Claude Oestges and Bruno Clerckx

MIMO WIRELESS COMMUNICATIONS

From Real-World Propagation to Space–Time Code Design

MIMO WIRELESS COMMUNICATIONS

Reviewers' Comments

Finally, a book devoted to MIMO from a new perspective that bridges traditional borders between propagation, channel modeling, signal processing, and space-time coding. Covering the latest European and American research results, the book gives a comprehensive treatment of general MIMO topics and provides surprising insights into advanced features, such as polarization, the finite-SNR behavior of space-time coding schemes in correlated channels, as well as correlation situations leading to higher mutual information than the i.i.d. Rayleigh channel. This is a book of high reference value combining intuitive and conceptual explanations with detailed, stringent derivations of basic facts of MIMO.

<div align="right">

Ernst Bonek
Emeritus Professor
Technische Universität Wien, Austria, former Chairman of Working Groups on
Propagation and Antennas in COST 259 and COST273

</div>

Claude Oestges and Bruno Clerckx have put together a survey of real-world MIMO propagation channel models and the performance of space-time coding and pre-coding schemes in such channels. This book offers important insights into how space-time coding can be tailored for real-world MIMO channels. The discussion of MIMO propagation models is also intuitive and well developed.

The book is both valuable and timely for wireless engineers since MIMO is now entering all the emerging broadband standards such as WiFi, WiMAX and 3GPPLTE.

<div align="right">

Professor Arogyaswami J. Paulraj
Stanford University, CA

</div>

MIMO Wireless Communications

From real-world propagation to space–time code design

Claude Oestges and Bruno Clerckx

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER   BOOK AID   Sabre Foundation
              International

Contents

List of Figures

List of Tables

This page intentionally left blank

Preface

When we started thinking about writing this book, we had been working together for more than five years on the borderline between propagation and signal processing. Therefore, it is not surprising that this book deals with propagation models and design tools for MIMO wireless communications. Yet, this book should constitute more than a simple combination of these two domains. It hopefully conveys our integrated understanding of MIMO, which results from endless controversial discussions on various multi-antenna related issues, as well as various interactions with numerous colleagues. Obviously, this area of technology is so large that it was beyond our aim to cover all aspects in details. Rather, our goal has been to provide researchers, R&D engineers and graduate students with a comprehensive coverage of radio propagation models and space–time coding techniques.

Much has been written about MIMO. Still, the present issue is to 'make the thing work' in real-word wireless channels and under realistic power constraints. Indeed, both the antenna size and the transmit power are limited, which imposes some limitations on system design. As an example, space–time coding designs relying on idealistic propagation models may lose many of their advantages in more realistic radio channels. Therefore, a true challenge consists of proposing design methodologies that take MIMO propagation into account. Often, propagation models are considered by many as the simple combination of various 'cooking recipes', typically path-loss laws and tap-delay lines. Similarly, it is sometimes thought that a wireless transmission scheme does not need to care about the physics of the radio channel, but only about its impact on the received signals. In other words, propagation models would mostly be useful for *a posteriori* testing and modifying specific designs. Whereas this is certainly an important aspect of propagation models, which is well covered by IEEE standards, it may not be the only one, especially when dealing with multi-antenna wireless systems. Not only can radio propagation provide answers to some crucial questions of MIMO related problems, but it is also of outstanding importance for designing efficient transmission schemes. As a consequence, our approach builds on a deep study of MIMO propagation aspects, naturally including, but not limited to, classical simulation models. It shall enable designers to develop robust space–time coding schemes based on solid theoretical developments allied to a strong propagation-motivated intuition.

This book combines a description of the most recent space–time coding techniques with complete coverage of MIMO propagation models. Under both information theory and error probability perspectives, we emphasize how real-world propagation affects the capacity and error performance of MIMO transmission schemes. We then describe innovative and practical designs of robust space–time codes. We also cover, in detail, precoding schemes, antenna selection techniques and multi-carrier transmissions. We have attempted to build the book content as a logical flow, pointing out important theoretical results and providing

various examples. Although we have tried to supply detailed and clearly indicated proofs for most results, the reader is sometimes guided to references for greater detail.

Now that this project has come to an end, there are several people who deserve our warmest thanks. We are first deeply grateful to Professor Arogyaswami J. Paulraj at Stanford University for introducing both of us to the challenging area of MIMO communications. Professor Ernst Bonek at TU Wien also deserves our gratitude for initiating this project, as well as for his long-lasting support and careful reading of the manuscript. We also acknowledge the help of many at UCL during these last years, especially Professor Danielle Vanhoenacker-Janvier and Professor Luc Vandendorpe. We thank all the past and present members of the Microwave Laboratory and the Digital Communication Group for their friendly encouragement. We also acknowledge the help of Tim Pitts, Kate Dennis, Helen Eaton and Jackie Holding at Elsevier for their kind and efficient handling of this publication project.

Last but not least, we wish to heartily thank a number of anonymous and not so anonymous reviewers for their careful reading and most useful advices: Bertrand Devillers, Dr Mischa Dohler, Dr Maxime Guillaud, Professor Are Hjørungnes, Marios Kountouris, and Harold Sneessens.

Claude Oestges
Bruno Clerckx

List of Abbreviations

2D, 3D	two or three dimensional
3G, 4G	third or fourth generation
3GPP	third generation partnership project
AWGN	additive white Gaussian noise
BBC	BER balancing criterion
BER	bit error rate
BPSK	binary phase-shift key
BS	base station
CCI	co-channel interference
CDD	cyclic delay-diversity
CDF	cumulative distribution function
CDI(T)	channel distribution information (at the transmitter)
CDMA	code division multiple access
CIR	channel impulse response
CP	cyclic prefix
CSI(T)	channel state information (at the transmitter)
D-BLAST	diagonal Bell Labs layered space–time
DFT	discrete Fourier transform
DoA	direction-of-arrival
DoD	direction-of-departure
DPS	direction power spectrum
EGC	equal gain combining
EVD	eigenvalue decomposition
FDD	frequency division duplexing
FEC	forward error correction
FER	frame error rate
GDD	generalized delay-diversity
HS/MRC	hybrid selection/maximal ratio combining
IDFT	inverse discrete Fourier transform
iff	if and only if
i.i.d.	independent and identically distributed
ISI	inter symbol interference
LAST	lattice space–time
LDC	linear dispersion code
LDPC	Low Density Parity Check
LOS	line of sight
LP	Lindskog-Paulraj
LSR	local scattering ratio

MC	multi-carrier
MFB	matched filter bound
MGF	moment generating function
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
ML	maximum likelihood
MLSE	maximum-likelihood sequence estimation
MMSE	minimum mean square error
MRC	maximum ratio combining
MT	mobile terminal
OFDM	orthogonal frequency division multiplexing
OSIC	ordered successive interference canceller
O-SFBC	orthogonal space-frequency block code
O-STBC	orthogonal space–time block code
PAM	Pulse Amplitude modulation
PDF	probability density function
PDS	power-delay spectrum
PEP	pairwise error probability
PSK	phase shift keying
QAM	quadrature amplitude modulation
QoS	quality of service
QO-STBC	quasi-orthogonal space–time block code
QPSK	quadrature phase shift keying
RF	radio frequency
RMS	root mean square
Rx	receiver
SC	single-carrier
SER	symbol error rate
SIC	successive interference canceler
SIMO	single-input multiple-output
SINR	signal-to-interference-and-noise ratio
SISO	single-input single-output
SM	spatial multiplexing
SNR	signal-to-noise ratio
ST	space–time
STBC	space–time block code
STTC	space–time trellis code
SU	subscriber unit
SUI	Stanford University Interim
SVD	singular value decomposition
SVER	symbol vector error rate
TAST	threaded algebraic space–time
TCM	trellis coded modulation

TDD	time division duplexing
TIMO	two-input multiple-output
Tx	transmitter
ULA	Uniform linear arrays
UMTS	universal mobile telecommunication system
UTD	Uniform Theory of Diffraction
V-BLAST	vertical Bell Labs layered space–time
WLAN	wireless local area networks
WLL	wireless local loop
WMAN	wireless metropolitan area networks
WSSUH	wide-sense stationary uncorrelated scattering homogeneous
XPD	cross-polar discrimination
ZF	zero-forcing

This page intentionally left blank

List of Symbols

$\triangleq$	variable definition
$\mathcal{E}$	expectation operator
$\otimes$	Kronecker product
$\odot$	Hadamard (element-wise) product
$\star$	convolution product
$\doteq$	exponential equality, $f(x) \doteq x^n \Leftrightarrow \lim_{x \to \infty} \frac{\log_2(f(x))}{\log_2(x)} = b$
$\mathbb{R}$	real field
$\mathbb{Z}$	integer field
$\mathbb{C}$	complex field
ρ	signal-to-noise ratio
a, α	scalar
$\mathbf{a}$	column or row vector
$\mathbf{A}$	matrix
$\mathbf{A}(m, n)$	element in m^{th} row and n^{th} column of matrix $\mathbf{A}$
$\mathbf{A}(m, :)$	m^{th} row of matrix $\mathbf{A}$
$\mathbf{A}(:, n)$	n^{th} column of matrix $\mathbf{A}$
$\mathbf{A}^T$	transpose
$\mathbf{A}^*$	conjugate
$\mathbf{A}^H$	conjugate transpose (Hermitian)
$\mathbf{A}^\dagger$	pseudo-inverse
$\mathbf{A} \succeq 0$	means that $\mathbf{A}$ is positive semidefinite
$\det(\mathbf{A})$	determinant of $\mathbf{A}$
$\mathrm{Tr}\{\mathbf{A}\}$	trace of $\mathbf{A}$
$\Re[\mathbf{A}]$	real part of $\mathbf{A}$
$\Im[\mathbf{A}]$	imaginary part of $\mathbf{A}$
$|a|$	absolute value of scalar a
$\|\mathbf{A}\|_F$	Frobenius norm of $\mathbf{A}$
$|\mathbf{A}|$	element-wise absolute value of matrix $\mathbf{A}$
$\|\mathbf{a}\|$	norm of $\mathbf{a}$
$\mathrm{vec}(\mathbf{A})$	stacks $\mathbf{A}$ into $mn \times 1$ vector columnwise
$\mathrm{diag}\{a_1, a_2, \ldots, a_n\}$	$n \times n$ diagonal matrix with $(m, m)^{th}$ element $= a_m$
$r(\mathbf{A})$	rank of the matrix $\mathbf{A}$
$\sigma_k(\mathbf{A})$	k^{th} singular value of matrix $\mathbf{A}$
$\lambda_k(\mathbf{A})$	k^{th} eigenvalue of matrix $\mathbf{A}$ (if Hermitian) or of matrix $\mathbf{A}\mathbf{A}^H$ (if non Hermitian)
$\mathbf{0}_{m \times n}$	$m \times n$ matrix of zeros
$\mathbf{1}_{m \times n}$	$m \times n$ matrix of ones

$\mathbf{I}_m$	$m \times m$ identity matrix
$\lceil x \rceil$	$\lceil x \rceil = \begin{cases} 1 \text{ if } x > 0 \\ 0 \text{ if } x = 0 \end{cases}$
mod x	modulo x operation
$\delta[x]$	Dirac delta (unit impulse) function
$\sharp \tau$	cardinality of the set τ
$\psi(n)$	digamma function, $\psi(n) = \sum_{k=1}^{n-1} \frac{1}{k} - \gamma$, $\gamma \approx 0.57721566$ is Euler's constant
$\mathcal{Q}(x)$	$\mathcal{Q}$-function, $\mathcal{Q}(x) = 1/\sqrt{2\pi} \int_x^\infty e^{-t^2/2} dt$
$E_p(x)$	exponential integral of order p, $E_p(x) = \int_1^\infty e^{-xu} u^{-p} \, du$, $\Re[x] > 0$
$\Gamma(x, y)$	incomplete gamma function, $\Gamma(x, y) = \int_y^\infty u^{x-1} e^{-u} \, du$, $\Re[x] > 0$
$\Gamma(x)$	gamma function, $\Gamma(x) = \Gamma(x, 0) = \int_0^\infty u^{x-1} e^{-u} \, du$, $\Re[x] > 0$
$\binom{n}{p}$	binomial coefficient, $\binom{n}{p} = \frac{n!}{p!(n-p)!}$
$p_x(x)$	probability density function of random variable x
$p_{x_1,\dots,x_n}(x_1, \dots, x_n)$	joint probability density function of random variables $x_1, \dots, x_n$
$\mathcal{M}_x(\tau)$	moment generating function of random variable x, $\mathcal{M}_x(\tau) = \mathcal{E}\{e^{\tau x}\}$

About the Authors

Claude Oestges received the MS degree and the PhD degree in Applied Science from the Université catholique de Louvain (UCL, Louvain-la-Neuve, Belgium), in 1996 and 2000, respectively. From 1996 to 2000, he was an Assistant Lecturer in the Microwave Laboratory UCL. In 2001, he joined (for one year) the Smart Antennas Research Group (Information Systems Laboratory) of Stanford University (California, USA) as a post-doctoral scholar. Since October 2005, Claude Oestges has been a Research Associate of the Belgian National Science Foundation (FRS – Fonds de la Recherche Scientifique) and a part-time Associate Professor at UCL. His research interests cover wireless and satellite communications, with a specific focus on the propagation channel and its impact on system performance. Claude Oestges is the author or co-author of more than 60 papers in IEEE/IEE journals and conference proceedings. He was a member of the IEEE 802.11 Standardization Working Group on '*Multiple antenna channel modeling'*. He received the IEE Marconi Premium Award in 2001 and the IEEE Vehicular Technology Society 2004 Neal Shepherd Award.

Bruno Clerckx received the MS degree and the PhD degree in Applied Science from the Université catholique de Louvain (UCL, Louvain-la-Neuve, Belgium), in 2000 and 2005, respectively. From September 2000 to August 2006, he was with the Microwave Laboratory UCL, as a Research Assistant and a PhD student funded by the Belgian National Science Foundation for Industrial Research (until September 2005) and as a Post-Doctoral Scholar (until August 2006). He held several visiting research positions at the Smart Antennas Research Group (Information Systems Laboratory), Stanford University (CA, USA), and at the Mobile Communication Department of Eurecom Institute (Sophia-Antipolis, France). Bruno Clerckx is currently with the Communication & Network Laboratory of Samsung Advanced Institute of Technology (SAIT) in Korea. He is the author or co-author of more than 25 research papers and communications. He received the IEEE Symposium on Communications and Vehicular Technology Best Student Paper Award in 2002.

This page intentionally left blank

1

Introduction to multi-antenna communications

1.1 Brief history of array processing

Wireless system designers are facing a number of challenges. These include the limited availability of the radio frequency spectrum and a complex space–time varying wireless environment. In addition, there is an increasing demand for higher data rates, better quality of service, and higher network capacity. In recent years, Multiple-Input Multiple-Output (MIMO) systems have emerged as a most promising technology in these measures. MIMO communication systems can be defined intuitively [GSS$^+$03, PNG03] by considering that multiple antennas are used at the transmitting end as well as at the receiving end. The core idea behind MIMO is that signals sampled in the spatial domain at both ends are combined is such a way that they either create effective multiple parallel spatial data pipes (therefore increasing the data rate), and/or add diversity to improve the quality (bit-error rate or BER) of the communication.

Clearly, the benefits from multiple antennas arise from the use of a new dimension – space. Hence, because the spatial dimension comes as a complement to time (the natural dimension of digital communication data), MIMO technology is also known as 'space–time' wireless or 'smart' antennas. Until the 1990s, the use of antenna arrays at one end of the link was mainly oriented to the estimation of directions of arrival as well as diversity, leading to beamforming and spatial diversity. Beamforming is a powerful technique which increases the link signal-to-noise ratio (SNR) through focusing the energy into desired directions. The concept of spatial diversity is that, in the presence of random fading caused by multipath propagation, the SNR is significantly improved by combining the output of decorrelated antenna elements. The early 1990s witnessed new proposals for using antenna arrays to increase the capacity of wireless links, creating enormous opportunities beyond just diversity. It turned out that diversity was only a first step to mitigate multipath propagation. With the emergence of MIMO systems, multipaths were effectively converted into a benefit for the communication system. MIMO indeed takes advantage of random fading, and possibly delay spread, to multiply transfer

rates. Paulraj and Kailath [PK94] introduced a technique for increasing the capacity of a wireless link using multiple antennas at both ends. The prospect of dramatic improvements in wireless communication performance at no cost of extra spectrum was further illustrated in the now famous paper by Telatar [Tel95]. Simultaneously, Bell Labs developed the so-called BLAST architecture [Fos96] that achieved spectral efficiencies up to 10–20 bits/s/Hz, while the first space–time coding architectures appeared [TSC98]. The MIMO success story had begun. Today, MIMO appears as an ideal technology for large-scale commercial wireless products such as wireless local area and third generation networks.

1.2   Space–time wireless channels for multi-antenna systems

Chapters 2 and 3 are dedicated to the modeling of MIMO wireless channels. The term *channel* is usually employed to describe the impulse response of the linear time-varying communication system between a transmitter and a receiver.

Let us first consider Single-Input Single-Output (SISO) transmissions. The digital signal is defined in discrete-time by the complex time series $\{c_k\}_{k \in \mathbb{Z}}$ and is transmitted at the symbol rate T. The transmitted signal is then represented by

$$c(t) = \sum_{l=-\infty}^{\infty} \sqrt{E_s} c_l \delta(t - lT), \tag{1.1}$$

where E_s is the transmitted symbol energy, assuming that the average energy constellation is normalized to unity.

Starting with time-varying SISO channels, we define a function $\mathsf{h}_B(t, \tau)$ as the time-varying (along variable t) impulse response of the channel (along τ) over the system bandwidth $B = 1/T$. This means that $\mathsf{h}_B(t, \tau)$ is the response at time t to an impulse at time $t - \tau$. If the signal $c(t)$ is transmitted, the received signal $\mathsf{r}(t)$ is given by

$$\mathsf{r}(t) = \mathsf{h}_B(t, \tau) \star c(t) + \mathsf{n}(t) \tag{1.2}$$

$$= \int_0^{\tau_{max}} \mathsf{h}_B(t, \tau) c(t - \tau) d\tau + \mathsf{n}(t) \tag{1.3}$$

where $\star$ denotes the convolution product, $\mathsf{n}(t)$ is the additive noise of the system (at the output of the receive filter) and τ_{max} is the maximal length of the impulse response. Note that h_B is a scalar quantity, which can be further decomposed into three main terms

$$\mathsf{h}_B(t, \tau) = w_r(\tau) \star \mathsf{h}(t, \tau) \star w_t(\tau), \tag{1.4}$$

where

- $w_t(\tau)$ is the pulse-shaping filter
- $h(t, \tau)$ is the electromagnetic propagation channel (including the transmit and receive antennas) at time t
- $w_r(\tau)$ is the receive filter.

In digital communication theory [Pro01], it is generally assumed that the cascade $w(\tau) = w_r(\tau) \star w_t(\tau)$ does not create inter-symbol interference when $r(t)$ is sampled at rate T. This is known as the Nyquist criterion and implies that $w(\tau)$ is a Nyquist filter.

Note that, in practice, it is highly difficult to model $h(t, \tau)$, as this would require to sound the channel over an infinite bandwidth. As a consequence, $h_B(t, \tau)$ is usually the modeled quantity, although it is often written as $h(t, \tau)$ through a slight abuse of notation. In what follows, we take on the same notational approximation and write the channel impulse response as $h(t, \tau)$ or $h_t[\tau]$. The input–output relationship therefore reads as

$$r(t) = h(t, \tau) \star c(t) + n(t) \tag{1.5}$$

$$= \int_0^{\tau_{max}} h(t, \tau)c(t - \tau)d\tau + n(t) \tag{1.6}$$

$$= \sum_{l=-\infty}^{\infty} \sqrt{E_s}c_l h_t[t - lT] + n(t). \tag{1.7}$$

Sampling the received signal at the symbol rate T ($r_k = r(t_0 + kT)$, using the epoch t_0) yields

$$r_k = \sum_{l=-\infty}^{\infty} \sqrt{E_s}c_l h_{kT}[t_0 + (k - l)T] + n(t_0 + kT) \tag{1.8}$$

$$= \sum_{l=-\infty}^{\infty} \sqrt{E_s}c_l h_k[k - l] + n_k \tag{1.9}$$

If a communication system is such that the symbol duration T is much larger than the temporal length of the electromagnetic channel τ_{max}, the convolutions in (1.4) cause $h_B(t, \tau)$ to be modeled by a single dependence on t: the channel is then said to be flat fading or narrowband and written as $h_B(t)$ or, using the same abuse of notation as above, as $h(t)$. In the sampled domain, $h_k = h(kT)$ is simply the sampled channel representation of $h(t)$. If the channel is not flat fading, we will observe in Chapter 2 that a frequency selective channel is usually represented as a combination of several narrowband channels shifted in the delay domain (see Section 2.1.7).

As a consequence, we focus now on narrowband channels. A complete model for h_k, given specific transmit and receive locations, is

$$h_k = \alpha \, \beta \, h_k, \tag{1.10}$$

where

- α is a real-valued attenuation term depending only on the distance R between the transmitter and the receiver (otherwise known as the range): this term is known as the path-loss
- β is a real-valued additional attenuation term, which, for a given range, depends on the specific location of the transmitter and the receiver: this random term is known as the shadowing.
- h_k is a complex variable representing the fading term caused by the combination of non-coherent multipaths; by definition of α and β, the fading on a SISO link is a random variable such that $\mathcal{E}\{|h_k|^2\} = 1$.

A number of models of α and β have been developed [Par00, Cor01], and their application is identical for both single- and multi-antenna systems. Therefore, this book will not deal with path-loss and shadowing models, but will focus on multi-antenna fading models. Note that these models are not unimportant, especially when considering the normalization of the transmit or the receive power. Let us simply assume that

- $\alpha \propto R^{-\eta/2}$, η designating the path-loss exponent (for power)
- β is a lognormal variable, i.e. $20 \log_{10}(\beta)$ is a zero-mean normal variable of given standard deviation σ_β.

Hence, rewriting (1.5) in its sampled version without considering path-loss and shadowing, we get

$$y = \sqrt{E_s}hc + n, \tag{1.11}$$

where the time index is removed for better legibility and n is usually taken as white Gaussian distributed, $\mathcal{E}\{n_k n_l^*\} = \sigma_n^2 \delta(k-l)$. The average SNR is then defined as $\rho \triangleq E_s/\sigma_n^2$.

We still have to characterize the fading behavior of h. Assuming that the signal reaches the receiver via a large number of paths of similar energy, the application of the central limit theorem yields that h is a circularly complex variable with zero mean and given variance. This highly popular model has been used by cellular system designers for many years. As detailed in Appendix B, the channel amplitude $s \triangleq |h|$ follows a Rayleigh distribution

$$p_s(s) = \frac{s}{\sigma_s^2} \exp\left(-\frac{s^2}{2\sigma_s^2}\right), \tag{1.12}$$

whose first two moments are

$$\mathcal{E}\{s\} = \sigma_s \sqrt{\frac{\pi}{2}} \tag{1.13}$$

$$\mathcal{E}\{s^2\} = 2\sigma_s^2. \tag{1.14}$$
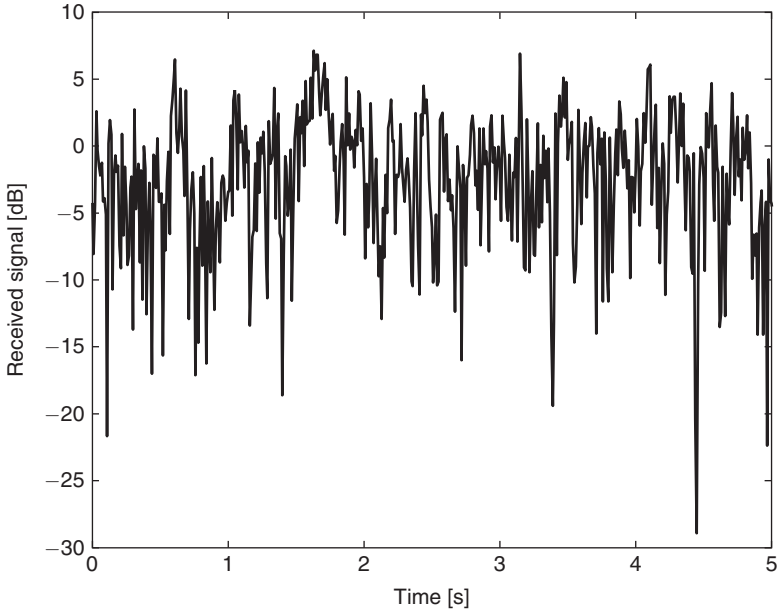
Figure 1.1 *Typical received signal strength in a Rayleigh fading channel*

The phase of h is uniformly distributed over $[0, 2\pi)$. We will come back in Chapter 2 to a more detailed justification of the Rayleigh model. Note, however, that the channel normalization introduced above implies that $2\sigma_s^2 = \mathcal{E}\{|h|^2\} = 1$.

Figure 1.1 illustrates a given realization of a Rayleigh fading channel over a certain time interval. It is interesting to observe that the signal level randomly fluctuates, with some sharp declines of power known as *fades*. Because the noise variance is usually constant, the instantaneous received SNR fluctuates similarly to the channel energy s^2, and may drop dramatically during certain periods. When the channel is in a deep fade, a reliable decoding of the transmitted signal may no longer be possible, resulting in an error. These negative effects of fading are addressed by means of diversity techniques, as detailed in the next section.

In multi-antenna systems, the transmitter and/or the receiver consist of arrays, i.e. they are made of several closely-spaced antennas. The fading channel between each transmit-receive antenna pair can be modeled as a SISO channel. Yet each SISO channel constituting of the MIMO channel may be characterized by a different shadowing (however, the size of the arrays is such that the path-loss is usually identical on all links). For uni-polarized transmissions and reduced inter-element spacings, this is nevertheless a not-so-common situation. If shadowing is identical, the channel matrix $\mathbf{H}$ may be written as follows for a MIMO system with n_t antennas forming the transmit array, and n_r antennas at the receive side. Stacking all inputs and outputs in vectors $\mathbf{c}_k = [c_{1,k}, \ldots, c_{n_t,k}]^T$

and $\mathbf{y}_k = [y_{1,k}, \ldots, y_{n_r,k}]^T$, the input–output relationship at any given time instant k reads as

$$\mathbf{y}_k = \sqrt{E_s}\mathbf{H}_k\mathbf{c}_k + \mathbf{n}_k, \tag{1.15}$$

where

- $\mathbf{H}_k$ is defined as the $n_r \times n_t$ MIMO channel matrix, $\mathbf{H}_k(n, m) = h_{nm,k}$ with h_{nm} denoting the narrowband channel between transmit antenna m ($m = 1, \ldots, n_t$) and receive antenna n ($n = 1, \ldots, n_r$)
- $\mathbf{n}_k = [n_{1,k}, \ldots, n_{n_r,k}]^T$ is the sampled noise vector, containing the noise contribution at each receive antenna, such that the noise is white in both time and spatial dimensions, $\mathcal{E}\{\mathbf{n}_k\mathbf{n}_l^*\} = \sigma_n^2\mathbf{I}_{n_r}\delta(k - l)$.

From now on, we assume that the channel remains constant over a symbol duration, and drop the time index k for better legibility. It must be clear from the context that m and n designate antennas and not sampled time instants. Using the same channel normalization as above, we also have that the average squared Frobenius norm of $\mathbf{H}$ is equal to

$$\mathcal{E}\{\|\mathbf{H}\|_F^2\} = n_t n_r. \tag{1.16}$$

Naturally, modeling only the individual SISO channels is not a complete representation of the multi-antenna channel behavior. The statistical correlations between all matrix elements have to be characterized as well. The following chapters will provide ample information on how to characterize wireless channels, in particular MIMO channels. In the meantime, let us assume that the various elements of the channel matrix are circularly symmetric complex Gaussian independent variables with equal unit variances. The so-called independent identically distributed (i.i.d.) Rayleigh assumption has been traditionally used when developing multi-antenna signal processing techniques, mostly because of its convenience and because it represents an ideal scenario. Hence, we will also use it extensively in this introductory chapter, whereas Chapters 2 and 3 deal with the modeling of correlated and/or non-Rayleigh MIMO channels.

1.3 Exploiting multiple antennas in wireless systems

1.3.1 Diversity techniques

Impact of fading on system performance

As already illustrated by Figure 1.1, the particularity of wireless links is that they are impaired by random fluctuations of the signal level not only across time, but also across space or frequency. This behavior is known as fading, and impacts the performance (in terms of symbol or bit error rate) of any wireless system. As an example, consider the simple case of binary phase-shift keying (BPSK) transmission through a SISO Rayleigh fading channel. In the absence of fading ($h = 1$), the symbol-error rate (SER) in an additive

white Gaussian noise (AWGN) channel is given by

$$\bar{P} = \mathcal{Q}\left(\sqrt{\frac{2E_s}{\sigma_n^2}}\right) = \mathcal{Q}\left(\sqrt{2\rho}\right) \tag{1.17}$$

When fading is considered, the received signal level fluctuates as $s\sqrt{E_s}$. As a result, the error rate is obtained through the following integration [Pro01],

$$\bar{P} = \int_0^\infty \mathcal{Q}\left(\sqrt{2\rho s}\right) p_s(s) \, ds \tag{1.18}$$

where $p_s(s)$ is the fading distribution. In Rayleigh fading, the integration in (1.18) yields

$$\bar{P} = \frac{1}{2}\left(1 - \sqrt{\frac{\rho}{1 + \rho}}\right). \tag{1.19}$$

At large SNR, the error rate in (1.19) simplifies to

$$\bar{P} \cong \frac{1}{4\rho}. \tag{1.20}$$

Strikingly, the error rate decreases only inversely with the SNR (with an asymptotic slope of one). By contrast, the decrease in error rate in non-fading AWGN channels is exponential with the SNR (see (1.17)).

Principle of diversity
To combat the impact of fading on the error rate, diversity techniques are usually employed. The principle of diversity is to provide the receiver with multiple versions of the same transmitted signal. Each of these versions is defined as a diversity branch. If these versions are affected by independent fading conditions, the probability that all branches are in a fade at the same time reduces dramatically. Hence, diversity helps stabilize the link through channel hardening which leads to improved performance in terms of error rate.

Because fading may take place in time, frequency and space, diversity techniques may similarly be exploited in each of these domains. As an example, time diversity can be obtained via appropriate coding and interleaving. Frequency diversity exploits the temporal spreading of the channel (in the τ domain) through equalization techniques [Pro01] or multi-carrier modulations. Naturally, both time and frequency diversity techniques incur a loss in time or bandwidth to allow for the introduction of redundancy. By contrast, spatial or polarization diversity does not sacrifice time and bandwidth, since it is provided by the use of multiple antennas at one or both sides of the link. Yet the spatial dimensions are increased by the use of antenna arrays.

Array and diversity gains

When discussing diversity schemes, two gains are classically introduced. It is important to make a clear distinction between these two gains, as they characterize two different improvements obtained from diversity.

One way of characterizing the merits of a diversity scheme is to evaluate the increase in average output SNR (i.e. at the input of the detector) relative to the single-branch average SNR ρ. Denoting the output SNR as ρ_{out}, we define the *array gain* as

$$g_a \triangleq \frac{\rho_{out}}{\rho},\tag{1.21}$$

which translates by a decrease of the error rate for a fixed transmit power.

A second figure of merit is the increase in the error rate slope as a function of the SNR. We define the *diversity gain* as the negative slope of the log–log plot of the average error probability $\bar{P}$ versus SNR

$$g_d^o(\rho) \triangleq -\frac{\log\left(\bar{P}\right)}{\log\left(\rho\right)}.\tag{1.22}$$

Note that classically, the diversity gain is taken as the asymptotic slope, i.e. for $\rho \to \infty$. Both gains are illustrated in Figures 1.2 and 1.3. Note that the only difference between both graphs is the SNR reference. In Figure 1.2, the error probability is plotted against the
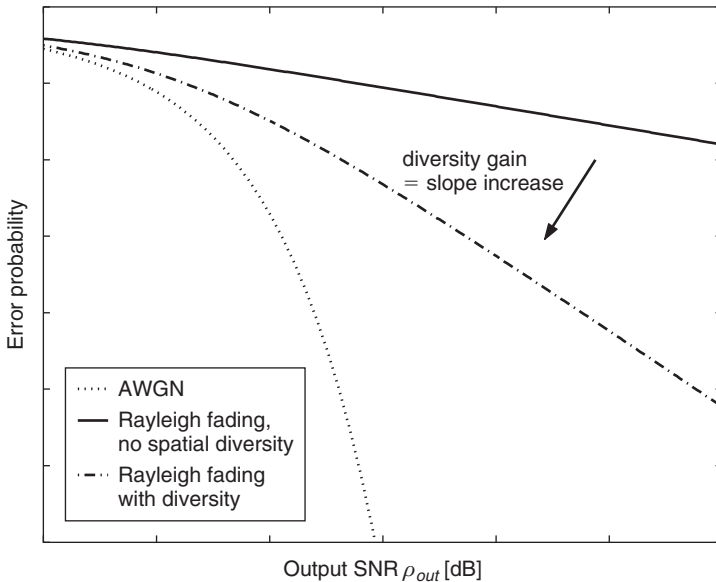
Figure 1.2   *Diversity gain in Rayleigh fading channels*

output average SNR, whereas in Figure 1.3, it is plotted against the single-branch SNR. The diversity curves have exactly the same shape (the slope is the diversity gain) but are shifted from one another by a SNR difference equal to the array gain.

Furthermore, it is important to note that the array gain does not depend on the degree of correlation between the branches, whereas the diversity gain is maximal for independent branches and decreases as the correlation between branches increases.

A third gain which should be differentiated with the above gains is known as the coding gain (see Section 5.3 in Chapter 5 for a formal definition). The latter manifests itself by a shift of the error curve (error rate vs. SNR) to the left. By contrast, we have observed that the diversity gain increases the slope of the error rate curve. Yet it may seem that the coding gain is very similar to the array gain. There is, however, a fundamental difference. If the error rate is plotted against the average receive SNR ρ_{out}, any variation of the array gain is invisible, as already observed in Figure 1.2. This is not the case with the coding gain, as two schemes with different coding gains will yield different (yet parallel) error curves: for a given SNR level ρ_{out} at the input of the detector, the error rates will differ.

1.3.2 Multiplexing capability

When employed at both sides of the link, multiple antennas may also be used to increase the transmission rate (or the *capacity*) of communication systems. In favorable channels,
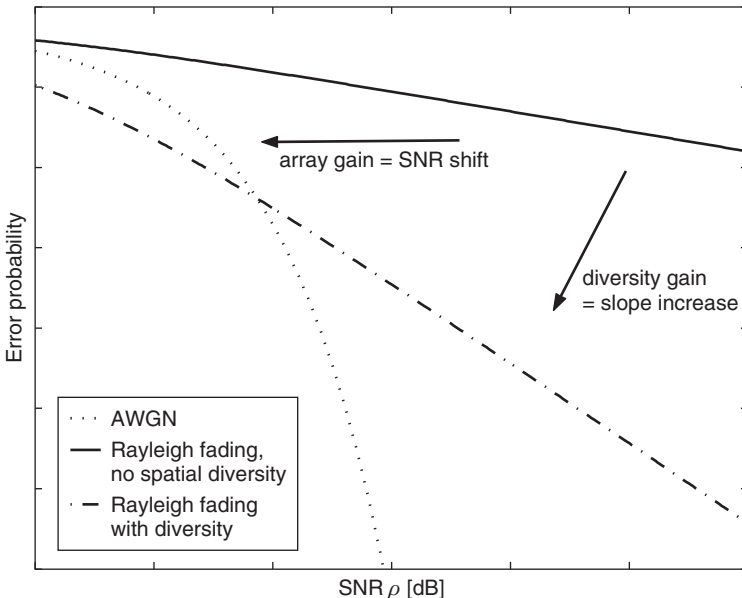
Figure 1.3 *Diversity and array gains in Rayleigh fading channels*

we will show how the rate may be increased proportionally to the minimum number of antenna elements, $\min\{n_t, n_r\}$.

This leverage of multi-antenna systems is known as spatial (or polarization) multiplexing. It is characterized by a so-called multiplexing gain g_s. Asymptotically (at high SNR), this gain is defined as

$$g_s \triangleq \lim_{\rho \longrightarrow \infty} \frac{R(\rho)}{\log_2(\rho)}, \tag{1.23}$$

where $R(\rho)$ is the transmission rate (a more precise definition valid at arbitrary SNR will be given in Chapter 4).

1.4 Single-Input Multiple-Output systems

Single-Input Multiple-Output (SIMO) systems rely on the use of $n_r \geq 2$ antennas at the receiver to achieve diversity. If these multiple antennas are sufficiently spaced (say, by one wavelength), one might expect that the different diversity branches fade independently, provided that the physical channel shows favorable properties (see Chapter 2). Receive diversity may be implemented via two rather different combining methods:

- selection combining: the combiner selects the branch with the highest SNR among the n_r receive signals, which is then used for detection
- gain combining: the signal used for detection is a linear combination of all branches, $z = \mathbf{w}^T \mathbf{y}$, where $\mathbf{w} = [w_1, \ldots, w_{n_r}]^T$ is the combining vector.

Note that we assume in the following that the receiver is able to acquire the perfect knowledge of the channel.

1.4.1 Receive diversity via selection combining

In all generality, selection combining may use other metrics than the SNR (highest absolute power, error rate, etc.), at the cost of a more complex implementation. Assume that the n_r channels are i.i.d. Rayleigh with unit energy and that the noise levels are equal on each antenna. Hence, the selection algorithm comes to compare the instantaneous amplitude of each channel s_n for $n = 1, \ldots, n_r$ and to choose the branch with the largest amplitude $s_{max} = \max\{s_1, \ldots, s_{n_r}\}$. The probability that s_{max} falls below a certain level S is given by

$$P[s_{max} < S] = P[s_1, \ldots, s_{n_r} \leq S],$$
$$= \left[1 - e^{-S^2}\right]^{n_r}, \tag{1.24}$$

using the results of Appendix B. The corresponding distribution of s_{max} is then simply obtained by derivation of (1.24)

$$p_{s_{max}}(s) = n_r \, 2s \, e^{-s^2} \left[1 - e^{-s^2}\right]^{n_r - 1}. \tag{1.25}$$

The average SNR at the output of the combiner ρ_{out} is eventually given by [Jan00]

$$\rho_{out} = \int_0^\infty \rho s^2 p_{s_{max}}(s)\,ds = \rho \sum_{n=1}^{n_r} \frac{1}{n}. \tag{1.26}$$

For large values of n_r, the array gain is approximated as

$$g_a = \sum_{n=1}^{n_r} \frac{1}{n} \cong \gamma + \log(n_r) + \frac{1}{2n_r}, \tag{1.27}$$

where $\gamma \approx 0.57721566$ is Euler's constant. We observe that g_a is of the order of $\log(n_r)$.

The diversity achieved by selection combining can be estimated by computing the error rate using the fading distribution of (1.25). For BPSK and a two-branch diversity, the SER as a function of the average SNR per channel ρ may be written as outlined in [SA00]

$$\begin{aligned}
\bar{P} &= \int_0^\infty \mathcal{Q}\left(\sqrt{2\rho s}\right) p_{s_{max}}(s)\,ds \\
&= \frac{1}{2} - \sqrt{\frac{\rho}{1+\rho}} + \frac{1}{2}\sqrt{\frac{\rho}{2+\rho}}.
\end{aligned} \tag{1.28}$$

At high SNR, we have

$$\bar{P} \cong \frac{3}{8\rho^2}. \tag{1.29}$$

The slope of the bit error rate curve is equal to 2. In general, it can be shown that the diversity gain g_d^o of a n_r-branch selection diversity scheme is equal to n_r, indicating that selection diversity extracts all the possible diversity out of the channel.

1.4.2 Receive diversity via gain combining

In gain combining, the signal z used for detection is a linear combination of all branches

$$z = \mathbf{w}^T \mathbf{y} = \sum_{n=1}^{n_r} w_n y_n, \tag{1.30}$$

where the w_n's are the combining weights and $\mathbf{w} \triangleq [w_1, \ldots, w_{n_r}]^T$. Depending on the choice of these weights, different gain combining methods have been developed. Let us assume that the data symbol c is sent through the channel and received by n_r antennas. Each antenna is characterized by the channel $h_n = |h_n|e^{j\phi_n}$, $n = 1, \ldots, n_r$, assumed to be Rayleigh distributed with unit variance, all the channels being independent. The signals from all antennas are then combined and the detection variable is expressed as

$$z = \sqrt{E_s}\mathbf{w}^T \mathbf{h}c + \mathbf{w}^T \mathbf{n}, \tag{1.31}$$

where $\mathbf{h} \triangleq [h_1, \ldots, h_{n_r}]^T$.

Equal gain combining

The first method, known as equal gain combining (EGC), fixes the weights as $w_n = e^{-j\phi_n}$, which means that signals from the different antennas are co-phased and added together. This method requires the perfect knowledge at the combiner of the signal phases. The post-combiner signal of (1.31) becomes

$$z = \sqrt{E_s} \sum_{n=1}^{n_r} |h_n|c + n',$$

(1.32)

where $n' = \sum_{m=1}^{n_r} n_m e^{-j\phi_m}$ remains a white Gaussian noise.

The mean value of the output SNR ρ_{out} is obtained fairly easily when the channel is Rayleigh distributed

$$
\begin{aligned}
\rho_{out} &= \frac{\mathcal{E}\left\{\left[\sum_{n=1}^{n_r} \sqrt{E_s}|h_n|\right]^2\right\}}{\mathcal{E}\{|\tilde{n}|^2\}} \\
&= \frac{E_s}{n_r \sigma_n^2} \mathcal{E}\left\{\left[\sum_{n=1}^{n_r} |h_n|\right]^2\right\} \\
&= \frac{\rho}{n_r}\left[\mathcal{E}\left\{\sum_{n=1}^{n_r} |h_n|^2\right\} + \sum_{n=1}^{n_r}\sum_{\substack{m=1 \\ m\neq n}}^{n_r} \mathcal{E}\{|h_n|\}\,\mathcal{E}\{|h_m|\}\right] \\
&= \frac{\rho}{n_r}\left[n_r + n_r(n_r - 1)\frac{\pi}{4}\right] \\
&= \rho\left[1 + (n_r - 1)\frac{\pi}{4}\right],
\end{aligned}
$$

(1.33)

where the two last equalities result from the properties of Rayleigh fading channels outlined in (1.13) or (B.2) (see Appendix B). We observe that the array gain grows linearly with n_r, and is therefore larger than the array gain of selection combining. Furthermore, it can also be shown that the diversity gain of equal gain combining is equal to n_r analogous to selection [Yac93].

Maximal ratio combining

A second gain combining method relies on the knowledge of the complex channel gains, so that the weights are chosen as $w_n = h_n^*$. In this case, the post-combining signal reads as

$$z = \sqrt{E_s}\|\mathbf{h}\|^2 c + n',$$

(1.34)

where $n' = \mathbf{h}^H \mathbf{n}$. This scheme is known as maximal ratio combining (MRC), for the reason that it maximizes the output SNR ρ_{out}. The latter is indeed equal to

$$
\begin{aligned}
\rho_{out} &= \frac{1}{\sigma_n^2} \mathcal{E} \left\{ \frac{E_s \|\mathbf{h}\|^4}{\|\mathbf{h}\|^2} \right\} \\
&= \rho \mathcal{E} \left\{ \|\mathbf{h}\|^2 \right\} \\
&= \rho n_r.
\end{aligned}
\tag{1.35}
$$

It is interesting to note that the last equality holds true irrespective of the correlation between the different h_n. In MRC diversity schemes, the array gain g_a is thus always equal to n_r, or equivalently, the output SNR is the sum of the SNR levels of all branches.

As far as the MRC diversity gain is concerned, let us first consider the case of BPSK transmission. Denoting $u = \|\mathbf{h}\|^2$, it is well-known that u follows a χ^2 distribution with $2n_r$ degrees of freedom when the different channels are i.i.d. Rayleigh (see Appendix B)

$$
p_u(u) = \frac{1}{(n_r - 1)!} u^{n_r - 1} e^{-u}.
\tag{1.36}
$$

The symbol error rate is then given by

$$
\bar{P} = \int_0^\infty \mathcal{Q}\left(\sqrt{2\rho u}\right) p_u(u) \, du
\tag{1.37}
$$

$$
= \left[\frac{1 - \sqrt{\rho/(1+\rho)}}{2} \right]^{n_r} \sum_{n=1}^{n_r} \binom{n_r + n - 2}{n - 1} \left[\frac{1 + \sqrt{\rho/(1+\rho)}}{2} \right]^{n-1}
\tag{1.38}
$$

which at high SNR, becomes

$$
\bar{P} = (4\rho)^{-n_r} \binom{2n_r - 1}{n_r}.
\tag{1.39}
$$

We observe that the diversity gain is again equal to n_r.

For alternative constellations, the error probability is given, assuming maximum likelihood (ML) detection [Pro01], by

$$
\bar{P} \approx \int_0^\infty \bar{N}_e \mathcal{Q}\left(d_{min} \sqrt{\frac{\rho u}{2}} \right) p_u(u) \, du,
\tag{1.40}
$$

where $\bar{N}_e$ and d_{min} are, respectively, the number of nearest neighbors and minimum distance of separation of the underlying constellation. The above expression can be solved analytically analogous to (1.38). Yet, it is often useful to apply the Chernoff bound (see Appendix A) to upper-bound the symbol error rate. Indeed, (1.40) may be rewritten as

$$
\bar{P} \approx \bar{N}_e \mathcal{E} \left\{ \mathcal{Q}\left(d_{min} \sqrt{\frac{\rho u}{2}} \right) \right\}
\tag{1.41}
$$

$$
\leq \bar{N}_e \mathcal{E} \left\{ e^{-\frac{d_{min}^2 \rho u}{4}} \right\}.
\tag{1.42}
$$

Since u is a χ^2 variable, the above average upper bound is given (see Appendix B) by

$$\bar{P} \leq \bar{N}_e \prod_{n=1}^{n_r} \frac{1}{1 + \rho d_{min}^2/4}. \tag{1.43}$$

In the high SNR regime, (1.43) simplifies into

$$\bar{P} \leq \bar{N}_e \left(\frac{\rho d_{min}^2}{4} \right)^{-n_r}. \tag{1.44}$$

Analogous to the BPSK case, the diversity gain g_d^o is equal to the number of receive branches in i.i.d. Rayleigh channels.

Minimum mean square error combining

When the noise is spatially correlated or in the presence of non-Gaussian interference, MRC is no longer optimal. An optimal gain combining technique in this case is the minimum mean square error (MMSE) combining [Win84], where the weights are chosen in order to minimize the mean square error between the transmitted symbol c and the combiner output z, i.e.

$$\mathbf{w}^\star = \arg \min_{\mathbf{w}} \mathcal{E} \left\{ |\mathbf{w}^T \mathbf{y} - c|^2 \right\}. \tag{1.45}$$

It can easily be shown that the optimal weight vector $\mathbf{w}^\star$ is given by

$$\mathbf{w}^\star = \mathbf{R}_{NI}^{-1} \mathbf{h}^*, \tag{1.46}$$

where $\mathbf{R}_{NI}$ is the correlation matrix of noise and interference. In the absence of interference, $\mathbf{R}_{NI} = \mathcal{E}\{\mathbf{nn}^H\}$. Furthermore, when the noise is spatially white (across the antennas), $\mathbf{R}_{NI} = \sigma_n^2 \mathbf{I}_{n_r}$ and MMSE diversity reduces to MRC diversity up to a scaling factor.

1.4.3   Receive diversity via hybrid selection/gain combining

A hybrid method combines the selection algorithm with MRC [WW99]. At each time instant, the receiver first selects the n_r' branches (out of n_r) with the largest SNR, and combines them using the MRC algorithm. The scheme is also known as generalized selection.

It can be shown that the average SNR at the output of the combiner is remarkably the sum of two terms. The first one corresponds to a n_r'-branch MRC. The second term results from the generalization of (1.26) to the selection of n_r'-branches out of n_r. The total array gain reads therefore as

$$g_a = n_r' + n_r' \sum_{n=n_r'+1}^{n_r} \frac{1}{n}. \tag{1.47}$$

Similarly, to selection combining ($n'_r = 1$), the diversity gain of the hybrid selection/MRC (HS/MRC) scheme is equal to n_r and not to n'_r [AS00, WBS$^+$03]. The advantage of HS/MRC is then clear: the full diversity is extracted, yet only n'_r RF chains are needed at the receiver.

1.5 Multiple-Input Single-Output systems

Multiple-Input Single-Output (MISO) systems exploit diversity at the transmitter through the use of n_t transmit antennas in combination with pre-processing or precoding. A significant difference with receive diversity is that the transmitter might not have the knowledge of the MISO channel. Indeed, at the receiver, the channel is easily estimated. This is not the case at the transmit side, where feedback from the receiver is required to inform the transmitter. There are basically two different ways of achieving *direct transmit diversity*:

- when the transmitter has a perfect channel knowledge, beamforming can be performed using various optimization metrics (SNR, SINR, etc.) to achieve both diversity and array gains
- when the transmitter has no channel knowledge, pre-processing known as space–time coding is used to achieve a diversity gain, but no array gain.

In this section, we evaluate different beamformers and introduce a very simple space–time coding technique known as the Alamouti scheme. We also present a few *indirect transmit diversity* techniques, which convert spatial diversity to time or frequency diversity.

1.5.1 Switched multibeam antennas

Before dealing with true beamformers, it is worth mentioning that a first and simple solution for the downlink is to implement a switching technique between fixed beams. Multiple non-overlapping beams are formed to cover a given angular sector. First, a sounding procedure determines the strongest uplink signal in each of the beams. In the downlink, the same beam is then used for transmission. If the chosen beam remains the best beam for a sufficiently long period of time, the control algorithm can be very simple. However, this technique assumes that the up and downlinks are reciprocal.

The major drawback of the method lies in its reduced array and diversity gains in non line-of-sight scenarios (in LOS scenarios, $g_a = n_t$ but there is no diversity to be exploited).

1.5.2 Transmit diversity via matched beamforming

This beamforming technique, also known as transmit MRC, assumes that the transmitter has perfect knowledge of the channel. To exploit diversity, the signal c is weighted

adequately before being transmitted on each antenna. At the receiver, the signal reads as

$$y = \sqrt{E_s}\mathbf{h}\mathbf{w}c + n, \tag{1.48}$$

where $\mathbf{h} \triangleq [h_1, \ldots, h_{n_t}]$ represents the MISO channel vector, and $\mathbf{w}$ is the weight vector. The choice that maximizes the receive SNR is given by [God97]

$$\mathbf{w} = \frac{\mathbf{h}^H}{\|\mathbf{h}\|}, \tag{1.49}$$

where the denominator guarantees that the average total transmit energy remains equal to E_s. This choice comes to transmit along the direction of the matched channel, hence it is also known as *matched beamforming* or *conventional beamforming*. Similarly, to receive MRC the average output SNR is equal to $\rho_{out} = n_t\rho$, hence the array gain is equal to the number of transmit antennas. The diversity gain is also equal to n_t as the symbol error rate is upper-bounded at high SNR by

$$\bar{P} \leq \bar{N}_e \left(\frac{\rho d_{min}^2}{4}\right)^{-n_t}. \tag{1.50}$$

Therefore, matched beamforming presents the same performance as receive MRC, but requires perfect transmit channel knowledge, which implies feedback from the receiver in time-duplexing systems. If frequency duplexing is used, the channel reciprocity may no longer be ensured, compromising the quality of the channel knowledge at the transmitter. Furthermore, while matched beamforming is optimal in the absence of interfering signals, it does not offer any protection against interference.

Matched beamforming may also be combined with a selection algorithm, analogous to the generalized selection algorithm described above for SIMO systems. The transmitter selects n_t' antennas out of n_t to be used in the beamformer. Quite naturally, this technique achieves a full diversity gain of n_t, but a reduced transmit array gain.

1.5.3   Null-steering and optimal beamforming

We have seen that matched beamforming maximizes the SNR in AWGN channels. Yet, in the presence of co-channel interference, matched beamforming is not optimal anymore. A first solution [God97] is to use a *null-steering beamformer*. This technique is able to produce nulls in up to $n_t - 1$ directions, i.e. it can cancel the interfering signal in the directions of $n_I \leq n_t - 1$ interfered users. However, the weights estimated by this scheme do not maximize the output SNR at the intended user. This limitation is overcome by the use of an *optimal beamformer* [God97], which trades the signal power delivered to the intended user for interference generated to the n_I other users. The transmit diversity offered at the intended user by these two schemes is naturally equal to $n_t - n_I$.

1.5.4 Transmit diversity via space–time coding

The beamforming techniques described above all required the knowledge of the channel at the transmitter to derive the optimal weights. By contrast, Alamouti has developed a particularly simple but ingenious transmit diversity scheme for two transmit antennas, known as the Alamouti scheme, which does not require transmit channel knowledge [Ala98]. Consider that two symbols c_1 and c_2 are transmitted simultaneously from antennas 1 and 2 during the first symbol period, followed by symbols $-c_2^*$ and c_1^*, transmitted from antennas 1 and 2 during the next symbol period. Assume that the flat fading channel remains constant over the two successive symbol periods, and is denoted by $\mathbf{h} = [h_1 \ h_2]$ (note that the subscripts here denote the antenna number and not the symbol periods). The symbol y_1 received at the first symbol period is

$$y_1 = \sqrt{E_s} h_1 \frac{c_1}{\sqrt{2}} + \sqrt{E_s} h_2 \frac{c_2}{\sqrt{2}} + n_1, \tag{1.51}$$

and the symbol y_2 received at the second symbol period is

$$y_2 = -\sqrt{E_s} h_1 \frac{c_2^*}{\sqrt{2}} + \sqrt{E_s} h_2 \frac{c_1^*}{\sqrt{2}} + n_2, \tag{1.52}$$

where each symbol is divided by $\sqrt{2}$ so that the vector $\mathbf{c} = \begin{bmatrix} c_1/\sqrt{2} & c_2/\sqrt{2} \end{bmatrix}$ has a unit average energy (assuming that c_1 and c_2 are drawn from a unit average energy constellation) and n_1 and n_2 are the additive noise contributions at each symbol period (here, the subscripts denote the symbol periods and not the antennas). We may express the combination of (1.51) and (1.52) as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \sqrt{E_s} \underbrace{\begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}}_{\mathbf{H}_{\mathit{eff}}} \underbrace{\begin{bmatrix} c_1/\sqrt{2} \\ c_2/\sqrt{2} \end{bmatrix}}_{\mathbf{c}} + \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix}. \tag{1.53}$$

We observe that the two symbols are spread over two antennas and over two symbol periods, hence $\mathbf{H}_{\mathit{eff}}$ appears as a *space–time* channel. Applying the matched filter $\mathbf{H}_{\mathit{eff}}^H$ to the received vector $\mathbf{y}$ effectively decouples the transmitted symbols as shown below

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \mathbf{H}_{\mathit{eff}}^H \begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \sqrt{E_s} \left[|h_1|^2 + |h_2|^2 \right] \mathbf{I}_2 \begin{bmatrix} c_1/\sqrt{2} \\ c_2/\sqrt{2} \end{bmatrix} + \mathbf{H}_{\mathit{eff}}^H \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix}$$

$$\mathbf{z} = \sqrt{E_s} \|\mathbf{h}\|^2 \mathbf{I}_2 \mathbf{c} + \mathbf{n}', \tag{1.54}$$

where $\mathbf{n}'$ is such that $\mathcal{E}\{\mathbf{n}'\} = \mathbf{0}_{2 \times 1}$ and $\mathcal{E}\{\mathbf{n}'\mathbf{n}'^H\} = \|\mathbf{h}\|^2 \sigma_n^2 \mathbf{I}_2$. The average output SNR is thus equal to

$$\rho_{out} = \frac{1}{\sigma_n^2} \mathcal{E} \left\{ \frac{E_s \left[\|\mathbf{h}\|^2 \right]^2}{2\|\mathbf{h}\|^2} \right\}$$

$$= \rho, \tag{1.55}$$

illustrating that the Alamouti scheme does not provide any array gain owing to the lack of transmit channel knowledge (remember that $\mathcal{E}\{\|\mathbf{h}\|^2\} = n_t = 2$).

However, for i.i.d. Rayleigh channels, it can be shown, as above, that the average symbol error rate at high SNR can be upper-bounded according to

$$\bar{P} \leq \bar{N}_e \left(\frac{\rho d_{min}^2}{8} \right)^{-2}, \tag{1.56}$$

i.e. the diversity gain is equal to $n_t = 2$ analogous to transmit MRC, despite the lack of transmit channel knowledge. Globally, the Alamouti scheme performs worse than transmit or receive MRC owing to the null array gain. As observed in Figure 1.4 (for BPSK modulation), it clearly incurs a 3dB loss when the SER is plotted against the branch SNR. Furthermore, analogous to MRC, the full diversity is only achieved in independent Rayleigh fading channels. When the spatial channels h_1 and h_2 are correlated (e.g. take $h_1 = h_2$ as an extreme case), the Alamouti scheme performs similarly to (1.20).

The above example is restricted to $n_t = 2$. Yet, transmit diversity techniques not requiring transmit channel knowledge may be designed for $n_t > 2$, and are known under the general denomination of space–time codes, for they spread symbols over both space (antennas) and time. They can also be generalized to MIMO systems as shown in Section 1.6.2. Space–time coding is also reviewed in great details in Chapter 5.
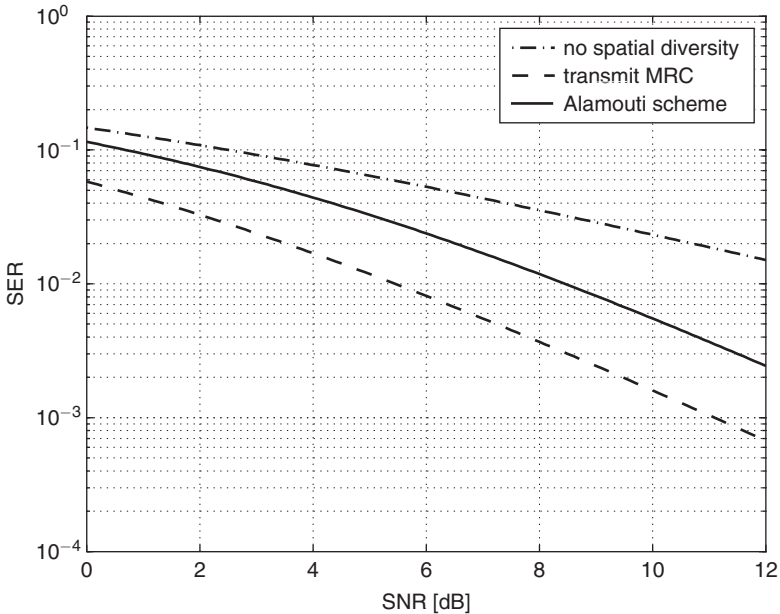
Figure 1.4 *Performance of transmit MRC and Alamouti schemes with two transmit antennas in i.i.d. Rayleigh fading channels (for BPSK modulation)*

1.5.5 Indirect transmit diversity

So far, we have presented direct transmit diversity techniques, where the spatial diversity is extracted based on combining or space–time coding. It is also possible to convert spatial diversity to time or frequency diversity, which are then exploited using well-known SISO techniques.

Assume that $n_t = 2$ and that the signal on the second transmit branch is either delayed by one symbol period [SW94], either phase-rotated [HAN92] through a well-chosen frequency-shift. If the channels h_1 and h_2 are i.i.d. Rayleigh, the spatial diversity (i.e. the use of two antennas) is converted into frequency and time diversity, respectively. Indeed, the effective SISO channel resulting from the addition of the two branches seen by the receiver now fades over frequency or time, and this selective fading can be exploited by conventional diversity techniques, e.g. FEC/interleaving for frequency diversity [PNG03, BG-PvdV06]. Note that the first method is widely known as delay-diversity and detailed further in Chapter 5.

1.6 Multiple-Input Multiple-Output systems

With multiple antennas at both ends of the link comes the ability to exploit other leverages than diversity and array gains – it is now possible to increase the transmission throughput via the spatial multiplexing capability of MIMO channels. However, we will also observe that it is not possible to maximize both the spatial multiplexing and the diversity gains. This trade-off will be further formalized in Chapter 4. Similarly, we will show that the array gain in Rayleigh channels is also limited and is actually smaller than $n_t n_r$. In this section, we classify MIMO techniques with respect to the quality of channel knowledge at the transmitter. A similar classification is used in this book: Chapters 5 to 7 deal with space–time coding techniques requiring no transmit channel knowledge, whereas Chapter 8 is dedicated to techniques available when the transmitter has a partial (e.g. stochastic) knowledge of the channel.

1.6.1 MIMO with perfect transmit channel knowledge

Dominant eigenmode transmission

Analogous to the previous sections, let us first focus on maximizing the diversity gain of a $n_r \times n_t$ MIMO system. Intuitively, this can be done through transmitting the same signal from all transmit antennas after weighting by a $n_t \times 1$ vector $\mathbf{w}_t$. At the receive array, the antenna outputs are combined into a scalar signal z through a weighted summation according to a $n_r \times 1$ vector $\mathbf{w}_r$. Subsequently, the transmission is described by

$$\mathbf{y} = \sqrt{E_s}\mathbf{H}\mathbf{w}_t c + \mathbf{n}, \tag{1.57}$$

$$z = \mathbf{w}_r^H \mathbf{y},$$

$$= \sqrt{E_s}\mathbf{w}_r^H \mathbf{H}\mathbf{w}_t c + \mathbf{w}_r^H \mathbf{n}. \tag{1.58}$$

Maximizing the receive SNR comes to maximizing $\left\| \mathbf{w}_r^H \mathbf{H} \mathbf{w}_t \right\|_F^2 / \left\| \mathbf{w}_r \right\|_F^2$. To solve this problem, we need to use the singular value decomposition (SVD) of $\mathbf{H}$ as

$$\mathbf{H} = \mathbf{U_H} \mathbf{\Sigma_H} \mathbf{V_H}^H, \tag{1.59}$$

where $\mathbf{U_H}$ and $\mathbf{V_H}$ are $n_r \times r(\mathbf{H})$ and $n_t \times r(\mathbf{H})$ unitary matrices, $r(\mathbf{H})$ being the rank of $\mathbf{H}$ and

$$\mathbf{\Sigma_H} = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_{r(\mathbf{H})}\} \tag{1.60}$$

is the diagonal matrix containing the singular values of $\mathbf{H}$. Using this particular decomposition of the channel matrix, it is easily shown [PNG03] that the receive SNR is maximized when $\mathbf{w}_t$ and $\mathbf{w}_r$ are the transmit and receive singular vectors corresponding to the maximum singular value of $\mathbf{H}$, $\sigma_{max} = \max\{\sigma_1, \sigma_2, \ldots, \sigma_{r(\mathbf{H})}\}$. This technique is known as the dominant eigenmode transmission, and (1.58) may be rewritten as

$$z = \sqrt{E_s} \sigma_{max} c + \tilde{n}, \tag{1.61}$$

where $\tilde{n} = \mathbf{w}_r^H \mathbf{n}$ has a variance equal to σ_n^2.

From (1.61), it is easily observed that the array gain is equal to $\mathcal{E}\{\sigma_{max}^2\} = \mathcal{E}\{\lambda_{max}\}$, where λ_{max} is the largest eigenvalue of $\mathbf{H}\mathbf{H}^H$. The array gain for i.i.d. Rayleigh channels is thus bounded as follows

$$\max\{n_t, n_r\} \leq g_a \leq n_t n_r. \tag{1.62}$$

Furthermore, the results in Appendix B indicate that in the i.i.d. Rayleigh case, the asymptotic array gain of a dominant eigenmode transmission (i.e., for large n_t, n_r) is given by

$$g_a = \left(\sqrt{n_t} + \sqrt{n_r}\right)^2. \tag{1.63}$$

Finally, the diversity gain is obtained by upper and lower-bounding the error rate at high SNR [PNG03] (assuming that the Chernoff bound is a good approximation of the SER at high SNR)

$$\bar{N}_e \left(\frac{\rho d_{min}^2}{4\min\{n_t, n_r\}}\right)^{-n_t n_r} \geq \bar{P} \geq \bar{N}_e \left(\frac{\rho d_{min}^2}{4}\right)^{-n_t n_r}. \tag{1.64}$$

The above equation implies that the error rate maintains a slope of $n_t n_r$ as a function of the SNR: the dominant eigenmode transmission extracts a full diversity gain of $n_t n_r$.

Dominant eigenmode transmission with antenna selection

The dominant eigenmode transmission introduced above may be generalized to include a selection algorithm at either the transmitter or the receiver. We focus in the following on transmit selection, but all considerations remain valid if the receiver performs the selection (remember that both the transmitter and the receiver have a perfect channel knowledge).

The dominant eigenmode transmission with antenna selection [MWCW05] works as follows. We define a set of matrices $\mathbf{H}'$ created by removing $n_t - n_t'$ columns from $\mathbf{H}$. The set of all possible $\mathbf{H}'$ is denoted as $\mathcal{S}\{\mathbf{H}'\}$ and its cardinality is $\binom{n_t}{n_t'}$. At each time instant, the scheme comes to perform a dominant eigenmode transmission using the matrix $\mathbf{H}'$ offering the largest $\sigma'_{\max} = \max\{\sigma'_1, \sigma'_2, \ldots, \sigma'_{r(\mathbf{H}')}\}$. The output SNR thereby reads as

$$\rho_{out} = \rho \max_{\mathcal{S}\{\mathbf{H}'\}} \{\sigma'_{\max}\}. \tag{1.65}$$

The average SNR can be calculated as described in [GP02], yielding the following array gain

$$g_a = \sum_{k=n_t'}^{n_t-n_t'+1} X_k \tag{1.66}$$

with

$$X_k = \frac{n_t!}{(k-1)!\,(n_t-k)!\,(n_r-1)!} \sum_{l=0}^{k-1} \left[(-1)^l \binom{k-1}{l} \right.$$

$$\left. \times \sum_{m=0}^{(n_r-1)(n_t-k+l)} a_{n_t-k+l} \frac{\Gamma\,(1+n_r+m)}{(n_t-k+l+1)^{1+n_r+m}} \right] \tag{1.67}$$

where a_s is the coefficient of u^m in the expansion of $\sum_{i=0}^{n_r-1} \left(u^i/u! \right)^s$.

Analogous to the classical dominant eigenmode transmission, it is shown in [GP02] that antenna selection extracts the same diversity gain as if all n_t transmit antennas were used, i.e. the diversity gain is equal to $n_t n_r$.

Multiple eigenmode transmission

The dominant eigenmode transmission naturally achieves no multiplexing gain as the same symbol is sent over all transmit antennas. As an alternative, one may desire to increase the system throughput by maximizing the spatial multiplexing gain. To this end, symbols are spread over all non-zero eigenmodes of the channel. We assume in the following that $n_r \geq n_t$ and that the channel matrix is i.i.d. Rayleigh with an SVD given by (1.59). If the transmitter multiplies the input vector $\mathbf{c}$ ($n_t \times 1$) using $\mathbf{V_H}$ as a precoding matrix, and the receiver multiplies the received vector by $\mathbf{U_H^H}$, the effective input–output relationship reads as

$$\mathbf{y} = \sqrt{E_s}\mathbf{U_H^H H V_H c} + \mathbf{U_H^H n}$$

$$= \sqrt{E_s}\mathbf{\Sigma_H c} + \tilde{\mathbf{n}}. \tag{1.68}$$

We observe that the channel has been decomposed into n_t parallel SISO channels given by $\{\sigma_1, \ldots, \sigma_{n_t}\}$. What should be noticed is that all these channels are totally decoupled as

if n_t virtual data pipes had been created. The mutual information of the MIMO channel is therefore the sum of the SISO channel capacities

$$\mathcal{I} = \sum_{k=1}^{n_t} \log_2\left(1 + \rho p_k \sigma_k^2\right), \tag{1.69}$$

where $\{p_1, \ldots, p_{n_t}\}$ is the power allocation on each of the channel eigenmodes, normalized such that $\sum_{k=1}^{n_t} p_k = 1$. We will show in Chapter 4 that the MIMO capacity is obtained by finding the optimal power allocation maximizing the mutual information of (1.69), and the capacity scales linearly in n_t, hence the spatial multiplexing gain is equal to n_t. By contrast, this transmission does not necessarily achieve the full diversity gain of $n_t n_r$ but does at least provide n_r-fold array and diversity gains (still assuming $n_t \leq n_r$).

In general, the capacity scales linearly with the rank of $\mathbf{H}$ in arbitrarily correlated channels, as analyzed in detail in Chapter 4. One consequence is that in highly correlated channels, only the dominant eigenmode is used for transmission, reducing the spatial multiplexing gain to one. In such channels, there is of course no available diversity gain, though a MIMO array gain of $\mathcal{E}\{\sigma_{max}^2\}$ is obtained.

Naturally, the multiple eigenmode transmission may be combined with antenna selection at the receiver. As long as $n_r' \geq n_t$, the multiplexing gain remains equal to n_t, but the array and diversity gains are reduced.

Finally, hybrid schemes based on both multiple and dominant eigenmode transmissions can also be used. As an example, it is always possible to achieve some diversity by grouping antennas subsets for diversity gain and operate a multiplexing on the new channel with reduced dimension [HL05].

1.6.2   MIMO without transmit channel knowledge

When the transmitter has no channel knowledge, the presence of multiple antennas at both sides may allow to extract diversity and/or increase the capacity. This is achieved through the use of so-called space–time codes, which expand symbols over the antennas (i.e. over space) and over time. In the following, we introduce space–time block and trellis codes, which will be analyzed in more detail in Chapters 5 and 7.

Space–time block coding
Let us start with a simple example by extending the Alamouti scheme of Section 1.5.4 to MIMO 2×2 transmissions.

Analogous to the MISO case, consider that two symbols c_1 and c_2 are transmitted simultaneously from transmit antennas 1 and 2 during the first symbol period, while symbols $-c_2^*$ and c_1^* are transmitted from antennas 1 and 2 during the next symbol period.

Assume that the flat fading channel remains constant over the two successive symbol periods, and that the 2×2 channel matrix reads as

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \tag{1.70}$$

Note that the subscripts here denote the receive and transmit antenna index and not the symbol period. The vector signal received at the receive array at the first symbol period is

$$\mathbf{y}_1 = \sqrt{E_s} \mathbf{H} \begin{bmatrix} c_1/\sqrt{2} \\ c_2/\sqrt{2} \end{bmatrix} + \mathbf{n}_1, \tag{1.71}$$

and the vector signal received at the second symbol period is

$$\mathbf{y}_2 = \sqrt{E_s} \mathbf{H} \begin{bmatrix} -c_2^*/\sqrt{2} \\ c_1^*/\sqrt{2} \end{bmatrix} + \mathbf{n}_2, \tag{1.72}$$

where $\mathbf{n}_1$ and $\mathbf{n}_2$ are the additive noise contributions at each symbol period over the receive antenna array (so the subscripts here denote the symbol periods, and not the antennas). The receiver forms a combined signal vector $\mathbf{y}$ as

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2^* \end{bmatrix} = \underbrace{\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \\ h_{12}^* & -h_{11}^* \\ h_{22}^* & -h_{21}^* \end{bmatrix}}_{\mathbf{H}_{\mathit{eff}}} \underbrace{\begin{bmatrix} c_1/\sqrt{2} \\ c_2/\sqrt{2} \end{bmatrix}}_{\mathbf{c}} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2^* \end{bmatrix}. \tag{1.73}$$

Analogous to the MISO system, both symbols c_1 and c_2 are spread over the two transmit antennas and over the two symbol periods. Furthermore, $\mathbf{H}_{\mathit{eff}}$ is orthogonal for all channel realizations, i.e. $\mathbf{H}_{\mathit{eff}}^H \mathbf{H}_{\mathit{eff}} = \|\mathbf{H}\|_F^2 \mathbf{I}_2$. If we compute $\mathbf{z} = \mathbf{H}_{\mathit{eff}}^H \mathbf{y}$, we get

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \mathbf{H}_{\mathit{eff}}^H \mathbf{y} = \|\mathbf{H}\|_F^2 \mathbf{I}_2 \, \mathbf{c} + \mathbf{n}', \tag{1.74}$$

where $\mathbf{n}'$ is such that $\mathcal{E}\{\mathbf{n}'\} = \mathbf{0}_{2 \times 1}$ and $\mathcal{E}\{\mathbf{n}' \mathbf{n}'^H\} = \|\mathbf{H}\|_F^2 \, \sigma_n^2 \mathbf{I}_2$. The above equation illustrates that the transmission of c_1 and c_2 is fully decoupled, i.e.

$$z_k = \sqrt{E_s/2} \, \|\mathbf{H}\|_F^2 c_k + \tilde{n}_k \qquad k = 1, 2 \tag{1.75}$$

with the average output SNR given by

$$\rho_{out} = \frac{1}{\sigma_n^2} \mathcal{E} \left\{ \frac{E_s \left[\|\mathbf{H}\|_F^2 \right]^2}{2\|\mathbf{H}\|_F^2} \right\}$$

$$= 2\rho, \tag{1.76}$$

illustrating that the Alamouti scheme in a 2×2 configuration provides a receive array gain ($g_a = n_r = 2$) but no transmit array gain (since the transmitter has no channel knowledge).
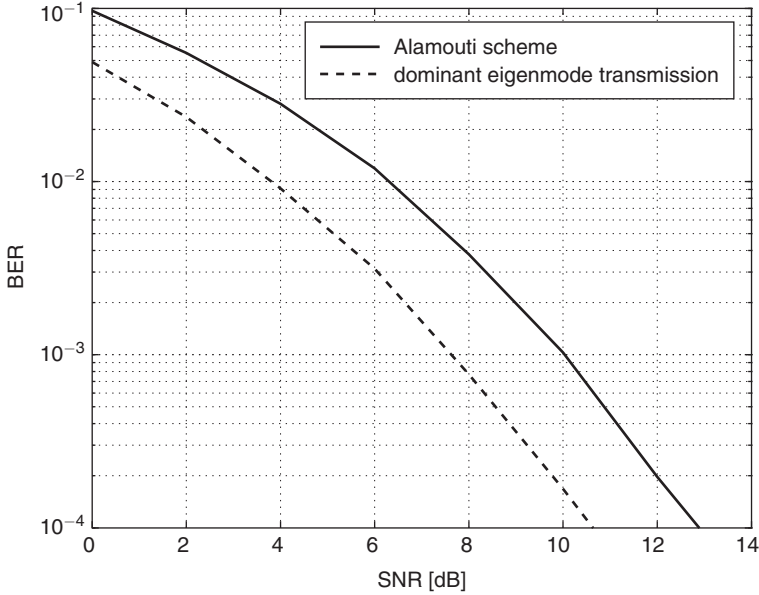
Figure 1.5 *Performance of dominant eigenmode and Alamouti transmissions in a 2×2 i.i.d. Rayleigh fading channel (with QPSK modulation)*

However, it may extract the full diversity ($g_d^o = n_t n_r = 4$) as shown in [PNG03],

$$\bar{P} \leq \bar{N}_e \left(\frac{\rho d_{min}^2}{8} \right)^{-4}. \tag{1.77}$$

A comparison between the Alamouti scheme and the dominant eigenmode transmission is depicted in Figure 1.5 as a function of the SNR. Clearly, the diversity gain is equal to 4 in both cases, but the array gain is larger by 3 dB for the dominant eigenmode transmission. Note that the Alamouti scheme may also be used with any number of receive antennas ($g_a = n_r$ and $g_d^o = 2n_r$), but cannot be applied to systems with more than two transmit antennas.

Figure 1.5 has illustrated the possibility of extracting the full diversity of the MIMO channel without transmit channel knowledge. The principle of spreading symbols over space and time is generalized through the concept of space–time block codes (STBCs). In general, these map Q symbols onto a codeword $\mathbf{C}$ of size $n_t \times T$, where T is thus the duration of the codewords. The codeword $\mathbf{C}$ is usually normalized such that $\mathcal{E}\{\mathrm{Tr}\{\mathbf{CC}^H\}\} = T$. As an example, the 2×2 Alamouti scheme ($T = 2, n_t = 2, Q = 2$) is represented by the following codeword matrix

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix}. \tag{1.78}$$

The spatial multiplexing rate of a space–time block code is then defined as $r_s \triangleq \frac{Q}{T}$, and a space–time block code is full-rate when $r_s = n_t$. The Alamouti scheme is therefore characterized by $r_s = 1$.

A first class of STBCs is constituted by the orthogonal STBCs (O-STBCs), which include the Alamouti scheme described above. O-STBCs transmit one or less independent symbol per symbol period over the n_t transmit antennas. They provide an array gain of n_r and extract the full diversity gain of $n_t n_r$. Furthermore, they allow for a direct detection since vector detections are converted into much less complex scalar detections, as illustrated by (1.75). However, for complex constellations, O-STBCs with $r_s = 1$ only exist for $n_t = 2$. Otherwise, complex O-STBCs for arbitrary n_t offer spatial multiplexing rates $r_s < 1$.

Instead of extracting the full diversity with O-STBCs, it is possible to transmit n_t independent symbols per symbol period, achieving a spatial multiplexing rate of n_t. Such full-rate schemes are known as Spatial Multiplexing (SM). For uncoded SM (a.k.a. V-BLAST, standing for Vertical-Bell Laboratories Layered Space–Time), each codeword expands onto one symbol duration (there is no temporal encoding): the achievable array and diversity gains with ML decoding are equal to n_r, while the spatial multiplexing gain equals $\min\{n_t, n_r\}$. By contrast, coded SM transmissions such as D-BLAST (Diagonal-BLAST) may deliver the full diversity ($g_d^o = n_t n_r$) by optimally coding over time. Note however that these gains are extremely sensitive to the detection algorithm. As an example, we will show in Chapter 5 that the diversity gain of uncoded SM with zero-forcing or MMSE detection is only $n_r - n_t + 1$ (assuming $n_r \geq n_t$). Naturally, it is also possible to combine Spatial Multiplexing with receive antenna selection (assuming $n_r \geq n_t$). The capacity loss induced by the selection is discussed in Chapter 4.

So far, we have seen that O-STBCs exploit the full diversity but with a limited spatial multiplexing rate. On the other hand, uncoded Spatial Multiplexing enables higher throughput but does not succeed in leveraging transmit diversity. If an increase of the receiver complexity is authorized, it is possible to improve the data rates while still providing transmit diversity. A first step was made in that direction by D-BLAST. Alternatively, this can be realized by so-called linear dispersion codes, which appear as an intermediate solution between O-STBCs and SM. Finally, codes known as algebraic codes have been developed with the same objective in mind. Given their relative complexity, the analysis of linear dispersion and algebraic codes is presented in Chapter 5.

A final remark: because space–time codes do not require the channel knowledge at the transmitter, they should be designed in such a way that their performance is not too sensitive to the actual channel state (at any time instant). Designing robust space–time codes is a topic in itself which is dealt with extensively in Chapters 6 and 7.

Space–time trellis coding

Space–time trellis codes (STTCs) were actually proposed in [TSC98] before STBCs. They are an extension of classical convolutional codes to multi-antenna transmit arrays. The

difference with STBCs lies in the fact that the encoder output in space–time trellis coding is not only a function of the input bits but also of the state of the encoder, which is related to the previous input bits. This memory is inherent to the trellis approach and provides an additional coding gain (see Chapter 6).

Space-frequency coding

In frequency selective channels, it is possible to exploit the additional frequency diversity by coding not only across space (i.e. across antennas) but also across the frequency band, e.g. using orthogonal frequency division multiplexing (OFDM). This technique, known as SF MIMO-OFDM, together with alternative coding schemes over frequency selective channels, is considered in detail in Chapter 9.

1.6.3   MIMO with partial transmit channel knowledge

The exploitation of the array gain may also be possible if the transmitter has only a partial channel knowledge. Perfect channel knowledge at the transmitter has been covered in Section 1.6.1, but requires a high rate feedback link between the receiver and the transmitter to keep the latter continuously informed about the channel state. By contrast, exploiting only the channel statistics or a quantized version of the channel at the transmitter requires a much lower rate feedback link.

Precoding techniques generally consist in combining a multi-mode beamformer spreading the codewords in orthogonal directions related to the channel distribution with a constellation shaper, or more simply, a power allocation scheme. There are naturally many similarities with the various eigenmode transmissions of Section 1.6.1, the difference being that the eigenbeams are now based on the statistics of $\mathbf{H}$ rather than on the instantaneous value of $\mathbf{H}$.

Similarly, antenna selection techniques may rely only on partial channel knowledge, choosing transmit or receive antennas based on the first and second-order statistics of $\mathbf{H}$ [GP02]. Intuitively, this comes to choose the antenna pairs with the lowest correlation. Naturally, such a technique does not minimize the instantaneous error performance, but only the average error rate. As a result, it leads mostly to a coding gain and small diversity advantage.

A generalization of antenna selection consists of exploiting a limited amount of feedback at the transmitter through quantized precoding. This technique relies on a codebook of precoding matrices, i.e. a finite set of precoders, designed off-line and known to both the transmitter and receiver. The receiver estimates the best precoder as a function of the current channel and then feeds back the index of the best precoder in the codebook.

All these techniques are dealt with comprehensively in Chapter 8.

1.7 Multiple antenna techniques in commercial wireless systems

In this last section, we briefly examine current or future commercial implementations of MIMO techniques. Specifically, multiple antennas have been integrated into Third Generation (3G) cellular systems, and broadband fixed/mobile wireless access networks (IEEE 802.16e) also known as WiMax. Finally, MIMO will be included into the future IEEE 802.11n release.

Regarding 3G systems, the current *cdma2000* standard provides transmit diversity options, via an extension of the Alamouti scheme. The wideband CDMA Universal Mobile Telephony Standard (UMTS) and its future versions developed by the third generation partnership project (3GPP) also allow for implementing space–time transmit diversity schemes, in combination with transmit beamforming at the base station.

In wireless metropolitan area networks (WMAN), the IEEE 802.16d/e standard also known as *WiMax* makes use of MIMO-OFDMA techniques, combining multiple antennas with OFDM modulation. The IEEE 802.16e standard operates in non line-of-sight between 2 and 11 GHz. Data rates of around 40 and 15 Mbps will be available for fixed and mobile users respectively.

In wireless local area networks (WLAN), the IEEE 802.11n (*Wi-Fi*) standard considers peak data rates of 150 Mbps, with 500 Mbps as an option. MIMO technology is implemented in the standard using three different techniques: antenna selection, space–time coding (e.g. the Alamouti scheme) and possibly beamforming (though advanced beamforming techniques requiring transmit channel knowledge are likely to be introduced only in medium-term products).

This page intentionally left blank

2

Physical MIMO channel modeling

As seen in Chapter 1, many signal processing techniques have been developed with very simple assumptions regarding physical wireless channels, including multi-antenna channels. In particular, we have so far considered that the elements of the MIMO channel matrix are uncorrelated variables. In practical scenarios, these assumptions may be far from realistic. Yet the characteristics of radio propagation environments dictate the ultimate performance of wireless systems, irrespective of the assumptions used to design the system. Therefore, the goal of Chapters 2 and 3 is to investigate realistic models and analytical representations of wireless channels for MIMO systems.

When dealing with MIMO channels, space comes as an additional dimension, which needs to be modeled on its own in the same way as time and frequency variations have been modeled for wideband SISO systems. This multi-dimensionality should always be kept in mind when modeling MIMO channels or when designing space-time codes. As an example, let us consider that the angular distribution of the energy has to be modeled at both the transmit and receive sides. This leads to a so-called *double-directional* description:

- the term *directional* indicates that the channel description includes a model of the angular distribution of the energy at the antennas, by contrast to a non-directional model, which deals only with the temporal spreading;
- the term *double* means that the spatial description concerns both sides, i.e. that there are multiple antennas at transmit and receive sides.

Naturally, the more usual notions, such as the channel stationarity and the channel fading dynamics, must also be considered from a spatial perspective. On the one hand, stationarity addresses the question of how long the channel statistical behavior remains essentially constant. On the other hand, the channel dynamic properties (represented by the so-called Doppler spectrum) indicate how fast the stationary channel parameters change over time. These two topics are critical when the signal processing requires the knowledge of the channel at the transmitter.

We will find that an important aspect of MIMO channel modeling consists of describing the correlations between pairs of single-antenna channels. These correlations, which are directly linked to the directional spread of the energy, are largely responsible for the performance loss of space-time codes. Therefore, significant parts of Chapter 2 and especially Chapter 3 are dedicated to correlation modeling.

More philosophically, multiple antennas also challenge the way we should consider the propagation channel. With MIMO techniques, signal processing is able to benefit from the channel, rather than having to remove its impact. The consequence is that channel models have to be accurate, avoiding conservative rules without being too optimistic.

In this chapter, we develop the theoretical background behind MIMO channel modeling and we review a large number of physical models as well as a few standardized channel models. In doing so, we will try to highlight the main advantages and weaknesses of each model, identifying the required assumptions and input parameters. We will also dedicate several sections to multi-polarized channels and to antenna coupling models.

