# 3

# Point-to-point communication: detection, diversity, and channel uncertainty

In this chapter we look at various basic issues that arise in communication over fading channels. We start by analyzing uncoded transmission in a narrowband fading channel. We study both coherent and non-coherent detection. In both cases the error probability is much higher than in a non-faded AWGN channel. The reason is that there is a significant probability that the channel is in a deep fade. This motivates us to investigate various *diversity* techniques that improve the performance. The diversity techniques operate over time, frequency or space, but the basic idea is the same. By sending signals that carry the same information through different paths, multiple independently faded replicas of data symbols are obtained at the receiver end and more reliable detection can be achieved. The simplest diversity schemes use *repetition coding*. More sophisticated schemes exploit channel diversity and, at the same time, efficiently use the degrees of freedom in the channel. Compared to repetition coding, they provide *coding gains* in addition to *diversity gains*. In space diversity, we look at both transmit and receive diversity schemes. In frequency diversity, we look at three approaches:

- single-carrier with inter-symbol interference equalization,
- direct-sequence spread-spectrum,
- orthogonal frequency division multiplexing.

Finally, we study the impact of channel uncertainty on the performance of diversity combining schemes. We will see that, in some cases, having too many diversity paths can have an adverse effect due to channel uncertainty.

To familiarize ourselves with the basic issues, the emphasis of this chapter is on concrete techniques for communication over fading channels. In Chapter 5 we take a more fundamental and systematic look and use information theory to derive the *best* performance one can achieve. At that fundamental level, we will see many of the issues discussed here recur.

The derivations in this chapter make repeated use of a few key results in vector detection under Gaussian noise. We develop and summarize the basic results in Appendix A, emphasizing the underlying geometry. The reader is

encouraged to take a look at the appendix before proceeding with this chapter and to refer back to it often. In particular, a thorough understanding of the canonical detection problem in Summary A.2 will be very useful.

# 3.1 Detection in a Rayleigh fading channel

## 3.1.1 Non-coherent detection

We start with a very simple detection problem in a fading channel. For simplicity, let us assume a flat fading model where the channel can be represented by a single discrete-time complex filter tap $h_0[m]$, which we abbreviate as $h[m]$:

$$y[m] = h[m]x[m] + w[m], \tag{3.1}$$

where $w[m] \sim \mathcal{CN}(0, N_0)$. We suppose Rayleigh fading, i.e., $h[m] \sim \mathcal{CN}(0, 1)$, where we normalize the variance to be 1. For the time being, however, we do not specify the dependence between the fading coefficients $h[m]$ at different times $m$ nor do we make any assumption on the prior knowledge the receiver might have of $h[m]$. (This latter assumption is sometimes called *non-coherent* communication.)

First consider uncoded binary antipodal signaling (or binary phase-shift-keying, BPSK) with amplitude $a$, i.e., $x[m] = \pm a$, and the symbols $x[m]$ are independent over time. This signaling scheme fails completely, even in the absence of noise, since the phase of the received signal $y[m]$ is uniformly distributed between 0 and $2\pi$ regardless of whether $x[m] = a$ or $x[m] = -a$ is transmitted. Further, the received amplitude is independent of the transmitted symbol. Binary antipodal signaling is binary phase modulation and it is easy to see that phase modulation in general is similarly flawed. Thus, signal structures are required in which either different signals have different magnitudes, or coding between symbols is used. Next we look at orthogonal signaling, a special type of coding between symbols.

Consider the following simple orthogonal modulation scheme: a form of binary pulse-position modulation. For a pair of time samples, transmit either

$$\mathbf{x}_A := \begin{pmatrix} x[0] \\ x[1] \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix}, \tag{3.2}$$

or

$$\mathbf{x}_B := \begin{pmatrix} 0 \\ a \end{pmatrix}. \tag{3.3}$$

We would like to perform detection based on

$$\mathbf{y} := \begin{pmatrix} y[0] \\ y[1] \end{pmatrix}. \tag{3.4}$$

This is a simple hypothesis testing problem, and it is straightforward to derive the maximum likelihood (ML) rule:

$$\Lambda(\mathbf{y}) \underset{\mathbf{x}_B}{\overset{\mathbf{x}_A}{\gtrless}} 0, \tag{3.5}$$

where $\Lambda(\mathbf{y})$ is the log-likelihood ratio

$$\Lambda(\mathbf{y}) := \ln\left\{\frac{f(\mathbf{y}|\mathbf{x}_A)}{f(\mathbf{y}|\mathbf{x}_B)}\right\}. \tag{3.6}$$

It can be seen that, if $\mathbf{x}_A$ is transmitted, $y[0] \sim \mathcal{CN}(0, a^2 + N_0)$ and $y[1] \sim \mathcal{CN}(0, N_0)$ and $y[0], y[1]$ are independent. Similarly, if $\mathbf{x}_B$ is transmitted, $y[0] \sim \mathcal{CN}(0, N_0)$ and $y[1] \sim \mathcal{CN}(0, a^2 + N_0)$. Further, $y[0]$ and $y[1]$ are independent. Hence the log-likelihood ratio can be computed to be

$$\Lambda(\mathbf{y}) = \frac{\left\{|y[0]|^2 - |y[1]|^2\right\} a^2}{(a^2 + N_0)N_0}. \tag{3.7}$$

The optimal rule is simply to decide $\mathbf{x}_A$ is transmitted if $|y[0]|^2 > |y[1]|^2$ and decide $\mathbf{x}_B$ otherwise. Note that the rule does not make use of the phases of the received signal, since the random unknown phases of the channel gains $h[0], h[1]$ render them useless for detection. Geometrically, we can interpret the detector as projecting the received vector $\mathbf{y}$ onto each of the two possible transmit vectors $\mathbf{x}_A$ and $\mathbf{x}_B$ and comparing the energies of the projections (Figure 3.1). Thus, this detector is also called an *energy* or a *square-law* detector. It is somewhat surprising that the optimal detector does not depend on how $h[0]$ and $h[1]$ are correlated.

We can analyze the error probability of this detector. By symmetry, we can assume that $\mathbf{x}_A$ is transmitted. Under this hypothesis, $y[0]$ and $y[1]$ are
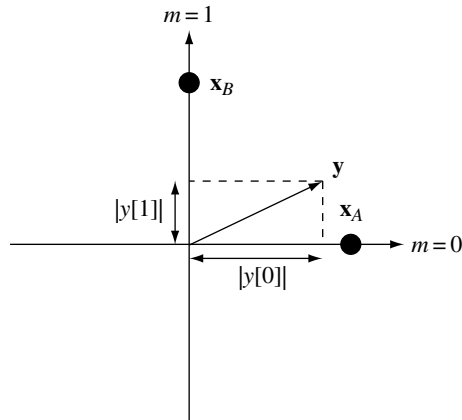


**Figure 3.1** The non-coherent detector projects the received vector $\mathbf{y}$ onto each of the two orthogonal transmitted vectors $\mathbf{x}_A$ and $\mathbf{x}_B$ and compares the lengths of the projections.

independent circular symmetric complex Gaussian random variables with variances $a^2 + N_0$ and $N_0$ respectively. (See Section A.1.3 in the appendices for a discussion on circular symmetric Gaussian random variables and vectors.) As shown there, $|y[0]|^2$, $|y[1]|^2$ are exponentially distributed with mean $a^2 + N_0$ and $N_0$ respectively.[1] The probability of error can now be computed by direct integration:

$$p_e = \mathbb{P}\left\{|y[1]|^2 > |y[0]|^2 | \mathbf{x}_A\right\} = \left[2 + \frac{a^2}{N_0}\right]^{-1}. \tag{3.8}$$

We make the general definition

$$\boxed{\mathsf{SNR} := \frac{\text{average received signal energy per (complex) symbol time}}{\text{noise energy per (complex) symbol time}}} \tag{3.9}$$

which we use consistently throughout the book *for any modulation scheme.* The noise energy per complex symbol time is $N_0$.[2] For the orthogonal modulation scheme here, the average received energy per symbol time is $a^2/2$ and so

$$\mathsf{SNR} := \frac{a^2}{2N_0}. \tag{3.10}$$

Substituting into (3.8), we can express the error probability of the orthogonal scheme in terms of $\mathsf{SNR}$:

$$p_e = \frac{1}{2(1 + \mathsf{SNR})}. \tag{3.11}$$

This is a very discouraging result. To get an error probability $p_e = 10^{-3}$ one would require $\mathsf{SNR} \approx 500$ (27 dB). Stupendous amounts of power would be required for more reliable communication.

### 3.1.2 Coherent detection

Why is the performance of the non-coherent maximum likelihood (ML) receiver on a fading channel so bad? It is instructive to compare its performance with detection in an AWGN channel without fading:

$$y[m] = x[m] + w[m]. \tag{3.12}$$

---

[1]  Recall that a random variable $U$ is exponentially distributed with mean $\mu$ if its pdf is $f_U(u) = \frac{1}{\mu}e^{-u/\mu}$.

[2]  The orthogonal modulation scheme considered here uses only real symbols and hence transmits only on the I channel. Hence it may seem more natural to define the SNR in terms of noise energy per *real* symbol, i.e., $N_0/2$. However, later we will consider modulation schemes that use complex symbols and hence transmit on both the I and Q channels. In order to be consistent throughout, we choose to define SNR this way.

For antipodal signaling (BPSK), $x[m] = \pm a$, a sufficient statistic is $\Re\{y[m]\}$ and the error probability is

$$p_e = Q\left(\frac{a}{\sqrt{N_0/2}}\right) = Q\left(\sqrt{2\mathsf{SNR}}\right), \qquad (3.13)$$

where $\mathsf{SNR} = a^2/N_0$ is the received signal-to-noise ratio per symbol time, and $Q(\cdot)$ is the complementary cumulative distribution function of an $N(0, 1)$ random variable. This function decays exponentially with $x^2$; more specifically,

$$Q(x) < e^{-x^2/2}, \qquad x > 0 \qquad (3.14)$$

and

$$Q(x) > \frac{1}{\sqrt{2\pi}x}\left(1 - \frac{1}{x^2}\right)e^{-x^2/2}, \qquad x > 1. \qquad (3.15)$$

Thus, *the detection error probability decays exponentially in SNR in the AWGN channel while it decays only inversely with the SNR in the fading channel*. To get an error probability of $10^{-3}$, an SNR of only about 7 dB is needed in an AWGN channel (as compared to 27 dB in the non-coherent fading channel). Note that $2\sqrt{\mathsf{SNR}}$ is the separation between the two constellation points as a multiple of the standard deviation of the Gaussian noise; the above observation says that when this separation is much larger than 1, the error probability is very small.

Compared to detection in the AWGN channel, the detection problem considered in the previous section has two differences: the channel gains $h[m]$ are random, and the receiver is assumed not to know them. Suppose now that the channel gains are tracked at the receiver so that they are known at the receiver (but still random). In practice, this is done either by sending a known sequence (called a *pilot* or training sequence) or in a decision directed manner, estimating the channel using symbols detected earlier. The accuracy of the tracking depends, of course, on how fast the channel varies. For example, in a narrowband 30-kHz channel (such as that used in the North American TDMA cellular standard IS-136) with a Doppler spread of 100 Hz, the coherence time $T_c$ is roughly 80 symbols and in this case the channel can be estimated with minimal overhead expended in the pilot.[3] For our current purpose, let us suppose that the channel estimates are perfect.

Knowing the channel gains, *coherent* detection of BPSK can now be performed on a symbol by symbol basis. We can focus on one symbol time and drop the time index

$$y = hx + w \qquad (3.16)$$

---

[3]  The channel estimation problem for a broadband channel with many taps in the impulse response is more difficult; we will get to this in Section 3.5.

Detection of $x$ from $y$ can be done in a way similar to that in the AWGN case; the decision is now based on the sign of the real sufficient statistic

$$r := \Re\{(h/|h|)^* y\} = |h|x + z, \tag{3.17}$$

where $z \sim N(0, N_0/2)$. If the transmitted symbol is $x = \pm a$, then, for a given value of $h$, the error probability of detecting $x$ is

$$Q\left(\frac{a|h|}{\sqrt{N_0/2}}\right) = Q\left(\sqrt{2|h|^2 \mathsf{SNR}}\right) \tag{3.18}$$

where $\mathsf{SNR} = a^2/N_0$ is the average received signal-to-noise ratio per symbol time. (Recall that we normalized the channel gain such that $\mathbb{E}[|h|^2] = 1$.) We average over the random gain $h$ to find the overall error probability. For Rayleigh fading when $h \sim \mathcal{CN}(0, 1)$, direct integration yields

$$p_\mathrm{e} = \mathbb{E}\left[Q\left(\sqrt{2|h|^2 \mathsf{SNR}}\right)\right] = \frac{1}{2}\left(1 - \sqrt{\frac{\mathsf{SNR}}{1 + \mathsf{SNR}}}\right). \tag{3.19}$$

(See Exercise 3.1.) Figure 3.2 compares the error probabilities of coherent BPSK and non-coherent orthogonal signaling over the Rayleigh fading channel, as well as BPSK over the AWGN channel. We see that while the error probability for BPSK over the AWGN channel decays very fast with the SNR, the error probabilities for the Rayleigh fading channel are much worse,
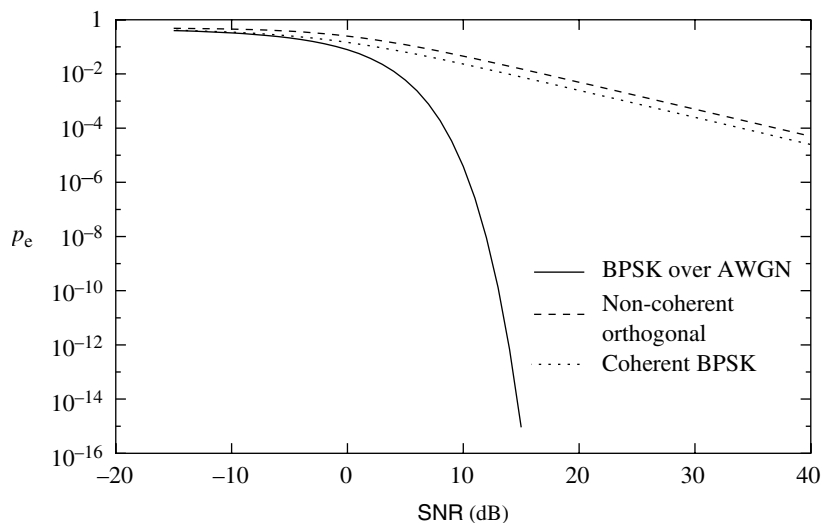


**Figure 3.2** Performance of coherent BPSK vs. non-coherent orthogonal signaling over Rayleigh fading channel vs. BPSK over AWGN schannel.

whether the detection is coherent or non-coherent. At high SNR, Taylor series expansion yields

$$\sqrt{\frac{\mathsf{SNR}}{1+\mathsf{SNR}}} = 1 - \frac{1}{2\mathsf{SNR}} + O\left(\frac{1}{\mathsf{SNR}^2}\right). \tag{3.20}$$

Substituting into (3.19), we get the approximation

$$p_\mathrm{e} \approx \frac{1}{4\mathsf{SNR}}, \tag{3.21}$$

which decays inversely proportional to the SNR, just as in the non-coherent orthogonal signaling scheme (cf. (3.11)). There is only a 3 dB difference in the required SNR between the coherent and non-coherent schemes; in contrast, at an error probability of $10^{-3}$, there is a 17 dB difference between the performance on the AWGN channel and coherent detection on the Rayleigh fading channel.[4]

We see that the main reason why detection in the fading channel has poor performance is not because of the lack of knowledge of the channel at the receiver. It is due to the fact that the channel gain is random and there is a significant probability that the channel is in a "deep fade". At high SNR, we can in fact be more precise about what a "deep fade" means by inspecting (3.18). The quantity $|h|^2\mathsf{SNR}$ is the instantaneous received SNR. Under typical channel conditions, i.e., $|h|^2\mathsf{SNR} \gg 1$, the conditional error probability is very small, since the tail of the $Q$-function decays very rapidly. In this regime, the separation between the constellation points is much larger than the standard deviation of the Gaussian noise. On the other hand, when $|h|^2\mathsf{SNR}$ is of the order of 1 or less, the separation is of the same order as the standard deviation of the noise and the error probability becomes significant. The probability of this event is

$$\mathbb{P}\{|h|^2\mathsf{SNR} < 1\} = \int_0^{1/\mathsf{SNR}} \mathrm{e}^{-x}\mathrm{d}x \tag{3.22}$$

$$= \frac{1}{\mathsf{SNR}} + O\left(\frac{1}{\mathsf{SNR}^2}\right). \tag{3.23}$$

This probability has the same order of magnitude as the error probability itself (cf. (3.21)). Thus, we can define a "deep fade" via an order-of-magnitude approximation:

$$\boxed{\begin{aligned} &\text{Deep fade event}: |h|^2 < \frac{1}{\mathsf{SNR}}. \\ &\qquad \mathbb{P}\{\text{deep fade}\} \approx \frac{1}{\mathsf{SNR}}. \end{aligned}}$$

---

[4] Communication engineers often compare schemes based on the difference in the required SNR to attain the same error probability. This corresponds to the horizontal gap between the error probability versus SNR curves of the two schemes.

We conclude that high-SNR error events most often occur because the channel is in deep fade and not as a result of the additive noise being large. In contrast, in the AWGN channel the only possible error mechanism is for the additive noise to be large. Thus, the error probability performance over the AWGN channel is much better.

We have used the explicit error probability expression (3.19) to help identify the typical error event at high SNR. We can in fact turn the table around and use it as a basis for an approximate analysis of the high-SNR performance (Exercises 3.2 and 3.3). Even though the error probability $p_e$ can be directly computed in this case, the approximate analysis provides much insight as to how typical errors occur. Understanding typical error events in a communication system often suggests how to improve it. Moreover, the approximate analysis gives some hints as to how robust the conclusion is to the Rayleigh fading model. In fact, the only aspect of the Rayleigh fading model that is important to the conclusion is the fact that $\mathbb{P}\{|h|^2 < \epsilon\}$ is proportional to $\epsilon$ for $\epsilon$ small. This holds whenever the pdf of $|h|^2$ is positive and continuous at 0.

### 3.1.3 From BPSK to QPSK: exploiting the degrees of freedom

In Section 3.1.2, we have considered BPSK modulation, $x[m] = \pm a$. This uses only the real dimension (the I channel), while in practice both the I and Q channels are used simultaneously in coherent communication, increasing spectral efficiency. Indeed, an extra bit can be transmitted by instead using QPSK (quadrature phase-shift-keying) modulation, i.e., the constellation is

$$\{a(1+j), a(1-j), a(-1+j), a(-1-j)\}; \tag{3.24}$$

in effect, a BPSK symbol is transmitted on each of the I and Q channels simultaneously. Since the noise is independent across the I and Q channels, the bits can be detected separately and the bit error probability on the AWGN channel (cf. (3.12)) is

$$Q\left(\sqrt{\frac{2a^2}{N_0}}\right), \tag{3.25}$$

the same as BPSK (cf. (3.13)). For BPSK, the SNR (as defined in (3.9)) is given by

$$\text{SNR} = \frac{a^2}{N_0}, \tag{3.26}$$

while for QPSK,

$$\text{SNR} = \frac{2a^2}{N_0}, \tag{3.27}$$

is twice that of BPSK since both the I and Q channels are used. Equivalently, for a given SNR, the bit error probability of BPSK is $Q(\sqrt{2\mathsf{SNR}})$ (cf. (3.13)) and that of QPSK is $Q(\sqrt{\mathsf{SNR}})$. The error probability of QPSK under Rayleigh fading can be similarly obtained by replacing $\mathsf{SNR}$ by $\mathsf{SNR}/2$ in the corresponding expression (3.19) for BPSK to yield

$$p_e = \frac{1}{2}\left(1 - \sqrt{\frac{\mathsf{SNR}}{2+\mathsf{SNR}}}\right) \approx \frac{1}{2\mathsf{SNR}}. \tag{3.28}$$

at high SNR. For expositional simplicity, we will consider BPSK modulation in many of the discussions in this chapter, but the results can be directly mapped to QPSK modulation.

One important point worth noting is that it is much more energy-efficient to use both the I and Q channels rather than just one of them. For example, if we had to send the two bits carried by the QPSK symbol on the I channel alone, then we would have to transmit a 4-PAM symbol. The constellation is $\{-3b, -b, b, 3b\}$ and the average error probability on the AWGN channel is

$$\frac{3}{2}Q\left(\sqrt{\frac{2b^2}{N_0}}\right). \tag{3.29}$$

To achieve approximately the same error probability as QPSK, the argument inside the $Q$-function should be the same as that in (3.25) and hence $b$ should be the same as $a$, i.e., the same minimum separation between points in the two constellations (Figure 3.3). But QPSK requires a transmit energy of $2a^2$ per symbol, while 4-PAM requires a transmit energy of $5b^2$ per symbol. Hence, for the same error probability, approximately 2.5 times more transmit energy is needed: a 4 dB worse performance. Exercise 3.4 shows that this loss is even more significant for larger constellations. The loss is due to the fact that it is more energy efficient to pack, for a desired minimum distance separation, a
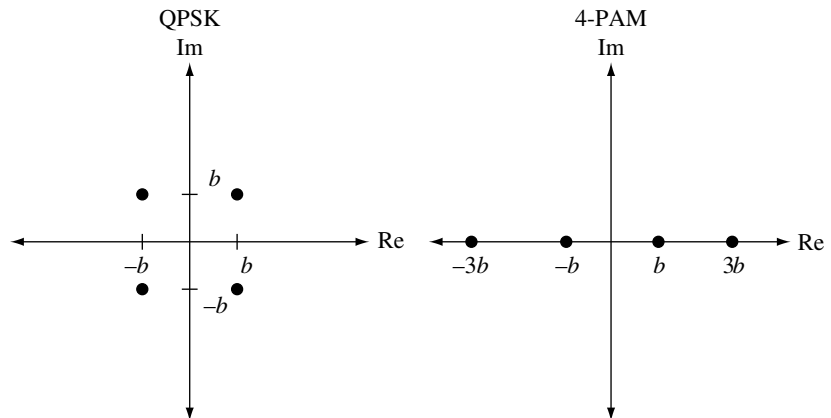


**Figure 3.3** QPSK versus 4-PAM: for the same minimum separation between constellation points, the 4-PAM constellation requires higher transmit power.

given number of constellation points in a higher-dimensional space than in a lower-dimensional space. We have thus arrived at a general design principle (cf. Discussion 2.1):

> A good communication scheme exploits all the available degrees of freedom in the channel.

This important principle will recur throughout the book, and in fact will be shown to be of a fundamental nature as we talk about channel capacity in Chapter 5. Here, the choice is between using just the I channel and using both the I and Q channels, but the same principle applies to many other situations. As another example, the non-coherent orthogonal signaling scheme discussed in Section 3.1.1 conveys one bit of information and uses one real dimension per two symbol times (Figure 3.4). This scheme does not assume any relationship between consecutive channel gains, but if we assume that they do not change much from symbol to symbol, an alternative scheme is *differential* BPSK, which conveys information in the relative phases of consecutive transmitted symbols. That is, if the BPSK information symbol is $u[m]$ at time $m$ ($u[m] = \pm 1$), the transmitted symbol at time $m$ is given by

$$x[m] = u[m]x[m-1]. \tag{3.30}$$

Exercise 3.5 shows that differential BPSK can be demodulated non-coherently at the expense of a 3-dB loss in performance compared to coherent BPSK (at high SNR). But since non-coherent orthogonal modulation also has a 3-dB worse performance compared to coherent BPSK, this implies that differential BPSK and non-coherent orthogonal modulation have the *same* error probability performance. On the other hand, differential BPSK conveys one
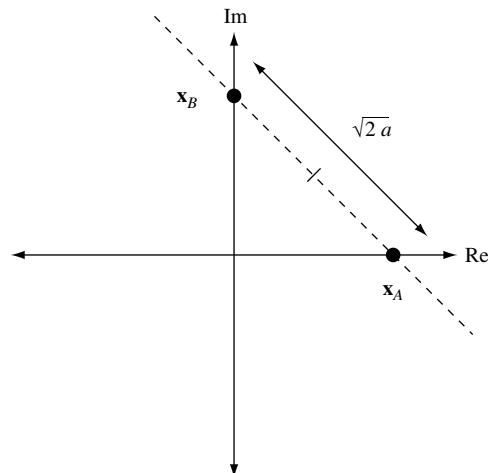


**Figure 3.4** Geometry of orthogonal modulation. Signaling is performed over one real dimension, but two (complex) symbol times are used.

bit of information and uses one real dimension per *single* symbol time, and therefore has twice the spectral efficiency of orthogonal modulation. Better performance is achieved because differential BPSK uses more efficiently the available degrees of freedom.

