_k \tag{6.20}$$

The term "maximal-ratio combiner" has been coined to describe the combiner of Fig. 6.4 that produces the optimum result given in Eq. (6.20). Indeed, if follows from this result that the instantaneous output signal-to-noise ratio of the maximal-ratio combiner can be large even when the signal-to-noise ratios of the individual branches are small. The selection combiner of Section 6.2.1 is clearly inferior in performance to the maximal-ratio combiner, since the instantaneous signal-to-noise ratio produced by the selection combiner is simply the largest among the N_r terms of Eq. (6.20).

The maximal signal-to-noise ratio γ_{mrc} is the sample value of a random variable denoted by Γ_{mrc}. According to Eq. (6.19), γ_{mrc} is equal to the sum of N_r exponentially distributed random variables for a frequency-flat, slowly fading Rayleigh channel. From probability theory, the probability density function of such a sum is known to be *chi-square with* $2N_r$ *degrees of freedom* (see Appendix C); that is,

$$f_{\Gamma_{\text{mrc}}}(\gamma_{\text{mrc}}) = \frac{1}{(N_r\text{-}1)!}\,\frac{\gamma_{\text{mrc}}^{N_r\text{-}1}}{\gamma_{\text{av}}^{N_r}}\exp\left(-\frac{\gamma_{\text{mrc}}}{\gamma_{\text{av}}}\right) \tag{6.21}$$

Note that for $N_r = 1$, Eqs. (6.11) and (6.21) reduce to the same value, which is to be expected.

Figure 6.5 plots the scaled probability density function $f_X(x) = \gamma_{\text{av}} f_{\Gamma_{\text{mrc}}}(\gamma_{\text{mrc}})$ against the normalized variable $x = \gamma_{\text{mrc}}/\gamma_{\text{av}}$ for varying N_r. On basis of this figure, we may make observations similar to those for the selection combiner, except for the fact that, for any N_r, the scaled probability density function for the maximal-ratio combiner is radically different from that for the selection combiner.

EXAMPLE 6.2 Outage Probability for Maximal-Ratio Combiner

The cumulative distribution function for the maximal-ratio combiner is defined by

$$\begin{aligned}\text{Prob}(\gamma_{\text{mrc}} < x) &= \int_0^x f_{\Gamma_{\text{mrc}}}(\gamma_{\text{mrc}})\,d\gamma_{\text{mrc}} \\ &= 1 - \int_x^\infty f_{\Gamma_{\text{mrc}}}(\gamma_{\text{mrc}})\,d\gamma_{\text{mrc}}\end{aligned} \tag{6.22}$$

where the probability density function $f_{\Gamma_{\text{mrc}}}(\gamma_{\text{mrc}})$ is itself defined by Eq. (6.21). Using Eq. (6.22), Fig. 6.6 plots the outage probability for the maximal-ratio combiner with N_r as a running parameter. Comparing this figure with Fig. 6.3 for selection combining, we see that the outage-probability curves for these two diversity techniques are superficially similar. The *diversity gain*, defined as the savings in E/N_0 at a given bit error rate, provides a measure of the effectiveness of a diversity technique on an outage-probability basis. ■

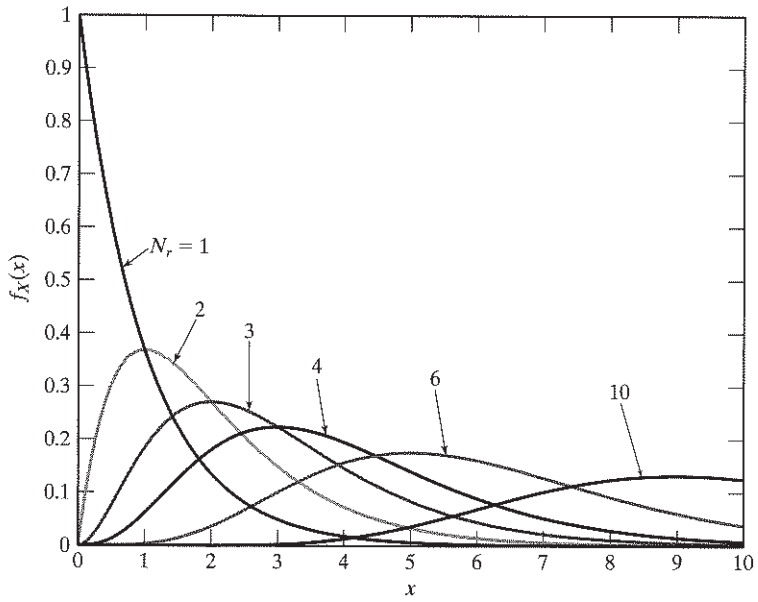


FIGURE 6.5 Normalized probability density function $f_X(x) = \frac{1}{(N_r\text{-}1)!} x^{N_r\text{-}1} \exp(-x)$ for a varying number N_r of receive antennas.

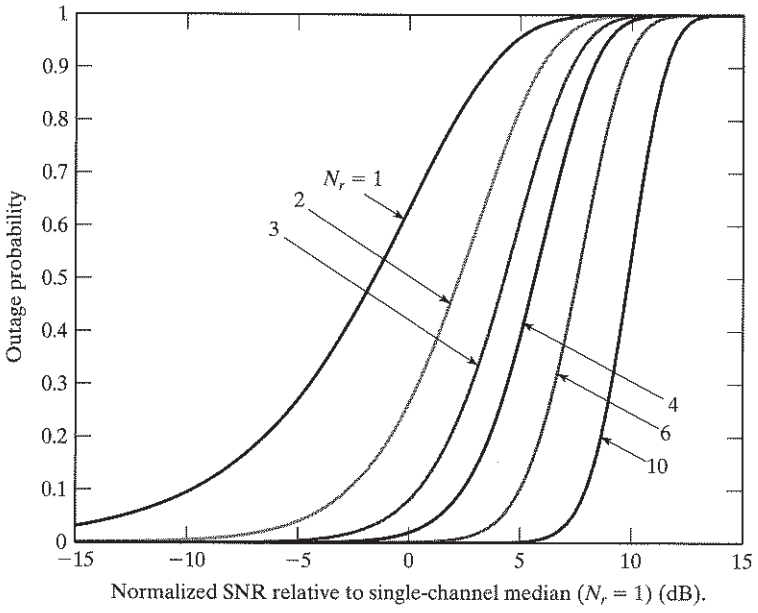


FIGURE 6.6 Outage probability of maximal-ratio combiner for a varying number N_r of receive antennas.

As a point of comparison of the outage performances of selection combining and maximal-ratio combining, consider a diversity-on-receive system with $N_r = 6$ and normalized output signal-to-noise ratio = 5dB. Examination of Figs. 6.3 and 6.6 for these settings reveals the following outage probabilities:

Selection combiner:	50%
Maximal-ratio combiner:	10%

These numbers clearly illustrate the highly superior outage performance of the maximal-ratio combiner over the selection combiner.

EXAMPLE 6.3 Bit Error Rate of Coherent Binary FSK

In this example, we determine the average probability of symbol error for the case of coherent binary frequency-shift keying (BFSK) over a frequency-flat, slowly fading Rayleigh channel. The use of maximal-ratio combining at the receiver is assumed. This simple case is amenable to an analytic formulation. (Note that the requirement for a coherent phase reference for coherent analytic evaluation detection makes the use of selection combining somewhat meaningless—hence the interest in only maximal-ratio combining in the context of coherent BFSK.)

Adapting the formula for the probability of symbol error for BFSK over an additive white Gaussian-noise channel for the problem at hand, we may write

$$\text{Prob}(\text{error}|\gamma_{\text{mrc}}) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{1}{2}\gamma_{\text{mrc}}}\right) \tag{6.23}$$

which is obtained by substituting γ_{mrc} for the signal energy-to-noise spectral density ratio E/N_0 in the formula for coherent BFSK in Table 3.4.

We next recognize that the instantaneous output signal-to-noise ratio γ_{mrc} is in fact a random variable. To determine the average probability of symbol error, we must average the conditional probability of error of Eq. (6.23) with respect to γ_{mrc}, or

$$P_e = \mathbf{E}[\text{Prob}(\text{errorr}|\gamma_{\text{mrc}})] \tag{6.24}$$

where the expectation also is with respect to γ_{mrc}. This expectation is found by multiplying the conditional probability $\text{Prob}(\text{error}|\gamma_{\text{mrc}})$ by the probability density function of γ_{mrc} and then integrating the product with respect to γ_{mrc}. That is, we write

$$P_e = \int_0^\infty \text{Prob}(\text{error}|\gamma_{\text{mrc}}) f_\Gamma(\gamma_{\text{mrc}}) d\gamma_{\text{mrc}} \tag{6.25}$$

Substituting Eqs. (6.21) and (6.23) into Eq. (6.25) yields

$$\begin{aligned} P_e &= \frac{1}{2(N_r\text{-}1)!}\int_0^\infty \text{erfc}\left(\sqrt{\frac{1}{2}\gamma_{\text{mrc}}}\right)\frac{\gamma_{\text{mrc}}^{N_r\text{-}1}}{\gamma_{\text{av}}^{N_r}}\exp\left(-\frac{\gamma_{\text{mrc}}}{\gamma_{\text{av}}}\right)d\gamma_{\text{mrc}} \\ &= \frac{1}{2(N_r\text{-}1)!}\int_0^\infty \text{erfc}\left(\sqrt{\frac{1}{2}\gamma_{\text{av}}x}\right)x^{N_r\text{-}1}\exp(-x)dx \end{aligned} \tag{6.26}$$

where $x = \gamma_{\text{mrc}}/\gamma_{\text{av}}$ and erfc(·) denotes the *complementary error function,* discussed in Appendix E. ■

Unfortunately, there is no exact closed-form solution of Eq. (6.26). To proceed further, we may use numerical integration or seek an approximate solution. The reader is referred to Problem 6.25 for the derivation of an approximate formula for P_e.

6.2.3 Equal-Gain Combining

In a theoretical context, the maximal-ratio combiner is the *optimum* among linear diversity-combining techniques, in the sense that it produces the largest possible value of instantaneous output signal-to-noise ratio. However, in practical terms, there are three important issues to keep in mind:[3]

1. Significant instrumentation is needed to adjust the complex weighting parameters of the maximal-ratio combiner to their exact values, in accordance with Eq. (6.18).
2. The additional improvement in output signal-to-noise ratio gained by the maximal-ratio combiner over the selection-combiner of Section 6.2.1 is not that large, and it is quite likely that the additional improvement in receiver performance is lost in the inability to achieve the exact setting of the maximal ratio combiner.
3. So long as a linear combiner uses the diversity branch with the strongest signal, other details of the combiner may result in a minor improvement in overall receiver performance.

Issue 3 points to the formulation of the so-called equal-gain combiner, in which all the complex weighting parameters a_k have their phase angles set opposite to those of their respective multipath branches in accordance with Eq. (6.18), but, unlike the a_k in the maximal-ratio combiner, their magnitudes are set equal to some constant value—unity, for convenience of use.

We may reach a similar conclusion by examining Eq. (6.16), in which we see that the summation term $\sum_k |a_k|^2$ involving the magnitudes of the complex weighting parameters is common to both the numerator and denominator of the instantaneous output signal-to-noise ratio. That is, whether we set $a_k = ca_k e^{-j\theta_k}$, in accordance with Eq. (6.18) for the maximal-ratio combiner, or we simply set $a_k = e^{-j\theta_k}$, for all k in the equal-gain combiner, the instantaneous output signal-to-noise ratio is unchanged and hence constitutes further justification for using the equal-gain combiner in preference to the maximal-ratio combiner.

6.2.4 Square-Law Combining

Maximal-ratio combining—and, for that matter, equal-gain combining—relies on the ability to estimate the phase of the different diversity branches and to combine the signals coherently. Often, such a procedure is not practical, due to the physical separation of the diversity receivers or to hardware limitations. In this case, square-law combining offers the opportunity to obtain an advantage in diversity without requiring phase estimation.

Unlike maximal-ratio combining, square-law combining is applicable only to certain modulation techniques. In particular, it is applicable to orthogonal modulation, including modulations such as FSK or direct-sequence CDMA signals, in which

different (approximately orthogonal) frequencies or sequences are used to represent different data symbols. With binary orthogonal signalling, the receiver generates the two decision variables

$$Q_{0k} = \frac{1}{\sqrt{N_o}} \int_0^T \tilde{x}_k(t)\tilde{s}_0^*(t)dt \tag{6.27}$$

and

$$Q_{1k} = \frac{1}{\sqrt{N_o}} \int_0^T \tilde{x}_k(t)\tilde{s}_1^*(t)dt \tag{6.28}$$

where $\tilde{s}_0(t)$ and $\tilde{s}_1(t)$ are normalized versions of the two possible binary symbols. In *orthogonal modulation,* the two signaling waveforms approximately satisfy the conditions

$$\int_0^T \tilde{s}_i(t)\tilde{s}_j^*(t)dt = \begin{cases} E_b & i = j \\ 0 & i \neq j \end{cases} \tag{6.29}$$

where E_b is the transmitted signal energy per bit. If the binary symbol 0 is transmitted, it turns out that the two decision variables are equivalent to

$$Q_{0k} = \frac{\sqrt{E_b}\alpha_k e^{j\theta_k}}{\sqrt{N_0}} + \frac{w_{0k}}{\sqrt{N_0}} \tag{6.30}$$

and

$$Q_{1k} = \frac{w_{1k}}{\sqrt{N_0}} \tag{6.31}$$

The w_{0k} and w_{1k} are independent Gaussian random variables of zero mean and variance N_0. The square-law receiver makes a decision between the binary symbols 0 and 1 as follows:

$$\text{If} \quad \begin{array}{ll} |Q_{0k}|^2 > |Q_{1k}|^2 & \text{say } 0 \\ \text{otherwise} & \text{say } 1 \end{array} \tag{6.32}$$

The structure of the receiver is illustrated in Fig. 6.7.

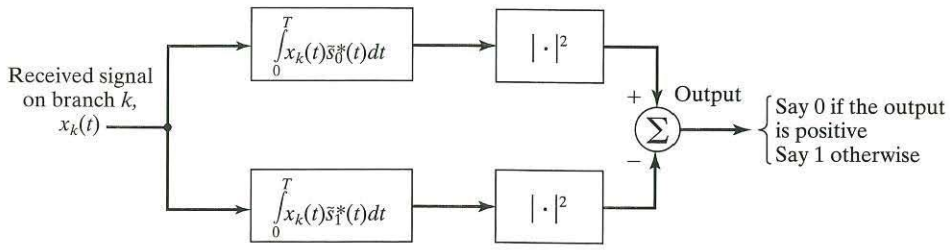


FIGURE 6.7 Illustration of receiver for orthogonal waveforms.

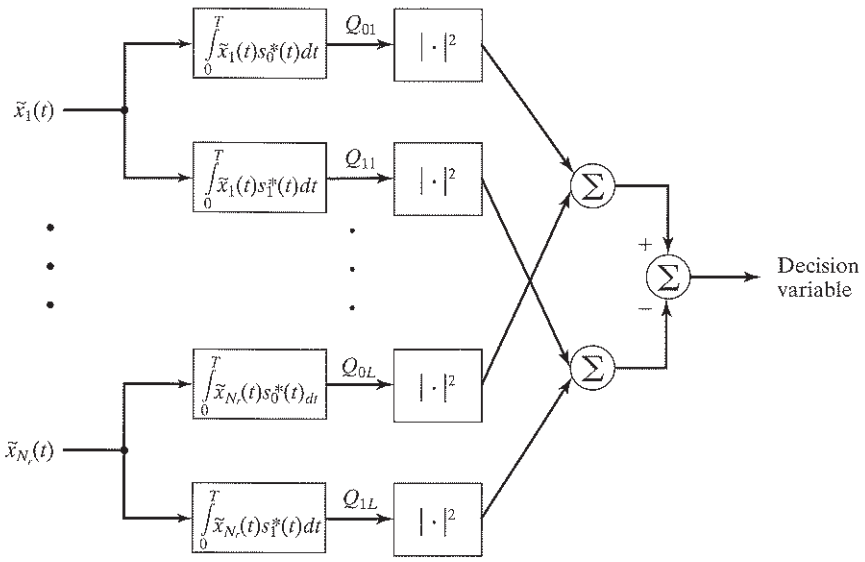


FIGURE 6.8 Block diagram of diversity receiver based on square-law combining.

With receive diversity, the detector outputs are added before being compared against the decision variables. This form of diversity, referred to as *square-law combining*, is illustrated in Fig. 6.8. With square-law combining, we form the decision variables

$$Q_0 = \sum_{k=1}^{N_r} |Q_{0k}|^2 \tag{6.33}$$

and

$$Q_1 = \sum_{k=1}^{N_r} |Q_{1k}|^2 \tag{6.34}$$

and then perform the same test as in Eq.(6.32) to decide whether a 0 or a 1 is transmitted.

Closer inspection of Eqs. (6.33) and (6.34) indicates that both Q_0 and Q_1 are the sum of squares of complex Gaussian random variables. Consequently, they both have a chi-square distribution with $2N_r$ degrees of freedom, analogous to Eq. (6.19). (The chi-square distribution is discussed in Appendix C.) The difference between the random variables Q_{0k} and Q_{1k} lies in their variance. If $s_0(t)$ was transmitted, then the variances are given by

$$\begin{aligned} \text{Var}(Q_{0k}) &= \frac{E_b}{N_0}\text{Var}(\alpha_k e^{j\theta_k}) + \frac{\text{Var}(w_{0k})}{N_0} \\ &= \frac{E_b}{N_0}(\mathbf{E}[\alpha_k^2]) + \frac{\mathbf{E}[w_{0k}^2]}{N_0} \\ &= \gamma_{\text{av}} + 1 \end{aligned} \tag{6.35}$$

and

$$\begin{aligned} \mathrm{Var}(Q_{1k}) &= \frac{1}{N_0}\mathrm{Var}(w_{1k}) \\ &= 1 \end{aligned} \tag{6.36}$$

where the fading is assumed to be Rayleigh distributed with average signal-to-noise ratio γ_{av}. In this case, the probability density function for the correct symbol is given by

$$f_{Q_0}(q) = \frac{1}{(N_r-1)!}\,\frac{q^{N_r-1}}{(\gamma_{\mathrm{av}}+1)^{N_r}}\exp\left(-\frac{q}{\gamma_{\mathrm{av}}+1}\right) \tag{6.37}$$

and that for the incorrect symbol is given by

$$f_{Q_1}(q) = \frac{1}{(N_r-1)!}q^{N_r-1}\exp(-q) \tag{6.38}$$

Equation (6.37) follows from Eq. (6.21), with γ_{av} replaced by $\gamma_{\mathrm{av}}+1$, in accordance with Eq. (6.35), under the assumption that $s_0(t)$ is transmitted. Similarly, Eq. (6.38) follows from Eq. (6.21), with γ_{av} set equal to unity, in accordance with Eq. (6.36), under the assumption that $s_1(t)$ is transmitted.

Since Q_0 and Q_1 are independent random variables, the probability of error is the probability that $Q_0 < Q_1$ when symbol 0 is transmitted. Mathematically, this probability is obtained from the double integral

$$\mathrm{Prob}(Q_0 < Q_1) = \int_0^\infty f_{Q_0}(q_0)\left(\int_{q_o}^\infty f_{Q_1}(q_1)dq_1\right)dq_0 \tag{6.39}$$

A similar result holds when a 1 is transmitted.

In Fig. 6.9, we compare the probability of error for coherent BFSK with square-law combining to maximal-ratio combining for diversity orders 1, 2, and 4 with binary signaling over a Rayleigh-fading channel. With BFSK, maximal-ratio combining offers approximately a 3-dB improvement over square-law combining, a performance difference that is quite small compared with the overall gains that either diversity technique usually provides.

Problem 6.2 Using Eq. (6.29) in (6.28), show that Eq. (6.31) follows, within a scaling factor. Hence, show that Q_{0k} and Q_{1k} are independent random variables. ■

Problem 6.3 For a diversity $N_r = 2$ system and $\gamma_{\mathrm{av}} = 20$ dB, compute the fraction of the time the decision variable Q_0 drops below 10 dB when 0 has been transmitted. What is the probability that Q_1 is greater than 10 dB under the same conditions?

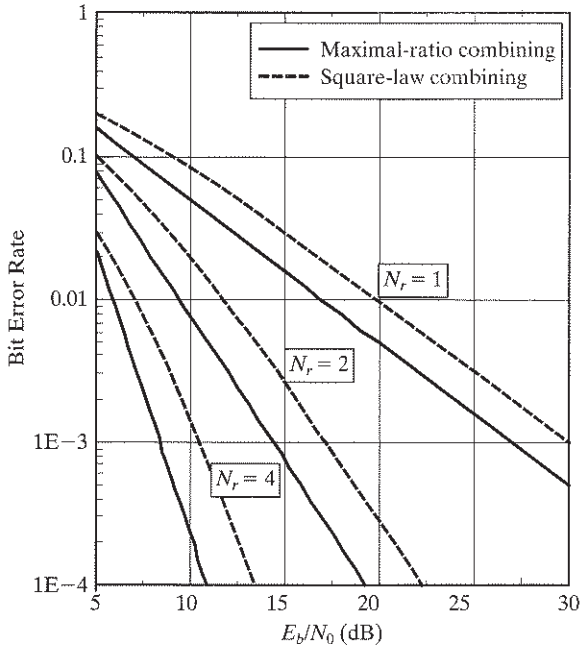


FIGURE 6.9 Comparison of diversity performance: square-law combining versus maximal-ratio combining.

Ans. Define the incomplete gamma function as

$$\Gamma(x, a) = \frac{1}{\Gamma(a)}\int_0^x s^{a-1}\exp(-s)ds$$

where

$$\Gamma(a) = \int_0^\infty s^{a-1}\exp(-s)ds$$

and $\Gamma(a) = (a-1)!$ *if a is a positive integer. Then show*

$$P(Q_0 < 10\text{dB}) = \Gamma(10/101, 2) = 0.0046$$

and

$$P(Q_1 > 10\text{dB}) = 1 - \Gamma(10, 2) = 5 \times 10^{-4}$$ ■

6.3 MULTIPLE-INPUT, MULTIPLE-OUTPUT ANTENNA SYSTEMS [4]

In Section 6.2, we studied space-diversity wireless communication systems employing multiple receive antennas to combat the multipath fading problem. In effect, fading was treated as a source that degrades performance, necessitating the use of space diversity on receive to mitigate it. In this section, we discuss *multiple-input, multiple-output (MIMO) wireless communications*, also referred to in the literature as *multiple-transmit, multiple-receive (MTMR) wireless communications.*

MIMO wireless communications include space diversity on receive as a special case. Most important, however, are the following three points:

1. The fading phenomenon is viewed not as a nuisance, but rather as an environmental source of possible enrichment.
2. Space diversity at both the transmit and receive ends of the wireless communications link provides the basis for a significant increase in channel capacity or spectral efficiency.
3. Unlike increasing capacity with conventional techniques, increasing channel capacity with MIMO is achieved by increasing computational complexity while maintaining the primary communication resources (i.e., total transmit power and channel bandwidth) fixed.

6.3.1 Coantenna Interference

Figure 6.10 shows the block diagram of a MIMO wireless link. The signals transmitted by the N_t transmit antennas over the wireless channel are all chosen to lie inside a common frequency band. Naturally, the transmitted signals are scattered differently by the channel. Moreover, due to multiple signal transmissions, the system experiences a spatial form of signal-dependent interference referred to as *coantenna interference (CAI)*.

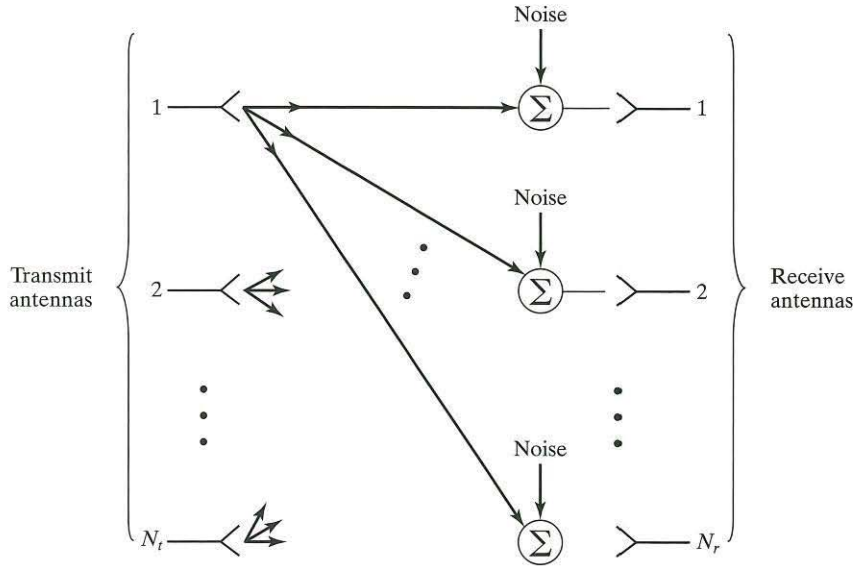


FIGURE 6.10 Block diagram of MIMO wireless link with N_t transmit antennas and N_r receive antennas.

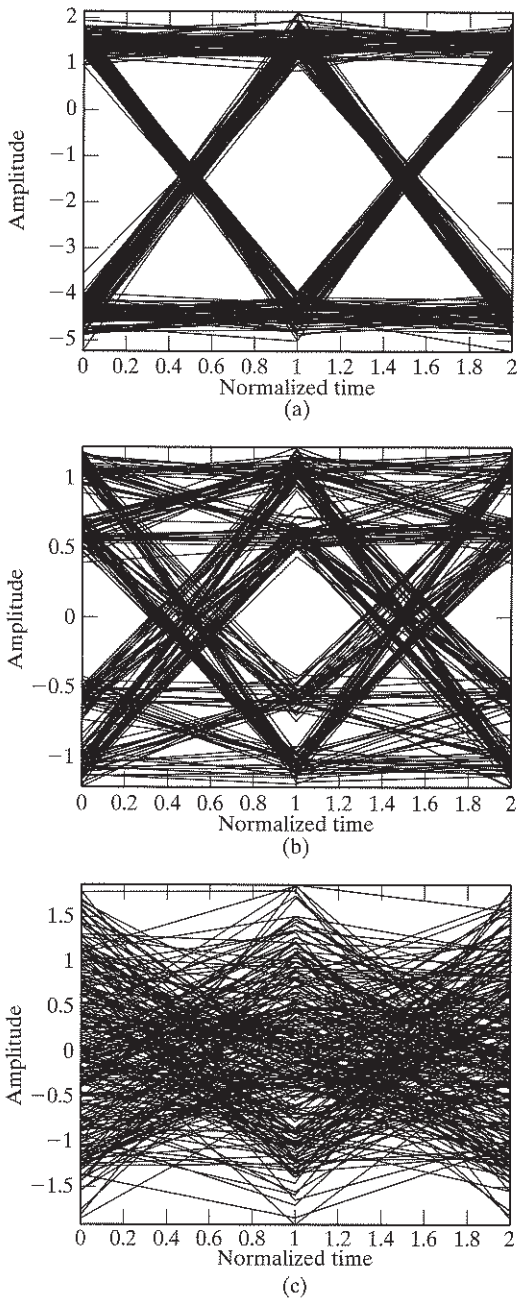


FIGURE 6.11 Effect of coantenna interference on the eye diagram for one receive antenna and different numbers of transmit antennas. (a) $N_t = 1$, (b) $N_t = 2$, (c) $N_t = 8$.

Figure 6.11 illustrates the effect of CAI for one, two, and eight simultaneous transmissions and a single receive antenna (i.e., $N_t = 1, 2, 8$ and $N_r = 1$), using BPSK; the transmitted BPSK signals are different, but they all have the same average power and occupy the same bandwidth. The figure clearly shows the difficulty that arises from CAI when the number N_t of transmit antennas is large. In particular, with eight simultaneous signal transmissions, the *eye pattern* of the received signal is practically closed. The pattern, commonly used in the study and design of digital communications, derives its name from the fact that it resembles the human eye for the transmission of binary data; the interior region of the eye pattern is called the *eye opening*. The challenge for the receiver is how to mitigate the CAI problem and thereby make it possible to provide a spectacular increase in spectral efficiency.

In a theoretical context, the spectral efficiency of a communication system is intimately linked to the channel capacity of the system. To proceed with evaluation of the channel capacity of MIMO wireless communications, we begin by formulating a baseband channel model for the system.

6.3.2 Basic Baseband Channel Model

Consider a MIMO narrowband wireless communication system built around a flat-fading channel and with N_t transmit antennas and N_r receive antennas. The antenna configuration is hereafter referred to as the pair (N_t, N_r). For a statistical analysis of the MIMO system in what follows, we use baseband representations of the transmitted and received signals, as well as of the channel. In particular, we introduce the following notation:

- The spatial parameter

$$N = \min\{N_t, N_r\} \tag{6.40}$$

  defines a *new degree of freedom* introduced into the wireless communication system by using a MIMO channel with N_t transmit antennas and N_t receive antennas.

- The N_t-by-1 vector

$$\mathbf{s}(n) = [\tilde{s}_1(n), \tilde{s}_2(n), \ldots, \tilde{s}_{N_t}(n)]^T \tag{6.41}$$

  denotes the complex signal vector transmitted by the N_t antennas at discrete time n. The symbols constituting the vector $\mathbf{s}(n)$ are assumed to have zero mean and common variance σ_s^2. The *total* transmit power is *fixed* at the value

$$P = N_t \sigma_s^2 \tag{6.42}$$

  For P to be maintained constant, the variance σ_s^2 (i.e., the power radiated by each transmit antenna) must be inversely proportional to N_t.