2.1   Multidimensional channel modeling

2.1.1   *The double-directional channel impulse response*

In any wireless network, signals arrive at a receiver via various propagation mechanisms, and the existence of multiple scattered paths (multipaths) with different time-varying time delays, directions of departure and arrival, phases and attenuations gives rise to a highly complex transmission channel. The possible propagation mechanisms can be categorized into five basic mechanisms: (i) free-space (or line-of-sight, LOS) propagation, (ii) transmission (and absorption), (iii) specular reflection, (iv) diffraction, and (v) diffusion (also known as diffuse scattering), as well as any combination between the last four. Except for the line-of-sight, all these mechanisms imply the interaction of the propagating wave with one or more arbitrary obstacles (walls, trees, cars, human beings, etc.), which we generally refer to as *scatterers* or interacting objects. Figure 2.1 illustrates this multipath propagation concept. The line-of-sight path experiences free-space loss only. Specular reflection occurs when a propagating electromagnetic wave impinges upon a plane and smooth surface whose dimensions are very large compared to the wavelength. Diffraction appears when the path between the transmitter and the receiver is obstructed by a discontinuity, such as an edge, wedge or cylinder. Transmission through an obstacle also causes partial absorption of energy. Finally, diffusion is caused by interactions of the wave with objects the dimensions of which are on the order of the wavelength, e.g. foliage or rough obstacles. Note that, in this case, the resulting scattered wave is most often non-coherent: its phase is not a deterministic quantity and is therefore only known in a stochastic manner [Ish98].

Neglecting path-loss and shadowing (see Chapter 1), the local radio propagation channel between a transmitter located at $\mathbf{p}_t$ and a receiver located in the three-dimensional space
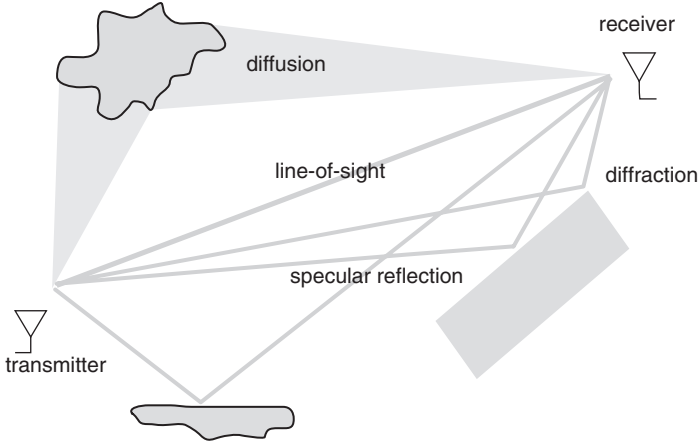
Figure 2.1 *A typical multipath scenario*

at $\mathbf{p}_r$ can therefore be expressed as a double-directional *impulse response* [SMB01] based on n_s paths

$$h(\mathbf{p}_t, \mathbf{p}_r, \tau, \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_r) = \sum_{k=0}^{n_s-1} h_k\left(\mathbf{p}_t, \mathbf{p}_r, \tau, \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_r\right), \qquad (2.1)$$

where τ, $\boldsymbol{\Omega}_t$, and $\boldsymbol{\Omega}_r$ respectively, represent the delay, the direction of departure (DoD) and the direction of arrival (DoA) (both in 3-D space). The directional vectors are defined as follows. At the transmit side, $\boldsymbol{\Omega}_t$ is uniquely determined by its spherical coordinates (i.e. the azimuth Θ_t and elevation ψ_t) on a sphere of unit radius according to the relationship [Fle00]

$$\boldsymbol{\Omega}_t = [\cos\Theta_t \sin\psi_t, \sin\Theta_t \sin\psi_t, \cos\psi_t]^T \qquad (2.2)$$

The DoA (at the receiver) is defined analogously.

Arbitrary waveforms can be modeled by choosing appropriate functions for $h_k(\mathbf{p}_t, \mathbf{p}_r, \tau, \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_r)$. However, since Maxwell's equations are linear, arbitrary waveforms may alternatively be obtained by superposition of a number of plane waves. In the case of a plane wave, $h_k(\mathbf{p}_t, \mathbf{p}_r, \tau, \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_r)$ can be expressed as

$$h_k\left(\mathbf{p}_t, \mathbf{p}_r, \tau, \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_r\right) \triangleq \alpha_k\, e^{j\phi_k}\, \delta\left(\tau - \tau_k\right)\, \delta(\boldsymbol{\Omega}_t - \Omega_{t,k})\, \delta(\boldsymbol{\Omega}_r - \Omega_{r,k}), \qquad (2.3)$$

where

- α_k is the amplitude of the kth contribution
- ϕ_k is the phase of the kth contribution
- τ_k is the time delay of the kth contribution

- $\Omega_{t,k}$ is the DoD of the kth contribution
- $\Omega_{r,k}$ is the DoA of the kth contribution.

The values of τ, Ω_t, and Ω_r are determined by the relative location of the transmitter, receiver, and scatterer. That means that, in Figure 2.1, one would assign values of τ, Ω_t, and Ω_r for the line-of-sight, and for each of the reflected, diffracted and diffuse components.

When the transmitter and/or receiver and/or scatterers are moving, the above variables are in all generality time-variant. This also concerns the number of physical contributions n_s, as some interactions may disappear while new interactions appear. Hence, over long periods of time (or large distances), some of the above variables may change drastically. Yet the rates at which these variables vary may significantly differ. A compact notation for the time-variant double-directional channel is obtained as

$$h\left(t, \tau, \Omega_t, \Omega_r\right) = \sum_{k=0}^{n_s(t)-1} h_k\left(t, \tau, \Omega_t, \Omega_r\right), \qquad (2.4)$$

where $h_k(t, \tau, \Omega_t, \Omega_r)$ is expressed in the same way as in (2.3). Note that all temporal variations (i.e. the variations of the transmit and receive locations, and those of the scatterers) are grouped into a unique temporal dependence under the variable t. For this reason, $\mathbf{p}_t$ and $\mathbf{p}_r$ have been dropped in (2.4).

The double-directional channel impulse response is thus a function of four parameters (time, delay, directions of departure and arrival):

- **time**: this dependence represents the variation of the channel with time or, similarly, with the motion of the receiver (related with time via the receiver speed v); ϕ_k usually changes rapidly over one wavelength λ (i.e. during λ/v), while n_s, τ_k, $\Omega_{t,k}$, $\Omega_{r,k}$, and possibly α_k may remain constant over longer distances (i.e. periods of time);
- **(time) delay**: at any given time t, the channel (i.e. the received power) varies as a function of the delay τ, since each contribution arrives with a delay proportional to its path length; note that the scale of τ is typically much smaller than that of t;
- **directions of departure and arrival**: each contribution departs from the transmitter and arrives at the receiver in certain directions, so the channel also depends on Ω_t and Ω_r; we can interpret this as a directional distribution of the energy at both sides of the link.

We emphasize that the description of (2.4) only represents the propagation channel, and is independent of antenna patterns and system bandwidth. In other words, (2.4) assumes isotropic antennas at both sides, and an infinite system bandwidth. If antennas and filtering over a given bandwidth B are accounted for, (2.4) becomes

$$h_{a,B}\left(t, \tau, \Omega_t, \Omega_r\right) = w_r(\tau) \star \left[g_r\left(\Omega_r\right)\, h(t, \tau, \Omega_t, \Omega_r)\, g_t(\Omega_t)\right] \star w_t(\tau), \qquad (2.5)$$

where $g_t(\mathbf{\Omega}_t)$ and $g_r(\mathbf{\Omega}_r)$ are the complex field-patterns of the transmit and receive antennas, respectively, and $w_t(\tau)$ and $w_r(\tau)$ are the transmit and receive filters. We will return to the description of band-limited channels in Section 2.1.7. Note already that

- all channel measurements (and therefore models) are by nature antenna-dependent and band-limited, although the measurement/modeling bandwidth is typically much higher than the actual communication system bandwidth B
- in a very convenient abuse of notations, Chapters 4 and above therefore denote by h the global channel $h_{a,B}$.

Besides the directional impulse response, alternative functions may be used to describe the multidimensional radio channel. In non-directional models, it is classical to characterize the channel in the different Fourier domains, using frequency variables instead of the temporal variables t and τ [Par00, Pro01]. This characterization is based on a transfer function, whose parameters are the Doppler frequency ν (in the Fourier domain, ν is the frequency variable corresponding to t) and the frequency f (corresponding to τ):

- the Doppler frequency ν is related to the *time selectivity*, and is linked to the maximal rate at which the instantaneous channel changes
- the variations over f measure the *frequency selectivity* over the transmission bandwidth.

In double-directional models, the following functions are the direct extensions of Bello's system functions [Pro01] to include the direction domains:

- $h(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$ is the time-variant direction-resolved impulse response
- $T(t, f, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$ is the time-variant direction-resolved transfer function
- $B(\nu, f, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$ is the Doppler-variant direction-resolved transfer function
- $S(\nu, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$ is the Doppler-variant direction-resolved impulse response.

Furthermore, both directional (angular) domains can similarly be seen as spatial frequency domains at the transmit and receive sides. The correspondence with the antenna or aperture domains also results from a Fourier transform. Figure 2.2 illustrates this concept, restricting the discussion to

- the transmit side
- the 2-dimensional space: the direction vector $\mathbf{\Omega}_t$ can be replaced by the scalar azimuth angle Θ_t, defined with respect to the link axis, i.e. the line joining the transmitter to the receiver
- a linear transmit array: because of this geometry, the array aperture can be represented in one dimension relative to the wavelength, which we denote as l_t/λ.

It is well-known from antenna theory that the directional channel at given time and frequency $T(\sin \Theta_t)$ is related to the spatial channel variation along the transmitting
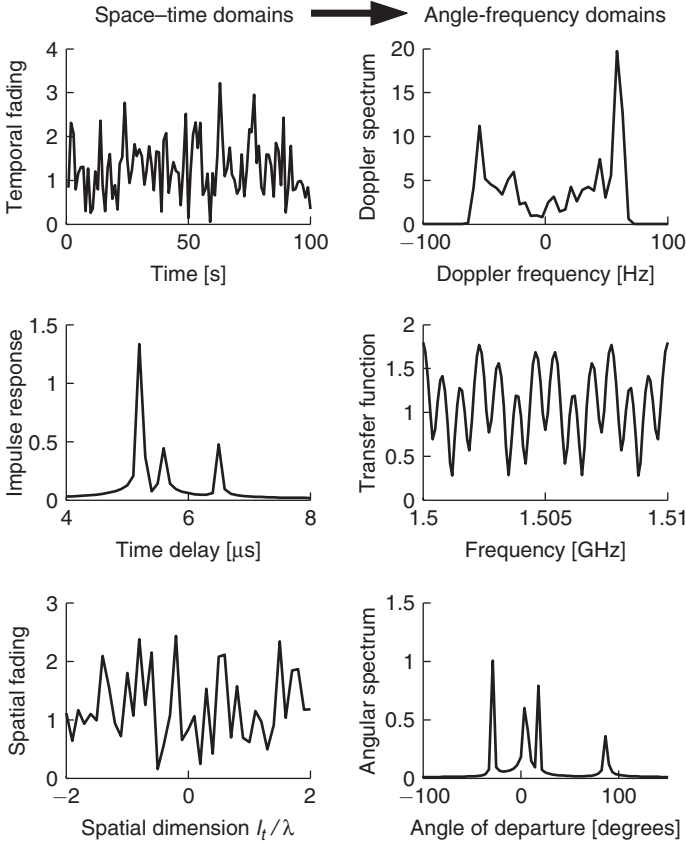
Figure 2.2   *Extension of Bello's functions to the spatial and angular domains*

aperture $T_a(l_t/\lambda)$ [Bal05, Kat02] through the Fourier transform

$$T(\sin \Theta_t) = \int_{-\infty}^{+\infty} T_a(l_t/\lambda) \, e^{j2\pi(l_t/\lambda)\sin \Theta_t} \, d(l_t/\lambda),$$

$$T_a(l_t/\lambda) = \int_{-\infty}^{+\infty} T(\sin \Theta_t) \, e^{-j2\pi(l_t/\lambda)\sin \Theta_t} \, d(\sin \Theta_t). \tag{2.6}$$

In (2.6), the description of the channel in the $\sin \Theta_t$-domain is simply the spectrum of the spatial channel variation along the antenna aperture. Note that the Fourier relationship is not directly between the antenna and the angular domains, although there exists a non-linear monotonic relationship between l_t and Θ_t (so we will use Θ_t instead of $\sin \Theta_t$). We also use this example to define an alternative angular domain through $\varphi_t = 2\pi(l_t/\lambda)\sin \Theta_t$. The usefulness of this definition will appear later in Chapter 3.

The above reasoning can be easily extended to

- the receive array
- the full 3-dimensional space (with azimuth and elevation angles)
- non-linear arrays (although the physical interpretation of the spatial frequency is not as direct as above) in generalized aperture domains measured along vectors ℓ_t and ℓ_r.

As a consequence, the Fourier relationships enable us to define sixteen four-dimensional system functions

- in the time t or Doppler frequency ν domains
- in the delay τ or frequency f domains
- in the transmit antenna ℓ_t or DoD $\mathbf{\Omega}_t$ domains
- in the receive antenna ℓ_r or DoA $\mathbf{\Omega}_r$ domains.

Hence, functions similar to h, T, B and S may be expressed in the transmit and receive antenna domains. As an example, the time-variant aperture-resolved transfer function $T_a(t, f, \ell_t, \ell_r)$ will play an important role in MIMO channel modeling.

2.1.2 Multidimensional correlation functions and stationarity

It is naturally not realistic to characterize the time-variant channel with an infinite precision: the precise location of the transmitter, receiver and scatterers are not known, so that ϕ_k can be considered as a random variable, even locally. On a larger scale, the same holds true for n_s, τ_k, $\Omega_{t,k}$, $\Omega_{r,k}$ and α_k. As a consequence, time-variant directional channels can be considered as random variables, so that the transfer functions become stochastic processes. Their statistical characterization necessitates the knowledge of their multidimensional joint probability density functions. This would require a great deal of effort, which is unlikely to happen in practice. A more convenient approach consists of modeling only the multidimensional correlation functions. As an example the correlation function of the time-variant direction-resolved impulse response is given by

$$R_h\left(t, t', \tau, \tau', \mathbf{\Omega}_t, \mathbf{\Omega}_t', \mathbf{\Omega}_r, \mathbf{\Omega}_r'\right) = \mathcal{E}\left\{h\left(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right), h^*\left(t', \tau', \mathbf{\Omega}_t', \mathbf{\Omega}_r'\right)\right\}, \qquad (2.7)$$

while the Doppler-delay transmit-receive direction correlation is outlined by

$$R_S\left(\nu, \nu', \tau, \tau', \mathbf{\Omega}_t, \mathbf{\Omega}_t', \mathbf{\Omega}_r, \mathbf{\Omega}_r'\right) = \mathcal{E}\left\{S\left(\nu, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right), S^*\left(\nu', \tau', \mathbf{\Omega}_t', \mathbf{\Omega}_r'\right)\right\}. \qquad (2.8)$$

This description is accurate for Gaussian channels, since, in this case, the mean and covariance provide a statistically complete description. For non-Gaussian channels, it is only an approximate but nevertheless useful representation.

Finally, the correlation function of the time-variant aperture-resolved transfer function is of particular interest in MIMO studies, and reads as

$$R_{T_a}\left(t, t', f, f', \ell_t, \ell_t', \ell_r, \ell_r'\right) = \mathcal{E}\left\{T_a\left(t, f, \ell_t, \ell_r\right), T_a^*\left(t', f', \ell_t', \ell_r'\right)\right\}. \qquad (2.9)$$

Note that the expectation operator in (2.7)–(2.9) represents an ensemble average over all possible channel realizations. Expressions in (2.7)–(2.9) are still very complex, so further assumptions might be needed in terms of stationarity in one or more dimensions [Her04, PNG03].

Definition 2.1   *A process is called wide-sense stationary (WSS) with respect to time when the time correlation only depends on the time difference $\Delta t = t - t'$. In the Doppler domain, the Cramer-Loève theorem [Pro01] implies that the correlation function can be expressed as a delta-function of the Doppler frequency difference, $R_S \propto \delta(\Delta v)$, i.e., signals arriving with different Doppler frequencies are uncorrelated.*

Definition 2.2   *A process is called uncorrelated scattering (US) when the frequency correlation only depends on the frequency difference $\Delta f = f - f'$. In the delay domain, this implies by analogy that the corresponding correlation function can be expressed as a delta-function of the delay difference, $R_S \propto \delta(\Delta \tau)$. Signals arriving with different delays are therefore uncorrelated.*

Definition 2.3   *A directional process is called homogeneous with respect to space when the spatial correlation only depends on the spatial difference at both transmit and receive sides, $\Delta \ell_t = \ell_t - \ell'_t$ and $\Delta \ell_r = \ell_r - \ell'_r$. In the angular domain, this implies that the corresponding correlation function can be expressed as a delta-function of the direction difference, and that signals departing/arriving from/to different directions are therefore uncorrelated.*

Definition 2.4   *A directional process is called wide-sense stationary uncorrelated scattering homogeneous (WSSUSH) when the Doppler-delay transmit-receive direction correlation function is white in all dimensions*

$$R_S\left(v, v', \tau, \tau', \mathbf{\Omega}_t, \mathbf{\Omega}'_t, \mathbf{\Omega}_r, \mathbf{\Omega}'_r\right) = \mathcal{C}_S\left(v, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) \delta(\Delta v)\, \delta(\Delta \tau)\, \delta(\Delta \mathbf{\Omega}_t)\, \delta(\Delta \mathbf{\Omega}_r), \quad (2.10)$$

where $\mathcal{C}_S(v, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$ is defined as the Doppler-delay joint direction power spectrum. Signals with different delays, Doppler-shifts, DoDs and DoAs are uncorrelated. The corresponding time-frequency transmit-receive antenna correlation function therefore only depends on shifts in all dimensions, but no longer on the absolute values.

Full stationarity in all dimensions (WSSUSH) is a very restrictive condition, particularly in the transmit and receive antenna domains. Indeed, antenna stationarity implies that the correlation only depends on the antenna spacing, but not on the precise location of the antennas within an array. While this may be achieved easily in uniform linear arrays (ULAs), it is far from granted in arbitrary antenna arrays. Also, this requires all antennas to share the same polarization. Hence, it is sensible to introduce a partial stationary concept, separating time and frequency on one side (i.e. using the classical WSSUS assumption) and antennas on the other side. We will come back to this notion through the definition of

the spatial correlation matrix, as well as at the end of Chapter 3 where we deal with non-stationarity measures. From now on, we assume that channels are WSSUSH (including antenna stationary), which is often a valid assumption for ULAs.

Finally, WSSUSH processes are usually assumed to be ergodic so an ensemble average at any time t can be accurately estimated by a temporal average. That further implies that random generators can be used to simulate time series of the channel, e.g. assuming complex Gaussian statistics (see Chapter 3).

2.1.3 Channel fading, K-factor and Doppler spectrum

For a stationary channel, further essential aspects of wireless networks are the channel fading statistics and dynamics. Consider the narrowband field amplitude of a WSSUSH channel

$$h(t) \triangleq \iiint h(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r) \, d\tau \, d\mathbf{\Omega}_t \, d\mathbf{\Omega}_r \qquad (2.11)$$

The term *fading* describes the variation of the local channel $h(t)$ due to the varying phases and amplitudes of scatterers. One may further make a distinction between slow and fast fading processes, but this is a more system-wise approach, which is dealt with in Chapter 4. Several statistical distributions are generally used to describe the fading behavior of $|h(t)|$: Rayleigh, Rice, Nakagami, Weibull, etc. [Par00]. The first two distributions are easily interpreted within a physical perspective, as detailed below.

Rayleigh fading

When the number of independent scatterers n_s is sufficiently large (theoretically infinite), and all scattered contributions are non-coherent and of approximately equal energy, then by the central limit theorem $h(t)$ is a complex Gaussian variable with zero mean and independent quadrature components. The distribution of $|h(t)| \triangleq s(t)$ is then given by the Rayleigh probability density function (already used in Chapter 1)

$$p_s(s) = \frac{s}{\sigma_s^2} \exp\left(-\frac{s^2}{2\sigma_s^2}\right), \qquad (2.12)$$

where $2\sigma_s^2$ is the average power of $s(t)$.

Ricean fading

The presence of a non-fading or coherent component leads to a Ricean distribution. A coherent field component is characterized by the absence of any fading over the stationary period. What is a coherent contribution in practice? Here, we need to make a distinction between mobile and fixed wireless scenarios.

In mobile scenarios, the coherent contribution is often associated with the LOS component. Since any scattered contribution is generally the combination of smaller

micro-contributions clustered together, any motion of the receiver will modify the phases ϕ_k associated with all micro-contributions. So the global scattered contribution will see its phase fade over time. Still, some strong specular reflections may also appear as coherent over the stationarity period. Therefore, the coherent contribution may sometimes represent more than just the LOS.

By contrast, in fixed wireless access, many scattered contributions are coherent, as transmitter, receiver and many scatterers are static. Only a few scattered contributions are non-coherent; typically, those characterized with small motions, e.g. under the effect of the wind blowing in the trees. In such scenarios, equating the coherent contribution exclusively with the LOS is thus far from correct.

The relative strength of the dominant coherent component is characterized by an important parameter, known as the K-factor K. As the channel contains a coherent component, its amplitude can be written as

$$|h(t)| = \bar{h} + \tilde{h}(t), \tag{2.13}$$

where $\bar{h}$ is the amplitude of the coherent component, and $\tilde{h}(t)$ is the non-coherent part, whose energy is denoted as $2\sigma_s^2$. The K-factor is defined as

$$K = \frac{|\bar{h}|^2}{2\sigma_s^2}. \tag{2.14}$$

As a consequence of (2.13), $|h(t)| \triangleq s'$ is Ricean distributed, and its distribution is given, as a function of K, as

$$p_{s'}(s') = \frac{2s'K}{|\bar{h}|^2} \exp\left[-K\left(\frac{s'^2}{|\bar{h}|^2} + 1\right)\right] I_0\left(\frac{2s'K}{|\bar{h}|}\right). \tag{2.15}$$

Doppler spectrum
The channel dynamic behavior is usually represented by the Doppler spectrum. The latter is defined for a stationary channel as

$$S(\nu) = \iiint \mathcal{C}_S\left(\nu, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) d\tau \, d\mathbf{\Omega}_t \, d\mathbf{\Omega}_r$$

$$= \mathcal{E}\left\{\left|\iiint S\left(\nu, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) d\tau \, d\mathbf{\Omega}_t \, d\mathbf{\Omega}_r\right|^2\right\}. \tag{2.16}$$

Naturally, the Doppler spectrum is also the power spectral density of $h(t)$ and the Fourier transform of the channel temporal auto-correlation function $R_h(\Delta_t)$

$$R_h(\Delta_t) = \iiint R_h\left(\Delta_t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) d\tau \, d\mathbf{\Omega}_t \, d\mathbf{\Omega}_r. \tag{2.17}$$
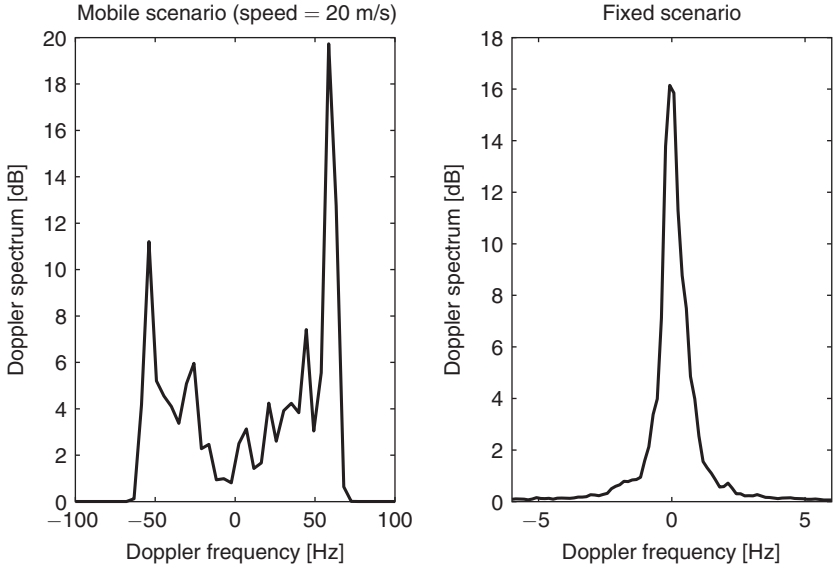
Figure 2.3 *Typical Doppler spectra for mobile and fixed scenarios*

The Doppler spectrum is limited by the ratio of the maximal speed of the transmitter/ receiver to the wavelength in mobile scenarios [Par00]. For fixed access networks, the Doppler spectrum appears to be a two-sided exponential function, with a strong peak around 0 Hz [BGN⁺00]. Typical Doppler spectra at 900 MHz are depicted in Figure 2.3 for both cases.

Assume now that the angular domains are reduced to the azimuthal domain and that the receiving antenna is a half-wave dipole. If equi-energy scatterers are uniformly distributed around the receiver so that Θ_r is uniformly distributed over $[0, 2\pi)$, the Doppler spectrum reduces in the absence of coherent component to the so-called Jakes spectrum

$$S(v) = \frac{3}{2\pi v_m \sqrt{1 - \frac{v^2}{v_m^2}}} \text{ for } |v| < v_m \tag{2.18}$$

and $S(v) = 0$ for $|v| \geq v_m$, with $v_m = f_c v/c$ (f_c is the carrier frequency and v is the receiver's speed). We actually observe that the mobile scenario of Figure 2.3 presents such a typical U-shape of the so-called Jakes spectrum, with $v_m = 60$ Hz.

The coherence time of the channel is defined as the time over which the channel is assumed to be constant. It is naturally inversely proportional to the width of the Doppler spectrum. For the Jakes spectrum of (2.18), a good approximation of the coherence time is given by

$$T_{coh} = \frac{9}{19\pi v_m}. \tag{2.19}$$

Comparing the coherence time to the symbol/frame duration T provides two widely used concepts:

- if $T \ll T_{coh}$, the channel is said to be slow fading, and approximated as constant over T
- if $T \gg T_{coh}$, the channel is said to be fast fading (or simply time-varying over T), so that the complete channel distribution is explored over the symbol/frame duration.

2.1.4 Power delay and direction spectra

We now introduce several power spectra, which are largely used in channel modeling. To obtain those spectra, let us consider an ergodic WSSUSH channel and take $\Delta t = 0$ in (2.7), so that at any given time t, the correlation function, may be rewritten as

$$R_h\big(t, t', \tau, \tau', \mathbf{\Omega}_t, \mathbf{\Omega}'_t, \mathbf{\Omega}_r, \mathbf{\Omega}'_r\big)_{WSSUSH,\, t=t'} = \mathcal{P}_h\big(\tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\, \delta(\Delta\tau)\, \delta(\Delta\mathbf{\Omega}_t)\, \delta(\Delta\mathbf{\Omega}_r), \tag{2.20}$$

where $\mathcal{P}_h\big(\tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)$ represents the power-delay joint direction spectrum of the channel at any time t as a direct interpretation of (2.7),

$$\mathcal{P}_h\big(\tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big) = \mathcal{E}\big\{\big|h\big(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\big|^2\big\}. \tag{2.21}$$

In (2.21) it is important to note that $\mathcal{E}$ represents an ensemble average, independent of t because the channel is stationary, but also equivalent to a temporal average by ergodicity.

Naturally, marginal power spectra may be further defined [Fle00]:

- the power-delay spectrum

$$\mathcal{P}(\tau) = \mathcal{E}\left\{\left|\underbrace{\iint h\big(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\, d\mathbf{\Omega}_t\, d\mathbf{\Omega}_r}_{\triangleq h(t,\tau)}\right|^2\right\},$$

$$= \int \int \mathcal{P}_h\big(\tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\, d\mathbf{\Omega}_t\, d\mathbf{\Omega}_r, \tag{2.22}$$

- the joint direction power spectrum

$$\mathcal{A}(\mathbf{\Omega}_t, \mathbf{\Omega}_r) = \mathcal{E}\left\{\left|\int h\big(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\, d\tau\right|^2\right\},$$

$$= \int \mathcal{P}_h\big(\tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\, d\tau, \tag{2.23}$$

- the transmit direction power spectrum

$$\mathcal{A}_t(\mathbf{\Omega}_t) = \mathcal{E}\left\{\left|\iint h\big(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\, d\tau\, d\mathbf{\Omega}_r\right|^2\right\},$$

$$= \int \int \mathcal{P}_h\big(\tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\big)\, d\tau\, d\mathbf{\Omega}_r, \tag{2.24}$$

- the receive direction power spectrum

$$
\mathcal{A}_r(\mathbf{\Omega}_r) = \mathcal{E}\left\{ \left| \iint h(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)\, d\tau\, d\mathbf{\Omega}_t \right|^2 \right\},
$$

$$
= \int \int \mathcal{P}_h(\tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)\, d\tau\, d\mathbf{\Omega}_t, \tag{2.25}
$$

Several statistical moments of the power-delay and power-direction profiles are of practical interest, allowing for an evaluation of the channel frequency or spatial selectivity. The RMS delay and transmit/receive directional spreads are the square-root of the second-order moments of $\mathcal{P}(\tau)$, $\mathcal{A}_t(\mathbf{\Omega}_t)$ and $\mathcal{A}_r(\mathbf{\Omega}_r)$, respectively. The RMS delay-spread is therefore given by

$$
\tau_{RMS} = \sqrt{ \frac{\int_0^\infty (\tau - \tau_M)^2 \mathcal{P}(\tau)\, d\tau}{\int_0^\infty \mathcal{P}(\tau)\, d\tau} }, \tag{2.26}
$$

where

$$
\tau_M = \frac{\int_0^\infty \tau \mathcal{P}(\tau)\, d\tau}{\int_0^\infty \mathcal{P}(\tau)\, d\tau} \tag{2.27}
$$

is the average excess delay.

In the direction domains, the RMS directional spreads are expressed similarly [Fle00], e.g. at the transmit side:

$$
\mathbf{\Omega}_{t,M} = \frac{\int \mathbf{\Omega}_t \mathcal{A}_t(\mathbf{\Omega}_t)\, d\mathbf{\Omega}_t}{\int \mathcal{A}_t(\mathbf{\Omega}_t)\, d\mathbf{\Omega}_t} \tag{2.28}
$$

$$
\Omega_{t,RMS} = \sqrt{ \frac{\int \|\mathbf{\Omega}_t - \mathbf{\Omega}_{t,M}\|^2 \mathcal{A}_t(\mathbf{\Omega}_t)\, d\mathbf{\Omega}_t}{\int \mathcal{A}_t(\mathbf{\Omega}_t)\, d\mathbf{\Omega}_t} }, \tag{2.29}
$$

Note that the average direction $\mathbf{\Omega}_{t,M}$ is a vector, while $\Omega_{t,RMS}$ is a scalar (but not in angular units), corresponding to a RMS Euclidean distance.

The vectors $\mathbf{\Omega}_t$ and $\mathbf{\Omega}_r$ may also be divided into azimuth and elevation angles, (Θ_t, ψ_t) and (Θ_r, ψ_r). Spreads are then defined for each of these four angles, e.g. the transmit azimuth-spread is given by

$$
\Theta_{t,RMS} = \sqrt{ \frac{\int_0^{2\pi} (\Theta_t - \Theta_{t,M})^2 \mathcal{A}_t(\Theta_t)\, d\Theta_t}{\int_0^{2\pi} \mathcal{A}_t(\Theta_t)\, d\Theta_t} }, \tag{2.30}
$$

where

$$
\Theta_{t,M} = \frac{\int_0^{2\pi} \Theta_t \mathcal{A}_t(\Theta_t)\, d\Theta_t}{\int_0^{2\pi} \mathcal{A}_t(\Theta_t)\, d\Theta_t} \tag{2.31}
$$

is the average transmit azimuth angle.

2.1.5   From double-directional propagation to MIMO channels

Deriving the MIMO channel matrix

So far, we have described the double-directional channel. This is still not the classical description of a MIMO channel as introduced in Chapter 1, where the multi-antenna channel was represented as a $n_r \times n_t$ matrix, $\mathbf{H}(n, m)$ being the channel between the mth transmit antenna to the nth receive antenna. However, it is straightforward to convert the double-directional channel to a MIMO channel, by writing

$$\mathbf{H}(t, \tau) = \begin{bmatrix} h_{11}(t, \tau) & h_{12}(t, \tau) & \dots & h_{1n_t}(t, \tau) \\ h_{21}(t, \tau) & h_{22}(t, \tau) & \dots & h_{2n_t}(t, \tau) \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_r 1}(t, \tau) & h_{n_r 2}(t, \tau) & \dots & h_{n_r n_t}(t, \tau) \end{bmatrix}, \tag{2.32}$$

where

$$h_{nm}(t, \tau) \triangleq \iint h_{nm}\left(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) d\mathbf{\Omega}_t \, d\mathbf{\Omega}_r$$

$$\triangleq \iint h\left(t, \mathbf{p}_t^{(m)}, \mathbf{p}_r^{(n)}, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) d\mathbf{\Omega}_t \, d\mathbf{\Omega}_r, \tag{2.33}$$

$\mathbf{p}_t^{(m)}$ and $\mathbf{p}_r^{(n)}$ designating the locations of the mth transmit antenna and the nth receive antenna. Again, (2.33) is written for isotropic antennas at both ends. However, antenna patterns are easily accounted for by means of (2.5). Finally, let us introduce $\mathbf{H}_t[\tau] \triangleq \mathbf{H}(t, \tau)$ as a compact alternative notation for the MIMO channel matrix

$$\mathbf{H}(t, \tau) = \mathbf{H}_t[\tau] = \sum_{k=0}^{n_s-1} \mathbf{H}\left(t, \tau_k\right) \delta\left(\tau - \tau_k\right). \tag{2.34}$$

This compact notation will be used when we describe design methods for space-time codes over frequency-selective channels.

Introducing the steering vectors

Under three assumptions (the narrowband and balanced array assumptions as well as the plane wave assumption [PNG03, Her04]), we will now show how $h_{nm}(t, \tau)$ can be easily written as a function of $h_{11}\left(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right)$.

Definition 2.5   *An antenna array is assumed to be narrowband when the system bandwidth B is much smaller than the reciprocal of the transit time across the array, i.e. the product of the bandwidth by the array aperture is much smaller than the velocity of light[1].*

---

[1] This assumption does not imply that the channel is frequency-flat for the considered system.

Definition 2.6 *An antenna array is balanced when at given time and delay and for given DoD and DoA, the energy of the channel between any transmit and receive antenna pair does not depend on the particular pair.*

These first assumptions imply that $h_{nm}(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$ has the same energy as $h_{11}(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$ and that for a given contribution, the delay τ is the same for both channels. Hence, $h_{nm}(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$ can be simply written as a shifted version of $h_{11}(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$.

Definition 2.7 *A wave propagating from any scatterer to an antenna array can be approximated as plane if the distance between the scatterer and the array is much larger than the dimension of the array aperture.*

This assumption means that the phase-shift between $h_{nm}(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$ and $h_{11}(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$ is easily expressed based on simple geometrical considerations using plane wavefronts.

When these three assumptions are met, we can directly express $h_{nm}(t, \tau)$ as

$$h_{nm}(t, \tau) = \int \int h_{11}(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r) \, e^{-j\mathbf{k}_r^T(\mathbf{\Omega}_r)\left[\mathbf{p}_r^{(n)} - \mathbf{p}_r^{(1)}\right]} \, e^{-j\mathbf{k}_t^T(\mathbf{\Omega}_t)\left[\mathbf{p}_t^{(m)} - \mathbf{p}_t^{(1)}\right]} \, d\mathbf{\Omega}_t \, d\mathbf{\Omega}_r \quad (2.35)$$

where $\mathbf{k}_t(\mathbf{\Omega}_t)$ and $\mathbf{k}_r(\mathbf{\Omega}_r)$ are the transmit and receive wave propagation 3×1 vectors.

Let us now consider a two-dimensional propagation scenario, so that the angular variable $\mathbf{\Omega}_t$ simply becomes the azimuth angle Θ_t. For a transmit ULA oriented broadside to the link axis,

$$e^{-j\mathbf{k}_t^T(\mathbf{\Omega}_t) \cdot \left[\mathbf{p}_t^{(m)} - \mathbf{p}_t^{(1)}\right]} \triangleq e^{-js_m(\mathbf{\Omega}_t)} = \underbrace{e^{-j(m-1)\varphi_t(\theta_t)}}_{\text{for a ULA}}, \quad (2.36)$$

where $\varphi_t(\theta_t) = 2\pi(d_t/\lambda)\cos\theta_t = 2\pi(d_t/\lambda)\sin\Theta_t$, and $d_t = \|\mathbf{p}_t^{(m)} - \mathbf{p}_t^{(m-1)}\|$ denotes the inter-element spacing of the transmit array. The second equality only holds true for ULAs. Note that Θ_t is referenced with respect to the link axis (it is therefore independent of the array orientation). By contrast, θ_t is defined relatively to the array orientation (so $\theta_t = \pi/2$ corresponds to the link axis for a broadside array). The convenient use of this variable is thus restricted to linear arrays. The same operation can be carried out for all transmit antennas, as well as for all receive antennas. So, we define a transmit steering vector (expressed here for a ULA) in the relative direction θ_t by

$$\mathbf{a}_t(\theta_t) = \begin{bmatrix} 1 & e^{-j\varphi_t(\theta_t)} & \dots & e^{-j(n_t-1)\varphi_t(\theta_t)} \end{bmatrix}^T. \quad (2.37)$$

The receive steering vector $\mathbf{a}_r(\theta_r)$ in any relative direction of arrival θ_r is defined similarly as

$$\mathbf{a}_r(\theta_r) = \begin{bmatrix} 1 & e^{-j\varphi_r(\theta_r)} & \dots & e^{-j(n_r-1)\varphi_r(\theta_r)} \end{bmatrix}^T, \quad (2.38)$$

with $\varphi_r(\theta_r) = 2\pi(d_r/\lambda)\cos\theta_r \, (=2\pi(d_r/\lambda)\sin\Theta_r$ for a broadside array), d_r being the inter-element spacing at the receive array. It is important to stress that

- the definition of a steering vector does not necessarily require the array to be balanced (for an unbalanced array, the various elements of the steering vector simply have different amplitudes)

- for a balanced array, the steering vectors are classically normalized so that $\|\mathbf{a}_t\|^2 = n_t$ and $\|\mathbf{a}_r\|^2 = n_r$
- the definition of steering vectors is not restricted to ULAs, although their expressions for arbitrary arrays may not be explicit.

To sum up, under the plane wave and narrowband array assumptions, the channel matrix in (2.32) can be rewritten as a function of steering vectors as

$$\mathbf{H}(t, \tau) = \int \int h\left(t, \mathbf{p}_t^{(1)}, \mathbf{p}_r^{(1)}, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r\right) \mathbf{a}_r\left(\mathbf{\Omega}_r\right) \mathbf{a}_t^T\left(\mathbf{\Omega}_t\right) d\mathbf{\Omega}_t \, d\mathbf{\Omega}_r. \tag{2.39}$$

In particular, any coherent contribution with DoD $\mathbf{\Omega}_{t,c}$ and DoA $\mathbf{\Omega}_{r,c}$ is simply proportional to $\mathbf{a}_r(\mathbf{\Omega}_{r,c}) \mathbf{a}_t^T(\mathbf{\Omega}_{t,c})$.

A finite scatterer representation

For a WSSUSH narrowband channel with scattering in the azimuthal plane only, (2.39) can be rewritten in a more compact way. The fundamental assumption of the finite scatterer representation [SMB01, Bur03] is that the transmitter and receiver are coupled via a finite number of scattering paths with $n_{s,t}$ DoDs at the transmitter and $n_{s,r}$ DoAs at the receiver, so that the integral in (2.39) can be replaced by a summation, yielding

$$\mathbf{H}(t, \tau) = \mathbf{A}_r \mathbf{H}_s(t, \tau) \mathbf{A}_t^T, \tag{2.40}$$

where $\mathbf{A}_r$ and $\mathbf{A}_t$ represent the $n_r \times n_{s,r}$ and $n_t \times n_{s,t}$ matrices whose columns are the steering vectors related to the directions of each path observed at Rx and Tx (these matrices are in general not unitary), while the central matrix $\mathbf{H}_s(t, \tau)$ is a $n_{s,r} \times n_{s,t}$ matrix whose elements are the complex path gains between all DoDs and DoAs at time instant t and delay τ. If the uncorrelated scattering is indeed verified, the elements of $\mathbf{H}_s$ are independent variables.

2.1.6   Statistical properties of the channel matrix

Spatial correlation

The originality of MIMO techniques lies in the exploitation of the spatial or double-directional structure of the channel. This structure is also largely responsible for the performance of MIMO systems. Therefore, it does make sense to introduce a way to characterize the spatial properties of a multi-antenna channel, or more specifically, the space-only correlation of the channel. With this goal in mind, let us consider the expression in (2.9) when defining R_{T_a}. Reducing to zero the time and frequency shifts, we end up with the correlation between two narrowband spatial channels at time t and frequency f: the first channel links a transmit antenna located at ℓ_t to a receive antenna at ℓ_r, and the second channel links a transmit antenna located at ℓ_t' to a receive antenna at ℓ_r'. For a MIMO

channel with a limited number of antennas, we can switch from the double-directional notation to a matrix notation, and therefore define the spatial correlation matrix as

$$\mathbf{R} = \mathcal{E}\left\{\text{vec}\left(\mathbf{H}^H\right)\ \text{vec}(\mathbf{H}^H)^H\right\}. \tag{2.41}$$

This matrix is a $n_t n_r \times n_t n_r$ positive semi-definite Hermitian matrix, which describes the correlation between all pairs of transmit-receive channels. Depending on which channel pairs are chosen, several correlations can be defined:

- if both channels share the same transmit and receive antennas m and n, $\mathcal{E}\{\mathbf{H}(n, m)\mathbf{H}^*(n, m)\}$ simply represents the average energy of the channel between antenna m and antenna n;
- if both channels share the same transmit antenna, $r_m^{(nq)} = \mathcal{E}\{\mathbf{H}(n, m)\mathbf{H}^*(q, m)\}$ represents the receive correlation between channels originating from transmit antenna m and impinging upon receive antennas n and q;
- if both channels share the same receive antenna, $t_n^{(mp)} = \mathcal{E}\{\mathbf{H}(n, m)\mathbf{H}^*(n, p)\}$ represents the transmit correlation between channels originating from transmit antennas m and p and arriving at receive antenna n;
- if both channels originate from different transmit antennas, and arrive at different receive antennas, $\mathcal{E}\{\mathbf{H}(n, m)\mathbf{H}^*(q, p)\}$ is simply defined as the cross-channel correlation between channels (m, n) and (q, p).

We draw the attention of the reader to the fact that all these covariances are correlations between *channels*, despite the fact that some are called transmit or receive *antenna* correlations for simplicity. It is also convenient to define transmit and receive correlation matrices $\mathbf{R}_t$ and $\mathbf{R}_r$ as

$$\mathbf{R}_t = \mathcal{E}\left\{\mathbf{H}^H\mathbf{H}\right\} \tag{2.42}$$

$$\mathbf{R}_r = \mathcal{E}\left\{\left(\mathbf{H}\mathbf{H}^H\right)^T\right\} \tag{2.43}$$

To conclude, we would like to stress that, for homogeneous channels, the correlation coefficient between two individual channels is directly related to the joint angular power spectrum $\mathcal{A}(\mathbf{\Omega}_t, \mathbf{\Omega}_r)$, while the transmit (respectively receive) correlation is related to $\mathcal{A}_t(\mathbf{\Omega}_t)$ (respectively $\mathcal{A}_r(\mathbf{\Omega}_r)$). As an example, assuming 2-D propagation, the transmit correlation between two antennas can be written as

$$t = \int_0^{2\pi} e^{j\varphi_t(\theta_t)} \mathcal{A}_t(\theta_t)\, d\theta_t \tag{2.44}$$

where, for convenience, we expressed the transmit azimuth power spectrum with respect to the relative azimuth.

Two extreme cases are worth mentioning here. The first case corresponds to a very rich scattering environment around the transmitter with a uniform distribution of the energy,

so that $\mathcal{A}_t(\theta_t) \cong 1/2\pi$. The transmit correlation then reads as

$$
\begin{aligned}
t &= \frac{1}{2\pi} \int_0^{2\pi} e^{j\varphi_t(\theta_t)} d\theta_t \\
&= \frac{1}{2\pi} \int_0^{2\pi} e^{j2\pi(d_t/\lambda)\cos\theta_t} d\theta_t \\
&= J_0\left(2\pi\frac{d_t}{\lambda}\right).
\end{aligned}
\tag{2.45}
$$

The transmit correlation only depends on the spacing between the two considered antennas.

The second extreme case occurs when scatterers around the transmit array are concentrated along a narrow direction $\theta_{t,0}$, i.e. $\mathcal{A}_t(\theta_t) \to \delta(\theta_t - \theta_{t,0})$. Such channel is also known as degenerate and causes a very high transmit correlation approaching one

$$
t \to e^{j\varphi_t(\theta_{t,0})} = e^{j2\pi(d_t/\lambda)\cos\theta_{t,0}}.
\tag{2.46}
$$

Interestingly, the scattering direction is directly related to the phase of the transmit correlation.

Using the finite scatterer representation of (2.40), and assuming that all paths are independent and have the same average power normalized to one, the antenna covariances are directly related to the steering vectors as

$$
\mathbf{R}_t = \left(\mathbf{A}_t\mathbf{A}_t^H\right)^T = \mathbf{A}_t^*\mathbf{A}_t^T
\tag{2.47}
$$

$$
\mathbf{R}_r = \left(\mathbf{A}_r\mathbf{A}_r^H\right)^T = \mathbf{A}_r^*\mathbf{A}_r^T.
\tag{2.48}
$$

It is straight forward to prove that the first equation above is indeed equivalent to (2.44) in this case.

Finally, it is interesting to consider the case when the energy spreading is very large at both sides or when the antenna inter-element spacings (d_t and d_r) are sufficiently large. In such scenarios, the various elements of $\mathbf{H}$ become uncorrelated, and $\mathbf{R}$ becomes diagonal.

Singular values and eigenvalues

The channel matrix $\mathbf{H}$ is not always full rank. Denoting by $r(\mathbf{H})$ the rank of $\mathbf{H}$, the singular value decomposition (SVD) of the $n_r \times n_t$ channel matrix reads as

$$
\mathbf{H} = \mathbf{U}_\mathbf{H}\mathbf{\Sigma}_\mathbf{H}\mathbf{V}_\mathbf{H}^H,
\tag{2.49}
$$

where $\mathbf{U}_\mathbf{H}$ and $\mathbf{V}_\mathbf{H}$ are $n_r \times r(\mathbf{H})$ and $n_t \times r(\mathbf{H})$ unitary matrices, and

$$
\mathbf{\Sigma}_\mathbf{H} = \mathrm{diag}\left\{\sigma_1, \sigma_2, \ldots, \sigma_{r(\mathbf{H})}\right\}
\tag{2.50}
$$

is the diagonal matrix containing the ordered singular values of $\mathbf{H}$.

Denoting by n the minimum of $\{n_t, n_r\}$, the matrix $\mathbf{W} = \mathbf{H}\mathbf{H}^H$ (for $n_t > n_r$) or $\mathbf{W} = \mathbf{H}^H\mathbf{H}$ (for $n_t < n_r$) is an $n \times n$ positive semi-definite Hermitian matrix, whose eigenvalue decomposition (EVD) is given by

$$\mathbf{W} = \mathbf{U_W}\mathbf{\Lambda_W}\mathbf{U}_\mathbf{W}^H, \tag{2.51}$$

where $\mathbf{\Lambda_W} = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ contains the eigenvalues of $\mathbf{W}$ (i.e. the squared singular values of $\mathbf{H}$), $r(\mathbf{H})$ of which are non-zero and equal to $\{\lambda_1, \ldots, \lambda_{r(\mathbf{H})}\} = \{\sigma_1^2, \sigma_2^2, \ldots, \sigma_{r(\mathbf{H})}^2\}$. For notational convenience, we always write $\lambda_k \triangleq \lambda_k(\mathbf{W})$. These eigenvalues are naturally random variables, whose distribution will be studied in Chapter 4.

Frobenius norm
The squared Frobenius norm of $\mathbf{H}$ is given by

$$\|\mathbf{H}\|_F^2 = \text{Tr}(\mathbf{H}\mathbf{H}^H) = \sum_{n=1}^{n_r}\sum_{m=1}^{n_t} |\mathbf{H}(n, m)|^2 = \sum_{k=1}^{n} \lambda_k. \tag{2.52}$$

This represents the total energy of the channel, and is often normalized. Similarly to the eigenvalues of $\mathbf{W}$, $\|\mathbf{H}\|_F^2$ is also a random variable, whose distribution is analyzed for various scenarios in [NBE+02] (for those two points, see Chapter 4).

2.1.7 Discrete channel modeling: sampling theorem revisited

Before going into the more detailed description of various channel models, we need to address the question of sampled representations of the directional channel impulse response. It seems evident that the resolution of any channel model does not have to be arbitrarily high. Indeed, the transmitter/receiver/scatterer speed and the system bandwidth are limited, and arrays are not continuous, as individual antenna elements occupy discrete locations.

The representation of Doppler and bandwidth-limited channels has been studied for many years. Regarding the time selectivity, it is well-known that the channel should be sampled according to the Nyquist criterion, based on the maximum Doppler frequency (we denote by $\mathbf{H}_k$ the narrowband channel matrix $\mathbf{H}(t)$ at the kth discrete time instant). Usually, the symbol rate is much higher than the maximum Doppler frequency, so the discretization is such that the kth discrete time instant corresponds to the kth symbol, or, for very high symbol rates, to the kth frame (i.e. the channel remains constant during one symbol or during one frame). The same criterion applies in the delay domain, the minimum resolution being given by the system bandwidth. Consider (2.5) with omnidirectional antennas. The filtering by $w(\tau) = w_r(\tau) \star w_t(\tau)$ will combine several physical paths indexed by k into a smaller number of paths (which we index by l). Because $w(\tau)$ is related to the symbol duration (it is ideally a Nyquist filter), this filtering operation leads to the well-known definition of so-called taps, corresponding to a discrete set of equi-spaced delays $\{\tau_l, l = 0, \ldots, L-1\}$, where L is related to the maximum excess delay. Note that L is usually much smaller

than n_s (yet, the true value of n_s is not necessarily known, as it would require an infinite measurement bandwidth, as already discussed in Chapter 1).

Therefore, any tap consists of several physical contributions, in the same way as the narrowband channel $h(t)$. These contributions are either coherent or fade independently, and the central limit theorem applies: each tap is the addition of a complex Gaussian variable with a constant non-fading term (see (2.13)). As a consequence, the tap amplitude is Ricean distributed, becoming Rayleigh distributed in the absence of coherent contribution. The discrete version of the power-delay profile is known as a tap-delay profile

$$\mathcal{P}(\tau) = \sum_{l=0}^{L-1} \beta_l \, \delta(\tau - \tau_l). \tag{2.53}$$

Each tap is then modeled as a narrowband channel on its own; the wideband channel is eventually obtained by adding all taps together, as already outlined in (2.4). Note that we deliberately use different indexes for taps or contributions, although both could be identical if (2.4) is measured over the system bandwidth. The tap spacing in the delay domain inversely depends on the bandwidth of the considered communication system. Specifically, for a given maximum excess delay, a larger bandwidth requires a higher delay resolution, hence a larger number of taps.

A final remark concerns the US assumption. Because of the filter side lobes, taps may no longer be uncorrelated, in contrast with the original physical paths. Despite the fact that this correlation may be significant, many practical channel models do not account for the correlation between taps as this may represent a very difficult task where the result still depends on the filtering operation.

Regarding the spatial properties, the antenna domain at each side is also a discrete set, whose cardinality is equal to the number of antennas. This implies that the required resolutions in the angular domains are equal to n_t and n_r, respectively. Therefore, the directional MIMO channel only requires such a resolution in the angular domain. Typically, for a 2×2 channel, the channel is described by 2 DoDs and 2 DoAs only, acting like *spatial taps*. This is the basis of so-called beamspace models (a.k.a. virtual channel representations, see Section 3.2.2).

2.1.8 Physical versus analytical models

In the rest of this chapter, we will review some existing physical models of the MIMO channel, while Chapter 3 will describe the so-called analytical representations of the channel. The former essentially characterizes the double-directional impulse response $h(t, \tau, \mathbf{\Omega}_t, \mathbf{\Omega}_r)$, while the latter allows for describing the spatial correlation properties of the MIMO channel matrix $\mathbf{H}(t, \tau)$. Although both types of approaches model the full channel (including antenna patterns and array configurations), physical models more easily accommodate different antenna patterns and/or array configurations, in two or three dimensions.

However, they cannot directly be used in space-time code design, as they do not offer analytical expressions of $\mathbf{H}(t, \tau)$. There is naturally a link between a physical model and its analytical representation, as illustrated in detail in the next chapter. However, we insist that analytical models are only a formalism, and do not give any quantitative information. In particular, an analytical model does not link the spatial correlation with geometrical or antenna parameters. Physical models can also be used in *a posteriori* simulations of signal processing algorithms. Indeed, these are often derived based on simplified analytical representations. It is therefore important to be able to test these algorithms against realistic propagation scenarios, and not just in the scenarios they have been designed for. For all these reasons, physical or standardized models are essential in understanding MIMO propagation and using MIMO channel modeling tools.

2.2   Electromagnetic models

2.2.1   *Ray-based deterministic methods*

For sometime now, full-wave parabolic and integral equation models have been used to model refraction and diffraction by the troposphere and terrain. An alternative solution consists of ray methods, which have been widely applied to the mobile terrestrial channel. Indeed, as long as the wavelength is small compared to the size of obstacles, we can consider that each contribution behaves like a ray, or more generally, like a narrow beam. So, the electromagnetic mechanisms may be described by extensions of geometrical optics. The basic problem consists in the calculation of the narrowband electric field received at the receiver, in amplitude, phase and polarization. This field, regarded as a complex vector, results from the combination of the direct component with several contributions due to specular reflection, diffraction and transmission through obstacles, as well as multiple combinations of these mechanisms.

Note that ray-tracing methods usually support only four types of components: the line-of-sight component, the component transmitted though obstacles, as well as single and multiple reflected and diffracted contributions. Because it is non-coherent (the phase is not deterministic), diffuse scattering is not easily dealt with in ray-based methods (the wavelength is not small compared to the size or the roughness of obstacles), and is therefore neglected on the basis that it is small with respect to the specular component.

The LOS contribution then consists of an emitted field E_0, which is successively:

- weighted by the radiation pattern of the transmit antenna
- attenuated by the usual free-space transmission loss, and a possible additional attenuation through an obstacle
- phase-shifted according to the relative length of the LOS path with respect to the wavelength
- weighted by the radiation pattern and the polarization mismatch at the receive terminal.

Both reflected and diffracted contributions are calculated similarly to the LOS component, considering the total length of the path and accounting for specular reflection and diffraction by means of complex reflection and diffraction coefficients [ANM00, OCRVJ02]. The geometrical relationships between the incident and the reflected/diffracted rays are deduced from Snell and Keller [MPM90] laws, respectively. The reflection and diffraction coefficients are obtained using the Fresnel formulae for reflection [MPM90] and the Uniform Theory of Diffraction (UTD) [Kel62, KP74, Lue84]. Subsequently, any single contribution at receive location $\mathbf{p}_r$ can be expressed as

$$E_k(\mathbf{p}_r) = \frac{1}{\xi_k} \Lambda_k'\left(s_k, s_k'\right) e^{-jks_k'} \, \mathbf{g}_r^*\left(\Omega_{r,k}\right) \mathbf{T}_k \, \mathbf{g}_t\left(\Omega_{t,k}\right) \Lambda_k\left(s_k, s_k'\right) e^{-jks_k} \, E_0 \qquad (2.54)$$

where

- E_0 is the emitted by the transmitter located at $\mathbf{p}_t$
- $\mathbf{g}_t$ and $\mathbf{g}_r$ are complex vectors accounting for the transmit/receive antenna polarizations and amplitude gains in the direction of the transmitted/received waves [Bal05]
- s_k is the path length from the transmitter to the scatterer
- s_k' is the path length from the scatterer to the receiver
- Λ_k and Λ_k' are the free-space transmitter/scatterer and scatterer/receiver spreading factors ($\Lambda_k = 1/s_k$ for a spherical wave front) respectively
- ξ_k is a possible additional attenuation along the path (caused by the transmission through an obstacle)
- $\mathbf{T}_k$ is a complex matrix modeling the interaction mechanism responsible for the considered contribution. We will come back to the electromagnetic modeling of $\mathbf{T}_k$ in Section 2.2.2.

Multiple contributions are easily expressed in the same way as in (2.54). If all single and multiple contributions are deterministic, the total field amplitude is eventually obtained by the following complex addition

$$E(\mathbf{p}_r) = \sum_k E_k(\mathbf{p}_r) + \sum_{k,l} E_{kl}(\mathbf{p}_r) + \sum_{k,l,p} E_{klp}(\mathbf{p}_r) + \dots, \qquad (2.55)$$

where $E_{kl}(\mathbf{p}_r)$ and $E_{klp}(\mathbf{p}_r)$ represent double and triple interactions respectively, e.g. double and triple reflections/diffractions, or a combination between these.

Although they were used in propagation long before the emergence of multi-antenna systems, ray-based techniques are intrinsically multidimensional (double-directional, polarization-sensitive, and including channel dynamics). The prediction of the MIMO channel matrix is then obtained directly from the discussion in Section 2.1.5.

As far as the accuracy of ray-tracing methods is concerned, four points appear to be crucial:

- the availability of up-to-date high-resolution databases
- the efficiency of the computational techniques, to trace back all paths between the transmitter and receiver quickly, for a sufficient order of reflection and diffraction

(investigations suggest that up to seven successive reflections and double diffraction should be accounted for [AN00], although five reflections and single diffraction may appear as a good compromise)
- the modeling of moving objects, such as cars and buses [FMK$^+$06]
- the accurate knowledge of electrical characteristic parameters for all obstacles and materials (default values are 10^{-3} S/m for conductivity σ and 5 for relative permittivity ϵ_r in typical suburban areas [Par00]).

2.2.2 Multi-polarized channels

Up to now, we have considered MIMO systems employing arrays of spatially separated elements at both ends. However, the exploitation of the polarization dimension might lead to more compact implementations. Indeed, orthogonal polarizations ideally offer a complete separation between channels, with a full decorrelation at both transmit and receive sides [SZM$^+$06]. In this section, we limit ourselves to 2×2 dual-polarized MIMO channels (both the transmit and receive arrays are made of two antennas with orthogonal polarizations). These do not need to be identical at Tx and Rx. For example, the transmit polarization scheme may be vertical-horizontal (denoted as VH), while the received scheme could be chosen as slanted (± 45 degrees). Furthermore, orthogonally-polarized antennas may be spatially separated, in order to combine both spatial and polarization diversity/multiplexing, e.g. two separated dual-polarized arrays forming a 4-antenna array. Finally, the extension to $n_r \times n_t$ MIMO schemes (with even n_t and n_r) is quite straightforward, assuming that antenna arrays are made of $n_t/2$ and $n_r/2$ dual-polarized sub-arrays.

Ideally, all cross-polar transmissions (e.g. from a vertically-polarized Tx antenna to an horizontally-polarized Rx antenna) should be zero. This is not the case in real-world conditions, owing to two depolarization mechanisms.

- Antenna depolarization: a linearly polarized antenna has a non-zero pattern for the cross-polar field. Therefore the signal at the output of, say, a vertically-polarized antenna receiving an horizontally-polarized wave will not be zero.
- Scatterer-based depolarization: reflection, diffraction and diffuse scattering usually modify the polarization of the incident wave.

The first mechanism is well-known in antenna theory [Bal05]. It is easily taken into account by means of the cross-polar antenna pattern. Such patterns usually present a minimum around the direction of the maximum co-polar gain, as illustrated in Figure 2.4, but may present maxima in other directions. We observe, however, that the amplitude of the cross-polar pattern is significantly smaller than that of the co-polar pattern.

The second mechanism is more complex. To better understand it, let us look at how electromagnetic models such as ray-tracing provide an intrinsic scattering depolarization model. In (2.54), we have introduced the matrix complex scattering coefficient **T**, which is a 3×3 matrix, corresponding to the 3 dimensions (x, y, z) of the space. For given surface/edge
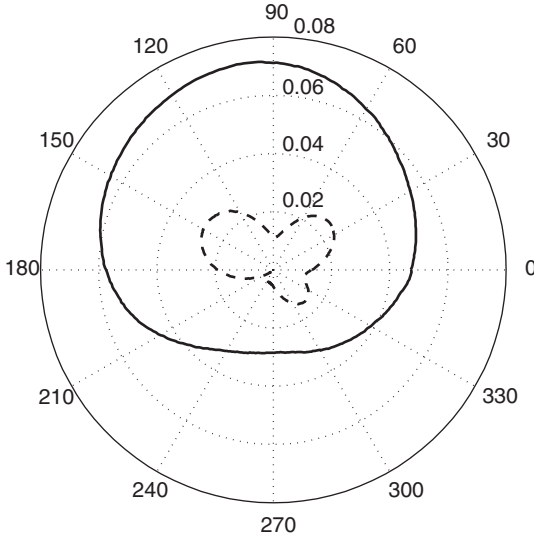
Figure 2.4   *Copolar (solid) and cross-polar (dashed) field-radiation patterns (in magnitude) of a plane monopolar antenna at 2.4 GHz*

and direction of incidence, it can be shown that there exist two orthogonal polarizations of the incident field, denoted as h (hard) and s (soft), which are reflected/diffracted without any depolarization. These directions are known as the characteristic polarizations of the interacting object.

In the case of a specular reflection, the directions of the characteristic polarizations are, respectively, given by unit vectors parallel (hard) and perpendicular (soft) to the plane of incidence (i.e. the plane containing the incident wave and the normal to the surface). When the incident wave is polarized along one of these two directions, it is reflected without depolarization and the reflection coefficients are given by the classical Fresnel coefficients

$$\Sigma_h = \frac{\epsilon_r' \cos \alpha - \sqrt{\epsilon_r' - \sin^2 \alpha}}{\epsilon_r' \cos \alpha + \sqrt{\epsilon_r - \sin^2 \alpha}} \tag{2.56}$$

$$\Sigma_s = \frac{\cos \alpha - \sqrt{\epsilon_r' - \sin^2 \alpha}}{\cos \alpha + \sqrt{\epsilon_r' - \sin^2 \alpha}} \tag{2.57}$$

where α is the angle of incidence, and ϵ_r' is the effective relative permittivity

$$\epsilon_r' = \epsilon_r - j \frac{\sigma}{2\pi f \epsilon_0}, \tag{2.58}$$

ϵ_r being the material permittivity relative to the vacuum permittivity $\epsilon_0 = 8.85 10^{-12}$ F/m, σ, the conductivity, and f, the frequency. When the incident wave is not polarized along

those directions, it is depolarized, i.e. its polarization after reflection is different than before reflection. The matrix reflection coefficient $\mathbf{T}$ is then easily expressed as a function of the characteristic polarizations, and of Σ_h and Σ_s (see [MPM90] for more details).

As far as diffraction is concerned, the existence of two characteristic polarizations allows for a similar description, based on two diffraction coefficients D_h and D_s.

In both cases, the expression of $\mathbf{T}$ results in a transfer from the incident polarization to the cross-polarization as soon as the former is no longer aligned with one of the characteristic directions. Furthermore, when the incident polarization is aligned with a principal direction, another effect is worth mentioning: Σ_h and Σ_s (equivalently, D_h and D_s) are not equal. This means that orthogonal co-polar components will not experience the same scattering coefficient, resulting in a co-polar imbalance. As an example, successive reflections between vertical walls attenuate horizontal polarization more than vertical polarization.

What about depolarization in diffuse scattering mechanisms? As already mentioned, diffuse or rough surface scattering cannot be included in ray-tracing tools. Still, what can be said about the depolarization properties of a diffuse scatterer? In general, the latter is characterized by the absence of principal directions, resulting in crosstalk for any incident polarization. For particular random surfaces (e.g. when the roughness characteristics are well approximated by Gaussian statistics), it is nevertheless possible (though beyond our scope) to model the dyadic scattering coefficient as a function of the incident/scattered angles and the surface properties [Ish98].

2.3   Geometry-based models

Geometry-based models are stochastic models of the channel using a simplified ray-based approach. However, instead of modeling the environment exactly, the scatterers whose scattering properties are often taken as complex Gaussian, are only specified by a spatial distribution. Depending on this distribution, several models have been developed, but in all cases, the channel between any transmit and receive antennas is obtained as

$$h = \lim_{n_s \to \infty} \sum_{k=0}^{n_s-1} e^{-jks'_k} \, g_r^* \left(\Theta_{r,k} \right) \, \Gamma_k \, g_t \left(\Omega_{t,k} \right) \, e^{-jks_k} \tag{2.59}$$

where

- g_t and g_r are the complex scalar amplitude gains of the transmit/receive antennas in the direction of the transmitted/received waves
- s_k is the path length from the transmit antenna to the kth scatterer
- s'_k is the path length from the kth scatterer to the receive antenna
- Γ_k is the complex scattering coefficient of the kth scatterer.
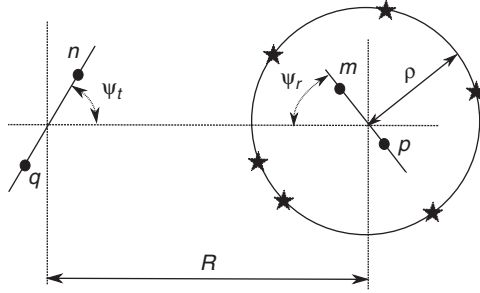
Figure 2.5   *One-ring model*

2.3.1   One-ring model

The one-ring model is a popular geometry-based stochastic model, which has been pro-
posed in the context of MIMO systems by [SFGK00]. It represents Rayleigh-fading
channels where single-bounce scatterers located around the mobile terminal (MT) dom-
inate. This is true when the base station (BS) is elevated, and not obstructed by local
scattering. For notational convenience, we assign the roles of transmitter and receiver to
the mobile terminal and base station, respectively (i.e. we consider an uplink scenario).
The radius of the scatterer ring is denoted as ρ, the distance from transmitter to receiver,
as R, and the ratio R/ρ, as Δ (see Figure 2.5). The one-ring model is then basically a
ray-tracing model, using the following assumptions.

- Every actual scatterer is represented by a corresponding effective scatterer located at
  the same angle on the scatterer ring. Effective scatterers are assumed to be uniformly
  distributed on the ring, and are associated with a random phase $\phi(\Theta_r)$ representing the
  scattering properties. All scatterers are omnidirectional. Statistically, $\phi(\Theta_r)$ is modeled
  as uniformly distributed in $[0, 2\pi)$ and independent in Θ_r. The radius of the scatterer
  ring is determined by the root-mean-square (RMS) delay spread of the microcellular
  channel delay-spread, as τ_{RMS} is proportional to ρ.
- There is no line-of-sight path, and only rays that are reflected by effective scatterers
  exactly once are considered.
- All rays are equal in amplitude.

Proposition 2.1   *Based on the set of assumptions highlighted above, and further
assuming that Δ is large, the correlation matrix $\mathbf{R}$ of the one-ring model is approxi-
mated by*

$$\mathcal{E}\{\mathbf{H}(n,m)\mathbf{H}^*(q,p)\} \approx \exp\left[-j2\pi \cos \psi_t D_{mp}/\lambda\right]$$

$$J_0\left(2\pi/\lambda\sqrt{\left(\frac{\sin \psi_t\, D_{mp}}{\Delta} + \sin \psi_r\, D_{nq}\right)^2 + (\cos \psi_r\, D_{nq})^2}\right) \qquad (2.60)$$

where D_{mp} is the spacing between antennas m and p.

PROOF: Based on (2.36), we can easily write

$$\mathbf{H}(n,m)\mathbf{H}^*(q,p) = \sum_{\Theta_t,\Theta_r} e^{-j\left[s_m(\Theta_t)-s_p(\Theta_t)\right]}e^{-j\left[s_n(\Theta_r)-s_q(\Theta_r)\right]}. \tag{2.61}$$

The argument of the exponential above is nothing else than the path length difference between channels (n,m) and (q,p). The latter is given by simple geometrical relationships as

$$\begin{aligned}(s_m - s_p) + (s_n - s_q) &= \frac{2\pi}{\lambda}\left[D_{mp}\cos(\psi_t - \Theta_t) + D_{nq}\cos(\psi_r - \Theta_r)\right]\\ &= \frac{2\pi}{\lambda}\left[D_{mp}(\cos\Theta_t\cos\psi_t + \sin\Theta_t\sin\psi_t)\right.\\ &\quad \left.+D_{nq}(\cos\Theta_r\cos\psi_r + \sin\Theta_r\sin\psi_r)\right]\end{aligned}$$

Note that Θ_t and Θ_r are unequivocally related, as the one-ring model is single bounce. So, the sum in (2.61) only needs to be performed on uniformly distributed Θ_r. Furthermore, if Δ is large, Θ_t and Θ_r can be approximated by

$$\sin\Theta_t \approx \Delta^{-1}\sin\Theta_r$$

$$\cos\Theta_t \approx 1 - \frac{1}{4\Delta^2} + \frac{1}{2\Delta^2}\cos 2\Theta_r \approx 1.$$

Finally, taking the expectation, and using the fact that $\frac{1}{2\pi}\int_0^{2\pi}\exp(jx\cos\Theta + jy\sin\Theta)\,d\Theta = J_0\left(\sqrt{x^2+y^2}\right)$, we are able to rewrite the correlation as (2.60). ∎

One last remark about (2.60): D_{nq} and D_{mp} are not necessarily positive distances. Their sign translates the fact that n can be smaller or larger than q (and the same holds true for m and p), so that $D_{nq} = -D_{qn}$, and $D_{mp} = -D_{pm}$. Antennas can, however, be indexed arbitrarily, yet both arrays must be indexed in a similar way (e.g. both from top to bottom in Figure 2.5).

The one-ring model has been extended in [AK02], by considering a limited angle-spread at the MT, through the use of a Von Mises distribution for the DoA, centered around a mean direction μ,

$$p_{\Theta_r}(\Theta_r) = \frac{\exp\left[\kappa\cos(\Theta_r - \mu)\right]}{2\pi I_0(\kappa)}, \tag{2.62}$$

where κ controls the receive angular spread, κ varying between 0 (isotropic scattering, full one-ring) and ∞ (extremely directional scattering). The correlation in (2.60) then becomes

$$\mathcal{E}\{\mathbf{H}(n,m)\mathbf{H}^*(q,p)\} \approx \frac{\exp\left[j2\pi\cos\psi_t D_{mp}\lambda\right]}{I_0(\kappa)}$$

$$I_0\left(\sqrt{\kappa^2 - \frac{(\sin\psi_t D_{mp}/\Delta + \sin\psi_r D_{nq})^2 - (\cos\psi_r D_{nq})^2}{\lambda^2}}\right). \tag{2.63}$$

Figure 2.6   *Two-ring model (symmetric)*

It is trivial to verify that (2.63) reduces to (2.60) for $\kappa = 0$.

2.3.2   Two-ring model

Several models are based on two rings, in order to deal with more complex situations than the one-ring model. The symmetrical double-bounce model considers that both the transmitter and the receiver are surrounded by a local scatterer ring, as illustrated in Figure 2.6. This model appears to fit indoor scenarios better, where both the BS and the MT are located at similar heights. The main drawback of this model is that the channel coefficients are no longer complex Gaussian variables, as all paths are scattered twice. The result is that the amplitude s of the channel matrix elements actually follows a double-Rayleigh distribution

$$p_s(s) = \frac{s}{\sigma_s^4} K_0 \left(\frac{s}{\sigma_s^2} \right), \tag{2.64}$$

where $K_0(\cdot)$ is a zeroth ordered modified Bessel Function of the second kind. The consequence is that the channel correlation matrix does not completely describe the MIMO channel.

A second asymmetrical single-bounce model considers that scatterers are located on two different rings

- a local ring, surrounding the MT (similarly to the one-ring model), whose radius is much smaller than the range R
- a remote ring, surrounding the link, whose radius is much larger than the range R.

The correlation resulting from remote scatterers is simply written as

$$\mathcal{E}\{\mathbf{H}(n,m)\mathbf{H}^*(q,p)\} = J_0 \left(2\pi \sqrt{D_{mp}^2 + D_{nq}^2 \pm 2D_{mp}D_{nq} \cos(\psi_t - \psi_r)} \right). \tag{2.65}$$

2.3.3   Combined elliptical-ring model

The goal of this model [OEP03] is to derive a representation of the MIMO channel, starting from the knowledge of the tap-delay profile $\mathcal{P}(\tau)$ and the angular-spread at the user terminal

at a specified range. A power-delay profile at a given range R^* consists of L of taps, each tap being characterized by an average power β_l relative to the power of the first tap, a relative delay τ_l and a K-factor K_l (possibly equal to 0).

In most empirical tap-delay line models (such as those described in Section 2.4), the tap amplitudes $|h(t, \tau_l)|$ are considered as Rayleigh-distributed, except for the first tap. This first tap is Ricean-distributed if a strong direct contribution exists, e.g. in line-of-sight or quasi line-of-sight links.

Since scattering mechanisms in macrocellular scenarios are mostly two-dimensional processes, scatterers causing echoes with identical delays are situated on an ellipse, the foci of which are given by transmit and receive locations. Consequently, any tap-delay profile can be spatially represented by a finite set of L ellipses, each one containing S_l scatterers. Assuming that the model is valid for LOS or quasi-LOS links, the first ellipse degenerates onto the link axis. Considering omnidirectional antennas at both ends, the lth tap is expressed as the sum of a coherent contribution $\bar{h}(t, \tau_l)$ and of a Rayleigh contribution $\tilde{h}(t, \tau_l)$,

$$h(t, \tau_l) = \bar{h}(t, \tau_l) + \tilde{h}(t, \tau_l) \tag{2.66}$$

The Rayleigh contribution is written as

$$\tilde{h}(t, \tau_l) = \frac{\lambda}{4\pi R_l^{\eta/2}} \sum_{m=1}^{S_l-1} g_r(\Theta_{r,lm}) \Gamma_{lm} e^{j\phi_{lm}} g_t(\Theta_{t,lm}) \tag{2.67}$$

where g_t and g_r are the transmit and receive antenna field-patterns, $R_l = c\tau_l + R$ is the path length related to the lth ellipse, η is the path-loss exponent, $\Gamma_{lm} \exp(j\phi_{lm})$, $\Theta_{t,lm}$ and $\Theta_{r,lm}$ are respectively the complex reflection coefficient, the DoD and the DoA of the mth scatterer on the lth ellipse. The coherent contribution corresponds here to the propagation along the link axis, and reads for $K_0 > 0$, $l = 0$, as

$$\bar{h}(t, \tau_l) = g_t(\pi/2) g_r(\pi/2) \frac{\lambda}{4\pi R^{\eta/2}}, \tag{2.68}$$

whereas for $K_0 = 0$ or $l \geq 0$, $\bar{h}(t, \tau_l) = 0$.

In order to build the geometrical representation of the channel, the size of the ellipses and the respective numbers of scatterers S_l is inferred at any range R from the input parameters $\{\beta_l, \tau_l; l = 0, \ldots, L-1\}$ and K_0 corresponding to the reference range. The set of values $\{S_0, \ldots, S_{L-1}\}$ related to the maximal BS-to-MT distance R^* is denoted as $\{S_0^*, \ldots, S_{L-1}^*\}$.

So far, scatterers have been uniformly distributed on ellipses so that the channel impulse response fits a desired power-delay profile at a reference range. Such channels are symmetrical, in the sense that the angle-spread – as seen from the base station – is the same as the angle-spread at the subscriber unit. However, macrocellular communication links with similar power-delay profiles may significantly differ as far as the spatial distribution
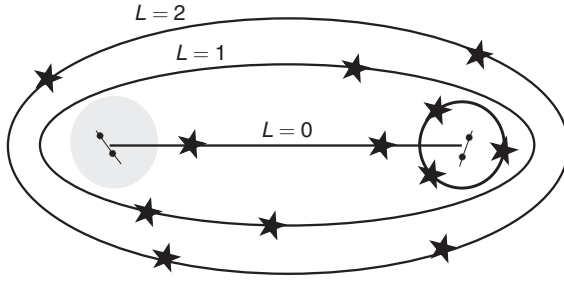
Figure 2.7   *Combined elliptical-ring model for 3 ellipses and a local scatterer ring (the size of the disc and the circle have been increased for better legibility)*

of scatterers is concerned. More specifically, the difference between the relative heights of base station and mobile terminal will cause the angle-spread to differ at the BS and the MT. Unlike the BS, the MT can *see* a larger number of scatterers owing to its lower height and larger elevation beamwidth. In order to account for such effects, two local rings are simultaneously introduced:

- a disc of exclusion, representing a scatterer-free area (radius R_B) around the BS
- a circular ring (radius R_M) surrounding the MT and bearing a subset of $S_0'(\leq S_0)$ scatterers taken among the estimated number of scatterers on the first ellipse to keep the same tap-delay profile.

The local scattering ratio (LSR) at the reference range is defined as S_0'/S_0^*, i.e. it gives the ratio between the amount of local scatterers surrounding the MT to the total amount of scatterers corresponding to the first ellipse (i.e. those surrounding the MT and those situated along the axis between the MT and the BS). Local scattering ratios of 0 and 1, correspond at the MT to a low multipath richness with highly directional scattering and to a rich omnidirectional multipath environment respectively. Therefore, the LSR is directly related to the azimuth-spread at the MT. A graphical representation of the model is given in Figure 2.7, for $L = 2$. A complete and detailed description of the model can be found in [OEP03].

2.3.4   Elliptical and circular models

The geometry-based models described above have scatterers randomly located on rings or ellipses. A different approach consists of distributing scatterers within a region, i.e. a scatterer probability density function has to be specified. Depending on the shape of the region, we obtain so-called elliptical or circular models. Both models use the same assumptions as all geometry-based models, i.e.

- propagation takes place in a 2-D plane
- scatterers are assumed to be omnidirectional re-radiating elements
- only single scattering is accounted for.

In the elliptical model [LR96], the scatterers are uniformly distributed within an ellipse, the foci of which correspond to the transmitter and receiver, so any ellipse is iso-delay. This model has been developed for microcellular propagation, where multipaths originate from both ends of the link. The size of the ellipse is directly related to the maximum excess delay, τ_{max}. There is also a direct relationship between τ_{max} and the channel RMS delay-spread τ_{RMS} [ECS$^+$98].

The circular model [PRR02] is an extension of the one-ring model, scatterers now being located in the entire circular region inside the local ring. As the one-ring model, the circular model better fits macrocellular propagation, characterized by a strong asymmetry between the local environments at the base station and the mobile terminal. Note that elliptical and circular models allow for a path-loss exponent to be defined, so that scatterers with longer delays also suffer larger attenuation. That is not the case for models presented in Sections 2.3.1 and 2.3.2. Finally, more details about the derivation of the joint DoA-ToA distribution may be found in [ER99].

2.3.5 Extension of geometry-based models to dual-polarized channels

In the geometry-based models detailed above, the scattering coefficients are scalar parameters (usually modeled by a complex Gaussian distribution). In order to account for different transmit/receive polarizations, these coefficients must become matrix parameters.

In this way, any scatterer is now characterized by a scattering matrix $\mathbf{\Gamma}$, which depends on the polarization scheme. Let us first introduce the statistical properties of the scattering matrix $\mathbf{\Gamma}_0$ for the particular case of a vertical-horizontal scheme at both link ends

$$\mathbf{\Gamma}_0 = \begin{bmatrix} \Gamma_{vv} & \Gamma_{vh} \\ \Gamma_{hv} & \Gamma_{hh} \end{bmatrix} \tag{2.69}$$

where subscripts v and h refer to vertical and horizontal polarizations respectively. A model for this matrix is proposed in [OEP04], where effective scattering matrices (i.e. corresponding to various aggregate physical paths) for downlink scenarios are characterized for different transmit and receive locations in an urban environment at a frequency of 2.5 GHz. A more general version of this model is as follows:

1. Γ_{vv} is complex Gaussian distributed, with the additional condition that $|\Gamma_{vv}| \leq 1$.
2. Γ_{hh} can be expressed as a phase-shifted and attenuated version of Γ_{vv}

$$\Gamma_{hh} = \Gamma_{vv} \mu e^{-j\epsilon} \tag{2.70}$$

where ϵ is a zero-mean random variable and μ is the gain imbalance between vertical and horizontal scattering amplitudes. This imbalance results from the fact that the major propagation mechanisms, such as street waveguiding (i.e. reflection by vertical walls)

and rooftop diffraction by horizontal wedges, are polarization-selective in favor of the vertical polarization (see the discussion in Section 2.2.2).

3. Γ_{hv} and Γ_{hv} are similarly proportional to Γ_{vv} and Γ_{hh}, the proportionality factor being denoted as χ

$$\Gamma_{hv} = \Gamma_{vv}\chi e^{-j\phi} \tag{2.71}$$

$$\Gamma_{vh} = \Gamma_{hh}\chi e^{-j\phi} \tag{2.72}$$

with ϕ being independently uniformly distributed over $[0, 2\pi)$.

Quantitatively speaking, ϵ has a standard deviation of 0.3 radians, while μ and χ appear lognormally distributed, with average values of $-8\,\mathrm{dB}$ and $-13\,\mathrm{dB}$, respectively, and a standard deviation of 0.5 [OEP04].

From the scattering matrix Γ_0, it is easy to obtain the scattering matrix Γ corresponding to any dual-polarized scheme $(\zeta, \pi/2 - \zeta) \to (\gamma, \pi/2 - \gamma)$, ζ and γ designating one of the polarization angles of the scheme relative to the vertical direction, respectively at transmit and receive sides (the second polarization is respectively $\pi/2 - \zeta$ and $\pi/2 - \gamma$). The scattering matrix Γ reads as

$$\Gamma = \begin{bmatrix} \cos\gamma & \sin\gamma \\ \sin\gamma & -\cos\gamma \end{bmatrix} \Gamma_0 \begin{bmatrix} \cos\zeta & \sin\zeta \\ \sin\zeta & -\cos\zeta \end{bmatrix}. \tag{2.73}$$

An alternative model is also proposed in [SZM^{+}06], using a geometry-based formalism.

2.4 Empirical models

Empirical models are based on experimental results. Either they generalize the tap-delay-line concept to include the directional domains, or use a combination of tap-delay line and geometry-based models with prescribed parameters. As a consequence, it would be pretty difficult to design a space-time code based on these models. Still, these models are very useful for *a posteriori* simulation purposes. They also constitute the background of many standardized channel models.

2.4.1 *Extended Saleh-Valenzuela model*

In [SV87], a non-directional propagation tap-delay model has been proposed for indoor scenarios. The multipath components are experimentally found to arrive in groups, or clusters of scatterers. In [SJJS00,WJ02], it is further observed that clustering also takes place in the angular domain. This results in the following channel impulse response model:

$$h(t, \tau, \Theta_t, \Theta_r) = \sum_{l=0}^{L-1} \sum_{k=0}^{K-1} \alpha_{kl}\, \delta(\tau - \tau_l - T_{kl})\, \delta(\Theta_t - \Theta_{t,l} - \varpi_{t,kl})\, \delta(\Theta_r - \Theta_{r,l} - \varpi_{r,kl}) \tag{2.74}$$
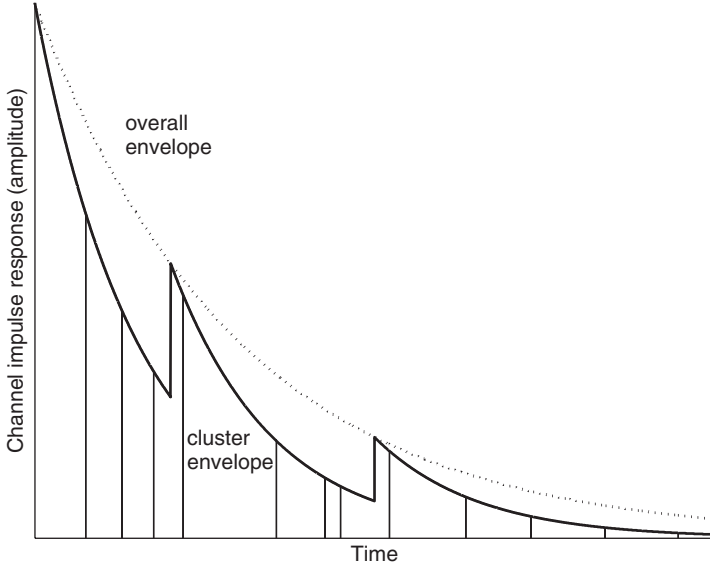
Figure 2.8 *Illustration of exponential decay of mean cluster amplitude and ray amplitude within clusters*

where τ_l, $\Theta_{t,l}$ and $\Theta_{r,l}$ are the time of arrival (ToA), DoD and DoA of the lth cluster respectively, while T_{kl}, $\varpi_{t,kl}$ and $\varpi_{r,kl}$ are the relative ToA, DoD and DoA of the kth ray within the lth cluster. This double-exponential decay is illustrated in Figure 2.8.