## 3.1.4 Diversity

The performance of the various schemes considered so far for fading channels is summarized in Table 3.1. Some schemes are spectrally more efficient than others, but from a practical point of view, they are all bad: the error probabilities all decay very slowly, like 1/SNR. From Section 3.1.2, it can be seen that the root cause of this poor performance is that reliable communication depends on the strength of a single signal path. There is a significant probability that this path will be in a deep fade. When the path is in a deep fade, any communication scheme will likely suffer from errors. A natural solution to improve the performance is to ensure that the information symbols pass through multiple signal paths, each of which fades independently, making sure that reliable communication is possible as long as one of the paths is strong. This technique is called *diversity*, and it can dramatically improve the performance over fading channels.

There are many ways to obtain diversity. Diversity over **time** can be obtained via *coding* and *interleaving*: information is coded and the coded symbols are dispersed over time in different coherence periods so that different parts of the codewords experience independent fades. Analogously, one can also exploit diversity over **frequency** if the channel is frequency-selective. In a channel with multiple transmit or receive antennas spaced sufficiently, diversity can be obtained over **space** as well. In a cellular network, **macro-diversity** can be exploited by the fact that the signal from a mobile can be received at two base-stations. Since diversity is such an important resource, a wireless system typically uses several types of diversity.

In the next few sections, we will discuss diversity techniques in time, frequency and space. In each case, we start with a simple scheme based on *repetition coding*: the same information symbol is transmitted over several signal paths. While repetition coding achieves the maximal diversity gain, it is usually quite wasteful of the degrees of freedom of the channel. More sophisticated schemes can increase the data rate and achieve a *coding gain* along with the diversity gain.

To keep the discussion simple we begin by focusing on the coherent scenario: the receiver has perfect knowledge of the channel gains and can coherently combine the received signals in the diversity paths. As discussed in the previous section, this knowledge is learnt via training (pilot) symbols and the accuracy depends on the coherence time of the channel and the received power of the transmitted signal. We discuss the impact of channel measurement error and non-coherent diversity combining in Section 3.5.

**Table 3.1** Performance of coherent and non-coherent schemes under Rayleigh fading. The data rates are in bits/s/Hz, which is the same as bits per complex symbol time. The performance of differential QPSK is derived in Exercise 3.5. It is also 3-dB worse than coherent QPSK.

| Scheme | Bit error prob. (High SNR) | Data rate (bits/s/Hz) |
|---|---|---|
| Coherent BPSK | 1/(4SNR) | 1 |
| Coherent QPSK | 1/(2SNR) | 2 |
| Coherent 4-PAM | 5/(4SNR) | 2 |
| Coherent 16-QAM | 5/(2SNR) | 4 |
| Non-coherent orth. mod. | 1/(2SNR) | 1/2 |
| Differential BPSK | 1/(2SNR) | 1 |
| Differential QPSK | 1/SNR | 2 |

## 3.2 Time diversity

Time diversity is achieved by averaging the fading of the channel over time. Typically, the channel coherence time is of the order of tens to hundreds of symbols, and therefore the channel is highly correlated across consecutive symbols. To ensure that the coded symbols are transmitted through independent or nearly independent fading gains, *interleaving* of codewords is required (Figure 3.5). For simplicity, let us consider a flat fading channel. We transmit a codeword $\mathbf{x} = [x_1, \ldots, x_L]^t$ of length $L$ symbols and the received signal is given by

$$y_\ell = h_\ell x_\ell + w_\ell, \qquad \ell = 1, \ldots, L. \tag{3.31}$$

Assuming ideal interleaving so that consecutive symbols $x_\ell$ are transmitted sufficiently far apart in time, we can assume that the $h_\ell$ are independent. The parameter $L$ is commonly called the number of *diversity branches*. The additive noises $w_1, \ldots, w_L$ are i.i.d. $\mathcal{CN}(0, N_0)$ random variables.
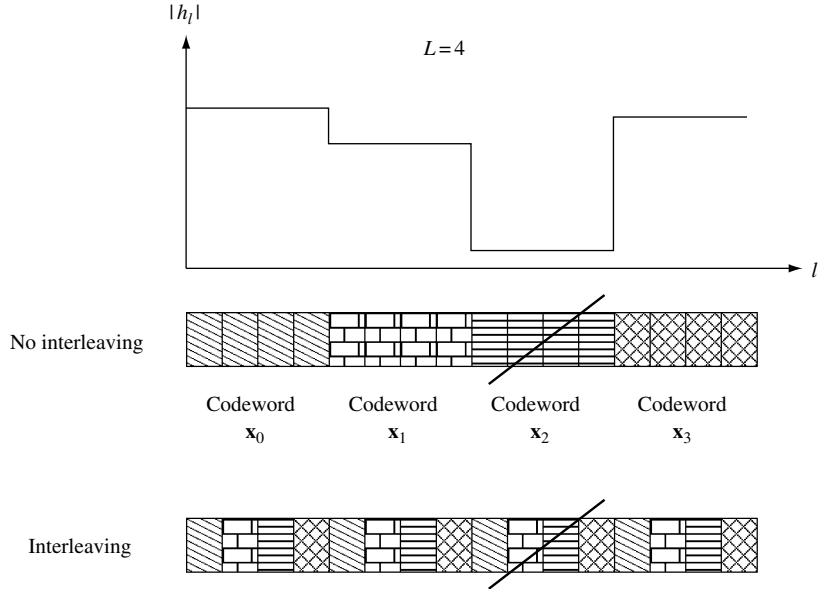
### 3.2.1 Repetition coding

The simplest code is a *repetition code*, in which $x_\ell = x_1$ for $\ell = 1, \ldots, L$. In vector form, the overall channel becomes

$$\mathbf{y} = \mathbf{h}x_1 + \mathbf{w}, \tag{3.32}$$

where $\mathbf{y} = [y_1, \ldots, y_L]^t$, $\mathbf{h} = [h_1, \ldots, h_L]^t$ and $\mathbf{w} = [w_1, \ldots, w_L]^t$.

$|h_l|$

$L = 4$

$l$

No interleaving

Codeword $\mathbf{x}_0$    Codeword $\mathbf{x}_1$    Codeword $\mathbf{x}_2$    Codeword $\mathbf{x}_3$

Interleaving

Consider now coherent detection of $x_1$, i.e., the channel gains are known to the receiver. This is the canonical vector Gaussian detection problem in Summary A.2 of Appendix A. The scalar

$$\frac{\mathbf{h}^*}{\|\mathbf{h}\|}\mathbf{y} = \|\mathbf{h}\|x_1 + \frac{\mathbf{h}^*}{\|\mathbf{h}\|}\mathbf{w} \tag{3.33}$$

is a sufficient statistic. Thus, we have an equivalent scalar detection problem with noise $(\mathbf{h}^*/\|\mathbf{h}\|)\mathbf{w} \sim \mathcal{CN}(0, N_0)$. The receiver structure is a *matched filter* and is also called a *maximal ratio combiner*: it weighs the received signal in each branch in proportion to the signal strength and also aligns the phases of the signals in the summation to maximize the output SNR. This receiver structure is also called *coherent combining*.

Consider BPSK modulation, with $x_1 = \pm a$. The error probability, conditional on $\mathbf{h}$, can be derived exactly as in (3.18):

$$Q\left(\sqrt{2\|\mathbf{h}\|^2\mathsf{SNR}}\right) \tag{3.34}$$

where as before $\mathsf{SNR} = a^2/N_0$ is the average received signal-to-noise ratio per (complex) symbol time, and $\|\mathbf{h}\|^2\mathsf{SNR}$ is the received SNR for a given channel vector $\mathbf{h}$. We average over $\|\mathbf{h}\|^2$ to find the overall error probability. Under Rayleigh fading with each gain $h_\ell$ i.i.d. $\mathcal{CN}(0, 1)$,

$$\|\mathbf{h}\|^2 = \sum_{\ell=1}^{L} |h_\ell|^2 \tag{3.35}$$

is a sum of the squares of $2L$ independent real Gaussian random variables, each term $|h_\ell|^2$ being the sum of the squares of the real and imaginary parts of $h_\ell$. It is Chi-square distributed with $2L$ degrees of freedom, and the density is given by

$$f(x) = \frac{1}{(L-1)!} x^{L-1} e^{-x}, \qquad x \geq 0. \tag{3.36}$$

The average error probability can be explicitly computed to be (cf. Exercise 3.6)

$$p_e = \int_0^\infty Q\left(\sqrt{2x\mathsf{SNR}}\right) f(x) \, dx$$

$$= \left(\frac{1-\mu}{2}\right)^L \sum_{\ell=0}^{L-1} \binom{L-1+\ell}{\ell} \left(\frac{1+\mu}{2}\right)^\ell, \tag{3.37}$$
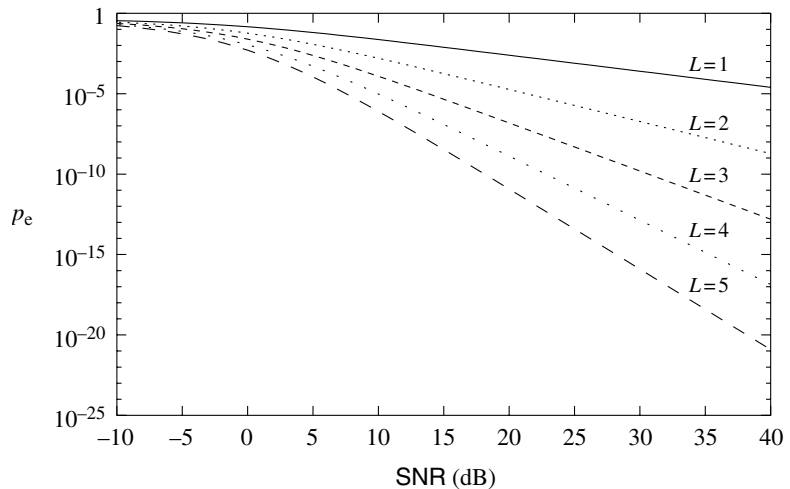
where

$$\mu := \sqrt{\frac{\mathsf{SNR}}{1+\mathsf{SNR}}}. \tag{3.38}$$

The error probability as a function of the SNR for different numbers of diversity branches $L$ is plotted in Figure 3.6. Increasing $L$ dramatically decreases the error probability.

At high SNR, we can see the role of $L$ analytically: consider the leading term in the Taylor series expansion in $1/\mathsf{SNR}$ to arrive at the approximations

$$\frac{1+\mu}{2} \approx 1, \quad \text{and} \quad \frac{1-\mu}{2} \approx \frac{1}{4\mathsf{SNR}}. \tag{3.39}$$



**Figure 3.6** Error probability as a function of SNR for different numbers of diversity branches $L$.

Furthermore,

$$\sum_{\ell=0}^{L-1} \binom{L-1+\ell}{\ell} = \binom{2L-1}{L}. \tag{3.40}$$

Hence,

$$p_e \approx \binom{2L-1}{L} \frac{1}{(4\mathsf{SNR})^L} \tag{3.41}$$

at high SNR. In particular, the error probability decreases as the $L$th power of SNR, corresponding to a slope of $-L$ in the error probability curve (in dB/dB scale).

To understand this better, we examine the probability of the deep fade event, as in our analysis in Section 3.1.2. The typical error event at high SNR is when the overall channel gain is small. This happens with probability

$$\mathbb{P}\{\|\mathbf{h}\|^2 < 1/\mathsf{SNR}\}. \tag{3.42}$$

Figure 3.7 plots the distribution of $\|\mathbf{h}\|^2$ for different values of $L$; clearly the tail of the distribution near zero becomes lighter for larger $L$. For small $x$, the probability density function of $\|\mathbf{h}\|^2$ is approximately

$$f(x) \approx \frac{1}{(L-1)!} x^{L-1} \tag{3.43}$$

and so

$$\mathbb{P}\{\|\mathbf{h}\|^2 < 1/\mathsf{SNR}\} \approx \int_0^{\frac{1}{\mathsf{SNR}}} \frac{1}{(L-1)!} x^{L-1} \mathrm{d}x = \frac{1}{L!} \frac{1}{\mathsf{SNR}^L}. \tag{3.44}$$



**Figure 3.7** The probability density function of $\|\mathbf{h}\|^2$ for different values of $L$. The larger the $L$, the faster the probability density function drops off around 0.
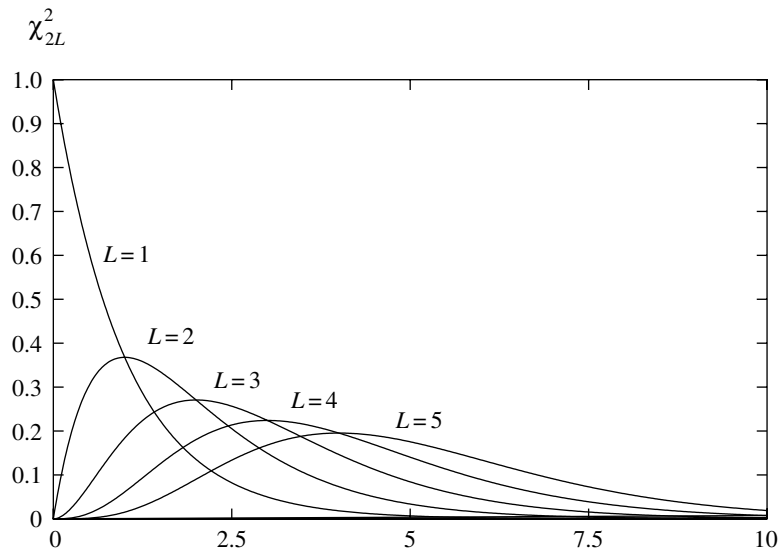
This analysis is too crude to get the correct constant before the $1/\mathsf{SNR}^L$ term in (3.41), but does get the correct exponent $L$. Basically, an error occurs when $\sum_{\ell=1}^{L} |h_\ell|^2$ is of the order of or smaller than $1/\mathsf{SNR}$, and this happens when *all* the magnitudes of the gains $|h_\ell|^2$ are small, of the order of $1/\mathsf{SNR}$. Since the probability that each $|h_\ell|^2$ is less than $1/\mathsf{SNR}$ is approximately $1/\mathsf{SNR}$ and the gains are independent, the probability of the overall gain being small is of the order $1/\mathsf{SNR}^L$. Typically, $L$ is called the *diversity gain* of the system.

## 3.2.2 Beyond repetition coding

The repetition code is the simplest possible code. Although it achieves a diversity gain, it does not exploit the degrees of freedom available in the channel effectively because it simply repeats the same symbol over the $L$ symbol times. By using more sophisticated codes, a coding gain can also be obtained beyond the diversity gain. There are many possible codes that one can use. We first focus on the example of a *rotation code* to explain some of the issues in code design for fading channels.

Consider the case $L = 2$. A repetition code which repeats a BPSK symbol $u = \pm a$ twice obtains a diversity gain of 2 but would only transmit one bit of information over the two symbol times. Transmitting two independent BPSK symbols $u_1, u_2$ over the two times would use the available degrees of freedom more efficiently, but of course offers no diversity gain: an error would be made whenever one of the two channel gains $h_1, h_2$ is in deep fade. To get both benefits, consider instead a scheme that transmits the vector

$$\mathbf{x} = \mathbf{R} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{3.45}$$

over the two symbol times, where

$$\mathbf{R} := \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3.46}$$

is a rotation matrix (for some $\theta \in (0, 2\pi)$). This is a code with four codewords:

$$\mathbf{x}_A = \mathbf{R} \begin{bmatrix} a \\ a \end{bmatrix}, \qquad \mathbf{x}_B = \mathbf{R} \begin{bmatrix} -a \\ a \end{bmatrix}, \qquad \mathbf{x}_C = \mathbf{R} \begin{bmatrix} -a \\ -a \end{bmatrix}, \qquad \mathbf{x}_D = \mathbf{R} \begin{bmatrix} a \\ -a \end{bmatrix}; \tag{3.47}$$

they are shown in Figure 3.8(a).[5] The received signal is given by

$$y_\ell = h_\ell x_\ell + w_\ell, \qquad \ell = 1, 2. \tag{3.48}$$

---

[5] Here communication is over the (real) I channel since both $x_1$ and $x_2$ are real, but as in Section 3.1.3, the spectral efficiency can be doubled by using both the I and the Q channels. Since the two channels are orthogonal, one can apply the same code separately to the symbols transmitted in the two channels to get the same performance gain.

**Figure 3.8** (a) Codewords of rotation code. (b) Codewords of repetition code.
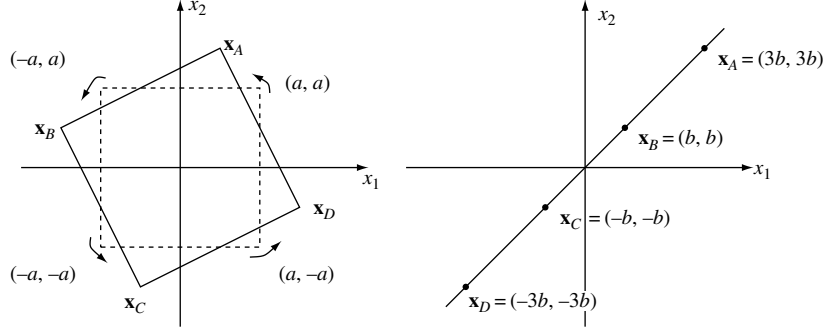


It is difficult to obtain an explicit expression for the exact error probability. So, we will proceed by looking at the union bound. Due to the symmetry of the code, without loss of generality we can assume $\mathbf{x}_A$ is transmitted. The union bound says that

$$p_e \leq \mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_B\} + \mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_C\} + \mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_D\}, \tag{3.49}$$

where $\mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_B\}$ is the pairwise error probability of confusing $\mathbf{x}_A$ with $\mathbf{x}_B$ when $\mathbf{x}_A$ is transmitted and when these are the only two hypotheses. Conditioned on the channel gains $h_1$ and $h_2$, this is just the binary detection problem in Summary A.2 of Appendix A, with

$$\mathbf{u}_A = \begin{bmatrix} h_1 x_{A1} \\ h_2 x_{A2} \end{bmatrix} \quad \text{and} \quad \mathbf{u}_B = \begin{bmatrix} h_1 x_{B1} \\ h_2 x_{B2} \end{bmatrix}. \tag{3.50}$$

Hence,

$$\mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_B | h_1, h_2\} = Q\left( \frac{\|\mathbf{u}_A - \mathbf{u}_B\|}{2\sqrt{N_0/2}} \right) = Q\left( \sqrt{\frac{\mathsf{SNR}(|h_1|^2|d_1|^2 + |h_2|^2|d_2|^2)}{2}} \right), \tag{3.51}$$

where $\mathsf{SNR} = a^2/N_0$ and

$$\mathbf{d} := \frac{1}{a}(\mathbf{x}_A - \mathbf{x}_B) = \begin{bmatrix} 2\cos\theta \\ 2\sin\theta \end{bmatrix} \tag{3.52}$$

is the normalized difference between the codewords, normalized such that the transmit energy is 1 per symbol time. We use the upper bound $Q(x) \leq e^{-x^2/2}$, for $x > 0$, in (3.51) to get

$$\mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_B | h_1, h_2\} \leq \exp\left( \frac{-\mathsf{SNR}(|h_1|^2|d_1|^2 + |h_2|^2|d_2|^2)}{4} \right). \tag{3.53}$$

Averaging with respect to $h_1$ and $h_2$ under the independent Rayleigh fading assumption, we get

$$
\mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_B\} \leq \mathbb{E}_{h_1,h_2}\left[\exp\left(\frac{-\mathsf{SNR}(|h_1|^2|d_1|^2 + |h_2|^2|d_2|^2)}{4}\right)\right]
$$

$$
= \left(\frac{1}{1 + \mathsf{SNR}|d_1|^2/4}\right)\left(\frac{1}{1 + \mathsf{SNR}|d_2|^2/4}\right). \tag{3.54}
$$

Here we have used the fact that the moment generating function for a unit mean exponential random variable $X$ is $\mathbb{E}[e^{sX}] = 1/(1-s)$ for $s < 1$. While it is possible to get an exact expression for the pairwise error probability, this upper bound is more explicit; moreover, it is asymptotically tight at high SNR (Exercise 3.7).

We first observe that if $d_1 = 0$ or $d_2 = 0$, then the diversity gain of the code is only 1. If they are both non-zero, then at high SNR the above bound on the pairwise error probability becomes

$$
\mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_B\} \leq \frac{16}{|d_1 d_2|^2}\mathsf{SNR}^{-2}, \tag{3.55}
$$

Call

$$
\boxed{\delta_{AB} := |d_1 d_2|^2,} \tag{3.56}
$$

the *squared product distance* between $\mathbf{x}_A$ and $\mathbf{x}_B$, when the average energy of the code is normalized to be 1 per symbol time (cf. (3.52)). This determines the pairwise error probability between the two codewords. Similarly, we can define $\delta_{ij}$ to be the squared product distance between $\mathbf{x}_i$ and $\mathbf{x}_j$, $i, j = A, B, C, D$. Combining (3.55) with (3.49) yields a bound on the overall error probability:

$$
p_e \leq 16\left(\frac{1}{\delta_{AB}} + \frac{1}{\delta_{AC}} + \frac{1}{\delta_{AD}}\right)\mathsf{SNR}^{-2}
$$

$$
\leq \frac{48}{\min_{j=B,C,D}\delta_{Aj}}\mathsf{SNR}^{-2}. \tag{3.57}
$$

We see that as long as $\delta_{ij} > 0$ for all $i, j$, we get a diversity gain of 2. The minimum squared product distance $\min_{j=B,C,D}\delta_{Aj}$ then determines the *coding gain* of the scheme beyond the diversity gain. This parameter depends on $\theta$, and we can optimize over $\theta$ to maximize the coding gain. Here

$$
\delta_{AB} = \delta_{AD} = 4\sin^2 2\theta, \qquad \text{and} \qquad \delta_{AC} = 16\cos^2 2\theta. \tag{3.58}
$$

The angle $\theta^*$ that maximizes the minimum squared product distance makes $\delta_{AB}$ equal $\delta_{AC}$, yielding $\theta^* = (1/2)\tan^{-1}2$ and $\min \delta_{ij} = 16/5$. The bound in (3.57) now becomes

$$p_e \leq 15\,\mathsf{SNR}^{-2}. \tag{3.59}$$

To get more insight into why the product distance is important, we see from (3.51) that the typical way for $\mathbf{x}_A$ to be confused with $\mathbf{x}_B$ is for the squared Euclidean distance $|h_1|^2|d_1|^2 + |h_2|^2|d_2|^2$ between the *received* codewords to be of the order of $1/\mathsf{SNR}$. This event holds roughly when both $|h_1|^2|d_1|^2$ and $|h_2|^2|d_2|^2$ are of the order of $1/\mathsf{SNR}$, and this happens with probability approximately

$$\left(\frac{1}{|d_1|^2\mathsf{SNR}}\right)\left(\frac{1}{|d_2|^2\mathsf{SNR}}\right) = \frac{1}{|d_1|^2|d_2|^2}\mathsf{SNR}^{-2}. \tag{3.60}$$

Thus, it is important that both $|d_1|^2$ and $|d_2|^2$ are large to ensure diversity against fading in both components.

