CHAPTER

9 MIMO III: diversity–multiplexing tradeoff and universal space-time codes

In the previous chapter, we analyzed the performance benefits of MIMO communication and discussed architectures that are designed to reap those benefits. The focus was on the *fast fading* scenario. The story on *slow fading* MIMO channels is more complex. While the communication capability of a fast fading channel can be described by a single *number*, its capacity, that of a slow fading channel has to be described by the outage probability curve $p_{\text{out}}(\cdot)$, as a function of the target rate. This curve is in essence a tradeoff between the data rate and error probability. Moreover, in addition to the power and degree-of-freedom gains in the fast fading scenario, multiple antennas provide a diversity gain in the slow fading scenario as well. A clear characterization of the performance benefits of multiple antennas in slow fading channels and the design of good space-time coding schemes that reap those benefits are the subjects of this chapter.

The outage probability curve $p_{\text{out}}(\cdot)$ is the natural benchmark for evaluating the performance of space-time codes. However, it is difficult to characterize analytically the outage probability curves for MIMO channels. We develop an approximation that captures the dual benefits of MIMO communication in the high SNR regime: increased data rate (via an increase in the spatial degrees of freedom or, equivalently, the multiplexing gain) and increased reliability (via an increase in the diversity gain). The dual benefits are captured as a fundamental *tradeoff* between these two types of gains.[1] We use the optimal *diversity–multiplexing tradeoff* as a benchmark to compare the various space-time schemes discussed previously in the book. The tradeoff curve also suggests how optimal space-time coding schemes should look. A powerful idea for the design of tradeoff-optimal schemes is *universality*, which we discuss in the second part of the chapter.

We have studied an approach to space-time code design in Chapter 3. Codes designed using that approach have small error probabilities, averaged over

[1] The careful reader will note that we saw an inkling of the tension between these two types of gains in our study of the 2×2 MIMO Rayleigh fading channel in Chapter 3.

the distribution of the fading channel gains. The drawback of the approach is that the performance of the designed codes may be sensitive to the supposed fading distribution. This is problematic, since, as we mentioned in Chapter 2, accurate statistical modeling of wireless channels is difficult. The outage formulation, however, suggests a different approach. The operational interpretation of the outage performance is based on the existence of *universal codes*: codes that *simultaneously* achieve reliable communication over *every* MIMO channel that is not in outage. Such codes are robust from an engineering point of view: they achieve the best possible outage performance for every fading distribution. This result motivates a universal code design criterion: instead of using the pairwise error probability averaged over the fading distribution of the channel, we consider the *worst-case* pairwise error probability over all channels that are not in outage. Somewhat surprisingly, the universal code-design criterion is closely related to the product distance, which is obtained by averaging over the Rayleigh distribution. Thus, the product distance criterion, while seemingly tailored for the Rayleigh distribution, is actually more fundamental. Using universal code design ideas, we construct codes that achieve the optimal diversity–multiplexing tradeoff.

Throughout this chapter, the receiver is assumed to have perfect knowledge of the channel matrix while the transmitter has none.

9.1 Diversity–multiplexing tradeoff

In this section, we use the outage formulation to characterize the performance capability of slow fading MIMO channels in terms of a tradeoff between diversity and multiplexing gains. This tradeoff is then used as a unified framework to compare the various space-time coding schemes described in this book.

9.1.1 Formulation

When we analyzed the performance of communication schemes in the slow fading scenario in Chapters 3 and 5, the emphasis was on the *diversity gain*. In this light, a key measure of the performance capability of a slow fading channel is the *maximum diversity gain* that can be extracted from it. For example, a slow i.i.d. Rayleigh faded MIMO channel with n_t transmit and n_r receive antennas has a maximum diversity gain of $n_t \cdot n_r$: i.e., for a fixed target rate R, the outage probability $p_{out}(R)$ decays like $1/\mathsf{SNR}^{n_t n_r}$ at high SNR.

On the other hand, we know from Chapter 7 that the key performance benefit of a *fast fading* MIMO channel is the spatial multiplexing capability it provides through the additional degrees of freedom. For example, the capacity of an i.i.d. Rayleigh fading channel scales like $n_{\min} \log \mathsf{SNR}$, where $n_{\min} := \min(n_t, n_r)$ is the number of spatial degrees of freedom in the channel. This fast fading (ergodic) capacity is achieved by averaging over the variation of the channel over time. In the slow fading scenario, no such averaging is possible and one cannot communicate at this rate reliably. Instead, the information rate allowed through the channel is a random variable fluctuating around the fast fading capacity. Nevertheless, one would still expect to be able to benefit from the increased degrees of freedom even in the slow fading scenario. Yet the maximum diversity gain provides no such indication; for example, both an $n_t \times n_r$ channel and an $n_t n_r \times 1$ channel have the same maximum diversity gain and yet one would expect the former to allow better spatial multiplexing than the latter. One needs something more than the maximum diversity gain to capture the spatial multiplexing benefit.

Observe that to achieve the maximum diversity gain, one needs to communicate at a *fixed* rate R, which becomes vanishingly small compared to the fast fading capacity at high SNR (which grows like $n_{\min} \log \mathsf{SNR}$). Thus, one is actually sacrificing all the spatial multiplexing benefit of the MIMO channel to maximize the reliability. To reclaim some of that benefit, one would instead want to communicate at a rate $R = r \log \mathsf{SNR}$, which is a *fraction* of the fast fading capacity. Thus, it makes sense to formulate the following *diversity–multiplexing tradeoff* for a slow fading channel.

A diversity gain $d^*(r)$ is achieved at multiplexing gain r if

$$R = r \log \mathsf{SNR} \tag{9.1}$$

and

$$p_{\text{out}}(R) \approx \mathsf{SNR}^{-d^*(r)}, \tag{9.2}$$

or more precisely,

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log p_{\text{out}}(r \log \mathsf{SNR})}{\log \mathsf{SNR}} = -d^*(r). \tag{9.3}$$

The curve $d^*(\cdot)$ is the diversity–multiplexing tradeoff of the slow fading channel.

The above tradeoff characterizes the slow fading performance limit of the *channel*. Similarly, we can formulate a diversity–multiplexing tradeoff for any space-time coding *scheme*, with outage probabilities replaced by *error probabilities*.

A space-time coding scheme is a family of codes, indexed by the signal-to-noise ratio SNR. It attains a multiplexing gain r and a diversity gain d if the data rate scales as

$$R = r \log \mathsf{SNR} \tag{9.4}$$

and the error probability scales as

$$p_{\mathrm{e}} \approx \mathsf{SNR}^{-d}, \tag{9.5}$$

i.e.,

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log p_{\mathrm{e}}}{\log \mathsf{SNR}} = -d. \tag{9.6}$$

The diversity–multiplexing tradeoff formulation may seem abstract at first sight. We will now go through a few examples to develop a more concrete feel. The tradeoff performance of specific coding schemes will be analyzed and we will see how they perform compared to each other and to the optimal diversity–multiplexing tradeoff of the channel. For concreteness, we use the i.i.d. Rayleigh fading model. In Section 9.2, we will describe a general approach to tradeoff-optimal space-time code based on universal coding ideas.

9.1.2 Scalar Rayleigh channel

PAM and QAM

Consider the scalar slow fading Rayleigh channel,

$$y[m] = hx[m] + w[m], \tag{9.7}$$

with the additive noise i.i.d. $\mathcal{CN}(0, 1)$ and the power constraint equal to SNR. Suppose h is $\mathcal{CN}(0, 1)$ and consider uncoded communication using PAM with a data rate of R bits/s/Hz. We have done the error probability analysis in Section 3.1.2 for $R = 1$; for general R, the analysis is similar. The average error probability is governed by the minimum distance between the PAM points. The constellation ranges from approximately $-\sqrt{\mathsf{SNR}}$ to $+\sqrt{\mathsf{SNR}}$, and since there are 2^R constellation points, the minimum distance is approximately

$$D_{\min} \approx \frac{\sqrt{\mathsf{SNR}}}{2^R}, \tag{9.8}$$

and the error probability at high SNR is approximately (cf. (3.28)),

$$p_e \approx \frac{1}{2}\left(1-\sqrt{\frac{D_{\min}^2}{4+D_{\min}^2}}\right) \approx \frac{1}{D_{\min}^2} \approx \frac{2^{2R}}{\mathsf{SNR}}. \tag{9.9}$$

By setting the data rate $R = r\log \mathsf{SNR}$, we get

$$p_e \approx \frac{1}{\mathsf{SNR}^{1-2r}}, \tag{9.10}$$

yielding a diversity–multiplexing tradeoff of

$$\boxed{d_{\text{pam}}(r) = 1-2r, \qquad r \in \left[0, \frac{1}{2}\right].} \tag{9.11}$$

Note that in the approximate analysis of the error probability above, we focus on the scaling of the error probability with the SNR and the data rate but are somewhat careless with constant multipliers: they do not matter as far as the diversity–multiplexing tradeoff is concerned.

We can repeat the analysis for QAM with data rate R. There are now $2^{R/2}$ constellation points in each of the real and imaginary dimensions, and hence the minimum distance is approximately

$$D_{\min} \approx \frac{\sqrt{\mathsf{SNR}}}{2^{R/2}}, \tag{9.12}$$

and the error probability at high SNR is approximately

$$p_e \approx \frac{2^R}{\mathsf{SNR}}, \tag{9.13}$$

yielding a diversity–multiplexing tradeoff of

$$\boxed{d_{\text{qam}}(r) = 1-r, \qquad r \in [0, 1].} \tag{9.14}$$

The tradeoff curves are plotted in Figure 9.1.

Let us relate the two endpoints of a tradeoff curve to notions that we already know. The value $d_{\max} := d(0)$ can be interpreted as the SNR exponent that describes how fast the error probability can be decreased with the SNR for a *fixed* data rate; this is the *classical diversity gain* of a scheme. It is 1 for both PAM and QAM. The decrease in error probability is due to an increase in $D_{\min}$. This is illustrated in Figure 9.2.

In a dual way, the value $r_{\max}$ for which $d(r_{\max}) = 0$ describes how fast the data rate can be increased with the SNR for a *fixed* error probability. This number can be interpreted as the number of (complex) degrees of freedom that are exploited by the scheme. It is 1 for QAM but only $1/2$ for PAM.

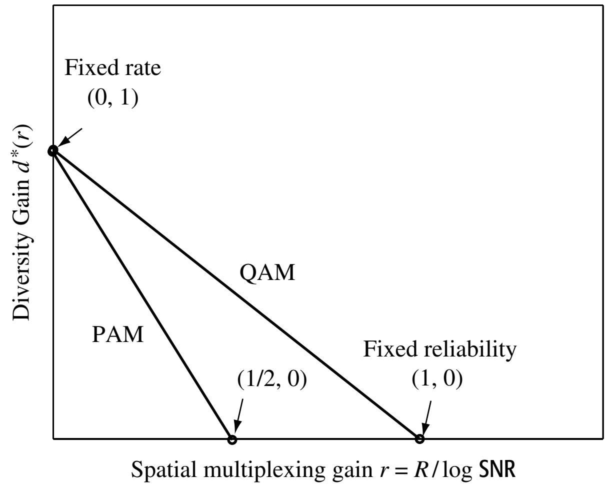

Figure 9.1 Tradeoff curves for the single antenna slow fading Rayleigh channel.

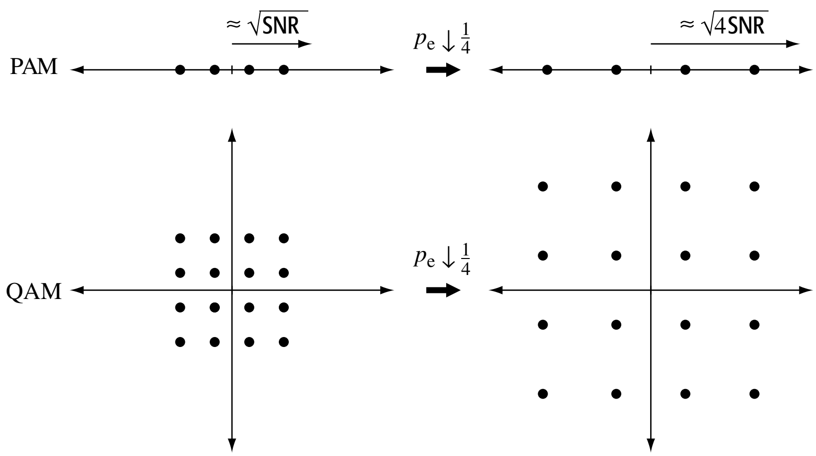

Figure 9.2 Increasing the SNR by 6 dB decreases the error probability by 1/4 for both PAM and QAM due to a doubling of the minimum distance.

This is consistent with our observation in Section 3.1.3 that PAM uses only half the degrees of freedom of QAM. The increase in data rate is due to the packing of more constellation points for a given $D_{\min}$. This is illustrated in Figure 9.3.