- For the flat-fading, and therefore memoryless, channel, we may use $\tilde{h}_{ik}(n)$ to denote the sampled complex gain of the channel from transmit antenna k to

receive antenna i at discrete time n, where $i = 1,2,...,N_r$ and $k = 1,2,...,N_t$. We may thus express the N_r-by-N_t *complex channel matrix* as

$$\mathbf{H}(n) = \underbrace{\left.\begin{bmatrix} \tilde{h}_{11}(n) & \tilde{h}_{21}(n) & \dots & \tilde{h}_{N_t1}(n) \\ \tilde{h}_{21}(n) & \tilde{h}_{22}(n) & \dots & \tilde{h}_{N_t2}(n) \\ \vdots & \vdots & & \vdots \\ \tilde{h}_{N_r1}(n) & \tilde{h}_{N_r2} & \dots & h_{N_rN_t}(n) \end{bmatrix}\right\}}_{N_t \text{ transmit antennas}} \begin{matrix} N_r \\ \text{receive} \\ \text{antennas} \end{matrix} \tag{6.43}$$

- The system of equations

$$\tilde{x}_i(n) = \sum_{k=1}^{N_t} \tilde{h}_{ik}(n)\tilde{s}_k(n) + \tilde{w}_i(n) \qquad \begin{matrix} i = 1, 2, \dots, N_r \\ k = 1, 2, \dots, N_t \end{matrix} \tag{6.44}$$

defines the complex signal received at the ith antenna due to the transmitted symbol $\tilde{s}_k(n)$ radiated by the kth antenna. The term $w_i(n)$ denotes the additive complex channel noise perturbing $x_i(n)$. Let the N_r-by-1 vector

$$\mathbf{x}(n) = [\tilde{x}_1(n), \tilde{x}_2(n), \dots, \tilde{x}_{N_r}(n)] \tag{6.45}$$

denote the complex received signal vector and the N_r-by-1 vector

$$\mathbf{w}(n) = [\tilde{w}_1(n), \tilde{w}_2(n), \dots, \tilde{w}_{N_r}(n)]^T \tag{6.46}$$

denote the complex channel noise vector. We may then rewrite the system of equations (6.44) in the compact matrix form

$$\mathbf{x}(n) = \mathbf{H}(n)\mathbf{s}(n) + \mathbf{w}(n) \tag{6.47}$$

Equation (6.47) describes the *basic complex channel model for MIMO wireless communications*, assuming the use of a flat-fading channel. The equation describes the input–output behavior of the channel at discrete time n. To simplify the exposition, hereafter we suppress the dependence on time n by writing

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{w} \tag{6.48}$$

where it is understood that all four vector–matrix terms of the equation, namely, **s**, **H**, **w**, and **x**, are in fact dependent on the discrete time n. Figure 6.12 depicts the basic channel model of Eq. (6.48).

For mathematical tractability, we assume a *Gaussian model* made up of three elements relating to the transmitter, channel, and receiver, respectively:

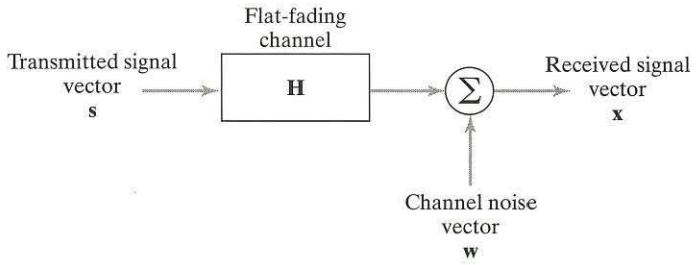


FIGURE 6.12 Depiction of the basic channel model of Eq. (6.48).

1. The N_t symbols constituting the transmitted signal vector $\mathbf{s}$ are drawn from a *white complex Gaussian codebook*; that is, the symbols $\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_{N_t}$ are independently and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and common variance σ_s^2. Hence, the correlation matrix of the transmitted signal vector $\mathbf{s}$ is defined by

$$\begin{aligned}\mathbf{R}_{\mathbf{s}} &= \mathbf{E}[\mathbf{s}\mathbf{s}^{\dagger}] \\ &= \sigma_s^2 \mathbf{I}_{N_t}\end{aligned} \tag{6.49}$$

where $\mathbf{I}_{N_t}$ is the N_t-by-N_t identity matrix.

2. The $N_t \times N_r$ elements of the channel matrix $\mathbf{H}$ are drawn from an ensemble of i.i.d. complex random variables with zero mean and unit variance, as shown by the complex distribution

$$h_{ik}: \quad N(0, 1/\sqrt{2}) + jN(0, 1/\sqrt{2}) \qquad \begin{matrix} i = 1, 2, \ldots, N_r \\ k = 1, 2, \ldots, N_t \end{matrix} \tag{6.50}$$

where $\mathcal{N}(.,.)$ denotes a real Gaussian distribution. On this basis, we find that the amplitude component h_{ik} is *Rayleigh* distributed, so we sometimes speak of the MIMO channel as a *rich Rayleigh scattering environment*. By the same token, we also find that the squared amplitude component, namely, $|h_{ik}|^2$, is a *chi-square random variable* with mean

$$\mathbf{E}[|h_{ik}|^2] = 1 \qquad \text{for all } i \text{ and } k \tag{6.51}$$

3. The N_r elements of the channel noise vector $\mathbf{w}$ are i.i.d. complex Gaussian random variables with zero mean and common variance $\sigma_{\mathbf{w}}^2$; that is, the correlation matrix of the noise vector $\mathbf{w}$ is given by

$$\begin{aligned}\mathbf{R}_{\mathbf{w}} &= \mathbf{E}[\mathbf{w}\mathbf{w}^{\dagger}] \\ &= \sigma_{\mathbf{w}}^2 \mathbf{I}_{N_r}\end{aligned} \tag{6.52}$$

where $\mathbf{I}_{N_r}$ is the N_r-by-N_r identity matrix.

In light of Eq. (6.42) and the assumption that h_{ik} is a normalized random variable with zero mean and unit variance, the *average signal-to-noise ratio* (SNR) at each receiver input is given by

$$\begin{aligned}\rho &= \frac{P}{\sigma_{\mathbf{w}}^2} \\ &= \frac{N_t \sigma_{\mathbf{s}}^2}{\sigma_{\mathbf{w}}^2}\end{aligned} \tag{6.53}$$

which, for a prescribed noise variance $\sigma_{\mathbf{w}}^2$, is fixed once the total transmit power P is fixed. Note also that (1) all the N_t transmitted signals occupy a common channel bandwidth and (2) the SNR ρ is independent of N_r.

The idealized Gaussian model described herein is applicable to indoor local area networks and other wireless environments where the mobility of the user's terminals is limited. The model, however, ignores the unavoidable ambient noise, which, as a result of experimental measurements, is known to be decidedly non-Gaussian due to the *impulse nature* of human-made electromagnetic interference as well as natural noise.[5]

6.4 MIMO CAPACITY FOR CHANNEL KNOWN AT THE RECEIVER

With the basic complex channel model at hand, we are now ready to focus the discussion on the primary issue of interest: the channel capacity of a MIMO wireless link. Two cases will be considered. The first case, discussed in Section 6.4.1, considers a link that is stationary and therefore ergodic. The second case, presented in Section 6.4.2, considers a nonergodic link, assuming quasi-stationarity from one data burst to another.

6.4.1 Ergodic Capacity

In Section 4.3, we stated that the information capacity of a *real* additive white Gaussian noise (AWGN) channel, subject to the constraint of a fixed transmit power P, is defined by

$$C = B \log_2\left(1 + \frac{P}{\sigma_w^2}\right) \text{ bits/s} \tag{6.54}$$

where B is the channel bandwidth and σ_w^2 is the noise variance measured over B.

Given a time-invariant channel, Eq. (6.54) defines the maximum data rate that can be transmitted over the channel with an arbitrarily small probability of error incurred as a result of the transmission. With the channel used K times for the transmission of K symbols in, say, T seconds, the transmission capacity per unit time is (T/K) times the formula for C given in Eq. (6.54). Recognizing that $K = 2BT$, in accordance

with the sampling theorem discussed in Chapter 4, we may express the information capacity of the AWGN channel in the equivalent form

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma_w^2}\right) \text{ bits/s/Hz} \tag{6.55}$$

Note that 1 bit per second per hertz corresponds to 1 bit per transmission.

With wireless communications as the medium of interest, consider next the case of a *complex*, flat-fading channel with the receiver having perfect knowledge of the channel state. The capacity of such a channel is given by[6]

$$C = \mathbf{E}\left[\log_2\left(1 + \frac{|h|^2 P}{\sigma_w^2}\right)\right] \text{ bits/s/Hz} \tag{6.56}$$

where the expectation is taken over the gain $h(n)$ of the channel, and the channel is assumed to be stationary and ergodic. In recognition of this assumption, C is commonly referred to as the *ergodic capacity* of the flat-fading channel, and the channel coding is applied across fading intervals (i.e., over an "ergodic" interval of channel variation with time).

It is important to note that the scaling factor of 1/2 is missing from the capacity formula of Eq. (6.56). The reason for this omission is the fact that that equation refers to a complex baseband channel, whereas Eq. (6.54) refers to a real channel. The fading channel covered by Eq. (6.56) operates on a complex signal—a signal with in-phase and quadrature components. Therefore, such a complex channel is equivalent to two real channels with equal capacities and operating in parallel—hence the result presented in that equation.

Equation (6.56) applies to the simple case of a *single-input, single-output (SISO) flat-fading channel*. Generalizing this formula to the case of a multiple-input, multiple-output (MIMO) flat-fading channel, governed by the Gaussian model described in Section 6.3.1, we find that the ergodic capacity of the MIMO channel is given by

$$C = \mathbf{E}\left[\log_2\left\{\frac{\det(\mathbf{R}_\mathbf{w} + \mathbf{H}\mathbf{R}_\mathbf{s}\mathbf{H}^\dagger)}{\det(\mathbf{R}_\mathbf{w})}\right\}\right] \text{ bits/s/Hz} \tag{6.57}$$

which is subject to the constraint

$$\max_{\mathbf{R}_\mathbf{s}} \operatorname{tr}(\mathbf{R}_\mathbf{s}) \le P$$

where P is the constant transmit power. The expectation in Eq. (6.57) is over the random channel matrix $\mathbf{H}$, and the superscript $\dagger$ denotes Hermitian transposition; $\mathbf{R}_\mathbf{s}$ and $\mathbf{R}_\mathbf{w}$ are, respectively, the correlation matrices of the transmitted signal vector $\mathbf{s}$ and channel noise vector $\mathbf{w}$. A detailed derivation of Eq. (6.57) is presented in Appendix G.

In general, it is difficult to evaluate Eq. (6.57), except for the Gaussian model described in Section 6.5. In particular, substituting Eqs. (6.49) and (6.52) into Eq. (6.57) and simplifying yields

$$C = \mathbf{E}\left[\log_2\left\{\det\left(\mathbf{I}_{N_r} + \frac{\sigma_s^2}{\sigma_w^2}\mathbf{H}\mathbf{H}^\dagger\right)\right\}\right] \text{ bits/s/Hz} \tag{6.58}$$

Invoking the definition of the average signal-to-noise ratio ρ introduced in Eq. (6.53), we may rewrite Eq. (6.58) in the equivalent form

$$C = \mathbf{E}\left[\log_2\left\{\det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t}\mathbf{H}\mathbf{H}^\dagger\right)\right\}\right] \text{ bits/s/Hz} \tag{6.59}$$

Equation (6.59), defining the ergodic capacity of a MIMO flat-fading channel, involves the determinant of an N_r-by-N_r sum matrix followed by the logarithm to the base 2. Accordingly, we refer to this formula as the *log-det capacity formula* for a Gaussian MIMO channel.[7]

Problem 6.4 The log-det capacity formula of Eq. (6.59) assumes that $N_t \geq N_r$ for the N_r-by-N_r matrix product $\mathbf{H}\mathbf{H}^\dagger$ to be of full rank. Show that, for the alternative case, $N_t \leq N_r$, which makes the N_t-by-N_t matrix product $\mathbf{H}^\dagger\mathbf{H}$ to be of full rank, the log-det capacity formula of the MIMO link is defined by

$$C = \mathbf{E}\left[\log_2\left\{\det\left(\mathbf{I}_{N_t} + \frac{\rho}{N_r}\mathbf{H}^\dagger\mathbf{H}\right)\right\}\right] \text{ bit/s/Hz} \tag{6.60}$$

where, as before, the expectation is over the channel matrix $\mathbf{H}$. ■

Note that Eqs. (6.59) and (6.60) are equivalent, in that either one of them applies to all $\{N_t, N_r\}$ antenna configurations. The two formulas differentiate themselves only when the full-rank issue is of concern, as explained in Problem 6.4.

Clearly, Eq. (6.56), pertaining to a conventional flat-fading link with a single antenna at both ends, is a special case of the log-det capacity formula. Specifically, for $N_t = N_r = 1$ (i.e., no spatial diversity), $\rho = P/\sigma_w^2$, and $\mathbf{H} = h$ (with dependence on discrete time n suppressed as stated on page 361), Eq. (6.58) reduces to Eq. (6.56).

Another insightful result that follows from the log-det capacity formula is that if $N_t = N_r = N$, then as N approaches infinity, the capacity C defined in Eq. (6.58) grows asymptotically (at least) linearly with N; see Problem 6.32. That is,

$$\lim_{N \to \infty} \frac{C}{N} \geq \text{constant} \tag{6.61}$$

The asymptotic formula of Eq. (6.61) may be stated in words as follows:

> *The ergodic capacity of a MIMO flat-fading wireless link with an equal number N of transmit and receive antennas grows roughly proportionately with N.*

What this statement teaches us is that, by increasing the computational complexity resulting from the use of multiple antennas at both the transmit and receive ends of a wireless link, we are able to increase the spectral efficiency of the link in a far greater manner than is possible by conventional means (e.g., increasing the transmit signal-to-noise ratio). The potential for this very sizable increase in the spectral efficiency of a MIMO wireless communication system is attributed to the key parameter $N = \min\{N_t, N_r\}$, which, in accordance with Eq. (6.40), defines the *number of spatial degrees of freedom* provided by the system. (Later in Section 6.10, we show that N is equal to the maximal multiplexing gain.)

Problem 6.5 Show that, at high signal-to-noise ratios, the capacity gain of a MIMO wireless communication system with the channel state known to the receiver is $N = \min\{N_t, N_r\}$ bits per second per hertz for every 3-dB increase in signal-to-noise ratio. ■

6.4.2 Two Other Special Cases of the Log-Det Formula: Capacities of Receive and Transmit Diversity Links

Naturally, the log-det capacity formula of Eq. (6.59) for the channel capacity of an (N_t, N_r) wireless link includes the channel capacities of receive and transmit diversity links as special cases:

1. *Diversity-on-receive channel.* The log-det capacity formula of Eq. (6.60) applies to this case. Specifically, for $N_t = 1$, the channel matrix $\mathbf{H}$ reduces to a column vector, and with it, Eq. (6.60) reduces to

$$C = \mathbf{E}\left[\log_2\left\{\left(1 + \rho \sum_{i=1}^{N_r} |h_i|^2\right)\right\}\right] \text{ bits/s/Hz} \tag{6.62}$$

Compared with the channel capacity of Eq. (6.56), for a single-input, single-output fading channel with $\rho = P/\sigma_w^2$, the squared channel gain $|h|^2$ is replaced by the sum of squared amplitudes $|h_i|^2$, $i = 1,2,...,N_r$. Equation (6.62) expresses the ergodic capacity due to the *linear combination* of the receive-antenna outputs, which is designed to maximize the information contained in the N_r received signals about the transmitted signal. This is simply a restatement of the maximal-ratio combining principle discussed in Section 6.2.

Problem 6.6 As pointed out previously, the selection combiner is a special case of the maximal-ratio combiner. What is the channel capacity of a wireless diversity channel using the selection combiner?

Ans. $\mathbf{E}[\log_2(1 + \rho h^2_{\max})]$ *where* $h_{\max} = \max\{|h_i|\}_{i=1}^{N_r}$ ■

2. *Diversity-on-transmit channel.* The log-det capacity formula of Eq. (6.59) applies to this second case. Specifically, for $N_r = 1$, the channel matrix $\mathbf{H}$ reduces to a row

vector, and with it, Eq. (6.59) reduces to

$$C = \mathbf{E}\left[\log_2\left(1 + \frac{\rho}{N_t}\sum_{k=1}^{N_t}|h_k|^2\right)\right] \text{ bits/s/Hz} \tag{6.63}$$

where the matrix product $\mathbf{H}\mathbf{H}^\dagger$ is replaced by the sum of squared amplitudes $|h_k|^2$, $k = 1,2,...,N_t$. Compared with Case 1 on receive diversity, the capacity of the diversity-on-transmit channel is reduced because the total transmit power is held constant, independently of the number of N_t transmit antennas.

6.4.3 Outage Capacity

To realize the log-det capacity formula of Eq. (6.59), the MIMO channel code needs to see an *ergodic process* of the random-channel processes, which, in turn, results in a hardening of the rate of reliable transmission to the $\mathbf{E}[\log_2\{\det(\cdot)\}]$ information rate (i.e., the channel capacity approaches the log-det formula). As in all information-theoretic arguments, the bit error rate would go to zero asymptotically in the block length of the code, thereby entailing a long transportation delay from the sender to the sink. In practice, however, the MIMO wireless channel is often nonergodic, and the requirement is to operate the channel under *delay constraints*. The issue of interest is then summed up as follows:

> *How much information can be transmitted across a nonergodic channel, particularly if the channel code is long enough to see just one random-channel matrix?*

In the situation described here, the rate of reliable information transmission (i.e., the strict Shannon-sense capacity) is zero, since, for any positive rate, the probability that the channel would not support such a rate is nonzero.

To get around this serious conceptual difficulty, the notion of *outage* is introduced into the characterization of the MIMO link. (Outage was discussed in Section 6.2 in the context of diversity-on-receive.) Specifically, *the outage probability* of a MIMO link is defined as *the probability for which the link is in a state of outage (i.e., failure) for data transmitted across the link at a certain rate R*, measured in bits per second per hertz. To proceed on this probabilistic basis, it is customary to operate the MIMO link by transmitting data in the form of bursts or frames, invoking a *quasi-static model* governed by four points:

1. The burst is *long* enough to accommodate the transmission of a large number of symbols, which, in turn, permits the use of an idealized *infinite-time horizon* that is basic to information theory.
2. Yet the burst is *short* enough that the wireless link can be treated as *quasi static* during each burst; the slow variation is used to justify the assumption that the receiver can acquire perfect knowledge of the channel state.
3. The channel matrix is permitted to change, say, from burst k to the next burst, $k + 1$, thereby accounting for statistical variations of the link.

4. The different realizations of the transmitted signal vector $\mathbf{s}$ are drawn from a *white Gaussian codebook*; that is, the correlation matrix of $\mathbf{s}$ is defined by Eq. (6.49).

Points 1 and 4 pertain to signal transmission, while points 2 and 3 pertain to the channel.

To proceed with the evaluation of outage probability, we first note that points 1 through 4 of the stochastic model just described for a nonstationary wireless link permit us to build on some of the results discussed in Section 6.4.1. In particular, in light of the log-det capacity formula of Eq. (6.59), we may view the random variable

$$C_k = \log_2\left\{\det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t}\mathbf{H}_k\mathbf{H}_k^{\dagger}\right)\right\} \text{ bits/s/Hz for burst } k \tag{6.64}$$

as the expression for a sample of the wireless link. In other words, with the random-channel matrix $\mathbf{H}_k$ varying from one burst to the next, C_k will itself vary in a corresponding way. A consequence of this random behavior is that, occasionally, a draw from the cumulative distribution function of the wireless link results in a value for C_k that is inadequate to support reliable communication over the link, in which case the link is said to be in an *outage* state. Correspondingly, for a given transmission strategy, we define the *outage probability at rate R* as

$$P_{\text{outage}}(R) = \text{Prob}\{C_k < R \text{ for some burst } k\} \tag{6.65}$$

or, equivalently,

$$P_{\text{outage}}(R) = \text{Prob}\left\{\log_2\left\{\det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t}\mathbf{H}_k\mathbf{H}_k^{\dagger}\right)\right\} < R \text{ for some burst } k\right\} \tag{6.66}$$

On this basis, we may define the *outage capacity* of the MIMO link as the *maximum bit rate that can be maintained across the link for all bursts of data transmissions (i.e., all possible channel states) for a prescribed outage probability.*

Problem 6.7 To calculate the outage probability, we use the complementary cumulative distribution function of the random-channel matrix $\mathbf{H}$, rather than the cumulative probability function itself. Explain. (By definition, the *complementary cumulative distribution function* (ccdf) is equal to unity minus the cumulative distribution function.) ■

EXAMPLE 6.4 Outage Capacity for Different Antenna Configurations and Varying Signal-to-Noise Ratios

In light of the random nature of the channel matrix $\mathbf{H}$, the outage capacity is evaluated with the use of Monte Carlo simulation by computing the cumulative distribution function of the wireless link for a large number of statistically different realizations of $\mathbf{H}$. To illustrate the simulation procedure, suppose we wish to calculate the outage capacity $C_{15\%}$ for error-free transmission for 100 – 15 = 85 percent of the time. The calculation is performed for a (2,2) antenna configuration operating at a signal-to-noise ratio of 10 dB (i.e., $\rho = 10$).

We first obtain the *cumulative distribution function* (cdf) for this wireless link by generating a large number of Rayleigh-distributed random transfer functions under the flat-fading assumption.

According to Eq. (6.64) with $\rho = 10$ and $N_r = N_t = 2$, the capacity is given by

$$C_k = \log_2\left\{\det\left(\mathbf{I}_2 + \frac{10}{2}\mathbf{H}_k\,\mathbf{H}_k^{\dagger}\right)\right\} \text{ bits/s/Hz}$$

for realization k of the channel transfer function. Fig. 6.13(a) plots the histogram (i.e., probability density function) of the resulting channel capacity data. This histogram is closely approximated by a *Gaussian distribution*, which should not be surprising when it is realized that an extensive amount of averaging could be involved in computing the log-det capacity formula.[8]

Integrating the probability density function curve of Fig. 6.13(a) and then subtracting the result from unity yields the complementary cumulative distribution function of the link, which is plotted in Fig. 6.13(b). Such a plot indicates the probability that a sample capacity will be greater

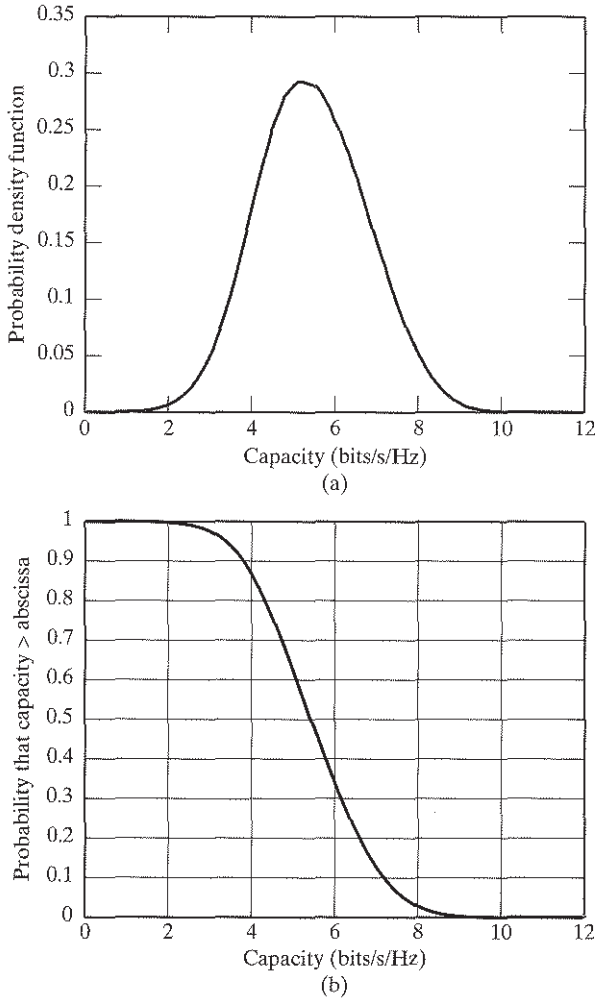


FIGURE 6.13 (a) Histogram (probability density function) of channel data for signal-to-noise ratio $\rho = 10$ dB. (b) Complementary cumulative probability distribution function corresponding to the histogram of part (a).

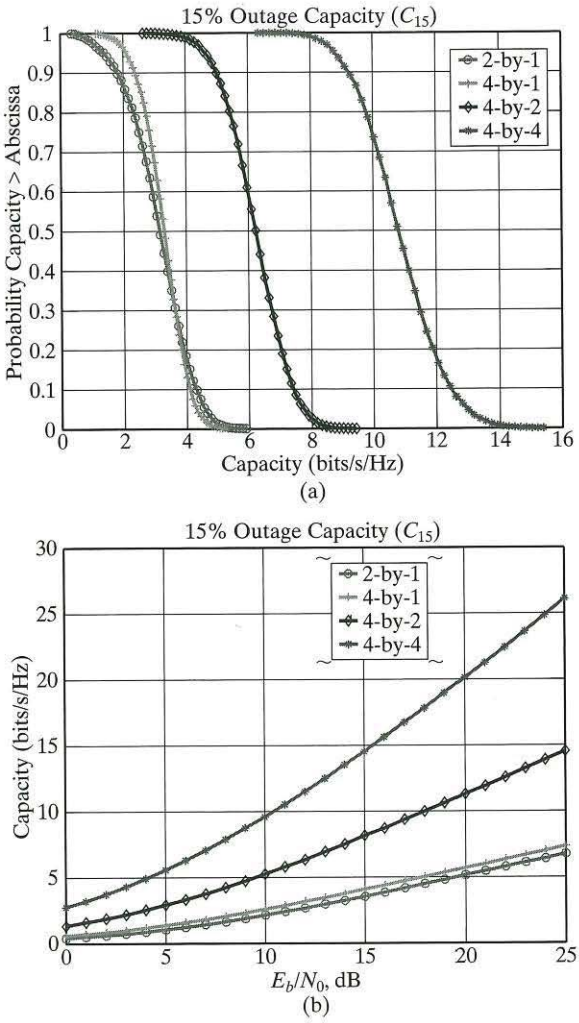


FIGURE 6.14 (a) Plots of the probability that the channel capacity is greater than the abscissa for four different antenna configurations:

$$N_t = 2, N_r = 1$$
$$N_t = 4, N_r = 1$$
$$N_t = 4, N_r = 2$$
$$N_t = 4, N_r = 4$$

(b) Plots of the outage capacity versus the signal-to-noise ratio for the four antenna configurations given in part (a).

than a threshold value (i.e., outage capacity) at a particular SNR (10 dB in this example) and for a specific antenna configuration (2 by 2 in this example). Thus, from Fig. 6.13(b), we can find the probability that the channel maintains a capacity of 4.2 bits/s/Hz for 85% of the time.

Figure 6.14 displays two different plots of the outage capacity for different signal-to-noise ratios and antenna configurations. Part (a) plots the probability that the channel capacity is greater than the abscissa versus the capacity in bits/s/Hz for a signal-to-noise ratio $\rho = 10$ dB. Part (b) of the figure portrays the picture differently by plotting the outage capacity in bits/s/Hz versus the signal-to-noise ratio in dB for varying antenna configurations.

Figure 6.14 clearly demonstrates two important points:

1. For $N_r = 1$, we see that increasing the number of transmit antennas by using $N_r = 2,4$, results in a modest increase in the outage capacity for a fixed SNR.
2. For $N_t = 4$ and fixed SNR, there is a significant increase in the outage capacity in going from the antenna configuration (4,1) to (4,2) and a much bigger increase in the outage capacity in going from the antenna configuration (4,2) to (4,4). ■

Note that, at high signal-to-noise ratios, the outage probability $P_{\text{outage}}(R)$ as defined in Eq. (6.66) is approximately the same as the *frame (burst)-error probability* in terms of the signal-to-noise ratio exponent.[9] Accordingly, we may use an analysis based on the outage probability to evaluate the performance of practical space–time block coding techniques. (Space-time block codes are discussed later in Section 6.6.) That is, for a prescribed rate R, we may evaluate how the performance of a certain space–time block coding technique compares with that predicted through an outage analysis or measurement.

6.4.4 Channel Known at the Transmitter

The log-det capacity formula of Eq. (6.59) is based on the premise that the transmitter has *no* knowledge of the channel state. Knowledge of the channel state, however, can be made available to the transmitter by first estimating the channel matrix $\mathbf{H}$ at the receiver and then sending this estimate to the transmitter via a *feedback channel*.[10] In such a scenario, the capacity is optimized over the correlation matrix of the transmitted signal vector $\mathbf{s}$, subject to the power constraint; that is, the trace of this correlation matrix is less than or equal to the constant transmit power P. Details of this optimization are presented in Appendix G.

From a practical perspective, it is important to note that the capacity gain provided by knowledge of the channel state at the transmitter over the log-det formula of Eq. (6.59) is significant only at low signal-to-noise ratios; the gain reduces to zero as the signal-to-noise ratio increases.

6.5 SINGULAR-VALUE DECOMPOSITION OF THE CHANNEL MATRIX

We may gain further insight into the behavior of a MIMO wireless communication system by applying what is known as the singular-value decomposition to the channel matrix of the system. The relationship between this algebraic decomposition of a rectangular matrix and the eigendecomposition of a Hermitian matrix formed by multiplying the matrix by its Hermitian transpose is discussed in Appendix H.

To begin the exposition, consider the matrix product $\mathbf{HH}^\dagger$ in the log-det capacity formula of Eq. (6.59). This product satisfies the Hermitian property for all $\mathbf{H}$. We may therefore diagonalize $\mathbf{HH}^\dagger$ by invoking the *eigendecomposition* of a Hermitian matrix and so write

$$\mathbf{U}^\dagger\mathbf{H}\mathbf{H}^\dagger\mathbf{U} = \Lambda \tag{6.67}$$

where the two new matrices $\mathbf{U}$ and Λ are described as follows:

- The matrix Λ is a *diagonal matrix* whose N_r elements are the *eigenvalues* of the matrix product $\mathbf{HH}^\dagger$.
- The matrix $\mathbf{U}$ is a *unitary matrix* whose N_r columns are the *eigenvectors* associated with the eigenvalues of $\mathbf{HH}^\dagger$.