It is interesting to see how this code compares to the repetition scheme. To keep the bit rate the same (2 bits over 2 real-valued symbols), the repetition scheme would be using 4-PAM modulation $\{-3b, -b, b, 3b\}$. The codewords of the repetition scheme are shown in Figure 3.8(b). From (3.51), the pairwise error probability between two adjacent codewords (say, $\mathbf{x}_A$ and $\mathbf{x}_B$) is

$$\mathbb{P}\{\mathbf{x}_A \rightarrow \mathbf{x}_B\} = \mathbb{E}\left[Q\left(\sqrt{\mathsf{SNR}/2 \cdot (|h_1|^2|d_1|^2 + |h_2|^2|d_2|^2)}\right)\right]. \tag{3.61}$$

But now $\mathsf{SNR} = 5b^2/N_0$ is the average SNR per symbol time for the 4-PAM constellation,[6] and $d_1 = d_2 = 2/\sqrt{5}$ are the normalized component differences between the adjacent codewords. The minimum squared product distance for the repetition code is therefore $16/25$ and we can compare this to the minimum squared product distance of $16/5$ for the previous rotation code. Since the error probability is proportional to $\mathsf{SNR}^{-2}$ in both cases, we conclude that the rotation code has an improved *coding gain* over the repetition code in terms of a saving in transmit power by a factor of $\sqrt{5}$ (3.5 dB) for the same product distance. This improvement comes from increasing the overall product distance, and this is in turn due to spreading the codewords in the two-dimensional space rather than packing them on a single-dimensional line as in the repetition code. This is the same reason that QPSK is more efficient than BPSK (as we have discussed in Section 3.1.3).

We summarize and generalize the above development to any time diversity code.

---

[6]  As we have seen earlier, the 4-PAM constellation requires five times more energy than BPSK for the same separation between the constellation points.

## Summary 3.1 Time diversity code design criterion

Ideal time-interleaved channel

$$y_\ell = h_\ell x_\ell + w_\ell, \qquad \ell = 1, \ldots, L, \tag{3.62}$$

where $h_\ell$ are i.i.d. $\mathcal{CN}(0, 1)$ Rayleigh faded channel gains.

$\mathbf{x}_1, \ldots, \mathbf{x}_M$ are the codewords of a time diversity code with block length $L$, normalized such that

$$\frac{1}{ML} \sum_{i=1}^{M} \|\mathbf{x}_i\|^2 = 1. \tag{3.63}$$

Union bound on overall probability of error:

$$p_e \leq \frac{1}{M} \sum_{i \neq j} \mathbb{P}\{\mathbf{x}_i \to \mathbf{x}_j\} \tag{3.64}$$

Bound on pairwise error probability:

$$\mathbb{P}\{\mathbf{x}_i \to \mathbf{x}_j\} \leq \prod_{\ell=1}^{L} \frac{1}{1 + \mathsf{SNR}|x_{i\ell} - x_{j\ell}|^2/4} \tag{3.65}$$

where $x_{i\ell}$ is the $\ell$th component of codeword $\mathbf{x}_i$, and $\mathsf{SNR} := 1/N_0$.

Let $L_{ij}$ be the number of components on which the codewords $\mathbf{x}_i$ and $\mathbf{x}_j$ differ. Diversity gain of the code is

$$\min_{i \neq j} L_{ij}. \tag{3.66}$$

If $L_{ij} = L$ for all $i \neq j$, then the code achieves the full diversity $L$ of the channel, and

$$p_e \leq \frac{4^L}{M} \sum_{i \neq j} \frac{1}{\delta_{ij}} \mathsf{SNR}^{-L} \leq \frac{4^L(M-1)}{\min_{i \neq j} \delta_{ij}} \mathsf{SNR}^{-L} \tag{3.67}$$

where

$$\delta_{ij} := \prod_{\ell=1}^{L} |x_{i\ell} - x_{j\ell}|^2 \tag{3.68}$$

is the squared product distance between $\mathbf{x}_i$ and $\mathbf{x}_j$.

The rotation code discussed above is specifically designed to exploit time diversity in *fading* channels. In the AWGN channel, however, rotation of the constellation does not affect performance since the i.i.d. Gaussian noise is invariant to rotations. On the other hand, codes that are designed for the AWGN channel, such as linear block codes or convolutional codes, can be used to extract time diversity in fading channels when combined with interleaving. Their performance can be analyzed using the general framework above. For example, the diversity gain of a binary linear block code where the coded symbols are ideally interleaved is simply the *minimum Hamming distance* between the codewords or equivalently the minimum weight of a codeword; the diversity gain of a binary convolutional code is given by the *free distance* of the code, which is the minimum weight of the coded sequence of the convolutional code. The performance analysis of these codes and various decoding techniques is further pursued in Exercise 3.11.

It should also be noted that the above code design criterion is derived assuming i.i.d. Rayleigh fading across the symbols. This can be generalized to the case when the coded symbols pass through *correlated* fades of the channel (see Exercise 3.12). Generalization to the case when the fading is Rician is also possible and is studied in Exercise 3.18. Nevertheless these code design criteria all depend on the specific channel statistics assumed. Motivated by information theoretic considerations, we take a completely different approach in Chapter 9 where we seek a *universal* criterion which works for *all* channel statistics. We will also be able to define what it means for a time-diversity code to be *optimal*.

---

**Example 3.1 Time diversity in GSM**

Global System for Mobile (GSM) is a digital cellular standard developed in Europe in the 1980s. GSM is a frequency division duplex (FDD) system and uses two 25-MHz bands, one for the uplink (mobiles to base-station) and one for the downlink (base-station to mobiles). The original bands set aside for GSM are the 890–915 MHz band (uplink) and the 935–960 MHz band (downlink). The bands are further divided into 200-kHz sub-channels and each sub-channel is shared by eight users in a time-division fashion (time-division multiple access (TDMA)). The data of each user are sent over time slots of length 577 microseconds ($\mu$s) and the time slots of the eight users together form a frame of length 4.615 ms (Figure 3.9).

Voice is the main application for GSM. Voice is coded by a speech encoder into speech frames each of length 20 ms. The bits in each speech frame are encoded by a convolutional code of rate 1/2, with the two generator polynomials $D^4 + D^3 + 1$ and $D^4 + D^3 + D + 1$. The number of coded bits for each speech frame is 456. To achieve time diversity, these coded bits are interleaved across eight consecutive time slots assigned to that specific user: the 0th, 8th, ..., 448th bits are put into the first time slot, the 1st, 9th, ..., 449th bits are put into the second time slot, etc.
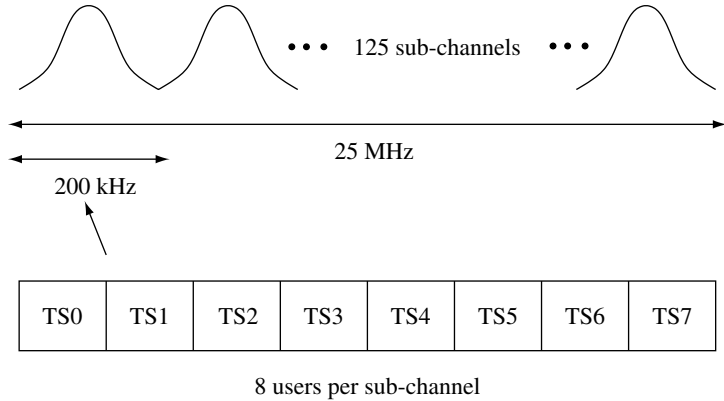
**Figure 3.9** The 25-MHz band of a GSM system is divided into 200-kHz sub-channels, which are further divided into time slots for eight different users.

Since one time slot occurs every 4.615 ms for each user, this translates into a delay of roughly 40 ms, a delay judged tolerable for voice. The eight time slots are shared between two 20-ms speech frames. The interleaving structure is summarized in Figure 3.10.

The maximum possible time diversity gain is 8, but the actual gain that can be obtained depends on how fast the channel varies, and that depends primarily on the mobile speed. If the mobile speed is $v$, then the largest possible Doppler spread (assuming full scattering in the environment) is $D_s = 2f_c v/c$, where $f_c$ is the carrier frequency and $c$ is the speed of light. (Recall the example in Section 2.1.4.) The coherence time is roughly $T_c = 1/(4D_s) = c/(8f_c v)$ (cf. (2.44)). For the channel to fade more or less independently across the different time slots for a user, the coherence time should be less than 5 ms. For $f_c = 900$ MHz, this translates into a mobile speed of at least 30 km/h.
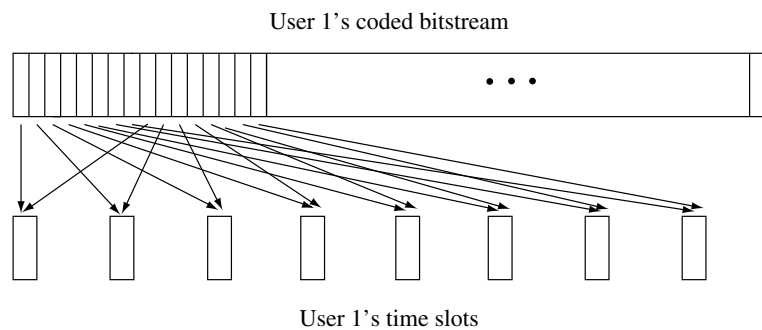


**Figure 3.10** How interleaving is done in GSM.

> For a walking speed of say 3 km/h, there may be too little time diversity. In this case, GSM can go into a *frequency hopping* mode, where consecutive frames (each composed of the time slots of the eight users) can hop from one 200-kHz sub-channel to another. With a typical delay spread of about 1 μs, the coherence bandwidth is 500 kHz (cf. Table 2.1). The total bandwidth equal to 25 MHz is thus much larger than the typical coherence bandwidth of the channel and the consecutive frames can be expected to fade independently. This provides the same effect as having time diversity. Section 3.4 discusses other ways to exploit frequency diversity.

## 3.3 Antenna diversity

To exploit time diversity, interleaving and coding over several coherence time periods is necessary. When there is a strict delay constraint and/or the coherence time is large, this may not be possible. In this case other forms of diversity have to be obtained. Antenna diversity, or spatial diversity, can be obtained by placing multiple antennas at the transmitter and/or the receiver. If the antennas are placed sufficiently far apart, the channel gains between different antenna pairs fade more or less independently, and independent signal paths are created. The required antenna separation depends on the local scattering environment as well as on the carrier frequency. For a mobile which is near the ground with many scatterers around, the channel decorrelates over shorter spatial distances, and typical antenna separation of half to one carrier wavelength is sufficient. For base-stations on high towers, larger antenna separation of several to tens of wavelengths may be required. (A more careful discussion of these issues is found in Chapter 7.)

We will look at both *receive diversity*, using multiple receive antennas (single input multiple output or SIMO channels), and *transmit diversity*, using multiple transmit antennas (multiple input single output or MISO channels). Interesting coding problems arise in the latter and have led to recent excitement in *space-time codes*. Channels with multiple transmit *and* multiple receive antennas (so-called multiple input multiple output or MIMO channels) provide even more potential. In addition to providing diversity, MIMO channels also provide additional *degrees of freedom* for communication. We will touch on some of the issues here using a $2 \times 2$ example; the full study of MIMO communication will be the subject of Chapters 7 to 10.

### 3.3.1 Receive diversity

In a flat fading channel with 1 transmit antenna and $L$ receive antennas (Figure 3.11(a)), the channel model is as follows:

$$y_\ell[m] = h_\ell[m]x[m] + w_\ell[m] \qquad \ell = 1, \ldots, L \qquad (3.69)$$

(a)                                    (b)                                    (c)
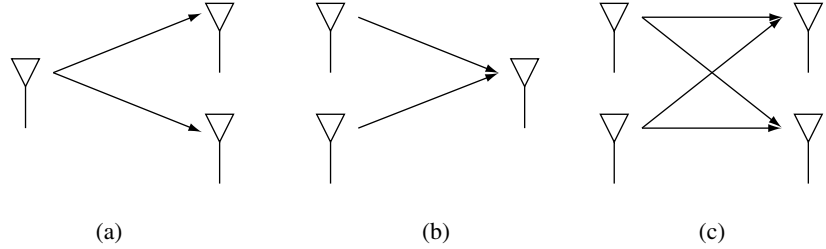
where the noise $w_\ell[m] \sim \mathcal{CN}(0, N_0)$ and is independent across the antennas. We would like to detect $x[1]$ based on $y_1[1], \ldots, y_L[1]$. This is exactly the same detection problem as in the use of a repetition code and interleaving over time, with $L$ diversity branches now over space instead of over time. If the antennas are spaced sufficiently far apart, we can assume that the gains $h_\ell[1]$ are independent Rayleigh, and we get a diversity gain of $L$.

With receive diversity, there are actually two types of gain as we increase $L$. This can be seen by looking at the expression (3.34) for the error probability of BPSK conditional on the channel gains:

$$Q\left(\sqrt{2\|\mathbf{h}\|^2 \mathsf{SNR}}\right). \tag{3.70}$$

We can break up the total received SNR conditioned on the channel gains into a product of two terms:

$$\|\mathbf{h}\|^2 \mathsf{SNR} = L\mathsf{SNR} \cdot \frac{1}{L}\|\mathbf{h}\|^2. \tag{3.71}$$

The first term corresponds to a *power gain* (also called *array gain*): by having multiple receive antennas and coherent combining at the receiver, the effective total received signal power increases linearly with $L$: doubling $L$ yields a 3-dB power gain.[7] The second term reflects the *diversity gain*: by averaging over multiple independent signal paths, the probability that the overall gain is small is decreased. The diversity gain $L$ is reflected in the SNR exponent in (3.41); the power gain affects the constant before the $1/\mathsf{SNR}^L$. Note that if the channel gains $h_\ell[1]$ are fully correlated across all branches, then we only get a power gain but no diversity gain as we increase $L$. On the other hand, even when all the $h_\ell$ are independent there is a diminishing marginal return as $L$ increases: due to the law of large numbers, the second term in (3.71),

$$\frac{1}{L}\|\mathbf{h}\|^2 = \frac{1}{L}\sum_{\ell=1}^{L} |h_\ell[1]|^2, \tag{3.72}$$

---

[7] Although mathematically the same situation holds in the time diversity repetition coding case, the increase in received SNR there comes from increasing the total *transmit* energy required to send a single bit; it is therefore not appropriate to call that a power gain.

converges to 1 with increasing $L$ (assuming each of the channel gains is normalized to have unit variance). The power gain, on the other hand, suffers from no such limitation: a 3-dB gain is obtained for every doubling of the number of antennas.[8]

## 3.3.2 Transmit diversity: space-time codes

Now consider the case when there are $L$ transmit antennas and 1 receive antenna, the MISO channel (Figure 3.11(b)). This is common in the downlink of a cellular system since it is often cheaper to have multiple antennas at the base-station than to have multiple antennas at every handset. It is easy to get a diversity gain of $L$: simply transmit the same symbol over the $L$ different antennas during $L$ symbol times. At any one time, only one antenna is turned on and the rest are silent. This is simply a repetition code, and, as we have seen in the previous section, repetition codes are quite wasteful of degrees of freedom. More generally, *any* time diversity code of block length $L$ can be used on this transmit diversity system: simply use one antenna at a time and transmit the coded symbols of the time diversity code successively over the different antennas. This provides a coding gain over the repetition code. One can also design codes specifically for the transmit diversity system. There have been a lot of research activities in this area under the rubric of *space-time coding* and here we discuss the simplest, and yet one of the most elegant, space-time code: the so-called Alamouti scheme. This is the transmit diversity scheme proposed in several third-generation cellular standards. The Alamouti scheme is designed for two transmit antennas; generalization to more than two antennas is possible, to some extent.

### Alamouti scheme

With flat fading, the two transmit, single receive channel is written as

$$y[m] = h_1[m]x_1[m] + h_2[m]x_2[m] + w[m], \qquad (3.73)$$

where $h_i$ is the channel gain from transmit antenna $i$. The Alamouti scheme transmits two complex symbols $u_1$ and $u_2$ over two symbol times: at time 1, $x_1[1] = u_1$, $x_2[1] = u_2$; at time 2, $x_1[2] = -u_2^*$, $x_2[2] = u_1^*$. If we assume that the channel remains constant over the two symbol times and set $h_1 = h_1[1] = h_1[2]$, $h_2 = h_2[1] = h_2[2]$, then we can write in matrix form:

$$\begin{bmatrix} y[1] & y[2] \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \end{bmatrix} \begin{bmatrix} u_1 & -u_2^* \\ u_2 & u_1^* \end{bmatrix} + \begin{bmatrix} w[1] & w[2] \end{bmatrix}. \qquad (3.74)$$

---

[8] This will of course ultimately not hold since the received power cannot be larger than the transmit power, but the number of antennas for our model to break down will have to be humongous.

We are interested in detecting $u_1, u_2$, so we rewrite this equation as

$$\begin{bmatrix} y[1] \\ y[2]^* \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} w[1] \\ w[2]^* \end{bmatrix}. \tag{3.75}$$

We observe that the columns of the square matrix are orthogonal. Hence, the detection problem for $u_1, u_2$ decomposes into two separate, orthogonal, scalar problems. We project **y** onto each of the two columns to obtain the sufficient statistics

$$r_i = \|\mathbf{h}\| u_i + w_i, \qquad i = 1, 2, \tag{3.76}$$

where $\mathbf{h} = [h_1, h_2]^t$ and $w_i \sim \mathcal{CN}(0, N_0)$ and $w_1, w_2$ are independent. Thus, the diversity gain is 2 for the detection of each symbol. Compared to the repetition code, two symbols are now transmitted over two symbol times instead of one symbol, but with half the power in each symbol (assuming that the total transmit power is the same in both cases).

The Alamouti scheme works for any constellation for the symbols $u_1, u_2$, but suppose now they are BPSK symbols, thus conveying a total of two bits over two symbol times. In the repetition scheme, we need to use 4-PAM symbols to achieve the same data rate. To achieve the same minimum distance as the BPSK symbols in the Alamouti scheme, we need five times the energy per symbol. Taking into account the factor of 2 energy saving since we are only transmitting one symbol at a time in the repetition scheme, we see that the repetition scheme requires a factor of 2.5 (4 dB) more power than the Alamouti scheme. Again, the repetition scheme suffers from an inefficient utilization of the available degrees of freedom in the channel: over the two symbol times, bits are packed into only one dimension of the received signal space, namely along the direction $[h_1, h_2]^t$. In contrast, the Alamouti scheme spreads the information onto two dimensions – along the orthogonal directions $[h_1, h_2^*]^t$ and $[h_2, -h_1^*]^t$.

### The determinant criterion for space-time code design

In Section 3.2, we saw that a good code exploiting time diversity should maximize the minimum product distance between codewords. Is there an analogous notion for space-time codes? To answer this question, let us think of a space-time code as a set of complex codewords $\{\mathbf{X}_i\}$, where each $\mathbf{X}_i$ is an $L$ by $N$ matrix. Here, $L$ is the number of transmit antennas and $N$ is the block length of the code. For example, in the Alamouti scheme, each codeword is of the form

$$\begin{bmatrix} u_1 & -u_2^* \\ u_2 & u_1^* \end{bmatrix}, \tag{3.77}$$

with $L = 2$ and $N = 2$. In contrast, each codeword in the repetition scheme is of the form

$$\begin{bmatrix} u & 0 \\ 0 & u \end{bmatrix}. \tag{3.78}$$

More generally, any block length $L$ time diversity code with codewords $\{\mathbf{x}_i\}$ translates into a block length $L$ transmit diversity code with codeword matrices $\{\mathbf{X}_i\}$, where

$$\mathbf{X}_i = \text{diag}\{x_{i1}, \ldots, x_{iL}\}. \tag{3.79}$$

For convenience, we normalize the codewords so that the average energy per symbol time is 1, hence $\mathsf{SNR} = 1/N_0$. Assuming that the channel remains constant for $N$ symbol times, we can write

$$\mathbf{y}^t = \mathbf{h}^*\mathbf{X} + \mathbf{w}^t, \tag{3.80}$$

where

$$\mathbf{y} := \begin{bmatrix} y[1] \\ \vdots \\ y[N] \end{bmatrix}, \qquad \mathbf{h} := \begin{bmatrix} h_1^* \\ \vdots \\ h_L^* \end{bmatrix}, \qquad \mathbf{w} := \begin{bmatrix} w[1] \\ \vdots \\ w[N] \end{bmatrix}. \tag{3.81}$$

To bound the error probability, consider the pairwise error probability of confusing $\mathbf{X}_B$ with $\mathbf{X}_A$, when $\mathbf{X}_A$ is transmitted. Conditioned on the fading gains $\mathbf{h}$, we have the familiar vector Gaussian detection problem (see Summary A.2): here we are deciding between the vectors $\mathbf{h}^*\mathbf{X}_A$ and $\mathbf{h}^*\mathbf{X}_B$ under additive circular symmetric white Gaussian noise. A sufficient statistic is $\Re\{\mathbf{v}^*\mathbf{y}\}$, where $\mathbf{v} := \mathbf{h}^*(\mathbf{X}_A - \mathbf{X}_B)$. The conditional pairwise error probability is

$$\mathbb{P}\{\mathbf{X}_A \to \mathbf{X}_B \,|\, \mathbf{h}\} = Q\left(\frac{\|\mathbf{h}^*(\mathbf{X}_A - \mathbf{X}_B)\|}{2\sqrt{N_0/2}}\right). \tag{3.82}$$

Hence, the pairwise error probability averaged over the channel statistics is

$$\mathbb{P}\{\mathbf{X}_A \to \mathbf{X}_B\} = \mathbb{E}\left[Q\left(\sqrt{\frac{\mathsf{SNR}\,\mathbf{h}^*(\mathbf{X}_A - \mathbf{X}_B)(\mathbf{X}_A - \mathbf{X}_B)^*\mathbf{h}}{2}}\right)\right]. \tag{3.83}$$

The matrix $(\mathbf{X}_A - \mathbf{X}_B)(\mathbf{X}_A - \mathbf{X}_B)^*$ is Hermitian[9] and is thus diagonalizable by a unitary transformation, i.e., we can write $(\mathbf{X}_A - \mathbf{X}_B)(\mathbf{X}_A - \mathbf{X}_B)^* = \mathbf{U}\Lambda\mathbf{U}^*$,

---

[9] A complex square matrix $\mathbf{X}$ is Hermitian if $\mathbf{X}^* = \mathbf{X}$.

where $\mathbf{U}$ is unitary[10] and $\Lambda = \text{diag}\{\lambda_1^2, \ldots, \lambda_L^2\}$. Here $\lambda_\ell$ are the *singular values* of the codeword difference matrix $\mathbf{X}_A - \mathbf{X}_B$. Therefore, we can rewrite the pairwise error probability as

$$\mathbb{P}\{\mathbf{X}_A \to \mathbf{X}_B\} = \mathbb{E}\left[ Q\left( \sqrt{\frac{\text{SNR} \sum_{\ell=1}^{L} |\tilde{h}_\ell|^2 \lambda_\ell^2}{2}} \right) \right], \qquad (3.84)$$

where $\tilde{\mathbf{h}} := \mathbf{U}^*\mathbf{h}$. In the Rayleigh fading model, the fading coefficients $h_\ell$ are i.i.d. $\mathcal{CN}(0, 1)$ and then $\tilde{\mathbf{h}}$ has the same distribution as $\mathbf{h}$ (cf. (A.22) in Appendix A). Thus we can bound the average pairwise error probability, as in (3.54),

$$\mathbb{P}\{\mathbf{X}_A \to \mathbf{X}_B\} \leq \prod_{\ell=1}^{L} \frac{1}{1 + \text{SNR}\,\lambda_\ell^2/4}. \qquad (3.85)$$

If all the $\lambda_\ell^2$ are strictly positive for all the codeword differences, then the maximal diversity gain of $L$ is achieved. Since the number of positive eigenvalues $\lambda_\ell^2$ equals the rank of the codeword difference matrix, this is possible only if $N \geq L$. If indeed all the $\lambda_\ell^2$ are positive, then,

$$\begin{aligned}\mathbb{P}\{\mathbf{X}_A \to \mathbf{X}_B\} &\leq \frac{4^L}{\text{SNR}^L \prod_{\ell=1}^{L} \lambda_\ell^2} \\ &= \frac{4^L}{\text{SNR}^L \det[(\mathbf{X}_A - \mathbf{X}_B)(\mathbf{X}_A - \mathbf{X}_B)^*]},\end{aligned} \qquad (3.86)$$

and a diversity gain of $L$ is achieved. The coding gain is determined by the minimum of the determinant $\det[(\mathbf{X}_A - \mathbf{X}_B)(\mathbf{X}_A - \mathbf{X}_B)^*]$ over all codeword pairs. This is sometimes called the *determinant criterion*.