The two endpoints represent two extreme ways of using the increase in the resource (SNR): increasing the reliability for a fixed data rate, or increasing the data rate for a fixed reliability. More generally, we can simultaneously increase the data rate (positive multiplexing gain r) and increase the reliability (positive diversity gain $d > 0$) but there is a tradeoff between how much of each type of gain we can get. The diversity–multiplexing curve describes this tradeoff. Note that the classical diversity gain only describes the rate of decay of the error probability for a fixed data rate, but does not provide any information on how well a scheme exploits the available degrees of freedom. For example, PAM and QAM have the same classical diversity

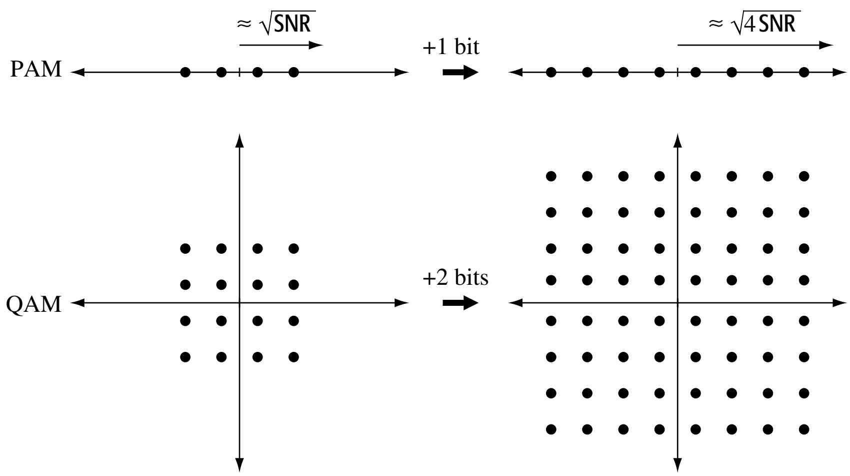

Figure 9.3 Increasing the SNR by 6 dB increases the data rate for QAM by 2 bits/s/Hz but only increases the data rate of PAM by 1 bit/s/Hz.

gain, even though clearly QAM is more efficient in exploiting the available degrees of freedom. The tradeoff curve, by treating error probability and data rate in a symmetrical manner, provides a more complete picture. We see that in terms of their tradeoff curves, QAM is indeed superior to PAM (see Figure 9.1).

Optimal tradeoff

So far, we have considered the tradeoff between diversity and multiplexing in the context of two specific schemes: uncoded PAM and QAM. What is the fundamental diversity–multiplexing tradeoff of the scalar channel itself? For the slow fading Rayleigh channel, the outage probability at a target data rate $R = r \log \mathsf{SNR}$ is

$$\begin{aligned} p_{\text{out}} &= \mathbb{P}\{\log(1+|h|^2\mathsf{SNR}) < r\log \mathsf{SNR}\} \\ &= \mathbb{P}\left\{|h|^2 < \frac{\mathsf{SNR}^r - 1}{\mathsf{SNR}}\right\} \\ &\approx \frac{1}{\mathsf{SNR}^{1-r}}, \end{aligned} \tag{9.15}$$

at high SNR. In the last step, we used the fact that for Rayleigh fading, $\mathbb{P}\{|h|^2 < \epsilon\} \approx \epsilon$ for small ϵ. Thus

$$\boxed{d^*(r) = 1 - r, \qquad r \in [0, 1].} \tag{9.16}$$

Hence, the uncoded QAM scheme trades off diversity and multiplexing gains optimally.

The tradeoff between diversity and multiplexing gains can be viewed as a *coarser* way of capturing the fundamental tradeoff between error probability and data rate over a fading channel at high SNR. Even very simple, low-complexity schemes can trade off optimally in this coarser context (the uncoded QAM achieved the tradeoff for the Rayleigh slow fading channel). To achieve the *exact* tradeoff between outage probability and data rate, we need to code over long block lengths, at the expense of higher complexity.

9.1.3 Parallel Rayleigh channel

Consider the slow fading parallel channel with i.i.d. Rayleigh fading on each sub-channel:

$$y_\ell[m] = h_\ell x_\ell[m] + w_\ell[m], \qquad \ell = 1, \ldots, L. \tag{9.17}$$

Here, the w_ℓ are i.i.d. $\mathcal{CN}(0, 1)$ additive noise and the transmit power per sub-channel is constrained by SNR. We have seen that L Rayleigh faded sub-channels provide a (classical) diversity gain equal to L (cf. Section 3.2 and Section 5.4.4): this is an L-fold improvement over the basic single antenna slow fading channel. In the parlance we introduced in the previous section, this says that $d^*(0) = L$. How about the diversity gain at any positive multiplexing rate?

Suppose the target data rate is $R = r \log \mathsf{SNR}$ bits/s/Hz per sub-channel. The optimal diversity $d^*(r)$ can be calculated from the rate of decay of the outage probability with increasing SNR. For the i.i.d. Rayleigh fading parallel channel, the outage probability at rate per sub-channel $R = r \log \mathsf{SNR}$ is (cf. (5.83))

$$p_{\text{out}} = \mathbb{P}\left\{ \sum_{\ell=1}^{L} \log(1 + |h_\ell|^2 \mathsf{SNR}) < Lr \log \mathsf{SNR} \right\}. \tag{9.18}$$

Outage typically occurs when each of the sub-channels cannot support the rate R (Exercise 9.1): so we can write

$$p_{\text{out}} \approx (\mathbb{P}\{\log(1 + |h_1|^2 \mathsf{SNR}) < r \log \mathsf{SNR}\})^L \approx \frac{1}{\mathsf{SNR}^{L(1-r)}}. \tag{9.19}$$

So, the optimal diversity–multiplexing tradeoff for the parallel channel with L diversity branches is

$$\boxed{d^*(r) = L(1 - r), \qquad r \in [0, 1],} \tag{9.20}$$

an L-fold gain over the scalar single antenna performance (cf. (9.16)) at *every* multiplexing gain r; this performance is illustrated in Figure 9.4.

One particular scheme is to transmit the same QAM symbol over the L sub-channels; the repetition converts the parallel channel into a scalar channel with squared amplitude $\sum_\ell |h_\ell|^2$, but with the rate reduced by a factor of $1/L$.

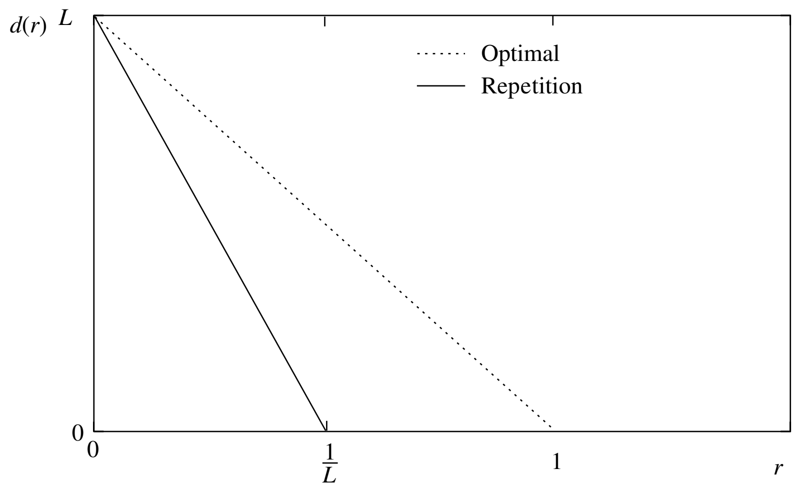

Figure 9.4 The diversity–multiplexing tradeoff of the i.i.d. Rayleigh fading parallel channel with L sub-channels together with that of the repetition scheme.

The diversity–multiplexing tradeoff achieved by this scheme can be computed to be

$$d_{\text{rep}}(r) = L(1 - Lr), \qquad r \in \left[0, \frac{1}{L}\right], \tag{9.21}$$

(Exercise 9.2). The classical diversity gain $d_{\text{rep}}(0)$ is L, the full diversity of the parallel channel, but the number of degrees of freedom per sub-channel is only $1/L$, due to the repetition.

9.1.4 MISO Rayleigh channel

Consider the n_t transmit and single receive antenna MISO channel with i.i.d. Rayleigh coefficients:

$$y[m] = \mathbf{h}^*\mathbf{x}[m] + w[m]. \tag{9.22}$$

As usual, the additive noise $w[m]$ is i.i.d. $\mathcal{CN}(0, 1)$ and there is an overall transmit power constraint of SNR. We have seen that the Rayleigh fading MISO channel with n_t transmit antennas provides the (classical) diversity gain of n_t (cf. Section 3.3.2 and Section 5.4.3). By how much is the diversity gain increased at a positive multiplexing rate of r?

We can answer this question by looking at the outage probability at target data rate $R = r \log \mathsf{SNR}$ bits/s/Hz:

$$p_{\text{out}} = \mathbb{P}\left\{\log\left(1 + \|\mathbf{h}\|^2 \frac{\mathsf{SNR}}{n_t}\right) < r \log \mathsf{SNR}\right\}. \tag{9.23}$$

Now $\|\mathbf{h}\|^2$ is a χ^2 random variable with $2n_t$ degrees of freedom and we have seen that $\mathbb{P}\{\|\mathbf{h}\|^2 < \epsilon\} \approx \epsilon^{n_t}$ (cf. (3.44)). Thus, p_{out} decays as $\mathsf{SNR}^{-n_t(1-r)}$ with increasing SNR and the optimal diversity–multiplexing tradeoff for the i.i.d. Rayleigh fading MISO channel is

$$\boxed{d^*(r) = n_t(1-r), \qquad r \in [0, 1].} \tag{9.24}$$

Thus the MISO channel provides an n_t-fold increase in diversity at all multiplexing gains.

In the case of $n_t = 2$, we know that the Alamouti scheme converts the MISO channel into a scalar channel with the same outage behavior as the original MISO channel. Hence, if we use QAM symbols in conjunction with the Alamouti scheme, we achieve the diversity–multiplexing tradeoff of the MISO channel. In contrast, the repetition scheme that transmits the same QAM symbol from each of the two transmit antennas one at a time achieves a diversity–multiplexing tradeoff curve of

$$d_{\text{rep}}(r) = 2(1-2r), \qquad r \in \left[0, \frac{1}{2}\right]. \tag{9.25}$$

The tradeoff curves of these schemes as well as that of the 2×1 MISO channel are shown in Figure 9.5.

9.1.5 2 × 2 MIMO Rayleigh channel

Four schemes revisited

In Section 3.3.3, we analyzed the (classical) diversity gains and degrees of freedom utilized by four schemes for the 2×2 i.i.d. Rayleigh fading

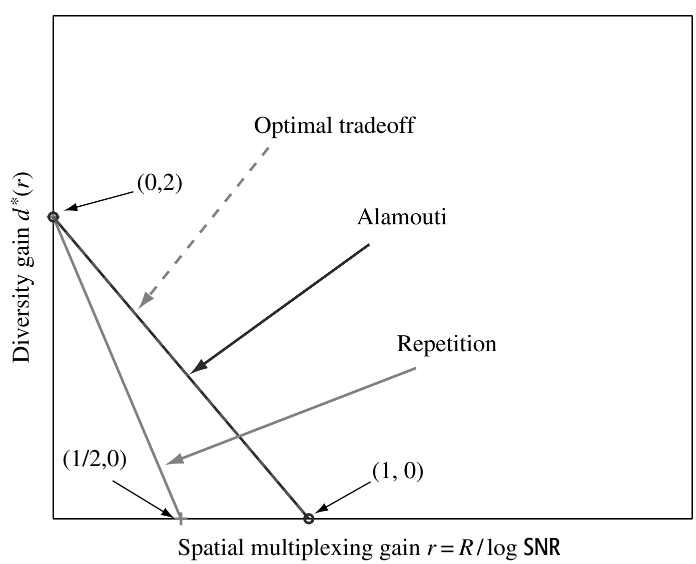

Figure 9.5 The diversity–multiplexing tradeoff of the 2×1 i.i.d. Rayleigh fading MISO channel along with those of two schemes.

Table 9.1 A summary of the performance of the four schemes for the 2 × 2 channel.

	Classical diversity gain	Degrees of freedom utilized	D–M tradeoff
Repetition	4	1/2	$4-8r,\ r\in[0,1/2]$
Alamouti	4	1	$4-4r,\ r\in[0,1]$
V-BLAST (ML)	2	2	$2-r,\ r\in[0,2]$
V-BLAST (nulling)	1	2	$1-r/2,\ r\in[0,2]$
Channel itself	4	2	$4-3r,\ r\in[0,1]$ $2-r,\ r\in[1,2]$

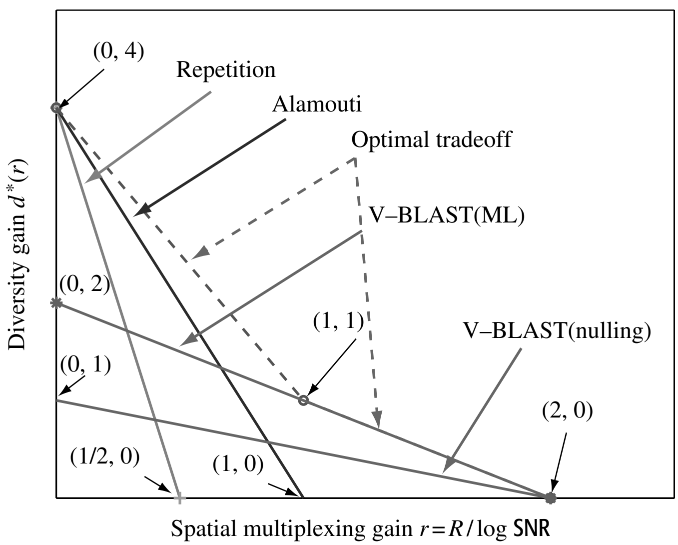

Figure 9.6 The diversity–multiplexing tradeoff of the 2 × 2 i.i.d. Rayleigh fading MIMO channel along with those of four schemes.

MIMO channel (with the results summarized in Summary 3.2). The diversity–multiplexing tradeoffs of these schemes when used in conjunction with uncoded QAM can be computed as well; they are summarized in Table 9.1 and plotted in Figure 9.6. The classical diversity gains and degrees of freedom utilized correspond to the endpoints of these curves.