By definition, the inverse of a unitary matrix is equal to the Hermitian transpose of the matrix, as shown by

$$\mathbf{U}^{-1} = \mathbf{U}^\dagger \tag{6.68}$$

or, equivalently,

$$\mathbf{U}\mathbf{U}^\dagger = \mathbf{U}^\dagger\mathbf{U} = \mathbf{I}_{N_r} \tag{6.69}$$

where $\mathbf{I}_{N_r}$ is the N_r-by-N_r identity matrix.

Let the N_t-by-N_t matrix $\mathbf{V}$ be another unitary matrix; that is,

$$\mathbf{V}\mathbf{V}^\dagger = \mathbf{V}^\dagger\mathbf{V} = \mathbf{I}_{N_t} \tag{6.70}$$

where $\mathbf{I}_{N_t}$ is the N_t-by-N_t identity matrix. Since the multiplication of a matrix by the identity matrix leaves the matrix unchanged, we may inject the matrix product $\mathbf{VV}^\dagger$ into the center of the left-hand side of Eq. (6.67), thus:

$$\mathbf{U}^\dagger\mathbf{H}(\mathbf{V}\mathbf{V}^\dagger)\mathbf{H}^\dagger\mathbf{U} = \Lambda \tag{6.71}$$

The left-hand side of Eq. (6.71), representing a square matrix, is recognized as the product of two rectangular matrices: the N_r-by-N_t matrix $\mathbf{U}^\dagger\mathbf{HV}$ and the N_t-by-N_r matrix $\mathbf{V}^\dagger\mathbf{H}^\dagger\mathbf{U}$, which are the Hermitian of each other. Let the N_t-by-N_t matrix $\mathbf{D}$ denote a new diagonal matrix related to the N_r-by-N_r diagonal matrix Λ with $N_r \leq N_t$ by

$$\Lambda = [\mathbf{D}\ \ \mathbf{0}][\mathbf{D}\ \ \mathbf{0}]^\dagger \tag{6.72}$$

where the null matrix $\mathbf{0}$ is added to maintain proper overall matrix dimensionality of the equation. Except for some zero elements, $\mathbf{D}$ is the *square root* of Λ. Then, examining Eqs. (6.71) and (6.72) and comparing terms, we deduce the new decomposition

$$\mathbf{U}^\dagger\mathbf{HV} = [\mathbf{D}\ \ \mathbf{0}] \tag{6.73}$$

Equation (6.73) is a mathematical statement of the *singular-value decomposition (SVD) theorem*, according to which we have the following descriptions:

- The elements of the diagonal matrix

$$\mathbf{D} = \text{diag}(d_1, d_2, \ldots, d_{N_t}) \tag{6.74}$$

  are the *singular values* of the channel matrix **H**.

- The columns of the unitary matrix

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{N_r}] \tag{6.75}$$

  are the *left singular vectors* of matrix **H**.

- The columns of the second unitary matrix

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{N_t}] \tag{6.76}$$

  are the *right singular vectors* of matrix **H**.

Problem 6.8 Applying the singular-value decomposition of Eq. (6.73) to the basic channel model of Eq. (6.48), show that for $N_r \leq N_t$:

$$\bar{\mathbf{x}} = [\mathbf{D}, \mathbf{0}]\bar{\mathbf{s}} + \bar{\mathbf{w}} \tag{6.77}$$

where

$$\bar{\mathbf{x}} = \mathbf{U}^\dagger \mathbf{x} \tag{6.78}$$

$$\bar{\mathbf{s}} = \mathbf{V}^\dagger \mathbf{s} \tag{6.79}$$

and

$$\bar{\mathbf{w}} = \mathbf{U}^\dagger \mathbf{w} \tag{6.80}$$

■

Using the definitions of Eqs. (6.74) through (6.76), we may rewrite the decomposed channel model of Eq. (6.77) in the scalar form

$$\bar{x}_i = d_i \bar{s}_i + \bar{w}_i \qquad i = 1,2,\ldots,N_r \tag{6.81}$$

According to Eq. (6.81), singular-value decomposition of the channel matrix **H** has transformed the MIMO wireless link with $N_r \leq N_t$ into N_r *virtual channels*, as illustrated in Fig. 6.15. (Note that $\bar{s}_i = 0$ for $N_r < i \leq N_t$.)The virtual channels are all *decoupled* from each other in that they constitute a parallel set of N_r *single-input, single-output (SISO) channels*, with each channel being described by the scalar input–output relation of Eq. (6.81). A comparison of the channel models of Figs. 6.12 and 6.15 immediately reveals the decoupling facilitated in the virtual model of Fig. 6.15 by the singular-value decomposition of the channel matrix **H**.

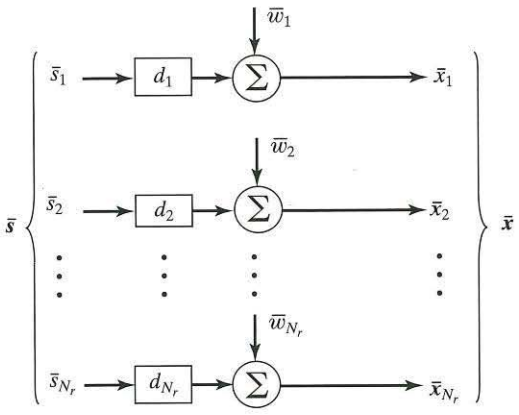


FIGURE 6.15 Set of N_r virtual decoupled channels resulting from the singular-value decomposition of the channel matrix $\mathbf{H}$, assuming that $N_t \leq N_r$.

6.5.1 Eigendecomposition of the Log-det Capacity Formula

The log-det formula of Eq. (6.59) for the ergodic capacity of a MIMO link involves the matrix product $\mathbf{HH}^\dagger$. Substituting Eq. (6.59) into Eq. (6.67) leads to the *spectral decomposition* of $\mathbf{HH}^\dagger$ in terms of N_r *eigenmodes*, with each eigenmode corresponding to virtual data transmission using a pair of right- and left-singular vectors of the channel matrix $\mathbf{H}$ as the transmit and receive antenna weights, respectively. Thus, we may write

$$\begin{aligned} \mathbf{HH}^\dagger &= \mathbf{U}\Lambda\mathbf{U}^\dagger \\ &= \sum_{i=1}^{N_r} \lambda_i \mathbf{u}_i \mathbf{u}_i^\dagger \end{aligned} \tag{6.82}$$

where the outer product $\mathbf{u}_i\mathbf{u}_i^\dagger$ is an N_r-by-N_r matrix with a rank equal to unity. Moreover, substituting the first line of this decomposition into the determinant part of Eq. (6.59) yields

$$\det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t}\mathbf{HH}^\dagger\right) = \det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t}\mathbf{U}\Lambda\mathbf{U}^\dagger\right) \tag{6.83}$$

Next, invoking the *determinant identity*

$$\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA}) \tag{6.84}$$

and then using the defining equation (6.69), we may rewrite Eq. (6.83) in the equivalent form

$$\begin{aligned}\det\left(\mathbf{I}_{N_r}+\frac{\rho}{N_t}\mathbf{H}\mathbf{H}^{\dagger}\right) &= \det\left(\mathbf{I}_{N_r}+\frac{\rho}{N_t}\mathbf{U}^{\dagger}\mathbf{U}\Lambda\right)\\ &= \det\left(\mathbf{I}_{N_r}+\frac{\rho}{N_t}\Lambda\right)\\ &= \prod_{i=1}^{N_r}\left(1+\frac{\rho}{N_t}\lambda_i\right)\end{aligned} \tag{6.85}$$

where λ_i is the ith eigenvalue of $\mathbf{H}\mathbf{H}^{\dagger}$. Finally, substituting Eq. (6.85) into Eq. (6.59) yields

$$C = \mathbf{E}\left[\sum_{i=1}^{N_r}\log_2\left(1+\frac{\rho}{N_t}\lambda_i\right)\right]\text{bits/s/Hz} \tag{6.86}$$

which is subject to constant transmit power; the expectation is over the eigenvalues of the matrix product $\mathbf{H}\mathbf{H}^{\dagger}$. Equation (6.86) shows that, thanks to the properties of the logarithm, *the ergodic capacity of a MIMO wireless communication system is the sum of capacities* of N_r *virtual single-input, single-output channels defined by the spatial eigenmodes of the matrix product* $\mathbf{H}\mathbf{H}^{\dagger}$.

According to Eq. (6.86), the channel capacity C attains its maximum value when equal signal-to-noise ratios ρ/N_t are allocated to each virtual channel in Fig. 6.15 (i.e., the N_r eigenmodes of the channel matrix $\mathbf{H}$ are all equally effective). By the same token, the capacity C is minimum when there is a single virtual channel (i.e., all the eigenmodes of the channel matrix $\mathbf{H}$ are zero except for one). The capacities of actual wireless links lie somewhere between these two extremes.

Specifically, as a result of *fading correlation* encountered in practice, it is possible for there to be a large disparity amongst the eigenvalues of $\mathbf{H}\mathbf{H}^{\dagger}$; that is, one or more of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{N_r}$ may be small. Such a disparity has quite a detrimental effect on the capacity of the wireless link, compared with the maximal condition under which all the eigenvalues of $\mathbf{H}\mathbf{H}^{\dagger}$ are equal. A similar effect may also arise in a Rician fading environment when the line-of-sight (LOS) component is quite strong (i.e., the Rician factor is greater than, say, 10 dB), in which case one eigenmode of the channel is dominant. For example, when the angle spread of the incoming radio waves impinging on a linear array of receive antennas is reduced from 60° to 0.6°, the complementary cumulative distribution function (ccdf) of a MIMO wireless communication system with $(N_t, N_r) = (7,7)$ antenna configuration degenerates effectively to that of a (7,1) system.[11]

Equation (6.86), based on the log-det capacity formula of Eq. (6.59), assumes that $N_t \geq N_r$. Using the log-det capacity formula of Eq. (6.60) for the alternative case,

$N_t \leq N_r$, and following a procedure similar to that used to derive Eq. (6.86), we may show that

$$C = \mathbf{E}\left[\sum_{i=1}^{N_t} \log_2\left(1 + \frac{\rho}{N_r}\lambda_i\right)\right] \text{ bits/s/Hz} \tag{6.87}$$

where the expectation is over the λ_i, which denote the eigenvalues of the N_t-by-N_t matrix product $\mathbf{H}^\dagger\mathbf{H}$.

Problem 6.9 For the special case of $N_t = N_r = N$, show that the ergodic capacity scales linearly, rather than logarithmically, with increasing SNR as N approaches infinity.

Ans. $$C \to \mathbf{E}\sum_{i=1}^{N}\left(\frac{\rho}{N\log_e 2}\right)\lambda_i = \left(\frac{\lambda_{\text{av}}}{\log_e 2}\right)\rho$$

where

$$\lambda_{\text{av}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\lambda_i$$

is the average eigenvalue of $\mathbf{H}\mathbf{H}^\dagger = \mathbf{H}^\dagger\mathbf{H}$. ■

6.6 SPACE–TIME CODES FOR MIMO WIRELESS COMMUNICATIONS

Transmission techniques for MIMO wireless communications may be considered under two broadly defined categories:

1. *Unconstrained signaling techniques*, exemplified by the so-called BLAST architectures, whose aim is to increase the channel capacity by using standard channel codes.
2. *Space–time codes*, whose aim is the joint channel encoding of multiple transmit antennas.

BLAST architectures are considered under Theme Example 1 in Section 6.9. Space–time codes are discussed in this section.

As with ordinary channel codes, space–time codes employ *redundancy* for the purpose of providing protection against channel fading, noise, and interference. They may also be used to minimize the outage probability or, equivalently, maximize the outage capacity. Depending on the level of redundancy introduced into the design of space–time codes, the degree of statistical independence among the transmitted signals is correspondingly reduced.

Space–time codes may themselves be classified into two types—space–time trellis codes and space–time block codes—depending on how transmission of the signal over the wireless channel takes place.

A *space–time trellis code* permits the *serial* transmission of symbols by combining signal processing at the receiver with coding techniques that are appropriate to the use of multiple antennas at the transmitter. Space–time trellis codes, designed for two to four transmit antennas, perform extremely well in a slow-fading environment, exemplified by indoor data transmission. For decoding, a multidimensional (i.e., vector) version of the Viterbi algorithm is required. Accordingly, for a fixed number of transmit antennas, the *decoding complexity of space–time trellis codes (measured in terms of trellis states at the decoder) increases exponentially* as a function of the spectral efficiency.[12]

In a *space–time block code*, by contrast, transmission of the signal takes place in *blocks*. The code is defined by a *transmission matrix*, the formulation of which involves three parameters:

- The number of transmitted symbols, denoted by l
- The number of transmit antennas, denoted by N_t, which defines the *size* of the transmission matrix
- The number of time slots in a data block, denoted by m

With m time slots involved in the transmission of l symbols, the ratio l/m defines the *rate* of the code, which is denoted by κ.

For efficient transmission, the transmitted symbols are expressed in *complex* form. Moreover, in order to facilitate the use of *linear processing* to estimate the transmitted symbols at the receiver and thereby simplify the receiver design, orthogonality is introduced into the design of the transmission matrix. Here, we may identify two different design procedures:

- *Complex orthogonal design*, in which the transmission matrix is square, satisfying the condition for complex *orthogonality* in both the spatial and temporal sense.
- *Generalized complex orthogonal design*, in which the transmission matrix is nonsquare, satisfying the condition for complex orthogonality only in the temporal sense; the code rate is less than unity.

In other words, *complex orthogonality of the transmission matrix in the temporal sense is a sufficient condition for linear processing at the receiver.*

A complex orthogonal design of size N_t exists if, and only if, $N_t = 2$ (i.e., the transmitter uses two transmit antennas), and the *Alamouti code* is that code with a code rate of unity. In contrast, generalized complex orthogonal designs permit the use of more than two transmit antennas, with the result that the code rate is less than unity.

The Alamouti code is much less complex than a space–time trellis code for the same antenna configuration (i.e., two transmit antennas and a single receive antenna), but it does not perform as well as the same space–time trellis code. Nonetheless, the Alamouti code is the preferred choice, essentially because of its remarkable computational simplicity and satisfactory performance capability. Indeed, the discovery of the code by Alamouti was motivated by addressing the issue of exponentially increasing decoding complexity of space–time trellis codes, and this discovery, in turn, motivated

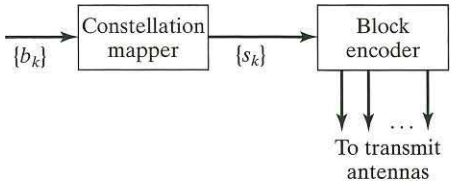


FIGURE 6.16 Block diagram of orthogonal space–time block encoder.

the search by Tarokh *et al.* for its generalization so as to accommodate the use of more than two transmit antennas.

Before proceeding to discuss the Alamouti code and generalized complex orthogonal designs of space–time block codes, we present some preliminary considerations.

6.6.1 Preliminaries

Figure 6.16 shows the baseband diagram of a *space–time block encoder*, which consists of two functional units: a mapper and the block encoder itself. The *mapper* takes the incoming binary data stream $\{b_k\}$, $b_k = \pm 1$, and generates a new *sequence of blocks*, with each block made up of multiple symbols that are complex. For example, the mapper may be in the form of an *M-ary PSK mapper* or an *M-ary QAM mapper*, which are illustrated for $M = 16$ in the signal-space diagrams of Fig. 6.17. All the symbols in a particular column of the transmission matrix are pulse shaped (in accordance with the criteria described in Section 3.4) and then modulated into a form suitable for simultaneous transmission over the channel by the transmit antennas. The pulse shaper and modulator are not shown in Fig. 6.16, as the basic issue of interest is that of baseband data transmission with an emphasis on the formulation of space–time block codes. The *block encoder* converts each block of complex symbols produced by the mapper into an l-by-N_t transmission matrix $\mathbf{S}$, where l and N_t are the *temporal* dimension and *spatial* dimension, respectively, of the transmission matrix. The individual elements of the transmission matrix $\mathbf{S}$ are made up of the complex symbols, say, $\tilde{s}_k$, generated by the mapper, their complex conjugates $\tilde{s}_k^*$, and linear combinations of $\tilde{s}_k$ and $\tilde{s}_k^*$, where the asterisk denotes complex conjugation.

EXAMPLE 6.5 Quadriphase-Shift Keying

As a simple example, consider the map portrayed by QPSK with $M = 4$. The map is described in Table 6.1, where E is the transmitted signal energy.

The input dibits (pairs of adjacent bits) are *Gray* encoded, wherein only one bit is flipped as we move from one symbol to the next. (Gray encoding was discussed in Section 3.6.) The mapped signal points lie on a circle of radius $\mathbf{E}$ centered at the origin of the signal-space diagram. ■

Problem 6.10 Construct the map describing the M-ary PSK for $M = 16$ in Fig. 6.17(a). ■

TABLE 6.1 Gray-encoded QPSK mapper.

Dibit	Coordinates of mapped signal points
10	$\sqrt{E/2}(1, -1) = \sqrt{E}e^{j7\pi/4}$
11	$\sqrt{E/2}(-1, -1) = \sqrt{E}e^{j5\pi/4}$
01	$\sqrt{E/2}(-1, +1) = \sqrt{E}e^{j3\pi/4}$
00	$\sqrt{E/2}(+1, +1) = \sqrt{E}e^{j\pi/4}$

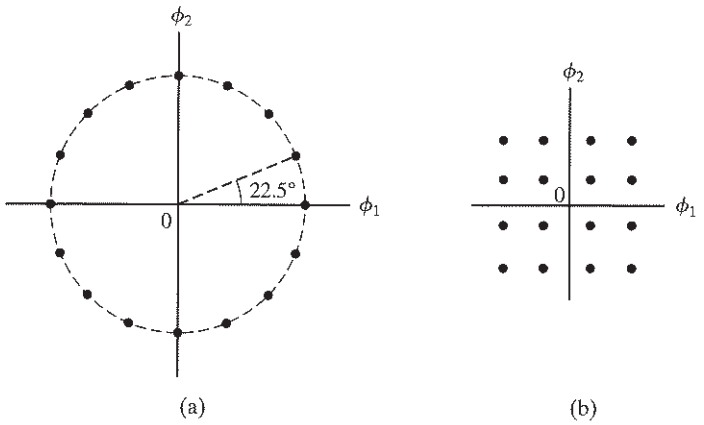


FIGURE 6.17 (a) Signal constellation of 16-PSK. (b) Signal constellation of 16-QAM.

Problem 6.11 Construct the map describing the M-ary QAM for $M = 16$ in Fig. 6.17(b). ■

Henceforth, the discussion of space–time block codes is confined to two-dimensional mappers exemplified by those portrayed in Fig. 6.17. That is, the output of the mapper is represented by a complex number, as illustrated in Table 6.1 for QPSK.

6.6.2 Alamouti Code[13]

The *Alamouti code* is a two-by-one orthogonal space–time block code. That is, it uses two transmit antennas and a single receive antenna, as shown in the block diagram of Fig. 6.18. Let $\tilde{s}_1$ and $\tilde{s}_2$ denote the complex symbols (signals) produced by the mapper, which are to be transmitted over the wireless channel. Signal transmission over the channel proceeds as follows:

- At some arbitrary time t, antenna 1 transmits $\tilde{s}_1$, and simultaneously, antenna 2 transmits $\tilde{s}_2$.
- At time $t + T$, where T is the symbol duration, signal transmission is switched, with $-\tilde{s}_2^*$ transmitted by antenna 1 and $\tilde{s}_1^*$ simultaneously transmitted by antenna 2.

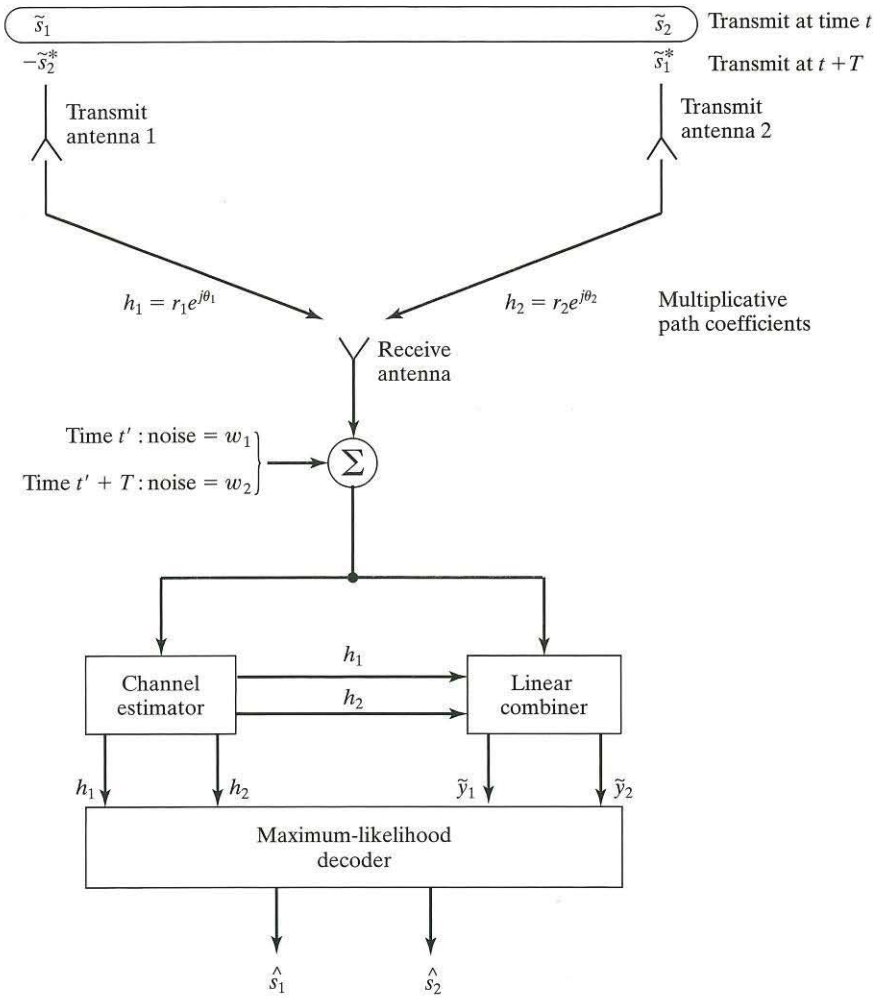


FIGURE 6.18 Block diagram of the transceiver (transmitter and receiver) for the Alamouti code. Note that $t' > t$ to allow for propagation delay.

The two-by-two space–time block code, just described, is formally written in matrix form as

$$\mathbf{S} = \begin{bmatrix} \tilde{s}_1 & \tilde{s}_2 \\ -\tilde{s}_2^* & \tilde{s}_1^* \end{bmatrix} \begin{matrix} \Rightarrow \text{Space} \\ \Downarrow \\ \text{Time} \end{matrix} \tag{6.88}$$

The *transmission matrix* $\mathbf{S}$ is a *complex-orthogonal matrix* (quaternion), in that it satisfies the condition for orthogonality in both the spatial and temporal sense. To demonstrate this important property, let

$$\mathbf{S}^\dagger = \begin{bmatrix} \tilde{s}_1^* & -\tilde{s}_2 \\ \tilde{s}_2^* & \tilde{s}_1 \end{bmatrix} \quad \begin{matrix} \Rightarrow \text{ Time} \\ \Downarrow \\ \text{Space} \end{matrix} \tag{6.89}$$

denote the *Hermitian transpose* of $\mathbf{S}$, involving both transposition and complex conjugation. To demonstrate orthogonality in a spatial sense, we multiply the code matrix $\mathbf{S}$ by its Hermitian transpose $\mathbf{S}^\dagger$ on the right, obtaining

$$\begin{aligned} \mathbf{S}\mathbf{S}^\dagger &= \begin{bmatrix} \tilde{s}_1 & \tilde{s}_2 \\ -\tilde{s}_2^* & \tilde{s}_1^* \end{bmatrix} \begin{bmatrix} \tilde{s}_1^* & -\tilde{s}_2 \\ \tilde{s}_2^* & \tilde{s}_1 \end{bmatrix} \\ &= \begin{bmatrix} |\tilde{s}_1|^2 + |\tilde{s}_2|^2 & -\tilde{s}_1\tilde{s}_2 + \tilde{s}_2\tilde{s}_1 \\ -\tilde{s}_2^*\tilde{s}_1^* + \tilde{s}_1^*\tilde{s}_2^* & |\tilde{s}_2|^2 + |\tilde{s}_1|^2 \end{bmatrix} \\ &= (|\tilde{s}_1|^2 + |\tilde{s}_2|^2) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{6.90}$$

which equals the two-by-two identity matrix, multiplied by the scaling factor $(|\tilde{s}_1|^2 + |\tilde{s}_2|^2)$.

This same result also holds for the alternative matrix product $\mathbf{S}^\dagger\mathbf{S}$, which is proof of *orthogonality* in the temporal sense. Thus, the transmission matrix of the Alamouti code satisfies the unique condition

$$\mathbf{S}\mathbf{S}^\dagger = \mathbf{S}^\dagger\mathbf{S} = (|\tilde{s}_1|^2 + |\tilde{s}_2|^2)\mathbf{I} \tag{6.91}$$

where $\mathbf{I}$ is the two-by-two identity matrix. Note that

$$\mathbf{S}^{-1} = \frac{1}{|\tilde{s}_1|^2 + |\tilde{s}_2|^2}\,\mathbf{S}^\dagger \tag{6.92}$$

In light of Eqs. (6.63), (6.90), and (6.91), we may summarize four important properties of the Alamouti code:

Property 1. Unitarity (Complex Orthogonality)

The Alamouti code is an orthogonal space–time block code in that the product of its transmission matrix with its Hermitian transpose is equal to the two-by-two identity matrix scaled by the sum of the squared amplitudes of the transmitted symbols.

Property 2. Full-Rate Complex Code

The Alamouti code (with two transmit antennas) is the only complex space–time block code with a code rate of unity. For any signal constellation, it therefore achieves full diversity at the full transmission rate.

Property 3. Linearity

The Alamouti code is linear in the transmitted symbols. That is, we may expand the code matrix $\mathbf{S}$ as a linear combination of the transmitted symbols and their complex conjugates. Doing so, we obtain

$$\mathbf{S} = \tilde{s}_1\Gamma_{11} + \tilde{s}_1^*\Gamma_{12} + \tilde{s}_2\Gamma_{21} + \tilde{s}_2^*\Gamma_{22} \tag{6.93}$$

where

$$\left.\begin{aligned}\Gamma_{11} &= \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\\ \Gamma_{12} &= \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\\ \Gamma_{21} &= \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\\ \Gamma_{22} &= \begin{bmatrix}0 & 0\\ -1 & 0\end{bmatrix}\end{aligned}\right\} \tag{6.94}$$

Property 4. Optimality of Capacity

For two transmit antennas and a single receive antenna, the Alamouti code is the only optimal space–time block code that satisfies the log-det capacity formula of Eq. (6.63).

Turning next to the design of the receiver, we assume that the channel is frequency flat and slowly varying, such that the complex multiplicative distortion introduced by the channel at time t is essentially the same as that at time $t + T$, where T is the symbol duration. This multiplicative distortion is denoted by $\alpha_k e^{j\theta_k}$ where $k = 1,2$, as indicated in Fig. 6.18. Thus, with $\tilde{s}_1$ and $\tilde{s}_2$ transmitted simultaneously at time t, the complex received signal at time $t' > t$, allowing for propagation delay, is

$$\tilde{x}_1 = \alpha_1 e^{j\theta_1}\tilde{s}_1 + \alpha_2 e^{j\theta_2}\tilde{s}_2 + w_1 \tag{6.95}$$

where w_1 is the complex channel noise at time t'. Next, with $-\tilde{s}_2^*$ and $\tilde{s}_1^*$ transmitted simultaneously at time $t + T$, the corresponding complex signal received at time $t' + T$ is

$$\tilde{x}_2 = -\alpha_1 e^{j\theta_1}\tilde{s}_2^* + \alpha_2 e^{j\theta_2}\tilde{s}_1^* + w_2 \tag{6.96}$$

where w_2 is the complex channel noise at time $t' + T$.

In the course of time, from time t' to $t' + T$, the channel estimator in the receiver produces estimates of the multiplicative distortion represented by $\alpha_k e^{j\theta_k}$, where $k = 1, 2$. Hereafter, we assume that these two estimates are accurate enough for them to be treated as essentially exact. In other words, $\alpha_1 e^{j\theta_1}$ and $\alpha_2 e^{j\theta_2}$ are both known to the receiver. Accordingly, we may reformulate the variable $\tilde{x}_1$ defined by Eq. (6.95) and the complex conjugate of the second variable $\tilde{x}_2$ defined in Eq. (6.96) in matrix form, as shown by

$$\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2^* \end{bmatrix} = \begin{bmatrix} \alpha_1 e^{j\theta_1} & \alpha_2 e^{j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{-j\theta_1} \end{bmatrix} \begin{bmatrix} \tilde{s}_1 \\ \tilde{s}_2 \end{bmatrix} + \begin{bmatrix} \tilde{w}_1 \\ \tilde{w}_2^* \end{bmatrix} \tag{6.97}$$

The nice thing about this equation is that the original complex signals s_1 and s_2 appear as a vector of two unknowns. It is with this goal in mind that $\tilde{x}_1$ and $\tilde{x}_2^*$ were used for the elements of the two-by-one received signal vector on the left-hand side of Eq. (6.97).