In the special case when the transmit diversity code comes from a time diversity code, the space-time code matrices are diagonal (cf. (3.79)), and $\lambda_\ell = |d_\ell|^2$, the squared magnitude of the component difference between the corresponding time diversity codewords. The determinant criterion then coincides with the squared product distance criterion (3.68) we already derived for time diversity codes.

We can compare the coding gains obtained by the Alamouti scheme with the repetition scheme. That is, how much less power does the Alamouti scheme consume to achieve the same error probability as the repetition scheme? For the Alamouti scheme with BPSK symbols $u_i$, the minimum determinant is 4. For the repetition scheme with 4-PAM symbols, the minimum determinant is 16/25. (Verify!) This translates into the Alamouti scheme having a coding

---

[10] A complex square matrix $\mathbf{U}$ is unitary if $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I}$.

gain of roughly a factor of 6 over the repetition scheme, consistent with the analysis above.

The Alamouti transmit diversity scheme has a particularly simple receiver structure. Essentially, a linear receiver allows us to decouple the two symbols sent over the two transmit antennas in two time slots. Effectively, both symbols pass through non-interfering *parallel* channels, both of which afford a diversity of order 2. In Exercise 3.16, we derive some properties that a code construction must satisfy to mimic this behavior for more than two transmit antennas.

### 3.3.3 MIMO: a 2 × 2 example

Degrees of freedom

Consider now a MIMO channel with two transmit and two receive antennas (Figure 3.11(c)). Let $h_{ij}$ be the Rayleigh distributed channel gain from transmit antenna $j$ to receive antenna $i$. Suppose both the transmit antennas and the receive antennas are spaced sufficiently far apart that the fading gains, $h_{ij}$, can be assumed to be independent. There are four independently faded signal paths between the transmitter and the receiver, suggesting that the maximum diversity gain that can be achieved is 4. The same repetition scheme described in the last section can achieve this performance: transmit the same symbol over the two antennas in two consecutive symbol times (at each time, nothing is sent over the other antenna). If the transmitted symbol is $x$, the received symbols at the two receive antennas are

$$y_i[1] = h_{i1}x + w_i[1], \qquad i = 1, 2 \tag{3.87}$$

at time 1, and

$$y_i[2] = h_{i2}x + w_i[2], \qquad i = 1, 2 \tag{3.88}$$

at time 2. By performing maximal-ratio combining of the four received symbols, an effective channel with gain $\sum_{i=1}^{2} \sum_{j=1}^{2} |h_{ij}|^2$ is created, yielding a four-fold diversity gain.

However, just as in the case of the $2 \times 1$ channel, the repetition scheme utilizes the degrees of freedom in the channel poorly; it only transmits one data symbol per two symbol times. In this regard, the Alamouti scheme performs better by transmitting two data symbols over two symbol times. Exercise 3.20 shows that the Alamouti scheme used over the $2 \times 2$ channel provides effectively two independent channels, analogous to (3.76), but with the gain in each channel equal to $\sum_{i=1}^{2} \sum_{j=1}^{2} |h_{ij}|^2$. Thus, *both* the data symbols see a diversity gain of 4, the same as that offered by the repetition scheme.

But does the Alamouti scheme utilize *all* the available degrees of freedom in the $2 \times 2$ channel? How many degrees of freedom does the $2 \times 2$ channel have anyway?

In Section 2.2.3 we have defined the degrees of freedom of a channel as the dimension of the received signal space. In a channel with two transmit and a single receive antenna, this is equal to *one* for every symbol time. The repetition scheme utilizes only half a degree of freedom per symbol time, while the Alamouti scheme utilizes all of it.

With $L$ receive, but a single transmit antenna, the received signal lies in an $L$-dimensional vector space, but it does not span the full space. To see this explicitly, consider the channel model from (3.69) (suppressing the symbol time index $m$):

$$\mathbf{y} = \mathbf{h}x + \mathbf{w}, \qquad (3.89)$$

where $\mathbf{y} := [y_1, \ldots, y_L]^t$, $\mathbf{h} = [h_1, \ldots, h_L]^t$ and $\mathbf{w} = [w_1, \ldots, w_L]^t$. The signal of interest, $\mathbf{h}x$, lies in a one-dimensional space.[11] Thus, we conclude that the degrees of freedom of a multiple receive, single transmit antenna channel is still 1 per symbol time.

But in a $2 \times 2$ channel, there are potentially *two* degrees of freedom per symbol time. To see this, we can write the channel as

$$\mathbf{y} = \mathbf{h}_1 x_1 + \mathbf{h}_2 x_2 + \mathbf{w}, \qquad (3.90)$$

where $x_j$ and $\mathbf{h}_j$ are the transmitted symbol and the vector of channel gains from transmit antenna $j$ respectively, and $\mathbf{y} = [y_1, y_2]^t$ and $\mathbf{w} = [w_1, w_2]^t$ are the vectors of received signals and $\mathcal{CN}(0, N_0)$ noise respectively. As long as $\mathbf{h}_1$ and $\mathbf{h}_2$ are linearly independent, the signal space dimension is 2: the signal from transmit antenna $j$ arrives in its own direction $\mathbf{h}_j$, and with two receive antennas, the receiver can distinguish between the two signals. Compared to a $2 \times 1$ channel, there is an additional degree of freedom coming from *space*. Figure 3.12 summarizes the situation.



**Figure 3.12** (a) In the $1 \times 2$ channel, the signal space is one-dimensional, spanned by $\mathbf{h}$. (b) In the $2 \times 2$ channel, the signal space is two-dimensional, spanned by $\mathbf{h}_1$ and $\mathbf{h}_2$.
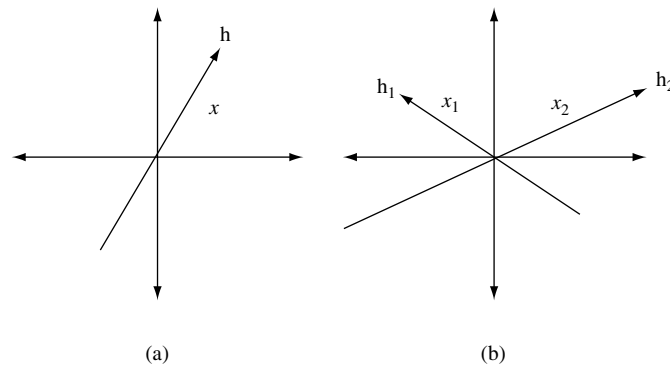
[11] This is why the *scalar* $(\mathbf{h}^*/\|\mathbf{h}\|)\mathbf{y}$ is a sufficient statistic to detect $x$ (cf. (3.33)).

## Spatial multiplexing

Now we see that neither the repetition scheme nor the Alamouti utilizes all the degrees of freedom in a $2 \times 2$ channel. A very simple scheme that does is the following: transmit independent uncoded symbols over the different antennas as well as over the different symbol times. This is an example of a *spatial multiplexing* scheme: independent data streams are multiplexed in space. (It is also called V-BLAST in the literature.) To analyze the performance of this scheme, we extend the derivation of the pairwise error probability bound (3.85) from a single receive antenna to multiple receive antennas. Exercise 3.19 shows that with $n_r$ receive antennas, the corresponding bound on the probability of confusing codeword $\mathbf{X}_B$ with codeword $\mathbf{X}_A$ is

$$\mathbb{P}\{\mathbf{X}_A \to \mathbf{X}_B\} \leq \left[ \prod_{\ell=1}^{L} \frac{1}{1 + \mathsf{SNR}\, \lambda_\ell^2/4} \right]^{n_r}. \tag{3.91}$$

where $\lambda_\ell$ are the singular values of the codeword difference $\mathbf{X}_A - \mathbf{X}_B$. This bound holds for space-time codes of general block lengths. Our specific scheme does not code across time and is thus "space-only". The block length is 1, the codewords are two-dimensional vectors $\mathbf{x}_1, \mathbf{x}_2$ and the bound simplifies to

$$\begin{aligned}
\mathbb{P}\{\mathbf{x}_1 \to \mathbf{x}_2\} &\leq \left[ \frac{1}{1 + \mathsf{SNR}\, \|\mathbf{x}_1 - \mathbf{x}_2\|^2/4} \right]^2 \\
&\leq \frac{16}{\mathsf{SNR}^2 \, \|\mathbf{x}_1 - \mathbf{x}_2\|^4}.
\end{aligned} \tag{3.92}$$

The exponent of the SNR factor is the diversity gain: the spatial multiplexing scheme achieves a diversity gain of 2. Since there is no coding across the transmit antennas, it is clear that no transmit diversity can be exploited; thus the diversity comes entirely from the dual receive antennas. The factor $\|\mathbf{x}_1 - \mathbf{x}_2\|^4$ plays a role analogous to the determinant $\det[(\mathbf{X}_A - \mathbf{X}_B)(\mathbf{X}_A - \mathbf{X}_B)^*]$ in determining the coding gain (cf. (3.86)).

Compared to the Alamouti scheme, we see that V-BLAST has a smaller diversity gain (2 compared to 4). On the other hand, the full use of the spatial degrees of freedom should allow a more efficient packing of bits, resulting in a better coding gain. To see this concretely, suppose we use BPSK symbols in the spatial multiplexing scheme to deliver 2 bits/s/Hz. Assuming that the average transmit energy per symbol time is normalized to be 1 as before, we can use (3.92) to explicitly calculate a bound on the *worst-case* pairwise error probability:

$$\max_{i \neq j} \mathbb{P}\{\mathbf{x}_i \to \mathbf{x}_j\} \leq 4 \cdot \mathsf{SNR}^{-2}. \tag{3.93}$$

On the other hand, the corresponding bound for the Alamouti scheme using 4-PAM symbols to deliver the same 2 bits/s/Hz can be calculated from (3.86) to be

$$\max_{i \neq j} \mathbb{P}\{\mathbf{x}_i \to \mathbf{x}_j\} \leq 1600 \cdot \mathsf{SNR}^{-4}. \tag{3.94}$$

We see that indeed the bound for the Alamouti scheme has a much poorer constant before the factor that decays with SNR.

We can draw two lessons from the V-BLAST scheme. First, we see a new role for multiple antennas: in addition to diversity, they can also provide additional degrees of freedom for communication. This is in a sense a more powerful view of multiple antennas, one that will be further explored in Chapter 7. Second, the scheme also reveals limitations in our performance analysis framework for space-time codes. In the earlier sections, our approach has always been to seek schemes which extract the maximum diversity from the channel and then compare them on the basis of the coding gain, which is a function of how efficiently the schemes utilize the available degrees of freedom. This approach falls short in comparing V-BLAST and the Alamouti scheme for the $2 \times 2$ channel: V-BLAST has poorer diversity than the Alamouti scheme but is more efficient in exploiting the spatial degrees of freedom, resulting in a better coding gain. A more powerful framework combining the two performance measures into a unified metric is needed; this is one of the main subjects of Chapter 9. There we will also address the issue of what it means by an *optimal* scheme and whether it is possible to find a scheme which achieves the full diversity *and* the full degrees of freedom of the channel.

### Low-complexity detection: the decorrelator

One advantage of the Alamouti scheme is its low-complexity ML receiver: the decoding decouples into two orthogonal single-symbol detection problems. ML detection of V-BLAST does not enjoy the same advantage: joint detection of the two symbols is required. The complexity grows exponentially with the number of antennas. A natural question to ask is: what performance can suboptimal single-symbol detectors achieve? We will study MIMO receiver architectures in depth in Chapters 7 and 9, but here we will give an example of a simple detector, the *decorrelator*, and analyze its performance in the $2 \times 2$ channel.

To motivate the definition of this detector, let us rewrite the channel (3.90) in matrix form:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{3.95}$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2]$ is the channel matrix. The input $\mathbf{x} := [x_1, x_2]^t$ is composed of two independent symbols $x_1, x_2$. To decouple the detection of the two symbols, one idea is to invert the effect of the channel:

$$\tilde{\mathbf{y}} = \mathbf{H}^{-1}\mathbf{y} = \mathbf{x} + \mathbf{H}^{-1}\mathbf{w} = \mathbf{x} + \tilde{\mathbf{w}} \tag{3.96}$$

and detect each of the symbols separately. This is in general suboptimal compared to joint ML detection, since the noise samples $\tilde{w}_1$ and $\tilde{w}_2$ are correlated. How much performance do we lose?

Let us focus on the detection of the symbol $x_1$ from transmit antenna 1. By direct computation, the variance of the noise $\tilde{w}_1$ is

$$\frac{|h_{22}|^2 + |h_{21}|^2}{|h_{11}h_{22} - h_{21}h_{12}|^2} N_0. \tag{3.97}$$

Hence, we can rewrite the first component of the vector equation in (3.96) as

$$\tilde{y}_1 = x_1 + \frac{\sqrt{|h_{22}|^2 + |h_{21}|^2}}{|h_{11}h_{22} - h_{21}h_{12}|} z_1, \tag{3.98}$$

where $z_1 \sim \mathcal{CN}(0, N_0)$, the scaled version of $\tilde{w}_1$, is independent of $x_1$. Equivalently, the scaled output can be written as

$$\begin{aligned} y_1' &:= \frac{h_{11}h_{22} - h_{21}h_{12}}{\sqrt{|h_{22}|^2 + |h_{21}|^2}} \tilde{y}_1 \\ &= (\phi_2^* \mathbf{h}_1) x_1 + z_1, \end{aligned} \tag{3.99}$$

where

$$\mathbf{h}_i := \begin{bmatrix} h_{i1} \\ h_{i2} \end{bmatrix}, \qquad \phi_i := \frac{1}{\sqrt{|h_{i2}|^2 + |h_{i1}|^2}} \begin{bmatrix} h_{i2}^* \\ -h_{i1}^* \end{bmatrix}, \qquad i = 1, 2. \tag{3.100}$$

Geometrically, one can interpret $\mathbf{h}_j$ as the "direction" of the signal from transmit antenna $j$ and $\phi_j$ as the direction orthogonal to $\mathbf{h}_j$. Equation (3.99) says that when demodulating the symbol from antenna 1, channel inversion eliminates the interference from transmit antenna 2 by projecting the received signal $\mathbf{y}$ in the direction orthogonal to $\mathbf{h}_2$ (Figure 3.13). The signal part is $(\phi_2^* \mathbf{h}_1) x_1$. The scalar gain $\phi_2^* \mathbf{h}_1$ is circular symmetric Gaussian, being the projection of a two-dimensional i.i.d. circular symmetric Gaussian random vector ($\mathbf{h}_1$) onto an independent unit vector ($\phi_2$) (cf. (A.22) in Appendix A). The scalar channel (3.99) is therefore Rayleigh faded like a $1 \times 1$ channel and has only unit diversity. Note that if there were no interference from antenna 2, the diversity gain would have been 2: the norm $\|\mathbf{h}_1\|^2$ of the entire vector $\mathbf{h}_1$ has to be small for poor reception of $x_1$. However, here, the component of $\mathbf{h}_1$ perpendicular to $\mathbf{h}_2$ being small already wreaks havoc; this is the price paid for nulling out the interference from antenna 2. In contrast, the ML detector, by jointly detecting the two symbols, retains the diversity gain of 2.

We have discussed V-BLAST in the context of a point-to-point link with two transmit antennas. But since there is no coding across the antennas, we can equally think of the two transmit antennas as two distinct users each with a single antenna. In the multiuser context, the receiver described above is sometimes called the *interference nuller*, *zero-forcing receiver* or

**Figure 3.13** Demodulation of $x_1$: the received vector **y** is projected onto the direction $\phi_2$ orthogonal to $\mathbf{h}_2$. The effective channel for $x_1$ is in deep fade whenever the projection of $\mathbf{h}_1$ onto $\phi_2$ is small.
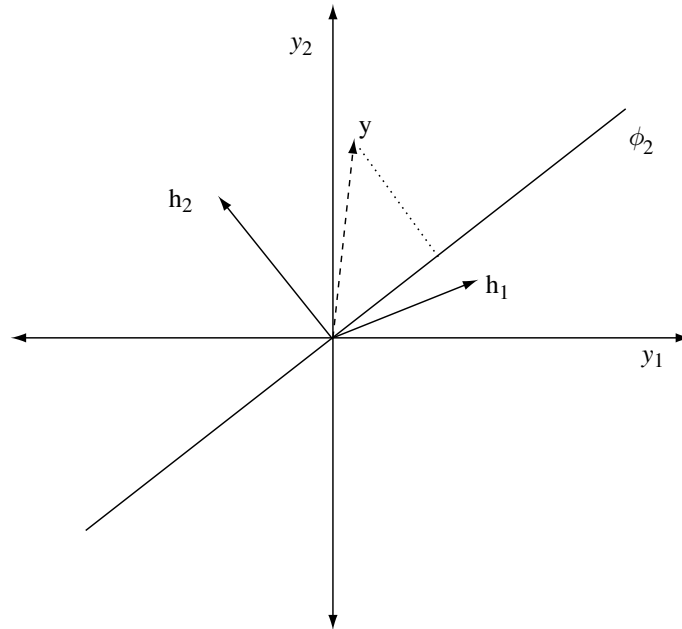


the *decorrelator*. It nulls out the effect of the other user (interferer) while demodulating the symbol of one user. Using this receiver, we see that dual receive antennas can perform one of two functions in a wireless system: they can *either* provide a two-fold diversity gain in a point-to-point link when there is no interference, *or* they can be used to null out the effect of an interfering user but provide no diversity gain more than 1. But they cannot do *both*. This is however not an intrinsic limitation of the *channel* but rather a limitation of the decorrelator; by performing joint ML detection instead, the two users can in fact be simultaneously supported with a two-fold diversity gain each.

## Summary 3.2  $2 \times 2$ MIMO schemes

The performance of the various schemes for the $2 \times 2$ channel is summarized below.

|                  | Diversity gain | Degrees of freedom utilized per symbol time |
|------------------|:--------------:|:-------------------------------------------:|
| Repetition       | 4              | 1/2                                         |
| Alamouti         | 4              | 1                                           |
| V-BLAST (ML)     | 2              | 2                                           |
| V-BLAST (nulling)| 1              | 2                                           |
| Channel itself   | 4              | 2                                           |

## 3.4 Frequency diversity

### 3.4.1 Basic concept

So far we have focused on narrowband flat fading channels. These channels are modeled by a single-tap filter, as most of the multipaths arrive during one symbol time. In wideband channels, however, the transmitted signal arrives over multiple symbol times and the multipaths can be resolved at the receiver. The frequency response is no longer flat, i.e., the transmission bandwidth $W$ is greater than the coherence bandwidth $W_c$ of the channel. This provides another form of diversity: frequency.

We begin with the discrete-time baseband model of the wireless channel in Section 2.2. Recalling (2.35) and (2.38), the sampled output $y[m]$ can be written as

$$y[m] = \sum_{\ell} h_\ell[m]x[m-\ell] + w[m]. \qquad (3.101)$$

Here $h_\ell[m]$ denotes the $\ell$th channel filter tap at time $m$. To understand the concept of frequency diversity in the simplest setting, consider first the one-shot communication situation when one symbol $x[0]$ is sent at time 0, and no symbols are transmitted after that. The receiver observes

$$y[\ell] = h_\ell[\ell]x[0] + w[\ell], \qquad \ell = 0, 1, 2, \ldots \qquad (3.102)$$

If we assume that the channel response has a finite number of taps $L$, then the delayed replicas of the signal are providing $L$ branches of diversity in detecting $x[0]$, since the tap gains $h_\ell[\ell]$ are assumed to be independent. This diversity is achieved by the ability of resolving the multipaths at the receiver due to the wideband nature of the channel, and is thus called *frequency diversity*.

A simple communication scheme can be built on the above idea by sending an information symbol every $L$ symbol times. The maximal diversity gain of $L$ can be achieved, but the problem with this scheme is that it is very wasteful of degrees of freedom: only one symbol can be transmitted every delay spread. This scheme can actually be thought of as analogous to the repetition codes used for both time and spatial diversity, where one information symbol is repeated $L$ times. In this setting, once one tries to transmit symbols more frequently, *inter-symbol interference* (ISI) occurs: the delayed replicas of previous symbols interfere with the current symbol. The problem is then how to deal with the ISI while at the same time exploiting the inherent frequency diversity in the channel. Broadly speaking, there are three common approaches:

- **Single-carrier systems with equalization** By using linear and non-linear processing at the receiver, ISI can be mitigated to some extent. Optimal ML detection of the transmitted symbols can be implemented using the Viterbi algorithm. However, the complexity of the Viterbi algorithm grows

exponentially with the number of taps, and it is typically used only when the number of significant taps is small. Alternatively, linear equalizers attempt to detect the current symbol while linearly suppressing the interference from the other symbols, and they have lower complexity.

- **Direct-sequence spread-spectrum** In this method, information symbols are modulated by a pseudonoise sequence and transmitted over a bandwidth $W$ much larger than the data rate. Because the symbol rate is very low, ISI is small, simplifying the receiver structure significantly. Although this leads to an inefficient utilization of the total degrees of freedom in the system from the perspective of one user, this scheme allows multiple users to share the total degrees of freedom, with users appearing as pseudonoise to each other.

- **Multi-carrier systems** Here, transmit precoding is performed to convert the ISI channel into a set of non-interfering, orthogonal sub-carriers, each experiencing narrowband flat fading. Diversity can be obtained by coding across the symbols in different sub-carriers. This method is also called Discrete Multi-Tone (DMT) or Orthogonal Frequency Division Multiplexing (OFDM). Frequency-hop spread-spectrum can be viewed as a special case where one carrier is used at a time.

For example, GSM is a single-carrier system, IS-95 CDMA and IEEE 802.11b (a wireless LAN standard) are based on direct-sequence spread-spectrum, and IEEE 802.11a is a multi-carrier system,

Below we study these three approaches in turn. An important conceptual point is that, while frequency diversity is something *intrinsic* in a wideband channel, the presence of ISI is not, as it depends on the modulation technique used. For example, under OFDM, there is no ISI, but sub-carriers that are separated by more than the coherence bandwidth fade more or less independently and hence frequency diversity is still present.