Variables τ_l and T_{kl} are described by independent inter-arrival probability density functions

$$p\left(\tau_l|\tau_{l-1}\right) = \upsilon \exp\left[-\upsilon(\tau_l - \tau_{l-1})\right] \tag{2.75}$$

$$p\left(T_{kl}|T_{(k-1)l}\right) = \Upsilon \exp\left[-\Upsilon(T_{kl} - T_{(k-1)l})\right], \tag{2.76}$$

where, by definition, $\tau_0 = 0$ and $T_{0l} = 0$. The angular variables $\Theta_{t,l}$ and $\Theta_{r,l}$ are modeled as uniformly distributed over $[0, 2\pi)$, while the relative angles $\varpi_{t,kl}$ and $\varpi_{r,kl}$ have been experimentally found to follow a two-sided Laplacian distribution

$$p_\varpi(\varpi) = \frac{1}{\sqrt{2}\sigma_\varpi} \exp\left[-|\sqrt{2}\varpi/\sigma_\varpi|\right], \tag{2.77}$$

where σ_ϖ is the angular standard deviation. Multipaths α_{kl} are complex Gaussian variables, whose mean power relative to the first path is given by

$$\frac{\mathcal{E}\left\{|\alpha_{kl}|^2\right\}}{\mathcal{E}\left\{|\alpha_{00}|^2\right\}} = e^{-\tau_l/\Gamma} e^{-T_{kl}/\gamma}, \tag{2.78}$$

where Γ and γ are power-delay time constants, respectively, for clusters and rays. Typical values for the model parameters are listed in Table 2.1.

Table 2.1   *Typical parameters for the extended Saleh-Valenzuela model*

Parameter	Experimental value [SJJS00]	Experimental value [SV87]
Γ [ns]	35–80	60
γ [ns]	30–80	20
$1/\Upsilon$ [ns]	17	300
$1/\upsilon$ [ns]	5–7	5
σ_ϖ [degrees]	22–26	–

Table 2.2   *SUI-4 channel parameters (omnidirectional antennas) parameters (omnidirectional antennas)*

	Tap 1	Tap 2	Tap 3
Delay [μs]	0.0	1.5	4.0
Power [dB]	0	−4	−8
90% K-factor	0	0	0
Doppler ν_m [Hz]	0.20	0.15	0.25
Envelope antenna correlation	0.3	0.3	0.3

2.4.2   Stanford University Interim channel models

Stanford University Interim (SUI) channel models have been developed for fixed wireless access networks at 2.5 GHz. They are basically tap-delay line models, with assigned envelope antenna correlation coefficients[2]. Antennas are assumed to be omnidirectional at both sides, and the distance between Tx and Rx is 7 km. The Doppler spectrum is assumed to be given by

$$S(\nu) = 1 - 1.720(\nu/\nu_m)^2 + 0.785(\nu/\nu_m)^4, \nu \leq \nu_m \qquad (2.79)$$
$$= 0, \nu > \nu_m$$

where ν_m is the maximum Doppler frequency.

The models describe six types of channels, numbered from 1 to 6, as detailed in Appendix C. As an example, the parameters of SUI channel 4 are outlined in Table 2.2. What should be emphasized is that each SUI channel is characterized by a single correlation coefficient, irrespective of the antenna configuration, and that no angular information is specified. Therefore, it is always possible to use the non-directional parameters of SUI channel models, and combine them with directional information, such as angular spreads or power spectra, to obtain different values for the channel correlations (as an example, the model of Section 2.3.3 is able to use the non-directional SUI parameters as input in order to estimate the correlations for various values of the LSR).

---

[2] Antenna correlation have been defined in Section 2.1.6.

2.4.3   COST models

The COST 259 directional channel model (DCM) was originally established for the simulation of systems with multiple antenna elements at either the base station or the mobile terminal. The model describes the joint impacts of small-scale as well as large-scale effects, and covers macro, micro and picocellular scenarios. It consists of superposing three different layers.

- At the top layer, there is a distinction between different radio environments, which represent groups of environments that have similar propagation characteristics (Typical Urban, etc.).
- The goal of the second layer is to model non-stationary large-scale effects, i.e. the variations of channel characteristics as the mobile terminal moves over large distances (typically, 100 wavelengths or more). These effects include the appearance/disappearance of remote scattering clusters, shadowing, changes of DoAs or in the delay-spread. The large-scale effects are described by their probability density functions, whose parameters differ for the different radio environments.
- The bottom layer deals with small-scale fading, caused by interference of various multipath components. Statistics of the small-scale fading are determined by the large-scale effects.

At the beginning of a simulation, clusters of scatterers (one local cluster, around the MT, and several remote scatterer clusters) are distributed at random fixed locations in the coverage area, according to a specified probability density function. Each of the clusters has a small-scale averaged delay-angle power spectrum assumed to be separable on the cluster level

$$\mathcal{P}_{h,c}\left(\tau, \boldsymbol{\Omega}_r\right) = \mathcal{P}_c(\tau)\mathcal{A}_{r,c}\left(\boldsymbol{\Omega}_r\right), \tag{2.80}$$

where $\mathcal{P}_c(\tau)$ is taken as exponential and $\mathcal{A}_{r,c}(\boldsymbol{\Omega}_r)$ is Laplacian in azimuth and elevation[3]. Note that (2.80) does not imply that the total delay-angle power spectrum is separable. The scatterer clusters (i.e. $\mathcal{P}_c(\tau)$ and $\mathcal{A}_{r,c}(\boldsymbol{\Omega}_r)$) are parameterized by a cluster RMS delay and angular spread. These intra-cluster spreads are correlated random variables characterized by their joint probability density functions [AMSM02, APM02]. Finally, each scatterer is characterized by a random complex scattering coefficient, usually taken as complex Gaussian distributed.

As the MT moves, the delays and angles between the clusters are obtained deterministically from their position and the positions of BS and MT, while large-scale effects (including changes in the intra-cluster spreads) are obtained stochastically. The angularly resolved complex impulse responses are then calculated similarly as in ray-tracing tools, but with point-like Gaussian scatterers. Analogous to ray-tracing, the COST 259 model provides continuous directional impulse responses (within a certain range of validity). A concise

---

[3] We consider an uplink scenario, with multiple antennas at the BS.

description of the model can be found in [SMB01], while references to detailed descriptions of the final version are given in [Cor01].

The COST 273 model [Cor06] can be seen as the double-directional extension of the COST 259 model. Because it is double-directional, the joint transmit-receive angular power spectrum needs to be accounted for. Hence, the generation of scattering clusters is implemented as to include different scattering mechanisms: local clusters around Tx and/or Rx (with a large angle-spread), single-interaction clusters and twin-clusters for multiple interactions. Single-interaction clusters are located at randomly chosen locations in the 2-D plane, and DoD, DoA and delay are computed by means of geometrical relationships. Multiple interactions are modeled thanks to a twin-cluster concept. Each physical cluster is divided into two clusters, one corresponding to the Tx side, the other corresponding to the Rx side. The advantage is that the angular dispersion can be modeled independently at Tx and Rx, based on the marginal angular power spectra. Note that this does not mean that the joint angular power spectrum can be modeled as the product of the marginal spectra. Indeed, each DoD is related to one DoA, and vice-versa. However, there is no geometrical relationship between DoD, DoA and delay (unlike the case of single-interaction clusters).

2.5   Standardized models

When comparing the performance of different system implementations, it is convenient to use standardized models established by several international organizations. We shall now briefly review three recent efforts in the context of MIMO systems, although it must be emphasized that these models do not help with understanding MIMO propagation concepts.

2.5.1   IEEE 802.11 TGn models

This set of models is an improved and standardized version of the extended Saleh-Valenzuela model, with overlapping clusters in the delay domain. The models have been designed for indoor MIMO wireless local area networks at both 2 and 5 GHz and for bandwidths up to 100 MHz.

Six canonical channels are modeled, covering flat fading, residential, small office, typical office, large office and large open spaces. Tap-delay profiles are represented by means of overlapping clusters (in the delay domain). The number of clusters, the values of DoD and DoA and the cluster angular spreads (as seen from Tx and Rx) are fixed for each cluster of the different canonical channels. Typically, the number of clusters varies from 2 to 6, and the cluster azimuthal spreads are selected in the range of 20 to 40 degrees, and are correlated with the cluster delay-spread. The power angular profile of each cluster is then modeled as Laplacian, as in the original Saleh-Valenzuela model. Note that the global power angular profiles at Tx and Rx are then calculated separately, and assumed to be

statistically independent when computing the channel correlation matrix. A more detailed outline of the model construction, as well as the current version of the detailed parameters, can be found in [Eea04].

2.5.2   IEEE 802.16d/e models

These models [Eea01] are intended for macrocellular fixed wireless access. The targeted scenario is as follows:

- cell size is less than 10 km in radius
- user's antenna is installed under-the-eaves or on the rooftop
- base station height is 15 to 40 m.

In fact, IEEE 802.16 models are an improved version of the SUI channel models, valid for both omnidirectional and directional antennas. The use of directional antennas will cause the global K-factor to increase while the delay-spread will decrease. As an example, Table 2.3 indicates how SUI channel 4 is modified when the terminal antenna has a 30 degree beamwidth. Curiously, correlations are not modified, although one might have expected the correlations to increase as the beamwidth decreases.

An additional feature of IEEE 802.16 standard is a model for the narrowband Ricean K-factor

$$K = K_0 F_s F_h F_b R^\gamma u, \tag{2.81}$$

where

- F_s is a seasonal factor, $F_s = 1.0$ in summer (leaves) and 2.5 in winter (no leaves)
- F_h is the receive antenna height factor, $F_h = 0.46(h/3)$ (h is the receive antenna height in meters)
- F_b is the beamwidth factor, $F_b = -0.62(b/17)$ (b in degrees)
- K_0 and γ are regression coefficients, $K_0 = 10$ and $\gamma = -0.5$
- u is a lognormal variable, i.e., $10\log_{10}(u)$ is a zero-mean normal variable with a standard deviation of 8 dB.

Table 2.3   *SUI-4 channel parameters for a 30 degree beamwidth antenna at the user's end*

	Tap 1	Tap 2	Tap 3
Delay [μs]	0.0	1.5	4.0
Power [dB]	0	-10	-20
90% K-factor	1	0	0
Doppler v_m [Hz]	0.20	0.15	0.25
Envelope antenna correlation	0.3	0.3	0.3

2.5.3　3GPP/3GPP2 spatial channel models

The 3GPP/3GPP2 channel models [CCG$^+$on] are established specifically for the simulation of 5-MHz bandwidth third generation networks in urban and suburban macrocells as well as in urban microcells. The modeling structure is similar to that of the COST 259 directional channel model, although it differs by two important points. First, it is not defined as a continuous model, but prescribes a specific discretized implementation, similar to the 802.11 TGn models (i.e. DoDs, DoAs and azimuth spreads are fixed). However, the model has been constructed so that correlations between the different parameters can be incorporated. Secondly, it does not allow for continuous, large-scale movements of the mobile terminal, but considers different possible positions of the mobile terminal within the cell. The provided implementation is a tap-delay line model. Each tap consists of several sub-paths that share the same delay, but have different directions of arrival and departure. Several options have been defined that can be switched on to obtain a better agreement with real-world data, such as polarized antennas, far scatterer clusters, line-of-sight, and urban canyons. Interference is handled by modeling strong interferers as spatially correlated, while weak interferers are taken as spatially white.

2.6　Antennas in MIMO systems

2.6.1　About antenna arrays

Antenna arrays naturally constitute an essential part of a MIMO system. A large number of different array configurations exist. It is not our goal to study antenna array theory in depth, but it is interesting to highlight a few points related to antenna arrays.

- We have already defined the concept of narrowband and balanced arrays (see (2.5) and (2.6)). Usual arrays are often narrowband for classical systems and are also balanced, although local scattering from the antenna mounting structure may cause some unbalance, even in relatively compact arrays.
- In the above, many expressions are also given for uniform linear arrays, although alternative shapes such as circular arrays are met in practice. As already mentioned, the definition of steering vectors is not restricted to linear arrays. Only two conditions are needed to define a steering vectors, as long as the channel is WSSUSH: a narrowband array and a plane wave incidence.
- However, arrays with complex configurations might not support the homogeneous channel assumption. As an example, an array made of three directional antennas oriented in different directions does not yield a MIMO matrix whose elements describe an homogeneous channel. What does this mean in practice? For such arrays, the definitions discussed in Section 2.1.4 are no longer valid, as the power delay-angle profiles will not be the same across the arrays. For such MIMO systems, the definition of steering vectors is naturally not possible.
- Antennas may be chosen as omnidirectional or directional. Omnidirectional antennas may see more scatterers, but directional antennas would provide more gain. So, there is

no guarantee that omnidirectional antennas yield better performance for a constrained transmit power.

- Antenna spacing might play an important role, which leads to a compromise between the size of the array and the captured diversity. Small spacing is also responsible for antenna coupling, which is dealt with in the next section.

2.6.2 Mutual coupling

When several antenna elements are located closely to each other, the electrical field generated by one antenna alters the current distribution on the other antennas. As a consequence, the radiation pattern and input impedance of each array element are disturbed by the presence of the other elements. This effect is known as antenna or mutual coupling and can be evaluated through the use of a coupling matrix, as shown in Appendix D, for antennas that are minimum scatterers with respect to impedance or admittance parameters. The coupling matrix is a simple approach by which the channel matrix is multiplied by the coupling matrices $\mathbf{M}_t$ and $\mathbf{M}_r$ at transmit and receive sides, respectively

$$\mathbf{H}_{mc} = \mathbf{M}_r \mathbf{H} \mathbf{M}_t \tag{2.82}$$

where $\mathbf{H}_{mc}$ denotes the new channel matrix accounting for mutual coupling effects.

Antenna coupling model
We now deal with antennas that are minimum scatterers with respect to impedance parameters. It is shown in Appendix D that the coupling matrix (at either transmit or receive sides) can be written for two antennas as

$$\mathbf{M} = \begin{bmatrix} a & b \\ b & a \end{bmatrix} = (Z_T + Z_A)(\mathbf{Z} + Z_T \mathbf{I})^{-1} \tag{2.83}$$

where Z_A is the antenna impedance, Z_T is the termination impedance of each element and $\mathbf{Z}$ is the mutual impedance matrix. Note that the proportionality factor $(Z_T + Z_A)$ ensures that $\mathbf{H}_{mc} = \mathbf{H}$ in the hypothetical absence of coupling. As an example, antennas that are minimum scatterers with respect to impedance parameters are half-wavelength dipoles. The expressions of the mutual impedance matrix for such antennas are well-known [Bal05]

$$\mathbf{Z}(1, 1) = \mathbf{Z}(2, 2) = 73.1 + j45.2$$
$$\mathbf{Z}(1, 2) = \mathbf{Z}(2, 1) = r + jq \tag{2.84}$$

with

$$r = 30 \left(2 \mathcal{C}_i \left(\frac{2\pi d}{\lambda} \right) - \mathcal{C}_i \left(\frac{2\pi}{\lambda} \left(\sqrt{d^2 + h^2} + h \right) \right) \right.$$
$$\left. - \mathcal{C}_i \left(\frac{2\pi}{\lambda} \left(\sqrt{d^2 + h^2} - h \right) \right) \right) \tag{2.85}$$
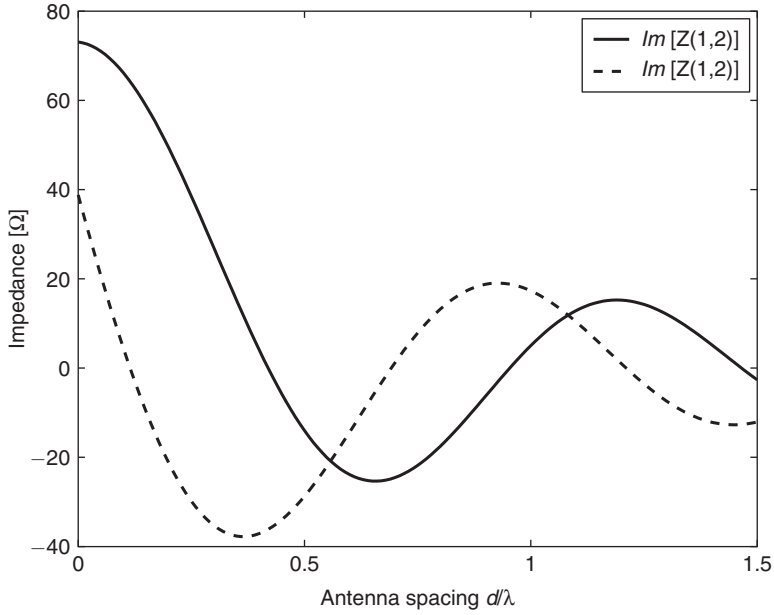
Figure 2.9   *Mutual dipole impedance as a function of antenna spacing relative to the wavelength*

$$q = -30 \left(2\mathcal{S}_i \left(\frac{2\pi d}{\lambda} \right) - \mathcal{S}_i \left(\frac{2\pi}{\lambda} \left(\sqrt{d^2 + h^2} + h \right) \right) \right.$$

$$\left. -\mathcal{S}_i \left(\frac{2\pi}{\lambda} \left(\sqrt{d^2 + h^2} - h \right) \right) \right) \tag{2.86}$$

where $\mathcal{S}_i$ and $\mathcal{C}_i$ are the sine and cosine integrals

$$\mathcal{C}_i(x) = -\int_x^\infty \frac{\cos(y)}{y} dy \tag{2.87}$$

$$\mathcal{S}_i(x) = -\int_0^x \frac{\sin(y)}{y} dy \tag{2.88}$$

and d and h denote the array inter-element spacing and the length of the antenna, respectively. The mutual impedance $\mathbf{Z}(1,2)$ is represented in Figure 2.9 for a half-wave dipole. A high degree of coupling can be observed, with the mutual impedance remaining significant even for spacings larger than one wavelength.

It is important to note that, while we focus on minimum scatterers with respect to impedance parameters here, the expressions given in the following as a function of a and b are valid for all antennas that are minimum scatterers. In particular, the derivation of $\mathbf{M}$ for minimum scatterers with respect to admittance parameters is outlined in Appendix D. An exact minimum scatterer antenna with respect to admittance parameters is the slot-type antenna.
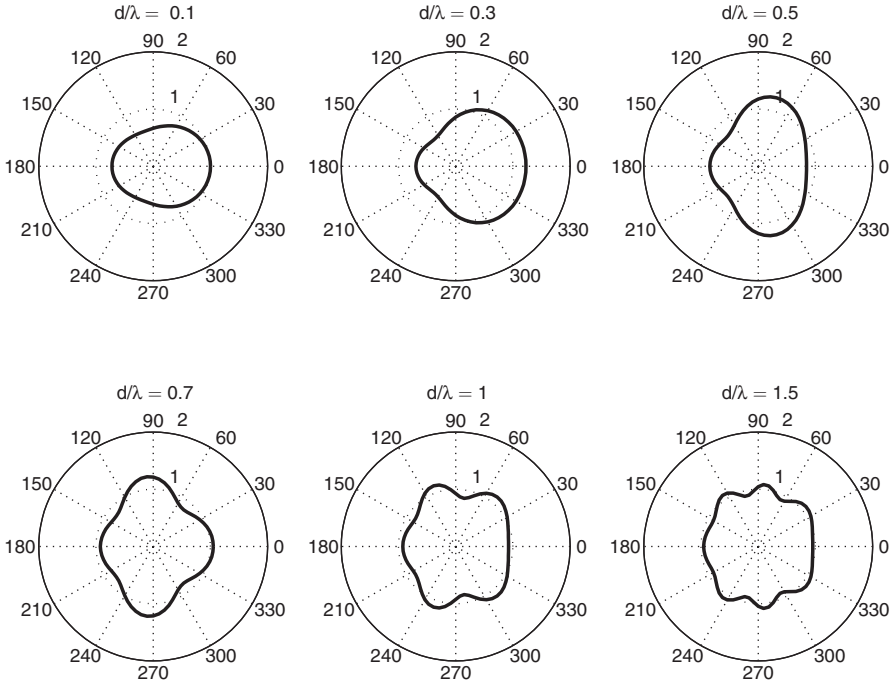
Figure 2.10 *Field-radiation pattern (in magnitude) of the right-hand side antenna for several inter-element spacings*

To facilitate the physical interpretation of coupling, it is interesting to consider antenna coupling as a distortion of the original omnidirectional radiation patterns. The coupling matrix yields two new patterns, $g_1(\theta)$ and $g_2(\theta)$ (see Appendix D), with θ being the local angular coordinate relative to the array baseline. Figure 2.10 presents the radiation pattern (in magnitude) of the right-hand side antenna, say antenna 2, for several inter-element spacings and $Z_T = Z_0 = 50$ ohms. The radiation pattern of the left-hand side antenna (antenna 1) is naturally symmetrical with respect to the 90–270° axis to that of antenna 2.

The impacts of antenna coupling are clearly visible:

- the antenna gain is modified, which will affect the radiation efficiency [RK05] and the received power for a fixed transmit power constraint
- the pattern is no longer omnidirectional; this will modify the channel correlations.

Both effects are concurrent, and both will likely impact the system performance, e.g. the capacity or the transmission error rate.

Impact of antenna termination

So far, we have assumed that antenna are terminated by an impedance Z_T, which we fixed to Z_0 in the simulations of Figure 2.10. However, a true upper-bound on system

performance may be achieved through optimal loading networks. In all generality, we evaluate the impact of inserting a matching network between the output ports of the receive antenna and the termination impedance chosen as $Z_T = Z_0$. A natural representation of the communication system is thus obtained by using the scattering parameter[4] (S-parameter). This S-parameter expresses the relationship between the complex envelopes of generalized inward and outward-propagating waves, denoted as **a** and **b**,

$$\mathbf{b} = \mathbf{Sa} \qquad (2.89)$$

where **S** is the element scattering matrix, which is unequivocally related to the impedance matrix, while waves **a** and **b** are related to the input/output currents and voltages at the different ports [Poz98].

The antenna/channel combination is represented by S-matrix $\mathbf{S}_H$, which is partitioned into four sub-matrices [WJ04] as follows

$$\mathbf{S}_H = \begin{bmatrix} \mathbf{S}_t & \mathbf{S}_{tr} \\ \mathbf{S}_{rt} & \mathbf{S}_r \end{bmatrix}, \qquad (2.90)$$

where $\mathbf{S}_t$ and $\mathbf{S}_r$ represent the Tx and Rx antenna reflection coefficients (in the formalism of transmission lines), $\mathbf{S}_{rt}$ is the classical MIMO channel matrix (including antennas) expressed in terms of a S-matrix, and $\mathbf{S}_{tr}$ is assumed to be equal to a null matrix. This last assumption basically means that power reflected by the receive antennas does not couple back into the transmit antennas, and holds true if Tx and Rx arrays are sufficiently separated, presumably in the far-field of each other. For $Z_T = Z_0$, the channel scattering matrix (including mutual coupling effects) is given by

$$\mathbf{S}_{rt} = (Z_0\mathbf{I} + \mathbf{Z}_r)^{-1}\,\mathbf{Z}_{rt}\,(\mathbf{I} - \mathbf{S}_t), \qquad (2.91)$$

where $\mathbf{Z}_t$ and $\mathbf{Z}_r$ are the mutual impedance matrices of transmit and receive arrays, and $\mathbf{Z}_{rt}$ is the channel transimpedance matrix (i.e. between transmit currents and received voltages). Note that this expression is more general than the combination of (2.82) and (2.83), as it is also valid for antennas which are not minimum scatterers. However, in this case, the derivation of $\mathbf{Z}_{rt}$ requires numerical computation techniques.

For minimum scatterer antennas with respect to impedance parameters, (2.91) is equivalent to (2.82) where $\mathbf{M}_t$ and $\mathbf{M}_r$ are given by (2.83). Indeed, $\mathbf{Z}_{rt}$ is independent from coupling for minimum scatterer antennas with respect to impedance parameters, as each element is defined by canceling the transmitted currents on all other antennas (by definition of minimum scatterer antennas with respect to impedance parameters, these antennas are

---

[4] Although its definition uses the term *scattering*, this parameter should not be confused with our scattering coefficient defined in the previous section. Naturally, both definitions are not unrelated but that is beyond the scope of this section.

therefore invisible). Hence, $\mathbf{Z}_{rt}$ is simply proportional to the channel transfer matrix, $\mathbf{Z}_{rt} \propto \mathbf{H}$. On the other hand

$$\mathbf{S}_t = \frac{\mathbf{Z}_t - Z_0\mathbf{I}}{\mathbf{Z}_t - Z_0\mathbf{I}}, \tag{2.92}$$

by definition of the S-matrix. Hence, inserting (2.92) into (2.91), one obtains

$$\mathbf{H}_{mc} \propto \mathbf{S}_{rt} \propto (Z_0\mathbf{I} + \mathbf{Z}_r)^{-1}\,\mathbf{H}(Z_0\mathbf{I} + \mathbf{Z}_t)^{-1}. \tag{2.93}$$

The advantage of (2.91) is that it allows us to account for a matching network. The latter is characterized by S-matrix $\mathbf{S}_M$

$$\mathbf{S}_M = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}, \tag{2.94}$$

where subscripts 1 and 2, respectively, refer to input and output ports. The channel matrix in the presence of mutual coupling (at both Tx and Rx) and a matching network at the receiver is then given [WJ04] by

$$\mathbf{H}_{mcm} = \mathbf{S}_{21}(\mathbf{I} - \mathbf{S}_{rr}\mathbf{S}_{11})^{-1}\,\mathbf{S}_{rt}. \tag{2.95}$$

Based on the above expression, true upper-bounds of the system performance can be derived as the matching network is then optimized to achieve capacity.

This page intentionally left blank

<div style="text-align: right">**3**</div>

Analytical MIMO channel representations for system design

Physical channel models described in Chapter 2 are mostly used for simulating system performance. However, they do not allow for the explicit design of space–time coding techniques. In this respect, mathematical representations are required. While physical models reproduce the properties of the MIMO channel by specifying the locations of obstacles and the array configuration, so-called analytical models provide a mathematical representation of the channel matrix. A major difference between both types of modeling approaches lies in the fact that analytical models only provide a framework, but do not give fully quantitative information, unless they are parametrized by means of extrinsic information. By contrast, models such as the COST 273 model, or a ray-tracing tool, directly yield values of essential characteristics of the MIMO channel (correlations, mutual information, etc.), as a function of geometrical and array parameters. A typical analytical model will mathematically express the MIMO channel matrix as a function of a random Gaussian-fading matrix and various channel correlations or steering vectors, without specifying any value for these parameters. Since the correlations depend on the actual antenna configuration used, these models cannot be generalized easily to other configurations, unless new correlation coefficients are extracted from experimental results or physical models. Yet, analytical models are particularly useful when analyzing mathematically the impact of correlations on any performance parameter, since the relationship is then explicit. However, the link between system performance and antenna configuration can only be established if the relationship between antenna configuration and correlations in the environment under study is known.

3.1 General representations of correlated MIMO channels

Before detailing a number of analytical models, it is important to emphasize a few assumptions used in this chapter as well as in Chapter 4.

1. We deal with wide-sense stationary uncorrelated scattering and generally homogeneous processes. Non-stationarity is addressed in Section 3.3.3.
2. Normalization matters. We usually normalize the average squared channel Frobenius norm of uni-polarized channels as follows

$$\mathcal{E}\left\{\|\mathbf{H}\|_F^2\right\} = n_t n_r. \tag{3.1}$$

   This normalization has no impact on the channel model itself, as it only represents a scaling factor of the whole matrix. It is nevertheless highly relevant when considering transmit power constraints, path-loss effects, especially when estimating the mutual information or when comparing uni-polarized with dual-polarized schemes.
3. We consider, for the sake of simplicity, that all individual identically-polarized channels are characterized by equal average power. This last assumption might not be true in real-world correlated channels, and should not be taken as a general rule. Still, it greatly simplifies the presentation and comparison of different models.

Furthermore, it must be mentioned that analytical models are by essence stochastic and narrowband models. They represent the distribution of either $\mathbf{H}(t)$ or $\mathbf{H}_k$, denoted simply as $\mathbf{H}$ for better legibility. Frequency-selective representations are naturally obtained in the sampled-delay domain by combining L narrowband representations, also known as *taps*, as denoted by (2.34). The lth tap is then characterized on its own by a narrowband channel model, which implies that each tap is characterized by a spatial correlation matrix $\mathbf{R}_l$.

3.1.1   Rayleigh fading channels

As in the case of single-antenna channels, the Rayleigh fading assumption is often used by MIMO system designers, mostly because it is realistic in environments rich in scatterers. It corresponds to modeling the narrowband transmission between a transmit and a receive antenna as the sum of a large number of contributions with random and statistically independent phases, directions of departure and directions of arrival [Par00]. In WSSUSH Rayleigh fading channels, each individual channel is thus a zero-mean complex circularly symmetric Gaussian variable, or equivalently a complex variable whose amplitude and phase respectively, are, Rayleigh and uniformly distributed.

I.i.d. Rayleigh channels

When the antenna spacings and/or the angular spreading of the energy at both sides of the link are large enough, the various channel correlations become very small, and can be assumed as equal to zero (as observed in Section 2.1.6). Furthermore, if all individual channels are characterized by the same average power (i.e. the antenna arrays should be balanced), then the correlation matrix $\mathbf{R}$ is proportional to the identity matrix. Using the normalization of (3.1), we actually have $\mathbf{R} = \mathbf{I}_{n_t n_r}$, and we denote the channel matrix as $\mathbf{H}_w$, which is therefore a random fading matrix with unit variance and circularly

symmetric complex Gaussian entries. The so-called independent identically distributed (i.i.d.) Rayleigh assumption has been (and still is) used extensively when designing space–time codes. However, we must stress that real-world channels sometimes significantly deviate from this ideal channel owing to a number of reasons:

- limited angular spread and/or reduced array sizes cause the channels to become correlated (channels are not independent anymore)
- a coherent contribution may induce the channel statistics to become Ricean (channels are not Rayleigh distributed anymore)
- the use of multiple polarizations creates gain imbalances between the various elements of the channel matrix (channels are no longer identically distributed).

Correlated Rayleigh channels

As a consequence of the complex Gaussian (Rayleigh) assumption, the $n_t n_r \times n_t n_r$ positive semi-definite Hermitian correlation matrix $\mathbf{R}$ defined by (2.41) constitutes a sufficient description of the stochastic behavior of the MIMO channel. In a 2×2 system with equal power channels, $\mathbf{R}$ is a 4×4 matrix, outlined as

$$\mathbf{R} = \mathcal{E}\big\{\mathrm{vec}\big(\tilde{\mathbf{H}}^H\big)\mathrm{vec}\big(\tilde{\mathbf{H}}^H\big)^H\big\} = \begin{bmatrix} 1 & t_1^* & r_1^* & s_1^* \\ t_1 & 1 & s_2^* & r_2^* \\ r_1 & s_2 & 1 & t_2^* \\ s_1 & r_2 & t_2 & 1 \end{bmatrix} \tag{3.2}$$

where $\tilde{\mathbf{H}}$ is the notation used to represent the particular case of Rayleigh fading channel matrices, and

- $t_1 = \mathcal{E}\big\{\tilde{\mathbf{H}}(1,1)\tilde{\mathbf{H}}^*(1,2)\big\}$ and $t_2 = \mathcal{E}\big\{\tilde{\mathbf{H}}(2,1)\tilde{\mathbf{H}}^*(2,2)\big\}$ are defined as the transmit correlations as observed respectively from receive antennas 1 and 2
- $r_1 = \mathcal{E}\big\{\tilde{\mathbf{H}}(1,1)\tilde{\mathbf{H}}^*(2,1)\big\}$ and $r_2 = \mathcal{E}\big\{\tilde{\mathbf{H}}(1,2)\tilde{\mathbf{H}}^*(2,2)\big\}$ are defined as the receive correlations as observed respectively from transmit antennas 1 and 2
- $s_1 = \mathcal{E}\big\{\tilde{\mathbf{H}}(1,1)\tilde{\mathbf{H}}^*(2,2)\big\}$ and $s_2 = \mathcal{E}\big\{\tilde{\mathbf{H}}(1,2)\tilde{\mathbf{H}}^*(2,1)\big\}$ are defined as the cross-channel correlations [OP04].

From (2.41), any channel realization is obtained by

$$\mathrm{vec}\big(\tilde{\mathbf{H}}^H\big) = \mathbf{R}^{1/2}\mathrm{vec}\big(\mathbf{H}_w^H\big), \tag{3.3}$$

where $\mathbf{H}_w$ is one realization of an i.i.d. channel matrix.

While the representation of (3.3) is the most general for identically-distributed Rayleigh channels, some channels do not match the identically-distributed Rayleigh assumption. Analytical formulations for Ricean fading, dual-polarized, and keyhole channels are outlined on the next pages.

3.1.2   Ricean fading channels

The Rayleigh assumption typically holds in mobile scenarios, as well in deeply-shadowed fixed wireless links [Eea01]. However, there are situations in mobile cellular networks where there may exist a strong coherent component. This component, which does not experience any fading over time, consists of, for example, a line-of-sight field, or one or several specular contributions. Similarly, in fixed wireless access scenarios, most reflected and diffracted contributions add coherently, as the transmitter and the receiver are fixed. Non-coherent scattered contributions are only caused by the motion of a few obstacles (car, people, tree leaves, etc.).

All these situations lead to a Ricean distribution of the received field amplitude. The matrix $\bar{\mathbf{H}}$ corresponding to the coherent component(s) has fixed phase-shift-only entries [PNG03]

$$\bar{\mathbf{H}} = \begin{bmatrix} e^{j\alpha_{11}} & e^{j\alpha_{12}} \\ e^{j\alpha_{21}} & e^{j\alpha_{22}} \end{bmatrix}. \tag{3.4}$$

The values of the phase-shifts α_{nm} are strongly related to the array configuration and its orientation with respect to the direction of the dominant component(s). When there is a single dominant component, such as a line-of-sight, $\bar{\mathbf{H}}$ is poorly conditioned, thereby decreasing the multiplexing gain for a given received SNR. Indeed, with only one coherent contribution with given DoD and DoA ($\Omega_{t,c}$ and $\Omega_{r,c}$),

$$\bar{\mathbf{H}} = \mathbf{a}_r\left(\Omega_{r,c}\right) \mathbf{a}_t^T\left(\Omega_{t,c}\right) \tag{3.5}$$

if we assume a sufficiently large separation between Tx and Rx. The above expression clearly denotes a rank decrease. As an example, for broadside arrays, the Ricean matrix corresponding to a pure line-of-sight component is given by $\bar{\mathbf{H}} = \mathbf{1}_{n_r \times n_t}$.

Combining $\bar{\mathbf{H}}$ with the non-coherent contribution $\tilde{\mathbf{H}}$ yields [RFLFV00]

$$\mathbf{H} = \sqrt{\frac{K}{1+K}}\bar{\mathbf{H}} + \sqrt{\frac{1}{1+K}}\tilde{\mathbf{H}} \tag{3.6}$$

where $\mathcal{E}\{\mathbf{H}\} = \sqrt{\frac{K}{1+K}}\,\bar{\mathbf{H}}$ and K is the so-called Ricean K-factor. Note that the homogeneous behavior of the channel is essential to write (3.6), as it allows for the definition of a unique K-factor K.

3.1.3   Dual-polarized channels

As already mentioned in Sections 2.2.2 and 2.3.5, the use of antennas with different polarizations may lead to power and correlation imbalance between the elements of the channel matrix. Analogous to Section 2.3.5, we initially restrict our analysis to 2×2 MIMO channels, and we denote the dual-polarized channel matrix as $\mathbf{H}_{\times,a}$. This channel matrix corresponds to a system for which both the transmit and receive arrays are made of

two antennas, co-localized or not, with orthogonal polarizations. The polarizations at both ends do not need to be identical. For example, the transmit polarization scheme may be vertical–horizontal (denoted as VH), while the received scheme is chosen as slanted (± 45 degrees).

To account for depolarization in analytical representations, we must consider depolarization caused by the non-ideal antennas as well as by the scattering medium. Regarding the first effect, it has been illustrated earlier (see Section 2.2.2) that antenna XPD is easily included in a physical model by means of the cross-polar antenna pattern. Analytically, this can be approximated on the average by a scalar antenna depolarization factor χ_a. The latter is used to build an antenna depolarization matrix $\mathbf{X}_a$, which pre or post-multiplies the channel matrix (this is similar to the coupling matrix introduced in Section 2.6.2)

$$\mathbf{X}_a = \begin{bmatrix} 1 & \sqrt{\chi_a} \\ \sqrt{\chi_a} & 1 \end{bmatrix}. \tag{3.7}$$

As an example, considering antenna coupling at the receiver yields the following model

$$
\begin{aligned}
\mathbf{H}_{\times,a} &= \mathbf{X}_a \mathbf{H}_\times \\
&= \sqrt{\frac{K}{1+K}} \underbrace{\mathbf{X}_a \bar{\mathbf{H}}_\times}_{\bar{\mathbf{H}}_{\times,a}} + \sqrt{\frac{1}{1+K}} \underbrace{\mathbf{X}_a \tilde{\mathbf{H}}_\times}_{\tilde{\mathbf{H}}_{\times,a}}
\end{aligned}
\tag{3.8}
$$

In what follows, we assume that antenna cross-polar coupling is negligible (i.e. $\chi_a = 0$) at both link ends, and concentrate on modeling $\tilde{\mathbf{H}}_\times$ and $\bar{\mathbf{H}}_\times$.

Dual-polarized Rayleigh fading channels

Let us start with the Rayleigh (non-coherent) fading part of the dual-polarized channel matrix, denoted as $\tilde{\mathbf{H}}_\times$. Any model should account for three mechanisms

- the spatial correlation arising from the finite spacing between the antennas (if dual-polarized antennas are co-located, this correlation is equal to one)
- the gain imbalance between the various co and cross-polar components
- the (de)correlation between all pairs of co and cross-polar antennas arising only from the polarization difference (i.e. for co-located dual-polarized antennas).

In a first approach [PNG03, LTV03], $\tilde{\mathbf{H}}_\times$ is decomposed by extracting the impact of depolarization on the channel gains, yielding

$$\tilde{\mathbf{H}}_\times = |\tilde{\mathbf{X}}| \odot \tilde{\mathbf{H}}', \tag{3.9}$$

where $|\tilde{\mathbf{X}}|$ depends on the polarization scheme. For a slanted-to-slanted scheme (± 45 degrees at both Tx and Rx), it is naturally given by

$$\left|\tilde{\mathbf{X}}_{\pm 45° \to \pm 45°}\right| = \begin{bmatrix} 1 & \sqrt{\chi} \\ \sqrt{\chi} & 1 \end{bmatrix}, \tag{3.10}$$

where χ is the global real-valued depolarization factor ($0 < \chi < 1$) for a slanted-to-slanted scheme. What is important to notice is that $\tilde{\mathbf{H}}'$ still includes two correlation mechanisms (space and polarization). Hence, it is generally not equal to an equivalent uni-polarized transmission matrix $\tilde{\mathbf{H}}$ (i.e. with the same antenna spacings, all polarizations then being identical). As a result, $\tilde{\mathbf{H}}'$ is a hybrid matrix, modeling the correlation aspects of both spacing and polarization. How the correlations in $\tilde{\mathbf{H}}'$ depend on the spatial and polarization correlations is not known *a priori*, except in one particular case. Indeed, if we assume that the use of orthogonal polarizations totally decorrelates all individual channels, $\tilde{\mathbf{H}}' = \mathbf{H}_w$, where $\mathbf{H}_w$ is the classical i.i.d. complex Gaussian matrix, irrespective of the antenna spacing.

In a more general modeling approach, a second model explicitly separates spacing-related and polarization-related effects. The separation is thus operated based on the physical mechanisms (space versus polarization) rather than on their impact (gain versus correlation). Subsequently, the dual-polarized Rayleigh channel matrix may be rewritten as

$$\tilde{\mathbf{H}}_{\times} = \tilde{\mathbf{X}} \odot \tilde{\mathbf{H}}. \tag{3.11}$$

In (3.11), $\tilde{\mathbf{H}}$ is modeled as a uni-polarized correlated Rayleigh channel, while $\tilde{\mathbf{X}}$ models both the correlation and power imbalance impacts of scattering-induced depolarization. This principle is represented in Figure 3.1. A relatively general model of the channel matrix for VH-to-VH downlink transmissions is given by

$$\mathrm{vec}\left(\tilde{\mathbf{X}}_{\mathrm{VH} \to \mathrm{VH}}^{H}\right) =$$

$$\begin{bmatrix} 1 & \sqrt{\mu\chi}\vartheta^* & \sqrt{\chi}\sigma^* & \sqrt{\mu}\delta_1^* \\ \sqrt{\mu\chi}\vartheta & \mu\chi & \sqrt{\mu}\chi\delta_2^* & \mu\sqrt{\chi}\sigma^* \\ \sqrt{\chi}\sigma & \sqrt{\mu}\chi\delta_2 & \chi & \sqrt{\mu\chi}\vartheta^* \\ \sqrt{\mu}\delta_1 & \mu\sqrt{\chi}\sigma & \sqrt{\mu\chi}\vartheta & \mu \end{bmatrix}^{1/2} \mathrm{vec}\left(\mathbf{X}_w^H\right), \tag{3.12}$$

where

- μ and χ represent the co-polar imbalance and the scattering XPD, respectively, and are assumed to be constant
- σ and ϑ are the receive and transmit correlation coefficients (i.e. the correlation coefficients between VV and HV, HH and HV, VV and VH or HH and VH)
- δ_1 and δ_2 are the cross-channel correlation coefficients caused by the use of orthogonal polarizations, i.e. δ_1 is the correlation between the VV and the HH components, and δ_2 is the correlation between the VH and the HV components
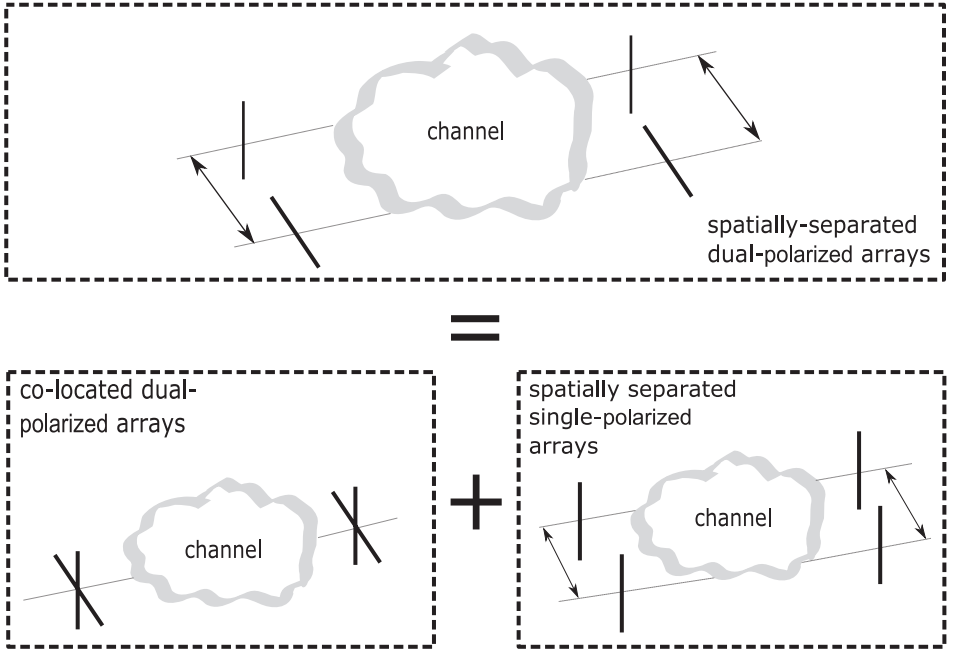
Figure 3.1 *XPD modeling: separation between space and polarization*

- $\mathbf{X}_w$ is a 2×2 matrix whose four elements are independent circularly symmetric complex exponentials of unit amplitude, $e^{j\phi_k}$, $k = 1, \ldots, 4$, the angles ϕ_k being uniformly distributed over $[0, 2\pi)$.

In this book, we use a simplified version of the above model. We indeed assume that the correlation coefficients σ and ϑ between any cross-polar component (VH or HV) and any co-polar component (VV or HH) are equal to zero, although they might actually be slightly higher [KST06]. Furthermore, if $\delta_1 = \delta_2 = 0$, we may also write that $\mathbf{H}_\times = |\tilde{\mathbf{X}}| \odot \mathbf{H}_w$.

Note that while we use the same notations as in Section 2.3.5 for μ and χ, the variables are not identical. In Section 2.3.5, they have been used to describe the matrix scattering coefficient $\mathbf{\Gamma}_0$, while (3.12) aims at modeling the global narrowband dual-polarized channel matrix for co-located antennas. Both models are naturally related, as the narrowband channel is actually the sum over all scattered contributions.

For alternative polarization schemes, the Rayleigh channel matrix is simply obtained by applying adequate rotations

$$\tilde{\mathbf{X}}_{(\zeta, \pi/2 - \zeta) \to (\gamma, \pi/2 - \gamma)} = \begin{bmatrix} \cos \gamma & \sin \gamma \\ \sin \gamma & -\cos \gamma \end{bmatrix} \tilde{\mathbf{X}}_{VH \to VH} \begin{bmatrix} \cos \zeta & \sin \zeta \\ \sin \zeta & -\cos \zeta \end{bmatrix} \tag{3.13}$$

where ζ and γ are one of the polarization angles of the scheme relative to the vertical direction at transmit and receive sides, respectively (the second polarization is $\pi/2 - \zeta$

and $\pi/2 - \gamma$). The correlation matrix $\mathbf{R}_{\tilde{\mathbf{X}}}$ of $\tilde{\mathbf{X}}$ can be directly calculated from the above modeling. In particular, when $\delta_1 = \delta_2 = 0$, $\mathbf{R}_{\tilde{\mathbf{X}}}$ becomes for $\pm 45° \to \pm 45°$ schemes

$$
\mathbf{R}_{\tilde{\mathbf{X}}} = \frac{1}{4} \begin{bmatrix} (1+\mu)^2(1+\chi^2) & (1-\mu^2)(1+\chi^2) \\ (1-\mu^2)(1+\chi^2) & (1-\mu)^2(1+\chi^2) \\ (1-\mu^2)(1-\chi^2) & (1-\mu)^2(1-\chi^2) \\ (1+\mu)^2(1-\chi^2) & (1-\mu^2)(1-\chi^2) \end{bmatrix}
$$

$$
\begin{bmatrix} (1-\mu^2)(1-\chi^2) & (1+\mu)^2(1-\chi^2) \\ (1-\mu)^2(1-\chi^2) & (1-\mu^2)(1-\chi^2) \\ (1-\mu)^2(1+\chi^2) & (1-\mu^2)(1+\chi^2) \\ (1-\mu^2)(1+\chi^2) & (1+\mu)^2(1+\chi^2) \end{bmatrix} \tag{3.14}
$$

Note that if $\mu \neq 1$, the resulting transmit and receive antenna correlations are higher in $\pm 45° \to \pm 45°$ schemes than in $VH \to VH$ schemes, in favor of a lower gain imbalance between the co-polarized components.

A more elaborate model might use μ and χ as random variables, but that would significantly increase the modeling complexity.

Arbitrary $n_r \times n_t$ schemes (for even values of n_t and n_r) are modeled by considering that the transmit (resp. receive) array is made of $n_t/2$ (resp. $n_r/2$) dual-polarized sub-arrays. Hence, the global channel matrix is represented as

$$
\tilde{\mathbf{H}}_{\times, n_r \times n_t} = \tilde{\mathbf{H}}_{n_r/2 \times n_t/2} \otimes \tilde{\mathbf{X}}, \tag{3.15}
$$

where $\tilde{\mathbf{H}}_{n_r/2 \times n_t/2}$ is related to the spacing between the sub-arrays, and $\tilde{\mathbf{X}}$ is the 2×2 dual-polarized matrix modeled by (3.12).

Dual-polarized Ricean fading channels
So far, we have assumed Rayleigh fading, which corresponds to non-line-of-sight mobile scenarios or large range fixed scenarios. In Ricean fading, we express the Ricean (LOS) dual-polarized matrix for an arbitrary 2×2 symmetric scheme $(\zeta, \pi/2 - \zeta) \to (\zeta, \pi/2 - \zeta)$ neglecting antenna XPD as

$$
\bar{\mathbf{H}}_{\times} = \bar{\mathbf{X}}_{(\zeta, \pi/2-\zeta) \to (\zeta, \pi/2-\zeta)} = \begin{bmatrix} e^{j\alpha_{11}} & 0 \\ 0 & e^{j\alpha_{22}} \end{bmatrix}, \tag{3.16}
$$

where α_{11} and α_{22} depend on the array orientation (e.g. $\alpha_{11} = \alpha_{22} = 0$ for broadside arrays at both ends, so $\bar{\mathbf{X}}_{(\zeta, \pi/2-\zeta) \to (\zeta, \pi/2-\zeta)} = \mathbf{I}_2$ in such cases). When $\gamma \neq \zeta$, the Ricean matrix becomes fully structured analogous to uni-polarized schemes.

3.1.4   Double-Rayleigh fading model for keyhole channels

Models based on Ricean/Rayleigh fading statistics are not capable of reproducing the keyhole effect. The keyhole effect is characterized by a rank-deficient channel matrix, even though the latter shows low transmit and receive correlations. Physically, this would correspond to scenarios with rich scattering conditions at both Tx and Rx, but where all paths propagating from the transmit scattering area to the receive scattering area are forced to a small aperture. This effect was predicted theoretically in [CLW+03,GBGP02], and experimentally demonstrated in artificial scenarios [ATM03].

This mechanism can be modeled by replacing the i.i.d. complex Gaussian matrix by a rank-deficient matrix [GBGP02]

$$\mathbf{H}_{kh} = \left(\mathbf{H}_{w,r}\right)^{1/2} \mathbf{\Omega}_{kh} \left(\mathbf{H}_{w,t}\right)^{1/2}, \qquad (3.17)$$

where $\mathbf{\Omega}_{kh}$ denotes the scattering correlation matrix, i.e. modeling the propagation from the scatterers near the transmitter to the scatterers near the receiver, and the two other matrices are i.i.d. random complex Gaussian matrices. This model highlights two important properties of keyholes: the amplitude statistics are not Rayleigh (so second-order characteristics are not sufficient) but double-Rayleigh (see (2.64)), and the rank of the instantaneous transfer matrix is limited by the rank of the scattering correlation matrix, while the marginal correlation matrices at the link ends are not affected.

An interesting point about keyholes concerns the frequency of occurrence of such channels in real-world scenarios. In [CLW+03], experimentally observed keyholes have been reported. However, these are not true keyholes, but pseudo-keyholes, i.e. situations where the transmit and receive correlations are apparently small, while the rank of $\mathbf{H}$ was simultaneously rather low. In [ATM03], an artificial keyhole has been created at 2.5 GHz by placing a waveguide operating in the dominant mode between two adjacent rooms (one of them shielded). Note that the keyhole only appears when using a waveguide. If the waveguide is replaced by a hole (or a larger waveguide), multi-mode propagation takes place, and the keyhole vanishes. Apart from this rather curious situation, no true keyhole has so far, ever been measured, indoors or outdoors. Therefore, despite their theoretical interest, the occurrence of keyholes in real-world channels may be neglected when designing space–time codes.

3.2   Simplified representations of Gaussian MIMO channels

The vectorized formulation of (3.3) is not easily tractable in analytical expressions of the channel matrix. Furthermore, it requires the characterization of the entire correlation matrix, whose dimensions rapidly increase as the array sizes increase. As a consequence, simplified models relying on various assumptions have been developed for Rayleigh channels, as illustrated below.

3.2.1   The Kronecker model

Introduced by [CKT98, SFGK00, YBO$^+$01], the Kronecker model simplifies the expression of the full correlation matrix by using a separability assumption

$$\mathbf{R} = \mathbf{R}_r \otimes \mathbf{R}_t, \tag{3.18}$$

where $\mathbf{R}_t$ and $\mathbf{R}_r$ are the transmit and receive correlation matrices, respectively, introduced in (2.42).

Mathematically, the Kronecker model is valid if and only if two conditions are jointly met, although contradictory statements can be found in the literature [KSP$^+$02]. The first condition is that the transmit (resp. receive) correlation coefficients are (in magnitude) independent from the considered receive (resp. transmit) antenna. In a 2×2 channel, this yields $r_1 = r_2$ and $t_1 = t_2$. While this condition is easily fulfilled for most usual antenna arrays with reasonably sized antenna spacings (within the spatial stationarity region of the channel), there are cases where this condition is not met: the correlations at one end then depend on the antenna considered at the other end of the link. As an example, consider the case of mutually-coupled antennas. We have seen in Section 2.6.2 that the radiation patterns of closely located antennas are distorted because of coupling, becoming asymmetrical relative to the boresight. In such scenarios, the correlations at the other end of the link will depend on the considered antenna at the coupled end.

In [KSP$^+$02], it is claimed that the above condition is the only one required by the Kronecker assumption of (3.18). However, there is an additional condition [CKT98,CTKV02]: the cross-channel correlations must be equal to the product of corresponding transmit and receive correlations. So, in a 2×2 channel, $\mathbf{R}$ is further simplified by considering that $s_1 = rt$ and $s_2 = rt^*$. Note that for real-valued correlations, a single cross-channel correlation can be defined as $s = s_1 = s_2 = rt$. We formalize the Kronecker model assumptions in the following proposition.

Proposition 3.1 *A $n_r \times n_t$ Rayleigh fading channel matrix $\tilde{\mathbf{H}}$ is Kronecker separable iff*

1. *$\mathcal{E}\{\tilde{\mathbf{H}}(n,m)\tilde{\mathbf{H}}^*(n,p)\}$ and $\mathcal{E}\{\tilde{\mathbf{H}}(n,m)\tilde{\mathbf{H}}^*(q,m)\}$ are independent of n and m respectively,*
2. *$\mathcal{E}\{\tilde{\mathbf{H}}(n,m)\tilde{\mathbf{H}}^*(q,p)\} = \mathcal{E}\{\tilde{\mathbf{H}}(n,u)\tilde{\mathbf{H}}^*(q,u)\} \, \mathcal{E}\{\tilde{\mathbf{H}}(v,m)\tilde{\mathbf{H}}^*(v,p)\},$*

$\forall m, p, u = 1, \ldots, n_t$ and $\forall n, q, v = 1, \ldots, n_r$.

Again, we stress that both conditions are necessary. So how does the second condition of Proposition 3.1 translates into a more propagation-oriented condition?

Proposition 3.2 *When the transmit and receive correlation coefficients are (in magnitude) independent of the considered receive and transmit antenna respectively, the*

Kronecker model is valid, irrespective of antenna configurations and intra-array spacings, iff all DoDs couple with the same power profile into all DoAs, and vice-versa.

The above proposition is related to the physical directions of departure and arrival. Its proof is straightforward and left as an exercise for the reader. However, let us stress that the Kronecker model is then valid for all array configurations. An alternative formulation is that the joint angular power spectrum has to be easily separated into marginal spectra [BOH⁺03].

Finally, inserting (3.18) into (3.3), the channel matrix may be expressed as

$$\tilde{\mathbf{H}} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t^{1/2}. \tag{3.19}$$

The advantage of the Kronecker model is readily apparent: operations on vec(**H**) have been replaced by matrix manipulations on **H**. This greatly simplifies the expressions of many parameters, such as the mutual information, the error probability, etc. It is also easy to understand that the Kronecker model allows for separate transmit and receive optimizations.

3.2.2   The virtual channel representation

The original idea of this model is to relax the correlation-separability restriction of the Kronecker model (i.e. the product assumption on the cross-channel correlations) and to allow for any arbitrary coupling between the transmit and receive ends of the link, while preserving a relative ease-of-use. We have seen in Chapter 2 that an intuitive description of the channel can be obtained from a finite scatterer assumption, with steering vectors pointing in the directions of every scatterer. However, there is theoretically no need to describe the propagation in all these directions. Indeed, the spatial signal space has a finite dimension, due to the finite array dimensions and finite number of antennas. This observation can be exploited in order to develop a linear virtual channel representation [Say02, HLHS03], valid for uniform linear arrays at both ends transmitting/receiving only in the half-space, (i.e. only scatterers located between Tx and Rx are seen by the antennas, or equivalently, $0 \leq \theta_t, \theta_r \leq \pi$). The idea behind this modeling approach is that the finite number of antennas is nothing less than a discrete sampling in the aperture domains. As described in Section 2.1.1, these domains are related to the angular domains via a Fourier transform. By applying Nyquist's sampling theorem, we find that the channel can be equivalently represented in the angular domains by a finite number of angular beams on either side (equal to n_t and n_r) in fixed directions. The number of beams and the virtual angular resolution (hence the total angular aperture as well) only depend on the size and antenna spacings of the arrays. The use of the Fourier relationship also justifies why the model is only valid for ULAs, as the sampling needs to be uniform in the aperture domains.

Coming back to our intuitive reasoning, the model can also be thought of as a sampled representation of the finite scatterer model with discrete values of $\varphi_t(\theta_t)$ and $\varphi_r(\theta_r)$ in the
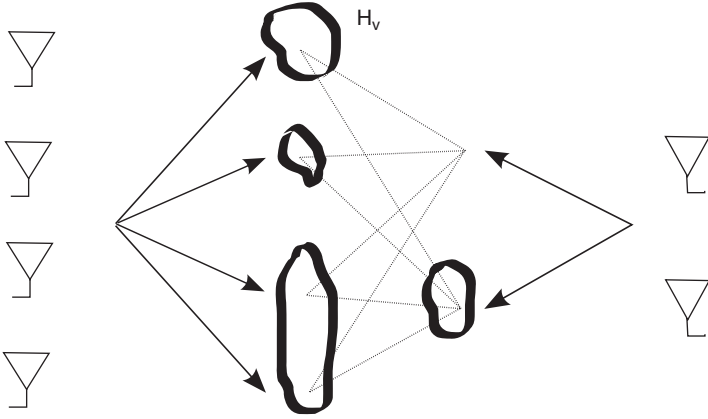
Figure 3.2   *Virtual channel representation of a 2×4 channel*

φ-domain. So-called virtual angles are arbitrarily defined as $\hat{\varphi}_{t_p} = 2\pi(p - 1/2)/n_t - \pi$ and $\hat{\varphi}_{r_q} = 2\pi(q - 1/2)/n_r - \pi$, with $1 \leq p \leq n_t$ and $1 \leq q \leq n_r$. These angles define fixed beams with uniform spacing in the φ-domain, but the angular resolution in the physical DoD–DoA θ-domain is non-uniform, since $\varphi_t = 2\pi \frac{d_t}{\lambda} \cos \theta_t$ and $\varphi_r = 2\pi \frac{d_r}{\lambda} \cos \theta_r$ (d_t and d_r are the Tx and Rx antenna spacings). A schematic representation of the model is depicted in Figure 3.2. In the same way as contributions in the delay domain are grouped into taps when filtered over a certain bandwidth, each physical angular contribution (or scatterer) is grouped into a finite number of directions. Naturally, any scatterer located between two virtual angles is somehow represented in each virtual direction owing to the spatial filtering sidelobes. It is also interesting to note that two degrees of freedom remain in choosing the first directions $\hat{\varphi}_{t_1}$ and $\hat{\varphi}_{r_1}$: the choice proposed above might therefore not be optimal, but no other choice can be made without first having knowledge of the spatial structure of the channel.

From the definition of virtual angles, the channel representation can be expressed as a Fourier decomposition

$$\tilde{\mathbf{H}} = \hat{\mathbf{A}}_r \mathbf{H}_v \hat{\mathbf{A}}_t^T \tag{3.20}$$

where $\mathbf{H}_v$ ($n_r \times n_t$) is the channel matrix representation in the Fourier domain and contains the complex gain of each virtual path, i.e. the weights of the various branches in the central part of Figure 3.2. Virtual steering matrices $\hat{\mathbf{A}}_t$ ($n_t \times n_t$) and $\hat{\mathbf{A}}_r$ ($n_r \times n_r$) are formed (columnwise) by the normalized virtual steering vectors $\hat{\mathbf{a}}_t$ and $\hat{\mathbf{a}}_r$, which are constructed similarly as (2.37) and (2.38) but in the virtual directions

$$\hat{\mathbf{a}}_t = \frac{1}{\sqrt{n_t}} \left[1 \ e^{-j\hat{\varphi}_t} \ \dots \ e^{-j(n_t-1)\hat{\varphi}_t} \right]^T, \tag{3.21}$$

$$\hat{\mathbf{a}}_r = \frac{1}{\sqrt{n_r}} \left[1 \ e^{-j\hat{\varphi}_r} \ \dots \ e^{-j(n_r-1)\hat{\varphi}_r} \right]^T. \tag{3.22}$$

In contrast to the steering matrices in (2.40), matrices $\hat{\mathbf{A}}_t$ and $\hat{\mathbf{A}}_r$ are unitary, which implies that their columns are orthogonal. As a result, they form basis vectors in the transmit and receive spatial domains. Another consequence of this property is that $\mathbf{H}_v$ is related to the physical channel matrix $\tilde{\mathbf{H}}$ by

$$\mathbf{H}_v = \hat{\mathbf{A}}_r^H \tilde{\mathbf{H}} \hat{\mathbf{A}}_t^*. \tag{3.23}$$

The correlation matrix of $\text{vec}(\mathbf{H}_v^H)$ is then given by

$$\mathbf{R}_v = \left[\hat{\mathbf{A}}_r^T \otimes \hat{\mathbf{A}}_t^T \right] \mathbf{R} \left[\hat{\mathbf{A}}_r^* \otimes \hat{\mathbf{A}}_t^* \right]. \tag{3.24}$$

So far we have represented the MIMO channel matrix into another space, without any approximation. The fundamental assumption of the virtual channel model is that $\mathbf{R}_v$ is a diagonal matrix, which means that the elements of $\mathbf{H}_v$ are uncorrelated, i.e. the transmission/reception in each virtual direction is independent. Equivalently, that means that the complex weights of each branch in the central part of Figure 3.2 are all independent. Hence, $\mathbf{H}_v$ can be rewritten as $\boldsymbol{\Omega}_v \odot \mathbf{H}_w$, which makes the model particularly elegant and easy-to-use. Furthermore, $\mathbf{R}_v$ and $\boldsymbol{\Omega}_v$ are often sparse in realistic propagation scenarios. How realistic this assumption is will be dealt with later, when we compare several analytical models. The positive and real-valued elements of

$$[\boldsymbol{\Omega}_v \odot \boldsymbol{\Omega}_v] (q, p) = \mathbf{R}_v(n_r (q - 1) + p, n_r (q - 1) + p) \tag{3.25}$$

determine the average power-coupling between the pth virtual transmit angle and the qth virtual receive angle. Note that the propagation channel is then entirely characterized by the coupling matrix $\boldsymbol{\Omega}_v$.

Finally, let us note that the assumption of uncorrelated virtual steering vectors between Tx and Rx may be far from true in some cases (as we will see later), but should intuitively improve with increasing n_t and n_r. It can also be thought of as a separability condition on the eigenvectors of $\mathbf{R}$, assuming that $\mathbf{U_R}$ can be written as $\hat{\mathbf{A}}_t^* \otimes \hat{\mathbf{A}}_r^*$, where $\mathbf{U_R}$ results from the EVD of $\mathbf{R}$

$$\mathbf{R} = \mathbf{U_R} \boldsymbol{\Lambda_R} \mathbf{U_R}^H, \tag{3.26}$$

and the elements of $\boldsymbol{\Lambda_R}$ are given (if the virtual steering vectors are indeed uncorrelated) by the elements of $\boldsymbol{\Omega}_v$.

3.2.3 The eigenbeam model

As detailed above, the unitary matrices of the virtual channel model are chosen irrespective of the channel correlations, i.e. irrespective of the actual directions where the physical scatterers are located. Furthermore, remember that two degrees of freedom remain in the choice of the first virtual angles, and that the virtual channel representation is only valid for ULAs. The eigenbeam model is a generalization to arbitrary arrays where the left and right unitary matrices are well chosen. Indeed, it assumes that they are now determined by the eigendirections of the transmit and receive correlation(s), i.e. they correspond to

sampling the angular domains as to match the positions of the scatterers. Therefore, the eigenbeam model can only yield a more accurate representation than the virtual channel representation. Naturally, the price for this increased performance is the knowledge of the transmit and receive correlation matrices, which is not required to determine $\hat{\mathbf{A}}_t$ and $\hat{\mathbf{A}}_r$.

To derive the eigenbeam model of a Rayleigh channel, let us define the eigenvalue decompositions of the receive and transmit correlation matrices as

$$\mathbf{R}_r = \mathbf{U}_{\mathbf{R}_r} \mathbf{\Lambda}_{\mathbf{R}_r} \mathbf{U}_{\mathbf{R}_r}^H$$
$$\mathbf{R}_t = \mathbf{U}_{\mathbf{R}_t} \mathbf{\Lambda}_{\mathbf{R}_t} \mathbf{U}_{\mathbf{R}_t}^H. \tag{3.27}$$

We may always decompose $\tilde{\mathbf{H}}$ along the eigenmodes (a.k.a. eigendirections or eigenbeams) at both transmit and receive sides, and write

$$\tilde{\mathbf{H}} = \mathbf{U}_{\mathbf{R}_r}^* \mathbf{H}_e \mathbf{U}_{\mathbf{R}_t}^H, \tag{3.28}$$

where $\mathbf{H}_e$ is a fading matrix relating receive and transmit eigenmodes. Note that this expression slightly differs from the original expression of Weichselberger [Wei03,WHOB06], owing to our different definitions of $\mathbf{R}$, $\mathbf{R}_r$ and $\mathbf{R}_t$.

The eigenbeam model [Wei03,WHOB06] is then based on the assumption that all transmit and receive eigenmodes are mutually completely uncorrelated, so that (3.28) can be expressed as

$$\tilde{\mathbf{H}} = \mathbf{U}_{\mathbf{R}_r}^* \left(\mathbf{\Omega}_e \odot \mathbf{H}_w \right) \mathbf{U}_{\mathbf{R}_t}^H. \tag{3.29}$$

In (3.29), $\mathbf{H}_w$ is the usual i.i.d. complex Gaussian random fading matrix and $\mathbf{\Omega}_e \odot \mathbf{\Omega}_e$ is the power coupling matrix whose positive and real-valued elements determine the average power-coupling between the pth transmit eigenmode and the qth receive eigenmode

$$[\mathbf{\Omega}_e \odot \mathbf{\Omega}_e] (q, p) = \mathbf{R}_e(n_r (q - 1) + p, n_r(q - 1) + p), \tag{3.30}$$

with

$$\mathbf{R}_e = \left[\mathbf{U}_{\mathbf{R}_r}^H \otimes \mathbf{U}_{\mathbf{R}_t}^H \right] \mathbf{R} \left[\mathbf{U}_{\mathbf{R}_r} \otimes \mathbf{U}_{\mathbf{R}_t} \right]. \tag{3.31}$$

As with the virtual channel representation, the eigenbeam model assumption may also be rewritten as a separability condition on the eigenvectors of $\mathbf{R}$, assuming that $\mathbf{U}_{\mathbf{R}}$ in (3.26) can be written as $\mathbf{U}_{\mathbf{R}_t} \otimes \mathbf{U}_{\mathbf{R}_r}$, and that the eigenvalues of $\mathbf{R}$ are given by the elements of $\mathbf{\Omega}_e$. As already mentioned, both models differ in choice of transmit and receive unitary matrices. Of course, the more adequate choice of the eigenbeam model necessitates a knowledge of the transmit and receive correlation matrices, while this knowledge is not required by the virtual channel representation.

What about the structure of $\mathbf{\Omega}_e$? It is strongly linked to the radio environment. A matrix $\mathbf{\Omega}_e$ with only one non-zero valued column corresponds to rich scattering conditions at the receiver and a reduced angle-spread at the transmitter (typically, a one-ring scenario).

A diagonal structure of $\mathbf{\Omega}_e$ models an environment where each single transmit eigenmode is linked to a single and distinctive receive eigenmode. A full structure of $\mathbf{\Omega}_e$ is equivalent of having a rich scattering cluster around each link end. If, additionally, $\mathbf{\Omega}_e$ is of rank one, the coupling between all modes is identical, and the eigenbeam model reduces to the Kronecker model, which yields the following corollary of Proposition 3.2.

Corollary 3.1   *When the transmit and receive correlation coefficients are (in magnitude) independent of the considered receive and transmit antenna respectively, the Kronecker model is valid iff all transmit eigenvectors couple with the same power profile into all receive eigenvectors, and vice-versa.*

By contrast to Proposition 3.2, this corollary deals with eigenvectors, which already combine the array configurations and the physical directions.

For fully-structured coupling matrices, the rank is therefore a decisive parameter, since it allows mitigation not only of the weakness of the Kronecker model (which implies a full coupling), but also the assumptions of single-bounce propagation (which is equivalent to uniquely-linked modes). The fact that partial coupling is thus allowed and easily controlled makes this model highly flexible.

To sum up, the eigenbeam model input parameters are the eigenbasis of the receive and transmit correlation matrices and the coupling matrix. In Section 3.4.2 we illustrate how to practically derive these parameters. It is also important to remember that all three simplified models assume a sort of separability, either on the correlation or on the eigenvector matrix of the correlation.

3.3   Propagation-motivated MIMO metrics

Having described various models, it would be useful to investigate how we can compare these models. Similarly, for given MIMO channel realizations (obtained through simulations or resulting from experimental campaigns), one may desire to evaluate the MIMO potentials of this channel, such as the multiplexing and diversity gains. Finally, evaluating the space–time stationarity of MIMO channels also requires the use of adequate measures. To achieve these purposes, we need to define metrics. Hopefully, these metrics should be propagation-motivated and give insight into the MIMO channel structure. This is a very ambitious goal and we will rapidly notice hereafter that most metrics cannot cover all MIMO aspects simultaneously. Finally, it is important to be aware that MIMO metrics strongly depend on antenna patterns and configurations, since a MIMO channel matrix combines antenna arrays with a double-directional channel.

3.3.1   Comparing models and correlation matrices

A first natural characteristic parameter of MIMO channels is the correlation matrix $\mathbf{R}$, which is highly representative of the spatial richness of the channel. Therefore, it is sensible

to compare models (or models with measured channels) by comparing their correlation matrices. Introduced in [MBF02, YBO^{+}04], a measure for the difference or the error between an approximated correlation matrix $\mathbf{R}'$ and the channel correlation matrix $\mathbf{R}$ is given by

$$\Psi = \frac{\|\mathbf{R} - \mathbf{R}'\|_F}{\|\mathbf{R}\|_F}. \tag{3.32}$$

Another interesting measure of the distance between correlation matrices is introduced by [Her04]. To obtain a metric that is equal to zero when the correlation matrices are identical (apart from a scalar factor) and equal to one if they differ in the largest possible way, the correlation matrix distance between $\mathbf{R}$ and $\mathbf{R}'$ is defined as

$$d_{corr} = 1 - \frac{\mathrm{Tr}\left\{\mathbf{R}\mathbf{R}'\right\}}{\|\mathbf{R}\|_F \|\mathbf{R}'\|_F}, \tag{3.33}$$

Both parameters will be employed later when we compare the simplified models described in Section 3.2 with synthetic finite scatterer channels.

3.3.2   Characterizing the multipath richness

The notion of multipath richness is not a precise concept, although everyone understands what is meant by this otherwise vague expression. Correlation is clearly an indication of how rich the scattering of the double-directional channel may be. Alternative metrics are however often used, each highlighting a particular aspect of MIMO communications.

Directional spreading

As already mentioned, there is a direct relationship between the joint angular spectrum and the various correlation coefficients. Because transmit and receive sides are often considered separately (e.g. in the Kronecker model), we may want to characterize the direction or angular spreading separately at transmit or receive sides. The RMS direction, azimuth and elevation spreads as the square-root of the second-order moments of the power direction, azimuth and elevation spectra were discussed earlier in Section 2.1.4. The last two metrics are given in angular units (radians or degrees). By contrast, [DR98] proposes a measurement of the azimuth-spread by means of a dimensionless metric

$$\Lambda_{\Theta} = \sqrt{1 - \frac{\left|\int_0^{2\pi} \mathcal{A}(\Theta)e^{j\Theta}d\Theta\right|^2}{\left|\int_0^{2\pi} \mathcal{A}(\Theta)d\Theta\right|^2}}, \tag{3.34}$$

which can be estimated either at the transmitter or at the receiver. If we use this definition, azimuth-spreads range from 0 to 1, with $\Lambda_{\Theta} = 0$ denoting the extreme case of a single component departing to/arriving from a single direction and $\Lambda_{\Theta} = 1$ denoting no clear bias in the azimuthal power distribution. Naturally, a similar definition can be used as far as elevation spreads are concerned.

The metrics introduced in [PBN06] are valid for three-dimensional directional spreads and rely on the second-order statistics of the spatial fading. Let us denote by ∇r^2 (3×1) the first-order partial derivatives (along spatial coordinates x, y and z) of the received power $r^2 = |g|^2$ at location (x, y, z). The correlation matrix of ∇r^2 is given by

$$\mathbf{R}_{\nabla r^2} = \mathcal{E}\left\{\nabla r^2 \left(\nabla r^2\right)^T\right\}. \tag{3.35}$$

The directional spread is then characterized by two quantities, $\text{Tr}\{\mathbf{R}_{\nabla r^2}\}$ and $\det\{\mathbf{R}_{\nabla r^2}\}$. The former provides a measure of the second-order spatial fading statistics (i.e. the spatial fading rate) averaged over all spherical directions, while the latter represents the level of variation in second-order fading statistics with direction. Both parameters increase as azimuth and elevation spreads increase, although $\text{Tr}\{\mathbf{R}_{\nabla r^2}\}$ is upper-bounded by 0.5.

Multiplexing capability

A number of metrics can be used to quantify the multiplexing capability of a MIMO channel.

The multiplexing advantage of a MIMO channel is classically measured by the *ergodic mutual information* of the channel when an equal power allocation is used at the transmitter

$$\bar{\mathcal{I}}_e = \mathcal{E}\left\{\log_2 \det\left[\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H\right]\right\}, \tag{3.36}$$

where $\rho = E_s/\sigma_n^2$ is the SNR. Clearly, the mutual information depends on the SNR, so any relative difference between measured and simulated ergodic mutual informations should always be related to the SNR; which makes comparisons SNR-dependent. Alternatively, a few metrics have been derived which only depend on the distribution of the eigenvalues.

The above expression of the ergodic mutual information may not be easily tractable, and does not allow direct analysis of the impact of correlation. To palliate this problem, an *upper-bound* can be obtained by application of Jensen's inequality [OP04]

$$\log_2(\bar{\kappa}) = \log_2\left(\mathcal{E}\left\{\det\left[\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H\right]\right\}\right) \geq \bar{\mathcal{I}}_e. \tag{3.37}$$

An explicit closed-form linking this upper bound with the channel correlations can be derived, noting that for any $2m$ identically distributed zero-mean complex Gaussian correlated variables $x_1, \ldots, x_{2m}$

$$\begin{aligned}
\mathcal{E}\{x_1 \, x_2 \, x_3 \ldots x_{2m-1} \, x_{2m}\} = {} & \mathcal{E}\{x_1 \, x_2\} \, \mathcal{E}\{x_3 \, x_4\} \ldots \mathcal{E}\{x_{2m-1} \, x_{2m}\} \\
& + \mathcal{E}\{x_2 \, x_3\} \, \mathcal{E}\{x_4 \, x_5\} \ldots \mathcal{E}\{x_1 \, x_{2m}\} + \ldots \\
& \text{(all permutations).}
\end{aligned} \tag{3.38}$$

For relatively small antenna arrays, the analysis of the closed form is particularly revealing. Let us start with 2×2 systems. Hence, $\bar{\kappa}$ reads as

$$
\begin{aligned}
\bar{\kappa} = 1 + \frac{\rho}{2} & \sum_{k,l=1}^{2} \mathcal{E}\left\{ |\mathbf{H}(l,k)|^2 \right\} \\
& + \left(\frac{\rho}{2} \right)^2 \left[\mathcal{E}\left\{ |\mathbf{H}(1,1)|^2 |\mathbf{H}(2,2)|^2 \right\} + \mathcal{E}\left\{ |\mathbf{H}(2,1)|^2 |\mathbf{H}(1,2)|^2 \right\} \right. \\
& \left. - \mathcal{E}\left\{ \mathbf{H}^*(1,1)\mathbf{H}(1,2)\mathbf{H}(2,1)\mathbf{H}^*(2,2) \right\} \right. \\
& \left. - \mathcal{E}\left\{ \mathbf{H}(1,1)\mathbf{H}^*(1,2)\mathbf{H}^*(2,1)\mathbf{H}(2,2) \right\} \right].
\end{aligned}
\tag{3.39}
$$

The term proportional to the SNR per transmit antenna is simply the average squared Frobenius norm of the channel, equal to $n_t n_r = 4$. The term proportional to $\rho^2/4$ is calculated using the property identified by (3.38), and observing that $\mathcal{E}\{\mathbf{H}(l,k)\mathbf{H}(p,m)\} = \mathcal{E}\{\mathbf{H}^*(l,k)\mathbf{H}^*(p,m)\} = 0$ for any k, l, m, and p. We can now rewrite (3.39) as

$$
\bar{\kappa} = 1 + 2\gamma + \left(\frac{\rho}{2} \right)^2 \left[\sum_{k=1}^{2} \left(1 + |s_k|^2 - |t_k|^2 - |r_k|^2 \right) \right],
\tag{3.40}
$$

where s_k, t_k and r_k, $k = 1, 2$ are defined by (3.2), respectively, as the cross-channel, transmit and receive correlations. While the transmit and receive correlations appear to degrade the multiplexing advantage, cross-channel correlations seem to have a positive impact. We have illustrated this curious behavior for 2×2 systems, but further analysis reveals that it also holds true for larger arrays (see Chapter 4, also concerning the interpretation of this result).

The *richness vector* is defined as the mean cumulative log-sum of the channel ordered non-zero eigenvalues values of $\mathbf{H}\mathbf{H}^H$, $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ (introduced in Section 2.1.6)

$$
N(p) = \sum_{k=1}^{p} \log_2(\lambda_k).
\tag{3.41}
$$

Its relationship with the mutual information is pretty straightforward, as the latter equals the cumulated richness vector at high SNR, apart from a constant term.

The *Demel condition number* is the ratio of the largest to the smallest singular value of $\mathbf{W}$. Its average value is also an indicator of the multiplexing gain of a MIMO channel

$$
\nu_D = \max_k(\lambda_k) / \min_k(\lambda_k).
\tag{3.42}
$$

The smaller the mean condition number (lower-bounded by 1), the larger the average multiplexing gain.

Effective diversity measure

The degree of diversity offered by a MIMO channel is intuitively related to the number of degrees of freedom of the channel, i.e. to the distribution of the eigenvalues of $\mathbf{R}$, or equivalently to the variance of the effective channel $\|\mathbf{H}\|_F^2$. A compact metric characterizing the spread of this distribution, inversely proportional to the variance of $\|\mathbf{H}\|_F^2$ and known as the *effective diversity measure*, has been proposed in [Nab03, IN03] for Rayleigh channels

$$
\begin{aligned}
N_{div} &= \left[\frac{\mathrm{Tr}\{\mathbf{R}\}}{\|\mathbf{R}\|_F} \right]^2 \\
&= \frac{[\mathrm{Tr}\{\mathbf{\Lambda_R}\}]^2}{\mathrm{Tr}\{\mathbf{\Lambda_R^2}\}} \\
&= \frac{\left[\sum_k \lambda_k(\mathbf{R}) \right]^2}{\sum_k \lambda_k^2(\mathbf{R})}
\end{aligned}
\tag{3.43}
$$

The smaller the eigenvalue spread (i.e. the variance of $\|\mathbf{H}\|_F^2$), the higher the diversity measure. We may also express N_{div} as a function of the channel correlations

$$
N_{div} = \frac{n_t^2 n_r^2}{n_t n_r + \sum_{k,l=1\ k \neq l}^{n_t n_r} |\mathbf{R}(k,l)|^2}
\tag{3.44}
$$

We have expressly written the denominator as made of two parts: the first one is the average squared Frobenius norm of the channel, while the second term is a summation over all correlation coefficients, i.e. transmit, receive and cross-channel correlations. In Rayleigh channels, the effective diversity measure is maximized by i.i.d. fading ($N_{div} = n_t n_r$), as illustrated in Figure 3.3 for a 2×2 Kronecker-structured system. We stress that any correlation will decrease the effective diversity measure, by increasing the variance of $\|\mathbf{H}\|_F^2$. This is in contrast with the average mutual information with equal power allocation, for which we have just seen that cross-channel correlations are playing a constructive role.

Equation (3.43) is generalized to Ricean fading channels [Nab03, NBP05] as

$$
N_{div} = \frac{\left[\mathrm{Tr}\{\mathbf{\Lambda_R}\} + K \mathrm{vec}(\bar{\mathbf{H}})^H \mathrm{vec}(\bar{\mathbf{H}}) \right]^2}{\mathrm{Tr}\{\mathbf{\Lambda_R^2}\} + 2K \mathrm{vec}(\bar{\mathbf{H}})^H \mathbf{R} \mathrm{vec}(\bar{\mathbf{H}})}
\tag{3.45}
$$

where the additional term takes into account the geometry of the Ricean component relative to the geometry of $\mathbf{R}$. As an example, Figure 3.4 depicts the variation of N_{div} for 2×2 systems as a function of the Ricean K-factor and correlation levels, considering that $\bar{\mathbf{H}} = \mathbf{1}_{2\times 2}$ and $\mathbf{R} = \mathbf{R}_r \otimes \mathbf{R}_t$.

The definition of (3.45) is also compatible with the usual definition of the diversity gain corresponding to the high-SNR slope of the error probability in diversity systems with respect to AWGN channels. Indeed, for i.i.d. Rayleigh channels, we have seen that $N_{div} = n_t n_r$, while for AWGN channels (i.e. for $K \to \infty$), we get $N_{div} = \infty$ as expected. More specifically, the diversity g_d^o is the nearest integer larger than N_{div}.
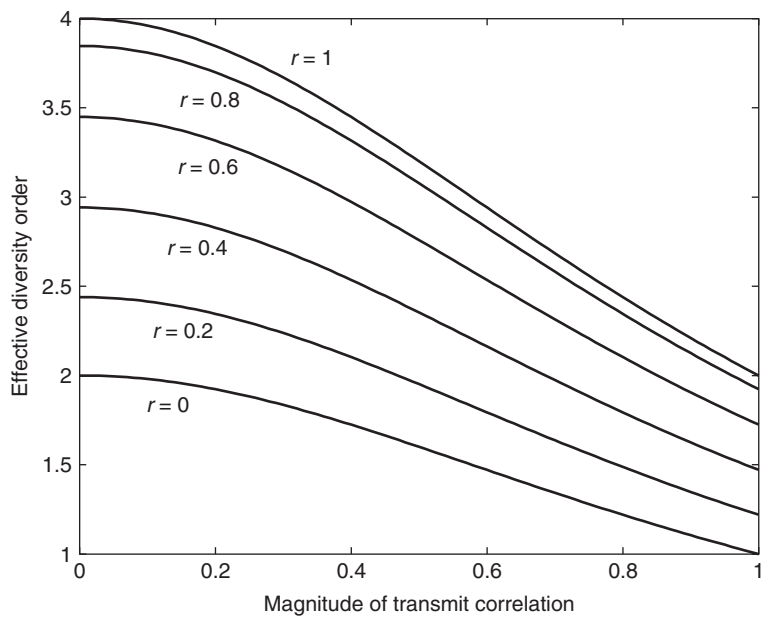
Figure 3.3   *Effective diversity measure in a 2×2 Kronecker-structured system as a function of transmit and receive correlations*
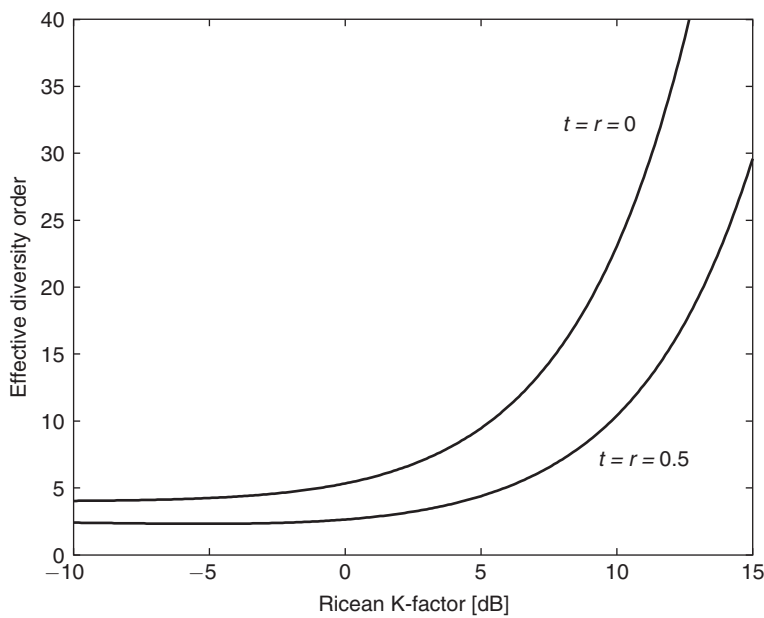
Figure 3.4   *Effective diversity measure in several Ricean MIMO channels*

3.3.3 Measuring the non-stationarity of MIMO channels

For MIMO channels fully described by second-order statistics (such as correlated Rayleigh channels), the space–time-frequency correlation functions introduced in Chapter 2 are a complete (though complex) stochastic description. To simplify the analysis, we have considered a space-only characterization of MIMO channels, which has led to the definition of a spatial correlation matrix $\mathbf{R}(t, f)$. For narrowband channels, the latter naturally reduces to $\mathbf{R}(t)$. This correlation matrix does not vary as long as the channel remains WSS, since for WSS channels, the various correlations do not depend on the absolute time. We introduce the notion of temporal stationarity region as the period of time during which the spatial characteristics of the channel (i.e. $\mathbf{R}$, or $\mathbf{A}_t$ and $\mathbf{A}_r$, or $\mathbf{U}_{\mathbf{R}_t}$ and $\mathbf{U}_{\mathbf{R}_r}$, etc.) remain essentially constant. The knowledge of the temporal stationarity region is critical in adaptive transmission schemes, since the channel statistics are estimated at the receiver, and then fed back to the transmitter. If the temporal stationarity region is too small, the channel statistics will have changed by the time the transmitter learns the channel.