Problem 6.12 Show that the two-by-two channel matrix in Eq. (6.97) defined by the multiplicative fading factors $\alpha_1 e^{j\theta_1}$ and $\alpha_2 e^{j\theta_2}$ is an orthogonal matrix; that is, show that

$$\begin{bmatrix} \alpha_1 e^{j\theta_1} & \alpha_2 e^{j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{-j\theta_1} \end{bmatrix}^{\dagger} \begin{bmatrix} \alpha_1 e^{j\theta_1} & \alpha_2 e^{j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{-j\theta_1} \end{bmatrix} = (\alpha_1^2 + \alpha_2^2) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

■

Motivated by the result of Problem 6.12, we multiply both sides of Eq. (6.97) by the Hermitian transpose of the two-by-two channel matrix, namely,

$$\begin{bmatrix} \alpha_1 e^{-j\theta_1} & \alpha_2 e^{j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{j\theta_1} \end{bmatrix}$$

and thus obtain the following pair of simultaneous equations, written in matrix form:

$$\begin{aligned} \begin{bmatrix} \alpha_1 e^{-j\theta_1} & \alpha_2 e^{j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{j\theta_1} \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2^* \end{bmatrix} &= (\alpha_1^2 + \alpha_2^2) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{s}_1 \\ \tilde{s}_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 e^{-j\theta_1} & \alpha_2 e^{j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{j\theta_1} \end{bmatrix} \begin{bmatrix} \tilde{w}_1 \\ \tilde{w}_2^* \end{bmatrix} \\ &= (\alpha_1^2 + \alpha_2^2) \begin{bmatrix} \tilde{s}_1 \\ \tilde{s}_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 e^{-j\theta_1} \tilde{w}_1 + \alpha_2 e^{j\theta_2} \tilde{w}_2^* \\ \alpha_2 e^{-j\theta_2} \tilde{w}_1 - \alpha_1 e^{j\theta_1} \tilde{w}_2^* \end{bmatrix} \end{aligned} \tag{6.98}$$

The two-by-one *vector product* on the left-hand side of Eq. (6.98) defines the two complex outputs computed by the linear combiner in Fig. 6.18 in terms of the complex signals $\tilde{x}_1$ and $\tilde{x}_2$ received at times t' and $t' + T$. Thus, let

$$\begin{aligned}\begin{bmatrix}\tilde{y}_1\\ \tilde{y}_2\end{bmatrix} &= \begin{bmatrix}\alpha_1 e^{-j\theta_1} & \alpha_2 e^{j\theta_2}\\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{j\theta_1}\end{bmatrix}\begin{bmatrix}\tilde{x}_1\\ \tilde{x}_2^*\end{bmatrix}\\ &= \begin{bmatrix}\alpha_1 e^{-j\theta_1}\tilde{x}_1 + \alpha_2 e^{j\theta_2}\tilde{x}_2^*\\ \alpha_2 e^{-j\theta_2}\tilde{x}_1 - \alpha_1 e^{j\theta_1}\tilde{x}_2^*\end{bmatrix}\end{aligned} \tag{6.99}$$

define the two-by-one vector of complex signals at the output of the combiner. Correspondingly, let

$$\begin{bmatrix}\tilde{\nu}_1\\ \tilde{\nu}_2\end{bmatrix} = \begin{bmatrix}\alpha_1 e^{-j\theta_1}\tilde{w}_1 + \alpha_2 e^{j\theta_2}\tilde{w}_2^*\\ \alpha_2 e^{-j\theta_2}\tilde{w}_1 - \alpha_1 e^{j\theta_1}\tilde{w}_2^*\end{bmatrix} \tag{6.100}$$

define the two-by-one vector of additive complex noise contributions in the combiner outputs. Accordingly, we may recast Eq. (6.98) in a matrix form of input–output relations describing the overall behavior of the Alamouti code, structured as in Fig. 6.18, as follows:

$$\begin{bmatrix}\tilde{y}_1\\ \tilde{y}_2\end{bmatrix} = (\alpha_1^2 + \alpha_2^2)\begin{bmatrix}\tilde{s}_1\\ \tilde{s}_2\end{bmatrix} + \begin{bmatrix}\tilde{\nu}_1\\ \tilde{\nu}_2\end{bmatrix} \tag{6.101}$$

In expanded form,

$$\tilde{y}_k = (\alpha_1^2 + \alpha_2^2)\tilde{s}_k + \tilde{\nu}_k \qquad k = 1, 2 \tag{6.102}$$

Note that, due to complex orthogonality of the Alamouti code, the unwanted symbol $\tilde{s}_2$ is cancelled out in the equation for $\tilde{y}_1$ and the unwanted symbol $\tilde{s}_1$ is cancelled out in the equation for $\tilde{y}_2$. *It is these cancellations that are responsible for the simplification of the receiver.*

Note also that the scaling factor $(\alpha_1^2 + \alpha_2^2)$ can be small only if the fading coefficients α_1 and α_2 are both small. In other words, the diversity paths linking the receive antenna to the two transmit antennas must simultaneously undergo fading in order for $(\alpha_1^2 + \alpha_2^2)$ to be small. Hence, the detrimental effect of fading arises if, and only if, both diversity paths suffer from it. We may therefore state that a wireless communication system based on the Alamouti code enjoys a *two-level diversity gain*.

Problem 6.13 Ordinarily, the complex channel noise terms $\tilde{w}_1$ and $\tilde{w}_2$ are Gaussian distributed with zero mean and a common variance. Assume that $\alpha_k e^{j\theta_k}$, $k = 1, 2$, are known multiplicative constants.

(a) Show that the complex noise terms $\tilde{v}_1$ and $\tilde{v}_2$ at the combiner output are also Gaussian distributed with zero mean.

(b) Given that the real and imaginary components of $\tilde{v}_1$ and $\tilde{v}_2$ have a common variance σ_0^2, determine the common variance of $\tilde{v}_1$ and $\tilde{v}_2$.

(c) Show that the probability density function of v_1 or v_2 is

$$f_{V_k}(\tilde{v}_k) = \frac{1}{\pi\sigma_0^2}\exp\left(-\frac{|\tilde{v}_k|^2}{\sigma_0^2}\right) \qquad k = 1, 2 \tag{6.103}$$

■

Equation (6.102) is a partial description of the receiver structure depicted in Fig. 6.18 for an Alamouti-encoded system. The next and final issue to be considered is, given the noisy linear combiner output $\tilde{y}_k$ expressed in terms of the transmitted symbol $\tilde{s}_k$, how to provide an optimal estimate of $\tilde{s}_k$ for $k = 1,2$. To provide insight into this symbol-estimation problem, Fig. 6.19 illustrates the signal-space scenario for an Alamouti-encoded system based on the QPSK signal constellation. At time t, the symbols $\tilde{s}_1$ and $\tilde{s}_2$ drawn from this constellation are transmitted, followed by transmission of the modified symbols $-\tilde{s}_2^*$ and $\tilde{s}_1^*$ at time $t + T$. To account for the multipath fading phenomenon, the signal points $\tilde{s}_1$ and $\tilde{s}_2$ are weighted by the multiplying factor $(\alpha_1^2 + \alpha_2^2)$ in accordance with Eq. (6.102). The complex Gaussian noise clouds centered on the weighted signal points represent the effect of additive complex noise terms, whose intensity decreases as we move away from the weighted signal points. Figure 6.19 also includes the linear combiner outputs signified as the observations $\tilde{y}_1$ and $\tilde{y}_2$. Given the scenario pictured in this figure, what is the optimal decision for the receiver to make?

The answer to this fundamental question lies in the *maximum-likelihood decoding rule*, which assumes that the transmitted symbols $\tilde{s}_1$ and $\tilde{s}_2$ are *equally probable*. Let $\hat{s}_1$ and $\hat{s}_2$ denote the maximum-likelihood estimates of $\tilde{s}_1$ and $\tilde{s}_2$, respectively. One further item that we need to introduce is

$$d^2(\tilde{z}, \tilde{\xi}) = (\tilde{z} - \tilde{\xi})(\tilde{z} - \tilde{\xi})^*$$

which defines the squared Euclidean distance between two complex signal points denoted by $\tilde{z}$ and $\tilde{\xi}$.

To proceed then with the formulation of the maximum-likelihood decoding rule for an Alamouti-encoded system, we see from Fig. 6.18 that, given knowledge of the complex channel coefficients h_1 and h_2, the maximum-likelihood decoder performs its decision-making in response to the linear combiner outputs $\tilde{y}_1$ and $\tilde{y}_2$, which play the role of *decision statistics*. (By definition, sufficient statistics summarize the whole of the

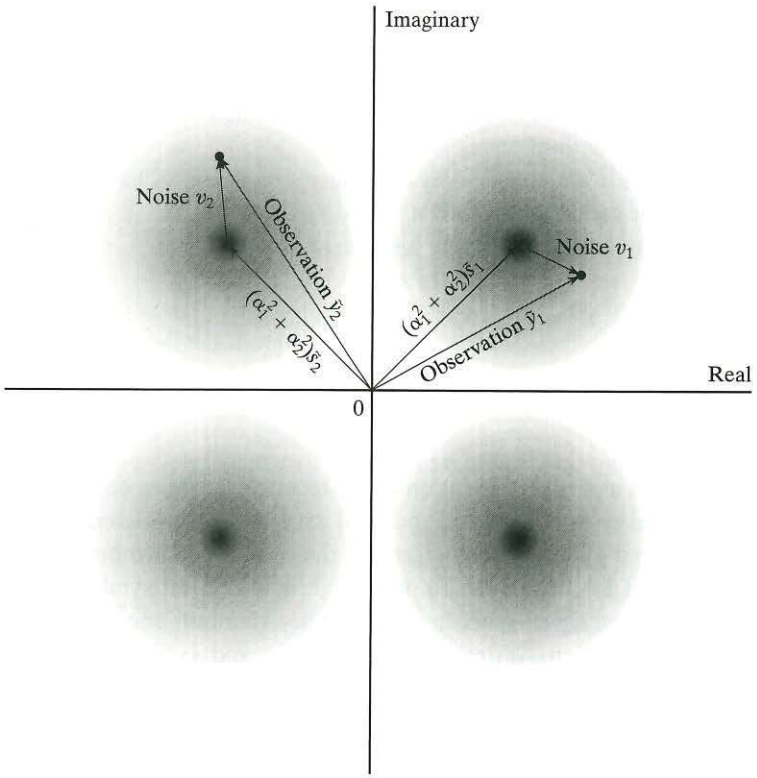


FIGURE 6.19 Signal-space diagram for Alamouti code, using the QPSK signal constellation. The signal point $\tilde{s}_1$ and $\tilde{s}_2$ are weighted by $(\alpha_1^2 + \alpha_2^2)$. The corresponding linear combiner outputs $\tilde{y}_1$ and $\tilde{y}_2$ are noisy versions of these weighted signals.

relevant information supplied by observables.) It is therefore logical that we formulate the maximum-likelihood decoding rule in terms of $\tilde{y}_1$ and $\tilde{y}_2$. To do so, we use two things: definition of the squared Euclidean distance, and the formula of Eq. (6.102) for the linear combiner outputs $\tilde{y}_k$, $k = 1,2$. In particular, we see from Eq. (6.102) that, except for the statistics of additive complex Gaussian noise component $\tilde{v}_k$, the decision statistic $\tilde{y}_k$ is uniquely determined by the transmitted symbol $\tilde{s}_k$, $k = 1,2$. Accordingly, we may simplify implementation of the maximum-likelihood decoder by decomposing it into two independent decoders, one operating on the decision statistic $\tilde{y}_1$ and the second one operating on the other decision statistic $\tilde{y}_2$, which are observed T seconds apart. On this basis, we may now formally state the maximum-likelihood decoding rule as follows:

> *Given that the receiver has knowledge of* (i) *the channel fading coefficients* α_1 *and* α_2, *and* (ii) *the set of all possible transmitted symbols in*

the mapper's constellation denoted by S, the maximum-likelihood estimates of the transmitted symbols $\tilde{s}_1$ *and* $\tilde{s}_2$ *are respectively defined by*

$$\hat{s}_1 = \arg \min_{\varphi \in S} \left\{ d^2(\tilde{y}_1, (\alpha_1^2 + \alpha_2^2)\varphi) \right\} \tag{6.104}$$

and

$$\hat{s}_2 = \arg \min_{\varphi \in S} \left\{ d^2(\tilde{y}_2, (\alpha_1^2 + \alpha_2^2)\varphi) \right\} \tag{6.105}$$

where the φ *denote the different hypotheses for the linear combiner output* $\tilde{y}_1$ *and* $\tilde{y}_2$.

Note that the right-hand sides of Eqs. (6.104) and (6.105) do not include any references to the actual transmitted symbols $\tilde{s}_1$ and $\tilde{s}_2$, which is how it should be since they are both unknown at the receiver. Obviously, the receiver makes the correct decisions on the transmitted symbols if both $\hat{s}_1 = \tilde{s}_1$ and $\hat{s}_2 = \tilde{s}_2$.

Returning to the scenario pictured in Fig. 6.19, we note the following

$$S = \{\tilde{s}_j\}_{j=1}^{4}$$

and

$$\varphi = \tilde{s}_1, \tilde{s}_2, \tilde{s}_3, \text{ or } \tilde{s}_4$$

Application of the maximum-likelihood decoding rules of Eqs. (6.104) and (6.105) to this example scenario yields $\hat{s}_1 = \tilde{s}_1$ and $\hat{s}_2 = \tilde{s}_2$, both of which are correct.

Problem 6.14 For M-ary PSK, the mapper's signal constellation consists of M points uniformly distributed on a circle with center at the origin and radius $\sqrt{E}$, where E is the signal energy per symbol. Show that for this particular modulation scheme, the maximum-likelihood decoding rule of Eqs. (6.104) and (6.105) respectively reduce to

$$\hat{s}_1 = \arg \min_{\varphi_1 \in S} d^2(\tilde{y}_1, \varphi_1)$$

and

$$\hat{s}_2 = \arg \min_{\varphi_2 \in S} d^2(\tilde{y}_2, \varphi_2)$$

■

6.6.3 Performance Comparison of Diversity-on-Receive and Diversity-on-Transmit Schemes

Figure 6.20 presents a computer simulation comparing the bit error rate performance of coherent binary PSK over an uncorrelated Rayleigh-fading channel for three different schemes:

(a) No diversity (i.e., one transmit antenna and one receive antenna)

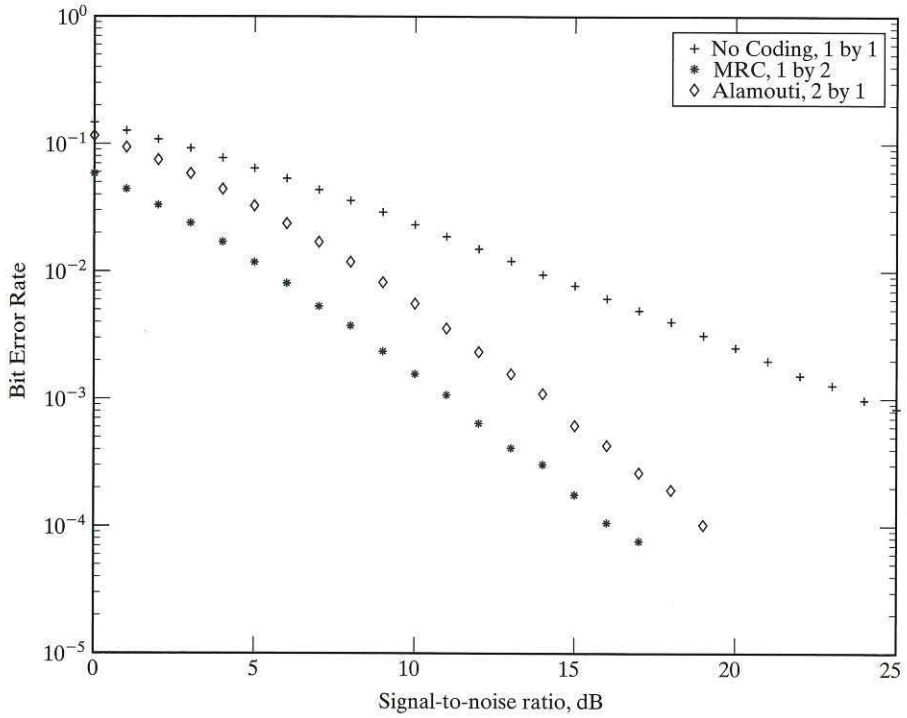


FIGURE 6.20 Comparison of the bit error rate performance of coherent BPSK over flat-fading Rayleigh channel for three configurations:
(a) No diversity.
(b) Maximal-ratio combiner ($N_t = 1, N_r = 2$).
(c) Alamouti code ($N_t = 2, N_r = 1$).

(b) The maximal-ratio combiner (i.e., one transmit antenna and two receive antennas)
(c) The Alamouti code (i.e., two transmit antennas and one receive antenna)

It is assumed that the total transmit power is the same for all three schemes, and in the case of the two diversity schemes (b) and (c), there is perfect knowledge of the channel(s) at the receiver(s).

From the figure, we see that the performance of the Alamouti code is worse by about 3 dB, compared with the maximal-ratio combiner. This 3-dB penalty is attributed to the fact that, in the space-diversity-on-transmit scheme using the Alamouti code, the transmit power in each of the two antennas is one-half of the transmit power in the space-diversity-on-receive scheme using the maximal-ratio combiner. Indeed, the diversity schemes based on the Alamouti code and maximal-ratio combiners would behave in the same way if each transmit antenna in the Alamouti code was to radiate the same power as the single transmit antenna in the maximal-ratio combiner.

Note also that, in an asymptotic sense, the plots for the MRC and Alamouti code have the same slope. This slope provides a measure of the *diversity order*. For the case at hand, the slope, and therefore the diversity order, is two, a result that confirms what we already know: the Alamouti code uses two transmit antennas, and the MRC uses two receive antennas. The issue of diversity order is considered in greater detail in Section 6.10.

6.6.4 Generalized Complex Orthogonal Space–Time Block Codes[14]

As mentioned previously, it was the remarkable computational simplicity of the Alamouti code and its capability to deliver a satisfactory performance that motivated the search for complex space–time block codes using more than two antennas. This search started with the equally pioneering work of Tarokh, Jafarkhani, and Calderbank. Building on the classic work of a number of theorists, including Radon and Hurwitz, Tarokh et al. introduced a *theory of generalized complex orthogonal designs*. This new theory applies to the construction of nonsquare orthogonal complex space–time block codes that combine coding at the transmitter with linear processing at the receiver.

Generalized complex orthogonal designs of space–time block codes distinguish themselves from the Alamouti code in three respects:

1. *A nonsquare transmission matrix*, which accommodates the use of more than two transmit antennas
2. *A fractional code rate*
3. *Orthogonality of the transmission matrix only in the temporal sense*, which is sufficient for maximum-likelihood decoding implemented in the form of a linear receiver in a manner analogous to the way the Alamouti code is implemented.

To define what we mean by a generalized complex orthogonal design, let $\boldsymbol{G}$ denote an m-by-N_t matrix, with N_t denoting the number of transmit antennas and m denoting the number of time slots. Let the entries of the matrix be designated

$$0, \pm s_1, \pm s_1^*, \pm s_2, \pm s_2^*, \ldots, \pm s_l, \pm s_l^*$$

where the number of transmitted symbols $l < m$. Then the matrix $\boldsymbol{G}$ is said to be a generalized complex orthogonalized design of size N_t and code rate $\kappa = l/m$ if it satisfies the condition for orthogonality in the temporal sense—that is, if

$$\boldsymbol{G}^\dagger \boldsymbol{G} = \left(\sum_{j=1}^{l} |s_j|^2 \right) \mathbf{I} \tag{6.106}$$

where $\mathbf{I}$ is the N_t-by-N_t identity matrix.

The construction of space–time block codes using the generalized complex orthogonal design is exemplified by the rate-1/2 codes for transmission over a wireless channel. We have two cases:

(a) For three transmit antennas ($l = 4, m = 8$),

$$\boldsymbol{G}_3 = \begin{bmatrix} \tilde{s}_1 & \tilde{s}_2 & \tilde{s}_3 \\ -\tilde{s}_2 & \tilde{s}_1 & -\tilde{s}_4 \\ -\tilde{s}_3 & \tilde{s}_4 & \tilde{s}_1 \\ -\tilde{s}_4 & -\tilde{s}_3 & \tilde{s}_2 \\ \tilde{s}_1^* & \tilde{s}_2^* & \tilde{s}_3^* \\ -\tilde{s}_2^* & \tilde{s}_1^* & -\tilde{s}_4^* \\ -\tilde{s}_3^* & \tilde{s}_4^* & \tilde{s}_1^* \\ -\tilde{s}_4^* & -\tilde{s}_3^* & \tilde{s}_2^* \end{bmatrix} \begin{array}{l} \Rightarrow \text{Space} \\ \\ \\ \\ \\ \\ \\ \\ \Downarrow \text{Time} \end{array} \tag{6.107}$$

(b) For four transmit antennas ($l = 4, m = 8$),

$$\boldsymbol{G}_4 = \begin{bmatrix} \tilde{s}_1 & \tilde{s}_2 & \tilde{s}_3 & \tilde{s}_4 \\ -\tilde{s}_2 & \tilde{s}_1 & -\tilde{s}_4 & \tilde{s}_3 \\ -\tilde{s}_3 & \tilde{s}_4 & \tilde{s}_1 & -\tilde{s}_2 \\ -\tilde{s}_4 & -\tilde{s}_3 & \tilde{s}_2 & \tilde{s}_1 \\ \tilde{s}_1^* & \tilde{s}_2^* & \tilde{s}_3^* & \tilde{s}_4^* \\ -\tilde{s}_2^* & \tilde{s}_1^* & -\tilde{s}_4^* & \tilde{s}_3^* \\ -\tilde{s}_3^* & \tilde{s}_4^* & \tilde{s}_1^* & -\tilde{s}_2^* \\ -\tilde{s}_4^* & -\tilde{s}_3^* & \tilde{s}_2^* & \tilde{s}_1^* \end{bmatrix} \begin{array}{l} \Rightarrow \text{Space} \\ \\ \\ \\ \\ \\ \\ \\ \Downarrow \text{Time} \end{array} \tag{6.108}$$

(The subscripts in $\boldsymbol{G}_3$ and $\boldsymbol{G}_4$ refer to the numbers of transmit antennas.)

The symbol $\boldsymbol{G}$ used for the transmission matrices of these two codes is different from the corresponding symbol $\mathbf{S}$ for the Alamouti code, to emphasize the basic difference between them. For the Alamouti code, we have $\mathbf{S}\mathbf{S}^\dagger = \mathbf{S}^\dagger\mathbf{S}$, whereas the codes of Eqs. (6.107) and (6.108) do *not* satisfy the condition $\boldsymbol{G}\boldsymbol{G}^\dagger = \boldsymbol{G}^\dagger\boldsymbol{G}$.

Compared with the Alamouti code defined by the transmission matrix of Eq. (6.88), the space–time codes $\boldsymbol{G}_3$ and $\boldsymbol{G}_4$ are at a disadvantage in two respects:

1. The bandwidth efficiency is reduced by a factor of two.

2. The number of time slots across which the channel is required to have a constant fading envelope is increased by a factor of four.

To improve the bandwidth efficiency, we may use rate-3/4 generalized complex linear-processing orthogonal designs referred to as *sporadic codes*. The construction of these codes differs from that of $\boldsymbol{G}_3$ and $\boldsymbol{G}_4$—hence the use of the symbol $\boldsymbol{H}$ to denote their transmission matrices. Two cases of sporadic codes are as follows:

(a) For three transmit antennas ($l = 3, m = 4$),

$$\boldsymbol{H}_3 = \begin{bmatrix} \tilde{s}_1 & \tilde{s}_2 & \tilde{s}_3/\sqrt{2} \\ -\tilde{s}_2^* & \tilde{s}_1^* & \tilde{s}_3/\sqrt{2} \\ \tilde{s}_3^*/\sqrt{2} & \tilde{s}_3^*/\sqrt{2} & (-\tilde{s}_1 - \tilde{s}_1^* + \tilde{s}_2 - \tilde{s}_2^*)/2 \\ \tilde{s}_3^*/\sqrt{2} & -\tilde{s}_3^*/\sqrt{2} & (\tilde{s}_2 + \tilde{s}_2^* + \tilde{s}_1 - \tilde{s}_1^*)/2 \end{bmatrix} \Rightarrow \text{Space}$$

$$\Downarrow$$

$$\text{Time} \tag{6.109}$$

(b) For four transmit antennas ($l = 3, m = 4$),

$$\boldsymbol{H}_4 = \begin{bmatrix} \tilde{s}_1 & \tilde{s}_2 & \tilde{s}_3/\sqrt{2} & \tilde{s}_3/\sqrt{2} \\ -\tilde{s}_2^* & \tilde{s}_1^* & \tilde{s}_3/\sqrt{2} & -\tilde{s}_3/\sqrt{2} \\ \tilde{s}_3^*/\sqrt{2} & \tilde{s}_3^*/\sqrt{2} & (-\tilde{s}_1 - \tilde{s}_1^* + \tilde{s}_2 - \tilde{s}_2^*)/2 & (-\tilde{s}_2 - \tilde{s}_2^* + \tilde{s}_1 - \tilde{s}_1^*)/2 \\ \tilde{s}_3^*/\sqrt{2} & -\tilde{s}_3^*/\sqrt{2} & (\tilde{s}_2 + \tilde{s}_2^* + \tilde{s}_1 - \tilde{s}_1^*)/2 & -(\tilde{s}_1 + \tilde{s}_1^* + \tilde{s}_2 - \tilde{s}_2^*)/2 \end{bmatrix} \Rightarrow \text{Space}$$

$$\Downarrow$$

$$\text{Time} \tag{6.110}$$

Table 6.2 summarizes the parameters of five space–time block codes: the Alamouti code **S** of Eq. (6.88), the generalized complex orthogonal codes of Eqs. (6.107) and (6.108), and the generalized complex orthogonal codes of the sporadic variety given in Eqs. (6.109) and (6.110).

TABLE 6.2 Summary of the Parameters of Different Space–Time Block Codes.

Space–time code	Number of transmit antennas, N_t	Number of transmitted symbols, l	Number of time slots, m	Rate, $\kappa = \frac{l}{m}$
S	2	2	2	1
$\boldsymbol{G}_3$	3	4	8	1/2
$\boldsymbol{G}_4$	4	4	8	1/2
$\boldsymbol{H}_3$	3	3	4	3/4
$\boldsymbol{H}_4$	4	3	4	3/4

Maximum-likelihood decoding of the space–time block codes $\boldsymbol{G}_3$, $\boldsymbol{G}_4$, $\boldsymbol{H}_3$, and $\boldsymbol{H}_4$ is achieved by using only linear processing at the receiver by virtue of the complex orthogonality of these four codes in the temporal sense. (For details of these decoding algorithms, see Problems 6.40 and 6.41; and Note 15.)

6.6.5 Performance Comparisons of Different Space–Time Block Codes Using a Single Receiver

In this section, we present computer simulation results that compare the five different space–time block codes **S**, $\boldsymbol{G}_3$, $\boldsymbol{G}_4$, $\boldsymbol{H}_3$, and $\boldsymbol{H}_4$, for modulation schemes that produce 3 bits/s/Hz, 2 bits/s/Hz, and 1 bit/s/Hz, assuming a single receive antenna. The incoming binary data stream is space–time block encoded in accordance with the scheme shown in Fig. 6.16.

The simulation results presented in Fig. 6.21 compare the symbol error rates for a data transmission of 3 bits/s/Hz for the following coding–modulation schemes:

(a) No coding and 8-PSK

(b) Alamouti code **S**, using two transmit antennas and 8-PSK

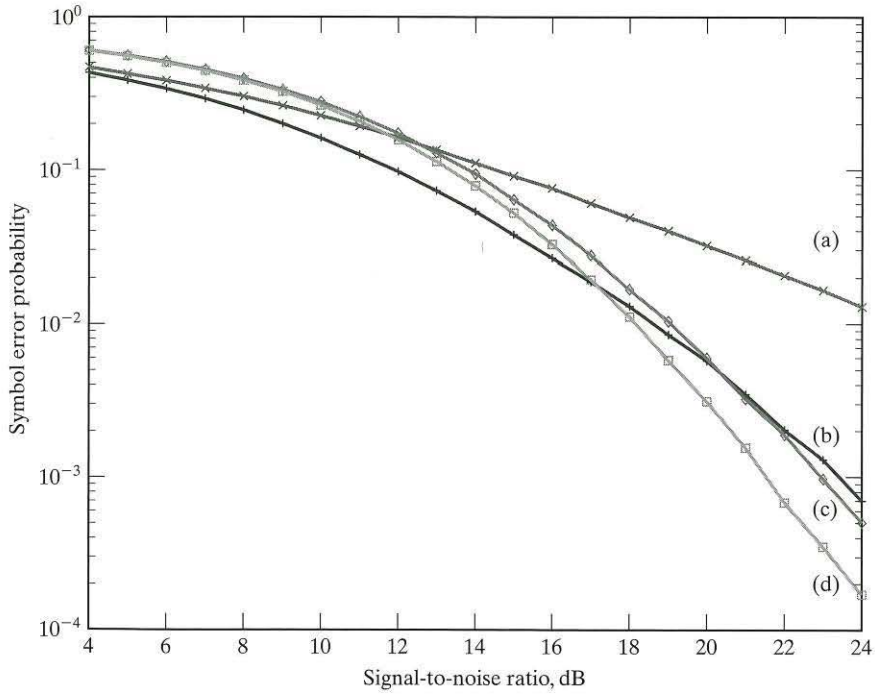


FIGURE 6.21 Comparison of the bit error rate performance of different space–time block codes:
(a) No space-time coding.
(b) Alamouti code (8-PSK).
(c) Sporadic code $\mathbf{H}_3$ (16-QAM).
(d) Sporadic code $\mathbf{H}_4$ (16-QAM).

(c) Sporadic space–time block code $\boldsymbol{H}_3$, using three transmit antennas and 16-QAM

(d) Sporadic space–time block code H_4, using four transmit antennas and 16-QAM

Recognizing that the code rate for both $\boldsymbol{H}_3$ and $\boldsymbol{H}_4$ is 3/4, we find that the effective transmission rate is 3 bits/s/Hz. From the figure, we see that, at a symbol error rate of 10^{-3}, the 16-QAM, rate-3/4 code $\mathbf{H}_4$ combination provides a gain of about 2 dB over the 8-PSK, full-rate Alamouti combination.