Narrowband systems typically operate in a relatively high SNR regime. In contrast, the energy is spread across many degrees of freedom in many wideband systems, and the impact of the channel uncertainty on the ability of the receiver to extract the inherent diversity in frequency-selective channels becomes more pronounced. This point will be discussed in Section 3.5, but in the present section, we assume that the receiver has a perfect estimate of the channel.

### 3.4.2 Single-carrier with ISI equalization

Single-carrier with ISI equalization is the classic approach to communication over frequency-selective channels, and has been used in wireless as well as wireline applications such as voiceband modems. Much work has been done in this area but here we focus on the diversity aspects.

Starting at time 1, a sequence of *uncoded* independent symbols $x[1], x[2], \ldots$ is transmitted over the frequency-selective channel (3.101).

Assuming that the channel taps do not vary over these $N$ symbol times, the received symbol at time $m$ is

$$y[m] = \sum_{\ell=0}^{L-1} h_\ell x[m-\ell] + w[m], \qquad (3.103)$$

where $x[m] = 0$ for $m < 1$. For simplicity, we assume here that the taps $h_\ell$ are i.i.d. Rayleigh with equal variance $1/L$, but the discussion below holds more generally (see Exercise 3.25).

We want to detect each of the transmitted symbols from the received signal. The process of extracting the symbols from the received signal is called *equalization*. In contrast to the simple scheme in the previous section where a symbol is sent every $L$ symbol times, here a symbol is sent *every* symbol time and hence there is significant ISI. Can we still get the maximum diversity gain of $L$?

### Frequency-selective channel viewed as a MISO channel

To analyze this problem, it is insightful to transform the frequency-selective channel into a *flat* fading *MISO* channel with $L$ transmit antennas and a single receive antenna and channel gains $h_0, \ldots, h_{L-1}$. Consider the following transmission scheme on the MISO channel: at time 1, the symbol $x[1]$ is transmitted on antenna 1 and the other antennas are silent. At time 2, $x[1]$ is transmitted at antenna 2, $x[2]$ is transmitted on antenna 1 and the other antennas remain silent. At time $m$, $x[m-\ell]$ is transmitted on antenna $\ell+1$, for $\ell = 0, \ldots, L-1$. See Figure 3.14. The received symbol at time $m$ in this MISO channel is precisely the same as that in the frequency-selective channel under consideration.

Once we transform the frequency-selective channel into a MISO channel, we can exploit the machinery developed in Section 3.3.2. First, it is clear that if we want to achieve full diversity on a symbol, say $x[N]$, we need to observe the received symbols up to time $N + L - 1$. Over these symbol times, we can write the system in matrix form (as in (3.80)):
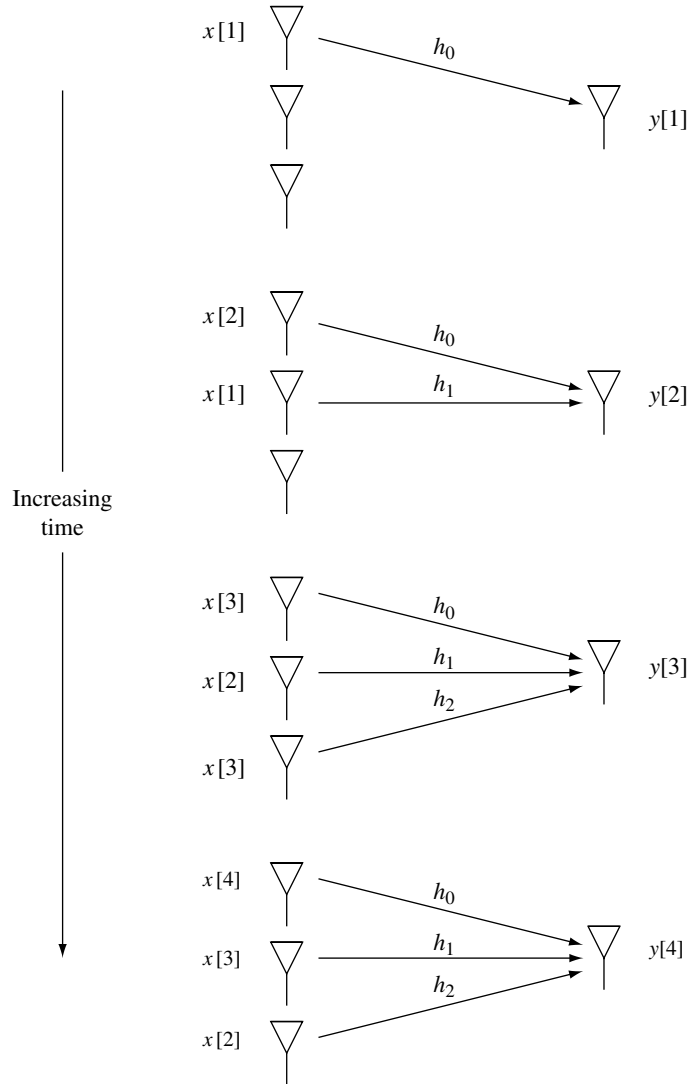
$$\mathbf{y}^t = \mathbf{h}^* \mathbf{X} + \mathbf{w}^t, \qquad (3.104)$$

where $\mathbf{y}^t := [y[1], \ldots, y[N+L-1]]$, $\mathbf{h}^* := [h_0, \ldots, h_{L-1}]$, $\mathbf{w}^t := [w[1], \ldots w[N+L-1]]$ and the $L$ by $N+L-1$ space-time code matrix

$$\mathbf{X} = \begin{bmatrix} x[1] & x[2] & \cdot & \cdot & \cdot & x[N] & \cdot & \cdot & x[N+L-1] \\ 0 & x[1] & x[2] & \cdot & \cdot & \cdot & x[N] & \cdot & x[N+L-2] \\ 0 & 0 & x[1] & x[2] & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & x[1] & x[2] & \cdot & \cdot & x[N] \end{bmatrix} \qquad (3.105)$$

corresponds to the transmitted sequence $\mathbf{x} = [x[1], \ldots, x[N+L-1]]^t$.

**Figure 3.14** The MISO scenario equivalent to the frequency- selective channel.

### Error probability analysis

Consider the maximum likelihood detection of the *sequence* **x** based on the received vector **y** (MLSD). With MLSD, the pairwise error probability of confusing $\mathbf{x}_A$ with $\mathbf{x}_B$, when $\mathbf{x}_A$ is transmitted is, as in (3.85),

$$\mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_B\} \le \prod_{\ell=1}^{L} \frac{1}{1 + \mathsf{SNR}\lambda_\ell^2/4}, \qquad (3.106)$$

where $\lambda_\ell^2$ are the eigenvalues of the matrix $(\mathbf{X}_A - \mathbf{X}_B)(\mathbf{X}_A - \mathbf{X}_B)^*$ and $\mathsf{SNR}$ is the total received SNR per received symbol (summing over all paths). This

error probability decays like $\mathsf{SNR}^{-L}$ whenever the difference matrix $\mathbf{X}_A - \mathbf{X}_B$ is of rank $L$.

By a union bound argument, the probability of detecting the particular symbol $x[N]$ incorrectly is bounded by

$$\sum_{\mathbf{x}_B : x_B[N] \neq x_A[N]} \mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_B\}, \tag{3.107}$$

summing over all the transmitted vectors $\mathbf{x}_B$ which differ with $\mathbf{x}_A$ in the $N$th symbol.[12] To get full diversity, the difference matrix $\mathbf{X}_A - \mathbf{X}_B$ must be full rank for every such vector $\mathbf{x}_B$ (cf. (3.86)). Suppose $m^*$ is the symbol time in which the vectors $\mathbf{x}_A$ and $\mathbf{x}_B$ *first* differ. Since they differ at least once within the first $N$ symbol times, $m^* \leq N$ and the difference matrix is of the form

$$\mathbf{X}_A - \mathbf{X}_B = \begin{bmatrix} 0 \cdot 0 \; x_A[m^*] - x_B[m^*] & \cdot & \cdot & \cdot & \cdot \\ 0 \cdot \cdot & 0 & x_A[m^*] - x_B[m^*] \cdot & \cdot & \cdot \\ 0 \cdot \cdot & \cdot & 0 & \cdot & \cdot & \cdot \\ \cdot \cdot \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 \cdot \cdot & \cdot & \cdot & 0 \; x_A[m^*] - x_B[m^*] \cdot \end{bmatrix}. \tag{3.108}$$

By inspection, all the rows in the difference matrix are linearly independent. Thus $\mathbf{X}_A - \mathbf{X}_B$ is of full rank (i.e., the rank is equal to $L$). We can summarize:

> Uncoded transmission combined with maximum likelihood sequence detection achieves full diversity on symbol $x[N]$ using the observations up to time $N + L - 1$, i.e., a delay of $L - 1$ symbol times.

Compared to the scheme in which a symbol is transmitted every $L$ symbol times, the same diversity gain of $L$ is achieved and yet an independent symbol can be transmitted every symbol time. This translates into a significant "coding gain" (Exercise 3.26).

In the analysis here it was convenient to transform the frequency-selective channel into a MISO channel. However, we can turn the transformation around: if we transmit the space-time code of the form in (3.105) on a MISO channel, then we have converted the MISO channel into a frequency-selective

---

[12] Strictly speaking, the MLSD only minimizes the *sequence* error probability, not the *symbol* error probability. However, this is the standard detector implemented for ISI equalization via the Viterbi algorithm, to be discussed next. In any case, the symbol error probability performance of the MLSD serves as an upper bound to the optimal symbol error performance.

channel. This is the *delay diversity* scheme and it was one of the first proposed transmit diversity schemes for the MISO channel.

## Implementing MLSD: the Viterbi algorithm

Given the received vector **y** of length $n$, MLSD requires solving the optimization problem

$$\max_{\mathbf{x}} \mathbb{P}\{\mathbf{y}|\mathbf{x}\}. \tag{3.109}$$

A brute-force exhaustive search would require a complexity that grows exponentially with the block length $n$. An efficient algorithm needs to exploit the structure of the problem and moreover should be recursive in $n$ so that the problem does not have to be solved from scratch for every symbol time. The solution is the ubiquitous *Viterbi algorithm*.

The key observation is that the memory in the frequency-selective channel can be captured by a finite state machine. At time $m$, define the state (an $L$-dimensional vector)

$$\mathbf{s}[m] := \begin{bmatrix} x[m-L+1] \\ x[m-L+2] \\ \cdot \\ x[m] \end{bmatrix} \tag{3.110}$$

An example of the finite state machine when the $x[m]$ are BPSK symbols is given in Figure 3.15. The number of states is $M^L$, where $M$ is the constellation size for each symbol $x[m]$.
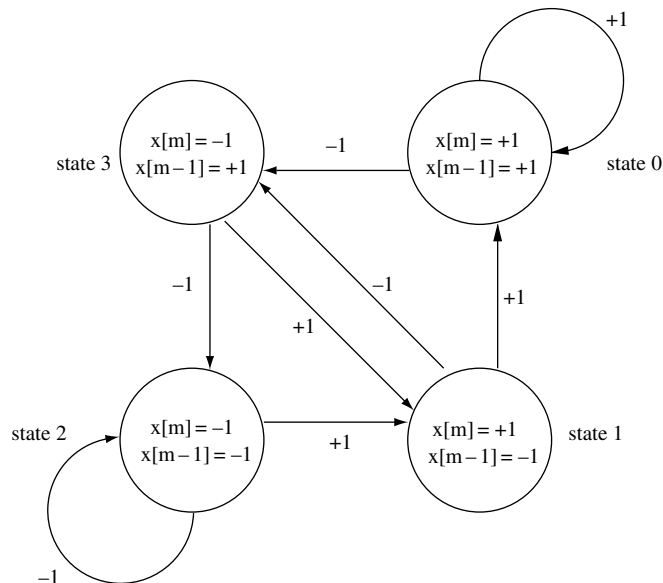


**Figure 3.15** A finite state machine when $x[m]$ are $\pm 1$ BPSK symbols and $L = 2$. There is a total of four states.

The received symbol $y[m]$ is given by

$$y[m] = \mathbf{h}^*\mathbf{s}[m] + w[m], \tag{3.111}$$

with $\mathbf{h}$ representing the frequency-selective channel, as in (3.104). The MLSD problem (3.109) can be rewritten as

$$\min_{\mathbf{s}[1],\ldots,\mathbf{s}[n]} -\log \mathbb{P}\{y[1],\ldots,y[n] \,|\, \mathbf{s}[1],\ldots,\mathbf{s}[n]\}, \tag{3.112}$$

subject to the transition constraints on the state sequence (i.e., the second component of $s[m]$ is the same as the first component of $s[m+1]$). Conditioned on the state sequence $\mathbf{s}[1],\ldots,\mathbf{s}[n]$, the received symbols are independent and the log-likelihood ratio breaks into a sum:

$$\log \mathbb{P}\{y[1],\ldots,y[n] \,|\, \mathbf{s}[1],\ldots,\mathbf{s}[n]\} = \sum_{m=1}^{n} \log \mathbb{P}\{y[m] \,|\, \mathbf{s}[m]\}. \tag{3.113}$$

The optimization problem in (3.112) can be represented as the problem of finding the shortest path through an $n$-stage *trellis*, as shown in Figure 3.16. Each state sequence $(\mathbf{s}[1],\ldots,\mathbf{s}[n])$ is visualized as a path through the trellis, and given the received sequence $y[1],\ldots,y[n]$, the cost associated with the $m$th transition is

$$c_m(\mathbf{s}[m]) := -\log \mathbb{P}\{y[m] \,|\, \mathbf{s}[m]\}. \tag{3.114}$$
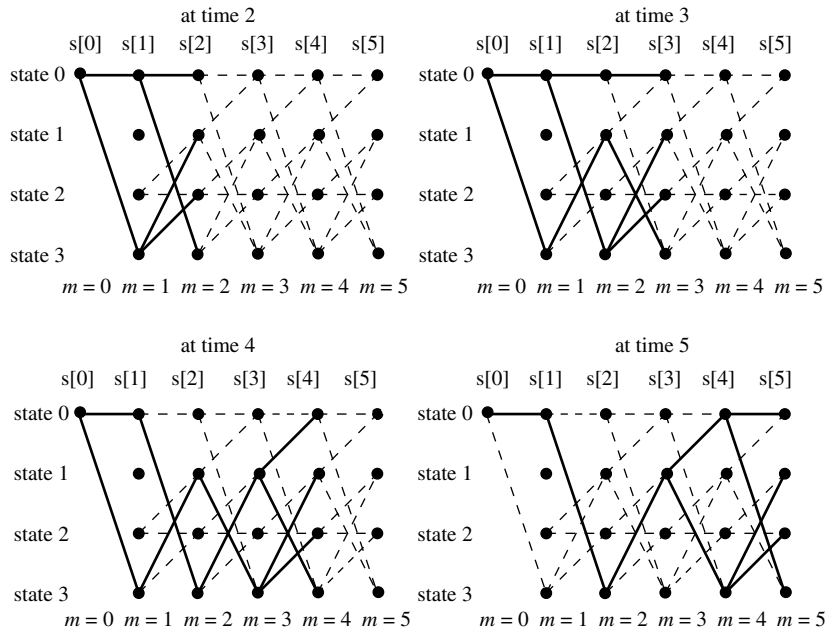


**Figure 3.16** The trellis representation of the channel.

The solution is given recursively by the *optimality principle* of dynamic programming. Let $V_m(\mathbf{s})$ be the cost of the shortest path to a given state $\mathbf{s}$ at stage $m$. Then $V_m(\mathbf{s})$ for all states $\mathbf{s}$ can be computed recursively:

$$V_1(\mathbf{s}) = c_1(\mathbf{s}) \tag{3.115}$$

$$V_m(\mathbf{s}) = \min_{\mathbf{u}}[V_{m-1}(\mathbf{u}) + c_m(\mathbf{s})], \qquad m > 1. \tag{3.116}$$

Here the minimization is over all possible states $\mathbf{u}$, i.e., we only consider the states that the finite state machine can be in at stage $m-1$ and, further, can still end up at state $\mathbf{s}$ at stage $m$. The correctness of this recursion is based on the following intuitive fact: if the shortest path to state $\mathbf{s}$ at stage $m$ goes through the state $\mathbf{u}^*$ at stage $m-1$, then the part of the path up to stage $m-1$ must itself be the shortest path to state $\mathbf{u}^*$. See Figure 3.17. Thus, to compute the shortest path up to stage $m$, it suffices to augment only the shortest paths up to stage $m-1$, and these have already been computed.

Once $V_m(\mathbf{s})$ is computed for all states $\mathbf{s}$, the shortest path to stage $m$ is simply the minimum of these values over all states $\mathbf{s}$. Thus, the optimization problem (3.112) is solved. Moreover, the solution is recursive in $n$.

The complexity of the Viterbi algorithm is linear in the number of stages $n$. Thus, the cost is constant per symbol, a vast improvement over brute-force exhaustive search. However, its complexity is also proportional to the size of the state space, which is $M^L$, where $M$ is the constellation size of each symbol. Thus, while MLSD can be done for channels with a small number of taps, it becomes impractical when $L$ becomes large.

The computational complexity of MLSD leads to an interest in seeking suboptimal equalizers which yield comparable performance. Some candidates are *linear* equalizers (such as the zero-forcing and minimum mean square error (MMSE) equalizers, which involve simple linear operations on the received symbols followed by simple hard decoders), and their decision-feedback versions (DFE), where previously detected symbols are removed from the received signal before linear equalization is performed. We will discuss these equalizers further in Discussion 8.1, where we exploit
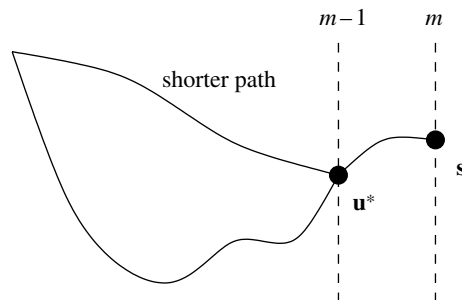


**Figure 3.17** The dynamic programming principle. If the first $m-1$ segments of the shortest path to state $\mathbf{s}$ at stage $m$ were not the shortest path to state $\mathbf{u}^*$ at stage $m-1$, then one could have found an even shorter path to state $\mathbf{s}$.

a correspondence between the MIMO channel and the frequency-selective channel.

### 3.4.3 Direct-sequence spread-spectrum

A common communication system that employs a wide bandwidth is the direct-sequence (DS) spread-spectrum system. Its basic components are shown in Figure 3.18. Information is encoded and modulated by a pseudonoise (PN) sequence and transmitted over a bandwidth $W$. In contrast to the system we analyzed in the last section where an independent symbol is sent at each symbol time, the data rate $R$ bits/s in a spread-spectrum system is typically much smaller than the transmission bandwidth $W$ Hz. The ratio $W/R$ is sometimes called the *processing gain* of the system. For example, IS-95 (CDMA) is a direct-sequence spread-spectrum system. The bandwidth is 1.2288 MHz and a typical data rate (voice) is 9.6 kbits/s, so the processing gain is 128. Thus, very few bits are transmitted per degree of freedom per user. In spread-spectrum jargon, each sample period is called a *chip*, and another way of describing a spread-spectrum system is that the chip rate is much larger than the data rate.

Because the symbol rate per user is very low in a spread-spectrum system, ISI is typically negligible and equalization is not required. Instead, as we will discuss next, a much simpler receiver called the *Rake* receiver can be used to extract frequency diversity. In the cellular setting, multiple spread-spectrum users would share the large bandwidth so that the aggregate bit rate can be high even though the rate of each user is low. The large processing gain of a user serves to mitigate the interference from other users, which appears as random noise. In addition to providing frequency diversity against multipath fading and allowing multiple access, spread-spectrum systems serve other purposes, such as anti-jamming from intentional interferers, and achieving message privacy in the presence of other listeners. We will discuss the multiple access aspects of spread-spectrum systems in Chapter 4. For now, we focus on how DS spread-spectrum systems can achieve frequency diversity.

#### The Rake receiver

Suppose we transmit one of two $n$-chips long pseudonoise sequences $\mathbf{x}_A$ or $\mathbf{x}_B$. Consider the problem of binary detection over a wideband multipath channel. In this context, a binary symbol is transmitted over $n$ chips. The received signal is given by

$$y[m] = \sum_\ell h_\ell[m]x[m-\ell] + w[m]. \qquad (3.117)$$

We assume that $h_\ell[m]$ is non-zero only for $\ell = 0, \ldots, L-1$, i.e., the channel has $L$ taps. One can think of $L/W$ as the delay spread $T_\mathrm{d}$. Also, we assume
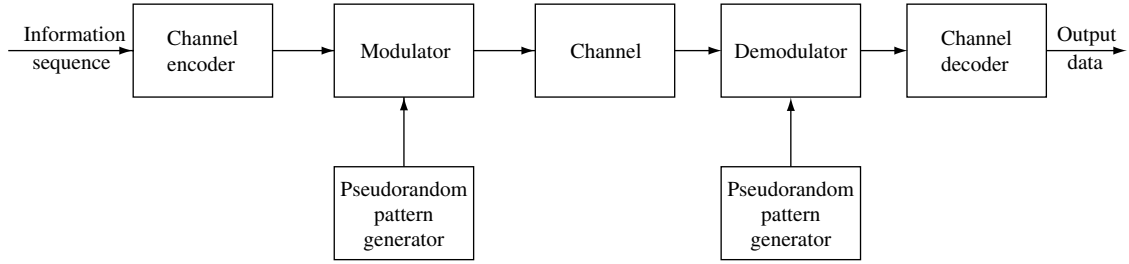
**Figure 3.18** Basic elements of a direct sequence spread-spectrum system.

that $h_\ell[m]$ does not vary with $m$ during the transmission of the sequence, i.e., the channel is considered time-invariant. This holds if $n \ll T_c W$, where $T_c$ is the coherence time of the channel. We also assume that there is negligible interference between consecutive symbols, so that we can consider the binary detection problem in isolation for each symbol. This assumption is valid if $n \gg L$, which is quite common in a spread-spectrum system with high processing gain. Otherwise, ISI between consecutive symbols becomes significant, and an equalizer would be needed to mitigate the ISI. Note however we assume that simultaneously $n \gg T_d W$ and $n \ll T_c W$, which is possible only if $T_d \ll T_c$. In a typical cellular system, $T_d$ is of the order of microseconds and $T_c$ of the order of tens of milliseconds, so this assumption is quite reasonable. (Recall from Chapter 2, Table 2.2 that a channel satisfying this condition is called an *underspread* channel.)

With the above assumptions, the output is just a convolution of the input with the LTI channel plus noise

$$y[m] = (h * x)[m] + w[m], \qquad m = 1, \dots, n + L \qquad (3.118)$$

where $h_\ell$ is the $\ell$th tap of the time-invariant channel filter response, with $h_\ell = 0$ for $\ell < 0$ and $\ell > L - 1$. Assuming the channel $h$ is known to the receiver, two sufficient statistics, $r_A$ and $r_B$, can be obtained by projecting the received vector $\mathbf{y} := [y[1], \dots, y[n+L]]^t$ onto the $n + L$ dimensional vectors $\mathbf{v}_A$ and $\mathbf{v}_B$, where $\mathbf{v}_A := [(h * x_A)[1], \dots, (h * x_A)[n+L]]^t$ and $\mathbf{v}_B := [(h * x_B)[1], \dots, (h * x_B)[n+L]]^t$, i.e.,

$$r_A := \mathbf{v}_A^* \mathbf{y}, \qquad r_B := \mathbf{v}_B^* \mathbf{y}. \qquad (3.119)$$

The computation of $r_A$ and $r_B$ can be implemented by first matched filtering the received signal to $\mathbf{x}_A$ and to $\mathbf{x}_B$. The outputs of the matched filters are passed through a filter matched to the channel response $h$ and then sampled at time $n + L$ (Figure 3.19). This is called the *Rake* receiver. What the Rake actually does is taking inner products of the received signal with shifted versions at the candidate transmitted sequences. Each output is then weighted by the channel tap gains at the appropriate delays and summed. The signal path associated with a particular delay is sometimes called a *finger* of the Rake receiver.