There are numerous causes of non-stationarity in radio propagation [Her04]. If the receiver and/or transmitter are moving, directions of departure and arrival typically change, components get shadowed or disappear while new ones appear. Non-stationarity may also result from moving obstacles, such as passing cars and people. The strength of non-stationarity is directly related to the number of moving scatterers. In the case of people moving indoors, a significant degree of variation can be expected.

Evaluation of stationarity regions
Evaluation of stationarity regions has been investigated in many aspects in the past, and a large number of metrics exist to characterize the non-stationarity of radio channels. For wideband SISO systems, an intuitive metric is the correlation coefficient between consecutive power delay profiles. However, these SISO metrics do not give any indication on how much the spatial properties change over time. Therefore, a further step is to consider non-stationarity metrics for diversity (SIMO/MISO) or beamforming systems, e.g. the F-eigen ratio [VHU02]

$$q^{(F)} = \frac{\mathrm{Tr}\left\{\mathbf{U}_{\mathbf{R}'(F)}^{H} \mathbf{R} \mathbf{U}_{\mathbf{R}'(F)}\right\}}{\mathrm{Tr}\left\{\mathbf{U}_{\mathbf{R}(F)}^{H} \mathbf{R} \mathbf{U}_{\mathbf{R}(F)}\right\}}, \qquad (3.46)$$

where $\mathbf{U}_{\mathbf{R}(F)}$ and $\mathbf{U}_{\mathbf{R}'(F)}$ contain the eigenvectors corresponding to the F largest eigenvalues of correlation matrices $\mathbf{R}$ and $\mathbf{R}'$ at time instants t and t'. The F-eigen ratio measures the penalty of using out-dated long-term statistics in beamforming application instead of the current statistics. By definition, the F-eigen ratio is only useful for systems where transmission is performed over a reduced set of eigenvectors. In MIMO systems exploiting all eigenmodes, the F-eigen ratio is thus less appealing.

Therefore, it is preferable in a MIMO context to quantify the variation of the full correlation matrix, as it is the most complete description for both narrowband and wideband systems

(for wideband systems, the description is based on one correlation matrix for each tap). The goal is thus to evaluate how much $\mathbf{R}$ changes over time, i.e. how much $\mathbf{R}'$ at a given instant t' is similar to $\mathbf{R}$ at instant t. We have already defined two metrics characterizing the closeness of two correlation matrices (see Section 3.3.1). In particular, the correlation matrix distance shares the property of being comprised between 0 and 1 with a correlation coefficient, which makes it a good candidate for evaluating the non-stationarity of the channel [Her04]: $d_{corr} = 0$ when $\mathbf{R} = \mathbf{R}'$, and is equal to one when the difference between the two correlation matrices is maximum, i.e. when the spatial structure (the eigenvectors) of $\mathbf{R}$ is (are) orthogonal to the spatial structure (the eigenvectors) of $\mathbf{R}'$. Naturally, the correlation matrix distance can be applied separately at the transmit and receive sides (respectively on $\mathbf{R}_t$ and $\mathbf{R}_r$).

Example 3.1   *Consider a MIMO 2×2 system with half-wavelength spacing at both ends ($d_t = d_r = \lambda/2$). The system operates in an outdoor environment providing the two following propagation scenarios.*

- *At the transmitter, there are three relative angles of departure of $\theta_{t,1} = \pi/6$, $\theta_{t,2} = \pi/2$ and $\theta_{t,3} = 2\pi/3$ ($\pi/2$ is the direction of the link axis).*
- *At the receiver, there are three relative angles of arrival of $\theta_{r,1} = \pi/4$, $\theta_{r,2} = \pi/3$ and $\theta_{r,3} = 3\pi/4$ ($\pi/2$ is the direction of the link axis).*
- *The scattering mechanisms are described by complex Gaussian statistics (analogous to geometry-based models).*
- *DoDs/DoAs are either uniquely coupled (scenario A; first DoD with first DoA, etc.) through the propagation channel or fully-coupled (scenario B: each DoD is coupled with all DoAs, and vice-versa).*

Both scenarios are illustrated in Figures 3.5 and 3.6. Figure 3.7 evaluates how the correlation matrix distance increases with respect to the initial scenario when $\theta_{r,3}$, and possibly $\theta_{t,2}$, vary smoothly over 80 time units (time is arbitrarily indexed between 0 and 80), respectively from $\pi/4$ to 0 and from $\pi/2$ to $\pi/4$. We observe that the correlation matrix distance remains limited, although it is up to five times larger when two scatterers are moving.

Impact of non-stationarity region on real-world MIMO system design

For indoor mobile scenarios (in office environments), experimental results [Her04] show that the spatial structure at the access point is relatively stable when the mobile terminal moves over a few meters, typically within an office room (d_{corr} remains below 0.2, but may reach 0.5 in some cases). The spatial structure at the user's terminal is by contrast more variable, with values of d_{corr} reaching 0.7 even over small distances.

What is the significance of the above results on the design of adaptive MIMO transmission techniques? Typically, the receiver estimates the channel, and then feedbacks some form of estimated statistics to the adaptive transmitter. The accuracy of this information is therefore crucial. In indoor mobile downlink schemes, the receiver (i.e. the mobile terminal)
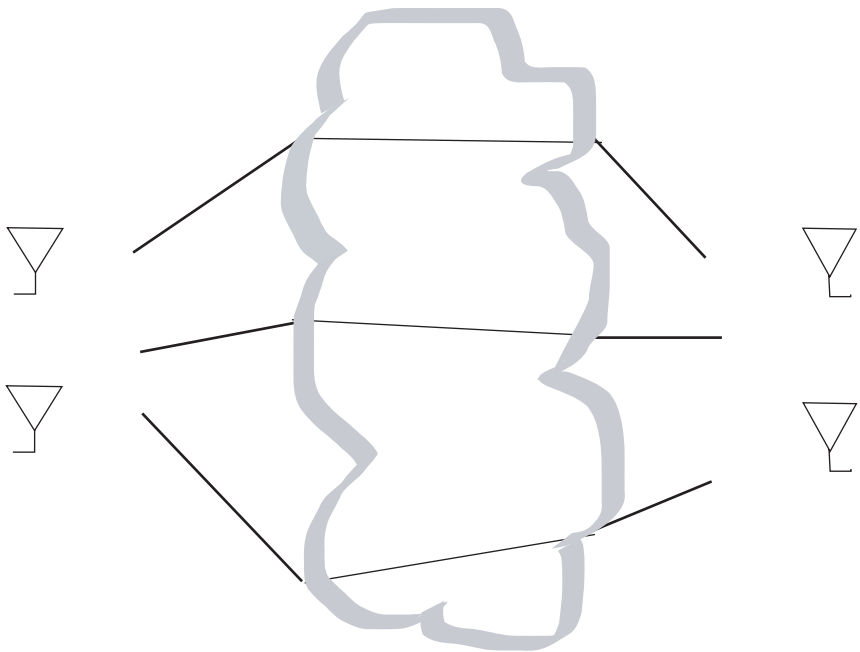
Figure 3.5 *Path configuration for scenario A*
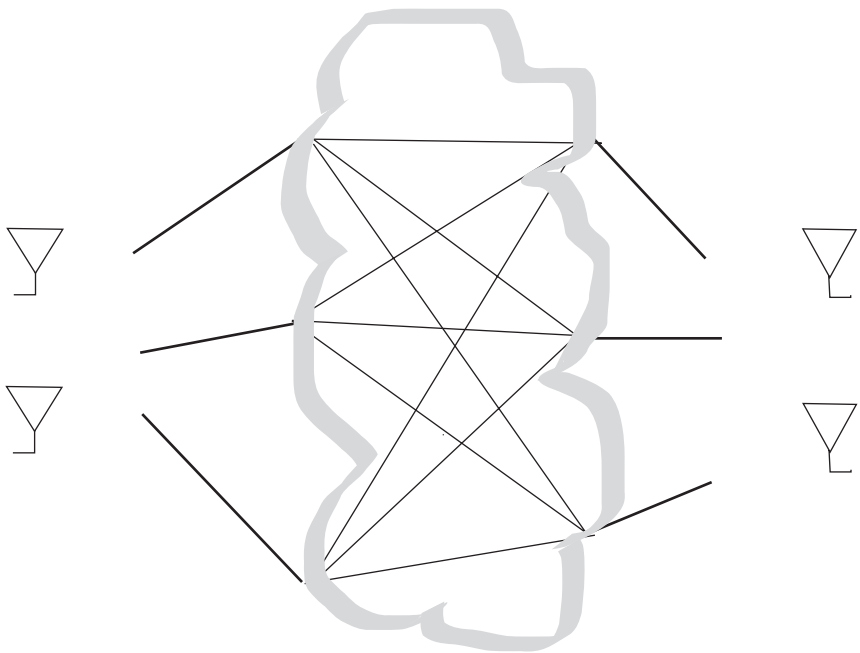
Figure 3.6 *Path configuration for scenario B*
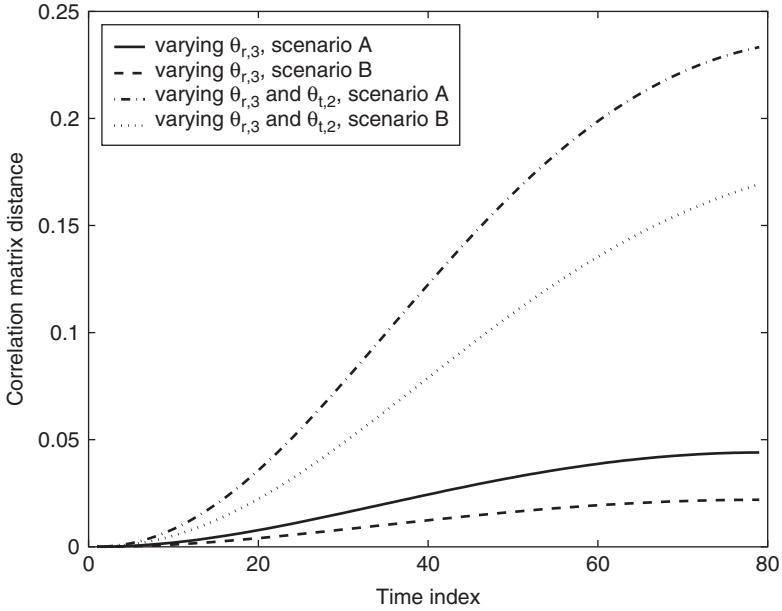
Figure 3.7   *Correlation matrix distance for different scenarios*

sends back the correlation matrix to the access point, which can extract the transmit corre-lation and use it to adaptively optimize the transmission. By contrast, in the uplink, if the receiver (i.e. the access point) feeds back the correlation matrix to the mobile terminal, the use of this information by the latter is more questionable. Indeed, by the time the transmit spatial information reaches the transmitter (i.e. the mobile terminal), this structure may have already changed.

The use of adaptive transmission techniques may, therefore, be compromised in indoor mobile uplink schemes. It can be expected that similar conclusions can be found in high-mobility outdoor uplink scenarios.

3.4   Relationship between physical models and analytical representations

3.4.1   The Kronecker model paradox

As previously mentioned, analytical models are only symbolic representations which need to be parametrized by physical sources (data or models). In this respect, a given analytical model might not be able to represent a physical model. As an example, let us consider the paradox of the extraordinarily popular Kronecker model. For many, the Kronecker model is indeed representative of most typical cases, irrespective of the actual propagation mechanisms and antenna arrays. But is this actually true?

Most papers using the Kronecker model cite [CKT98, SFGK00, YBO^{+}01, KSP^{+}02, CTKV02, YBO^{+}04] as a justification. However, the motivation in [CKT98, CTKV02] is primarily to simplify the expression of the channel matrix, and based on *a posteriori* simulations of s_1 and rt. The discussion in [SFGK00] is based on the one-ring model, introduced in Chapter 2. We will show here that [SFGK00] does not actually prove that single-bounce geometry-based channel models are adequately represented by the Kronecker structure for arbitrary antenna array configurations. The same remark holds true for the experimental results of [YBO^{+}01, YBO^{+}04]. The point is that many geometry-based or experimental results rely on large antenna spacings at one side of the link (i.e. low antenna correlation at one side), which is a particular case of array configuration. The consequence is that the Kronecker model has often been validated, but with these particular antenna configurations. Finally, [KSP^{+}02] claims that only the first mathematical condition has to be fulfilled. Proposition 3.1 outlines that this first condition is necessary, but not sufficient.

What happens when some DoDs are uniquely coupled with specific DoAs? Clearly, the joint direction power spectrum is not separable, unless all DoDs (or all DoAs) are identical. However, the Kronecker representation might still be a very good approximation of the channel correlation. If uniquely-coupled propagation modes with significant power in a Rayleigh-fading channel[1] exist, the channel correlation is Kronecker-structured when sufficiently large antenna spacings at one side at least allow for low antenna correlation across the array at that side. This condition forces the antenna spacing to be sufficiently large, so that this is a stronger condition than just canceling the antenna correlation at that side. To illustrate this point, let us take a few examples relying on geometry-based models.

The first example considers the one-ring model, with a 2×2 uplink system and broadside configuration. The transmit antennas are omnidirectional antennas separated by a distance d while the receive antennas (i.e. at the base station/access point) are omnidirectional antennas and spaced by D. When the ring radius is much smaller than the transmit-to-receive distance, the different correlation coefficients can be analytically calculated by means of (2.60) as

$$r = J_0\left[\frac{2\pi}{\lambda}D\Delta\right]$$

$$t = J_0\left[\frac{2\pi}{\lambda}d\right]$$

$$s_1, s_2 = J_0\left[\frac{2\pi}{\lambda}(D\Delta \pm d)\right] \tag{3.47}$$

where Δ is the ratio between the ring radius and the Tx-Rx distance.

---

[1] That implies that the angle-spread does not approach zero at any side. Indeed, if the angular power spectrum at one side only can be approximated by a δ function (so, the angle-spread is equal to zero at that side), Proposition 3.2 is automatically met. Channels with a very small angle-spread at one side are thus well represented by the Kronecker model.

In particular, when d/λ and D/λ are chosen as the smallest possible values forcing the transmit and receive correlations to zero, the cross-channel correlations s_1 and s_2 are equal to -0.24 and 1, respectively. This is clearly in violation with the product assumption of the Kronecker model. However, if d is large enough, t, s_1 and s_2 will all tend toward zero, irrespective of the values of D and r. The dual result holds true if D is large enough. The consequence is that large antenna spacings at least at one side yields a Kronecker-structured channel (the same results would be obtained with arbitrary $n_r \times n_t$ systems with any orientation). This also has an indirect implication: zero correlations at the transmitter and receiver are not a sufficient condition for the one-ring channel matrix to tend to $\mathbf{H}_w$, whereas choosing large antenna spacings at both ends is a sufficient condition of statistical independence between elements. It may also be observed that if $\Delta = 0$, the Rx angle-spread is null, resulting in $r = 1$, and $s_1, s_2 = t = rt$. This particular case of the one-ring model is well represented by the Kronecker assumption (it actually fully meets the conditions of Proposition 3.2).

The second example deals with the more complex combined elliptical-ring model applied in a 2×2 downlink system with broadside configuration. Again, the transmit antennas are omnidirectional antennas separated by a distance D while the omnidirectional receive antennas are spaced by d. A typical Rayleigh-fading power-delay profile (IEEE 802.16 model 4 [Eea01], see Table 2.2) valid at a reference range of 7 km is simulated at a range of 4 km, with the radius of the local ring around the receiver equal to 30 m, and $D = 40d$. Correlations are estimated for a local scattering ratio of 0.7. In Figure 3.8, we observe that the Kronecker approximation acts as an average between both diagonal correlations s_1 and s_2, and that deviations are smaller than those observed with the one-ring model. Still, for
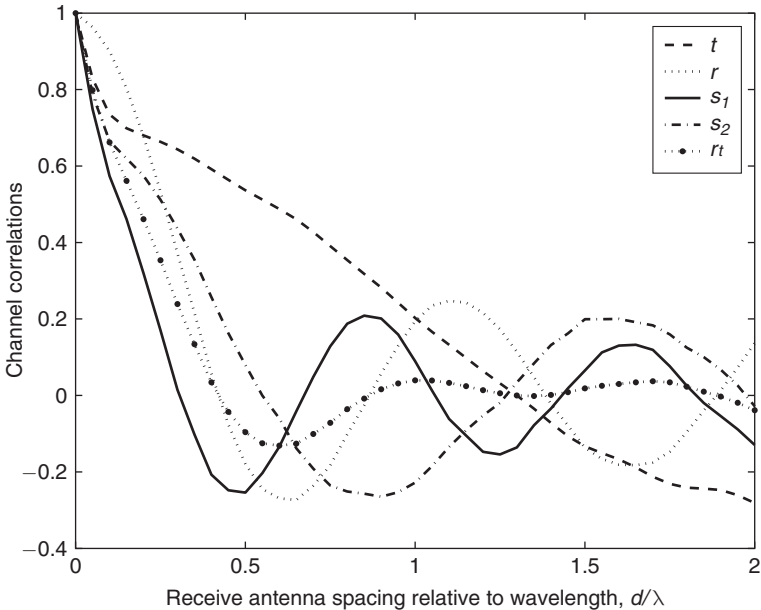
Figure 3.8 *Channel correlations (including the Kronecker approximation of s_1 and s_2 as equal to rt) vs. receive antenna spacing d/λ*

$d > \lambda$, the Kronecker model predicts a reduced cross-channel correlation (in absolute value) with respect to the exact correlations. How important the impact of this underestimation is remains to be seen later, as it depends on the chosen metric. As a consequence, this second geometry-based model appears to be better (although not perfectly!) represented by the Kronecker approximation, owing to the combination of one ring with various ellipses of scatterers (one DoD might correspond to more than one DoA).

The above discussion and examples help us introduce a paradox of the Kronecker model: unless it is rigorously valid (it is then perfectly valid for any antenna spacing), it appears to be approximately valid only for low-correlated channels (at one side at least). Yet the Kronecker model is widely used to account precisely for arbitrary correlation effects. The larger the coupling between DoDs and DoAs, the worse the approximation of the diagonal correlations. That brings us to a final question: how important is the error introduced by the Kronecker model? The answer is not straightforward, as it depends on which output or performance parameter is considered. Elements toward a nuanced answer are given in [Oez04, OCB05], and will be investigated further in Section 3.4.3 in a global comparison between all three simplified channel models and measured data.

3.4.2   Numerical examples

Consider the channel introduced in Example 3.1 (Section 3.3.3). The description can actually be formalized through the finite scatterer representation of Section 2.1.5, with $n_{s,t} = n_{s,r} = n_s = 3$. The central matrix $\mathbf{H}_s$ is equal to $\mathbf{I}_3 \odot \mathbf{H}_w / \sqrt{n_s}$ and $\mathbf{1}_{3 \times 3} \odot \mathbf{H}_w / n_s$, for scenarios A and B respectively. Note that the dimension of $|\mathbf{H}_s|$ is 3×3, irrespective of the number of antennas.

Full correlation
For both proposed scenarios, let us first compute the full correlation matrix for a 2×2 system. Because antenna spacings are small ($d_t = d_r = \lambda/2$), we may safely assume that $t_1 = t_2$ and $r_1 = r_2$ in (3.2). These coefficients can easily be expressed based on (2.44) and similar developments, using the fact that the discrete azimuth-power spectra may be written as the sum of Dirac functions with equal weights. Hence

- for scenario A

$$t = \frac{1}{3} \sum_{l=1}^{3} \exp\left(-j\pi \cos \theta_{t,l}\right) \cong 0.03 - 0.20j$$

$$r = \frac{1}{3} \sum_{l=1}^{3} \exp\left(-j\pi \cos \theta_{r,l}\right) \cong -0.40 + 0.33j$$

$$s_1 = \frac{1}{3} \sum_{l=1}^{3} \exp\left(-j\pi \cos \theta_{t,l} - j\pi \cos \theta_{r,l}\right) \cong -0.19 + 0.21j$$

$$s_2 = \frac{1}{3} \sum_{l=1}^{3} \exp\left(-j\pi \cos \theta_{t,l} + j\pi \cos \theta_{r,l}\right) \cong 0.56 + 0.03j$$

- for scenario *B*

$$t = \frac{1}{3} \sum_{l=1}^{3} \exp\left(-j\pi \cos \theta_{t,l}\right) \cong 0.03 - 0.20j,$$

$$r = \frac{1}{3} \sum_{l=1}^{3} \exp\left(-j\pi \cos \theta_{r,l}\right) \cong -0.40 + 0.33j,$$

$$s_1 = rt \cong 0.05 + 0.09j,$$

$$s_2 = rt^* \cong -0.08 - 0.07i.$$

Kronecker model

The Kronecker model is based only on t and r, and its parameters are

$$\mathbf{R}_t = \begin{bmatrix} 1 & t^* \\ t & 1 \end{bmatrix}$$

$$\mathbf{R}_r = \begin{bmatrix} 1 & r^* \\ r & 1 \end{bmatrix}.$$

Note that it is identical for both considered scenarios, although we know from Proposition 3.2 that only scenario *B* is accurately represented by the separability assumption.

Virtual channel representation

The virtual channel representation of the channel relies on two virtual angles, identical at each end, $\hat{\varphi}_t = \hat{\varphi}_r = 0$ and -0.5, yielding unitary virtual steering matrices

$$\hat{\mathbf{A}}_t = \hat{\mathbf{A}}_r = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -j & j \end{bmatrix}.$$

The virtual correlation matrix is given by (3.24)

- for scenario *A*

$$\mathbf{R}_v = \begin{bmatrix} 1.51 & -0.15j & 0.31j & -0.18 \\ 0.15j & 1.16 & 0.18 & 0.49j \\ -0.31j & 0.18 & 0.10 & 0.09j \\ -0.18 & -0.49j & -0.09j & 1.24 \end{bmatrix},$$

- for scenario *B*

$$\mathbf{R}_v = \begin{bmatrix} 1.07 & -0.04j & 0.32j & 0.01 \\ 0.04j & 1.60 & -0.01 & 0.48j \\ -0.32j & -0.01 & 0.54 & -0.02j \\ 0.01 & -0.48j & 0.02j & 0.80 \end{bmatrix}.$$

Assuming that elements of $\mathbf{H}_v$ are uncorrelated is equivalent to canceling all elements in $\mathbf{R}_v$ except for the main diagonal, whose elements form the power-coupling matrix $\mathbf{\Omega}_v \odot \mathbf{\Omega}_v$. As a consequence

$$\mathbf{\Omega}_{v,A} = \begin{bmatrix} 1.23 & 1.08 \\ 0.33 & 1.11 \end{bmatrix}$$

$$\mathbf{\Omega}_{v,B} = \begin{bmatrix} 1.03 & 1.26 \\ 0.73 & 0.89 \end{bmatrix}.$$

Eigenbeam model

To build the eigenbeam model of the channel, we need to compute the EVD of $\mathbf{R}_t$ and $\mathbf{R}_r$ and extract the eigenvectors

$$\mathbf{U}_{\mathbf{R}_t} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ e^{j\pi 0.55} & e^{j\pi 0.55} \end{bmatrix}$$

$$\mathbf{U}_{\mathbf{R}_r} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -1 \\ e^{-j\pi 0.22} & -e^{-j\pi 0.22} \end{bmatrix}.$$

These two matrices are unitary, and independent of the scattering scenario (A or B). However, the coupling matrices $\mathbf{\Omega}_e$ differ, as they are built from the square-root of the four diagonal elements of $\mathbf{R}_e$,

- for scenario A

$$\mathbf{R}_e = \begin{bmatrix} 1.55 & 0.04j & 0.31j & -0.26 \\ -0.04j & 1.50 & 0.26 & -0.31j \\ -0.31j & 0.26 & 0.85 & -0.04j \\ -0.26 & 0.32j & 0.04j & 0.10 \end{bmatrix},$$

- for scenario B

$$\mathbf{R}_e = \begin{bmatrix} 1.83 & 0 & 0 & 0 \\ 0 & 1.22 & 0 & 0 \\ 0 & 0 & 0.57 & 0 \\ 0 & 0 & 0 & 0.38 \end{bmatrix}.$$

Note that in scenario B, $\mathbf{R}_e$ is diagonal, indicating that the eigenbeam model represents channel B without any approximation.

So,

$$\mathbf{\Omega}_{e,A} = \begin{bmatrix} 1.24 & 1.22 \\ 0.92 & 0.32 \end{bmatrix}$$

$$\mathbf{\Omega}_{e,B} = \begin{bmatrix} 1.35 & 1.10 \\ 0.76 & 0.62 \end{bmatrix}.$$

For scenario A, $\mathbf{\Omega}_e$ is of rank 2, while it is only of rank 1 for scenario B.

Model comparison with three scatterers and a 2×2 system

Tables 3.1 and 3.2 summarize the deviations of the modeled correlation matrices with respect to the original correlation matrices for all three models. We observe that all three models perform relatively similarly in scenario A, while both the Kronecker and the eigenbeam models fit precisely in scenario B (that should not be surprising!). The correlation matrix distance is in any case lower than 8%.

Model comparison with three scatterers and an 8×8 system

Using this example, let us now investigate the impact of the array sizes. Although this scenario is slightly academic (since 8 antennas are not theoretically needed to exploit the benefit of only three scatterers, see Chapter 4), it is particularly convenient to highlight some of the modeling weaknesses. Therefore, we consider the same channel, but with an 8×8 system. The different matrices are not given here ($\mathbf{R}$ is a 64×64 matrix), but Tables 3.3 and 3.4 summarize the different results.

We observe that all models provide degraded fit between the exact and the modeled correlation s for scenario A. Analogous to 2×2 systems, the Kronecker and eigenbeam models

Table 3.1 $\Psi(\mathbf{R}, \mathbf{R}_{model})$ *for different models in scenarios A and B for 2×2 system*

Scenario/model	Kronecker	Virtual	Eigenbeam
A	0.40	0.38	0.33
B	0	0.36	0

Table 3.2   *Correlation matrix distance $d_{corr}(\mathbf{R}, \mathbf{R}_{model})$ in scenarios A and B for 2×2 system*

Scenario/model	Kronecker	Virtual	Eigenbeam
A	0.08	0.08	0.06
B	0	0.07	0

Table 3.3 $\Psi(\mathbf{R}, \mathbf{R}_{model})$ *for different models in scenarios A and B for 8×8 system*

Scenario/model	Kronecker	Virtual	Eigenbeam
A	0.80	0.85	0.79
B	0	0.76	0.06

Table 3.4   *Correlation matrix distance $d_{corr}(\mathbf{R}, \mathbf{R}_{model})$ in scenarios A and B for 8×8 system*

Scenario/model	Kronecker	Virtual	Eigenbeam
A	0.40	0.47	0.39
B	0	0.35	0

allow for a correct representation of propagation scenario *B*. In scenario *A*, the eigenbeam model appears to be more robust than the virtual channel representation, which performs worst among the three models. To better understand the limitations of the virtual channel model in such a scenario (a few scatterers with unique coupling and a medium number of antennas), let us have a look at the coupling matrix for scenario *A*

$$
\Omega_{v,A} = \begin{bmatrix}
1.91 & 0.20 & 0.26 & 0.67 & 0.69 & 0.57 & 0.55 & 0.24 \\
3.87 & 0.47 & 0.67 & 1.89 & 1.89 & 0.76 & 0.58 & 0.42 \\
1.06 & 0.45 & 0.67 & 1.89 & 1.89 & 0.74 & 0.55 & 0.39 \\
0.64 & 0.17 & 0.24 & 0.67 & 0.67 & 0.40 & 0.36 & 0.18 \\
0.53 & 0.13 & 0.18 & 0.45 & 0.46 & 0.43 & 0.41 & 0.17 \\
0.53 & 0.16 & 0.19 & 0.40 & 0.44 & 0.65 & 0.65 & 0.24 \\
0.82 & 0.50 & 0.51 & 0.69 & 0.95 & 2.48 & 2.48 & 0.87 \\
0.87 & 0.26 & 0.29 & 0.53 & 0.62 & 1.23 & 1.23 & 0.44
\end{bmatrix}.
$$

We observe that eight of its elements are significant (say, above 1.50), although the channel is only made of three paths. Because of the decorrelation assumption, these eight elements will represent eight independent paths, hence the degradation of the metrics.

Another way of considering the artifact of the virtual channel model is illustrated in Figure 3.9, which depicts the imaging of scattering scenario *A* provided by the virtual channel representation. The graph represents $\Omega_{v,A} = \mathcal{E}\{|H_v|^2\}$ as a function of the physical angles θ_t and θ_r. We observe that the three scatterers are well reproduced. Yet, the point-like scatterers are enlarged because of the reduced angular resolution (but a finer resolution is
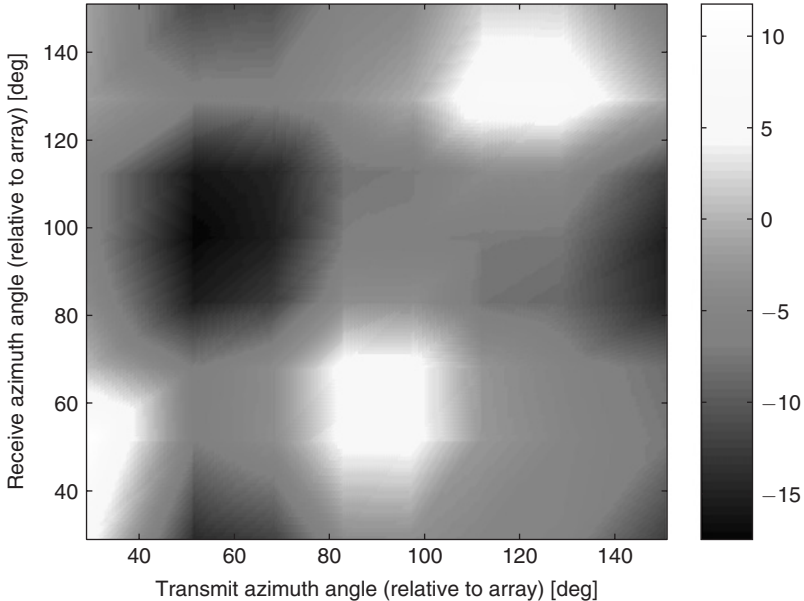
Figure 3.9 *Imaging of the scattering environment for scenario A via* $\mathcal{E}\{|H_v|^2\}$ *(in dB scale)*
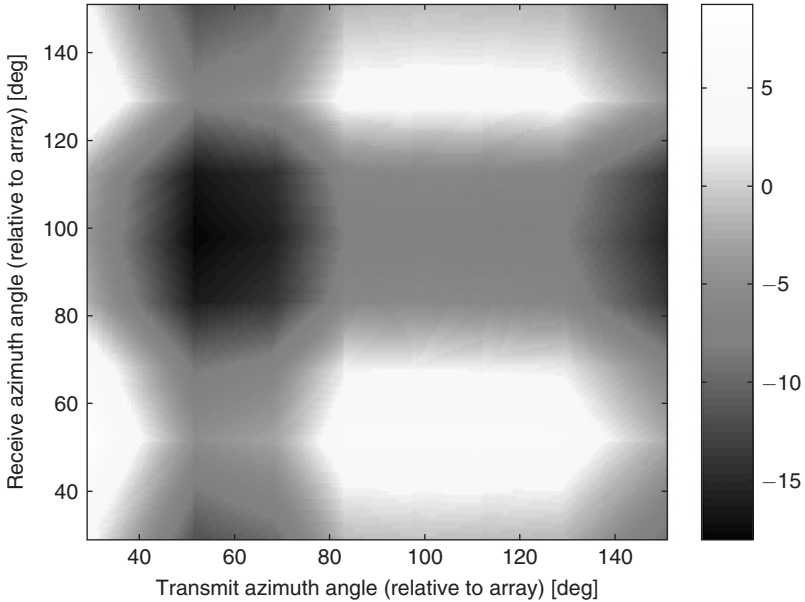
Figure 3.10   *Imaging of the scattering environment for scenario B via $\mathcal{E}\{|H_v|^2\}$ (in dB scale)*

theoretically not needed as we only have 8 antennas). However, the decorrelation assumption causes each enlarged scatterer to be represented by independent sub-scatterers. In this case, the virtual channel model will make the channel richer than it actually is. Naturally, if we had optimized the choice of the virtual directions, the performance would have been better, but still worse than the eigenbeam model. Indeed, the latter chooses the optimal left and right unitary matrices, based on the knowledge of $\mathbf{R}_r$ and $\mathbf{R}_t$, so that the eigendirections are aligned with the most significant scatterers. By contrast, even if we choose the best virtual eigendirections, the virtual partitioning is fixed once one direction at each end is fixed. The imaging of scattering scenario B is depicted in Figure 3.10. Although we should observe nine spots corresponding to the nine pairs of DoD-DoA, we only see four of them, because:

- two DoAs are almost equal ($\pi/4$ and $\pi/3$), corresponding to the same virtual receive beam
- two DoDs are not sufficiently far apart for the angular resolution, so that the respective side lobes meet, yielding two enlarged scattering spots.

The consequence is that the error of the virtual representation caused by the decorrelation assumption (the increase of independent paths) may be slightly compensated by this blurring effect reducing the global number of scattered paths.

Model comparison with twenty scatterers
Finally, we investigate the impact of increasing the number of scatterers n_s to 20. In this case, the DoAs and DoDs are fixed to arbitrary values randomly chosen between 0 and

Table 3.5 $\Psi(\mathbf{R}, \mathbf{R}_{model})$ *for different models in scenarios A and B*

Scenario/model	Kronecker	Virtual	Eigenbeam
A (2×2)	0.14	0.27	0.07
A (8×8)	0.68	0.71	0.61
B (2×2)	0	0.27	0.01
B (8×8)	0	0.59	0.11

Table 3.6 *Correlation matrix distance $d_{corr}(\mathbf{R}, \mathbf{R}_{model})$ in scenarios A and B*

Scenario/model	Kronecker	Virtual	Eigenbeam
A (2×2)	0.01	0.04	0.03
A (8×8)	0.27	0.30	0.21
B (2×2)	0	0.04	0
B (8×8)	0	0.19	0.01

2π. Tables 3.5 and 3.6 outline the different results. The compensation effect of the virtual channel model is more pronounced here, reducing the error on the correlation matrix distance.

3.4.3 Comparison between analytical models: a system viewpoint

When comparing simplified models of the correlated Rayleigh-fading MIMO channel using a performance viewpoint rather than the correlation matrix, we must first derive a few performance metrics for the channels (scenarios A and B, with 3 and 20 scatterers) and systems (2×2 and 8×8) for which we have already analyzed the prediction of the correlation matrix.

Regarding 2×2 systems, all models perform similarly, which could have been expected from the above analysis; as an example, relative errors on the mutual information range from 0 to 2%. The largest errors are found for the virtual channel model with 3 scatterers (2.2 and 2.5% for scenarios A and B, respectively).

For 8×8 systems in scenario A with 3 scatterers (see Figure 3.11), the Kronecker and eigenbeam models are pretty much equivalent regarding the prediction of mutual information, which they underestimate by 3.5% at 25 dB SNR. By contrast, the virtual channel model overestimates the mutual information by up to 88% at 25 dB SNR. This results directly from the imaging issue previously discussed, as the virtual channel representation tends to model the channel with more independent multipaths than it actually has, therefore overestimating the mutual information. Results are pretty similar for scenario B. When increasing the number of discrete scatterers from 3 to 20, the relative errors on the mutual
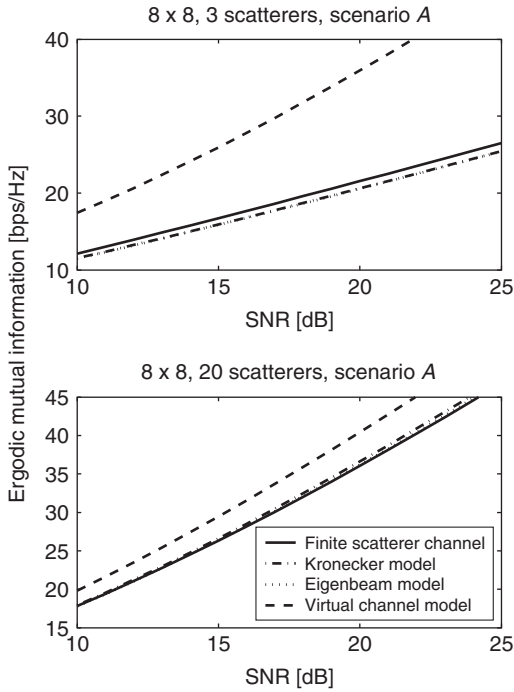
Figure 3.11 *Mutual information vs. SNR for different models in two 8×8 scattering scenarios*

information (with respect to that of the finite scatterer channel) reduce to 2, 1 and 12% respectively for the Kronecker, eigenbeam and virtual channel models in scenario A, and to 0, 0 and 10% in scenario B.

If we use the effective diversity measure as performance metric, we obtain similar results. For 2×2 systems, all simplified models perform almost equally, with the largest errors (14%) being observed with the virtual channel representation for the 3-scatterer channel. For 8×8 systems, all models overestimate the diversity measure in scenarios A, sometimes by more than 100%, while only the virtual channel model fails in scenarios B. Again, the largest errors are produced by the virtual channel model, which overestimates the diversity measure by a factor of 3 to 4 in the worst cases (i.e. for the 8×8 3-scatterer channel A).

As a conclusion, we observe that in the case of discrete Rayleigh fading scatterers, both the Kronecker model and the eigenbeam model provide good representations of the channel for array sizes ranging from 2 to 8. Further interesting results have been produced in the framework of European COST Action 273 [Oez04, OCB05, Cor06]. Based on an extensive indoor experimental MIMO campaign and several system performance indicators (mutual information, diversity order, beamforming gain), the results can be summarized as follows.

- The eigenbeam performs very well for evaluation of the mutual information, while the Kronecker model still remains a credible alternative, especially for MIMO systems

with a limited spatial resolution, such as 2×2, and 3×3 (but the relative error on the mutual information remains below 10% for 8×8 systems, being maximum for high correlation values). By contrast, the virtual channel representation may yield relative errors much higher than 10% for large arrays such as 8×8.

- For beamforming purposes on very large arrays (typically much more than 8), the virtual channel representation model might be the simplest choice, given its angle-based approach, although no validation has been conducted for very large arrays.
- None of the simplified models reproduce the diversity measure correctly. While the eigenbeam model outperforms both the Kronecker model (though only slightly) and the virtual channel representation, it may still show errors up to 100%.

Coming back to the Kronecker model paradox, we may now conclude that the Kronecker model may apparently be used for modeling Gaussian MIMO channels if one of the following situations is met (these situations are not mutually exclusive):

- DoDs and DoAs are uncoupled (that also includes the case of a very small angle-spread at one side only)
- arrays of limited size are considered, typically less than 2 to 3 although the use of 8×8 arrays does not seem to provide huge errors
- antennas at one side of the link (usually at the base station) are sufficiently spaced so that decorrelation is achieved at that side.

In all other scenarios (including MIMO systems with larger compact arrays in channels characterized by strong DoD-DoA coupling), there is no straightforward answer, although the eigenbeam model tends to perform better (though not always well) in both experimental and simulated scenarios. The Kronecker model does still perform satisfactorily in many cases. By contrast, the discrete angular sampling of the virtual channel model may bring significant artifacts, and might not be really advantageous in designing robust space–time codes. Given its simplicity, the Kronecker model appears, therefore, as a good compromise.

This page intentionally left blank

Mutual information and capacity of real-world random MIMO channels

The capacity C is usually defined as the maximum error-free data rate that a channel is able to support. A classical assumption when evaluating the capacity of MIMO channels is to consider the fading coefficients between the pairs of transmit-receive antennas as independent and identically Rayleigh distributed (i.i.d.). This so-called i.i.d. model ($\mathbf{H} = \mathbf{H}_w$) has been detailed in Section 3.1.1. It is also usual to assume a block fading channel, i.e. that the channel is constant over a block of length equal to the coherence time of the channel (T_{coh}) and changes independently from block to block.

When T_{coh} is much smaller than the codeword duration T, the channel is defined as fast fading. As any codeword experiences a large number of different channel realizations (at the limit, the whole channel distribution is explored), the achievable rate is well represented by the so-called ergodic capacity $\bar{C}$, which is the ensemble average of the information rate over the statistics of the channel matrix $\mathbf{H}$, i.e. $\bar{C} = \mathcal{E}\{C(\mathbf{H})\}$.

In slow fading (quasi-static) channels, the coherence time T_{coh} is much larger than the codeword duration T. The channel thereby remains constant over the duration of a codeword but changes independently from block to block. A useful notion in this case is the outage capacity, i.e. the information rate guaranteed for a given percentage of the channel realizations. There is indeed a certain probability P_{out} that a given rate R is not supported by a random channel realization, resulting in the declaration of an outage.

In this chapter we will address the following issues.

- We start by evaluating the ergodic capacity with perfect channel state information at the transmitter (CSIT), and we introduce the popular water-filling algorithm.
- Then we consider with more details the cases when the system has, at most, a partial (stochastic) knowledge of the channel at the transmitter. Typically, the transmitter is

informed about the channel distribution, hence we use the denomination CDIT for Channel Distribution Information at the Transmitter. Both uncorrelated and correlated channels are investigated from an ergodic point of view.

- The impact of several propagation scenarios on the ergodic mutual information is also considered.
- Finally, we deal with outage capacity considerations leading to the diversity-multiplexing trade-off.

As far as notations are concerned, we will often observe that expressions depend upon $\max\{n_t, n_r\}$ and $\min\{n_t, n_r\}$ rather than n_t and n_r directly. Hence, we rewrite when needed the sizes of the arrays as $N = \max\{n_t, n_r\}$ and $n = \min\{n_t, n_r\}$.

4.1   Capacity of fading channels with perfect transmit channel knowledge

For a given MIMO channel realization $\mathbf{H}$, the mutual information is written as [Tel95]

$$\mathcal{I}(\mathbf{H}, \mathbf{Q}) = \log_2 \det\left[\mathbf{I}_{n_r} + \rho \mathbf{H}\mathbf{Q}\mathbf{H}^H\right] \tag{4.1}$$

where $\rho = E_s/\sigma_n^2$ is the SNR and $\mathbf{Q}$ is the input covariance matrix whose trace is normalized to unity.

In particular, $\mathbf{Q} = \mathbf{I}_{n_t}/n_t$ corresponds to an equal power allocation scheme over the transmit antennas, and the corresponding mutual information is denoted as $\mathcal{I}_e(\mathbf{H})$. When the transmitter has the full instantaneous knowledge of the channel state information, the capacity is defined as follows.

Definition 4.1   *The capacity of a $n_r \times n_t$ MIMO channel with perfect channel state information at the transmitter is*

$$C_{CSIT}(\mathbf{H}) = \max_{\mathbf{Q} \geq 0:\mathrm{Tr}\{\mathbf{Q}\}=1} \log_2 \det\left[\mathbf{I}_{n_r} + \rho \mathbf{H}\mathbf{Q}\mathbf{H}^H\right] \tag{4.2}$$

To obtain the optimum input covariance matrix $\mathbf{Q} = \mathbf{Q}^\star$, the transmission is first decoupled along the individual channel modes, so that n parallel data pipes are formed in the directions of the singular vectors of the channel matrix $\mathbf{H}$ at both the transmitter and the receiver. Along these modes, deriving $\mathbf{Q}^\star$ requires to find the optimal power allocation $\{p_1^\star, \ldots, p_n^\star\}$ across these modes and eventually express $\mathbf{Q}^\star$ as

$$\mathbf{Q}^\star = \mathbf{V}_\mathbf{H} \,\mathrm{diag}\{p_1^\star, \ldots, p_n^\star\}\mathbf{V}_\mathbf{H}^H, \tag{4.3}$$

where $\mathbf{V}_\mathbf{H}$ is given by the SVD of $\mathbf{H}$, already introduced in previous chapters

$$\mathbf{H} = \mathbf{U}_\mathbf{H}\mathbf{\Sigma}_\mathbf{H}\mathbf{V}_\mathbf{H}^H, \tag{4.4}$$

and $\mathbf{\Sigma_H} = \mathrm{diag}\{\sigma_1, \ldots, \sigma_n\}$, and $\sigma_k^2 \triangleq \lambda_k \triangleq \lambda_k(\mathbf{W})$ with $\mathbf{W} = \mathbf{HH}^H$ (for $n_t > n_r$) or $\mathbf{W} = \mathbf{H}^H\mathbf{H}$ (for $n_t < n_r$). The capacity is then written as

$$C_{CSIT}(\mathbf{H}) = \max_{\{p_k\}_{k=1}^n} \sum_{k=1}^n \log_2\left[1 + \rho p_k \lambda_k\right] \tag{4.5}$$

$$= \sum_{k=1}^n \log_2\left[1 + \rho p_k^\star \lambda_k\right]. \tag{4.6}$$

and the optimal power allocation $\{p_1^\star, \ldots, p_n^\star\}$, resulting from the power-constrained maximization, is given by the well-known water-filling algorithm [CT91, CKT98, Tel99].

Proposition 4.1 *The power allocation strategy $\{p_1, \ldots, p_n\} = \{p_1^\star, \ldots, p_n^\star\}$ that maximizes $\sum_{k=1}^n \log_2(1 + \rho\lambda_k p_k)$ under the power constraint $\sum_{k=1}^n p_k = 1$, is given by the water-filling solution*

$$p_k^\star = \left(\mu - \frac{1}{\rho\lambda_k}\right)^+, \quad k = 1, \ldots, n \tag{4.7}$$

where μ is chosen so as to satisfy the power constraint $\sum_{k=1}^n p_k^\star = 1$.

Figure 4.1 illustrates the above equation. For each level $\rho\lambda_k$ larger than $1/\mu$, the optimal power allocation corresponds to filling up the mode up to the water level μ. If $\rho\lambda_k \leq 1/\mu$, no power is allocated to the kth mode.

In practice, the optimal power allocation is iteratively estimated [PNG03]. First set the counter i to 1. At each iteration, the constant μ is evaluated from the power constraint

$$\mu(i) = \frac{1}{n - i + 1}\left(1 + \sum_{k=1}^n \frac{1}{\rho\lambda_k}\right). \tag{4.8}$$
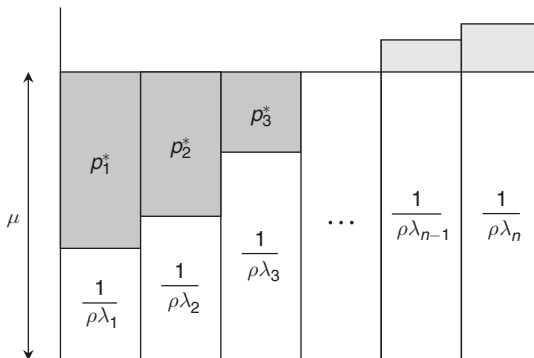
Figure 4.1 *Principle of water-filling algorithm*

The power allocated to each mode at the ith iteration is then computed using

$$p_k(i) = \left(\mu - \frac{1}{\rho \lambda_k} \right)^+, \quad k = 1, \ldots, n - i + 1. \tag{4.9}$$

If the power allocated to the weakest mode is negative (e.g. $p_{n-i+1} < 0$), this mode is dropped out by setting $p_{n-i+1}^\star = 0$ and the power allocation on the other modes is re-calculated with the value of i incremented by 1. This process is iterated until the power allocated on each mode is non-negative.

This highly important result in information and communication theory is extensively used in the following chapters, hence its proof is given below.

PROOF: The maximization problem may be rewritten as

$$\min_{\{p_k\}_{k=1}^n} - \sum_{k=1}^{n} \log_2 \left(1 + \rho \lambda_k p_k \right)$$

$$\text{subject to } p_k \geq 0, \ k = 1, \ldots, n \text{ and } \sum_{k=1}^{n} p_k = 1.$$

This problem can be solved by Lagrange optimization. The proof follows the development made in [BV04]. We write the unconstrained Lagrangian function as

$$\mathcal{L}(p_k, \xi_k, \nu) = - \sum_{k=1}^{n} \log_2 \left(1 + \rho \lambda_k p_k \right) - \xi_k p_k + \nu \left(\sum_{k=1}^{n} p_k - 1 \right) \tag{4.10}$$

with ξ_k being the Lagrange multiplier associated to the inequality constraint $p_k \geq 0$ and ν the Lagrange multiplier for the power constraint $\sum_{k=1}^{n} p_k = 1$. The Karush-Kuhn-Tucker conditions are then written for each value of $k = 1, \ldots, n$ as

$$p_k^\star \geq 0 \tag{4.11}$$

$$\sum_{k=1}^{n} p_k^\star = 1 \tag{4.12}$$

$$\xi_k^\star \geq 0 \tag{4.13}$$

$$\xi_k^\star p_k^\star = 0 \tag{4.14}$$

$$\frac{\partial \mathcal{L}(p_k, \xi_k, \nu)}{\partial p_k} \bigg|_{p_k^\star, \xi_k^\star, \nu^\star} = \frac{\rho \lambda_k}{1 + \rho \lambda_k p_k^\star} - \xi_k^\star + \nu^\star = 0 \tag{4.15}$$

where the superscript * denotes an optimal value. Based on the above equations, we have

$$\nu^\star \geq \frac{\rho \lambda_k}{1 + \rho \lambda_k p_k^\star} \qquad k = 1, \ldots, n \tag{4.16}$$

$$\left(\nu^\star - \frac{\rho \lambda_k}{1 + \rho \lambda_k p_k^\star} \right) p_k^\star = 0 \quad k = 1, \ldots, n \tag{4.17}$$

- If $v^\star < \rho\lambda_k$, (4.16) holds only if $p_k^\star > 0$, which implies by (4.17) that $v^\star = \rho\lambda_k / (1 + \rho\lambda_k p_k^\star)$ or equivalently $p_k^\star = \frac{1}{v^\star} - \frac{1}{\rho\lambda_k}$.
- If $v^\star \geq \rho\lambda_k$, we cannot have $p_k^\star > 0$ since this would imply $v^\star \geq \rho\lambda_k > \rho\lambda_k / (1 + \rho\lambda_k p_k^\star)$, thereby violating (4.17). Hence, $p_k^\star$ must be equal to 0 in this case.

As a result, we may write that

$$p_k^\star = \begin{cases} \frac{1}{v^\star} - \frac{1}{\rho\lambda_k}, & \text{if } v^\star < \rho\lambda_k \\ 0, & \text{if } v^\star \geq \rho\lambda_k \end{cases} \tag{4.18}$$

or equivalently

$$p_k^\star = \left(\frac{1}{v^\star} - \frac{1}{\rho\lambda_k} \right)^+ \tag{4.19}$$

with $x^+ = \max\{x, 0\}$. The level $v^\star$ is determined from the power constraint $\sum_{k=1}^{n} p_k^\star = 1$. By taking $\mu = \frac{1}{v^\star}$, (4.7) is proved. ∎

When the channel is fading, the capacity C is naturally a random variable. If the maximal Doppler frequency is sufficiently high to allow for coding over many channel realizations, the channel is said to be fast fading and the transmission capability is then represented by a single quantity known as the ergodic capacity, i.e.

$$\bar{C}_{CSIT} = \mathcal{E}\left\{ \max_{\mathbf{Q} \geq 0: \text{Tr}\{\mathbf{Q}\}=1} \log_2 \det\left[\mathbf{I}_{n_r} + \rho\mathbf{HQH}^H \right] \right\}$$

$$= \sum_{k=1}^{n} \mathcal{E}\left\{ \log_2 \left[1 + \rho p_k^\star \lambda_k \right] \right\}, \tag{4.20}$$

where $p_k^\star$ and λ_k, $k = 1, \ldots, n$ are now random variables. Indeed, the expression of (4.6) implies that at each time instant (over one fading block), the transmitter adapts the power allocation to achieve the MIMO capacity. This assumes a perfect instantaneous CSIT, i.e. that reciprocity between uplink and downlink applies (e.g. in a TDD scheme) or that the feedback rate is able to cope with the channel variations (e.g. in an FDD scheme).

Low SNR regime
At low SNR, the water-filling algorithm of (4.7) reduces to allocating all the available power to the strongest or dominant eigenmode (characterized by $\lambda_{max} = \max\{\lambda_k\}$), so the ergodic capacity becomes

$$\bar{C}_{CSIT} = \mathcal{E}\left\{ \log_2 \left[1 + \rho\lambda_{max} \right] \right\}$$

$$\cong \rho\mathcal{E}\{\lambda_{max}\} \log_2(e) \tag{4.21}$$

If $\mathbf{H} = \mathbf{H}_w$, we may further develop (4.21) using results of Appendix B and get

$$\bar{C}_{CSIT} \cong \rho \frac{n}{N} \left(\sqrt{N} + \sqrt{n} \right)^2 \log_2(e). \tag{4.22}$$

High SNR regime

In the high SNR regime, the water-filling solution approaches a uniform power allocation on all non-zero eigenmodes, so $p_k^\star = 1/r(\mathbf{H})$ for all values of $k = 1, \ldots, r(\mathbf{H})$. In i.i.d. Rayleigh channels, $\mathbf{H}$ is usually not rank-deficient so that $n = \min\{n_t, n_r\} = r(\mathbf{H})$. The capacity is thus given by

$$\bar{C}_{CSIT} \cong \sum_{k=1}^{n} \mathcal{E} \left\{ \log_2 \left[1 + \frac{\rho}{n} \lambda_k \right] \right\} \tag{4.23}$$

$$\cong n \log_2 \left(\frac{\rho}{n} \right) + \mathcal{E} \left\{ \sum_{k=1}^{n} \log_2(\lambda_k) \right\}. \tag{4.24}$$

Using again results of Appendix B, we may rewrite (4.24) as

$$\bar{C}_{CSIT} = n \log_2 \left(\frac{\rho}{n} \right) + \sum_{k=1}^{n} \mathcal{E} \left\{ \log_2 \left(\chi^2_{2(N-n+k)} \right) \right\},$$

$$= n \log_2 \left(\frac{\rho}{n} \right) + \frac{1}{\log 2} \left(\sum_{k=1}^{n} \sum_{l=1}^{N-k} \frac{1}{l} - n\gamma \right), \tag{4.25}$$

where $\chi^2_{2(N-n+k)}$ designates a χ^2 variable with $2(N - n + k)$ degrees of freedom (see also (4.29) and (4.30) on next page). Note that this does not mean that each λ_k is individually χ^2 distributed, as explained in Appendix B (even if this would result in the same expression). We highlight here the important result that $\bar{C}_{CSIT}$ scales linearly with n. Recalling the definition of the spatial multiplexing gain g_s (1.23), we see that $g_s = n$.

4.2 Ergodic capacity of i.i.d. Rayleigh fast fading channels with partial transmit channel knowledge

Often the transmitter is not able to acquire the instantaneous knowledge of the channel state. Yet it can have partial knowledge, typically the statistical distribution. The ergodic capacity is then defined as follows.

Definition 4.2 *The ergodic capacity of a $n_r \times n_t$ MIMO channel with channel distribution information at the transmitter (CDIT) is given by*

$$\bar{C}_{CDIT} \triangleq \bar{C} = \max_{\mathbf{Q} \geq 0 : \text{Tr}(\mathbf{Q}) = 1} \mathcal{E} \left\{ \log_2 \det \left[\mathbf{I}_{n_r} + \rho \mathbf{H} \mathbf{Q} \mathbf{H}^H \right] \right\}, \tag{4.26}$$

where $\mathbf{Q}$ *is the input covariance matrix optimized as to maximize the ergodic mutual information.*

In i.i.d. Rayleigh fading channels, it has been shown [Tel99] that the ergodic capacity is achieved under an equal power allocation scheme $\mathbf{Q} = \mathbf{I}_{n_t}/n_t$, i.e.

$$\bar{C}_{CDIT} = \bar{\mathcal{I}}_e = \mathcal{E}\left\{ \log_2 \det\left[\mathbf{I}_{n_r} + \frac{\rho}{n_t} \mathbf{H}_w \mathbf{H}_w^H \right] \right\}, \tag{4.27}$$

or equivalently,

$$\bar{C}_{CDIT} = \bar{\mathcal{I}}_e = \mathcal{E}\left\{ \sum_{k=1}^{n} \log_2\left[1 + \frac{\rho}{n_t} \lambda_k \right] \right\}, \tag{4.28}$$

where $n = r(\mathbf{H}_w)$ is the rank of $\mathbf{H}_w$ and $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ are the non-zero eigenvalues of $\mathbf{H}_w \mathbf{H}_w^H$.

Let us first consider a few properties of $\mathbf{H}_w$. As detailed in Appendix B, $\mathbf{W} = \mathbf{H}_w \mathbf{H}_w^H$ (for $n_t > n_r$) or $\mathbf{W} = \mathbf{H}_w^H \mathbf{H}_w$ (for $n_t < n_r$) follow a central Wishart distribution with parameters n and N. From Theorem B.1, we know that the product of the eigenvalues of $\mathbf{W}$ has the distribution $\prod_{k=1}^{n} \chi_{2(N-n+k)}^2$. Furthermore, from [Lee97], we also know that

$$\mathcal{E}\left\{ \log \chi_{2\nu}^2 \right\} = \psi(\nu), \tag{4.29}$$

where $\psi(\nu)$ is the digamma function, which, for integer ν, may be expressed as

$$\psi(\nu) = \sum_{k=1}^{\nu-1} \frac{1}{k} - \gamma, \tag{4.30}$$

where $\gamma \approx 0.57721566$ is Euler's constant.

Coming back to the capacity formula, let us first consider the case of an arbitrary SNR. A lower bound of $\bar{C}_{CDIT}$ directly derived from Theorem B.1 (see Appendix B) is proven in [FG98] and given by two equivalent expressions

$$\bar{C}_{CDIT} \geq \sum_{k=1}^{n} \mathcal{E}\left\{ \log_2\left[1 + \frac{\rho}{n_t} \chi_{2(N-n+k)}^2 \right] \right\}, \tag{4.31}$$

$$\geq \sum_{k=1}^{n} \mathcal{E}\left\{ \log_2\left[1 + \frac{\rho}{n_t} \chi_{2(N-k+1)}^2 \right] \right\}. \tag{4.32}$$

We observe that the MIMO capacity is lower-bounded by the capacity of n parallel SIMO channels whose diversity orders range from $N - n + 1$ up to N. Indeed, on the kth channel, the SNR is equivalent to the SNR achieved by a SIMO MRC scheme with $N - n + k$ independent branches (see Chapter 1 and Appendix B).

The above lower bound has been further developed in [ONBP03] as

$$\bar{C}_{CDIT} \geq \sum_{k=1}^{n} \log_2 \left[1 + \frac{\rho}{n_t} \exp \left(\sum_{l=1}^{N-n+k-1} \frac{1}{l} - \gamma \right) \right], \tag{4.33}$$

which can be rewritten without loss of generality as

$$\bar{C}_{CDIT} \geq \sum_{k=1}^{n} \log_2 \left[1 + \frac{\rho}{n_t} \exp \left(\sum_{l=1}^{N-k} \frac{1}{l} - \gamma \right) \right]. \tag{4.34}$$

Finally, an upper-bound can be obtained (for $n_t \geq n_r$) considering an idealized situation where there is a separate array of receive antennas for each transmit antenna [FG98]. For such system, the total capacity is the capacity of n_t independent SIMO systems, each achieving an equivalent SNR which is χ^2 distributed with $2n_r$ degrees of freedom (see Appendix B). Hence, the upper-bound reads as

$$\bar{C}_{CDIT} \leq \sum_{k=1}^{n_t} \log_2 \left[1 + \frac{\rho}{n_t} \chi^2_{2n_r} \right]. \tag{4.35}$$

Low SNR regime

Considering the low SNR regime (i.e. such that $\rho \lambda_k / n_t << 1$ for all k), we may now expand (4.28) as

$$\bar{C}_{CDIT} = \mathcal{E} \left\{ \log_2 \left[1 + \frac{\rho}{n_t} \sum_{k=1}^{n} \lambda_k + \left(\frac{\rho}{n_t} \right)^2 (\ldots) + \ldots \right. \right.$$

$$\left. \left. + \left(\frac{\rho}{n_t} \right)^n \prod_{k=1}^{n} \lambda_k \right] \right\} \tag{4.36}$$

$$\geq \mathcal{E} \left\{ \log_2 \left[1 + \frac{\rho}{n_t} \|\mathbf{H}_w\|_F^2 \right] \right\}. \tag{4.37}$$

We observe that the ergodic capacity is directly determined by the energy of the channel. Additionally, it is interesting to note that (4.37) is not restricted to i.i.d. channels, but may also be used for correlated channels (see Section 4.3.1). However, in i.i.d. Rayleigh channels, $\|\mathbf{H}_w\|_F^2$ is the sum of $n_t n_r = Nn$ squared independent Rayleigh variables, i.e. it is a χ^2 random variable with $2n_t n_r = 2Nn$ degrees of freedom (see Appendix B). Hence, using Jensen's inequality (see Appendix A) and similar to (4.33), (4.37) becomes

$$\bar{C}_{CDIT} \geq \log_2 \left[1 + \frac{\rho}{n_t} \exp \left(\sum_{k=1}^{Nn-1} \frac{1}{k} - \gamma \right) \right]. \tag{4.38}$$

High SNR regime

At high SNR (such that $\rho\lambda_k/n_t >> 1$ for all k), we may approximate $\bar{C}_{CDIT}$ by

$$\bar{C}_{CDIT} \approx \mathcal{E}\left\{\sum_{k=1}^{n} \log_2\left[\frac{\rho}{n_t}\lambda_k\right]\right\}, \tag{4.39}$$

$$= n\log_2\left(\frac{\rho}{n_t}\right) + \mathcal{E}\left\{\sum_{k=1}^{n}\log_2(\lambda_k)\right\},$$

$$= n\log_2\left(\frac{\rho}{n_t}\right) + \sum_{k=1}^{n}\mathcal{E}\left\{\log_2\left(\chi^2_{2(N-n+k)}\right)\right\}, \tag{4.40}$$

$$= n\log_2\left(\frac{\rho}{n_t}\right) + \frac{1}{\log 2}\left(\sum_{k=1}^{n}\sum_{l=1}^{N-n+k-1}\frac{1}{l} - n\gamma\right), \tag{4.41}$$

$$= n\log_2\left(\frac{\rho}{n_t}\right) + \frac{1}{\log 2}\left(\sum_{k=1}^{n}\sum_{l=1}^{N-k}\frac{1}{l} - n\gamma\right). \tag{4.42}$$

Again, $\chi^2_{2(N-n+k)}$ designates a χ^2 variable with $2(N-n+k)$ degrees of freedom. Using Jensen's inequality (see Appendix A) and Corollary B.1, the last term in (4.42) can be upper-bounded by $\log_2\left[\frac{N!}{(N-n)!}\right]$. The important result is that the ergodic capacity $\bar{C}_{CDIT}$ scales linearly with n. The multiplexing gain g_s is equal to n, similarly to the CSIT case.

It is interesting to compare (4.42) corresponding to CDIT with (4.25) corresponding to CSIT. We already know from (4.20) that the asymptotic capacity achieved through the water-filling algorithm also scales linearly with n, though with a larger rate of growth [CTKV02]. Naturally, the capacity achieved by (4.20) is larger than that for (4.28). Yet, the difference $\bar{C}_{CSIT} - \bar{C}_{CDIT}$ is constant [Gui05] at high SNR, as the water-filling solution for uncorrelated channel approaches a uniform allocation at high SNR. From (4.25) and (4.42), it is evident that this constant equals $n\log_2(n_t/n)$:

- if $n_r \geq n_t$, the constant is close to zero
- if $n_r < n_t$, it equals $n_r\log_2(n_t/n_r)$; intuitively, the gain offered by CSIT corresponds to the amount of energy saved by the transmitter not emitting in the subspace not seen by the receiver.

Ergodic capacity at any SNR

The full calculation of $\bar{C}_{CDIT}$ [Tel99, SL03] at any SNR is formalized in the following proposition.

Proposition 4.2 *The ergodic capacity of i.i.d. Rayleigh fast fading channels is given by*

$$\bar{C}_{CDIT} = \bar{\mathcal{I}}_e = n \int_0^\infty \log_2 (1 + \rho\lambda/n_t) p_\lambda(\lambda) d\lambda, \tag{4.43}$$

where $p_\lambda(\lambda)$ is the distribution of a randomly selected (non-ordered) eigenvalue of $\mathbf{W}$ detailed in Appendix B.

A closed-form expression of (4.43) is obtained in [SL03]

$$\bar{C}_{CDIT} = e^{n_t/\rho} \log_2(e) \sum_{k=0}^{n-1} \sum_{l=0}^{k} \sum_{m=0}^{2l} \left\{ \frac{(-1)^m (2l)! (N-n+m)!}{2^{2k-m} \, l! \, m! \, (N-n+l)!} \right. \tag{4.44}$$

$$\left. \times \binom{2k - 2l}{k - l} \binom{2l + 2N - 2n}{2l - m} \sum_{p=1}^{N-n+m+1} E_p \left(\frac{n_t}{\rho} \right) \right\},$$

where $E_p(z)$ is the exponential integral of order p,

$$E_p(z) = \int_1^\infty e^{-zx} x^{-p} \, dx, \quad \Re[z] > 0. \tag{4.45}$$

Let us now apply (4.44) to three particular cases of MIMO systems:

1. $n_t = N$ and $n_r = 1$ (MISO)

$$\bar{C}_{CDIT} = e^{N/\rho} \log_2(e) \sum_{p=1}^{N} E_p \left(\frac{N}{\rho} \right), \tag{4.46}$$

2. $n_t = 1$ and $n_r = N$ (SIMO)

$$\bar{C}_{CDIT} = e^{1/\rho} \log_2(e) \sum_{p=1}^{N} E_p \left(\frac{1}{\rho} \right), \tag{4.47}$$

3. $n_t = n_r = n = N$

$$\bar{C}_{CDIT} = e^{N/\rho} \log_2(e) \sum_{k=0}^{N-1} \sum_{l=0}^{k} \sum_{m=0}^{2l} \left\{ \frac{(-1)^m}{2^{2k-m}} \binom{2l}{l} \right.$$

$$\left. \times \binom{2k - 2l}{k - l} \binom{2l}{m} \sum_{p=1}^{m+1} E_p \left(\frac{N}{\rho} \right) \right\}. \tag{4.48}$$
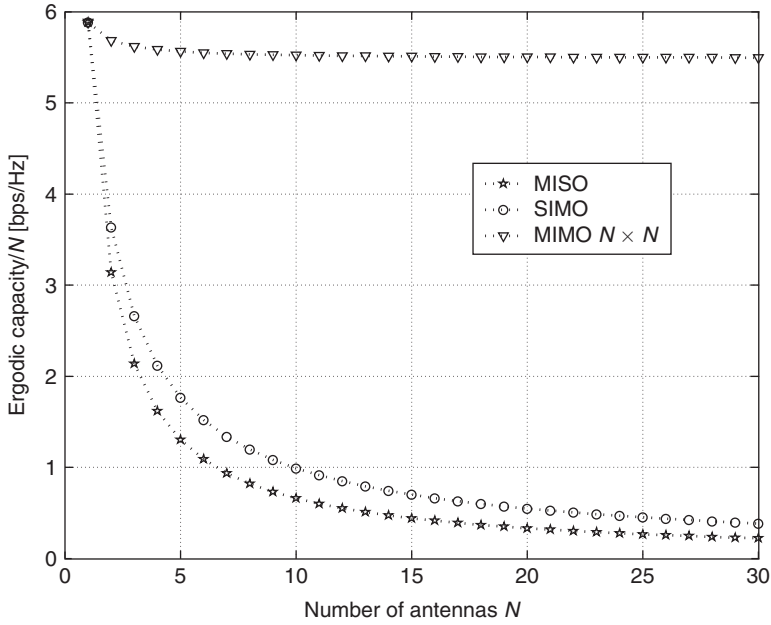
Figure 4.2 *Capacity of various i.i.d. channels at 20 dB SNR*

The capacities per antenna $\bar{C}_{CDIT}/N$ achieved in these three particular cases are depicted in Figure 4.2 as a function of N. Asymptotically (for $N \to \infty$), the ergodic capacity in (4.48) scales linearly with N, while for MISO and SIMO systems, $\bar{C}_{CDIT}/N$ decreases toward zero with increasing N (and the decrease is faster for MISO than for SIMO). This implies that if the number of antennas is increased only at one side, the asymptotic rate of the capacity growth is equal to zero. For $n_t = n_r = N$, we may further write [CTKV02, SL03]

$$\lim_{N\to\infty} \frac{\bar{C}_{CDIT}}{N} = 2\log_2\left(1 + \sqrt{4\rho + 1}\right) - \frac{\log_2(e)}{4\rho}\left(\sqrt{4\rho + 1} - 1\right)^2 - 2. \qquad (4.49)$$

Finally, for $n_t = n_r = N$, a particularly useful approximation of (4.48) is derived in [SL03]

$$\bar{C}_{CDIT} \approx e^{1/\rho}\log_2(e)E_1\left(\frac{1}{\rho}\right)$$

$$+ (N - 1)\left\{2\log_2\left(1 + \sqrt{4\rho + 1}\right) - \frac{\log_2(e)}{4\rho}\left(\sqrt{4\rho + 1} - 1\right)^2 - 2\right\}. \qquad (4.50)$$

Figure 4.3 compares the ergodic capacity of 4×4 i.i.d. Rayleigh fading channels for the CSIT and CDIT cases. Naturally, the CSIT capacity is higher, although the gap reduces
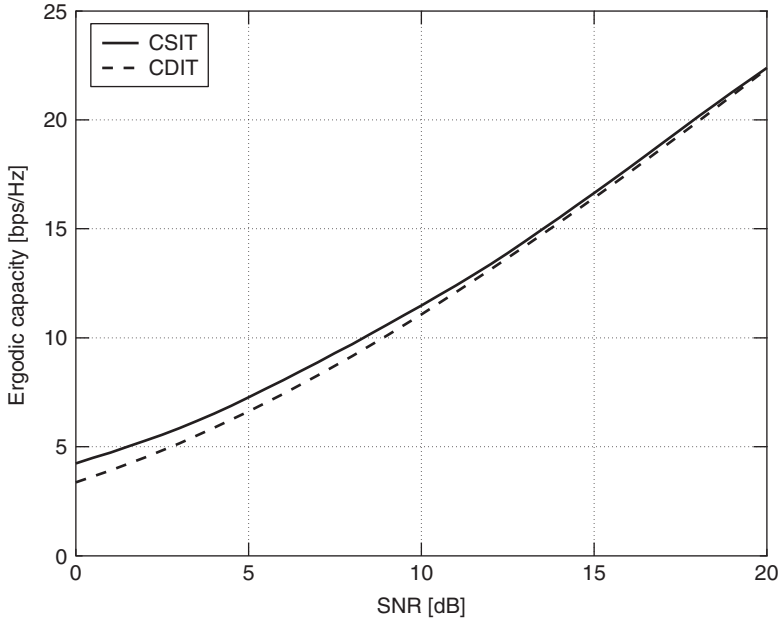
Figure 4.3   *Ergodic capacity of 4×4 i.i.d. Rayleigh channels with full and partial channel knowledge*

when the SNR increases (note that this holds true because $n_t = n_r$, as discussed when considering the high regime).

Asymptotic results

Let us now assume that the number of antennas is large. In this case, the use of the Stieljes transform allows for an alternative analytical formulation of the mutual information [DLLD05]. More specifically, we assume that there are more antennas at the transmitter ($n_t = N$, $n_r = n$), and that $n \to \infty$, $N \to \infty$ in a constant ratio $N/n > 1$.

Proposition 4.3   *Under certain conditions, the empirical distribution of the eigenvalues of W scaled by $1/N$ converges almost surely to the limit probability density $p_{\lambda'}(\lambda')$ (where $\lambda' \triangleq \lambda/N$) which implies that*

$$\frac{\bar{\mathcal{I}}_e}{n} \to \bar{\mathcal{I}}_e^\circ = \int \log_2 \left[1 + \rho\lambda'\right] p_{\lambda'}(\lambda') \, d\lambda'. \tag{4.51}$$

The derivation of the limit distribution $p_{\lambda'}(\lambda')$ relies on the Stieljes transform of $p_{\lambda'}(\lambda')$,

$$G(z) = \int_{\mathbb{R}^+} \frac{1}{\lambda' - z} p_{\lambda'}(\lambda') \, d\lambda', \tag{4.52}$$

which can be thought of as the expected value of $\frac{1}{\lambda'-z}$ for any z. Knowing $G(z)$, the density function is retrieved through the following inversion formula

$$p_{\lambda'}(\lambda') = \frac{1}{\pi} \lim_{\xi \to 0} \Im\left[G(\lambda' + j\xi)\right]. \tag{4.53}$$

For independent Rayleigh fading channels, the Marčenko-Pastur theorem [MP67] yields the following result on $G(z)$

$$G(z) = \frac{1}{\rho^{-1} - z + \frac{n/N}{1+G(z)}}. \tag{4.54}$$

Using the inversion formula in (4.53), we finally obtain

$$p_{\lambda'}(\lambda') = \frac{N}{n} \frac{\sqrt{\left[\left(1 + \sqrt{n/N}\right)^2 - \lambda'\right]\left[\lambda' - \left(1 - \sqrt{n/N}\right)^2\right]}}{2\pi\lambda'} \tag{4.55}$$

if $\lambda' \in \left[\left(1 - \sqrt{n/N}\right)^2, \left(1 + \sqrt{n/N}\right)^2\right]$ and $p_\lambda(\lambda') = 0$ elsewhere. The asymptotic ergodic capacity per receive antenna is eventually given by

$$\lim_{n \to \infty} \frac{\bar{C}_{CDIT}}{n} = \log_2\left(1 + \rho + \rho\frac{n}{N} - \rho\beta\right) + \left(1 - \frac{N}{n}\right)\log_2\left(1 - \beta\right) - \log_2(e)\frac{N}{n}\beta \tag{4.56}$$

with

$$\beta = \frac{1}{2}\left[1 + \frac{n}{N} + \frac{1}{\rho} - \sqrt{\left(1 + \frac{n}{N} + \frac{1}{\rho}\right)^2 - 4\frac{n}{N}}\right]. \tag{4.57}$$

Note that when $N = n$, (4.56) exactly reduces to (4.49).

Antenna selection

A last important example concerns the hybrid selection/MRC scheme, in which n'_r receive antennas are selected among n_r (see Chapter 1). Because the channel is i.i.d. Rayleigh fading, we know that the input covariance matrix is the scaled identity $\mathbf{Q} = \mathbf{I}_{n_t}/n_t$. Yet, the mutual information must be optimized over the receive antenna selection. Denoting as $\mathbf{H}'_w$ the $n'_r \times n_t$ channel matrix obtained by removing $n_r - n'_r$ rows from $\mathbf{H}_w$, the capacity reads as

$$C_{CDIT,HS/MRC} = \max_{\mathcal{S}(\mathbf{H}'_w)} \log_2 \det\left[\mathbf{I}_{n'_r} + \frac{\rho}{n_t}\mathbf{H}'_w\mathbf{H}'^H_w\right] \tag{4.58}$$

where $\mathcal{S}(\mathbf{H}'_w)$ is the set of all possible $\mathbf{H}'_w$.

For $n'_r \leq n_t$, it is relatively easy to derive an upper-bound using the same rationale as for the upper bound of (4.35). Simply exchanging the role of transmitter and receiver in (4.35)

shows that each receive antenna has its own set of n_t transmit antennas. Selecting then the best n_r' receive antennas, the capacity is upper-bounded [MWW01, MWCW05] by

$$C_{CDIT,HS/MRC} \leq \sum_{k=1}^{n_r'} \log_2 \left[1 + \frac{\rho}{n_t} \chi_{2n_t}^2(k) \right] \tag{4.59}$$

where the $\chi_{2n_t}^2(k)$ are the n_r' largest variables among a set of n_r independent χ^2 random variables with $2n_t$ degrees of freedom (see Appendix B). Note that the χ^2 variables are now indexed by (k), in contrast to (4.35), as they must be sorted in order to extract the n_r' largest of them.

Taking $n_r \geq n_r' > n_t$ only increases the diversity order of the channel (or equivalently, the slope of the capacity cumulative distribution). In this case, the above upper-bound is rather loose. To derive a tighter upper-bound [MW04], we see that, in the ideal system leading to (4.35), each SIMO sub-system performs separately HS/MRC. As a consequence, the equivalent SNR of the lth sub-system ρ_l is not $\chi_{2n_r}^2$ but reads as

$$\rho_l = \sum_{k=1}^{n_r'} \chi_2^2(k), \tag{4.60}$$

where the $\chi_2^2(k)$ are obtained by ordering a set of n_r independent χ^2 random variables (with two degrees of freedom, see Appendix B). The upper bound is then simply given by

$$C_{CDIT,HS/MRC} \leq \sum_{l=1}^{n_t} \log_2 \left[1 + \frac{\rho}{n_t} \sum_{k=1}^{n_r'} \chi_2^2(k) \right]. \tag{4.61}$$

Finally, the case $n_r' = n_t$ has been dealt with extensively in [GGP03b, GGP03a]. A lower bound on the capacity is expressed as the sum of two independent quantities, the first corresponding to the capacity with all receive antennas and the second corresponding to a loss term. Defining $\mathbf{L}_w$ as a $n_r \times n_t$ orthonormal basis of the column space of $\mathbf{H}_w$, the lower bound reads as

$$C_{CDIT,HS/MRC} \geq \log_2\det \left[\mathbf{I}_{n_r} + \frac{\rho}{n_t} \mathbf{H}_w \mathbf{H}_w^H \right] + \log_2\det \left[\mathbf{L}_w'^H \mathbf{L}_w' \right], \tag{4.62}$$

where $\mathbf{L}_w'$ is the $n_r' \times n_t$ submatrix of $\mathbf{L}_w$ corresponding to the selected antennas and depends on the selection algorithm. Using an incremental loss minimization algorithm developed in [GGP03a], it can be shown that

$$\log_2\det \left[\mathbf{L}_w'^H \mathbf{L}_w' \right] = \sum_{k=1}^{n_r'} \beta_k^2, \tag{4.63}$$

where the β_k^2 are statistically independent and distributed as detailed in [GGP03a]. Taking $n_r' = n_t = n$ and $n_r = N$, the lower bound of (4.62) is eventually rewritten as

$$C_{CDIT,HS/MRC} \geq \sum_{k=1}^{n} \log_2 \left[1 + \frac{\rho}{n_t} \chi^2_{2(N-k+1)} \beta_k^2 \right]. \tag{4.64}$$

The above expression has a nice physical interpretation: the capacity is lower-bounded by the capacity of n parallel channels whose SNR levels are decreased by a factor β_k^2 ($k = 1, \ldots, n$) with regard to the case with all receive antennas. Yet, it is also clear in (4.64) that the diversity order of each branch is maintained, so that the diversity gain is the same as without selection.

4.3 Mutual information and capacity of correlated Rayleigh channels with partial transmit channel knowledge

4.3.1 *Mutual information with equal power allocation*

Kronecker correlated Rayleigh channels
Let us first consider that the channel covariance matrix $\mathbf{R}$ can be represented by the Kronecker assumption

$$\mathbf{R} = \mathbf{R}_r \otimes \mathbf{R}_t. \tag{4.65}$$

In the low SNR regime, the developments leading to (4.37) are also valid, hence

$$\bar{\mathcal{I}}_e \geq \mathcal{E} \left\{ \log_2 \left[1 + \frac{\rho}{n_t} \|\mathbf{H}\|_F^2 \right] \right\}. \tag{4.66}$$

The statistics of the squared Frobenius norm are not χ^2 anymore. However, they are easily estimated from the moment generating function (MGF) of $\|\mathbf{H}\|_F^2$, given by [PNG03]

$$\mathcal{M}_{\|\mathbf{H}\|_F^2}(\tau) \triangleq \mathcal{E}\left\{ e^{\tau\|\mathbf{H}\|_F^2} \right\}$$

$$= \frac{1}{\det(\mathbf{I}_{n_t n_r} - \tau\mathbf{R})}. \tag{4.67}$$

Substituting $\mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t^{1/2}$ to $\mathbf{H}$, the mutual information with equal-power allocation is well approximated in the high SNR regime by [PNG03]

$$\bar{\mathcal{I}}_e \approx \mathcal{E} \left\{ \log_2 \det \left[\frac{\rho}{n_t} \mathbf{H}_w \mathbf{H}_w^H \right] \right\} + \log_2 \det(\mathbf{R}_r) + \log_2 \det(\mathbf{R}_t). \tag{4.68}$$

Because the sum of the eigenvalues of $\mathbf{R}_r$ is limited by n_r, their product, equal to $\det(\mathbf{R}_r)$, is ≤ 1. That implies that $\log_2\det(\mathbf{R}_r) \leq 0$ (equality is reached if $\mathbf{R}_r = \mathbf{I}_{n_r}$). The same reasoning holds true as far as $\mathbf{R}_t$ is concerned. The conclusion is that receive and transmit correlations always degrade the mutual information (with power uniform allocation) with

respect to the i.i.d. case, and that the high-SNR loss is given by $\log_2\det(\mathbf{R}_r) + \log_2\det(\mathbf{R}_t)$ bits/s/Hz.

We may accurately observe that $\mathbf{W}$ is always a full-rank matrix for Kronecker-structured channels (under the assumption that $\mathbf{R}_t$ and $\mathbf{R}_r$ are full rank), and its n non-zero ordered eigenvalues are characterized by the following joint distribution [KA06a]

$$p_{\lambda_1,\lambda_2,\dots,\lambda_n}(\lambda_1, \lambda_2, \dots, \lambda_n) = \frac{\pi^{n(n-1)}}{\tilde{\Gamma}_n(n)\tilde{\Gamma}_n(N)\det(\mathbf{R}_t)^{n_r}\det(\mathbf{R}_r)^{n_t}} \tag{4.69}$$

$$\prod_{k=1}^{n} \lambda_k^{N-n} \prod_{l<k}^{n} (\lambda_k - \lambda_l)^2 \, _0\tilde{F}_0^{(N)(n)}(\mathbf{R}_t^{-1}, \mathbf{R}_r^{-1}, \mathbf{W}),$$

where $\tilde{\Gamma}_m(p) = \pi^{m(m-1)/2} \prod_{k=1}^{m} \Gamma(p - k + 1)$ is the multivariate gamma function, and the hypergeometric function $_0\tilde{F}_0^{(N)(n)}$ of three matrix arguments is defined as detailed in [KA06a], based on zonal polynomials. However, this method is quite complex and is not presented in this book. Yet, a rigorous solution will be discussed for semi-correlated channels.

