

Full text available at: <http://dx.doi.org/10.1561/0100000002>

**Transmission and
Reception with
Multiple Antennas:
Theoretical
Foundations**

Full text available at: <http://dx.doi.org/10.1561/0100000002>

Full text available at: <http://dx.doi.org/10.1561/0100000002>

Transmission and Reception with Multiple Antennas: Theoretical Foundations

Ezio Biglieri

*Politecnico di Torino
Corso Duca degli Abruzzi 24
Torino I-10129, Italy
ezio.biglieri@polito.it*

Giorgio Taricco

*Politecnico di Torino
Corso Duca degli Abruzzi 24
Torino I-10129, Italy
giorgio.taricco@polito.it*

now

the essence of knowledge

Boston – Delft

Full text available at: <http://dx.doi.org/10.1561/0100000002>

Foundations and Trends[®] in Communications and Information Theory

Published, sold and distributed by:

now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1 781 871 0245
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:

now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

A Cataloging-in-Publication record is available from the Library of Congress

Printed on acid-free paper

ISBN: 1-933019-01-8; ISSNs: Paper version 1567-2190; Electronic version 1567-2328

© 2004 E. Biglieri and G. Taricco

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Contents

1	Introduction	1
2	Preliminaries	3
3	Channel models	9
3.1	Fading channel	9
3.2	Dynamics of the fading channel	12
3.3	MIMO channel	15
3.4	Narrowband multiple-input, multiple-output channel models	17
3.5	Channel state information	20
4	Channel capacity	21
4.1	Deterministic channel	22
4.2	Ergodic Rayleigh fading channel	25
4.3	Correlated channels	36

vi *Contents*

4.4	A critique to asymptotic analysis [29]	38
4.5	Nonergodic Rayleigh fading channel	40
5	Influence of channel-state information	53
5.1	Perfect CSI at the receiver	53
5.2	Imperfect CSI	54
5.3	Imperfect CSI at the receiver: General guidelines	57
5.4	CSI at the transmitter and at the receiver	66
6	Coding for multiple-antenna systems	71
6.1	Maximum likelihood detection	72
7	Some practical coding schemes	79
7.1	Delay diversity	79
7.2	Alamouti scheme	79
7.3	Alamouti scheme revisited: Orthogonal designs	82
7.4	Linear space–time codes	83
7.5	Trellis space–time codes	85
7.6	Space–time codes when CSI is not available	85
8	Suboptimum receiver interfaces	87
8.1	Linear interfaces	88
8.2	Nonlinear interfaces	92
9	The fundamental tradeoff	101
9.1	2×2 schemes	104
9.2	Orthogonal designs	105
9.3	Zero-forcing vertical BLAST	105
9.4	Zero-forcing diagonal BLAST	106
9.5	MIMO transmission in a multiuser environment	106
9.6	Derivation of the diversity–rate tradeoff curve	107
A	Complex random variables and vectors	111

A.1	Gaussian random variables	112
A.2	Real random vectors	112
A.3	Complex random vectors	113
B	Results from information theory	117
B.1	A basic inequality	117
B.2	Capacity of the Gaussian MIMO channel	118
B.3	Ergodic capacity	120
C	Random matrices	123
D	Numerical calculation of error probabilities	127
D.1	Application: MIMO channel	129
E	Two proofs	133
E.1	Proof of (4.23)	133
E.2	Proof of (4.51)	137
	Acknowledgements	145
	References	147
	Notations and Acronyms	157

Full text available at: <http://dx.doi.org/10.1561/0100000002>

1

Introduction

The fundamental characteristic of wireless communication, that is, its ability to connect users in motion, represents an attractive feature for users, but at the same time a strong challenge to systems designers. The appeal of wireless systems is documented by their commercial success, with projections indicating that in the near future the number of their subscribers will exceed that of wireline subscribers. Concurrently, users expect a growing variety of sophisticated services along with increasingly higher quality, which contrasts with the difficulties involved with transmission on a mobile radio channel. On one hand, multimedia services require high data rates, while mobility requires networking with seamless connectivity. On the other hand, bandwidth and power are limited, interferences from other users impair transmission quality, mobility causes rapid variations of the physical channel, portability demands a long battery life, etc. All this makes it very difficult to provide high-quality mobile services.

Up until recently, it was accepted that the degrees of freedom to be allocated to system design could only be obtained from time or bandwidth. Of late, it was discovered that *space*, obtained by increasing the number of transmit and receive antennas, can also effectively generate

2 Introduction

degrees of freedom, and hence expand the range of choices made available to the design. The use of many antennas at the receiver side is not a new proposition, but the option of simultaneously transmitting information from multiple antennas was originally left aside, because of the *spatial* interference that degrades the received signal. The use of multiple transmit antennas was initially examined in the context of “beam-forming” techniques [20, 81]. More recently, the discovery of the stunningly high channel capacities that multiple antennas can yield [39, 40, 106], of receiver architectures that make it possible to mitigate spatial interference with limited complexity [39], and of “space–time” codes that can be used to approach capacity [104], made it clear that the introduction of multiple antennas in the wireless link, and consequently of signal processing in the space and time domains, is not only a useful addition to existing systems, but may be needed to provide the necessary transmission quality.

This paper focuses on the main aspects of single-user multiple-antenna theory, with the goal of presenting a comprehensive, yet compact, survey, emphasizing its mathematical aspects. After describing channel models, we compute the capacities they achieve, we briefly overview “space–time” codes, and we describe how suboptimum architectures can be employed to simplify the receiver. Since each one of the several topics presented here might be expanded to a full article length, some of them could only receive here a cursory treatment, or no treatment at all. We regret in particular that no discussion was possible of the bandwidth–power tradeoff in the wideband regime characterized by low spectral efficiency and low energy per bit [112], of multiuser channels [47], and of questions regarding applications of multiple antennas to wireless LANs and to cellular systems ([45] and references therein). The effects of channel state information known at the transmitter, or not known at all, would also warrant a more thorough discussion.

2

Preliminaries

We consider a radio transmission whereby t antennas simultaneously transmit one signal each, and r antennas receive these signals (Fig. 2.1). Assume two-dimensional elementary constellations throughout, and the additive presence of Gaussian noise only.¹ The channel input–output relationship is

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \quad (2.1)$$

where $\mathbf{x} \in \mathbb{C}^t$, $\mathbf{y} \in \mathbb{C}^r$, $\mathbf{H} \in \mathbb{C}^{r \times t}$ (i.e., \mathbf{H} is an $r \times t$ complex matrix whose entries h_{ij} describe the fading gains from the j th transmit to the i th receive antenna), and \mathbf{z} is zero-mean circularly-symmetric complex Gaussian noise. The component x_i , $i = 1, \dots, t$, of vector \mathbf{x} is the signal transmitted from antenna i ; the component y_j , $j = 1, \dots, r$, of vector \mathbf{y} is the signal received by antenna j . We also assume that the noise components affecting the different receivers are independent with variance N_0 , i.e.,

$$\mathbb{E}[\mathbf{z}\mathbf{z}^\dagger] = N_0\mathbf{I}_r, \quad (2.2)$$

where \mathbf{I}_r is the $r \times r$ identity matrix, and the signal energy is constrained by $\mathbb{E}[\mathbf{x}^\dagger\mathbf{x}] = tE_s$ where E_s denotes the energy per symbol.

¹ The presence of external interferences is considered, among others, in [16, 17, 77].

4 Preliminaries

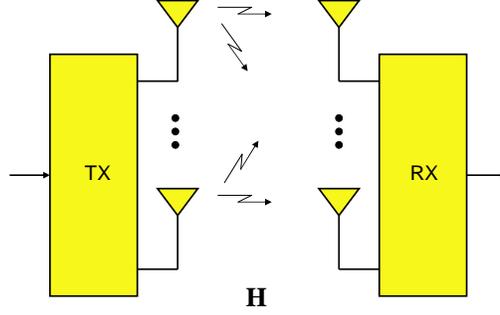


Fig. 2.1 Transmission and reception with multiple antennas. The gains of each propagation path are described by the $r \times t$ matrix \mathbf{H} .

The additional assumption that $\mathbb{E}[|h_{ij}|^2] = 1$ for all i, j ² yields the average signal-to-noise ratio (SNR) at the receiver

$$\rho = \frac{tE_sR_s}{N_0W}, \quad (2.3)$$

where R_s denotes the symbol rate and W the signal bandwidth. In the following we further assume that $R_s = W$ (Nyquist signalling rate), so that the relation

$$\rho = \frac{tE_s}{N_0} \quad (2.4)$$

holds. Then, rather than assuming a power or energy constraint, we will refer to an SNR constraint, i.e.,

$$\mathbb{E}[\mathbf{x}^\dagger \mathbf{x}] \leq \rho N_0. \quad (2.5)$$

Finally, we define the energy per bit $E_b \triangleq tE_s/\mu_b$, where μ_b denotes the number of information bits transmitted per symbol interval.

Explicitly, we have from (2.1):

$$y_j = \sum_{i=1}^t h_{ij}x_i + z_j, \quad j = 1, \dots, r \quad (2.6)$$

² The assumption of equal second-order moments for the channel coefficients facilitates the analysis, but is somewhat restrictive, as it does not allow consideration of antennas differing in their radiation patterns, etc. See, e.g., [34, 78].

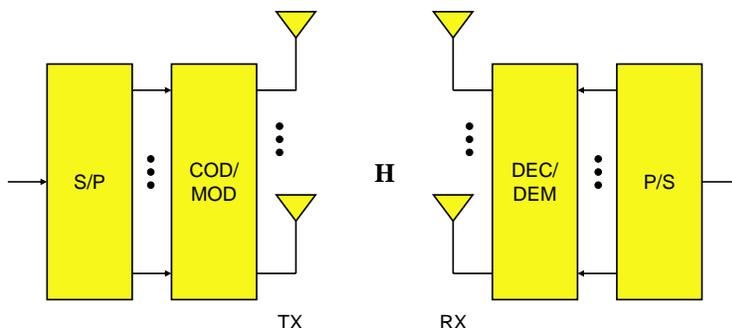


Fig. 2.2 Spatial multiplexing and diversity obtained by transmission and reception with multiple antennas.

which shows how every component of the received signal is a linear combination of the signals emitted by each transmit antenna. We say that \mathbf{y} is affected by spatial interference from the signals transmitted by the various antennas. This interference has to be removed, or controlled in some way, in order to sort out the multiple transmitted signals. We shall see in the following how this can be done: for the moment we may just observe that the tools for the analysis of multiple-antenna transmission have much in common with those used in the study of other disciplines centering on interference control, like digital equalization of linear dispersive channels [5] (where the received signals are affected by *intersymbol* interference) or multiuser detection [111] (where the received signals are affected by *multiple-access* interference).

The upsides of using multiple antennas can be summarized by defining two types of gain. As we shall see in the following, in the presence of fading a multiplicity of transmit antennas creates a set of parallel channels, that can be used to potentially increase the data rate up to a factor of $\min\{t, r\}$ (with respect to single-antenna transmission) and hence generate a *rate gain*. This corresponds to the *spatial multiplexing* depicted in Fig. 2.2. Here the serial-to-parallel converter S/P distributes the stream of data across the transmit antennas; after reception, the

6 Preliminaries

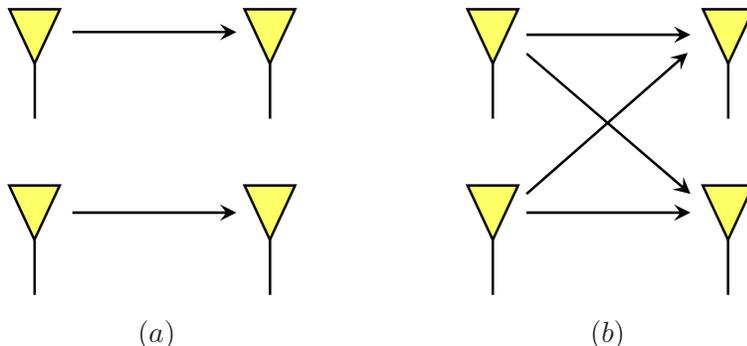


Fig. 2.3 Spatial multiplexing through transmission and reception with multiple antennas.

original stream is reconstituted by the parallel-to-serial converter P/S^3 . The other gain is due to the combination of received signals that are independently faded replicas of a single transmitted signal, which allows a more reliable reception. We call *diversity gain* the number of independent paths traversed by each signal, which has a maximum value rt . We hasten to observe here that these two gains are not independent, but there is a fundamental tradeoff between the two: and actually it can be said that the problem of designing a multiple-antenna system is based on this tradeoff. As an example, Fig. 2.3 illustrates the diversity–rate tradeoff for a multiple-input multiple-output (MIMO) system with $t = 2$ transmit and $r = 2$ receive antennas. Fig. 2.3(a) assumes the channels are orthogonal, so that the rate is maximum (twice as large as the single-channel rate) but there is no diversity gain since each symbol is transmitted on one channel only. Fig. 2.3(b) assumes that the transmitter replicates the same signal over the two channels, so that there is no rate gain, but the diversity is now four, since the signal travels four independent paths. We shall further discuss this point in Section 9.

³ Here we limit ourselves to considering only transmissions with the same rate on all antennas. However, different (and possibly adaptive) modulation rates can also be envisaged [26].

The problems we address in this paper are the following:

- (1) What is the limiting performance (channel capacity) of this multiple-antenna system?
- (2) What is its error probability?
- (3) How can we design codes matched to the channel structure (*space-time* codes)?
- (4) How can we design architectures allowing simple decoding of space-time codes, and what is their performance?

As mentioned in the Introduction, the above points will be addressed in different depths. In particular, (3) and (4) deserve a treatment more thorough than we could provide here.

Full text available at: <http://dx.doi.org/10.1561/0100000002>

3

Channel models

In a wireless communication system the transmitted signal is assumed to arrive at the receiver after propagation through several different paths. This phenomenon is referred to as *multipath propagation*. Signal propagation depends on different electromagnetic effects such as scattering, refraction, reflection, or diffraction caused by other objects present in the communication scenario. All these combine together, and produce an overall attenuation of the signal amplitude accompanied by a phase rotation which is accounted for by a complex gain coefficient. Due to the difficulty of modeling the geometry of the propagation, it is assumed that these coefficients change randomly: the common term used to denote this physical effect is *fading*. We start dealing with fading from the point of view of single-input, single-output (SISO) communication systems, and further extend the results to MIMO systems.

3.1 Fading channel

Fading occurs in a multipath communication scenario over different time and frequency scales. Its effects are perceived depending on the time and frequency characteristics of the transmitted signal. The fading

attenuation is commonly decomposed as the sum of three components: *i*) path loss; *ii*) macroscopic fading; and *iii*) microscopic fading.

Path loss. This is an attenuation that depends on the distance d between transmitter and receiver, and on the antenna parameters. In the open space, the received power is approximately given by the Friis formula

$$P_r = P_t \left(\frac{\lambda_c}{4\pi d} \right)^2 G_t G_r, \quad (3.1)$$

where P_t is the transmitted power, λ_c is the carrier wavelength, and G_t , G_r are the transmit and receive antenna gains, respectively. Assuming that the wavelength λ_c is much smaller than the size of the obstacles encountered in the signal path, in a land-mobile communication environment there are usually two main propagation paths: direct and surface-reflected. The received power is approximately given by [64]:

$$P_r = P_t \left(\frac{h_t h_r}{d^2} \right)^2 G_t G_r, \quad (3.2)$$

where h_t and h_r are the transmit and the receive antenna heights, respectively. Notice that in this case the path loss is independent of the carrier wavelength, in contrast with the case of free-space propagation.

Macroscopic fading. This is mainly due to the presence of obstacles in the propagation path causing an effect called *shadowing*. This attenuation is called macroscopic because its time scale is longer than the time scale of microscopic fading due to the relatively slow changes in the communication scenario.¹ Experimental results show that the logarithm of this attenuation is approximately Gaussian distributed. A typical value for the standard deviation is 8 dB [64].

Microscopic fading. This is the result of scattered signal components received from the (possibly) moving objects populating the communication scenario. The presence of many scatterers produces the

¹ Time scales here are a consequence of motion, and hence it may be more appropriate to talk about *spatial* scales. These are on the order of the wavelength for microscopic fading, and on the order of size of the obstacles encountered for macroscopic fading. See [33, 64].

superposition of a large number of independent path gains that tends to be distributed as a complex Gaussian random variable. Thus, assuming the presence of a direct path, the received signal complex amplitude gain can be modeled, in the presence of L propagation paths, as

$$\Gamma = A_d + \sum_{i=1}^L A_i e^{j\Theta_i},$$

where A_d is a constant due to the presence of a direct path while the amplitudes A_i and phases Θ_i are random variables (RVs) depending on the individual path gains. These RVs are reasonably assumed to be independent, as they are due to different scatterers, with phases uniformly distributed over $(0, 2\pi)$. Then, their sum

$$\Gamma_s \triangleq \sum_{i=1}^L A_i e^{j\Theta_i}$$

has independent, identically distributed (iid) real and imaginary parts which, by the Central Limit Theorem, converge in distribution, as $L \rightarrow \infty$, to a pair of independent Gaussian random variables.

Summarizing, the fading gain is modeled by a complex Gaussian random variable with independent real and imaginary parts having the same variance and (possibly) different means. After normalization of the second moment of the fading gain, it is assumed that the real part has mean $(K/(K+1))^{1/2}$ and the common variance is equal to $1/(2(K+1))$, so that K represents the direct-to-diffuse power ratio, i.e., the ratio of the squared mean to the sum of the two variances, which represent the diffuse power component. This is the *Rician* fading model; under the assumption that $\mathbb{E}[R^2] = 1$, the probability density function (pdf) of the random channel gain R , taking values $r \geq 0$, is given by

$$p_{\text{Rice}}(r) = 2r(1+K) \exp\{-(1+K)r^2 - K\} I_0\left(2r\sqrt{K(1+K)}\right) u(r), \quad (3.3)$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function defined as

$$I_0(x) \triangleq \frac{1}{\pi} \int_0^\pi e^{x \cos \phi} d\phi \quad (3.4)$$

and $u(\cdot)$ denotes the unit-step function: $u(x) = 1$ for $x > 0$, and $u(x) = 0$ otherwise. As $K \rightarrow 0$, i.e., as the direct path reduces its power, and hence only diffuse power is received, since $I_0(0) = 1$ the Rice pdf tends to the *Rayleigh* pdf

$$p_{\text{Rayleigh}}(r) = 2re^{-r^2}u(r). \quad (3.5)$$

Conversely, if K grows, i.e., the fixed-path power becomes considerably higher than the diffuse power, then the Gaussian pdf is a good approximation for the Rician density. In particular, as $K \rightarrow \infty$ (no diffuse power) we obtain the unfaded-channel pdf:

$$p_{\text{unfaded}}(r) = \delta(r - 1). \quad (3.6)$$

Another popular statistical model for the envelope R of the fading is the Nakagami- m distribution. The probability density function of R is

$$p_{\text{Nakagami}}(r) = \frac{2m^m}{\Gamma(m)}r^{2m-1}e^{-mr^2}u(r). \quad (3.7)$$

The parameter $m \geq 1/2$, called the *fading figure*, is the inverse of the variance of R^2 . As special cases, the choice $m = 1$ yields the Rayleigh distribution, $m = 1/2$ yields a single-sided Gaussian distribution, and $m \rightarrow \infty$ yields the unfaded-channel pdf.

3.2 Dynamics of the fading channel

Consider the complex baseband equivalent model of a fading channel. Let $h(\theta; \tau)$ be the response of a fading channel at time θ to an impulse transmitted at time $\theta - \tau$, namely, $\delta(t - (\theta - \tau))$, in the absence of receiver noise. We interpret θ as *time* and τ as *delay*. Denoting the transmitted and received signals as $x(t)$ and $y(t)$, respectively, we have the following channel equation:

$$y(t) = \int_{-\infty}^{\infty} h(t; \tau)x(t - \tau) d\tau + z(t), \quad (3.8)$$

where $z(t)$ is a complex white Gaussian random process with independent real and imaginary parts having power spectral density $N_0/2$.

A typical representation of the fading channel impulse response is

$$h(t; \tau) = \sum_{\ell=1}^L h_{\ell}(t)\delta(\tau - \tau_{\ell}), \quad (3.9)$$

representing the linear superposition of L signal paths with gain $h_\ell(t)$ and delay τ_ℓ . The received signal corresponding to the transmission of $x(t)$ is given by

$$y(t) = \sum_{\ell=1}^L h_\ell(t)x(t - \tau_\ell). \quad (3.10)$$

Further simplifications of the model consist of assuming that the impulse response is wide-sense stationarity with respect to time, i.e.,

$$\mathbb{E}[h(t_1; \tau)h^*(t_2; \tau)] = R_h(t_1 - t_2; \tau), \quad (3.11)$$

and uncorrelated for different delays, i.e.,

$$\mathbb{E}[h(t; \tau_1)h^*(t; \tau_2)] = 0 \quad \text{if } \tau_1 \neq \tau_2. \quad (3.12)$$

The latter assumption yields the so-called *uncorrelated scattering*. As a further simplification comprising the above assumptions we set

$$\mathbb{E}[h(t_1; \tau_1)h^*(t_2; \tau_2)] \triangleq R_h(t_1 - t_2; \tau_1)\delta(\tau_1 - \tau_2), \quad (3.13)$$

which implies stationarity in the frequency domain. In fact, defining the Fourier transform

$$H(t; f) \triangleq \int_{-\infty}^{\infty} h(t; \tau)e^{-j2\pi f\tau} d\tau, \quad (3.14)$$

we obtain

$$\begin{aligned} & \mathbb{E}[H(t_1; f_1)H^*(t_2; f_2)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{E}[h(t_1; \tau_1)h^*(t_2; \tau_2)]e^{-j2\pi f_1\tau_1 + j2\pi f_2\tau_2} d\tau_1 d\tau_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_h(t_1 - t_2; \tau_1)\delta(\tau_1 - \tau_2)e^{-j2\pi f_1\tau_1 + j2\pi f_2\tau_2} d\tau_1 d\tau_2 \\ &= \int_{-\infty}^{\infty} R_h(t_1 - t_2; \tau_1)e^{-j2\pi(f_1 - f_2)\tau_1} d\tau_1 \\ &\triangleq R_H(t_1 - t_2; f_1 - f_2). \end{aligned} \quad (3.15)$$

A double Fourier transform of this autocorrelation function yields the *scattering function* of the fading channel:

$$S_H(\tau; \lambda) \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_H(\Delta t; \Delta f)e^{-j2\pi(\lambda\Delta t - \tau\Delta f)} d\Delta t d\Delta f, \quad (3.16)$$

where τ represents the delay and λ the frequency offset. Marginalization with respect to τ and λ yields the spectra

$$S_H(\tau) \triangleq \int_{-\infty}^{\infty} S_H(\tau; \lambda) d\lambda \quad (3.17)$$

and

$$S_H(\lambda) \triangleq \int_{-\infty}^{\infty} S_H(\tau; \lambda) d\tau. \quad (3.18)$$

The delay spread T_D and the Doppler-bandwidth spread B_D can be defined as the sizes of the intervals where $S_H(\tau) \neq 0$ and $S_H(\lambda) \neq 0$, respectively, if $S_H(\tau)$ and $S_H(\lambda)$ have bounded support. Otherwise, T_D and B_D can be defined as standard deviations, or 3-dB fallouts, of the complete support of $S_H(\tau)$ and $S_H(\lambda)$. These two quantities, whose operational meaning will be discussed soon, describe the delay and Doppler-frequency variations due to fading.

Other key parameters are the coherence time T_c , the coherence bandwidth B_c , and the coherence angle θ_c . Qualitatively, T_c , B_c , and θ_c are, respectively, the duration of a time interval, a bandwidth, and a range of angles of arrival, where the fading does not alter significantly the spectrum of the received signal. It can be shown [64] that the Doppler-bandwidth spread and the delay spread are linked to the coherence time and bandwidth by the approximate relations

$$B_c \approx \frac{1}{T_D} \quad \text{and} \quad T_c \approx \frac{1}{B_D}. \quad (3.19)$$

Comparison of these parameters against the symbol time and the signal bandwidth allows a classification of fading channels. Assume transmission of a signal with duration T and bandwidth W . Then:

- If $T_c < T$, the fading channel is *fast*, or *time-selective*. Otherwise, it is *slow*, or *time-nonselective*.
- If $B_c < W$, the fading channel is *frequency-selective*. Otherwise, it is *frequency-flat*, or *frequency-nonselective*.

Since the signal bandwidth and duration are approximately the inverse of each other: $W \approx 1/T$, and since for most channels of interest $B_D \ll B_c$, then we have a classification of the fading channel based on the signal bandwidth:

- If $W < B_D$, the fading channel is *fast* and *frequency-nonselective*.
- If $B_D < W < B_c$, the fading channel is *slow* and *frequency-nonselective*.
- If $B_c < W$, the fading channel is *slow* and *frequency-selective*.

If the fading channel is selective both in time and frequency, the Doppler-bandwidth spread has to exceed the coherence bandwidth, namely, $B_D > B_c$. This situation occurs, for example, in avionic communications, where the mobile speed and the transmission delay are very large. Some typical values for indoor wireless communications are: carrier frequency of 1 GHz, delay spread of 10 ns, and mobile speed of 3 m/s. We have

$$B_D = \frac{v}{\lambda} = 10 \text{ Hz} \qquad B_c \approx \frac{1}{T_m} = 100 \text{ MHz}$$

Thus, if the signal bandwidth W ranges from 100 Hz to 10 MHz, the fading process is slow and frequency-nonselective.

3.3 MIMO channel

According to the classification given in the previous section we can list equivalent discrete-time channel models depending on the type of fading considered. The reader should be warned that the classification here is rather coarse, as it hinges upon models simple enough to allow tractable analysis.

For fast, frequency-nonselective channels we have

$$\mathbf{y}_n = \mathbf{H}_n \mathbf{x}_n + \mathbf{z}_n \tag{3.20}$$

with \mathbf{H}_n , $-\infty < n < \infty$, an independent, identically distributed random process. This fading model is usually referred to as *ergodic*. The reason for this classification lies in the fact that, during transmission, a long enough code word experiences essentially all states of the channel, and hence it “averages out” the channel randomness.

For slow, frequency-nonselective channels the model becomes

$$\mathbf{y}_n = \mathbf{H} \mathbf{x}_n + \mathbf{z}_n \tag{3.21}$$

and each code word, however long, experiences only one channel state. In this case we talk about *nonergodic fading*.

For slow, frequency-selective channels we have

$$\mathbf{y}_n = \sum_{\ell=1}^L \mathbf{H}_\ell \mathbf{x}_{n-\ell} + \mathbf{z}_n. \quad (3.22)$$

In this case, the delay spread T_D is large compared to the symbol duration T , which causes several delayed copies of the input signal to combine linearly to produce the output. It is assumed that the channel is so slow that the matrix coefficients \mathbf{H}_ℓ remain constant throughout the transmission of the code word. In other words, it is required that $NT \ll T_c$, or $W \gg NB_D$, which adds to the frequency-selectivity condition yielding

$$W \gg \max\{B_c, NB_D\}. \quad (3.23)$$

This model applies, for instance, to spread-spectrum communications.

In the following we focus our attention on narrowband communications only (for treatments of MIMO broadband fading channels, and in particular of the impact of frequency selectivity on capacity and on receiver structures, see, e.g., [1, 19]).

3.3.1 More on ergodicity: Interleaving, and the block-fading model

The nonergodic channel model is applicable to a wireless system with mobile terminals moving no faster than walking speed, so that the channel gain, albeit random, varies so slowly with time that it can be assumed as constant along transmission of a long block of data (see also [31, 39, 40, 84, 104, 106]). For each transmitted block, the channel matrix takes on an independent realization. Generally, blocks can be transmitted in separate times (e.g., in a time-division system [80]), in separate frequencies (e.g., in a multicarrier system), or in separate times and frequencies (e.g., with slow time-frequency hopping [23, 67, 68]). This observation elicits the definition of the *block-fading* (BF) channel model (see [8] and references therein).

Assume for example a wireless transmission system whose Doppler bandwidth spread ranges from 1 to 100 Hz, corresponding to coherence

times from 0.01 to 1 s, and whose transmission rates go from $2 \cdot 10^4$ to $2 \cdot 10^6$ symbols per second. Here, blocks whose lengths range from $2 \cdot 10^4 \times 0.01 = 200$ symbols to $2 \cdot 10^6 \times 1 = 2 \cdot 10^6$ symbols are affected by approximately the same fading gain. This channel can be made ergodic if interleaving is introduced. Notice that interleaving and deinterleaving involve a delay that for some applications may not be acceptable. In this case, separation of coded symbols by more than the coherence time of the channel is not possible, and therefore a length- N block is affected by a number of independent fading gains which is less than N . In the block-fading model, each word is split into a number of blocks which is a fraction of N , and over each block the channel fading is correlated so highly that we may model it as constant.

When delay constraints are present, the block-fading model turns out to be the sensible choice in many instances. This model assumes that the fading gain process is piecewise constant, and can be described through a sequence of independent random variables, each of which is the fading gain in a block of ν elementary signals. A code word of length N is spread over F blocks of length ν symbols each, so that $N = \nu F$. If $\nu = N$, and hence $F = 1$ (no interleaving), we have a channel in which the entire code word is affected by the same fading gain. If $\nu = 1$, and hence $F = N$ (ideal interleaving) each symbol is affected by an independent fading gain, which shows that the independent fading channel model (3.20) is a special case of this model.

The delay constraint to which the communication system is subject determines the maximum number F of blocks over which a code word of length $N = \nu F$ can be spread. The choice $F \rightarrow \infty$ makes the channel ergodic.

3.4 Narrowband multiple-input, multiple-output channel models

Assume again that the $r \times t$ channel matrix \mathbf{H} remains constant during the transmission of an entire code word. Analysis of this channel requires the joint pdf of the rt entries of \mathbf{H} . A number of relatively simple models for this pdf have been proposed in the technical literature, based on experimental results and analyses. Among these we consider

the following:

Rich scattering. The entries of \mathbf{H} are independent circularly-symmetric complex zero-mean Gaussian random variables [39, 40].

Fully correlated. The entries of \mathbf{H} are correlated circularly-symmetric complex zero-mean Gaussian random variables. The correlation coefficients of all the pairs of elements are required to specify this model.

Separately correlated. The entries of \mathbf{H} are correlated circularly-symmetric complex zero-mean Gaussian random variables, with the correlation between two entries of \mathbf{H} separated in two factors accounting for the receive and transmit correlation:

$$\mathbb{E}[(\mathbf{H})_{i,j}(\mathbf{H})_{i',j'}^*] = (\mathbf{R})_{i,i'}(\mathbf{T})_{j,j'} \quad (3.24)$$

for two given Hermitian nonnegative definite matrices \mathbf{R} ($r \times r$) and \mathbf{T} ($t \times t$). This model is justified by the fact that only the surrounding objects at the receiver and at the transmitter cause the local antenna-elements correlation, while they have no impact on the correlation at the other end of the link [76,94]. The channel matrix can be expressed as

$$\mathbf{H} = \mathbf{R}^{1/2}\mathbf{H}_u\mathbf{T}^{1/2}, \quad (3.25)$$

where \mathbf{H}_u is a matrix of uncorrelated, circularly-symmetric complex zero-mean Gaussian random variables with unit variance, and $(\cdot)^{1/2}$ denotes matrix square root.²

For a fair comparison of different correlation cases, we assume that the total average received power is constant, i.e.,

$$\begin{aligned} \mathbb{E}[\text{Tr}(\mathbf{T}\mathbf{H}_u\mathbf{R}\mathbf{H}_u^\dagger)] &= \sum_{i,j,k,\ell} \mathbb{E}[(\mathbf{T})_{ij}(\mathbf{H}_u)_{jk}(\mathbf{R})_{k\ell}(\mathbf{H}_u)_{i\ell}^*] \\ &= \sum_{i,k} (\mathbf{T})_{ii}(\mathbf{R})_{kk} \\ &= \text{Tr}(\mathbf{T})\text{Tr}(\mathbf{R}) \\ &= tr \end{aligned} \quad (3.26)$$

² The square root of matrix $\mathbf{A} \geq \mathbf{0}$ whose singular-value decomposition (SVD) [62] is $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^\dagger$ is defined as $\mathbf{A}^{1/2} \triangleq \mathbf{U}\mathbf{D}^{1/2}\mathbf{V}^\dagger$.

Since \mathbf{H} is not affected if \mathbf{T} is scaled by a factor $\alpha \neq 0$ and \mathbf{R} by a factor α^{-1} , we assume without loss of generality that

$$\text{Tr}(\mathbf{T}) = t \quad \text{and} \quad \text{Tr}(\mathbf{R}) = r. \quad (3.27)$$

Uncorrelated keyhole. The rank of \mathbf{H} may be smaller than $\min\{t, r\}$. A special case occurs when \mathbf{H} has rank one (“key-hole” channel). Assume $\mathbf{H} = \mathbf{h}_r \mathbf{h}_t^\dagger$, with the entries of the vectors \mathbf{h}_r and \mathbf{h}_t being independent, circularly-symmetric, complex zero-mean Gaussian random variables [30, 44]. This model applies in the presence of walls that the propagating signal passes through a small aperture, such as a keyhole. In this way, the incident electric field is a linear combination of the electric fields arriving from the transmit antennas, and irradiates through the hole after scalar multiplication by the scattering cross-section of the keyhole. As a result, the channel matrix can be written as the product of a column vector by a row vector. Similar phenomena arise in indoor propagation through hallways or tunnels, which can be thought of as overmoded waveguides at microwave frequencies [30].

Notice that the channel matrix entries are mutually uncorrelated but not independent (in fact, \mathbf{h}_r and \mathbf{h}_t are independent, and hence $\mathbb{E}[\mathbf{H}_{ij} \mathbf{H}_{i'j'}^*] = \mathbb{E}[\mathbf{h}_{r,i} \mathbf{h}_{t,j}^* \mathbf{h}_{r,i'}^* \mathbf{h}_{t,j'}] = \mathbb{E}[\mathbf{h}_{r,i} \mathbf{h}_{r,i'}^*] \mathbb{E}[\mathbf{h}_{t,j}^* \mathbf{h}_{t,j'}] = 0$ whenever $i \neq i'$ or $j \neq j'$). The magnitude of each matrix entry is given by the product of two independent Rayleigh-distributed random variables. Denoting by h a generic entry of \mathbf{H} , and assuming that the vectors \mathbf{h}_r and \mathbf{h}_t have unit variances, we have

$$\mathbb{P}(|h|^2 \leq \rho) = \int_0^\infty e^{-x} \int_0^{\rho/x} e^{-y} dy dx = 1 - \int_0^\infty e^{-x-\rho/x} dx. \quad (3.28)$$

Hence, $|h|^2$ has pdf

$$p(\rho) = \int_0^\infty \frac{1}{x} e^{-x-\rho/x} dx = 2K_0(2\sqrt{\rho}), \quad (3.29)$$

where $K_0(x)$ is the zeroth-order modified Bessel function of the

second kind, defined as

$$K_0(x) \triangleq \int_0^\infty \cos(x \sinh t) dt = \int_0^\infty \frac{\cos(xt)}{\sqrt{1+t^2}} dt \quad (3.30)$$

Rician channel. The channel models listed above are zero mean. However, for certain applications, the channel matrix \mathbf{H} should be modeled as having a nonzero mean. Rician MIMO channels are examined in [99].

3.5 Channel state information

As we shall discuss with more detail in the following, a crucial factor in determining the performance of a multi-antenna system is the availability, at the transmitting or at the receiving terminal, of the *channel-state information* (CSI), that is, the set of values taken on by the fading gains in each one of the transmission paths.

In a fixed wireless environment, the fading gains can be expected to vary slowly, so their estimate can be obtained by the receiver with a reasonable accuracy, even in a system with a large number of antennas, and possibly relayed to the transmitter. In some cases, we may assume that a partial knowledge of the CSI is available. One way of obtaining this estimate is by periodically sending pilot signals on the same channel used for data (these pilot signals are used in wireless systems also for acquisition, synchronization, etc.). We shall address this issue in Section 5.

4

Channel capacity

In this section we evaluate the capacity of the multiple-antenna, or MIMO (multiple-input, multiple-output) transmission system described by (2.1). Several models for the matrix \mathbf{H} will be considered [106]:

- (a) \mathbf{H} is deterministic.
- (b) \mathbf{H} is a random matrix, and each channel use (viz., the transmission of one symbol from each of the t transmit antennas) corresponds to an independent realization of \mathbf{H} (ergodic channel).
- (c) \mathbf{H} is a random matrix, but once it is chosen it remains fixed for the whole transmission (nonergodic channel).

When \mathbf{H} is random (cases (b) and (c) above) we assume here that its entries are $\sim \mathcal{N}_c(0, 1)$, i.e., iid Gaussian with zero-mean, independent real and imaginary parts, each with variance 1/2. Equivalently, each entry of \mathbf{H} has uniform phase and Rayleigh magnitude. This choice models Rayleigh fading with enough separation within antennas such that the fades for each transmit/receive antenna pair are independent. We also assume, unless otherwise stated, that the CSI (that is, the

realization of \mathbf{H}) is known at the receiver, while the distribution of \mathbf{H} is known at the transmitter (the latter assumption is necessary for capacity computations, since the transmitter must choose an optimum code for that specific channel).

As discussed in previous Section, model (b) is applicable to fast fading, or to slow fading with deep-enough interleaving, while model (c) is applicable to a slow fading channel in which delay constraints do not allow ergodicity to be achieved.

We will measure capacity in bits per dimension pair. This translates into bit/s/Hz if signal design allows one dimension pair per second to be transmitted in a 1-Hz bandwidth. Unless otherwise specified, $\log(\cdot)$ will denote the base-2 logarithm of (\cdot) .

4.1 Deterministic channel

Here we assume that the nonrandom value of \mathbf{H} is known at both transmitter and receiver. We derive the capacity by maximizing the average mutual information $I(\mathbf{x}; \mathbf{y})$ between input and output of the channel over the choice of the distribution of \mathbf{x} . Singular-value decomposition of the matrix \mathbf{H} yields [62]

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^\dagger, \quad (4.1)$$

where $\mathbf{U} \in \mathbb{C}^{r \times r}$ and $\mathbf{V} \in \mathbb{C}^{t \times t}$ are unitary, and $\mathbf{D} \in \mathbb{R}^{r \times t}$ is diagonal. We can write

$$\mathbf{y} = \mathbf{U}\mathbf{D}\mathbf{V}^\dagger \mathbf{x} + \mathbf{z}. \quad (4.2)$$

Premultiplication of (4.2) by \mathbf{U}^\dagger shows that the original channel is equivalent to the channel described by the input–output relationship

$$\tilde{\mathbf{y}} = \mathbf{D}\tilde{\mathbf{x}} + \tilde{\mathbf{z}}, \quad (4.3)$$

where $\tilde{\mathbf{y}} \triangleq \mathbf{U}^\dagger \mathbf{y}$, $\tilde{\mathbf{x}} \triangleq \mathbf{V}^\dagger \mathbf{x}$ (so that $\mathbb{E}[\tilde{\mathbf{x}}^\dagger \tilde{\mathbf{x}}] = \mathbb{E}[\mathbf{x}^\dagger \mathbf{x}]$), and $\tilde{\mathbf{z}} \triangleq \mathbf{U}^\dagger \mathbf{z} \sim \mathcal{N}_c(0, N_0 \mathbf{I}_r)$. Now, the rank of \mathbf{H} is at most $m \triangleq \min\{t, r\}$, and hence at most m of its singular values are nonzero. Denote these by $\sqrt{\lambda_i}$, $i = 1, \dots, m$, and rewrite (4.3) componentwise in the form

$$\tilde{y}_i = \sqrt{\lambda_i} \tilde{x}_i + \tilde{z}_i, \quad i = 1, \dots, m \quad (4.4)$$

which shows that transmission takes place on a set of m parallel equivalent channels. The remaining components of $\tilde{\mathbf{y}}$ (if any) are equal to

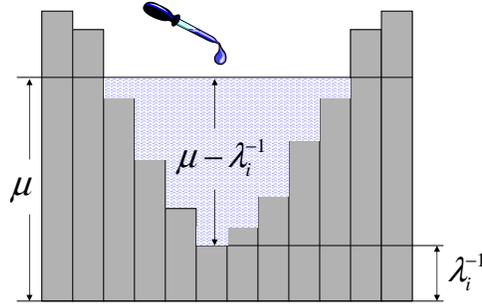


Fig. 4.1 Illustration of “water-filling.”

the corresponding components of the noise vector $\tilde{\mathbf{z}}$: we see that, for $i > m$, \tilde{y}_i is independent of the transmitted signal, and \tilde{x}_i does not play any role.

Maximization of the mutual information requires independent \tilde{x}_i , $i = 1, \dots, m$, each with independent Gaussian, zero-mean real and imaginary parts. Their variances should be chosen, as indicated in Appendix B, via “water-filling” (Fig. 4.1):

$$\mathbb{E}[\text{Re } \tilde{x}_i]^2 = \mathbb{E}[\text{Im } \tilde{x}_i]^2 = \frac{1}{2} (\mu - \lambda_i^{-1})_+ \quad (4.5)$$

where $(\cdot)_+ \triangleq \max(0, \cdot)$. This comes from a result of Information Theory concerning parallel channels [43]. With μ chosen so as to meet the SNR constraint, we see that the SNR, as parametrized by μ , is

$$\rho(\mu) = \sum_i (\mu - \lambda_i^{-1})_+ \quad (4.6)$$

and the capacity takes on the value (in bits per dimension pair)

$$C(\mu) = \sum_i (\log(\mu \lambda_i))_+. \quad (4.7)$$

Observation 4.1. Since the nonzero eigenvalues of $\mathbf{H}^\dagger \mathbf{H}$ are the same as those of $\mathbf{H} \mathbf{H}^\dagger$, the capacities of the channels corresponding to \mathbf{H} and to \mathbf{H}^\dagger are the same. A sort of “reciprocity” holds in this case [106].

Observation 4.2. Section B.2 illustrates a different method of obtaining the “water-filling” solution (4.5). The method is based on Hadamard’s inequality instead of *channel diagonalization* as in (4.4).