**Figure 3.19** The Rake receiver. Here, $\tilde{\mathbf{h}}$ is the filter matched to $\mathbf{h}$, i.e., $\tilde{h}_\ell = h^*_{-\ell}$. Each tap of $\tilde{\mathbf{h}}$ represents a finger of the Rake.
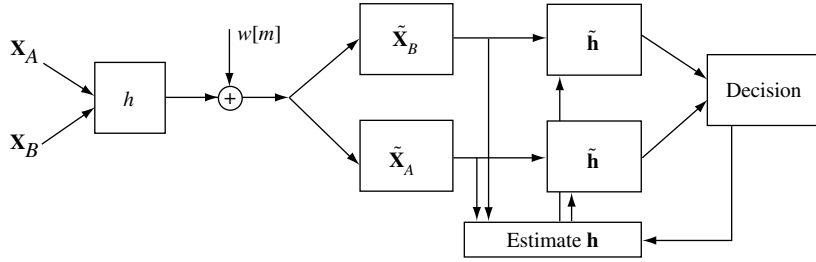


As discussed earlier, we are continuing with the assumption that the channel gains $h_\ell$ are known at the receiver. In practice, these gains have to be estimated and tracked from either a pilot signal or in a decision-directed mode using the previously detected symbols. (The channel estimation problem will be discussed in Section 3.5.2.) Also, due to hardware limitations, the actual number of fingers used in a Rake receiver may be less than the total number of taps $L$ in the range of the delay spread. In this case, there is also a tracking mechanism in which the Rake receiver continuously searches for the strong paths (taps) to assign the limited number of fingers to.

## Performance analysis

Let us now analyze the performance of the Rake receiver. To simplify our notation, we specialize to antipodal modulation (i.e., $\mathbf{x}_A = -\mathbf{x}_B = \mathbf{u}$); the analysis for other modulation schemes is similar. One key aspect of spread-spectrum systems is that the transmitted signal ($\pm\mathbf{u}$) has a *pseudonoise* characteristic. The defining characteristic of a pseudonoise sequence is that its shifted versions are nearly orthogonal to each other. More precisely, if we write $\mathbf{u} = [u[1], \ldots, u[n]]$, and

$$\mathbf{u}^{(\ell)} := [0, \ldots, 0, u[1], \ldots, u[n], 0, \ldots, 0]^t \qquad (3.120)$$

as the $n + L$ dimensional version of $\mathbf{u}$ shifted by $\ell$ chips (hence there are $\ell$ zeros preceding $\mathbf{u}$ and $L - \ell$ zeros following $\mathbf{u}$ above), the pseudonoise property means that for every $\ell = 0, \ldots, L-1$,

$$|(\mathbf{u}^{(\ell)})^*(\mathbf{u}^{(\ell')})| \ll \sum_{i=1}^{n} |u[i]|^2, \qquad \ell \neq \ell'. \qquad (3.121)$$

To simplify the analysis, we assume full orthogonality: $(\mathbf{u}^{(\ell)})^*(\mathbf{u}^{(\ell')}) = 0$ if $\ell \neq \ell'$.

We will now show that the performance of the Rake is the same as that in the diversity model with $L$ branches for repetition coding described in Section 3.2. We can see this by looking at a set of sufficient statistics for the

detection problem different from the ones we used earlier. First, we rewrite
the channel model in vector form

$$\mathbf{y} = \sum_{\ell=0}^{L-1} h_\ell \mathbf{x}^{(\ell)} + \mathbf{w}, \qquad (3.122)$$

where $\mathbf{w} := [w[1], \dots, w[n+L]]^t$ and $\mathbf{x}^{(\ell)} = \pm \mathbf{u}^{(\ell)}$, the version of the trans-
mitted sequence (either $\mathbf{u}$ or $-\mathbf{u}$) shifted by $\ell$ chips. The received signal
(without the noise) therefore lies in the span of the $L$ vectors $\{\mathbf{u}^{(\ell)}/\|\mathbf{u}\|\}_\ell$. By
the pseudonoise assumption, all these vectors are orthogonal to each other.
A set of $L$ sufficient statistics $\{r^{(\ell)}\}_\ell$ can be obtained by projecting $\mathbf{y}$ onto
each of these vectors

$$r^{(\ell)} = h_\ell x + w^{(\ell)}, \qquad \ell = 0, \dots, L-1, \qquad (3.123)$$

where $x = \pm\|\mathbf{u}\|$. Further, the orthogonality of $\mathbf{u}^{(\ell)}$ implies that $w^{(\ell)}$ are i.i.d.
$\mathcal{CN}(0, N_0)$. Comparing with (3.32), this is exactly the same as the $L$-branch
diversity model for the case of repetition code interleaved over time. *Thus, we
see that the Rake receiver in this case is nothing more than a maximal ratio
combiner of the signals from the $L$ diversity branches.* The error probability
is given by

$$p_e = \mathbb{E}\left[ Q\left( \sqrt{2\|\mathbf{u}\|^2 \sum_{\ell=1}^{L} |h_\ell|^2 / N_0} \right) \right]. \qquad (3.124)$$

If we assume a Rayleigh fading model such that the tap gains $h_\ell$ are i.i.d.
$\mathcal{CN}(0, 1/L)$, i.e., the energy is spread equally among all the $L$ taps (normaliz-
ing such that the $\mathbb{E}[\sum_\ell |h_\ell|^2] = 1$), then the error probability can be explicitly
computed (as in (3.37)):

$$p_e = \left(\frac{1-\mu}{2}\right)^L \sum_{\ell=0}^{L-1} \binom{L-1+\ell}{\ell} \left(\frac{1+\mu}{2}\right)^\ell, \qquad (3.125)$$

where

$$\mu := \sqrt{\frac{\mathsf{SNR}}{1+\mathsf{SNR}}} \qquad (3.126)$$

and $\mathsf{SNR} := \|\mathbf{u}\|^2/(N_0 L)$ can be interpreted as the average signal-to-noise ratio
*per diversity branch*. Noting that $\|\mathbf{u}\|^2$ is the average total energy received
per bit of information, we can define $\mathcal{E}_b := \|\mathbf{u}\|^2$. Hence, the SNR per branch
is $1/L \cdot \mathcal{E}_b/N_0$. Observe that the factor of $1/L$ accounts for the splitting of
energy due to spreading: the larger the spread bandwidth $W$, the larger $L$ is,

and the more diversity one gets, but there is less energy in each branch.[13] As $L \to \infty$, $\sum_{\ell=1}^{L} |h_\ell|^2$ converges to 1 with probability 1 by the law of large numbers, and from (3.124) we see that

$$p_e \to Q\left(\sqrt{2\mathcal{E}_b/N_0}\right),\qquad(3.127)$$

i.e., the performance of the AWGN channel with the same $\mathcal{E}_b/N_0$ is asymptotically achieved.

The above analysis assumes an equal amount of energy in each tap. In a typical multipath delay profile, there is more energy in the taps with shorter delays. The analysis can be extended to the cases when the $h_\ell$ have unequal variances as well. (See Section 14.5.3 in [96]).

### 3.4.4 Orthogonal frequency division multiplexing

Both the single-carrier system with ISI equalization and the DS spread-spectrum system with Rake reception are based on a time-domain view of the channel. But we know that if the channel is linear time-invariant, sinusoids are eigenfunctions and they get transformed in a particularly simple way. ISI occurs in a single-carrier system because the transmitted signals are not sinusoids. This suggests that if the channel is underspread (i.e., the coherence time is much larger than the delay spread) and is therefore approximately time-invariant for a sufficiently long time-scale, then transformation into the frequency domain can be a fruitful approach to communication over frequency-selective channels. This is the basic idea behind OFDM.

We begin with the discrete-time baseband model

$$y[m] = \sum_\ell h_\ell[m]x[m-\ell] + w[m].\qquad(3.128)$$

For simplicity, we first assume that for each $\ell$, the $\ell$th tap is not changing with $m$ and hence the channel is linear time-invariant. Again assuming a finite number of non-zero taps $L := T_d W$, we can rewrite the channel model in (3.128) as

$$y[m] = \sum_{\ell=0}^{L-1} h_\ell x[m-\ell] + w[m].\qquad(3.129)$$

Sinusoids are eigenfunctions of LTI systems, but they are of infinite duration. If we transmit over only a finite duration, say $N_c$ symbols, then the sinusoids are no longer eigenfunctions. One way to restore the eigenfunction

---

[13] This is assuming a very rich scattering environment, leading to many paths, all of equal energy. In reality, however, there are just a few paths that are strong enough to matter.

property is by adding a *cyclic prefix* to the symbols. For every block of symbols of length $N_c$, denoted by

$$\mathbf{d} = [d[0], d[1], \ldots, d[N_c - 1]]^t,$$

we create an $N_c + L - 1$ input block as

$$\mathbf{x} = [d[N_c - L + 1], d[N_c - L + 2], \ldots, d[N_c - 1], d[0], d[1], \ldots, d[N_c - 1]]^t,$$
(3.130)

i.e., we add a *prefix* of length $L - 1$ consisting of data symbols rotated cyclically (Figure 3.20). With this input to the channel (3.129), consider the output

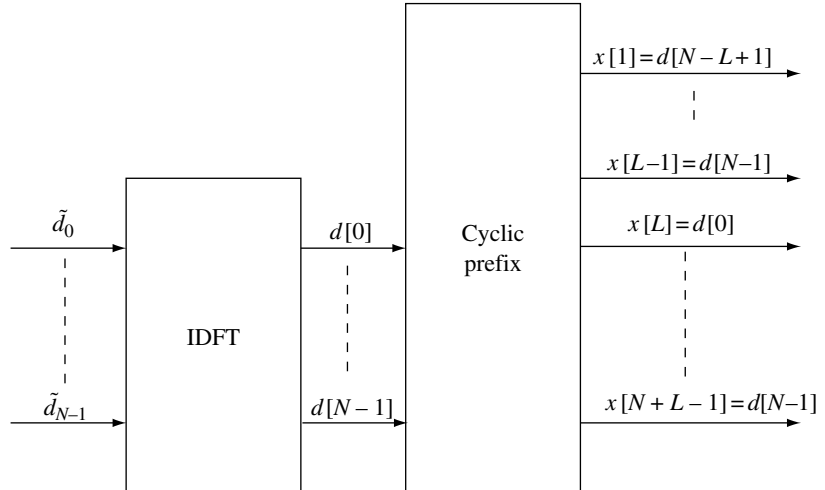$$y[m] = \sum_{\ell=0}^{L-1} h_\ell x[m - \ell] + w[m], \qquad m = 1, \ldots, N_c + L - 1.$$

The ISI extends over the first $L - 1$ symbols and the receiver ignores it by considering only the output over the time interval $m \in [L, N_c + L - 1]$. Due to the additional cyclic prefix, the output over this time interval (of length $N_c$) is

$$y[m] = \sum_{\ell=0}^{L-1} h_\ell d[(m - L - \ell) \text{ modulo } N_c] + w[m].$$
(3.131)

See Figure 3.21.

Denoting the output of length $N_c$ by

$$\mathbf{y} = [y[L], \ldots, y[N_c + L - 1]]^t,$$



**Figure 3.20** The cyclic prefix operation.

**Figure 3.21** Convolution between the channel (**h**) and the input (**x**) formed from the data symbols (**d**) by adding a cyclic prefix. The output is obtained by multiplying the corresponding values of **x** and **h** on the circle, and outputs at different times are obtained by rotating the *x*-values with respect to the *h*-values. The current configuration yields the output *y*[*L*].
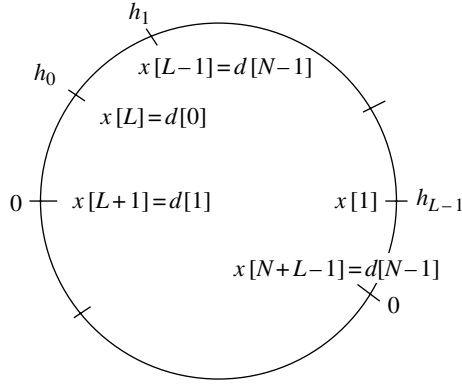


and the channel by a vector of length $N_c$

$$\mathbf{h} = [h_0, h_1, \ldots, h_{L-1}, 0, \ldots, 0]^t, \tag{3.132}$$

(3.131) can be written as

$$\mathbf{y} = \mathbf{h} \otimes \mathbf{d} + \mathbf{w}. \tag{3.133}$$

Here we denoted

$$\mathbf{w} = [w[L], \ldots, w[N_c + L - 1]]^t, \tag{3.134}$$

as a vector of i.i.d. $\mathcal{CN}(0, N_0)$ random variables. We also used the notation of $\otimes$ to denote the *cyclic convolution* in (3.131). Recall that the discrete Fourier transform (DFT) of **d** is defined to be

$$\tilde{d}_n := \frac{1}{\sqrt{N_c}} \sum_{m=0}^{N_c-1} d[m] \exp\left(\frac{-j2\pi nm}{N_c}\right), \qquad n = 0, \ldots, N-1. \tag{3.135}$$

Taking the discrete Fourier transform (DFT) of both sides of (3.133) and using the identity

$$\text{DFT}(\mathbf{h} \otimes \mathbf{d})_n = \sqrt{N_c} \text{DFT}(\mathbf{h})_n \cdot \text{DFT}(\mathbf{d})_n, \qquad n = 0, \ldots, N_c - 1, \tag{3.136}$$

we can rewrite (3.133) as

$$\tilde{y}_n = \tilde{h}_n \tilde{d}_n + \tilde{w}_n, \qquad n = 0, \ldots, N_c - 1. \tag{3.137}$$

Here we have denoted $\tilde{w}_0, \ldots, \tilde{w}_{N_c-1}$ as the $N_c$-point DFT of the noise vector $w[1], \ldots, w[N_c]$. The vector $[\tilde{h}_0, \ldots, \tilde{h}_{N_c-1}]^t$ is defined as the DFT of the $L$-tap channel **h**, multiplied by $\sqrt{N_c}$,

$$\tilde{h}_n = \sum_{\ell=0}^{L-1} h_\ell \exp\left(\frac{-j2\pi n\ell}{N_c}\right). \tag{3.138}$$

Note that the $n$th component $\tilde{h}_n$ is equal to the frequency response of the channel (see (2.20)) at $f = nW/N_c$.

We can redo everything in terms of matrices, a viewpoint which will prove particularly useful in Chapter 7 when we will draw a connection between the frequency-selective channel and the MIMO channel. The circular convolution operation $\mathbf{u} = \mathbf{h} \otimes \mathbf{d}$ can be viewed as a linear transformation

$$\mathbf{u} = \mathbf{Cd}, \tag{3.139}$$

where

$$\mathbf{C} := \begin{bmatrix} h_0 & 0 & \cdot & 0 & h_{L-1} & h_{L-2} & \cdot & h_1 \\ h_1 & h_0 & 0 & \cdot & 0 & h_{L-1} & \cdot & h_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & h_{L-1} & h_{L-2} & \cdot & h_1 & h_0 \end{bmatrix} \tag{3.140}$$

is a *circulant* matrix, i.e., the rows are cyclic shifts of each other. On the other hand, the DFT of $\mathbf{d}$ can be represented as an $N_c$-length vector $\mathbf{Ud}$, where $\mathbf{U}$ is the unitary matrix with its $(k, n)$th entry equal to

$$\frac{1}{\sqrt{N_c}} \exp\left(\frac{-j2\pi kn}{N_c}\right), \qquad k, n = 0, \ldots, N_c - 1. \tag{3.141}$$

This can be viewed as a coordinate change, expressing $\mathbf{d}$ in the basis defined by the rows of $\mathbf{U}$. Equation (3.136) is equivalent to

$$\mathbf{Uu} = \Lambda \mathbf{Ud}, \tag{3.142}$$

where $\Lambda$ is the diagonal matrix with diagonal entries $\sqrt{N_c}$ times the DFT of $\mathbf{h}$, i.e.,

$$\Lambda_{nn} = \tilde{h}_n := \left(\sqrt{N_c} \mathbf{Uh}\right)_n, \qquad n = 0, \ldots, N_c - 1.$$

Comparing (3.139) and (3.142), we come to the conclusion that

$$\mathbf{C} = \mathbf{U}^{-1} \Lambda \mathbf{U}. \tag{3.143}$$

Equation (3.143) is the matrix version of the key DFT property (3.136). In geometric terms, this means that the circular convolution operation is diagonalized in the coordinate system defined by the rows of $\mathbf{U}$, and the eigenvalues of $\mathbf{C}$ are the DFT coefficients of the channel $\mathbf{h}$. Equation (3.133) can thus be written as

$$\mathbf{y} = \mathbf{Cd} + \mathbf{w} = \mathbf{U}^{-1} \Lambda \mathbf{Ud} + \mathbf{w}. \tag{3.144}$$
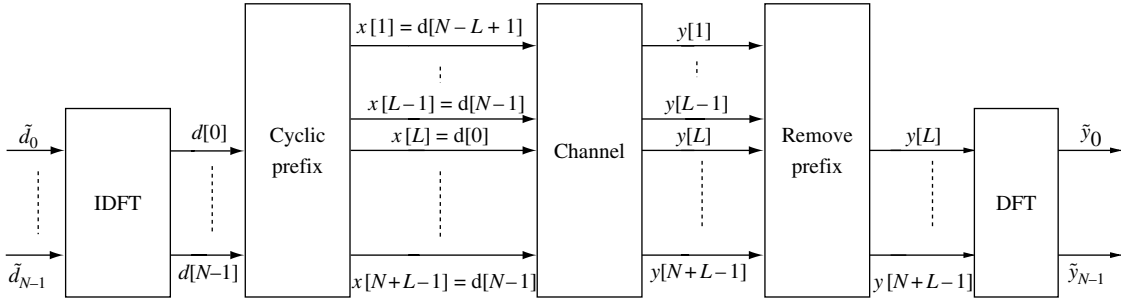
**Figure 3.22** The OFDM transmission and reception schemes.

This representation suggests a natural rotation at the input and at the output to convert the channel to a set of non-interfering channels with no ISI. In particular, the actual data symbols (denoted by the length $N_c$ vector $\check{\mathbf{d}}$) in the frequency domain are rotated through the IDFT (inverse DFT) matrix $\mathbf{U}^{-1}$ to arrive at the vector $\mathbf{d}$. At the receiver, the output vector of length $N_c$ (obtained by ignoring the first $L$ symbols) is rotated through the DFT matrix $\mathbf{U}$ to obtain the vector $\tilde{\mathbf{y}}$. The final output vector $\tilde{\mathbf{y}}$ and the actual data vector $\check{\mathbf{d}}$ are related through

$$\tilde{y}_n = \tilde{h}_n \check{d}_n + \tilde{w}_n, \qquad n = 0, \ldots, N_c - 1. \tag{3.145}$$

We have denoted $\tilde{\mathbf{w}} := \mathbf{U}\mathbf{w}$ as the DFT of the random vector $\mathbf{w}$ and we see that since $\mathbf{w}$ is isotropic, $\tilde{\mathbf{w}}$ has the same distribution as $\mathbf{w}$, i.e., a vector of i.i.d. $\mathcal{CN}(0, N_0)$ random variables (cf. (A.26) in Appendix A).

These operations are illustrated in Figure 3.22, which affords the following interpretation. The data symbols modulate $N_c$ *tones* or *sub-carriers*, which occupy the bandwidth $W$ and are uniformly separated by $W/N_c$. The data symbols on the sub-carriers are then converted (through the IDFT) to time domain. The procedure of introducing the cyclic prefix before transmission allows for the removal of ISI. The receiver ignores the part of the output signal containing the cyclic prefix (along with the ISI terms) and converts the length $N_c$ symbols back to the frequency domain through a DFT. The data symbols on the sub-carriers are maintained to be orthogonal as they propagate through the channel and hence go through narrowband *parallel* sub-channels. This interpretation justifies the name of OFDM for this communication scheme. Finally, we remark that DFT and IDFT can be very efficiently implemented (using Fast Fourier Transform) whenever $N_c$ is a power of 2.

### OFDM block length

The OFDM scheme converts communication over a multipath channel into communication over simpler parallel narrowband sub-channels. However, this simplicity is achieved at a cost of underutilizing two resources, resulting in a loss of performance. First, the cyclic prefix occupies an amount of time which cannot be used to communicate data. This loss amounts to a fraction

$L/(N_c + L)$ of the total time. The second loss is in the power transmitted. A fraction $L/(N_c + L)$ of the average power is allocated to the cyclic prefix and cannot be used towards communicating data. Thus, to minimize the overhead (in both time and power) due to the cyclic prefix we prefer to have $N_c$ as large as possible. The time-varying nature of the wireless channel, however, constrains the largest value $N_c$ can reasonably take.

We started the discussion in this section by considering a simple channel model (3.129) that did not vary with time. If the channel is slowly time-varying (as discussed in Section 2.2.1, this is a reasonable assumption) then the coherence time $T_c$ is much larger than the delay spread $T_d$ (the *under-spread* scenario). For underspread channels, the block length of the OFDM communication scheme $N_c$ can be chosen significantly larger than the multi-path length $L = T_d W$, but still much smaller than the coherence block length $T_c W$. Under these conditions, the channel model of linear time invariance approximates a slowly time-varying channel over the block length $N_c$, while keeping the overhead small.

The constraint on the OFDM block length can also be understood in the frequency domain. A block length of $N_c$ corresponds to an inter-sub-carrier spacing equal to $W/N_c$. In a wireless channel, the Doppler spread introduces uncertainty in the frequency of the received signal; from Table 2.1 we see that the Doppler spread is inversely proportional to the coherence time of the channel: $D_s = 1/4T_c$. For the inter-sub-carrier spacing to be much larger than the Doppler spread, the OFDM block length $N_c$ should be constrained to be much smaller than $T_c W$. This is the same constraint as above.

Apart from an underutilization of time due to the presence of the cyclic prefix, we also mentioned the additional power due to the cyclic prefix. OFDM schemes that put a zero signal instead of the cyclic prefix have been proposed to reduce this loss. However, due to the abrupt transition in the signal, such schemes introduce harmonics that are difficult to filter in the overall signal. Further, the cyclic prefix can be used for timing and frequency acquisition in wireless applications, and this capability would be lost if a zero signal replaced the cyclic prefix.

## Frequency diversity

Let us revert to the non-overlapping narrowband channel representation of the ISI channel in (3.145). The correlation between the channel frequency coefficients $\tilde{h}_0, \ldots, \tilde{h}_{N_c-1}$ depends on the coherence bandwidth of the channel. From our discussion in Section 2.3, we have learned that the coherence bandwidth is inversely proportional to the multipath spread. In particular, we have from (2.47) that

$$W_c = \frac{1}{2T_d} = \frac{W}{2L},$$

where we use our notation for $L$ as denoting the length of the ISI. Since each sub-carrier is $W/N_c$ wide, we expect approximately

$$\frac{N_c W_c}{W} = \frac{N_c}{2L}$$

as the number of neighboring sub-carriers whose channel coefficients are heavily correlated (Exercise 3.28). One way to exploit the frequency diversity is to consider ideal interleaving across the sub-carriers (analogous to the time-interleaving done in Section 3.2) and consider the model of (3.31)

$$y_\ell = h_\ell x_\ell + w_\ell, \qquad \ell = 1, \ldots, L.$$

The difference is that now $\ell$ represents the sub-carriers while it is used to denote time in (3.31). However, with the ideal frequency interleaving assumption we retain the same independent assumption on the channel coefficients. Thus, the discussion of Section 3.2 on schemes harnessing diversity is directly applicable here. In particular, an $L$-fold diversity gain (proportional to the number of ISI symbols $L$) can be obtained. Since the communication scheme is over sub-carriers, the form of diversity is due to the frequency-selective channel and is termed *frequency diversity* (as compared to the time diversity discussed in Section 3.2 which arises due to the time variations of the channel).