Figure 6.22 compares the bit error rates for a data transmission of 2 bits/s/Hz for the following coding–modulation schemes:

(a) No coding and QPSK

(b) Alamouti code **S**, using two transmit antennas and QPSK

(c) A generalized orthogonal space–time block code $\boldsymbol{G}_3$, using three transmit antennas and 16-QAM

(d) A generalized orthogonal space–time block code $\boldsymbol{G}_4$, using four transmit antennas and 16-QAM

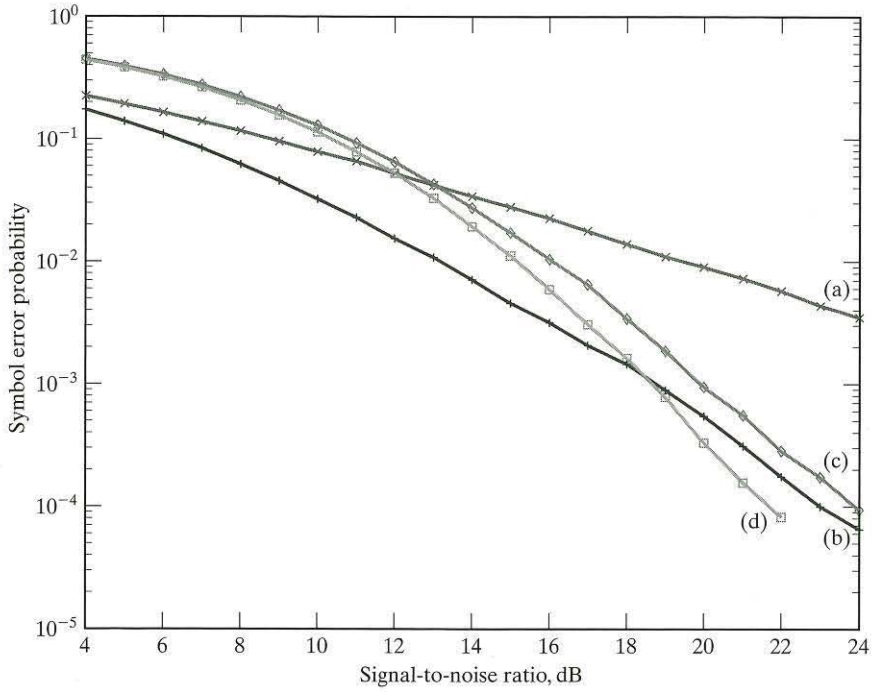


FIGURE 6.22 Comparison of bit error rate performance of different space–time block codes:
(a) No space-time coding.
(b) Alamouti code (QPSK).
(c) Code $\mathbf{G}_3$ (16-QAM).
(d) Code $\mathbf{G}_4$ (16-QAM).

With the chosen constellations, the effective transmit rate in each case is 2 bits/s/Hz. The figure shows that, at a symbol error rate of 10^{-4}, the 16-QAM, rate-1/2 code $\boldsymbol{G}_4$ combination produces a gain of about 1 dB over the QPSK, full-rate Alamouti combination. At a symbol error rate of 10^{-5}, the gain advantage grows to about 5 dB (not shown in Fig.6.22).

Finally, Fig. 6.23 compares the bit error rates for a data transmission of 1 bit/s/Hz for the following coding–modulation schemes:

- No coding and BPSK
- Alamouti code **S**, using two transmit antennas and BPSK
- A generalized space–time block code $\boldsymbol{G}_3$, using three transmit antennas and QPSK
- A generalized space–time block code $\boldsymbol{G}_4$, using four transmit antennas and QPSK

In each case, the effective transmission rate is 1 bit/s/Hz. For a bit error rate of 10^{-4}, the figure shows that the QPSK, $\boldsymbol{G}_4$ combination provides a gain of about 5 dB over the BPSK, **S** combination. At a symbol error rate of 10^{-5}, the gain advantage grows to about 7.5 dB (not shown in Fig. 6.23).

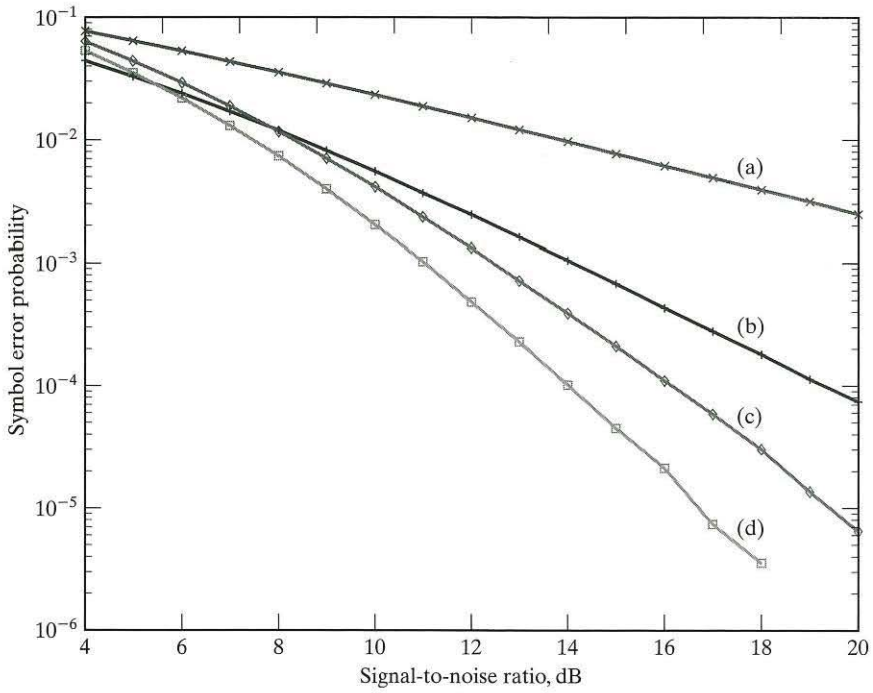


FIGURE 6.23 Comparison of bit error rate performance of different space–time block codes:
(a) No space-time coding.
(b) Alamouti code (binary PSK).
(c) Generalized code $\mathbf{G}_3$ (QPSK).
(d) Generalized code $\mathbf{G}_4$ (QPSK).

6.7 DIFFERENTIAL SPACE–TIME BLOCK CODES

In light of the material presented in Section 6.5 on the ergodic capacity of a MIMO wireless channel, the case for a substantial increase in spectral efficiency of the system rests on the premise that the receiver has "knowledge," or, more precisely, a "near-perfect estimate," of the channel matrix $\mathbf{H}$. Clearly, such a provision imposes a certain amount of overhead on the design and implementation of the receiver. (The topic of channel estimation was discussed in Section 3.11.) We may eliminate the need for channel estimation and thereby simplify the receiver design by using a *differential space–time block coding scheme.*[16]

The encoding system we have in mind builds on two concepts:

1. Alamouti's space–time transmit-diversity block code
2. Differential space–time block coding

The Alamouti code, involving the use of two transmit antennas and a single receive antenna in its most basic form, was discussed in Section 6.6. The space–time version of differential coding is discussed next.

6.7.1 Differential Space–Time Block Coding

In the Alamouti code, the two signals (symbols) $\tilde{s}_{1,t}$ and $\tilde{s}_{2,t}$ are transmitted on two separate antennas at time t, followed by the transmission of the related pair of signals $-\tilde{s}^*_{2,t+1}$ and $\tilde{s}^*_{1,t+1}$, at time $t+1$. Note that

- We have appended the subscripts t and $t+1$ to the pertinent signals to distinguish a signal transmission from the one following it.
- We have set the transmitted signal duration T equal to unity to simplify the presentation.

We assume that the total transmitted power, namely, $|\tilde{s}_{1,t}|^2 + |\tilde{s}_{2,t}|^2$, is maintained constant. To further simplify matters without loss of generality, we normalize the total transmitted power to unity. Under this assumption, it follows that the row vectors $[\tilde{s}_{1,t}, \tilde{s}_{2,t}]$ and $[-\tilde{s}^*_{2,t+1}, \tilde{s}^*_{1,t+1}]$ form an *orthonormal set.* Consequently, *any other pair of transmitted signals can be represented as a linear combination of this orthonormal set.* Specifically, given the new pair of complex signals $\tilde{s}_{3,t+2}$ and $\tilde{s}_{4,t+2}$ to be transmitted at time $t+2$, we may express them as the row vector

$$[\tilde{s}_{3,t+2}, \tilde{s}_{4,t+2}] = a_{1,t+2}[\tilde{s}_{1,t}, \tilde{s}_{2,t}] + a_{2,t+2}[-\tilde{s}^*_{2,t}, \tilde{s}^*_{1,t}] \tag{6.111}$$

where $a_{1,t+2}$ and $a_{2,t+2}$ are coefficients of the linear combination. These two coefficients are defined as the *inner products* of the row vector $[\tilde{s}_{3,t+2}, \tilde{s}_{4,t+2}]$ with the previously transmitted row vectors $[\tilde{s}_{1,t}, \tilde{s}_{2,t}]$ and $[-\tilde{s}^*_{2,t+1}, \tilde{s}^*_{1,t+1}]$, respectively;

that is,

$$\begin{aligned}[a_{1,t+2}, a_{2,t+2}] &= [\tilde{s}_{3,t+2}, \tilde{s}_{4,t+2}]\left[[\tilde{s}_{1,t}, \tilde{s}_{2,t}]^{\dagger}, [-\tilde{s}^{*}_{2,t+1}, \tilde{s}^{*}_{1,t+1}]^{\dagger}\right] \\ &= [\tilde{s}_{3,t+2}, \tilde{s}_{4,t+2}]\begin{bmatrix} \tilde{s}^{*}_{1,t} & -\tilde{s}_{2,t+1} \\ \tilde{s}^{*}_{2,t} & \tilde{s}_{1,t+1} \end{bmatrix} \\ &= [\tilde{s}_{3,t+2}, \tilde{s}_{4,t+2}]\mathbf{S}^{\dagger}_{t,t+1}\end{aligned} \tag{6.112}$$

where $\mathbf{S}^{\dagger}_{t,t+1}$ is the Hermitian transpose of the two-by-two transmitted signal matrix of the Alamouti code constructed at times t and $t+1$; see Eq. (6.89). Similarly, we may write

$$[-a^{*}_{2,t+3}, a^{*}_{1,t+3}] = [-\tilde{s}^{*}_{4,t+3}, \tilde{s}^{*}_{3,t+3}]\mathbf{S}^{\dagger}_{t,t+1} \tag{6.113}$$

Combining Eqs. (6.112) and (6.113) into a single matrix relation, we may formulate the *coefficients matrix*

$$\begin{aligned}\mathbf{A}_{t+2,t+3} &= \begin{bmatrix} a_{1,t+2} & a_{2,t+2} \\ -a^{*}_{2,t+3} & a^{*}_{1,t+3} \end{bmatrix} \\ &= \mathbf{S}_{t+2,t+3}\mathbf{S}^{\dagger}_{t,t+1}\end{aligned} \tag{6.114}$$

where $\mathbf{S}_{t+2,t+3}$ is the transmitted signal matrix constructed at times $t+2$ and $t+3$; see Eq. (6.88). Equation (6.114) states that the coefficients matrix is a *product of two orthogonal Alamouti (quaternionic) matrices.*

Problem 6.15 Show that the coefficients matrix $\mathbf{A}_{t+2,t+3}$ is an orthonormal matrix; that is, show that the inverse of $\mathbf{A}_{t+2,t+3}$ equals its Hermitian transpose. ■

Since $\mathbf{S}_{t,t+1}$ is a unitary matrix by virtue of orthonormality of its two constituent row vectors, it follows that

$$\mathbf{S}^{-\dagger}_{t,t+1} = \mathbf{S}_{t,t+1}$$

We may therefore solve Eq. (6.114) for the *transmission matrix* $\mathbf{S}_{t+2,t+3}$ by writing

$$\begin{aligned}\mathbf{S}_{t+2,t+3} &= \mathbf{A}_{t+2,t+3}\mathbf{S}^{-\dagger}_{t,t+1} \\ &= \mathbf{A}_{t+2,t+3}\mathbf{S}_{t,t+1}\end{aligned} \tag{6.115}$$

Equation (6.115) provides the *basis for differential space–time block encoding at the transmitter.*

The corresponding basis for differential decoding at the receiver lies in Eq. (6.114), modified by the channel. To be specific, we assume that the wireless link is quasi-static with the channel matrix $\mathbf{H}$ remaining essentially constant over two consecutive blocks of signal transmission (i.e., over the time interval $(t,t+3)$). Then, in the absence of channel noise, the received signal matrix in response to the transmitted signal matrix $\mathbf{S}_{t,\,t+1}$ is given by (see Problem 6.43)

$$\mathbf{X}_{t,\,t+1} = \mathbf{S}_{t,\,t+1}\mathbf{H} \tag{6.116}$$

Similarly, the received signal matrix in response to the next transmitted signal matrix $\mathbf{S}_{t+2,\,t+3}$ is given by

$$\mathbf{X}_{t+2,\,t+3} = \mathbf{S}_{t+2,\,t+3}\mathbf{H} \tag{6.117}$$

With the mathematical structure of Eq. (6.114) in mind, we use Eqs. (6.116) and (6.117) to construct the new two-by-two matrix

$$\begin{aligned}\mathbf{Y}_{t+2,\,t+3} &= \mathbf{X}_{t+2,\,t+3}\mathbf{X}_{t,\,t+1}^{\dagger}\\ &= \mathbf{S}_{t+2,\,t+3}\mathbf{H}(\mathbf{S}_{t,\,t+1}\mathbf{H})^{\dagger}\\ &= \mathbf{S}_{t+2,\,t+3}\mathbf{H}\mathbf{H}^{\dagger}\mathbf{S}_{t,\,t+1}^{\dagger}\end{aligned} \tag{6.118}$$

(Note that the elements of the matrix $\mathbf{Y}$ should not be confused with the y's defined in Eq. (6.99).)

From the solution to Problem 6.12, we note that

$$\mathbf{H} = \begin{bmatrix} \alpha_1 e^{j\theta_1} & \alpha_2 e^{j\theta_2} \\ \alpha_2 e^{-j\theta_2} & -\alpha_1 e^{-j\theta_1} \end{bmatrix} \tag{6.119}$$

and

$$\mathbf{H}^{\dagger}\mathbf{H} = \mathbf{H}\mathbf{H}^{\dagger} = (\alpha_1^2 + \alpha_2^2)\mathbf{I} \tag{6.120}$$

where $\mathbf{I}$ is the two-by-two identity matrix. Accordingly, Eq. (6.118) reduces to

$$\begin{aligned}\mathbf{Y}_{t+2,\,t+3} &= (\alpha_1^2 + \alpha_2^2)\mathbf{S}_{t+2,\,t+3}\mathbf{S}_{t,\,t+1}^{\dagger}\\ &= (\alpha_1^2 + \alpha_2^2)\mathbf{A}_{t+2,\,t+3}\end{aligned} \tag{6.121}$$

where, in the second line, we made use of Eq. (6.114).

Equation (6.121) forms the *basis for differential space–time block decoding at the receiver*. Here again, in light of the scaling factor $(\alpha_1^2 + \alpha_2^2)$, the receiver exhibits a two-level diversity gain in the same way that the coherent Alamouti receiver discussed in Section 6.6 does.

Signal transmission begins by sending an arbitrary pair of signals $\tilde{s}_{1,0}$ and $\tilde{s}_{2,0}$ at time $t = 0$, followed by sending the related pair of signals $-\tilde{s}_{2,1}^*$ and $\tilde{s}_{1,1}^*$ at time $t = 1$. No information is conveyed by sending these signals; rather, the two transmissions at times $t = 0$ and $t = 1$ provide the receiver with a known *frame of reference* for facilitating the differential space–time block decoding process. The transmission of information-bearing data begins at $t = 2$, at which time the receiver commences the differential decoding process.

In what follows, we restrict the discussion to M-ary PSK. This method of modulation, involving M signal points in a two-dimensional space, befits its use for constructing the two-by-two transmitted signal matrix characterizing the Alamouti code. With two sets of signal points involved in the formulation of Eqs. (6.115) and (6.121), we may identify two signal spaces. One, denoted by $\mathcal{A}$, is spanned by the pair of complex coefficients (a_1, a_2) constituting the matrix $\mathbf{A}$. The second, denoted by $\mathcal{S}$, is spanned by the complex signals $\tilde{s}_1, \tilde{s}_2$ constituting the matrix $\mathbf{S}$. These two signal spaces have the following properties:

Property 1.

With M-ary PSK as the method of modulation used to transmit the Alamouti code, the points representing the signal space $\mathcal{S}$ are uniformly distributed on a circle of unit radius (assuming that the transmitted signal power is normalized to unity). Correspondingly, the points representing the signal space $\mathcal{A}$ constitute a quadrature amplitude modulation (QAM) constellation.

Property 2.

The minimum distance between the points in the signal space $\mathcal{S}$ is equal to the minimum distance between the points in the signal space $\mathcal{A}$.

Property 3.

The constellation of points in the input space $\mathcal{A}$ involves an expansion in the size of the alphabet, compared with that in the constellation of points in the transmitted signal space $\mathcal{S}$.

To construct the matrix $\mathbf{A}$, we need a *bijective mapping* of blocks of $2b$ bits onto the signal space $\mathcal{A}$. The mapping is bijective in the sense that it is *one-to-one and onto*, as shown in the next two examples. The first example illustrates the constellation expansion property; the second illustrates the use of Gray coding as a principled way of constructing the bijective mapping.

EXAMPLE 6.6 Constellation Expansion

Let $\tilde{s}_0$ and $\tilde{s}_1$ denote the pair of symbols transmitted at times $t = 0$ and $t = 1$, respectively, followed by the transmission of symbols $\tilde{s}_2$ and $\tilde{s}_3$ at times $t = 2$ and $t = 3$. We may thus express the

corresponding matrices as

$$\mathbf{S}_{0,1} = \begin{bmatrix} \tilde{s}_0 & \tilde{s}_1 \\ -\tilde{s}_1^* & \tilde{s}_0^* \end{bmatrix}$$

and

$$\mathbf{S}_{2,3} = \begin{bmatrix} \tilde{s}_2 & \tilde{s}_3 \\ -\tilde{s}_3^* & \tilde{s}_2^* \end{bmatrix}$$

Accordingly, the use of Eq. (6.114) yields

$$\begin{aligned} \mathbf{A}_{2,3} &= \mathbf{S}_{2,3}\mathbf{S}_{0,1}^{\dagger} \\ &= \begin{bmatrix} \tilde{s}_2 & \tilde{s}_3 \\ -\tilde{s}_3^* & \tilde{s}_2^* \end{bmatrix} \begin{bmatrix} \tilde{s}_0^* & -\tilde{s}_1 \\ \tilde{s}_1^* & \tilde{s}_0 \end{bmatrix} \\ &= \begin{bmatrix} \tilde{s}_2\tilde{s}_0^* + \tilde{s}_3\tilde{s}_1^* & -\tilde{s}_2\tilde{s}_1 + \tilde{s}_3\tilde{s}_0 \\ -\tilde{s}_3^*\tilde{s}_0^* + \tilde{s}_2^*\tilde{s}_1^* & \tilde{s}_3^*\tilde{s}_1 + \tilde{s}_2^*\tilde{s}_0 \end{bmatrix} \end{aligned}$$

To proceed further, consider the simple example of BPSK as the method of modulation used to transmit the Alamouti code, exemplified by the matrices $\mathbf{S}_{0,1}$ and $\mathbf{S}_{2,3}$. Specifically, suppose the symbols $\tilde{s}_0$, $\tilde{s}_1$, $\tilde{s}_2$, and $\tilde{s}_3$ come from the constellation $\{-1, +1\}$ depicted in Fig. 6.24(a). Then, clearly, the four elements of the coefficients matrix $\mathbf{A}_{2,3}$ come from the constellation $\{-2, 0, +2\}$ depicted in Fig. 6.24(b). Comparing these two constellations, we immediately see that there is an expansion in the size of the alphabet, in accordance with Property 3.

The constellation expansion illustrated in Fig. 6.44 arises from the fact that, although the BPSK forms a multiplicative group (e.g., the QPSK constellation is the product of two BPSK constellations), it does not form an additive group over the integers. In contrast, the matrix $\mathbf{A}$ involves both multiplicative and additive groups—hence the constellation expansion.

It is also noteworthy that the minimum distance between the signal points in the input constellation $\mathcal{A}$ of Fig. 6.24(b) is the same as the minimum distance between the transmitted signal points in the constellation $\mathcal{S}$ of Fig. 6.24(a) in accordance with Property 2. In both cases, the minimum distance is 2. ■

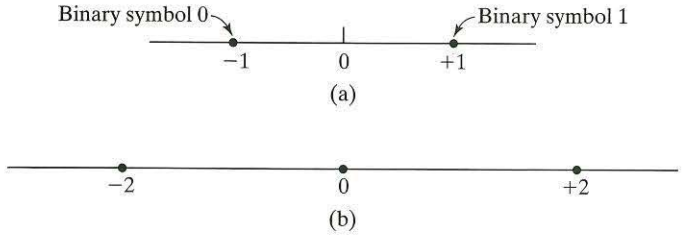


FIGURE 6.24 a) Signal constellation for BPSK.
(b) Expanded constellation for input data.

For a demonstration of Property 1, the reader is referred to Problem 6.45, in which it is requested that the reader start with a 4-QAM (i.e., QPSK) constellation for signal space $\mathcal{S}$ and arrive at a 9-QAM constellation for signal space $\mathcal{A}$. The solution to Problem 6.45 provides further confirmation of Properties 2 and 3.

EXAMPLE 6.7 Gray Coding for Bijective Mapping

To continue with the differential space–time coding using BPSK, we need a frame of reference for signal transmission at times $t = 0$ and $t = 1$. With the constellation of Fig. 6.44(a) as the polar basis for representing the BPSK, we choose the dibit 00 as the frame of reference. Then, with the symbol 1 representing the first 0 bit transmitted on antenna 1 and the symbol –1 representing the second 0 bit transmitted on antenna 2, the corresponding signal matrix is

$$\mathbf{S}_{0,1} = \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix} \quad \text{dibit } 00$$

With this frame of reference fixed, we may now go on to determine the values assumed by the coefficients matrix $\mathbf{A}_{2,3}$ for the transmission of Gray-encoded dibits 00, 01, 11, and 10, using Eq. (6.114). The results of this calculation are summarized as follows:

1. *Transmitted data dibit* 00

$$\mathbf{S}_{2,3} = \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}$$

$$\mathbf{A}_{2,3} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

2. *Transmitted data dibit* 01

$$\mathbf{S}_{2,3} = \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}$$

$$\mathbf{A}_{2,3} = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}$$

3. *Transmitted data dibit* 11

$$\mathbf{S}_{2,3} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$\mathbf{A}_{2,3} = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$$

4. *Transmitted data dibit* 10

$$\mathbf{S}_{2,3} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{A}_{2,3} = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$$

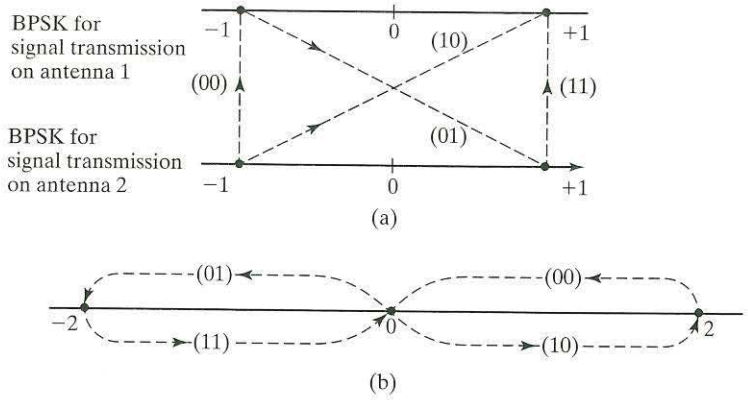


FIGURE 6.25 (a) Gray encoding of input dibits.
(b) Gray encoding of transmitted dibits.

TABLE 6.3 Signal Transmissions, Assuming 00 as the Frame of Reference.

Transmitted dibit	Coefficients at time $t = 2$		Transmitted signals at time $t = 2$	
	a_2	a_3	$\tilde{s}_2$	$\tilde{s}_3$
00	2	0	−1	−1
01	0	−2	−1	1
11	−2	0	1	1
10	0	2	1	−1

Figure 6.25 illustrates the bijective mapping of the input signal space $\mathcal{A}$ onto the transmitted signal space $\mathcal{S}$. Part (a) refers to the input dibits, and part (b) refers to the corresponding transmitted pair of bits transmitted on antennas 1 and 2.

Table 6.3 summarizes (1) the pair of coefficients a_2 and a_3 representing an incoming dibit, and (2) the corresponding pair of symbols $\tilde{s}_2$ and $\tilde{s}_3$ transmitted on antennas 1 and 2, assuming the dibit 00 as the frame of reference. ■

Problem 6.16 Fill in the details leading to the formulation of Fig. 6.25 and Table 6.3. ■

Problem 6.17 The results presented in Fig. 6.25 and Table 6.3 assume the frame of reference 00. Repeat Problem 6.16, this time using the frame of reference 11. ■

6.7.2 Transmitter and Receiver Structures

As mentioned previously, Eq. (6.115) forms the mathematical basis for constructing the transmitter. On this basis, we may construct the block diagram depicted in Fig. 6.26(a), which consists of the following parts:

1. *A mapper*, which generates the entries that make up matrix $\mathbf{A}_{t+2,t+3}$ in response to the incoming pair of data bits sent by the source of information at times $t+2$ and $t+3$.
2. *A differential encoder*, which transforms the matrix $\mathbf{A}_{t+2,t+3}$ into the matrix $\mathbf{S}_{t+2,t+3}$. The two row vectors of matrix $\mathbf{S}_{t+2,t+3}$ are transmitted on a pair of antennas at times $t+2$ and $t+3$. The differential encoder itself consists of two components:
   2.1. *A delay unit* $z^{-2}\mathbf{I}$, which feeds back the matrix $\mathbf{S}_{t,t+1}$ (i.e., the previous value of $\mathbf{S}_{t+2,t+3}$) to the input of the differential encoder.
   2.2. *A multiplier*, which multiplies the matrix inputs $\mathbf{A}_{t+2,t+3}$ and $\mathbf{S}_{t,t+1}$ to provide the transmitted signal matrix $\mathbf{S}_{t,t+3}$ in accordance with Eq. (6.115).

To construct the receiver, we return to Eq. (6.121). On the basis of this equation, we may construct the block diagram of the receiver of Fig. 6.26(b), which consists of the following functional blocks:

1. *A differential decoder*, which itself consists of two components:
   1.1. *A delay unit* $z^{-2}\mathbf{I}$, which feeds forward the matrix $\mathbf{X}_{t,t+1}$ (i.e., the previous value of the received signal matrix $\mathbf{X}_{t+2,t+3}$).
   1.2. A *multiplier*, which multiplies the matrices $\mathbf{X}_{t+2,t+3}$ and $\mathbf{X}_{t,t+1}$ to produce the new matrix $\mathbf{Y}_{t+2,t+3}$.
2. *A signal estimator*, which computes the matrix $\hat{\mathbf{A}}_{t+2,t+3}$ that is closest to $\mathbf{Y}_{t+2,t+3}$ in terms of Euclidean distance; the estimate so computed is denoted by $\hat{\mathbf{A}}_{t+2,t+3}$.
3. *An inverse mapper*, which operates on the estimate $\hat{\mathbf{A}}_{t+2,t+3}$ to produce corresponding estimates of the original pair of data bits transmitted at times $t+2$ and $t+3$.

6.7.3 Noise Performance

In deriving Eq. (6.121), we ignored the presence of additive channel noise. It is in recognition of the unavoidable presence of channel noise that we spoke of "estimates" of transmitted data bits in describing the structure of the receiver.

How, then, does channel noise affect the performance of the receiver depicted in Fig. 6.26? To answer this question, we remind ourselves of the classical problem pertaining to the scalar case of differential phase-shift keying (DPSK). In this context, it is well known that the performance of a DPSK receiver corrupted by additive white Gaussian noise is worse than that of coherent BPSK. (See Section 3.11) So it is also for the differential space–time block coding system of Fig. 6.26, compared with the coherent version of the system discussed in Section 6.6.

Using computer simulations, Fig. 6.27 compares the noise performance of the differential space–time block coding system of Fig. 6.26 with that of the coherent version of the system.The simulations are based on the following configuration:

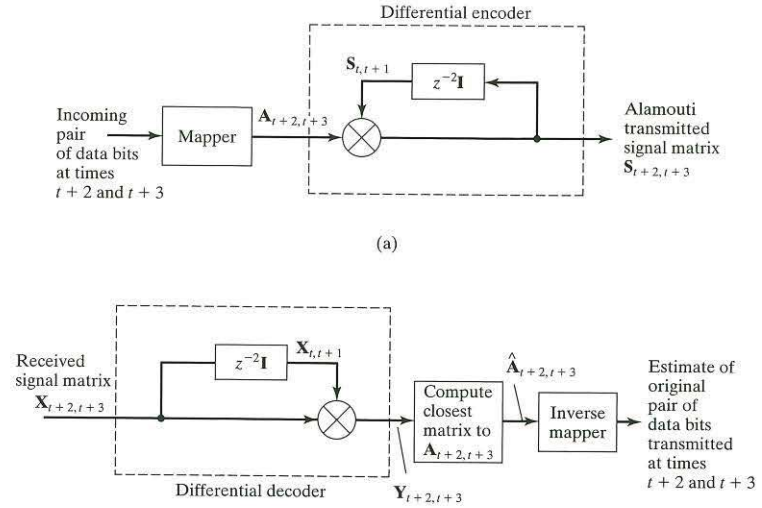


FIGURE 6.26 Block diagrams of (a) differential space–time block encoder. (b) differential space–time block decoder, where z^{-1} = unit-delay operator and $\mathbf{I}$ = two-by-two identity matrix.