Example 4.3. Take $t = r = m$, and $\mathbf{H} = \mathbf{I}_m$. Due to the structure of \mathbf{H} , there is no spatial interference here, and transmission occurs over m parallel additive white Gaussian noise (AWGN) channels, each with SNR ρ/m and hence with capacity $\log(1 + \rho/m)$ bit/dimension pair. Thus,

$$C = m \log(1 + \rho/m). \quad (4.8)$$

We see here that we have a rate gain, as the capacity is proportional to the number of transmit antennas. Notice also that, as $m \rightarrow \infty$, the capacity tends to the limiting value

$$C = \rho \log e.$$

Example 4.4. Consider as \mathbf{H} the all-1 matrix. This corresponds to a maximum of spatial interference. Its SVD is

$$\mathbf{H} = \begin{bmatrix} \sqrt{1/r} \\ \sqrt{1/r} \\ \vdots \\ \sqrt{1/r} \end{bmatrix} (\sqrt{rt}) [\sqrt{1/t} \cdots \sqrt{1/t}]. \quad (4.9)$$

In this case we have $m = 1$, $\sqrt{\lambda_1} = \sqrt{rt}$, and hence $\lambda_1 = rt$. Thus, for $\rho > 0$,

$$\rho = \left(\mu - \frac{1}{rt} \right)_+ = \mu - \frac{1}{rt} \quad (4.10)$$

and hence the capacity is

$$C = \log \left[\left(\rho + \frac{1}{rt} \right) rt \right] = \log(1 + rt \rho). \quad (4.11)$$

The signals achieving this capacity can be described as follows. We have that $\tilde{\mathbf{x}}$ has only one component, and

$$\mathbf{V} = \sqrt{\frac{1}{t}} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (4.12)$$

Thus, the components of $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$ are all equal, i.e., the transmit antennas all send the same signal. If P denotes the total transmitted power, each transmit antenna sends a power P/t . Because of the structure of \mathbf{H} the signals *add coherently* at the receiver, so that at each receiver we have the voltage $t\sqrt{P/t}$, and hence the power $t^2P/t = Pt$. Since each receiver sees the same signal, and the noises are uncorrelated, the overall SNR is $rt\rho$, as shown by the capacity formula (4.11). In this case we see no rate gain, but a diversity gain is obtained through proper combination of the received signals. This result can be interpreted in the context of beam-forming [18].

4.2 Ergodic Rayleigh fading channel

We assume here that \mathbf{H} is independent of both \mathbf{x} and \mathbf{z} , with entries $\sim \mathcal{N}_c(0, 1)$. We also assume that for each channel use an independent realization of \mathbf{H} is drawn, so that the channel is ergodic (see Section 3.3). If the receiver has perfect CSI, the mutual information between the channel input (the vector \mathbf{x}) and its output (the pair \mathbf{y}, \mathbf{H}), is [32]:

$$I(\mathbf{x}; \mathbf{y}, \mathbf{H}) = I(\mathbf{x}; \mathbf{H}) + I(\mathbf{x}; \mathbf{y} \mid \mathbf{H}). \quad (4.13)$$

Since \mathbf{H} and \mathbf{x} are independent, then $I(\mathbf{x}; \mathbf{H}) = 0$, and hence

$$I(\mathbf{x}; \mathbf{y}, \mathbf{H}) = I(\mathbf{x}; \mathbf{y} \mid \mathbf{H}) = \mathbb{E}_{\tilde{\mathbf{H}}}[I(\mathbf{x}; \mathbf{y} \mid \mathbf{H} = \tilde{\mathbf{H}})], \quad (4.14)$$

where $\tilde{\mathbf{H}}$ denotes a realization of the random matrix \mathbf{H} . The maximum of $I(\mathbf{x}; \mathbf{y}, \mathbf{H})$, taken with respect to \mathbf{x} , yields the channel capacity C . From the results of Appendix B we know that the capacity, achieved by a transmitted signal $\mathbf{x} \sim \mathcal{N}_c(0, (\rho/t)\mathbf{I}_t)$, is equal to

$$C = \mathbb{E} \left[\log \det \left(\mathbf{I}_r + \frac{\rho}{t} \mathbf{H} \mathbf{H}^\dagger \right) \right]. \quad (4.15)$$

The exact computation of (4.15) will be examined soon. For the moment, note that for fixed r and as $t \rightarrow \infty$, the strong law of large numbers yields

$$\frac{1}{t} \mathbf{H} \mathbf{H}^\dagger \rightarrow \mathbf{I}_r \quad \text{a.s.} \quad (4.16)$$

Thus, as $t \rightarrow \infty$ the capacity tends to

$$\begin{aligned} \log \det(\mathbf{I}_r + \rho \mathbf{I}_r) &= \log(1 + \rho)^r \\ &= r \log(1 + \rho) \end{aligned} \quad (4.17)$$

so that it increases *linearly* with r . Compare this result with (4.11), where C increases with r only *logarithmically*.

A simplistic interpretation of the above would qualify fading as *beneficial* to MIMO transmission, as independent path gains generate r independent spatial channels. More accurately, it should be realized that high capacity is generated by a multiplicity of nonzero singular values in \mathbf{H} , which is typically achieved if \mathbf{H} is random matrix, but not if it is deterministic. For a quantitative expression of the above statement, see [61]. There, it is proved that if \mathbf{H} is the sum of a deterministic (line-of-sight) matrix \mathbf{D} and a random matrix whose entries are iid zero-mean unit-variance complex circularly-symmetric Gaussian random variables, then the capacity of the MIMO channel is monotonically non-decreasing in the singular values of \mathbf{D} .

4.2.1 A simple upper bound to C

The exact computation of C is rather intricate and yields an integral expression that is not easily amenable to analysis. Before examining it, we introduce an exceedingly simple upper bound to the value of C . Observe that the log-det function is concave over the set of nonnegative matrices [32, Th. 16.8.1]:

$$\log \det(\lambda \mathbf{K}_1 + (1 - \lambda) \mathbf{K}_2) \geq \lambda \log \det \mathbf{K}_1 + (1 - \lambda) \log \det \mathbf{K}_2 \quad (4.18)$$

for all $\lambda \in (0, 1)$. Therefore, by Jensen's inequality, we have

$$C = \mathbb{E} \left[\log \det \left(\mathbf{I}_r + \frac{\rho}{t} \mathbf{H} \mathbf{H}^\dagger \right) \right]$$

$$\begin{aligned} &\leq \log \det \left(\mathbf{I}_r + \frac{\rho}{t} \mathbb{E}[\mathbf{H}\mathbf{H}^\dagger] \right) \\ &= r \log(1 + \rho). \end{aligned} \quad (4.19)$$

Also, recalling that the matrices $\mathbf{H}\mathbf{H}^\dagger$ and $\mathbf{H}^\dagger\mathbf{H}$ share the same set of nonzero eigenvalues, and hence

$$\det(\mathbf{I}_r + (\rho/t)\mathbf{H}\mathbf{H}^\dagger) = \det(\mathbf{I}_t + (\rho/t)\mathbf{H}^\dagger\mathbf{H}), \quad (4.20)$$

we obtain the bound

$$\begin{aligned} C &= \mathbb{E} \left[\log \det \left(\mathbf{I}_t + \frac{\rho}{t} \mathbf{H}^\dagger \mathbf{H} \right) \right] \\ &\leq \log \det \left(\mathbf{I}_t + \frac{\rho}{t} \mathbb{E}[\mathbf{H}^\dagger \mathbf{H}] \right) \\ &= t \log(1 + r\rho/t). \end{aligned} \quad (4.21)$$

The upper limits are attained for $\mathbf{H}\mathbf{H}^\dagger = \mathbf{I}_r$ and $\mathbf{H}^\dagger\mathbf{H} = \mathbf{I}_t$, respectively.

By combining (4.19) and (4.21), we obtain the upper bound

$$C \leq \min \{ r \log(1 + \rho), t \log(1 + r\rho/t) \}. \quad (4.22)$$

Fig. 4.2 plots (4.22) in terms of capacity per transmit antenna (C/t), for $\rho = 0, 10$, and 20 dB, versus the ratio t/r . It also shows the asymptotic ($t, r \rightarrow \infty$ with $t/r \rightarrow$ the finite limit on the abscissa) capacity per transmit antenna of an ergodic independent Rayleigh fading MIMO channel (to be derived soon). It can be seen that the highest gap to the upper bound occurs when $t = r$, where the upper bound is from 20% (SNR=0 or 20 dB) to 27% (SNR=10 dB) higher than the asymptotic capacity per transmit antenna of the ergodic independent Rayleigh fading MIMO channel (Eq. (4.32) adapted).

4.2.2 Exact computation of C

We have the exact result (see Appendix E for its proof)

$$C = \log(e) \frac{m!}{(n-1)!} \sum_{\ell=0}^{m-1} \sum_{\mu=0}^m \sum_{p=0}^{\ell+\mu+n-m}$$

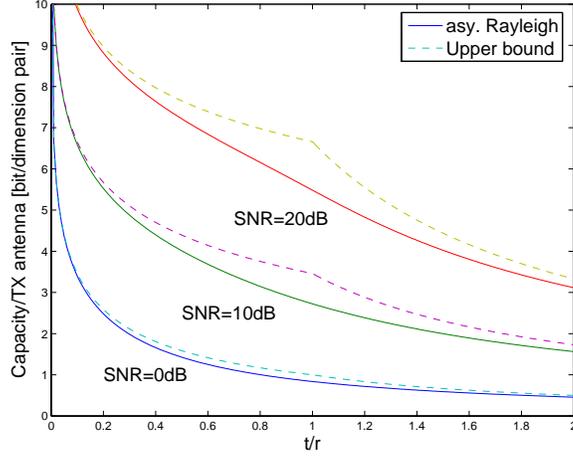


Fig. 4.2 Comparison of the capacity upper bound (4.22) with the (asymptotic) capacity of the independent Rayleigh fading channel when the SNR is 0, 10, and 20 dB.

$$\frac{(-1)^{\ell+\mu}(\ell + \mu + n - m)!}{\ell!\mu!} e^{t/\rho} E_{p+1}(t/\rho) \left[\binom{n-1}{m-1-\ell} \binom{n}{m-1-\mu} - \binom{n-1}{m-2-\ell} \binom{n}{m-\mu} \right], \quad (4.23)$$

where

$$E_n(x) \triangleq \int_1^\infty e^{-xy} y^{-n} dy$$

is the exponential integral function of order n .

Some exact capacity values for $\rho = 20$ dB are plotted in Fig. 4.3 and 4.4. Special cases, as well as asymptotic approximations to the values of C , are examined in the examples that follow.

Example 4.5. ($r \gg t$) Consider first $t = 1$, so that $m = 1$ and $n = r$. Application of (4.23) yields

$$C = \log(e) \sum_{k=1}^r e^{1/\rho} E_k(1/\rho). \quad (4.24)$$

This is plotted in Fig. 4.5. An asymptotic expression of C , valid as $r \rightarrow \infty$, can be obtained as follows. Using in (4.24) the approximation,

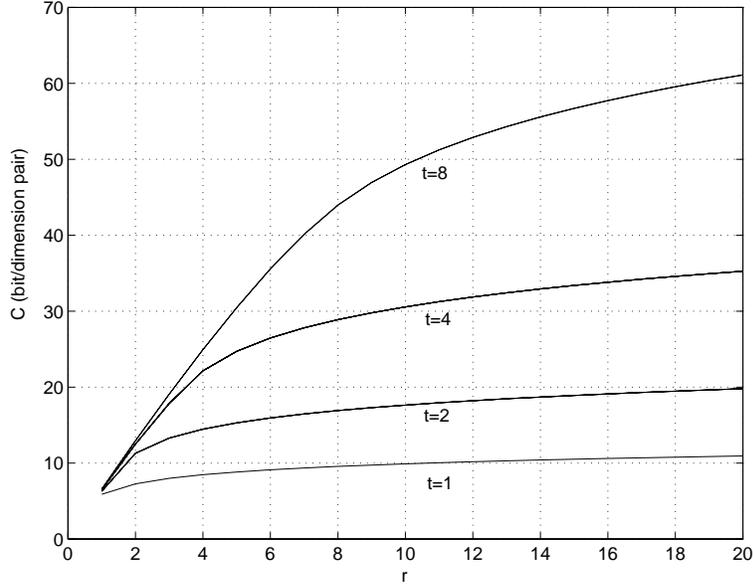


Fig. 4.3 Capacity of the ergodic Rayleigh MIMO channel with $\rho = 20$ dB.

valid for large k ,

$$e^x E_k(x) \sim \frac{1}{x + k}, \quad (4.25)$$

we obtain

$$C \sim \log(e) \sum_{k=1}^r \frac{1}{1/\rho + k} \approx \log(e) \int_0^r \frac{1}{1/\rho + x} dx = \log(1 + r\rho). \quad (4.26)$$

This approximation to the capacity is also plotted in Fig. 4.5. We see here that if $t = 1$ the capacity increases only logarithmically as the number of receive antennas is increased—a quite inefficient way of boosting capacity.

For finite $t > 1$ (and $r \rightarrow \infty$), we set $\mathbf{W} = \mathbf{H}^\dagger \mathbf{H} \rightarrow r \mathbf{I}_t$ a.s. Hence, the following asymptotic expression holds:

$$C = \log \det(\mathbf{I}_t + (\rho/t)\mathbf{W}) \rightarrow t \log(1 + (\rho/t)r). \quad (4.27)$$

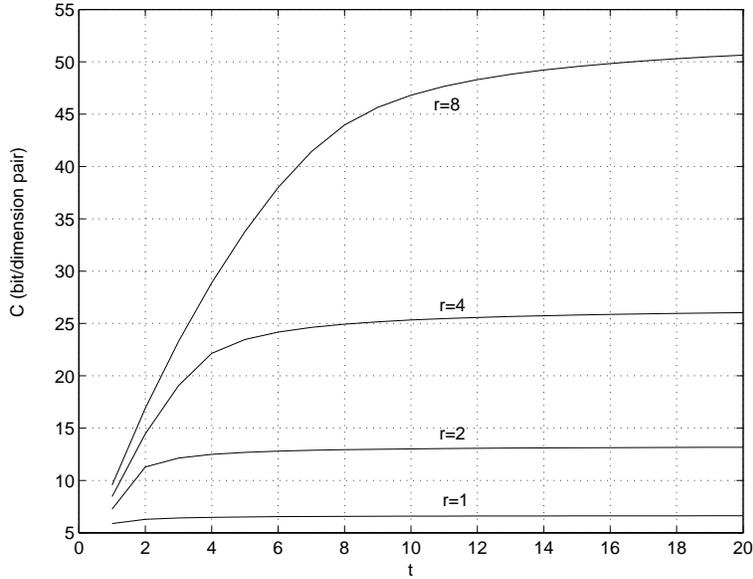


Fig. 4.4 Capacity of the ergodic Rayleigh MIMO channel with $\rho = 20$ dB.

Example 4.6. ($t \gg r$) Consider first $r = 1$, so that $m = 1$ and $n = t$. Application of (4.23) yields

$$C = \log(e) \sum_{k=1}^t e^{t/\rho} E_k(t/\rho). \quad (4.28)$$

This is plotted in Fig. 4.6. Proceeding as in Example 4.5, an asymptotic expression of C as $t \rightarrow \infty$ can be obtained, yielding $C \sim \log(1 + \rho)$. This approximation to the capacity is also plotted in Fig. 4.6.

For finite $r > 1$ (and $t \rightarrow \infty$), we set $\mathbf{W} = \mathbf{H}\mathbf{H}^\dagger \rightarrow t\mathbf{I}_r$ a.s. Hence, the following asymptotic expression holds:

$$C = \log \det(\mathbf{I}_r + (\rho/t)\mathbf{W}) \rightarrow r \log(1 + \rho). \quad (4.29)$$

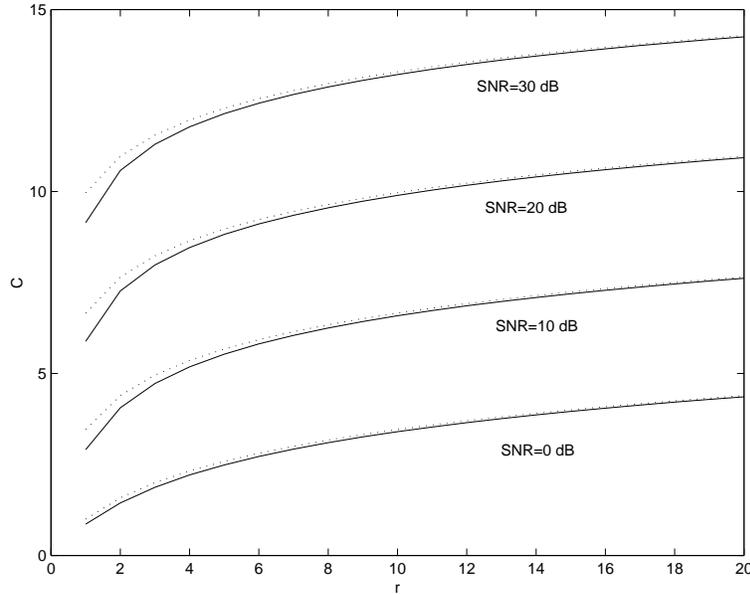


Fig. 4.5 Capacity of the ergodic Rayleigh MIMO channel with $t = 1$ (continuous line). The asymptotic approximation $C \sim \log(1 + \rho r)$ is also shown (dotted line).

Example 4.7. ($r = t$) With $r = t$ we have $m = n = r$, so that application of (4.23) yields

$$C = r \log(e) \sum_{\ell=0}^{r-1} \sum_{\mu=0}^r \sum_{p=0}^{\ell+\mu} \frac{(-1)^{\ell+\mu} (\ell + \mu)!}{\ell! \mu!} e^{t/\rho} E_{p+1}(t/\rho) \left[\binom{r-1}{\ell} \binom{r}{\mu+1} - \binom{r-1}{\ell+1} \binom{r}{\mu} \right].$$

The capacity is plotted in Fig. 4.7.

The results of Fig. 4.7 show that capacity increases almost linearly with m . This fact can be analyzed in a general setting by showing that when t and r both grow to infinity the capacity per antenna tends to a constant. To prove this, observe that with the usual definitions

32 Channel capacity

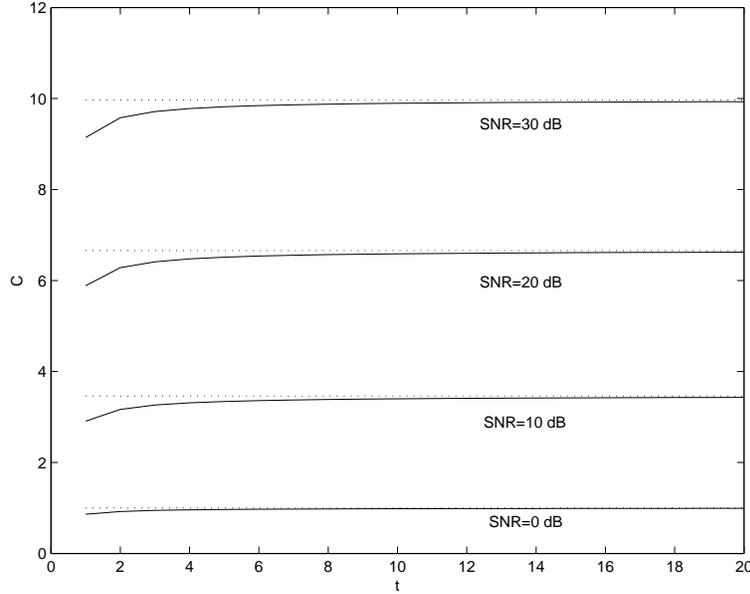


Fig. 4.6 Capacity of the ergodic Rayleigh MIMO channel with $r = 1$ (continuous line). The asymptotic approximation $C \sim \log(1 + \rho)$ is also shown (dotted line).

$m \triangleq \min\{t, r\}$ and $n \triangleq \max\{t, r\}$, (E.2) becomes

$$\frac{C}{m} \sim \mathbb{E} \left[\log \left(1 + \rho \frac{m}{t} \nu \right) \right], \quad (4.30)$$

where $\nu \triangleq \lambda_1/m$ is now a random variable whose pdf is known (see Theorem C.2): as $m \rightarrow \infty$ and n/m approaches a limit $\tau \geq 1$,

$$p(\nu) = \frac{1}{2\pi\nu} \sqrt{(\nu_+ - \nu)(\nu - \nu_-)} \quad (4.31)$$

with

$$\nu_{\pm} \triangleq (1 \pm \sqrt{\tau})^2$$

for $\nu_- \leq \nu \leq \nu_+$. The expectation in (4.30) can be computed in closed form [85, 11, 113], yielding

$$\frac{C}{m} = (\log(w_+\rho) + (1 - \alpha) \log(1 - w_-) - (w_- \alpha) \log e) \cdot \max\{1, 1/\alpha\}, \quad (4.32)$$

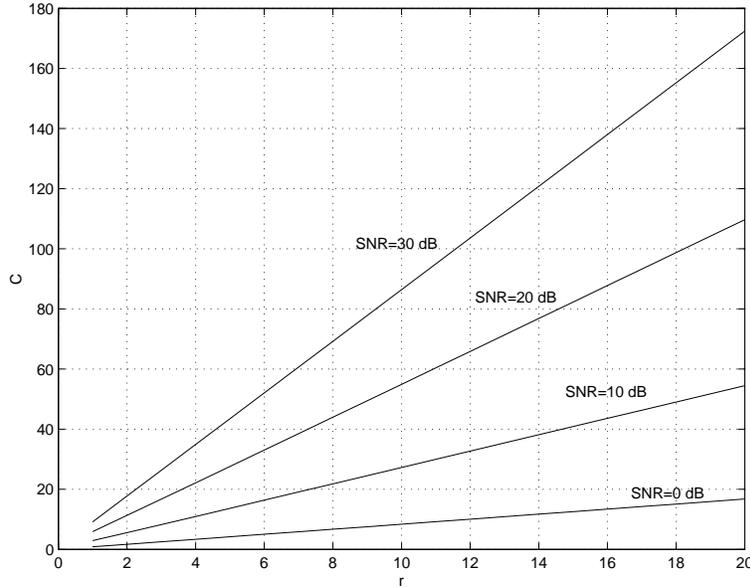


Fig. 4.7 Capacity with independent Rayleigh fading and $t = r$ antennas.

where

$$w_{\pm} \triangleq (w \pm \sqrt{w^2 - 4/\alpha})/2 \tag{4.33}$$

and

$$w \triangleq 1 + \frac{1}{\alpha} + \frac{1}{\rho}. \tag{4.34}$$

This asymptotic result can be used to approximate the value of C for finite r, t by setting $\alpha = t/r$. This approximation provides values very close to the true capacity even for small r and t , as shown in Figs. 4.8 and 4.9. The figures show the asymptotic value of C/m (for $t, r \rightarrow \infty$ with $t/r \rightarrow \alpha$) versus α and the nonasymptotic values of C/m corresponding to $r = 2$ and 4, respectively.

Observation 4.8. We can observe from (4.32) how, for large SNR, i.e., for $\rho \rightarrow \infty$, the ergodic capacity is asymptotically equal to $m \log \rho$: comparing this result with the asymptotic capacity of the single-input, single-output channel $C \sim \log \rho$, we see that use of multiple antennas

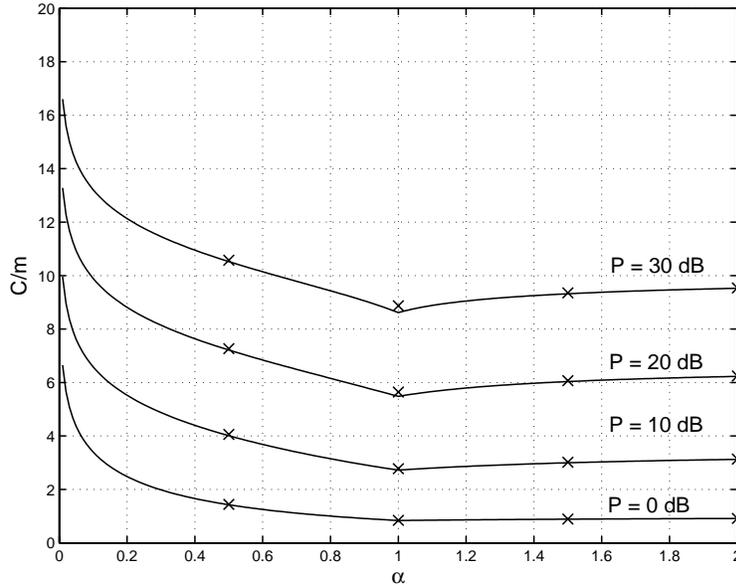


Fig. 4.8 Asymptotic ergodic capacity per antenna (C/m) with independent Rayleigh fading as $t, r \rightarrow \infty$ and $t/r \rightarrow \alpha$ (solid curves). The exact ergodic capacity per antenna for $r = 2$ is also shown for comparison (\times).

increases the capacity by a factor m . That is, multiple antennas generate m independent parallel channels. This explains why m is sometimes called the *number of degrees of freedom* generated by the MIMO system.

Observation 4.9. In [93], it is proved that for an uncorrelated keyhole channel

$$C \leq \log(1 + r\zeta).$$

This result can be interpreted by saying that this channel, regardless of the number of antennas, offers no rate gain.

Observation 4.10. For the validity of (4.32), it is not necessary to assume that the entries of \mathbf{H} are Gaussian, as needed for the preceding

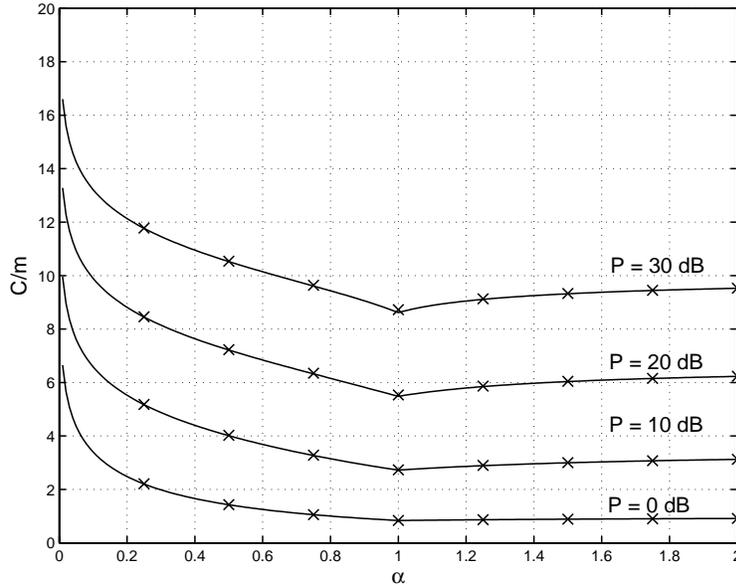


Fig. 4.9 Same as Fig. 4.8, but $r = 4$.

nonasymptotic results: a sufficient condition is that \mathbf{H} have iid entries with unit variance (Theorem C.2).

Observation 4.11. The reciprocity valid for deterministic channels (Observation 4.1) does not hold in this case. If $C(r, t, \rho)$ denotes the capacity of a channel with t transmit and r receive antennas, and SNR ρ , we have

$$C(a, b, \hat{\rho}b) = C(b, a, \hat{\rho}a). \quad (4.35)$$

Thus, for example, $C(r, 1, \hat{\rho}) = C(1, r, r\hat{\rho})$, which shows that with transmit rather than receive diversity we need r times as much transmit power to achieve the same capacity.

Observation 4.12. Choose $t = r = 1$ as the baseline; this yields *one* more bit per dimension pair for every 3 dB of SNR increase. In fact,

for large ρ ,

$$C = \log(1 + \rho) \sim \log \rho \quad (4.36)$$

and hence, if $\rho \rightarrow 2\rho$, we have

$$\log(2\rho) = 1 + \log \rho. \quad (4.37)$$

For multiple antennas with $t = r$, (4.30) shows that for every 3 dB of SNR increase we have t more bits per dimension pair.

4.3 Correlated channels

The separately-correlated MIMO channel model has been introduced in Section 3.4. The entries of the channel matrix are correlated, circularly-symmetric, complex zero-mean Gaussian random variables, and the channel matrix can be written as

$$\mathbf{H} = \mathbf{R}^{1/2} \mathbf{H}_u \mathbf{T}^{1/2}, \quad (4.38)$$

where \mathbf{H}_u is a matrix of independent, circularly-symmetric complex zero-mean Gaussian random variables with unit variance. The ergodic capacity of this channel is given by:

$$C = \mathbb{E} \left[\log \det \left(\mathbf{I}_r + \frac{\rho}{t} \mathbf{H}_u \mathbf{T} \mathbf{H}_u^\dagger \mathbf{R} \right) \right]. \quad (4.39)$$

Applying Jensen's inequality we obtain the following upper bound [93]:

$$\begin{aligned} C &\leq \log \mathbb{E} \left[\det \left(\mathbf{I}_r + \frac{\rho}{t} \mathbf{H}_u \mathbf{T} \mathbf{H}_u^\dagger \mathbf{R} \right) \right] \\ &= \log \left\{ \sum_{k=0}^m \left(\frac{\rho}{t} \right)^k k! \sum_{1 \leq i_1 < \dots < i_k \leq t} \det \left(\mathbf{T}_{i_1, \dots, i_k} \right) \sum_{1 \leq j_1 < \dots < j_k \leq r} \det \left(\mathbf{R}_{j_1, \dots, j_k} \right) \right\}, \end{aligned}$$

where $\mathbf{A}_{i_1, \dots, i_k}$ denotes the submatrix of \mathbf{A} obtained by extracting the rows and columns of indices i_1, \dots, i_k (i.e., $(\mathbf{A}_{i_1, \dots, i_k})_{rs} = A_{i_r, i_s}$ for $r, s = 1, \dots, k$). In the special case $t = r$, the upper bound can be simplified by taking only the highest power term ($k = m = t = r$) and yields the asymptotic approximation:

$$C \sim m \log(\rho/t) + \log(m!) + \log \det(\mathbf{TR}) \quad (4.40)$$

(in fact, the only set of indices $1 \leq i_1 < \dots < i_m \leq m$ is $i_1 = 1, \dots, i_m = m$, and hence the submatrices coincide with the full matrices themselves).

This result is tantamount to saying that, in the case of equal number of transmit and receive antennas, m , the asymptotic power loss of the separately correlated MIMO channel with respect to the uncorrelated channel is $-10 \log_{10} \det(\mathbf{TR})/m$ dB. Let $t_i, i = 1, \dots, m$ denote the positive eigenvalues of \mathbf{T} , and recall the trace constraint (3.27). We obtain

$$\det(\mathbf{T})^{1/m} = \prod_i t_i^{1/m} \leq \frac{1}{m} \sum_i t_i = 1.$$

A similar result applies to \mathbf{R} , so that we conclude that the power loss

$$-10 \log_{10} \det(\mathbf{TR})/m \geq 0$$

with equality if and only if $\mathbf{T} = \mathbf{R} = \mathbf{I}_m$. This confirms that, under the “fair comparison” conditions dictated by (3.27), the asymptotic power loss due to separate correlation is always positive, and zero only in the uncorrelated case. This proves the following asymptotic (in the SNR) statements:

- (Separate) correlation degrades system performance.
- The linear growth of capacity with respect to the minimum number of transmit/receive antennas is preserved.

The above can be extended, with the help of some algebra, to the case $t \neq r$.

Example 4.13. Consider the case of a constant separately-correlated $m \times m$ MIMO fading channel with correlation matrices

$$\mathbf{T} = \begin{pmatrix} 1 & \rho_T & \dots & \rho_T \\ \rho_T & 1 & \dots & \rho_T \\ \vdots & \vdots & \ddots & \vdots \\ \rho_T & \rho_T & \dots & 1 \end{pmatrix} \quad (4.41)$$

and

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_R & \cdots & \rho_R \\ \rho_R & 1 & \cdots & \rho_R \\ \vdots & \vdots & \ddots & \vdots \\ \rho_R & \rho_R & \cdots & 1 \end{pmatrix}. \quad (4.42)$$

Simple algebra leads to the asymptotic approximation

$$\begin{aligned} C \sim & m \log(\rho/t) + \log(m!) + (m-1) \log(1 - \rho_T) \\ & + \log(1 - \rho_T + m\rho_T) + (m-1) \log(1 - \rho_R) + \log(1 - \rho_T + m\rho_R). \end{aligned}$$

When m is large, the asymptotic power loss is about

$$-10 \log_{10}((1 - \rho_T)(1 - \rho_R)) \text{ dB}.$$

4.4 A critique to asymptotic analysis [29]

The previous results derived under the assumption $r \rightarrow \infty$ should be accepted *cum grano salis*. Our assumption that the entries of the channel-gain matrix \mathbf{H} are independent random variables becomes increasingly questionable as r increases. In fact, for this assumption to be justified the antennas should be separated by some multiple of the wavelength, which cannot be obtained when a large number of antennas is packed in a finite volume. Thus, as r increases the effects of correlation invalidates the assumption of independent channel gains. In addition, if the variance of the entries of \mathbf{H} does not depend on r , increasing r leads to an increased total received power, which becomes physically unacceptable beyond a certain value. It follows that capacity calculations for large r and a finite volume become quite involved. A simple, yet instructive, analysis is possible if the effects of varying correlation are disregarded, and a MIMO system is assumed whereby not only the total transmit power remains constant as t increases, but also the average received power remains constant when r increases. This is obtained by rescaling \mathbf{H} by a factor $r^{-1/2}$, so that the capacity (4.15)-(E.2) becomes

$$C = \mathbb{E} \left[\log \det \left(\mathbf{I}_r + \frac{\rho}{rt} \mathbf{H} \mathbf{H}^\dagger \right) \right] = \sum_{i=1}^m \mathbb{E} \log \left(1 + \frac{\rho}{rt} \lambda_i \right). \quad (4.43)$$

One simple heuristic way of dealing with this situation consists of rewriting C in the form

$$C = \mathbb{E} \left[\log \det \left(\mathbf{I}_t + \frac{\rho}{rt} \mathbf{H}^\dagger \mathbf{H} \right) \right]$$

and observing that, due to the strong law of large numbers, $(1/r)\mathbf{H}^\dagger \mathbf{H} \rightarrow \mathbf{I}_t$ almost surely. Thus,

$$C \rightarrow t \log(1 + \rho/t), \quad (4.44)$$

that is, the channel is transformed into a set of t independent parallel channels, each with capacity $\log(1 + \rho/t)$. As also t grows to infinity, from (4.44) we obtain

$$C \rightarrow \rho \log e, \quad (4.45)$$

a conclusion in contrast with our previous result that capacity increases linearly with the number of antennas. In a more rigorous fashion, using the inequality $\ln(1 + x) \leq x$, and observing that the trace of a matrix equals the sum of its eigenvalues, we obtain from (4.36)

$$C \leq \frac{\rho}{rt} \text{Tr} [\mathbb{E}(\mathbf{H}\mathbf{H}^\dagger)] \log e \rightarrow \rho \log e.$$

Conversely, observing that $x \geq 0$ implies $\ln(1 + x) \geq x - x^2/2$, and that $\sum_i \lambda_i^2 = \text{Tr}(\mathbf{H}\mathbf{H}^\dagger \mathbf{H}\mathbf{H}^\dagger)$, we obtain

$$C \geq \rho \left\{ 1 - \frac{1}{2r^2 t^2} \text{Tr} [\mathbb{E}(\mathbf{H}\mathbf{H}^\dagger \mathbf{H}\mathbf{H}^\dagger)] \right\} \log e. \quad (4.46)$$

The expectation can be calculated as follows:

$$\begin{aligned} & \mathbb{E}[\text{Tr}(\mathbf{H}\mathbf{H}^\dagger \mathbf{H}\mathbf{H}^\dagger)] \\ &= \sum_{i=1}^r \sum_{j=1}^r \sum_{k=1}^t \sum_{\ell=1}^t \mathbb{E}[(\mathbf{H})_{i,k} (\mathbf{H})_{j,k}^* (\mathbf{H})_{i,\ell}^* (\mathbf{H})_{j,\ell}] \\ &= \sum_{i=1}^r \sum_{k=1}^t \sum_{\ell=1}^t \mathbb{E}[|(\mathbf{H})_{i,k}|^2 |(\mathbf{H})_{i,\ell}|^2] + \sum_{i=1}^r \sum_{\substack{j=1 \\ j \neq i}}^r \sum_{k=1}^t \mathbb{E}[|(\mathbf{H})_{i,k}|^2 |(\mathbf{H})_{j,k}|^2] \\ &= [rt(t-1) + r(r-1)t] \{\mathbb{E}[|(\mathbf{H})_{i,k}|^2]\}^2 + rt \mathbb{E}[|(\mathbf{H})_{i,k}|^4] \\ &= tr(t+r), \end{aligned} \quad (4.47)$$

since the pdf of $|(\mathbf{H})_{i,k}|^2$ is $e^{-x}u(x)$ and hence $\mathbb{E}[|(\mathbf{H})_{i,k}|^{2m}] = m!$. Inserting (4.47) in (4.46) we get

$$C \geq \rho \left[1 - \frac{\rho(r+t)}{2rt} \right] \log e. \quad (4.48)$$

In conclusion, for a fixed ρ , as r and t increase we have again $C \rightarrow \rho \log e$.

4.5 Nonergodic Rayleigh fading channel

When \mathbf{H} is chosen randomly at the beginning of the transmission, and held fixed for all channel uses, average capacity has no meaning, as the channel is nonergodic. In this case the quantity to be evaluated is outage probability, that is, the probability that the transmission rate R exceeds the mutual information of the channel [39, 40, 42, 106].

Under these conditions, the mutual information, that we shall call (with a suggestive but theoretically inaccurate terminology) *instantaneous capacity*, is the random variable

$$C(\mathbf{H}) = \log \det \left(\mathbf{I}_r + \frac{\rho}{t} \mathbf{H} \mathbf{H}^\dagger \right) \quad (4.49)$$

and the outage probability is defined as

$$P_{\text{out}}(R) \triangleq \mathbb{P}(C(\mathbf{H}) < R). \quad (4.50)$$

The maximum rate that can be supported by the channel with a given outage probability is referred to as the *outage capacity*.

The evaluation of (4.50) should be done by Monte Carlo simulation. However, one can profitably use an asymptotic result which states that, as t and r grow to infinity, the instantaneous capacity $C(\mathbf{H})$ tends to a Gaussian random variable. This result has been recently obtained independently by several authors under slightly different technical assumptions [46, 58, 91, 77], and its value is strongly enhanced by the fact that $C(\mathbf{H})$ is very well approximated by a Gaussian random variable even for small numbers of antennas. Following in part the approach proposed in [91, 77], which is based on a technique borrowed from statistical mechanics and called the *Replica Method*, we prove in Appendix E that $C(\mathbf{H})$ is asymptotically Gaussian. Thus, by computing its asymptotic mean and variance, we characterize its asymptotic behavior.

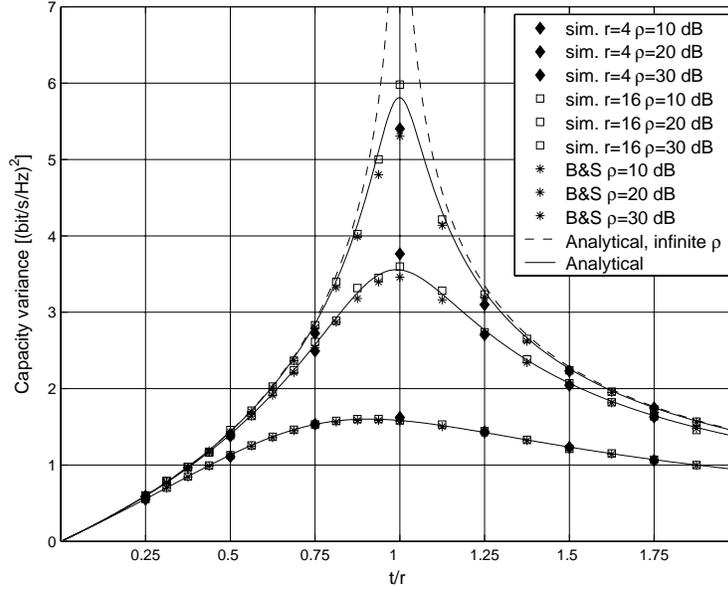


Fig. 4.10 Variance of the nonergodic capacity. Continuous line: Asymptotic variance from [91, 77]. (*): Numerical integration of Bai-Silverstein [3] expression for the variance. (◆) and (□): Monte Carlo simulation.

In practice, the asymptotic mean and variance of $C(\mathbf{H})$ yield a close approximation to the statistics of $C(\mathbf{H})$ even for very small r and t . Thus, the outage probability can be closely approximated for any pair t, r in the form

$$P_{\text{out}}(R) \approx Q\left(\frac{\mu_C - R}{\sigma_C}\right), \quad (4.51)$$

where, from Appendix E,

$$\begin{aligned} \mu_C &\triangleq -t\left\{(1 + \beta) \log w + q_0 r_0 \log e + \log r_0 + \beta \log(q_0/\beta)\right\} \\ \sigma_C^2 &\triangleq -\log e \cdot \log(1 - q_0^2 r_0^2/\beta) \end{aligned}$$

expressed in bit/dimension pair and (bit/dimension pair)², respectively, with $w \triangleq \sqrt{1/\rho}$, $\beta \triangleq \alpha^{-1}$, and q_0, r_0 defined in (E.41). (An alternative expression for the asymptotic capacity variance in the form of an integral was obtained in [3]). Numerical results are shown in Fig. 4.10.

42 Channel capacity

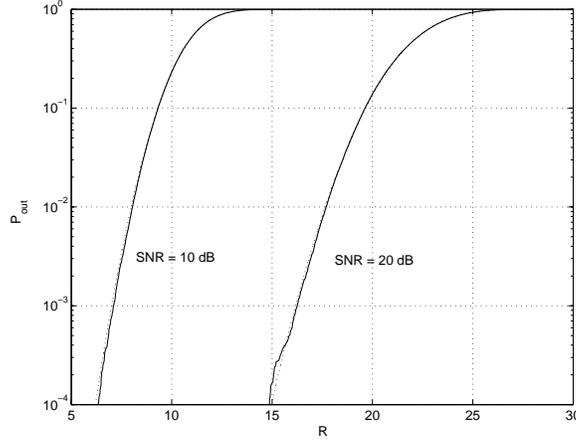


Fig. 4.11 Outage probability for $r = t = 4$ and a nonergodic Rayleigh channel vs. R , the transmission rate in bits per dimension pair. The continuous line shows the results obtained by Monte Carlo simulation, while the dashed line shows the normal approximation.

Fig. 4.11, which plots P_{out} versus ρ for $r = t = 4$ and two values of SNR, shows the quality of the Gaussian approximation for $r = t = 4$ and a Rayleigh channel.