## Summary 3.3 Communication over frequency-selective channels

We have studied three approaches to extract frequency diversity in a frequency-selective channel (with $L$ taps). We summarize their key attributes and compare their implementational complexity.

### 1 Single-carrier with ISI equalization
Using maximum likelihood sequence detection (MLSD), full diversity of $L$ can be achieved for uncoded transmission sent at symbol rate.

MLSD can be performed by the Viterbi algorithm. The complexity is constant per symbol time but grows exponentially with the number of taps $L$.

The complexity is entirely at the receiver.

### 2 Direct-sequence spread-spectrum
Information is spread, via a pseudonoise sequence, across a bandwidth much larger than the data rate. ISI is typically negligible.

The signal received along the $L$ nearly orthogonal diversity paths is maximal-ratio combined using the Rake receiver. Full diversity is achieved.

Compared to MLSD, complexity of the Rake receiver is much lower. ISI is avoided because of the very low spectral efficiency per user, but the spectrum is typically shared between many interfering users. Complexity is thus shifted to the problem of interference management.

**3 Orthogonal frequency division multiplexing**
Information is modulated on non-interfering sub-carriers in the frequency domain.

The transformation between the time and frequency domains is done by means of adding/subtracting a cyclic prefix and IDFT/DFT operations. This incurs an overhead in terms of time and power.

Frequency diversity is attained by coding over independently faded sub-carriers. This coding problem is identical to that for time diversity.

Complexity is shared between the transmitter and the receiver in performing the IDFT and DFT operations; the complexity of these operations is insensitive to the number of taps, scales moderately with the number of sub-carriers $N_c$ and is very manageable with current implementation technology.

Complexity of diversity coding across sub-carriers can be traded off with the amount of diversity desired.

## 3.5 Impact of channel uncertainty

In the past few sections we assumed perfect channel knowledge so that coherent combining can be performed at the receiver. In fast varying channels, it may not be easy to estimate accurately the phases and magnitudes of the tap gains before they change. In this case, one has to understand the impact of estimation errors on performance. In some situations, non-coherent detection, which does not require an estimate of the channel, may be the preferred route. In Section 3.1.1, we have already come across a simple non-coherent detector for fading channels without diversity. In this section, we will extend this to channels with diversity.

When we compared coherent and non-coherent detection for channels without diversity, the difference was seen to be relatively small (cf. Figure 3.2). An important question is what happens to that difference as the number of diversity paths $L$ increases. The answer depends on the specific diversity scenario. We first focus on the situation where channel uncertainty has the most impact: DS spread-spectrum over channels with frequency diversity. Once we understand this case, it is easy to extend the insights to other scenarios.

## 3.5.1 Non-coherent detection for DS spread-spectrum

We considered this scenario in Section 3.4.3, except now the receiver has no knowledge of the channel gains $h_\ell$. As we saw in Section 3.1.1, no information can be communicated in the phase of the transmitted signal in conjunction with non-coherent detection (in particular, antipodal signaling cannot be used). Instead, we consider binary orthogonal modulation,[14] i.e., $\mathbf{x}_A$ and $\mathbf{x}_B$ are orthogonal and $\|\mathbf{x}_A\| = \|\mathbf{x}_B\|$.

Recall that the central pseudonoise property of the transmitted sequences in DS spread-spectrum is that the shifted versions are nearly orthogonal. For simplicity of analysis, we continue with the assumption that shifted versions of the transmitted sequence are exactly orthogonal; this holds for both $\mathbf{x}_A$ and $\mathbf{x}_B$ here. We make the further assumption that versions of the two sequences with different shifts are also orthogonal to each other, i.e., $(\mathbf{x}_A^{(\ell)})^*(\mathbf{x}_B^{(\ell')}) = 0$ for $\ell \neq \ell'$ (the so-called zero cross-correlation property). This approximately holds in many spread-spectrum systems. For example, in the uplink of IS-95, the transmitted sequence is obtained by multiplying the selected codeword of an orthogonal code by a (common) pseudonoise $\pm 1$ sequence, so that the low cross-correlation property carries over from the auto-correlation property of the pseudonoise sequence.

Proceeding as in the analysis of coherent detection, we start with the channel model in vector form (3.122) and observe that the projection of $\mathbf{y}$ onto the $2L$ orthogonal vectors $\{\mathbf{x}_A^{(\ell)}/\|\mathbf{x}_A\|, \mathbf{x}_B^{(\ell)}/\|\mathbf{x}_B\|\}_\ell$ yields $2L$ sufficient statistics:

$$r_A^{(\ell)} = h_\ell x_1 + w_A^{(\ell)}, \qquad \ell = 0, \ldots, L-1,$$
$$r_B^{(\ell)} = h_\ell x_2 + w_B^{(\ell)}, \qquad \ell = 0, \ldots, L-1,$$

where $w_A^{(\ell)}$ and $w_B^{(\ell)}$ are i.i.d. $\mathcal{CN}(0, N_0)$, and

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} \|\mathbf{x}_A\| \\ 0 \end{pmatrix} & \text{if } \mathbf{x}_A \text{ is transmitted,} \\ \begin{pmatrix} 0 \\ \|\mathbf{x}_B\| \end{pmatrix} & \text{if } \mathbf{x}_B \text{ is transmitted.} \end{cases} \tag{3.146}$$

This is essentially a generalization of the non-coherent detection problem in Section 3.1.1 from 1 branch to $L$ branches. Just as in the 1 branch case, a

---

[14] Typically $M$-ary orthogonal modulation is used. For example, the uplink of IS-95 employs non-coherent detection of 64-ary orthogonal modulation.

square-law type detector is the optimal non-coherent detector: decide in favor of $\mathbf{x}_A$ if

$$\sum_{\ell=0}^{L-1} |r_A^{(\ell)}|^2 \geq \sum_{\ell=0}^{L-1} |r_B^{(\ell)}|^2, \tag{3.147}$$

otherwise decide in favor of $\mathbf{x}_B$. The performance can be analyzed as in the 1 branch case: the error probability has the same form as in (3.125), but with $\mu$ given by

$$\mu = \frac{1/L \cdot \mathcal{E}_{\mathrm{b}}/N_0}{2 + 1/L \cdot \mathcal{E}_{\mathrm{b}}/N_0}, \tag{3.148}$$

where $\mathcal{E}_{\mathrm{b}} := \|\mathbf{x}_A\|^2$. (See Exercise 3.31.) As a basis of comparison, the performance of coherent detection of binary orthogonal modulation can be analyzed as for the antipodal case; it is again given by (3.125) but with $\mu$ given by (Exercise 3.33):

$$\mu = \sqrt{\frac{1/L \cdot \mathcal{E}_{\mathrm{b}}/N_0}{2 + 1/L \cdot \mathcal{E}_{\mathrm{b}}/N_0}}. \tag{3.149}$$

It is interesting to compare the performance of coherent and non-coherent detection as a function of the number of diversity branches. This is shown in Figures 3.23 and 3.24. For $L = 1$, the gap between the performance of both schemes is small, but they are bad anyway, as there is a lack of diversity. This point has already been made in Section 3.1. As $L$ increases, the performance of coherent combining improves monotonically and approaches the performance of an AWGN channel. In contrast, the performance of non-coherent detection first improves with $L$ but then degrades as $L$ is increased further.
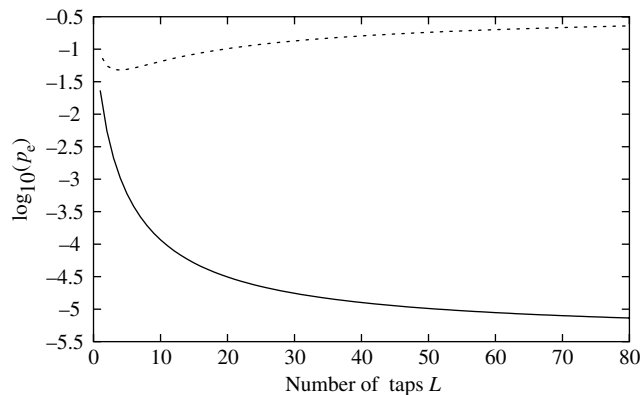


**Figure 3.23** Comparison of error probability under coherent detection (—) and non-coherent detection (---), as a function of the number of taps $L$. Here $\mathcal{E}_{\mathrm{b}}/N_0 = 10\,\mathrm{dB}$.

**Figure 3.24** Comparison of error probability under coherent detection (—) and non-coherent detection (---), as a function of the number of taps $L$. Here $\mathcal{E}_b/N_0 = 15\,\mathrm{dB}$.
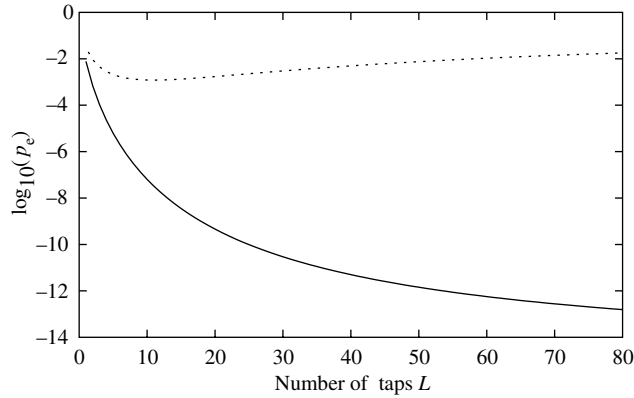


The initial improvement comes from a diversity gain. There is however a law of diminishing return on the diversity gain. At the same time, when $L$ becomes too large, the SNR per branch becomes very poor and non-coherent combining cannot effectively exploit the available diversity. This leads to an ultimate degradation in performance. In fact, it can be shown that as $L \to \infty$ the error probability approaches 1/2.

## 3.5.2 Channel estimation

The significant performance difference between coherent and non-coherent combining when the number of branches is large suggests the importance of channel knowledge in wideband systems. We assumed perfect channel knowledge when we analyzed the performance of the coherent Rake receiver, but in practice, the channel taps have to be estimated and tracked. It is therefore important to understand the impact of channel measurement errors on the performance of the coherent combiner. We now turn to the issue of channel estimation.

In data detection, the transmitted sequence is one of several possible sequences (representing the data symbol). In channel estimation, the transmitted sequence is assumed to be known at the receiver. In a pilot-based scheme, a known sequence (called a pilot, sounding tone, or training sequence) is transmitted and this is used to estimate the channel.[15] In a decision-feedback scheme, the previously detected symbols are used instead to update the channel estimates. If we assume that the detection is error free, then the development below applies to both pilot-based and decision-directed schemes.

---

[15] The downlink of IS-95 uses a pilot, which is assigned its own pseudonoise sequence and transmitted superimposed on the data.

Focus on one symbol duration, and suppose the transmitted sequence is a known pseudonoise sequence $\mathbf{u}$. We return to the channel model in vector form (cf. (3.122))

$$\mathbf{y} = \sum_{\ell=0}^{L-1} h_\ell \mathbf{u}^{(\ell)} + \mathbf{w}. \tag{3.150}$$

We see that since the shifted versions of $\mathbf{u}$ are orthogonal to each other and the taps are assumed to be independent of each other, projecting $\mathbf{y}$ onto $\mathbf{u}^{(\ell)}/\|\mathbf{u}^{(\ell)}\|$ will yield a sufficient statistic to estimate $h_\ell$ (see Summary A.3)

$$r^{(\ell)} := (\mathbf{u}^{(\ell)})^* \mathbf{y} = h_\ell \|\mathbf{u}^{(\ell)}\| + w^{(\ell)} = \sqrt{\mathcal{E}} h_\ell + w^{(\ell)}, \tag{3.151}$$

where $\mathcal{E} := \|\mathbf{u}^{(\ell)}\|^2$. This is implemented by filtering the received signal by a filter matched to $\mathbf{u}$ and sampling at the appropriate chip time. This operation is the same as the first stage of the Rake receiver, and the channel estimator can in fact be combined with the Rake receiver if done in a decision-directed mode. (See Figure 3.19.)

Typically, channel estimation is obtained by averaging $K$ such measurements over a coherence time period in which the channel is constant:

$$r_k^{(\ell)} := \sqrt{\mathcal{E}} h_\ell + w_k^{(\ell)}, \qquad k = 1, \dots, K. \tag{3.152}$$

Assuming that $h_\ell \sim \mathcal{CN}(0, 1/L)$, the minimum mean square estimate of $h_\ell$ given these measurements is (cf. (A.84) in Summary A.3)

$$\hat{h}_\ell = \frac{\sqrt{\mathcal{E}}}{K\mathcal{E} + L N_0} \sum_{k=1}^{K} r_k^{(\ell)}. \tag{3.153}$$

The mean square error associated with this estimate is (cf. (A.85) in Summary A.3)

$$\frac{1}{L} \cdot \frac{1}{1 + K\mathcal{E}/(L N_0)}. \tag{3.154}$$

the same for all branches.

The key parameter affecting the estimation error is

$$\mathsf{SNR}_{\text{est}} := \frac{K\mathcal{E}}{L N_0}. \tag{3.155}$$

When $\mathsf{SNR}_{\text{est}} \gg 1$, the mean square estimation error is much smaller than the variance of $h_\ell$ (equal to $1/L$) and the impact of the channel estimation error on the performance of the coherent Rake receiver is not significant; perfect

channel knowledge is a reasonable assumption in this regime. On the other hand, when $\mathsf{SNR}_{est} \ll 1$, the mean square error is close to $1/L$, the variance of $h_\ell$. In this regime, we hardly have any information about the channel gains and the performance of the coherent combiner cannot be expected to be better than the non-coherent combiner, which we know has poor performance whenever $L$ is large.

How should we interpret the parameter $\mathsf{SNR}_{est}$? Since the channel is constant over the coherence time $T_c$, we can interpret $K\mathcal{E}$ as the total received energy over the channel coherence time $T_c$. We can rewrite $\mathsf{SNR}_{est}$ as

$$\mathsf{SNR}_{est} = \frac{PT_c}{LN_0} \qquad (3.156)$$

where $P$ is the received power of the signal from which channel measurements are obtained. Hence, $\mathsf{SNR}_{est}$ can be interpreted as the signal-to-noise ratio available to estimate the channel per coherence time per tap. Thus, channel uncertainty has a significant impact on the performance of the Rake receiver whenever this quantity is significantly below $0\,\mathrm{dB}$.

If the measurements are done in a decision-feedback mode, $P$ is the received power of the data stream itself. If the measurements are done from a pilot, then $P$ is the received power of the pilot. On the downlink of a CDMA system, one can have a pilot common to all users, and the power allocated to the pilot can be larger than the power of the signals for the individual users. This results in a larger $\mathsf{SNR}_{est}$, and thus makes coherent combining easier. On the uplink, however, it is not possible to have a common pilot, and the channel estimation will have to be done with a weaker pilot allotted to the individual user. With a lower received power from the individual users, $\mathsf{SNR}_{est}$ can be considerably smaller.

### 3.5.3 Other diversity scenarios

There are two reasons why wideband DS spread-spectrum systems are significantly impacted by channel uncertainty:

- the amount of energy per resolvable path decreases inversely with increasing number of paths, making their gains harder to estimate when there are many paths;
- the number of diversity paths depends both on the bandwidth and the delay spread and, given these parameters, the designer has no control over this number.

What about in other diversity scenarios?

In antenna diversity with $L$ receive antennas, the received energy per antenna is the same regardless of the number of antennas, so the channel

measurement problem is the same as with a single receive antenna and does not become harder. The situation is similar in the time diversity scenario. In antenna diversity with $L$ *transmit* antennas, the received energy per diversity path *does* decrease with the number of antennas used, but certainly we can restrict the number $L$ to be the number of different channels that can be reliably learnt by the receiver.

How about in OFDM systems with frequency diversity? Here, the designer has control over how many sub-carriers to spread the signal energy over. Thus, while the number of *available* diversity branches $L$ may increase with the bandwidth, the signal energy can be restricted to a fixed number of sub-carriers $L' < L$ over any one OFDM time block. Such communication can be restricted to concentrated time-frequency blocks and Figure 3.25 visualizes one such scheme (for $L' = 2$), where the choice of the $L'$ sub-carriers is different for different OFDM blocks and is hopped over the entire bandwidth. Since the energy in each OFDM block is concentrated within a fixed number of sub-carriers at any one time, coherent reception is possible. On the other hand, the maximum diversity gain of $L$ can still be achieved by coding across the sub-carriers within one OFDM block as well as across different blocks.

One possible drawback is that since the total power is only concentrated within a subset of sub-carriers, the total degrees of freedom available in the system are not utilized. This is certainly the case in the context of point-to-point communication; in a system with other users sharing the same band-width, however, the other degrees of freedom can be utilized by the other users and need not go wasted. In fact, one key advantage of OFDM over DS spread-spectrum is the ability to maintain orthogonality across multiple users in a multiple access scenario. We will return to this point in Chapter 4.
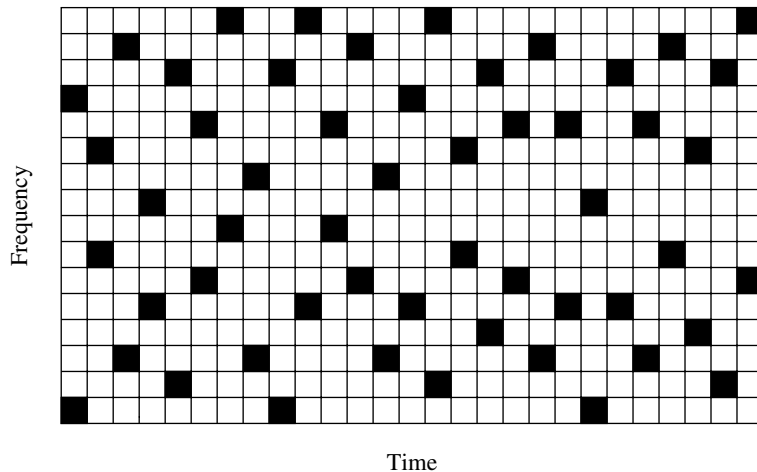
**Figure 3.25** An illustration of a scheme that uses only a fixed part of the bandwidth at every time. Here, one small square denotes a single sub-carrier within one OFDM block. The time-axis indexes the different OFDM blocks; the frequency-axis indexes the different sub-carriers.

## Chapter 3 The main plot

**Baseline**

We first looked at detection on a narrowband slow fading Rayleigh channel. Under both coherent and non-coherent detection, the error probability behaves like

$$p_{\mathrm{e}} \approx \mathsf{SNR}^{-1} \qquad (3.157)$$

at high SNR. In contrast, the error probability decreases *exponentially* with the SNR in the AWGN channel. The typical error event for the fading channel is due to the channel being in deep fade rather than the Gaussian noise being large.

**Diversity**

*Diversity* was presented as an effective approach to improve performance drastically by providing redundancy across independently faded branches. Three modes of diversity were considered:

- time – the interleaving of coded symbols over different coherence time periods;
- space – the use of multiple transmit and/or receive antennas;
- frequency – the use of a bandwidth greater than the coherence bandwidth of the channel.

In all cases, a simple scheme that repeats the information symbol across the multiple branches achieves full diversity. With $L$ i.i.d. Rayleigh branches of diversity, the error probability behaves like

$$p_{\mathrm{e}} \approx c \cdot \mathsf{SNR}^{-L} \qquad (3.158)$$

at high SNR.

Examples of repetition schemes:

- repeating the same symbol over different coherence periods;
- repeating the same symbol over different transmit antennas one at a time;
- repeating the same symbol across OFDM sub-carriers in different coherence bands;
- transmitting a symbol once every delay spread in a frequency-selective channel so that multiple delayed replicas of the symbol are received without interference.

**Code design and degrees of freedom**

More sophisticated schemes cannot achieve higher diversity gain but can provide a *coding gain* by improving the constant $c$ in (3.158). This is

achieved by utilizing the available *degrees of freedom* better than in the repetition schemes.

Examples:
- rotation and permutation codes for time diversity and for frequency diversity in OFDM;
- Alamouti scheme for transmit diversity;
- uncoded transmission at symbol rate in a frequency-selective channel with ISI equalization.

Criteria to design schemes with good coding gain were derived for the different scenarios by using the union bound (based on pairwise error probabilities) on the actual error probability:
- product distance between codewords for time diversity;
- determinant criterion for space-time codes.

**Channel uncertainty**

The impact of channel uncertainty is significant in scenarios where there are many diversity branches but only a small fraction of signal energy is received along each branch. Direct-sequence spread-spectrum is a prime example.

The gap between coherent and non-coherent schemes is very significant in this regime. Non-coherent schemes do not work well as they cannot combine the signals along each branch effectively.

Accurate channel estimation is crucial. Given the amount of transmit power devoted to channel estimation, the efficacy of detection performance depends on the key parameter $\mathsf{SNR}_{\text{est}}$, the received SNR per coherence time per diversity branch. If $\mathsf{SNR}_{\text{est}} \gg 0\,\text{dB}$, then detection performance is near coherent. If $\mathsf{SNR}_{\text{est}} \ll 0\,\text{dB}$, then effective combining is impossible.

Impact of channel uncertainty can be ameliorated in some schemes where the transmit energy can be focused on smaller number of diversity branches. Effectively $\mathsf{SNR}_{\text{est}}$ is increased. OFDM is an example.

## 3.6 Bibliographical notes

Reliable communication over fading channels has been studied since the 1960s. Improving the performance via diversity is also an old topic. Standard digital communication texts contain many formulas for the performance of coherent and non-coherent diversity combiners, which we have used liberally in this chapter (see Chapter 14 of Proakis [96], for example).

Early works recognizing the importance of the product distance criterion for improving the coding gain under Rayleigh fading are Wilson and Leung [144] and Divsalar

and Simon [30], in the context of trellis-coded modulation. The rotation example is taken from Boutros and Viterbo [13]. Transmit antenna diversity was studied extensively in the late 1990s code design criteria were derived by Tarokh *et al.* [115] and by Guey *et al.* [55]; in particular, the determinant criterion is obtained in Tarokh *et al.* [115]. The delay diversity scheme was introduced by Seshadri and Winters [107]. The Alamouti scheme was introduced by Alamouti [3] and generalized to orthogonal designs by Tarokh *et al.* [117]. The diversity analysis of the decorrelator was performed by Winters *et al.* [145], in the context of a space-division multiple access system with multiple receive antennas.

The topic of equalization has been studied extensively and is covered comprehensively in standard textbooks on communication theory; for example, see the book by Barry *et al.* [4]. The Viterbi algorithm was introduced in [139]. The diversity analysis of MLSD is adopted from Grokop and Tse [54].

The OFDM approach to communicate over a wideband channel was first used in military systems in the 1950s and discussed in early papers in the 1960s by Chang [18] and Saltzberg [104]. Circular convolution and the DFT are classical undergraduate material in digital signal processing (Chapter 8, and Section 8.7.5, in particular, of [87]).

The spread-spectrum approach to harness frequency diversity has been well summarized by Viterbi [140]. The Rake receiver was designed by Price and Green [95]. The impact of channel uncertainty on the performance has been studied by various authors, including Médard and Gallager [85], Telatar and Tse [120] and Subramanian and Hajek [113].

## 3.7 Exercises

**Exercise 3.1** Verify (3.19) and the high SNR approximation (3.21). *Hint*: Write the expression as a double integral and interchange the order of integration.