Asymptotic results

Before dealing with semi-correlated channels, let us first consider that the number of antennas is large enough. The use of the Stieljes transform allows here for an implicit analytical formulation of the mutual information [DLLD05]. Using Proposition 4.3, and after some (non-trivial) calculations, we eventually obtain that the asymptotic average mutual information per receive antenna $\bar{\mathcal{I}}_e^\circ$ is given by

$$\bar{\mathcal{I}}_e^\circ = \frac{1}{n}\log_2\det(\mathbf{I}_n + \delta_t\mathbf{R}_r) + \frac{1}{n}\log_2\det(\mathbf{I}_N + \delta_r\mathbf{R}_t) - \frac{1}{n}\frac{N}{\rho}\delta_t\delta_r, \tag{4.70}$$

where δ_t and δ_r are the solutions of

$$\delta_t = \frac{\rho}{n}\mathrm{Tr}\left[\boldsymbol{\Lambda}_{\mathbf{R}_t}(\mathbf{I}_N + \delta_r\boldsymbol{\Lambda}_{\mathbf{R}_t})^{-1}\right] \tag{4.71}$$

$$\delta_r = \frac{\rho}{n}\mathrm{Tr}\left[\boldsymbol{\Lambda}_{\mathbf{R}_r}(\mathbf{I}_n + \delta_t\boldsymbol{\Lambda}_{\mathbf{R}_r})^{-1}\right] \tag{4.72}$$

The convergence rate is in $\mathcal{O}(1/n)$ for $\bar{\mathcal{I}}_e$, i.e.

$$\bar{\mathcal{I}}_e = n\bar{\mathcal{I}}_e^\circ + \mathcal{O}(1/n), \tag{4.73}$$

which results in asymptotic approximations being already reasonably good for array sizes as low as 3×3. However, the most important conclusion of this analysis is that even with correlation, the mutual information scales linearly with n (assuming that $\mathbf{R}_t$ and $\mathbf{R}_r$ are full rank).

Semi-correlated Rayleigh channels

Although it is possible to obtain exact results in the form of an infinite series based on (4.69), it is much easier to deal with semi-correlated channels, i.e. to assume that either $\mathbf{R}_t$ or $\mathbf{R}_r$ is the identity matrix[1]. In this case, we denote by $\boldsymbol{\Phi}$ the correlation matrix at the correlated side, while its eigenvalues are denoted as $\phi_1 < \ldots < \phi_n$ (by contrast to $\{\lambda_1, \ldots, \lambda_n\}$, these are non-random parameters). The ergodic mutual information statistics are derived from the moment generating function (MGF) $\mathcal{M}_{\mathcal{I}_e}(\tau) = \mathcal{E}\{e^{\tau \mathcal{I}_e}\}$ by

$$\mathcal{E}\left\{[\mathcal{I}_e]^k\right\} = \left.\frac{d^k \mathcal{M}_{\mathcal{I}_e}(\tau)}{d\tau^k}\right|_{\tau=0}. \tag{4.74}$$

The MGF of the mutual information can be derived from the eigenvalues distribution [KA06a]. From the MGF, the ergodic mutual information under equal power allocation is simply obtained via a first-order derivative, and is eventually expressed (when $\boldsymbol{\Phi}$ is $n \times n$) as

$$\bar{\mathcal{I}}_e = \frac{1}{\log(2)d_v \prod_{k=1}^n \Gamma(N-k+1)} \sum_{k=1}^n \det(\boldsymbol{\Psi}_k), \tag{4.75}$$

where

$$d_v = \left(\prod_{k=1}^n \phi_k^N\right) \prod_{1 \le l < k \le n} \left(\frac{1}{\phi_k} - \frac{1}{\phi_l}\right), \tag{4.76}$$

and $\boldsymbol{\Psi}_k, k = 1, \ldots, n$ are $n \times n$ matrices,

$$\boldsymbol{\Psi}_k(u, v) = \begin{cases} \int_0^\infty \log(1 + y\rho/n_t) y^{N-u} e^{-y/\phi_v} dy, & \text{if } k = u \\ \phi_v^{N-u+1} \Gamma(N-u+1), & \text{if } k \ne u \end{cases} \tag{4.77}$$

When $\boldsymbol{\Phi}$ is $N \times N$, a similar expression can be obtained as detailed in [KA06a].

Non-Kronecker correlated Rayleigh channels

When the channel correlation is not Kronecker separable, closed-forms of the mutual information do not exist. In this case, it is convenient to resort to the use of an upper bound introduced in Section 3.3.2

$$\bar{\mathcal{I}}_e \le \log_2(\bar{\kappa}) = \log_2\left(\mathcal{E}\left\{\det\left[\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H\right]\right\}\right). \tag{4.78}$$

An explicit closed-form linking this upper bound with the channel correlations has been derived in Section 3.3.2, and we have observed that transmit and receive correlations degrade the ergodic mutual information upper bound. By contrast, cross-channel

---

[1] We have observed in Chapter 3 that the Kronecker representation of (4.65) is better verified under this last assumption.
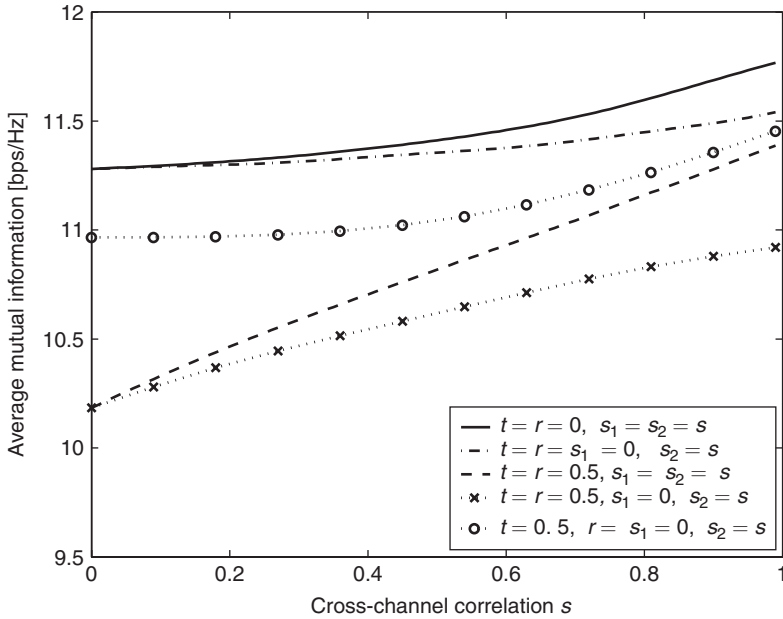
Figure 4.4 *Average mutual information of various correlated channels at 20 dB SNR as a function of one/both cross-channel correlation(s)*

correlations appear to increase the mutual information upper bound. That is confirmed by Monte Carlo simulations of $\bar{\mathcal{I}}_e$.

In Figure 4.4, $\bar{\mathcal{I}}_e$ is depicted for a 2×2 system as a function of s_1 and/or s_2, for various receive and transmit correlations and a SNR $\rho = 20$ dB (the full correlation matrix representation is used). A relative gain of 2 to 4% is observed with respect to the i.i.d. scenario, when increasing one or both cross-channel correlations. In moderately correlated channels, the gain between $s = 0$ and $s = 1$ may be as large as 11%. Furthermore, for channels moderately correlated at the receive/transmit sides, large cross-channel correlations ensure that the ergodic mutual information almost reaches the i.i.d. capacity. As an example, at 20 dB SNR, the following scenario ($t = r = 0.5$; $s_1 = s_2 = 0.95$) yields the same ergodic mutual information as the i.i.d. channel.

Diagonal Rayleigh channels

For 2×2 systems, the above results show that i.i.d. Rayleigh channels do not provide the highest mutual information when the power is equally allocated among all transmit antennas. Based on an analytical upper bound, we have observed that a particular correlation structure (unit cross-channel correlations and zero antenna correlations) increase the ergodic mutual information beyond the capacity of independent Rayleigh channels. This behavior has also been confirmed by simulations. How can we generalize this correlation structure to larger arrays? In this case, not all cross-channel correlations can be set to

one, as then the correlation matrix is no longer positive semidefinite. For $n_t = n_r = N$, it is illustrated in [OO05] that the complex Gaussian channel matrix maximizing the ergodic mutual information under equal power allocation is then given by

$$|\mathbf{H}| = \begin{bmatrix} a_1 & a_2 & \ldots & a_{N-1} & a_N \\ a_2 & a_3 & \ldots & a_N & a_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_N & a_1 & \ldots & a_{N-2} & a_{N-1} \end{bmatrix} \tag{4.79}$$

where $\{a_k\}$, $k = 1, \ldots, N$ are N (real-valued) Rayleigh i.i.d. variables of unit variance, and the notation $|\cdot|$ means that $|\mathbf{H}|(p,q) = |\mathbf{H}(p,q)|$. Such a channel matrix is unitarily equivalent to a diagonal channel matrix $\mathbf{D} = \sqrt{N} \, \text{diag}\{d_1, \ldots, d_N\}$, where $\{d_k\}$ are i.i.d. complex Gaussian variables of unit variance

$$\mathbf{H} = \mathbf{U}_r \mathbf{D} \mathbf{U}_t^T \tag{4.80}$$

with $\mathbf{U}_t$ and $\mathbf{U}_r$ being two unitary matrices. As a consequence, we define these MIMO channels as diagonal channels. Physically speaking, diagonal channels are uncorrelated at transmit and receive sides, but some of their cross-channel correlations are equal to unity. Alternatively, the steering vectors at both transmit and receive sides are orthogonal, since (4.80) is equivalent to the finite scatterer model iff this orthogonality condition is met ($\mathbf{A}_t = \mathbf{U}_t$ and $\mathbf{A}_r = \mathbf{U}_r$). Naturally, the N eigenvalues of $\mathbf{HH}^H$ are independent identically χ^2 distributed with two degrees of freedom [OO05], so the ergodic mutual information is obtained as

$$\bar{\mathcal{I}}_e = N \int_0^\infty \log_2 \left(1 + \frac{\rho}{N} \lambda^2 \right) \frac{2\lambda}{N} \, e^{-\frac{\lambda^2}{N}} d\lambda$$

$$= N \log_2(e) \, e^{\frac{1}{\rho}} \, \text{E}_1 \left(\frac{1}{\rho} \right), \tag{4.81}$$

which is exactly N times the ergodic capacity of a SISO Rayleigh channel (note that $E_1(z)$ is the exponential integral of order 1). For a given SNR, it is easily verified that (4.81) also provides a higher mutual information than the i.i.d. channel capacity of (4.48), as shown in Figure 4.5. However, diagonal channels only present an effective diversity order of N (while for i.i.d. channels, $N_{div} = N^2$). Equivalently, that means that, despite the higher average mutual information offered by diagonal channels, the slope of the cumulative distribution of $\bar{\mathcal{I}}_e$ is smaller. Hence, the outage mutual information is lower for low outage levels P_{out} (approximately for $P_{out} < 15\%$ in a 2×2 channel at $15 \, \text{dB}$ SNR) [OO05].

Finite scatterer channels

So far, it seems that the capacity increase is always linear in the number of antennas, although the rate of growth depends on the channel correlations. That actually assumes that the transmitted signals reach the receive array through a large number of paths. What
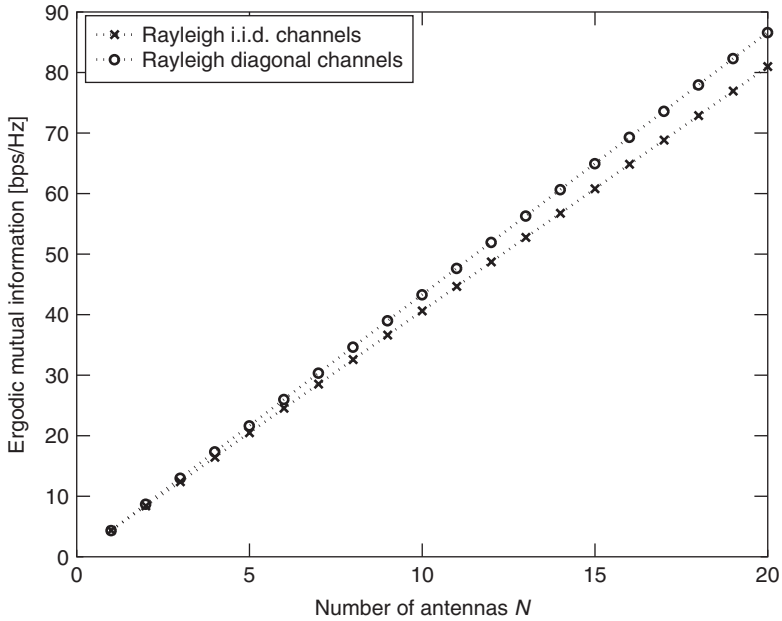
Figure 4.5 *Average mutual information of diagonal channels at 15 dB SNR as a function of the number of antennas at each side*

happens when the number of significant scatterers is limited? In that case, it can be shown that the rank of **H** is equal to the number of scatterers, and the capacity grows only as a function of that same number rather than scaling in the number of antennas. To show this, we resort to the finite scatterer representation of (2.40) introduced in Chapter 2, which gives the channel matrix as

$$\tilde{\mathbf{H}} = \mathbf{A}_r \mathbf{H}_s \mathbf{A}_t^T, \tag{4.82}$$

where $\mathbf{A}_t$ and $\mathbf{A}_r$ represent the $n_t \times n_{s,t}$ and $n_r \times n_{s,r}$ steering matrices, whose columns are the steering vectors related to the directions of each path observed at Tx and Rx, and $\mathbf{H}_s$ is a $n_{s,r} \times n_{s,t}$ matrix whose elements are the complex path gains between all DoDs and DoAs. For uncorrelated scatterers, $\mathbf{H}_s$ can be further developed as $\mathbf{\Omega}_s \odot \mathbf{H}_w$, $\mathbf{\Omega}_s$ denoting the coupling matrix between the different directions (DoDs and DoAs).

Figure 4.6 illustrates the variation of the ergodic mutual information with equal power allocation as a function of $n_t = n_r = N$ for two numbers of scatterers ($n_{s,t} = n_{s,r} = n_s = 3$ or 20). Clearly, the mutual information in the first scenario ($n_s = 3$) does not grow linearly with N but saturates for $N > 3$. To derive asymptotic explicit formulas of the saturation level, [DM05] assumes that

- the antennas are spaced by $\lambda/2$
- $n_{s,t} \le n_t$ and $n_{s,r} \le n_r$ (i.e. one is interested in the saturation level)
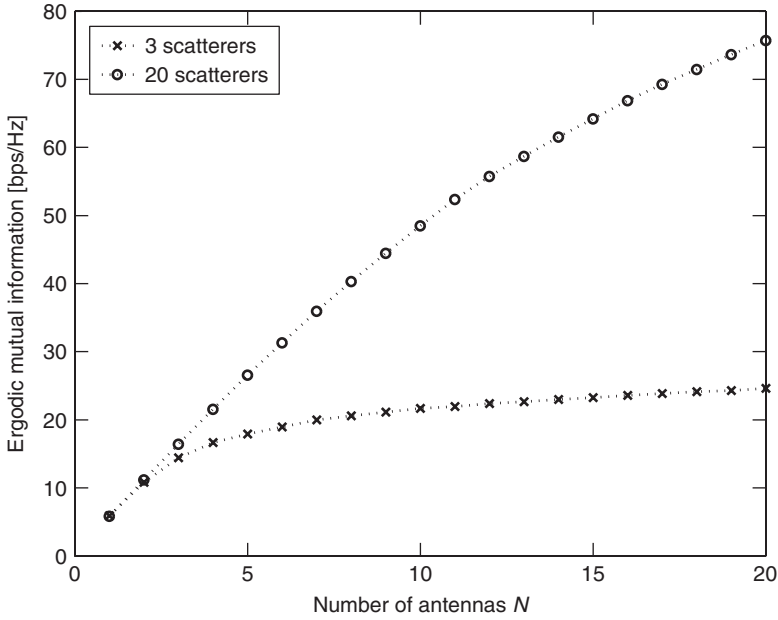
Figure 4.6 *Average mutual information of two finite scatterer channels at 20 dB SNR as a function of the number of antennas at each side*

- the scatterers are maximally distant from each other, i.e. the DoDs and DoAs are respectively given by $\theta_{t,k} = 2k\pi/n_{s,t}$, $k = 1, \ldots, n_{s,t}$ and $\theta_{t,l} = 2l\pi/n_{s,r}$, $l = 1, \ldots, n_{s,r}$
- all paths are coupled with each other and have equal average power, hence $\mathbf{\Omega}_s = \mathbf{1}_{n_{s,r} \times n_{s,t}} / \sqrt{n_{s,r} n_{s,t}}$.

Denoting $\xi_t = n_t/n_{s,t} \geq 1$ and $\xi_r = n_r/n_{s,r} \geq 1$, it is then shown in [DM05] that the asymptotic ergodic mutual information per receive antenna is equal to

$$\frac{\bar{\mathcal{I}}_e}{n_r} = \frac{1}{\xi_r} \log \left[1 + \rho\xi_r - \rho\xi_r\alpha \right] + \frac{1}{\xi_t \nu} \log \left[1 + \rho\nu\xi_t - \rho\xi_r\alpha \right] - \frac{1}{\xi_t \nu} \alpha \qquad (4.83)$$

where $\nu = n_r/n_t$ and

$$\alpha = \frac{1}{2} \left[1 + \frac{\xi_t}{\xi_r} + \frac{1}{\rho\xi_r} - \sqrt{\left(1 + \frac{\xi_t}{\xi_r} + \frac{1}{\rho\xi_r} \right)^2 - \frac{4\xi_t}{\xi_r}} \right]. \qquad (4.84)$$

4.3.2 Ergodic capacity of correlated Rayleigh channels with partial transmit channel knowledge

We now assume that the transmitter is informed about the channel correlation matrix $\mathbf{R}$. That implies that this information is fed back from the receiver to the transmitter, which is

naturally less costly in terms of feedback bits than acquiring the full CSIT. Indeed, we have seen that the channel spatial structure changes at a much slower rate than fading. When $\mathbf{H} = \mathbf{H}_w$, $\mathbf{Q} = \mathbf{I}_{n_t}/n_t$ is the optimal strategy. The choice of the optimal $\mathbf{Q}$ when $\mathbf{H}$ is correlated is examined below. To simplify the analysis, it is usually assumed that the Kronecker channel model is valid, which implies that the channel knowledge reduces to the correlation matrices at the transmitter and/or the receiver ($\mathbf{R}_t$ and/or $\mathbf{R}_r$). Before dealing with general Kronecker channels, we first consider that fading at the receiver is spatially white, $\mathbf{R}_r = \mathbf{I}_{n_r}$.

Semi-correlated Rayleigh channels

In this case, it is shown in [JG01] that the optimal input covariance matrix can always be expressed as

$$\mathbf{Q} = \mathbf{U}_{\mathbf{R}_t} \mathbf{\Lambda}_{\mathbf{Q}} \mathbf{U}_{\mathbf{R}_t}^H, \tag{4.85}$$

where $\mathbf{U}_{\mathbf{R}_t}$ is a unitary matrix formed by the eigenvectors of $\mathbf{R}_t$ (arranged to correspond to decreasing eigenvalues of $\mathbf{R}_t$), and $\mathbf{\Lambda}_{\mathbf{Q}}$ is a diagonal matrix whose elements are also arranged in decreasing order. The maximization problem of (4.26) can be rewritten as

$$\bar{C}_{CDIT} = \max_{\mathbf{\Lambda}_{\mathbf{Q}} \geq 0 : \mathrm{Tr}\{\mathbf{\Lambda}_{\mathbf{Q}}\} \leq 1} \mathcal{E} \left\{ \log_2 \det \left[\mathbf{I}_{n_r} + \rho \mathbf{H}_w \mathbf{\Lambda}_{\mathbf{R}_t}^{1/2} \mathbf{\Lambda}_{\mathbf{Q}} \mathbf{\Lambda}_{\mathbf{R}_t}^{1/2} \mathbf{H}_w^H \right] \right\}, \tag{4.86}$$

where $\mathbf{\Lambda}_{\mathbf{R}_t}$ contains the eigenvalues of $\mathbf{R}_t$ (see (3.27)).

The solution of the maximization is reasonably well approximated by a water-filling solution in the sense that stronger channel modes get allocated more power than weaker modes. To better understand this, let us upper bound (4.86) thanks to Jensen's inequality:

$$\bar{C}_{CDIT} \leq \max_{\mathbf{\Lambda}_{\mathbf{Q}} \geq 0 : \mathrm{Tr}\{\mathbf{\Lambda}_{\mathbf{Q}}\} \leq 1} \sum_{k=1}^{n_t} \log_2 \left(1 + \rho n_r \lambda_k(\mathbf{R}_t) \lambda_k(\mathbf{Q}) \right). \tag{4.87}$$

The Lagrangian optimization of this upper bound yields the classical water-filling solution

$$\lambda_k(\mathbf{Q}) = \left[\mu - \frac{1}{\rho n_r \lambda_k(\mathbf{R}_t)} \right]^+ \tag{4.88}$$

where μ is chosen such that $\mathrm{Tr}\{\mathbf{\Lambda}_{\mathbf{Q}}\} = 1$. Using an input covariance based on (4.88) leads to a mutual information that is a lower bound on the ergodic channel capacity (4.86). However, the lower bound is quite tight when n_r is large compared to n_t. Actually, the gap reduces as the ratio n_r/n_t approaches infinity. When n_t is larger than n_r, such lower bound is pretty loose, especially at high SNR. Indeed, the optimal power allocation may sensibly differ from the power allocation obtained based on Jensen's upper bound [Vu06]. At low SNR, however, transmission should only be done over the dominant mode, which is suggested by (4.88). Note that the use of Jensen's inequality to find an estimate of the optimal power allocation is not limited to semi-correlated Rayleigh channels, but can

also be used for more general channels, as those highlighted in the next sections. More generally, $n_r \lambda_k(\mathbf{R}_t)$ is then simply replaced by $\lambda_k\left(\mathcal{E}\{\mathbf{H}^H\mathbf{H}\}\right)$ in (4.88).

Rigorously, the exact optimal scheme must be obtained through numerical optimization of $\mathbf{\Lambda_Q}$. To get more insight, let us consider the simple case of TIMO (two-input multiple-output) systems with correlation t at the transmit side (and no correlation at the receive side), as proposed in [SM03]. We will thus compare three strategies:

- the optimal power allocation obtained through numerical optimization (the latter is trivial, as only one parameter is free since $\text{Tr}(\mathbf{\Lambda_Q}) = 1$), yielding $\bar{C}_{CDIT}$
- the equal-power allocation $\mathbf{Q} = \mathbf{I}_2/2$, which should be optimal for low t, yielding $\bar{\mathcal{I}}_e$
- the beamforming strategy, i.e. the first element of $\mathbf{\Lambda_Q}$ is equal to one and the other one is zero, i.e. we transmit only on the strongest eigenmode of the channel; the mutual information under that scheme is denoted as $\bar{\mathcal{I}}_{bf}$.

Figure 4.7 shows that, at the considered SNR ($\rho = 0\,\text{dB}$), using beamforming at low transmit correlation or using equal power allocation at high transmit correlation are both extremely inefficient strategies. Beamforming is indeed favored at high transmit correlation, as a high transmit correlation basically implies via (2.44) that $\mathcal{A}_t(\Theta_t)$ is highly directional. Therefore, transmitting only in that preferred direction is unsurprisingly optimal. Similarly, a low transmit correlation is equivalent of a large transmit angle spread, so equal-power allocation is favored. The optimal strategy naturally outperforms both equal-power allocation and
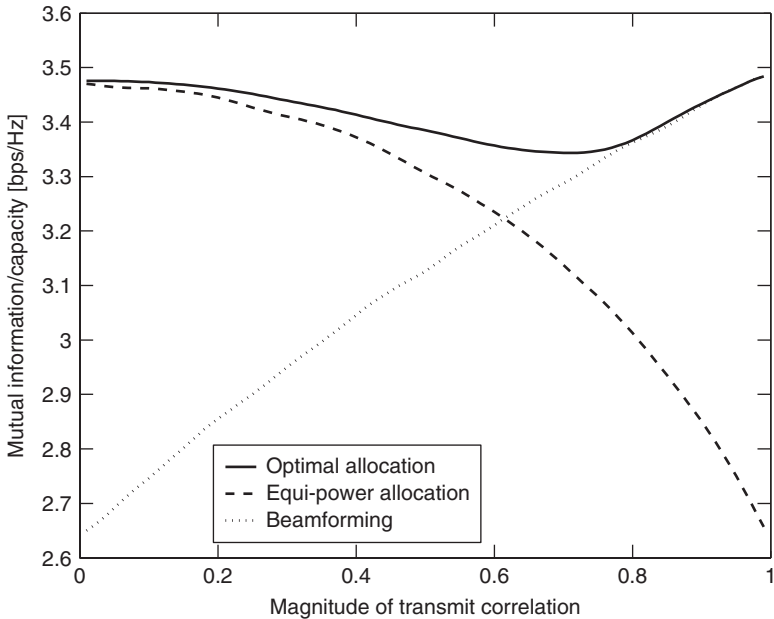
Figure 4.7 *Mutual information of various strategies at 0 dB SNR as a function of the transmit correlation $|t|$*

beamforming at intermediate values of the transmit correlation. Still, a strategy switching between beamforming and equal-power allocation, achieving a mutual information of $\max\{\bar{\mathcal{I}}_{bf}, \bar{\mathcal{I}}_e\}$ comes reasonably close (within roughly 5%) to the capacity $\bar{C}_{CDIT}$. That apparently holds true for any TIMO system at any SNR [SM03]. Interestingly, $\bar{C}_{CDIT}$ is not a decreasing function of $|t|$ (by contrast to $\bar{\mathcal{I}}_e$). In Figure 4.7 ($\rho = 0$dB), $\bar{C}_{CDIT}(|t| = 1)$ is even slightly higher than $\bar{C}_{CDIT}(|t| = 0)$. This observation strongly nuances the popular belief that antenna correlations necessarily degrade the system performance, and also that i.i.d. Rayleigh fading conditions maximize the capacity performance (remember that cross-channel correlations may also be favorable, even with equal-power allocation).

If using a switching strategy, one must know when to switch, i.e. for given n_r and ρ, at which transmit correlation does beamforming achieve larger mutual information than equal-power allocation. In the above example ($n_r = 2$ and $\rho = 0$ dB), beamforming is optimal for $t \geq 0.61$. In [JG01,SM03], a beamforming optimality condition is derived for arbitrary MIMO systems.

Proposition 4.4   *In a $n_r \times n_t$ channel, beamforming is the optimal strategy achieving capacity iff*

$$\Gamma\left[-n_r, \frac{1}{\rho\lambda_1(\mathbf{R}_t)}\right] \geq \frac{[\rho\lambda_1(\mathbf{R}_t)]^{n_r} \rho\lambda_2(\mathbf{R}_t)}{1 + \rho\lambda_2(\mathbf{R}_t)} \exp\left[-\frac{1}{\rho\lambda_1(\mathbf{R}_t)}\right], \qquad (4.89)$$

where Γ denotes the incomplete gamma function and $\{\lambda_1(\mathbf{R}_t) \geq \lambda_2(\mathbf{R}_t)\}$ are the two largest eigenvalues of $\mathbf{R}_t$.

Note that for TIMO systems, $\mathbf{R}_t$ has precisely two eigenvalues, which are equal to $1 \pm |t|$. We should further observe that the transmit correlation for which beamforming becomes optimal depends on both n_r and ρ. This is shown in Figure 4.8 as far as ρ is concerned. The curves represent the optimal fraction of power to non-beamforming mode (a value of zero corresponds to pure beamforming, while a value of 1 corresponds to equal power allocation). The larger ρ (or equivalently, n_r), the smaller the range of $|t|$ over which beamforming is optimal. This could also be inferred from (4.88) where we may observe that ρn_r acts as an effective SNR.

Kronecker correlated Rayleigh channels
In this case, it is shown in [JB04] that the optimal input covariance matrix can again be expressed as

$$\mathbf{Q} = \mathbf{U}_{\mathbf{R}_t} \mathbf{\Lambda}_\mathbf{Q} \mathbf{U}_{\mathbf{R}_t}^H, \qquad (4.90)$$

where $\mathbf{U}_{\mathbf{R}_t}$ is a unitary matrix formed by the eigenvectors of $\mathbf{R}_t$ (arranged in such order that they correspond to decreasing eigenvalues of $\mathbf{R}_t$), and $\mathbf{\Lambda}_\mathbf{Q}$ is a diagonal matrix whose elements are also arranged in decreasing order. That is, the optimal strategy is to employ independent complex circular Gaussian inputs along the eigenvectors of $\mathbf{R}_t$. The optimal
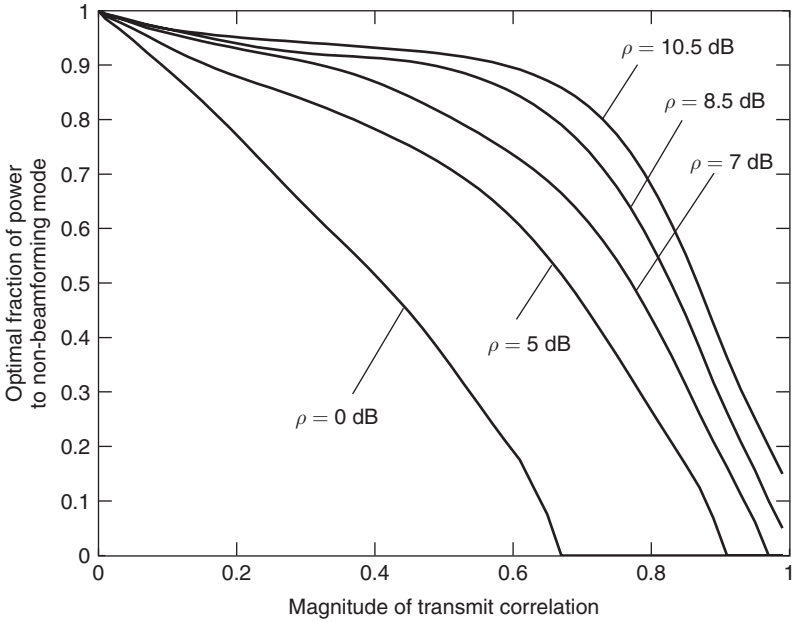
Figure 4.8 *Optimal fraction of power to non-beamforming mode as a function of |t| and ρ for $n_r = 2$*

eigenvalues forming $\mathbf{\Lambda_Q}$ have to be computed numerically. As with Proposition 4.4, a beamforming optimality condition is derived in [JB04]. Its analysis reveals that, in Rayleigh channels, the beamforming optimality range is larger with fully correlated receive antennas than with fully uncorrelated receive antennas. In other words, the larger the receive correlation, the larger the beamforming optimality range.

Finally, it is interesting to discuss whether CDIT may increase the ergodic capacity of Kronecker-structured correlated Rayleigh channels beyond the i.i.d. Rayleigh capacity [JB04, BGPvdV06].

- Receive correlations degrade both the mutual information $\bar{\mathcal{I}}_e$ and the capacity with CDIT.
- Transmit correlations always decrease $\bar{\mathcal{I}}_e$ but may increase $\bar{C}_{CDIT}$ at low SNR (irrespective of n_t and n_r) or at higher SNR when $n_t > n_r$.

4.4 Mutual information and capacity of Ricean channels with partial transmit channel knowledge

4.4.1 Mutual information with equal power allocation

The impact of a coherent component on the mutual information is often dealt with assuming that the non-coherent (Rayleigh) contribution is i.i.d. and that $\bar{\mathbf{H}}\bar{\mathbf{H}}^H$ has only one non-zero

eigenvalue equal to $n_t n_r = Nn$, e.g. $\bar{\mathbf{H}} = \mathbf{1}_{n_r \times n_t}$. If we assume that the K-factor is sufficiently high as to neglect the Rayleigh part, the ergodic mutual information scales as

$$\bar{\mathcal{I}}_e = \log_2(1 + n_r\rho), \tag{4.91}$$

which is exactly the capacity of an AWGN channel whose array gain g_a is equal to n_r. The multiplexing advantage totally vanishes as the rank of the channel matrix is one: the mutual information asymptotic growth scales as $\log_2(n_r)$ instead of $n = \min\{n_t, n_r\}$ for the i.i.d. Rayleigh fading case.

For arbitrary K-factors and assuming that the Rayleigh part is i.i.d., it is shown in [KA06b] that the exact ergodic mutual information with equal power allocation reads as

$$\bar{\mathcal{I}}_e = \frac{e^{-KNn}}{\log(2)\Gamma(N-n+1)(KNn)^{n-1}\prod_{k=1}^{n-1}\Gamma(N-k)\Gamma(n-k)}$$

$$\times \sum_{k=1}^{n} \det(\mathbf{\Psi}'_k), \tag{4.92}$$

where $\mathbf{\Psi}'_k$, $k = 1, \ldots, n$ are $n \times n$ matrices

$$\mathbf{\Psi}'_k(u,1) = \begin{cases} \int_0^\infty \log(1 + y\rho/n_t)y^{N-u}e^{-y}{}_0F_1(N-n+1, KNny)\,dy, & \text{if } k = 1 \\ \Gamma(N-u+1){}_1F_1(n-u+1, N-n+1, KNn), & \text{if } k \neq 1 \end{cases} \tag{4.93}$$

$$\mathbf{\Psi}'_k(u,v) = \Gamma(N+n-u-v+1)$$

$$\times \begin{cases} e^{n_t/\rho}\sum_{k=1}^{N+n-u-v+1}\frac{\Gamma(k-N-n+u+v-1,n_t/\rho)}{(\rho/n_t)^{N+n-u-v+1-k}}, & \text{if } k = v, v \neq 1 \\ 1, & \text{if } k \neq v, v \neq 1 \end{cases} \tag{4.94}$$

The above expressions assume that the Rayleigh contribution is uncorrelated. If this is not the case, we may use our mutual information upper bound of (4.78). In the case of 2×2 Ricean channels with a correlated Rayleigh contribution and $\bar{\mathbf{H}} = \mathbf{1}_{n_r \times n_t}$, the upper bound reads as

$$\log_2(\bar{\kappa}) = \log_2\left\{1 + 2\rho + \left(\frac{\rho}{2}\right)^2\frac{1}{(K+1)^2}\left[2K\left(2 + \Re[s_1 + s_2 - 2r - 2t]\right)\right.\right.$$

$$\left.\left. + 2 + |s_1|^2 + |s_2|^2 - |t|^2 - |r|^2\right]\right\}. \tag{4.95}$$

Figure 4.9 illustrates the impact of the Ricean K-factor K on the ergodic mutual information obtained by (4.92) (note that for $K > 1$, it can be shown that the upper bound is very tight, so we might have used (4.95) instead). For a given received SNR, the existence of a Ricean component clearly decreases the mutual information. However, if we consider a fixed
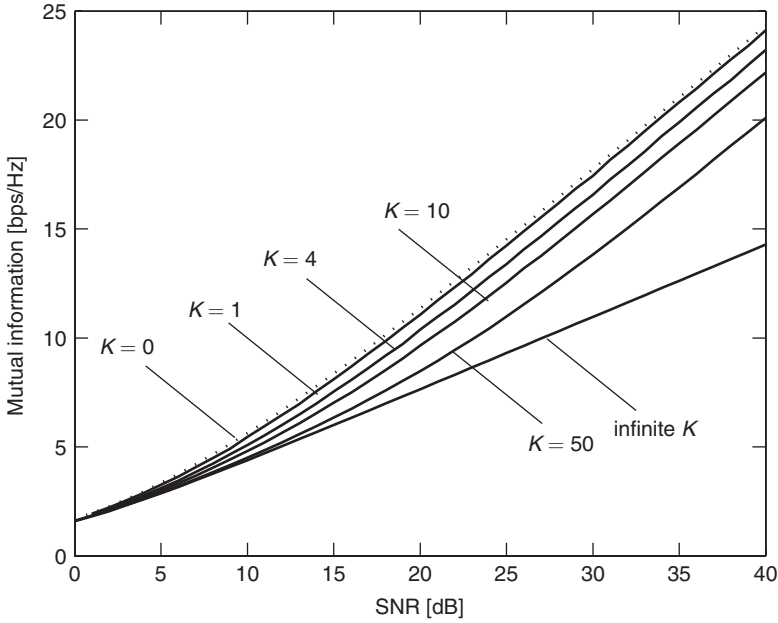
Figure 4.9 *Mutual information of Ricean 2×2 channels for different K-factors ($\tilde{\mathbf{H}} = \mathbf{H}_w$, so $K = 0$ corresponds to a Rayleigh i.i.d. channel)*

transmit power, it is most likely that the existence of a Ricean component (e.g. due to a line-of-sight) implies a higher SNR than Rayleigh fading scenarios, since Rayleigh fading is more often encountered in deeply shadowed conditions.

So far, we have considered that $\bar{\mathbf{H}} = \mathbf{1}_{n_r \times n_t}$. When $\bar{\mathbf{H}}$ has more than one non-zero singular value, the impact of the K-factor can be significantly different. In particular, if the various phases in (3.4) are such that $\bar{\mathbf{H}}$ is orthogonal (but how physical is this?), the mutual information will increase as the K-factor increases, as illustrated in [PNG03].

4.4.2 Ergodic capacity with partial transmit channel knowledge

As above, we first assume that the non-coherent (Rayleigh) contribution is i.i.d. and that $\bar{\mathbf{H}}\bar{\mathbf{H}}^H$ has only one non-zero eigenvalue equal to $n_t n_r = Nn$. In this case, it is shown in [JG04] that the optimal input covariance matrix has an eigenvalue decomposition $\mathbf{Q} = \mathbf{U}_{\mathbf{Q}} \mathbf{\Lambda}_{\mathbf{Q}} \mathbf{U}_{\mathbf{Q}}^H$, where the first column of $\mathbf{U}_{\mathbf{Q}}$ is the transposed conjugate of the first row of $\bar{\mathbf{H}}$ normalized to a unit norm (assuming without loss of generality that the first row of $\bar{\mathbf{H}}$ corresponds to the non-zero eigenvalue)

$$\mathbf{U}_{\mathbf{Q}}(:,1) = \frac{\bar{\mathbf{H}}(1,:)^H}{\|\bar{\mathbf{H}}(1,:)\|}. \tag{4.96}$$

The remaining columns of $\mathbf{U_Q}$ are arbitrarily chosen, except for the restriction that $\mathbf{U_Q}$ is unitary. The eigenvalues $\lambda_2(\mathbf{Q}) = \ldots = \lambda_{n_t}(\mathbf{Q})$ are all equal to $(1 - \lambda_1(\mathbf{Q}))/(n_t - 1)$, and $\lambda_1(\mathbf{Q})$ has to be determined numerically. Note that a closed-form expression of $\bar{C}_{CDIT}$ derived in [LTV05] at high SNR shows that the capacity decreases with increasing K at a fixed SNR (and still assuming that $\bar{\mathbf{H}}\bar{\mathbf{H}}^H$ has only one non-zero eigenvalue). As with Proposition 4.4, an optimality condition of beamforming (i.e. $\lambda_1(\mathbf{Q}) = 1$) is detailed in [JG04], yielding the expected result that beamforming is optimal when $K \to \infty$.

When the Rayleigh contribution is not i.i.d. or when the rank of $\bar{\mathbf{H}}$ is larger than one, there is no analytical solution for $\mathbf{U_Q}$. It is shown in [VP05] that a good approximation of the optimal input covariance can be obtained through the application of Jensen's inequality at high SNR, resulting in the following result:

- $\mathbf{U_Q}$ is given by the eigenvector matrix of $\bar{\mathbf{W}} = \mathcal{E}\{\mathbf{H}_H \mathbf{H}\}$,
- $\lambda_1(\mathbf{Q}) = \ldots = \lambda_{n_t}(\mathbf{Q})$ being given by the standard water-filling solution applied on the eigenvalues of $\bar{\mathbf{W}}$.

The power allocation is related to a large number of parameters: the rank of $\bar{\mathbf{H}}$, the SNR, the K-factor, the correlations, etc. It is therefore difficult to draw general conclusions. However, for high K-factors, the Jensen-resulting power allocation is the optimal allocation, as $\mathbf{H} \cong \bar{\mathbf{H}}$ is deterministic. It is then easy to realize that beamforming on a single mode of $\bar{\mathbf{H}}$ is not optimal (unless $\bar{\mathbf{H}}$ has only one singular value, as seen above). It should also be noticed that when $\bar{\mathbf{H}}$ is full rank, the capacity increases as K increases.

4.5   Mutual information in some particular channels

4.5.1   *Dual-polarized channels*

For very high K-factors, the channel matrix only depends on the Ricean component. Assuming that the latter is dominated by the line-of-sight contribution, the dual-polarized channel matrix reads as

$$\mathbf{H}_{\times,a} \approx \bar{\mathbf{H}}_{\times,a} = \begin{bmatrix} 1 & \sqrt{\chi_a} \\ \sqrt{\chi_a} & 1 \end{bmatrix}, \tag{4.97}$$

where $0 \leq \chi_a \leq 1$ reflects the level of antenna XPD (for good antennas, $\chi_a \to 0$). Let us consider two particular cases, i.e. $\chi_a = 0$ and $\chi_a = 1$. The mutual information is given by

$$\bar{\mathcal{I}}_e(\chi_a = 0) = 2\log_2\left(1 + \frac{\rho}{2}\right) \tag{4.98}$$

$$\bar{\mathcal{I}}_e(\chi_a = 1) = \log_2(1 + 2\rho) \tag{4.99}$$

Both cases are compared in Figure 4.10. It is clear that a good antenna XPD restores the multiplexing gain at high SNR levels.
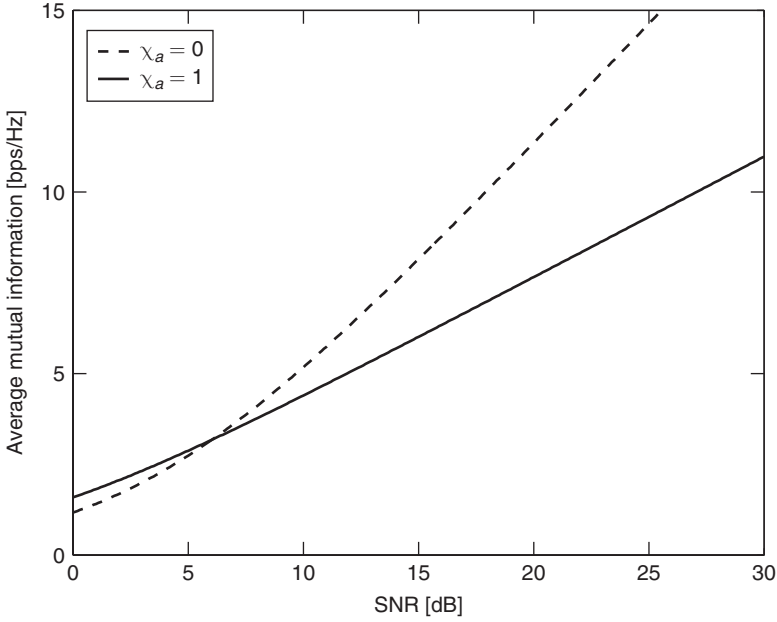
Figure 4.10   *Mutual information of Ricean 2×2 uni- and dual-polarized channels*

Let us now consider that the K-factor is finite, and use the model of (3.12) to highlight the impact of scattering XPD and gain imbalance on the mutual information upper bound. For convenience, we assume analogous to Chapter 3 that μ and χ are non-random, and also take $\chi_a = 0$ ($\mathbf{H}_{\times,a} = \mathbf{H}_\times$). We investigate two scenarios:

- spatially correlated dual-polarized antennas and $\delta_1 = \delta_2 = 1$

$$\bar{\mathbf{H}}_\times = \mathbf{I}_2$$
$$\tilde{\mathbf{H}}_\times = \begin{bmatrix} 1 & \sqrt{\mu\chi}\, e^{j\phi} \\ \sqrt{\chi}\, e^{j\phi} & \sqrt{\mu} \end{bmatrix} \odot \tilde{\mathbf{H}} \tag{4.100}$$

  where $\tilde{\mathbf{H}}$ is given by (3.3)

- well separated (spatially uncorrelated) dual-polarized antennas

$$\bar{\mathbf{H}}_\times = \mathbf{I}_2$$
$$\tilde{\mathbf{H}}_\times = \begin{bmatrix} 1 & \sqrt{\mu\chi}e^{j\phi} \\ \sqrt{\chi}e^{j\phi} & \sqrt{\mu} \end{bmatrix} \odot \mathbf{H}_w \tag{4.101}$$

  which is valid irrespective of δ_1 and δ_2.

Two alternative scenarios (i.e. $\delta_1 = \delta_2 = 0$ with correlated or well separated antennas) do not need to be considered, as they are indeed both covered by the second scenario (all three scenarios cause elements of $\tilde{\mathbf{H}}_\times$ to become fully uncorrelated). For both analyzed scenarios, the mutual information upper bound is written as detailed in [Oes06]

- for $\delta_1 = \delta_2 = 1$

$$\log_2(\bar{\kappa}) = \log_2 \left\{ 1 + \frac{\rho}{2} \left[\frac{2K + (1 + \mu)(1 + \chi)}{K + 1} \right] \right.$$
$$+ \left(\frac{\rho}{2} \right)^2 \frac{1}{(K + 1)^2} \left[K \left(K + 2\Re[s_1]\sqrt{\mu} + \mu + 1 \right) \right.$$
$$\left. \left. + \mu \left(1 + |s_1|^2 + \chi^2 \left(1 + |s_2|^2 \right) \right) \right] \right\}, \tag{4.102}$$

- for well separated dual-polarized antennas with arbitrary values of δ_1 and δ_2 or for $\delta_1 = \delta_2 = 0$

$$\log_2(\bar{\kappa}) = \log_2 \left\{ 1 + \frac{\rho}{2} \left[\frac{2K + (1 + \mu)(1 + \chi)}{K + 1} \right] \right.$$
$$\left. + \left(\frac{\rho}{2} \right)^2 \frac{1}{(K + 1)^2} \left[K(K + \mu + 1) + \mu \left(1 + \chi^2 \right) \right] \right\}. \tag{4.103}$$

We note that if $K \to \infty$, both (4.102) and (4.103) approach the result of (4.98). Then, we also note that the mutual information is higher in the first scenario (i.e. $\delta_1 = \delta_2 = 1$). Indeed, in the latter, the channel matrix behaves similarly as a diagonal channel (for $s_1 = s_2 = \mu = \chi = 1$ and $K = 0$, $\mathbf{H}_\times = \tilde{\mathbf{H}}_\times$ is exactly a 2×2 diagonal channel). Note, however, that the difference between achieved mutual informations in both scenarios remains limited.

Finally, the ergodic mutual information of Rayleigh channels increases in both (4.102) and (4.103) as μ and χ increase, as shown in Figure 4.11 for χ (for $\mu = 0.5$ and close antennas with $\delta_1 = \delta_2 = 1$). We also note that dual-polarized transmissions only offer a larger mutual information for Ricean or highly correlated Rayleigh fading channels.

4.5.2   Impact of antenna coupling on mutual information

To understand the dual impact of mutual coupling on $\bar{\mathcal{I}}_e$, let us investigate how it affects the mutual information upper bound of (4.78). When coupling is accounted for, the latter reads as

$$\log_2 \bar{\kappa}_c = \log_2 \left\{ 1 + 2\rho \left[|a|^2 + |b|^2 + 2\Re \left[ab^* \right] \Re[t] \right] \right.$$
$$+ \left(\frac{\rho}{2} \right)^2 \left[|a|^4 + |b|^4 - 2\Re \left[a^2 b^{*2} \right] \right]$$
$$\left. \left[1 + |s_1|^2 + 1 + |s_2|^2 - 2|r|^2 - 2|t|^2 \right] \right\} \tag{4.104}$$

The above expression implicitly assumes that the transmitted power is kept constant, irrespective of antenna coupling. It may be observed that the actual mutual information shift $\Delta \bar{\mathcal{I}}_e = \bar{\mathcal{I}}_{e,c} - \bar{\mathcal{I}}_e$ is well approximated by $\log_2 (\bar{\kappa}_c/\bar{\kappa})$, so that the following discussion is based on $\log_2 (\bar{\kappa}_c)$ and $\log_2 (\bar{\kappa}_c/\bar{\kappa})$.
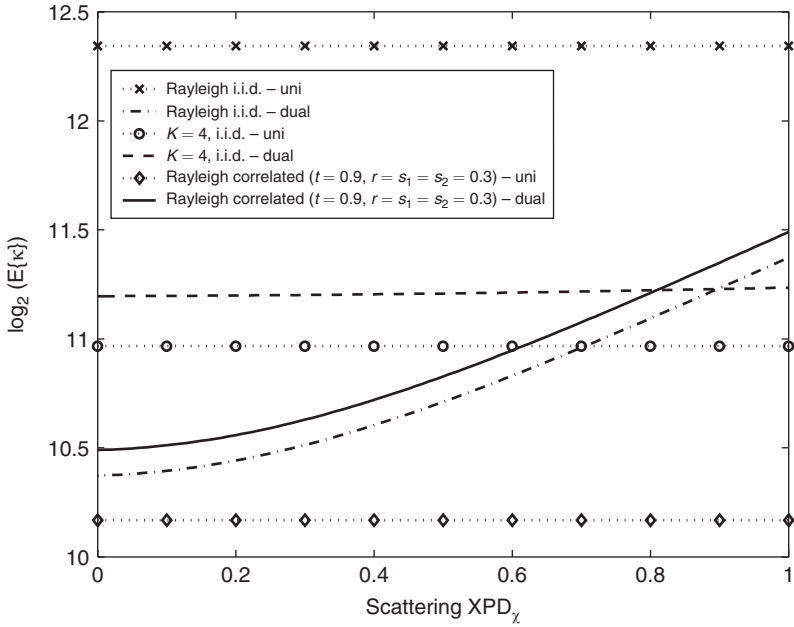
Figure 4.11 *Mutual information of uni- and dual-polarized 2×2 channels for different K-factors and correlations*

At low SNR, the term in (4.104) proportional to ρ dominates, so that the impact of coupling can be analyzed through the Frobenius norm. The ratio ς between the average channel Frobenius norm with coupling and the average channel Frobenius norm without coupling is easily expressed as

$$\varsigma = \frac{\mathcal{E}\left\{\|\mathbf{H}_c\|_F^2\right\}}{\mathcal{E}\left\{\|\mathbf{H}\|_F^2\right\}} = |a|^2 + |b|^2 + 2\Re\left[ab^*\right]\Re[t]. \tag{4.105}$$

Since ς depends on the transmit correlation coefficient t, it is expected that the effect of mutual coupling on the channel energy will only be related to the antenna spacing and orientation and to the transmit angle-spread. In order to analyze the behavior of ς, Figure 4.12(a) depicts the variation of $|a|^2 + |b|^2$ and $2\Re[ab^*]$ as a function of the transmit antenna spacing.

In omnidirectional scattering scenarios, we have seen in Chapter 2 that t is real-valued and equal to $J_0(2\pi d_t/\lambda)$. That causes $2\Re[ab^*]\Re[t]$ to be negative and larger (in magnitude) than $|a|^2 + |b|^2 - 1$, so that $\varsigma \le 1$ for all antenna spacings.

To analyze the directional scattering case, let us assume an extreme situation, where $\mathcal{A}_t(\Theta_t) \to \delta[\Theta_t - \Theta_{t,0}]$. In such cases, $t \to \exp\left(-j2\pi d_t \cos\left(\psi_t - \Theta_{t,0}\right)\right)$, hence $\Re[t] \to \cos\left(2\pi d_t/\lambda \cos\left(\psi_t - \Theta_{t,0}\right)\right)$. Inserting this expression into (4.105) leads to the conclusion that ς can be ≥ 1 for spacings d_t between 0.4 and 0.9λ when $|\psi_t - \Theta_{t,0}| \approx \pi/2$.
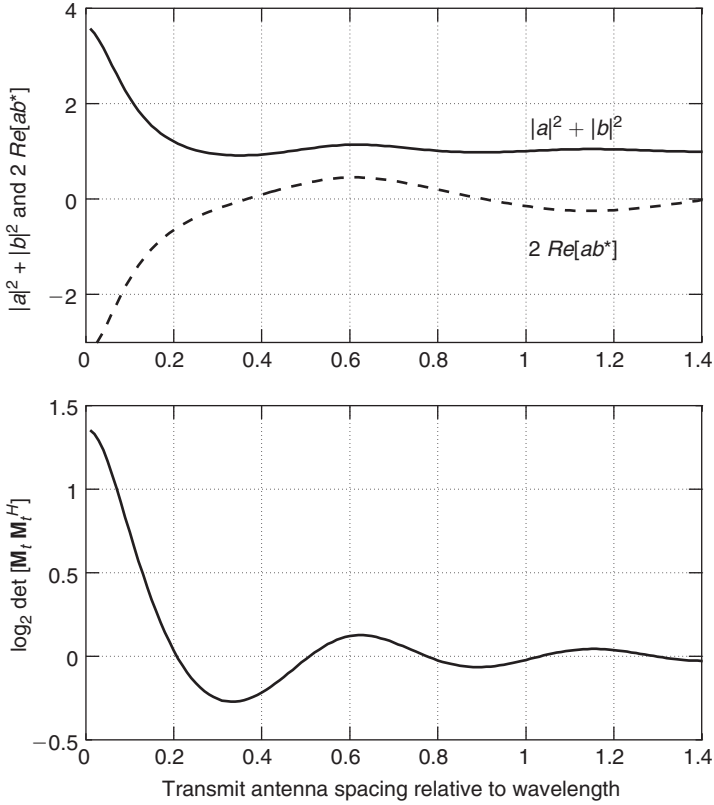
Figure 4.12   *Variation of $|a|^2 + |b|^2$, $2\Re[ab^*]$ and $\log_2 \det [\mathbf{M}_t \mathbf{M}_t^H]$ as a function of d_t*

The behavior of ς is confirmed by analyzing the orientation of the main lobes of the transmit coupled antennas with respect to the distribution of scatterers. If the coupled antenna patterns (such as those of Figure 2.10) present a minimum (resp. maximum) in the direction of the main lobes of $\mathcal{A}_t(\Theta_t)$, ς will be low (resp. high).

Therefore, mutual coupling at the transmitter has a positive impact on the channel energy (and so the low-SNR mutual information) only when three conditions are jointly met:

- for transmit antenna spacings between 0.4 and 0.9λ
- in directional scattering conditions at the transmitter
- with the transmit antenna array being oriented orthogonally to the maximum of the angular power spectrum.

When the SNR is sufficiently high so that the last term in (4.104) dominates, the ergodic mutual information (both the upper bound and the true value) are shifted by

$$\log_2 \left(\det \left(\mathbf{M}_t \mathbf{M}_t^H \right) \right) = \log_2 \left(|a|^4 + |b|^4 - 2\Re\left[a^2 b^{*2}\right] \right) \tag{4.106}$$

when coupling is accounted for. This shift is independent of the channel correlations without coupling. It is shown in Figure 4.12(b) that $\log_2\left(\det\left(\mathbf{M}_t\mathbf{M}_t^H\right)\right)$ is positive for antenna spacings less than 0.2λ and between 0.5 and 0.8λ, while it is negative for spacings between 0.2 and 0.5λ and between 0.8 and λ. Note that the magnitude of the resulting high-SNR shift, as directly read on the graph, decreases as the spacing increases, and is equal to zero for spacings larger than one wavelength.

4.6 Outage capacity and diversity-multiplexing trade-off in i.i.d. Rayleigh slow fading channels

In the previous sections, we have dealt with the ergodic capacity to analyze the transmission capability of MIMO fast fading channels. In slow fading channels, the achievable rate cannot be described by means of a single (ergodic) quantity, because it might not be possible to fully average out the randomness of the channel. We thus resort to the concept of outage probability P_{out}, which is defined as the probability that a given channel realization cannot support a given rate R

$$P_{out}(R) = \inf_{\mathbf{Q}\geq 0:\mathrm{Tr}\{\mathbf{Q}\}\leq 1} P(\log_2 \det(\mathbf{I}_{n_r} + \rho\mathbf{HQH}^H) < R). \tag{4.107}$$

For a given transmission rate R, we are interested to qualify how P_{out} behaves as a function of the SNR ρ. To this end, we formalize the important concepts of spatial multiplexing and diversity gains introduced in Chapter 1.

Definition 4.3 *The spatial multiplexing gain is the ratio of the transmission rate $R(\rho)$ to the capacity of an AWGN channel with array gain g_a*

$$g_s \triangleq \frac{R(\rho)}{\log_2 (1 + g_a\rho)}. \tag{4.108}$$

The diversity gain is the negative slope of the log-log plot of the outage probability versus SNR

$$g_d^\star(g_s, \rho) \triangleq -\frac{\partial \log (P_{out}(R))}{\partial \log (\rho)}. \tag{4.109}$$

The curve $g_d^\star(g_s, \rho)$ as function of g_s and ρ is known as the diversity-multiplexing trade-off of the channel for a SNR equal to ρ.

Note that the array gain g_a is usually chosen such that $g_a = \frac{1}{n_t}\mathcal{E}\{\|\mathbf{H}\|_F^2\} = n_r$ (see Section 4.4.1).

Intuitively, the multiplexing gain indicates how fast the transmission rate increases with the SNR, whereas the diversity gain represents how fast the outage probability decays with the SNR. Clearly, the multiplexing gain cannot exceed $n = \min\{n_t, n_r\}$, while the diversity gain cannot exceed $n_t n_r$ in i.i.d. Rayleigh channels. If we let the transmission rate

R increase too quickly with ρ, it is easy to imagine that the outage probability might not decay very rapidly and vice-versa. As a consequence, the diversity-multiplexing trade-off provides how much diversity can be traded for multiplexing, and inversely, at any given SNR ρ.

We further split this study into two sections. The first section is the original diversity-multiplexing trade-off derived in [ZT03] when the SNR approaches infinity. The second section constitutes a generalization of these preliminary results at any finite SNR [Nar05].

4.6.1   Infinite SNR

In the case of infinite SNR, $g_d^\star$ and g_s in Definition 4.3 reduce to these more intuitive definitions.

Definition 4.4   *A diversity gain $g_d^\star(g_s,\infty)$ is achieved at multiplexing gain g_s if*

$$\lim_{\rho \longrightarrow \infty} \frac{R(\rho)}{\log_2(\rho)} = g_s \tag{4.110}$$

$$\lim_{\rho \longrightarrow \infty} \frac{\log_2(P_{out}(R))}{\log_2(\rho)} = -g_d^\star(g_s,\infty). \tag{4.111}$$

The curve $g_d^\star(g_s,\infty)$ as function of g_s is known as the asymptotic diversity-multiplexing trade-off of the channel.

The diversity-multiplexing trade-off [ZT03] has been solved for i.i.d. slow Rayleigh fading channels. Note that to simplify notations, it is common to use the symbol $\doteq$ to denote the exponential equality, i.e. $f(\rho) \doteq \rho^b$ actually means

$$\lim_{\rho \to \infty} \frac{\log_2(f(\rho))}{\log_2(\rho)} = b \tag{4.112}$$

suggesting that function $f(\rho)$ is always of the form

$$f(\rho) = \delta\rho^b + g\left(\rho^\beta\right) \tag{4.113}$$

with $\beta < b$ and $\delta > 0$. Symbols $\dot{\geq}$ and $\dot{\leq}$ are similarly defined. Based on those notations, we can equivalently write (4.110) and (4.111) as

$$2^R \doteq \rho^{g_s} \tag{4.114}$$

$$P_{out}(R) \doteq \rho^{-g_d^\star(g_s,\infty)} \tag{4.115}$$

Proposition 4.5   *The asymptotic diversity-multiplexing trade-off $g_d^\star(g_s,\infty)$ of the i.i.d. Rayleigh fading channel is the piecewise-linear function joining the points $(g_s, g_d^\star(g_s,\infty))$ with $g_s = 0,\ldots,$ min $\{n_t,n_r\}$ and $g_d^\star(g_s,\infty) = (n_t - g_s)(n_r - g_s)$.*
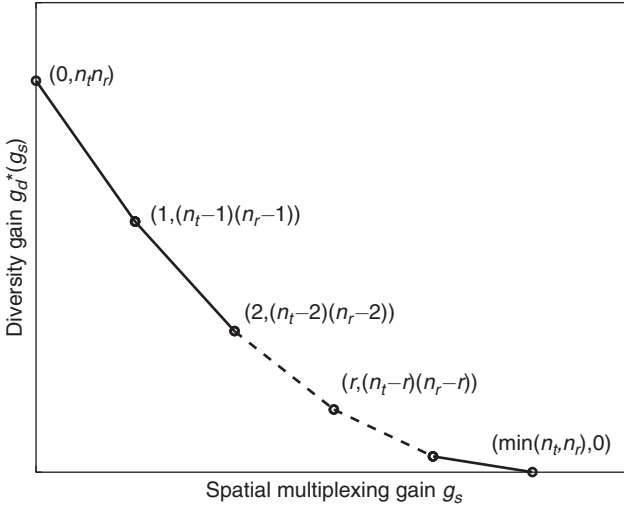
Figure 4.13 *Asymptotic diversity-multiplexing trade-off $g_d^\star(g_s)$ in i.i.d. Rayleigh fading channels*

This trade-off is illustrated in Figure 4.13. Both extreme points of the curve have a particularly interesting signification.

- The point $(0, n_t n_r)$ means that for a spatial multiplexing gain of zero (i.e. when the transmission rate is kept fixed), the maximal diversity gain achievable is the diversity order of the channel $n_t n_r$.
- The point $(\min\{n_t, n_r\}, 0)$ reveals that transmitting at diversity gain $g_d^\star = 0$ (i.e. when the outage probability is kept fixed) allows the data rate to increase with SNR as $n = \min\{n_t, n_r\}$. This corresponds to situations where a rate close to ergodic capacity is transmitted over the channel, thereby without protection against fading (the diversity gain becomes null).

Intermediate points on the curve suggest that it is possible to transmit at non-zero diversity and multiplexing gains but that any increase of one of those quantities leads to a decrease of the other quantity.

In Figure 4.14, for 2×2 i.i.d. Rayleigh fading channels, the outage probability is displayed versus SNR in light gray lines for fixed rates $R = 1, 2, \ldots, 30$ bits/s/Hz. The asymptotic slope of each curve is four and matches the maximum diversity gain $g_d^\star(0, \infty)$. The horizontal separation is 2 bits/s/Hz per 3 dB, which corresponds to the maximum multiplexing gain equal to $n(=2)$. The bold lines give the outage probability when the rate scales with ρ: as the rate increases more rapidly with SNR (i.e. as the multiplexing gain g_s increases), the slope of the outage probability curve (given by the diversity gain $g_d^\star$) vanishes. This is the diversity-multiplexing trade-off. The dashed lines represent the asymptotic slope of the outage probability at high SNR.
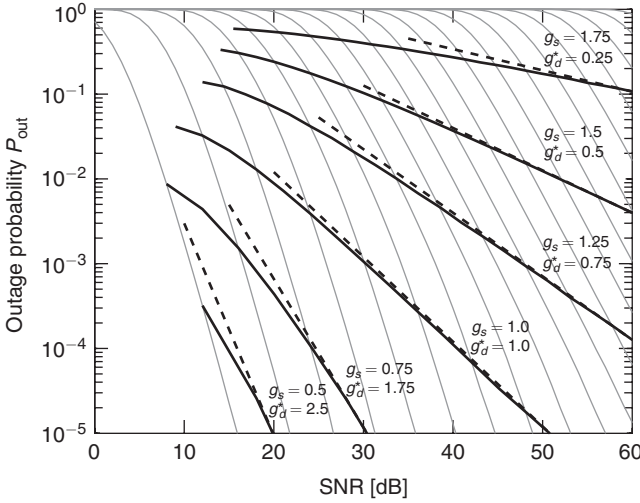
Figure 4.14   *Outage probability $P_{out}(R)$ as a function of the transmission rate R for both fixed and variable rate scaling as $R = g_s \log_2 (\rho)$ in 2×2 MIMO i.i.d. Rayleigh fading channels (courtesy of H. Yao [YW03])*

The rigorous proof of the asymptotic diversity-multiplexing trade-off curve for i.i.d. Rayleigh fading channels is rather lengthy and complex, and falls beyond the scope of this book. Rather, we provide a simple example to get some insight. Afterwards, we sketch the main ideas of the proof and give some more examples.

Example 4.1   *We first derive the diversity-multiplexing trade-off of a scalar Rayleigh channel h [ZT03]. Finding the diversity-multiplexing trade-off comes to determine, for a transmission rate R scaling with ρ as $g_s \log_2 (\rho)$, the rate at which the outage probability decreases with ρ as ρ increases. The outage probability of this channel is written as*

$$P_{out}(R) = P\Big(\log_2 \left[1 + \rho|h|^2\right] < g_s \log_2 (\rho)\Big)$$
$$= P\Big(1 + \rho|h|^2 < \rho^{g_s}\Big) \tag{4.116}$$

such that at high SNR, P_{out} is of the order of

$$P_{out}(R) \approx P\Big(|h|^2 \leq \rho^{-(1-g_s)}\Big)$$
$$\doteq \rho^{-(1-g_s)}. \tag{4.117}$$

The last expression comes from the fact that $|h|^2$ is exponentially distributed, i.e. $P(|h|^2 \leq \epsilon) \approx \epsilon$ for small ϵ (see Appendix B). It tells us that an outage occurs at high SNR when $|h|^2 \leq \rho^{-(1-g_s)}$ with a probability $\rho^{-(1-g_s)}$. The diversity-multiplexing trade-off for the scalar Rayleigh fading channel is thus given by $g_d^\star(g_s, \infty) = 1 - g_s$ for $g_s \in [0, 1]$. This corresponds exactly to the results of Proposition 4.5.

Let us now examine the proof of Proposition 4.5 as derived in [ZT03]. For the full and rigorous proof, the interested reader is referred to the original paper [ZT03].

PROOF: Finding $g_d^\star(g_s, \infty)$ comes to find, for a transmission rate $g_s \log_2(\rho)$, the rate at which the outage probability decreases as the SNR increases. While it has not been proved that the optimal covariance matrix $\mathbf{Q}$ maximizing the outage probability of an i.i.d. Rayleigh fading channel is the identity matrix, it has been conjectured so. Hence, the outage probability obtained with equal power allocation is often used as a good upper bound on the actual outage probability at high SNR. The asymptotic diversity-multiplexing trade-off of the channel is obtained by analyzing the following expression

$$P_{out}(R) = P\left(\log_2 \det\left[\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H\right] < g_s \log_2(\rho)\right) \tag{4.118}$$

and by deriving the exponent of the SNR term as $\rho \to \infty$. The behavior of the singular values of $\mathbf{H}$ when an outage occurs, in terms of transmission rate and SNR, needs to be quantified. Let us arrange in increasing order the non-zero eigenvalues $\lambda_i(\mathbf{H}\mathbf{H}^H)$ of $\mathbf{H}\mathbf{H}^H$. In Example 4.1, we have observed that an outage occurs if the amplitude of the scalar channel is too small w.r.t. to the SNR. This has been expressed as $|h|^2 \leq \rho^{-(1-g_s)}$. In a MIMO channel, we have to consider the eigenvalues of the channel matrix. Therefore, let us denote $\lambda_i(\mathbf{H}\mathbf{H}^H) = \rho^{-\alpha_i}$. The outage probability is then rewritten as

$$P_{out}(R) \doteq P\left(\prod_{i=1}^{n}\left(1 + \rho\lambda_i(\mathbf{H}\mathbf{H}^H)\right) < \rho^{g_s}\right)$$

$$\doteq P\left(\prod_{i=1}^{n} \rho^{(1-\alpha_i)^+} < \rho^{g_s}\right)$$

$$= P\left(\sum_{i=1}^{n}(1-\alpha_i)^+ < g_s\right) \tag{4.119}$$

where $(x)^+$ denotes $\max\{0, x\}$. An outage occurs when $\mathbf{H}$ becomes close to singular, i.e. when the singular values of $\mathbf{H}$ are close to zero. Values α_i are also an information about the singularity of $\mathbf{H}$, since the larger these values, the more singular $\mathbf{H}$. Defining the outage event $\mathcal{V} = \left\{\alpha = [\alpha_1, \ldots, \alpha_n] \,\middle|\, \sum_{i=1}^{n}(1-\alpha_i)^+ < g_s\right\}$, the outage probability comes to find the probability that $\alpha \in \mathcal{V}$, i.e. $P_{out}(R) \doteq \int_{\mathcal{V}} p(\alpha)d\alpha$. Evaluating the joint probability density function $p(\alpha)$ and considering only the dominant terms as $\rho \to \infty$ with $R = g_s \log_2(\rho)$, $0 \leq g_s \leq n$, the outage probability satisfies

$$P_{out}(g_s \log_2(\rho)) \doteq \rho^{-g_{d,out}(g_s)} \tag{4.120}$$

where

$$g_{d,out}(g_s) = \inf_{\alpha \in \mathcal{V}'} \sum_{i=1}^{n} 2i - 1 + |n_t - n_r|\alpha_i \tag{4.121}$$

and

$$\mathcal{V}' = \left\{ \alpha \in \mathcal{R}^{n+} | \alpha_1 \geq \ldots \geq \alpha_n \geq 0, \sum_{i=1}^{n} (1 - \alpha_i)^+ < g_s \right\} \tag{4.122}$$

with $\mathcal{R}^{n+}$ the set of real n-vectors with non-negative elements.

The resulting $g_{d,out}(g_s)$ can be computed and matches $g_d^\star(g_s, \infty)$ in Proposition 4.5 for all g_s. ■

To better understand the concept of diversity-multiplexing trade-off, let us consider a few more examples.

Example 4.2 *Assume that a transmission scheme transforms the original i.i.d. MIMO channel* **H** *into an equivalent SISO channel* $\|\mathbf{H}\|_F^2$ *by a simple matched filtering at the receiver (e.g. O-STBC). For a total rate R that scales with ρ such as $R = g_s \log_2(\rho)$, the outage probability of this equivalent channel is written as*

$$P_{out}(R) = P\left(\log_2 \left[1 + \frac{\rho}{n_t} \|\mathbf{H}\|_F^2 \right] < g_s \log_2(\rho) \right)$$
$$= P\left(1 + \frac{\rho}{n_t} \|\mathbf{H}\|_F^2 < \rho^{g_s} \right) \tag{4.123}$$

such that at high SNR, P_{out} is of the order of

$$P_{out}(R) \doteq P\left(\|\mathbf{H}\|_F^2 \leq \rho^{-(1-g_s)} \right)$$
$$\doteq \rho^{-n_t n_r (1-g_s)}. \tag{4.124}$$

The last expression comes from the fact that $\|\mathbf{H}\|_F^2$ is $\chi_{2n_t n_r}^2$ distributed, i.e. $P(\|\mathbf{H}\|_F^2 \leq \epsilon) \approx \epsilon^{n_t n_r}$ for small ϵ (see Appendix B). Hence, the diversity-multiplexing trade-off of this equivalent SISO channel is given by $g_d^\star(g_s, \infty) = n_t n_r (1 - g_s)$, for $g_s \in [0, 1]$. This equivalent channel has the same diversity-multiplexing trade-off as SIMO/MISO channel with $n_t n_r$ receive/transmit antennas, respectively. This suggests that transforming the original MIMO channel into an equivalent SISO channel is not efficient from a diversity-multiplexing trade-off perspective.

Example 4.3 *Let us now consider the particular case of a parallel slow fading channel with i.i.d. Rayleigh entries. The transmission is thus performed over L channels in parallel*

and there is no interference between those channels. The transmission at the kth time instant on the lth sub-channel $h^{(l)}$ reads as

$$r_k^{(l)} = \sqrt{E_s} h^{(l)} c_k^{(l)} + n_k^{(l)}. \tag{4.125}$$

Deriving the optimal diversity-multiplexing trade-off of this channel is quite straightforward if we assume a uniform input distribution, i.e. $\mathbf{Q} = \mathbf{I}_L/L$. For a total transmission rate R, an outage occurs when

$$\sum_{l=1}^{L} \log_2 \left(1 + \frac{\rho}{L} |h^{(l)}|^2\right) \leq R. \tag{4.126}$$

The outage probability is thus given by

$$\begin{aligned}
P_{out}(R) &= P \left(\sum_{l=1}^{L} \log_2 \left[1 + \frac{\rho}{L} |h^{(l)}|^2 \right] \leq g_s \log_2 (\rho) \right) \\
&\stackrel{(a)}{\doteq} \left[P \left(\frac{\rho}{L} |h^{(l)}|^2 \leq \rho^{\frac{g_s}{L}} \right) \right]^L \\
&\doteq \rho^{-L \left(1 - \frac{g_s}{L}\right)}.
\end{aligned} \tag{4.127}$$

where in (a) we use the fact that the dominant outage event occurs when all the sub-channels are in outage [TV05]. The optimal diversity-trade-off of the parallel fading channel reads therefore as

$$g_d^{\star}(g_s, \infty) = L \left(1 - \frac{g_s}{L}\right), \quad g_s = [0, L]. \tag{4.128}$$

Example 4.4 *We finally derive the diversity-multiplexing trade-off for 2×2 i.i.d. Rayleigh fading channels, using the approach proposed in [YW03]. The mutual information with equal power allocation of 2×2 MIMO channels is expressed as*

$$\begin{aligned}
\mathcal{I}_e(\mathbf{H}) &= \log_2 \det \left(\mathbf{I}_2 + \frac{\rho}{2} \mathbf{H} \mathbf{H}^H \right) \\
&= \log_2 \det \left(1 + \frac{\rho}{2} \|\mathbf{H}\|_F^2 + \left(\frac{\rho}{2}\right)^2 |\det(\mathbf{H})|^2 \right) \\
&= \log_2 \left(1 + \frac{\rho}{2} \left(r_{11}^2 + |r_{12}|^2 + r_{22}^2 \right) + \left(\frac{\rho}{2}\right)^2 r_{11}^2 r_{22}^2 \right)
\end{aligned} \tag{4.129}$$

where in (4.129) we use a QR decomposition of $\mathbf{H} = \mathbf{QR}$ with

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{bmatrix}.$$

Since $r_{11}^2 = \|\mathbf{H}(:,1)\|^2$, r_{11}^2 is approximately χ_4^2 distributed. Similarly, $|r_{12}|^2$ and r_{22}^2 are χ_2^2 variables. From (4.118), we get

$$P_{out} \doteq P\left(1 + \frac{\rho}{2}\left(r_{11}^2 + |r_{12}|^2 + r_{22}^2\right) + \left(\frac{\rho}{2}\right)^2 r_{11}^2 r_{22}^2 < \rho^{g_s}\right). \tag{4.130}$$

For both ranges $1 \le g_s \le 2$ and $0 \le g_s \le 1$, let us focus only on dominant terms at high SNR (as in the proof of Proposition 4.5).

- $1 \le g_s \le 2$: keeping only the second order term yields

$$P_{out} \doteq P\left(\left(\frac{\rho}{2}\right)^2 r_{11}^2 r_{22}^2 < \rho^{g_s}\right). \tag{4.131}$$

Since r_{11}^2 is a higher order random variable than r_{22}^2, low values of $r_{11}^2 r_{22}^2$ are mostly caused by small r_{22}^2, suggesting that the dominant event of $\left(\frac{\rho}{2}\right)^2 r_{11}^2 r_{22}^2 < \rho^{g_s}$ is $(r_{11}^2 < 1) \cup (r_{22}^2 < \rho^{g_s-2})$

$$P_{out} \doteq P\left(\left(r_{11}^2 < 1\right) \cup \left(r_{22}^2 < \rho^{g_s-2}\right)\right)$$
$$\doteq \rho^{g_s-2}. \tag{4.132}$$

So $g_d^\star(g_s, \infty) = g_s - 2$ for $1 \le g_s \le 2$.

- $0 \le g_s \le 1$: keeping the first and second order terms, we now get

$$P_{out} \doteq P\left(\frac{\rho}{2}\left(r_{11}^2 + |r_{12}|^2 + r_{22}^2\right) + \left(\frac{\rho}{2}\right)^2 r_{11}^2 r_{22}^2 < \rho^{g_s}\right). \tag{4.133}$$

The dominant event of $\frac{\rho}{2}\left(r_{11}^2 + |r_{12}|^2 + r_{22}^2\right) + \left(\frac{\rho}{2}\right)^2 r_{11}^2 r_{22}^2 < \rho_s^g$ is now given by $\left(r_{11}^2 < \rho^{g_s-1}\right) \cup \left(|r_{12}|^2 < \rho^{g_s-1}\right) \cup \left(r_{22}^2 < \rho^{-1}\right)$

$$P_{out} \doteq P\left(\left(r_{11}^2 < \rho^{g_s-1}\right) \cup \left(|r_{12}|^2 < \rho^{g_s-1}\right) \cup \left(r_{22}^2 < \rho^{-1}\right)\right)$$
$$\doteq \rho^{2(g_s-1)}\rho^{g_s-1}\rho^{-1}$$
$$\doteq \rho^{3r-4}. \tag{4.134}$$

So $g_d^\star(g_s, \infty) = 3r - 4$ for $0 \le g_s \le 1$.

This matches the diversity-multiplexing trade-off predicted by Proposition 4.5.

4.6.2   Finite SNR

We generalize the above results to realistic values of SNR. In order to obtain the diversity-multiplexing trade-off, the derivation of the outage probability is first required. Analogous

to the previous section, we assume a scaled identity covariance matrix $\mathbf{Q}$ so that the outage probability is written as

$$P_{out}(R) = P\left(\log_2 \det\left[\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{HH}^H\right] < g_s \log_2\left(1 + g_a\rho\right)\right).$$ (4.135)

Mathematically averaging over the fading distribution is a rather complex problem. Therefore, lower bounds on outage probability have been proposed to keep the problem tractable [Nar05], although proofs of those results are not provided here. If we assume for simplicity that $n_r \geq n_t$, the following result is obtained in [Nar05].

Proposition 4.6 *In i.i.d. Rayleigh fading channels, a lower bound on P_{out} is given by*

$$P_{out}(g_s, \rho) > \prod_{l=1}^{n_t} \frac{\Gamma(n_r - n_t + 2l - 1, \alpha_l)}{\Gamma(n_r - n_t + 2l - 1)}$$ (4.136)

where $g_s = \sum_{l=1}^{n_t} a_l$, $a_0 = 0$, $\alpha_l = \frac{n_t}{\rho}((1 + g_a\rho)^{a_l} - (1 + g_a\rho)^{a_{l-1}})$.

To improve the accuracy of the outage probability lower bound, parameters $(a_1, \ldots, a_{n_t})$ are optimized for each value of g_s and ρ. Using the conditions $\alpha_l \geq 0$, $a_0 = 0$ and $g_s = \sum_{l=1}^{n_t} a_l$, the feasible set $\mathcal{V}_r$ for $(a_1, \ldots, a_{n_t})$ is given by

$$\mathcal{V}_r = \left\{(a_1, \ldots, a_{n_t}) | a_{l-1} < a_l < \frac{g_s - \sum_{k=1}^{l-1} a_k}{n_t - l + 1}, \; l = 1, \ldots, n_t - 1; \right.$$

$$\left. a_{n_t} = g_s - \sum_{k=1}^{n_t - 1} a_k \right\}.$$ (4.137)

Non-linear programming techniques are then used to optimize the set of parameters $(a_1, \ldots, a_{n_t}) \in \mathcal{V}_r$ which maximizes the lower bound.

Building on the definition of $g_d^\star(g_s, \rho)$, the following proposition provides an estimate of the diversity-multiplexing trade-off at finite SNR.

Proposition 4.7 *In i.i.d. Rayleigh fading channels at finite SNR, an estimate $\hat{g}_d^\star(g_s, \rho)$ of the optimal diversity gain $g_d^\star(g_s, \rho)$ achievable with a multiplexing gain g_s is given by*

$$\hat{g}_d^\star(g_s, \rho) = n_t \sum_{l=1}^{n_t} \frac{\alpha_l^{n_r - n_t + 2l - 2} \exp(-\alpha_l)}{\Gamma(n_r - n_t + 2l - 1, \alpha_l)} \left[\frac{1}{\rho}\left((1 + g_a\rho)^{a_l} - (1 + g_a\rho)^{a_{l-1}}\right)\right.$$

$$\left. - g_a\left(a_l (1 + g_a\rho)^{a_l - 1} - a_{l-1} (1 + g_a\rho)^{a_{l-1} - 1}\right)\right]$$ (4.138)

where $(a_1, \ldots, a_{n_t}) \in \mathcal{V}_r$ is chosen to maximize the outage probability lower bound.
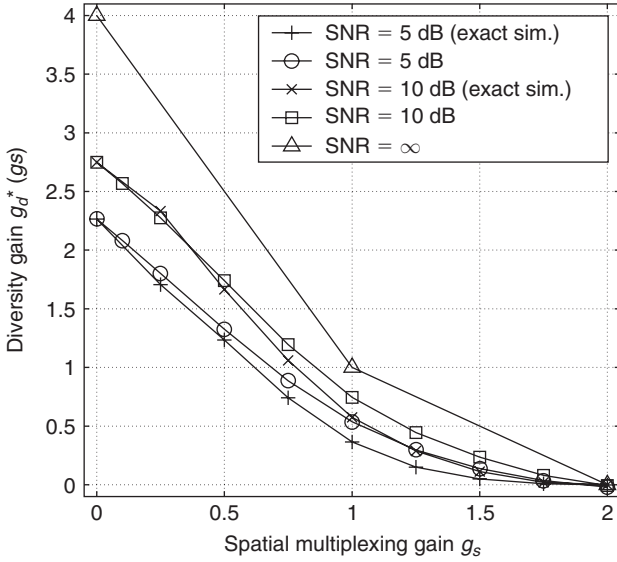
Figure 4.15   *Diversity-multiplexing trade-off at realistic SNR (5 and 10 dB) of a 2 × 2 MIMO i.i.d. Rayleigh fading channel (courtesy of R. Narasimhan [Nar05])*

Figure 4.15 illustrates the diversity-multiplexing trade-off of 2×2 Rayleigh i.i.d. MIMO channels at an SNR of 5 and 10 dB. The asymptotic curve is also displayed for comparison. It is clear from the graph that the finite SNR diversity-multiplexing trade-off is significantly lower than the asymptotic trade-off. Note that the finite SNR trade-off is evaluated thanks to Monte Carlo simulation and from (4.138). This last method slightly overestimates the trade-off since it is based on a lower bound on the outage probability. However, both methods yield pretty close results.

As the SNR increases, it may be shown that $\lim_{\rho \to \infty} \hat{g}_d^\star (g_s, \rho) = g_d^\star(g_s, \infty)$ is satisfied. At very high SNR, it is also clear that the maximum achievable diversity gain is $g_d^\star(0, \infty)$. From Proposition 4.7, evaluating the diversity gain for a zero multiplexing gain yields the maximum achievable diversity gain as a function of the SNR.

Proposition 4.8   *The maximum diversity gain achievable as a function of the SNR and for an array gain g_a is given by*

$$\lim_{g_s \to 0} \hat{g}_d^\star(g_s, \rho) = n_r n_t \left(1 - \frac{g_a \rho}{(1 + g_a \rho) \log (1 + g_a \rho)} \right). \qquad (4.139)$$

Figure 4.16 shows the normalized maximum diversity gain $\lim_{g_s \to 0} \hat{g}_d^\star(g_s, \rho)/(n_t n_r)$ given by (4.139) for the cases when $n_t = n_r = 2$, $n_t = n_r = 3$ and $n_t = n_r = 4$. The results suggest that the maximum diversity gain can only be achieved at very high SNR, which is beyond the SNR used in practice.
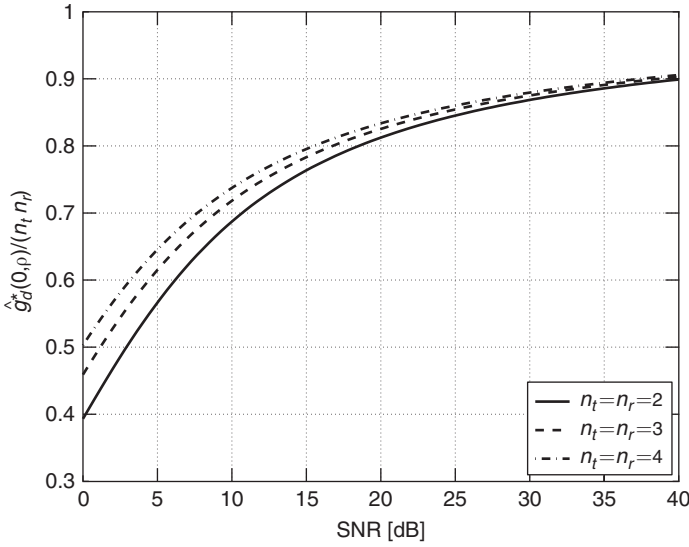
Figure 4.16 *Normalized maximum diversity gain $\hat{g}_d^\star(0,\rho)/(n_t n_r)$ as a function of SNR in i.i.d. Rayleigh fading channels*

4.7 Outage capacity and diversity-multiplexing trade-off in semi-correlated Rayleigh and Ricean slow fading channels

The aforementioned diversity-multiplexing trade-off deals with i.i.d. Rayleigh fading channels. We may further generalize the previous results to correlated Rayleigh and Ricean fading channels by means of lower bounds on the outage probability. However, getting those bounds is much more complex than for the i.i.d. case. We only briefly summarize one key result of [Nar06], assuming that $n_r \geq n_t$ and that the Rayleigh contribution is semi-correlated with transmit correlation only. The interested reader is referred to [Nar06] for more details.