Based on these results, we can approximate closely the outage probabilities as in Figs. 4.12 and 4.13. These figures show the rate that can be supported by the channel for a given SNR and a given outage probability, that is, from (4.51):

$$R = \mu_C - \sigma_C Q^{-1}(P_{\text{out}}). \quad (4.52)$$

Notice how as r, t increase the outage probabilities curves come closer to each other: in fact, as r and t grow to infinity the channel tends to an ergodic channel.

Fig. 4.14 shows the outage capacity (at $P_{\text{out}} = 0.01$) of an independent Rayleigh fading MIMO channel.

An asymptotic expression for R as $\rho \rightarrow \infty$ can be obtained. Observe that

$$q_0 = \begin{cases} \frac{\beta}{1-\beta}w + O(w^3) & \beta < 1 \\ 1 - w/2 + O(w^2) & \beta = 1 \\ \frac{\beta-1}{w} + \frac{1}{\beta-1}w + O(w^3) & \beta > 1 \end{cases} \quad (4.53)$$

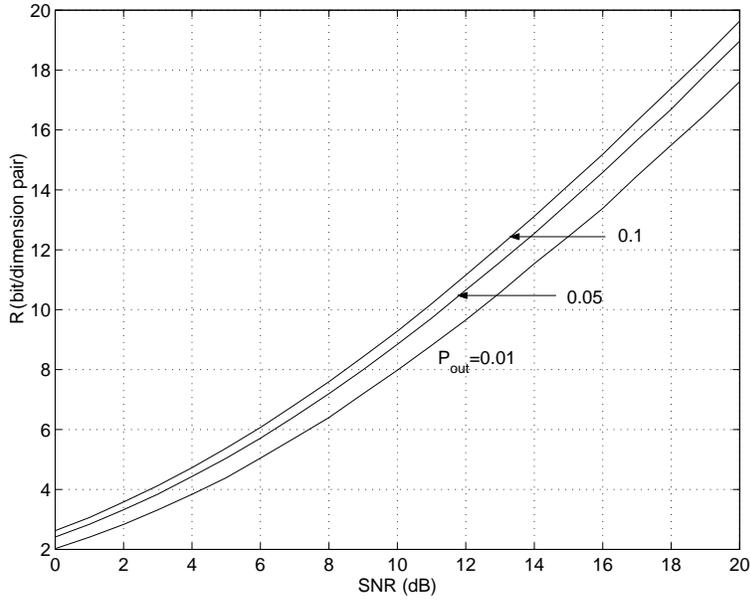


Fig. 4.12 Transmission rate that can be supported with $r = t = 4$ and a given outage probability by a nonergodic Rayleigh channel. The results are based on the Gaussian approximation.

and

$$r_0 = \begin{cases} \frac{1-\beta}{w} + \frac{\beta}{1-\beta}w + O(w^3) & \beta < 1 \\ 1 - w/2 + O(w^2) & \beta = 1 \\ \frac{1}{\beta-1}w + O(w^3) & \beta > 1 \end{cases} \quad (4.54)$$

Since from the above we obtain

$$1 - \frac{q_0^2 r_0^2}{\beta} = \begin{cases} 1 - \beta + O(w^2) & \beta < 1 \\ 1 - w/2 + O(w^2) & \beta = 1 \\ 1 - \frac{1}{\beta} + O(w^2) & \beta > 1 \end{cases}, \quad (4.55)$$

44 Channel capacity

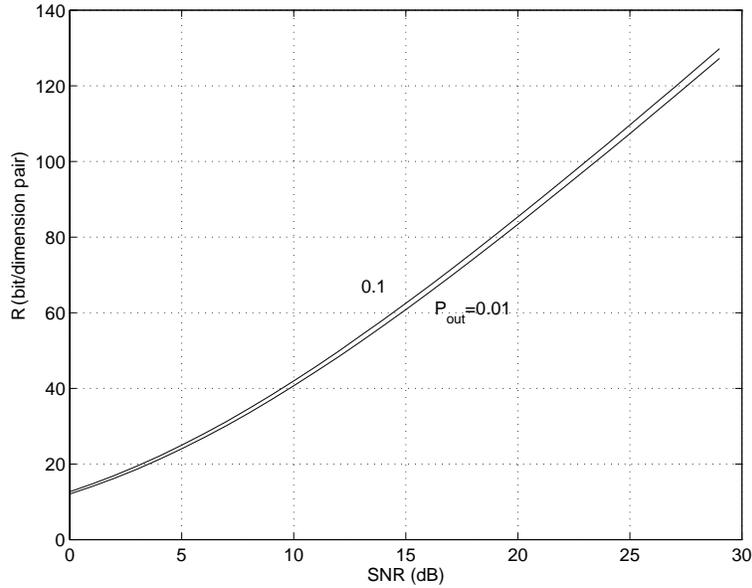


Fig. 4.13 Transmission rate that can be supported with $r = t = 16$ and a give outage probability by a nonergodic Rayleigh channel. The results are based on the Gaussian approximation.

then, as $\rho \rightarrow \infty$, we have

$$\sigma_C^2 \sim \begin{cases} \ln(1 - \beta) & \beta < 1 \\ \ln(2w) & \beta = 1 \\ \ln(1 - 1/\beta) & \beta > 1 \end{cases} . \quad (4.56)$$

Similarly, we can obtain, as $\rho \rightarrow \infty$,

$$\mu_C \sim m \log \rho, \quad (4.57)$$

a result compatible with previous observations. Using (4.56) and (4.57) in (4.52), we see that $R \sim m \log \rho$, which shows that the outage capacity, for $\rho \rightarrow \infty$, behaves as the ergodic capacity.

4.5.1 Block-fading channel

Here we take the approach of choosing a block-fading channel model, introduced before (Section 3.3.1) and shown in Fig. 4.15. Here the

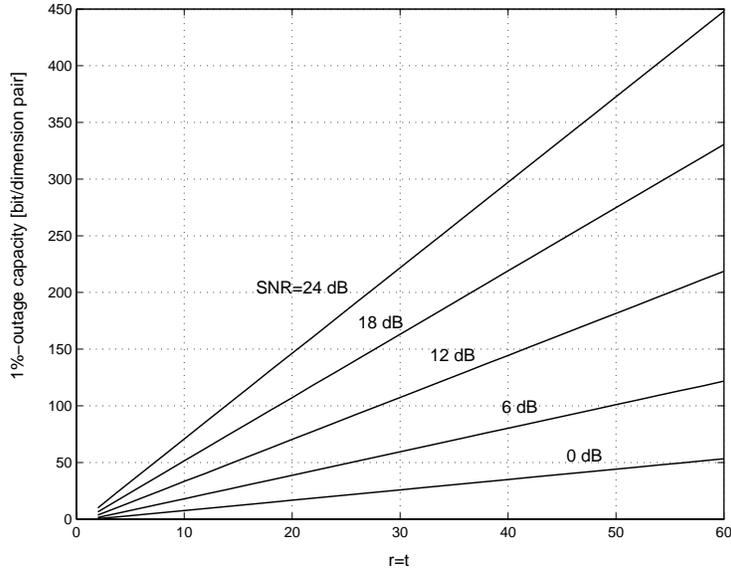


Fig. 4.14 Outage capacity (at $P_{\text{out}} = 0.01$) with independent Rayleigh fading and $r = t$ antennas.

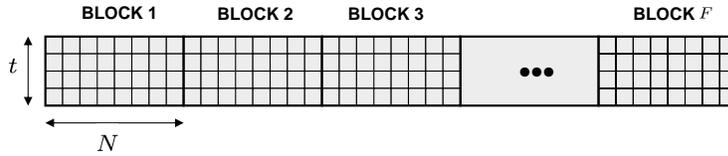


Fig. 4.15 One code word in an F -block fading channel.

channel is described by the F matrices \mathbf{H}_k , $k = 1, \dots, F$, each describing the fading gains in a block. The channel input–output equation is

$$\mathbf{y}_k[n] = \mathbf{H}_k \mathbf{x}_k[n] + \mathbf{z}_k[n] \quad (4.58)$$

for $k = 1, \dots, F$ (block index) and $n = 1, \dots, N$ (symbol index along a block), $\mathbf{y}_k, \mathbf{z}_k \in \mathbb{C}^r$, and $\mathbf{x}_k \in \mathbb{C}^t$. Moreover, the additive noise $\mathbf{z}_k[n]$ is a vector of circularly-symmetric complex Gaussian RVs with zero mean

and variance N_0 : hence,

$$\mathbb{E}[\mathbf{z}_k[n]\mathbf{z}_k[n]^\dagger] = N_0\mathbf{I}_r.$$

It is convenient to use the SVD

$$\mathbf{H}_k = \mathbf{U}_k\mathbf{D}_k\mathbf{V}_k^\dagger, \quad (4.59)$$

where \mathbf{D}_k is an $r \times t$ real matrix whose main-diagonal entries are the ordered singular values $\sqrt{\lambda_{k,1}} \geq \dots \geq \sqrt{\lambda_{k,m}}$, with $\lambda_{k,i}$ the i th largest eigenvalue of the Hermitian matrix $\mathbf{H}_k\mathbf{H}_k^\dagger$, and $m \triangleq \min\{r, t\}$. Since \mathbf{U}_k and \mathbf{V}_k are unitary, by premultiplying $\mathbf{y}_k[n]$ by \mathbf{U}_k^\dagger the input-output relation (4.58) can be rewritten in the form

$$\tilde{\mathbf{y}}_k[n] = \mathbf{D}_k\tilde{\mathbf{x}}_k[n] + \tilde{\mathbf{z}}_k[n], \quad (4.60)$$

where $\tilde{\mathbf{y}}_k[n] \triangleq \mathbf{U}_k^\dagger\mathbf{y}_k[n]$, $\tilde{\mathbf{x}}_k[n] \triangleq \mathbf{V}_k^\dagger\mathbf{x}_k[n]$, $\tilde{\mathbf{z}}_k[n] \triangleq \mathbf{U}_k^\dagger\mathbf{z}_k[n]$, and $\tilde{\mathbf{z}}_k[n] \sim \mathcal{N}_c(\mathbf{0}, N_0\mathbf{I}_r)$ since

$$\mathbb{E}[\tilde{\mathbf{z}}_k[n]\tilde{\mathbf{z}}_k[n]^\dagger] = \mathbf{U}^\dagger\mathbb{E}[\mathbf{z}_k[n]\mathbf{z}_k[n]^\dagger]\mathbf{U} = N_0\mathbf{I}_r.$$

No delay constraints. Since the random matrix process $\{\mathbf{H}_k\}_{k=1}^F$ is iid, as $F \rightarrow \infty$ the channel is ergodic, and the average capacity is the relevant quantity. When perfect CSI is available to the receiver only, this is given by

$$C = \mathbb{E} \left[\sum_{i=1}^m \log \left(1 + \frac{\rho}{t} \lambda_i \right) \right]. \quad (4.61)$$

If perfect CSI is available to transmitter and receiver,

$$C = \mathbb{E} \left[\sum_{i=1}^m (\log(\mu\lambda_i))_+ \right], \quad (4.62)$$

where μ is the solution of the “water-filling” equation

$$\mathbb{E} \left[\sum_{i=1}^m (\mu - 1/\lambda_i)_+ \right] = \rho. \quad (4.63)$$

For all block lengths $N = 1, 2, \dots$, the capacities (4.61) and (4.62) are achieved by code sequences with length FNt with $F \rightarrow \infty$. Capacity (4.61) is achieved by random codes whose symbols are iid complex

$\sim \mathcal{N}_c(0, \rho/t)$. Thus, all antennas transmit the same average energy per symbol. Capacity (4.62) can be achieved by generating a random code with iid components $\sim \mathcal{N}_c(0, 1)$ and having each code word split into F blocks of N vectors $\tilde{\mathbf{x}}_k[n]$ with t components each. For block k , the optimal linear transformation

$$\mathbf{W}_k = \mathbf{V}_k \text{diag}(\sqrt{\rho_{k,1}}, \dots, \sqrt{\rho_{k,m}}, \underbrace{0, \dots, 0}_{t-m}) \quad (4.64)$$

is computed, where $\rho_{k,i} \triangleq (\mu - 1/\lambda_{k,i})_+$. The vectors $\mathbf{x}_k[n] = \mathbf{W}_k \tilde{\mathbf{x}}_k[n]$ are transmitted from the t antennas. This optimal scheme can be viewed as the concatenation of an optimal encoder for the unfaded AWGN channel, followed by a beamformer with weighting matrix \mathbf{W}_k varying from block to block [6].

Delay constraints. Consider now a delay constraint that forces F to take on a finite value. Define $\mathbf{\Lambda} \triangleq \{\lambda_{k,i}\}_{k=1,i=1}^{F,m}$, $\mathbf{\Gamma} \triangleq \{\rho_{k,i}\}_{k=1,i=1}^{F,m}$, the *instantaneous* mutual information

$$I(\mathbf{\Lambda}, \mathbf{\Gamma}) \triangleq \frac{1}{F} \sum_{k=1}^F \sum_{i=1}^m \log(1 + \lambda_{k,i} \rho_{k,i}) \quad (4.65)$$

and the *instantaneous* SNR per block

$$\rho_F \triangleq \frac{1}{F} \sum_{k=1}^F \sum_{i=1}^m \rho_{k,i}. \quad (4.66)$$

Assuming that the receiver has perfect knowledge of the CSI (and hence of $\mathbf{\Lambda}$) we can define a power allocation rule depending on $\mathbf{\Lambda}$ so that $\rho_{k,i}$ and ρ_F are functions of $\mathbf{\Lambda}$. Then, we consider two power constraints:

$$\rho_F(\mathbf{\Lambda}) \leq \rho \quad (\text{short-term}) \quad (4.67)$$

$$\mathbb{E}[\rho_F(\mathbf{\Lambda})] \leq \rho \quad (\text{long-term}). \quad (4.68)$$

The optimum power allocation rules minimizing the outage probability

$$P_{\text{out}}(R) \triangleq \mathbb{P}(I(\mathbf{\Lambda}, \mathbf{\Gamma}) < R) \quad (4.69)$$

under constraints (4.67) and (4.68) are derived in [6] and reported as follows.

(1) With the short-term power constraint,

$$\mathbf{\Gamma}(\mathbf{\Lambda}) = \begin{cases} \mathbf{\Gamma}^{\text{st}}(\mathbf{\Gamma}, \rho) & \text{if } \mathbf{\Lambda} \in \mathcal{R}_{\text{on}}(R, \rho) \\ \mathbf{G}(\mathbf{\Gamma}) & \text{if } \mathbf{\Lambda} \in \mathcal{R}_{\text{off}}(R, \rho) \end{cases}, \quad (4.70)$$

where

(a) The (k, i) -th SNR is given by

$$\rho_{k,i}^{\text{st}} = (\mu^{\text{st}}(\mathbf{\Lambda}, \rho) - 1/\lambda_{k,i})_+, \quad (4.71)$$

where

$$\mu^{\text{st}}(\mathbf{\Lambda}, \rho) = \frac{F}{|\mathcal{F}(\rho)|} \rho + \frac{1}{|\mathcal{F}(\rho)|} \sum_{(k,i) \in \mathcal{F}(\rho)} \frac{1}{\lambda_{k,i}} \quad (4.72)$$

and $\mathcal{F}(\rho)$ is the unique set of indexes (k, i) such that $1/\lambda_{k,i} \leq \mu^{\text{st}}(\mathbf{\Lambda}, \rho)$ for all $(k, i) \in \mathcal{F}(\rho)$ and $1/\lambda_{k,i} > \mu^{\text{st}}(\mathbf{\Lambda}, \rho)$ for all $(k, i) \notin \mathcal{F}(\rho)$.

(b) The set

$$\mathcal{R}_{\text{on}}(R, \rho) \triangleq \{\mathbf{\Lambda} : I(\mathbf{\Lambda}, \mathbf{\Gamma}^{\text{st}}(\mathbf{\Lambda}, \rho)) \geq R\} \quad (4.73)$$

is called *power-on region*.

(c) The set

$$\mathcal{R}_{\text{off}}(R, \rho) \triangleq \{\mathbf{\Lambda} : I(\mathbf{\Lambda}, \mathbf{\Gamma}^{\text{st}}(\mathbf{\Lambda}, \rho)) < R\} \quad (4.74)$$

is called *outage* or *power-off region*.

(d) $\mathbf{G}(\mathbf{\Lambda})$ is an arbitrary power allocation function satisfying the short-term constraint, i.e., $\rho_F(\mathbf{G}) \leq \rho$.

(2) With the long-term power constraint,

$$\mathbf{\Gamma}(\mathbf{\Lambda}) = \begin{cases} \mathbf{\Gamma}^{\text{lt}}(\mathbf{\Lambda}, R) & \text{if } \mathbf{\Lambda} \in \mathcal{R}_{\text{on}}^*(R, \rho^*) \\ \mathbf{0} & \text{if } \mathbf{\Lambda} \in \mathcal{R}_{\text{off}}^*(R, \rho^*) \end{cases}, \quad (4.75)$$

where

(a) The (k, i) -th SNR is given by

$$\rho_{k,i}^{\text{lt}} = (\mu^{\text{lt}}(\mathbf{\Lambda}, R) - 1/\lambda_{k,i})_+, \quad (4.76)$$

where

$$\mu^{\text{lt}}(\mathbf{\Lambda}, R) = \left(\frac{2^{FR}}{\prod_{(k,i) \in \mathcal{F}^*(R)} \lambda_{k,i}} \right)^{1/|\mathcal{F}^*(R)|} \quad (4.77)$$

and $\mathcal{F}^*(R)$ is the unique set of indexes (k, i) such that $1/\lambda_{k,i} \leq \mu^{\text{lt}}(\mathbf{\Lambda}, R)$ for all $(k, i) \in \mathcal{F}^*(R)$ and $1/\lambda_{k,i} > \mu^{\text{lt}}(\mathbf{\Lambda}, R)$ for all $(k, i) \notin \mathcal{F}^*(R)$.

(b) The set

$$\mathcal{R}_{\text{on}}^*(R, \rho^*) \triangleq \left\{ \mathbf{\Lambda} : \rho_F(\mathbf{\Gamma}^{\text{lt}}(\mathbf{\Lambda}, R)) \leq \rho^* \right\} \quad (4.78)$$

is called *power-on region*.

(c) The set

$$\mathcal{R}_{\text{off}}^*(R, \rho^*) \triangleq \left\{ \mathbf{\Lambda} : \rho_F(\mathbf{\Gamma}^{\text{lt}}(\mathbf{\Lambda}, R)) > \rho^* \right\} \quad (4.79)$$

is called *outage or power-off region*.

(d) The *threshold* $\rho^* > 0$ is set in order to satisfy the long-term constraint (4.68) with equality, i.e., it is the solution of

$$\mathbb{E}[\rho_F(\mathbf{\Gamma}^{\text{lt}}(\mathbf{\Lambda}, R)) 1\{\mathbf{\Lambda} \in \mathcal{R}_{\text{on}}(R, \rho^*)\}] = \rho,$$

where $1\{\mathcal{A}\} \triangleq 1$ if \mathcal{A} is true and 0 otherwise.

In other words, the outage probability is minimized under a long-term power constraint by setting a threshold ρ^* . If the instantaneous SNR per block necessary to avoid an outage exceeds ρ^* , then transmission is turned off and an outage is declared. If it is below ρ^* , transmission is turned on, and power is allocated to the blocks according to a rule that depends on the fading statistics only through the threshold value ρ^* (see [6]).

Fig. 4.16 illustrates the concept of outage region for a single transmit and receive antenna system ($t = r = 1$) with $F = 2$, $R = 1$ bit/dimension pair, and $\rho = 1$ dB. The outage region is the inner region corresponding to smaller values of the channel matrix eigenvalues ($\sqrt{\lambda_{k,1}}$, $k = 1, 2$) reflecting the occurrence of a deep fade.

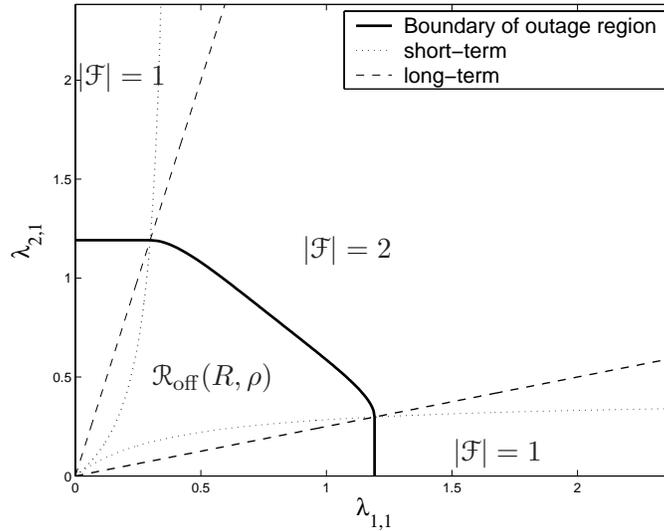


Fig. 4.16 Outage region $\mathcal{R}_{\text{off}}(R, \rho)$ of a single transmit and receive antenna system ($t = r = 1$) with $F = 2$, $R = 1$ bit/dimension pair, and $\rho = 1$ dB. The boundaries of constant- $|\mathcal{F}|$ regions are also indicated for the short-term and long-term constraints as dotted and dashed lines, respectively.

It is interesting to note that the outage regions $\mathcal{R}_{\text{off}}(R, \rho)$ (short-term) and $\mathcal{R}_{\text{off}}^*(R, \rho)$ (long-term) exhibit the same functional dependence on R and ρ in spite of their very different definitions of eqs. (4.74) and (4.79) [6]. This is again illustrated by Fig. 4.16. The figure also shows that though the outage regions $\mathcal{R}_{\text{off}}(1, 10^{0.1})$ and $\mathcal{R}_{\text{off}}^*(1, 10^{0.1})$ coincide, the boundaries of constant- $|\mathcal{F}|$ regions differ in the two cases (short-term and long-term) [6].

Another important concept related to outage probability is given in the following.

Definition 4.14. The *zero-outage capacity*, sometimes also referred to as *delay-limited capacity*, is the maximum rate for which the minimum outage probability is zero under a given power constraint [6, 24, 25, 107].

It can be shown [6] that, under a positive long-term power constraint, the zero-outage capacity of a block-fading channel is positive if the channel is *regular*. A regular channel is defined as follows.

Definition 4.15. A block fading channel is said to be *regular* if the fading distribution is continuous, and

$$\mathbb{E}[1/\bar{\lambda}_F] < \infty, \quad (4.80)$$

where $\bar{\lambda}_F$ is the geometric mean of the $\lambda_{i,k}$:

$$\bar{\lambda}_F \triangleq \prod_{k,i} \lambda_{k,i}^{1/mF}, \quad (4.81)$$

where $m \triangleq \min\{t, r\}$.

Example 4.16. The Rayleigh fading channel with $F = m = 1$ is not regular, and its zero-outage capacity is null. The Rayleigh block fading channel is regular if $mF > 1$ (see [6] for a proof). For example, if $F > 1$ and $m = 1$ we have

$$\mathbb{E}[1/\bar{\lambda}_F] = (\mathbb{E}[\lambda_1^{-1/F}])^F = [\Gamma(1 - 1/F)]^F < \infty,$$

where $\Gamma(x) \triangleq \int_0^\infty u^{x-1} e^{-u} du$ is the standard Gamma function.

4.5.2 Asymptotics

Under a long-term power constraint and with optimal transmit power allocation the zero-outage capacity of a regular block-fading channel as $\rho \rightarrow \infty$ is given by [6]

$$C_{\text{zero-outage}} \sim m \log \left(\frac{\rho}{m \mathbb{E}[1/\bar{\lambda}_F]} \right). \quad (4.82)$$

As $m \rightarrow \infty$ and $\max\{t, r\}/m \rightarrow \alpha > 0$, the limiting value of the normalized zero-outage capacity per degree of freedom C/m coincides with the limiting normalized ergodic capacity [6].

Full text available at: <http://dx.doi.org/10.1561/0100000002>

5

Influence of channel-state information

A crucial factor in determining the performance of a multi-antenna system is the availability of the channel-state information (CSI), that is, the knowledge of the values of the fading gains in each one of the transmission paths. As we have seen, in a system with t transmit and r receive antennas and an ergodic Rayleigh fading channel modeled by a $t \times r$ matrix with random iid complex Gaussian entries, the average channel capacity with perfect CSI at the receiver is about $m \triangleq \min\{t, r\}$ times larger than that of a single-antenna system for the same transmitted power. The capacity increases by about m bits per dimension pair for every 3-dB increase in SNR. Due to the assumption of perfect CSI available at the receiver, this result can be viewed as a fundamental limit for coherent multiple-antenna systems [120].

5.1 Perfect CSI at the receiver

The most commonly studied situation is that of perfect CSI available at the receiver, which is the assumption under which we developed our study of multiple-antenna systems above.

5.2 Imperfect CSI

Fundamental limits of noncoherent communication, i.e., one taking place in an environment where estimates of the fading coefficients are not available, were derived by Marzetta and Hochwald in [73, 56]. They use a block-fading channel model, whereby the fading gains are Rayleigh and remain constant for N time instants before changing to an independent realization. To compute the capacity of this channel, we assume that coding is performed using blocks, each of them consisting of tN elementary symbols being transmitted by t antennas in N time instants. Each block is represented by the $t \times N$ matrix \mathbf{X} . We further assume that the $r \times N$ noise matrix \mathbf{Z} has iid $\mathcal{N}_c(0, N_0)$ entries. With the assumptions of Section 2 on the statistics of \mathbf{H} , the received signal is the $r \times N$ matrix

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}. \quad (5.1)$$

The entries of \mathbf{Y} have the explicit expression

$$y_{in} = \sum_{j=1}^t h_{ij}x_{jn} + z_{in}, \quad i = 1, \dots, r, \quad n = 1, \dots, N. \quad (5.2)$$

Given \mathbf{X} , these are random variables whose mean value is zero and whose covariance is

$$\mathbb{E}[y_{in}y_{i'n'}^* | \mathbf{X}] = \sum_{j=1}^t \sum_{j'=1}^t \mathbb{E}[h_{ij}h_{i'j'}^*]x_{jn}x_{j'n'}^* + \mathbb{E}[z_{in}z_{i'n'}^*]. \quad (5.3)$$

Now, under our assumptions \mathbf{H} and \mathbf{Z} are temporally and spatially white, that is,

$$\mathbb{E}[h_{ij}h_{i'j'}^*] = \delta_{ii'} \quad \mathbb{E}[z_{in}z_{i'n'}^*] = \delta_{ii'}\delta_{nn'}, \quad (5.4)$$

so we have

$$\mathbb{E}[y_{in}y_{i'n'}^* | \mathbf{X}] = \delta_{ii'} \left[\sum_{j=1}^t x_{jn}x_{j'n'}^* + \delta_{nn'} \right]. \quad (5.5)$$

The previous equality expresses the fact that the rows of \mathbf{Y} are independent, while the columns have a nonzero correlation. This observation

allows us to write down the relation connecting the rows $(\mathbf{Y})_i$ of \mathbf{Y} with those of \mathbf{H} , denoted $(\mathbf{H})_i$, and those of \mathbf{Z} , denoted $(\mathbf{Z})_i$, so that

$$(\mathbf{Y})_i = (\mathbf{H})_i \mathbf{X} + (\mathbf{Z})_i \quad i = 1, \dots, r. \quad (5.6)$$

Each row of \mathbf{Y} is a zero-mean Gaussian vector with covariance matrix

$$\mathbb{E}[(\mathbf{Y})_i^\dagger (\mathbf{Y})_i | \mathbf{X}] = \mathbf{X}^\dagger \mathbf{X} + \mathbf{I}_N, \quad (5.7)$$

and writing the pdf of matrix \mathbf{Y} as the product of the pdfs of its rows, we obtain

$$\begin{aligned} p(\mathbf{Y} | \mathbf{X}) &= \prod_{i=1}^r p(\mathbf{Y}_i | \mathbf{X}) \\ &= \frac{1}{\pi^r \det^r[\mathbf{X}^\dagger \mathbf{X} + \mathbf{I}_N]} \prod_{i=1}^r \exp\{-(\mathbf{Y})_i (\mathbf{X}^\dagger \mathbf{X} + \mathbf{I}_N)^{-1} (\mathbf{Y})_i^\dagger\} \\ &= \frac{1}{\pi^r \det^r[\mathbf{X}^\dagger \mathbf{X} + \mathbf{I}_N]} \exp\{-\text{Tr}((\mathbf{X}^\dagger \mathbf{X} + \mathbf{I}_N)^{-1} \mathbf{Y}^\dagger \mathbf{Y})\}. \end{aligned} \quad (5.8)$$

Observe the following:

- (a) The pdf of \mathbf{Y} depends on its argument only through the product $\mathbf{Y}^\dagger \mathbf{Y}$, which consequently plays the role of a sufficient statistic. If $N < r$, the $N \times N$ matrix $\mathbf{Y}^\dagger \mathbf{Y}$ provides a representation of the received signals which is more economical than the $r \times N$ matrix \mathbf{Y} .
- (b) The pdf (5.8) depends on the transmitted signal \mathbf{X} only through the $N \times N$ matrix $\mathbf{X}^\dagger \mathbf{X}$.

Observation (b) above is the basis of the following theorem, which, in its essence, says that there is no increase in capacity if we have $t > N$, and hence there is no point in making the number of transmit antennas greater than N if there is no CSI. In particular, if $N = 1$ (an independent fade occurs at each symbol period) only one transmit antenna is useful. Note how this result contrasts sharply with its counterpart of CSI known at the receiver, where the capacity grows linearly with $\min\{t, r\}$.

Theorem 5.1. The channel capacity for $t > N$ equals the capacity for $t = N$.

Proof. Suppose that the capacity is achieved for a particular pdf of matrix \mathbf{X} with $t > N$. Recalling observation (b) above, the capacity is determined by the matrix $\mathbf{X}^\dagger \mathbf{X}$: if we prove that a transmitted signal can be found that generates the same matrix with only N transmit antennas, then the theorem is proved. Now, perform the Cholesky factorization [62, p. 407] $\mathbf{X}^\dagger \mathbf{X} = \mathbf{L} \mathbf{L}^\dagger$, with \mathbf{L} an $N \times N$ lower-triangular matrix. Using N transmit antennas with a signal matrix that has the same pdf as \mathbf{L}^\dagger , we obtain the same pdf that achieves capacity. In fact, if \mathbf{X} satisfies that average-power constraint

$$\frac{1}{N} \mathbb{E}[\text{Tr } \mathbf{X}^\dagger \mathbf{X}] = \rho N_0, \quad (5.9)$$

so does \mathbf{L}^\dagger . □

From [73], the signal matrix that achieves capacity can be written in the form:

$$\mathbf{X} = \mathbf{D} \Phi \quad (5.10)$$

where Φ is a $t \times N$ matrix such that $\Phi \Phi^\dagger = \mathbf{I}_t$. Moreover, Φ has a pdf which is unchanged when the matrix is multiplied by a deterministic unitary matrix (this is the matrix counterpart of a complex scalar having unit magnitude and uniformly-distributed phase). \mathbf{D} is a $t \times t$ real nonnegative diagonal matrix independent of Φ , whose role is to scale \mathbf{X} to meet the power constraint. In general, the optimizing \mathbf{D} is unknown, as is the exact expression of capacity. However, for the high-SNR regime ($\rho \gg 1$), the following results [73, 120] are available:

- (a) If $N \gg t$ and $t \leq \min\{N/2, r\}$, then capacity is attained when $\mathbf{D} = \sqrt{\rho N N_0 / t} \mathbf{I}_t$, so that $\mathbf{X} = \sqrt{\rho N N_0 / t} \Phi$.
- (b) For every 3-dB increase of ρ , the capacity increase is $\check{t}(1 - \check{t}/N)$, where $\check{t} \triangleq \min\{t, r, \lfloor N/2 \rfloor\}$.
- (c) If $N \geq 2r$, there is no capacity increase by using $r > t$.

An obvious upper bound to capacity can be obtained if we assume that the receiver is provided with perfect knowledge of the realization of \mathbf{H} . Hence, the bound to capacity per block of N symbols is

$$C \leq N \log \det \left[\mathbf{I}_t + \frac{\rho}{t} \mathbf{H}^\dagger \mathbf{H} \right]. \quad (5.11)$$

We can reasonably expect that the actual capacity tends to the RHS of previous inequality, because a certain (small) fraction of the coherence time can be reserved for sending training data to be used by the receiver for its estimate of \mathbf{H} .

5.3 Imperfect CSI at the receiver: General guidelines

In the real world the receiver has an imperfect knowledge of the CSI.

Here we assume that CSI is obtained by transmission of a preamble in the form of a known $t \times N_p$ code matrix \mathbf{X}_p with total energy $\text{Tr}(\mathbf{X}_p \mathbf{X}_p^\dagger) = t N_p E_p$, with E_p the average symbol energy. Since to estimate the $r \times t$ matrix \mathbf{H} we need at least rt measurements, and each symbol time yields r measurements at the receiver, we need $N_p \geq t$. Moreover, the matrix \mathbf{X}_p must have full rank t , since otherwise t linearly independent columns would not be available to yield rt independent measurements. As a consequence, $\mathbf{X}_p \mathbf{X}_p^\dagger$ must be nonsingular. The corresponding received signal is denoted by

$$\mathbf{Y}_p = \mathbf{H} \mathbf{X}_p + \mathbf{Z}_p. \quad (5.12)$$

Among the several receiver structures that can be envisaged we focus on the following:

- (a) The simplest receiver inserts directly the maximum-likelihood (ML) estimate of the channel into the ML metric conditioned on \mathbf{H} . The detection problem consists of computing first

$$\hat{\mathbf{H}} \triangleq \arg \max_{\mathbf{H}} p(\mathbf{Y}_p | \mathbf{X}_p, \mathbf{H}) \quad (5.13)$$

and then

$$\hat{\mathbf{X}} \triangleq \arg \max_{\mathbf{X}} \tilde{\mu}(\mathbf{X}), \quad (5.14)$$

where

$$\tilde{\mu}(\mathbf{X}) \triangleq \|\mathbf{Y} - \hat{\mathbf{H}} \mathbf{X}\|^2. \quad (5.15)$$

Since (5.15) is commonly referred to as *mismatched* metric, we call this a *mismatched receiver*.

- (b) The receiver estimates the channel matrix $\hat{\mathbf{H}}$ from \mathbf{Y}_p and \mathbf{X}_p by an ML criterion, and uses this result to detect the transmitted signal \mathbf{X} . The detection problem consists of computing

$$\hat{\mathbf{H}} \triangleq \arg \max_{\mathbf{H}} p(\mathbf{Y}_p | \mathbf{X}_p, \mathbf{H}) \quad (5.16)$$

and

$$\hat{\mathbf{X}} \triangleq \arg \max_{\mathbf{X}} p(\mathbf{Y} | \mathbf{X}, \mathbf{H} = \hat{\mathbf{H}}), \quad (5.17)$$

where $p(\mathbf{Y} | \mathbf{X}, \mathbf{H} = \hat{\mathbf{H}})$ denotes the probability density function of \mathbf{Y} given \mathbf{X} and \mathbf{H} , with \mathbf{H} equal to $\hat{\mathbf{H}}$.

- (c) The receiver detects the transmitted signal \mathbf{X} by jointly processing \mathbf{Y} , \mathbf{Y}_p , and \mathbf{X}_p (as suggested in [53]) without explicit estimation of \mathbf{H} . In this case, the detection problem can be written as

$$\begin{aligned} \hat{\mathbf{X}} &\triangleq \arg \max_{\mathbf{X}} p(\mathbf{Y}, \mathbf{Y}_p | \mathbf{X}, \mathbf{X}_p) \\ &= \mathbb{E}_{\mathbf{H}} \left[p(\mathbf{Y} | \mathbf{X}, \mathbf{H}) p(\mathbf{Y}_p | \mathbf{X}_p, \mathbf{H}) \right], \end{aligned} \quad (5.18)$$

since, conditionally on \mathbf{H} , \mathbf{X} , and \mathbf{X}_p , the received signals \mathbf{Y} and \mathbf{Y}_p are independent.

Approach (a) is plainly the simplest. Approach (b) is more efficient (see [27] for the single-input, single-output case) and allows one to study the impairments caused by imperfect knowledge of \mathbf{H} and by the presence of noise in the received pilot signal \mathbf{Y}_p . Approach (c) is the optimum: disregarding CSI recovery, it focuses on the detection of the transmitted signal \mathbf{X} . We discuss in detail the second and third receiver types.

Approach (b): Receiver metric based on channel matrix estimate. The ML estimate of \mathbf{H} based on the observation of \mathbf{Y}_p is obtained by maximizing $p(\mathbf{Y}_p | \mathbf{H}, \mathbf{X}_p)$ or, equivalently, by minimizing $\|\mathbf{Y}_p - \mathbf{H}\mathbf{X}_p\|$ with respect to \mathbf{H} , yielding:

$$\hat{\mathbf{H}} = \mathbf{Y}_p \mathbf{X}_p^\dagger (\mathbf{X}_p \mathbf{X}_p^\dagger)^{-1} = \mathbf{H} + \mathbf{E}, \quad (5.19)$$

where

$$\mathbf{E} \triangleq \mathbf{Z}_p \mathbf{X}_p^\dagger (\mathbf{X}_p \mathbf{X}_p^\dagger)^{-1} \quad (5.20)$$

is the matrix error on the estimate $\hat{\mathbf{H}}$. Plainly, \mathbf{H} and \mathbf{E} are independent, and denoting by $(\cdot)_i$ the i th row of a matrix (\cdot) , we can write

$$\mathbf{E}_i = (\mathbf{Z}_p)_i \mathbf{X}_p^\dagger (\mathbf{X}_p \mathbf{X}_p^\dagger)^{-1}. \quad (5.21)$$

Thus, the rows of \mathbf{E} are independent vectors of zero-mean circularly-symmetric complex Gaussian random variables with covariance matrix

$$\begin{aligned} \boldsymbol{\Sigma}_e &\triangleq \mathbb{E}[\mathbf{E}_i^\dagger \mathbf{E}_i] \\ &= \mathbb{E}[(\mathbf{X}_p \mathbf{X}_p^\dagger)^{-1} \mathbf{X}_p (\mathbf{Z}_p)_i^\dagger (\mathbf{Z}_p)_i \mathbf{X}_p^\dagger (\mathbf{X}_p \mathbf{X}_p^\dagger)^{-1}] \\ &= N_0 (\mathbf{X}_p \mathbf{X}_p^\dagger)^{-1}. \end{aligned} \quad (5.22)$$

The simplest choice for \mathbf{X}_p is a matrix with orthogonal rows and average symbol energy E_p . In this case, the entries of \mathbf{E} are independent circularly-symmetric complex Gaussian random variables with mean zero and variance $N_0/(N_p E_p)$. In [120] a diagonal \mathbf{X}_p is advocated, which corresponds to having only one transmit antenna active at a time, since orthogonality, albeit attractive, may be incompatible with standard signal constellations such as PSK. In fact, it is desirable that the rows of \mathbf{X}_p have good autocorrelation and cross-correlation properties. This is achieved by *perfect root-of-unity sequences* only in some special cases (see [35] and references therein).