**Exercise 3.2** In Section 3.1.2 we studied the performance of antipodal signaling under coherent detection over a Rayleigh fading channel. In particular, we saw that the error probability $p_e$ decreases like $1/\mathsf{SNR}$. In this question, we study a deeper characterization of the behavior of $p_e$ with increasing SNR.

1. A precise way of saying that $p_e$ decays like $1/\mathsf{SNR}$ with increasing SNR is the following:

$$\lim_{\mathsf{SNR} \to \infty} p_e \cdot \mathsf{SNR} = c,$$

   where $c$ is a constant. Identify the value of $c$ for the Rayleigh fading channel.

2. Now we want to test how robust the above result is with respect to the fading distribution. Let $h$ be the channel gain, and suppose $|h|^2$ has an arbitrary continuous pdf $f$ satisfying $f(0) > 0$. Does this give enough information to compute the high SNR error probability like in the previous part? If so, compute it. If not, specify what other information you need. *Hint*: You may need to interchange limit and integration in your calculations. You can assume that this can be done without worrying about making your argument rigorous.

3. Suppose now we have $L$ independent branches of diversity with gains $h_1, \ldots, h_L$, and $|h_\ell|^2$ having an arbitrary distribution as in the previous part. Is there enough

information for you to find the high SNR performance of repetition coding and coherent combining? If so, compute it. If not, what other information do you need?
4. Using the result in the previous part or otherwise, compute the high SNR performance under Rician fading. How does the parameter $\kappa$ affect the performance?

**Exercise 3.3** This exercise shows how the high SNR slope of the probability of error (3.19) versus SNR curve can be obtained using a typical error event analysis, without the need for directly carrying out the integration.

Fix $\epsilon > 0$ and define the $\epsilon$-typical error events $\mathbb{E}_\epsilon$ and $\mathbb{E}_{-\epsilon}$, where

$$\mathbb{E}_\epsilon := \{h : |h|^2 < 1/\mathsf{SNR}^{1-\epsilon}\}. \tag{3.159}$$

1. By conditioning on the event $\mathbb{E}_\epsilon$, show that at high SNR

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log p_\mathrm{e}}{\log \mathsf{SNR}} \leq -(1-\epsilon). \tag{3.160}$$

2. By conditioning on the event $\mathbb{E}_{-\epsilon}$, show that

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log p_\mathrm{e}}{\log \mathsf{SNR}} \geq -(1+\epsilon). \tag{3.161}$$

3. Hence conclude that

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log p_\mathrm{e}}{\log \mathsf{SNR}} = -1. \tag{3.162}$$

This says that the asymptotic slope of the error probability versus SNR plot (in dB/dB scale) is $-1$.

**Exercise 3.4** In Section 3.1.2, we saw that there is a 4-dB energy loss when using 4-PAM on only the I channel rather than using QPSK on both the I and the Q channels, although both modulations convey two bits of information. Compute the corresponding loss when one wants to transmit $k$ bits of information using $2^k$-PAM rather than $2^k$-QAM. You can assume $k$ to be even. How does the loss depend on $k$?

**Exercise 3.5** Consider the use of the differential BPSK scheme proposed in Section 3.1.3 for the Rayleigh flat fading channel.
1. Find a natural non-coherent scheme to detect $u[m]$ based on $y[m-1]$ and $y[m]$, assuming the channel is constant across the two symbol times. Your scheme does not have to be the ML detector.
2. Analyze the performance of your detector at high SNR. You may need to make some approximations. How does the high SNR performance of your detector compare to that of the coherent detector?
3. Repeat your analysis for differential QPSK.

**Exercise 3.6** In this exercise we further study coherent detection in Rayleigh fading.
1. Verify Eq. (3.37).
2. Analyze the error probability performance of coherent detection of binary orthogonal signaling with $L$ branches of diversity, under an i.i.d. Rayleigh fading assumption (i.e., verify Eq. (3.149)).

**Exercise 3.7** In this exercise, we study the performance of the rotated code in Section 3.2.2.

1. Give an explicit expression for the *exact* pairwise error probability $\mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_B\}$ in (3.49). *Hint*: The techniques from Exercise 3.1 will be useful here.
2. This pairwise error probability was upper bounded in (3.54). Show that the product of SNR and the difference between the upper bound and the actual pairwise error probability goes to zero with increasing SNR. In other words, the upper bound in (3.54) is tight up to the leading term in $1/$SNR.

**Exercise 3.8** In the text, we mainly use real symbols to simplify the notation. In practice, complex constellations are used (i.e., symbols are sent along both the I and Q components). The simplest complex constellation is QPSK: the constellation is $\{a(1+j), a(1-j), a(-1-j), a(-1+j)\}$.

1. Compute the error probability of QPSK detection for a Rayleigh fading channel with repetition coding over $L$ branches of diversity. How does the performance compare to a scheme which uses only real symbols?
2. In Section 3.2.2, we developed a diversity scheme based on rotation of real symbols (thus using only the I channel). One can develop an analogous scheme for QPSK complex symbols, using a $2 \times 2$ complex unitary matrix instead. Find an analogous pairwise code-design criterion as in the real case.
3. Real orthonormal matrices are special cases of complex unitary matrices. Within the class of real orthonormal matrices, find the optimal rotation to maximize your criterion.
4. Find the optimal unitary matrix to maximize your criterion. (This may be difficult!)

**Exercise 3.9** In Section 3.2.2, we rotate two BPSK symbols to demonstrate the possible improvement over repetition coding in a time diversity channel with two diversity paths. Continuing with the same model, now consider transmitting at a higher rate using a $2n$-PAM constellation for each symbol. Consider rotating the resulting 2D constellation by a rotation matrix of the form in (3.46). Using the performance criterion of the minimum squared product distance, construct the optimal rotation matrix.

**Exercise 3.10** In Section 3.2.2, we looked at the example of the rotation code to achieve time diversity (with the number of branches, $L$, equal to 2). In the text, we use real symbols and in Exercise 3.8 we extend to complex symbols. In the latter scenario, another coding scheme is the *permutation code*. Shown in Figure 3.26 are two 16-QAM constellations. Each codeword in the permutation code for $L = 2$ is obtained by picking a pair of points, one from each constellation, which are represented by the same icon. The codeword is transmitted over two (complex) symbol times.

1. Why do you think this is called a permutation code?
2. What is the data rate of this code?
3. Compute the diversity gain and the minimum product distance for this code.
4. How does the performance of this code compare to the rotation code in Exercise 3.8, part (3), in terms of the transmit power required?

**Exercise 3.11** In the text, we considered the use of rotation codes to obtain time diversity. Rotation codes are designed specifically for fading channels. Alternatively, one can use standard AWGN codes like binary linear block codes. This question looks at the diversity performance of such codes.

**Figure 3.26** A permutation code.
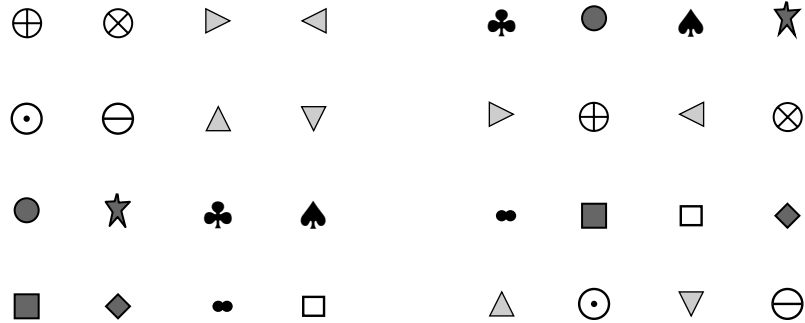


Consider a perfectly interleaved Rayleigh fading channel:

$$y_\ell = h_\ell x_\ell + w_\ell, \qquad \ell = 1, \ldots, L$$

where $h_\ell$ and $w_\ell$ are i.i.d. $\mathcal{CN}(0, 1)$ and $\mathcal{CN}(0, N_0)$ random variables respectively. A $(L, k)$ binary linear block code is specified by a $k$ by $L$ generator matrix $\mathbf{G}$ whose entries are 0 or 1. $k$ information bits form a $k$-dimensional binary-valued vector $\mathbf{b}$ which is mapped into the binary codeword $\mathbf{c} = \mathbf{G}^t \mathbf{b}$ of length $L$, which is then mapped into $L$ BPSK symbols and transmitted over the fading channel.[16] The receiver is assumed to have a perfect estimate of the channel gains $h_\ell$.

1. Compute a bound on the error probability of ML decoding in terms of the SNR and parameters of the code. Hence, compute the diversity gain in terms of code parameter(s).

2. Use your result in (1) to compute the diversity gain of the (3, 2) code with generator matrix:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}. \tag{3.163}$$

How does the performance of this code compare to the rate 1/2 repetition code?

3. The ML decoding is also called *soft decision decoding* as it takes the entire received vector $\mathbf{y}$ and finds the transmitted codeword closest in Euclidean distance to it. Alternatively, a suboptimal but lower-complexity decoder uses *hard decision decoding*, which for each $\ell$ first makes a hard decision $\hat{c}_\ell$ on the $\ell$th transmitted coded symbol based only on the corresponding received symbol $y_\ell$, and then finds the codeword that is closest in Hamming distance to $\hat{\mathbf{c}}$. Compute the diversity gain of this scheme in terms of basic parameters of the code. How does it compare to the diversity gain achieved by soft decision decoding? Compute the diversity gain of the code in part (2) under hard decision decoding.

4. Suppose now you still do hard decision decoding except that you are allowed to also declare an "erasure" on some of the transmitted symbols (i.e., you can refuse to make a hard decision on some of the symbols). Can you design a scheme that

---

[16] Addition and multiplication are done in the binary field.

yields a better diversity gain than the scheme in part (3)? Can you do as well as soft decision decoding? Justify your answers. Try your scheme out on the example in part (2). *Hint*: the trick is to figure out when to declare an erasure. You may want to start thinking of the problems in terms of the example in part (2). The typical error event view in Exercise 3.3 may also be useful here.

**Exercise 3.12** In our study of diversity models (cf. (3.31)), we have modeled the $L$ branches to have independent fading coefficients. Here we explore the impact of correlation between the $L$ diversity branches. In the time diversity scenario, consider the correlated model: $h_1, \ldots, h_L$ are jointly circular symmetric complex Gaussian with zero mean and covariance $\mathbf{K}_h$ ($\mathcal{CN}(0, \mathbf{K}_h)$ in our notation).

1. Redo the diversity calculations for repetition coding (Section 3.2.1) for this correlated channel model by calculating the rate of decay of error probability with SNR. What is the dependence of the asymptotic (in SNR) behavior of the typical error event on the correlation $\mathbf{K}_h$? You can answer this by characterizing the rate of decay of (3.42) at high SNR (as a function of $\mathbf{K}_h$).
2. We arrived at the product distance code design criterion to harvest coding gain along with time diversity in Section 3.2.2. What is the analogous criterion for correlated channels? *Hint*: Jointly complex Gaussian random vectors are related to i.i.d. complex Gaussian vectors via a linear transformation that depends on the covariance matrix.
3. For transmit diversity with independent fading across the transmit antennas, we have arrived at the generalized product distance code design criterion in Section 3.3.2. Calculate the code design criterion for the correlated fading channel here (the channel $\mathbf{h}$ in (3.80) is now $\mathcal{CN}(0, \mathbf{K}_h)$).

**Exercise 3.13** The optimal coherent receiver for repetition coding with $L$ branches of diversity is a maximal ratio combiner. For implementation reasons, a simpler receiver one often builds is a *selection combiner*. It does detection based on the received signal along the branch with the strongest gain only, and ignores the rest. For the i.i.d. Rayleigh fading model, analyze the high SNR performance of this scheme. How much of the inherent diversity gain can this scheme get? Quantify the performance loss from optimal combining. *Hint*: You may find the techniques developed in Exercise 3.2 useful for this problem.

**Exercise 3.14** It is suggested that full diversity gain can be achieved over a Rayleigh faded MISO channel by simply transmitting the same symbol at each of the transmit antennas simultaneously. Is this correct?

**Exercise 3.15** An $L \times 1$ MISO channel can be converted into a time diversity channel with $L$ diversity branches by simply transmitting over one antenna at a time.

1. In this way, any code designed for a time diversity channel with $L$ diversity branches can be used for a MISO (multiple input single output) channel with $L$ transmit antennas. If the code achieves $k$-fold diversity in the time diversity channel, how much diversity can it obtain in the MISO channel? What is the relationship between the minimum product distance metric of the code when viewed as a time diversity code and its minimum determinant metric when viewed as a transmit diversity code?

2. Using this transformation, the rotation code can be used as a transmit diversity scheme. Compare the performance of this code and the Alamouti scheme in a $2 \times 1$ Rayleigh fading channel, using BPSK symbols. Which one is better? How about using QPSK symbols?

3. Use the permutation code (cf. Figure 3.26) from Exercise 3.10 on the $2 \times 1$ Rayleigh fading channel and compare (via a numerical simulation) its performance with the Alamouti scheme using QPSK symbols (so the rate is the same in both the schemes).

**Exercise 3.16** In this exercise, we derive some properties a code construction must satisfy to mimic the Alamouti scheme behavior for more than two transmit antennas. Consider communication over $n$ time slots on the $L$ transmit antenna channel (cf. (3.80)):

$$\mathbf{y}^t = \mathbf{h}^* \mathbf{X} + \mathbf{w}^t. \tag{3.164}$$

Here $\mathbf{X}$ is the $L \times n$ space-time code. Over $n$ time slots, we want to communicate $L$ independent constellation symbols, $d_1, \ldots, d_L$; the space-time code $\mathbf{X}$ is a deterministic function of these symbols.

1. Consider the following property for every channel realization $\mathbf{h}$ and space-time codeword $\mathbf{X}$

$$(\mathbf{h}^* \mathbf{X})^t = \mathbf{A}\mathbf{d}. \tag{3.165}$$

   Here we have written $\mathbf{d} = [d_1, \ldots, d_L]^t$ and $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_L]$, a matrix with *orthogonal* columns. The vector $\mathbf{d}$ depends solely on the codeword $\mathbf{X}$ and the matrix $\mathbf{A}$ depends solely on the channel $\mathbf{h}$. Show that, if the space-time codeword $\mathbf{X}$ satisfies the property in (3.165), the joint receiver to detect $\mathbf{d}$ separates into individual linear receivers, each separately detecting $d_1, \ldots, d_L$.

2. We would like the effective channel (after the linear receiver) to provide each symbol $d_m$ $(m = 1, \ldots, L)$ with full diversity. Show that, if we impose the condition that

$$\|\mathbf{a}_m\| = \|\mathbf{h}\|, \qquad m = 1, \ldots L, \tag{3.166}$$

   then each data symbol $d_m$ has full diversity.

3. Show that a space-time code $\mathbf{X}$ satisfying (3.165) (the linear receiver property) and (3.166) (the full diversity property) must be of the form

$$\mathbf{X}\mathbf{X}^* = \|\mathbf{d}\|^2 \mathbf{I}_L, \tag{3.167}$$

   i.e., the columns of $\mathbf{X}$ must be orthogonal. Such an $\mathbf{X}$ is called an *orthogonal design*. Indeed, we observe that the codeword $\mathbf{X}$ in the Alamouti scheme (cf. (3.77)) is an orthogonal design with $L = n = 2$.

**Exercise 3.17** This exercise is a sequel to Exercise 3.16. It turns out that if we require $n = L$, then for $L > 2$ there are no orthogonal designs. (This result is proved in Theorem 5.4.2 in [117].) If we settle for $n > L$ then orthogonal designs exist for

$L > 2$. In particular, Theorem 5.5.2 of [117] constructs orthogonal designs for all $L$ and $n \geq 2L$. This does not preclude the existence of orthogonal designs with rate larger than 0.5. A reading exercise is to study [117] where orthogonal designs with rate larger than 0.5 are constructed.

**Exercise 3.18** The pairwise error probability analysis for the i.i.d. Rayleigh fading channel has led us to the product distance (for time diversity) and generalized product distance (for transmit diversity) code design criteria. Extend this analysis for the i.i.d. Rician fading channel.

1. Does the diversity order change for repetition coding over a time diversity channel with the $L$ branches i.i.d. Rician distributed?
2. What is the new code design criterion, analogous to product distance, based on the pairwise error probability analysis?

**Exercise 3.19** In this exercise we study the performance of space-time codes (the subject of Section 3.3.2) in the presence of multiple receive antennas.

1. Derive, as an extension of (3.83), the pairwise error probability for space-time codes with $n_r$ receive antennas.
2. Assuming that the channel *matrix* has i.i.d. Rayleigh components, derive, as an extension of (3.86), a simple upper bound for the pairwise error probability.
3. Conclude that the code design criterion remains *unchanged* with multiple receive antennas.

**Exercise 3.20** We have studied the performance of the Alamouti scheme in a channel with two transmit and one receive antenna. Suppose now we have an additional receive antenna. Derive the ML detector for the symbols based on the received signals at both receive antennas. Show that the scheme effectively provides two independent scalar channels. What is the gain of each of the channels?

**Exercise 3.21** In this exercise we study some expressions for error probabilities that arise in Section 3.3.3.

1. Verify Eqs. (3.93) and (3.94). In which SNR range is (3.93) smaller than (3.94)?
2. Repeat the derivation of (3.93) and (3.94) for a general target rate of $R$ bits/s/Hz (suppose that $R$ is an integer). How does the SNR range in which the spatial multiplexing scheme performs better depend on $R$?

**Exercise 3.22** In Section 3.3.3, the performance comparison between the spatial multiplexing scheme and the Alamouti scheme is done for PAM symbols. Extend the comparison to QAM symbols with the target data rate $R$ bits/s/Hz (suppose that $R \geq 4$ is an even integer).

**Exercise 3.23** In the text, we have developed code design criteria for pure time diversity and pure spatial diversity scenarios. In some wireless systems, one can get both time and spatial diversity simultaneously, and we want to develop a code design criterion for that. More specifically, consider a channel with $L$ transmit antennas and 1 receive antenna. The channel remains constant over blocks of $k$ symbol times, but changes to an independent realization every $k$ symbols (as a result of interleaving, say). The channel is assumed to be independent across antennas. All channel gains are Rayleigh distributed.

1. What is the maximal diversity gain that can be achieved by coding over $n$ such blocks?
2. Develop a pairwise code design criterion over this channel. Show how this criterion reduces to the special cases we have derived for pure time and pure spatial diversity.

**Exercise 3.24** A mobile having a single receive antenna sees a Rayleigh flat fading channel

$$y[m] = h[m]x[m] + w[m],$$

where $w[m] \sim \mathcal{CN}(0, N_0)$ and i.i.d. and $\{h[m]\}$ is a complex circular symmetric stationary Gaussian process with a given correlation function $R[m]$ which is monotonically decreasing with $m$. (Recall that $R[m]$ is defined to be $\mathbb{E}[h[0]h[m]^*]$.)

1. Suppose now we want to put an extra antenna on the mobile at a separation $d$. Can you determine, from the information given so far, the joint distribution of the fading gains the two antennas see at a particular symbol time? If so, compute it. If not, specify any additional information you have to assume and then compute it.
2. We transmit uncoded BPSK symbols from the base-station to the mobile with dual antennas. Give an expression for the average error probability for the ML detector.
3. Give a back-of-the-envelope approximation to the high SNR error probability, making explicit the effect of the correlation of the channel gains across antennas. What is the diversity gain from having two antennas in the correlated case? How does the error probability compare to the case when the fading gains are assumed to be independent across antennas? What is the effect of increasing the antenna separation $d$?

**Exercise 3.25** Show that full diversity can still be obtained with the maximum likelihood sequence equalizer in Section 3.4.2 even when the channel taps $h_\ell$ have different variances (but are still independent). You can use a heuristic argument based on typical error analysis.

**Exercise 3.26** Consider the maximum likelihood sequence detection described in Section 3.4.2. We computed the achieved diversity gain but did not compute an explicit bound on the error probability on detecting each of the symbol $x[m]$. Below you can assume that BPSK modulation is used for the symbols.

1. Suppose $N = L$. Find a bound on the error probability of the MLSD incorrectly detecting $x[0]$. *Hint*: finding the worst-case pairwise error probability does not require much calculation, but you should be a little careful in applying the union bound.
2. Use your result to estimate the coding gain over the scheme that completely avoids ISI by sending a symbol every $L$ symbol times. How does the coding gain depend on $L$?
3. Extend your analysis to general block length $N \geq L$ and the detection of $x[m]$ for $m \leq N - L$.

**Exercise 3.27** Consider the equalization problem described in Section 3.4.2. We studied the performance of MLSD. In this exercise, we will look at the performance of a linear equalizer. For simplicity, suppose $N = L = 2$.

1. Over the two symbol times (time 0 and time 1), one can think of the ISI channel as a $2 \times 2$ MIMO channel between the input and output symbols. Identify the channel matrix $\mathbf{H}$.
2. The MIMO point of view suggests using, as an alternative to MLSD, the zero-forcing (decorrelating) receiver to detect $x[0]$ based on completely inverting the

channel. How much diversity gain can this equalizer achieve? How does it compare to the performance of MLSD?

**Exercise 3.28** Consider a multipath channel with $L$ i.i.d. Rayleigh faded taps. Let $\tilde{h}_n$ be the complex gain of the $n$th carrier in the OFDM modulation at a particular time. Compute the joint statistics of the gains and lend evidence to the statement that the gains of the carriers separated by more than the coherence bandwidths are approximately independent.

**Exercise 3.29** Argue that for typical wireless channels, the delay spread is much less than the coherence time. What are the implications of this observation on: (1) an OFDM system; (2) a direct-sequence spread-spectrum system with Rake combining? (There may be multiple implications in each case.)

**Exercise 3.30** Communication takes place at passband over a bandwidth $W$ around a carrier frequency of $f_c$. Suppose the baseband equivalent discrete-time model has a finite number of taps. We use OFDM modulation. Let $\tilde{h}_n[i]$ be the complex gain for the $n$th carrier and the $i$th OFDM symbol. We typically assume there are a large number of reflectors so that the tap gains of the discrete-time model can be modeled as Gaussian distributed, but suppose we do not make this assumption here. Only relying on natural assumptions on $f_c$ and $W$, argue the following. State your assumptions on $f_c$ and $W$ and make your argument as clear as possible.
1. At a fixed symbol time $i$, the $\tilde{h}_n[i]$ are identically distributed across the carriers.
2. More generally, the *processes* $\{\tilde{h}_n[i]\}_n$ have the same statistics for different $n$.

**Exercise 3.31** Show that the square-law combiner (given by (3.147)) is the optimal non-coherent ML detector for a channel with i.i.d. Rayleigh faded branches, and analyze the non-coherent error probability performance (i.e., verify (3.148)).

**Exercise 3.32** Consider the problem of Rake combining under channel measurement uncertainty, discussed in Section 3.4.3. Assume a channel with $L$ i.i.d. Rayleigh faded branches. Suppose the channel estimation is as given in Eqs. (3.152) and (3.153). We communicate using binary orthogonal signaling. The receive is coherent with the channel estimates used in place of the true channel gains $h_\ell$. It is not easy to compute explicitly the error probability of this detector, but through either an approximate analysis, numerical computation or simulation, get an idea of its performance as a function of $L$. In particular, give evidence supporting the intuitive statement that, when $L \gg K\mathcal{E}/N_0$, the performance of this detector is very poor.

**Exercise 3.33** We have studied coherent performance of antipodal signaling of the Rake receiver in Section 3.4.3. Now consider binary orthogonal modulation: we either transmit $\mathbf{x}_A$ or $\mathbf{x}_B$, which are both orthogonal and their shifts are also orthogonal with each other. Calculate the error probability with the coherent Rake (i.e., verify (3.149)).