The asymptotic high SNR diversity-multiplexing trade-off is given as follows in semi-correlated Rayleigh and Ricean fading channels.

Proposition 4.9 *The asymptotic high SNR diversity-multiplexing trade-off in semi-correlated Rayleigh fading channels with full-rank transmit correlation matrix is given by*

$$\lim_{\rho \to \infty} \hat{g}_d^\star{}_{Rayleigh}(g_s, \rho) = g_d^\star(g_s, \infty), \tag{4.140}$$

and in Ricean fading channels by

$$\lim_{\rho \to \infty} \hat{g}_d^\star{}_{Rice}(g_s, \rho) = g_d^\star(g_s, \infty). \tag{4.141}$$
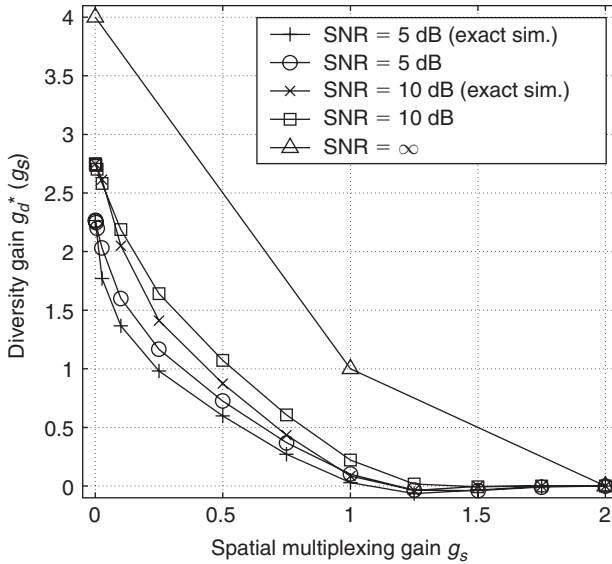
Figure 4.17   *Diversity-multiplexing trade-off at realistic SNR (5 and 10 dB) in 2 × 2 MIMO transmit correlated Rayleigh fading channels (courtesy of R. Narasimhan [Nar06])*

Quite interestingly, the spatial (transmit) correlation and the coherent component do not affect the high-SNR diversity-multiplexing trade-off! Intuitively, the high SNR trade-off is only a function of the rank of the channel matrix, i.e. the number of non-zero eigenvalues, which is not modified by full-rank transmit correlation matrix or by a coherent component with a finite K-factor (these only modify the channel condition number). However, at realistic SNR, both the rank of the channel matrix and the distribution of the eigenvalues (measured in terms of the condition number, for example) affect the trade-off. We will highlight in Chapter 6 that it is also the case when dealing with the error performance of space–time codes over correlated channels. Note also that the larger the transmit correlation and/or the Ricean K-factor, the larger the SNR required to observe the asymptotic high-SNR diversity-multiplexing trade-off.

Figure 4.17 shows the diversity-multiplexing trade-off achievable at realistic SNR over 2×2 correlated Rayleigh fading channels. It is assumed that the transmit correlation matrix is given by

$$\mathbf{R}_t = \begin{bmatrix} 1 & 0.0043 + j0.9789 \\ 0.0043 - j0.9789 & 1 \end{bmatrix} \quad (4.142)$$

corresponding to a single scattering cluster with an angle of departure of 60° with regard to the link axis and a uniform power azimuth spectrum with an angle-spread of 15°. At realistic SNR, the diversity gain in the presence of transmit correlation is reduced compared to the diversity gain observed in i.i.d. Rayleigh fading channels for the same SNR (see Figure 4.15) and is much lower than the gain obtained at asymptotically high SNR.
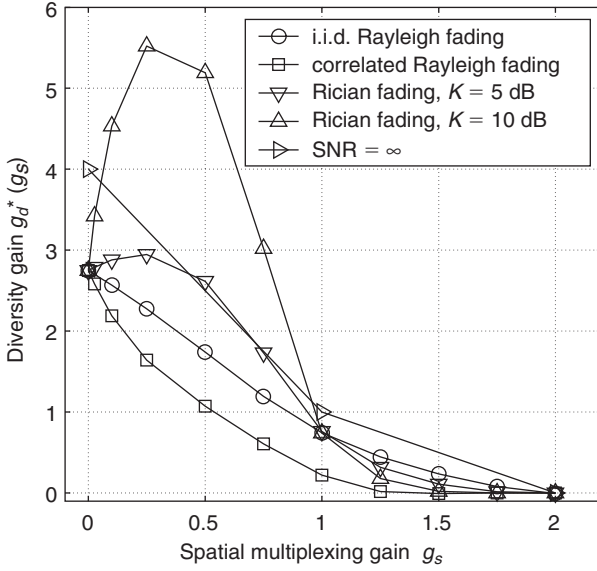
Figure 4.18 *Diversity-multiplexing trade-off at realistic SNR (10 dB) in 2×2 MIMO transmit correlated Rayleigh and Ricean (K = 5 and 10 dB) fading channels (courtesy of R. Narasimhan [Nar06])*

The impact of Ricean K-factor on the diversity-multiplexing trade-off at 10 dB SNR is shown in Figure 4.18. Results are radically different from the Rayleigh fading case. The presence of a coherent component (e.g. a LOS component) increases the diversity gain beyond the high SNR diversity gain for a spatial multiplexing gain $0 \leq g_s \leq 1$ and a sufficiently large K-factor (compared to the SNR). As the SNR increases, the performance is not only determined by the coherent component but also by the minimum eigenvalue of $\mathbf{H}^H \mathbf{H}$ which lowers the diversity gain. Note that for a given SNR and $g_s = 0$, i.e. for a transmission rate approaching zero (note the difference with an infinite SNR, where for $g_s = 0$, the transmission rate is constant), all curves converge toward the same diversity gain irrespective of the transmit correlation or the K-factor. This diversity gain is given by (4.139). It can actually be shown that (4.139) also holds true for semi-correlated Rayleigh and Ricean fading channels [Nar06]. Finally, for a spatial multiplexing gain larger than one, the diversity gain vanishes rapidly due to the lack of degrees of freedom caused by the rank-deficient coherent component.

This page intentionally left blank

5

Space–time coding over i.i.d. Rayleigh flat fading channels

As detailed in the previous chapter, information theory predicts that MIMO channels are able to provide huge gains in terms of reliability and transmission rate. In this chapter,

- we derive practical methodologies to achieve these gains
- we address how link performance and data rates may be increased through coding across space and time
- we present two design methodologies motivated by an error rate optimization perspective or by information theory
- we give an overview of schemes known as space–time codes and we discuss how they perform in terms of both error and transmission rates.

Two groups of encoding schemes are reviewed: space–time block coding (encoding is based on a block definition) and space–time trellis coding (codes are described by a trellis). Among space–time block codes, we discuss the broad class of Spatial Multiplexing (V-BLAST and D-BLAST) schemes, orthogonal and quasi-orthogonal codes, linear dispersion codes as well as the more recently developed algebraic codes. Regarding trellis codes, we consider the classical space–time trellis codes as well as the super-orthogonal space–time trellis codes.

5.1 Overview of a space–time encoder

In Figure 5.1, a general encoding scheme used in MIMO systems can be seen as a sequence of two black boxes where temporal coding, time interleaving and symbol mapping are performed in the first box while space–time coding is performed in the second box. The input of the first box is a sequence of B bits and its output is a sequence of Q symbols. This sequence of symbols is then spread in space (over the n_t transmit antennas) and in time (over T symbol durations) in order to form a codeword represented by a matrix $\mathbf{C}$ of size $n_t \times T$. The ratio B/T is the signaling rate of the transmission, whereas the ratio Q/T is defined as the spatial multiplexing rate of the space–time codes. The latter is representative of how many symbols are packed within a codeword per unit of time.
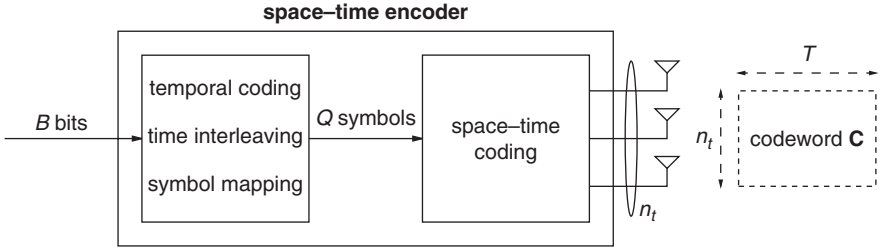
Figure 5.1   *General overview of a space–time encoder of a MIMO system*

Briefly, the goal of the first black box is to combat the randomness created by the noise at the receiver. The time interleaver improves the error correction performance of the temporal code by spreading burst errors caused by deep fades extending on several codeword lengths. The space–time code present in the second black box can be thought of as a spatial interleaver which spreads symbols over several antennas in order to mitigate the spatial selective fading. It is worth noting that for some classes of codes, the symbol mapper operation is done in the space–time coding box, namely for space–time trellis coding.

5.2   System model

In this chapter, we consider a MIMO system with n_t transmit and n_r receive antennas communicating through a frequency flat-fading channel. A codeword $\mathbf{C} = [\mathbf{c}_0 \ldots \mathbf{c}_{T-1}]$ of size $n_t \times T$ contained in the codebook $\mathcal{C}$ (i.e. the set of all possible transmitted codewords) is transmitted over T symbol durations via n_t transmit antennas. At the kth time instant, the transmitted and received signals are related by

$$\mathbf{y}_k = \sqrt{E_s}\mathbf{H}_k\mathbf{c}_k + \mathbf{n}_k \tag{5.1}$$

where $\mathbf{y}_k$ is the $n_r \times 1$ received signal vector, $\mathbf{H}_k$ is the $n_r \times n_t$ channel matrix and $\mathbf{n}_k$ is a $n_r \times 1$ zero mean complex additive white Gaussian noise (AWGN) vector with $\mathcal{E}\left\{\mathbf{n}_k\mathbf{n}_l^H\right\} = \sigma_n^2\mathbf{I}_{n_r}\delta(k-l)$. The parameter E_s is the energy normalization factor, so that the ratio E_s/σ_n^2 represents the SNR denoted as ρ.

Power normalization
In order to match the normalization of the input covariance matrix $\mathbf{Q}$ introduced in Chapter 4 ($\text{Tr}\{\mathbf{Q}\} = 1$), we normalize the codeword average transmit power such that $\mathcal{E}\left\{\text{Tr}\left\{\mathbf{CC}^H\right\}\right\} = T$. The quantity E_s can thus be thought of as the total average energy available at the transmitter over a symbol duration. We should stress that alternative normalizations may be found in the literature.

As an example, the dependence in n_t may be extracted out of the codeword normalization by considering the system model $\mathbf{y}_k = \sqrt{E_s/n_t}\mathbf{H}_k\mathbf{c}_k + \mathbf{n}_k$ with an average transmit power normalization fixed to $\mathcal{E}\left\{\text{Tr}\left\{\mathbf{CC}^H\right\}\right\} = n_tT$. Naturally, both normalizations are equivalent.

Regarding the channel energy normalization, we again assume for simplicity that $\mathcal{E}\left\{\|\mathbf{H}\|_F^2\right\} = n_t n_r$.

Slow and fast Rayleigh fading channels

In i.i.d. Rayleigh *slow fading* (*quasi-static*) channels, the coherence time T_{coh} is much larger than the codeword length, i.e. $T_{coh} \gg T$. Hence, the channel over one block is written as $\{\mathbf{H}_k = \mathbf{H}_w\}_{k=0}^{T-1}$, with $\mathbf{H}_w$ denoting an independent and identically distributed random fading matrix with unit variance circularly symmetric complex Gaussian entries.

When the channel is time varying over the duration of a block ($T \gg T_{coh}$), any codeword experiences many different channel realizations and the channel is said to be fast fading. In order to reduce T_{coh}, interleaving is typically used to break the residual correlation between successive time instants. For an ideally time interleaved channel, the channel gains vary independently between time instants. When the channel is ideally time interleaved, spatially white and Rayleigh distributed, $\mathbf{H}_k = \mathbf{H}_{k,w}$, where $\{\mathbf{H}_{k,w}\}_{k=0}^{T-1}$ are uncorrelated matrices, each $\{\mathbf{H}_{k,w}\}$ being an i.i.d. random fading matrix with unit variance circularly symmetric complex Gaussian entries. In this book, we always refer to fast fading channels as soon as $T \gg T_{coh}$, although we do not always specify whether the channel is ideally time interleaved or not (it should, however, be clear from the context).

5.3 Error probability motivated design methodology

Maximum-likelihood (ML) decoding is generally assumed when deriving space–time code designs. With instantaneous channel realizations perfectly known at the receive side, the ML decoder computes an estimate of the transmitted codeword according to

$$\hat{\mathbf{C}} = \arg\min_{\mathbf{C}} \sum_{k=0}^{T-1} \left\| \mathbf{y}_k - \sqrt{E_s}\mathbf{H}_k\mathbf{c}_k \right\|^2 \tag{5.2}$$

where the minimization is performed over all possible codeword vectors $\mathbf{C}$. When a codeword $\mathbf{C}$ is transmitted through n_t antennas, we are interested in the probability that the ML decoder decodes the codeword $\mathbf{E} = [\mathbf{e}_0 \ldots \mathbf{e}_{T-1}]$ instead of codeword $\mathbf{C}$. This probability is known as the pairwise error probability (PEP) and is classically studied as a measure of error performance. When the PEP is conditioned on the channel realizations $\{\mathbf{H}_k\}_{k=0}^{T-1}$, it is defined as the conditional PEP [Pro01],

$$P\left(\mathbf{C} \to \mathbf{E}|\{\mathbf{H}_k\}_{k=0}^{T-1}\right) = \mathcal{Q}\left(\sqrt{\frac{\rho}{2} \sum_{k=0}^{T-1} \left\|\mathbf{H}_k(\mathbf{c}_k - \mathbf{e}_k)\right\|_F^2}\right) \tag{5.3}$$

where $\mathcal{Q}(x)$ is the Gaussian $\mathcal{Q}$-function (see Appendix A). The average PEP, denoted as $P(\mathbf{C} \to \mathbf{E})$, is finally obtained by averaging the conditional PEP of (5.3) over the probability

distribution of the channel gains. Quite naturally, the system performance is in general, especially at high SNR, dominated by the couples of codewords that lead to the worst PEP. This worst PEP is referred to as the worst-case PEP or the maximum PEP, and is a fundamental concept in code design.

However, we stress that it is always better to account for the whole PEP spectrum, since the total error performance may also be highly affected by the occurrence of some codeword pairs not responsible for the worst-case PEP. Because the exact total error performance is not easily predictable, an upper bound of the average total error rate P_e may be obtained through the use of a union bound. Assuming that all codewords $\mathbf{C}$ are equally likely, the average union bound P_u reads as

$$P_e \leq P_u = \frac{1}{\sharp \mathcal{C}} \sum_{\mathbf{C} \in \mathcal{C}} \sum_{\substack{\mathbf{E} \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{E}}} P(\mathbf{C} \to \mathbf{E}). \tag{5.4}$$

where $\sharp \mathcal{C}$ denotes the cardinality of the codebook $\mathcal{C}$. Equivalently, a conditional union bound may be written similarly to (5.4) but with $P\big(\mathbf{C} \to \mathbf{E}|\{\mathbf{H}_k\}_{k=0}^{T-1}\big)$ instead of $P(\mathbf{C} \to \mathbf{E})$.

Our objective is now to derive criteria which codes have to satisfy in order to minimize the pairwise error probability when the rate of the transmission is fixed, i.e. when the codebook does not change with the SNR, or equivalently when the spatial multiplexing gain g_s is equal to zero. Before deriving the criteria for fast and slow fading channels, let us recall the definitions of the diversity and coding gains.

Definition 5.1   *The diversity gain g_d^o achieved by a pair of codewords $\{\mathbf{C}, \mathbf{E}\} \in \mathcal{C}$ is defined as the slope of $P(\mathbf{C} \to \mathbf{E})$ as a function of the SNR on a log-log scale, usually evaluated at very high SNR, i.e.*

$$g_d^o(\infty) = \lim_{\rho \to \infty} g_d^o(\rho) = -\lim_{\rho \to \infty} \frac{\log_2(P(\mathbf{C} \to \mathbf{E}))}{\log_2 \rho}. \tag{5.5}$$

A higher diversity gain g_d^o leads to an increased slope of the $P(\mathbf{C} \to \mathbf{E})$ vs. ρ curve. The quantity $g_d^o(\infty)$ is thus the diversity gain achieved by a fixed rate code at high SNR. In the slow fading case, $g_d^o(\infty)$ is thus similar to the diversity gain $g_d^\star(0, \infty)$ (introduced in Chapter 4) but is measured on the $P(\mathbf{C} \to \mathbf{E})$ curve rather than on the P_{out} curve.

Definition 5.2   *The coding gain achieved by a pair of codewords $\{\mathbf{C}, \mathbf{E}\} \in \mathcal{C}$ is defined as the magnitude of the left shift of the $P(\mathbf{C} \to \mathbf{E})$ vs. ρ curve evaluated at very high SNR.*

If $P(\mathbf{C} \to \mathbf{E})$ is well approximated at high SNR by

$$P(\mathbf{C} \to \mathbf{E}) \approx c(g_c \rho)^{-g_d^o} \tag{5.6}$$

with c being a constant; g_c is identified as the coding gain.

5.3.1 Fast fading MIMO channels: the distance-product criterion

In i.i.d. Rayleigh fast fading channels, it is shown in Appendix E that the average PEP may be expressed as

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \prod_{k=0}^{T-1} \left(1 + \eta \left\| \mathbf{c}_k - \mathbf{e}_k \right\|^2\right)^{-n_r} d\beta \tag{5.7}$$

where $\eta = \rho/(4 \sin^2 \beta)$. Because the integral over β is usually not straightforward, an upper bound of (5.7) is based on the Chernoff bound $\mathcal{Q}(x) \le e^{-x^2/2}$ (see Appendix E)

$$P(\mathbf{C} \to \mathbf{E}) \le \prod_{k=0}^{T-1} \left(1 + \frac{\rho}{4} \left\| \mathbf{c}_k - \mathbf{e}_k \right\|^2\right)^{-n_r}. \tag{5.8}$$

Hence, the integral is removed and $\sin^2 \beta$ is taken as equal to 1. In the high SNR regime, the average PEP is further upper-bounded by

$$P(\mathbf{C} \to \mathbf{E}) \le \left(\frac{\rho}{4}\right)^{-n_r l_{\mathbf{C},\mathbf{E}}} \prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \left\| \mathbf{c}_k - \mathbf{e}_k \right\|^{-2n_r} \tag{5.9}$$

with $l_{\mathbf{C},\mathbf{E}}$ the effective length of the pair of codewords $\{\mathbf{C}, \mathbf{E}\}$, i.e. $l_{\mathbf{C},\mathbf{E}} = \sharp \tau_{\mathbf{C},\mathbf{E}}$ with $\tau_{\mathbf{C},\mathbf{E}} = \{k | \mathbf{c}_k - \mathbf{e}_k \neq 0\}$. From Definitions 5.1 and 5.2, the diversity gain achieved by the pair of codewords $\{\mathbf{C}, \mathbf{E}\}$ is given by $n_r l_{\mathbf{C},\mathbf{E}}$ and the coding gain is proportional to $\prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \left\| \mathbf{c}_k - \mathbf{e}_k \right\|^{-2n_r}$.

At high SNR, the error probability is naturally dominated by the worst-case PEP (the distribution of the noise covers almost exclusively the nearest neighbors). As a consequence, the design criterion focuses on the maximization of the worst-case PEP [TSC98]. This worst-case PEP is due to error events whose effective length is equal to the minimum effective length of the code. Those codewords are responsible for the smallest diversity gain, as it can be deduced from (5.9).

Design criterion 5.1 *(**Distance-product criterion**) Over i.i.d. Rayleigh fast fading channels,*

1. **distance criterion**: *maximize the minimum effective length L_{min} of the code over all pairs of codewords $\{\mathbf{C}, \mathbf{E}\}$ with $\mathbf{C} \neq \mathbf{E}$*

$$L_{min} = \min_{\substack{\mathbf{C},\mathbf{E} \\ \mathbf{C} \neq \mathbf{E}}} l_{\mathbf{C},\mathbf{E}} \tag{5.10}$$

2. **product criterion**: *maximize the minimum product distance d_p of the code over all pairs of codewords $\{\mathbf{C}, \mathbf{E}\}$ with $\mathbf{C} \neq \mathbf{E}$*

$$d_p = \min_{\substack{\mathbf{C},\mathbf{E} \\ \mathbf{C} \neq \mathbf{E} \\ l_{\mathbf{C},\mathbf{E}} = L_{min}}} \prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \left\| \mathbf{c}_k - \mathbf{e}_k \right\|^2. \tag{5.11}$$

As deduced from (5.9), the presence of multiple antennas at the transmitter does not impact the achievable diversity gain $g_d^o = n_r L_{min}$ but improves the coding gain $g_c = d_p$.

5.3.2   Slow fading MIMO channels: the rank-determinant and rank-trace criteria

In i.i.d. Rayleigh slow fading, the average PEP reads as (see Appendix E)

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \left[\det \left(\mathbf{I}_{n_t} + \eta \tilde{\mathbf{E}} \right) \right]^{-n_r} d\beta \tag{5.12}$$

where $\tilde{\mathbf{E}} \triangleq (\mathbf{C} - \mathbf{E})(\mathbf{C} - \mathbf{E})^H$. Expressing the determinant of a matrix as the product of its eigenvalues, we may equivalently express (5.12) as

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{r(\tilde{\mathbf{E}})} \left(1 + \eta \lambda_i(\tilde{\mathbf{E}}) \right)^{-n_r} d\beta \tag{5.13}$$

with $r(\tilde{\mathbf{E}})$ denoting the rank of the error matrix $\tilde{\mathbf{E}}$ and $\{\lambda_i(\tilde{\mathbf{E}})\}$ for $i = 1, \ldots, r(\tilde{\mathbf{E}})$ the set of its non-zero eigenvalues. Again using the Chernoff bound, an upper bound of (5.12) is given by

$$P(\mathbf{C} \to \mathbf{E}) \leq \left[\det \left(\mathbf{I}_{n_t} + \frac{\rho}{4} \tilde{\mathbf{E}} \right) \right]^{-n_r} \tag{5.14}$$

$$= \prod_{i=1}^{r(\tilde{\mathbf{E}})} \left(1 + \frac{\rho}{4} \lambda_i(\tilde{\mathbf{E}}) \right)^{-n_r}. \tag{5.15}$$

Rank-determinant criterion in the high SNR regime
At high SNR, $\frac{\rho}{4} \lambda_i(\tilde{\mathbf{E}}) \gg 1$ and (5.15) becomes

$$P(\mathbf{C} \to \mathbf{E}) \leq \left(\frac{\rho}{4} \right)^{-n_r r(\tilde{\mathbf{E}})} \prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i^{-n_r}(\tilde{\mathbf{E}}) \tag{5.16}$$

where $\lambda_i^x(\tilde{\mathbf{E}})$ denotes $\lambda_i(\tilde{\mathbf{E}})$ to the xth power. For a given pair of codewords $\{\mathbf{C}, \mathbf{E}\}$, the diversity gain is given by the rank of the error matrix $\tilde{\mathbf{E}}$ multiplied by the number of receive antennas, i.e. $n_r r(\tilde{\mathbf{E}})$. The coding gain is directly proportional to the quantity

$$\prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i(\tilde{\mathbf{E}}). \tag{5.17}$$

The maximization of the worst-case PEP leads to the following design criterion [TSC98].

Design criterion 5.2   *(Rank-determinant criterion) Over i.i.d. Rayleigh slow fading channels,*

1. **rank criterion**: *maximize the minimum rank r_{min} of $\tilde{\mathbf{E}}$ over all pairs of codewords $\{\mathbf{C}, \mathbf{E}\}$ with $\mathbf{C} \neq \mathbf{E}$*

$$r_{min} = \min_{\substack{\mathbf{C}, \mathbf{E} \\ \mathbf{C} \neq \mathbf{E}}} r(\tilde{\mathbf{E}}) \tag{5.18}$$

2. **determinant criterion**: *over all pairs of codewords $\{\mathbf{C}, \mathbf{E}\}$ with $\mathbf{C} \neq \mathbf{E}$, maximize the minimum of the product d_λ of the non-zero eigenvalues of $\tilde{\mathbf{E}}$*

$$d_\lambda = \min_{\substack{\mathbf{C}, \mathbf{E} \\ \mathbf{C} \neq \mathbf{E}}} \prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i(\tilde{\mathbf{E}}). \tag{5.19}$$

If $r_{min} = n_t$, the product of the non-zero eigenvalues of the error matrix is equal to the determinant of the error matrix and the determinant criterion comes to maximize the minimum determinant of the error matrix over all pairs of codewords $\{\mathbf{C}, \mathbf{E}\}$ with $\mathbf{C} \neq \mathbf{E}$

$$d_\lambda = \min_{\substack{\mathbf{C}, \mathbf{E} \\ \mathbf{C} \neq \mathbf{E}}} \det(\tilde{\mathbf{E}}). \tag{5.20}$$

The diversity gain is maximized first, and the coding gain is maximized only in a second step.

Definition 5.3 *A full-rank (also known as full-diversity) code is characterized by $r_{min} = n_t$. A rank-deficient code is characterized by $r_{min} < n_t$.*

To guarantee that $r(\tilde{\mathbf{E}}) = n_t$, it is mandatory that the number of non-zero columns of $\mathbf{C} - \mathbf{E}$ be larger than the number of lines, i.e. n_t. A necessary but not sufficient condition to guarantee $r(\tilde{\mathbf{E}}) = n_t$ is that $l_{\mathbf{C},\mathbf{E}} \geq n_t$. Therefore, a necessary but not sufficient condition to guarantee that a code be full rank is that $L_{min} \geq n_t$. Note that $L_{min} \geq n_t$ is a much stronger condition than $T \geq n_t$ since $L_{min} \leq T$. When $r_{min} = n_t$ and $T = n_t$, the code is of minimum length.

Rank-trace criterion in the low SNR regime

At high SNR, classical design criteria are based on an upper bound of the average PEP. However, it has been shown that if the SNR is small and/or if the number of receive antennas is large [CYV01, BTT02], the determinant in (5.15) is mostly a function of the trace of $\tilde{\mathbf{E}}$. Indeed

$$\det\left(\mathbf{I}_{n_t} + \frac{\rho}{4}\tilde{\mathbf{E}}\right) = 1 + \frac{\rho}{4}\mathrm{Tr}\{\tilde{\mathbf{E}}\} + \ldots + \left(\frac{\rho}{4}\right)^{r(\tilde{\mathbf{E}})} \prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i(\tilde{\mathbf{E}}) \tag{5.21}$$

and the design criterion at low SNR and/or for large n_r in i.i.d. Rayleigh slow fading channels becomes [CYV01, BTT02] as follows.

Design criterion 5.3   (*Rank-trace criterion*) *Over i.i.d. Rayleigh slow fading channels at low SNR and/or for large n_r,*

1. **rank criterion**: *maximize the minimum rank r_{min} of $\tilde{\mathbf{E}}$ over all pairs of codewords $\{\mathbf{C}, \mathbf{E}\}$ with $\mathbf{C} \neq \mathbf{E}$.*
2. **trace criterion**: *maximize the minimum trace d_e of $\tilde{\mathbf{E}}$ over all pairs of codewords $\{\mathbf{C}, \mathbf{E}\}$ with $\mathbf{C} \neq \mathbf{E}$*

$$d_e = \min_{\substack{\mathbf{C}, \mathbf{E} \\ \mathbf{C} \neq \mathbf{E}}} \mathrm{Tr}\{\tilde{\mathbf{E}}\}. \tag{5.22}$$

While (5.21) is not only valid for full-diversity codes, simulations in [CYV01] have shown that, for a realistic number of receive antennas, taking full-rank codes and maximizing the trace of the error matrices lead to trellis codes outperforming trellis codes inferred from the rank-determinant criterion. Actually, the rank-trace criterion is generally recommended as soon as $n_r n_t \geq 4$.

Note that $\mathrm{Tr}\{\tilde{\mathbf{E}}\} = \|\mathbf{C} - \mathbf{E}\|_F^2$ can be thought of as an Euclidean distance. The trace criterion is therefore nothing else than an Euclidean distance maximization when the SNR is small or the number of antennas is large, which makes it very similar to the classical Euclidean distance maximization in AWGN channels. However, this is not surprising. With a sufficiently large number of receive antennas, a large amount of receive diversity is available and the fading is therefore reduced. In the asymptotic case of an infinite number of receive antennas, the channel is constant and behaves as if it were indeed AWGN.

Remark 5.1   *Since i.i.d. Rayleigh fading channels are spatially white, it is important to note that once an optimal code has been identified, an infinity of equally optimal codes are obtained by multiplying the codewords by a unitary precoder $\mathbf{P}$ $\left(\mathbf{P}^H \mathbf{P} = \mathbf{I}_{n_t}\right)$: if the set of codewords $\{\mathbf{C}\}$ is a good candidate, all sets of codewords $\{\mathbf{PC}\}$ with $\mathbf{P}$ unitary are equally good.*

Antenna selection

The selection of n_r' among the n_r receive antennas (resp. n_t' among the n_t transmit antennas) has been shown in Chapter 1 to be an efficient method for exploiting the full diversity gain n_r (resp. n_t) when the channel is perfectly known by the receiver (resp. transmitter). Here, we want to investigate whether receive antenna selection may extract the full receive diversity when space–time coding is used at the transmitter. It is actually shown in [BDA03] that the average PEP at high SNR in i.i.d. Rayleigh slow fading channels with a selection algorithm based on the instantaneous receive SNR at each antenna (i.e. based on the squared norm of each row of the channel matrix $\mathbf{H}$) is given by

$$P(\mathbf{C} \to \mathbf{E}) \leq \frac{n_r!}{(n_r - n_r')! n_r'! (n_t!)^{n_r - n_r'}} \prod_{i=1}^{n_t} \lambda_i^{-n_r'}(\tilde{\mathbf{E}}) \xi_0 \left(\frac{\rho}{4}\right)^{-n_t n_r} \tag{5.23}$$

for full-rank codes, and by

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{n_r!(n_r - n_r' + 1)^{r(\tilde{\mathbf{E}}) - n_t n_r}}{(n_r - n_r')! n_r'! (n_t!)^{n_r - n_r'}} \prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i^{-n_r'}(\tilde{\mathbf{E}}) \xi_1 \left(\frac{\rho}{4}\right)^{-r(\tilde{\mathbf{E}}) n_r'} \tag{5.24}$$

for rank-deficient codes. Coefficients ξ_0 and ξ_1 are function of the eigenvalues of the error matrix and therefore only affect the coding gain. Interestingly, (5.23) and (5.24) suggest that in a MIMO system with space–time coding at the transmitter, a SNR-based receive antenna selection does provide the full diversity gain only if the code is full rank. If the code is rank-deficient, the diversity achieved with receive antenna selection is equal to $n_r' r(\tilde{\mathbf{E}})$, while it is equal to $n_r r(\tilde{\mathbf{E}})$ without antenna selection. The use of receive antenna selection should therefore be restricted to full-rank codes.

This is a rather surprising result, as we might intuitively have thought that receive diversity comes for free and does not depend on the transmit scheme. But a rank-deficient code can be seen as a $r(\tilde{\mathbf{E}})$-rank code transmitted over $r(\tilde{\mathbf{E}})$ antennas, augmented by $n_t - r(\tilde{\mathbf{E}})$ lines transmitted over the $n_t - r(\tilde{\mathbf{E}})$ remnant antennas. Those lines are linear combinations of the lines of the $r(\tilde{\mathbf{E}})$-rank code and do not enable the receiver to improve the diversity gain. However, those $n_t - r(\tilde{\mathbf{E}})$ remnant antennas affect the selection of the subset of receive antennas. There is absolutely no guarantee that the chosen subset is the same as the one obtained if the $n_t - r(\tilde{\mathbf{E}})$ remnant antennas were not present. Hence, those antennas may induce the selection of a wrong receive antenna subset.

5.4   Information theory motivated design methodology

In the previous section, the design criteria maximize the diversity and coding gains for a fixed transmission rate and are based on the average PEP. We take here an alternative approach exploiting information theory results and aim to achieve the diversity-multiplexing trade-off, whose theoretical background has been outlined in Chapter 4.

5.4.1   *Fast fading MIMO channels: achieving the ergodic capacity*

Fast fading MIMO channels constitute a relatively simple situation with respect to slow fading channels. The reason for this is that the transmission capability over fast fading channels is described by a single quantity, which is the ergodic capacity

$$\bar{C} = \max_{\mathbf{Q}:\mathrm{Tr}\{\mathbf{Q}\}=1} \mathcal{E}\left\{\log_2 \det\left(\mathbf{I}_{n_r} + \rho \mathbf{H}\mathbf{Q}\mathbf{H}^H\right)\right\}. \tag{5.25}$$

When the channel realizations are known to the transmitter (CSIT scenario), we already know that the transmission of independent streams in the directions of the eigenvectors of the channel matrix $\mathbf{H}$ decouples the system into $n \triangleq \min\{n_t, n_r\}$ parallel data pipes. For a total transmission rate R, each layer k can then be encoded using a capacity-achieving

Gaussian code with rate R_k such that $\sum_{k=1}^{n} R_k = R$, ascribed a power $\lambda_k(\mathbf{Q})$ (the eigenvalues of $\mathbf{Q}$ can be seen as the powers attached to each layer) and can be decoded independently of the other layers. The optimal power allocation $\{\lambda_k^*(\mathbf{Q})\}$ is given by the water-filling allocation strategy in (4.7) described in Section 4.1. The resulting ergodic capacity is then written as (4.6)

$$\bar{C} = \sum_{k=1}^{n} \mathcal{E} \left\{ \log_2 \left(1 + \rho \lambda_k^*(\mathbf{Q}) \lambda_k \right) \right\} \tag{5.26}$$

with $\lambda_k \triangleq \lambda_k(\mathbf{HH}^H)$.

When the channel is i.i.d. Rayleigh fading and the instantaneous channel realizations are unknown to the transmitter (CDIT), it is shown in [Tel95] that the optimal transmit covariance matrix $\mathbf{Q}$ equals the identity matrix, i.e. $\mathbf{Q} = (1/n_t)\,\mathbf{I}_{n_t}$. The ergodic capacity is equal to

$$\bar{C} = \mathcal{E} \left\{ \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{n_t} \mathbf{HH}^H \right) \right\} \tag{5.27}$$

which at high SNR, has been shown to be approximated as (4.39)

$$\bar{C} \approx n \, \log_2 \left(\frac{\rho}{n_t} \right) + \sum_{k=1}^{n} \mathcal{E} \left\{ \log_2 \left(\chi_{2(N-n+k)}^2 \right) \right\} \tag{5.28}$$

suggesting that the transmission of independent information symbols may be performed in parallel over n virtual spatial channels. The transmitter is very similar to the CSIT case except that all eigenmodes now receive the same amount of power. However, all layers are also subject to interferences from the other layers at the receiver, hence the independent decoding of all layers is clearly suboptimal. Actually, a transmission scheme with uniform power allocation over n_t independent layers, each layer using an AWGN capacity-achieving code and performing joint ML decoding, achieves the ergodic capacity of fast fading MIMO channels [Fos96, FG98, HH02a]. This particular scheme is widely known as Spatial Multiplexing and will be reviewed later.

Achieving the ergodic capacity requires coding over an extremely large number of channel realizations to average out the effects of both fading and noise. Therefore, the channel needs to be sufficiently time-varying, such that interleaving is able to convert any channel into a fast fading channel. In practice, however, the codeword length may not be long enough to assume an ergodic channel. When coding is only performed over a single block, the channel becomes slow fading, and averaging out the randomness of the channel is not possible anymore. Achieving the diversity-multiplexing trade-off is the objective to design efficient space–time codes.

5.4.2 Slow fading MIMO channels: achieving the diversity-multiplexing trade-off

From information theory to code design

With CSIT, a dynamic rate/power allocation with separate coding avoids outages thanks to the decoupling into parallel data pipes. Without transmit channel knowledge, the rate adaptation is not possible anymore. Natural questions in slow fading channels are as follows. Is the transmission of independent streams sufficient to achieve some capacity optimality? Is coding across antenna necessary? Since the channel is slowly fading, it is impossible to code over a large number of independent channel realizations. Ultimately, we might end up coding over a single channel realization. Therefore, separate coding would lead to an outage as soon as one of the subchannels is in deep fade. This is the reason why joint coding across all subchannels is necessary in the absence of transmit channel knowledge, although this makes the decoding a significantly more complex task.

The goal is now to minimize the probability that the information rate falls below a given rate, i.e. to derive architectures achieving the outage performance. By contrast, in Section 5.3, the objective was to maximize the reliability (diversity gain). For a given fixed transmission rate, we had to simultaneously maximize the diversity and coding gains irrespective of the spatial multiplexing rate, namely implying that the ratio between the fixed transmission rate and the ergodic capacity vanishes as the SNR increases. By contrast, we need now to design codes achieving the diversity-multiplexing trade-off. Therefore, we want to transmit at a rate $R = g_s \log_2(\rho)$ and achieve a non-vanishing fraction of the ergodic capacity as the SNR increases. Similarly to Section 4.6.1, we define to this end the diversity-multiplexing trade-off achieved by a family of codes indexed by the SNR, denoted as $\{\mathcal{C}(\rho)\}$. This family of codes works similarly to adaptive modulation and coding. The idea is to use a particular codebook for a given SNR. As the SNR changes, the codebook changes in order to increase or decrease the total transmission rate.

Definition 5.4 *A scheme $\{\mathcal{C}(\rho)\}$, i.e. a family of codes indexed by the SNR ρ, is said to achieve a diversity gain g_d and a multiplexing gain g_s if*

$$\lim_{\rho\to\infty} \frac{R(\rho)}{\log_2(\rho)} = g_s \tag{5.29}$$

$$\lim_{\rho\to\infty} \frac{\log_2(P_e(\rho))}{\log_2(\rho)} = -g_d \tag{5.30}$$

or equivalently (based on the notations introduced in Chapter 4)

$$2^R \doteq \rho^{g_s} \tag{5.31}$$

$$P_e(\rho) \doteq \rho^{-g_d} \tag{5.32}$$

where $R(\rho)$ is the data rate and $P_e(\rho)$ the error probability averaged over the additive noise, the i.i.d. channel statistics and the transmitted codewords. The curve

$g_d(g_s)$ *is the diversity-multiplexing trade-off achieved by the scheme in the high SNR regime.*

It is important to highlight the significant distinction with the methodology of Section 5.3. Indeed, the diversity gain $g_d^o(\infty)$ defined in Section 5.3 is the slope of the pairwise error probability $P(\mathbf{C} \to \mathbf{E})$ at high SNR for a fixed code ($g_s = 0$). By contrast, the diversity defined in Definition 5.4 is the slope of the actual error probability P_e for a rate increasing with the SNR. Thereby, the difference between $g_d(0)$ and g_d^o lies in the fact that we use P_e instead of $P(\mathbf{C} \to \mathbf{E})$ in the definition of the diversity gain.

The extreme point of the curve $(0, g_d(0))$ has a particular signification with regard to the design criteria of Section 5.3. Indeed, this point indicates that, for a spatial multiplexing gain of 0 (i.e. when the transmission rate is kept fixed), the achievable diversity gain is $g_d(0)$. From the channel diversity-multiplexing trade-off, we know that $g_d(0) \leq n_t n_r$. Achieving this upper bound was actually our goal in Section 5.3, as the rank-criterion aims at extracting the maximal diversity gain possible out of $n_t n_r$ for a fixed data rate. The other extreme point $(g_{s,max}, 0)$ tells us that when transmitting at a diversity gain $g_d = 0$ (i.e. when the error probability is kept fixed), the data rate can be increased with SNR proportionally to $g_{s,max} \leq \min\{n_t, n_r\}$. When $g_s = \min\{n_t, n_r\}$, we transmit at a rate close to ergodic capacity, which suggests that there is no protection against the randomness of the channel, and the diversity gain is equal to zero. Intermediate points on the trade-off curve indicate how much diversity can be traded off for multiplexing and vice-versa. Information theory motivated design methodologies provide a much broader view of the possible coding performance than the error probability motivated methodologies of Section 5.3.

Design criteria
The following result is discussed in [ZT03].

Proposition 5.1 *The optimal asymptotic diversity-multiplexing trade-off of Proposition 4.5 is achievable with Gaussian random codes of length $T \geq n_t + n_r - 1$.*

Because the proof is quite lengthy, we only proceed here with a sketch of the proof.

PROOF: It consists in showing that P_e can be lower-bounded and upper-bounded as

$$P_{out}(R) \stackrel{(a)}{\leq} P_e(\rho) \stackrel{(b)}{\leq} P_{out}(R) + P(\text{error, no outage}) \tag{5.33}$$

and that under the assumption $T \geq n_t + n_r - 1$, the SNR exponent in the upper bound is not smaller than the SNR exponent in the outage probability as $\rho \to \infty$. Inequality (a) should not be surprising since it suggests that the error probability with a finite block length T is always lower-bounded by the outage probability: achieving the outage probability requires indeed infinite long codes and the performance improves as the block length increases. The problem with finite length codes is that errors may be caused by different effects: an ill-conditioned channel matrix $\mathbf{H}$, large noise or too close codewords. The last two effects

disappear as the block length increases since the so-created randomness is averaged out. Therefore, an upper bound of P_e is obtained in (b) by taking into account the probability that the channel is in outage, (the channel is ill-conditioned) and the probability that the channel is not in outage, but that an error occurs due to the other effects (i.e. a large noise and/or too close codewords). The first term $P_{out}(R)P(\text{error}|\text{outage}) \leq P_{out}(R)$ accounts for the case of a channel in outage. The probability that the channel is not in outage but that an error occurs is accounted for in the term $P(\text{error, no outage})$.

In order to evaluate the SNR exponent of $P(\text{error, no outage})$, a union bound approach is considered. At a transmission rate of $R = g_s \log_2(\rho)$, the total number of codewords is $2^{RT} = \rho^{g_s T}$, so that $P(\text{error}|\mathbf{H}) \leq \rho^{g_s T} p$ where p is the conditional PEP averaged over the ensemble of random codes (analogous to (5.15) but averaging over Gaussian codes instead of the channel)

$$p \leq \left[\det \left(\mathbf{I}_{n_r} + \frac{\rho}{4} \mathbf{H}\mathbf{H}^H \right) \right]^{-T}. \tag{5.34}$$

Finally, averaging over the non-outage events yields $P(\text{error, no outage})$. It can be shown that its SNR exponent matches the SNR exponent in P_{out}, indicating that the dominant error is caused by the outage event.

For Gaussian random codes with $T < n_t + n_r - 1$, it is not possible to match the lower and upper bounds. There is a significant probability that some codewords become close to each other: the outage event is not the dominating error event anymore. ∎

Remark 5.2 *Two important points are noteworthy. First, it should be remembered that achieving the optimal diversity-multiplexing trade-off does not require coding over an infinite block length (whereas it would be needed to achieve outage probability). Indeed, no additional diversity gain is obtained by coding over block lengths larger than $n_t + n_r - 1$. Second, the previous result is valid for Gaussian random codes. Remarkably, we will show later that structured codes with block length $T < n_t + n_r - 1$ can also achieve the optimal-diversity multiplexing trade-off.*

Example 5.1 *In Example 4.1, we have illustrated that the diversity-multiplexing trade-off of a scalar Rayleigh fading channel is $g_d^\star(g_s, \infty) = 1 - g_s$ for $g_s \in [0, 1]$.*

Let us now investigate whether a QAM constellation trades off diversity and multiplexing optimally. The error probability at high SNR may be approximated as

$$P_e \doteq \frac{1}{\rho d_{min}^2} \tag{5.35}$$

with d_{min}^2 the minimum Euclidean distance of the constellation. For a rate R that scales with SNR as $g_s \log_2(\rho)$, the number of points in the constellation is $2^{R/2}$ per

dimension. For a unit average energy constellation, $d^2_{min} \approx 1/2^R$ and the error probability is approximately

$$P_e \doteq \rho^{-(1-g_s)} \tag{5.36}$$

showing that QAM constellations optimally achieve the asymptotic diversity-multiplexing trade-off in scalar Rayleigh fading channels.

Intuitively, the diversity-multiplexing trade-off indicates that

- as the rate increases, more constellation symbols have to be packed together
- if the distance between constellations points shrinks too much with increasing SNR, then the achievable diversity gain decreases.

Much research has been carried out in order to design schemes achieving the diversity-multiplexing trade-off in i.i.d. Rayleigh fading channels. In particular, it has been investigated whether it is possible to achieve this trade-off with structured minimum-delay codes (i.e. $T = n_t$). The following result has been proved for the two-transmit antenna case in [YW03] and generalized to any number of transmit antennas in [ERKP$^+$06].

Proposition 5.2 *A minimum-delay $(T = n_t)$ code characterized by a codebook $\mathcal{C}$ of size $\sharp\mathcal{C}$*

- *which is linear over an alphabet $\mathcal{B}$, i.e. the elements of the codewords can be expressed as a linear combination of complex numbers with the weights given by the symbols of $\mathcal{B}$.*
- *whose rate scales with the SNR as*

$$\sharp\mathcal{B} = 2^{\frac{R}{n_t}} \doteq \rho^{\frac{g_s}{n_t}} \tag{5.37}$$

$$\max_{a \in \mathcal{B}} |a|^2 \dot{\leq} 1 \tag{5.38}$$

$$E_s \doteq \rho^{\frac{g_s}{n_t}} \tag{5.39}$$

- *which is full-rate, i.e. whose size satisfies $\sharp\mathcal{C} = (\sharp\mathcal{B})^{n_t T}$,*

achieves the diversity-multiplexing trade-off for any number of receive antennas if it satisfies the non-vanishing determinant criterion.

Design criterion 5.4   *(**Non-vanishing determinant criterion**) Design the codebook $\mathcal{C}$ whose rate scales with the SNR in such a way that*

$$\min_{\substack{\mathbf{C}, \mathbf{E} \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{E}}} \det(\tilde{\mathbf{E}}) \dot{\geq} \rho^{-g_s} \tag{5.40}$$

or equivalently

$$\forall \rho, \exists \epsilon > 0, \min_{\substack{\mathbf{C}, \mathbf{E} \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{E}}} \det\left(E_s \tilde{\mathbf{E}}\right) \geq \epsilon. \tag{5.41}$$

Example 5.2 *An alphabet $\mathcal{B}$ that satisfies (5.37) is the QAM constellation*

$$\mathcal{B}_M = \{x + jy | -M + 1 \leq x, y \leq M - 1, \ x, y \ odd\} \tag{5.42}$$

with M^2 elements where $M^2 = \rho^{\frac{g_s}{n_t}}$.

To get some insight into the reason why this non-vanishing determinant criterion is so important to achieve the diversity-multiplexing trade-off, we consider a 2×2 MIMO system with $T = 2$. The following proof follows directly from derivations in [YW03].

PROOF: Let us examine how error probability behaves at high SNR. To this end, an upper bound of the conditional error probability is derived and averaged over the channel distributions. The upper bound on $P(\mathbf{C} \rightarrow \mathbf{E})$ is obtained by means of a lower bound on the minimum distance at high SNR when the channel is not in outage. It is shown in [YW03] that the squared distance is large with respect to the noise variance, indicating that codewords tend to be far apart when the channel is not in outage. An outage event is thus the dominant error event and the error probability is very close to the outage probability. For two arbitrary codewords $\mathbf{C}$ and $\mathbf{E}$, the squared distance $\|\mathbf{H}(\mathbf{C} - \mathbf{E})\|_F^2$ is lower-bounded in terms of $\det(\mathbf{H})$ and $\|\mathbf{H}\|_F^2$. To proceed, we first evaluate the singular values of $\mathbf{C} - \mathbf{E}$

$$\sum_{i=1}^{2} \sigma_i^2(\mathbf{C} - \mathbf{E}) = \|\mathbf{C} - \mathbf{E}\|_F^2 \overset{\cdot}{\leq} \max \|\mathbf{C}\|_F^2 \overset{(a)}{\underset{\cdot}{\leq}} 1 \doteq \rho^0 \tag{5.43}$$

$$\prod_{i=1}^{2} \sigma_i^2(\mathbf{C} - \mathbf{E}) = |\det(\mathbf{C} - \mathbf{E})|^2 \overset{(b)}{\underset{\cdot}{\geq}} \rho^{-g_s} \tag{5.44}$$

where (a) and (b) are hypotheses required by Proposition 5.2, and $\sigma_i(\mathbf{C} - \mathbf{E})$, $i = 1, 2$ are the singular values of matrix $\mathbf{C} - \mathbf{E}$ ranked in decreasing order. Those last equations imply that $\rho^{-\frac{g_s}{2}} \overset{\cdot}{\leq} \sigma_1^2(\mathbf{C} - \mathbf{E}) \overset{\cdot}{\leq} \rho^0$ and $\rho^{-g_s} \overset{\cdot}{\leq} \sigma_2^2(\mathbf{C} - \mathbf{E}) \overset{\cdot}{\leq} \rho^0$. We are thereby able to lower-bound $\|\mathbf{H}(\mathbf{C} - \mathbf{E})\|_F^2$ using $\det(\mathbf{H})$

$$\|\mathbf{H}(\mathbf{C} - \mathbf{E})\|_F^2 \overset{(a)}{\geq} |\text{Tr}\{\mathbf{H}(\mathbf{C} - \mathbf{E})\}||\sigma_2(\mathbf{H}(\mathbf{C} - \mathbf{E}))| \geq |\det(\mathbf{H}(\mathbf{C} - \mathbf{E}))|$$
$$= |\det(\mathbf{H})||\det(\mathbf{C} - \mathbf{E})| \geq |\det(\mathbf{H})|\rho^{-\frac{g_s}{2}} \tag{5.45}$$

and using $\|\mathbf{H}\|_F^2$

$$\|\mathbf{H}(\mathbf{C} - \mathbf{E})\|_F^2 \overset{(a)}{\geq} \sigma_2^2(\mathbf{C} - \mathbf{E})\|\mathbf{H}\|_F^2 \overset{\cdot}{\geq} \rho^{-g_s}\|\mathbf{H}\|_F^2. \tag{5.46}$$

In (a), we make use of the matrix property $\text{Tr}\{\mathbf{AB}\} \geq \text{Tr}\{\mathbf{A}\}\lambda_{min}(\mathbf{B})$ with $\sigma_{min}(\mathbf{B}) = \sqrt{\lambda_{min}(\mathbf{B})}$ the smallest singular value of $\mathbf{B}$. We may write from the two previous lower bounds

$$\rho\|\mathbf{H}(\mathbf{C} - \mathbf{E})\|_F^2 \; \dot{\geq} \; \max\left\{\rho^{1-\frac{g_s}{2}}|\det(\mathbf{H})|, \rho^{1-g_s}\|\mathbf{H}\|_F^2\right\}. \tag{5.47}$$

Since the channel mutual information with equal power allocation reads as

$$\mathcal{I}_e(\mathbf{H}) = \log_2\left[1 + \|\mathbf{H}\|_F^2 \frac{\rho}{2} + (\det(\mathbf{H}))^2 \left(\frac{\rho}{2}\right)^2\right], \tag{5.48}$$

we have

$$2^{\mathcal{I}_e(\mathbf{H})-R} \; \dot{=} \; \left(\rho^{1-\frac{g_s}{2}}\det(\mathbf{H})\right)^2 + \rho^{1-g_s}\|\mathbf{H}\|_F^2. \tag{5.49}$$

When the channel is not in outage, $\det(\mathbf{H})$ and $\|\mathbf{H}\|_F^2$ cannot be simultaneously small. Expressed otherwise, when the channel is not in outage, $C(\mathbf{H}) > R$ implies that $\rho^{1-\frac{g_s}{2}}\det(\mathbf{H}) \; \dot{\geq} \; 1$ or $\rho^{1-g_s}\|\mathbf{H}\|_F^2 \; \dot{\geq} \; 1$. This leads to the conclusion in (5.47) that

$$\rho\|\mathbf{H}(\mathbf{C} - \mathbf{E})\|_F^2 \; \dot{\geq} \; 1. \tag{5.50}$$

When the channel is not in outage, all codewords are far apart from each other relative to the noise power. The dominant error event is thus the outage event. This is also the conclusion of the proof of Proposition 5.1. Afterwards, expressing $\det(\mathbf{H})$ and $\text{Tr}(\mathbf{H})$ as a function of $\sigma_i(\mathbf{H})$, and deriving the error probability for high SNR by averaging over the noise and i.i.d. Rayleigh channel realizations, we observe that the error probability has the same SNR exponent as the outage probability. ∎

5.5 Space–time block coding

Space–time block codes (STBCs) have considerably evolved over the years. While they originally attracted much attention thanks to their low decoding complexity, they have been the object of a renewed interest at the light of the diversity-multiplexing trade-off.

STBCs can be seen as a mapping of Q symbols (complex or real) onto a codeword $\mathbf{C}$ of size $n_t \times T$. Those codewords are uncoded in the sense that no error correcting code is contained in the STBC. Theoretically, STBCs may take several forms, but practically, linear STBCs are by far the most widely used. The idea behind linear STBCs is to spread information symbols in space and time in order to improve either the diversity gain, or the spatial multiplexing rate, or both the diversity gain and the spatial multiplexing rate. By packing more symbols into a given codeword, i.e. by increasing Q, the data rate is increased. In the following, we first present a general framework of linear STBCs together with some general properties, before particularizing this framework to some important subclasses of STBCs.

We assume that the channel is constant over the duration of the STBC codeword T. Over such duration, we may drop the subscript k in $\mathbf{H}_k$ and simply denote the channel matrix

as **H**. Bear in mind that in practice, STBCs are always concatenated with outer coding in order to improve error correcting capabilities and to mitigate the noise. Even if the channel is constant over the STBC codeword length, the outer code spreads across several STBC blocks, and enables to several channel realizations if the channel is varying between STBC blocks. As a consequence, the channel seen by the outer code may be fast fading.

5.5.1 A general framework for linear STBCs

Codeword structure
A linear STBC is expressed in its general form as [HH01]

$$\mathbf{C} = \sum_{q=1}^{Q} \boldsymbol{\Phi}_q \Re[c_q] + \boldsymbol{\Phi}_{q+Q} \Im[c_q] \tag{5.51}$$

where $\boldsymbol{\Phi}_q$ are complex basis matrices of size $n_t \times T$, c_q stands for the complex information symbol (taken for example from PSK or QAM constellations), Q is the number of complex symbols c_q transmitted over a codeword, $\Re$ and $\Im$ stand for the real and imaginary parts.

Definition 5.5 *The spatial multiplexing rate of a space–time block code is defined as* $r_s = \frac{Q}{T}$. *A full rate space–time block code is characterized by* $r_s = n_t$.

Applying the vec operator to (5.1) and making use of (5.51), a linear STBC-based transmission may be rewritten as follows

$$\mathcal{Y} = \mathcal{H}\mathcal{X}\mathcal{S} + \mathcal{N} \tag{5.52}$$

where $\mathcal{Y}[2n_r T \times 1]$ is the channel output vector, $\mathcal{H}[2n_r T \times 2n_t T]$ is the block diagonal channel, $\mathcal{X}[2n_t T \times 2Q]$ is the linear code matrix, $\mathcal{S}[2Q \times 1]$ is a block of uncoded input symbols and $\mathcal{N}[2n_r T \times 1]$ is the noise vector

$$\mathcal{Y} = \text{vec}\left(\begin{bmatrix} \Re[\mathbf{y}_0 \cdots \mathbf{y}_{T-1}] \\ \Im[\mathbf{y}_0 \cdots \mathbf{y}_{T-1}] \end{bmatrix}\right),$$

$$\mathcal{H} = \mathbf{I}_T \otimes \mathbf{H}', \quad \text{where} \quad \mathbf{H}' = \begin{bmatrix} \Re[\mathbf{H}] & -\Im[\mathbf{H}] \\ \Im[\mathbf{H}] & \Re[\mathbf{H}] \end{bmatrix},$$

$$\mathcal{X} = \begin{bmatrix} \text{vec}\left(\begin{bmatrix} \Re[\boldsymbol{\Phi}_1] \\ \Im[\boldsymbol{\Phi}_1] \end{bmatrix}\right) & \cdots & \text{vec}\left(\begin{bmatrix} \Re[\boldsymbol{\Phi}_{2Q}] \\ \Im[\boldsymbol{\Phi}_{2Q}] \end{bmatrix}\right) \end{bmatrix},$$

$$\mathcal{S} = \begin{bmatrix} \Re[c_1] & \cdots & \Re[c_Q] & \Im[c_1] & \cdots & \Im[c_Q] \end{bmatrix}^T,$$

$$\mathcal{N} = \text{vec}\left(\begin{bmatrix} \Re[\mathbf{n}_0 \cdots \mathbf{n}_{T-1}] \\ \Im[\mathbf{n}_0 \cdots \mathbf{n}_{T-1}] \end{bmatrix}\right). \tag{5.53}$$

Power normalization of basis matrices

As explained in the beginning of this chapter, we resort to normalize the transmitted codewords such that $\mathcal{E}\{\mathrm{Tr}\{\mathbf{CC}^H\}\} = T$. Assuming that $\mathcal{E}\{c_q\} = 0$ and $\mathcal{E}\{|c_q|^2\} = 1$, the basis matrices have therefore to satisfy the power constraint

$$\sum_{q=1}^{2Q} \mathrm{Tr}\{\mathbf{\Phi}_q\mathbf{\Phi}_q^H\} = 2T. \tag{5.54}$$

A stronger constraint consists in normalizing the average power of each basis matrices

$$\{\mathrm{Tr}\{\mathbf{\Phi}_q\mathbf{\Phi}_q^H\} = T/Q\}_{q=1}^{2Q}. \tag{5.55}$$

In the sequel, we will mostly deal with unitary basis matrices and make an extensive use of the following definitions.

Definition 5.6 *Tall $(T \leq n_t)$ unitary basis matrices are such that $\mathbf{\Phi}_q^H\mathbf{\Phi}_q = \frac{1}{Q}\mathbf{I}_T$ $\forall q = 1, \ldots, 2Q$. Wide $(T \geq n_t)$ unitary basis matrices are such that $\mathbf{\Phi}_q\mathbf{\Phi}_q^H = \frac{T}{Q n_t}\mathbf{I}_{n_t}$ $\forall q = 1, \ldots, 2Q$.*

Average pairwise error probability of STBCs

Investigating how the structure of the basis matrices of STBCs impacts the error rate performance provides some information about how to design linear STBCs in order to minimize the average error probability.

Proposition 5.3 *A PSK/QAM-based linear STBC consisting of unitary basis matrices minimizes the worst-case PEP in (5.12) averaged over i.i.d. Rayleigh slow fading channels if (sufficient condition) the unitary basis matrices $\{\mathbf{\Phi}_q\}_{q=1}^{2Q}$ satisfy the conditions*

$$\mathbf{\Phi}_q\mathbf{\Phi}_p^H + \mathbf{\Phi}_p\mathbf{\Phi}_q^H = \mathbf{0}_{n_t}, \quad q \neq p \text{ for wide } \{\mathbf{\Phi}_q\}_{q=1}^{2Q},$$

$$\mathbf{\Phi}_q^H\mathbf{\Phi}_p + \mathbf{\Phi}_p^H\mathbf{\Phi}_q = \mathbf{0}_T, \quad q \neq p \text{ for tall } \{\mathbf{\Phi}_q\}_{q=1}^{2Q}. \tag{5.56}$$

Those conditions are sufficient and necessary in order to minimize the Chernoff upper bound of the union bound in (5.4) averaged over i.i.d. Rayleigh slow fading channels.

PROOF: Using Hadamard's inequality (see Appendix A), we write

$$\det(\mathbf{I}_{n_t} + \eta\tilde{\mathbf{E}}) \leq \det(\mathbf{I}_{n_t} + \eta\mathbf{I}_{n_t} \odot \tilde{\mathbf{E}}). \tag{5.57}$$

Denoting $d_q \triangleq c_q - e_q$ with e_q the complex symbols of the codeword $\mathbf{E}$, the diagonal elements of $\tilde{\mathbf{E}}$ are expressed as

$$\tilde{\mathbf{E}}(n, n) = \left\| \sum_{q=1}^{Q} \mathbf{\Phi}_q(n, :)\Re[d_q] + \mathbf{\Phi}_{q+Q}(n, :)\Im[d_q] \right\|^2$$

$$\leq \left[\sum_{q=1}^{Q} \|\mathbf{\Phi}_q(n, :)\Re[d_q] + \mathbf{\Phi}_{q+Q}(n, :)\Im[d_q]\| \right]^2. \tag{5.58}$$

Since

$$
\begin{aligned}
\|\mathbf{\Phi}_q(n,:)\Re[d_q] + \mathbf{\Phi}_{q+Q}(n,:)\Im[d_q]\|^2 &= \left(\mathbf{\Phi}_q\mathbf{\Phi}_q^H\right)(n,n)|\Re[d_q]|^2 \\
&+ \left(\mathbf{\Phi}_{q+Q}\mathbf{\Phi}_{q+Q}^H\right)(n,n)\left|\Im[d_q]\right|^2 + \left(\mathbf{\Phi}_q\mathbf{\Phi}_{q+Q}^H + \mathbf{\Phi}_{q+Q}\mathbf{\Phi}_q^H\right)(n,n)\Re[d_q]\Im[d_q], \quad (5.59)
\end{aligned}
$$

using (5.57) and the unitarity of the wide basis matrices, we have

$$
\min_{q=1,\dots,Q} \min_{d_q} \det\left(\mathbf{I}_{n_t} + \eta\tilde{\mathbf{E}}\right)
$$

$$
\leq \min_{q=1,\dots,Q} \min_{d_q} \det\Big(\mathbf{I}_{n_t} + \zeta\Big[|\Re[d_q]|^2 + |\Im[d_q]|^2\Big]
$$

$$
+ \eta\big(\mathbf{I}_{n_t} \odot \big(\mathbf{\Phi}_q\mathbf{\Phi}_{q+Q}^H + \mathbf{\Phi}_{q+Q}\mathbf{\Phi}_q^H\big)\big)\Re[d_q]\Im[d_q]\Big)
$$

$$
\overset{(a)}{\leq} \det\Big(\mathbf{I}_{n_t} + \zeta\mathbf{I}_{n_t} \min_{d_q}\Big[|\Re[d_q]|^2 + |\Im[d_q]|^2\Big]\Big)
$$

$$
= \det\big(\mathbf{I}_{n_t} + \zeta\mathbf{I}_{n_t} d_{min}^2\big) = \big(1 + \zeta d_{min}^2\big)^{n_t} \qquad (5.60)
$$

where $\zeta = \eta\frac{T}{Qn_t}$. In (a) we use the fact that for QAM or PSK modulations, $\Re[d_q]$ and $\Im[d_q]$ can be positive or negative irrespective of $\mathbf{I}_{n_t} \odot \big(\mathbf{\Phi}_q\mathbf{\Phi}_{q+Q}^H + \mathbf{\Phi}_{q+Q}\mathbf{\Phi}_q^H\big)$. Expanding $\tilde{\mathbf{E}}$ using (5.51), it is clear that a sufficient condition to achieve (5.60) is that the unitary basis matrices be skew-hermitian $\mathbf{\Phi}_q\mathbf{\Phi}_l^H + \mathbf{\Phi}_l\mathbf{\Phi}_q^H = 0$ for $q \neq l$. It is also a necessary condition to minimize the union bound [San02].

Equivalently, for tall basis matrices, using the unitarity of the tall basis matrices and applying Hadamard's inequality, we may express that

$$
\min_{q=1,\dots,Q} \min_{d_q} \det\left(\mathbf{I}_T + \eta\tilde{\mathbf{E}}^H\right) \leq \left(1 + \frac{\eta}{Q}d_{min}^2\right)^T. \qquad (5.61)
$$

The equality in (5.60) is obtained if $\mathbf{\Phi}_q^H\mathbf{\Phi}_l + \mathbf{\Phi}_l^H\mathbf{\Phi}_q = 0$ for $q \neq l$. ∎

Proposition 5.4 *At asymptotically low SNR or for asymptotically large n_r in i.i.d. Rayleigh fading channels, a PSK/QAM-based linear STBC minimizes the worst-case PEP if (sufficient conditions) the basis matrices $\{\mathbf{\Phi}_q\}_{q=1}^{2Q}$ satisfy the conditions*

$$
\mathrm{Tr}\big\{\mathbf{\Phi}_q\mathbf{\Phi}_p^H + \mathbf{\Phi}_p\mathbf{\Phi}_q^H\big\} = 0, \quad q \neq p \qquad (5.62)
$$

or equivalently

$$
\mathcal{X}^T\mathcal{X} = \frac{T}{Q}\mathbf{I}_{2Q}. \qquad (5.63)
$$

Those conditions are sufficient and necessary in order to minimize the Chernoff upper bound of the union bound in (5.4) averaged over i.i.d. Rayleigh channels under the same asymptotic conditions.

PROOF: Observe that

$$
\min_{\substack{\mathbf{C},\mathbf{E}\\ \mathbf{C}\neq\mathbf{E}}} \|\mathbf{C}-\mathbf{E}\|_F^2 \;\leq\; \min_{q=1,\dots,Q}\;\min_{d_q}\left[\sum_{q=1}^{Q}\|(\mathbf{\Phi}_q\Re[d_q]+\mathbf{\Phi}_{q+Q}\Im[d_q])\|_F\right]^2
$$

$$
\overset{(a)}{\leq}\;\min_{q=1,\dots,Q}\;\min_{d_q}\frac{T}{Q}\left[|\Re[d_q]|^2+|\Im[d_q]|^2\right]
$$

$$
+\mathrm{Tr}\left\{\mathbf{\Phi}_q\mathbf{\Phi}_{q+Q}^H+\mathbf{\Phi}_{q+Q}\mathbf{\Phi}_q^H\right\}\Re[d_q]\Im[d_q]
$$

$$
\overset{(b)}{\leq}\;\frac{T}{Q}\min_{d_q}\left[|\Re[d_q]|^2+|\Im[d_q]|^2\right]
$$

$$
=\;\frac{T}{Q}d_{\min}^2 \tag{5.64}
$$

where in (a) we exploit the fact that $\left\{\mathrm{Tr}\{\mathbf{\Phi}_q\mathbf{\Phi}_q^H\}=T/Q\right\}_{q=1}^{2Q}$ and in (b) we use the fact that in PSK or QAM, $\Re[d_q]$ and $\Im[d_q]$ can be positive or negative independently of $\mathrm{Tr}\{\mathbf{\Phi}_q\mathbf{\Phi}_{q+Q}^H+\mathbf{\Phi}_{q+Q}\mathbf{\Phi}_q^H\}$. Expanding $\tilde{\mathbf{E}}$ via (5.51) and equalling $\min_{\substack{\mathbf{C},\mathbf{E}\\ \mathbf{C}\neq\mathbf{E}}}\mathrm{Tr}\{\tilde{\mathbf{E}}\}$ to (5.64) lead to the sufficient conditions in (5.62). The necessary conditions to minimize the union bound result from [San02]. ∎

Example 5.3   *A code such that $T=1$, $n_t=2$, $Q=2$ with the following basis matrices*

$$
\mathbf{\Phi}_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\end{bmatrix},\; \mathbf{\Phi}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\end{bmatrix},\; \mathbf{\Phi}_3=\frac{1}{\sqrt{2}}\begin{bmatrix}j\\0\end{bmatrix},\; \mathbf{\Phi}_4=\frac{1}{\sqrt{2}}\begin{bmatrix}0\\j\end{bmatrix}, \tag{5.65}
$$

or equivalently, with the following matrix $\mathcal{X}$

$$
\mathcal{X}=\frac{1}{\sqrt{2}}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} \tag{5.66}
$$

is an example of a linear STBC based on tall unitary basis matrices which satisfy Proposition 5.3.

Example 5.4   *A code such that $T=2$, $n_t=2$, $Q=2$ with the following basis matrices*

$$
\mathbf{\Phi}_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1&0\\0&1\end{bmatrix}, \qquad\qquad \mathbf{\Phi}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}0&-1\\1&0\end{bmatrix}, \tag{5.67}
$$

$$
\mathbf{\Phi}_3=\frac{1}{\sqrt{2}}\begin{bmatrix}j&0\\0&-j\end{bmatrix}, \qquad\qquad \mathbf{\Phi}_4=\frac{1}{\sqrt{2}}\begin{bmatrix}0&j\\j&0\end{bmatrix}, \tag{5.68}
$$

or equivalently, with the following matrix $\mathcal{X}$

$$
\mathcal{X} = \frac{1}{\sqrt{2}}
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & -1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & -1 & 0
\end{bmatrix}
\tag{5.69}
$$

is an example of a linear STBC based on wide unitary basis matrices which satisfy Proposition 5.3.

Ergodic capacity in i.i.d. Rayleigh fading channels

Achieving the ergodic capacity of the MIMO i.i.d. channel using a capacity-efficient linear STBC requires correctly designing the basis matrices and the input covariance matrix. By capacity-efficient, we mean a code that maximizes the average mutual information in the sense that it preserves the capacity without inducing any loss of capacity if linked with capacity-achieving outer codes such as turbo or LDPC codes. Naturally this would require to code with outer coding across several STBC blocks, each block experiencing different channel realizations.

The expression of the ergodic MIMO channel capacity in (5.52) for capacity optimal linear STBCs is then given by

$$
\bar{C} = \max_{\text{Tr}\{\mathcal{X}\mathcal{X}^T\} \leq 2T} \frac{1}{2T} \mathcal{E}_{\mathcal{H}} \left\{ \log \det \left(\mathbf{I}_{2n_r T} + \frac{\rho}{2} \mathcal{H} \mathcal{X} \mathcal{X}^T \mathcal{H}^T \right) \right\}
\tag{5.70}
$$

where it is assumed without loss of generality that $\mathcal{E}\{SS^H\} = \mathbf{I}_{2Q}$ [HP02, San02].

Proposition 5.5 *In i.i.d. Rayleigh fading channels, linear STBCs with wide $(Q \geq n_t T)$ matrices $\mathcal{X}$ that satisfy*

$$
\mathcal{X}\mathcal{X}^T = \frac{1}{n_t} \mathbf{I}_{2n_t T}
\tag{5.71}
$$

are capacity-efficient.

PROOF: This results from [Tel95] which shows that the optimal covariance matrix over i.i.d. Rayleigh fading channels is proportional to the identity matrix. ∎

Proposition 5.5 provides a necessary and sufficient condition for linear STBCs to be capacity-efficient. In order to satisfy (5.71), it is necessary to have $r_s = n_t$, confirming

that sending n_t independent data streams achieves the ergodic capacity of the fast fading MIMO channel as explained in Section 5.4.1. It is important to note that there is an infinity of codes satisfying the condition shown in (5.71).

Example 5.5   *The code given by the basis matrices displayed in Example 5.3 is capacity-efficient. However, the code of Example 5.4 is not capacity-efficient.*

From (5.63), we get this interesting result that if $Q = n_t T$, a capacity-efficient linear STBC becomes optimal under an error probability perspective at low SNR or as the number of receive antennas increases. Moreover, Jensen's inequality leads to

$$\bar{C} \leq \max_{\text{Tr}\{\mathcal{X}\mathcal{X}^T\} \leq 2T} \frac{1}{2T} \log \det\left(\mathbf{I}_{2n_r T} + \frac{\rho}{2}\mathcal{E}_\mathcal{H}\left\{\mathcal{H}\mathcal{X}\mathcal{X}^T\mathcal{H}^T\right\}\right)$$
$$= \max_{\text{Tr}\{\mathcal{X}\mathcal{X}^T\} \leq 2T} \frac{1}{2T} \log \det\left(\mathbf{I}_{2n_t T} + \frac{\rho\, n_r}{2}\mathcal{X}^T\mathcal{X}\right), \tag{5.72}$$

indicating that the gap between the ergodic capacity and this last upper bound decreases as the number of receive antennas increases. A code that is asymptotically optimal from the error probability point of view is also asymptotically optimal capacity-wise. We thus obtain the following result.

Proposition 5.6   *Asymptotically (at low SNR or with large n_r), a linear STBC designed such that*

$$\mathcal{X}^T\mathcal{X} = \frac{T}{Q}\mathbf{I}_{2Q} \tag{5.73}$$

is capacity-efficient and minimizes the worst-case PEP and the Chernoff bound of the union bound in i.i.d. Rayleigh fading channels.

A tall matrix $\mathcal{X}$ that satisfies (5.73) is known as a tight frame. An interesting property [HP02] is that the mutual information is also lower-bounded with frame-based codes.

Decoding
The complexity of maximum likelihood (ML) decoding of linear STBCs grows exponentially with n_t and Q. It is thus useful to investigate whether the basis matrices could not be adequately chosen in order to reduce the complexity of the ML receiver. The next proposition gives necessary and sufficient conditions to decouple the transmit symbols via a space–time matched filter. Under those conditions, each symbol can be decoded independently of the presence of the other symbols while still ensuring an optimal performance.

Proposition 5.7 *Applying the space–time matched filter $\mathcal{X}^H \mathcal{H}^H$ to the output vector $\mathcal{Y}$ decouples the transmitted symbols*

$$\mathcal{X}^H \mathcal{H}^H \mathcal{Y} = \frac{\sqrt{E_s}T}{Qn_t} \|\mathbf{H}\|_F^2 \mathbf{I}_{2T} \mathcal{S} + \mathcal{X}^H \mathcal{H}^H \mathcal{N} \tag{5.74}$$

if and only if the basis matrices are wide unitary

$$\mathbf{\Phi}_q \mathbf{\Phi}_q^H = \frac{T}{Qn_t} \mathbf{I}_{n_t}, \ \forall q = 1, \dots, 2Q \tag{5.75}$$

and pairwise skew-hermitian

$$\mathbf{\Phi}_q \mathbf{\Phi}_p^H + \mathbf{\Phi}_p \mathbf{\Phi}_q^H = \mathbf{0}_{n_t}, \ \forall q \neq p. \tag{5.76}$$

PROOF: Let us express the (k, l) element of $\mathcal{X}^H \mathcal{H}^H \mathcal{H} \mathcal{X}$, denoted as $\mathcal{X}^H \mathcal{H}^H \mathcal{H} \mathcal{X}(k, l)$. Using simple matrix manipulations, we write

$$\mathcal{X}^H \mathcal{H}^H \mathcal{H} \mathcal{X}(k, l) = \sum_{m=1}^{T} \Re\left[(\mathbf{\Phi}_k(\,:,m))^H \mathbf{H}^H \mathbf{H} \mathbf{\Phi}_l(\,:,m) \right] \tag{5.77}$$

$$= \frac{1}{2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} \mathcal{H}^H \mathcal{H}(i,j) \big(\mathbf{\Phi}_k \mathbf{\Phi}_l^H + \mathbf{\Phi}_l \mathbf{\Phi}_k^H \big)(j, i) \tag{5.78}$$

In order to decouple the symbols, $\mathcal{X}^H \mathcal{H}^H \mathcal{H} \mathcal{X}(k, l)$ must be equal to zero for $k \neq l$. Since this should be valid for all $\mathcal{H}$, it is necessary and sufficient that $\mathbf{\Phi}_k \mathbf{\Phi}_l^H + \mathbf{\Phi}_l \mathbf{\Phi}_k^H = 0$ for $k \neq l$. Furthermore, to ensure that the diagonal elements have a magnitude equal to $\frac{T}{Qn_t} \|\mathbf{H}\|_F^2$, the basis matrices must be wide unitary. ∎

Example 5.6 *The code given by the basis matrices displayed in Example 5.3 does not allow for decoupling. However, the code of Example 5.4 is very easily decoded thanks to the decoupling property.*

Codes which do not decouple the received streams may lead to a very large receiver complexity if ML decoding is used. In order to reduce this complexity, one solution is to exploit the linearity of the space–time block codes and to propose suboptimal decoders based on ZF, MMSE and successive interference cancelers (SIC) to replace the ML decoder. However, those decoders may induce a considerable performance degradation (in terms of diversity and coding gains). Therefore, efficient algorithms with low complexity but high performance have been developed, e.g. the popular sphere decoding technique. We deal with all those decoders in the next sections. Note that while we mainly discuss the performance of such decoders for Spatial Multiplexing, the reader should keep in mind that those decoders may also be used with any linear space–time block code, using the equivalent MIMO representation of (5.52).

5.5.2   Spatial Multiplexing/V-BLAST

Spatial Multiplexing (SM), also called V-BLAST, is a full rate code ($r_s = n_t$) that consists in transmitting independent data streams on each transmit antenna. Those data streams may be independently coded or left uncoded.

In uncoded transmissions, each codeword $\mathbf{C}$ expands on only one symbol duration ($T = 1$) and is thus written as a vector symbol of size $n_t \times 1$. Using results of the previous section, we know that a code with tall basis matrices such as SM minimizes the maximum average PEP if $\boldsymbol{\Phi}_q^H \boldsymbol{\Phi}_p + \boldsymbol{\Phi}_p^H \boldsymbol{\Phi}_q = 0$ for $q \neq p$. Since $T = 1$, this becomes $\mathrm{Tr}\left\{\boldsymbol{\Phi}_q \boldsymbol{\Phi}_p^H + \boldsymbol{\Phi}_p \boldsymbol{\Phi}_q^H\right\} = 0$ for $q \neq p$ or equivalently $\mathcal{X}^T \mathcal{X} = \frac{1}{n_t}\mathbf{I}_{2n_t}$. Since n_t independent symbols are transmitted, the matrix $\mathcal{X}$ is square and codes transmitting n_t independent symbols over 1 symbol duration and satisfying $\mathcal{X}^T \mathcal{X} = \frac{1}{n_t}\mathbf{I}_{2n_t}$ are capacity-efficient and minimize the worst-case average error probability (Propositions 5.3 and 5.5). This is summarized in the following Proposition.

Proposition 5.8   *Spatial Multiplexing with basis matrices characterized by square $\mathcal{X}$ such that*

$$\mathcal{X}^T \mathcal{X} = \frac{1}{n_t}\mathbf{I}_{2n_t} \tag{5.79}$$

is capacity-efficient (Proposition 5.5) and optimal from an error rate minimization perspective (Proposition 5.3).

Spatial Multiplexing can thus be constructed from one of those matrices $\mathcal{X}$.

Example 5.7   *The simplest code is obtained as*

$$\mathbf{C} = \frac{1}{\sqrt{n_t}}\begin{bmatrix} c_1 & \ldots & c_{n_t} \end{bmatrix}^T = \frac{1}{\sqrt{n_t}}\sum_{q=1}^{n_t}\mathbf{I}_{n_t}(:,q)\Re[c_q] + j\mathbf{I}_{n_t}(:,q)\Im[c_q]. \tag{5.80}$$

Each element c_q is a symbol chosen from a given constellation. Our notation $\mathbf{I}_{n_t}(:,q)$ stands for the qth column of the identity matrix $\mathbf{I}_{n_t}$. For $n_t = 2$ this code has already been presented in Example 5.3.

From Proposition 5.7, it is clear that SM does not decouple the streams at the receiver. Since the complexity of ML decoding is quite high, it is of practical interest to look for suboptimal receivers. In the sequel, we investigate several commonly used decoders in the context of SM schemes: ML and sphere decoders, linear receivers based on ZF and MMSE filtering and successive interference cancelers. For each of them, we discuss their

average PEP performance, their achievable maximum rate over fast fading channels and how well they achieve the diversity-multiplexing trade-off.

ML decoding

At each time instant, the receiver uses ML decoding in order to detect which vector symbol was effectively transmitted. Since the transmitted codewords expand on one symbol duration, the rank of the error matrix is equal to one and so is the achievable transmit diversity. From (5.12), the exact PEP is evaluated as follows

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \left(1 + \frac{\eta}{n_t} \sum_{q=1}^{Q} |c_q - e_q|^2 \right)^{-n_r} d\beta \tag{5.81}$$

$$\stackrel{(a)}{=} \frac{1}{2} \left[1 - \sqrt{\frac{\rho_s}{1 + \rho_s}} \sum_{i=0}^{n_r-1} \binom{2i}{i} \left(\frac{1}{4(1 + \rho_s)} \right)^i \right] \tag{5.82}$$

where $\rho_s \triangleq \frac{\rho}{4n_t} \sum_{q=1}^{Q} |c_q - e_q|^2$. In (a), the integral was evaluated thanks to the results of [Sim01]. The Chernoff upper bound of the average PEP at high SNR can be expressed as

$$P(\mathbf{C} \to \mathbf{E}) \leq \left(\frac{\rho}{4n_t} \right)^{-n_r} \left(\sum_{q=1}^{Q} |c_q - e_q|^2 \right)^{-n_r}. \tag{5.83}$$

The SNR exponent is equal to n_r. Due to the lack of coding across transmit antennas, no transmit diversity is achieved and the only receive diversity may be exploited.

Over fast fading channels, we know that it is not necessary to code across antennas to achieve the ergodic capacity. Indeed, transmitting n_t independent streams with uniform power allocation (each stream being coded with a capacity-achieving code over a very large number of channel realizations) and applying joint decoding at the receiver constitute a sufficient strategy to achieve the ergodic capacity. This leads to the following result.

Proposition 5.9 *Spatial Multiplexing with ML decoding and equal power allocation achieves the ergodic capacity of i.i.d. Rayleigh fast fading channels.*

Note that it has already been mentioned that the basis matrices satisfy Proposition 5.5.

Let us now investigate how ML-decoded Spatial Multiplexing with QAM symbols performs in terms of the multiplexing-diversity trade-off over i.i.d. Rayleigh slow fading channels.

Proposition 5.10 *For $n_r \geq n_t$, the diversity-multiplexing trade-off achieved by Spatial Multiplexing with ML decoding and QAM constellation over i.i.d. Rayleigh fading channels is given by*

$$g_d(g_s) = n_r \left(1 - \frac{g_s}{n_t} \right), \quad g_s \in [0, n_t]. \tag{5.84}$$

PROOF: Let us denote by R the total data rate achieved by the scheme. Assuming that the same data rate $R_q = R/n_t$ is assigned to all substreams, symbols c_q are chosen from a QAM constellation carrying $2^{R_q/2}$ symbols per dimension. For a unit average energy, the minimum distance between two points of such a constellation is of the order $1/2^{R_q/2}$. Using the union bound in (5.4), we assume without loss of generality that $\mathbf{E} = \mathbf{0}$ and bound the overall error probability as

$$P_e \leq \sum_{\mathbf{C} \neq \mathbf{0}} P(\mathbf{C} \rightarrow \mathbf{0}) \leq \sum_{\mathbf{C} \neq \mathbf{0}} \left(\frac{\rho}{4n_t} \right)^{-n_r} \left(\sum_{q=1}^{Q} |c_q|^2 \right)^{-n_r} \tag{5.85}$$

which is based on (5.83). Denoting as d_{min} the minimum Euclidean distance of the QAM constellation, we write $c_q = (a_q + jb_q) d_{min}$ where $a_q, b_q \in \mathbb{Z}$. Since the rate on each substream increases, such as $R_q = g_s/n_t \log_2(\rho)$, the minimum squared distance of the constellation is of the order ρ^{-g_s/n_t} and we can further bound P_e as

$$P_e \dot{\leq} \sum_{\mathbf{C} \neq \mathbf{0}} \left(\frac{\rho}{4n_t} \right)^{-n_r} \frac{1}{\left(\sum_{q=1}^{n_t} |a_q|^2 + |b_q|^2 \right)^{n_r}} \rho^{n_r \frac{g_s}{n_t}}$$

$$\dot{\leq} (4n_t)^{n_r} \rho^{-n_r \left(1 - \frac{g_s}{n_t} \right)} \sum_{\mathbf{C} \neq \mathbf{0}} \frac{1}{\left(\sum_{q=1}^{n_t} |a_q|^2 + |b_q|^2 \right)^{n_r}}. \tag{5.86}$$

It can be shown [TV06] that as the SNR increases and $n_r \geq n_t$,

$$\sum_{\{a_1, \ldots, a_{n_t}, b_1, \ldots, b_{n_t}\} \neq \mathbf{0}} \frac{1}{\left(\sum_{q=1}^{n_t} |a_q|^2 + |b_q|^2 \right)^{n_r}} \doteq 1, \tag{5.87}$$

which leads to the announced result.   ∎

When $n_r < n_t$ it is more complicated to get an exact estimation of the achieved diversity-multiplexing trade-off. The previous result points to the interesting fact that the dominant error event does not come from inter-stream interference but from the channel being in outage. Indeed, this trade-off is the same as if there were no other streams. The resulting trade-off is always suboptimal compared to the optimal MIMO diversity-multiplexing trade-off. SM with ML decoding and QAM constellation achieves a part of this optimal trade-off only if $n_t = n_r = n$. Indeed it achieves the last segment of the optimal MIMO diversity-multiplexing trade-off curve between points $(n-1, 1)$ and $(n, 0)$. However, it does not satisfy the non-vanishing determinant design criterion.