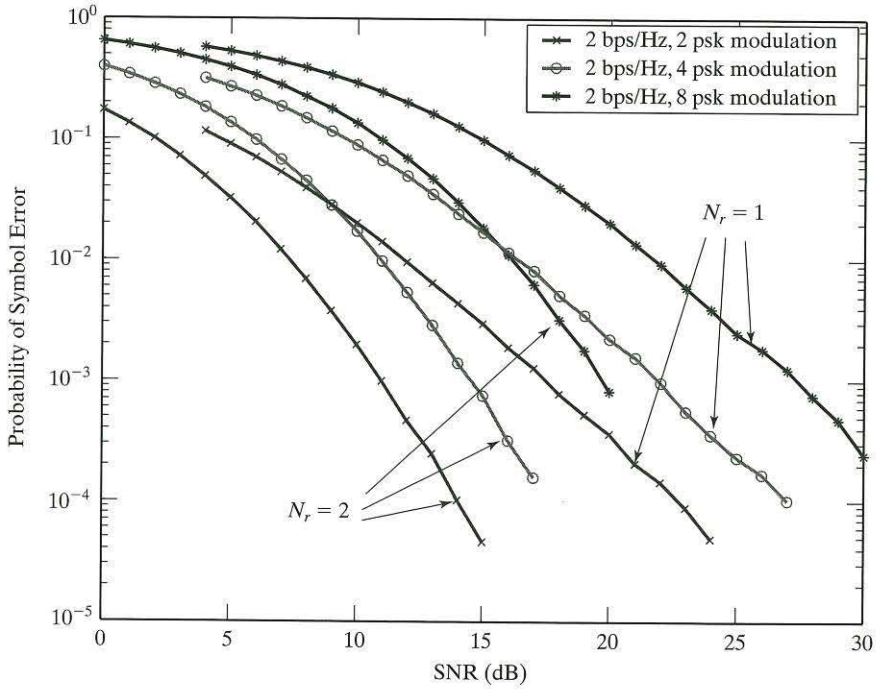


FIGURE 6.27 Noise–performance comparison of differential space–time block coding for specific M-ary PSK modulation strategies in a Rayleigh-fading environment.

- A single bit of data sent on each of the two transmit antennas
- BPSK modulation for transmitting the complex symbols $\tilde{s}_1$ and $\tilde{s}_2$ at time t, followed by the transmission of $-\tilde{s}_2^*$ and $\tilde{s}_1^*$ at time $t + 1$

The simulation results presented in Fig. 6.27 indicate a loss of 3 dB in receiver performance that is incurred through the use of differential space–time block coding. This loss in performance may be justified by formulating the additive noise corrupting the received signal matrix $\mathbf{X}$. The formulation is discussed in Problem 6.47, where it is shown that, under the assumption of a high signal-to-noise ratio, the average power of the additive noise is increased by about 3 dB, compared with that of the corresponding coherent receiver.

6.8 SPACE-DIVISION MULTIPLE ACCESS AND SMART ANTENNAS

The material presented up to this point has focused on the use of *space* as a basis for antenna diversity, the use of which improves spectral efficiency by counteracting the multipath-channel-fading problem. In this section, we discuss the role of space in cellular systems, the development of which has also resulted in a multifold increase in spectral efficiency by using directional antennas.

In *cellular systems*, a few channels are broadcast by the base station or shared by all users on the uplink. The majority of the traffic-bearing (and revenue-generating) channels are point-to-point, between a base station and a single user terminal, in which case we speak of the communications being *directional* in nature. The recognition that *user terminals can be spatially separated by virtue of their angular directions is the basis of space-division multiple access (SDMA)*, a technique that relies on the use of *directional antennas* to distinguish among users. A simple example is presented in Fig. 6.28, which shows a base station in a cellular system with sector antennas; that is, each antenna covers one sector—in this case, 120°— of the cell. In the illustration, each base station requires three nonoverlapping antennas, each with a field of view of 120°.

Consider the case in which there is a one-in-seven reuse pattern, as suggested by the figure. If there are N users per cell, then the power radiated on one of the sector antennas is $(N/3)P_T G_T$, where P_T is the average power per user and G_T is the antenna gain. To close the communications link, $P_T G_T$ must be the same whether the antenna is directional or omnidirectional. Consequently, we conclude that the total power radiated with a sector antenna is one-third of that produced by an omnidirectional antenna. By extension, a user terminal receiver suffers only one-third of the interference that would be produced by omnidirectional base-station antennas with the same number of users.

In the uplink, all user terminals have omnidirectional antennas, but only one-third of them are in the field of view of the base-station antenna. So the interference is reduced by two-thirds in this direction as well.

The general conclusion from this discussion is that, with 120° sector antennas at the base station, the number of user terminals can be tripled relative to the omnidirectional case and still maintain the same interference levels. There are many advantages to this approach:

- It can be applied with any of the multiple-access strategies discussed previously: FDMA, TDMA, or CDMA.
- It allows multiple users to operate on the same frequency and/or time slot in the same cell.
- It thus leads to more users in the same spectrum and improved capacity.
- The technology can be applied at the base station without affecting the mobile terminals.

Although the same idea may also be applied to the mobile terminal, at present it is more difficult to build a commercially acceptable mobile terminal that has a directional antenna. However, this is an area of active investigation.

With conventional base stations using omnidirectional antennas, when user density grows beyond the capacity of a single cell, the growth is accommodated by dividing the initial cell into a number of smaller cells in a process known as *cell splitting*. Power control is used to reduce the interference among these smaller cells. Although sector antennas are more expensive than omnidirectional antennas, it is still more economical to add sector antennas than it is to add new base stations.

The same technology can also be applied to satellite systems in which the field of view of the satellite is divided between a number of spotbeams and, depending upon the isolation between the beams, frequency and/or time slots may be reused in all or some of the beams.

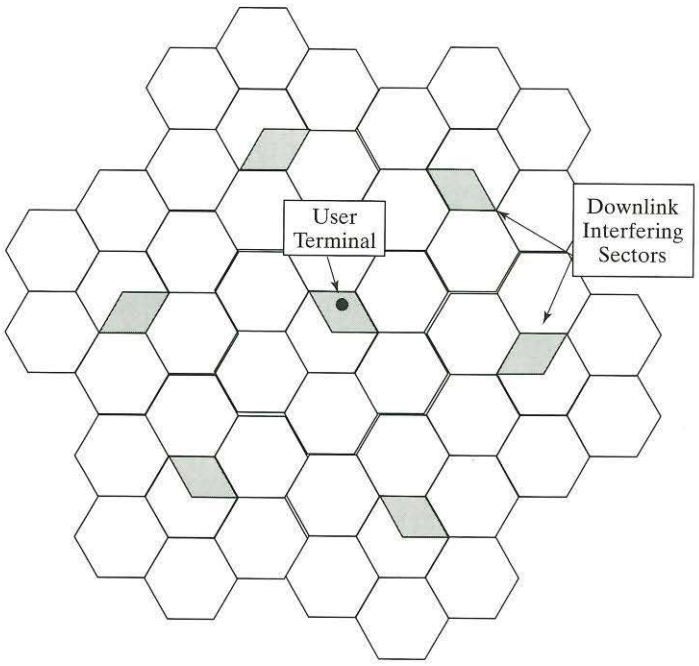


FIGURE 6.28 Cellular system with 120° sector antennas.

SDMA relies on *smart antennas*, in the sense that it takes advantage of the directional nature of radio communications. Some examples of smart antennas are the following:

1. The simplest example of a smart antenna is the *sector antenna* just described. This antenna provides significant capacity gains simply by dividing the service area of each base station into three (or more) angular sections with a significant amount of isolation between them.
2. *Switched-beam antennas* are the next step in the evolution of smart antennas. These antennas have a number of fixed beams that cover 360°. Switched-beam antennas are typically narrower than sector antennas. The receiver selects the beam that provides the best signal and interference reduction.
3. The most advanced example is the *adaptive antenna*. This antenna dynamically adjusts its pattern to minimize the effects of noise, interference, and multipath. With adaptive antennas, there is one beam for each user.

In Section 6.8.1, we will discuss how an adaptive antenna may be implemented. There are many advantages to smart antennas for mobile applications:

- *Greater range.* Since the antennas are directional, they have larger gains and can therefore provide a stronger signal strength for the same transmit power.
- *Fewer base stations.* For areas with a low user density, fewer base stations are required, because the existing base station has a greater range. In areas with a high user density, there is less interference, due to the greater user isolation provided by the directional antennas. Hence, a single base station can serve more users.
- *Better building penetration.* This potential benefit is due to the greater signal strength and increased transmitter gain.
- *Less sensitivity to power control errors.* This additional benefit is due to the better isolation among different user signals.
- *More responsive to traffic hot spots.* In areas such as airports and conference centers, user densities can become quite high, and directional antennas allow one or a small number of base stations to service these areas better.

SDMA improves system capacity by allowing greater spectrum reuse, through (1) minimization of the effects of interference and, (2) increasing signal strength for both the user terminal and the base station.

6.8.1 Antenna Arrays[17]

In this section, we analyze the behavior of a linear array of antenna elements receiving a signal, as illustrated in Fig. 6.29. For simplicity, we assume that the transmit antenna is omni-directional, but, in general, it could be directional. The key assumption is that the distance between the transmitter and receiver is large enough for the emitted wave to be viewed as a planar wavefront for the purposes of antenna analysis.

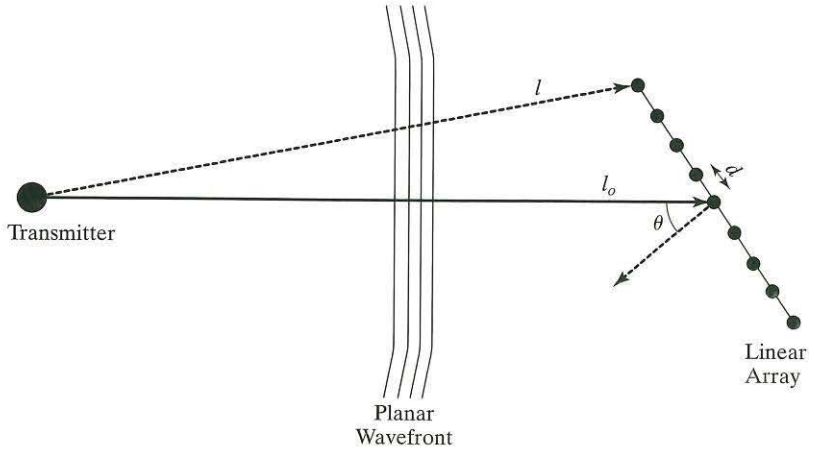


FIGURE 6.29 Plane wave incident on a linear antenna array.

The receiving antenna is positioned at an angle θ with respect to the linear array, as shown in the figure. Let the complex envelope of the transmitted signal be denoted by

$$\tilde{s}(t) = m(t)e^{j2\pi ft} \tag{6.122}$$

where $m(t)$ is the modulating signal and f is the transmission frequency. Then the received symbols at a distance l from the transmitter is given by

$$\tilde{r}(t, l) = A(l)m(t - l/c)e^{j2\pi f(l/c)} \tag{6.123}$$

where c is the speed of light and $A(l)$ is the path attenuation as a function of the distance l. (The symbol l used here to denote distance should not be confused with that used in section 6.6 to denote the number of transmitted symbols.) At this point, we make several key assumptions:

- The incident field is a plane wave. This assumption relies on the source being sufficiently distant and the array being physically small enough.
- The spacing of the antenna elements is small enough that there is no variation in amplitude among the signals received at the different elements. That is, the attenuation $A(l) \approx A(l_0) \equiv A_0$ is the same for all antenna elements.
- The bandwidth of the modulating signal is small compared with the carrier frequency f. This assumption implies that there will be little variation in the modulation over the physical dimensions of the antenna; that is,

  $$m(t - l/c) \approx m(t - l_0/c) \equiv m_0(t)$$

  for each element of the array. (c denotes the speed of light.)
- There is no mutual coupling between the antenna elements; that is, they can be treated independently.

Under these assumptions, the analysis depends solely on the phase relationship of the different elements, in a manner similar to the manner in which Rayleigh fading depends on the phase relationship of the different multipath rays.

Assume that there are $N+1$ uniformly spaced antenna elements, where N is even and d is the antenna spacing. Let the antenna elements be numbered from $-N/2$ to $N/2$. If antenna element k is at distance l_k from the transmitting antenna, then the corresponding carrier phase is given by

$$\begin{aligned}\phi(t) &= 2\pi f(t - l_k/c) \\ &= 2\pi f(t - l_0/c) - 2\pi f(l_k - l_0)/c \\ &\equiv \phi_0(t) - 2\pi f \Delta l_k/c \\ &\equiv \phi_0(t) - \Delta\phi_k\end{aligned} \tag{6.124}$$

Since the relative distance of the kth element is $\Delta l_k = kd\sin\theta$, the corresponding phase offset at antenna element k is given by

$$\begin{aligned}\Delta\phi_k &= (2\pi f/c)(kd\sin\theta) \\ &= 2\pi\left(\frac{kd}{\lambda}\right)\sin\theta\end{aligned} \tag{6.125}$$

where λ is the wavelength of the radio transmission. Consequently, from the assumptions underlying Eq.(6.123), the received signal at element k is given by

$$\begin{aligned}s_k(t) &= s_0(t)e^{j2\pi\left(\frac{kd}{\lambda}\right)\sin\theta} \\ &= a_k(\theta)s_0(t)\end{aligned} \tag{6.126}$$

where

$$s_0(t) = A_0 m_0(t) e^{j\phi_0(t)} \tag{6.127}$$

is the attenuated and delayed version of the transmitted signal; in Eq. (6.127),

$$a_k(\theta) = e^{j2\pi\left(\frac{kd}{\lambda}\right)\sin\theta} \tag{6.128}$$

is a complex rotation. A phased array computes a linear sum of the signals received at each element shown in Fig. 6.29, yielding the received signal

$$\begin{aligned}r(t) &= \sum_{k=-N/2}^{N/2} w_k^* s_k(t) \\ &= (\mathbf{w}^\dagger \mathbf{a}) s_0(t)\end{aligned} \tag{6.129}$$

where

$$\mathbf{w} = [w_{-N/2}, ..., w_{N/2}]^T$$

is the *weighting vector* and

$$\mathbf{a}(\theta) = [a_{-N/2}(\theta),...,a_{N/2}(\theta)]^T$$

We have thus explicitly demonstrated the dependence of $\mathbf{a}$ on θ, the angle of arrival of the wavefront. The term $\mathbf{w}^\dagger\mathbf{a}$ is a *complex gain* that is applied to the received signal; it depends on the choice of the weighting vector $\mathbf{w}$ and angle θ.

EXAMPLE 6.8 Antenna Pattern with Uniform Weighting

Suppose that we select the weighting vector $\mathbf{w} = [1,1,...,1]^T$. Then, in this case, the complex gain of the receiving array antenna as a function of θ is written as

$$\begin{aligned} G(\theta) &= \mathbf{w}^\dagger\mathbf{a}(\theta) \\ &= \sum_{k=-N/2}^{N/2} e^{j2\pi\left(\frac{kd}{\lambda}\right)\sin\theta} \end{aligned} \tag{6.130}$$

If we set

$$\Theta = e^{j2\pi\left(\frac{d}{\lambda}\right)\sin\theta} \tag{6.131}$$

then we recognize that Eq. (6.130) is the sum of a geometric series, as shown by

$$\begin{aligned} G(\theta) &= \sum_{k=-N/2}^{N/2} \Theta^k \\ &= \Theta^{-N/2}\sum_{k=0}^{N} \Theta^k \\ &= \Theta^{-N/2}\left(\frac{1-\Theta^{N+1}}{1-\Theta}\right) \\ &= \frac{\Theta^{1/2}(\Theta^{-(N+1)/2} - \Theta^{(N+1)/2})}{\Theta^{1/2}(\Theta^{-1/2} - \Theta^{1/2})} \\ &= \frac{\Theta^{-(N+1)/2} - \Theta^{(N+1)/2}}{\Theta^{-1/2} - \Theta^{1/2}} \end{aligned} \tag{6.132}$$

The difference terms in the last line of Eq. (6.132) are proportional to the exponential representation of the sine function and may thus be written as

$$G(\theta) = \frac{\sin\left(\pi(N+1)\left(\frac{d}{\lambda}\right)\sin\theta\right)}{\sin\left(\pi\left(\frac{d}{\lambda}\right)\sin\theta\right)} \tag{6.133}$$

Figure 6.30 plots the gain $G(\theta)$ as a function of θ, assuming that $N = 6$ and the element spacing $d = \lambda/4$. The figure shows that the antenna has a strong gain in the direction $\theta = 0^{\circ}$ and falls off quickly in other directions. ■

If the signal is arriving from a direction θ_0, then it is desirable to have the maximum signal strength in that direction. That is, we wish to maximize $\mathbf{w}^{\dagger}\mathbf{a}(\theta_0)$. Accordingly, we may use the Cauchy–Schwarz inequality of Eq. (6.15) to show that

$$\mathbf{w} = c\mathbf{a}(\theta_0) \tag{6.134}$$

is the optimum weight vector, which is aligned with $\mathbf{a}(\theta_0)$; the scalar c is not to be confused with the speed of light. Without loss of generality, we choose the scalar $c = 1$. In Fig. 6.31, we plot the optimum gain,

$$\begin{aligned} G(\theta) &= \mathbf{w}^{\dagger}\mathbf{a}(\theta) \\ &= \mathbf{w}(\theta_0)^{\dagger}\mathbf{a}(\theta) \end{aligned} \tag{6.135}$$

for $\theta_0 = 45^{\circ}$. As desired, we have produced an antenna that points in the direction of θ_0.

Figure 6.32 illustrates how we may produce an antenna that points in any direction simply by adjusting the weights. For a linear array, the quality of the pointing does depend somewhat on the angle θ_0. In the *end-fire* directions (i.e., large θ), however, the quality of pointing is not as good as that of other directions for small θ.

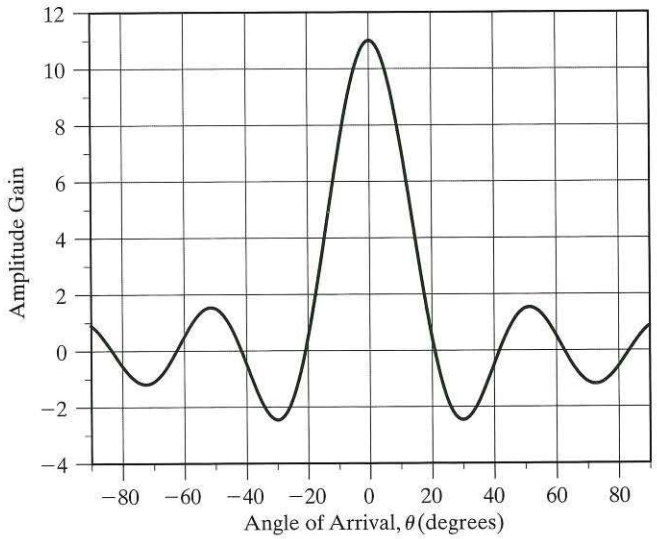


FIGURE 6.30 Antenna gain for constant weighting with $N = 6$ and $d = \lambda/4$.

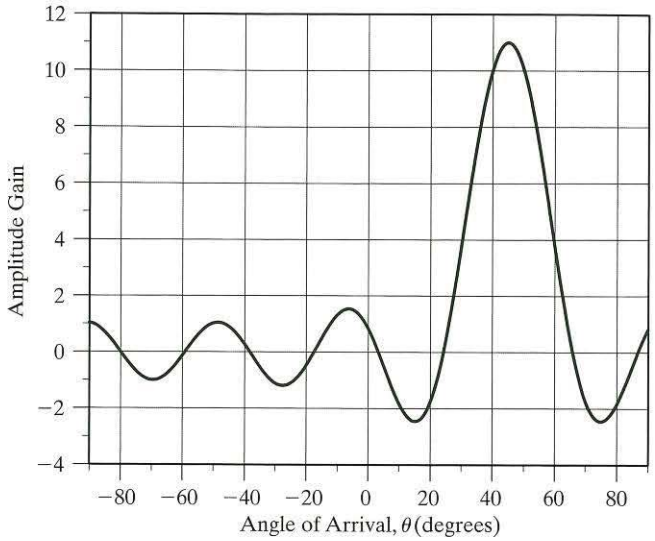


FIGURE 6.31 Gain for maximum signal strength with $\theta_0 = 45°$.

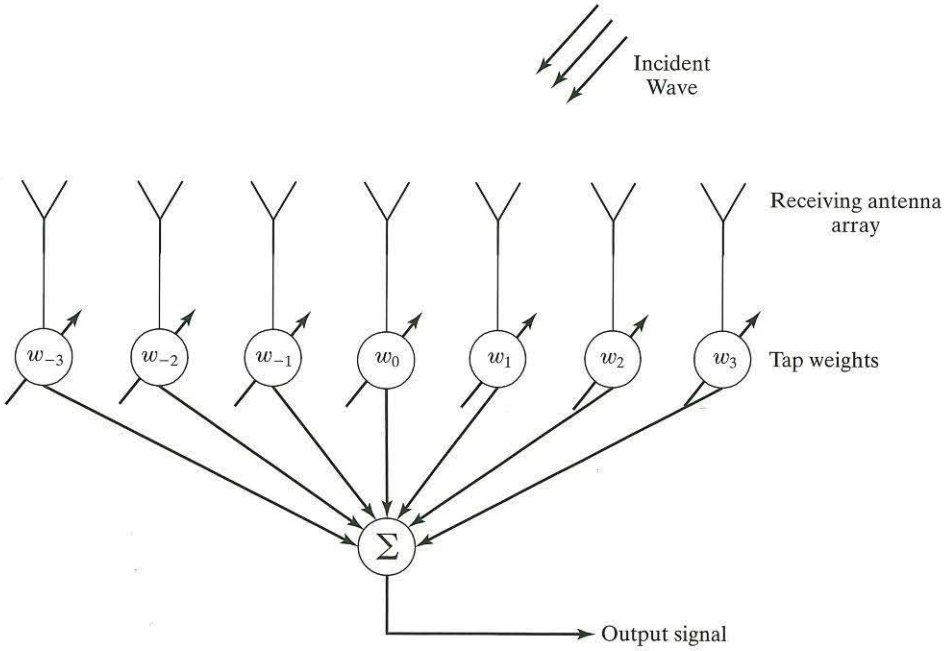


FIGURE 6.32 Antenna with adjustable elemental weights.

It is also clear, conceptually, how we could use the same array to generate multiple antennas; for example, we may use the same antenna elements with two different weighting vectors $\mathbf{w}_1$ and $\mathbf{w}_2$.

In addition, we may *adapt* the antenna to track a mobile terminal by adjusting the weights $\mathbf{w}(n)$ in discrete time n as a function of the user position. Such adaptive algorithms are discussed in Appendix I. The linear array is just one possible arrangement of antenna elements; more sophisticated arrangements (e.g., planar arrays) can provide higher gain and performance that is more uniform in all directions.

Problem 6.18 Suppose the received signal is known to consist of two strong multipath rays arriving from different directions. Describe how to design a phased array to capture the energy in both rays. ■

Problem 6.19 Derive an expression for the antenna gain $G(\theta)$ under the conditions of Example 6.18, assuming that the number of antenna elements, N, is even.

$$Ans.\ G(\theta) = \frac{\sin\left(\pi N\left(\frac{d}{\lambda}\right)\sin\theta\right)}{\sin\left(\pi\left(\frac{d}{\lambda}\right)\sin\theta\right)}$$ ■

6.8.2 Multipath with Directional Antennas

For large cells, referred to as *macrocells*, the base-station antenna is usually mounted on a tall mast and is free of most significant local multipath. In such an environment, almost all multipath is generated in the vicinity of the mobile terminal for both the uplink and downlink paths. Several models have been described in the literature; one of the original ones, *Lee's model*,[18] is illustrated in Fig. 6.33. With this model, we have a number of effective scatterers surrounding the mobile terminal in a circular pattern. Reflections from these scatterers cause the multipath seen at the base station. Measurements indicate that the radius of the scatterers is typically 100 to 200 wavelengths (λ), a value consistent with other measurements which indicate that the typical angle spreads for macrocells with transmitter–receiver separations of 1 km are from 2^o to 6^o when measured at 800 MHz. From this observation, two rules of thumb for spacing antennas to achieve good diversity have been developed: at the mobile, it is recommended that the antenna spacing be at least 0.2λ; at the base station, an antenna separation of approximately 40λ is recommended.

With *microcells*, the separation between the base station and the user terminal is much smaller. Consequently, if we were to apply the Lee model, the circle of effective scatterers could come close to, or even encompass, the base station. In addition, microcell base station antennas are often not mounted high, so there will be scattering effects around the base station as well as the mobile. These observations motivate a geometrical model, referred to as the *single-bounce elliptical model*,[19] for multipath propagation in microcells. With this model, we begin by considering multipath that consists of a single reflection having a fixed excess delay (relative to the direct path between the base station and user terminal). From geometrical considerations, this

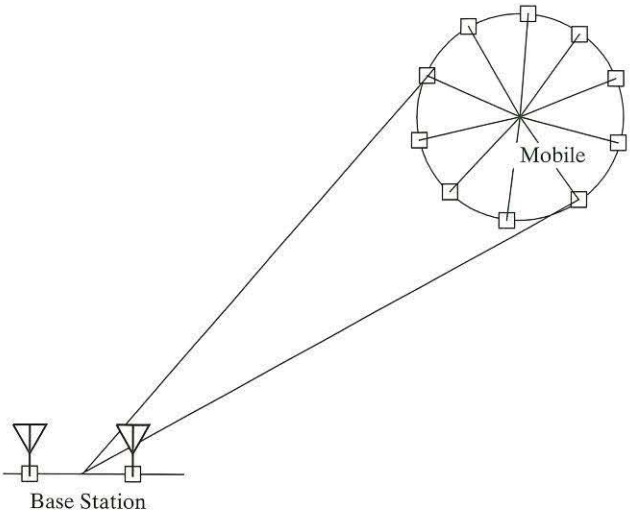


FIGURE 6.33 Lee's model for multipath propagation in macrocells.

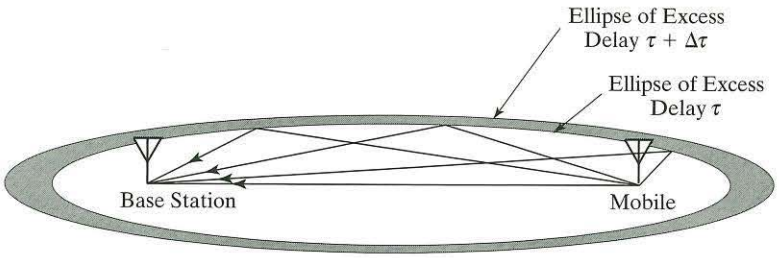


FIGURE 6.34 Single-bounce elliptical model for multipath propagation in microcells.

assumption defines an ellipse with the base station and mobile terminal situated at its foci and the reflectors located on its perimeter. We may then parameterize a series of ellipses with the excess delay parameter τ. All the sources of multipath with a delay between τ and $\tau + \Delta\tau$ are assumed to be due to reflectors in the area between the two corresponding ellipses, as shown in Fig. 6.34. By assuming the reflectors to be uniformly distributed, we can determine the distribution of the direction of arrival of multipath as a function of the time of arrival. This single-bounce elliptical model leads to some interesting observations:

- For a small number of reflectors, most multipath has a small excess delay and a small angular distribution; that is, paths with low excess delays and higher powers are clustered around the main path.

- As the number of reflectors increases, both the excess delay and the angular distribution increase.

The overall conclusion is that, with a small number of reflectors, a directional antenna may not reduce multipath as significantly as might be expected.

The previous results show that both smaller cells and directional antennas may reduce the number of multipath components that are received. A reduced number of multipath components will affect the signal's statistics. In Section 2.6.1, we considered the amplitude distribution of the received signal due to local reflections. The assumption behind that analysis was the existence of a large number of multipath rays, and the results indicated that the fading amplitude would have a Rayleigh distribution.

In this section, we will reevaluate the amplitude distribution for a small number of multipath components. Recall that the complex envelope of the received signal was modeled as (see Eq.(2.48))

$$\tilde{x}(t) = \sum_{k=1}^{N} \alpha_k e^{j\theta_k} \tilde{s}(t) \tag{6.136}$$

where α_k is the gain or attenuation of the kth path and θ_k is the relative phase of that path. The fading amplitude is given by

$$R = \left| \sum_{k=1}^{N} \alpha_k e^{j\theta_k} \right| \tag{6.137}$$

In Section 2.6.1, we argued that, for large N, the term on the right-hand side of Eq.(6.137) (before taking the magnitude) approaches a complex Gaussian random variable, in which case R approaches a Rayleigh-distributed random variable. With a directional antenna, N will be much smaller and a different result may follow. In Fig. 6.35, we plot the simulated distribution of R for different values of N; the Rayleigh distribution is also included in the figure for comparison. From the figure, we see that, for very small N, the probability of a deep fade is significantly larger than for large N. For $N = 6$ and larger, the distribution of R closely approximates the Rayleigh distribution. We conclude that it takes only a small number of paths to approximate a Rayleigh amplitude distribution, but for a *very* small number, the characteristics can be significantly worse.

The characteristics of the received multipath signal also depend on the bandwidth of the transmitted signal. If the transmitted signal is wideband, such as with direct-sequence spread spectrum, the delay spread is liable to be greater than a chip period. Consequently, the received signal may consist of two or more approximately independently faded signals. This is analogous to the diversity situation, and as a consequence, the variation in the received power (or envelope) is much less than is observed in the flat-fading case. The phenomenon affects directional antennas through the algorithms used for adaptation.

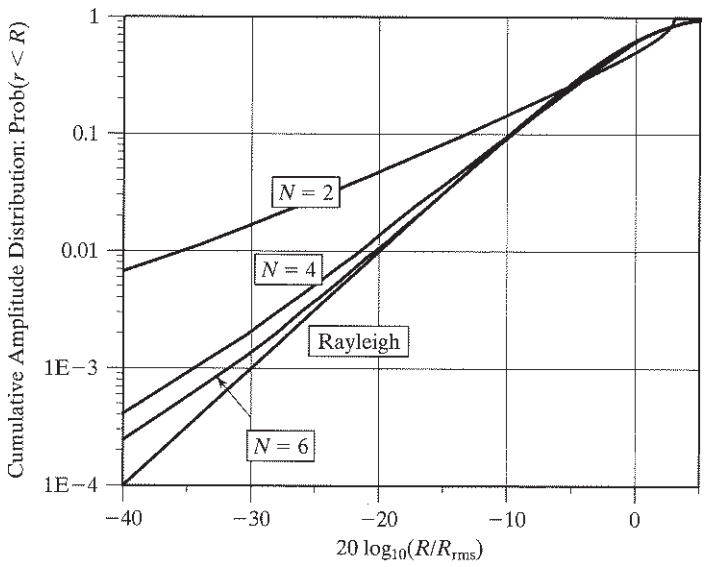


FIGURE 6.35 Amplitude distribution of multipath as a function of number of paths N.