Based on these assumptions, we calculate the ML metric from the *a posteriori* probability $p(\mathbf{Y} | \mathbf{X}, \hat{\mathbf{H}})$. First, we note that it can be written as

$$p(\mathbf{Y} | \mathbf{X}, \hat{\mathbf{H}}) = \prod_{i=1}^r p(\mathbf{Y}_i | \mathbf{X}, \hat{\mathbf{H}}_i), \quad (5.23)$$

where $\hat{\mathbf{H}}_i$ and \mathbf{Y}_i denote the i th rows of $\hat{\mathbf{H}}$ and \mathbf{Y} , respectively, since it is plain to see that, conditionally on \mathbf{X} , \mathbf{Y}_i depends only on \mathbf{H}_i and \mathbf{Z}_i . Thus, we can apply the following

Theorem 5.2 ([14]). Let \mathbf{z}_1 and \mathbf{z}_2 be circularly-symmetric complex Gaussian random vectors with zero means and full-rank covariance matrices $\boldsymbol{\Sigma}_{ij} \triangleq \mathbb{E}[\mathbf{z}_i \mathbf{z}_j^\dagger]$. Then, conditionally on \mathbf{z}_2 , the random vector \mathbf{z}_1

is circularly-symmetric complex Gaussian with mean $\Sigma_{12}\Sigma_{22}^{-1}\mathbf{z}_2$ and covariance matrix $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Letting

$$\mathbf{z}_1 = \mathbf{Y}_i^\dagger = \mathbf{X}^\dagger \mathbf{H}_i^\dagger + \mathbf{Z}_i^\dagger \quad \text{and} \quad \mathbf{z}_2 = \widehat{\mathbf{H}}_i^\dagger = \mathbf{H}_i^\dagger + \mathbf{E}_i^\dagger \quad (5.24)$$

in Theorem 5.2, we have

$$\begin{aligned} \Sigma_{11} &= N_0 \mathbf{I}_N + \mathbf{X}^\dagger \mathbf{X} \\ \Sigma_{12} &= \mathbf{X}^\dagger \\ \Sigma_{22} &= \mathbf{I}_t + N_0 (\mathbf{X}_p \mathbf{X}_p^\dagger)^{-1}. \end{aligned}$$

Then, the conditional probability density function of \mathbf{Y}_i^\dagger , given \mathbf{X} and $\widehat{\mathbf{H}}_i$, is a circularly-symmetric complex Gaussian distribution, with

$$\begin{aligned} \text{mean} &= \Sigma_{12}\Sigma_{22}^{-1}\mathbf{z}_2 \\ &= \mathbf{X}^\dagger (\mathbf{I}_t + \mathbf{X}_p \mathbf{X}_p^\dagger / N_0)^{-1} \mathbf{X}_p \mathbf{X}_p^\dagger / N_0 \widehat{\mathbf{H}}_i^\dagger \\ &= \mathbf{X}^\dagger (\mathbf{I}_t + \mathbf{X}_p \mathbf{X}_p^\dagger / N_0)^{-1} \mathbf{X}_p \mathbf{X}_p^\dagger / N_0 \widehat{\mathbf{H}}_i^\dagger \quad (5.25) \end{aligned}$$

$$\begin{aligned} \text{covariance matrix} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= N_0 \mathbf{I}_N + \mathbf{X}^\dagger \mathbf{X} - \mathbf{X}^\dagger (\mathbf{I}_t + \mathbf{X}_p \mathbf{X}_p^\dagger / N_0)^{-1} \mathbf{X} \\ &= N_0 [\mathbf{I}_N + \mathbf{X}^\dagger (\mathbf{I}_t + \mathbf{X}_p \mathbf{X}_p^\dagger / N_0)^{-1} \mathbf{X} / N_0]. \quad (5.26) \end{aligned}$$

In the special case of $\mathbf{X}_p \mathbf{X}_p^\dagger = N_p E_p \mathbf{I}_t$, the previous expressions simplify to

$$\text{mean} = \mu \mathbf{X}^\dagger \widehat{\mathbf{H}}_i^\dagger \quad (5.27)$$

$$\text{covariance matrix} = N_0 \mathbf{I}_N + (1 - \mu) \mathbf{X}^\dagger \mathbf{X}, \quad (5.28)$$

where we define

$$\mu \triangleq \frac{1}{1 + N_0 / (N_p E_p)}. \quad (5.29)$$

As a result, we have:

$$p(\mathbf{Y} | \mathbf{X}, \widehat{\mathbf{H}}) = \frac{\text{etr}(-(\mathbf{Y} - \mu \widehat{\mathbf{H}} \mathbf{X})(N_0 \mathbf{I}_N + (1 - \mu) \mathbf{X}^\dagger \mathbf{X})^{-1} (\mathbf{Y} - \mu \widehat{\mathbf{H}} \mathbf{X})^\dagger)}{\det(\pi(N_0 \mathbf{I}_N + (1 - \mu) \mathbf{X}^\dagger \mathbf{X}))^r} \quad (5.30)$$

corresponding to the metric

$$\begin{aligned} \mu(\mathbf{X}) &= \text{Tr}((\mathbf{Y} - \mu\hat{\mathbf{H}}\mathbf{X})(\mathbf{I}_N + (1 - \mu)\mathbf{X}^\dagger\mathbf{X}/N_0)^{-1}(\mathbf{Y} - \mu\hat{\mathbf{H}}\mathbf{X})^\dagger) \\ &\quad + rN_0 \ln \det(\mathbf{I}_N + (1 - \mu)\mathbf{X}^\dagger\mathbf{X}/N_0). \end{aligned} \quad (5.31)$$

For sequential implementation (required, e.g., by Viterbi decoding), setting $\mathbf{X} = (\mathbf{X}^-, \mathbf{x})$ and $\mathbf{Y} = (\mathbf{Y}^-, \mathbf{y})$, we can write the *branch* metric as

$$\begin{aligned} \Delta\mu(\mathbf{x}; \mathbf{X}^-, \mathbf{Y}) &\triangleq \mu(\mathbf{X}) - \mu(\mathbf{X}^-) \\ &= \|\mathbf{y} - \mu\hat{\mathbf{H}}\mathbf{x}\|^2 - rN_0 \ln(1 + (1 - \mu)\mathbf{x}^\dagger\boldsymbol{\Lambda}(\mathbf{X}^-)\mathbf{x}/N_0) \\ &\quad - \frac{1 - \mu}{N_0} \text{Tr}(\boldsymbol{\Xi}(\mathbf{X})^\dagger\boldsymbol{\Lambda}(\mathbf{X})\boldsymbol{\Xi}(\mathbf{X}) - \boldsymbol{\Xi}(\mathbf{X}^-)^\dagger\boldsymbol{\Lambda}(\mathbf{X}^-)\boldsymbol{\Xi}(\mathbf{X}^-)), \end{aligned} \quad (5.32)$$

where

$$\boldsymbol{\Xi}(\mathbf{X}) \triangleq \mathbf{X}(\mathbf{Y} - \rho\hat{\mathbf{H}}\mathbf{X})^\dagger = \boldsymbol{\Xi}(\mathbf{X}^-) + \mathbf{x}(\mathbf{y} - \rho\hat{\mathbf{H}}\mathbf{x})^\dagger \quad (5.33)$$

$$\begin{aligned} \boldsymbol{\Lambda}(\mathbf{X}) &\triangleq (\mathbf{I}_t + (1 - \rho)\mathbf{X}\mathbf{X}^\dagger/N_0)^{-1} \\ &= \boldsymbol{\Lambda}(\mathbf{X}^-) - \frac{1 - \mu}{N_0 + (1 - \mu)\mathbf{x}^\dagger\boldsymbol{\Lambda}(\mathbf{X}^-)\mathbf{x}}\boldsymbol{\Lambda}(\mathbf{X}^-)\mathbf{x}\mathbf{x}^\dagger\boldsymbol{\Lambda}(\mathbf{X}^-) \end{aligned} \quad (5.34)$$

(see [102] for further details).

Approach (c): Optimum receiver metric. In this case the receiver detects the transmitted word \mathbf{X} maximizing the probability density function $p(\mathbf{Y}, \mathbf{Y}_p | \mathbf{X}, \mathbf{X}_p)$ without any prior estimate of the channel matrix \mathbf{H} . We use the following

Theorem 5.3 ([87, App. B]). Given a Hermitian square matrix \mathbf{A} such that $\mathbf{I} + \mathbf{A} > 0$, a size-compatible complex matrix \mathbf{B} , and a matrix \mathbf{Z} of iid zero-mean circularly-symmetric complex Gaussian random variables with unit variance, the following identity holds:

$$\mathbb{E}[\text{etr}(-\mathbf{Z}\mathbf{A}\mathbf{Z}^\dagger - \mathbf{Z}\mathbf{B}^\dagger - \mathbf{B}\mathbf{Z}^\dagger)] = \det(\mathbf{I} + \mathbf{A})^{-r} \text{etr}[\mathbf{B}(\mathbf{I} + \mathbf{A})^{-1}\mathbf{B}^\dagger]. \quad (5.35)$$

Applying Theorem 5.3 we obtain

$$p(\mathbf{Y}, \mathbf{Y}_p | \mathbf{X}, \mathbf{X}_p)$$

$$\begin{aligned}
&= \mathbb{E}_{\mathbf{H}} \left[\frac{\exp(-(\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|^2 + \|\mathbf{Y}_p - \mathbf{H}\mathbf{X}_p\|^2)/N_0)}{(\pi N_0)^{(N_p+N)r}} \right] \\
&= (\pi N_0)^{-(N_p+N)r} \\
&\quad \mathbb{E}_{\mathbf{H}} \left[\text{etr} \left(-(\mathbf{H}^\dagger \mathbf{H}(\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}_p\mathbf{X}_p^\dagger) - \mathbf{H}(\mathbf{X}\mathbf{Y}^\dagger + \mathbf{X}_p\mathbf{Y}_p^\dagger) \right. \right. \\
&\quad \left. \left. - (\mathbf{Y}\mathbf{X}^\dagger + \mathbf{Y}_p\mathbf{X}_p^\dagger)\mathbf{H}^\dagger + (\mathbf{Y}\mathbf{Y}^\dagger + \mathbf{Y}_p\mathbf{Y}_p^\dagger) \right) / N_0 \right] \\
&= (\pi N_0)^{-(N_p+N)r} \det \left[\mathbf{I}_t + (\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}_p\mathbf{X}_p^\dagger) / N_0 \right]^{-r} \\
&\quad \text{etr} \left((\mathbf{Y}\mathbf{X}^\dagger + \mathbf{Y}_p\mathbf{X}_p^\dagger) [\mathbf{I}_t + (\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}_p\mathbf{X}_p^\dagger) / N_0]^{-1} \right. \\
&\quad \left. (\mathbf{X}\mathbf{Y}^\dagger + \mathbf{X}_p\mathbf{Y}_p^\dagger) / N_0^2 - (\mathbf{Y}\mathbf{Y}^\dagger + \mathbf{Y}_p\mathbf{Y}_p^\dagger) / N_0 \right). \quad (5.36)
\end{aligned}$$

The logarithm of (5.36) yields the corresponding metric to be minimized by the optimum receiver:

$$\begin{aligned}
\mu(\mathbf{X}) &= r \ln \det \left[\mathbf{I}_t + (\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}_p\mathbf{X}_p^\dagger) / N_0 \right] \\
&\quad - \text{Tr} \left\{ (\mathbf{Y}\mathbf{X}^\dagger + \mathbf{Y}_p\mathbf{X}_p^\dagger) [\mathbf{I}_t + (\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}_p\mathbf{X}_p^\dagger) / N_0]^{-1} \right. \\
&\quad \left. (\mathbf{X}\mathbf{Y}^\dagger + \mathbf{X}_p\mathbf{Y}_p^\dagger) / N_0^2 \right\}. \quad (5.37)
\end{aligned}$$

If the pilot matrix \mathbf{X}_p has orthogonal rows, i.e., $\mathbf{X}_p\mathbf{X}_p^\dagger = N_p E_p \mathbf{I}_t$, we have the surprising result that the metrics (5.37) and (5.31) are equivalent (see [102] for details).

For sequential implementation, we can write the *branch* metric as

$$\begin{aligned}
\Delta\mu(\mathbf{x}; \mathbf{X}^-) &\triangleq \mu(\mathbf{X}) - \mu(\mathbf{X}^-) \\
&= r \ln(1 + \mathbf{x}^\dagger \mathbf{\Lambda}(\mathbf{X}^-) \mathbf{x} / N_0) \\
&\quad + \text{Tr} [\mathbf{\Xi}(\mathbf{X}^-)^\dagger \mathbf{\Lambda}(\mathbf{X}^-) \mathbf{\Xi}(\mathbf{X}^-) - \mathbf{\Xi}(\mathbf{X})^\dagger \mathbf{\Lambda}(\mathbf{X}) \mathbf{\Xi}(\mathbf{X})], \quad (5.38)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{\Xi}(\mathbf{X}) &\triangleq (\mathbf{X}\mathbf{Y}^\dagger + \mathbf{X}_p\mathbf{Y}_p^\dagger) / N_0 = \mathbf{\Xi}(\mathbf{X}^-) + \mathbf{x}\mathbf{y}^\dagger / N_0 \\
\mathbf{\Lambda}(\mathbf{X}) &\triangleq [\mathbf{I}_t + (\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}_p\mathbf{X}_p^\dagger) / N_0]^{-1} \\
&= \mathbf{\Lambda}(\mathbf{X}^-) - \frac{\mathbf{\Lambda}(\mathbf{X}^-) \mathbf{x} \mathbf{x}^\dagger \mathbf{\Lambda}(\mathbf{X}^-)}{N_0 + \mathbf{x}^\dagger \mathbf{\Lambda}(\mathbf{X}^-) \mathbf{x}}
\end{aligned}$$

$$\ln \det \mathbf{\Lambda}(\mathbf{X}) = \ln \det \mathbf{\Lambda}(\mathbf{X}^-) - \ln(1 + \mathbf{x}^\dagger \mathbf{\Lambda}(\mathbf{X}^-) \mathbf{x} / N_0)$$

(see [102] for details).

Simulation results. Some simulation examples will show the performance of space–time codes with imperfect channel estimation with the suboptimum mismatched metric (5.15) and with the ML metric (5.31). The latter metric is computed by using branch metrics (5.32). Performance results relevant to an ML receiver with perfect CSI (provided by a *genie* without rate loss) are also reported for comparison. Two trellis space–time codes are considered. They are obtained by mapping the rate-2/4 binary convolutional codes with 4 and 16 states, whose generator matrices are [28]

$$\mathbf{G}_1 = \begin{pmatrix} 0 & 3 & 1 & 2 \\ 3 & 1 & 2 & 1 \end{pmatrix}$$

and

$$\mathbf{G}_2 = \begin{pmatrix} 3 & 7 & 1 & 6 \\ 4 & 7 & 6 & 3 \end{pmatrix}$$

onto QPSK, so that the 4 encoded bits are mapped to QPSK symbol pairs sent to the $t = 2$ transmit antennas. Hereafter we refer to these codes by the names STC-1 and STC-2.

Figures 5.1 and 5.2 show the performance of a 2×2 MIMO block Rayleigh fading channel with the trellis space–time code STC-1 with the mismatched and ML metrics (5.15) and (5.31). The frame length considered is $N = 130$ and trellis termination is assumed. The diagrams plot the frame-error rate (FER) versus the signal-to-noise ratio (SNR) $S/N = tE_s/N_0$ at constant values of the pilot-to-noise ratio (PNR) $(S/N)_p = N_p E_p/N_0$. These plots show that when the PNR exceeds 15 dB the ML receiver reaches the *genie* ML receiver performance. The rate loss due to pilot symbol insertion is not accounted for in these diagrams.

Figure 5.3 shows the performance of a 2×2 MIMO block Rayleigh fading channel with the trellis space–time code STC-1 with the mismatched and ML metrics (5.15) and (5.31). The frame length is $N = 130$ including trellis termination. The diagrams plot the FER versus E_b/N_0 . E_b denotes the average received energy per information bit accounting for the rate loss due to pilot symbols, i.e.,

$$E_b = \frac{N + N_p}{N} \frac{tE_s}{\mu_b}.$$

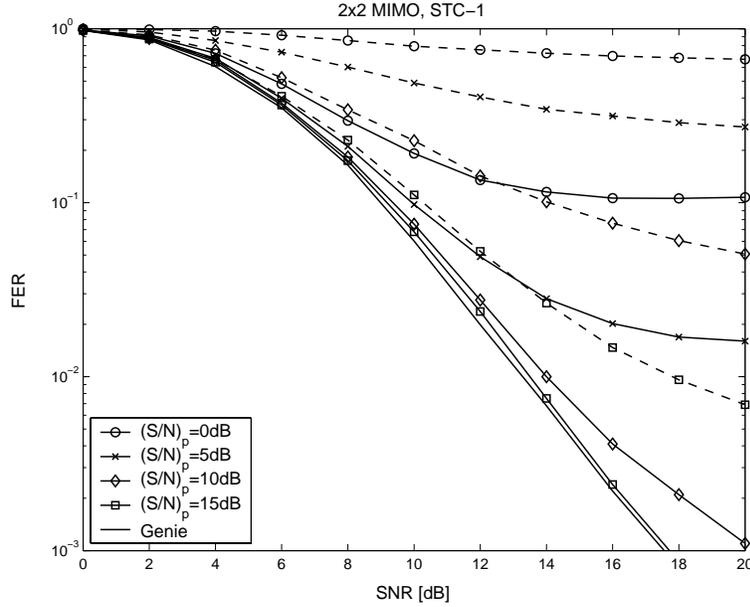


Fig. 5.1 Frame-error rate of a 2×2 ($t = 2, r = 2$) independent Rayleigh fading MIMO channel with the trellis space-time code STC-1. Solid curves show the performance with the ML metric (5.31). Dashed curves show the performance with the mismatched metric (5.15).

Here, μ_b denotes the number of information bits per symbol interval of the trellis space-time code. Moreover, it is assumed that the average pilot symbol energy is equal to the average data symbol energy (i.e., $E_p = E_s$) and N_p (number of pilot intervals per frame) takes on the values 2, 4, 8, 16, and 32.

Similarly, Figure 5.4 shows the performance of a 2×4 MIMO channel with the trellis space-time code STC-1. Figure 5.5 shows the performance of a 2×2 MIMO channel with the trellis space-time code STC-2.

It can be noticed that the improvement achieved by increasing the number of pilot symbols used with the ML metric is always limited and there is a gap to the lower bound performance of the *genie* ML receiver which has perfect CSI available at no expense. The gap depends on t , r , the code and the receiver considered. When $t = r = 2$ it is about

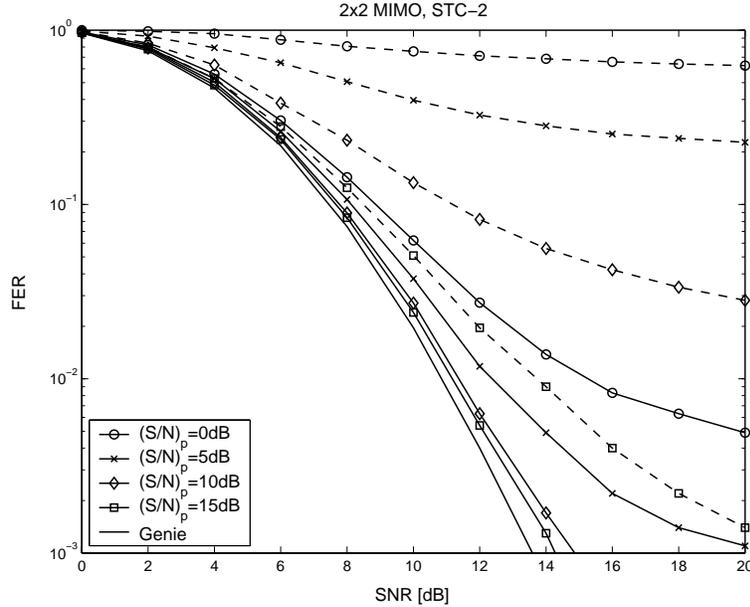


Fig. 5.2 Same as Fig. 5.1 for the code STC-2.

0.3 dB (STC-1 at $\text{FER} = 10^{-2}$ with $N_p = 4$, ML receiver) or 1.0 dB (STC-1 at $\text{FER} = 10^{-2}$ with $N_p = 16$, mismatched receiver). When $t = 2$ and $r = 4$ it is about 0.6 dB (STC-1 at $\text{FER} = 10^{-2}$ with $N_p = 4$, ML receiver) or 1.3 dB (STC-1 at $\text{FER} = 10^{-2}$ with $N_p = 16$, mismatched receiver). Finally, when $t = r = 2$ it is about 0.45 dB (STC-2 at $\text{FER} = 10^{-2}$ with $N_p = 4$, ML receiver) or 1.1 dB (STC-2 at $\text{FER} = 10^{-2}$ with $N_p = 16$, mismatched receiver).

Notice that the mismatched receiver attains the optimum performance when the number of pilot symbols per frame N_p is equal to 16 (about 11% of the overall frame of pilot and data symbols) while the ML receiver performance attains the optimum performance when the number of pilot symbols per frame N_p is equal to 4 (about 3% of the overall frame of pilot and data symbols).

Figures 5.6 to 5.8 show the FER performance versus the fraction of pilot symbols $N_p/(N_p + N)$ at fixed E_b/N_0 . They refer to 2×2 and 2×4 MIMO systems with trellis space-time codes STC-1 and STC-2. The

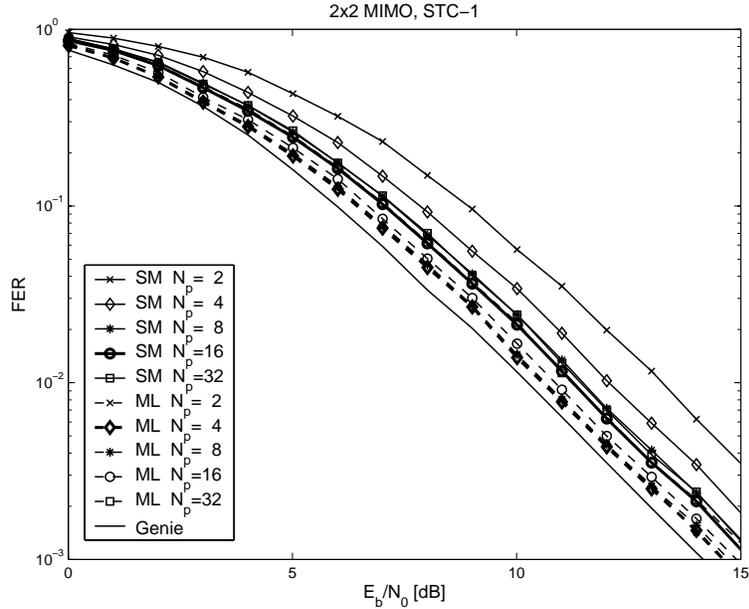


Fig. 5.3 Frame-error rate of a 2×2 ($t = 2, r = 2$) independent Rayleigh fading MIMO channel with the trellis space-time code STC-1 versus E_b/N_0 for several values of pilot intervals $N_p = 2, 4, 8,$ and 16 and frame length $N = 130$. Solid curves with markers show the performance with the suboptimum metric (5.15). Dashed curves with markers show the performance with the ML metric (5.31). The lowest solid curves shows the performance with perfect CSI.

ML receiver performs very close to the *genie* receiver and the optimum number of pilot symbols is about 4 for the ML receiver and 16 for the mismatched receiver.

5.4 CSI at the transmitter and at the receiver

It is also possible to envisage a situation in which channel state information is known to the receiver and to the transmitter: the latter can take the appropriate measures to counteract the effect of channel attenuations by suitably modulating its power. To assure causality, the assumption of CSI available at the transmitter is valid if it is applied to a multicarrier transmission scheme in which the available frequency band (over which the fading is selective) is split into a number of subbands, as

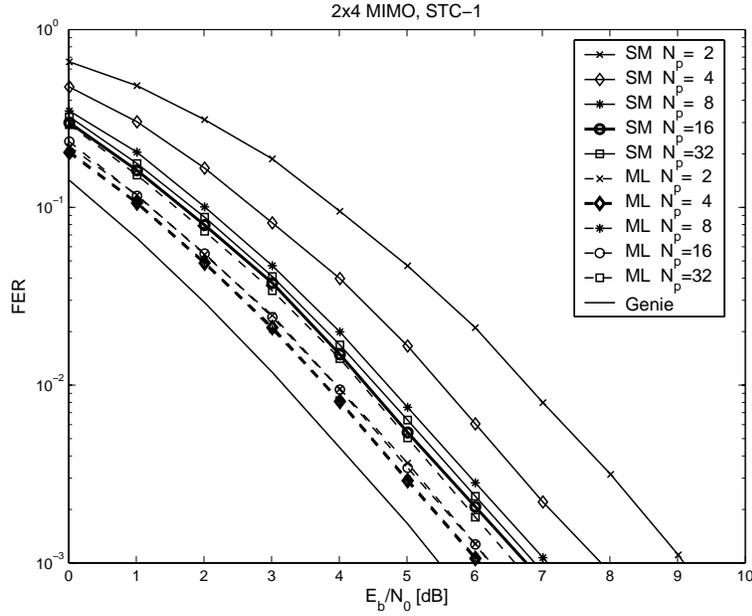


Fig. 5.4 Same as Fig. 5.3 for a 2×4 MIMO channel.

with OFDM. The subbands are so narrow that fading is frequency-flat in each of them, and they are transmitted simultaneously, via orthogonal subcarriers. From a practical point of view, the transmitter can obtain the CSI either from a dedicated feedback channel (some existing systems already implement a fast power-control feedback channel) or by time-division duplex, where the uplink and the downlink time-share the same subchannels and the fading gains can be estimated from the incoming signal.

Situations where a partial (parametric) CSI is available at the transmitter can also be envisaged. For example, for the Rician or Nakagami fading channels (see Section 3.1) the transmitter may be privy to the value of the parameter K or m , respectively. In another situation, only the Dammal condition number of \mathbf{H} may be known [55].

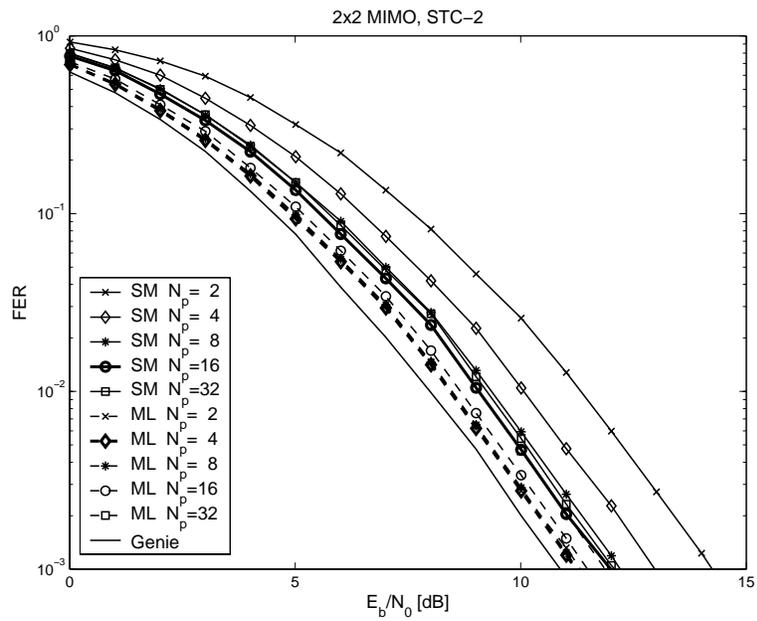


Fig. 5.5 Same as Fig. 5.3 for the code STC-2.

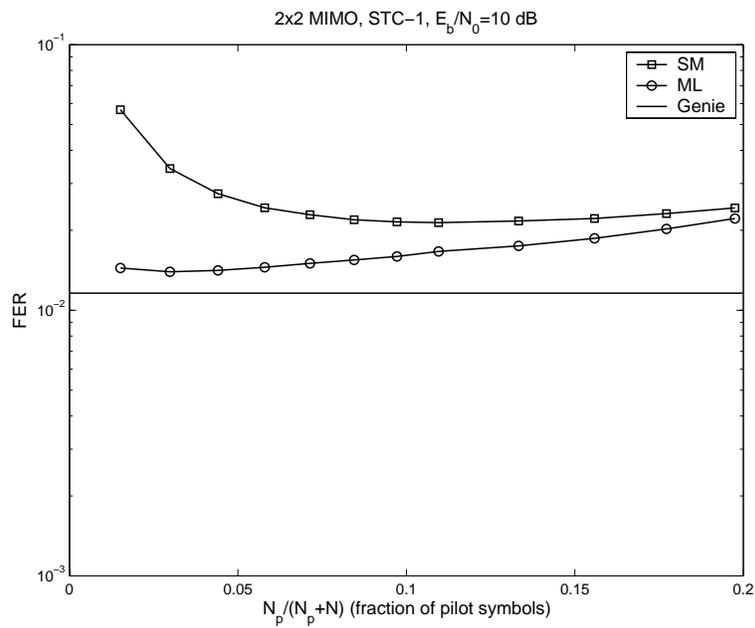


Fig. 5.6 Word error rate of a 2×2 ($t = 2, r = 2$) independent Rayleigh fading MIMO channel with the trellis space-time code STC-1 versus the fraction of pilot symbols $N_p/(N_p + N)$ at $E_b/N_0 = 10$ dB. Solid curves with \square show the performance with the suboptimum metric (5.15). Solid curves with \circ show the performance with the ML metric (5.31). The lowest straight line shows the performance of a genie-aided receiver, which has perfect CSI.

70 Influence of channel-state information

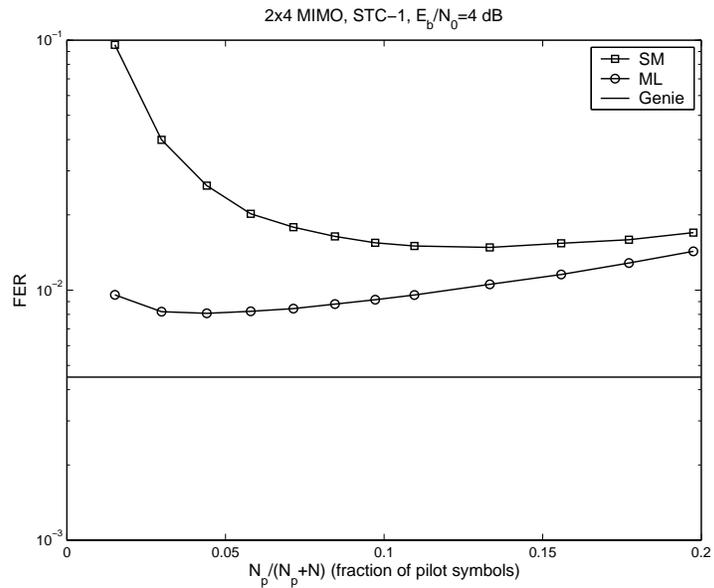


Fig. 5.7 Same as Fig. 5.6 for a 2×4 MIMO channel at $E_b/N_0 = 4$ dB.

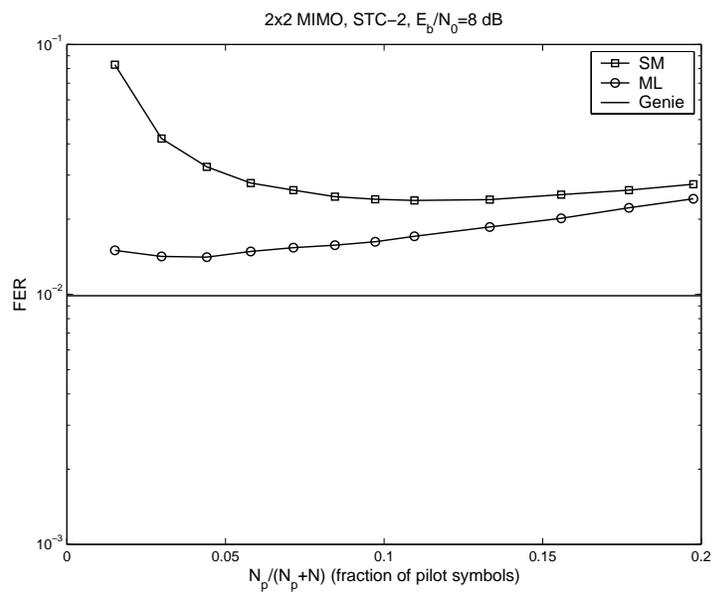


Fig. 5.8 Same as Fig. 5.6 for the code STC-2 at $E_b/N_0 = 8$ dB.

6

Coding for multiple-antenna systems

Given that considerable gains are achievable by a multi-antenna system, the challenge is to design coding schemes that perform close to capacity: space–time trellis codes, space–time block codes, and layered space–time codes have been advocated (see, e.g., [50, 95, 96, 100, 104, 105, 69]).

A space–time code with block length N is described by the $t \times N$ matrix $\mathbf{X} \triangleq (\mathbf{x}[1], \dots, \mathbf{x}[N])$. The code, which we denote \mathcal{X} , has $|\mathcal{X}|$ words. The row index of \mathbf{X} indicates space, while the column index indicates time: to wit, the i th component of the t -vector $\mathbf{x}[n]$, denoted $x_i[n]$, is a complex number representing the two-dimensional signal transmitted by the i th antenna at discrete time n , $n = 1, \dots, N$, $i = 1, \dots, t$. The received signal is the $r \times N$ matrix

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}, \quad (6.1)$$

where \mathbf{Z} is matrix of zero-mean circularly-symmetric complex Gaussian RVs with variance N_0 . Thus, the noise affecting the received signal is spatially and temporally independent, with $\mathbb{E}[\mathbf{Z}\mathbf{Z}^\dagger] = NN_0\mathbf{I}_r$, where \mathbf{I}_r denotes the $r \times r$ identity matrix and $(\cdot)^\dagger$ denotes Hermitian transposition. The channel is described by the $r \times t$ matrix \mathbf{H} . Here we assume that \mathbf{H} is independent of both \mathbf{X} and \mathbf{Z} , it remains constant

during the transmission of an entire code word (the “quasi-static”, or “block-fading”, assumption), and its realization (the CSI) is known at the receiver.

6.1 Maximum likelihood detection

Under the above assumptions, ML decoding corresponds to choosing the code word \mathbf{X} which minimizes the squared Frobenius norm $\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|^2$. Explicitly, ML detection and decoding corresponds to the minimization of the quantity

$$\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|^2 = \sum_{i=1}^r \sum_{n=1}^N \left| y_{in} - \sum_{j=1}^t h_{ij} x_{jn} \right|^2. \quad (6.2)$$

6.1.1 Pairwise error probability

For computations, as calculation of exact error probability is out of the question, we resort to the union bound

$$P(e) \leq \frac{1}{|\mathcal{X}|} \sum_{\mathbf{X} \in \mathcal{X}} \sum_{\widehat{\mathbf{X}} \neq \mathbf{X}} P(\mathbf{X} \rightarrow \widehat{\mathbf{X}}), \quad (6.3)$$

which is known to be asymptotically tight. The “pairwise error probability” (PEP) $\mathbb{P}(\mathbf{X} \rightarrow \widehat{\mathbf{X}})$ [5, p. 190], the basic building block for the evaluation of $P(e)$ in coded systems, admits a closed-form expression:

$$\begin{aligned} P(\mathbf{X} \rightarrow \widehat{\mathbf{X}}) &\triangleq \mathbb{P}(\|\mathbf{Y} - \mathbf{H}\widehat{\mathbf{X}}\|^2 < \|\mathbf{Y} - \mathbf{H}\mathbf{X}\|^2) \\ &= \mathbb{P}(\|\mathbf{H}\Delta + \mathbf{Z}\|^2 < \|\mathbf{Z}\|^2) \\ &= \mathbb{P}((\mathbf{H}\Delta + \mathbf{Z}, \mathbf{H}\Delta + \mathbf{Z}) - (\mathbf{Z}, \mathbf{Z})) < 0) \\ &= \mathbb{P}(\|\mathbf{H}\Delta\|^2 + 2(\mathbf{H}\Delta, \mathbf{Z}) < 0), \end{aligned} \quad (6.4)$$

where $\Delta \triangleq \mathbf{X} - \widehat{\mathbf{X}}$. The variance of the Gaussian random variable $\nu \triangleq (\mathbf{A}, \mathbf{Z})$ can be obtained as follows. Setting $\mathbf{A} = \mathbf{A}_1 + j\mathbf{A}_2$ and $\mathbf{Z} = \mathbf{Z}_1 + j\mathbf{Z}_2$ (where \mathbf{A}_1 , \mathbf{A}_2 , \mathbf{Z}_1 , and \mathbf{Z}_2 are real matrices) we have:

$$\begin{aligned} \mathbb{E}[\nu^2] &= \mathbb{E}[(\text{Tr}(\mathbf{A}_1\mathbf{Z}_1 - \mathbf{A}_2\mathbf{Z}_2))^2] \\ &= \mathbb{E}\left[\left(\sum_i \sum_j (\mathbf{A}_1)_{ij}(\mathbf{Z}_1)_{ji} - (\mathbf{A}_2)_{ij}(\mathbf{Z}_2)_{ji}\right)^2\right] \end{aligned}$$

$$\begin{aligned}
 &= \sum_i \sum_j ((\mathbf{A}_1)_{ij}^2 \mathbb{E}[(\mathbf{Z}_1)_{ji}^2] + (\mathbf{A}_2)_{ij}^2 \mathbb{E}[(\mathbf{Z}_2)_{ji}^2]) \\
 &= \frac{N_0}{2} \|\mathbf{A}\|^2
 \end{aligned} \tag{6.5}$$

since \mathbf{Z}_1 and \mathbf{Z}_2 are independent. Then, the pairwise error probability becomes

$$P(\mathbf{X} \rightarrow \hat{\mathbf{X}}) = \mathbb{E} \left[Q \left(\frac{\|\mathbf{H}\Delta\|}{\sqrt{2N_0}} \right) \right]. \tag{6.6}$$

By writing

$$\|\mathbf{H}\Delta\|^2 = \text{Tr} \left(\mathbf{H}^\dagger \mathbf{H} \Delta \Delta^\dagger \right), \tag{6.7}$$

we see that the exact pairwise error probability, and hence the union bound to $P(e)$, is given by the expected value of a function of the $t \times r$ matrix $\mathbf{H}^\dagger \mathbf{H}$. This matrix can be interpreted as representing the effect of the random spatial interference on error probability: in particular, if $\mathbf{H}^\dagger \mathbf{H} = \mathbf{I}_t$ then (6.6) becomes

$$P(\mathbf{X} \rightarrow \hat{\mathbf{X}}) = Q \left(\frac{\|\Delta\|}{\sqrt{2N_0}} \right), \tag{6.8}$$

which is the PEP we would obtain on a set of t parallel independent AWGN channels, each transmitting a code word consisting of a row of \mathbf{X} .

A useful approximation to the pairwise error probability (6.6) can be computed by substituting exponential functions for Q functions. This is obtained by applying the bound, asymptotically tight for large arguments:

$$Q \left(\frac{\|\mathbf{H}\Delta\|}{\sqrt{2N_0}} \right) \leq \exp \left(-\|\mathbf{H}\Delta\|^2 / 4N_0 \right). \tag{6.9}$$

Under the assumption of Rayleigh fading, that is, when $h_{ij} \sim \mathcal{N}_c(0, 1)$, we can compute the exact expectation of the RHS of (6.9) using Theorem C.1. We obtain

$$P(\mathbf{X} \rightarrow \hat{\mathbf{X}}) \leq \det \left[\mathbf{I}_t + \Delta \Delta^\dagger / 4N_0 \right]^{-r}. \tag{6.10}$$

6.1.2 The rank-and-determinant criterion

Since the determinant of a matrix is equal to the product of its eigenvalues, (6.10) yields

$$P(\mathbf{X} \rightarrow \widehat{\mathbf{X}}) \leq \prod_{j=1}^t (1 + \lambda_j/4N_0)^{-r}, \quad (6.11)$$

where λ_j denotes the j th eigenvalue of $\mathbf{\Delta}\mathbf{\Delta}^\dagger$. We can also write

$$P(\mathbf{X} \rightarrow \widehat{\mathbf{X}}) \leq \prod_{j \in \mathcal{J}} (\lambda_j/4N_0)^{-r}, \quad (6.12)$$

where \mathcal{J} is the index set of the nonzero eigenvalues of $\mathbf{\Delta}\mathbf{\Delta}^\dagger$. Denoting by ν the number of elements in \mathcal{J} , and rearranging the indexes so that $\lambda_1, \dots, \lambda_\nu$ are the nonzero eigenvalues, we have

$$P(\mathbf{X} \rightarrow \widehat{\mathbf{X}}) \leq \left(\prod_{j=1}^{\nu} \lambda_j \right)^{-r} \gamma^{-r\nu}, \quad (6.13)$$

where $\gamma \triangleq 1/4N_0$. From this expression we see that the total diversity order of the coded system is $r\nu_{\min}$, where ν_{\min} is the minimum rank of $\mathbf{\Delta}\mathbf{\Delta}^\dagger$ across all possible pairs $\mathbf{X}, \widehat{\mathbf{X}}$ ($r\nu_{\min}$ is the “diversity gain”). In addition, the pairwise error probability depends on the power r of the product of eigenvalues of $\mathbf{\Delta}\mathbf{\Delta}^\dagger$. This does not depend on the SNR (which is proportional to γ), and displaces the error probability curve instead of changing its slope. We call this the “coding gain.” Thus, for high enough SNR we can design a space–time code choosing as a criterion the maximization of the coding gain as well as of the diversity gain.

Notice that if $\nu_{\min} = t$, i.e., $\mathbf{\Delta}\mathbf{\Delta}^\dagger$ is full-rank for all code word pairs, we have

$$\prod_{j=1}^t \lambda_j = \det[\mathbf{\Delta}\mathbf{\Delta}^\dagger]. \quad (6.14)$$

An obvious necessary condition for $\mathbf{\Delta}\mathbf{\Delta}^\dagger$ to be full-rank is that $N \geq t$ (the code block length must be at least equal to the number of transmit antennas).

Observation 6.1. Note that, based on the above discussion, the maximum achievable diversity gain is rt . In Section 9 we shall discuss how this gain is generally not compatible with the maximum rate gain m .

6.1.3 The Euclidean-distance criterion

Observe that the term in the RHS of (6.10) can be written as

$$\det(\mathbf{I}_t + \gamma \mathbf{\Delta} \mathbf{\Delta}^\dagger) = 1 + \gamma \text{Tr}(\mathbf{\Delta} \mathbf{\Delta}^\dagger) + \dots + \gamma^t \det(\mathbf{\Delta} \mathbf{\Delta}^\dagger). \quad (6.15)$$

We see that if $\gamma \ll 1$ then the LHS of (6.15), and hence the PEP, depends essentially on $\text{Tr}(\mathbf{\Delta} \mathbf{\Delta}^\dagger)$, which is the squared Euclidean distance between \mathbf{X} and $\widehat{\mathbf{X}}$, while if $\gamma \gg 1$ it depends essentially on $\det(\mathbf{\Delta} \mathbf{\Delta}^\dagger)$, that is, on the product of the eigenvalues of $\mathbf{\Delta} \mathbf{\Delta}^\dagger$. This suggests that for low SNR the upper bound (6.15) to error probability depends on the Euclidean distance between code words, as one would expect because the system performance is dictated by additive noise rather than by fading. Conversely, as the SNR increases, the fading effects become more and more relevant, and the rank and determinant of $\mathbf{\Delta} \mathbf{\Delta}^\dagger$ dictate the behavior of the PEP.

A different perspective can be obtained by allowing the number r of receive antennas to grow to infinity. To do this, we first renormalize the entries of \mathbf{H} so that their variance is now $1/r$ rather than 1: this prevents the total receive power from diverging as $r \rightarrow \infty$. We obtain the following new form of (6.10):

$$P(\mathbf{X} \rightarrow \widehat{\mathbf{X}}) \leq \det \left[\mathbf{I}_t + \mathbf{\Delta} \mathbf{\Delta}^\dagger / 4rN_0 \right]^{-r}, \quad (6.16)$$

which yields, in lieu of (6.15):

$$\det(\mathbf{I}_t + (\gamma/r) \mathbf{\Delta} \mathbf{\Delta}^\dagger) = 1 + (\gamma/r) \text{Tr}(\mathbf{\Delta} \mathbf{\Delta}^\dagger) + \dots + (\gamma/r)^t \det(\mathbf{\Delta} \mathbf{\Delta}^\dagger). \quad (6.17)$$

This shows that as $r \rightarrow \infty$ the rank-and-determinant criterion is appropriate for a SNR increasing as fast as r , while the Euclidean-distance criterion is appropriate for finite SNRs.¹ This situation is illustrated

¹ Other design criteria can also be advocated. See, e.g., [51].