The repetition, Alamouti and V-BLAST with nulling schemes all convert the MIMO channel into scalar channels for which the diversity–multiplexing tradeoffs can be computed in a straightforward manner (Exercises 9.3, 9.4 and 9.5). The diversity–multiplexing tradeoff of V-BLAST with ML decoding can be analyzed starting from the pairwise error probability between two codewords $\mathbf{x}_A$ and $\mathbf{x}_B$ (with average transmit energy normalized to 1):

$$\mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_B|\mathbf{H}\} \le \frac{16}{\mathsf{SNR}^2\|\mathbf{x}_A-\mathbf{x}_B\|^4}, \qquad (9.26)$$

(cf. 3.92). Each codeword is a pair of QAM symbols transmitted on the two antennas, and hence the distance between the two closest codewords is that between two adjacent constellation points in one of the QAM constellation, i.e., $\mathbf{x}_A$ and $\mathbf{x}_B$ differ only in one of the two QAM symbols. With a total data rate of R bits/s/Hz, each QAM symbol carries $R/2$ bits, and hence each of the I and Q channels carries $R/4$ bits. The distance between two adjacent constellation points is of the order of $1/2^{R/4}$. Thus, the worst-case pairwise error probability is of the order

$$\frac{16 \cdot 2^R}{\mathsf{SNR}^2} = 16 \cdot \mathsf{SNR}^{-(2-r)}, \tag{9.27}$$

where the data rate $R = r \log \mathsf{SNR}$. This is the worst-case pairwise error probability, but Exercise 9.6 shows that the overall error probability is also of the same order. Hence, the diversity–multiplexing tradeoff of V-BLAST with ML decoding is

$$d(r) = 2 - r \qquad r \in [0, 2]. \tag{9.28}$$

As already remarked in Section 3.3.3, the (classical) diversity gain and the degrees of freedom utilized are not always sufficient to say which scheme is best. For example, the Alamouti scheme has a higher (classical) diversity gain than V-BLAST but utilizes fewer degrees of freedom. The tradeoff curves, in contrast, provide a clear basis for the comparison. We see that which scheme is better depends on the target diversity gain (error probability) of the operating point: for smaller target diversity gains, V-BLAST is better than the Alamouti scheme, while the situation reverses for higher target diversity gains.

Optimal tradeoff

Do any of the four schemes actually achieve the optimal tradeoff of the 2×2 channel? The tradeoff curve of the 2×2 i.i.d. Rayleigh fading MIMO channel turns out to be piecewise linear joining the points (0, 4), (1, 1) and (2, 0) (also shown in Figure 9.6). Thus, all of the schemes are tradeoff-suboptimal, except for V-BLAST with ML, which is optimal but only for $r > 1$.

The endpoints of the optimal tradeoff curve are (0, 4) and (2, 0), consistent with the fact that the 2×2 MIMO channel has a maximum diversity gain of 4 and 2 degrees of freedom. More interestingly, unlike all the tradeoff curves we have computed before, this curve is not a line but piecewise linear, consisting of two linear segments. V-BLAST with ML decoding sends two symbols per symbol time with (classical) diversity of 2 for each symbol, and achieves the gentle part, $2 - r$, of this curve. But what about the steep part, $4 - 3r$? Intuitively, there should be a scheme that sends 4 symbols over 3 symbol times (with a rate of $4/3$ symbols/s/Hz) and achieves the full diversity gain of 4. We will see such a scheme in Section 9.2.4.

9.1.6 $n_t \times n_r$ MIMO i.i.d. Rayleigh channel

Optimal tradeoff

Consider the $n_t \times n_r$ MIMO channel with i.i.d. Rayleigh faded gains. The optimal diversity gain at a data rate $r \log \mathsf{SNR}$ bits/s/Hz is the rate at which the outage probability (cf. (8.81)) decays with SNR:

$$p_{\text{out}}^{\text{mimo}}(r \log \mathsf{SNR}) = \min_{\mathbf{K}_x:\text{Tr}[\mathbf{K}_x]\le \mathsf{SNR}} \mathbb{P}\{\log\det(\mathbf{I}_{n_r} + \mathbf{H}\mathbf{K}_x\mathbf{H}^*) < r\log\mathsf{SNR}\}. \quad (9.29)$$

While the optimal covariance matrix $\mathbf{K}_x$ depends on the SNR and the data rate, we argued in Section 8.4 that the choice of $\mathbf{K}_x = \mathsf{SNR}/n_t \mathbf{I}_{n_t}$ is often used as a good approximation to the actual outage probability. In the coarser scaling of the tradeoff curve formulation, that argument can be made precise: the decay rate of the outage probability in (9.29) is the same as when the covariance matrix is the scaled identity. (See Exercise 9.8.) Thus, for the purpose of identifying the optimal diversity gain at a multiplexing rate r it suffices to consider the expression in (8.85):

$$p_{\text{out}}^{\text{iid}}(r \log \mathsf{SNR}) = \mathbb{P}\left\{\log\det\left(\mathbf{I}_{n_r} + \frac{\mathsf{SNR}}{n_t}\mathbf{H}\mathbf{H}^*\right) < r\log\mathsf{SNR}\right\}. \quad (9.30)$$

By analyzing this expression, the diversity–multiplexing tradeoff of the $n_t \times n_r$ i.i.d. Rayleigh fading channel can be computed. It is the piecewise linear curve joining the points

$$(k, (n_t - k)(n_r - k)), k = 0, \ldots, n_{\min}, \quad (9.31)$$

as shown in Figure 9.7.

The tradeoff curve summarizes succinctly the performance capability of the slow fading MIMO channel. At one extreme where $r \to 0$, the maximal diversity gain $n_t \cdot n_r$ is achieved, at the expense of very low multiplexing gain. At the other extreme where $r \to n_{\min}$, the full degrees of freedom are attained. However, the system is now operating very close to the fast fading capacity and there is little protection against the randomness of the slow fading channel; the diversity gain is approaching 0. The tradeoff curve bridges between the two extremes and provides a more complete picture of the slow fading performance capability than the two extreme points. For example, adding one transmit and one receive antenna to the system increases the degrees of freedom $\min(n_t, n_r)$ by 1; this corresponds to increasing the maximum possible multiplexing gain by 1. The tradeoff curve gives a more refined picture of the system benefit: for *any* diversity requirement d, the supported multiplexing gain is increased by 1.

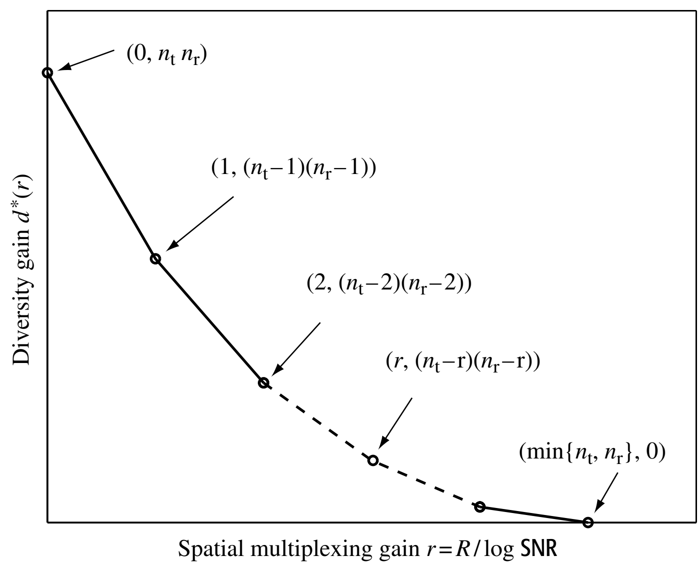

Figure 9.7 Diversity–multiplexing tradeoff, $d^*(r)$ for the i.i.d. Rayleigh fading channel.

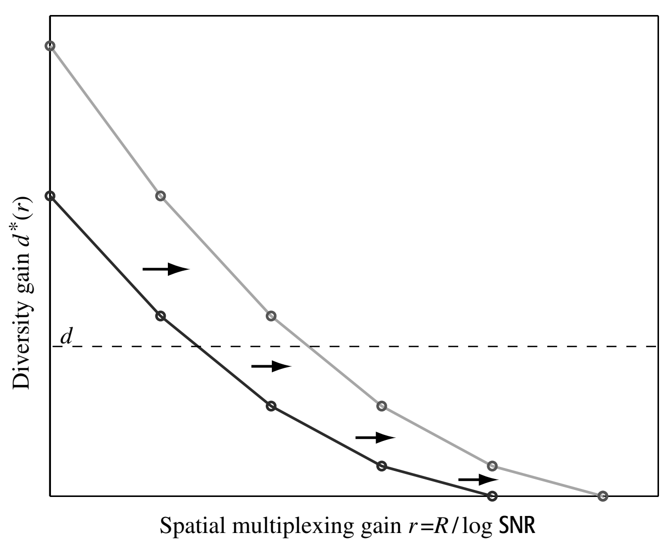

Figure 9.8 Adding one transmit and one receive antenna increases spatial multiplexing gain by 1 at each diversity level.

This is because the entire tradeoff curve is shifted by 1 to the right; see Figure 9.8.

The optimal tradeoff curve is based on the outage probability, so in principle arbitrarily large block lengths are required to achieve the optimal tradeoff curve. However, it has been shown that, in fact, space-time codes of block length $l = n_t + n_r - 1$ achieve the curve. In Section 9.2.4, we will see a scheme that achieves the tradeoff curve but requires arbitrarily large block lengths.

Geometric interpretation

To provide more intuition let us consider the geometric picture behind the optimal tradeoff for integer values of r. The outage probability is given by

$$p_{\text{out}}(r \log \mathsf{SNR}) = \mathbb{P}\left\{\log\det\left(\mathbf{I}_{n_r} + \frac{\mathsf{SNR}}{n_t}\mathbf{H}\mathbf{H}^*\right) < r\log\mathsf{SNR}\right\}$$
$$= \mathbb{P}\left\{\sum_{i=1}^{n_{\min}} \log\left(1 + \frac{\mathsf{SNR}}{n_t}\lambda_i^2\right) < r\log\mathsf{SNR}\right\}, \qquad (9.32)$$

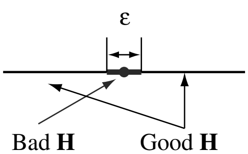

Figure 9.9 Geometric picture for the 1 × 1 channel. Outage occurs when $|h|$ is close to 0.

where λ_i are the (random) singular values of the matrix $\mathbf{H}$. There are $n_{\min}$ possible modes for communication but the effectiveness of mode i depends on how large the received signal strength $\mathsf{SNR}\lambda_i^2/n_t$ is for that mode; we can think of a mode as *fully effective* if $\mathsf{SNR}\lambda_i^2/n_t$ is of order SNR and not effective at all when $\mathsf{SNR}\lambda_i^2/n_t$ is of order 1 or smaller.

At low multiplexing gains ($r \to 0$), outage occurs when none of the modes are effective at all; i.e., *all* the squared singular values are small, of the order of $1/\mathsf{SNR}$. Geometrically, this event happens when the channel matrix $\mathbf{H}$ is close to the zero matrix; see Figure 9.9 and 9.10. Since $\sum_i \lambda_i^2 = \sum_{i,j} |h_{ij}|^2$, this event occurs only when *all* of the $n_t n_r$ squared magnitude channel gains, $|h_{ij}|^2$, are small, each on the order of $1/\mathsf{SNR}$. As the channel gains are independent and $\mathbb{P}\{|h_{ij}|^2 < 1/\mathsf{SNR}\} \approx 1/\mathsf{SNR}$, the probability of this event is on the order of $1/\mathsf{SNR}^{n_t n_r}$.

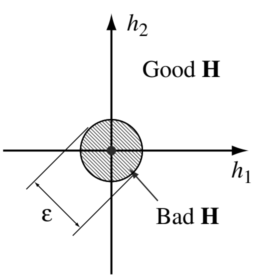

Figure 9.10 Geometric picture for the 1 × 2 channel. Outage occurs when $|h_1|^2 + |h_2|^2$ is close to 0.

Now consider the case when r is a positive integer. The situation is more complicated. For the outage event in (9.32) to occur, there are now many possible combinations of values that the singular values, λ_i, can take on, with modes taking on different shades of effectiveness. However, at high SNR, it can be shown that the *typical* way for outage to occur is when precisely r of the modes are fully effective and the rest completely ineffective. This means the largest r singular values of $\mathbf{H}$ are of order 1, while the rest are of the order $1/\mathsf{SNR}$ or smaller; geometrically, $\mathbf{H}$ is close to a rank r matrix. What is the probability of this event?

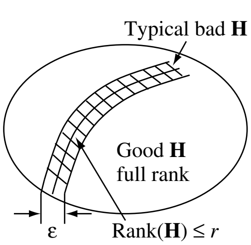

Figure 9.11 Geometric picture for the $n_t \times n_r$ channel at multiplexing gain r (r integer). Outage occurs when the channel matrix $\mathbf{H}$ is close to a rank r matrix.

In the case of $r = 0$, the outage event is when the channel matrix $\mathbf{H}$ is close to a rank 0 matrix. The channel matrix lies in the $n_t n_r$-dimensional space $\mathcal{C}^{n_r \times n_t}$, so for this to occur, there is a collapse in all $n_t n_r$ dimensions. This leads to an outage probability of $1/\mathsf{SNR}^{n_t n_r}$. At general multiplexing gain r (r positive integer), outage occurs when $\mathbf{H}$ is close to $\mathcal{V}_r$, the space of all rank r matrices. This requires a collapse in the component of $\mathbf{H}$ "orthogonal" to $\mathcal{V}_r$. Thus, one would expect the probability of this event to be approximately $1/\mathsf{SNR}^d$, where d is the number of such dimensions.[2] See Figure 9.11. It is

[2] $\mathcal{V}_r$ is not a linear space. So, strictly speaking, we cannot talk about the concept of orthogonal dimensions. However, $\mathcal{V}_r$ is a *manifold*, which means that the neighborhood of every point looks like a Euclidean space of the same dimension. So the notion of orthogonal dimensions (called the "co-dimension" of $\mathcal{V}_r$) still makes sense.

easy to compute d. A $n_r \times n_t$ matrix $\mathbf{H}$ of rank r is described by $rn_t + (n_r - r)r$ parameters: rn_t parameters to specify r linearly independent row vectors of $\mathbf{H}$ and $(n_r - r)r$ parameters to specify the remaining $n_r - r$ rows in terms of linear combinations of the first r row vectors. Hence $\mathcal{V}_r$ is $n_t r + (n_r - r)r$-dimensional and the number of dimensions orthogonal to $\mathcal{V}_r$ in $\mathcal{C}^{n_t n_r}$ is simply

$$n_t n_r - (n_t r + (n_r - r)r) = (n_t - r)(n_r - r).$$

This is precisely the SNR exponent of the outage probability in (9.32).