6.9 THEME EXAMPLE 1: BLAST ARCHITECTURES

For the first theme example of the chapter, we have chosen a family of MIMO wireless communication systems popularized as *BLAST architectures*; the acronym "BLAST" stands for "Bell Laboratories Layered Space–Time." BLAST architectures use standard one-dimensional forward error-correcting codes and low-complexity interference-cancellation schemes to construct and decode powerful two-dimensional space–time codes. These MIMO systems offer spectacular increases in spectral efficiency, provided that three conditions are met:

1. The system operates in a rich Rayleigh scattering environment.
2. Appropriate coding structures are used.
3. Error-free decisions are available in the interference-cancellation schemes. This condition assumes the combined use of arbitrarily long (and therefore powerful) FEC codes and perfect decoding.

The material presented herein focuses on three specific implementations of BLAST, depending on the type of coding employed:

1. Diagonal-BLAST (D-BLAST)
2. Vertical-BLAST (V-BLAST)
3. Turbo-BLAST

6.9.1 Diagonal-BLAST Architecture[20]

In 1996, Foschini pioneered a *diagonally layered space–time architecture* known as *diagonal-BLAST,* or simply *D-BLAST.* This architecture now provides the benchmark for MIMO wireless communications. The distinctive feature of D-BLAST is its use of a diagonally layered coding structure in which the code blocks are arranged across *diagonals* in a space–time fashion, as illustrated in Fig. 6.36 for the example of $N_t = 4$ transmit antennas. The figure depicts the incoming binary data stream demultiplexed and then processed for transmission over the wireless link as four independent data substreams with equal bit rates. A distinctive feature of D-BLAST is that each encoded–modulated substream is *cycled over time.*

To illustrate the diagonal nature of the layered space-time code generated by D-BLAST, consider the example of four transmit antennas. The transmission matrix produced by this structure has the form

$$\mathbf{S} = \begin{bmatrix} \tilde{s}_{1,1} & \tilde{s}_{1,2} & \tilde{s}_{1,3} & \tilde{s}_{1,4} & \tilde{s}_{1,5} & \cdots & \tilde{s}_{1,K-1} & \tilde{s}_{1,K} & 0 & 0 & 0 & 0 \\ 0 & \tilde{s}_{2,1} & \tilde{s}_{2,2} & \tilde{s}_{2,3} & \tilde{s}_{2,4} & \cdots & \tilde{s}_{2,K-2} & \tilde{s}_{2,K-1} & \tilde{s}_{2,K} & 0 & 0 & 0 \\ 0 & 0 & \tilde{s}_{3,1} & \tilde{s}_{3,2} & \tilde{s}_{3,3} & \cdots & \tilde{s}_{3,K-3} & \tilde{s}_{3,K-2} & \tilde{s}_{3,K-1} & \tilde{s}_{3,K} & 0 & 0 \\ 0 & 0 & 0 & \tilde{s}_{4,1} & \tilde{s}_{4,2} & \cdots & \tilde{s}_{4,K-4} & \tilde{s}_{4,K-3} & \tilde{s}_{4,K-2} & \tilde{s}_{4,K-1} & \tilde{s}_{4,K} & 0 \end{bmatrix}$$

where the encoded entries in the kth row are delayed by k-1 time units with respect to those in the first row, as the entries below the diagonal are padded with zeros. The entries on the first diagonal of the transmission matrix **S** are sent to the receiver by antenna 1, the entries on the second diagonal of **S** are sent by antenna 2, and so on, with the result that the encoded symbols are transmitted by different antennas.

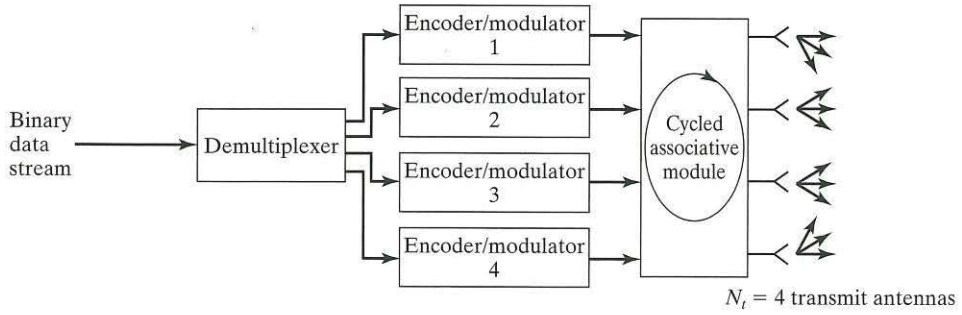


FIGURE 6.36 High-level diagram of D-BLAST transmitter for four transmit antennas.

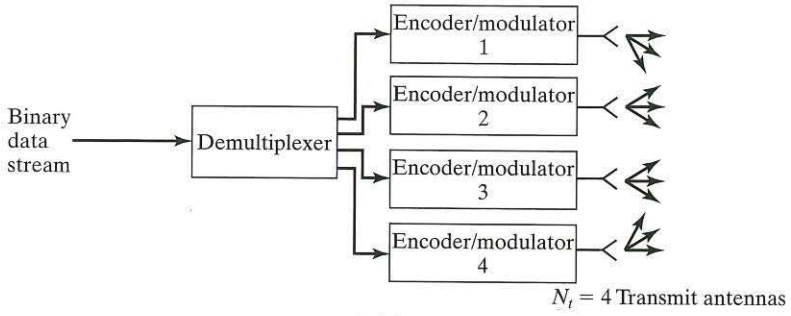
The diagonal coding structure of D-BLAST is not only highly elegant, but also capable of achieving the channel capacity based on Eq. (6.60) for an (N_t, N_r) antenna configuration with $N_t \leq N_r$, assuming that the channel is impaired by additive white Gaussian channel noise and that it operates in a quasi-static, flat-fading environment. However, a major drawback of D-BLAST lies in the need for independent coding at each diagonal layer, which implies the required use of short diagonal-layer coding schemes so as to reduce boundary space–time edge wastage; this form of wastage is illustrated in the structure of the transmission matrix $\mathbf{S}$. Unfortunately, the use of *short* diagonal codes suffers from the deficiency of not allowing adequate time to support powerful bandwidth-efficient codes with forward error-correcting capabilities.

6.9.2 Vertical-BLAST Architecture[21]

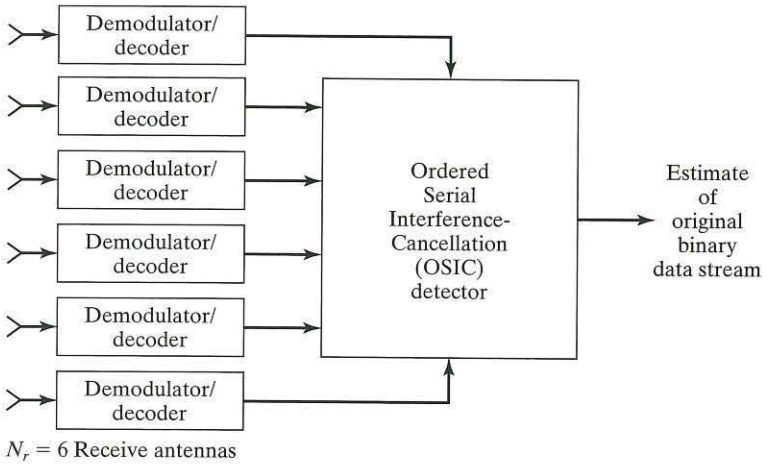
To mitigate the computational difficulty of D-BLAST, Wolniansky et al. (1998) proposed a simplified version of BLAST known as *vertical-BLAST*, or *V-BLAST*, the first practical implementation of MIMO wireless communications to demonstrate a spectral efficiency as high as 40 bits/s/Hz in real time. In V-BLAST, the incoming binary data stream is first demultiplexed into N_t substreams, each encoded independently and mapped onto an antenna of its own by a modulator for transmission over the channel, as depicted in Fig. 6.37(a). Insofar as the transmitter is concerned, the net result is the conversion of the incoming binary data stream into a vertical vector of encoded modulated substreams—hence the name "vertical-BLAST," or "V-BLAST." Comparing this figure with Fig. 6.36 for D-BLAST, we see that in V-BLAST there is no cycling over time—hence the significant reduction in system complexity. Moreover, in the V-BLAST transmitter, every antenna transmits its own independently coded substream of data. In so doing, V-BLAST eliminates the space–time edge wastage plaguing D-BLAST, but the outage capacity achieved by V-BLAST is substantially lower than that of D-BLAST.

Turning next to the receiver depicted in Fig. 6.37(b), the signals impinging on the receive antennas are individually demodulated and then channel decoded in accordance with the corresponding operations performed in the transmitter. The detection process, leading to an estimate of the original binary data stream, is performed by the functional block labeled *ordered serial interference-cancellation (OSIC) detector*, which exploits the timing synchronism between the V-BLAST receiver and transmitter. Specifically, the detection process involves the following sequence of operations:

1. *Order determination*, in which the N_r received substreams are to be detected, in accordance with the postdetection signal-to-noise ratios of the individual substreams.
2. *Detection* of the substream, starting with the *largest signal-to-noise ratio.*
3. *Signal cancellation*, wherein the effect of the detected substream is removed from subsequent substreams.
4. *Repetition* of steps 1 through 3 until all the N_r received substreams have been individually detected.



(a)



(b)

FIGURE 6.37 High-level diagram of V-BLAST.
(a) Block diagram of the transmitter four transmit antennas.
(b) Block diagram of the receiver for six receive antennas.
Note: The OSIC includes the multiplexing operation needed to restore the detected binary data stream to its original serial form.

The procedure is nonlinear, due to two factors: First, the estimated-signal-cancellation process is itself nonlinear; second, the detection step involves a quantization (slicing) operation that is appropriate to the signal constellation used in the transmitter.

Another noteworthy point is that the nulling-and-cancellation step is performed by exploiting the channel matrix **H**, which is estimated at the receiver through supervised training aided by sending a training sequence from each transmit antenna. The nulling process suppresses the interference and is followed by cancellation of the estimated signal in question.

It is important to note that for the OSIC detection process to work properly, the number of receive antennas N_r must be at least as large as the number of transmit antennas N_t, for reasons to be explained in Section 6.10.

One last comment is in order. In V-BLAST, each transmitted data stream is channel-encoded in its own way. Typically, the channel encoders in the N_t layers of the space-time system are identical. The structure of V-BLAST may be simplified by employing a single channel encoder before the demultiplexing operation in the transmitter; correspondingly, a single channel decoder is used after the multiplexing operation in the receiver.

Problem 6.20 A prototype V-BLAST system uses $N_t = 8$ transmit antennas and $N_r = 12$ receive antennas in an indoor environment with negligible delay spread. The system utilizes uncoded 16-QAM, where the term "uncoded" refers to the absence of channel coding. The system has a transmission rate of 24.3×10^3 symbols/s and a channel bandwidth of 30 kHz.

(a) Calculate the raw spectral efficiency of the system.

(b) Assuming that 80% of each burst is devoted to data transmission, calculate the payload efficiency of the system.

Ans. (a) 25.9 bits/s/Hz
(b) 20.7 bits/s/Hz ■

6.9.3 Turbo-BLAST Architecture[22]

The third BLAST architecture for high-throughput wireless communications exploits three basic ideas:

1. BLAST architecture, which uses multiple antennas on both transmit and receive.
2. *A random layered space–time (RLST) coding scheme*, which is based on the use of independent block coding and space–time interleaving.
3. *A Turbolike receiver*, also known as an *iterative detection and decoding (IDD) receiver*, which performs decoding of the RLST code and iterative estimation of the channel matrix.

In light of points 1 and 3, the new architecture is referred to as *Turbo-BLAST*. This architecture is also referred to as *Turbo-MIMO*. (The turbo coding principle was discussed in Chapter 4.)

Figure 6.38(a) gives a high-level picture of the Turbo-BLAST transmitter with $N_t = 4$ transmit antennas. The encoding process in the transmitter, responsible for generating a *serially concatenated RLST code*, involves the following steps:

- *Demultiplexing* the incoming bit stream into N_t substreams of equal bit rate
- *Independent block-encoding* of each data substream, using the same predetermined linear block FEC codes
- *Interleaving* the encoded substreams by means of a space–time permuter, which is independent of the incoming bit streams
- *Full-rate space–time encoding*, facilitated by the channel matrix; the term "full-rate" means that the symbol-constellation in the transmitter's mapper is designed to have a size equal to the spatial degree of freedom $N = \min(N_t, N_r)$.

The structure of the RLST encoder, designed on the basis of independent block coding of each transmit data substream and periodically cyclic space–time interleaving, is shown in Fig. 6.38(b). The transmitter pictured here is the multiple-antenna generalization of the serial turbo encoder discussed in Section 4.18 on joint equalization and decoding. For optimal performance of the RLST code, the receiver should use the *maximum a posteriori probability (MAP) decoding algorithm*, which is described in Appendix F. However, the computational complexity of this algorithm for the RLST code becomes increasingly unmanageable as the number of transmit or receive antennas is increased. Specifically, with K denoting the length of each layer in the RLST code, the MAP decoding algorithm needs to select one of $2^{N_t K}$ sequences, all of which increase exponentially with increasing N_t. To mitigate the computational

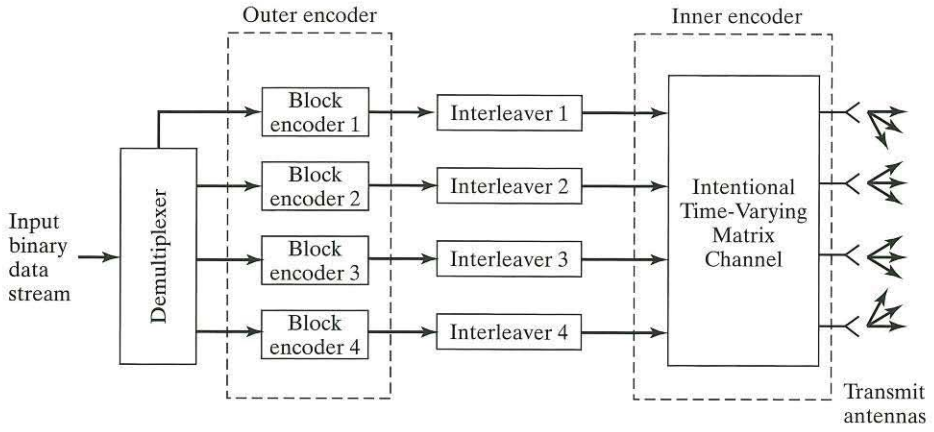


FIGURE 6.38 (a) High-level picture of the Turbo-BLAST transmitter with four transmit antennas. (b) Illustrating the structure of the random layered space-time (RLST) code generated in the transmitter.

complexity problem, we may use the near-optimal turbolike receiver depicted in Fig. 6.39. The interleavers used to design the RLST codes provide the basis for a near-optimal *iterative detection and decoding* (IDD) process with feasible computational complexity. The receiver has two stages, in accordance with the receive part of the turbo coding principle:

1. *The inner decoder*, consisting of a soft-input, soft-output (SISO) detector, which is designed to counteract the intersymbol interference (ISI) problem due to the multipath fading channel.
2. *The outer decoder*, consisting of N_t parallel SISO channel decoders, which is designed to correct symbol errors incurred during the course of transmission over the channel.

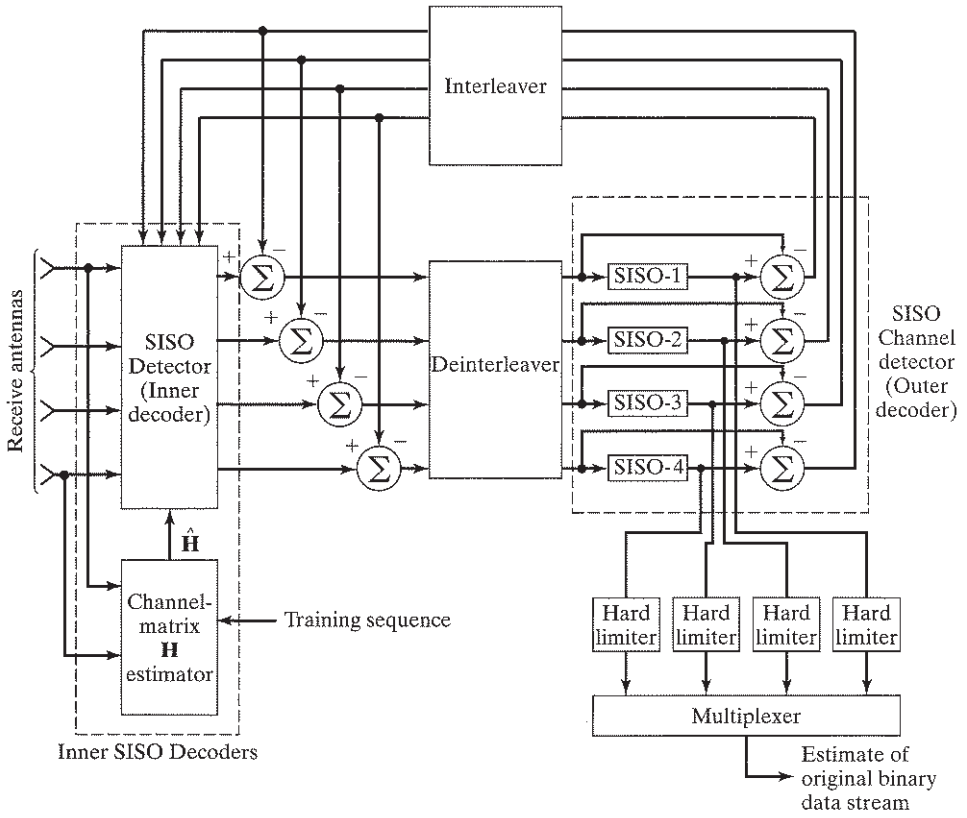


FIGURE 6.39 High-level block diagram of iterative decoder for Turbo-BLAST for four receive antennas. (Note: The abbreviation SISO stands for soft-input, soft output; it should not be confused with single-input, single-output antenna configuration.)

The detector and decoder stages of the receiver are separated by *space–time de-interleavers and interleavers*, which may be viewed as the spatial generalizations of their counterparts in the turbo decoder. The space–time deinterleavers and interleavers are used to compensate for the interleaving operations used in the transmitter and to de-correlate outputs before feeding them to the next decoding stage. The two sets of summers, where differences are computed, provide *extrinsic information* from the inner decoder to the outer decoder via the deinterleaver and from the outer decoder to the inner decoder via the interleaver, thereby forming a closed feedback loop around the two decoding stages in accordance with the turbo-coding principle. The iterative receiver produces new and better estimates at each iteration of the receiver and repeats the information-exchange process a number of times, to improve the channel decisions and channel estimates. The design of intersubstream coding at the transmitter, based on independent coding of each substream, leads to a simplification of the receiver, which now needs to select only one of 2^K sequences (where K is the length of each transmitted burst), separately for each transmitted sequence.

To reconstruct the original binary data stream, two things are done, as illustrated in Fig. 6.39:

1. When the IDD process for an information bit is terminated, the output of the SISO channel decoders are hard-limited.
2. The resulting hard-limiter outputs are multiplexed into a serial form, consistent with that of the original binary data stream.

During the *first iteration* of the receiver, a short training sequence is used to produce a preliminary estimate of the channel matrix **H**. Unfortunately, with a short training sequence, it may be difficult to achieve a good estimate of the time-varying MIMO channel, particularly in an outdoor environment. To mitigate this problem, the channel matrix is reestimated with the use of newly derived estimates of symbols at each subsequent iteration of the receiver. The *bootstrapping* technique described herein tends to extract maximal information out of each burst of data received.

6.9.4 Experimental Performance Evaluation of Turbo-BLAST versus V-BLAST[23]

We conclude Theme Example 1 by comparing the performance of QPSK-modulated Turbo-BLAST with that of a correspondingly horizontal-coded V-BLAST, using real-life indoor channel measurements on various MIMO configurations. The channel measurements were acquired by utilizing a narrowband testbed in an indoor environment with negligible delay spread. At the transmit end, each substream of 100 information bits was independently encoded, using a rate-1/2 convolutional code generator (7,5), and then interleaved by means of space–time interleavers. The space interleavers were designed with diagonal layering interleavers (Fig. 6.38); the time interleavers were chosen randomly, and no attempt was made to optimize their design. With regard to horizontal-coded V-BLAST, each of the substreams was first, independently coded via rate-1/2 convolutional code with generator (7,5) and, second, QPSK modulated. Using

the measured channel characteristics, and then, using various BLAST configurations, we evaluated the performance of Turbo-BLAST over a wide range of signal-to-noise ratios. For the first two experiments, we evaluated the Turbo-BLAST system with the exact channel matrix. In the third experiment, we evaluated the performance with channel estimation, using a short training sequence and an iteratively estimated channel.

Experiment 1. Turbo-BLAST versus V-BLAST for $N_t = 5, 6, 7, 8$ and $N_r = 8$.

We first consider BLAST configurations with fewer transmit antennas than receive antennas. Figure 6.40 compares the bit-error-rate performance of Turbo-BLAST (solid traces) and coded V-BLAST (broken traces) for antenna configurations of 8 receive and 5 to 8 transmit antennas. The Turbo-BLAST system produces the best receiver performance within the first 10 iterations of its operation. The bit-error-rate performance of both V-BLAST and Turbo-BLAST improves with a decreasing number of transmitters, with Turbo-BLAST outperforming V-BLAST in all four cases. In

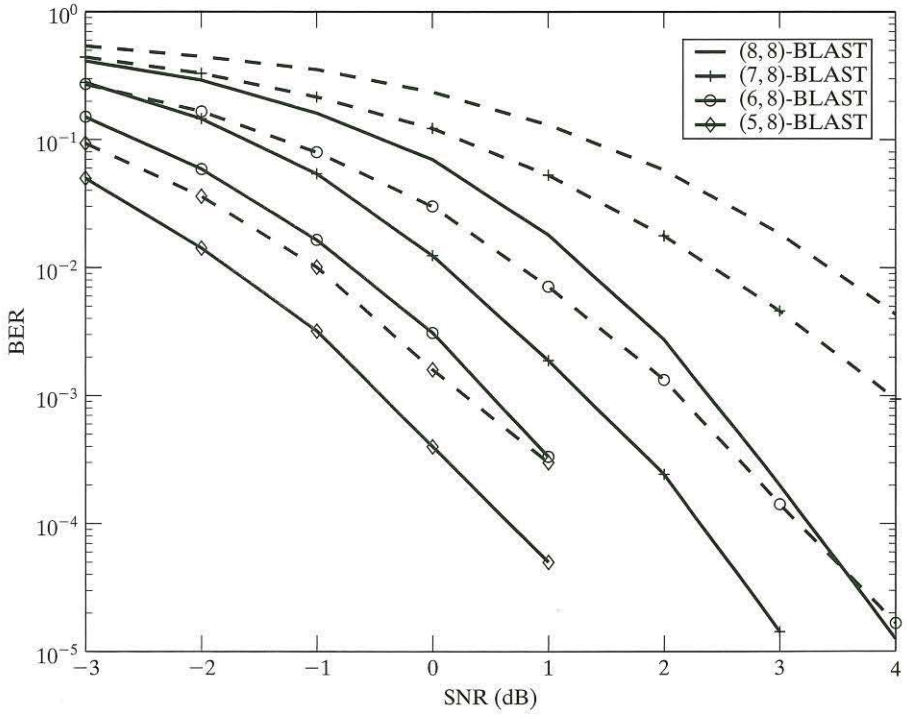


FIGURE 6.40 Bit-error-rate (BER) performance for $N_t = 5, 6, 7$, and 8 and $N_r = 8$, using convolutional code with rate $1/2$ and constraint length 3 and using QPSK modulation.
Solid traces: Turbo-BLAST.
Broken traces: V-BLAST.

terms of V-BLAST performance, a substantial gain in bit-error-rate performance is realized with fewer transmit antennas. For example, at a BER of 10^{-3}, Turbo-BLAST achieves 2–3 dB of gain over V-BLAST for $N_t = 7$ and $N_r = 8$, whereas only 0.5 dB of gain is attained when $N_t = 5$ and $N_r = 8$.

Experiment 2. Turbo-BLAST versus V-BLAST for $N_t = 8$ and $N_r = 5,6,7,8$.

We next consider BLAST configurations with fewer receive antennas than transmit antennas. Figure 6.41 compares the bit-error-rate performance of Turbo-BLAST (solid traces) with that of horizontal-coded V-BLAST (broken traces). With antenna configurations of eight transmit and five to eight receive antennas, here again we find that Turbo-BLAST gives the best overall performance within the first 10 iterations. The figure reveals a major limitation of the V-BLAST system, namely, the inability to work efficiently with fewer receive antennas than transmit antennas. In the context of Turbo-BLAST, we can make two observations from Fig. 6.41: First, the bit-error-rate performance of Turbo-BLAST improves with an increasing number of receivers, with Turbo-BLAST outperforming V-BLAST in all four cases; second, increasing the number N_r of receive antennas from 7 to 8 offers little benefit.

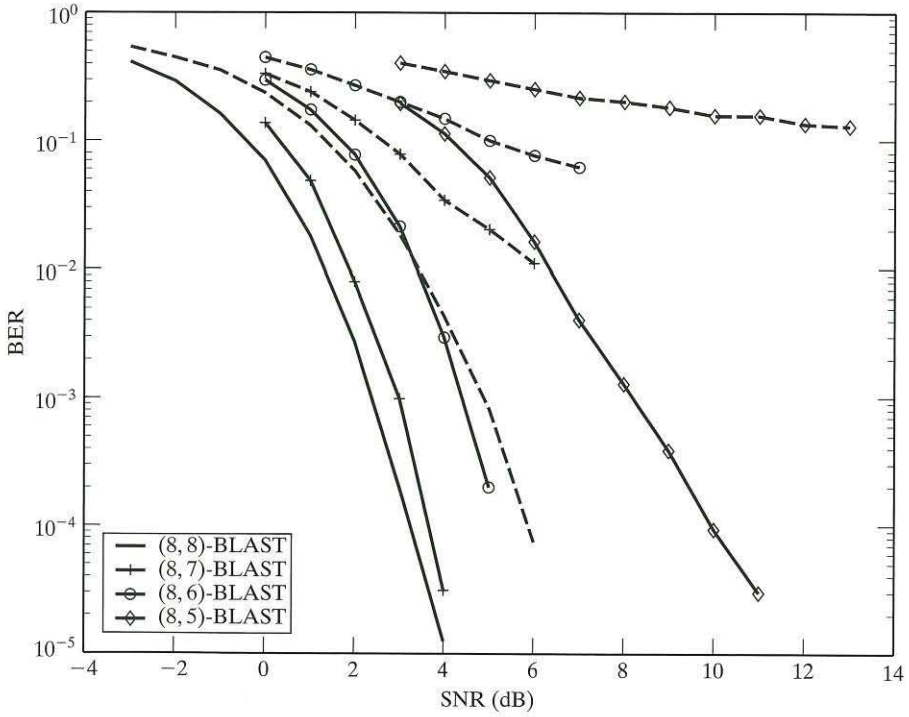


FIGURE 6.41 Bit-error-rate (BER) performance for $N_t = 8$ and $N_r = 5, 6, 7$, and 8, using convolutional code with rate $1/2$ and constraint length 3, and using QPSK modulation.
Solid traces: Turbo-BLAST.
Broken traces: V-BLAST.

Experiment 3. Turbo-BLAST for $N_t = N_r = 8$ and iterative channel estimation.

In Fig. 6.42, we compare the performances of two decoding schemes: (1) an iterative decoder with initial channel estimation using 16 training symbols only, and (2) an iterative decoder with initial channel estimation and iterative refined channel estimation. The bit-error-rate performance results are compared for Turbo-BLAST architectures with perfect channel knowledge and with perfect channel and interference knowledge. The figure shows the convergence behavior of the IDD receiver of Fig. 6.39 at a signal-to-noise ratio of 3 dB. Although the bit-error-rate performance of the IDD receiver with iterative channel estimation is initially (i.e., at the first iteration) worse than the decoder with perfect channel knowledge, at the fifth iteration of the decoding process it comes very close in performance to the decoder with knowledge of channel state information. Moreover, both decoders converge very close to the decoder that has knowledge of both the channel and the interference (dashed trace). Because of channel estimation errors, the bit-error-rate performance of the IDD receiver with initial channel estimates is about 2–4 dB worse than the schemes with channel knowledge only and the ideal case with knowledge of the channel and interference.

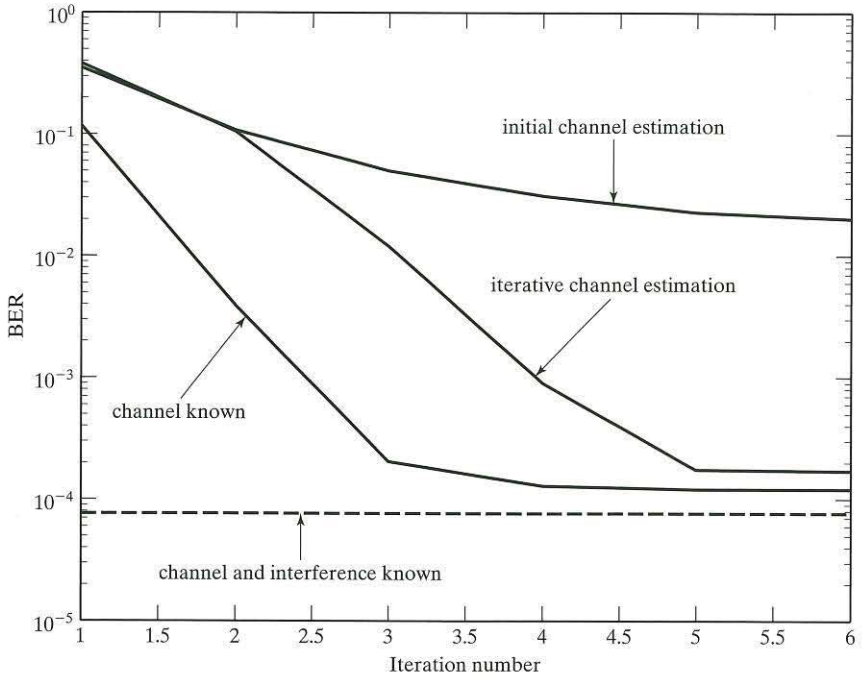


FIGURE 6.42 Convergence behaviors of the IDD receiver of Fig.6.39 under various conditions: bit-error-rate (BER) performance (at SNR = 3 dB) versus the number of iterations for $N_t = 8$ and $N_r = 8$, using convolutional code with rate 1/2, constraint length 3, and QPSK modulation.
Solid traces: Turbo-BLAST for varying channel conditions.
Broken traces: Ideal performance.