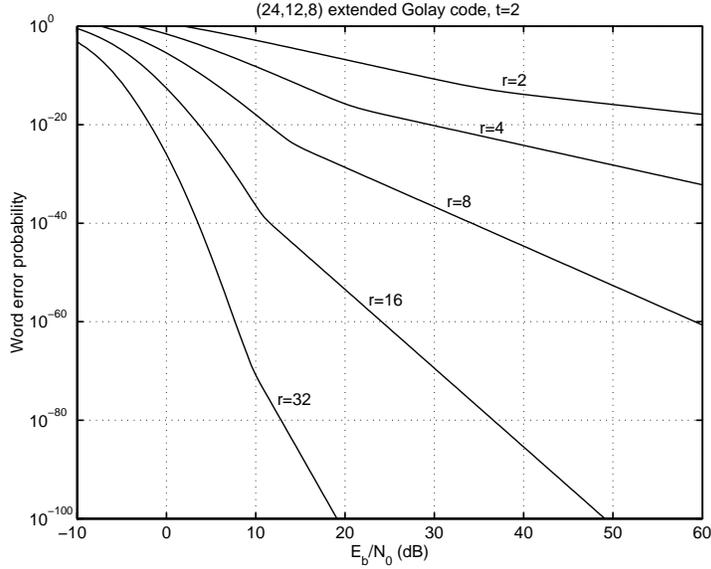


Fig. 6.1 Frame-error probability of the binary (24, 8, 12) extended Golay code with binary PSK over a channel with $t = 2$ transmit antennas and r receive antennas with ML decoding.

in the example of Fig. 6.1, which shows the union upper bound on the word-error probability $P(e)$ of the space-time code obtained by splitting evenly the code words of the (24, 8, 12) extended Golay binary code between two transmit antennas (the calculations are based on the techniques described in Appendix D). This space-time code has the minimum rank of Δ equal to 1, and hence a diversity gain r . Now, it is seen from Fig. 6.1 how the slope predicted by (6.10), and exhibited by a linear behavior in the $P(e)$ -vs.- E_b/N_0 chart, can be reached only for very small values of error probability (how small, generally depends on the weight distribution of the code under scrutiny). To justify this behavior, observe from Fig. 6.1 that for a given value of r the error-probability curve changes its behavior from a “waterfall” shape (for small to intermediate SNR) to a linear shape (high SNR). As the number of receive antennas grows, this change of slope occurs for values of $P(e)$ that are smaller and smaller as r increases. Thus, to study the error-probability curve in its waterfall region it makes sense to examine its asymptotic behavior as $r \rightarrow \infty$. The case $r \rightarrow \infty$, $t < \infty$

can easily be dealt with by using the strong law of large numbers: this yields $\mathbf{H}^\dagger \mathbf{H} \rightarrow \mathbf{I}_t$ a.s., \mathbf{I}_t the $t \times t$ identity matrix. We can see that, as $r \rightarrow \infty$,

$$\|\mathbf{H}\Delta\|^2 \rightarrow \|\Delta\|^2 \quad (6.18)$$

and hence

$$P(\mathbf{X} \rightarrow \hat{\mathbf{X}}) \rightarrow Q\left(\frac{\|\Delta\|}{\sqrt{2N_0}}\right). \quad (6.19)$$

This result shows that, as the number of receiving antennas grows large, the union bound on the error probability of the space-time code depends only on the Euclidean distances between pairs of code words. This is the result one would get with a transmission occurring over a non-fading additive white Gaussian noise (AWGN) channel whose transfer matrix \mathbf{H} has orthogonal columns, i.e., is such that $\mathbf{H}^\dagger \mathbf{H}$ is a scalar matrix. In this situation the smallest error probability, at the expense of a larger complexity, can be achieved by using a single code, optimized for the AWGN channel, whose words of length tN are equally split among the transmit antennas. Within this framework, the number of transmit antennas does not affect the PEP, but only the transmission rate which, expressed in bits per channel use, increases linearly with t .

For another example, observe Fig. 6.2. This shows how for intermediate SNRs the Euclidean criterion may yield codes better than the rank-and-determinant criterion. It compares the simulated performances, in terms of frame-error rate, of the 4-state, rate-1/2 space-time code of [103] and a comparable space-time code obtained by choosing a good binary, 4-state, rate-2/4 convolutional code [28] and mapping its symbols onto QPSK (the first and second encoded bits are mapped onto the QPSK symbol transmitted by the first antenna by Gray encoding, while the third and fourth encoded bits are mapped onto the QPSK symbol transmitted by the second antenna by Gray encoding). The frame length N is 130 symbols for both codes, including 1 symbol for trellis termination. It is seen that in the error-probability range of these two figures the “standard” convolutional code generally outperforms the space-time code of [103] even for small values of r .

Decoding is performed by using the Viterbi algorithm with perfect CSI at the receiver.

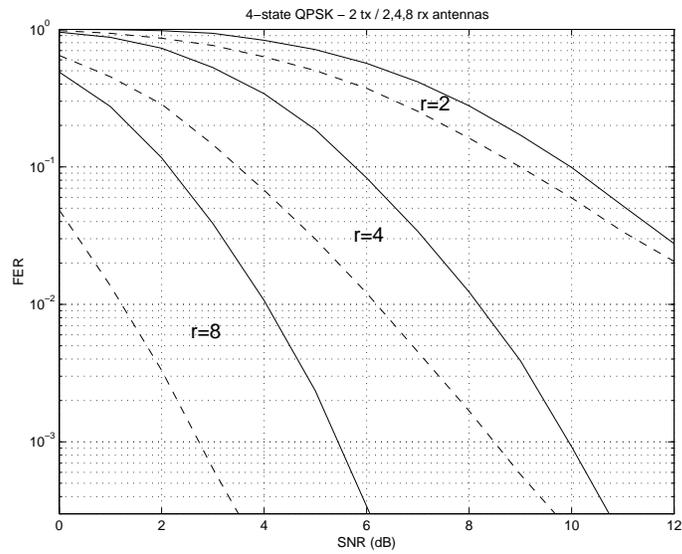


Fig. 6.2 Frame-error rates of two space-time codes with 4 states, rate 1/2, and QPSK. Number of transmit antennas: $t = 2$; number of receive antennas: $r = 2, 4, 8$. Continuous line: Code from [103]. Dashed line: Code obtained from a binary convolutional code good for the AWGN channel [28].

7

Some practical coding schemes

7.1 Delay diversity

One of the first coding schemes proposed [92, 118], is called *delay diversity*. This is a rate- $1/t$ repetition code, each of whose symbols is transmitted from a different antenna after being delayed. For example, with $t = 2$ the transmitted code matrix is

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots \\ 0 & x_1 & x_2 & \cdots \end{bmatrix}.$$

We can see that each symbol traverses rt paths, so that diversity rt is achieved. On the other hand, this comes at the cost of having a rate of only 1 symbol per channel use. Also, observe that delay diversity transforms the frequency-flat channel into an intersymbol-interference (and hence frequency-selective) channel. Optimum detection can be accomplished by using the Viterbi algorithm [5, Chap. 7] or standard equalization techniques (see *infra* the discussion of V-BLAST).

7.2 Alamouti scheme

We describe first this scheme by considering the simple case $t = 2$, $r = 1$, which yields the scheme illustrated in Fig. 7.1. The code matrix

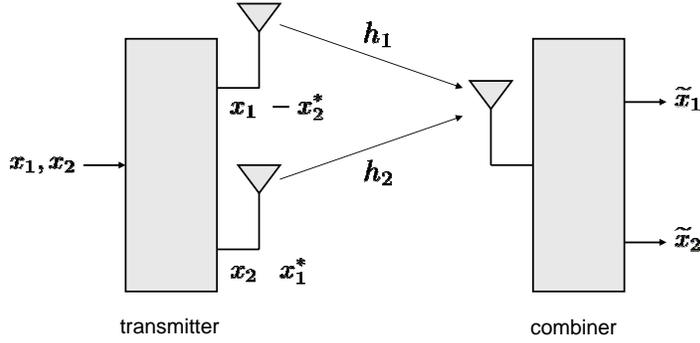


Fig. 7.1 Alamouti transmit-diversity scheme with $t = 2$ and $r = 1$.

\mathbf{X} has the form

$$\mathbf{X} = \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix}. \quad (7.1)$$

This means that during the first symbol interval the signal x_1 is transmitted from antenna 1, while signal x_2 is transmitted from antenna 2. During the next symbol period, antenna 1 transmits signal $-x_2^*$, and antenna 2 transmits signal x_1^* . Thus, the signals received in two adjacent time slots are

$$y_1 = h_1 x_1 + h_2 x_2 + z_1$$

and

$$y_2 = -h_1 x_2^* + h_2 x_1^* + z_2,$$

where h_1, h_2 denote the path gains from the two transmit antennas to the receive antenna. The combiner of Fig. 7.1, which has perfect CSI and hence knows the values of the path gains, generates the signals

$$\tilde{x}_1 = h_1^* y_1 + h_2 y_2^*$$

and

$$\tilde{x}_2 = h_2^* y_1 - h_1 y_2^*$$

so that

$$\begin{aligned} \tilde{x}_1 &= h_1^*(h_1 x_1 + h_2 x_2 + z_1) + h_2(-h_1^* x_2 + h_2^* x_1 + z_2^*) \\ &= (|h_1|^2 + |h_2|^2)x_1 + (h_1^* z_1 + h_2 z_2^*) \end{aligned} \quad (7.2)$$

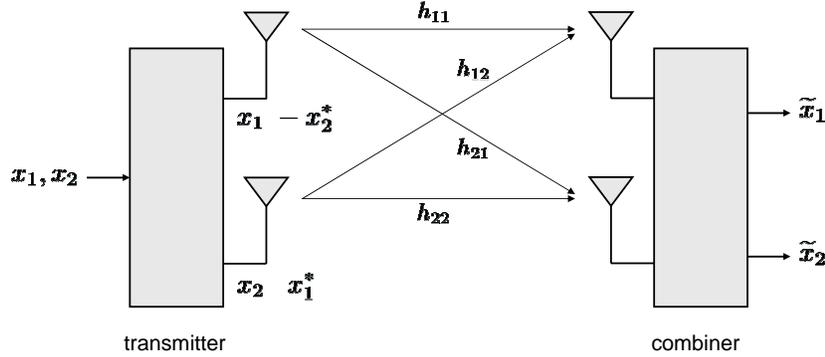


Fig. 7.2 Alamouti transmit-diversity scheme with $t = 2$ and $r = 2$.

and similarly

$$\tilde{x}_2 = (|h_1|^2 + |h_2|^2)x_2 + (h_2^*z_1 - h_1z_2^*). \quad (7.3)$$

Thus, x_1 is separated from x_2 . This scheme has the same performance as one with $t = 1$, $r = 2$, and maximal-ratio combining [5, Chap. 13] (provided that each transmit antenna transmits the same power as the single antenna for $t = 1$). To prove the last statement, observe that if the signal x_1 is transmitted, the two receive antennas observe $h_1x_1 + n_1$ and $h_2x_1 + n_2$, respectively, and after maximal-ratio combining the decision variable is

$$h_1^*(h_1x_1 + n_1) + h_2^*(h_2x_1 + n_2) = (|h_1|^2 + |h_2|^2)x_1 + (h_1^*z_1 + h_2z_2^*) = \tilde{x}_1.$$

This scheme can be generalized to other values of r . For example, with $t = r = 2$ and the same transmission scheme as before (see Fig. 7.2), we have, if $y_{11}, y_{12}, y_{21}, y_{22}$, denote the signals received by antenna 1 at time 1, by antenna 1 at time 2, by antenna 2 at time 1, and by antenna 2 at time 2, respectively,

$$\begin{aligned} \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} &= \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix} + \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \\ &= \begin{bmatrix} h_{11}x_1 + h_{12}x_2 + z_{11} & -h_{11}x_2^* + h_{12}x_1^* + z_{12} \\ h_{21}x_1 + h_{22}x_2 + z_{21} & -h_{21}x_2^* + h_{22}x_1^* + z_{22} \end{bmatrix}. \end{aligned}$$

The combiner generates

$$\tilde{x}_1 = h_{11}^* y_{11} + h_{12} y_{12}^* + h_{21}^* y_{21} + h_{22} y_{22}^*$$

and

$$\tilde{x}_2 = h_{12}^* y_{11} - h_{11} y_{12}^* + h_{22}^* y_{21} - h_{21} y_{22}^*,$$

which yields

$$\tilde{x}_1 = (|h_{11}|^2 + |h_{12}|^2 + |h_{21}|^2 + |h_{22}|^2)x_1 + \text{noise}$$

and

$$\tilde{x}_2 = (|h_{11}|^2 + |h_{12}|^2 + |h_{21}|^2 + |h_{22}|^2)x_2 + \text{noise}.$$

As above, it can easily be shown that the performance of this $t = 2$, $r = 2$ scheme is equivalent to that of a $t = 1$, $r = 4$ scheme with maximal-ratio combining (again, provided that each transmit antenna transmits the same power as with $t = 1$).

A general scheme with $t = 2$ and a general value of r can be exhibited: it has the same performance of a single-transmit-antenna scheme with $2r$ receive antennas and maximal-ratio combining.

7.3 Alamouti scheme revisited: Orthogonal designs

We can rewrite the transmitted signal in Alamouti scheme with $t = 2$ and $r = 1$ in the following equivalent form:

$$\begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}. \quad (7.4)$$

Now, if we define

$$\check{\mathbf{H}} \triangleq \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix},$$

we see that

$$\check{\mathbf{H}}^\dagger \check{\mathbf{H}} = (|h_1|^2 + |h_2|^2)\mathbf{I}_2. \quad (7.5)$$

Recalling (6.6)–(6.7), this shows that the error probability for this Alamouti scheme is the same as without spatial interference, and with a signal-to-noise ratio increased by a factor $(|h_1|^2 + |h_2|^2)$. For this reason Alamouti scheme is called an *orthogonal design*. There are also

orthogonal designs with $t > 2$. For example, with $t = 3$, $r = 1$, and $N = 4$ we have

$$\mathbf{X} = \begin{bmatrix} x_1 & -x_2^* & -x_3^* & 0 \\ x_2 & x_1^* & 0 & -x_3^* \\ x_3 & 0 & x_1^* & x_2^* \end{bmatrix}$$

so that the equation $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}$ can be rewritten in the equivalent form

$$\begin{bmatrix} y_1 \\ y_2^* \\ y_3^* \\ y_4^* \end{bmatrix} = \check{\mathbf{H}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \check{\mathbf{z}}, \quad (7.6)$$

where

$$\check{\mathbf{H}} \triangleq \begin{bmatrix} h_1 & h_2 & h_3 \\ h_2^* & -h_1^* & 0 \\ h_3^* & 0 & h_1^* \\ 0 & h_3^* & -h_2^* \end{bmatrix} \quad (7.7)$$

and $\check{\mathbf{z}}$ is a noise 4-vector. In this case we can verify that

$$\check{\mathbf{H}}^\dagger \check{\mathbf{H}} = (|h_1|^2 + |h_2|^2 + |h_3|^2)\mathbf{I}_3.$$

Notice that with this code we transmit three signals in four time intervals, so that its rate is 3/4 signals per channel use, while the original Alamouti schemes transmit 1 signal per channel use. In fact, orthogonal designs with $t > 2$ have rates that cannot exceed 3/4 [116].

To avoid the rate loss of orthogonal designs, algebraic codes can be designed that, for any number of transmit and receive antennas, achieve maximum diversity as Alamouti codes, while the rate is t symbols per channel use (see [71] and references therein).

7.4 Linear space–time codes

Alamouti codes and orthogonal designs share the property of having simple decoders, due to the linearity of their “space–time” map from symbols to transmit antennas. Schemes with this property form the class of *linear space–time codes* [52]. These can be used for any number of transmit and receive antennas, and may outperform orthogonal designs.

In these codes, the L symbols x_1, \dots, x_L are transmitted by t antennas in N time intervals. The code matrix \mathbf{X} has the form

$$\mathbf{X} = \sum_{\ell=1}^L (\alpha_{\ell} \mathbf{A}_{\ell} + j\beta_{\ell} \mathbf{B}_{\ell}), \quad (7.8)$$

where α_{ℓ} and β_{ℓ} are the real and imaginary part of x_{ℓ} , respectively, and $\mathbf{A}_{\ell}, \mathbf{B}_{\ell}$, $\ell = 1, \dots, L$, are $t \times N$ complex matrices.

Example 7.1. With Alamouti codes we may write

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix} \\ &= \begin{bmatrix} \alpha_1 + j\beta_1 & -\alpha_2 + j\beta_2 \\ \alpha_2 + j\beta_2 & \alpha_1 - j\beta_1 \end{bmatrix} \\ &= \alpha_1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + j\beta_1 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + j\beta_2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \end{aligned}$$

which shows them to be a special case of linear space-time codes.

Define the column vectors

$$\check{\mathbf{x}} \triangleq [\alpha_1 \quad \beta_1 \quad \dots \quad \alpha_L \quad \beta_L]', \quad \check{\mathbf{z}} \triangleq \text{vec}(\mathbf{Z})$$

and the $Nr \times 2L$ matrix

$$\check{\mathbf{H}} \triangleq [\text{vec}(\mathbf{H}\mathbf{A}_1) \quad \text{vec}(j\mathbf{H}\mathbf{B}_1) \quad \dots \quad \text{vec}(\mathbf{H}\mathbf{A}_L) \quad \text{vec}(j\mathbf{H}\mathbf{B}_L)].$$

Then we can write the received signal in the form

$$\begin{aligned} \check{\mathbf{y}} &\triangleq \text{vec}(\mathbf{Y}) = \text{vec}(\mathbf{H}\mathbf{X} + \mathbf{Z}) \\ &= \sum_{\ell=1}^L (\alpha_{\ell} \text{vec}(\mathbf{H}\mathbf{A}_{\ell}) + \beta_{\ell} \text{vec}(j\mathbf{H}\mathbf{B}_{\ell})) = \check{\mathbf{H}}\check{\mathbf{x}} + \check{\mathbf{z}}. \end{aligned}$$

Notice that, since L signals are transmitted and $\check{\mathbf{y}}$ has Nr components, to be able to recover $\check{\mathbf{x}}$ from $\check{\mathbf{y}}$ we must have $L \leq Nr$.

The observed signal $\check{\mathbf{y}}$ can be decoded as follows. Perform the QR factorization of $\check{\mathbf{H}}$:

$$\check{\mathbf{H}} = \check{\mathbf{Q}}\check{\mathbf{R}},$$

where $\check{\mathbf{Q}}$ is unitary, and $\check{\mathbf{R}}$ is an upper triangular matrix. Thus, if we operate a linear transformation on $\check{\mathbf{y}}$ consisting of its premultiplication by $\check{\mathbf{Q}}^\dagger$, we obtain (disregarding the noise for simplicity) a vector $\check{\mathbf{R}}\check{\mathbf{x}}$, whose last entry is proportional to β_L . From this, β_L can be detected. The next-to-last entry is a linear combination of α_L and β_L : thus, since β_L has already been detected, and hence its contribution to spatial interference can be canceled, we may use this entry to detect α_L . The third-from-last entry is a linear combination of β_{L-1} , α_L , and β_L . This can be used to detect β_{L-1} , and so on. This “nulling-and-canceling” idea will be reprised *infra*, with some additional details, in our discussion of zero-forcing V-BLAST. More generally, our treatment of V-BLAST can be applied, *mutatis mutandis*, to linear space–time codes.

7.5 Trellis space–time codes

Trellis space–time codes are trellis-coded modulation (TCM) schemes [7], in which every transition among states, as described by a trellis, is labeled by t signals, each being associated with one transmit antenna [104]. Trellis space–time codes can achieve higher rates than orthogonal designs, but they suffer from a complexity which grows exponentially in the number of transmit antennas.

Example 7.2. Examples of a space–time codes are shown in Figs. 7.3 and 7.4 through their trellises. The code in Fig. 7.3 has $t = 2$, four states, and transmits 2 bits per channel use by using 4PSK, whose signals are denoted 0, 1, 2, 3. With $r = 1$ its diversity is 2. Label xy means that signal x is transmitted by antenna 1, while signal y is simultaneously transmitted by antenna 2. The code in Fig. 7.4 has again $t = 2$, eight states, and transmits 3 bits per channel use by using 8PSK, whose signals are denoted 0, 1, . . . , 7. With $r = 1$ its diversity is 2.

7.6 Space–time codes when CSI is not available

In a rapidly-changing mobile environment, or when long training sequences are not allowed, the assumption of perfect CSI at the receiver

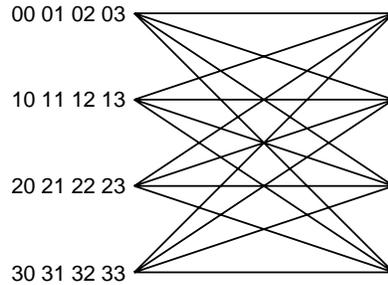


Fig. 7.3 A 4-PSK trellis space-time coding scheme with $t = 2$ and diversity $2r$.

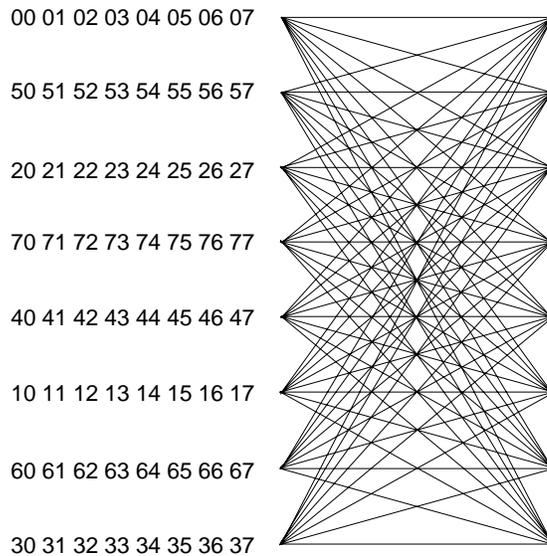


Fig. 7.4 An 8-PSK trellis space-time coding scheme with $t = 2$ and diversity $2r$.

may not be valid. In the absence of CSI at the receiver, [57, 74] advocate *unitary space-time modulation*, a technique which circumvents the use of training symbols (which entail a rate loss). Here the information is carried on the subspace that is spanned by orthonormal signals that are sent. This subspace survives multiplication by the unknown channel-gain matrix \mathbf{H} . A scheme based on *differential unitary space-time* signals is advocated in [59]. High-rate constellations with excellent performance, obtained via algebraic techniques, are described in [54].

8

Suboptimum receiver interfaces

From the capacity results described above we can see that extremely large spectral efficiencies can be achieved on a wireless link if the number of transmit and receive antennas is large. However, as t and r increase, the complexity of space-time coding with maximum-likelihood detection may become too large. This motivates the design of suboptimal receivers whose complexity is lower than ML, and yet perform close to it. In a receiver we distinguish an *interface*, which is a system accepting as its input the channel observation \mathbf{Y} and generating a “soft estimate” $\tilde{\mathbf{Y}}$ of the code matrix \mathbf{X} , and a decoder, whose input and output are $\tilde{\mathbf{Y}}$ and the decoded matrix $\hat{\mathbf{X}}$, respectively.

We describe here some of these interfaces, categorized as *linear* and *nonlinear*.¹

¹ Other reduced-complexity receiver interfaces can be envisaged. For example, in [75] a scheme is advocated where $r' < r$ antennas are used, by selecting the r' best received signals. As long as $r' \geq t$, the capacity achieved by this system is close to that of a full-complexity system.

8.1 Linear interfaces

A linear interface operates a linear transformation \mathbf{A} of the received signal, with $\mathbf{A} = \mathbf{A}(\mathbf{H})$ a $t \times r$ matrix chosen so as to allow a simplification of the metrics used in the Viterbi algorithm employed for soft decoding of \mathcal{X} . The conditional PEP for this linear interface with the metric $\|\mathbf{A}\mathbf{Y} - \mathbf{X}\|^2$ is given by

$$\begin{aligned} \mathbb{P}(\mathbf{X} \rightarrow \hat{\mathbf{X}} \mid \mathbf{H}) &= \mathbb{P}(\|\mathbf{A}\mathbf{Y} - \hat{\mathbf{X}}\|^2 < \|\mathbf{A}\mathbf{Y} - \mathbf{X}\|^2 \mid \mathbf{H}) \\ &= \mathbb{P}(\|\mathbf{A}\mathbf{H}\mathbf{X} - \hat{\mathbf{X}} + \mathbf{A}\mathbf{Z}\|^2 < \|\mathbf{A}\mathbf{H}\mathbf{X} - \mathbf{X} + \mathbf{A}\mathbf{Z}\|^2 \mid \mathbf{H}) \\ &= \mathbb{P}(\|\Delta\|^2 + 2((\mathbf{A}\mathbf{H} - \mathbf{I})\mathbf{X}, \Delta) + 2(\mathbf{A}\mathbf{Z}, \Delta) < 0 \mid \mathbf{H}). \end{aligned} \quad (8.1)$$

By observing that $(\mathbf{A}\mathbf{Z}, \Delta)$ is a zero-mean circularly-symmetric complex Gaussian RV with variance $N_0\|\mathbf{A}^\dagger\Delta\|^2$, the unconditional PEP becomes

$$\mathbb{P}(\mathbf{X} \rightarrow \hat{\mathbf{X}}) = \mathbb{E} \left[Q \left(\frac{\|\Delta\|^2 + 2((\mathbf{A}\mathbf{H} - \mathbf{I})\mathbf{X}, \Delta)}{\sqrt{2N_0\|\mathbf{A}^\dagger\Delta\|^2}} \right) \right]. \quad (8.2)$$

8.1.1 Zero-forcing interface

This consists of choosing $\mathbf{A} = \mathbf{H}^+$, where the superscript $+$ denotes the Moore–Penrose pseudoinverse of a matrix. For future reference, we note that we have

$$\mathbf{H}^+(\mathbf{H}^+)^\dagger = (\mathbf{H}^\dagger\mathbf{H})^{-1}. \quad (8.3)$$

If we assume $r \geq t$, then $\mathbf{H}^\dagger\mathbf{H}$ is invertible with probability 1, and we have

$$\mathbf{H}^+ = (\mathbf{H}^\dagger\mathbf{H})^{-1}\mathbf{H}^\dagger \quad (8.4)$$

so that

$$\mathbf{H}^+\mathbf{Y} = \mathbf{X} + \mathbf{H}^+\mathbf{Z}, \quad (8.5)$$

which shows that the spatial interference is completely removed from the received signal, thus justifying the name “zero forcing” associated with this interface. The metric used here is then $\|\mathbf{H}^+\mathbf{Y} - \mathbf{X}\|^2$.

From (8.1), the conditional PEP is given by

$$P(\mathbf{X} \rightarrow \hat{\mathbf{X}} \mid \mathbf{H}) = Q \left(\frac{\|\Delta\|^2}{2\sigma} \right), \quad (8.6)$$

where, by (8.3),

$$\begin{aligned}\sigma^2 &\triangleq \mathbb{V}[(\mathbf{\Delta}, \mathbf{H}^+ \mathbf{Z})] \\ &= \frac{N_0}{2} \text{Tr} [\mathbf{\Delta}^\dagger \mathbf{H}^+ (\mathbf{H}^+)^\dagger \mathbf{\Delta}] \\ &= \frac{N_0}{2} \text{Tr} [\mathbf{\Delta}^\dagger (\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{\Delta}].\end{aligned}\quad (8.7)$$

8.1.2 Linear MMSE interface

Here we choose the matrix \mathbf{A} that minimizes the mean-square value of the spatial interference plus noise. Define the MSE as

$$\varepsilon^2(\mathbf{A}) \triangleq \mathbb{E}[\|\mathbf{A}\mathbf{y} - \mathbf{x}\|^2] = \mathbb{E}[\text{Tr}((\mathbf{A}\mathbf{H} - \mathbf{I}_t)\mathbf{x} + \mathbf{A}\mathbf{z})((\mathbf{A}\mathbf{H} - \mathbf{I}_t)\mathbf{x} + \mathbf{A}\mathbf{z})^\dagger].\quad (8.8)$$

Using the standard simplifying assumption of iid zero-mean components of \mathbf{x} (with second moment E_s) we obtain the following expression:

$$\varepsilon^2(\mathbf{A}) = \text{Tr}(E_s(\mathbf{A}\mathbf{H} - \mathbf{I}_t)(\mathbf{A}\mathbf{H} - \mathbf{I}_t)^\dagger + N_0\mathbf{A}\mathbf{A}^\dagger).\quad (8.9)$$

The variation of $\varepsilon^2(\mathbf{A})$ with respect to \mathbf{A} is then given by

$$\delta(\varepsilon^2) = \text{Tr} \left\{ \delta\mathbf{A} [E_s\mathbf{H}(\mathbf{A}\mathbf{H} - \mathbf{I}_t)^\dagger + N_0\mathbf{A}^\dagger] + [E_s(\mathbf{A}\mathbf{H} - \mathbf{I}_t)\mathbf{H}^\dagger + N_0\mathbf{A}] \delta\mathbf{A}^\dagger \right\}.\quad (8.10)$$

The corresponding stationary point obtained by nulling this variation yields the MMSE solution:

$$\mathbf{A} = \mathbf{A}_{\text{mmse}} \triangleq \mathbf{H}^\dagger (\mathbf{H}\mathbf{H}^\dagger + \delta_s \mathbf{I}_r)^{-1} = (\mathbf{H}^\dagger \mathbf{H} + \delta_s \mathbf{I}_t)^{-1} \mathbf{H}^\dagger,\quad (8.11)$$

where $\delta_s \triangleq N_0/E_s$. From (8.2) we obtain

$$P(\mathbf{X} \rightarrow \widehat{\mathbf{X}}) = \mathbb{E} \left[Q \left(\frac{\|\mathbf{\Delta}\|^2 + 2((\mathbf{H}^\dagger \mathbf{H} + \delta_s \mathbf{I}_t)^{-1} \mathbf{H}^\dagger \mathbf{H} - \mathbf{I}_t) \mathbf{X}, \mathbf{\Delta}}{\sqrt{2N_0 \|\mathbf{H}(\mathbf{H}^\dagger \mathbf{H} + \delta_s \mathbf{I}_t)^{-1} \mathbf{\Delta}\|^2}} \right) \right].\quad (8.12)$$

Notice that as $\delta_s \rightarrow 0$ the right-hand side of (8.12) tends to the PEP of the zero-forcing detector, as it should.

8.1.3 Asymptotic performance of linear interfaces: Finite t , $r \rightarrow \infty$

Here we consider the case $r \gg t$ by examining the asymptotic performance obtained when $r \rightarrow \infty$ while t remains constant. By the strong

law of large numbers we have, as $r \rightarrow \infty$,

$$\mathbf{H}^\dagger \mathbf{H} \rightarrow \mathbf{I}_t \quad \text{a.s.} \quad (8.13)$$

and we have previously seen from (6.19) that with ML detection the pairwise error probability tends to that of a nonfading AWGN channel (no spatial interference). Using (8.13) in (8.7) and in (8.12), we see that, asymptotically, ZF and MMSE interfaces do not entail any loss of performance with respect to ML.

8.1.4 Asymptotic performance of linear interfaces: $t, r \rightarrow \infty$ with $t/r \rightarrow \alpha > 0$

Things change if both t and r grow to infinity while their ratio tends to a constant positive value α . In this case an SNR loss is expected, as we are going to illustrate for the ZF interface. We evaluate asymptotically (8.6), and compare this result to the corresponding PEP of the ML interface. Theorem C.2 shows that, as $t, r \rightarrow \infty$ with $t/r \rightarrow \alpha$, the cumulative empirical eigenvalue distribution of $\mathbf{H}^\dagger \mathbf{H}/r$ converges to a function $F(\lambda; \alpha)$ whose derivative is given by:

$$\frac{\partial}{\partial \lambda} F(\lambda; \alpha) = f(\lambda; \alpha) \triangleq (1 - \alpha^{-1})_+ \delta(\lambda) + \alpha^{-1} \frac{\sqrt{(\lambda - \lambda_-)_+ (\lambda_+ - \lambda)_+}}{2\pi\lambda}, \quad (8.14)$$

where $\lambda_\pm \triangleq (\sqrt{\alpha} \pm 1)^2$. In particular, when $\alpha = 0$ or ∞ , the pdf $f(\lambda; \alpha)$ tends to $\delta(\lambda - 1)$ or $\delta(\lambda)$, respectively.

The asymptotic PEP of the ML and ZF receivers can now be calculated by using results from Free Probability theory [114, 63]. We apply Theorem C.3 assuming that the matrix sequences \mathbf{A}_n and \mathbf{B}_n are given by $\mathbf{W} \triangleq \mathbf{H}^\dagger \mathbf{H}/r$ and $\mathbf{\Delta} \mathbf{\Delta}^\dagger$ as $r \rightarrow \infty$. Then, for the ML receiver we have

$$\begin{aligned} \frac{\|\mathbf{H}\mathbf{\Delta}\|^2}{2N_0} &= \frac{1}{2N_0} \text{Tr}(\mathbf{H}^\dagger \mathbf{H} \mathbf{\Delta} \mathbf{\Delta}^\dagger) \\ &= \frac{rt}{2N_0} \tau(\mathbf{W} \mathbf{\Delta} \mathbf{\Delta}^\dagger) \\ &\rightarrow \frac{rt}{2N_0} \mathbb{E}[\tau(\mathbf{W})] \tau(\mathbf{\Delta} \mathbf{\Delta}^\dagger) \quad (\text{a.s. as } t, r \rightarrow \infty, t/r \rightarrow \alpha), \end{aligned} \quad (8.15)$$

where $\tau(\mathbf{A}) \triangleq \text{Tr}(\mathbf{A})/n$ for every $n \times n$ matrix \mathbf{A} . Since

$$\mathbb{E}[\tau(\mathbf{W})] \rightarrow \int_a^b \lambda f(\lambda; \alpha) d\lambda = 1, \quad (8.16)$$

we obtain

$$\frac{\|\mathbf{H}\mathbf{\Delta}\|^2}{2N_0} \rightarrow \frac{r\|\mathbf{\Delta}\|^2}{2N_0} \quad (\text{a.s. as } t, r \rightarrow \infty, t/r \rightarrow \alpha) \quad (8.17)$$

and hence

$$\mathbb{P}(\mathbf{X} \rightarrow \widehat{\mathbf{X}}) \rightarrow Q\left(\frac{r\|\mathbf{\Delta}\|^2}{2N_0}\right) \quad (\text{a.s. as } t, r \rightarrow \infty, t/r \rightarrow \alpha). \quad (8.18)$$

For the ZF receiver we have from (8.7):

$$\sigma^2 = \frac{tN_0}{2r} \tau(\mathbf{W}^{-1} \mathbf{\Delta} \mathbf{\Delta}^\dagger) \rightarrow \frac{tN_0}{2r} \mathbb{E}[\tau(\mathbf{W}^{-1})] \tau(\mathbf{\Delta} \mathbf{\Delta}^\dagger). \quad (8.19)$$

Since

$$\mathbb{E}[\tau(\mathbf{W}^{-1})] \rightarrow \int_a^b \lambda^{-1} f(\lambda; \alpha) d\lambda = \frac{1}{1-\alpha}, \quad (8.20)$$

we obtain

$$\frac{\|\mathbf{\Delta}\|^4}{4\sigma^2} \rightarrow (1-\alpha) \frac{r\|\mathbf{\Delta}\|^2}{2N_0} \quad (\text{a.s. as } t, r \rightarrow \infty, t/r \rightarrow \alpha) \quad (8.21)$$

and hence

$$\mathbb{P}(\mathbf{X} \rightarrow \widehat{\mathbf{X}}) \rightarrow Q\left((1-\alpha) \frac{r\|\mathbf{\Delta}\|^2}{2N_0}\right) \quad (\text{a.s. as } t, r \rightarrow \infty, t/r \rightarrow \alpha). \quad (8.22)$$

Thus, the asymptotic SNR loss with respect to the ML interface is equal to $(1-\alpha)^{-1}$ for the ZF interface, which predicts that the choice $r = t$ with a large number of antennas yields a considerable loss in performance [12]. From this we may expect that these linear interfaces exhibit a PEP close to ML only for $r \gg t$; otherwise the performance loss may be substantial. This is validated by Fig. 8.1, which shows the error probability of a multiple-antenna system where the binary (8, 4, 4) Reed–Muller code is used by splitting its code words evenly between 2 transmit antennas. The word error probabilities shown are obtained through Monte Carlo simulation. Binary PSK is used, and the code rate is 1 bit per channel use. It is seen that for $r = 2$ both MMSE and ZF interface exhibit a considerable performance loss with respect to ML, while for $r = 8$ the losses are very moderate.

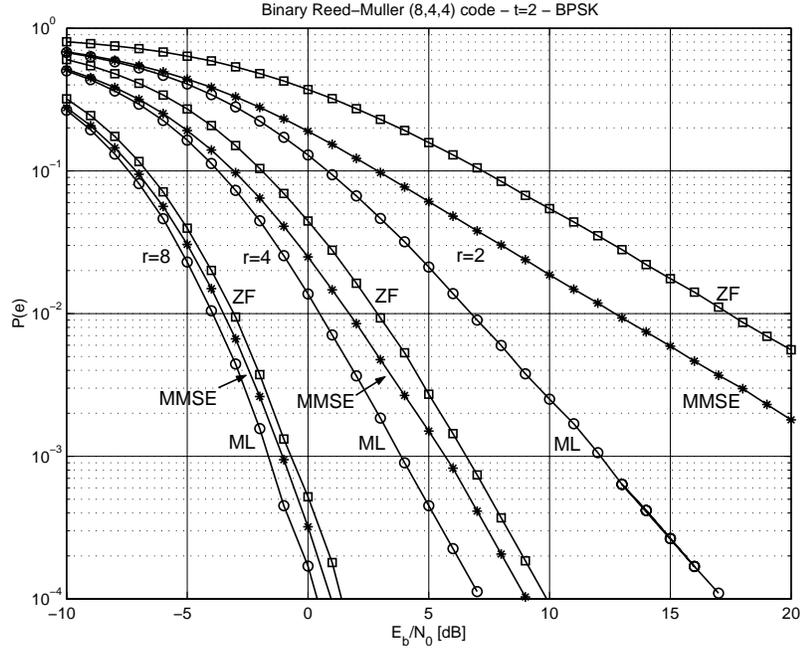


Fig. 8.1 Word error probability of the binary (8, 4, 4) Reed–Muller code with binary PSK over a channel with $t = 2$ transmit antennas and r receive antennas with ML, MMSE, and ZF interfaces (computer simulation results).

8.2 Nonlinear interfaces

The task of reducing the spatial interference affecting the received signal can be accomplished by first processing \mathbf{Y} linearly, then subtracting from the result an estimate of the spatial interference obtained from preliminary decisions on the transmitted code word. The metric used for decoding is $\|\tilde{\mathbf{Y}} - \mathbf{X}\|$, where $\tilde{\mathbf{Y}}$ is a “soft estimate” of \mathbf{X} given by

$$\tilde{\mathbf{Y}} \triangleq \mathbf{G}\mathbf{Y} - \mathbf{L}\hat{\mathbf{X}} \quad (8.23)$$

for a suitable choice of the two matrices \mathbf{G} and \mathbf{L} (Fig. 8.2). The diagonal entries of the matrix \mathbf{L} must be zero in order to have only spatial interference subtracted from $\mathbf{G}\mathbf{Y}$.

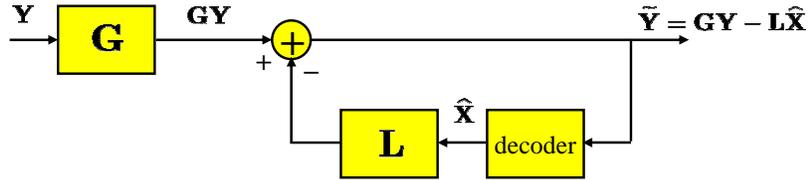


Fig. 8.2 General structure of a nonlinear interface.

8.2.1 Vertical BLAST interface

One nonlinear interface is called vertical BLAST (this stands for Bell Laboratories Layered Space–Time Architecture). With V-BLAST, the data are divided into t substreams to be transmitted on different antennas. The receiver pre-processes linearly the received signal by forming the matrix \mathbf{GY} , which has t rows. Then, it first decodes one of the substreams after reducing the spatial interference coming from the others. Next, the contribution of this substream is subtracted from the received signal, and the second substream is decoded after reducing the remaining spatial interference. This process is repeated t times.

Different implementations of the basic V-BLAST idea are possible, two of them being the zero-forcing *ZF V-BLAST* interface and the minimum-mean-square-error *MMSE V-BLAST* interface. These arise from the minimization of the mean-square error of the spatial interference without or with noise, respectively.

It should be observed that the performance of V-BLAST depends on the order in which the substreams are decoded (in the algorithm above, the actual numbering of the rows of \mathbf{GY} is arbitrary), and on the data rate associated with each substream. Several strategies are possible here (see [13] and references therein, and [121, 119, 41]): the decoding order may be predefined, and the data rates may be the same; or an ordering is chosen so as to maximize an SNR-related parameter, with equal data rates; or different data rates are assigned to different substreams (see, e.g., [70]).

ZF V-BLAST When the presence of noise is disregarded, the MSE of the disturbance can be written as

$$\varepsilon^2(\mathbf{G}, \mathbf{L}) = \mathbb{E}[\|\tilde{\mathbf{Y}} - \mathbf{X}\|^2] = \mathbb{E}[\|\mathbf{GHX} - \mathbf{L}\hat{\mathbf{X}} - \mathbf{X}\|^2]. \quad (8.24)$$

Under the approximations

$$\begin{aligned} \mathbb{E}[\mathbf{X}\hat{\mathbf{X}}^\dagger] &\approx \mathbb{E}[\mathbf{X}\mathbf{X}^\dagger] \\ \mathbb{E}[\hat{\mathbf{X}}\hat{\mathbf{X}}^\dagger] &\approx \mathbb{E}[\mathbf{X}\mathbf{X}^\dagger] \end{aligned} \quad (8.25)$$

(which are justified by the assumption of having $\hat{\mathbf{X}} \approx \mathbf{X}$ unless the error probability is high) we obtain

$$\begin{aligned} \varepsilon^2(\mathbf{G}, \mathbf{L}) &= \mathbb{E}[\|(\mathbf{GH} - \mathbf{L} - \mathbf{I}_t)\mathbf{X}\|^2] \\ &= \mathbb{E}[\text{Tr}\{(\mathbf{GH} - \mathbf{L} - \mathbf{I}_t)^\dagger(\mathbf{GH} - \mathbf{L} - \mathbf{I}_t)\mathbf{X}\mathbf{X}^\dagger\}] \\ &= NE_s\mathbb{E}[\|\mathbf{GH} - \mathbf{L} - \mathbf{I}_t\|^2], \end{aligned} \quad (8.26)$$

since $\mathbb{E}[\mathbf{X}\mathbf{X}^\dagger] = NE_s\mathbf{I}_t$. From the QR factorization [62] of \mathbf{H} ,

$$\underbrace{\mathbf{H}}_{r \times t} = \underbrace{\mathbf{Q}}_{r \times t} \underbrace{\mathbf{R}}_{t \times t}$$

(where \mathbf{R} is an upper triangular matrix), we see that the MSE $\varepsilon^2(\mathbf{G}, \mathbf{L})$ vanishes by setting

$$\begin{cases} \mathbf{G} = \text{diag}^{-1}(\mathbf{R})\mathbf{Q}^\dagger \\ \mathbf{L} = \text{diag}^{-1}(\mathbf{R})\mathbf{R} - \mathbf{I}_t. \end{cases} \quad (8.27)$$

The block diagram of Fig. 8.2 illustrates that ZF V-BLAST corresponds to having a strictly upper diagonal matrix \mathbf{L} . Explicitly, the steps of the ZF V-BLAST algorithm proceed as follows, indicating by $(\mathbf{A})_i$ the i th row of matrix \mathbf{A} , by $(\mathbf{A})_{ij}$ its entry in i th row and j th column, and by \implies the result of decoding:

$$\begin{cases} (\tilde{\mathbf{Y}})_t &= (\mathbf{G}\mathbf{Y})_t &\implies (\hat{\mathbf{X}})_t \\ (\tilde{\mathbf{Y}})_{t-1} &= (\mathbf{G}\mathbf{Y})_{t-1} - (\mathbf{L})_{t-1,t}(\hat{\mathbf{X}})_t &\implies (\hat{\mathbf{X}})_{t-1} \\ (\tilde{\mathbf{Y}})_{t-2} &= (\mathbf{G}\mathbf{Y})_{t-2} - (\mathbf{L})_{t-2,t}(\hat{\mathbf{X}})_t - (\mathbf{L})_{t-2,t-1}(\hat{\mathbf{X}})_{t-1} &\implies (\hat{\mathbf{X}})_{t-2} \\ &\vdots & \\ (\tilde{\mathbf{Y}})_1 &= (\mathbf{G}\mathbf{Y})_1 - (\mathbf{L})_{1,t}(\hat{\mathbf{X}})_t - \dots - (\mathbf{L})_{1,2}(\hat{\mathbf{X}})_2 &\implies (\hat{\mathbf{X}})_1. \end{cases}$$

The soft estimate of \mathbf{X} can be written as

$$\begin{aligned}\tilde{\mathbf{Y}} &= \text{diag}^{-1}(\mathbf{R})\mathbf{Q}^\dagger\mathbf{Y} - [\text{diag}^{-1}(\mathbf{R})\mathbf{R} - \mathbf{I}_t]\hat{\mathbf{X}} \\ &= \underbrace{\mathbf{X}}_{\textcircled{1}} + \underbrace{[\text{diag}^{-1}(\mathbf{R})\mathbf{R} - \mathbf{I}_t]\boldsymbol{\Delta}}_{\textcircled{2}} + \underbrace{\text{diag}^{-1}(\mathbf{R})\mathbf{Q}^\dagger\mathbf{Z}}_{\textcircled{3}}.\end{aligned}\quad (8.28)$$

The three terms in the last expression are: ① the useful term (which is free of spatial interference, thus justifying the name “zero forcing” associated with this interface); ② the interference due to past wrong decisions; and ③ colored noise.