9.2 Universal code design for optimal diversity–multiplexing tradeoff

The operational interpretation of the outage formulation is based on the existence of universal codes that can achieve arbitrarily small error whenever the channel is not in outage. To achieve such performance, arbitrarily long block lengths and powerful codes are required. In the high SNR regime, we have seen in Chapter 3 that the typical error event is the event that the channel is in a deep fade, where the deep-fade event depends on the channel as well as the scheme. This leads to a natural high SNR relaxation of the universality concept:

> A scheme is *approximately universal* if it is in deep fade only when the channel itself is in outage.

Being approximately universal is sufficient for a scheme to achieve the diversity–multiplexing tradeoff of the channel. Moreover, one can explicitly construct approximately universal schemes of short block lengths. We describe this approach towards optimal diversity–multiplexing tradeoff code design in this section. We start with the scalar channel and progress towards more complex models, culminating in the general $n_t \times n_r$ MIMO channel.

9.2.1 QAM is approximately universal for scalar channels

In Section 9.1.2 we have seen that uncoded QAM achieves the optimal diversity–multiplexing tradeoff of the scalar Rayleigh fading channel. One can obtain a deeper understanding of why this is so via a typical error event analysis. Conditional on the channel gain h, the probability of error of uncoded QAM at data rate R is approximately

$$Q\left(\sqrt{\frac{\mathsf{SNR}}{2}|h|^2 d_{\min}^2}\right), \tag{9.33}$$

where $d_{\min}$ is the minimum distance between two normalized constellation points, given by

$$d_{\min} \approx \frac{1}{2^{R/2}}. \tag{9.34}$$

When $\sqrt{\mathsf{SNR}}|h|d_{\min} \gg 1$, i.e. the separation of the constellation points at the receiver is much larger than the standard deviation of the additive Gaussian noise, errors occur very rarely due to the very rapid drop off of the Gaussian tail probability. Thus, as an order-of-magnitude approximation, errors typically occur due to:

$$\boxed{\text{Deep-fade event} : |h|^2 < \frac{2^R}{\mathsf{SNR}}.} \tag{9.35}$$

This deep-fade event is analogous to that of BPSK in Section 3.1.2. On the other hand, the *channel outage* condition is given by

$$\log\left(1+|h|^2\mathsf{SNR}\right) < R, \tag{9.36}$$

or equivalently

$$|h|^2 < \frac{2^R - 1}{\mathsf{SNR}}. \tag{9.37}$$

At high SNR and high rate, the channel outage condition (9.37) and the deep-fade event of QAM (9.35) coincide. Thus, *typically errors occur for QAM only when the channel is in outage*. Since the optimal diversity–multiplexing tradeoff is determined by the outage probability of the channel, this explains why QAM achieves the optimal tradeoff. (A rigorous proof of the tradeoff optimality of QAM based solely on this typical error event view is carried out in Exercise 9.9, which is the generalization of Exercise 3.3 where we used the typical error event to analyze classical diversity gain.)

In Section 9.1.2, the diversity–multiplexing tradeoff of QAM is computed by averaging the error probability over the Rayleigh fading. It happens to be equal to the optimal tradeoff. The present explanation based on relating the deep-fade event of QAM and the outage condition is more insightful. For one thing, this explanation is in terms of conditions on the channel gain h and has nothing to do with the distribution of h. This means that QAM achieves the optimal diversity–multiplexing tradeoff not only under Rayleigh fading but in fact under *any* channel statistics. This is the true meaning of universality. For example, for a channel with the near-zero behavior of $\mathbb{P}\{|h|^2 < \epsilon\} \approx \epsilon^k$, the optimal diversity–multiplexing tradeoff curve follows directly from (9.15): $d^*(r) = k(1-r)$. Uncoded QAM on this channel can achieve this tradeoff as well.

Note that the approximate universality of QAM depends only on a condition on its normalized minimum distance:

$$\boxed{d^2_{\min} > \frac{1}{2^R}.} \tag{9.38}$$

Any other constellation with this property is also approximately universal (Exercise 9.9).

Summary 9.1 Approximate universality

A scheme is *approximately universal* if it is in deep fade only when the channel itself is in outage.

Being approximately universal is sufficient for a scheme to achieve the diversity–multiplexing tradeoff of the channel.

9.2.2 Universal code design for parallel channels

In Section 3.2.2 we derived design criteria for codes that have a good coding gain while extracting the maximum diversity from the parallel channel. The criterion was derived based on averaging the error probability over the statistics of the fading channel. For example, the i.i.d. Rayleigh fading parallel channel yielded the product distance criterion (cf. Summary 3.1). In this section, we consider instead a *universal* design criterion based on considering the performance of the code over the *worst-case* channel that is not in outage. Somewhat surprisingly, this universal code design criterion reduces to the product distance criterion at high SNR. Using this universal design criterion, we can characterize codes that are approximately universal using the idea of typical error event used in the last section.

Universal code design criterion

We begin with the parallel channel with L diversity branches, focusing on just one time symbol (and dropping the time index):

$$y_\ell = h_\ell x_\ell + w_\ell \tag{9.39}$$

for $\ell = 1, \ldots, L$. Here, as before, the w_ℓ are i.i.d. $\mathcal{CN}(0, 1)$ noise. Suppose the rate of communication is R bits/s/Hz per sub-channel. Each codeword is a vector of length L. The ℓth component of any codeword is transmitted over the ℓth sub-channel in (9.39). Here, a codeword consists of one symbol for each of the L sub-channels; more generally, we can consider coding over multiple symbols for each of the sub-channels as well as coding across the different sub-channels. The derivation of a code design criterion for the more general case is done in Exercise 9.10.

The channels that are not in outage are those whose gains satisfy

$$\sum_{\ell=1}^{L} \log(1+|h_\ell|^2 \mathsf{SNR}) \geq LR. \tag{9.40}$$

As before, SNR is the transmit power constraint per sub-channel.

For a fixed pair of codewords $\mathbf{x}_A, \mathbf{x}_B$, the probability that $\mathbf{x}_B$ is more likely than $\mathbf{x}_A$ when $\mathbf{x}_A$ is transmitted, conditional on the channel gains $\mathbf{h}$, is (cf. (3.51))

$$\mathbb{P}\{\mathbf{x}_A \to \mathbf{x}_B | \mathbf{h}\} = Q\left(\sqrt{\frac{\mathsf{SNR}}{2}\sum_{\ell=1}^{L}|h_\ell|^2|d_\ell|^2}\right), \tag{9.41}$$

where d_ℓ is the ℓth component of the normalized codeword difference (cf. (3.52)):

$$d_\ell := \frac{1}{\sqrt{\mathsf{SNR}}}(x_{A\ell} - x_{B\ell}). \tag{9.42}$$

The *worst-case* pairwise error probability over the channels that are not in outage is the $Q(\sqrt{\cdot})$ function evaluated at the solution to the optimization problem

$$\min_{h_1,\ldots,h_L} \frac{\mathsf{SNR}}{2}\sum_{\ell=1}^{L}|h_\ell|^2|d_\ell|^2, \tag{9.43}$$

subject to the constraint (9.40). If we define $Q_\ell := \mathsf{SNR}\cdot|h_\ell|^2|d_\ell|^2$, then the optimization problem can be rewritten as

$$\min_{Q_1\geq 0,\ldots,Q_L\geq 0} \frac{1}{2}\sum_{\ell=1}^{L} Q_\ell \tag{9.44}$$

subject to the constraint

$$\sum_{\ell=1}^{L}\log\left(1+\frac{Q_\ell}{|d_\ell|^2}\right) \geq LR. \tag{9.45}$$

This is analogous to the problem of minimizing the total power required to support a target rate R bits/s/Hz per sub-channel over a parallel Gaussian channel; the solution is just standard waterfilling, and the worst-case channel is

$$|h_\ell|^2 = \frac{1}{\mathsf{SNR}}\cdot\left(\frac{1}{\lambda|d_\ell|^2}-1\right)^+, \qquad \ell = 1,\ldots,L. \tag{9.46}$$

Here λ is the Lagrange multiplier chosen such that the channel in (9.46) satisfies (9.40) with equality. The worst-case pairwise error probability is

$$Q\left(\sqrt{\frac{1}{2}\sum_{\ell=1}^{L}\left(\frac{1}{\lambda}-|d_\ell|^2\right)^+}\right), \tag{9.47}$$

where λ satisfies

$$\sum_{\ell=1}^{L}\left[\log\left(\frac{1}{\lambda|d_\ell|^2}\right)\right]^+ = LR. \tag{9.48}$$

Examples

We look at some simple coding schemes to better understand the universal design criterion, the argument of the $Q\left(\sqrt{\cdot/2}\right)$ function in (9.47):

$$\sum_{\ell=1}^{L}\left(\frac{1}{\lambda}-|d_\ell|^2\right)^+, \tag{9.49}$$

where λ satisfies the constraint in (9.48).

1. **No coding** Here symbols from L independent constellations (say, QAM), with 2^R points each, are transmitted separately on each of the sub-channels. This has very poor performance since all but one of the $|d_\ell|^2$ can be simultaneously zero. Thus the design criterion in (9.49) evaluates to zero.
2. **Repetition coding** Suppose the symbol is drawn from a QAM constellation (with 2^{RL} points) but the same symbol is repeated over each of the sub-channels. For the 2-parallel channel with $R = 2$ bits/s/Hz per sub-channel, the repetition code is illustrated in Figure 9.12. The smallest value of $|d_\ell|^2$ is 4/9. Due to the repetition, for any pair of codewords, the differences in the sub-channels are equal. With the choice of the worst pairwise differences, the universal criterion in (9.49) evaluates to 8/3 (see Exercise 9.12).
3. **Permutation coding** Consider the 2-parallel channel where the symbol on each of the sub-channels is drawn from a separate QAM constellation. This

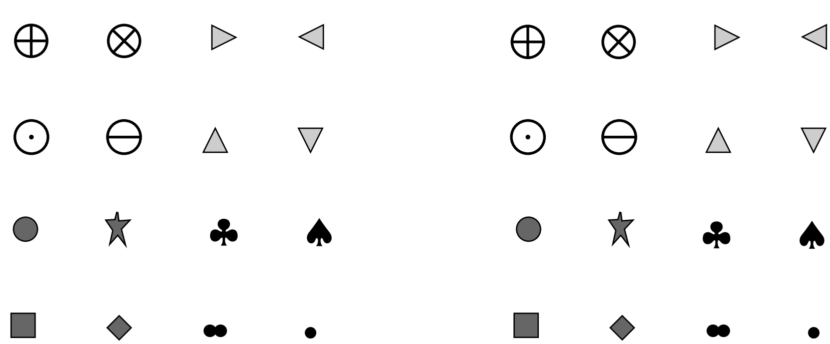

Figure 9.12 A repetition code for the 2-parallel channel with rate $R = 2$ bits/s/Hz per sub-channel.

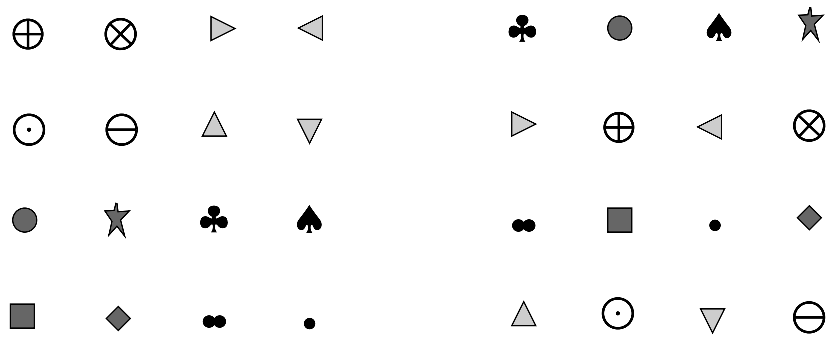

Figure 9.13 A permutation code for the 2-parallel channel with rate $R = 2$ bits/s/Hz per sub-channel.

is similar to the repetition code (Figure 9.12), but we consider different mappings of the QAM points in the sub-channels. In particular, we map the points such that if two points are close to each other in one QAM constellation, their images in the other QAM constellation are far apart. One such choice is illustrated in Figure 9.13, for $R = 2$ bits/s/Hz per sub-channel where two points that are nearest neighbors in one QAM constellation have their images in the other QAM constellation separated by at least double the minimum distance. With the choice of the worst pairwise differences for this code, the universal design criterion in (9.49) can be explicitly evaluated to be 44/9 (see Exercise 9.13).

This code involves a one-to-one map between the two QAM constellations and can be parameterized by a *permutation* of the QAM points. The repetition code is a special case of this class of codes: it corresponds to the *identity* permutation.

Universal code design criterion at high SNR

Although the universal criterion (9.49) can be computed given the codewords, the expression is quite complicated (Exercise 9.11) and is not amenable to use as a criterion for code design. We can however find a simple bound by relaxing the non-negativity constraint in the optimization problem (9.44). This allows the water depth to go negative, resulting in the following lower bound on (9.49):

$$L2^R|d_1 d_2 \cdots d_L|^{2/L} - \sum_{\ell=1}^{L} |d_\ell|^2. \tag{9.50}$$

When the rate of communication per sub-channel R is large, the water level in the waterfilling problem (9.44) is deep at every sub-channel for good codes, and this lower bound is tight. Moreover, for good codes the second term is small compared to the first term, and so in this regime the universal criterion is approximately

$$L2^R|d_1 d_2 \cdots d_L|^{2/L}. \tag{9.51}$$

Thus, the universal code design problem is to choose the codewords maximizing the pairwise product distance; in this regime, the criterion coincides with that of the i.i.d. Rayleigh parallel fading channel (cf. Section 3.2.2).