Sphere decoder

ML decoding leads to the best performance in terms of error rate but it is extremely demanding in terms of complexity. For a constellation of size M, ML decoding involves searching over M^{n_t} possible candidates. This is affordable when M and n_t are small, but not for large spectral efficiency systems. The increasing complexity is caused by the search
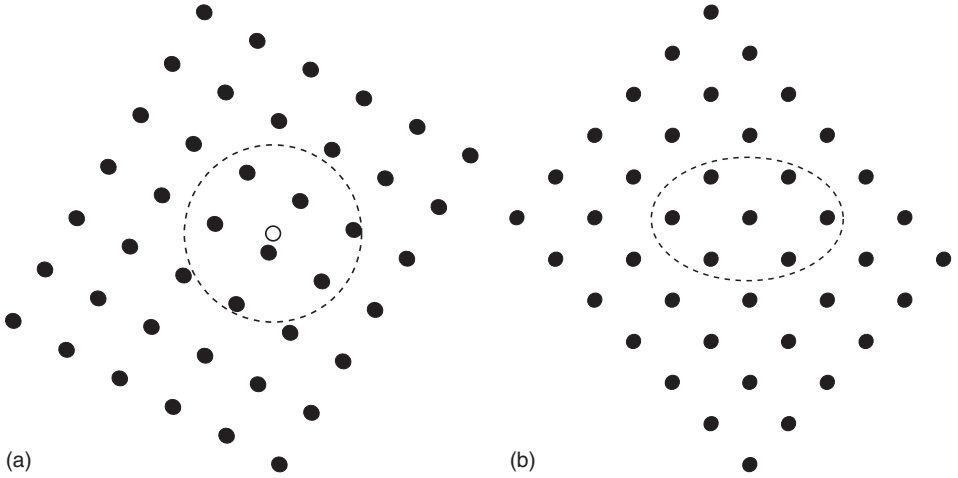
Figure 5.2   *(a) The sphere is centered at the received vector and contains the lattice points to be enumerated; (b) The sphere is transformed into an ellipsoid in the $\mathcal{T}$ coordinate system*

over all possible combinations, although many of them are most probably not the correct candidate: owing to the Gaussian distribution of noise, codewords that are far away from the received vector are much less probable than codewords close to the received vector.

The sphere decoder algorithm was originally developed in the 1980s [Poh81, FP85] but has recently attracted much attention in the MIMO community thanks to its similar performance to the ML decoder at a reasonable complexity [VB99, DCB00]. The main idea is to limit the search among the possible candidates to those located within a sphere of radius σ centered on the received vector (see Figure 5.2(a)). Hence, the question becomes how to identify the candidates inside the sphere. If this involved calculating the distance between each candidate and the received vector and comparing it to σ, the sphere decoder would not be very efficient. Fortunately, the sphere decoder proposes an efficient alternative. Note also that other algorithms where the region of interest is not spherical (e.g. a rectangular parallelogram) have been proposed; however, the sphere decoder remains the most popular.

The algorithm works as follows [VB99]. First express the classical complex-valued MIMO model in (5.1) as an equivalent real-valued model in (5.52)

$$\mathcal{Y} = \mathcal{H}_{eq}\mathcal{S} + \mathcal{N} \tag{5.88}$$

where $\mathcal{H}_{eq} = \mathcal{H}\mathcal{X}$ is an equivalent MIMO channel matrix. For SM, $\mathcal{X}$ is proportional to an identity matrix but the following development is valid for a general linear STBC. Assuming QAM constellations, the elements of $\mathcal{S} = \left[\Re[c_1] \cdots \Re[c_Q] \Im[c_1] \cdots \Im[c_Q]\right]^T$ are picked up in the set of integers $\mathbb{Z}$. A lattice is a discrete set of points expressed as a linear combination of basis elements with integer weights. The set of points $\Lambda = \left\{\mathcal{W} = \mathcal{H}_{eq}\mathcal{S}, \mathcal{S} \in \mathbb{Z}^{2Q}\right\}$ can be seen as a lattice of dimension $2Q$ where $\mathcal{H}_{eq}$ is the generator matrix of the lattice and the basis of the lattice is given by any basis of $\mathcal{H}_{eq}$. The problem in (5.2) consists thus in

finding the lattice point $\mathcal{W}$ which is the closest to the received vector $\mathcal{Y}$

$$\min_{\mathcal{W} \in \Lambda} \|\mathcal{Y} - \mathcal{W}\|^2 = \min_{\mathcal{V} \in \mathcal{Y} - \Lambda} \|\mathcal{V}\|^2 \qquad (5.89)$$

i.e. finding the vector $\mathcal{V}$ with the shortest norm in the translated lattice $\mathcal{Y} - \Lambda$. We assume that $2Q = 2n_r T$ (for SM, it comes to have $n_r = n_t$) and that $\mathcal{H}_{eq}$ is non singular. This is expressed as $\mathcal{Y} = \mathcal{H}_{eq}\mathcal{R}$, or equivalently $\mathcal{R} = \mathcal{H}_{eq}^{-1}\mathcal{Y}$ with the coordinates $\mathcal{R} \in \mathbb{R}^{2Q}$. Hence, we obtain points in the translated lattice as $\mathcal{V} = \mathcal{H}_{eq}\mathcal{U}$ with $\mathcal{U} = \mathcal{R} - \mathcal{S} \in \mathbb{R}^{2Q}$ the coordinates in the translated lattice.

We now come back to the identification of codewords (lattice points), which lie within the sphere. A lattice point lies in the sphere of radius σ centered on the received vector $\mathcal{Y}$ if $\|\mathcal{V}\|^2 = \mathcal{U}^T \mathcal{G} \mathcal{U} \leq \sigma^2$ with $\mathcal{G} = \mathcal{H}_{eq}^T \mathcal{H}_{eq}$. A Cholesky factorization of the matrix $\mathcal{G}$ yields $\mathcal{G} = \mathcal{L}^T \mathcal{L}$ with $\mathcal{L}$ an upper triangular matrix. Hence, we have

$$\|\mathcal{V}\|^2 = \|\mathcal{L}\mathcal{U}\|^2 = \sum_{i=1}^{2Q} \left| \mathcal{L}(i,i)\mathcal{U}(i) + \sum_{j=i+1}^{2Q} \mathcal{L}(i,j)\mathcal{U}(j) \right|^2 \qquad (5.90)$$

$$= \sum_{i=1}^{2Q} \mathcal{Q}(i,i)|\mathcal{T}(i)|^2 \leq \sigma^2 \qquad (5.91)$$

where we define a new coordinate system

$$\mathcal{T}(i) = \mathcal{U}(i) + \sum_{j=i+1}^{2Q} \mathcal{Q}(i,j)\mathcal{U}(j) \qquad (5.92)$$

by substituting $\mathcal{Q}(i,i) = |\mathcal{L}(i,i)|^2$ for $i = 1, \ldots, 2Q$ and $\mathcal{Q}(i,j) = \mathcal{L}(i,j)/\mathcal{L}(i,i)$ for $i = 1, \ldots, 2Q$, $j = i+1, \ldots, 2Q$. In this new coordinate system, $\mathcal{T}(i)$ defines an ellipsoid. We are interested by the lattice points contained within this ellipsoid (Figure 5.2(b)). For $i = 2Q$ down to $i = 1$, we find the limits of the coordinates $\mathcal{T}(i)$ as

$$-\sqrt{\frac{\sigma^2}{\mathcal{Q}(2Q,2Q)}} \quad \leq \quad \mathcal{T}(2Q) \quad \leq \quad \sqrt{\frac{\sigma^2}{\mathcal{Q}(2Q,2Q)}}$$

$$-\sqrt{\frac{\sigma^2 - \mathcal{Q}(2Q,2Q)|\mathcal{T}(2Q)|^2}{\mathcal{Q}(2Q-1,2Q-1)}} \leq \mathcal{T}(2Q-1) \leq \sqrt{\frac{\sigma^2 - \mathcal{Q}(2Q,2Q)|\mathcal{T}(2Q)|^2}{\mathcal{Q}(2Q-1,2Q-1)}} \qquad (5.93)$$

$$\vdots$$

Using (5.92) and the relationships $\mathcal{U}(i) = \mathcal{R}(i) - \mathcal{S}(i)$ for $i = 1, \ldots, 2Q$, we finally obtain the ranges for the integers $\mathcal{S}(i)$:

- for $\mathcal{S}(2Q)$

$$\left\lceil -\sqrt{\frac{\sigma^2}{\mathcal{Q}(2Q,2Q)}} + \mathcal{R}(2Q) \right\rceil \leq \mathcal{S}(2Q) \leq \left\lfloor \sqrt{\frac{\sigma^2}{\mathcal{Q}(2Q,2Q)}} + \mathcal{R}(2Q) \right\rfloor \qquad (5.94)$$

- for $\mathcal{S}(2Q - 1)$

$$\lceil -A_{(2Q-1)} + B_{(2Q-1)} \rceil \leq \mathcal{S}(2Q - 1) \leq \lfloor A_{(2Q-1)} + B_{(2Q-1)} \rfloor \qquad (5.95)$$

where

$$A_{(2Q-1)} = \sqrt{\frac{\sigma^2 - \mathcal{Q}(2Q, 2Q)|\mathcal{U}(2Q)|^2}{\mathcal{Q}(2Q - 1, 2Q - 1)}} \qquad (5.96)$$

$$B_{(2Q-1)} = \mathcal{R}(2Q - 1) + \mathcal{Q}(2Q - 1, 2Q)\mathcal{U}(2Q) \qquad (5.97)$$

- for the ith integer $\mathcal{S}(i)$

$$\lceil -A_{(i)} + B_{(i)} \rceil \leq \mathcal{S}(i) \leq \lfloor A_{(i)} + B_{(i)} \rfloor \qquad (5.98)$$

where

$$A_{(i)} = \sqrt{\frac{1}{\mathcal{Q}(i, i)} \left(\sigma^2 - \sum_{l=i+1}^{2Q} \mathcal{Q}(l, l) \left| \mathcal{U}(l) + \sum_{j=l+1}^{2Q} \mathcal{Q}(l, j)\mathcal{U}(j) \right|^2 \right)} \qquad (5.99)$$

$$B_{(i)} = \mathcal{R}(i) + \sum_{j=i+1}^{2Q} \mathcal{Q}(i, j)\mathcal{U}(j) \qquad (5.100)$$

with $\lceil x \rceil$ the smallest integer greater than x and $\lfloor x \rfloor$ the greatest integer smaller than x. Among all codewords $\mathcal{S}$ contained within those limits, the sphere decoder selects the one that is the closest to the received vector.

Note that the radius σ of the sphere should be chosen adequately. On the one hand, a large σ results in a complexity that is exponential in size, analogous to ML decoding. On the other hand, choosing σ too small, we might end up without any candidate codeword within the sphere. In practice, σ is adjusted as a function of the SNR. In any case, the radius σ should always be larger than the covering radius of the lattice, as the latter is the smallest radius for which the entire space is covered by the spheres [OV04]. This guarantees the existence of at least one point in the sphere, irrespective of the position of the received vector $\mathcal{Y}$.

Zero-forcing (ZF) linear receiver

The MIMO version of the ZF receiver acts similarly to a ZF equalizer in frequency selective channels. The MIMO channel is inverted at the receiver in order to totally suppress the interference from other transmitted symbols. The output of the ZF filter is thus only a function of the symbol to be detected and the noise. It is then fed into a ML decoder which estimates the transmitted symbol. The complexity of ZF decoding is similar to SISO ML decoding, but the inversion step is responsible for the noise enhancement. Assuming that

a symbol vector $\mathbf{C} = 1/\sqrt{n_t}\begin{bmatrix} c_1 & \ldots & c_{n_t} \end{bmatrix}^T$ is transmitted, the output of the ZF filter $\mathbf{G}_{ZF}$ is given by

$$\mathbf{z}_k = \mathbf{G}_{ZF}\mathbf{y}_k = \begin{bmatrix} c_1 & \ldots & c_{n_t} \end{bmatrix}^T + \mathbf{G}_{ZF}\mathbf{n}_k \tag{5.101}$$

where $\mathbf{G}_{ZF}$ inverts the channel

$$\mathbf{G}_{ZF} = \sqrt{\frac{n_t}{E_s}}\mathbf{H}_k^\dagger \tag{5.102}$$

with $\mathbf{H}_k^\dagger = \left(\mathbf{H}_k^H\mathbf{H}_k\right)^{-1}\mathbf{H}_k^H$ denoting the Moore-Penrose pseudo inverse. ZF filtering effectively decouples the channel into n_t parallel channels so that a scalar decoding may be performed on each of these channels. The ZF filter also maximizes the SNR under the constraint that the interferences from all other layers are nulled out. Specifically, the qth layer is detected through a projection of the output vector onto the direction closest to $\mathbf{H}_k(:,q)$ within the subspace orthogonal to the one spanned by the set of vectors $\mathbf{H}_k(:,p)$, $p \neq q$. Obviously, if $\mathbf{H}_k(:,q)$ is naturally orthogonal to all other columns of $\mathbf{H}_k$, this direction is simply given by $\mathbf{H}_k(:,q)$ and zero-forcing reduces to simple matched filtering, as if there were no other layers.

The additive noise contributions on each channel are enhanced with respect to the original noise and correlated between parallel channels. The covariance matrix of the noise at the output of the ZF filter reads as

$$\mathcal{E}\left\{\mathbf{G}_{ZF}\mathbf{n}_k(\mathbf{G}_{ZF}\mathbf{n}_k)^H\right\} = \frac{n_t}{\rho}\mathbf{H}_k^\dagger \left(\mathbf{H}_k^\dagger\right)^H \tag{5.103}$$

$$= \frac{n_t}{\rho}\left(\mathbf{H}_k^H\mathbf{H}_k\right)^{-1}. \tag{5.104}$$

The output SNR on the qth subchannel is thus given by

$$\rho_q = \frac{\rho}{n_t}\frac{1}{\left(\mathbf{H}_k^H\mathbf{H}_k\right)^{-1}(q,q)}. \tag{5.105}$$

Assuming that the channel is i.i.d. Rayleigh distributed, ρ_q is a χ^2 random variable with $2(n_r - n_t + 1)$ degrees of freedom [PNG03], denoted as $\chi^2_{2(n_r-n_t+1)}$. The average PEP on the qth subchannel is thus upper-bounded by

$$P(c_q \to e_q) \leq \left(\frac{\rho}{2n_t}\right)^{-(n_r-n_t+1)}\left|c_q - e_q\right|^{-2(n_r-n_t+1)}. \tag{5.106}$$

The lower complexity of the ZF receiver comes at the price of a diversity gain limited to $n_r - n_t + 1$. Clearly, the system is undetermined if $n_t > n_r$.

If coding over time is performed to average out the fading, the average maximum achievable rate $\bar{C}_{ZF}$ is equal to the sum of the maximum rates achievable by all layers

$$\bar{C}_{ZF} = \sum_{q=1}^{\min\{n_t, n_r\}} \mathcal{E}\{\log_2(1 + \rho_q)\} \tag{5.107}$$

$$= \min\{n_t, n_r\} \, \mathcal{E}\left\{\log_2\left(1 + \frac{\rho}{n_t}\chi^2_{2(n_r-n_t+1)}\right)\right\}. \tag{5.108}$$

At high SNR, $\bar{C}_{ZF}$ is well approximated by

$$\bar{C}_{ZF} \approx \min\{n_t, n_r\}\log_2\left(\frac{\rho}{n_t}\right) + \min\{n_t, n_r\}\mathcal{E}\left\{\log_2\left(\chi^2_{2(n_r-n_t+1)}\right)\right\}. \tag{5.109}$$

It is worth noting that the first term of (5.109) is similar to the first term in the expression in (4.40) of the high-SNR ergodic channel capacity with CDIT. This suggests that Spatial Multiplexing, in combination with a ZF decoder, allows for transmitting over $n = \min\{n_t, n_r\}$ independent data pipes. However, the second term in (5.109) is smaller than the corresponding term in (4.40). This comes from the fact that each layer achieves at most a diversity of $n_r - n_t + 1$, instead of $n_r - n_t + q$. This shifts $\bar{C}_{ZF}$ towards higher SNR with regard to $\bar{C}_{CDIT}$.

Let us now obtain the diversity-multiplexing trade-off achieved by this scheme in slow fading channels.

Proposition 5.11 *For $n_r \geq n_t$, the diversity-multiplexing trade-off achieved by Spatial Multiplexing with QAM constellation and ZF filtering over i.i.d. Rayleigh fading channels is given by*

$$g_d(g_s) = (n_r - n_t + 1)\left(1 - \frac{g_s}{n_t}\right), \, g_s \in [0, n_t]. \tag{5.110}$$

PROOF: Since the ZF filter decouples the MIMO channel into n_t parallel SISO channels, the error probability is upper-bounded by the worst-case PEP multiplied by the number of nearest neighbors. In QAM constellations, there are at most four nearest neighbors, hence

$$P_e(R) \leq 4 \min_{c_q, e_q} P(c_q \to e_q) \leq 4\left(\frac{\rho}{2n_t}\right)^{-(n_r-n_t+1)} d_{min}^{-2(n_r-n_t+1)} \tag{5.111}$$

where d_{min} is the minimum Euclidean distance of the QAM constellation. For a rate that scales with SNR, $(R = g_s \log_2 \rho)$, $d_{min}^2 \doteq \frac{1}{2^{R/n_t}}$ and the error probability is upper-bounded by

$$P_e(R) \doteq 4\left(\frac{1}{2n_t}\right)^{-(n_r-n_t+1)} \rho^{-(n_r-n_t+1)\left(1-\frac{g_s}{n_t}\right)} \tag{5.112}$$

which yields the announced result. ∎

This is naturally suboptimal with respect to the ML receiver, especially when the number of receive antennas is close to the number of transmit antennas.

Minimum mean square error (MMSE) linear receiver
The ZF receiver eliminates the interference but enhances noise. This might not be significant at high SNR, but at low SNR, it is both sensible and practical to design a filter maximizing the global signal to noise plus interference ratio (SNIR). One possibility is to minimize the total resulting noise, i.e. to find $\mathbf{G}$ such that $\left\| \mathbf{G}\mathbf{y}_k - \begin{bmatrix} c_1 & \dots & c_{n_t} \end{bmatrix}^T \right\|^2$ is minimum. The solution to this problem is given by the classical MMSE filter

$$\mathbf{G}_{MMSE} = \sqrt{\frac{n_t}{E_s}} \left(\mathbf{H}_k^H \mathbf{H}_k + \frac{n_t}{\rho} \mathbf{I}_{n_t} \right)^{-1} \mathbf{H}_k^H. \tag{5.113}$$

It can be shown [PNG03] that the output SNR on the qth subchannel is equal to

$$\rho_{ij} = \frac{1}{\left(\dfrac{\rho}{n_t} \mathbf{H}^H \mathbf{H} + \mathbf{I}_{n_t} \right)^{-1} (q, q)}$$

At low SNR, MMSE filtering reduces to matched filtering as $\mathbf{G}_{MMSE} \approx \frac{\sqrt{E_s}}{\sqrt{n_t}\sigma_n^2} \mathbf{H}_k^H$. This reduces the noise enhancement and outperforms ZF filtering. It is optimal from a capacity perspective. Indeed, the interference is much smaller than the noise, and the effective channel seen by the qth layer is $\mathbf{y}_k \approx \mathbf{H}_k(:, q)c_q + n_k$. The maximum achievable rate is naturally achieved by matched filtering. At high SNR, the MMSE filter is equivalent to ZF and the diversity achievable is thus limited to $n_r - n_t + 1$.

(Ordered) successive interference canceler
The idea of the successive interference canceler (SIC) is to successively decode one symbol (or usually one layer) and cancel the effect of this symbol from the received signal. The core of the decoder consists in three steps, as detailed later. The order by which the symbols are decoded is based on the SINR of each symbol/layer: the symbol/layer with the highest SINR is decoded first at each iteration, which requires the SINR evaluation. This first method is generally known as ordered SIC. Non-ordered SIC is obtained by decoding symbols 1 up to n_t, irrespective of their SINR. Spatial Multiplexing with (ordered) SIC is generally known as V-BLAST [GFVW99], and ZF and MMSE V-BLAST refer to Spatial Multiplexing with respectively ZF-SIC and MMSE-SIC receivers. The V-BLAST algorithm is summarized below for an ordered ZF-SIC (its MMSE version is similar).

1. *Initialization:*

$$i \longleftarrow 1 \tag{5.114}$$

$$\mathbf{y}_k^{(1)} = \mathbf{y}_k \tag{5.115}$$

$$\mathbf{G}^{(1)} = \mathbf{G}_{ZF}(\mathbf{H}) \tag{5.116}$$

$$q_1 \overset{(*)}{=} \arg \min_j \left\| \mathbf{G}^{(1)}(j, :) \right\|^2 \tag{5.117}$$

where $\mathbf{G}_{ZF}(\mathbf{H})$ is defined as the ZF filter of the matrix $\mathbf{H}$ in (5.102).

2. *Recursion:*
   (a) *step 1:* extract the q_i^{th} transmitted symbol from the received signal $\mathbf{y}_k^{(i)}$

$$\tilde{c}_{q_i} = \mathbf{G}^{(i)}(q_i,:)\mathbf{y}_k^{(i)} \tag{5.118}$$

   where $\mathbf{G}^{(i)}(q_i,:)$ is the q_i^{th} row of $\mathbf{G}^{(i)}$;
   (b) *step 2:* slice $\tilde{c}_{q_i}$ to obtain the estimated transmitted symbol $\hat{c}_{q_i}$;
   (c) *step 3:* assume that $\hat{c}_{q_i} = c_{q_i}$ and construct the received signal

$$\mathbf{y}_k^{(i+1)} = \mathbf{y}_k^{(i)} - \hat{c}_{q_i}\mathbf{H}_k(:,q_i) \tag{5.119}$$

$$\mathbf{G}^{(i+1)} = \mathbf{G}_{ZF}(\mathbf{H}_{\bar{q}_i}) \tag{5.120}$$

$$i \longleftarrow i+1 \tag{5.121}$$

$$q_{i+1} \overset{(*)}{=} \arg\min_{j\notin\{q_1,\ldots,q_i\}} \left\| \mathbf{G}^{(i+1)}(j,:) \right\|^2 \tag{5.122}$$

   where $\mathbf{H}_{\bar{q}_i}$ is the matrix obtained by zeroing columns $q_1,\ldots,q_i$ of $\mathbf{H}$. Here $\mathbf{G}_{ZF}(\mathbf{H}_{\bar{q}_i})$ denotes the ZF filter applied to $\mathbf{H}_{\bar{q}_i}$.

For the non-ordered SIC, the operations denoted by $(*)$ are simply not performed.

In terms of performance, the diversity order experienced by the decoded layer is increased by one at each iteration. Therefore, the symbol/layer detected at iteration i will achieve a diversity of $n_r - n_t + i$. In the absence of error propagation, the output SNR ρ_i at iteration i with ZF filtering, is $\chi^2_{2(n_r-n_t+i)}$ distributed. In the presence of error propagation, the error performance is mostly dominated by the weakest stream. The diversity order achieved by non-ordered SIC is approximately $n_r - n_t + 1$, whereas ordered SIC improves the performance by reducing the error propagation caused by the first decoded stream. However, even in this case, the diversity order remains lower than n_r.

Assuming that we are able to code over a large number of channel realizations (i.e. that the channel is fast fading), the maximum achievable rate with a ZF-SIC receiver is expressed as

$$\bar{C}_{ZF-SIC} = \sum_{q=1}^{\min\{n_t,n_r\}} \mathcal{E}\{\log_2(1+\rho_q)\} \tag{5.123}$$

$$= \sum_{q=1}^{\min\{n_t,n_r\}} \mathcal{E}\left\{\log_2\left(1 + \frac{\rho}{n_t}\chi^2_{2(n_r-n_t+q)}\right)\right\}. \tag{5.124}$$

At high SNR, we approximate $\bar{C}_{ZF-SIC}$ as

$$\bar{C}_{ZF-SIC} \approx \min\{n_t,n_r\}\log_2\left(\frac{\rho}{n_t}\right) + \sum_{q=1}^{\min\{n_t,n_r\}} \mathcal{E}\left\{\log_2\left(\chi^2_{2(n_r-n_t+q)}\right)\right\}. \tag{5.125}$$

Surprisingly, the maximum achievable rate at high SNR with ZF-SIC is very similar to the ergodic MIMO capacity over fast fading channels, outlined by (4.40). The loss that was observed with ZF filtering is now compensated because the successive interference cancellation improves the SNR of each decoded layer. This mechanism is summarized in the next proposition.

Proposition 5.12   *Spatial Multiplexing with ZF-SIC (ZF V-BLAST) and equal power allocation achieves the ergodic capacity of i.i.d. Rayleigh fast fading MIMO channels at asymptotically high SNR.*

It is important to keep in mind that this only holds true when error propagation is neglected.

We also know that MMSE filtering improves the performance at lower SNR as it does not enhance the noise. Using MMSE-SIC should thus lead to a good performance and achieve the ergodic MIMO capacity at both low and high SNR. It does actually perform much better than that, as it achieves the mutual information (with equal power allocation) for any $\mathbf{H}$ and SNR (if $n_r \geq n_t$)

$$\bar{C}_{MMSE-SIC} = \sum_{q=1}^{\min\{n_t,n_r\}} \log_2(1 + \rho_q) = \log_2 \det\left(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{HH}^H\right) \qquad (5.126)$$

where ρ_q denotes the SNR of stream q. A consequence of the above property in Rayleigh fading channels is summarized below.

Proposition 5.13   *Spatial Multiplexing with MMSE-SIC (MMSE V-BLAST) and equal power allocation achieves the ergodic capacity for all SNR in i.i.d. Rayleigh fast fading MIMO channels.*

This surprising result has two explanations [VG97, TV05].

1. The MMSE filter is information lossless. For a transmission $\mathbf{r} = \mathbf{h}c + \mathbf{n}$ with Gaussian input c and noise $\mathbf{n}$, MMSE filtering does preserve all information about c which is contained in the received vector $\mathbf{r}$

$$\mathcal{I}(c; \mathbf{r}) = \mathcal{I}(c; \mathbf{g}_{MMSE}\mathbf{r}) \qquad (5.127)$$

   with $\mathbf{g}_{MMSE}$ the MMSE filter constructed for the channel realization $\mathbf{h}$.
2. The SIC receiver implements the chain rule of mutual information. Assuming the input $\mathbf{C}$ to be Gaussian distributed, using the chain rule for mutual information, the mutual information between the input and output is given by

$$\mathcal{I}(\mathbf{c}_k; \mathbf{y}_k) = \mathcal{I}(c_1; \mathbf{y}_k) + \mathcal{I}(c_2; \mathbf{y}_k | c_1) + \ldots + \mathcal{I}(c_{n_t}; \mathbf{y}_k | c_1, \ldots, c_{n_t-1}) \qquad (5.128)$$

We clearly observe a parallelism between (5.128) and the SIC strategy, which performs, conditioned on previously detected symbols $(c_1, \ldots, c_{q-1})$

$$\mathbf{y}'_k = \mathbf{y}_k - \sum_{i=1}^{q-1} \mathbf{H}(:, i)c_i. \tag{5.129}$$

As a consequence

$$\mathcal{I}\left(c_q; \mathbf{y}_k | c_1, \ldots, c_{q-1}\right) = \mathcal{I}\left(c_q; \mathbf{y}'_k\right) = \mathcal{I}\left(c_q; \mathbf{G}_{MMSE}(q, :)\,\mathbf{y}'_k\right) \tag{5.130}$$

where the last inequality relies on the fact that MMSE is information lossless. Therefore, at each iteration, the rate is given by $\mathcal{I}\left(c_q; \mathbf{y}_k | c_1, \ldots, c_{q-1}\right)$. The total rate is thus $\mathcal{I}\left(\mathbf{c}_k; \mathbf{y}_k\right)$.

To complete the study of the Spatial Multiplexing algorithm with SIC decoding, let us obtain the diversity-multiplexing trade-off achieved by unordered ZF-SIC with a QAM constellation.

Proposition 5.14 *For $n_r \geq n_t$, the diversity-multiplexing trade-off achieved by Spatial Multiplexing with QAM constellation and unordered ZF-SIC receiver over i.i.d. Rayleigh fading channels is given by*

$$g_d(g_s) = (n_r - n_t + 1)\left(1 - \frac{g_s}{n_t}\right), \ g_s \in [0, n_t]. \tag{5.131}$$

PROOF: Since the equivalent channel gain seen by the qth decoded layer is $\chi^2_{2(n_r - n_t + q)}$ distributed, reasoning similarly to (5.112), the error probability achieved by the qth layer reads as

$$P_e^{(q)}(R) \doteq \rho^{-(n_r - n_t + q)\left(1 - \frac{g_s}{n_t}\right)}. \tag{5.132}$$

The total error probability is then lower and upper-bounded by [ZT03]

$$P_e^{(1)}(R) \leq P_e(R) \leq \sum_{q=1}^{n_t} P_e^{(q)}(R). \tag{5.133}$$

At high SNR, the SNR exponent of the lower and upper bounds are equal, hence

$$P_e(R) \doteq \rho^{-(n_r - n_t + 1)\left(1 - \frac{g_s}{n_t}\right)} \tag{5.134}$$

and the diversity-multiplexing trade-off is provided by the SNR exponent. ∎

Note the achieved trade-off is similar to the trade-off achieved by a simple ZF receiver. This is due to the fact that the first layer dominates the error probability since its error exponent is the smallest. When ordering is used, the distribution of the equivalent channel gains after ZF filtering is changed, which in turn improves the diversity-multiplexing trade-off. However, it is much more difficult to evaluate the latter.
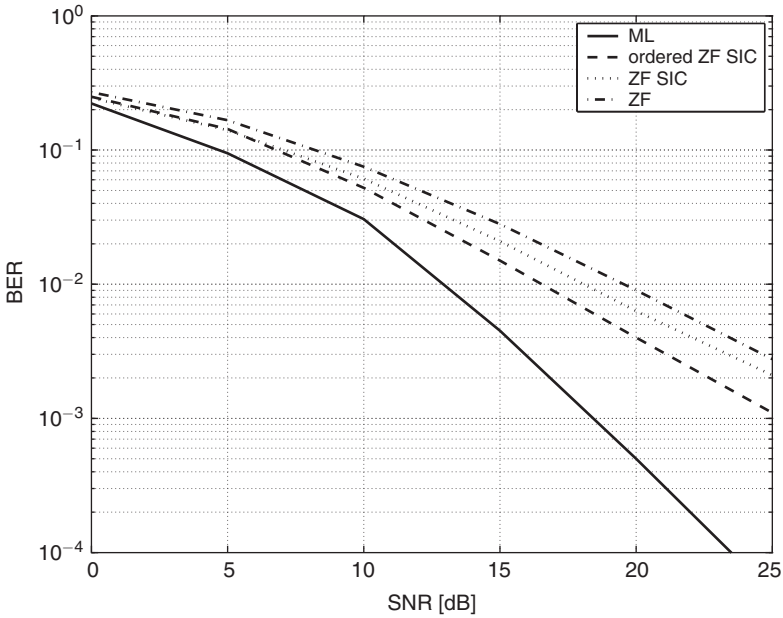
Figure 5.3 *Bit error rate (BER) of Spatial Multiplexing with various receivers (ML, (ordered) ZF SIC, ZF) in i.i.d. Rayleigh slow fading channels with $n_t = 2$ and $n_r = 2$ for 4 bits/s/Hz*

Impact of decoding strategy on error probability

In Figure 5.3, the performance of Spatial Multiplexing with the different decoding strategies (ML, ordered and non-ordered ZF SIC and simple ZF decoding) is illustrated for 2×2 i.i.d. Rayleigh fading channels. As expected, ML decoding highly outperforms all other strategies, because it is the only scheme exploiting the full receive diversity: the slope of the ML curve approaches 2. By contrast, ZF filtering only achieves a diversity order of $n_r - n_t + 1 = 1$. Ordered and non-ordered perform marginally better, but their diversity gain is still much lower than 2. Note that the advantage of ML decoding decreases as the number of receive antennas gets much larger than the number of transmit antennas.

5.5.3   D-BLAST

Using the error probability perspective of Section 5.5.2, it is clear that the transmission of independent data streams potentially exploits a diversity order of at most n_r out of $n_t n_r$. The lack of coding across antennas implies that the V-BLAST is in outage each time the SINR of a layer cannot support the rate allocated to that layer, e.g. when the corresponding subchannel is in deep fading. Therefore, it makes sense to use some sort of spatial interleaving so that each layer encounters all subchannels (i.e. all antennas), averaging out the randomness of each subchannel.

An improvement of the V-BLAST based on that principle is the D-BLAST algorithm, which is based on a stream rotation following the encoding of all layers. A layer is now

made of several streams, each stream experiencing a different subchannel (by contrast to V-BLAST, where each layer experiences the same subchannel during the whole transmission duration). Whereas the receiver is similar to V-BLAST, the transmitter requires an additional triangular block for the initialization. To illustrate the concept, consider two layers **a** and **b**, and two transmit antennas. Let us also assume that layer **a** is made of two streams $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ and layer **b** of two streams $\mathbf{b}^{(1)}$ and $\mathbf{b}^{(2)}$. Each stream can be seen as a block of symbols. The transmitted codeword **C** is now written as

$$\mathbf{C} = \begin{bmatrix} \mathbf{a}^{(1)} & \mathbf{b}^{(1)} \\ & \mathbf{a}^{(2)} & \mathbf{b}^{(2)} \end{bmatrix}. \tag{5.135}$$

During the first stream period, only $\mathbf{a}^{(1)}$ is transmitted on the first antenna. Antenna 2 sends nothing. Estimation of this stream can be performed through maximal ratio combining (MRC). Stream $\mathbf{a}^{(1)}$ experiences an equivalent channel gain $\mathbf{g}_1$. At the second stream period, $\mathbf{a}^{(2)}$ is estimated using a ZF or MMSE filter and considering $\mathbf{b}^{(1)}$ as an interference (analogous to V-BLAST). The channel gain experienced by $\mathbf{a}^{(2)}$ is denoted as $\mathbf{g}_2$. The estimates of $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ are then fed to the decoder to yield the first layer. The contribution of this layer is then canceled out from the received signal and the layer **b** is decoded (as in V-BLAST). Stream $\mathbf{b}^{(1)}$ is now left alone and takes the place of $\mathbf{a}^{(1)}$.

Let us focus on the detection of layer **a**. We may write that the equivalent channel experienced by this layer is

$$\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} = \sqrt{E_s} \begin{bmatrix} \mathbf{g}_1 & 0 \\ 0 & \mathbf{g}_2 \end{bmatrix} \begin{bmatrix} \mathbf{a}^{(1)} \\ \mathbf{a}^{(2)} \end{bmatrix} + \mathbf{n}. \tag{5.136}$$

From (5.136), the D-BLAST architecture appears as quite different from the V-BLAST. Indeed, in D-BLAST, $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ experience different channel gains, while in V-BLAST, one layer always experiences the same channel. Briefly, in D-BLAST, it is possible to code across streams $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ while in V-BLAST, $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ can be seen as two different layers.

Proposition 5.15 *Ignoring the initialization loss, ZF D-BLAST in i.i.d. Rayleigh fading channels with $n_r = n_t = n$ achieves the diversity-multiplexing trade-off given by the curve connecting points $(g_s, g_d) = (n - k, k(k + 1)/2)$ for $k = 0, \ldots, n$. MMSE D-BLAST achieves the entire optimal trade-off curve $g_d^\star (g_s, \infty)$.*

PROOF: The trade-off curve achievable by ZF D-BLAST results from an approach similar to the one used to prove Proposition 4.5. The optimality of MMSE D-BLAST is a consequence of the fact that MMSE SIC achieves the mutual information of the channel as explained in Section 5.5.2. More information may be found in [ZT03]. ∎

Achieving the diversity-multiplexing trade-off with MMSE D-BLAST requires coding with Gaussian random codes over a duration $T \geq n_t + n_r - 1$. We will find later that it is

possible to keep the MMSE D-BLAST architecture and use space-only codes ($T = 1$), also known as permutation codes, and still achieve the optimal diversity-multiplexing trade-off.

When the initialization loss is taken into account, MMSE D-BLAST does not achieve the optimal trade-off. For codewords of length T, only $T - n_t + 1$ symbols are transmitted on each antenna, leading to an effective transmission rate reduced by $(T - n_t + 1)/T$. Hence, the actual trade-off achieved by MMSE D-BLAST is given by $g_d^\star (T/(T - n_t + 1)g_s)$, i.e. the diversity obtained at a multiplexing gain g_s is the optimal diversity gain at the multiplexing gain $T/(T - n_t + 1)g_s$. For large codeword lengths, the loss is negligible. Note that in practice, error propagation mechanisms always take place, thereby highly limiting the number of layers which can be transmitted before re-initialization.

5.5.4   Orthogonal space–time block codes

Orthogonal space–time block codes (O-STBCs) are a very important sub-class of linear STBCs. They have remarkable properties which make them extremely easy to decode, while still achieving a full-diversity of $n_t n_r$. Indeed, O-STBCs are such that the MIMO ML decoding decouples into several SIMO ML decoding. Each transmitted symbol is thus decoded independently of the other transmitted symbols within the same block. However, O-STBCs have a much smaller spatial multiplexing rate than Spatial Multiplexing schemes.

Code properties
O-STBCs are linear STBCs structured as (5.51) and characterized by the two following properties:

1. The basis matrices are wide unitary

$$\boldsymbol{\Phi}_q \boldsymbol{\Phi}_q^H = \frac{T}{Qn_t} \mathbf{I}_{n_t} \quad \forall q = 1 \ldots 2Q \tag{5.137}$$

2. The basis matrices are pairwise skew-hermitian

$$\boldsymbol{\Phi}_q \boldsymbol{\Phi}_p^H + \boldsymbol{\Phi}_p \boldsymbol{\Phi}_q^H = 0, \quad q \neq p \tag{5.138}$$

or equivalently by this unique property

$$\mathbf{C}\mathbf{C}^H = \frac{T}{Qn_t} \left[\sum_{q=1}^{Q} |c_q|^2 \right] \mathbf{I}_{n_t}. \tag{5.139}$$

It has to be noted that (5.139) is fully equivalent to both (5.137) and (5.138), i.e. (5.139) is valid if and only if (5.137) and (5.138) are also valid.

Code construction

The first proposed O-STBC for two transmit antennas is attributed to Alamouti [Ala98]. The construction has then been extended to a larger number of antennas in [TJC99]. The code construction is based on the theory of amicable orthogonal designs developed by Radon-Hurwitz [TJC99]. There are two classes of orthogonal codes, using real symbol constellations (real O-STBCs) or complex symbol constellations (complex O-STBCs). Real O-STBCs with spatial multiplexing rate $r_s = 1$ can be designed for any number of transmit antennas. An example of real O-STBC is given below.

Example 5.8 *For three transmit antennas ($n_t = 3$), a real O-STBC expanding on four symbol durations ($T = 4$) with a spatial multiplexing rate equal to one is given by*

$$\mathbf{C} = \frac{1}{\sqrt{3}} \begin{bmatrix} c_1 & -c_2 & -c_3 & -c_4 \\ c_2 & c_1 & c_4 & -c_3 \\ c_3 & -c_4 & c_1 & c_2 \end{bmatrix}. \tag{5.140}$$

Symbols c_1, c_2, c_3 and c_4 originate from a real symbol constellation. This code is delay-optimal in the sense that $T = 4$ is the smallest delay achievable to guarantee a full-diversity real O-STBC with a spatial multiplexing rate of one for three transmit antennas.

By contrast, complex O-STBCs with $r_s = 1$ only exist for $n_t = 2$. For larger n_t, it is always possible to find codes with $r_s \leq 1/2$. For some particular values of $n_t > 2$, complex O-STBCs with $1/2 < r_s < 1$ have been developed. This is the case for $n_t = 3$ and $n_t = 4$ with $r_s = 3/4$. Examples of complex O-STBCs for $n_t = 2, 3, 4$ are given below.

Example 5.9 *The best known O-STBC is the Alamouti code [Ala98]. It is a complex O-STBC for two transmit antennas and characterized by a multiplexing rate $r_s = 1$. Codewords are written as*

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix}. \tag{5.141}$$

The basis matrices of the Alamouti code are easily found by identification with the structure of (5.51). They are actually those of Example 5.4. Clearly, those matrices are unitary and skew-hermitian, and $\mathbf{CC}^H = \frac{1}{2} \left[|c_1|^2 + |c_2|^2 \right] \mathbf{I}_2$. The Alamouti code has a spatial multiplexing rate $r_s = 1$ since two symbols are transmitted over two symbol durations.

Example 5.10   *For three transmit antennas, a complex O-STBC expanding on four symbol durations* $(T=4)$ *and transmitting three symbols on each block* $(Q=3)$ *is given by*

$$\mathbf{C} = \frac{2}{3} \begin{bmatrix} c_1 & -c_2^* & c_3^* & 0 \\ c_2 & c_1^* & 0 & c_3^* \\ c_3 & 0 & -c_1^* & -c_2^* \end{bmatrix}. \tag{5.142}$$

The spatial multiplexing rate r_s *is equal to 3/4.*

Example 5.11   *For four transmit antennas, a complex O-STBC expanding on four symbol durations* $(T=4)$ *and transmitting three symbols on each block* $(Q=3)$ *is given by*

$$\mathbf{C} = \frac{1}{\sqrt{3}} \begin{bmatrix} c_1 & -c_2^* & c_3^* & 0 \\ c_2 & c_1^* & 0 & c_3^* \\ c_3 & 0 & -c_1^* & -c_2^* \\ 0 & c_3 & c_2 & -c_1 \end{bmatrix}. \tag{5.143}$$

The spatial multiplexing rate r_s *is equal to 3/4.*

Interestingly, a codeword obtained by selecting n_t' lines out of the n_t lines of a complex O-STBC codeword is a complex O-STBC codeword for n_t' antennas. In the previous examples, the O-STBC for three antennas can indeed be obtained by selecting three lines out of the codeword of the four-antenna O-STBC.

Detection of O-STBCs
The following proposition is directly derived from Proposition 5.7.

Proposition 5.16   *O-STBCs enjoy the decoupling property of Proposition 5.7.*

O-STBCs are thereby extremely simply decoded, as illustrated in the following example.

Example 5.12   *Assume the following MISO transmission based on the Alamouti code with one receive antenna*

$$[\, y_1 \ y_2 \,] = \sqrt{\frac{E_s}{2}} [\, h_1 \ h_2 \,] \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix} + [\, n_1 \ n_2 \,] \tag{5.144}$$

with h_1 and h_2 the channel gains between the transmit antennas and the receive antenna. We may also rewrite (5.144) as

$$\begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \sqrt{\frac{E_s}{2}} \underbrace{\begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}}_{\mathbf{H}_{eff}} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix}. \tag{5.145}$$

Applying the space–time matched filter $\mathbf{H}_{eff}^H$ to the received vector decouples the transmitted symbols

$$
\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \mathbf{H}_{eff}^H \begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \sqrt{\frac{E_s}{2}} \Big[|h_1|^2 + |h_2|^2 \Big] \mathbf{I}_2 \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \mathbf{H}_{eff}^H \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix}. \tag{5.146}
$$

Expanding the original ML metric

$$
\left| y_1 - \sqrt{\frac{E_s}{2}} \, (h_1 c_1 + h_2 c_2) \right|^2 + \left| y_2 - \sqrt{\frac{E_s}{2}} \, (-h_1 c_2^* + h_2 c_1^*) \right|^2 \tag{5.147}
$$

and making use of z_1 and z_2, the decision metric for c_1 is

$$
\text{choose } c_i \text{ iff } \left| z_1 - \sqrt{\frac{E_s}{2}} \Big(|h_1|^2 + |h_2|^2 \Big) c_i \right|^2 \leq \left| z_1 - \sqrt{\frac{E_s}{2}} \Big(|h_1|^2 + |h_2|^2 \Big) c_k \right|^2 \; \forall i \neq k \tag{5.148}
$$

whereas the decision metric for c_2 is

$$
\text{choose } c_i \text{ iff } \left| z_2 - \sqrt{\frac{E_s}{2}} \Big(|h_1|^2 + |h_2|^2 \Big) c_i \right|^2 \leq \left| z_2 - \sqrt{\frac{E_s}{2}} \Big(|h_1|^2 + |h_2|^2 \Big) c_k \right|^2 \; \forall i \neq k \tag{5.149}
$$

Independent decoding of symbols c_1 and c_2 is so performed.

Error probability

Proposition 5.3 tells us that for a given set of parameters T, Q, and n_t, O-STBCs minimize the maximum average error probability over i.i.d. slow Rayleigh fading channels. The orthogonality properties of O-STBC make the calculation of the average PEP particularly simple. Indeed, from (5.12), the average PEP of an O-STBC in i.i.d. slow Rayleigh fading channels is given by

$$
P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \left(1 + \eta \frac{T}{Qn_t} \sum_{q=1}^{Q} |c_q - e_q|^2 \right)^{-n_r n_t} d\beta \tag{5.150}
$$

$$
= \frac{1}{2} \left[1 - \sqrt{\frac{\rho_s}{1 + \rho_s}} \sum_{i=0}^{n_t n_r - 1} \binom{2i}{i} \left(\frac{1}{4(1 + \rho_s)} \right)^i \right] \tag{5.151}
$$

with $\rho_s = \frac{\rho}{4} \frac{T}{Qn_t} \sum_{q=1}^{Q} |c_q - e_q|^2$. In (5.151), the integral was evaluated thanks to the results of [Sim01]. In the high SNR regime, this expression is generally upper-bounded by means of the Chernoff bound

$$
P(\mathbf{C} \to \mathbf{E}) \leq \left(\frac{\rho}{4} \frac{T}{Qn_t} \right)^{-n_r n_t} \left(\sum_{q=1}^{Q} |c_q - e_q|^2 \right)^{-n_r n_t}. \tag{5.152}
$$

As already mentioned, the PEP is the probability to confuse a codeword with another codeword. However, a codebook is generally made of many codewords and the total error rate is commonly upper bounded by a union bound, the summation being carried out over all possible pairs of codewords. However, with O-STBCs, the calculation of the average symbol error rate $\bar{P}$ is much easier thanks to the decoupling property: the error rate analysis becomes very similar to SISO systems, the equivalent channel gain being equal to the Frobenius norm of the channel matrix. As explained in [SA00], the conditional symbol error rate $P\left(\|\mathbf{H}\|^2\right)$ for PSK, PAM and QAM signaling is respectively given by

$$P_{PSK}\left(\|\mathbf{H}\|^2\right) = \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} e^{-\zeta\|\mathbf{H}\|_F^2} d\beta, \tag{5.153}$$

$$P_{PAM}\left(\|\mathbf{H}\|^2\right) = \frac{2}{\pi} \frac{M-1}{M} \int_0^{\frac{\pi}{2}} e^{-\zeta\|\mathbf{H}\|_F^2} d\beta, \tag{5.154}$$

$$P_{QAM}\left(\|\mathbf{H}\|^2\right) = \frac{4}{\pi} \frac{\sqrt{M}-1}{\sqrt{M}} \left[\frac{1}{\sqrt{M}} \int_0^{\frac{\pi}{4}} e^{-\zeta\|\mathbf{H}\|_F^2} d\beta + \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} e^{-\zeta\|\mathbf{H}\|_F^2} d\beta \right] \tag{5.155}$$

where $d_{min,PSK}^2 = 4\sin^2\left(\frac{\pi}{M}\right)$, $d_{min,PAM}^2 = \frac{12}{M^2-1}$ and $d_{min,QAM}^2 = \frac{6}{M-1}$ are the minimum squared Euclidean distance of M-PSK, M-PAM and M-QAM, respectively and $\zeta = \eta T \delta^2 / (Q n_t)$ with $\delta^2 = d_{min,PSK}^2$, $d_{min,PAM}^2$ or $d_{min,QAM}^2$. The average of the conditional symbol error rate $P\left(\|\mathbf{H}\|^2\right)$ over the fading distribution $p_{\|\mathbf{H}\|^2}$ gives the average symbol error rate

$$\bar{P} = \int_0^\infty P\left(\|\mathbf{H}\|^2\right) p_{\|\mathbf{H}\|^2}\left(\|\mathbf{H}\|^2\right) d\|\mathbf{H}\|^2. \tag{5.156}$$

Since the squared Frobenius norm of Rayleigh i.i.d. channels is a χ^2 variable with $2n_t n_r$ degrees of freedom (see Section 4.2), we obtain [SA00]

$$\bar{P}_{PSK} = \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} (1+\zeta)^{-n_r n_t} d\beta, \tag{5.157}$$

$$\bar{P}_{PAM} = \frac{2}{\pi} \frac{M-1}{M} \int_0^{\frac{\pi}{2}} (1+\zeta)^{-n_r n_t} d\beta, \tag{5.158}$$

$$\bar{P}_{QAM} = \frac{4}{\pi} \frac{\sqrt{M}-1}{\sqrt{M}} \left[\frac{1}{\sqrt{M}} \int_0^{\frac{\pi}{4}} (1+\zeta)^{-n_r n_t} d\beta + \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} (1+\zeta)^{-n_r n_t} d\beta \right]. \tag{5.159}$$

Mutual information properties of O-STBCs

From an information theoretic perspective, O-STBCs incur a capacity loss as they do not fulfill (5.71). Reasoning as in (5.21), we may write

$$\mathcal{I}_e\left(\mathbf{H}\right) = \log_2 \det\left(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{W}\right) = \log_2\left(1 + \frac{\rho}{n_t}\|\mathbf{H}\|_F^2 + \dots \right.$$

$$\left. + \left(\frac{\rho}{n_t}\right)^{r(\mathbf{H})} \prod_{k=1}^{r(\mathbf{H})} \lambda_k\left(\mathbf{H}\mathbf{H}^H\right)\right). \tag{5.160}$$

As previously highlighted, O-STBCs transform the MIMO channel into equivalent SISO channels

$$z_q = \sqrt{\frac{E_s T}{Q n_t}} \|\mathbf{H}\|_F^2 c_q + n_q' \tag{5.161}$$

with $n_q' \approx \mathcal{N}(0, \|\mathbf{H}\|_F^2 N_0/2)$. Since Q symbols are transmitted per T symbol durations, it results from (5.161) that O-STBCs achieve a mutual information equal to

$$\mathcal{I}_{O-STBC}(\mathbf{H}) = \frac{Q}{T} \log_2 \left(1 + \frac{\rho T}{Q n_t} \|\mathbf{H}\|_F^2 \right), \tag{5.162}$$

indicating that $\mathcal{I}_{O-STBC}(\mathbf{H}) \leq \mathcal{I}_e(\mathbf{H})$. This leads to the following proposition.

Proposition 5.17 *For a given channel realization* $\mathbf{H}$*, the mutual information achieved by any O-STBC is always upper-bounded by the channel mutual information with equal power allocation* $\mathcal{I}_e$*. Equality occurs if and only if both the rank of the channel and the spatial multiplexing rate of the code are equal to one.*

Corollary 5.1 *The Alamouti scheme is optimal with respect to the conditional channel mutual information when used with only one receive antenna.*

Diversity-multiplexing trade-off achieved by O-STBCs

The above analysis suggests that O-STBCs are generally not optimal from the diversity-multiplexing trade-off perspective. This is indeed the case as shown below. To this end, let us assume that symbols are drawn from a QAM constellation. The following proposition then holds true.

Proposition 5.18 *The diversity-multiplexing trade-off achieved by O-STBCs using QAM constellations in i.i.d. Rayleigh fading channels is given by*

$$g_d(g_s) = n_r n_t \left(1 - \frac{g_s}{r_s} \right), \ g_s \in [0, r_s]. \tag{5.163}$$

PROOF: In order to achieve a total rate R, each symbol should be chosen in a constellation with $2^{R/(2r_s)}$ points in each dimension. Making use of the decoupling property of O-STBCs and since the number of nearest symbols is at most four in any QAM constellation, the overall error probability P_e is upper-bounded by four times the worst-case pairwise error probability $P(c_q \rightarrow e_q)$

$$P_e(R) \leq 4 \min_{c_q, e_q} P(c_q \rightarrow e_q) \leq 4 \left(\frac{\rho}{4} \frac{T}{Q n_t} \right)^{-n_r n_t} d_{min}^{-2n_r n_t}$$

$$\doteq 4 \left(\frac{T}{4 Q n_t} \right)^{-n_r n_t} \rho^{-n_r n_t} 2^{\frac{R n_r n_t}{r_s}}$$

$$\doteq 4 \left(\frac{T}{4 Q n_t} \right)^{-n_r n_t} \rho^{-n_r n_t \left(1 - \frac{g_s}{r_s} \right)} \tag{5.164}$$

where d_{min} is the minimum Euclidean distance of the QAM constellation, of the order of $1/2^{R/r_s}$ with $R = g_s \log_2(\rho)$. The diversity-multiplexing trade-off achieved by O-STBCs with a QAM constellation is thus given by the SNR exponent in (5.164). ∎

Proposition 5.19   *The Alamouti code with any QAM constellation achieves the optimal diversity-multiplexing trade-off for two transmit and one receive antennas in i.i.d. Rayleigh fading channels.*

PROOF: The Alamouti code is the only O-STBC with $r_s = 1$. For $n_r = 1$, the diversity-multiplexing trade-off in (5.163) matches the diversity-multiplexing trade-off of the MISO 2×1 channel in Proposition 4.5 and Example 4.2. ∎

However, for a larger number of receive antennas, the diversity-multiplexing achievable with the Alamouti scheme is strictly suboptimal.

Performance of Alamouti code

Figure 5.4 depicts the performance of the Alamouti scheme in terms of block error rate for 4 different rates $R = 4, 8, 12, 16$ bits/s/Hz in 2×2 i.i.d. slow Rayleigh fading channels. The lighter curves represent the corresponding outage probability for comparison. The slope of each curve at high SNR yields the diversity gain achieved by the code, i.e. $g_d(g_s = 0) = g_d^o = 4$, reflecting that the Alamouti code exploits the full channel diversity. However, the growth of the multiplexing gain is slow: 12 dB separate the different curves, corresponding to a multiplexing gain $g_s = 1$, i.e. 1 bit/s/Hz increase per 3 dB SNR increase. Owing to this limited multiplexing gain, the Alamouti code performance might be far from the outage probability at high transmission rates. Nevertheless, for transmission rates as low as $R = 4$ bits/s/Hz, the achieved performance remains close to the outage probability with a codeword length of only two symbol durations.

5.5.5   Quasi-orthogonal space–time block codes

O-STBCs extract the full diversity at a low decoding complexity. However, they incur a high loss in spatial multiplexing rate w.r.t. Spatial Multiplexing, especially as the number of transmit antennas increases. Indeed, a complex O-STBC for $n_t = 2$ has a maximum spatial multiplexing rate of one, while for $n_t = 3, 4$, complex O-STBCs with a maximum spatial multiplexing rate of 3/4 exist. For a larger number of transmit antennas, it is getting more and more difficult to find O-STBCs with rates higher than 1/2.

In order to increase the spatial multiplexing rate while still partially enjoying the decoupling properties of O-STBCs, the relaxation on the orthogonality requirements has led to the development of quasi-orthogonal space–time block codes (QO-STBCs). The basic principle behind QO-STBCs is to use O-STBCs of reduced dimensions as the building blocks of a higher dimensional code, which retains the same spatial multiplexing rate of the O-STBC, as well as some orthogonality. Let us denote by $\mathcal{O}(c_1, \ldots, c_Q)$ an O-STBC made
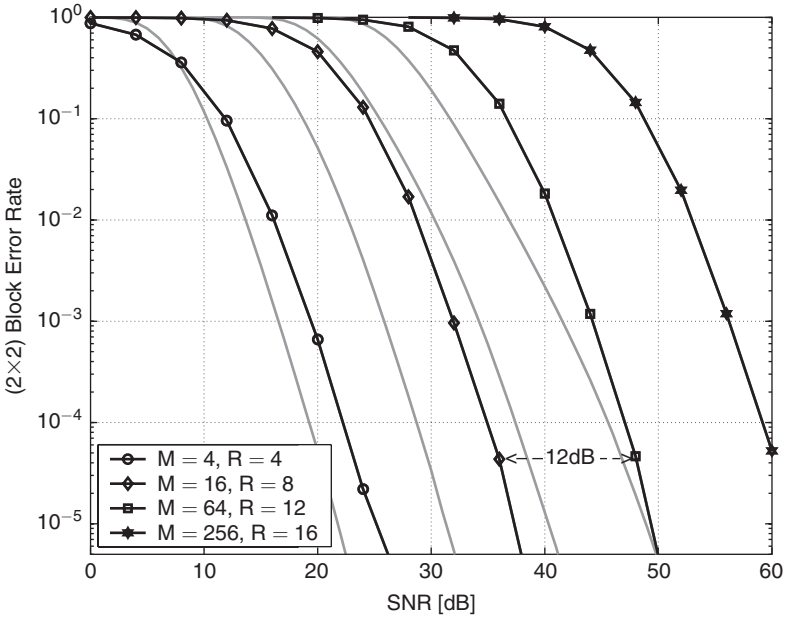
Figure 5.4 *Block error rate of the Alamouti code in i.i.d. Rayleigh slow fading channels with* $n_r = 2$ *in* $R = 2\log_2(M) = 4, 8, 12, 16$ *bit/s/Hz transmissions using* M^2*-QAM constellations of sizes* $M = 4, 16, 64, 256$ *(courtesy of H. Yao [YW03])*

of the Q symbols $c_1, \ldots, c_Q$. QO-STBC $\mathcal{Q}(c_1, \ldots, c_{2Q})$, made of $2Q$ complex symbols $c_1, \ldots, c_{2Q}$ is then derived based on one of the two following methods.

A first method, proposed in [TBH00], is outlined as follows

$$\mathcal{Q}(c_1, \ldots, c_{2Q}) = \begin{bmatrix} \mathcal{O}(c_1, \ldots, c_Q) & \mathcal{O}(c_{Q+1}, \ldots, c_{2Q}) \\ \mathcal{O}(c_{Q+1}, \ldots, c_{2Q}) & \mathcal{O}(c_1, \ldots, c_Q) \end{bmatrix}. \tag{5.165}$$

Originally, the code has been proposed with the Alamouti code as the orthogonal code component.

Example 5.13 *A $n_t = 4$ QO-STBC built from the Alamouti code is*

$$\mathcal{Q}(c_1, \ldots, c_4) = \frac{1}{2} \begin{bmatrix} c_1 & -c_2^* & c_3 & -c_4^* \\ c_2 & c_1^* & c_4 & c_3^* \\ c_3 & -c_4^* & c_1 & -c_2^* \\ c_4 & c_3^* & c_2 & c_1^* \end{bmatrix}. \tag{5.166}$$

Though not orthogonal, this code still retains some orthogonality, as it appears from the following relationship:

$$\mathcal{Q}(c_1,\ldots,c_4)\mathcal{Q}^H(c_1,\ldots,c_4) = \frac{1}{4} \begin{bmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ b & 0 & a & 0 \\ 0 & b & 0 & a \end{bmatrix}, \tag{5.167}$$

with

$$a = \sum_{q=1}^{4} |c_q|^2, \tag{5.168}$$

$$b = c_1 c_3^* + c_3 c_1^* - c_2 c_4^* - c_4 c_2^*. \tag{5.169}$$

The first column of $\mathcal{Q}(c_1,\ldots,c_4)$ is orthogonal to the second and fourth columns but not to the third column. Actually, the subspace created by the first and third columns of $\mathcal{Q}(c_1,\ldots,c_4)$ is orthogonal to the subspace created by the second and fourth columns of $\mathcal{Q}(c_1,\ldots,c_4)$. This indicates that the pair of symbols $\{c_1, c_3\}$ can be decoded independently of $\{c_2, c_4\}$. However, interference remains between the pairs $\{c_1, c_3\}$ and $\{c_2, c_4\}$. Note that the spatial multiplexing rate $r_s = 1$ while it was equal to 3/4 for (5.143).

An alternative derivation of QO-STBCs [Jaf01] consists in disposing the orthogonal codes $\mathcal{O}(c_1,\ldots,c_Q)$ and $\mathcal{O}(c_{Q+1},\ldots,c_{2Q})$ as if they were symbols of the Alamouti code, i.e.

$$\mathcal{Q}(c_1,\ldots,c_{2Q}) = \begin{bmatrix} \mathcal{O}(c_1,\ldots,c_Q) & -\mathcal{O}(c_{Q+1},\ldots,c_{2Q})^* \\ \mathcal{O}(c_{Q+1},\ldots,c_{2Q}) & \mathcal{O}(c_1,\ldots,c_Q)^* \end{bmatrix}. \tag{5.170}$$

Example 5.14   *A $n_t = 4$ QO-STBC built from the Alamouti code is*

$$\mathcal{Q}(c_1,\ldots,c_4) = \frac{1}{2} \begin{bmatrix} c_1 & -c_2^* & -c_3^* & c_4 \\ c_2 & c_1^* & -c_4^* & -c_3 \\ c_3 & -c_4^* & c_1^* & -c_2 \\ c_4 & c_3^* & c_2^* & c_1 \end{bmatrix}. \tag{5.171}$$

As is the previous code, this code is not orthogonal, since

$$\mathcal{Q}(c_1,\ldots,c_4)\mathcal{Q}^H(c_1,\ldots,c_4) = \begin{bmatrix} a & 0 & 0 & b \\ 0 & a & -b & 0 \\ 0 & -b & a & 0 \\ b & 0 & 0 & a \end{bmatrix}, \tag{5.172}$$

with

$$a = \sum_{q=1}^{4} |c_q|^2, \tag{5.173}$$

$$b = c_1 c_4^* + c_4 c_1^* - c_2 c_3^* - c_3 c_2^*. \tag{5.174}$$

The subspace created by the first and fourth columns of $\mathcal{Q}(c_1, \ldots, c_4)$ is orthogonal to the subspace created by the second and third columns of $\mathcal{Q}(c_1, \ldots, c_4)$. This indicates that the pair of symbols $\{c_1, c_4\}$ can be decoded independently of $\{c_2, c_3\}$.

A third similar scheme has also been proposed in [PF03].

Rotated QO-STBCs

The main drawback of QO-STBCs is their low achievable diversity. Consider as an example a QO-OSTBC of the first kind in (5.165) (results are similar for the second kind in (5.170)). For two possible different codewords $\mathbf{C} = \mathcal{Q}(c_1, \ldots, c_{2Q})$ and $\mathbf{E} = \mathcal{Q}(e_1, \ldots, e_{2Q})$, the determinant of the error matrix $\tilde{\mathbf{E}}$ reads as [SX04]

$$\det(\tilde{\mathbf{E}}) = \left(\sum_{q=1}^{Q} |(c_q - e_q) - (c_{q+Q} - e_{q+Q})|^2 \right)^{n_t}$$
$$\left(\sum_{q=1}^{Q} |(c_q - e_q) + (c_{q+Q} - e_{q+Q})|^2 \right)^{n_t}. \tag{5.175}$$

It is relatively easy to understand that QO-STBCs may be rank-deficient. When all symbols are chosen from the same PSK or QAM constellation, it is possible that the determinant of the error matrix be equal to zero. Typically, the minimum rank of the error matrix of codes in (5.166) and (5.171) is equal to two when symbols are all from the same PSK or QAM constellation.

In order to improve the diversity gain, it is possible to use different constellations for some pairs of symbols. Let us assume that symbols $(c_1, \ldots, c_Q)$ are chosen in a constellation $\mathcal{B}$ while symbols $(c_{Q+1}, \ldots, c_{2Q})$ are picked up in constellation $\mathcal{B}'$. A simple example consists in choosing in codes in (5.166) and (5.171) the symbols c_3 and c_4 from a rotated constellation instead of using the same constellation as symbols c_1 and c_2. This leads to rotated QO-STBCs [SP03, SX04, WX05, XL05]. In general, the second constellation $\mathcal{B}'$ is chosen as $\mathcal{B}' = e^{j\vartheta}\mathcal{B}$. Several studies have investigated the optimization of the rotation angle in order to maximize the minimum determinant. Using (5.175), the latter is expressed as

$$\min_{\substack{\mathbf{C}, \mathbf{E} \\ \mathbf{C} \neq \mathbf{E}}} \left[\det(\tilde{\mathbf{E}}) \right]^{1/(2n_t)} = \min_{\substack{u, u' \in \mathcal{B} \\ v, v' \in \mathcal{B}' = e^{j\vartheta}\mathcal{B} \\ (u,v) \neq (u',v')}} \left| (u - u')^2 - (v - v')^2 \right|. \tag{5.176}$$

It can be shown [XW05, XL05] that the optimal rotation angles for M-PSK constellation are $\vartheta = \pi/M$ and $\vartheta = \pi/(2M)$ respectively for M even and odd. QAM constellations have been studied in [SSL04]. If $\mathcal{B}$ is a signal constellation drawn from a square lattice, the optimum rotation angle is given by $\vartheta = \pi/4$, whereas if it is drawn from a lattice of equilateral triangles, the optimum rotation angle is given by $\vartheta = \pi/6$.

5.5.6   Linear dispersion codes

While Spatial Multiplexing achieves high data rates, it does not succeed in leveraging transmit diversity. On the other hand, orthogonal space–time block codes deliver the full diversity at the cost of a limited spatial multiplexing rate. However, while the skew-hermitian character is mandatory to decouple the streams at the receiver, it is not required to achieve full-diversity. Therefore, if a larger receiver complexity is authorized, it is possible to relax the skew-hermitian conditions and increase the data rates while still providing transmit diversity. Codes based on that principle are known as linear dispersion codes (LDCs).

Hassibi and Hochwald codes

Hassibi and Hochwald in [HH01] were the first to introduce the general framework presented in (5.51). The original idea of that framework is to derive the basis matrices that maximize the mutual information for a given set of parameters n_t, n_r, T, Q and ρ. Imposing a power constraint on the basis matrices, constraint gradient algorithms are used to derive the optimal basis matrices. Yet, the optimization being based on numerical methods, there is no guarantee that the codes obtained are a global solution.

Example 5.15   *An example of a LDC [HH02a] maximizing the mutual information for $T = 2$ and $Q = 4$ with 2 transmit antennas is given by*

$$\boldsymbol{\Phi}_1 = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad\qquad \boldsymbol{\Phi}_2 = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad (5.177)$$

$$\boldsymbol{\Phi}_3 = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad\qquad \boldsymbol{\Phi}_4 = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad (5.178)$$

$$\boldsymbol{\Phi}_5 = \frac{1}{2}\begin{bmatrix} j & 0 \\ 0 & j \end{bmatrix}, \qquad\qquad \boldsymbol{\Phi}_6 = \frac{1}{2}\begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix}, \qquad (5.179)$$

$$\boldsymbol{\Phi}_7 = \frac{1}{2}\begin{bmatrix} j & 0 \\ 0 & -j \end{bmatrix}, \qquad\qquad \boldsymbol{\Phi}_8 = \frac{1}{2}\begin{bmatrix} 0 & j \\ -j & 0 \end{bmatrix}. \qquad (5.180)$$

This code is capacity-efficient since $\mathcal{X}\mathcal{X}^T = \frac{1}{2}\mathbf{I}_8$ (from Proposition 5.5). However, this code has not been optimized with respect to the diversity and coding gains. This implies that it is actually rank-deficient and achieves only a transmit diversity of one,

like Spatial Multiplexing. Its coding gain is, however, slightly larger than the SM coding gain.

Heath codes and Sandhu codes

Other authors have tried to account for both the mutual information and error probability in the design of the basis matrices [San02, HP02, Hea01]. In [San02], codes that are capacity-efficient, i.e. with basis matrices such that $\mathcal{X}\mathcal{X}^T = \frac{1}{n_t}\mathbf{I}_{2n_tT}$ (Proposition 5.5), are randomly generated and the one that minimizes

$$\max_{\substack{q,p \\ q \neq p}} \left\| \mathbf{\Phi}_q \mathbf{\Phi}_p^H + \mathbf{\Phi}_p \mathbf{\Phi}_q^H \right\|_F^2$$

is chosen (Proposition 5.3). This design methodology provides efficient codes that are independent of any specific constellation. In [HP02, Hea01], codes with a tight frame structure $\mathcal{X}^T\mathcal{X} = \frac{T}{Q}\mathbf{I}_{2Q}$ are considered (see Propositions 5.4 and 5.6). Those with large minimum rank and minimum determinant are then picked up.

Error performance of linear dispersion codes

In Figure 5.5, the BERs achieved by Hassibi, Sandhu and Heath codes are compared. The performance of SM and O-STBC is also displayed for reference. In each case, ML decoding is performed at the receiver and a QPSK constellation is used for all codes except the Alamouti code where a 16-QAM keeps the same transmitted data rate. The Hassibi code has a slight improvement over SM: its minimum rank is equal to one (same as SM) but its coding gain is slightly larger. Heath and Sandhu codes offer a much better performance since they are capacity-efficient while also being full rank. The Alamouti code, while offering full diversity as well, is slightly outperformed by Heath and Sandhu codes. This comes from the fact that it is not capacity-efficient for $n_r > 1$. To compensate for its small spatial multiplexing rate, it uses less efficient signal constellations. As the number of receive antennas or the size of the constellation increase, the gain offered by full-rank capacity-efficient codes over O-STBCs increases. In general, O-STBCs are an attractive solution for MIMO systems with low data rates and a small number of antennas. For larger rates and larger number of antennas, high-rate capacity-efficient codes should be preferred, bearing in mind that those codes present an increased receiver complexity with respect to O-STBCs.

5.5.7 Algebraic space–time codes

With the elaboration of the diversity-multiplexing trade-off, many structured codes have been derived in order to maximize the worst-case PEP. Most of them, developed for TIMO (two-input multiple-output) systems, share a similar layering structure. A layer can be seen as a part of the information which is transmitted at each time instant by at most one antenna. Therefore, a layer is never found transmitting simultaneously on more than two antennas.
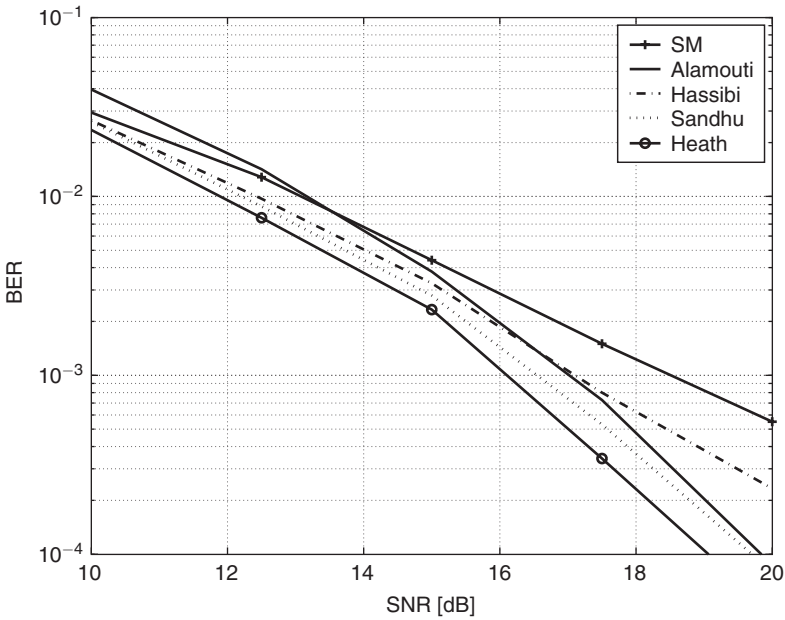
Figure 5.5 *Bit error rate (BER) of several LDCs in i.i.d. Rayleigh slow fading channels with $n_t = 2$ and $n_r = 2$ for 4 bits/s/Hz*

This layering principle was already present in D-BLAST and O-STBCs. In D-BLAST, a layer is defined as the sequence of outputs of a given encoder before entering the rotation part. After rotation, the same layer experiences different transmit antennas, one at a time.

Let us consider the Alamouti scheme as a first example. Two layers are found: the first one is made of symbols c_1 and c_1^*, while the second layer is made of symbols c_2 and $-c_2^*$. Clearly, symbols c_1 and c_2 (transmitted at the first time instant) are repeated at the second time instant, inducing a loss in spatial multiplexing rate. The presence of repetition coding within each layer is actually a general problem with O-STBCs. The framework proposed by algebraic codes consists in replacing the repetition of O-STBCs by a rotation, as illustrated below for two transmit antennas:

$$\mathbf{C} = \begin{bmatrix} u_1 & \varphi^{1/2} v_1 \\ \varphi^{1/2} v_2 & u_2 \end{bmatrix} \tag{5.181}$$

where $|\varphi| = 1$. The first layer is given by

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \mathcal{M}_1 \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \tag{5.182}$$

while the second layer is

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \mathcal{M}_2 \begin{bmatrix} c_3 \\ c_4 \end{bmatrix}. \tag{5.183}$$

Matrices $\mathcal{M}_1$ and $\mathcal{M}_2$ are unitary matrices. Symbols $\{c_k\}_{k=1}^{4}$ are points in a M^2-QAM constellation.

Several codes corresponding to different subclasses of the above framework have been investigated and optimized based on the rank-determinant criterion. The codes presented in the following do not induce any capacity loss, as they all satisfy Proposition 5.5. Moreover, all are full rank.

$\mathbf{B}_{2,\phi}$ *code*

The $B_{2,\phi}$ code [DTB02] relies on the following choice

$$\mathcal{M}_1 = \mathcal{M}_2 = \frac{1}{2}\begin{bmatrix} 1 & e^{j\varpi} \\ 1 & -e^{j\varpi} \end{bmatrix} \tag{5.184}$$

and $\varphi = e^{j\varpi}$. The value of ϖ leading to the worst-case PEP can be found through an exhaustive search as equal to $\varpi = 0.5$ and $\varpi = 0.521$ respectively for QPSK and 16-QAM constellations.

Threaded algebraic space–time (TAST) code

The TAST code [GD03] subsumes the $B_{2,\phi}$ as a special case, with

$$\mathcal{M}_1 = \mathcal{M}_2 = \frac{1}{2}\begin{bmatrix} 1 & e^{j\pi/4} \\ 1 & -e^{j\pi/4} \end{bmatrix}. \tag{5.185}$$

The optimization of the worst-case PEP leads to the optimal values of $\varphi = e^{j\frac{\pi}{6}}$ and $\varphi = e^{j\frac{\pi}{4}}$ respectively for QPSK and 16-QAM constellations.

Tilted QAM code

The tilted QAM code [YW03] is obtained by taking $\varphi = 1$ with two different rotation matrices

$$\mathcal{M}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} \cos\varpi_1 & \sin\varpi_1 \\ -\sin\varpi_1 & \cos\varpi_1 \end{bmatrix}, \tag{5.186}$$

$$\mathcal{M}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} \cos\varpi_2 & \sin\varpi_2 \\ -\sin\varpi_2 & \cos\varpi_2 \end{bmatrix}. \tag{5.187}$$

Optimization of the worst-case PEP leads to the optimal values $\varpi_1 = \frac{1}{2}\arctan\left(\frac{1}{2}\right)$ and $\varpi_2 = \frac{1}{2}\arctan(2)$.

By contrast to the two previous codes, the tilted QAM code has the remarkable property that the minimum determinant of the error matrix is lower-bounded for any constellation

QAM carved from $\mathbb{Z} + j\mathbb{Z}$. Indeed, it is shown in [YW03] that irrespective of the size of the constellation $\mathcal{B}_M$

$$\mathcal{B}_M = \{a + jb | - M + 1 \leq a, b \leq M - 1, a, b \text{ odd}\}, \tag{5.188}$$

the minimum determinant d_λ outlined by (5.20) is larger than 0.05. Note that the power of the constellation is not constrained to be ≤ 1 as previously. Taking a normalized constellation and calculating $\min_{\tilde{\mathbf{E}} \neq \mathbf{0}} \det(E_s\tilde{\mathbf{E}})$ would achieve the same result. Indeed, in (5.188), the average energy E_s is implicit in the constellation as well as in the codeword $\mathbf{C}$, whereas classically, $\mathbf{C}$ is normalized and the dependence on E_s is made explicit.

Actually, for any size of the constellation, the minimum determinant d_λ is equal to 0.05, being the largest achievable over the family of codes defined by two different rotations with $\varphi = 1$. Interestingly, the values of ϖ_1 and ϖ_2 are also optimal irrespective of the constellation size. The following property is thereby satisfied.

Proposition 5.20 *The tilted QAM code with $\varpi_1 = \frac{1}{2} \arctan\left(\frac{1}{2}\right)$ and $\varpi_2 = \frac{1}{2} \arctan (2)$ achieves the optimal diversity-multiplexing trade-off of Proposition 4.5 for QAM constellations, $n_t = T = 2$ and $g_s \in [0, \min\{2, n_r\}]$.*

PROOF: The code has non-vanishing determinants and therefore satisfies Proposition 5.2. ∎

It is important to note that this code is minimum delay ($T = n_t$) and still achieves the optimum diversity-multiplexing trade-off. From Proposition 5.1, this also shows that structured codes achieve the trade-off with a much smaller T than Gaussian random codes.

The performance achieved by the tilted-QAM code for various rates is shown in Figure 5.6, while lighter curves represent the outage probability behavior. Note that sphere decoding techniques are used in order to implement the ML decoding strategy with a reasonable complexity. Analogous to the Alamouti scheme in Figure 5.4, the tilted-QAM code achieves a diversity gain $g_d(g_s = 0)$ of 4. Unlike the Alamouti scheme, the curves are only separated by 6 dB, which corresponds to a multiplexing gain $g_s = 2$, i.e. 2 bits/s/Hz increase per 3 dB SNR increase. Note that this is the maximum achievable multiplexing gain $g_s = \min\{n_t, n_r\} = 2$. Furthermore, the tilted-QAM code performs quite close to the outage probability with a codeword length of only two, although it should be remembered that the outage probability is theoretically only achievable with infinite code length.

Dayal code
The Dayal code [DV05] has a unique rotation matrix

$$\mathcal{M}_1 = \mathcal{M}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos\varpi & \sin\varpi \\ -\sin\varpi & \cos\varpi \end{bmatrix}. \tag{5.189}$$

Maximization of the worst-case PEP leads to the optimal parameters $\varpi = \frac{1}{2} \arctan(2)$ and $\varphi = -j$, valid for all constellation sizes. As with to the tilted QAM code, this code has also a non-vanishing determinant, yielding the following proposition.
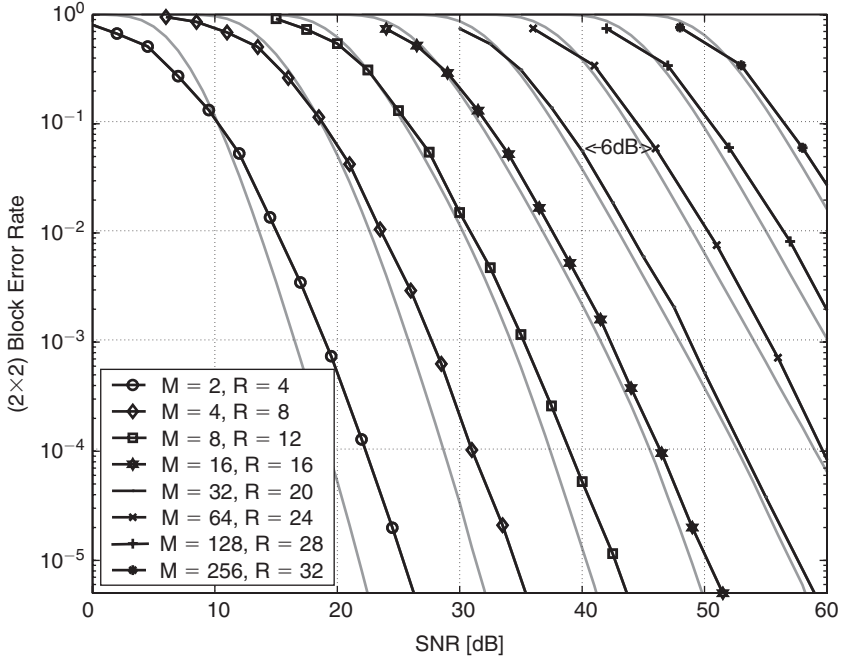
Figure 5.6 *Block error rate of the tilted-QAM code in i.i.d. Rayleigh slow fading channels with $n_r = 2$ in a $R = 2n_t \log_2(M) = 4, 8, \ldots, 32$-bit/s/Hz transmissions using M^2-QAM constellations of sizes $M = 2, 4, 8, \ldots, 256$ (courtesy of H. Yao [YW03])*

Proposition 5.21   *The Dayal code with $\varpi = \frac{1}{2} \arctan(2)$ and $\varphi = -j$ using QAM constellations achieves the optimal diversity-multiplexing trade-off of Proposition 5.1 for $n_t = T = 2$ and $g_s \in [0, \min\{2, n_r\}]$.*

However, this code offers a larger minimum determinant $d_\lambda = 0.2$ than the aforementioned codes, suggesting a reduced error probability. Figure 5.7 compares the Dayal code (with QPSK constellation) and the Alamouti scheme (with 16-QAM constellation to keep the same data rate) for one and two receive antennas in a 4-bit/s/Hz transmission (ML decoding is performed). As expected, the Dayal code outperforms the Alamouti scheme for $n_r = 2$. However, the opposite is observed for $n_r = 1$, as in that case, the Alamouti code is capacity optimal. This confirms once more that O-STBCs are attractive for MIMO systems with low rate and small arrays.

Golden code
The Golden code [BRV05] has been optimized by means of the algebraic number theory machinery. The codewords take the form

$$\mathbf{C} = \frac{1}{\sqrt{10}} \begin{bmatrix} \alpha(c_1 + c_2\theta) & \alpha(c_3 + c_4\theta) \\ j\bar{\alpha}(c_3 + c_4\bar{\theta}) & \bar{\alpha}(c_1 + c_2\bar{\theta}) \end{bmatrix} \tag{5.190}$$
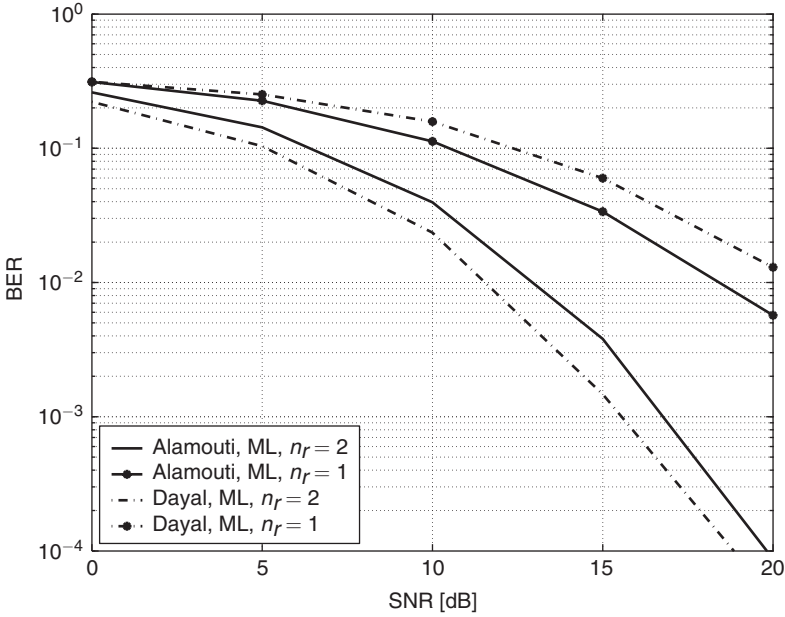
Figure 5.7   *Bit error rate (BER) of Dayal and Alamouti codes in i.i.d. Rayleigh slow fading channels with $n_r = 1$ and $n_r = 2$ in a 4-bit/s/Hz transmission*

where $\theta = \frac{1+\sqrt{5}}{2}$, $\bar{\theta} = \frac{1-\sqrt{5}}{2}$, $\alpha = 1 + j - j\theta$ and $\bar{\alpha} = 1 + j - j\bar{\theta}$. This code is known as the Golden code because its construction is based on the golden number. In our general algebraic coding framework, this code is equivalent to take $\varphi = 1$ and

$$\mathcal{M}_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} \alpha & \alpha\theta \\ \bar{\alpha} & \bar{\alpha}\bar{\theta} \end{bmatrix}, \ \mathcal{M}_2 = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 0 \\ 0 & j \end{bmatrix} \mathcal{M}_1. \tag{5.191}$$

The Golden code provides exactly the same performance as the Dayal code, since it can be obtained through left and right multiplications of the Dayal code by two complex diagonal and unitary matrices [DV05]. As with the tilted QAM and the Dayal codes, the Golden code satisfies the non-vanishing determinant criterion.