MMSE V-BLAST This minimizes the MSE of the disturbance $\tilde{\mathbf{Y}} - \mathbf{X}$ taking into account the presence of noise. Again, under the approximations (8.25) we can write the MSE as

$$\begin{aligned}\varepsilon^2(\mathbf{G}, \mathbf{L}) &= \mathbb{E}[\|\mathbf{G}\mathbf{Y} - \mathbf{L}\mathbf{X} - \mathbf{X}\|^2] \\ &= \mathbb{E}[\|(\mathbf{G}\mathbf{H} - \mathbf{L} - \mathbf{I}_t)\mathbf{X} + \mathbf{G}\mathbf{Z}\|^2] \\ &= NE_s \left[\|\mathbf{G}\mathbf{H} - \mathbf{L} - \mathbf{I}_t\|^2 + \delta_s \|\mathbf{G}\|^2 \right],\end{aligned}\quad (8.29)$$

where $\delta_s \triangleq N_0/E_s$. The minimum MSE can be found in two steps:

- i)* Minimizing $\varepsilon^2(\mathbf{G}, \mathbf{L})$ over the set of matrices $\mathbf{G} \in \mathbb{C}^{t \times r}$ leads to

$$\mathbf{G}_{\text{mmse}} = (\mathbf{L} + \mathbf{I}_t)(\mathbf{H}^\dagger\mathbf{H} + \delta_s\mathbf{I}_t)^{-1}\mathbf{H}^\dagger. \quad (8.30)$$

The corresponding minimum MSE is

$$\varepsilon_{\text{mmse}}^2(\mathbf{L}) = NN_0 \text{Tr} \left[(\mathbf{L} + \mathbf{I}_t)(\mathbf{H}^\dagger\mathbf{H} + \delta_s\mathbf{I}_t)^{-1}(\mathbf{L} + \mathbf{I}_t)^\dagger \right]. \quad (8.31)$$

- ii)* Next, $\varepsilon_{\text{mmse}}^2(\mathbf{L})$ is minimized over the set of $t \times t$ strictly upper triangular matrices (i.e., such that $[\mathbf{L}]_{ij} = 0$ whenever $i \geq j$). This can be done by using the Cholesky factorization [62] $\mathbf{H}^\dagger\mathbf{H} + \delta_s\mathbf{I}_t = \mathbf{S}^\dagger\mathbf{S}$, where \mathbf{S} is an upper triangular matrix. After using basic multiplication properties of triangular matrices, we obtain the following result:

$$\begin{aligned}\varepsilon_{\text{mmse}}^2(\mathbf{L}) &= NN_0 \text{Tr} \left[(\mathbf{L} + \mathbf{I}_t)(\mathbf{H}^\dagger\mathbf{H} + \delta_s\mathbf{I}_t)^{-1}(\mathbf{L} + \mathbf{I}_t)^\dagger \right] \\ &= NN_0 \|(\mathbf{L} + \mathbf{I}_t)\mathbf{S}^{-1}\|^2\end{aligned}$$

$$\begin{aligned} &\geq NN_0 \|\text{diag}((\mathbf{L} + \mathbf{I}_t)\mathbf{S}^{-1})\|^2 \\ &= NN_0 \|\text{diag}(\mathbf{S}^{-1})\|^2 = NN_0 \sum_{i=1}^t |[\mathbf{S}]_{i,i}|^{-2}. \end{aligned} \quad (8.32)$$

The minimum is attained by setting $\mathbf{L} = \text{diag}^{-1}(\mathbf{S})\mathbf{S} - \mathbf{I}_t$. Thus, $\varepsilon^2(\mathbf{G}, \mathbf{L})$ is minimized by setting

$$\begin{cases} \mathbf{G} = \mathbf{G}_{\text{mmse}} \triangleq \text{diag}^{-1}(\mathbf{S})\mathbf{S}^{-\dagger}\mathbf{H}^\dagger \\ \mathbf{L} = \mathbf{L}_{\text{mmse}} \triangleq \text{diag}^{-1}(\mathbf{S})\mathbf{S} - \mathbf{I}_t \end{cases} \quad (8.33)$$

and

$$\varepsilon_{\text{mmse}}^2 \triangleq \varepsilon^2(\mathbf{G}_{\text{mmse}}, \mathbf{L}_{\text{mmse}}) = NN_0 \sum_{i=1}^t |[\mathbf{S}]_{i,i}|^{-2}. \quad (8.34)$$

As a result, the soft estimate $\tilde{\mathbf{Y}}$ can be written as

$$\begin{aligned} \tilde{\mathbf{Y}} &= \text{diag}^{-1}(\mathbf{S})\mathbf{S}^{-\dagger}\mathbf{H}^\dagger\mathbf{Y} - (\text{diag}^{-1}(\mathbf{S})\mathbf{S} - \mathbf{I}_t)\hat{\mathbf{X}} \\ &= \underbrace{(\mathbf{I}_t - \text{diag}^{-1}(\mathbf{S})\mathbf{S}^{-\dagger})\mathbf{X}}_{\textcircled{1}} + \underbrace{(\text{diag}^{-1}(\mathbf{S})\mathbf{S} - \mathbf{I}_t)\mathbf{\Delta}}_{\textcircled{2}} \\ &\quad + \underbrace{\text{diag}^{-1}(\mathbf{S})\mathbf{S}^{-\dagger}\mathbf{H}^\dagger\mathbf{Z}}_{\textcircled{3}}, \end{aligned} \quad (8.35)$$

where the three terms in the last expression are: ① the (biased) useful term; ② the interference due to past wrong decisions; and ③ colored noise.

8.2.2 Diagonal BLAST interface

Consider the transmission scheme of Fig. 8.3, referred to as Diagonal BLAST (D-BLAST). Here, $\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots$ denote different data substreams. As discussed in Section 9 *infra*, this scheme differs from V-BLAST because each symbol in a data substream is transmitted by a different antenna, and hence is expected to achieve a larger diversity. To obtain this, the information stream is demultiplexed into t substreams, which are transmitted by t antennas through a *diagonal* interleaving scheme. The interleaver is designed so that the symbols of a given substream are cyclically sent over all the t antennas, in order to guarantee the

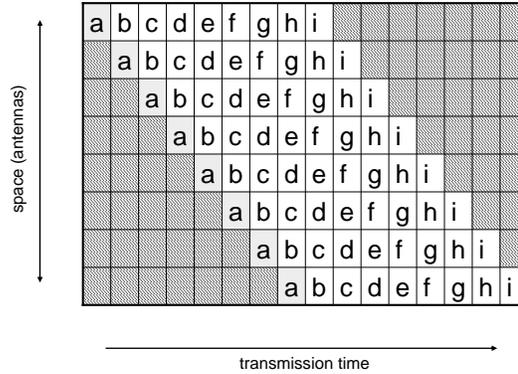


Fig. 8.3 An example of diagonal interleaving with $t = 8$.

necessary diversity order. Diagonals are written from top to bottom, and the letters in each rectangle denote the corresponding code symbol index, i.e., indicate the sequence in which diagonals are filled. Each rectangle in Fig. 8.3 may actually contain an arbitrary number $\kappa \geq 1$ of coded symbols. Each column of t symbols of the diagonal interleaver array is transmitted in parallel, from the t antennas.

To illustrate the operation of D-BLAST, consider a simple case with two transmit antennas. The transmitted matrix has the form

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots \\ 0 & x_{21} & x_{22} & \cdots \end{bmatrix}, \quad (8.36)$$

where x_{ij} is the signal transmitted by the i th antenna in the j th substream. The receiver first detects x_{11} , which is not affected by spatial interference. Then, it detects x_{21} ; this is affected by the spatial interference caused by x_{12} , which can be reduced or nulled, by using for example a zero-forcing filter. Next, the estimates of x_{11} and x_{21} are sent to the decoder of the first substream. Once this has been decoded, its contribution is subtracted out before decoding the second substream, and so forth. Notice that D-BLAST entails a rate loss due to the overhead symbols necessary to start the decoding process (these are shaded in Fig. 8.3).

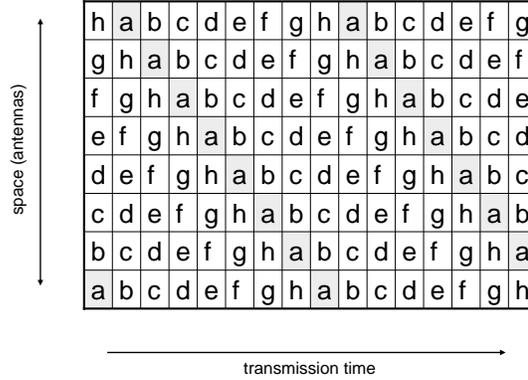


Fig. 8.4 An example of threading with $t = 8$ (each letter represents a layer).

8.2.3 Threaded space–time architecture

To avoid the rate loss implied by D-BLAST, the latter architecture can be generalized by wrapping substreams around as shown in Fig. 8.4. This Figure shows a simple special case of “threaded layering,” whereby the symbols are distributed in the code word matrix so as to achieve full spatial span t (which guarantees the right spatial diversity order) and full temporal span N (which guarantees the right temporal diversity order in the case of fast fading) [37].

8.2.4 BLAST with per-survivor processing

The layered space–time architecture described above is attractive, but exhibits some downsides. First, it requires several independent encoder/decoder pairs running in parallel. Moreover, if used with component trellis codes, not all symbols are decoded with the same decoding delay: this may pose a problem for the underlying Viterbi algorithm. Finally, the interference cancellation procedure on which BLAST is based is prone to error propagation. In order to solve these problems, while keeping the complexity limited, a modification of BLAST was proposed in [21, 22]. There, a single trellis encoder output is wrapped along the transmit antennas by a diagonal interleaver, and the detection scheme

is integrated into a per-survivor processing (PSP) receiver. Thanks to the PSP, the impact of unreliable decisions is greatly reduced.

8.2.5 Iterative interface

An alternative to BLAST consists of performing an iterative spatial interference cancellation (see, e.g., [9, 10, 88, 89, 90] and references therein). Referring again to the block diagram of Fig. 8.2, at iteration k , $k = 0, 1, \dots$, an estimate of the spatial interference is generated in the form

$$\mathbf{W}^{(k)} = (\mathbf{GH} - \text{diag}(\mathbf{GH}))\widehat{\mathbf{X}}^{(k)}. \quad (8.37)$$

Here $\widehat{\mathbf{X}}^{(k)}$ is the decoded word at iteration k , computed by minimizing the metric $\|\widetilde{\mathbf{Y}}^{(k)} - \mathbf{X}\|^2$, where

$$\begin{aligned} \widetilde{\mathbf{Y}}^{(k+1)} &= \widetilde{\mathbf{Y}} - \mathbf{W}^{(k)} \\ &= \widetilde{\mathbf{Y}} - (\mathbf{GH} - \text{diag}(\mathbf{GH}))\widehat{\mathbf{X}}^{(k)} \end{aligned} \quad (8.38)$$

and for $k = 0$ we define $\widehat{\mathbf{X}}^{(0)} \triangleq \mathbf{0}$. It can easily be seen that, if decoding is perfect (that is, if $\widehat{\mathbf{X}}^{(k)} = \mathbf{X}$ for some k), then

$$\widetilde{\mathbf{Y}}^{(k)} = \text{diag}(\mathbf{GH})\mathbf{X} + \mathbf{GZ}, \quad (8.39)$$

which shows that the spatial interference is completely removed.

Full text available at: <http://dx.doi.org/10.1561/0100000002>

9

The fundamental tradeoff

As we briefly mentioned in Section 2, the use of multiple antennas provides at the same time a *rate gain* and a *diversity gain*. The former is due to the fact that multiple, independent transmission paths generate a multiplicity of independent “spatial” channels that can simultaneously be used for transmission. The latter is obtained by exploiting the independent fading gains that affect the same signal, and that can be averaged through to increase the reliability of its detection. Here we examine, based on recent seminal work [108, 121], how the two quantities are related by fundamental performance limits that reflect the ubiquitous tradeoff between rate and transmission quality of a transmission system. Our discussion is qualitative and heuristic; the reader interested in the details of a rigorous proof is addressed to the above papers.

We focus our attention on the nonergodic fading channel of Section 4.5, with channel state information available at the receiver only, and to a high-SNR situation. The latter restriction refers to a system whose performance is not power-limited. We have seen (Observation 4.8) that, as the SNR $\rho \rightarrow \infty$, the capacity of an ergodic Rayleigh fading channel behaves as $C(\rho) \sim m \log \rho$, with $m \triangleq \min\{t, r\}$. Re-

calling the high-SNR expression of the capacity of the single-antenna ergodic Rayleigh fading channel, which is $\log \rho$, the result above can be interpreted by saying that the maximum number of independent parallel channels (or, in a different parlance, the number of degrees of freedom) created by t transmit and r receive antennas equals m , which is the maximum rate gain we can achieve. Consider next the number of independently faded paths: in our model this is equal to tr , which is indeed the maximum achievable diversity gain with maximum-likelihood detection (Observation 6.1).

We discuss here the fact that, while both gains can be achieved by MIMO systems, higher rate gains come at the expenses of diversity gains. We start our discussion by defining precisely what we mean by rate gain and diversity gain in the present context. In a situation where different data rates are involved, *a sequence of codes with increasing rate*, rather than a single code, must be considered. For a fair comparison among codes with different rates, the rate gain is defined by the ratio between the actual code rate $R(\rho)$ and the capacity of the scalar channel at that SNR:

$$\mu \triangleq \lim_{\rho \rightarrow \infty} \frac{R(\rho)}{C(\rho)}. \quad (9.1)$$

This indicates how far the system is operating from the capacity limit. Notice that the capacity increases with the SNR ρ , so to approach capacity the code rate $R(\rho)$ must also increase with ρ ; if a single code were used, the rate gain would vanish, because as ρ increases the ratio (9.1) would tend to zero. As for the diversity gain δ , this is defined as the exponent of ρ^{-1} in the expression of the average error probability of the system:

$$\delta \triangleq - \lim_{\rho \rightarrow \infty} \frac{\log P(e)}{\log \rho}. \quad (9.2)$$

The main point here is that the maximum values of rate gain and diversity gain *cannot be achieved simultaneously*; μ and δ are connected by a tradeoff curve that we are going to introduce and discuss. This curve plots, as a function of the rate gain μ , the maximum achievable diversity gain, denoted $\delta^*(\mu)$.

The tradeoff curve, in the special but important case of a code

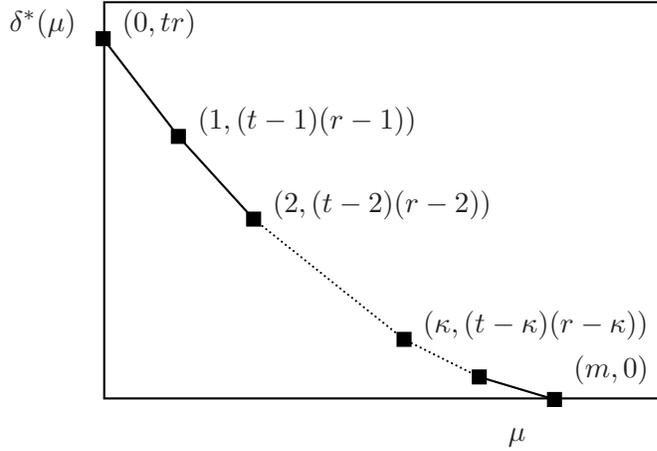


Fig. 9.1 Diversity–rate tradeoff for multiple-antenna systems with t transmit and r receive antennas.

with length $N \geq t + r - 1$,¹ is given by the piecewise-linear function connecting the points $(\kappa, \delta^*(\kappa))$, $\kappa \in \{0, 1, \dots, m\}$, where

$$\delta^*(\kappa) \triangleq (t - \kappa)(r - \kappa), \quad (9.3)$$

as shown in Fig. 9.1. We see that the maximum values that μ and δ can achieve are m and tr , respectively, as discussed before. Eq. (9.3) also shows that the maximum diversity gain can only be achieved for zero rate gain, and the maximum rate gain can only be achieved for zero diversity gain. More generally, (9.3) shows that, out of the total number of t transmit and r receive antennas, κ transmit and κ receive antennas are allocated to increase the rate, and the remaining $t - \kappa$ and $r - \kappa$ create diversity. A concise derivation of this result and an intuitive explanation based on it are given in Section 9.6.

This diversity–rate tradeoff curve can be used to compare different schemes, and to interpret their behavior, as shown in the examples that follow. In particular, we shall see how orthogonal schemes are attractive when high diversity gain is sought, while BLAST interfaces favor rate gain.

¹See [121] for lower values of N . It suffices to observe here that no more diversity gain can be obtained if the block length of the code exceeds $t + r - 1$, which consequently expresses the infinite-block-length performance.

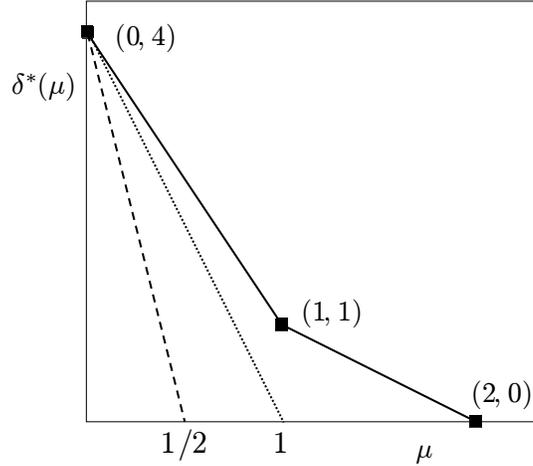


Fig. 9.2 Diversity–rate tradeoff for 2×2 systems. Continuous line: Optimal tradeoff. Dotted line: Alamouti scheme. Dashed line: Repetition scheme.

9.1 2×2 schemes

Consider two transmit and two receive antennas, and a block length chosen to comply with the condition of validity of (9.3), viz., $N \geq t + r - 1$. The maximum diversity gain is $tr = 4$, achieved if each transmitted signal passes through all the four propagation paths. The maximum rate gain is $t = r = 2$. The optimal tradeoff curve for this system is shown by the continuous line of Fig. 9.2.

A simple scheme that achieves maximum diversity is a repetition code:

$$\mathbf{X} = \begin{bmatrix} x_1 & 0 \\ 0 & x_1 \end{bmatrix}, \quad (9.4)$$

where x_1 is a signal from a suitable constellation (we may think of this scheme as an inner code concatenated with an outer code that generates x_1). Fig. 9.2 shows the tradeoff curve for this “repetition” system. Since it takes two channel uses to transmit one symbol, the maximum rate gain is $1/2$. When maximum diversity is achieved, the rate gain is 0. In fact, if a data rate $\mu \log \rho$ must be supported, the size of the constellation from which x_1 is drawn must increase, and consequently the minimum distance decreases, as does the achievable

diversity gain.

The Alamouti scheme can also be used on this channel. Here

$$\mathbf{X} = \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix}. \quad (9.5)$$

This achieves the full diversity gain. Two symbols are transmitted every 2 channel uses, hence the maximum rate gain is 1. Its tradeoff curve is shown in Fig. 9.2. Notice that, although both repetition and Alamouti scheme achieve the optimum diversity at $\mu = 0$, their behavior is markedly different when the diversity–rate tradeoff is taken into consideration.

9.2 Orthogonal designs

Consider first the special case $t = 2$, with \mathbf{X} given again by (9.5). The optimal tradeoff can be computed, and yields

$$\delta^*(\mu) = tr(1 - \mu)_+. \quad (9.6)$$

More generally, since orthogonal designs with full rate (that is, $\mu = 1$) do not exist for $t > 2$, one can observe that their maximum rate gain is strictly less than 1. Hence, although they achieve maximum diversity at $\mu = 0$, they are strictly suboptimum in terms of the diversity–rate tradeoff.

9.3 Zero-forcing vertical BLAST

Consider now zero-forcing vertical BLAST (ZF V-BLAST) with m transmit and receive antennas. Its performance, as discussed in Section 8.2.1, depends on the order of detection of the substreams and on the data rates of the substreams. For all versions of V-BLAST, the tradeoff curve is suboptimal, especially for low rate gains: in fact, every transmitted substream experiences only m independent fading gains, and, even with no spatial interference between substreams the tradeoff curve cannot exceed $\delta(\kappa) = m - \kappa$ substreams.

9.4 Zero-forcing diagonal BLAST

This system, which has coding over signals transmitted on different antennas, promises a higher diversity gain. Here, if the rate loss caused by the overhead symbols is disregarded, the tradeoff curve connects the points $(m - \kappa, \kappa(\kappa + 1)/2)$, $\kappa = 0, \dots, m$. Observe that the maximum diversity gain is now $m(m + 1)/2$, better than for V-BLAST but still short of the theoretical maximum m^2 . It is recognized [1,121] that this performance loss is caused by the zero-forcing step. If MMSE filtering is used instead of ZF, then D-BLAST achieves the optimum tradeoff curve (apart from the rate loss mentioned before). This behavior can be justified by observing that D-BLAST achieves the optimum mutual information of the MIMO channel for any realization of channel \mathbf{H} [82, Sec. 12.4.1].

9.5 MIMO transmission in a multiuser environment

Throughout this paper we have considered only point-to-point, single-user transmission. Suppose instead that K users, each equipped with t antennas, transmit data to an r -antenna receiver. If the latter wants to detect the signals transmitted by user 1 (say), multiple antennas can be used to discriminate among these signals. The optimum tradeoff can be determined even in this situation [108]. Assuming optimum multiuser detection [111], and a code block length $N > Kt + r - 1$, we have the following results:

- For $t \leq r/(K - 1)$, the tradeoff curve is the same as for single-user transmission: the presence of multiuser interference does not degrade the tradeoff curve.
- For $t > r/(K - 1)$, the tradeoff curve is the same as for single-user transmission up to $\kappa^* = r/(K - 1)$. For $\kappa^* < \kappa \leq \min\{r/K, t\}$, the tradeoff is as if the K users were pooled together into a single user transmitting with Kt antennas at a rate multiplied by K . In particular, $\delta^*(\min\{r/K, t\}) = 0$.

A simple result can also be obtained in the presence of a suboptimum detector. Assume $t = 1$ antenna per user, $K < r$, and a decorre-

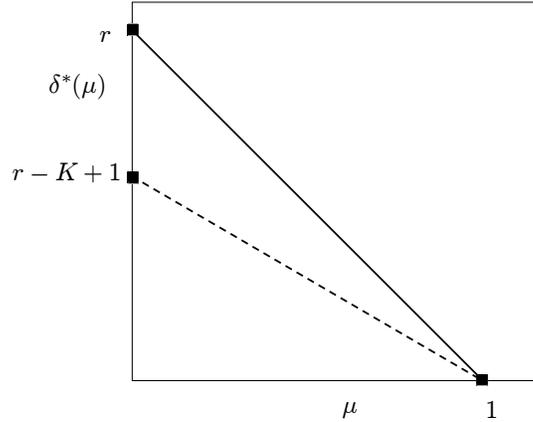


Fig. 9.3 Diversity–rate tradeoff for a multiuser scheme with 1 transmit antenna per user. Optimum detection (continuous line) and decorrelator detection (dashed line).

lating detector. In this case the optimum tradeoff curve is the real line connecting the point $(0, r - K + 1)$ to the point $(1, 0)$ (Fig. 9.3). This is consistent with the finding in [117] that it costs $K - 1$ diversity gain to null out $K - 1$ interferers.

9.6 Derivation of the diversity–rate tradeoff curve

Under the above assumptions, and in the asymptotic regime $\rho \rightarrow \infty$, the outage probability (see Section 4.5) corresponding to an information rate $R = \mu \log \rho$ can be written as

$$P_{\text{out}}(\mu, \rho) \triangleq \mathbb{P}\left(\sum_{i=1}^m \log(1 + \lambda_i \rho) \leq \mu \log \rho\right), \quad (9.7)$$

where λ_i is the i th ordered eigenvalue of the matrix $\mathbf{H}\mathbf{H}^\dagger$. If $\mu \geq m$ the outage probability is always 1 (since $\log \rho$ dominates asymptotically the other terms as $\rho \rightarrow \infty$), so that we restrict ourselves to the case $\mu < m$. The joint pdf of the λ_i 's is given by (C.5). Defining the new variables $\alpha_i \triangleq -\log \lambda_i / \log \rho$, we can write the outage probability (9.7) as follows:

$$P_{\text{out}}(\mu, \rho) = \frac{(\ln \rho)^m}{\Gamma_m(m)\Gamma_m(n)}. \quad (9.8)$$

$$\begin{aligned} & \cdot \int_{\boldsymbol{\alpha} \in \mathbb{R}^m, \alpha_1 \geq \dots \geq \alpha_m} \mathbf{1} \left\{ \sum_{i=1}^m \log(1 + \rho^{1-\alpha_i}) \leq \mu \log \rho \right\} \\ & \cdot \prod_{i=1}^m e^{-\rho^{-\alpha_i}} \rho^{-(n-m+1)\alpha_i} \prod_{i < j} (\rho^{-\alpha_i} - \rho^{-\alpha_j})^2 d\boldsymbol{\alpha}. \end{aligned}$$

Since $\rho \rightarrow \infty$, several simplifications can be used:

- The indicator function in the integral

$$\mathbf{1} \left\{ \sum_{i=1}^m \log(1 + \rho^{1-\alpha_i}) \leq \mu \log \rho \right\} \rightarrow \mathbf{1} \left\{ \sum_{i=1}^m (1 - \alpha_i)_+ \leq \mu \right\}. \quad (9.9)$$

- Since $\exp(-\rho^{-\alpha_i}) \rightarrow 0$ for $\alpha_i < 0$ and $\exp(-\rho^{-\alpha_i}) \rightarrow 1$ for $\alpha_i > 0$, the integration domain where the integrand is not asymptotically small reduces to \mathbb{R}_+^m .
- $(\rho^{-\alpha_i} - \rho^{-\alpha_j})^2 \rightarrow \rho^{-2\alpha_j}$, since $\alpha_i > \alpha_j$ for $i < j$ except for a set of measure zero.

Collecting the above observations we obtain, as $\rho \rightarrow \infty$,

$$\begin{aligned} P_{\text{out}}(\mu, \rho) & \rightarrow \frac{(\ln \rho)^m}{\Gamma_m(m)\Gamma_m(n)} \int_{\alpha_1 \geq \dots \geq \alpha_m \geq 0} \mathbf{1} \left\{ \sum_{i=1}^m (1 - \alpha_i)_+ \leq \mu \right\} \\ & \cdot \exp \left(-\ln \rho \sum_{i=1}^m (n - m + 2i - 1)\alpha_i \right) d\boldsymbol{\alpha}. \quad (9.10) \end{aligned}$$

Using Laplace's method of asymptotic multidimensional integral approximation [15], it can be shown that

$$P_{\text{out}}(\mu, \rho) \rightarrow \rho^{-d_{\text{out}}(\mu)}, \quad (9.11)$$

where

$$d_{\text{out}}(\mu) \triangleq \min_{\alpha_1 \geq \dots \geq \alpha_m \geq 0, \sum_{i=1}^m (1-\alpha_i)_+ \leq \mu} \sum_{i=1}^m (n - m + 2i - 1)\alpha_i. \quad (9.12)$$

This is a linear programming problem with nonlinear constraint, equivalent to the maximization problem

$$\bar{\boldsymbol{\alpha}}^* \triangleq \arg \max_{\bar{\alpha}_1 \leq \dots \leq \bar{\alpha}_m \leq 1, \sum_i (\bar{\alpha}_i)_+ \leq \mu} \sum_{i=1}^m (n - m + 2i - 1)\bar{\alpha}_i \quad (9.13)$$

Although such problems are very complex in general, this one admits a very simple solution. In order to maximize the objective function one sets $\bar{\alpha}_i = 1$ for $i = m, m - 1, \dots, m - \lfloor \mu \rfloor + 1$. Next, one has to set $\alpha_{m-\lfloor \mu \rfloor} = \mu - \lfloor \mu \rfloor < 1$. This corresponds to setting

$$\alpha_i = \begin{cases} 1 & i = 1, \dots, m - \lfloor \mu \rfloor - 1 \\ 1 - (\mu - \lfloor \mu \rfloor) & i = m - \lfloor \mu \rfloor \\ 0 & i = m - \lfloor \mu \rfloor + 1 \dots, m. \end{cases} \quad (9.14)$$

As a result we can write the following expression for the diversity:

$$d_{\text{out}} = (n - \lfloor \mu \rfloor)(m - \lfloor \mu \rfloor) - (\mu - \lfloor \mu \rfloor)(n + m - 1 - 2\lfloor \mu \rfloor). \quad (9.15)$$

Notice that, for integer μ , we have

$$d_{\text{out}} = (n - \mu)(m - \mu). \quad (9.16)$$

In this case, $\bar{\alpha}_i$ represents an indicator of the usage of the i th equivalent channel: $\bar{\alpha}_i = 0$ means that the i th channel is not used, and vice versa for $\bar{\alpha}_i = 1$. In fact, if $\bar{\alpha}_i = 0$ and hence $\alpha_i = 1$, the i th eigenvalue $\lambda_i = \rho^{-\alpha_i} \rightarrow 0$ as $\rho \rightarrow \infty$. That implies a rate loss due to the inability of using the i th channel. Meanwhile, the diversity $d_{\text{out}}(\mu)$ is increased by $(n - m + 2i - 1)$ units as shown by eq. (9.10).

Full text available at: <http://dx.doi.org/10.1561/0100000002>

APPENDICES

A

Complex random variables and vectors

A complex random variable (RV) has the form

$$Z = X + jY, \quad (\text{A.1})$$

where X and Y are real RVs; Z can also be viewed as a two-dimensional random vector RV (see below). Definition and basic properties of the mean value and the variance of Z are given as follows:

$$\mathbb{E}[Z] \triangleq \mathbb{E}[X] + j\mathbb{E}[Y] \quad (\text{A.2})$$

$$\begin{aligned} \mathbb{V}[Z] &\triangleq \mathbb{E}[|Z - \mathbb{E}[Z]|^2] && (\text{A.3}) \\ &= \mathbb{E}[(X - \mathbb{E}[X])^2 + (Y - \mathbb{E}[Y])^2] \\ &= \mathbb{V}[X] + \mathbb{V}[Y]. \end{aligned}$$

The *cross-covariance* of two complex RVs Z and W is defined as

$$R_{ZW} \triangleq \mathbb{E}[(Z - \mathbb{E}[Z])(W - \mathbb{E}[W])^*]. \quad (\text{A.4})$$

As a special case of this definition, we have $R_{ZZ} = \mathbb{V}[Z]$. Cross-covariance is Hermitian, viz., $R_{WZ} = R_{ZW}^*$.

A.1 Gaussian random variables

If X is real Gaussian with mean μ and variance σ^2 , i.e., its pdf is

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \quad (\text{A.5})$$

then we write $X \sim \mathcal{N}(\mu, \sigma^2)$, a notation stressing the fact that the pdf of X is completely specified by μ and σ^2 . If the real and imaginary part of the complex RV Z are independent with the same variance $\sigma^2/2$, and $\mu \triangleq \mathbb{E}(Z) \in \mathbb{C}$, then we say that Z is *circularly-symmetric*, and we write $Z \sim \mathcal{N}_c(\mu, \sigma^2)$. Its pdf is the product of those of its real and imaginary part:

$$p(z) = \frac{1}{\pi\sigma^2} e^{-|z-\mu|^2/\sigma^2}. \quad (\text{A.6})$$

A.2 Real random vectors

A real random vector is a column vector $\mathbf{x} = (X_1, X_2, \dots, X_n)'$ whose components are real random variables. Its mean value is defined as

$$\boldsymbol{\mu}_{\mathbf{x}} \triangleq (\mathbb{E}[X_1], \mathbb{E}[X_2], \dots, \mathbb{E}[X_n])'. \quad (\text{A.7})$$

The expectation of the squared norm of \mathbf{x} :

$$\mathbb{E}[\|\mathbf{x}\|^2] = \mathbb{E}[\mathbf{x}'\mathbf{x}] = \sum_{i=1}^n \mathbb{E}[X_i^2] \quad (\text{A.8})$$

is often referred to as the *energy* of \mathbf{x} . The *covariance matrix* of \mathbf{x} is defined as the nonnegative-definite $n \times n$ matrix

$$\mathbf{R}_{\mathbf{x}} \triangleq \mathbb{E}[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})'] = \mathbb{E}[\mathbf{x}\mathbf{x}'] - \boldsymbol{\mu}_{\mathbf{x}}\boldsymbol{\mu}_{\mathbf{x}}'. \quad (\text{A.9})$$

Notice that the diagonal entries of $\mathbf{R}_{\mathbf{x}}$ are the variances of the components of \mathbf{x} . The $n \times n$ matrix

$$\mathbf{C}_{\mathbf{x}} \triangleq \mathbb{E}[\mathbf{x}\mathbf{x}'] \quad (\text{A.10})$$

is called the *correlation matrix* of \mathbf{x} . We observe that the trace of $\mathbf{C}_{\mathbf{x}}$ equals the average energy of \mathbf{x} :

$$\mathbb{E}[\|\mathbf{x}\|^2] = \mathbb{E}[\mathbf{x}'\mathbf{x}] = \mathbb{E}[\text{Tr}(\mathbf{x}\mathbf{x}')] = \text{Tr} \mathbf{C}_{\mathbf{x}}. \quad (\text{A.11})$$

The *cross-covariance matrix* of the two random vectors \mathbf{x} and \mathbf{y} , with dimensions n and m , respectively, is defined as the $n \times m$ matrix

$$\mathbf{R}_{\mathbf{x},\mathbf{y}} \triangleq \mathbb{E}[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})'] = \mathbb{E}[\mathbf{x}\mathbf{y}'] - \boldsymbol{\mu}_{\mathbf{x}}\boldsymbol{\mu}_{\mathbf{y}}'. \quad (\text{A.12})$$

A.2.1 Real Gaussian random vectors

A real random vector $\mathbf{x} = (X_1, \dots, X_n)'$ is called *Gaussian* if its components are jointly Gaussian, that is, their joint probability density function (pdf) is

$$\begin{aligned} p(\mathbf{x}) &= \quad \quad \quad (\text{A.13}) \\ &= \det(2\pi\mathbf{R}_{\mathbf{x}})^{-n/2} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})'\mathbf{R}_{\mathbf{x}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \right\} \\ &= \det(2\pi\mathbf{R}_{\mathbf{x}})^{-n/2} \exp \left\{ -\frac{1}{2}\text{Tr} [\mathbf{R}_{\mathbf{x}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})'] \right\}. \end{aligned}$$

We write $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{R}_{\mathbf{x}})$, which stresses the fact that the pdf of a real Gaussian random vector is completely specified by its mean value and its covariance matrix.

A.3 Complex random vectors

A complex random vector is a column vector $\mathbf{z} = (Z_1, Z_2, \dots, Z_n)'$ whose components are complex random variables. The *covariance matrix* of \mathbf{z} is defined as the nonnegative-definite $n \times n$ matrix

$$\mathbf{R}_{\mathbf{z}} \triangleq \mathbb{E}[(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})^\dagger] = \mathbb{E}[\mathbf{z}\mathbf{z}^\dagger] - \boldsymbol{\mu}_{\mathbf{z}}\boldsymbol{\mu}_{\mathbf{z}}^\dagger. \quad (\text{A.14})$$

The diagonal entries of $\mathbf{R}_{\mathbf{z}}$ are the variances of the entries of \mathbf{z} . If $\mathbf{z} = \mathbf{x} + j\mathbf{y}$,

$$\mathbf{R}_{\mathbf{z}} = (\mathbf{R}_{\mathbf{x}} + \mathbf{R}_{\mathbf{y}}) + j(\mathbf{R}_{\mathbf{y}\mathbf{x}} - \mathbf{R}_{\mathbf{x}\mathbf{y}}). \quad (\text{A.15})$$

Thus, knowledge of $\mathbf{R}_{\mathbf{z}}$ does not yield knowledge of $\mathbf{R}_{\mathbf{x}}$, $\mathbf{R}_{\mathbf{y}}$, $\mathbf{R}_{\mathbf{y}\mathbf{x}}$, and $\mathbf{R}_{\mathbf{x}\mathbf{y}}$, i.e., of the complete second-order statistics of \mathbf{z} . The latter is completely specified if, in addition to $\mathbf{R}_{\mathbf{z}}$, the *pseudo-covariance matrix*

$$\begin{aligned} \tilde{\mathbf{R}}_{\mathbf{z}} &\triangleq \mathbb{E}[(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})'] \\ &= (\mathbf{R}_{\mathbf{x}} - \mathbf{R}_{\mathbf{y}}) + j(\mathbf{R}_{\mathbf{y}\mathbf{x}} + \mathbf{R}_{\mathbf{x}\mathbf{y}}) \end{aligned} \quad (\text{A.16})$$

is also known [79]. We have the following relations:

$$\begin{cases} \mathbf{R}_x &= \frac{1}{2}\text{Re}(\mathbf{R}_z + \tilde{\mathbf{R}}_z) \\ \mathbf{R}_y &= \frac{1}{2}\text{Re}(\mathbf{R}_z - \tilde{\mathbf{R}}_z) \\ \mathbf{R}_{xy} &= \frac{1}{2}\text{Im}(\mathbf{R}_z - \tilde{\mathbf{R}}_z) \\ \mathbf{R}_{yx} &= \frac{1}{2}\text{Im}(\mathbf{R}_z + \tilde{\mathbf{R}}_z). \end{cases} \quad (\text{A.17})$$

Proper complex random vectors endowed with the additional property

$$\tilde{\mathbf{R}}_z = 0 \quad (\text{A.18})$$

(see [79] for a justification of the term) are completely specified by \mathbf{R}_z as far as their second-order statistics are concerned.

Similar properties can be derived for the $n \times n$ matrix $\mathbf{C}_z \triangleq \mathbb{E}[\mathbf{z}\mathbf{z}^\dagger]$, which is called the *correlation matrix* of \mathbf{z} .

A.3.1 Complex Gaussian random vectors

A complex random vector $\mathbf{z} = \mathbf{x} + j\mathbf{y} \in \mathbb{C}^n$ is called *Gaussian* if its real part \mathbf{x} and imaginary part \mathbf{y} are jointly Gaussian, or, equivalently, if the real random vector

$$\check{\mathbf{z}} \triangleq \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \in \mathbb{R}^{2n}$$

is Gaussian.

Unlike Gaussian real random vectors, their complex counterparts are *not* completely specified by their mean values and covariance matrices (the pseudo-covariance matrices are also needed). In fact, to specify the pdf of \mathbf{z} , and hence of $\check{\mathbf{z}}$, we need, in addition to $\mathbb{E}[\mathbf{z}]$, the covariance matrix

$$\mathbf{R}_{\check{\mathbf{z}}} = \begin{bmatrix} \mathbf{R}_x & \mathbf{R}_{xy} \\ \mathbf{R}'_{xy} & \mathbf{R}_y \end{bmatrix} \quad (\text{A.19})$$

which is completely specified by \mathbf{R}_z and $\tilde{\mathbf{R}}_z$. In order to be able to uniquely determine \mathbf{R}_x , \mathbf{R}_y , and \mathbf{R}_{xy} from \mathbf{R}_z , we need to restrict our attention to the subclass of *proper* Gaussian random vectors, also called *circularly-symmetric*. The covariance matrix of $\check{\mathbf{z}}$ can be written as follows:

$$\mathbf{R}_{\check{\mathbf{z}}} = \frac{1}{2} \begin{bmatrix} \text{Re } \mathbf{R}_z & -\text{Im } \mathbf{R}_z \\ \text{Im } \mathbf{R}_z & \text{Re } \mathbf{R}_z \end{bmatrix}. \quad (\text{A.20})$$

Hence, a circularly-symmetric complex Gaussian random vector is characterized by $\boldsymbol{\mu}_{\mathbf{z}}$ and $\mathbf{R}_{\mathbf{z}}$. We write¹ $\mathbf{z} \sim \mathcal{N}_c(\boldsymbol{\mu}_{\mathbf{z}}, \mathbf{R}_{\mathbf{z}})$. The probability density function of \mathbf{z} is given by

$$\begin{aligned} p(\mathbf{z}) &\triangleq p(\check{\mathbf{z}}) \\ &= \det(2\pi\mathbf{R}_{\check{\mathbf{z}}})^{-1/2} \exp \left\{ -(\check{\mathbf{z}} - \boldsymbol{\mu}_{\check{\mathbf{z}}})^\dagger \mathbf{R}_{\check{\mathbf{z}}}^{-1} (\check{\mathbf{z}} - \boldsymbol{\mu}_{\check{\mathbf{z}}}) \right\} \\ &= \det(\pi\mathbf{R}_{\mathbf{z}})^{-1} \exp \left\{ -(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})^\dagger \mathbf{R}_{\mathbf{z}}^{-1} (\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}}) \right\}. \quad (\text{A.21}) \end{aligned}$$

The following theorem describes important properties of circularly-symmetric Gaussian random vectors.