Property of an approximately universal code

We can use the universal code design criterion developed above to characterize the property of a code that makes it approximately universal over the parallel channel at high SNR. Following the approach in Section 9.2.1, we first define a pairwise typical error event: this is when the argument of the $Q(\sqrt{\cdot/2})$ in (9.41) is less than 1:

$$\mathsf{SNR} \cdot \sum_{\ell=1}^{L} |h_\ell|^2 |d_\ell|^2 < 1. \tag{9.52}$$

For a code to be approximately universal, we want this event to occur only when the channel is in outage; equivalently, this event should not occur whenever the channel is not in outage. This translates to saying that the worst-case code design criterion derived above should be greater than 1. At high SNR, using (9.51), the condition becomes

$$\boxed{|d_1 d_2 \cdots d_L|^{2/L} > \frac{1}{L2^R}.} \tag{9.53}$$

Moreover, this condition should hold for any pair of codewords. It is verified in Exercise 9.14 that this is sufficient to guarantee that a coding scheme achieves the optimal diversity–multiplexing tradeoff of the parallel channel.

We saw the permutation code in Figure 9.13 as an example of a code with good universal design criterion value. This class of codes contains approximately universal codes. To see this, we first need to generalize the essential structure in the permutation code example in Figure 9.13 to higher rates and to more than two sub-channels. We consider codes of just a single block length to carry out the following generalization.

We fix the constellation from which the codeword is chosen in each sub-channel to be a QAM. Each of these QAM constellations contains the entire information to be transmitted: so, the total number of points in the QAM constellation is 2^{LR} if R is the data rate per sub-channel. The overall code is specified by the maps between the QAM points for each of the sub-channels. Since the maps are one-to-one, they can be represented by permutations of the QAM points. In particular, the code is specified by $L-1$ permutations $\pi_2, \ldots, \pi_L$: for each message, say m, we identify one of the QAM points, say q, in the QAM constellation for the first sub-channel. Then, to convey the message m, the transmit codeword is

$$(q, \pi_2(q), \ldots, \pi_L(q)),$$

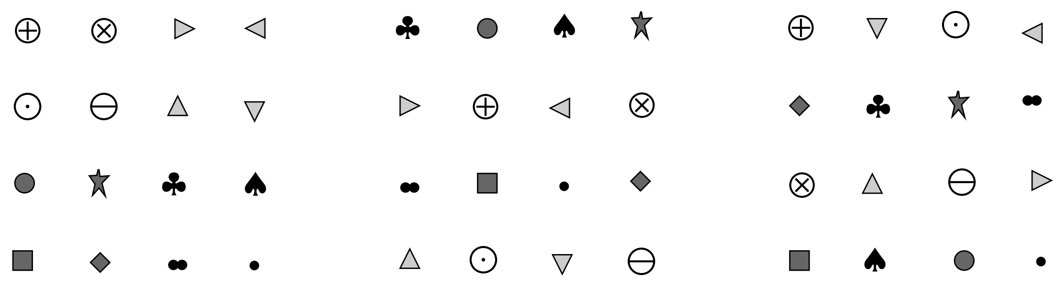

Figure 9.14 A permutation code for a parallel channel with three sub-channels. The entire information (4 bits) is contained in each of the QAM constellations.

i.e., the QAM point transmitted over the ℓth sub-channel is $\pi_\ell(q)$ with π_1 defined to be the identity permutation. An example of a permutation code with a rate of 4/3 bits/s/Hz per sub-channel for $L = 3$ (so the QAM constellation has 2^4 points) is illustrated in Figure 9.14.

Given the physical constraints (the operating SNR, the data rate, and the number of sub-channels), the engineer can now choose appropriate permutations to maximize the universal code design criterion. Thus permutation codes provide a framework within which specific codes can be designed based on the requirements. This framework is quite rich: Exercise 9.15 shows that even randomly chosen permutations are approximately universal with high probability.

Bit-reversal scheme: an operational interpretation of the outage condition

We can use the concept of approximately universal codes to give an operational interpretation of the outage condition for the parallel channel. To be able to focus on the essential issues, we restrict our attention to just two sub-channels, so $L = 2$. If we communicate at a total rate $2R$ bits/s/Hz over the parallel channel, the no-outage condition is

$$\log(1 + |h_1|^2 \mathsf{SNR}) + \log(1 + |h_2|^2 \mathsf{SNR}) > 2R. \tag{9.54}$$

One way of interpreting this condition is as though the first sub-channel provides $\log(1 + |h_1|^2\mathsf{SNR})$ bits of information and the second sub-channel provides $\log(1 + |h_2|^2\mathsf{SNR})$ bits of information, and as long as the total number of bits provided exceeds the target rate, then reliable communication is possible. In the high SNR regime, we exhibit below a permutation code that makes the outage condition concrete.

Suppose we independently code over the I and Q channels of the two sub-channels. So we can focus on only one of them, say, the I channel. We wish to communicate R bits over two uses of the I-channel. Analogous to the typical event analysis for the scalar channel, we can exactly recover all the R information bits from the first I sub-channel alone if

$$|h_1|^2 > \frac{2^{2R}}{\mathsf{SNR}}, \tag{9.55}$$

or

$$|h_1|^2\mathsf{SNR} > 2^{2R}. \tag{9.56}$$

However, we do not need to use just the first I sub-channel to recover all the information bits: the second I sub-channel also contains the same information and can be used in the recovery process. Indeed, if we create x_1^{I} by treating the ordered R bits as the binary representation of the points x_1^{I}, then one would intuitively expect that if

$$|h_1|^2\mathsf{SNR} > 2^{2R_1}, \tag{9.57}$$

then one should be able to recover at least R_1 of the most significant bits of information. Now, if we create x_2^{I} by treating the *reversal* of the R bits as its binary representation, then one should be able to recover at least R_2 of the most significant bits, if

$$|h_2|^2\mathsf{SNR} > 2^{2R_2}. \tag{9.58}$$

But due to the reversal, the most significant bits in the representation in the second I sub-channel are the least significant bits in the representation in the first I sub-channel. Hence, as long as $R_1 + R_2 \geq R$, then we can recover *all* R bits. This translates to the condition

$$\log(|h_1|^2\mathsf{SNR}) + \log(|h_2|^2\mathsf{SNR}) > 2R, \tag{9.59}$$

which is precisely the no-outage condition (9.54) at high SNR.

The bit-reversal scheme described here with some slight modifications can be shown to be approximately universal (Exercise 9.16). A simple variant of this scheme is also approximately universal (Exercise 9.17).

Summary 9.2 Universal codes for the parallel channel

A universal code design criterion between two codewords can be computed by finding the channel not in outage that yields the worst-case pairwise error probability.

At high SNR and high rate, the universal code design criterion becomes proportional to the product distance:

$$|d_1 \ldots d_L|^{2/L} \tag{9.60}$$

where L is the number of sub-channels and d_ℓ is the difference between the ℓth components of the codewords.

A code is approximately universal for the parallel channel if its product distance is large enough: for a code at a data rate of R bits/s/Hz per sub-channel, we require

$$|d_1 d_2 \cdots d_L|^2 > \frac{1}{(L2^R)^L}. \tag{9.61}$$

Simple bit-reversal schemes are approximately universal for the 2-parallel channel. Random permutation codes are approximately universal for the L-parallel channel with high probability.

9.2.3 Universal code design for MISO channels

The outage event for the $n_t \times 1$ MISO channel (9.22) is

$$\log\left(1 + \|\mathbf{h}\|^2 \frac{\mathsf{SNR}}{n_t}\right) < R. \tag{9.62}$$

In the case when $n_t = 2$, the Alamouti scheme converts the MISO channel to a scalar channel with gain $\|\mathbf{h}\|$ and SNR reduced by a factor of 2. Hence, the outage behavior is exactly the same as in the original MISO channel, and the Alamouti scheme provides a *universal* conversion of the 2×1 MISO channel to a scalar channel. Any approximately universal scheme for the scalar channel, such as QAM, when used in conjunction with the Alamouti scheme is also approximately optimal for the MISO channel and achieves its diversity–multiplexing tradeoff.

In the general case when the number of transmit antennas is greater than two, there is no equivalence of the Alamouti scheme. Here we explore two approaches to constructing universal schemes for the general MISO channel.

MISO channel viewed as a parallel channel

Using one transmit antenna at a time converts the MISO channel into a parallel channel. We have used this conversion in conjunction with repetition coding to argue the classical diversity gain of the MISO channel (cf. Section 3.3.2). Replacing the repetition code with an appropriate parallel channel code (such as the bit-reversal scheme from Section 9.2.2), we will see that converting the MISO channel into a parallel channel is actually tradeoff-optimal for the i.i.d. Rayleigh fading channel.

Suppose we want to communicate at rate $R = r \log \mathsf{SNR}$ bits/s/Hz on the MISO channel. Using one transmit antenna at a time yields a parallel channel with n_t diversity branches and the data rate of communication is R bits/s/Hz per sub-channel. The optimal diversity gain for the i.i.d. Rayleigh parallel fading channel is $n_t(1-r)$ (cf. (9.20)); thus, using one antenna at a

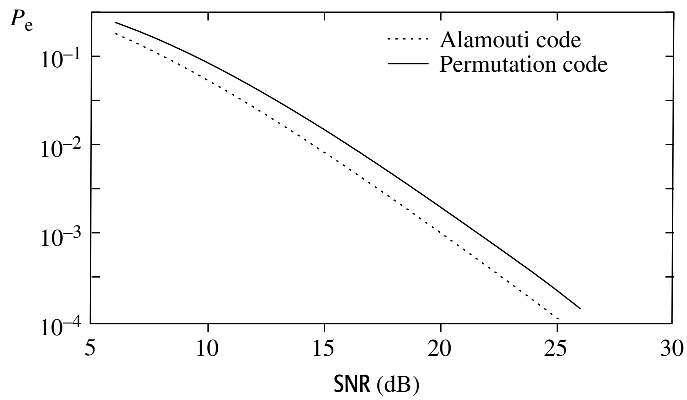

Figure 9.15 The error probability of uncoded QAM with the Alamouti scheme and that of a permutation code over one antenna at a time for the Rayleigh fading MISO channel with two transmit antennas: the permutation code is about 1.5 dB worse than the Alamouti scheme over the plotted error probability range.

time in conjunction with a tradeoff-optimal parallel channel code achieves the largest diversity gain over the i.i.d. Rayleigh fading MISO channel (cf. (9.24)).

To understand how much loss the conversion of the MISO channel into a parallel channel entails with respect to the optimal outage performance, we plot the error probabilities of two schemes with the same rate ($R = 2$ bits/s/Hz): uncoded QAM over the Alamouti scheme and the permutation code in Figure 9.13. This performance is plotted in Figure 9.15 where we see that the conversion of the MISO channel into a parallel channel entails a loss of about 1.5 dB in SNR for the same error probability performance.

Universality of conversion to parallel channel

We have seen that the conversion of the MISO channel into a parallel channel is tradeoff-optimal for the i.i.d. Rayleigh fading channel. Is this conversion universal? In other words, will a tradeoff-optimal scheme for the parallel channel also be tradeoff-optimal for the MISO channel, under *any* channel statistics? In general, the answer is no. To see this, consider the following MISO channel model: suppose the channels from all but the first transmit antenna are very poor. To make this example concrete, suppose $h_\ell = 0,\ \ell = 2, \dots, n_t$. The tradeoff curve depends on the outage probability (which depends only on the statistics of the first channel)

$$p_{\text{out}} = \mathbb{P}\left\{\log\left(1 + \mathsf{SNR}|h_1|^2\right) < R\right\}. \tag{9.63}$$

Using one transmit antenna at a time is a waste of degrees of freedom: since the channels from all but the first antenna are zero, there is no point in transmitting any signal on them. This loss in degrees of freedom is explicit in the outage probability of the parallel channel formed by transmitting from one antenna at a time:

$$p_{\text{out}}^{\text{parallel}} = \mathbb{P}\left\{\log\left(1 + \mathsf{SNR}|h_1|^2\right) < n_t R\right\}. \tag{9.64}$$

Comparing (9.64) with (9.63), we see clearly that the conversion to the parallel channel is not tradeoff-optimal for this channel model.

Essentially, using one antenna at a time equates temporal degrees of freedom with spatial ones. All temporal degrees of freedom are the same, but the spatial ones need not be the same: in the extreme example above, the spatial channels from all but the first transmit antenna are zero. Thus, it seems reasonable that when all the spatial channels are *symmetric* then the parallel channel conversion of the MIMO channel is justified. This sentiment is justified in Exercise 9.18, which shows that the parallel channel conversion is approximately *universal* over a restricted class of MISO channels: those with i.i.d. spatial channel coefficients.

Universal code design criterion

Instead of converting to a parallel channel, one can design universal schemes directly for the MISO channel. What is an appropriate code design criterion? In the context of the i.i.d. Rayleigh fading channel, we derived the determinant criterion for the codeword difference matrices in Section 3.3.2. What is the corresponding criterion for universal MISO schemes? We can answer this question by considering the worst-case pairwise error probability over all MISO channels that are not in outage.