6.10 THEME EXAMPLE 2: DIVERSITY, SPACE–TIME BLOCK CODES, AND V-BLAST

Theme example 1 was devoted to three different BLAST architectures, which constitute an important class of multiple-input, multiple-output (MIMO) wireless communications. But then, MIMO wireless systems also include antenna diversity and space–time block codes as special cases. It is therefore in order that we discuss two pertinent issues:

1. Evaluation of the practical merits of a diversity-on-receive antenna system versus its diversity-on-transmit counterpart.
2. Comparison of the receiver performance of a space–time block code with that of a BLAST architecture, both involving diversity at the transmit and receive ends of the wireless link.

The purpose of this second theme example is to address these two issues, using qualitative arguments for dealing with issue 1 and computer experiments for illustrating issue 2.

6.10.1 Diversity-on-Receive versus Diversity-on-Transmit

The diversity-on-receive technique is of long standing, with its roots dating back to the 1920s.[24] In contrast, the diversity-on-transmit technique is of recent origin, with its discovery in 1998 credited to Alamouti.[13]

These two antenna diversity techniques rely on the use of multiple antennas, on the receive side of a wireless communication system in one case and on the transmit side of the system in the other case. Therein lie the realities that shape the use of the two techniques in practice.

Consider first the case of diversity-on-receive. The need for multiple antennas operating at radio frequencies and for the ancillary instruments to select the strongest diversity branch in the case of selection combining, for example, makes the use of diversity-on-receive at the terminal (mobile) units in a wireless communication system rather awkward (requiring the use of dual antennas) and expensive to implement. The picture is, however, quite different at the base station, which is equipped, in terms of both transmit power and real estate, to serve thousands of terminal units. Indeed, it is for this reason that the diversity-on-receive technique is routinely included in the construction of base stations.

In a loose sense, diversity-on-transmit may be viewed as playing a *dual* role to that of diversity-on-receive in counteracting the multipath-channel-fading problem. Accordingly, the two techniques can serve complementary purposes in their use, side by side at the base station:

- Diversity-on-receive looks after the channel fading that affects incoming signals from terminal units operating inside the base station's coverage on the downlink.
- Diversity-on-transmit looks after the channel fading that affects outgoing signals transmitted by the base station to the terminal units in its coverage, *without any feedback from the mobile station to the base station* on the uplink.

To implement the diversity-on-transmit, for example, exemplified by the Alamouti code requires simply (i) the use of two transmit antennas at the base station and (ii) a linear decoding receiver with a single receive antenna at the terminal unit, both of which make sense in practical terms. Indeed, it is for this reason that the Alamouti code has been adopted in third-generation (3G) wireless systems.

In short, through the use of both diversity-on-receive and diversity-on-transmit at a base station, the reception, as well as the transmission, of information-bearing signals at the base station are made *reliable.*

6.10.2 Space–Time Block Codes versus V-BLAST

Consider next the issue of comparing the receiver performance of a space–time block code (STBC) system with that of a BLAST system. For the STBC system, we have picked the expanded version of the Alamouti code; see Problem 6.37. For the BLAST system, we have picked V-BLAST. Both of these multiple-antenna configurations are the simplest schemes in their respective classes of MIMO wireless communications. Moreover, in order to highlight the underlying differences between them, both systems are uncoded (i.e., no channel coding is used in either one of them).

In what follows, the performance of each of these two MIMO systems is evaluated in the content of a two-part experiment:

Experiment 1. Receiver performances

The objective of this experiment is to evaluate the effect of increasing signal-to-noise ratio (SNR) on the symbol error rate (SER) for a prescribed spectral efficiency.

Four specific $\{N_t, N_r\}$ wireless systems are evaluated:

(i) STBC: $\{2, 2\}$
V-BLAST: $\{2, 2\}$
(ii) STBC: $\{2, 4\}$
V-BLAST: $\{2, 4\}$

For all four multiple-antenna systems, the spectral efficiency is maintained at 4 bits/s/Hz. The constancy of spectral efficiency is achieved by using different modulation schemes: 4-PSK (QPSK) for V-BLAST and 16-PSK for STBC. V-BLAST sends independent symbols on its two transmit antennas, and with each transmission using 4-PSK, the spectral efficiency is therefore $2 \times \log_2 4 = 4$ bits/s/Hz. As with the STBC, since the Alamouti code is a rate-1 code, the use of 16-PSK yields the spectral efficiency $\log_2 16 = 4$ bits/s/Hz, as desired.

The results for the experiment are plotted in Fig. 6.43. On the basis of these results, we may make the following observations:

(i) *Two-by-two antenna systems*

- For low signal-to-noise ratios (5 to 17.02 dB), V-BLAST outperforms STBC. This result is in perfect agreement with statements reported in Section 4.11. Specifically, therein we made the point that an uncoded system outperforms a channel encoded system when the signal-to-noise ratio is low. In the context of our present experiments, recall that V-BLAST is uncoded whereas the STBC system

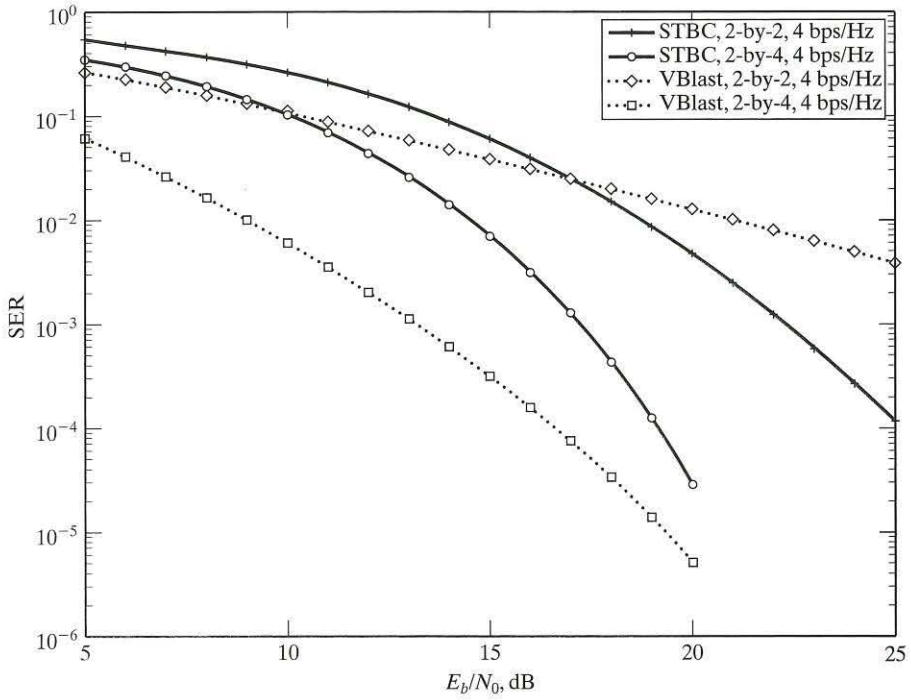


FIGURE 6.43 BER versus Eb/N_0 for four multiple-antenna systems studied under Experiment 1, assuming a fixed spectral efficiency of 4 bits/s/Hz.

uses a space-time encoder. Note, however, the SER (around 0.6 down to 0.023) is not small enough for either system to be of practical value in a wireless communications environment. But then recognize that if forward error-correction (FEC) channel codes are included in either system, then uncoded symbol error rates as high as 0.2 may be acceptable.

- For high signal-to-noise ratios (in excess of 17 dB), STBC begins to outperform V-BLAST. For example, at SNR = 25 dB, the SER produced by the STBC system is slightly larger than 10^{-4}. However, this improvement in receiver performance is attained at the cost of a significant increase in SNR.

(ii) *Two-by-four antenna systems*

- In doubling the number of receive antennas from two to four, the benefit of receive diversity is enhanced, thereby improving the performance of both STBC and V-BLAST. However, the important point to note here is the fact that, for low SNR, the improvement in receiver performance for V-BLAST is significantly better than that for STBC. For example, for SER = 10^{-3}, V-BLAST requires an SNR of 13.15 dBs. For this same SNR, the SER produced by STBC is 2.3×10^{-2}, which is more than an order of magnitude worse than the corresponding result for V-BLAST.

- The crossover SNR, at which STBC begins to outperform V-BLAST, is well in excess of 20 dB.

An important question is, for low SNR, how do we explain the significantly better receiver performance produced by V-BLAST over STBC in going from two to four receive antennas? The answer to this question resides in two factors:

1. The choice of modulation for a fixed spectral efficiency, 16-PSK for the STBC system and 4-PSK for V-BLAST, makes the former system much harder to demodulate than the latter. Asymptotically, there is a constant difference in receiver performance between these two modulation schemes, but, most important, the 16-PSK degrades more quickly than 4-PSK at low signal-to-noise ratios.
2. From Chapter 4, we recall that an uncoded system outperforms its coded counterpart at low signal-to-noise ratios. Thus, although both MIMO systems do not use channel coding, the STBC system does use the full-rate Alamouti code, which may be another possible factor for the observed difference in performance between the STBC and V-BLAST systems at low signal-to-noise ratios.

The influences of these two factors on the behavior of STBC and V-BLAST systems at low signal-to-noise ratios become more pronounced as the number of receive antennas is increased, hence the low-SNR results displayed in Fig. 6.43.

6.10.3 Diversity Order and Multiplexing Gain[25]

At various points in this chapter, we have made reference to diversity order. This subsection presents a brief discussion of the diversity order of MIMO wireless links and its experimental evaluation.

To proceed with the discussion, we first note that ultimately all the results relating to diversity order reflect in the behavior of the *average frame error rate* with respect to the signal-to-noise ratio; a frame or packet is another way of referring to a burst of data transmission across the wireless link. Consider then a space-time scheme whose average frame error rate, expressed as a function of the signal-to-noise ratio ρ, is denoted by $\text{FER}(\rho)$. On this basis, we may make the following statement:

A space-time coding scheme whose asymptotic behavior is described by

$$d_o = -\lim_{\rho \to \infty} \left\{ \frac{\log(\text{FER}(\rho))}{\log \rho} \right\} \tag{6.138}$$

is said to have a diversity order d_o.

In other words, the asymptotic slope of the average frame error rate plotted as a function of the signal-to-noise ratio, on a log-log scale, is equal to the diversity order d_o. As pointed out on page 371, the frame error rate is approximately the same as the outage probability in terms of the exponent signal-to-noise ratio. Hence, we may equally well formulate the definition of the diversity order d_o in terms of the outage probability P_{outage}.

An implication of Eq. (6.137) is that, in a Rayleigh fading environment, a space-time coding scheme with N_t transmit antennas and N_r receive antennas is capable of attaining the *maximal diversity order*

$$d_{o,\max} = N_t \times N_r \tag{6.139}$$

In this context, we may cite two simple examples of interest:

1. For the Alamouti code with $N_t = 2$ and $N_r = 1$, we have $d_o = 2$.
2. For the maximal-ratio combiner with $N_t = 1$ and $N_r = 2$, we again have $d_o = 2$.

Both of these results are intuitively satisfying.

What about BLAST systems that are designed to maximize ergodic capacity? Is there tension between maximizing capacity (i.e. transmission rate) and maximizing diversity order (i.e., reliability of communication)? To answer these two questions, we introduce the definition of multiplexing gain. Specifically, we make the following statement:

> *A space-time coding scheme, whose ergodic capacity expressed as a function of the signal-to-noise ratio ρ, denoted by $C(\rho)$, is said to have a multiplexing gain r if the asymptotic behavior (also see Eq. (6.40))*
>
> $$\lim_{\rho \to \infty} \frac{C(\rho)}{\log \rho} = r \tag{6.140}$$
>
> *holds.*

An implication of the formula of Eq. (6.140) is that in a high signal-to-noise ratio regime, the *maximal multiplexing gain* is defined by

$$r_{\max} = \min\{N_t, N_r\} \tag{6.141}$$

Continuing the discussion of BLAST systems, consider a BLAST system that uses the ordered serial interference cancellation (OSIC) system to deal with the co-antenna interference (CAI) problem. Unlike space-time-block code systems designed to maximize diversity order, the diversity order of a BLAST system varies from one layer (antenna) to the next one by virtue of the ordered way in which the OSIC detector solves the CAI problem. Specifically, to process the first layer output, $N_t - 1$ interferences have to be cancelled. Thus, with $N_t - 1$ of the transmit antennas (i.e., degree of freedom) committed to this cancellation process, the space diversity attainable by the first layer is

$$N_r - (N_t - 1) = N_r - N_t + 1$$

Next, when the OSIC detector processes the second layer output, there will now be $N_t - 2$ interferences to be cancelled, making the space diversity attainable by the second layer assume the value

$$N_r - (N_t - 2) = N_r - N_t + 2$$

and so on for the remaining layers. The overall diversity order of the BLAST system is defined by the *minimum* of the diversity orders for all the N_r layers of the system. On this basis, we may now define the diversity order of a BLAST system using the OSIC detector as

$$d_{o,\,\mathrm{BLAST}} = N_r - N_t + 1 \tag{6.142}$$

which is less than the diversity order attainable by the corresponding STBC system.

From the discussion presented herein, we see that MIMO wireless communication systems provide a *capacity-diversity trade-off*, depending on how the system is actually configured: increased diversity leads to improved reliability of the system, whereas increased capacity leads to improved data transmission across the wireless channel.

Experiment 2. Diversity Orders of STBC and V-BLAST Systems

The objective of this second experiment is to experimentally evaluate the diversity orders of STBC and V-BLAST systems in light of the material presented in Section 6.10.3. The parameters of the two classes of MIMO systems considered in the experiment are summarized in Table 6.4. The modulation strategies used in the experiment are as follows:

STBC: 16-PSK
V-BLAST: 4-PSK

For the experimental evaluations, we plot the symbol error rate (rather than the frame error rate) versus the signal-to-noise ratio. Although indeed the relationship between the symbol error rate and frame error rate is rather complicated, there is a strong correspondence between them. Heuristically, the two of them behave similarly in an asymptotic sense. Following a coarse argument, we may formulate the *probability of correct frame* as

$$\begin{aligned} P_{\mathrm{correct}}(\mathrm{Frame},\rho) &= 1 - \mathrm{FER}(\rho) \\ &= [1 - \mathrm{SER}(\rho)]^K \end{aligned} \tag{6.143}$$

where K is the number of symbols contained in a frame. Asymptotically, in the sense of increasing signal-to-noise ratio, ρ, and therefore decreasingly small symbol error rate, we may approximate Eq. (6.142) as follows:

$$1 - \mathrm{FER}(\rho) \approx 1 - K \times \mathrm{SER}(\rho)$$

or, equivalently,

$$\mathrm{FER}(\rho) \approx K \times \mathrm{SER}(\rho) \tag{6.144}$$

Equation (6.140) suggests that, in the SNR exponent, the frame and symbol error rates (probabilities) behave similarly. In other words, their respective diversity orders (i.e., the asymptotic slopes versus the signal-to-noise ratio on log-log scales) tend to behave similarly, which is confirmed by numerical evaluations. Accordingly, in light of Eqs. (6.137) and

TABLE 6.4 MIMO Configurations for Experiment 2.

MIMO configuration	Number of Antennas		Diversity order	
	N_t	N_r	Theory	Experiment
STBC	2	1	2	1.92
	2	2	4	3.99
V-BLAST	2	2	1	0.97
	2	3	2	1.91

(6.143), we may approximately reformulate the diversity order as

$$d_o \approx -\lim_{\rho \to \infty} \left\{ \frac{\log(\text{SER}(\rho))}{\log \rho} \right\} \tag{6.145}$$

Turning now to the experimental evaluations, we present two sets of results:

1. Figure 6.44(a) plots the symbol error rates versus the signal-to-noise ratio for the two STBC configurations listed in Table 6.4(a). The last two columns of this table present the theoretical values of diversity orders determined from Eq. (6.145) and the corresponding experimental values derived from the plots of Fig. 6.44(a).
2. Figure 6.44(b) plots the symbol error rates versus the signal-to-noise ratio for the two V-BLAST configurations listed in Table 6.4(b). Here again the last two columns of this table present the theoretical values of diversity orders determined from Eq. (6.145) and the corresponding experimental values derived from the plots of Fig. 6.44(b).

On the basis of the results on diversity order presented in the two parts of Table 6.4, we may state that there is good agreement between theory and experiment for both the STBC and BLAST systems considered herein.

6.11 THEME EXAMPLE 3: KEYHOLE CHANNELS [26]

In a MIMO channel, the ability to exploit space-division multiple-access techniques for spectrally efficient wireless communications is determined by the rank of the complex channel matrix **H**. (The rank of a matrix is defined by the number of independent columns in the matrix.) For a given (N_t, N_r) antenna configuration, it is desirable that the rank of **H** equal the minimum one of N_t transmit and N_r receive antennas, for it is only then that we are able to exploit the full potential of the MIMO antenna configuration. Under special conditions, however, the rank of the channel matrix **H** is reduced to unity, in which case the scattering (fading) energy flow across the MIMO link is effectively confined to a very narrow pipe, and with it, the channel capacity is severely degraded. The third theme example of the chapter explains the origin of this physical phenomenon, which is commonly referred to as the *keyhole channel* or *pinhole channel*.

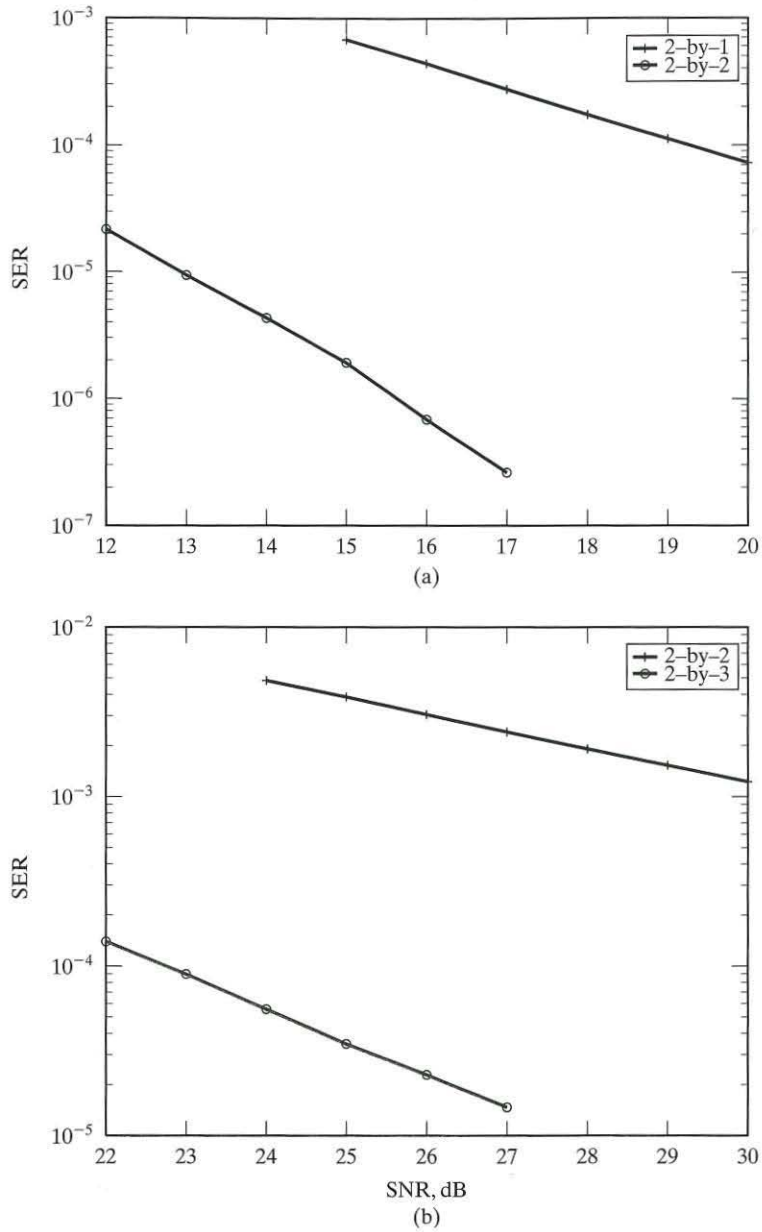


FIGURE 6.44 Experimental evaluation of diverstiy order for MIMO wireless links.
(a) Space-time block codes for two antenna configurations:
(i) $N_t = 2, N_r = 1$. (ii) $N_t = 2, N_r = 2$ Modulation scheme: 16-PSK
(b) V-BLAST architecture for two antenna configurations:
(ii) $N_t = 2, N_r = 1$. (ii) $N_t = 2, N_r = 3$ Modulation scheme: 4-PSK
[To achieve accurute statistics (i.e., asymptotic results), large amount of data were used in the experimental plots of Fig. 6.44.]

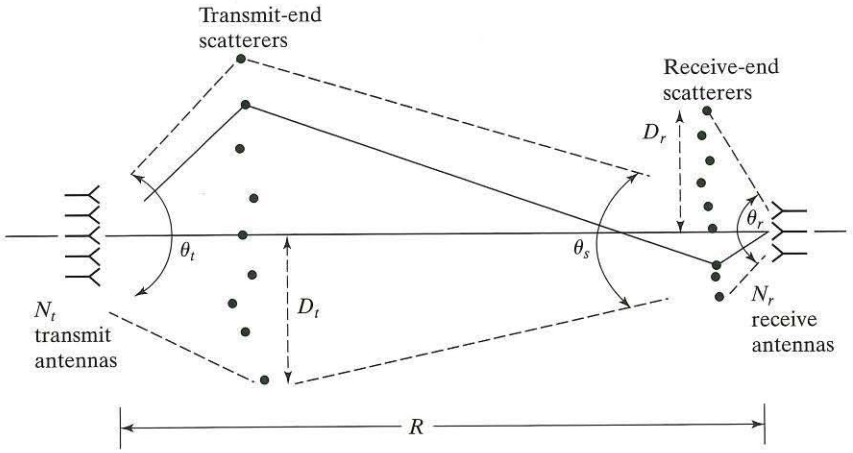


FIGURE 6.45 Propagation layout of MIMO communication link depicting two sets of scatterers, one in proximity to the transmit antennas and the other in proximity to the receive antennas.

Consider the idealized scattering environment pictured in Fig. 6.45. The environment embodies two sets of *significant scatterers*, one in proximity to the N_t transmit antennas and the other in proximity to the N_r receive antennas. Yet, the two sets of scatterers are located far enough away from their respective antenna arrays to justify the assumption of plane-wave propagation across the link. The *radial extents* of the two sets of scatterers are respectively denoted by D_t and D_r, both of which are assumed to be small compared with the range R (i.e., the distance between the transmit and receive antenna arrays). A few other assumptions pertaining to the scattering field are as follows:

- Both sets of scatterers, each numbering S, behave like *ideal reflectors.*
- The numbers of scatterers on both sides of the link are large enough to ensure the occurrence of random fading.
- The scatterers on the transmit side of the link are uniformly spaced, acting like an array of *virtual transmit antennas*, with each one of them receiving the transmitted radio signal and resending it (with no energy loss) toward the receiver.
- The scatterers on the receive side of the link are also uniformly spaced, acting like an array of *virtual receive antennas* with average spacing $2D_r/S$ and angle spread D_t/R.

Note that the scattering environment described in Fig. 6.45 is *spatially fixed and therefore time invariant.*

The *GBGP propagation model*,[27] named in recognition if its originators, Gesbert, Bölcskei, Gore, and Paulraj, is based on the scattering environment pictured

in Fig. 6.45. According to this model, the complex transfer function of the MIMO link is given by

$$\mathbf{H} = \frac{1}{\sqrt{S}}\mathbf{R}_{\theta_r, d_r}^{1/2}\mathbf{G}_r\mathbf{R}_{\theta_S, 2D_r/S}^{1/2}\mathbf{G}_t\mathbf{R}_{\theta_t, d_t}^{1/2} \tag{6.146}$$

where the five matrix terms of the equation (working backwards) are defined as follows:

$\mathbf{R}_{\theta_t, d_t}$ = N_t-by-N_t correlation matrix of the transmitted signal vector for transmit angle spread θ_t and antenna spacing d_t

$\mathbf{G}_t$ = S-by-N_t matrix of i.i.d. Rayleigh-fading coefficients produced by the set of S scatterers at the transmit end of the link

$\mathbf{R}_{\theta_S, 2D_r/S}$ = S-by-S matrix of the scatterer-to-scatterer cross-correlations between the two sets of scatterers with angle spread θ_S and scatterer spacing $2D_r/S$

$\mathbf{G}_r$ = N_r-by-S matrix of i.i.d. Rayleigh-fading coefficients produced by the set of S scatterers at the receive end of the link

$\mathbf{R}_{\theta_r, d_r}$ = N_r-by-N_r correlation matrix of the received signal vector for receive angle spread θ_r and antenna spacing d_r

In light of these definitions, decomposition of the N_r-by-N_t complex transfer function $\mathbf{H}$ of the (N_r, N_t) link, as shown in Eq. (6.146), is intuitively satisfying on the following grounds:

- The individual matrix terms are arranged in reverse order, starting from the receive antennas and progressing toward the transmitter, stage by stage, exactly in the same order as that shown in Fig. 6.45.
- The overall dimensions of the multiple matrix product are N_r by N_t, as is readily shown by the fact that

$$(N_r \times N_r) \cdot (N_r \times S) \cdot (S \times S) \cdot (S \times N_t) \cdot (N_t \times N_t) = (N_r \times N_t)$$

- Each element of the square matrices $\mathbf{R}_{\theta_t, d_t}$, $\mathbf{R}_{\theta_S, 2D_r/S}$ and $\mathbf{R}_{\theta_r, d_r}$ represents an autocorrelation function, the unit of which is the square of that for a transmitted or received signal. In contrast, each element of the channel transfer matrix $\mathbf{H}$ is defined as a signal ratio—hence the need for taking the square root of these three correlation matrices.
- Each element of the rectangular matrices $\mathbf{G}_t$ and $\mathbf{G}_r$ is a dimensionless ratio.

The rank of the channel matrix $\mathbf{H}$ is controlled primarily by the middle matrix term, $\mathbf{R}_{\theta_S, 2D_r/S}^{1/2}$, which is entirely due to scatterer-to-scatterer cross-correlations. Moreover, the matrix decomposition of Eq. (6.146) shows that it is indeed possible for the Rayleigh-fading processes at both the transmit and receive ends of the MIMO link to be completely uncorrelated (i.e., the matrices $\mathbf{G}_t$ and $\mathbf{G}_r$ to be of full rank), yet the overall rank of the channel matrix $\mathbf{H}$ may collapse to unity because the matrix

$\mathbf{R}_{\theta_S, 2D_r/S}^{1/2}$ is of unit rank. In such an eventuality, the MIMO link effectively assumes a *single degree of freedom* and, accordingly, exhibits *perfect cross-correlations* between the array of receive antennas and the array of transmit antennas. This special form of MIMO link is called a *keyhole channel* or *pinhole channel*, whose capacity is equivalent to that of a single-input, single-output link operating at the same signal-to-noise ratio as that of the MIMO link.

In physical terms, the keyhole channel arises when the product of the radial extents of the transmit and receive sets of scatterers, divided by the range between the transmit and receive antennas, is small compared with the wavelength λ of the transmitted radio signal; that is,

$$\frac{D_t D_r}{R} << \lambda \tag{6.147}$$

Under this condition, the angle spread θ_S of the scatterers at both ends of the link assumes a small value, and with it, the rank of the matrix $\mathbf{R}_{\theta_S, 2D_r/S}^{1/2}$ collapses.

The occurrence of keyhole channels has been confirmed by computer simulations and outdoor measurements. Fortunately, the loss in rank due to keyhole channels does not appear to be prevalent in most physical environments, because the condition of Eq. (6.147) is realized only infrequently.

6.12 SUMMARY AND DISCUSSION

In this chapter, we discussed different forms of space diversity, the main idea behind which is that two or more propagation paths connecting the receiver to the transmitter are better than a single propagation path. In historical terms, the first form of space diversity used to mitigate the multipath fading problem was that of *receive diversity*, involving a single transmit antenna and multiple receive antennas. Under receive diversity, we discussed the selection combiner, maximal-ratio combiner, equal-gain combiner, and square-law combiner. The selection combiner is the simplest form of receive diversity, operating on the principle that it is possible to select, among N_r receive-diversity branches, a particular diversity branch with the largest output signal-to-noise ratio; the branch so selected defines the desired received signal. The maximal-ratio combiner is more powerful than the selection combiner by virtue of the fact that it exploits the full information content of all the N_r receive-diversity branches about the transmitted signal of interest. The maximal-ratio combiner is characterized by a set of N_r receive-complex weighting factors, which are chosen to maximize the output signal-to-noise ratio of the combiner. The equal-gain combiner is a simplified version of the maximal-ratio combiner, with each weighting parameter set equal to unity. The square-law combiner differs from the selection combiner, maximal-ratio combiner, and equal-gain combiner in that it is nonlinear, but applicable only to orthogonal modulation techniques.

By far the most powerful form of space diversity is the use of multiple antennas at both the transmit and receive ends of the wireless link. The resulting configuration is