Theorem A.1 ([79]). If $\mathbf{z} \sim \mathcal{N}_c(\boldsymbol{\mu}_{\mathbf{z}}, \mathbf{R}_{\mathbf{z}})$, then every affine transformation

$$\mathbf{y} = \mathbf{A}\mathbf{z} + \mathbf{b}$$

yields a circularly-symmetric complex RV $\mathbf{y} \sim \mathcal{N}_c(\mathbf{A}\boldsymbol{\mu}_{\mathbf{z}} + \mathbf{b}, \mathbf{A}\mathbf{R}_{\mathbf{z}}\mathbf{A}^\dagger)$.

¹This notation is meant to avoid confusion with the real case, and to remind us that in our context circular symmetry is a property of *Gaussian* complex random vectors.

Full text available at: <http://dx.doi.org/10.1561/0100000002>

B

Results from information theory

B.1 A basic inequality

We state and prove a basic entropy inequality which allows the derivation of the capacity of the Gaussian MIMO channel.

Theorem B.1. Let \mathbf{z} be a complex random vector with covariance matrix $\mathbf{R}_{\mathbf{z}}$. Then, the differential entropy of \mathbf{z} satisfies the inequality

$$h(\mathbf{z}) \leq \log \det(\pi e \mathbf{R}_{\mathbf{z}}) \quad (\text{B.1})$$

with equality if and only if $\mathbf{z} \sim \mathcal{N}_c(\mathbf{a}, \mathbf{R}_{\mathbf{z}})$ for any vector \mathbf{a} having the same dimension as \mathbf{z} .

Proof. Let $\mathbf{z}_G \sim \mathcal{N}_c(\mathbf{0}, \mathbf{R}_{\mathbf{z}})$ and calculate its differential entropy by using the pdf (A.21):

$$\begin{aligned} h(\mathbf{z}_G) &= \mathbb{E}[\log \det(\pi \mathbf{R}_{\mathbf{z}_G}) + \mathbf{z}_G^\dagger \mathbf{R}_{\mathbf{z}_G}^{-1} \mathbf{z}_G \log e] \\ &= \log \det(\pi \mathbf{R}_{\mathbf{z}_G}) + \mathbb{E}[\mathbf{z}_G^\dagger \mathbf{R}_{\mathbf{z}_G}^{-1} \mathbf{z}_G] \log e \\ &= \log \det(\pi \mathbf{R}_{\mathbf{z}_G}) + \text{Tr}(\mathbf{R}_{\mathbf{z}_G}^{-1} \mathbb{E}[\mathbf{z}_G \mathbf{z}_G^\dagger]) \log e \\ &= \log \det(\pi e \mathbf{R}_{\mathbf{z}_G}). \end{aligned} \quad (\text{B.2})$$

Let $p(\mathbf{z})$ and $q(\mathbf{z})$ be the pdf's of \mathbf{z} and \mathbf{z}_G , respectively. The theorem follows by

$$\begin{aligned}
 h(\mathbf{z}) - h(\mathbf{z}_G) &= \int (q(\mathbf{z}) \log q(\mathbf{z}) - p(\mathbf{z}) \log p(\mathbf{z})) d\mathbf{z} \\
 &= \int p(\mathbf{z}) (\log q(\mathbf{z}) - \log p(\mathbf{z})) d\mathbf{z} \\
 &= \int p(\mathbf{z}) \log \frac{q(\mathbf{z})}{p(\mathbf{z})} d\mathbf{z} \\
 &\leq \int p(\mathbf{z}) \left(\frac{q(\mathbf{z})}{p(\mathbf{z})} - 1 \right) d\mathbf{z} \\
 &= 0,
 \end{aligned} \tag{B.3}$$

where we used the equality $\mathbb{E}[\log q(\mathbf{z})] = \mathbb{E}[\log q(\mathbf{z}_G)]$ and $\ln u \leq u - 1$ for $u > 0$ with equality if and only if $u = 1$.¹ Notice that equality holds if and only if $p(\mathbf{z}) = q(\mathbf{z})$, i.e., if and only if \mathbf{z} is circularly-symmetric Gaussian.

B.2 Capacity of the Gaussian MIMO channel

Let the channel input–output relationship be

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \tag{B.4}$$

where \mathbf{H} is a constant $r \times t$ matrix, \mathbf{x} is a t -vector, and \mathbf{y} and \mathbf{z} are r -vectors. Assume $\mathbf{x} \perp \mathbf{z}$ and $\mathbf{z} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{R}_z)$. From the mutual information

$$I(\mathbf{x}; \mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y} | \mathbf{x}) = h(\mathbf{y}) - h(\mathbf{z}), \tag{B.5}$$

we seek the channel capacity under the constraint

$$\text{Tr}(\mathbf{R}_x) \leq P. \tag{B.6}$$

For a given \mathbf{R}_x , the covariance matrix of \mathbf{y} is $\mathbf{R}_y = \mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger + \mathbf{R}_z$ and $h(\mathbf{y})$ is maximum for $\mathbf{y} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{R}_y)$ (Theorem B.1). Moreover, the maximum mutual information is given by

$$I(\mathbf{R}_x) = \log \det(\mathbf{I}_r + \mathbf{R}_x\mathbf{H}^\dagger\mathbf{R}_z^{-1}\mathbf{H}). \tag{B.7}$$

¹ Consider the function $f(u) \triangleq \ln u - u + 1$. Its derivative $f'(u) = u^{-1} - 1$ is positive for $0 < u < 1$ and negative for $u > 1$. Hence, it has a maximum in $u = 1$, i.e., $f(u) \leq f(1) = 0$, which proves the statement.

The channel capacity can be calculated according to different assumptions:

- (a) The receiver has perfect CSI and the transmitter has no CSI.
- (b) The receiver and the transmitter have perfect CSI.

With assumption (a), the transmitter divides the available power uniformly among the transmit antennas and the capacity is then

$$C_{\text{rx}} = \log \det \left(\mathbf{I}_r + \frac{P}{t} \mathbf{H}^\dagger \mathbf{R}_z^{-1} \mathbf{H} \right). \quad (\text{B.8})$$

With assumption (b), the capacity can be written as

$$C_{\text{tx/rx}} = \max_{\mathbf{R}_x \geq \mathbf{0}, \text{Tr } \mathbf{R}_x \leq P} \log \det(\mathbf{I}_r + \mathbf{R}_x \mathbf{H}^\dagger \mathbf{R}_z^{-1} \mathbf{H}). \quad (\text{B.9})$$

From Hadamard's inequality² [32, Th. 16.8.2] and the orthogonal diagonalization $\mathbf{H}^\dagger \mathbf{R}_z^{-1} \mathbf{H} = \mathbf{U} \mathbf{D} \mathbf{U}^\dagger$ (where \mathbf{U} is unitary and \mathbf{D} diagonal) we have

$$\begin{aligned} \log \det(\mathbf{I}_r + \mathbf{R}_x \mathbf{H}^\dagger \mathbf{R}_z^{-1} \mathbf{H}) &= \log \det(\mathbf{I}_r + \tilde{\mathbf{R}} \mathbf{D}) \\ &\leq \sum_{(\mathbf{D})_{i,i} > 0} \log(1 + (\tilde{\mathbf{R}})_{i,i} (\mathbf{D})_{i,i}), \end{aligned} \quad (\text{B.10})$$

where $\tilde{\mathbf{R}} \triangleq \mathbf{U}^\dagger \mathbf{R}_x \mathbf{U}$, with equality if and only if $\tilde{\mathbf{R}}$ is diagonal. Since the constraint (B.6) translates into $\text{Tr } (\tilde{\mathbf{R}}) \leq P$, the maximization problem admits the “water-filling” solution [32]:

$$(\tilde{\mathbf{R}})_{i,i} = \left(\mu - (\mathbf{D})_{i,i}^{-1} \right)_+ \quad (\text{B.11})$$

with μ obtained as the solution of

$$\sum_{(\mathbf{D})_{i,i} > 0} \left(\mu - (\mathbf{D})_{i,i}^{-1} \right)_+ = P. \quad (\text{B.12})$$

The channel input pdf achieving capacity is

$$\mathcal{N}_c(\mathbf{0}, \mathbf{U} \text{diag}((\mu - (\mathbf{D})_{i,i}^{-1})_+) \mathbf{U}^\dagger).$$

² For every nonnegative definite matrix \mathbf{K} , $\det \mathbf{K} \leq \prod_i (\mathbf{K})_{ii}$ with equality iff \mathbf{K} is diagonal.

B.3 Ergodic capacity

Consider the separately correlated fading channel described in Section 3.4 with channel matrix given by (3.25), additive white Gaussian noise $\mathbf{z} \sim \mathcal{N}_0(\mathbf{0}, N_0\mathbf{I}_r)$, and fast, frequency-nonselctive fading (see Section 3.3). Assume that the receiver has perfect CSI but the transmitter has no CSI. Here we calculate the average capacity under the power constraint $\mathbb{E}[\|\mathbf{x}\|^2] \leq P = \rho N_0$ in the case of $\mathbf{T} = \mathbf{I}_t$. This derivation follows the guidelines of [106] for the case of $\mathbf{R} = \mathbf{I}_r$ and [86] in the more general setting of $\mathbf{R} \neq \mathbf{I}_r$, although we restrict ourselves to consideration of iid, $\mathcal{N}_c(0, 1)$ entries of \mathbf{H}_u .

From the expression of the mutual information (B.5), capacity is given by

$$C = \max_{\mathbf{R}_x \geq \mathbf{0}, \text{Tr } \mathbf{R}_x \leq P} \mathbb{E}[\log \det(\mathbf{I}_r + \mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger/N_0)]. \quad (\text{B.13})$$

Using the orthogonal diagonalization $\mathbf{R}_x = \mathbf{U}\mathbf{D}\mathbf{U}^\dagger$ (with matrix \mathbf{U} unitary and \mathbf{D} diagonal) we notice that capacity can also be obtained as follows

$$C = \max_{\mathbf{D} \geq \mathbf{0}, \text{Tr } \mathbf{D} \leq \rho} \mathbb{E}[\log \det(\mathbf{I}_r + \mathbf{H}\mathbf{D}\mathbf{H}^\dagger)], \quad (\text{B.14})$$

where the maximum is sought over the set of diagonal matrices with nonnegative diagonal entries and trace upper bounded by ρ . The equivalence derives from the fact that \mathbf{H}_u and $\mathbf{H}_u\mathbf{U}$ (and hence \mathbf{H} and $\mathbf{H}\mathbf{U}$) have the same joint pdf. Let us write

$$\Psi(\mathbf{D}) \triangleq \mathbb{E}[\log \det(\mathbf{I}_r + \mathbf{H}\mathbf{D}\mathbf{H}^\dagger)]. \quad (\text{B.15})$$

Since the log-det function is concave [32, Th. 16.8.1], we have, for every vector (α_i) such that $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$,

$$\begin{aligned} \Psi\left(\sum_i \alpha_i \mathbf{D}_i\right) &= \mathbb{E}\left[\log \det\left(\sum_i \alpha_i (\mathbf{I}_r + \mathbf{H}\mathbf{D}_i\mathbf{H}^\dagger)\right)\right] \\ &\geq \sum_i \alpha_i \mathbb{E}[\log \det(\mathbf{I}_r + \mathbf{H}\mathbf{D}_i\mathbf{H}^\dagger)] \\ &= \sum_i \alpha_i \Psi(\mathbf{D}_i). \end{aligned} \quad (\text{B.16})$$

Now, let P_i denote the i th permutation matrix ($i = 1, \dots, t!$). For a given matrix \mathbf{D} such that $\text{Tr } \mathbf{D} = \rho$, define $\mathbf{D}_i \triangleq \mathbf{P}_i \mathbf{D} \mathbf{P}_i'$, i.e., the diagonal matrix obtained by applying the i th permutation on the diagonal elements. We notice that

(1) $\Psi(\mathbf{D}_i) = \Psi(\mathbf{D})$ since

$$\mathbb{E}[\log \det(\mathbf{I}_r + \mathbf{H} \mathbf{P}_i \mathbf{D} \mathbf{P}_i' \mathbf{H}^\dagger)] = \mathbb{E}[\log \det(\mathbf{I}_r + \mathbf{H} \mathbf{D} \mathbf{H}^\dagger)],$$

as \mathbf{H} and $\mathbf{H} \mathbf{P}_i$ have the same joint pdf.³

(2) $\sum_i \mathbf{D}_i / t! = (\rho/t) \mathbf{I}_t$ since every diagonal entry of \mathbf{D} appears the same number of times at each position of the matrix sum.

Hence, we have

$$\Psi((\rho/t) \mathbf{I}_t) \geq \frac{1}{t!} \sum_i \Psi(\mathbf{D}_i) = \Psi(\mathbf{D}), \quad (\text{B.17})$$

which proves that the maximum $\Psi(\mathbf{D})$, i.e. capacity, is attained for $\mathbf{D} = (\rho/t) \mathbf{I}_t$, i.e., uniform power allocation.

³ This point is not stated explicitly in [106]. It is implicit in the fact that \mathbf{H} is unitarily invariant in distribution.

Full text available at: <http://dx.doi.org/10.1561/0100000002>

C

Random matrices

A random matrix is a matrix whose entries are random variables. Consequently, a random matrix is described by assigning the joint probability density function (pdf) of its entries, which is especially easy when these are independent. For example, an $m \times n$ matrix \mathbf{A} , $m \leq n$, whose entries are independent identically distributed $\mathcal{N}(0, 1)$ real RVs, has pdf [36]

$$(2\pi)^{-mn/2} \text{etr}(-\mathbf{A}\mathbf{A}'/2). \quad (\text{C.1})$$

An $m \times n$ matrix \mathbf{B} , $m \leq n$, with iid complex Gaussian $\mathcal{N}_c(0, 1)$ entries has pdf [36]

$$\pi^{-mn} \text{etr}(-\mathbf{B}\mathbf{B}^\dagger). \quad (\text{C.2})$$

We have the following theorem [65].

Theorem C.1. If \mathbf{A} is a given $m \times m$ Hermitian matrix such that $\mathbf{I}_m + \mathbf{A} > \mathbf{0}$ and \mathbf{B} is an $m \times n$ matrix whose entries are iid as $\mathcal{N}_c(0, 1)$, then

$$\mathbb{E}[\text{etr}(-\mathbf{A}\mathbf{B}\mathbf{B}^\dagger)] = \det(\mathbf{I}_m + \mathbf{A})^{-n}. \quad (\text{C.3})$$

Proof. Splitting \mathbf{B} as $(\mathbf{b}_1, \dots, \mathbf{b}_n)$ we have

$$\begin{aligned} \mathbb{E}[\text{etr}(-\mathbf{A}\mathbf{B}\mathbf{B}^\dagger)] &= \pi^{-mn} \int \text{etr}(-(\mathbf{I}_m + \mathbf{A})\mathbf{B}\mathbf{B}^\dagger) d\mathbf{B} \\ &= \left\{ \pi^{-m} \int \text{etr}(-(\mathbf{I}_m + \mathbf{A})\mathbf{b}\mathbf{b}^\dagger) d\mathbf{b} \right\}^n \\ &= \det(\mathbf{I}_m + \mathbf{A})^{-n} \end{aligned} \quad (\text{C.4})$$

since the Jacobian of the transformation $\mathbf{b} = (\mathbf{I}_m + \mathbf{A})^{-1/2} \mathbf{c}$ is $\det(\mathbf{I}_m + \mathbf{A})^{-1}$. \square

The eigenvalues of a random matrix are random variables, and hence even these can be described by their joint probability density function. An important case occurs with the *complex Wishart* matrix, that is, a random complex Hermitian square $m \times m$ matrix $\mathbf{W} \triangleq \mathbf{B}\mathbf{B}^\dagger$, with \mathbf{B} as in (C.2). The pdf of the ordered eigenvalues $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$, $0 \leq \lambda_1 \leq \dots \leq \lambda_m$, of \mathbf{W} is given by [36]

$$p_{\text{ord}}(\boldsymbol{\lambda}) = \frac{1}{\Gamma_m(m)\Gamma_m(n)} \prod_{i=1}^m e^{-\lambda_i} \lambda_i^{n-m} \prod_{i < j} (\lambda_i - \lambda_j)^2, \quad (\text{C.5})$$

where $\Gamma_m(a) \triangleq \prod_{i=0}^{m-1} \Gamma(a - i)$. The joint pdf of the *unordered* eigenvalues is obtained from (C.5) by dividing it by $m!$.

It is interesting to observe the limiting distribution of the eigenvalues of a Wishart matrix as its dimensions grow to infinity. To do this, we define the *empirical distribution* of the eigenvalues of an $n \times n$ random matrix \mathbf{A} as the function $F(\lambda)$ which yields the fraction of eigenvalues of \mathbf{A} not exceeding λ , i.e.,

$$F(\lambda) \triangleq \frac{1}{n} |\{\lambda_i(\mathbf{A}) : \lambda_i(\mathbf{A}) < \lambda\}|. \quad (\text{C.6})$$

The empirical distribution is generally a random process. However, under certain mild technical conditions [97], as $n \rightarrow \infty$ the empirical distribution converges to a nonrandom cumulative distribution function. For a Wishart matrix we have the following theorem, a classic in random-matrix theory [4, 36, 66]:

Theorem C.2. Consider the sequence of $n \times m$ matrices \mathbf{A}_n , with iid entries having variances $1/n$; moreover, let $m = m(n)$, with $\lim_{n \rightarrow \infty} m(n)/n = c > 0$ and finite. Next, let $\mathbf{B}_n \triangleq \mathbf{A}_n \mathbf{A}_n^\dagger$. As $n \rightarrow \infty$, the empirical eigenvalue distribution of \mathbf{B}_n tends to a function $F(\lambda)$ whose derivative is

$$(1 - c)_+ \delta(\lambda) + \frac{1}{2\pi\lambda} \sqrt{(\lambda - \lambda_-)_+ (\lambda_+ - \lambda)_+} \quad (\text{C.7})$$

with $\lambda_\pm \triangleq (\sqrt{c} \pm 1)^2$.

The theorem that follows describes an important asymptotic property of a class of matrices. This is a special case of a general theory described in [63] (see also [109]).

Theorem C.3 ([12]). Let $(\mathbf{H}_n(s))_{s \in \mathcal{S}}$ be an independent family of $n \times n$ matrices whose entries are iid complex Gaussian random variables with independent, equally-distributed real and imaginary parts. Let $\mathbf{A}_n(s) \triangleq f(\mathbf{H}_n(s)^\dagger \mathbf{H}_n(s))$, where f is a real continuous function on \mathbb{R} . Let $(\mathbf{B}_n(t))_{t \in \mathcal{T}}$ be a family of deterministic matrices with eigenvalues $\lambda_1(n, t), \dots, \lambda_n(n, t)$ such that for all $t \in \mathcal{T}$

$$\sup_n \max_i \lambda_i(n, t) < \infty$$

and $(\mathbf{B}_n(t), \mathbf{B}_n^\dagger(t))_{t \in \mathcal{T}}$ has a limit distribution. Then, $\mathbf{A}_n(s)$ converges in distribution almost surely to a compactly supported probability measure on \mathbb{R} for each $s \in \mathcal{S}$ and, almost surely as $n \rightarrow \infty$,

$$\frac{1}{n} \text{Tr}(\mathbf{A}_n \mathbf{B}_n) \rightarrow \frac{1}{n} \mathbb{E}[\text{Tr}(\mathbf{A}_n)] \cdot \frac{1}{n} \mathbb{E}[\text{Tr}(\mathbf{B}_n)]. \quad (\text{C.8})$$

Full text available at: <http://dx.doi.org/10.1561/0100000002>

D

Numerical calculation of error probabilities

Consider the evaluation of the probability $P \triangleq \mathbb{P}(\nu > x)$, where ν and x are independent random variables whose moment-generating functions

$$\Phi_\nu(s) \triangleq \mathbb{E}[\exp(-s\nu)] \quad \text{and} \quad \Phi_x(s) \triangleq \mathbb{E}[\exp(-sx)]$$

are known. Defining $\Delta \triangleq x - \nu$, we have $P = \mathbb{P}(\Delta < 0)$. We describe a method for computing the value of P based on numerical integration. Assume that the moment-generating function of Δ , which due to the independence of ν and x can be written as

$$\Phi_\Delta(s) \triangleq \mathbb{E}[\exp(-s\Delta)] = \Phi_x(s) \Phi_\nu(-s), \quad (\text{D.1})$$

is analytically known. Using the Laplace inversion formula we obtain

$$\mathbb{P}(\Delta < 0) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{\Phi_\Delta(s)}{s} ds, \quad (\text{D.2})$$

where we assume that c is in the region of convergence (ROC) of $\Phi_\Delta(s)$. This is given by the intersection of the ROC of $\Phi_x(s)$ and the ROC of $\Phi_\nu(-s)$. Integral (D.2) can be computed exactly by using the method of residues [27, 110]. This method works well when the integrand exhibits simple poles, but becomes long and intricate when multiple poles or

essential singularities are present. Here we describe a general approach based on numerical calculation of the integral.

Now, expand the real and imaginary parts in (D.2). We have the following result:

$$\begin{aligned} \mathbb{P}(\Delta < 0) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\Phi_{\Delta}(c + j\omega)}{c + j\omega} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{c\operatorname{Re}\{\Phi_{\Delta}(c + j\omega)\} + \omega\operatorname{Im}\{\Phi_{\Delta}(c + j\omega)\}}{c^2 + \omega^2} d\omega. \end{aligned}$$

The change of variables $\omega = c\sqrt{1-x^2}/x$ yields

$$\begin{aligned} &\frac{1}{2\pi} \int_{-1}^1 \left\{ \operatorname{Re} \left[\Phi_{\Delta} \left(c + jc \frac{\sqrt{1-x^2}}{x} \right) \right] \right. \\ &\quad \left. + \frac{\sqrt{1-x^2}}{x} \operatorname{Im} \left[\Phi_{\Delta} \left(c + jc \frac{\sqrt{1-x^2}}{x} \right) \right] \right\} \frac{dx}{\sqrt{1-x^2}} \quad (\text{D.3}) \end{aligned}$$

and this integral can be approximated numerically by using a Gauss–Chebyshev numerical quadrature rule with L nodes [101]. We have

$$\begin{aligned} \mathbb{P}(\Delta < 0) & \quad (\text{D.4}) \\ &= \frac{1}{2L} \sum_{k=1}^L \left\{ \operatorname{Re} [\Phi_{\Delta}(c(1 + j\tau_k))] + \tau_k \operatorname{Im} [\Phi_{\Delta}(c(1 + j\tau_k))] \right\} + E_L, \end{aligned}$$

where $\tau_k \triangleq \tan((k - 1/2)\pi/L)$ and $E_L \rightarrow 0$ as $L \rightarrow \infty$. In numerical calculations, a rule-of-thumb choice yields $L = 64$.

Example D.1. As a special case of the above, consider the calculation of the expectation

$$P \triangleq \mathbb{E}[Q(\sqrt{\xi})], \quad (\text{D.5})$$

where $Q(y)$ is the Gaussian tail function, i.e., $Q(y) \triangleq \mathbb{P}(\nu > y)$ with $\nu \sim \mathcal{N}(0, 1)$, and ξ a nonnegative random variable. Defining $\Delta \triangleq \xi - \nu^2$, we have $P = (1/2)\mathbb{P}[\Delta < 0]$. Thus,

$$\Phi_{\Delta}(s) = \Phi_{\xi}(s)\Phi_{\nu^2}(-s) = \Phi_{\xi}(s)(1 - 2s)^{-1/2}.$$

Here the ROC of $\Phi_{\Delta}(s)$ includes the complex region defined by $\{0 < \text{Re}(s) < 1/2\}$. Therefore, we can safely assume $0 < c < 1/2$: a good choice is $c = 1/4$, corresponding to an integration line in the middle of the minimal ROC of $\Phi_{\Delta}(s)$. The latter integral can be evaluated numerically by using (D.4).

D.1 Application: MIMO channel

Here we apply the general technique outlined above to calculate pairwise error probabilities for MIMO channels affected by fading.

D.1.1 Independent-fading channel with coding

The discrete-time lowpass equivalent channel equation can be written as

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_i + \mathbf{z}_i \quad i = 1, \dots, N \quad (\text{D.6})$$

where N is the code block length, $\mathbf{H}_i \in \mathbb{C}^{rt}$ is the i th channel gain matrix, $\mathbf{x}_i \in \mathbb{C}^t$ is the i th transmitted symbol vector (each entry transmitted from a different antenna), $\mathbf{y}_i \in \mathbb{C}^r$ is the i th received sample vector (each entry received from a different antenna), and $\mathbf{z}_i \in \mathbb{C}^r$ is the i th received noise sample vector (each entry received from a different antenna). We assume that the channel gain matrices \mathbf{H}_i are element-wise independent and independent of each other with $[\mathbf{H}_i]_{jk} \sim \mathcal{N}_c(0, 1)$. Also, the noise samples are independent with $[\mathbf{z}]_i \sim \mathcal{N}_c(0, N_0)$.

It is straightforward to obtain the PEP associated with the two code words $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ and $\hat{\mathbf{X}} = (\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_N)$ as follows:

$$\begin{aligned} & P(\mathbf{X} \rightarrow \hat{\mathbf{X}}) \\ &= \mathbb{P} \left(\sum_{i=1}^N \{ \|\mathbf{y}_i - \mathbf{H}_i \hat{\mathbf{x}}_i\|^2 - \|\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_i\|^2 \} < 0 \right) \\ &= \mathbb{P} \left(\sum_{i=1}^N \{ \|\mathbf{H}_i(\mathbf{x}_i - \hat{\mathbf{x}}_i) + \mathbf{z}_i\|^2 - \|\mathbf{z}_i\|^2 \} < 0 \right) \\ &= \mathbb{E} \left[Q \left(\sqrt{\frac{1}{2N_0} \sum_{i=1}^N \|\mathbf{H}_i(\mathbf{x}_i - \hat{\mathbf{x}}_i)\|^2} \right) \right]. \end{aligned} \quad (\text{D.7})$$

Setting

$$\xi \triangleq \frac{1}{2N_0} \sum_{i=1}^N \|\mathbf{H}_i(\mathbf{x}_i - \hat{\mathbf{x}}_i)\|^2, \quad (\text{D.8})$$

a straightforward computation yields

$$\Phi_\xi(s) = \prod_{i=1}^N [1 + s\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2/(2N_0)]^{-r} \quad (\text{D.9})$$

and the result of Example D.1 applies.

D.1.2 Block-fading channel with coding

Here we assume that the channel gain matrices \mathbf{H}_i are independent of the time index i and are equal to \mathbf{H} : under this assumption the channel equation is

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}, \quad (\text{D.10})$$

where $\mathbf{H} \in \mathbb{C}^{rt}$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N) \in \mathbb{C}^{tN}$, $\mathbf{Y} \in \mathbb{C}^{rN}$, and $\mathbf{Z} \in \mathbb{C}^{rN}$. We assume iid entries $[\mathbf{H}]_{ij} \sim \mathcal{N}_c(0, 1)$ and iid $[\mathbf{Z}]_{ij} \sim \mathcal{N}_c(0, N_0)$. We obtain

$$\mathbb{P}(\mathbf{X} \rightarrow \hat{\mathbf{X}}) = \mathbb{E} \left[Q \left(\frac{\|\mathbf{H}\Delta\|}{\sqrt{2N_0}} \right) \right], \quad (\text{D.11})$$

where $\Delta \triangleq \mathbf{X} - \hat{\mathbf{X}}$.

Setting

$$\xi \triangleq \frac{\|\mathbf{H}\Delta\|^2}{2N_0}, \quad (\text{D.12})$$

we can evaluate the PEP by resorting to (D.5). In order to do so, we obtain the analytic expression of the moment-generating function (MGF). For the independent Rayleigh fading channel, straightforward computations yield

$$\Phi_\xi(s) = \prod_{i=1}^N [1 + s\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2/(2N_0)]^{-r}. \quad (\text{D.13})$$

Now, apply Theorem C.1. First, notice that ξ can be written in the form

$$\begin{aligned} \xi &= \frac{1}{2N_0} \sum_{i=1}^r \mathbf{h}_i \Delta \Delta^\dagger \mathbf{h}_i^\dagger \\ &= \frac{1}{2N_0} [\mathbf{h}_1, \dots, \mathbf{h}_r] [\mathbf{I}_r \otimes (\Delta \Delta^\dagger)] [\mathbf{h}_1, \dots, \mathbf{h}_r]^\dagger, \end{aligned} \quad (\text{D.14})$$

where \mathbf{h}_i denotes the i th row of matrix \mathbf{H} . Setting $\mathbf{z} = [\mathbf{h}_1, \dots, \mathbf{h}_r]^\dagger$, we have $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbb{E}[\mathbf{z}\mathbf{z}^\dagger] = \mathbf{I}_{rt}$. Finally, setting $\mathbf{A} = [\mathbf{I}_r \otimes (\boldsymbol{\Delta}\boldsymbol{\Delta}^\dagger)]/(2N_0)$ in (C.3), we obtain

$$\begin{aligned} \Phi_\xi(s) \triangleq \mathbb{E}[\exp(-s\xi)] &= \mathbb{E}[\exp(-s\mathbf{z}^\dagger \mathbf{A} \mathbf{z})] \\ &= \det(\mathbf{I} + s\boldsymbol{\Sigma}\mathbf{A})^{-1} \\ &= \det \left[\mathbf{I}_t + s\boldsymbol{\Delta}\boldsymbol{\Delta}^\dagger/2N_0 \right]^{-r}. \quad (\text{D.15}) \end{aligned}$$

Full text available at: <http://dx.doi.org/10.1561/0100000002>

E

Two proofs

E.1 Proof of (4.23)

Observe again (4.20), and define the $m \times m$ matrix

$$\mathbf{W} \triangleq \begin{cases} \mathbf{H}\mathbf{H}^\dagger, & r < t \\ \mathbf{H}^\dagger\mathbf{H}, & t \leq r, \end{cases} \quad (\text{E.1})$$

where again $m \triangleq \min\{t, r\}$. This is a nonnegative definite random matrix, and thus has real, nonnegative random eigenvalues. The joint pdf of the ordered eigenvalues of \mathbf{W} is known (see (C.5)). The expectation to be computed can be expressed in terms of one of the unordered eigenvalues of \mathbf{W} (say, λ_1) as follows:

$$\begin{aligned} C &= \mathbb{E} \log \prod_{i=1}^m \left(1 + \frac{\rho}{t} \lambda_i\right) = \mathbb{E} \sum_{i=1}^m \log \left(1 + \frac{\rho}{t} \lambda_i\right) \\ &= \sum_{i=1}^m \mathbb{E} \log \left(1 + \frac{\rho}{t} \lambda_i\right) \\ &= m \mathbb{E} \log \left(1 + \frac{\rho}{t} \lambda_1\right). \end{aligned} \quad (\text{E.2})$$

To compute the marginal pdf of λ_1 , use

$$p(\lambda_1) = \int \cdots \int p(\lambda_1, \lambda_2, \dots, \lambda_m) d\lambda_2 \cdots d\lambda_m. \quad (\text{E.3})$$

To perform this computation, we resort to an orthogonalization of the power sequence $1, \lambda, \lambda^2, \dots, \lambda^{m-1}$ in the Hilbert space of real functions defined in $(0, \infty)$ with inner product

$$(f, g) \triangleq \int_0^\infty f(\lambda)g(\lambda)\lambda^{n-m}e^{-\lambda} d\lambda. \quad (\text{E.4})$$

Explicitly, we express the pdf (E.3) in terms of the polynomials

$$\phi_{k+1}(\lambda) \triangleq \sqrt{\frac{k!}{(k+n-m)!}} L_k^{n-m}(\lambda). \quad (\text{E.5})$$

Here, $L_k^\alpha(\lambda)$ is an associated Laguerre polynomial [101], defined as

$$\begin{aligned} L_k^\alpha(\lambda) &= \frac{1}{k!} e^\lambda \lambda^{-\alpha} \frac{d^k}{d\lambda^k} (e^{-\lambda} \lambda^{k+\alpha}) \\ &= \sum_{\ell=0}^k (-1)^\ell \binom{k+\alpha}{k-\ell} \frac{\lambda^\ell}{\ell!}. \end{aligned} \quad (\text{E.6})$$

The polynomials $\phi_i(\lambda)$ satisfy the orthonormality relation

$$\int_0^\infty \phi_i(\lambda)\phi_j(\lambda)\lambda^{n-m}e^{-\lambda} d\lambda = \delta_{ij}. \quad (\text{E.7})$$

In order to calculate (E.3), we first observe that the term $\prod_{j=i+1}^m (\lambda_i - \lambda_j)$ appearing in (C.5) can be expressed as the determinant of the Vandermonde matrix

$$\mathbf{D}(\lambda_1, \lambda_2, \dots, \lambda_m) \triangleq \begin{bmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_m \\ \vdots & & \vdots \\ \lambda_1^{m-1} & \cdots & \lambda_m^{m-1} \end{bmatrix}, \quad (\text{E.8})$$

so that we can write

$$p(\lambda_1, \lambda_2, \dots, \lambda_m) = \frac{1}{\Gamma_m(m)\Gamma_m(n)} \det[\mathbf{D}(\lambda_1, \lambda_2, \dots, \lambda_m)]^2 \prod_{i=1}^m \lambda_i^{n-m} e^{-\lambda_i} \quad (\text{E.9})$$

with $\Gamma_m(a) \triangleq \prod_{i=0}^{m-1} \Gamma(a-i)$. Next, with row operations we transform matrix $\mathbf{D}(\lambda_1, \lambda_2, \dots, \lambda_m)$ into

$$\tilde{\mathbf{D}}(\lambda_1, \lambda_2, \dots, \lambda_m) \triangleq \begin{bmatrix} \phi_1(\lambda_1) & \cdots & \phi_1(\lambda_m) \\ \vdots & & \vdots \\ \phi_m(\lambda_1) & \cdots & \phi_m(\lambda_m) \end{bmatrix} \quad (\text{E.10})$$

so that the determinant of \mathbf{D} equals (apart from multiplicative constants generated by the row operations) the determinant of $\tilde{\mathbf{D}}$, that is,

$$\det \tilde{\mathbf{D}}(\lambda_1, \dots, \lambda_m) = \sum_{\alpha} (-1)^{\pi(\alpha)} \prod_{i=1}^m \phi_{\alpha_i}(\lambda_i), \quad (\text{E.11})$$

where the summation is over all permutations of $\{1, \dots, m\}$, and

$$\pi(\alpha) = \begin{cases} 0, & \alpha \text{ is an even permutation} \\ 1, & \text{otherwise.} \end{cases}$$

Thus, with $c(m, n)$ a normalization constant, we have

$$p(\lambda_1, \dots, \lambda_m) = c(m, n) \sum_{\alpha, \beta} (-1)^{\pi(\alpha) + \pi(\beta)} \prod_i \phi_{\alpha_i}(\lambda_i) \phi_{\beta_i}(\lambda_i) \lambda_i^{n-m} e^{-\lambda_i}, \quad (\text{E.12})$$

and, integrating over $\lambda_2, \dots, \lambda_m$, we obtain

$$\begin{aligned} p(\lambda_1) &= c(m, n) \sum_{\alpha, \beta} (-1)^{\pi(\alpha) + \pi(\beta)} \phi_{\alpha_1}(\lambda_1) \phi_{\beta_1}(\lambda_1) \lambda_1^{n-m} e^{-\lambda_1} \prod_{i=2}^m \delta_{\alpha_i \beta_i} \\ &= c(m, n) (m-1)! \sum_{i=1}^m \phi_i^2(\lambda_1) \lambda_1^{n-m} e^{-\lambda_1} \\ &= \frac{1}{m} \sum_{i=1}^m \phi_i^2(\lambda_1) \lambda_1^{n-m} e^{-\lambda_1}, \end{aligned} \quad (\text{E.13})$$

where the second equality follows from the fact that, if $\alpha_i = \beta_i$ for $i \geq 2$, then also $\alpha_1 = \beta_1$ (since both α and β are permutations of the same set) and thus $\alpha = \beta$. The last equality follows from the fact that $\phi_i^2(\lambda_1) \lambda_1^{n-m} e^{-\lambda_1}$ integrates to unity, which entails $c(m, n) = 1/m!$. In

conclusion, the capacity is given by

$$C = \int_0^\infty \log \left(1 + \frac{\rho}{t} \lambda \right) \sum_{k=0}^{m-1} \frac{k!}{(k+n-m)!} [L_k^{n-m}(\lambda)]^2 \lambda^{n-m} e^{-\lambda} d\lambda, \quad (\text{E.14})$$

where $m \triangleq \min\{t, r\}$, $n \triangleq \max\{t, r\}$.

Eq. (E.14) was derived in [106]. More recently, in [93] the integral in (E.14) was given the closed form

$$C = \log_2(e) e^{t/\rho} \sum_{i=0}^{m-1} \sum_{j=0}^i \sum_{\ell=0}^{2j} \left\{ \frac{(-1)^\ell (2j)! (n-m+\ell)!}{2^{2i-\ell} j! \ell! (n-m+j)!} \cdot \binom{2i-2j}{i-j} \binom{2j+2n-2m}{2j-\ell} \sum_{k=0}^{n-m+\ell} E_{k+1} \left(\frac{t}{\rho} \right) \right\} \quad (\text{E.15})$$

which involves a quadruple summation. Using a different approach, described below, a simpler expression can be obtained, resulting into a triple summation.

Consider the Christoffel–Darboux identity for orthonormal polynomials [101]:

$$\sum_{k=1}^m \phi_k(x) \phi_k(y) = \frac{A_m}{A_{m+1}} \frac{\phi_{m+1}(x) \phi_m(y) - \phi_m(x) \phi_{m+1}(y)}{x-y}, \quad (\text{E.16})$$

where A_k denotes the coefficient of x^{k-1} in $\phi_k(x)$. Taking the limit as $y \rightarrow x$, the above identity yields

$$\sum_{k=1}^m \phi_k(x)^2 = \frac{A_m}{A_{m+1}} [\phi_m(x) \phi'_{m+1}(x) - \phi_{m+1}(x) \phi'_m(x)]. \quad (\text{E.17})$$

When specialized to associated Laguerre polynomials, Eq. (E.5) and (E.6) yield

$$A_k = \frac{(-1)^{k-1}}{\sqrt{(k-1)!(k-1+\alpha)!}}$$

so that

$$\begin{aligned} \sum_{k=1}^m \phi_k(x)^2 &= \frac{m!}{(m+\alpha)!} [L_m^\alpha(x) (L_{m-1}^\alpha(x))' - L_{m-1}^\alpha(x) (L_m^\alpha(x))'] \\ &= \frac{m!}{(m+\alpha)!} [L_{m-1}^\alpha(x) L_{m-1}^{\alpha+1}(x) - L_{m-2}^{\alpha+1}(x) L_m^\alpha(x)] \quad (\text{E.18}) \end{aligned}$$

where we used the relation $[L_k^\alpha(x)]' = L_{k-1}^{\alpha+1}(x)$ [101]. Then, we can rewrite Eq. (E.14) as [48, 49]:

$$C = \frac{m!}{(n-1)!} \int_0^\infty [L_{m-1}^{n-m}(\lambda)L_{m-1}^{n-m+1}(\lambda) - L_{m-2}^{n-m+1}(\lambda)L_m^{n-m}(\lambda)] \log\left(1 + \frac{\rho}{t}\lambda\right) \lambda^{n-m} e^{-\lambda} d\lambda \quad (\text{E.19})$$

and further expand it using (E.6) as follows:

$$C = \frac{m!}{(n-1)!} \sum_{\ell=0}^{m-1} \sum_{\mu=0}^m \frac{(-1)^{\ell+\mu}}{\ell!\mu!} \int_0^\infty \log\left(1 + \frac{\rho}{t}\lambda\right) \lambda^{\ell+\mu+n-m} e^{-\lambda} d\lambda \left[\binom{n-1}{m-1-\ell} \binom{n}{m-1-\mu} - \binom{n-1}{m-2-\ell} \binom{n}{m-\mu} \right]. \quad (\text{E.20})$$

Finally, using a result in [93]:

$$\int_0^\infty \ln(1 + \rho\lambda) \lambda^\mu e^{-\lambda} d\lambda = \mu! e^{1/\rho} \sum_{p=0}^{\mu} E_{p+1}(1/\rho), \quad (\text{E.21})$$

we obtain:

$$C = \log(e) \frac{m!}{(n-1)!} \sum_{\ell=0}^{m-1} \sum_{\mu=0}^m \sum_{p=0}^{\ell+\mu+n-m} \frac{(-1)^{\ell+\mu} (\ell + \mu + n - m)!}{\ell!\mu!} e^{t/\rho} E_{p+1}(t/\rho) \left[\binom{n-1}{m-1-\ell} \binom{n}{m-1-\mu} - \binom{n-1}{m-2-\ell} \binom{n}{m-\mu} \right].$$

Notice that the number of terms added in (4.23) is approximately nm^2 against approximately $m^3(n-m/2)/3$ required in the capacity formula derived in [93].