The pairwise error probability (of confusing the transmit codeword matrix $\mathbf{X}_A$ with $\mathbf{X}_B$) conditioned on a specific MISO channel realization is (cf. (3.82))

$$\mathbb{P}\{\mathbf{X}_A \to \mathbf{X}_B|\mathbf{h}\} = Q\left(\frac{\|\mathbf{h}^*(\mathbf{X}_A - \mathbf{X}_B)\|}{\sqrt{2}}\right). \tag{9.65}$$

In Section 3.3.2 we averaged this quantity over the statistics of the MISO channel (cf. (3.83)). Here we consider the worst-case over all channels not in outage:

$$\max_{\mathbf{h}:\|\mathbf{h}\|^2 > \frac{n_t(2^R-1)}{\mathsf{SNR}}} Q\left(\frac{\|\mathbf{h}^*(\mathbf{X}_A - \mathbf{X}_B)\|}{\sqrt{2}}\right). \tag{9.66}$$

From a basic result in linear algebra, the worst-case pairwise error probability in (9.66) can be explicitly written as (Exercise 9.19)

$$Q\left(\sqrt{\frac{1}{2}\lambda_1^2 n_t (2^R - 1)}\right), \tag{9.67}$$

where λ_1 is the *smallest* singular value of the normalized codeword difference matrix

$$\frac{1}{\sqrt{\mathsf{SNR}}}(\mathbf{X}_A - \mathbf{X}_B). \tag{9.68}$$

Essentially, the worst-case channel aligns itself in the direction of the weakest singular value of the codeword difference matrix. So, the universal code design criterion for the MISO channel is to ensure that no singular value is too small; equivalently

maximize the minimum singular value of the codeword difference matrices. (9.69)

There is an intuitive explanation for this design criterion: a universal code has to protect itself against the worst channel that is not in outage. The condition of no-outage only puts a constraint on the *norm* of the channel vector **h** but not on its direction. So, the worst channel aligns itself to the "weakest direction" of the codeword difference matrix to create the most havoc. The corresponding worst-case pairwise error probability will be governed by the smallest singular value of the codeword difference matrix. On the other hand, the i.i.d. Rayleigh channel does not prefer any specific direction: thus the design criterion tailored to its statistics requires that the *average* direction be well protected and this translates to the determinant criterion. While the two criteria are different, codes with large determinant tend to also have a large value for the smallest singular value; the two criteria (based on worst-case and average-case) are related in this aspect.

We can use the universal code design criterion to derive a property that makes a code universally achieve the tradeoff curve (as we did for the parallel channel in the previous section). We want the typical error event to occur only when the channel is in outage. This corresponds to the argument of $Q(\sqrt{(\cdot)/2})$ in the worst-case error probability (9.67) to be greater than 1, i.e.,

$$\boxed{\lambda_1^2 > \frac{1}{n_{\mathrm{t}}(2^R - 1)} \approx \frac{1}{n_{\mathrm{t}} 2^R}.} \tag{9.70}$$

for every pair of codewords. We can explicitly verify that the Alamouti scheme with independent uncoded QAMs on the two data streams satisfies the approximate universality property in (9.70). This is done in Exercise 9.20.

Summary 9.3 Universal codes for the MISO channel

The MISO channel can be converted into a parallel channel by using one transmit antenna at a time. This conversion is approximately universal for the class of MISO channels with i.i.d. fading coefficients.

The universal code design criterion is to maximize the minimum singular value of the codeword difference matrices.

9.2.4 Universal code design for MIMO channels

We finally arrive at the multiple transmit *and* multiple receive antenna slow fading channel:

$$\mathbf{y}[m] = \mathbf{H}\mathbf{x}[m] + \mathbf{w}[m]. \tag{9.71}$$

The outage event of this channel is

$$\log\det(\mathbf{I}_{n_\mathrm{r}} + \mathbf{H}\mathbf{K}_x\mathbf{H}^*) < R, \tag{9.72}$$

where $\mathbf{K}_x$ is the optimizing covariance in (9.29).

Universality of D-BLAST

In Section 8.5, we have seen that the D-BLAST architecture with the MMSE–SIC receiver converts the MIMO channel into a parallel channel with n_t sub-channels. Suppose we pick the transmit strategy $\mathbf{K}_x$ in the D-BLAST architecture (the covariance matrix represents the combination of the power allocated to the streams and coordinate system under which they are mixed before transmitting, cf. (8.3)) to be the one in (9.72). The important property of this conversion is the conservation expressed in (8.88): denoting the effective SNR of the kth sub-channel of the parallel channel by SINR_k,

$$\log\det\left(\mathbf{I}_{n_\mathrm{r}} + \mathbf{H}\mathbf{K}_x\mathbf{H}^*\right) = \sum_{k=1}^{n_\mathrm{t}} \log(1+\mathsf{SINR}_k). \tag{9.73}$$

However, $\mathsf{SINR}_1, \ldots, \mathsf{SINR}_{n_\mathrm{t}}$, across the sub-channels are *correlated*. On the other hand, we saw codes (with just block length 1) that universally achieve the tradeoff curve for any parallel channel (in Section 9.2.2). This means that, using approximately universal parallel channel codes for each of the interleaved streams, the D-BLAST architecture with the MMSE–SIC receiver at a rate of $R = r\log\mathsf{SNR}$ bits/s/Hz per stream has a diversity gain determined by the decay rate of

$$\mathbb{P}\left\{\sum_{k=1}^{n_\mathrm{t}} \log(1+\mathsf{SINR}_k) < R\right\}, \tag{9.74}$$

with increasing SNR. With n interleaved streams, each having block length 1 (i.e., $N=1$ in the notation of Section 8.5.2), the initialization loss in D-BLAST reduces a data rate of R bits/s/Hz per stream into a data rate of $nR/(n+n_\mathrm{t}-1)$ bits/s/Hz on the MIMO channel (Exercise 8.27). Suppose we use the D-BLAST architecture in conjunction with a block length 1 universal parallel channel code for each of n interleaved streams. If this code operates at a multiplexing gain of r on the MIMO channel, the diversity gain obtained is, substituting for the rate in (9.74) and comparing with (9.73), the decay rate of

$$\mathbb{P}\left\{\log\det\left(\mathbf{I}_{n_r}+\mathbf{H}\mathbf{K}_x\mathbf{H}^*\right)<\frac{r(n+n_t-1)}{n}\log\mathsf{SNR}\right\}. \tag{9.75}$$

Now comparing this with the actual decay behavior of the outage probability (cf. (9.29)), we see that the D-BLAST/MMSE–SIC architecture with n interleaved streams used to operate at a multiplexing gain of r over the MIMO channel has a diversity gain equal to the decay rate of

$$p_{\text{out}}^{\text{mimo}}\left(\frac{r(n+n_t-1)}{n}\log\mathsf{SNR}\right). \tag{9.76}$$

Thus, with a large number, n, of interleaved streams, the D-BLAST/MMSE–SIC architecture achieves universally the tradeoff curve of the MIMO channel. With a finite number of streams, it is strictly tradeoff-suboptimal. In fact, the tradeoff performance can be improved by replacing the MMSE–SIC receiver by joint ML decoding of all the streams. To see this concretely, let us consider the 2×2 MIMO Rayleigh fading channel (so $n_t = n_r = 2$) with just two interleaved streams (so $n = 2$). The transmit signal lasts 3 time symbols:

$$\begin{bmatrix} 0 & x_B^{(1)} & x_B^{(2)} \\ x_A^{(1)} & x_A^{(2)} & 0 \end{bmatrix}. \tag{9.77}$$

With the MMSE–SIC receiver, the diversity gain obtained at the multiplexing rate of r is the optimal diversity gain at the multiplexing rate of $3r/2$. This scaled version of the optimal tradeoff curve is depicted in Figure 9.16. On the other hand, with the ML receiver the performance is significantly improved, also depicted in Figure 9.16. This achieves the optimal diversity performance for multiplexing rates between 0 and 1, and in fact is the scheme that sends 4 symbols over 3 symbol times that we were seeking in Section 9.1.5! The performance analysis of the D-BLAST architecture with the joint ML receiver is rather intricate and is carried out in Exercise 9.21. Basically, MMSE–SIC is suboptimal because it favors stream 1 over stream 2 while ML treats them equally. This asymmetry is only a small edge effect when there are many interleaved streams but does impact performance when there are only a small number of streams.

Universal code design criterion

We have seen that the D-BLAST architecture is a universal one, but how do we recognize when another space-time code also has good outage performance universally? To answer this question, we can derive a code design criterion based on the worst-case MIMO channel that is not in outage. Consider space-time code matrices with block length n_t. The worst-case channel aligns itself in the "weakest directions" afforded by a codeword pair difference matrix. With just

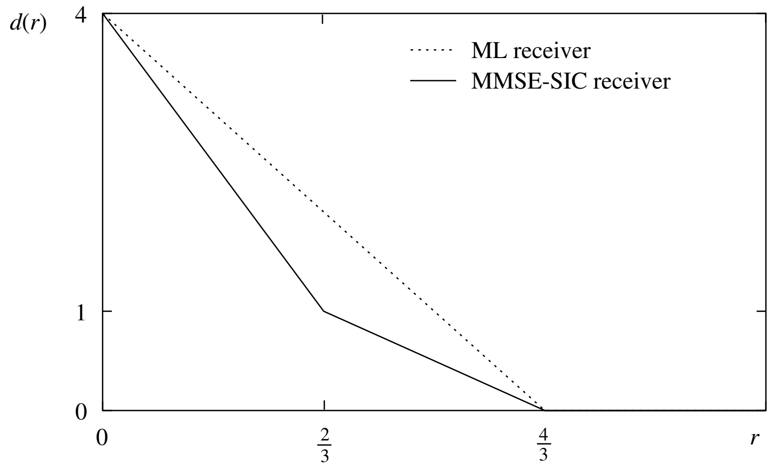

Figure 9.16 Tradeoff performance for the D-BLAST architecture with the ML receiver and with the MMSE–SIC receiver.

one receive antenna, the MISO channel is simply a row vector and it aligns itself in the direction of the smallest singular value of the codeword difference matrix (cf. Section 9.2.3). Here, there are n_{min} directions for the MIMO channel and the corresponding design criterion is an extension of that for the MISO channel: the universal code design criterion at high SNR is to maximize

$$\lambda_1 \lambda_2 \cdots \lambda_{n_{\text{min}}}, \tag{9.78}$$

where $\lambda_1, \ldots, \lambda_{n_{\text{min}}}$ are the smallest n_{min} singular values of the normalized codeword difference matrices (cf. (9.68)). The derivation is carried out in Exercise 9.22. With $n_{\text{t}} \leq n_{\text{r}}$, this is just the *determinant* criterion, derived in Chapter 3 by averaging the code performance over the i.i.d. Rayleigh statistics.

The exact code design criterion at an intermediate value of SNR is similar to the expression for the universal code design for the parallel channel (cf. (9.49)).

Property of an approximately universal code

Using exactly the same arguments as in Section 9.2.2, we can use the universal code design criterion developed above to characterize the property of a code that makes it approximately universal over the MIMO channel (see Exercise 9.23):

$$\boxed{|\lambda_1 \lambda_2 \cdots \lambda_{n_{\text{min}}}|^{2/n_{\text{min}}} > \frac{1}{n_{\text{min}} 2^{R/n_{\text{min}}}}.} \tag{9.79}$$

As in the parallel channel (cf. Exercise 9.14), this condition is only an order-of-magnitude one. A relaxed condition

$$|\lambda_1 \lambda_2 \cdots \lambda_{n_{\text{min}}}|^{2/n_{\text{min}}} > c \cdot \frac{1}{n_{\text{min}} 2^{R/n_{\text{min}}}}, \quad \text{for some constant } c > 0, \tag{9.80}$$

can also be used for approximate universality: it is sufficient to guarantee that the code achieves the optimal diversity–multiplexing tradeoff. We can make a couple of interesting observations immediately from this result.

- If a code satisfies the condition for approximate universality in (9.80) for an $n_t \times n_r$ MIMO channel with $n_r \geq n_t$, i.e., the number of receive antennas is equal to or larger than the number of transmit antennas, then it is also approximately universal for an $n_t \times l$ MIMO channel with $l \geq n_r$.
- The singular values of the normalized codeword matrices are upper bounded by $2\sqrt{n_t}$ (Exercise 9.24). Thus, a code that satisfies (9.80) for an $n_t \times n_r$ MIMO channel also satisfies the criterion in (9.80) for an $n_t \times l$ MIMO channel with $l \leq n_r$. Thus is it also approximately universal for the $n_t \times l$ MIMO channel with $l \leq n_r$.

We can conclude the following from the above two observations:

A code that satisfies (9.80) for an $n_t \times n_t$ MIMO channel is approximately universal for an $n_t \times n_r$ MIMO channel for *every* value of the number of receive antennas n_r.

Exercise 9.25 shows a rotation code that satisfies (9.80) for the 2×2 MIMO channel; so this code is approximately universal for every $2 \times n_r$ MIMO channel.

We have already observed that the D-BLAST architecture with approximately universal parallel channel codes for the interleaved streams is approximately universal for the MIMO channel. Alternatively, we can see its approximate universality by explicitly verifying that it satisfies the condition in (9.80) with $n_t = n_r$. Here, we will see this for the 2×2 channel with two interleaved streams in the D-BLAST transmit codeword matrix (cf. (9.77)). The normalized codeword difference matrix can be written as

$$\mathbf{D} = \begin{bmatrix} 0 & d_B^{(1)} & d_B^{(2)} \\ d_A^{(1)} & d_A^{(2)} & 0 \end{bmatrix}, \tag{9.81}$$

where $\left(d_B^{(\ell)}, d_A^{(\ell)}\right)$ is the normalized pairwise difference codeword for an approximately universal parallel channel code and satisfies the condition in (9.53):

$$|d_B^{(\ell)} d_A^{(\ell)}| > \frac{1}{4 \cdot 2^R}, \qquad \ell = 1, 2. \tag{9.82}$$

Here R is the rate in bits/s/Hz in each of the streams. The product of the two singular values of $\mathbf{D}$ is

$$\begin{aligned} \lambda_1^2 \lambda_2^2 &= \det(\mathbf{D}\mathbf{D}^*) \\ &= |d_B^{(1)} d_A^{(1)}|^2 + |d_B^{(2)} d_A^{(2)}|^2 + |d_B^{(2)} d_A^{(1)}|^2 \\ &> \frac{1}{4 \cdot 2^R}, \end{aligned} \tag{9.83}$$

where the last inequality follows from (9.82). A rate of R bits/s/Hz on each of the streams corresponds to a rate of $2R/3$ bits/s/Hz on the MIMO channel. Thus, comparing (9.83) with (9.79), we have verified the approximate universality of D-BLAST at a reduced rate due to the initialization loss. In other words, the diversity gain obtained by the D-BLAST architecture in (9.77) at a multiplexing rate of r over the MIMO channel is $d^*(3r/2)$.