E.2 Proof of (4.51)

For the proof we need the following general result:

Theorem E.1 ([38]). Given a square matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$, the improper integral

$$\int_{\mathbb{C}^n} \exp(-\pi \mathbf{x}^\dagger \mathbf{A} \mathbf{x}) d\mathbf{x} \quad (\text{E.22})$$

exists, in the Lebesgue sense, if and only if all eigenvalues of \mathbf{A} have positive real parts. In this case,

$$\int_{\mathbb{C}^n} \exp(-\pi \mathbf{x}^\dagger \mathbf{A} \mathbf{x}) d\mathbf{x} = \det \mathbf{A}^{-1}. \quad (\text{E.23})$$

We now write the determinant in (4.49) as follows:

$$\det \begin{pmatrix} \mathbf{I}_r & j\sqrt{\rho/t} \mathbf{H} \\ j\sqrt{\rho/t} \mathbf{H}^\dagger & \mathbf{I}_t \end{pmatrix} = \det \left(\mathbf{I}_r + \frac{\rho}{t} \mathbf{H} \mathbf{H}^\dagger \right). \quad (\text{E.24})$$

Then, applying Theorem E.1, we have:

$$\begin{aligned} & \det \left(\mathbf{I}_r + \frac{\rho}{t} \mathbf{H} \mathbf{H}^\dagger \right)^{-1} \\ &= \int_{\mathbb{C}^r} \int_{\mathbb{C}^t} \exp \left[-\pi (\mathbf{x}^\dagger, \mathbf{y}^\dagger) \begin{pmatrix} \mathbf{I}_r & j\sqrt{\rho/t} \mathbf{H} \\ j\sqrt{\rho/t} \mathbf{H}^\dagger & \mathbf{I}_t \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \right] d\mathbf{x} d\mathbf{y} \\ &= \int_{\mathbb{C}^r} \int_{\mathbb{C}^t} \exp \left[-\pi (\mathbf{x}^\dagger \mathbf{x} + \mathbf{y}^\dagger \mathbf{y}) + j\pi \sqrt{\rho/t} (\mathbf{x}^\dagger \mathbf{H} \mathbf{y} + \mathbf{y}^\dagger \mathbf{H}^\dagger \mathbf{x}) \right] d\mathbf{x} d\mathbf{y}. \end{aligned} \quad (\text{E.25})$$

Then, for integer s , we can write the s th power of the above LHS by “replicating” the integral s times with different integration variables to obtain

$$\begin{aligned} & \det \left(\mathbf{I}_r + \frac{\rho}{t} \mathbf{H} \mathbf{H}^\dagger \right)^{-s} \\ &= \prod_{a=1}^s \left\{ \int_{\mathbb{C}^r} \int_{\mathbb{C}^t} \exp \left[-\pi (\mathbf{x}_a^\dagger \mathbf{x}_a + \mathbf{y}_a^\dagger \mathbf{y}_a) \right. \right. \\ & \quad \left. \left. + j\pi \sqrt{\rho/t} (\mathbf{x}_a^\dagger \mathbf{H} \mathbf{y}_a + \mathbf{y}_a^\dagger \mathbf{H}^\dagger \mathbf{x}_a) \right] d\mathbf{x}_a d\mathbf{y}_a \right\}. \end{aligned} \quad (\text{E.26})$$

Averaging the result with respect to \mathbf{H} , we obtain

$$\begin{aligned} \Phi_C(s) &\triangleq \mathbb{E}[\exp(-sC)] \\ &= \int_{\mathbb{C}^{r \times t}} \pi^{-rt} \exp(-\|\mathbf{H}\|^2) d\mathbf{H} \prod_{a=1}^s \int_{\mathbb{C}^r} \int_{\mathbb{C}^t} \\ & \exp \left[-\pi (\mathbf{x}_a^\dagger \mathbf{x}_a + \mathbf{y}_a^\dagger \mathbf{y}_a) + j\pi \sqrt{\rho/t} (\mathbf{x}_a^\dagger \mathbf{H} \mathbf{y}_a + \mathbf{y}_a^\dagger \mathbf{H}^\dagger \mathbf{x}_a) \right] d\mathbf{x}_a d\mathbf{y}_a \end{aligned}$$

$$\begin{aligned}
 &= \int_{\mathbb{C}^{r \times t}} \pi^{-rt} d\mathbf{H} \int_{\mathbb{C}^{r \times s}} \int_{\mathbb{C}^{t \times s}} \text{etr} [-\pi(\mathbf{X}^\dagger \mathbf{X} + \mathbf{Y}^\dagger \mathbf{Y})] \\
 &\quad \text{etr} \left[j\pi\sqrt{\rho/t} (\mathbf{H}\mathbf{Y}\mathbf{X}^\dagger + \mathbf{X}\mathbf{Y}^\dagger\mathbf{H}^\dagger) - \mathbf{H}\mathbf{H}^\dagger \right] d\mathbf{X} d\mathbf{Y}, \tag{E.27}
 \end{aligned}$$

where $\mathbf{X} \triangleq (\mathbf{x}_1, \dots, \mathbf{x}_s)$ and $\mathbf{Y} \triangleq (\mathbf{y}_1, \dots, \mathbf{y}_s)$. Due to the fact that the integrand is bounded and continuous in all its variables, the integration order can be exchanged, and we integrate first over \mathbf{H} . Defining $\mathbf{A} \triangleq \pi\sqrt{\rho/t}\mathbf{X}\mathbf{Y}^\dagger$, we have:

$$\begin{aligned}
 &\int_{\mathbb{C}^{r \times t}} \pi^{-rt} \text{etr} \left[j(\mathbf{A}\mathbf{H}^\dagger + \mathbf{H}\mathbf{A}^\dagger) - \mathbf{H}\mathbf{H}^\dagger \right] d\mathbf{H} \\
 &= \prod_{i=1}^r \prod_{k=1}^t \pi^{-1} \int_{\mathbb{C}} \exp \left[2j \left(\text{Re}(\mathbf{A})_{i,k} \text{Re}(\mathbf{H})_{i,k} + \text{Im}(\mathbf{A})_{i,k} \text{Im}(\mathbf{H})_{i,k} \right) \right. \\
 &\quad \left. - (\text{Re}(\mathbf{H})_{i,k})^2 - (\text{Im}(\mathbf{H})_{i,k})^2 \right] d(\mathbf{H})_{i,k} \\
 &= \prod_{i=1}^r \prod_{k=1}^t \pi^{-1} \\
 &\quad \cdot \int_{\mathbb{R}} \exp \left[- \left(\text{Re}(\mathbf{H})_{i,k} - j\text{Re}(\mathbf{A})_{i,k} \right)^2 - (\text{Re}(\mathbf{A})_{i,k})^2 \right] d(\text{Re}(\mathbf{H})_{i,k}) \\
 &\quad \cdot \int_{\mathbb{R}} \exp \left[- \left(\text{Im}(\mathbf{H})_{i,k} - j\text{Im}(\mathbf{A})_{i,k} \right)^2 - (\text{Im}(\mathbf{A})_{i,k})^2 \right] d(\text{Im}(\mathbf{H})_{i,k}) \\
 &= \prod_{i=1}^r \prod_{k=1}^t \exp \left[- |(\mathbf{A})_{i,k}|^2 \right].
 \end{aligned}$$

Inserting (E.28) into (E.27) we obtain

$$\Phi_C(s) = \int_{\mathbb{C}^{r \times s}} \int_{\mathbb{C}^{t \times s}} \text{etr} \left[-\pi(\mathbf{X}^\dagger \mathbf{X} + \mathbf{Y}^\dagger \mathbf{Y}) - \pi^2 \frac{\rho}{t} \mathbf{X}^\dagger \mathbf{X} \mathbf{Y}^\dagger \mathbf{Y} \right] d\mathbf{X} d\mathbf{Y}. \tag{E.28}$$

From the equality

$$\frac{1}{j2\pi} \int_{q_0-j\infty}^{q_0+j\infty} dq \int_{r_0-\infty}^{r_0+\infty} \exp \left\{ q(\gamma r - x) - ry \right\} dr = \frac{\exp(-xy/\gamma)}{\gamma} \tag{E.29}$$

holding for any $x, y, \gamma, q_0, r_0 \in \mathbb{C}$, we have

$$\Phi_C(s) = \int_{\mathbb{C}^{r \times s}} \int_{\mathbb{C}^{r \times s}} \text{etr} \left[-\pi(\mathbf{X}^\dagger \mathbf{X} + \mathbf{Y}^\dagger \mathbf{Y}) \right]$$

140 Two proofs

$$\begin{aligned}
 & \prod_{a,b=1}^s \exp \left[-\pi^2 \frac{\rho}{t} (\mathbf{X}^\dagger \mathbf{X})_{a,b} (\mathbf{Y}^\dagger \mathbf{Y})_{b,a} \right] d\mathbf{X} d\mathbf{Y} \\
 = & \int_{\mathbb{C}^{r \times s}} \int_{\mathbb{C}^{r \times s}} \text{etr} \left[-\pi (\mathbf{X}^\dagger \mathbf{X} + \mathbf{Y}^\dagger \mathbf{Y}) \right] d\mathbf{X} d\mathbf{Y} \\
 & \prod_{a,b=1}^s \frac{1}{j2\pi} \int_{q_{b,a}-j\infty}^{q_{b,a}+j\infty} d(\mathbf{Q})_{b,a} \int_{r_{a,b}-\infty}^{r_{a,b}+\infty} \\
 & t \exp \left[(\mathbf{Q})_{b,a} \left(t(\mathbf{R})_{a,b} - \pi\sqrt{\rho}(\mathbf{X}^\dagger \mathbf{X})_{a,b} \right) - (\mathbf{R})_{a,b} \pi\sqrt{\rho}(\mathbf{Y}^\dagger \mathbf{Y})_{b,a} \right] \\
 & d(\mathbf{R})_{a,b} \\
 = & t^{s^2} \int_{\mathbb{C}^{r \times s}} \int_{\mathbb{C}^{r \times s}} \text{etr} \left[-\pi (\mathbf{X}^\dagger \mathbf{X} + \mathbf{Y}^\dagger \mathbf{Y}) \right] d\mathbf{X} d\mathbf{Y} \\
 & (j2\pi)^{-s^2} \int_{\mathbf{Q}_0 + (j\mathbb{R})^{s \times s}} d\mathbf{Q} \int_{\mathbf{R}_0 + \mathbb{R}^{s \times s}} \\
 & \text{etr} \left[t\mathbf{Q}\mathbf{R} - \mathbf{Q} \pi\sqrt{\rho}\mathbf{X}^\dagger \mathbf{X} - \mathbf{R} \pi\sqrt{\rho}\mathbf{Y}^\dagger \mathbf{Y} \right] d\mathbf{R} \\
 = & \left(\frac{t}{j2\pi} \right)^{s^2} \int_{\mathbf{Q}_0 + (j\mathbb{R})^{s \times s}} d\mathbf{Q} \int_{\mathbf{R}_0 + \mathbb{R}^{s \times s}} \\
 & \exp \left[t \text{Tr} (\mathbf{Q} \mathbf{R}) - t \ln \det(\mathbf{I}_s + \mathbf{Q}\sqrt{\rho}) - r \ln \det(\mathbf{I}_s + \mathbf{R}\sqrt{\rho}) \right] d\mathbf{R}. \tag{E.30}
 \end{aligned}$$

Here, \mathbf{Q}_0 and \mathbf{R}_0 are arbitrary $s \times s$ complex matrices that can be chosen to simplify the evaluation of the integral. Notice that the integration domain in the last integral is the Cartesian product of *one-dimensional paths* in \mathbb{C} . Now, we resort to multidimensional saddle-point integration [15] to calculate the integral asymptotically as $t, r \rightarrow \infty$ with $t/r \rightarrow \alpha$. First, we define

$$\varphi(\mathbf{Q}, \mathbf{R}) \triangleq \text{Tr} (\mathbf{Q} \mathbf{R}) - \ln \det(w\mathbf{I}_s + \mathbf{Q}) - \beta \ln \det(w\mathbf{I}_s + \mathbf{R}) \tag{E.31}$$

with $\beta \triangleq r/t$ and $w \triangleq \sqrt{1/\rho}$. Expanding the second-order variation of (E.31),

$$\begin{aligned}
 \delta\varphi(\mathbf{Q}, \mathbf{R}) = & \text{Tr} \left\{ [\mathbf{R} - (w\mathbf{I}_s + \mathbf{Q})^{-1}] \delta\mathbf{Q} + [\mathbf{Q} - \beta(w\mathbf{I}_s + \mathbf{R})^{-1}] \delta\mathbf{R} \right. \\
 & \left. + \frac{1}{2} [(w\mathbf{I}_s + \mathbf{Q})^{-1} \delta\mathbf{Q}]^2 + \delta\mathbf{Q} \delta\mathbf{R} + \frac{1}{2} \beta [(w\mathbf{I}_s + \mathbf{R})^{-1} \delta\mathbf{R}]^2 \right\}, \tag{E.32}
 \end{aligned}$$

we obtain the following critical points:

$$\mathbf{R} = (w\mathbf{I}_s + \mathbf{Q})^{-1} \quad \mathbf{Q} = \beta(w\mathbf{I}_s + \mathbf{R})^{-1} \quad (\text{E.33})$$

At any critical point we have

$$\delta\varphi(\mathbf{Q}, \mathbf{R}) = \frac{1}{2}\text{Tr} \left\{ (\mathbf{R}\delta\mathbf{Q})^2 + 2\delta\mathbf{Q}\delta\mathbf{R} + \beta^{-1}(\mathbf{Q}\delta\mathbf{R})^2 \right\} \quad (\text{E.34})$$

We now *assume* that the absolute maximum of $\text{Re} \{ \varphi(\mathbf{Q}, \mathbf{R}) \}$ occurs in correspondence of two scalar matrices $\mathbf{Q}_0 = \check{q}\mathbf{I}_s$ and $\mathbf{R}_0 = \check{r}\mathbf{I}_s$.¹ This leads to the following pair of equations:

$$\check{r} = \frac{1}{w + \check{q}} \quad \check{q} = \frac{\beta}{w + \check{r}} \quad (\text{E.35})$$

and the corresponding second-order expansion:

$$\begin{aligned} \varphi(\check{q}\mathbf{I}_s + \delta\mathbf{Q}, \check{r}\mathbf{I}_s + \delta\mathbf{R}) = & \quad (\text{E.36}) \\ \varphi(\check{q}\mathbf{I}_s, \check{r}\mathbf{I}_s) + \frac{1}{2}\text{Tr} \left\{ \check{r}^2(\delta\mathbf{Q})^2 + 2\delta\mathbf{Q}\delta\mathbf{R} + \beta^{-1}\check{q}^2(\delta\mathbf{R})^2 \right\}. \end{aligned}$$

Equations (E.34) have two possible solutions:

$$\check{q}_{1,2} = \frac{\beta - 1 - w^2 \pm \sqrt{(\beta - 1 - w^2)^2 + 4w^2\beta}}{2w} \quad (\text{E.37})$$

$$\check{r}_{1,2} = \frac{1 - \beta - w^2 \pm \sqrt{(1 - \beta - w^2)^2 + 4w^2}}{2w}. \quad (\text{E.38})$$

For the solution (q_0, r_0) we have

$$\begin{aligned} \varphi(q_0\mathbf{I}_s + \delta\mathbf{Q}, r_0\mathbf{I}_s + \delta\mathbf{R}) = & \quad (\text{E.39}) \\ \varphi(q_0\mathbf{I}_s, r_0\mathbf{I}_s) + \frac{1}{2}\text{Tr} \left\{ r_0^2(\delta\mathbf{Q})^2 + 2\delta\mathbf{Q}\delta\mathbf{R} + \beta^{-1}q_0^2(\delta\mathbf{R})^2 \right\} \end{aligned}$$

so that

$$\begin{aligned} w^{-ts(\beta+1)}\Phi_C(s) = & \quad (\text{E.40}) \\ \left(\frac{t}{j2\pi} \right)^{s^2} \int_{\mathcal{D}_q} d\check{\mathbf{Q}} \int_{\mathcal{D}_r} \exp[t\varphi(q_0\mathbf{I}_s + \check{\mathbf{Q}}, r_0\mathbf{I}_s + \check{\mathbf{R}})] d\check{\mathbf{R}} \end{aligned}$$

¹ The assumption of certain symmetries in the solution in order to avoid local minima, called the *Replica Symmetry Ansatz* in the statistical mechanics literature, lacks a general rigorous justification.

$$\begin{aligned} &\rightarrow \left(\frac{t}{j2\pi}\right)^{s^2} \exp[t\varphi(q_0\mathbf{I}_s, r_0\mathbf{I}_s)] \prod_{a=1}^s \prod_{b=1}^s \int_{\mathcal{D}_{a,b}^q} d(\check{\mathbf{Q}})_{a,b} \int_{\mathcal{D}_{a,b}^r} d(\check{\mathbf{R}})_{a,b} \\ &\exp\left[\frac{t}{2}\left(r_0^2(\check{\mathbf{Q}})_{a,b}(\check{\mathbf{Q}})_{b,a} + 2(\check{\mathbf{Q}})_{a,b}(\check{\mathbf{R}})_{b,a} + \beta^{-1}q_0^2(\check{\mathbf{R}})_{a,b}(\check{\mathbf{R}})_{b,a}\right)\right], \end{aligned}$$

where the integration domains \mathcal{D}_q and \mathcal{D}_r are chosen as the paths of steepest descent from the maximum in order to let the integrals converge [15]. These domains are the Cartesian products of $\mathcal{D}_{a,b}^q$ and $\mathcal{D}_{a,b}^r$, respectively, for $a, b = 1 \dots, s$. It can be shown that this condition requires that $q_0^2 r_0^2 > \beta$ and then

$$\begin{cases} q_0 = \frac{\beta - 1 - w^2 + \sqrt{(\beta - 1 - w^2)^2 + 4w^2\beta}}{2w} \\ r_0 = \frac{1 - \beta - w^2 + \sqrt{(1 - \beta - w^2)^2 + 4w^2}}{2w}. \end{cases} \quad (\text{E.41})$$

Finally, we obtain

$$\Phi_C(s) \rightarrow w^{st(1+\beta)} \exp[st(q_0 r_0 + \ln r_0 + \beta \ln(q_0/\beta))](1 - q_0^2 r_0^2/\beta)^{-s^2/2}. \quad (\text{E.42})$$

This result, which is a key ingredient of the Replica Method, confirms that the pdf of $C(\mathbf{H})$ approaches asymptotically the Gaussian distribution [46, 58, 91, 77]. It is valid provided that we are allowed to extend the range of s from nonnegative integers to the whole complex plane $\text{Re } s > 0$. More precisely, we assume that the expression (E.42) obtained for nonnegative integer values of s can be extended to a compact neighborhood of $s = 0^+$, and that the derivative with respect to s can be exchanged with the limit as $t \rightarrow \infty$. Then, as $t \rightarrow \infty$ and $t/r \rightarrow \alpha$, we can derive the asymptotic mean and variance as follows:

$$\begin{aligned} \frac{\mathbb{E}[C]}{t} &= \frac{1}{t} \frac{d}{ds} \ln \Phi_C(s) \Big|_{s=0} \\ &\rightarrow -\left\{ (1 + \beta) \ln w + q_0 r_0 + \ln r_0 + \beta \ln(q_0/\beta) \right\} \\ \mathbb{V}[C] &= \frac{d^2}{ds^2} \ln \Phi_C(s) \Big|_{s=0} \rightarrow -\ln(1 - q_0^2 r_0^2/\beta) \end{aligned}$$

expressed in nat/dimension pair and (nat/dimension pair)², respectively. The expectation $\mathbb{E}[C]$ is the ergodic capacity discussed in Section 4.2. In spite of the visual difference, this expression is equivalent

to (4.32). The variance $\mathbb{V}[C]$ was evaluated independently in the form of an integral for finite r and t in [98] and [115] (the latter reference actually derives the moment-generating function of $C(\mathbf{H})$, and hence all of its moments).

Full text available at: <http://dx.doi.org/10.1561/0100000002>

Acknowledgements

This work was supported by Cercom and the PRIMO project within FIRB.

The contribution of two anonymous reviewers, whose reports were as thorough and constructive as an author can dream of, is gratefully acknowledged.

Full text available at: <http://dx.doi.org/10.1561/0100000002>

References

- [1] S. L. Ariyavisitakul, "Turbo space-time processing to improve wireless channel capacity," *IEEE Trans. Commun.*, Vol. 48, No. 8, pp. 1347–1359, Aug. 2000.
- [2] Z.D. Bai and J.W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Annals of Probability*, Vol. 26, pp. 316–345, 1998.
- [3] Z. D. Bai and J. W. Silverstein, "CLT for linear spectral statistics of large dimensional sample covariance matrices," *Annals of Probability*, Vol. 32, No. 1A, pp. 553–605, 2004.
- [4] Z. D. Bai and Y. Q. Yin, "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," *Annals of Probability*, Vol. 21, pp. 1275–1294, 1993.
- [5] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*. New York: Kluwer/Plenum, 1999.
- [6] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, Vol. 47, No. 4, pp. 1273–1289, May 2001.
- [7] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*. New York: Macmillan, 1991.
- [8] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, 50th Anniversary Issue, Vol. 44, No. 6, pp. 2619–2692, October 1998.
- [9] E. Biglieri, A. Nordio, and G. Taricco, "Suboptimum receiver interfaces and space-time codes," *IEEE Trans. Signal Processing*, Vol. 51, No. 11, pp. 2720–2728, Nov. 2003.

- [10] E. Biglieri, A. Nardio, and G. Taricco, "Doubly-iterative decoding of space-time turbo codes with a large number of antennas," *Proc. IEEE ICC 2004*, Paris, France, Vol. 1, pp. 473-477, June 20-24, 2004.
- [11] E. Biglieri and G. Taricco "Large-system analyses of multiple-antenna system capacities," *Journal of Communications and Networks*, Vol. 5, No. 2, pp. 57-64, June 2003.
- [12] E. Biglieri, G. Taricco, and A. Tulino, "Performance of space-time codes for a large number of antennas," *IEEE Trans. Information Theory*, Vol. 48, No. 7, pp. 1794-1803, July 2002.
- [13] E. Biglieri, G. Taricco, and A. Tulino, "Decoding space-time codes with BLAST architectures," *IEEE Trans. Signal Processing*, Vol. 50, No. 10, pp. 2547-2552, October 2002.
- [14] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*. New York: Springer, 1999.
- [15] N. Bleistein and R.A. Handelsman, *Asymptotic Expansions of Integrals*. Dover, 1986
- [16] D. W. Bliss, K. W. Forsythe, A. O. Hero, III, and A. F. Yegulalp, "Environmental issues for MIMO capacity," *IEEE Trans. Signal Processing*, Vol. 50, No. 9, pp. 2128-2142, Sept. 2002.
- [17] R. S. Blum, "MIMO capacity with interference," *IEEE J. Select. Areas Commun.*, Vol. 21, No. 5, pp. 793-801, June 2003.
- [18] H. Boelcskei, D. Gesbert, C. Papadias, and A. J. van der Veen (Eds.), *space-time Wireless Systems: From Array Processing to MIMO Communications*. Cambridge University Press, to be published, 2005.
- [19] H. Boelcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, Vol. 50, No. 2, pp. 225-234, Feb. 2002.
- [20] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: The multivariate case," *Bell Syst. Tech. Journ.*, Vol. 53, pp. 745-778, May/June 1974.
- [21] G. Caire, "PSP-based space-time schemes for large spectral efficiency," *Proc. 2000 Conference on Information Sciences and Systems*, Princeton University, pp. TA6-30-TA6-35, March 15-17, 2000.
- [22] G. Caire and G. Colavolpe, "On low-complexity space-time coding for quasi-static channels," *IEEE Trans. Information Theory*, Vol. 49, No. 6, pp. 1400-1416, June 2003.
- [23] G. Caire, R. Knopp and P. Humblet, "System capacity of F-TDMA cellular systems," *IEEE Trans. Commun.*, Vol. 46, No. 12, pp. 1649-1661, December 1998.
- [24] G. Caire, G. Taricco and E. Biglieri, "Optimal power control for minimum outage rate in wireless communications," *Proc. IEEE ICC '98*, Vol. 1, pp. 58-62, Atlanta, GA, June 1998.
- [25] G. Caire, G. Taricco and E. Biglieri, "Optimal power control for the fading channel," *IEEE Trans. Information Theory*, Vol. 45, No. 5, pp. 1468-1489, July 1999.

- [26] S. Catreux, P.F. Driessens, and L.J. Greenstein, "Attainable throughput of an interference-limited multiple-input multiple-output (MIMO) cellular system," *IEEE Trans. Commun.*, Vol. 49, No. 8, pp. 1307–1311, Aug. 2001.
- [27] J. K. Cavers and P. Ho, "Analysis of the Error Performance of Trellis-Coded Modulations in Rayleigh Fading Channels," *IEEE Trans. Commun.*, Vol. 40, No. 1, pp. 74–83, Jan. 1992.
- [28] J.-J. Chang, D.-J. Hwang, and M.-C. Lin, "Some extended results on the search for good convolutional codes," *IEEE Trans. Information Theory*, Vol. 43, No. 5, pp. 1682–1697, Sept. 1997.
- [29] N. Chiurtu, B. Rimoldi, and E. Telatar, "Dense multiple antenna systems," *Proc. IEEE Information Theory Workshop 2001*, Cairns, Australia, pp. 108–109, Sept. 2–7, 2001.
- [30] D. Chizhik, G.J. Foschini, M.J. Gans, and R.A. Valenzuela, "Keyholes, correlations, and capacities of multielement transmit and receive antennas," *IEEE Trans. Wireless Commun.*, Vol. 1, No. 2, pp. 361–368, April 2002.
- [31] C.-N. Chuah, D. Tse, and J. M. Kahn, "Capacity of multi-antenna array systems in indoor wireless environment," *Proc. IEEE Globecom '98*, Sydney, Australia, Vol. 4, pp. 1894–1899, Nov. 8–12, 1998.
- [32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [33] D. C. Cox, "Universal digital portable radio communications," *Proc. IEEE*, Vol. 75, No. 4, pp. 463–477, Apr. 1987.
- [34] C. B. Dietrich, Jr., K. Dietze, J. R. Nealy, and W. L. Stutzman, "Spatial, polarization, and pattern diversity for wireless handheld terminals," *IEEE Trans. Antennas Propagat.*, Vol. 49, No. 9, pp. 1271–1281, Sept. 2001.
- [35] S. N. Diggavi, A. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, "Great expectations: The value of spatial diversity in wireless networks," *Proc. IEEE, Special Issue on Gigabit Wireless*, Vol. 92, No. 2, pp. 219–270, Feb. 2004.
- [36] A. Edelman, *Eigenvalues and condition numbers of random matrices*. PhD Thesis, Dept. Mathematics, Massachusetts Institute of Technology, Cambridge, MA, 1989.
- [37] H. El-Gamal and A. R. Hammons, Jr., "A new approach to layered space–time coding and signal processing," *IEEE Trans. Information Theory*, Vol. 47, No. 6, pp. 2321–2334, Sept. 2001.
- [38] M. Elia and G. Taricco, "Integration of the exponential function of a complex quadratic form," *Applied Mathematics E-Notes*, Vol. 3, pp. 95–98, 2003. (Available from <http://www.math.nthu.edu.tw/~amen/>).
- [39] G. J. Foschini, "Layered space–time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, Vol. 1, No. 2, pp. 41–59, Autumn 1996.
- [40] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, Vol. 6, No. 3, pp. 311–335, March 1998.
- [41] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employ-

150 References

- ing multi-element arrays," *IEEE J. Select. Areas Commun.*, Vol. 17, No. 11, pp. 1841–1852, Nov. 1999.
- [42] G. J. Foschini and R. A. Valenzuela, "Initial estimation of communication efficiency of indoor wireless channels," *Wireless Networks*, Vol. 3, No. 2, pp. 141–154, 1997.
- [43] R. G. Gallager, *Information Theory and Reliable Communication*. New York: J. Wiley & Sons, 1968.
- [44] D. Gesbert, H. Bölcskei, D. Gore, and A. Paulraj, "MIMO wireless channels: capacity and performance prediction," *Proc. IEEE Globecom 2000*, San Francisco, Vol. 2, pp. 1083–1088, Nov. 2000.
- [45] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space–time coded wireless systems," *IEEE J. Select. Areas Commun.*, Vol. 21, No. 3, pp. 281–302, April 2003.
- [46] V. L. Girko, "A refinement of the central limit theorem for random determinants," *Theory Probab. Appl.*, Vol. 42, No. 1, pp. 121–129, 1997.
- [47] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, Vol. 21, No. 5, pp. 684–702, June 2003.
- [48] A. Grant, "Rayleigh fading multiple antenna channels," *EURASIP Journal of Applied Signal Processing, Special Issue on Space–Time Coding and its Applications - Part I*, Vol. 2002, No. 3, pp. 316–329, March 2002.
- [49] A. Grant, "Correction to 'Rayleigh fading multiple antenna channels'," *EURASIP Journal of Applied Signal Processing*, Vol. 2003, No. 3, p. 317, March 2003.
- [50] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, Vol. 47, No. 4, pp. 527–537, April 1999.
- [51] A. R. Hammons, Jr., and H. El Gamal, "On the theory of space–time codes for PSK modulation," *IEEE Trans. Information Theory*, Vol. 46, No. 2, pp. 524–542, Mar. 2000.
- [52] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Information Theory*, Vol. 48, No. 7, pp. 1804–1824, July 2002.
- [53] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Information Theory*, Vol. 49, No. 4, pp. 951–963, April 2003.
- [54] B. Hassibi, B. M. Hochwald, A. Shokrollahi, and W. Sweldens, "Representation theory for high-rate multiple-antenna code design," *IEEE Trans. Information Theory*, Vol. 47, No. 6, pp. 2335–2367, Sept. 2001.
- [55] R.W. Heath, Jr., and A. Paulraj, "Characterization of MIMO channels for spatial multiplexing systems," *Proc. IEEE ICC 2001*, Helsinki, Finland, Vol. 2, pp. 591–595, June 11–14, 2001.
- [56] B. Hochwald and T. Marzetta, "Space–time modulation scheme for unknown Rayleigh fading environment," *Proc. 36th Annual Allerton Conference on Communication, Control and Computing*, pp. 431–440, Allerton House, Monticello, IL, September 23–25, 1998.

- [57] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communication in Rayleigh flat-fading," *IEEE Trans. Information Theory*, Vol. 46, No. 2, pp. 543–564, March 2000.
- [58] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel-hardening and its implications for rate feedback and scheduling," *IEEE Trans. Information Theory*, Vol. 50, No. 9, pp. 1893–1909, Sept. 2004.
- [59] B. Hochwald and W. Sweldens, "Differential unitary space time modulation," *IEEE Trans. Commun.*, Vol. 48, No. 12, pp. 2041–2052, December 2000.
- [60] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, Vol. 51, No. 3, pp. 389–399, Mar. 2003.
- [61] D. Hoesli and A. Lapidoth, "The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean," *Proc. 5th International ITG Conference on Source and Channel Coding (SCC)*, Erlangen, Germany, pp. 381–385, January 14–16, 2004.
- [62] R. Horn and C. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1985.
- [63] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. American Mathematical Society, 2000.
- [64] W. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
- [65] A.T. James, "Distribution of matrix variates and latent roots derived from normal samples," *Ann. Math. Statistics*, Vol. 35, pp. 475–501, 1964.
- [66] D. Jonsson, "Some limit theorems for the eigenvalues of of a sample covariance matrix," *J. Multivariate Analysis*, Vol. 12, No. 1, pp. 1–38, 1982.
- [67] R. Knopp, *Coding and Multiple-Accessing over Fading Channels*, PhD dissertation, EPFL, Lausanne (Switzerland), and Institut Eurécom, Sophia Antipolis (France), 1997.
- [68] R. Knopp and P. A. Humblet, "On coding for block-fading channels," *IEEE Trans. Information Theory*, Vol. 46, No. 1, pp. 189–205, Jan. 2000.
- [69] E.G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, UK: Cambridge Univ. Press, 2003.
- [70] A. Lozano, "Capacity-approaching rate function for layered multi-antenna architectures," *IEEE Trans. Wireless Commun.*, Vol. 2, No. 4, pp. 616–620, July 2003.
- [71] X. Ma and G. B. Giannakis, "Full-diversity full-rate complex-field space-time coding," *IEEE Trans. Signal Processing*, Vol. 51, No. 11, pp. 2917–2930, Nov. 2003.
- [72] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," *Proc. 37th Annual Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, IL, pp. 958–966. Sept. 22–24, 1999.
- [73] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Information Theory*, Vol. 45, No. 1, pp. 139–157, January 1999.
- [74] T. L. Marzetta, B. M. Hochwald, and B. Hassibi, "New approach to multiple-user multiple-antenna wireless communication," *Proc. 2000 Conference on In-*

- formation Sciences and Systems*, Princeton University, pp. WA4-16–WA4-21, March 15–17, 2000.
- [75] A.F. Molish, M.Z. Win, and J.H. Winters, “Capacity of MIMO systems with antenna selection,” *Proc. IEEE ICC 2001*, Helsinki, Finland, June 11-14, 2001.
- [76] A. L. Moustakas, H. U. Baranger, L. Balents, A. M. Sengupta, and S.H. Simon, “Communication through a diffusive medium: Coherence and capacity,” *Science*, Vol. 287, pp. 287–290, Jan. 2000.
- [77] A. L. Moustakas, S. H. Simon, A. M. Sengupta, “MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis,” *IEEE Trans. Information Theory* Vol. 49, No. 10, pp. 2545–2561, Oct. 2003.
- [78] R. U. Nabar, H. Boelcskei, V. Erceg, D. Gesbert, A. J. Paulraj, “Performance of multiantenna signaling techniques in the presence of polarization diversity,” *IEEE Trans. Signal Processing*, Vol. 50, No. 10, pp. 2553–2562, Oct. 2002.
- [79] F. D. Neeser and J. L. Massey, “Proper complex random processes with applications to information theory,” *IEEE Trans. Information Theory*, Vol. 39, No. 4, pp. 1293–1302, July 1993.
- [80] L. Ozarow, S. Shamai, and A. D. Wyner, “Information theoretic considerations for cellular mobile radio,” *IEEE Trans. Vehicular Technology*, Vol. 43, No. 2, pp. 359–378, May 1994.
- [81] A. Paulraj and T. Kailath. *Increasing Capacity in Wireless Broadcast Systems Using Distributed Transmission/Directional Reception*. US Patent, 5 345 599, 1994.
- [82] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time Wireless Communications*. Cambridge University Press, 2003.
- [83] J. G. Proakis, *Digital Communications*, Third Ed. New York: Mc-Graw-Hill, 1995.
- [84] G. Raleigh and J. Cioffi, “Spatio-temporal coding for wireless communication,” *IEEE Trans. Commun.*, Vol. 46, No. 3, pp. 357–366, March 1998.
- [85] P. B. Rapajic and D. Popescu, “Information capacity of a random signature multiple-input multiple-output channel,” *IEEE Trans. Commun.*, Vol. 48, No. 8, pp. 1245–1248, Aug. 2000.
- [86] W. Rhee and J.M. Cioffi, “On the capacity of multiuser wireless channels with multiple antennas,” *IEEE Trans. Information Theory*, Vol. 49, No. 10, pp. 2580–2595, Oct. 2003.
- [87] M. Schwartz, W. R. Bennett, and S. Stein, *Communications Systems and Techniques*. New York: McGraw-Hill, 1966.
- [88] M. Sellathurai and S. Haykin, “Turbo-BLAST for wireless communications: Theory and experiments,” *IEEE Trans. Signal Processing*, Vol. 50, No. 10, pp. 2538–2546, Oct. 2002.
- [89] M. Sellathurai and S. Haykin, “Turbo-BLAST: Performance evaluation in correlated Rayleigh-fading environment,” *IEEE J. Select. Areas Commun.*, Vol. 21, No. 3, pp. 340–349, Apr. 2003.
- [90] M. Sellathurai and S. Haykin, “T-BLAST for wireless communications: First experimental results,” *IEEE Trans. Vehicular Technology*, Vol. 52, No. 3, pp. 530–535, May 2003.

- [91] A. M. Sengupta and P. P. Mitra, "Capacity of multivariate channels with multiplicative noise: Random matrix techniques and large- N expansions for full transfer matrices," *LANL arXiv:physics*, 31 Oct. 2000.
- [92] N. Seshadri and J. H. Winters, "Two signaling schemes for improving the error performance of frequency-division duplex (FDD) transmission systems using transmitter antenna diversity," *Proc. 43rd IEEE Vehicular Technology Conference (VTC 1993)*, pp. 508–511, May 18–20, 1993.
- [93] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Information Theory*, Vol. 49, No. 10, pp. 2636–2647, Oct. 2003.
- [94] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems", *IEEE Trans. Commun.*, Vol. 48, No. 3, pp. 502–513, Mar. 2000.
- [95] D. Shiu and J. M. Kahn, "Design of high-throughput codes for multiple-antenna wireless systems," *IEEE Trans. Information Theory*, submitted for publication, January 1999.
- [96] D. Shiu and J. M. Kahn, "Layered space-time codes for wireless communications using multiple transmit antennas," *Proc. IEEE ICC'99*, Vancouver, BC, Vol. 1, pp. 436–440, June 6–10, 1999.
- [97] J. W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *Journal of Multivariate Analysis*, Vol. 55, pp. 331–339, 1995.
- [98] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," *Proc. IEEE ICC 2002*, pp. 406–410, New York, April 28–May 2, 2002.
- [99] P. Soma, D. S. Baum, V. Erceg, R. Krishnamoorthy, and A. J. Paulraj, "Analysis and modeling of multiple-input multiple-output (MIMO) radio channel based on outdoor measurements conducted at 2.5 GHz for fixed BWA applications," in *Proc. IEEE ICC 2002*, Vol. 1, pp. 272–276, New York, NY, May 28–April 2, 2002.
- [100] A. S. Stefanov and T. M. Duman, "Turbo coded modulation for systems with transmit and receive antenna diversity," *Proc. Globecom 99*, Rio de Janeiro, Brazil, Vol. 5, pp. 2336–2340, Dec. 5–9, 1999.
- [101] G. Szegő, *Orthogonal Polynomials*. Americal Mathematical Society, Providence, RI, 1939.
- [102] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, to be published, 2004.
- [103] V. Tarokh and T. K. Y. Lo, "Principal ratio combining for fixed wireless applications when transmitter diversity is employed," *IEEE Commun. Letters*, Vol. 2, No. 8, pp. 223–225, August 1998.
- [104] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Information Theory* Vol. 44, No. 2, pp. 744 – 765, March 1998.

154 *References*

- [105] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Information Theory*, Vol. 45, No. 5, pp. 1456–1467, July 1999.
- [106] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, Vol. 10, No. 6, pp. 585–595, November–December 1999.
- [107] D. Tse and V. Hanly, "Multi-access fading channels—Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Information Theory*, Vol. 44, No. 7, pp. 2796–2815, November 1998.
- [108] D. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff for multiaccess channels," *Proc. IEEE ISIT 2003*, Yokohama, Japan, p. 352, June 29–July 4, 2003.
- [109] A. Tulino and S. Verdú, "Random Matrix Theory and Wireless Communications," *Foundations and Trends in Communications and Information Theory*, Vol. 1, Issue 1, 2004.
- [110] M. Uysal and N.C. Georghiades, "Error performance analysis of space-time codes over Rayleigh fading channels," *Journal of Communications and Networks*, Vol. 2, No. 4, pp. 344–350, Dec. 2000.
- [111] S. Verdú, *Multiuser Detection*. Cambridge, UK: Cambridge Univ. Press, 1998.
- [112] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Information Theory*, Vol. 48, No. 6, pp. 1319–1343, June 2002.
- [113] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Information Theory*, Vol. 45, No. 2, pp. 622–640, Mar. 1999.
- [114] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables: A Noncommutative Probability Approach to Free Products with Applications to Random Matrices, Operator Algebras, and Harmonic Analysis on Free Groups*, Providence, NJ: American Mathematical Society, 1992.
- [115] Z. Wang and G. B. Giannakis, "Outage mutual information of space-time MIMO channels," *Fortieth Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, pp. 885–894, October 2–4, 2002.
- [116] H. Wang and X.-G. Xia, "Upper bounds of rates of space-time block codes from complex orthogonal designs," *Proc. IEEE ISIT 2002*, Lausanne, Switzerland, p. 303, June 30–July 5, 2002.
- [117] J. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Commun.*, Vol. 42, No. 2/3/4, pp. 1740–1751, Febr./March/April 1994.
- [118] A. Wittneben, "Basestation modulation diversity for digital SIMULCAST," *Proc. 41st Vehicular Technology Conference (VTC 1991)*, pp. 848–853, May 19–22, 1991.
- [119] D. Wuebben, R. Boehnke, J. Rinas, V. Kuehn, and K. D. Kammeyer, "Efficient algorithm for decoding layered space-time codes," *Electronics Letters*, Vol. 37, No. 22, pp. 1348–1349, 25 Oct. 2001.
- [120] L. Zheng and D. N. C. Tse, "Communication on the Grassman manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Information Theory*, Vol. 48, No. 2, pp. 359–383, February 2002.

Full text available at: <http://dx.doi.org/10.1561/0100000002>

References 155

- [121] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental trade-off in multiple antenna channels," *IEEE Trans. Information Theory*, Vol. 49, No. 5, pp. 1073–1096, May 2003.

Full text available at: <http://dx.doi.org/10.1561/0100000002>

Notations and Acronyms

\triangleq Equal by definition $(\cdot)_+ \triangleq \max(0, \cdot)$ $\mathbf{1}\{A\} \triangleq$ 1 if A is true and 0 otherwise (indicator function) $\Gamma(x) \triangleq \int_0^\infty u^{x-1} e^{-u} du$, The Gamma function \mathbf{A}^+ (Moore–Penrose) pseudoinverse of matrix \mathbf{A} \mathbf{A}' Transpose of matrix \mathbf{A} \mathbf{A}^\dagger Conjugate (or Hermitian) transpose of matrix \mathbf{A} $(\mathbf{A}, \mathbf{B}) \triangleq \text{Re Tr}(\mathbf{A}\mathbf{B}^\dagger)$, scalar matrix product $\ \mathbf{A}\ ^2 \triangleq (\mathbf{A}, \mathbf{A})$, squared Frobenius norm of matrix \mathbf{A} $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is nonnegative definite a.s. Almost surely AWGN Additive white Gaussian noise \mathbb{C} The set of complex numbers CSI Channel state information $\mathbb{E}[X]$ Expectation of the random variable X	$\text{etr}(\cdot) \triangleq \exp(\text{Tr}(\cdot))$ FER Frame-error rate \mathbf{I}_n The $n \times n$ identity matrix iid Independent and identically distributed Im Imaginary part LHS Left-hand side ln Natural logarithm log Logarithm in base 2 MGF Moment-generating function MIMO Multiple-input, multiple-output ML Maximum-likelihood MMSE Minimum mean-square error MSE Mean-square error pdf Probability density function PEP Pairwise error probability $Q(x) \triangleq (2\pi)^{-1/2} \int_x^\infty \exp(-z^2/2) dz$, the Gaussian tail function \mathbb{R} The set of real numbers Re Real part RHS Right-hand side ROC Region of convergence RV Random variable
---	--