Discussion 9.1 Universal codes in the downlink

Consider the downlink of a cellular system where the base-stations are equipped with multiple transmit antennas. Suppose we want to broadcast the same information to all the users in the cell in the downlink. We would like our transmission scheme to not depend on the number of receive antennas at the users: each user could have a different number of receive antennas, depending on the model, age, and type of the mobile device.

Universal MIMO codes provide an attractive solution to this problem. Suppose we broadcast the common information at rate R using a space-time code that satisfies (9.79) for an $n_t \times n_t$ MIMO channel. Since this code is approximately universal for every $n_t \times n_r$ MIMO channel, the diversity seen by each user is *simultaneously* the best possible at rate R. To summarize: the diversity gain obtained by each user is the best possible with respect to both

- the number of receive antennas it has, and
- the statistics of the fading channel the user is currently experiencing.

Chapter 9 The main plot

For a slow fading channel at high SNR, the tradeoff between data rate and error probability is captured by the tradeoff between multiplexing and diversity gains. The optimal diversity gain $d^*(r)$ is the rate at which outage probability decays with increasing SNR when the data rate is increasing as $r \log \mathsf{SNR}$. The classical diversity gain is the diversity gain at a fixed rate, i.e., the multiplexing gain $r = 0$.

The optimal diversity gain $d^*(r)$ is determined by the outage probability of the channel at a data rate of $r \log \mathsf{SNR}$ bits/s/Hz. The operational interpretation is via the existence of a universal code that achieves reliable communication simultaneously over all channels that are not in outage.

The universal code viewpoint provides a new code design criterion. Instead of averaging over the channel statistics, we consider the performance of a code over the worst-case channel that is not in outage.

- For the parallel channel, the universal criterion is to maximize the product of the codeword differences. Somewhat surprisingly, this is the same as the criterion arrived at by averaging over the Rayleigh channel statistics.
- For the MISO channel, the universal criterion is to maximize the smallest singular value of the codeword difference matrices.
- For the $n_t \times n_r$ MIMO channel, the universal criterion is to maximize the product of the n_{min} smallest singular values of the codeword difference matrices. With $n_r \geq n_t$, this criterion is the same as that arrived at by averaging over the i.i.d. Rayleigh statistics.

The MIMO channel can be transformed into a parallel channel via D-BLAST. This transformation is universal: universal parallel channel codes for each of the interleaved streams in D-BLAST serve as a universal code for the MIMO channel. The rate loss due to initialization in D-BLAST can be reduced by increasing the number of interleaved streams. For the MISO channel, however, the D-BLAST transformation with only one stream, i.e., using the transmit antennas one at a time, is approximately universal within the class of channels that have i.i.d. fading coefficients.

9.3 Bibliographical notes

The design of space-time codes has been a fertile area of research. There are books that provide a comprehensive view of the subject: for example, see the books by Larsson, Stoica and Ganesan [72], and Paulraj *et al.* [89]. Several works have recognized the tradeoff between diversity and multiplexing gains. The formulation of the coarser scaling of error probability and data rate and the corresponding characterization of their fundamental tradeoff for the i.i.d. Rayleigh fading channel is the work of Zheng and Tse [156].

The notion of universal communication, i.e., communicating reliably over a class of channel, was first formulated in the context of discrete memoryless channels by Blackwell *et al.* [10], Dobrushin [31] and Wolfowitz [146]. They showed the existence of universal codes. The results were later extended to Gaussian channels by Root and Varaiya [103]. Motivated by these information theoretic results, Wesel and his coauthors have studied the problem of universal code design in a sequence of works, starting with his Ph.D. thesis [142]. The worst-case code design metric for the parallel channel and a heuristic derivation of the product distance criterion were obtained in [143]. This was extended to MIMO channels in [67]. The general concept of approximate universality in the high SNR regime was formulated by Tavildar and Viswanath [118]; earlier, in the special case of the 2×2 MIMO channel, Yao and Wornell [152] used the determinant condition (9.80) to show the tradeoff-optimality of their rotation-based codes. The conditions derived for approximate universality, (cf. (9.38), (9.53), (9.70) and (9.80)) are also necessary; this is derived in Tavildar and Viswanath [118].

The design of tradeoff-optimal space-time codes is an active area of research, and several approaches have been presented recently. They include: rotation-based codes for the 2×2 channel, by Yao and Wornell [152] and Dayal and Varanasi [29]; lattice space-time (LAST) codes, by El Gamal *et al.* [34]; permutation codes for the parallel channel derived from D-BLAST, by Tavildar and Viswanath [118]; Golden code, by Belfiore *et al.* [5] for the 2×2 channel; codes based on cyclic divisional algebras, by Elia *et al.* [35]. The tradeoff-optimality of most of these codes is demonstrated by verifying the approximate universality conditions.

9.4 Exercises

Exercise 9.1 Consider the L-parallel channel with i.i.d. Rayleigh coefficients. Show that the optimal diversity gain at a multiplexing rate of r per sub-channel is $L - Lr$.

Exercise 9.2 Consider the repetition scheme where the same codeword is transmitted over the L i.i.d. Rayleigh sub-channels of a parallel channel. Show that the largest diversity gain this scheme can achieve at a multiplexing rate of r per sub-channel is $L(1 - Lr)$.

Exercise 9.3 Consider the repetition scheme of transmitting the same codeword over the n_t transmit antennas, one at a time, of an i.i.d. Rayleigh fading $n_t \times n_r$ MIMO channel. Show that the maximum diversity gain this scheme can achieve, at a multiplexing rate of r, is $n_t n_r (1 - n_t r)$.

Exercise 9.4 Consider using the Alamouti scheme over a $2 \times n_r$ i.i.d. Rayleigh fading MIMO channel. The transmit codeword matrix spans two symbol times $m = 1, 2$ (cf. Section 3.3.2):

$$\begin{bmatrix} u_1 & -u_2^* \\ u_2 & u_1^* \end{bmatrix}. \tag{9.84}$$

1. With this input to the MIMO channel in (9.71), show that we can write the output over the two time symbols as (cf. (3.75))

$$\begin{bmatrix} \mathbf{y}[1] \\ (\mathbf{y}[2]^*)^t \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 \\ (\mathbf{h}_2^*)^t & -(\mathbf{h}_1^*)^t \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} \mathbf{w}[1] \\ (\mathbf{w}[2]^*)^t \end{bmatrix}. \tag{9.85}$$

   Here we have denoted the two columns of $\mathbf{H}$ by $\mathbf{h}_1$ and $\mathbf{h}_2$.
2. Observing that the two columns of the effective channel matrix in (9.85) are orthogonal, show that we can extract simple sufficient statistics for the data symbols u_1, u_2 (cf. (3.76)):

$$r_i = \|\mathbf{H}\| u_i + w_i, \qquad i = 1, 2. \tag{9.86}$$

   Here $\|\mathbf{H}\|^2$ denotes $\|\mathbf{h}_1\|^2 + \|\mathbf{h}_2\|^2$ and the additive noises w_1 and w_2 are i.i.d. $\mathcal{CN}(0, 1)$.
3. Conclude that the maximum diversity gain seen by either stream (u_1 or u_2) at a multiplexing rate of r per stream is $2n_r(1 - r)$.

Exercise 9.5 Consider the V-BLAST architecture with a bank of decorrelators for the $n_t \times n_r$ i.i.d. Rayleigh fading MIMO channel with $n_r \geq n_t$. Show that the effective channel seen by each stream is a scalar fading channel with distribution $\chi^2_{2(n_r - n_t + 1)}$. Conclude that the diversity gain with a multiplexing gain of r is $(n_r - n_t + 1)\,(1 - r/n_t)$.

Exercise 9.6 Verify the claim in (9.28) by showing that the sum of the pairwise error probabilities in (9.26), with $\mathbf{x}_A, \mathbf{x}_B$ each a pair of QAM symbols (the union bound on the error probability) has a decay rate of $2-r$ with increasing SNR.

Exercise 9.7 The result in Exercise 9.6 can be generalized. Show that the diversity gain of transmitting uncoded QAMs (each at a rate of $R = r/n \log \mathsf{SNR}$ bits/s/Hz) on the n transmit antennas of an i.i.d. Rayleigh fading MIMO channel with n receive antennas is $n-r$.

Exercise 9.8 Consider the expression for $p_{\text{out}}^{\text{mimo}}$ in (9.29) and for $p_{\text{out}}^{\text{iid}}$ in (9.30). Suppose that the entries of the MIMO channel $\mathbf{H}$ have some joint distribution and are not necessarily i.i.d. Rayleigh.

1. Show that

$$p_{\text{out}}^{\text{iid}}(r \log \mathsf{SNR}) \geq p_{\text{out}}^{\text{mimo}}(r \log \mathsf{SNR}) \geq \mathbb{P}\{\log\det(\mathbf{I}_{n_r} + \mathsf{SNR}\,\mathbf{H}\mathbf{H}^*) < r \log \mathsf{SNR}\}. \tag{9.87}$$

2. Show that the lower bound above decays at the same polynomial rate as $p_{\text{out}}^{\text{iid}}$ with increasing SNR.
3. Conclude that the polynomial decay rates of both $p_{\text{out}}^{\text{mimo}}$ and $p_{\text{out}}^{\text{iid}}$ with increasing SNR are the same.

Exercise 9.9 Consider a scalar slow fading channel

$$y[m] = hx[m] + w[m], \tag{9.88}$$

with an optimal diversity–multiplexing tradeoff $d^*(\cdot)$, i.e.,

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log p_{\text{out}}(r \log \mathsf{SNR})}{\log \mathsf{SNR}} = -d^*(r). \tag{9.89}$$

Let $\epsilon > 0$ and consider the following event on the channel gain h:

$$\mathbb{E}_\epsilon := \{h : \log(1 + |h|^2 \mathsf{SNR}^{1-\epsilon}) < R\}. \tag{9.90}$$

1. Show, by conditioning on the event $\mathbb{E}_\epsilon$ or otherwise, that the probability of error $p_{\text{e}}(\mathsf{SNR})$ of QAM with rate $R = r \log \mathsf{SNR}$ bits/symbol satisfies

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log p_{\text{e}}(\mathsf{SNR})}{\log \mathsf{SNR}} \leq -d^*(r)(1-\epsilon). \tag{9.91}$$

   Hint: you should show that conditional on the $\mathbb{E}_\epsilon$ not happening, the probability of error decays very fast and is negligible compared to the probability of error conditional on $\mathbb{E}_\epsilon$ happening.
2. Hence, conclude that QAM achieves the diversity–multiplexing tradeoff of any scalar channel.
3. More generally, show that any constellation that satisfies the condition (9.38) achieves the diversity–multiplexing tradeoff curve of the channel.
4. Even more generally, show that any constellation that satisfies the condition

$$d_{\min}^2 > c \cdot \frac{1}{2^R} \qquad \text{for any constant } c > 0 \tag{9.92}$$

achieves the diversity–multiplexing tradeoff curve of the channel. This shows that the condition (9.38) is really only an order-of-magnitude condition. A slightly weaker version of this condition is also necessary for a code to be approximately universal; see [118].

Exercise 9.10 Consider coding over a block length N for communication over the parallel channel in (9.17). Derive the universal code design criterion, generalizing the derivation in Section 9.2.2 over a block length of 1.

Exercise 9.11 In this exercise we will try to explicitly calculate the universal code design criterion for the parallel fading channel; for given differences between a pair of normalized codewords, the criterion is to maximize the expression in (9.49).

1. Suppose the codeword differences on all the sub-channels have the same magnitude, i.e., $|d_1| = \cdots = |d_L|$. Show that in this case the worst case channel is the same over all the sub-channels and the universal criterion in (9.49) simplifies considerably to

$$L(2^R - 1)|d_1|^2. \tag{9.93}$$

2. Suppose the codeword differences are ordered: $|d_1| \leq \cdots \leq |d_L|$.
   (a) Argue that if the worst case channel h_ℓ on the ℓth sub-channel is non-zero, then it is also non-zero on all the sub-channels $1, \ldots, \ell - 1$.
   (b) Consider the largest k such that

$$|d_k|^{2k} \leq 2^{RL}|d_1 \cdots d_k|^2 \leq |d_{k+1}|^{2k}, \tag{9.94}$$

   with $|d_{L+1}|$ defined as $+\infty$. Argue that the worst-case channel is zero on all the sub-channels $k+1, \ldots, L$. Observe that $k = L$ when all the codeword differences have the same magnitude; this is in agreement with the result in part (1).
3. Use the results of the previous part (and the notation of k from (9.94)) to derive an explicit expression for λ in (9.49):

$$\lambda^k |d_1 \cdots d_k|^2 = 2^{-RL}. \tag{9.95}$$

   Conclude that the universal code design criterion is to maximize

$$\left(k(2^{RL}|d_1 d_2 \cdots d_k|^2)^{1/k} - \sum_{\ell=1}^{k} |d_\ell|^2 \right). \tag{9.96}$$

Exercise 9.12 Consider the repetition code illustrated in Figure 9.12. This code is for the 2-parallel channel with $R = 2$ bits/s/Hz per sub-channel. We would like to evaluate the value of the universal design criterion, minimized over all pairs of codewords. Show that this value is equal to 8/3. *Hint*: The smallest value is yielded by choosing the pair of codewords as nearest neighbors in the QAM constellation. Since this is a repetition code, the codeword differences are the same for both the channels; now use (9.93) to evaluate the universal design criterion.

Exercise 9.13 Consider the permutation code illustrated in Figure 9.13 (with $R = 2$ bits/s/Hz per sub-channel). Show that the smallest value of the universal design criterion, minimized over all choices of codeword pairs, is equal to 44/9.

Exercise 9.14 In this exercise we will explore the implications of the condition for approximate universality in (9.53).

1. Show that if a parallel channel scheme satisfies the condition (9.53), then it achieves the diversity–multiplexing tradeoff of the parallel channel. *Hint*: Do Exercise 9.9 first.
2. Show that the diversity–multiplexing tradeoff can still be achieved even when the scheme satisfies a more relaxed condition:

$$|d_1 d_2 \cdots d_L|^{2/L} > c \cdot \frac{1}{L2^R}, \qquad \text{for some constant } c > 0. \tag{9.97}$$

Exercise 9.15 Consider the class of permutatation codes for the L-parallel channel described in Section 9.2.2. The codeword is described as $(q, \pi_2(q), \ldots, \pi_L(q))$ where q belongs to a normalized QAM (so that each of the I and Q channels are peak constrained by ± 1) with 2^{LR} points; so, the rate of the code is R bits/s/Hz per sub-channel. In this exercise we will see that this class contains approximately universal codes.

1. Consider *random* permutations with the uniform measure; since there are $2^{LR}!$ of them, each of the permutations occurs with probability $1/2^{LR}!$. Show that the average inverse product of the pairwise codeword differences, averaged over both the codeword pairs and the random permutations, is upper bounded as follows:

$$\mathbb{E}_{\pi_2,\ldots,\pi_L}\left[\frac{1}{2^{LR}(2^{LR}-1)} \times \sum_{q_1 \neq q_2} \frac{1}{|q_1-q_2|^2|\pi_2(q_1)-\pi_2(q_2)|^2 \cdots |\pi_L(q_1)-\pi_L(q_2)|^2}\right] \leq L^L R^L. \tag{9.98}$$

2. Conclude from the previous part that there exist permutations $\pi_2, \ldots, \pi_L$ such that

$$\frac{1}{2^{LR}} \sum_{q_1} \left(\sum_{q_2 \neq q_1} \frac{1}{|q_1-q_2|^2|\pi_2(q_1)-\pi_2(q_2)|^2 \cdots |\pi_L(q_1)-\pi_L(q_2)|^2} \right) \leq L^L R^L 2^{LR}. \tag{9.99}$$

3. Now suppose we fix q_1 and consider the sum of the inverse product of all the possible pairwise codeword differences:

$$f(q_1) := \sum_{q_2 \neq q_1} \frac{1}{|q_1-q_2|^2|\pi_2(q_1)-\pi_2(q_2)|^2 \cdots |\pi_L(q_1)-\pi_L(q_2)|^2}. \tag{9.100}$$

Since $f(q_1) \geq 0$, argue from (9.99) that at least half the QAM points q_1 must have the property that

$$f(q_1) \leq 2L^L R^L 2^{LR}. \tag{9.101}$$

Further, conclude that for such q_1 (they make up at least half of the total QAM points) we must have for every $q_2 \neq q_1$ that

$$|q_1-q_2|^2|\pi_2(q_1)-\pi_2(q_2)|^2 \cdots |\pi_L(q_1)-\pi_L(q_2)|^2 \geq \frac{1}{2L^L R^L 2^{LR}}. \tag{9.102}$$

4. Finally, conclude that there exists a permutation code that is approximately universal for the parallel channel by arguing the following:
   - Expurgating no more than half the number of QAM points only reduces the total rate LR by no more than 1 bit/s/Hz and thus does not affect the multiplexing gain.
   - The product distance condition on the permutation codeword differences in (9.102) does not quite satisfy the condition for approximate universality in (9.97). Relax the condition in (9.97) to

$$|d_1 d_2 \cdots d_L|^{2/L} > c \cdot \frac{1}{R2^R}, \qquad \text{for some constant } c > 0, \tag{9.103}$$

   and show that this is sufficient for a code to achieve the optimal diversity–multiplexing tradeoff curve.

Exercise 9.16 Consider the bit-reversal scheme for the parallel channel described in Section 9.2.2. Strictly speaking, the condition in (9.57) is not true for every integer between 0 and $2^R - 1$. However, the set of integers for which this is not true is small (i.e., expurgating them will not change the multiplexing rate of the scheme). Thus the bit-reversal scheme with an appropriate expurgation of codewords is approximately universal for the 2-parallel channel. A reading exercise is to study [118] where the expurgated bit-reversal scheme is described in detail.

Exercise 9.17 Consider the bit-reversal scheme described in Section 9.2.2 but with every alternate bit *flipped* after the reversal. Then for every pair of normalized codeword differences, it can be shown that

$$|d_1 d_2|^2 > \frac{1}{64 \cdot 2^{2R}}, \tag{9.104}$$

where the data rate is R bits/s/Hz per sub-channel. Argue now that the bit-reversal scheme with alternate bit flipping is approximately universal for the 2-parallel channel. A reading exercise is to study the proof of (9.104) in [118]. *Hint*: Compare (9.104) with (9.53) and use the result derived in Exercise 9.14.

Exercise 9.18 Consider a MISO channel with the fading channels from the n_t transmit antennas, $h_1, \ldots, h_{n_\mathrm{t}}$, i.i.d.

1. Show that

$$\mathbb{P}\left\{\log\left(1 + \frac{\mathsf{SNR}}{n_\mathrm{t}} \sum_{\ell=1}^{n_\mathrm{t}} |h_\ell|^2\right) < r \log \mathsf{SNR}\right\} \tag{9.105}$$

   and

$$\mathbb{P}\left\{\sum_{\ell=1}^{n_\mathrm{t}} \log(1 + \mathsf{SNR}|h_\ell|^2) < n_\mathrm{t} r \log \mathsf{SNR}\right\} \tag{9.106}$$

   have the same decay rate with increasing SNR.
2. Interpret (9.105) and (9.106) with the outage probabilities of the MISO channel and that of a parallel channel obtained through an appropriate transformation of the MISO channel, respectively. Argue that the conversion of the MISO channel into a parallel channel discussed in Section 9.2.3 is approximately universal for the class of i.i.d. fading coefficients.

Exercise 9.19 Consider an $n_t \times n_t$ matrix $\mathbf{D}$. Show that

$$\min_{\mathbf{h}:\|\mathbf{h}\|=1} \mathbf{h}^*\mathbf{D}\mathbf{D}^*\mathbf{h} = \lambda_1^2, \tag{9.107}$$

where λ_1 is the smallest singular value of $\mathbf{D}$.

Exercise 9.20 Consider the Alamouti transmit codeword (cf. (9.84)) with u_1, u_2 independent uncoded QAMs with 2^R points in each.

1. For every codeword difference matrix

$$\begin{bmatrix} d_1 & -d_2^* \\ d_2 & d_1^* \end{bmatrix}, \tag{9.108}$$

show that the two singular values are the same and equal to $\sqrt{|d_1|^2 + |d_2|^2}$.

2. With the codeword difference matrix normalized as in (9.68) and each of the QAM symbols u_1, u_2 constrained in power of $\mathsf{SNR}/2$ (i.e., both the I and Q channels are peak constrained by $\pm\sqrt{\mathsf{SNR}/2}$), show that if the codeword difference d_ℓ is not zero, then it is

$$|d_\ell|^2 \geq \frac{2}{2^R}, \qquad \ell = 1, 2.$$

3. Conclude from the previous steps that the square of the smallest singular value of the codeword difference matrix is lower bounded by $2/2^R$. Since the condition for approximate universality in (9.70) is an order-of-magnitude one (the constant factor next to the 2^R term does not matter, see Exercises 9.9 and 9.14), we have explicitly shown that the Alamouti scheme with uncoded QAMs on the two streams is approximately universal for the two transmit antenna MISO channel.

Exercise 9.21 Consider the D-BLAST architecture in (9.77) with just two interleaved streams for the 2×2 i.i.d. Rayleigh fading MIMO channel. The two streams are independently coded at rate $R = r \log \mathsf{SNR}$ bits/s/Hz each and composed of the pair of codewords $\left(x_A^{(\ell)}, x_B^{(\ell)}\right)$ for $\ell = 1, 2$. The two streams are coded using an approximately universal parallel channel code (say, the bit-reversal scheme described in Section 9.2.2).

A union bound averaged over the Rayleigh MIMO channel can be used to show that the diversity gain obtained by each stream with joint ML decoding is $4 - 2r$. A reading exercise is to study the proof of this result in [118].

Exercise 9.22 [67] Consider transmitting codeword matrices of length at least n_t on the $n_t \times n_r$ MIMO slow fading channel at rate R bits/s/Hz (cf. (9.71)).

1. Show that the pairwise error probability between two codeword matrices $\mathbf{X}_A$ and $\mathbf{X}_B$, conditioned on a specific realization of the MIMO channel $\mathbf{H}$, is

$$Q\left(\sqrt{\frac{\mathsf{SNR}}{2}\|\mathbf{H}\mathbf{D}\|^2}\right), \tag{9.109}$$

where $\mathbf{D}$ is the normalized codeword difference matrix (cf. (9.68)).

2. Writing the SVDs $\mathbf{H} := \mathbf{U}_1 \mathbf{\Psi} \mathbf{V}_1^*$ and $\mathbf{D} := \mathbf{U}_2 \mathbf{\Lambda} \mathbf{V}_2^*$, show that the pairwise error probability in (9.109) can be written as

$$Q\left(\sqrt{\frac{\mathsf{SNR}}{2} \|\Psi \mathbf{V}_1^* \mathbf{U}_2 \Lambda\|^2}\right). \tag{9.110}$$

3. Suppose the singular values are increasingly ordered in Λ and decreasingly ordered in Ψ. For fixed $\Psi, \Lambda, \mathbf{U}_2$, show that the channel eigendirections $\mathbf{V}_1^*$ that minimize the pairwise error probability in (9.110) are

$$\mathbf{V}_1 = \mathbf{U}_2. \tag{9.111}$$

4. Observe that the channel outage condition depends only on the singular values Ψ of $\mathbf{H}$ (cf. Exercise 9.8). Use the previous parts to conclude that the calculation of the worst-case pairwise error probability for the MIMO channel reduces to the optimization problem

$$\min_{\psi_1, \ldots, \psi_{n_{\min}}} \frac{\mathsf{SNR}}{2} \sum_{\ell=1}^{L} |\psi_\ell|^2 |\lambda_\ell|^2, \tag{9.112}$$

subject to the constraint

$$\sum_{\ell=1}^{n_{\min}} \log\left(1 + \frac{\mathsf{SNR}}{n_\mathrm{t}} |\psi_\ell|^2\right) \geq R. \tag{9.113}$$

Here we have written

$$\mathbf{\Psi} := \mathrm{diag}\{\psi_1, \ldots, \psi_{n_{\min}}\}, \quad \text{and} \quad \mathbf{\Lambda} := \mathrm{diag}\{\lambda_1, \ldots, \lambda_{n_\mathrm{t}}\}.$$

5. Observe that the optimization problem in (9.112) and the constraint (9.113) are very similar to the corresponding ones in the *parallel* channel (cf. (9.43) and (9.40), respectively). Thus the universal code design criterion for the MIMO channel is the same as that of a parallel channel (cf. (9.47)) with the following parameters:
   - there are $n_{\min}$ sub-channels,
   - the rate per sub-channel is $R/n_{\min}$ bits/s/Hz,
   - the parallel channel coefficients are $\psi_1, \ldots, \psi_{n_{\min}}$, the singular values of the MIMO channel, and
   - the codeword differences are the smallest singular values, $\lambda_1, \ldots, \lambda_{n_{\min}}$, of the codeword difference matrix.

Exercise 9.23 Using the analogy between the worst-case pairwise error probability of a MIMO channel and that of an appropriately defined parallel channel (cf. Exercise 9.22), justify the condition for approximate universality for the MIMO channel in (9.79).

Exercise 9.24 Consider transmitting codeword matrices of length $l \geq n_\mathrm{t}$ on the $n_\mathrm{t} \times n_\mathrm{r}$ MIMO slow fading channel. The total power constraint is SNR, so for any transmit codeword matrix $\mathbf{X}$, we have $\|\mathbf{X}\|^2 \leq l\mathsf{SNR}$. For a pair of codeword matrices $\mathbf{X}_A$ and $\mathbf{X}_B$, let the normalized codeword difference matrix be $\mathbf{D}$ (normalized as in (9.68)).

1. Show that $\mathbf{D}$ satisfies

$$\|\mathbf{D}\|^2 \le \frac{2}{\mathsf{SNR}}(\|\mathbf{X}_A\|^2 + \|\mathbf{X}_B\|^2) \le 4l. \tag{9.114}$$

2. Writing the singular values of $\mathbf{D}$ as $\lambda_1, \ldots, \lambda_{n_t}$, show that

$$\sum_{\ell=1}^{n_t} \lambda_\ell^2 \le 4l. \tag{9.115}$$

Thus, each of the singular values is upper bounded by $2\sqrt{l}$, a constant that does not increase with SNR.

Exercise 9.25 [152] Consider the following transmission scheme (spanning two symbols) for the two transmit antenna MIMO channel. The entries of the transmit codeword matrix $\mathbf{X} := [x_{ij}]$ are defined as

$$\begin{bmatrix} x_{11} \\ x_{22} \end{bmatrix} := \mathbf{R}(\theta_1) \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} x_{21} \\ x_{12} \end{bmatrix} := \mathbf{R}(\theta_2) \begin{bmatrix} u_3 \\ u_4 \end{bmatrix}. \tag{9.116}$$

Here u_1, u_2, u_3, u_4 are independent QAMs of size $2^{R/2}$ each (so the data rate of this scheme is R bits/s/Hz). The rotation matrix $\mathbf{R}(\theta)$ is (cf. (3.46))

$$\mathbf{R}(\theta) := \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{9.117}$$

With the choice of the angles θ_1, θ_2 equal to $1/2\tan^{-1} 2$ and $1/2\tan^{-1}(1/2)$ radians respectively, Theorem 2 of [152] shows that the determinant of every normalized codeword difference matrix $\mathbf{D}$ satisfies

$$|\det \mathbf{D}|^2 \ge \frac{1}{10 \cdot 2^R}. \tag{9.118}$$

Conclude that the code described in (9.116), with the appropriate choice of the angles θ_1, θ_2 above, is approximately universal for every MIMO channel with two transmit antennas.