Proposition 5.22   *The Golden code achieves the optimal diversity-multiplexing trade-off of Proposition 4.5 for QAM constellations, $n_t = T = 2$ and $g_s \in [0, \min\{2, n_r\}]$.*

The Golden code is a particular case of a class of codes known as perfect space–time block codes [RBV04, ORBV06]. These codes are $n_t \times n_t$ linear block codes which are characterized by the following properties:

1. they are full rate ($r_s = n_t$)
2. they have a non-vanishing determinant
3. they are constructed from cyclic division algebra
4. they have a cubic constellation shaping (and are thus energy efficient).
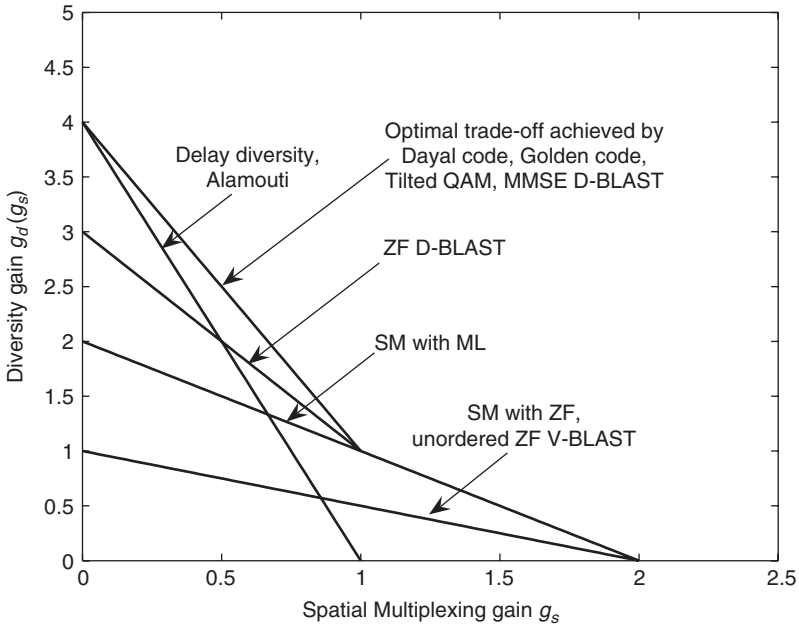
Figure 5.8 *Diversity-multiplexing trade-off achieved by several space–time codes in a 2×2 i.i.d. Rayleigh fading MIMO channel*

Other codes built from cyclic division algebra have also been proposed in [ERKP+06]. It is shown that cyclic division algebra-based space–time codes satisfying the non-vanishing determinant criterion also achieve the diversity-multiplexing trade-off (Proposition 5.2). Other constructions based on cyclic division algebra may be found in [SSRS03, BR03, KSR05].

Alternative code constructions
A few other codes achieving the diversity-multiplexing trade-off have been proposed. However, they do not usually satisfy the non-vanishing determinant criterion (unlike the codes described above). As an example, lattice space–time (LAST) codes [GCD04] present a structure similar to random Gaussian codes. Those codes are shown to achieve the diversity-multiplexing trade-off in i.i.d. Rayleigh channels with a decoder based on a generalized MMSE estimator.

5.5.8 Global performance comparison

In Figure 5.8, we illustrate the trade-off achieved by algebraic codes as well as by other codes presented in previous sections in a 2×2 i.i.d. Rayleigh fading channel.

Figure 5.9 compares the performance of some algebraic codes along with SM and Alamouti schemes. ML decoding is performed and a QPSK constellation is used, except for the
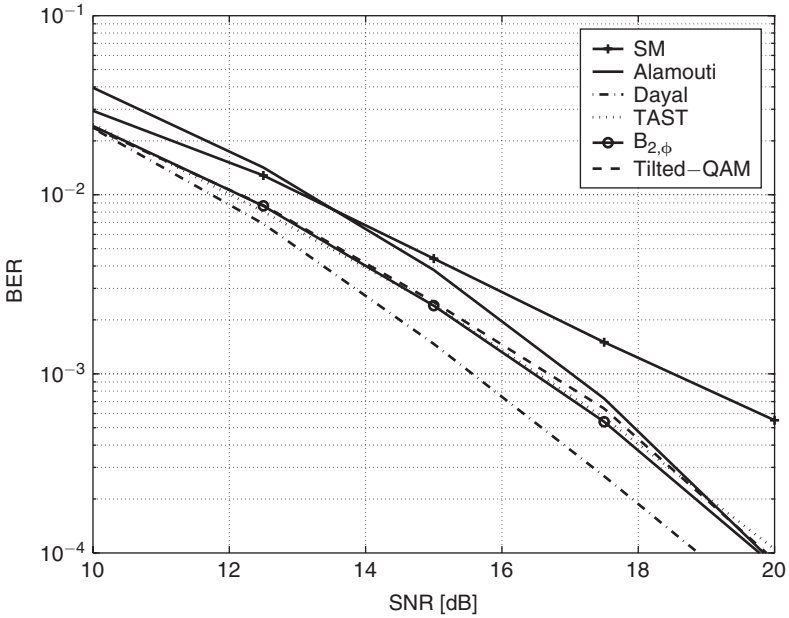
Figure 5.9　*Bit error rate (BER) of several algebraic space–time block codes in i.i.d. Rayleigh slow fading channels with $n_t = 2$ and $n_r = 2$ in a 4-bit/s/Hz transmission*

Alamouti code where a 16-QAM is used to keep the same data rate. All schemes are characterized by very similar error rates, with a slight advantage for the Dayal code. Moreover, the error probabilities in Figure 5.9 are also similar to those obtained above with Heath and Sandhu linear dispersion codes. Yet, the advantage of the tilted QAM and Dayal codes is the guarantee that they will perform well for any rate, and thus for any size of the QAM constellation. This does not necessarily hold true for the Sandhu code, even if this code construction does not depend upon the constellation.

An important drawback of algebraic and linear dispersion codes is the extremely high receiver complexity. The ML decoding necessitates an exhaustive search, that becomes cumbersome as the number of transmit antennas and the constellation size increase. Even if a receiver scheme such as the sphere decoder might reduce the complexity, it remains high. Naturally, algebraic and linear dispersion codes being linear codes, suboptimal receivers such as those proposed for SM can be used to help reduce the complexity. The problem is that those codes have been originally developed with ML decoding in mind and lose their advantage with suboptimal linear receivers. This is perfectly illustrated in Figure 5.10 for an ordered successive interference canceler based on ZF filtering. When used with SM, this performs ordered ZF V-BLAST. Naturally, since the Alamouti code is orthogonal, it achieves the same performance with ZF filtering as with ML decoding. However, the performance of the Dayal code with ordered ZF-SIC vanishes totally and becomes similar to that of V-BLAST. It is interesting to observe that high-rate codes perform better than low-rate
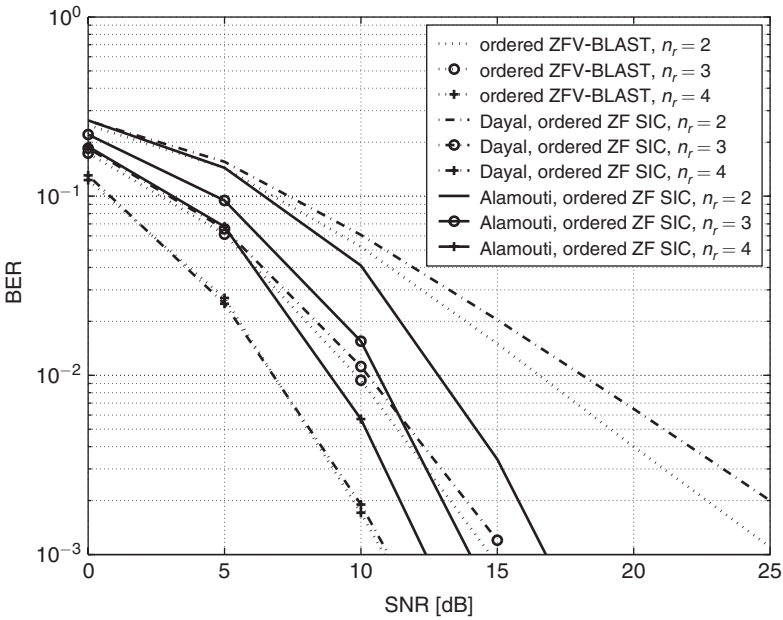
Figure 5.10 *Bit error rate (BER) of SM, Dayal and Alamouti codes in i.i.d. Rayleigh slow fading channels with $n_r = 2, 3, 4$ in a 4-bit/s/Hz transmission*

codes with suboptimal receivers as the number of receive antennas increases and the SNR decreases. This demonstrates how significantly the code design depends on the receiver.

5.6 Space–time trellis coding

By contrast to the popular belief, space–time trellis codes (STTCs) were actually the first proposed space–time codes [TSC98]. They can be thought of as extensions of conventional convolutional codes to n_t-antenna transmit arrays. While the encoder output in STBCs is only a function of the input bits, the encoder output in space–time trellis coding is a function of the input bits and the state of the encoder, the latter depending upon the previous input bits. This memory inherent to the trellis approach provides an additional coding gain with respect to space–time block coding.

5.6.1 Space–time trellis codes

Let us now describe the encoder strategies and discuss the performance of several popular STTCs as a function of the design criterion, the number of receive antennas and the complexity.

STTC encoder/decoder and generator matrix
At each time instant k, B bits are fed to the encoder. Analogous to classical convolutional encoders, each of the B bits enters a trellis encoder consisting in a shift register whose
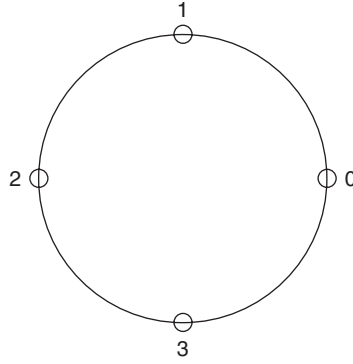
Figure 5.11   *Labeling of the QPSK constellation*

memory size is denoted as ν_j. The output of the STTC encoder is a function of the B input bits and the state of the STTC encoder, i.e. the ν $\left(=\sum_{j=1}^{B} \nu_j\right)$ previous bits. The number of states encountered by the encoder is thus equal to 2^ν. The encoder starts and finishes in state 0. Denoting by the ith column of $\mathbf{B}$ the $B \times 1$ input bit vector entering the encoder at time $k - i + 1$, the output of the encoder $\mathbf{x}_k$ (of size $n_t \times 1$) is given by

$$\mathbf{x}_k = \left[\mathbf{G}^T \text{vec}(\mathbf{B})\right] \quad \text{mod } M \tag{5.192}$$

where M is the size of the symbol constellation and $\mathbf{G}$ is the code generator matrix. The ith column of $\mathbf{G}$, denoted as $\mathbf{G}(:, i)$, can be seen as the generator matrix of the trellis code used on the ith transmit antenna. Elements of $\mathbf{G}$ are chosen in the set $\{0, \ldots, M - 1\}$. Each component of the output vector $\mathbf{x}_k$ is then mapped onto a constellation symbol to form the transmitted codeword $\mathbf{c}_k$. An example of mapping for a QPSK constellation is illustrated in Figure 5.11, where the elements of the output vector of the encoder $\mathbf{x}_k$ are mapped to symbols using the following labeling

$$\begin{aligned}
0 &\to 1 \\
1 &\to e^{j\frac{\pi}{2}} \\
2 &\to e^{j\frac{2\pi}{2}} \\
3 &\to e^{j\frac{3\pi}{2}}
\end{aligned} \tag{5.193}$$

and the corresponding codeword $\mathbf{C}$ will be transmitted through the channel.

At the receiver, the transmitted frame is decoded using maximum-likelihood sequence estimation (MLSE). This is carried out by means of the well-known Viterbi algorithm.

An illustration of a space–time trellis encoder for two transmit antenna is displayed in Figure 5.12. If the memory orders $\nu_1 = 1$ and $\nu_2 = 2$ in the Figure are considered, the code generator matrix $\mathbf{G}$ is given by

$$\mathbf{G}^T = \begin{bmatrix} a_0^1 & b_0^1 & a_1^1 & b_1^1 & b_2^1 \\ a_0^2 & b_0^2 & a_1^2 & b_1^2 & b_2^2 \end{bmatrix}. \tag{5.194}$$
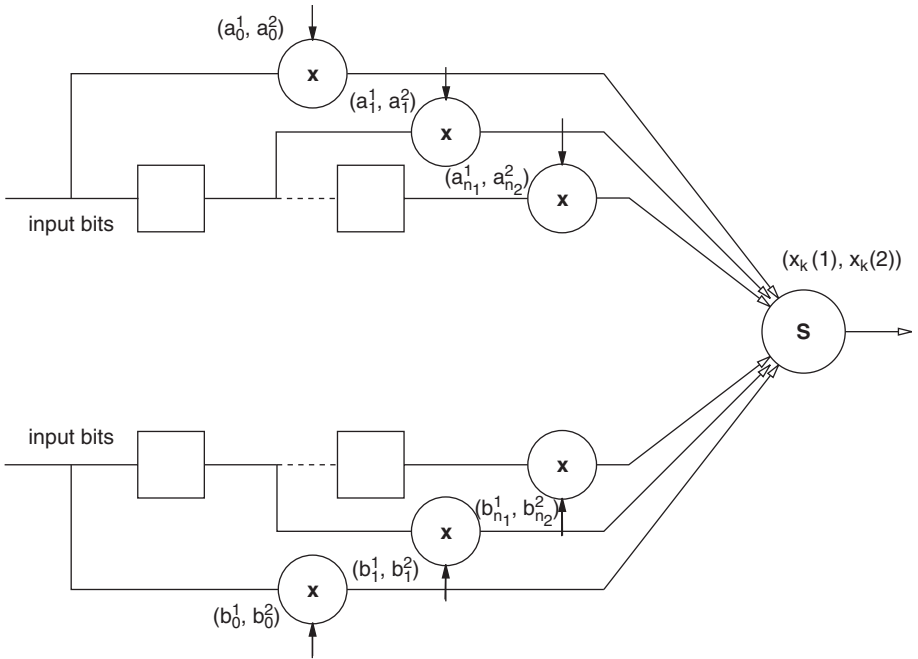
Figure 5.12 *STTC encoder for two transmit antennas*

Since $\nu = 3$, the number of states generated by this code is equal to 8.

Trellis representation

A common and alternative fashion to represent STTCs is to use a trellis, as illustrated for several QPSK based 4-state and 8-state codes with two transmit antennas (Figures 5.13 and 5.14; the labeling of Figure 5.11 is used). These trellis should be understood as follows. The number of nodes in the trellis corresponds to the number of states. In the following examples, there are four nodes in Figure 5.13 and eight nodes in Figure 5.14, corresponding to four and eight states, respectively. The trellis is made of two parallel columns of nodes, linked with lines. The left column represents the current states while the right column stands for the next states. A line between current node p and next node q indicates that the encoder is authorized to evolve from state p to state q if the correct input enters the encoder. Once in a node (i.e. a state), the possible outputs of the encoder are displayed on the left-hand side of the current state. There are 2^B possible groups of outputs since B bits simultaneously enter the encoder, each B-tuple corresponding to a point on the constellation used (see Figure 5.11). Assuming the nth constellation point enters the encoder, the nth group of outputs of the current state is made of the symbols sent on the antennas, the mth entry of the group corresponding to the symbol sent on the mth antenna. Afterward, the encoder evolves to the next state, given by the state linked to the current state via the nth line when counting clockwise.
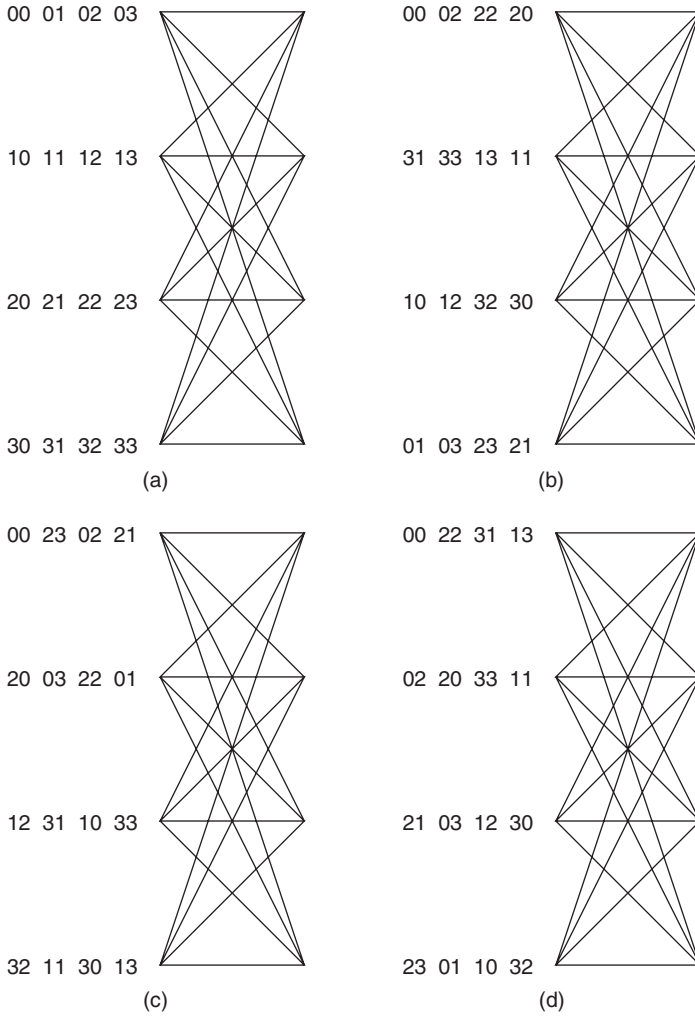
Figure 5.13   *Trellis representation of QPSK 4-state 2 bits/s/Hz space–time trellis codes for two transmit antennas: (a) 'TSC' code (delay-diversity code) [Wit93, SW94, TSC98]; (b) 'BBH' code [BBH00]; (c) 'CYV' code [CYV01]; (d) 'FVY' code [FVY01]*

Let us consider the following example to understand how to read the trellis.

Example 5.16   *Figure 5.13(a) is the trellis representation of the 4-state STTC for two transmit antennas proposed in [Wit93, SW94, TSC98] and denoted here as 'TSC'. Its generator matrix is given by*

$$\mathbf{G}^T = \begin{bmatrix} 0 & 0 & 2 & 1 \\ 2 & 1 & 0 & 0 \end{bmatrix}. \tag{5.195}$$

Let us understand how to read the trellis of this code. On the first node (i.e. the first state), the possible outputs are given by 00, 01, 02 and 03. This means that if the input symbols are 0, 1, 2 or 3, the output symbols are respectively 00 (0 on antenna 1 and 40 on antenna 2), 01 (0 on antenna 1 and 1 on antenna 2), 02 (0 on antenna 1 and 2 on antenna 2) or 03 (0 on antenna 1 and 3 on antenna 2). Moreover, the next state will be respectively 0, 1, 2 or 3. It is easy to realize that this trellis corresponds to a delay-diversity coding since the effect of the trellis is to send at time instant $k+1$ on antenna 1 the symbol sent at time instant k on antenna 2.

Trade-off between rate, diversity and complexity

The following result [TSC98] highlights the complexity of a trellis code as a function of its achievable diversity gain and its rate.

Proposition 5.23 *A STTC with a rate B bits/s/Hz and a minimum rank r_{min} holds at least $2^{B(r_{min}-1)}$ states in its trellis.*

The proof is not given here, but the interested reader is referred to [TSC98] for more details.

Delay-diversity scheme

Example 5.16 has introduced a very popular code. The latter is also known as the delay-diversity code because the codewords can be expressed as

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & c_2 & \dots & c_{T-1} & 0 \\ 0 & c_1 & c_2 & \dots & c_{T-1} \end{bmatrix}, \tag{5.196}$$

illustrating that the sequence of symbols c_k transmitted on the first antenna also appears on the second antenna with a delay of one symbol duration. This code actually transforms the spatial diversity into frequency diversity by converting the flat fading channel into a frequency selective channel. Its effect is indeed the same as if $\mathbf{C}(1,1:T-1)$ was transmitted over a frequency selective channel characterized by two taps of equal average power separated by one symbol duration. Equalization based on MLSE perfectly exploits the full frequency diversity offered by such a channel.

This code is not limited to QPSK and two transmit antennas as in Example 5.16, but can be used with any symbol constellations. Moreover, it is easily scalable as the number of transmit antennas changes. On antenna m, it is sufficient to delay the sequence of symbols $\{c_k\}$ by $m-1$ symbol durations. The equivalent channel looks like a frequency-selective channel with n_t taps.

The following proposition investigates how delay-diversity performs from the diversity-multiplexing trade-off point of view.

Proposition 5.24   *In the asymptotic case of an infinite codeword length T and a finite number of transmit antenna, the diversity-multiplexing trade-off achieved by delay-diversity coding with ML decoding and QAM constellation over i.i.d. Rayleigh fading channels is given by*

$$g_d(g_s) = n_r n_t (1 - g_s), \ g_s \in [0, 1]. \tag{5.197}$$

PROOF: Since delay-diversity converts a multi-antenna flat fading channel into a single antenna frequency-selective channel, the announced proposition results directly from [Gro05] where an ML detector is shown to achieve a diversity-multiplexing trade-off of $L(1 - g_s)$ over a L-tap frequency-selective channel.   ∎

Note that for a finite codeword length T, the effective transmission rate is reduced by $(T - n_t + 1)/T$ due to the $n_t - 1$ zeroes introduced at the end of the $T - n_t + 1$ symbols sent on the first antenna (in (5.196), one zero is introduced). This loss is similar to the initialization loss in D-BLAST. Hence, taking into account this loss, the actual trade-off achieved by delay-diversity is given by $g_d(T/(T - n_t + 1)g_s)$, i.e. the effective diversity gain achieved by a delay-diversity with codeword length T at a multiplexing gain g_s is the diversity gain that would have been obtained with an infinite codeword length at the multiplexing gain $T/(T - n_t + 1)g_s$.

Proposition 5.25   *In the asymptotic case of an infinite codeword length T and a finite number of transmit antenna, the delay-diversity code with QAM constellation achieves the optimal diversity-multiplexing trade-off in i.i.d. Rayleigh fading MISO channels.*

For $n_t = 2$, the diversity-multiplexing trade-off achieved by delay-diversity is similar to the Alamouti scheme (Figure 5.8). For $n_t > 2$, the delay-diversity code performs better than O-STBCs since the spatial multiplexing rate remains equal to one in delay-diversity for any transmit array size. However, delay-diversity coding requires a much longer codeword length than the Alamouti scheme.

Additional examples
Below are more examples of codes designed according to different design criteria for various numbers of states.

Example 5.17   (*Rank-determinant criterion based 4-state code*) *Figure 5.13(b) depicts the trellis representation of the 4-state STTC for two transmit antennas proposed in [BBH00] and denoted as 'BBH'. The corresponding generator matrix is given by*

$$\mathbf{G}^T = \begin{bmatrix} 2 & 0 & 1 & 3 \\ 2 & 2 & 0 & 1 \end{bmatrix}. \tag{5.198}$$

This code has been optimized for a given rate through a numerical search based on the rank-determinant criterion.

Example 5.18 *(**Rank-trace criterion based 4-state code**) Figure 5.13(c) is the trellis representation of the 4-state STTC for two transmit antennas proposed in [CYV01] and denoted as 'CYV'. Its generator matrix is given by*

$$\mathbf{G}^T = \begin{bmatrix} 0 & 2 & 1 & 2 \\ 2 & 3 & 2 & 0 \end{bmatrix}. \tag{5.199}$$

This code has also been optimized for a given rate through computer search, but according to the rank-trace criterion.

Example 5.19 *(**Distance-product criterion based 4-state code**) Figure 5.13(d) is the trellis representation of the 4-state STTC for two transmit antennas proposed in [FVY01], denoted as 'FVY'*

$$\mathbf{G}^T = \begin{bmatrix} 3 & 2 & 2 & 0 \\ 1 & 2 & 1 & 2 \end{bmatrix}. \tag{5.200}$$

The numerical optimization for a given rate is based on the distance-product criterion, hence with fast fading channels in mind.

Example 5.20 *(**Rank-trace criterion based 8-state code**) Figure 5.14 gives the trellis representation of an 8-state STTC for two transmit antennas. Its generator matrix is given by*

$$\mathbf{G}^T = \begin{bmatrix} 2 & 0 & 3 & 3 & 2 \\ 3 & 2 & 2 & 1 & 0 \end{bmatrix}. \tag{5.201}$$

This code has been optimized for a given rate through a numerical search using the rank-trace criterion.

Error performance
Let us finally analyze how STTCs perform as a function the design criterion, the number of states and the receive array size. In Table 5.1, several characteristics of the aforementioned 4-state and the 8-state STTCs are displayed.

In Figure 5.15, we compare the performance of the QPSK based 4-state 'TSC', 'BBH' and 'CYV' codes for various numbers of receive antennas over i.i.d. Rayleigh fading channels (the length of a frame is chosen as equal to 130 symbols). The gain offered by the 'CYV'

00 02 23 21

31 33 10 12

32 30 11 13

23 21 02 00

20 22 03 01

11 13 30 32

12 10 31 33

03 01 22 20
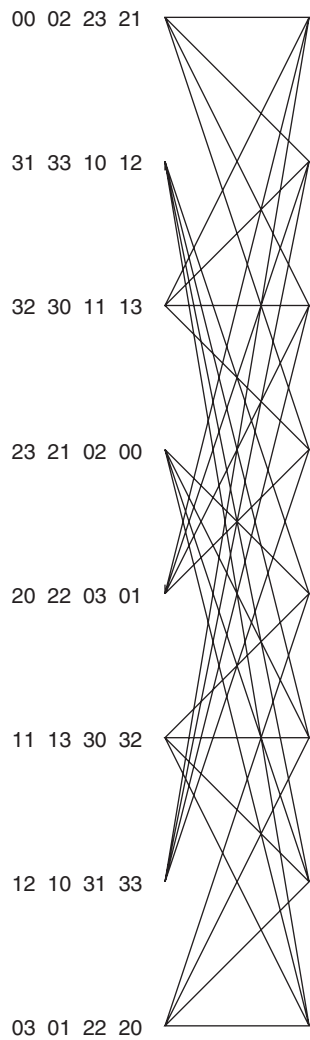
Figure 5.14   *Trellis representation and generator matrix of a new QPSK 8-state 2-transmit antenna space–time trellis code*

Table 5.1   *Performance of various STTCs*

	r_{min}	d_λ	d_e	L_{min}	d_p
4-state 'TSC'	2	4	4	2	4
4-state 'BBH'	2	8	6	2	8
4-state 'CYV'	2	4	10	2	24
4-state 'FVY'	2	4	10	2	24
8-state 'TSC'	2	12	8	2	16
8-state 'CYV'	2	8	12	2	48
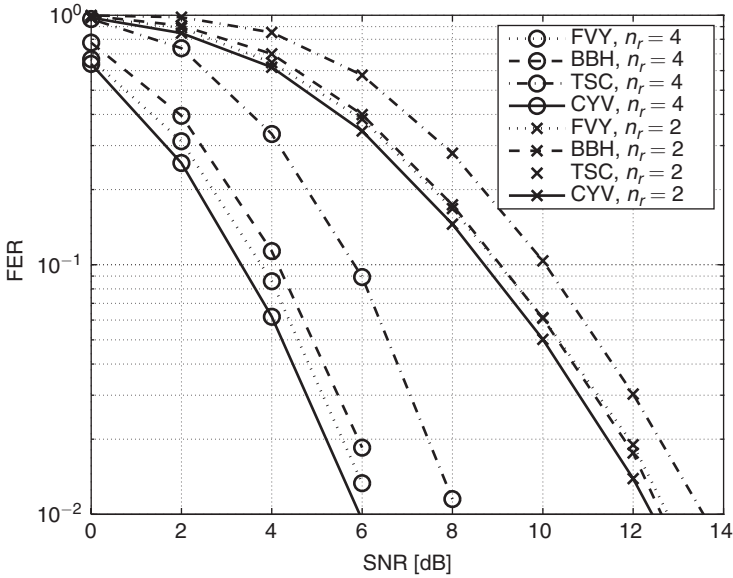
Figure 5.15 *Frame error rate of several 4-state STTC in i.i.d. Rayleigh slow fading channels with* $n_t = 2$ *and* $n_r = 2, 4$

code grows as the number of receive antennas n_r increases. The channel becoming AWGN as n_r increases, it is natural that codes with large minimum Euclidean distance d_e perform better. Consequently, for large n_r, STTC designs based upon the rank-trace criterion offer larger coding gains than rank-determinant criterion-based designs.

Let us now focus on the 4-state and 8-state 'CYV' and 'TSC' codes. The 'CYV' codes have been optimized for slow fading channels using the rank-trace criterion. We note that the larger the number of states, the larger d_e, suggesting that the larger the number of states, the larger the coding gain. Similarly, the 'TSC' codes have been designed based on the rank-determinant criterion. The larger the number of states, the larger d_λ. The number of states does not impact the diversity gain as long as the code is full rank, but only the coding gain in slow fading channels. This is confirmed by Figure 5.16 for i.i.d. Rayleigh slow fading channels with four receive antennas.

In fast fading channels, the situation is quite different. On the one hand, for well designed codes, the minimum effective length of the code L_{min} increases with the number of states. The larger the number of states, the larger the diversity and coding gains. On the other hand, the presence of multiple transmit antennas does not increase the diversity gain but only the coding gain, i.e. d_p. Figure 5.17 confirms that codes with larger product distances achieve better performance and that the diversity gain increases with the effective length (and so with the number of states). Note that for simulation, we assume that each frame consists here of 128 symbols out of each transmit antenna.
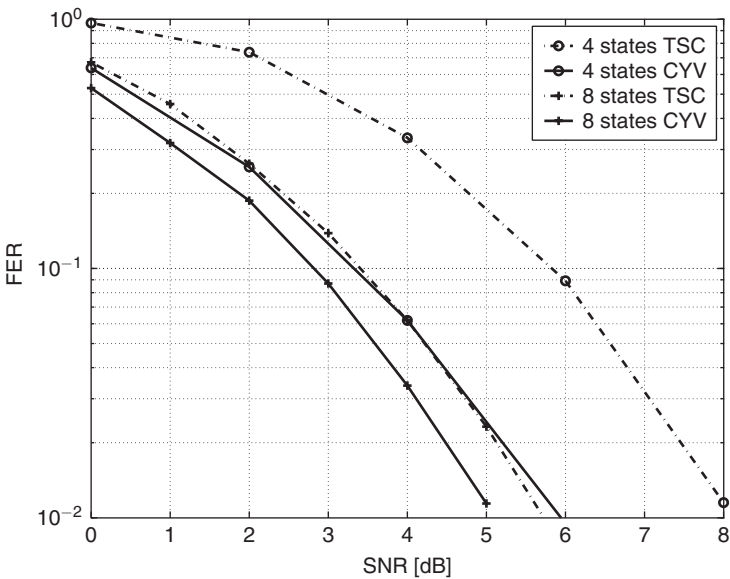
Figure 5.16 *Frame error rate of 4-state and 8-state 'CYV' and 'TSC' codes in i.i.d. Rayleigh slow fading channels with $n_t = 2$ and $n_r = 4$*
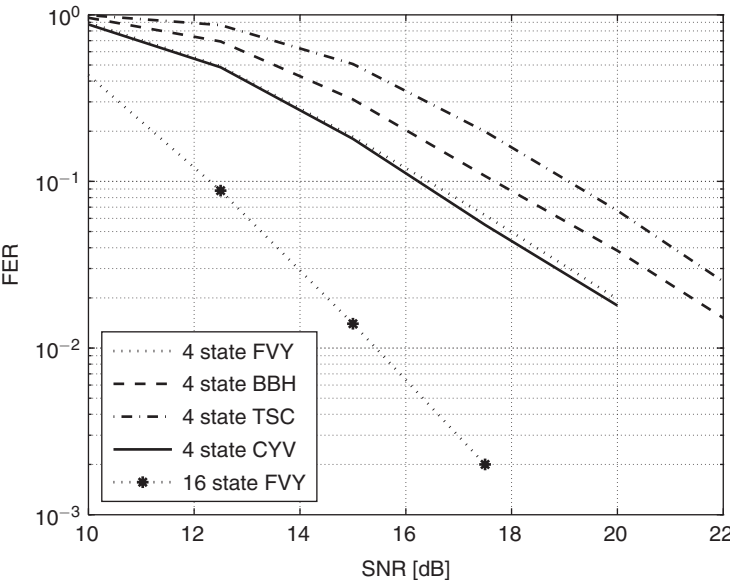
Figure 5.17 *Frame error rate of several 4-state STTC in i.i.d. fast fading channels with $n_t = 2$ and $n_r = 1$*

By contrast to O-STBCs, STTCs exhibit a diversity gain together with a coding gain inherent to the trellis. However, the receiver complexity is increased with respect to O-STBCs, as the decoding complexity grows exponentially with the number of states.

5.6.2 Super-orthogonal space–time trellis codes

The previous sections have analyzed two important classes of codes, O-STBC and STTC. On the one hand, O-STBC extracts the full diversity and allows for a very simple decoding scheme but lacks in providing coding gain. On the other hand, STTCs provide full diversity and coding gains but at the cost of higher decoding complexity. To further improve the coding gain of orthogonal coding schemes, an outer trellis code may be concatenated with an O-STBC. The latter may then be considered as the inner code, while the former may rely on trellis coded modulation (TCM) designs for AWGN channels.

In classical SIMO transmissions, the goal of the outer code is to select a point in the signal constellation based on the current state and the input bits. When concatenating an outer code with an O-STBC, the selected point actually corresponds to one of the possible O-STBC matrices. These matrices can be thought of as high dimensional points, that are also selected based on the current state and the input bits. However, the simple concatenation of an outer code with an O-STBC [SHP01, SF02b] incurs a rate loss. Indeed, the concatenated scheme cannot achieve a rate one transmission, since an O-STBC is at most rate one (achieved by Alamouti scheme) and the outer code introduces redundancy.

To increase the rate, one possibility is to increase the cardinality of the inner O-STBC, i.e. to enlarge the set of O-STBC matrices. As an example, consider the Alamouti code usually described as

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix}. \tag{5.202}$$

Yet, all orthogonal structures that use c_1, c_2 as well as their conjugates with positive or negative signs given by

$$\begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix}, \begin{bmatrix} -c_1 & c_2^* \\ c_2 & c_1^* \end{bmatrix}, \begin{bmatrix} c_1 & c_2^* \\ -c_2 & c_1^* \end{bmatrix}, \begin{bmatrix} c_1 & c_2^* \\ c_2 & -c_1^* \end{bmatrix},$$

$$\begin{bmatrix} -c_1 & c_2^* \\ -c_2 & -c_1^* \end{bmatrix}, \begin{bmatrix} -c_1 & -c_2^* \\ c_2 & -c_1^* \end{bmatrix}, \begin{bmatrix} c_1 & -c_2^* \\ -c_2 & -c_1^* \end{bmatrix}, \begin{bmatrix} -c_1 & -c_2^* \\ -c_2 & c_1^* \end{bmatrix},$$

enjoy similar properties as the original scheme. If the constellation is fixed, all possible orthogonal matrices cannot be created based on only one of these structures. For instance, with BPSK, the generation of all matrices requires two of the above structures. The idea

of super-orthogonal STTC [JS03, SF02a, IMYL01] is to pick up the multi-dimensional point in this enlarged set of O-STBC matrices. The underlying principle is reminiscent of Ungerboeck's idea [BDMS91], consisting in expanding signal constellations. Here, there is no expansion of the signal constellations but of the orthogonal matrices [JS03]. This makes it possible to design a rate-one trellis code that extracts the full diversity with an improved coding gain of about 2 dB with regard to STTC.

In [JS03], the set of orthogonal structures considered for $n_t = 2$ is expressed as

$$\mathcal{O}(c_1, c_2, \theta) = \begin{bmatrix} e^{j\theta} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix}. \tag{5.203}$$

Note that the multiplication by the diagonal unitary matrix does not modify the code properties but enlarges the set of orthogonal matrices. Taking $\theta = 0$, we obtain the O-STBC structure of (5.202). The rotation parameter θ is chosen such that the constellation is not expanded. Therefore, for BPSK, θ is chosen in the set $\{0, \pi\}$; for QPSK, it is chosen in the set $\{0, \pi/2, \pi, 3\pi/2\}$.

Analogous to TCM, set partitioning for orthogonal codes is performed analogous to Ungerboeck's set partitioning [BDMS91]. Nevertheless, instead of using an Euclidean distance, the set partitioning is based on the determinant of the error matrix $\tilde{\mathbf{E}}$ between two O-STBC codewords (remember that this is the coding gain measure for space–time coded transmissions).

The code construction is then based on the following observations. Two codewords that originate from the same O-STBC structure (i.e. with a same θ) are full rank and therefore extract the full diversity. Two codewords issued from different O-STBC structures (i.e. with a different θ) may not have full-rank error matrices and do not necessarily extract the full diversity. Hence, the code construction proposed in [JS03] guarantees that all pairs of codewords diverging from or merging to a state share the same O-STBC structure. Codewords that do not share the same O-STBC structure are assigned to different states. Consequently, the resulting code achieves the full diversity.

This is only a very brief overview of the principle of super-orthogonal trellis coding. Interested readers are referred to [JS03, SF02a, IMYL01] for more detailed discussions. Note that similar ideas have been exploited for alternative inner codes. Super quasi-orthogonal trellis coding schemes for quasi-orthogonal inner codes can be found in [JH05], while [HVBon] has proposed the concatenation of a trellis code with the Golden code.

6

Error probability in real-world MIMO channels

Chapter 5 has dealt with the design of space–time codes in i.i.d. Rayleigh fading channels. However, as highlighted in Chapters 2 and 3, real-world channels span a large variety of propagation conditions and some environments may highly deviate from the i.i.d. Rayleigh fading scenario. It is thus of great importance to understand

- how codes developed under the *i.i.d. Rayleigh* assumption behave in more realistic propagation conditions, i.e. how these codes are affected by non-ideal propagation conditions
- how a more adapted design criterion might significantly improve their performance.

In Section 6.1, we start from a physical perspective to understand how the channel scattering richness at the transmitter affects the performance of space–time codes. Through the application of intuitive concepts, we point out the problems faced by classical designs. These concepts are then formalized in Sections 6.2 to 6.5 corresponding to various realistic fading channels. Finally, Section 6.6 summarizes the important ideas of this chapter, which form the basis for developing the design criteria of Chapter 7.

6.1 A conditional pairwise error probability approach

6.1.1 Degenerate channels

On the one hand, we know from (5.3) in Chapter 5 that the conditional pairwise error probability (PEP) is a function of

$$\sum_{k=0}^{T-1} \left\| \mathbf{H}_k \left(\mathbf{c}_k - \mathbf{e}_k \right) \right\|_F^2 = \sum_{n=1}^{n_r} \sum_{k=0}^{T-1} \left| \sum_{m=1}^{n_t} \mathbf{H}_k(n,m)(\mathbf{C} - \mathbf{E})(m,k) \right|^2 . \qquad (6.1)$$

On the other hand, the conditional channel energy is a function of

$$\sum_{k=0}^{T-1} \left\| \mathbf{H}_k \mathbf{c}_k \right\|_F^2 = \sum_{n=1}^{n_r} \sum_{k=0}^{T-1} \left| \sum_{m=1}^{n_t} \mathbf{H}(n,m) \mathbf{C}(m,k) \right|^2 . \tag{6.2}$$

In order to understand how propagation mechanisms affect the performance of space–time coding, let us get some insight into how (6.1) and (6.2) behave as a function of the scattering richness around the transmitter. As noted in Chapters 2 and 3, the channel matrix $\mathbf{H}_k$ can always be decomposed into the summation of L multipath contributions $\mathbf{H}_k^{(l)}$. Assuming that each contribution is generated by a specific scatterer or a cluster of scatterers (i.e. each contribution is characterized by a given DoD) and using the narrowband array assumption (see Definition 2.5 in Section 2.1.5), we may write, similarly to (2.39)

$$\mathbf{H}_k = \sum_{l=0}^{L-1} \mathbf{H}_k^{(l)} = \sum_{l=0}^{L-1} \mathbf{H}_k^{(l)}(:,1) \mathbf{a}_t^T\left(\theta_{t,k}^{(l)} \right), \tag{6.3}$$

where $\mathbf{a}_t\left(\theta_{t,k}^{(l)} \right)$ is the transmit array response in the direction of departure $\theta_{t,k}^{(l)}$ (relative to the array axis) corresponding to the lth scatterer at the kth time instant. For a horizontal uniform linear array, similarly to (2.37)

$$\mathbf{a}_t(\theta_t) = \left[1 \ e^{-j\varphi_t} \ \dots \ e^{-j(n_t-1)\varphi_t} \right]^T \tag{6.4}$$

with $\varphi_t = 2\pi d_t/\lambda \cos \theta_t$, d_t being the inter-element spacing, λ, the wavelength and θ_t, the direction of departure. Combining (6.3) with (6.1) and (6.2), we have

$$\sum_{k=0}^{T-1} \left\| \mathbf{H}_k(\mathbf{c}_k - \mathbf{e}_k) \right\|_F^2 = \sum_{n=1}^{n_r} \sum_{k=0}^{T-1} \left| \sum_{l=0}^{L-1} \mathbf{H}_k^{(l)}(n,1) \right.$$
$$\left. \left[\sum_{m=1}^{n_t} \mathbf{a}_t\left(\theta_{t,k}^{(l)} \right)(m,1)(\mathbf{C}-\mathbf{E})(m,k) \right] \right|^2 \tag{6.5}$$

and

$$\sum_{k=0}^{T-1} \left\| \mathbf{H}_k \mathbf{c}_k \right\|_F^2 = \sum_{n=1}^{n_r} \sum_{k=0}^{T-1} \left| \sum_{l=0}^{L-1} \mathbf{H}_k^{(l)}(n,1) \left[\sum_{m=1}^{n_t} \mathbf{a}_t\left(\theta_{t,k}^{(l)} \right)(m,1)\mathbf{C}(m,k) \right] \right|^2 . \tag{6.6}$$

Equations (6.5) and (6.6) both illustrate that the original MIMO transmission can be considered as the SIMO transmission of an equivalent codeword, given at the kth time instant by

$$\sum_{m=1}^{n_t} \mathbf{a}_t\left(\theta_{t,k}^{(l)} \right)(m,1)\mathbf{C}(m,k) \tag{6.7}$$

while the equivalent error matrix reads as

$$\sum_{m=1}^{n_t} \mathbf{a}_t\big(\theta_{t,k}^{(l)}\big)(m,1)(\mathbf{C}-\mathbf{E})(m,k). \tag{6.8}$$

The equivalent codeword of (6.7) multiplies each path contribution. Instead of being defined as an equivalent codeword, it may be thought of as an effective array factor (as introduced in antenna theory), the wording effective meaning that it is function of the transmitted codewords. This indicates that, at every symbol period,

- the energy radiated in any direction varies as a function of the transmitted codewords
- for a given codeword and omnidirectional antennas, the radiated energy is not uniformly distributed in all directions, but may present maxima and minima in certain directions.

Moreover, reasoning similarly on (6.8), we may infer that the PEP evaluated for a given pair of $\mathbf{C},\mathbf{E}$ with $\mathbf{C}\neq\mathbf{E}$ depends on the distribution of the scatterers around the transmitter, especially when both the transmit angle-spread and the inter-element spacing are not large enough. Let us illustrate in Figure 6.1 this intuitive result by considering different transmit angle-spreads and inter-element spacings.

1. Assume first that the inter-element spacing is fixed and the channel becomes extremely rich in scatterers in such a way that the scatterers are located uniformly all around the transmitter (Figure 6.1(a)). Due to the limited inter-element spacing, the equivalent array factors mentioned above will have preferred directions depending on their weights, i.e. depending on the codeword and the error matrix respectively for the conditional channel energy and the conditional PEP. Since the scatterers are spatially uniformly distributed in rich scattering conditions, performance will not be affected by the directions of the maxima of the equivalent radiation patterns. This means that rotating the transmit array will not affect the performance, as the channel is spatially white for the transmitter.
2. Assume now that the channel only presents a few scatterers in sparse directions and that the inter-element spacing is very large (Figure 6.1(d)). A large inter-element spacing induces that the equivalent radiation patterns present a very large number of grating lobes. As the spacing increases, the number of grating lobes increases and the equivalent radiation patterns look as if they were omnidirectional. So the performance will not be affected by the positions of the scatterers. The transmit array may be rotated without modifying the performance.
3. If the inter-element spacing is large and the channel is rich in scatterers, the radiation patterns are omnidirectional and the propagation environment is spatially white (Figure 6.1(c)). It is then straightforward that irrespective of the scatterer spatial distribution and the array orientation, the performance is always good. This is an ideal propagation scenario, and the i.i.d. Rayleigh fading model encountered in previous chapters falls into that category.
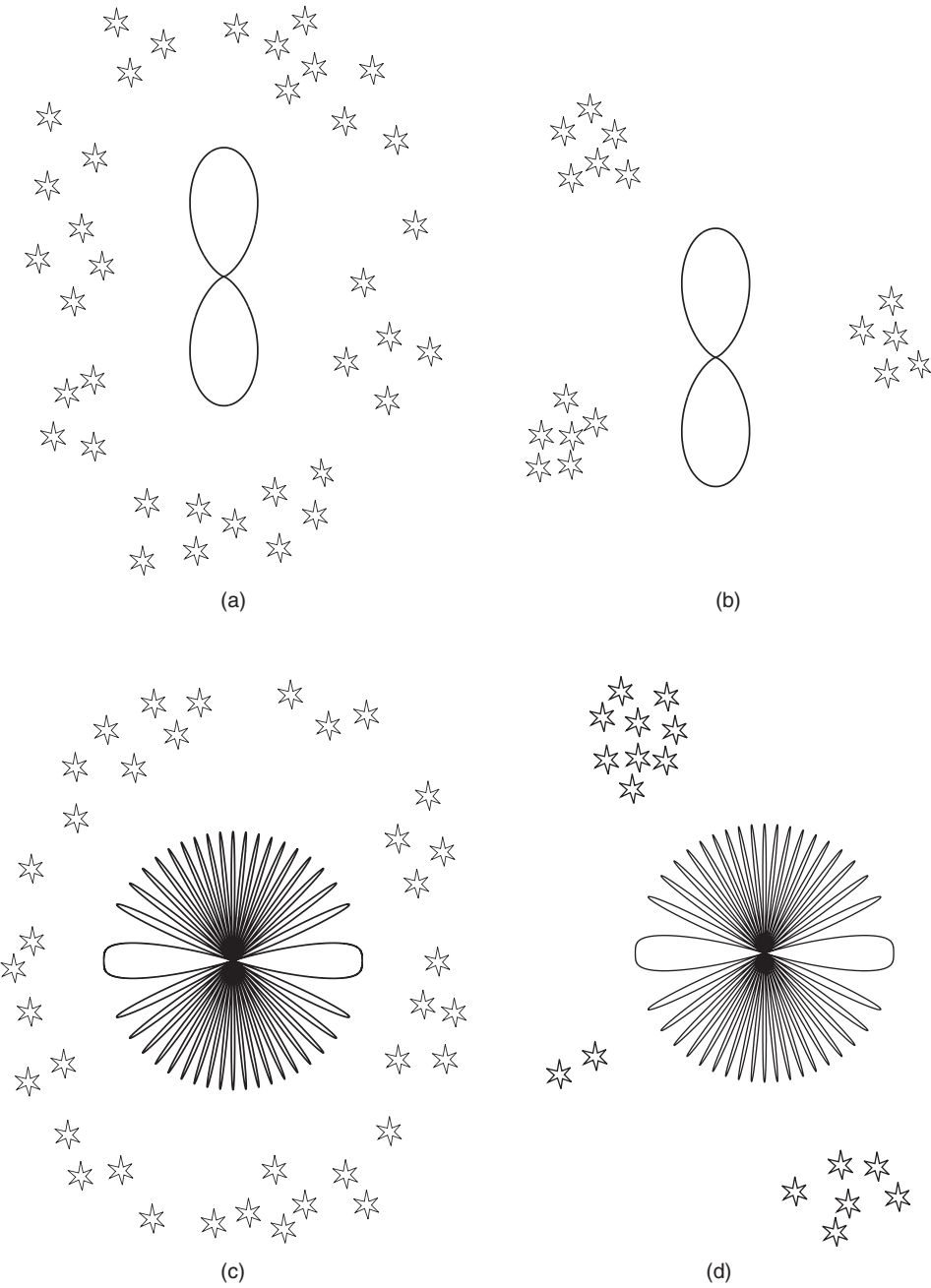
Figure 6.1   *Visualization of the impact of the scattering richness and inter-element spacing on MIMO system performance*

4. Problems occur when both the scattering richness and the inter-element spacing are not large enough (Figure 6.1(b)). Performance depends on how the maxima of the equivalent radiation patterns match the positions of the scatterers. If those maxima are oriented toward the scatterers, good performance may be expected. On the other hand, if radiation patterns are oriented toward directions where no scatterers are located, or equivalently, if the directions of scatterers correspond to the minima of the equivalent radiation patterns, degradation of performance is expected.

As an outcome of the above discussion, we categorize an environment where the angle-spread at the transmitter is very small by the following definition, which will be used extensively in the next chapters.

Definition 6.1 *A MIMO channel is said to be degenerate in the direction of departure* θ_t *if all scatterers surrounding the transmitter are located along the same direction* θ_t.

In practice, the channel gains vary much faster than the inner structure of the channel (see Section 3.3.3). This means that while the fading gains may vary much on the duration of several frames, the directions of departure of the different scatterers remain static. We may thus assume that $\theta_{t,k}^{(l)} = \theta_t^{(l)}$ for all k. As outlined above, situations become problematic when the transmit angle-spread decreases, leading to detrimental interactions between the channel and the code. Since robust codes should perform well whatever the channel conditions, they should ideally be robust even with very small transmit angle-spreads. The worst situation occurs when the channel becomes degenerate in any direction θ_t. In this case $\mathbf{a}_t(\theta_t^{(l)}) = \mathbf{a}_t(\theta_t)$ for all l and we get

$$\sum_{k=0}^{T-1} \left\| \mathbf{H}_k(\mathbf{c}_k - \mathbf{e}_k) \right\|_F^2 = \sum_{k=0}^{T-1} \left| (\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_t) \right|^2 \left[\sum_{n=1}^{n_r} \left| \mathbf{H}_k(n,1) \right|^2 \right] \qquad (6.9)$$

$$\sum_{k=0}^{T-1} \left\| \mathbf{H}_k \mathbf{c}_k \right\|_F^2 = \sum_{k=0}^{T-1} \left| \mathbf{c}_k^T \mathbf{a}_t(\theta_t) \right|^2 \left[\sum_{n=1}^{n_r} \left| \mathbf{H}_k(n,1) \right|^2 \right]. \qquad (6.10)$$

Equations (6.9) and (6.10) express that in the presence of small angle-spread at the transmit side, the MIMO channel degenerates into a SIMO channel where the $1 \times T$ transmitted codeword is given by $\mathbf{a}_t^T(\theta_t)\mathbf{C}$. Since a space-time code designed for i.i.d. channels is only concerned with $\mathbf{C}$ and $\mathbf{E}$, its interaction with $\mathbf{a}_t(\theta_t)$ is not taken into account. So, there is no guarantee that a space–time code designed for i.i.d. channels will perform well on arbitrary (correlated) channels.

6.1.2 The Spatial Multiplexing example

In order to illustrate our physical interpretation about the impact of channel correlations on the performance of space–time coding, we consider a Spatial Multiplexing scheme with
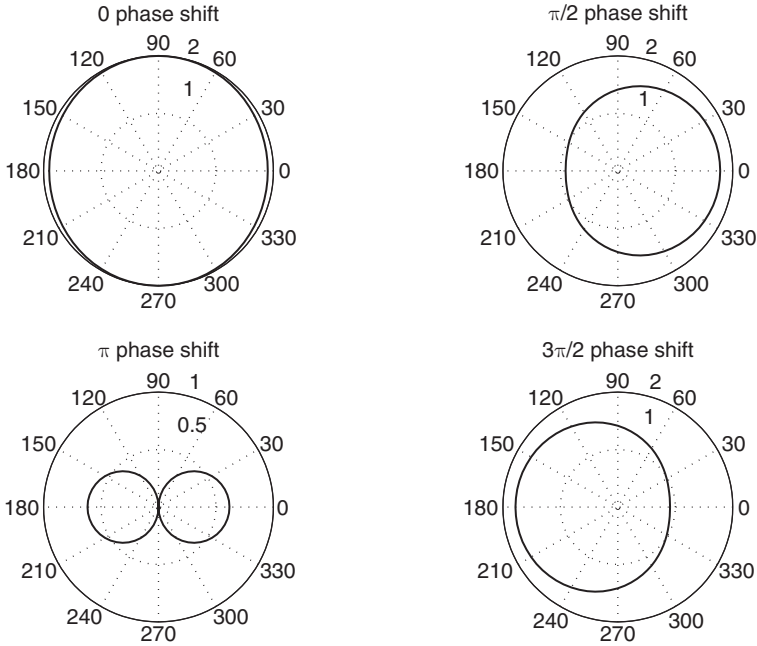
Figure 6.2 $G_t(\theta_t|\mathbf{c}_k)$ *(θ_t varying over 360°) for the four possible phase shifts between two transmitted QPSK symbols and MT inter-element spacing $d_t/\lambda = 0.1$*

2 transmit antennas, so $\mathbf{c}_k = [c_1[k] \;\; c_2[k]]^T$. Assuming a horizontal linear array, we define an effective array pattern $G_t(\theta_t|\mathbf{c}_k)$ as follows,

$$|\mathbf{c}_k^T\mathbf{a}_t(\theta_t)|^2 = c_1[k]\underbrace{\left[1 + \frac{c_2[k]}{c_1[k]}e^{-2\pi j\frac{d_t}{\lambda}\cos\theta_t}\right]}_{G_t(\theta_t|\mathbf{c}_k)}. \qquad (6.11)$$

Figures 6.2 and 6.3 depict $G_t(\theta_t|\mathbf{c}_k)$ as a function of θ_t (varying over 360°) for the 4 possible phase shifts between two transmitted QPSK symbols, respectively for inter-element spacings d_t/λ of 0.1 and 0.5 at the mobile terminal (MT). In a broadside configuration, the direction of the base station (BS) is given by $\theta_t = 90°$ (or 270°).

Figure 6.4 shows simulation results of the symbol error rate of a maximum likelihood (ML) detector as a function of the phase shift between the two simultaneously transmitted QPSK symbols at a SNR of 5 dB. We consider the combined elliptical-ring model of Section 2.3.3, applied to a 2-GHz uplink, with broadside antenna configurations at each side and a transmitter-receiver distance of 1 km. The antennas are two half-wavelength dipoles oriented side by side. At the BS, an inter-element distance of 20λ is assumed. No coherent component and a local scattering ratio (LSR) of 0.4 are assumed. For such propagation parameters, many scatterers are located along the link axis. Keeping that in mind, it is possible to understand the results of Figure 6.4, based on Figures 6.2 and 6.3.
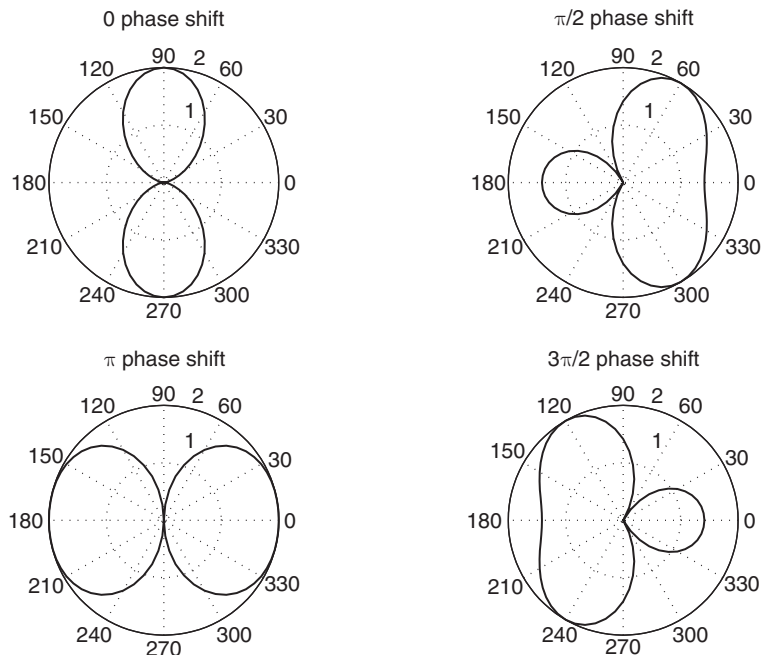
Figure 6.3 $G_t(\theta_t|\mathbf{c}_k)$ *(θ_t varying over $360°$) for the four possible phase shifts between two transmitted QPSK symbols and MT inter-element spacing $d_t/\lambda = 0.5$*
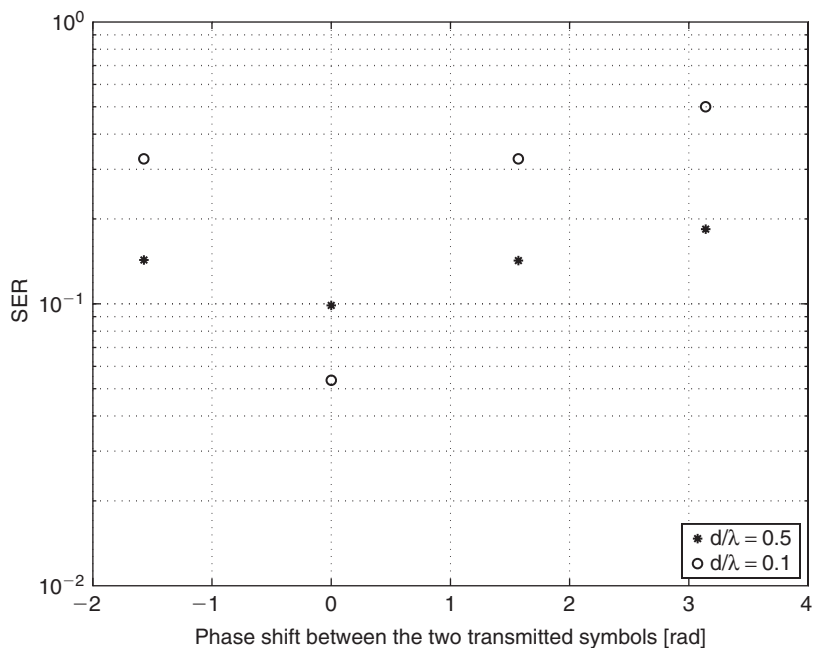
Figure 6.4 *Symbol error rate as a function of the phase shift (in radians) between the transmitted QPSK symbols*

It is worth noting that mutual coupling has not been taken into account when deriving the results of Figures 6.2, 6.3 and 6.4. However, mutual coupling should have been taken into account in the calculation of $G_t(\theta_t|\mathbf{c}_k)$. Indeed, for small inter-element spacing, it plays a key role in the effective radiated power. This is better understood by comparing, in Figure 6.2, the total radiated power (integrated over $[0, 2\pi)$) in the presence of a zero phase shift with the power radiated for any other phase shift. Clearly, the total radiated power obtained with zero phase shift is much larger than the power radiated with the other phase shifts, which is physically not correct. The radiated power should not depend on the phase shifts. This occurs because mutual coupling has been neglected. Taking into account the coupling, the power radiated in the presence of a zero phase shift would not be larger than the power radiated for any other phase shift.

When symbols have the same phase, $G_t(\theta_t|\mathbf{c}_k)$ at the transmitter (i.e. the MT) is oriented towards the receive direction (i.e. towards the BS). Hence, the received power is high and these symbols are better detected (Figure 6.4). By contrast, when the symbols have a phase difference of π, a null appears in $G_t(\theta_t|\mathbf{c}_k)$ in the direction of the BS (Figures 6.2 and 6.3). Hence, the highly reduced received power and performance (Figure 6.4). This variation of performance should be particularly significant for low LSR, corresponding to scenarios with obstacles mainly located along the link axis. The inter-element spacing also plays a role. For increasing inter-element spacings, the grating lobes cause $G_t(\theta_t|\mathbf{c}_k)$ to become almost omnidirectional, irrespective of the phase difference between symbol pairs.

To sum up, for directional scattering conditions and realistic (small) inter-element spacings at the transmitter, the number of lobes in $G_t(\theta_t|\mathbf{c}_k)$ is reduced, and nulls may appear in the direction of the strongest scatterers, reducing the signal-to-noise ratio at the receiver. Hence, in order to guarantee a large SNR in many fading scenarios, the magnitude of $G_t(\theta_t|\mathbf{c}_k)$ has to be maximized in the direction of the strongest scatterers. In a broadside array configuration, Figure 6.4 shows that the pairs of symbols with a phase difference of zero are better detected than the other pairs. Therefore, this suggests that one should guarantee that simultaneous transmitted symbols on each antenna have quasi-equal phases. This is the intuitive principle of beamforming and requires the transmitter to possess some information about the positions of the strongest scatterers.

6.2 Introduction to an average pairwise error probability approach

So far, we have analyzed the impact of the propagation environment on the conditional PEP through a detailed study of the interactions between a codeword and a given channel realization $\mathbf{H}$. Here, we take on a more classical statistical approach, by which we average the conditional PEP over given channel statistics. This was the approach we used in Chapter 5 for i.i.d. Rayleigh fading channels. By contrast, we now assume that the MIMO channels are possibly correlated and may present a dominant (Ricean) component. For a given set

of flat fading channel realizations $\{\mathbf{H}_k\}_{k=0}^{T-1}$, we recall that the PEP (5.3) is written as

$$P\left(\mathbf{C} \to \mathbf{E} \middle| \{\mathbf{H}_k\}_{k=0}^{T-1}\right) = \mathcal{Q}\left(\sqrt{\frac{\rho}{2} \sum_{k=0}^{T-1} \|\mathbf{H}_k(\mathbf{c}_k - \mathbf{e}_k)\|_F^2}\right) \qquad (6.12)$$

where $\mathcal{Q}(x)$ is the Gaussian $\mathcal{Q}$-function and $\rho = E_s/\sigma_n^2$ is the SNR. The mean PEP is then obtained through averaging (6.12) over the probability distribution of the channel gains. At this point, it is also useful to recall that the diversity gain g_d^o is the negative slope of the log-log plot of the average pairwise error probability $P(\mathbf{C} \to \mathbf{E})$ versus SNR at any given SNR.

Let us start our analysis with a very general case consisting of Ricean fading channels with joint spatial and temporal correlation. All fading channels analyzed further are simply particular cases of this general scenario. Defining

$$\mathbf{H} = \sqrt{\frac{K}{1+K}}\left(\mathbf{1}_{1\times T} \otimes \bar{\mathbf{H}}\right) + \sqrt{\frac{1}{1+K}}\left[\tilde{\mathbf{H}}_0 \ \tilde{\mathbf{H}}_1 \ \cdots \ \tilde{\mathbf{H}}_{T-1}\right],$$

$$\tilde{\mathcal{H}} = \text{vec}\left(\left[\tilde{\mathbf{H}}_0 \ \tilde{\mathbf{H}}_1 \ \cdots \ \tilde{\mathbf{H}}_{T-1}\right]^H\right)$$

$$\bar{\mathcal{H}} = \text{vec}\left(\left(\mathbf{1}_{1\times T} \otimes \bar{\mathbf{H}}\right)^H\right), \qquad (6.13)$$

$$\mathbf{D} = \text{diag}\{\mathbf{c}_0 - \mathbf{e}_0, \mathbf{c}_1 - \mathbf{e}_1, \ldots, \mathbf{c}_{T-1} - \mathbf{e}_{T-1}\},$$

$$\mathbf{\Delta} = \mathbf{I}_{n_r} \otimes \mathbf{D}\mathbf{D}^H,$$

and using results of Appendix E, the average PEP reads as

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\eta K \bar{\mathcal{H}}^H \mathbf{\Delta}(\mathbf{I}_{Tn_r n_t} + \eta \mathbf{\Xi}\mathbf{\Delta})^{-1}\bar{\mathcal{H}}\right)$$

$$\prod_{i=1}^{r(\mathbf{\Xi}\mathbf{\Delta})} (1 + \eta\lambda_i(\mathbf{\Xi}\mathbf{\Delta}))^{-1} d\beta \qquad (6.14)$$

where

- $r(\mathbf{\Xi}\mathbf{\Delta})$ stands for the rank of $\mathbf{\Xi}\mathbf{\Delta}$
- $\{\lambda_i(\mathbf{\Xi}\mathbf{\Delta})\}_{i=1}^{r(\mathbf{\Xi}\mathbf{\Delta})}$ are the non-zero eigenvalues of $\mathbf{\Xi}\mathbf{\Delta}$
- $\mathbf{\Xi} = \mathcal{E}\{\tilde{\mathcal{H}}\tilde{\mathcal{H}}^H\}$ denotes the spatio-temporal correlation matrix
- $\eta = \rho/(4\sin^2\beta(1+K))$ can be thought of as an effective SNR.

Note that (6.14) is really general and accounts for any spatial and temporal correlation profile. In particular, no assumption about a possible Kronecker structure of the correlation matrix has been made at this point. Still, to get more insight into how spatial correlation at the transmitter and/or the receiver affects the performance, we have seen that the Kronecker

representation is a good approximation of the actual channel. In that case, $\boldsymbol{\Xi}$ is then expressed as $\mathbf{R}_r \otimes \mathbf{I}_T \otimes \mathbf{R}_t$ for perfectly time interleaved fading channels and as $\mathbf{R}_r \otimes \mathbf{1}_{T \times T} \otimes \mathbf{R}_t$ for slow fading channels. As observed in Chapter 3, we recall that it does make sense to consider that $\mathbf{R}_r$ and $\mathbf{R}_t$ are full rank, i.e. $r(\mathbf{R}_r) = n_r$, $r(\mathbf{R}_t) = n_t$.

Let us now investigate a few particular cases of (6.14).

Ricean slow fading channels with spatial correlation

When the channel is constant over the duration of the frame and changes independently from frame to frame, we have seen in the previous chapters that the channel is said to be **slow fading**. The temporal correlation is one, so that the Rayleigh part of the channel matrix, $\tilde{\mathbf{H}}_k$, becomes constant for all k (it is denoted as $\tilde{\mathbf{H}}$ for simplicity). As highlighted in Appendix E, the average PEP is then rewritten as

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{r(\mathbf{C_R})} (1 + \eta \lambda_i(\mathbf{C_R}))^{-1}$$
$$\exp\left[-\eta K \mathrm{vec}\big(\bar{\mathbf{H}}^H\big)^H \big(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}\big)(\mathbf{I}_{n_r n_t} + \eta \mathbf{C_R})^{-1} \mathrm{vec}\big(\bar{\mathbf{H}}^H\big)\right] d\beta \qquad (6.15)$$

where

$$\mathbf{C_R} = \mathbf{R}\big(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}\big),$$
$$\tilde{\mathbf{E}} = (\mathbf{C} - \mathbf{E})(\mathbf{C} - \mathbf{E})^H,$$
$$\mathbf{R} = \mathcal{E}\left\{\mathrm{vec}\big(\tilde{\mathbf{H}}^H\big)\mathrm{vec}\big(\tilde{\mathbf{H}}^H\big)^H\right\}. \qquad (6.16)$$

We will also assume that $\mathbf{R}$ is in general full rank, i.e. $r(\mathbf{R}) = n_t n_r$. Yet, this is not always the case, e.g. for the diagonal channels of Section 4.3.1, although such situations are not very common. Note also that if the Rayleigh part is Kronecker-structured, one may expand $\mathbf{R}$ as $\mathbf{R}_r \otimes \mathbf{R}_t$.

Spatio-temporally correlated Ricean block fading channels

The block fading channel is another channel model commonly used in signal processing. In a classical block fading channel model

- the codeword length T extends over M channel blocks of length $N = T/M$
- the channel gains are assumed to be independent from block to block and constant within a block.

The channel matrix of the mth block is denoted as $\tilde{\mathbf{H}}_{(m)}$. Analogous to slow fading channels, we define the error matrix on the mth block as

$$\tilde{\mathbf{E}}_{(m)} = (\mathbf{C} - \mathbf{E})_{(m)}(\mathbf{C} - \mathbf{E})_{(m)}^H \qquad (6.17)$$

where

$$(\mathbf{C} - \mathbf{E})_{(m)} = \big[\mathbf{c}_{(m-1)N+1} - \mathbf{e}_{(m-1)N+1} \ \cdots \ \mathbf{c}_{mN} - \mathbf{e}_{mN}\big]. \qquad (6.18)$$

Referring to Appendix E, the average PEP is expressed as

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\eta K \bar{\mathcal{H}}_b^H \tilde{\mathbf{E}}_\mathbf{B} \left(\mathbf{I}_{n_r n_t M} + \eta \mathbf{C}_\mathbf{B}\right)^{-1} \bar{\mathcal{H}}_b\right)$$

$$\prod_{i=1}^{r(\mathbf{C_B})} \left(1 + \eta \lambda_i(\mathbf{C_B})\right)^{-1} d\beta \tag{6.19}$$

where

$$\mathbf{C_B} = \mathbf{B}\tilde{\mathbf{E}}_\mathbf{B},$$

$$\tilde{\mathbf{E}}_\mathbf{B} = \text{diag}\left\{\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}_{(1)}, \ldots, \mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}_{(M)}\right\},$$

$$\mathbf{B} = \mathcal{E}\left\{\tilde{\mathcal{H}}_b \tilde{\mathcal{H}}_b^H\right\}, \tag{6.20}$$

with the following quantities

$$\tilde{\mathcal{H}}_b^H = \left[\text{vec}\left(\tilde{\mathbf{H}}_{(1)}^H\right)^H \cdots \text{vec}\left(\tilde{\mathbf{H}}_{(M)}^H\right)^H\right] \tag{6.21}$$

$$\bar{\mathcal{H}}_b^H = \left[\mathbf{1}_{1 \times M} \otimes \text{vec}\left(\bar{\mathbf{H}}^H\right)^H\right]. \tag{6.22}$$

Since blocks are independent

$$\mathbf{B} = \text{diag}\left\{\mathbf{R}_{(1)}, \ldots, \mathbf{R}_{(M)}\right\} \tag{6.23}$$

with $\mathbf{R}_{(m)} = \mathcal{E}\left\{\text{vec}\left(\tilde{\mathbf{H}}_{(m)}^H\right) \text{vec}\left(\tilde{\mathbf{H}}_{(m)}^H\right)^H\right\}$.

Rayleigh fading channels
We have seen that Rayleigh fading channels are an important class of channels. When the K-factor K is equal to zero, the average PEP of (6.14) simplifies into

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \left(\det\left(\mathbf{I}_{Tn_r n_t} + \eta \mathbf{\Xi}\mathbf{\Delta}\right)\right)^{-1} d\beta$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{r(\mathbf{\Xi}\mathbf{\Delta})} \left(1 + \eta \lambda_i(\mathbf{\Xi}\mathbf{\Delta})\right)^{-1} d\beta \tag{6.24}$$

where the effective SNR η now simply reads as $\eta = \rho/(4\sin^2\beta)$.

In Rayleigh slow fading channels, the average PEP (6.15) becomes

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \left(\det\left(\mathbf{I}_{n_r n_t} + \eta \mathbf{C_R}\right)\right)^{-1} d\beta$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{r(\mathbf{C_R})} \left(1 + \eta \lambda_i(\mathbf{C_R})\right)^{-1} d\beta, \tag{6.25}$$

where $\mathbf{R} = \mathbf{R}_r \otimes \mathbf{R}_t$ if the Rayleigh channel is modeled by the Kronecker representation.

Finally, in Rayleigh block fading channels, the average PEP (6.19) is expressed as

$$P(\mathbf{C} \rightarrow \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{r(\mathbf{C_B})} (1 + \eta\lambda_i(\mathbf{C_B}))^{-1} d\beta$$

$$\stackrel{(a)}{=} \frac{1}{\pi} \int_0^{\pi/2} \prod_{m=1}^{M} \prod_{i=1}^{r(\mathbf{C}_{\mathbf{R}(m)})} \left(1 + \eta\lambda_i\left(\mathbf{C}_{\mathbf{R}(m)}\right)\right)^{-1} d\beta \qquad (6.26)$$

where equality (*a*) results from the independence assumption on the blocks and $\mathbf{C}_{\mathbf{R}(m)} = \mathbf{R}_{(m)}\left(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}_{(m)}\right)$.

We are now ready to discuss the impact of the spatial and/or temporal correlation and the K-factor on the average PEP. While doing this, we will pay specific attention to the careful use of the so-called high SNR regime. As explained in Chapter 5, this assumption has led to the rank-determinant criterion (Section 5.2) and the product-distance criterion (Section 5.3.1) for designing efficient codes respectively in i.i.d. Rayleigh slow and fast fading channels. However, Chapter 4 has highlighted that the K-factor and the spatial correlation significantly affect the achievable diversity-multiplexing trade-off at realistic SNR. From an error probability perspective, the situation is very similar, as shown in the rest of this chapter: at high SNR, the code design is not much affected by the value of the K-factor or by the spatial correlations, whereas this is not so at more realistic SNR. As a result, the general behavior of a code cannot be accurately predicted based on the high SNR behavior of that same code. This further suggests that the high SNR regime is not a realistic regime when designing space–time codes for correlated channels.

6.3 Average pairwise error probability in Rayleigh fading channels

6.3.1 High SNR regime

As in Chapter 5, a code is said to evolve in the high SNR regime when all eigenvalues of $\mathbf{C_R}$, $\mathbf{\Xi}\mathbf{\Delta}$ or $\mathbf{C_B}$ are much larger than η^{-1}. Physically, this means that each eigenvalue is large enough so that its effect is visible at the considered SNR. Let us successively consider slow, block and fast fading channels.

Rayleigh slow fading channels
Because the code evolves in the high SNR regime, $\lambda_i(\mathbf{C_R})\eta$ is much larger than one for each *i*. Hence, (6.25) may be rewritten as

$$P(\mathbf{C} \rightarrow \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-r(\mathbf{C_R})} \prod_{i=1}^{r(\mathbf{C_R})} \lambda_i^{-1}(\mathbf{C_R}) \, d\beta. \qquad (6.27)$$

Using the decomposition $\mathbf{I}_{n_r} \otimes (\mathbf{C} - \mathbf{E})^H = \mathbf{U}_{\mathbf{C}} \mathbf{\Lambda}_{\mathbf{C}} \mathbf{V}_{\mathbf{C}}^H$ and noting that $\mathbf{\Lambda}_{\mathbf{C}}$ is made of a non-zero principal submatrix $\mathbf{\Lambda}'_{\mathbf{C}}$ of size $n_r r(\tilde{\mathbf{E}})$, the final expression of (6.27) is summarized in the following proposition.

Proposition 6.1 *At asymptotically high SNR in Rayleigh slow fading channels with full-rank spatial correlation matrix $\mathbf{R}$, the average PEP is expressed as*

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(r(\mathbf{C_R}))} \prod_{i=1}^{r(\mathbf{C_R})} \lambda_i^{-1} \left(\mathbf{\Lambda}'_{\mathbf{C}} \mathbf{Q}' \mathbf{\Lambda}'_{\mathbf{C}} \right) d\beta \tag{6.28}$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r r(\tilde{\mathbf{E}})} \left(\det \left(\mathbf{Q}' \right) \right)^{-1} \prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i^{-n_r} (\tilde{\mathbf{E}}) d\beta \tag{6.29}$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_t n_r} (\det(\mathbf{R}))^{-1} \left(\det \left(\tilde{\mathbf{E}} \right) \right)^{-n_r} d\beta \tag{6.30}$$

$$\leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r r(\tilde{\mathbf{E}})} \prod_{i=1}^{n_r r(\tilde{\mathbf{E}})} \lambda_i^{-1}(\mathbf{R}) \prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i^{-n_r} (\tilde{\mathbf{E}}) d\beta \tag{6.31}$$

where $\mathbf{Q}'$ is a principal submatrix of $\mathbf{V}_{\mathbf{C}}^H \mathbf{R} \mathbf{V}_{\mathbf{C}}$ of size $n_r r \left(\tilde{\mathbf{E}} \right) \times n_r r \left(\tilde{\mathbf{E}} \right)$. Note that equality in (6.30) only holds true for full-rank codes, and that in (6.31), the eigenvalues of $\mathbf{R}$ are ordered by increasing value.

How should we understand Proposition 6.1? On the one hand, the diversity gain achieved by a space–time code is equal to $n_r r(\tilde{\mathbf{E}})$ and is not affected by the spatial correlation, i.e. it is the same as the diversity gain achieved in i.i.d. Rayleigh slow fading channels. For full-rank codes, this gain equals $n_t n_r$.

On the other hand, the coding gain is affected by the spatial correlations and can be decomposed into two multiplicative terms:

1. the first term is given by the product of the non-zero eigenvalues of $\tilde{\mathbf{E}}$, and equals the coding gain obtained in i.i.d. Rayleigh channels
2. the second term $\det(\mathbf{Q}')$ accounts for the interactions between the code and the channel through the projection of $\mathbf{V}_{\mathbf{C}}$ onto the space spanned by $\mathbf{R}$. For full-rank codes, $\det(\mathbf{Q}') = \det(\mathbf{R})$, resulting in (6.30).

For full-rank codes, spatial correlations decrease the coding gain by $\det(\mathbf{R})$, irrespective of the code structure. All full-rank codes are thus equally affected in the high SNR regime. As a result, the design criterion for a full-rank code in correlated channels is the rank-determinant criterion obtained in Chapter 5. By contrast, the impact of the spatial correlation on the coding gain of a rank-deficient code depends on the code via the projection of the error matrices onto the space spanned by the eigenvectors of $\mathbf{R}$. Therefore,

the performance in correlated channels depends on both the eigenvectors and the eigenvalues of the error matrices, although that was not the case in i.i.d. channels. Equivalently, two rank-deficient codes offering the same performance in i.i.d. channels may perform differently in correlated channels.

As a consequence, **designing codes using the rank-determinant criterion is not sufficient to guarantee a good performance in spatially correlated Rayleigh slow fading channels when the code is rank-deficient**.

Example 6.1   *In order to illustrate the above concepts, let us come back to the Spatial Multiplexing scheme: the codewords are vectors of size $n_t \times 1$, $\tilde{\mathbf{E}} = \mathbf{U_E}\mathbf{\Lambda_E}\mathbf{U_E^H}$ with $\mathbf{\Lambda_E} = \text{diag}\{\|\mathbf{C} - \mathbf{E}\|^2, 0, \ldots, 0\}$ and the average PEP can be easily calculated (assuming a Kronecker representation of the channel) as*

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r} (\det(\mathbf{R}_r))^{-1} \left| (\mathbf{C} - \mathbf{E})^H \mathbf{R}_t (\mathbf{C} - \mathbf{E}) \right|^{-n_r}$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r} (\det(\mathbf{R}_r))^{-1} \left| \mathbf{U_E^H}(1,:)\mathbf{R}_t\mathbf{U_E}(:,1) \right|^{-n_r} \|\mathbf{C} - \mathbf{E}\|^{-2n_r}. \quad (6.32)$$

There are clearly similarities with (6.29): the performance of SM in correlated channels depends on the projection of $\mathbf{U_E}(:,1)$ onto the space spanned by the eigenvectors of $\mathbf{R}_t$. When $\mathbf{U_E}(:,1)$ is parallel to the eigenvector of $\mathbf{R}_t$ corresponding to the lowest eigenvalue, the performance will be significantly affected. Remember that $\mathbf{R}_t$ can be thought of as a statistical description of the location of the strongest scatterers: the eigenvectors of $\mathbf{R}_t$ sample the 2D space and the eigenvalue attached to an eigenvector somehow indicates the density of scatterers in this direction. Hence, having $\mathbf{U_E}(:,1)$ lying in the space spanned by the eigenvector of $\mathbf{R}_t$ corresponding to the lowest eigenvalue is like transmitting all the information contained in the unique non-zero eigenvalue of the error matrix $\tilde{\mathbf{E}}$ in the direction of the space offering the lowest scatterer density. Clearly, for some matrices $\mathbf{R}_t$ (i.e. for some scattering distributions), the performance of a given error matrix may drastically change. This was already the conclusion of our conditional PEP approach. What was then described deterministically based on $\mathbf{a}_t(\theta_t)$ is now stochastically described using $\mathbf{R}_t$, yet, the physical interpretation remains exactly the same. Note that receive correlation only induces a coding gain loss independent of the error matrix.

In the high SNR regime, a full-rank code does not behave like a rank-deficient code such as SM. The reason is that a full-rank code spreads the information over all the eigenvalues of $\tilde{\mathbf{E}}$ (instead of one). Moreover, the high SNR regime implies that all directions of the space receive a sufficient amount of power so that should the scattering distribution become more directional, all directions of $\mathbf{R}_t$ receive the information contained in $\tilde{\mathbf{E}}$. Intuitively, it seems that at finite SNR, only the directions of $\mathbf{R}_t$ with large eigenvalues will be able to extract the information contained in $\tilde{\mathbf{E}}$. Would that suggest that a full-rank code behaves at finite SNR

as if it were rank-deficient? This will be further investigated after the detailed derivation of the results of Proposition 6.1, which formalizes the above discussion and example.

PROOF OF PROPOSITION 6.1:   Let us first consider that the code is ***full rank***, i.e. $r(\tilde{\mathbf{E}}) = n_t$. Since the spatial correlation matrices are full rank, the effects of the code can be separated from those caused by the spatial correlation

$$
\prod_{n=1}^{n_t n_r} \lambda_n(\mathbf{C_R}) = \prod_{n=1}^{n_t n_r} \lambda_n\left(\mathbf{R}\left(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}\right)\right)
$$

$$
= \det\left(\mathbf{R}\left(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}\right)\right)
$$

$$
= \left(\det\left(\tilde{\mathbf{E}}\right)\right)^{n_r} \det\left(\mathbf{R}\right). \tag{6.33}
$$

From (6.27), we simply get

$$
P(\mathbf{C} \rightarrow \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_t n_r} (\det(\mathbf{R}))^{-1} \left(\det(\tilde{\mathbf{E}})\right)^{-n_r} d\beta. \tag{6.34}
$$

Note that if we use the Kronecker model

$$
\det(\mathbf{R}) = (\det(\mathbf{R}_r))^{n_t} (\det(\mathbf{R}_t))^{n_r}. \tag{6.35}
$$

For full-rank codes and full-rank spatial correlation matrix $\mathbf{R}$ the diversity order in the high SNR regime is not modified and remains equal to $n_t n_r$. Yet, spatial correlations affect the coding gain, whose expression is made of two terms:

1. the first term $\left(\det(\tilde{\mathbf{E}})\right)^{-n_r}$ is the coding gain offered by the code itself and is exactly similar to what is obtained in i.i.d. Rayleigh slow fading channels
2. the second term $(\det(\mathbf{R}))^{n_t}$ is exclusively a function of the spatial correlation matrix. Since $\det(\mathbf{R}) \leq 1$, spatial correlation always decreases the total coding gain with respect to uncorrelated channels.

We stress that both parts of the coding gain are totally independent. The contributions of the code are separated from the negative effects of the channel. We say that there is no *interaction* between the channel and the code. As a result, the performance solely depends on the minimum distance code error matrix, i.e. $\arg\min_{\mathbf{C},\mathbf{E}} \det\left(\tilde{\mathbf{E}}\right)$, illustrating that the design of robust space–time full-rank codes in correlated channels is strictly identical to the design of space–time codes in i.i.d. Rayleigh slow fading channels (rank-determinant criterion 5.2) as originally derived in [BP00a].

If the code is ***rank-deficient***, the analysis becomes slightly more complicated. For full-rank space correlation matrix, the diversity gain is given by $r(\mathbf{C_R}) = n_r r(\tilde{\mathbf{E}})$ and remains equal to the gain achieved in i.i.d. Rayleigh slow fading channels. However, since (6.33) no longer holds true, the coding gain cannot be factorized into contributions inherent to the code and contributions caused by the spatial correlation. What does it imply on the

coding gain? Since the non-zero eigenvalues of $\mathbf{C_R}$ are equal to the non-zero eigenvalues of $\left(\mathbf{I}_{n_r} \otimes (\mathbf{C} - \mathbf{E})^H\right) \mathbf{R}(\mathbf{I}_{n_r} \otimes (\mathbf{C} - \mathbf{E}))$, the decompositions

$$\mathbf{I}_{n_r} \otimes (\mathbf{C} - \mathbf{E})^H = \mathbf{U_C}\mathbf{\Lambda_C}\mathbf{V}_\mathbf{C}^H,$$

$$\mathbf{R} = \mathbf{U_R}\mathbf{\Lambda_R}\mathbf{U}_\mathbf{R}^H,$$

$$\mathbf{U_Q} = \mathbf{V}_\mathbf{C}^H\mathbf{U_R}, \tag{6.36}$$

yield

$$\prod_{i=1}^{r(\mathbf{C_R})} \lambda_i(\mathbf{C_R}) = \prod_{i=1}^{r(\mathbf{C_R})} \lambda_i\left(\left(\mathbf{I}_{n_r} \otimes (\mathbf{C} - \mathbf{E})^H\right) \mathbf{R}(\mathbf{I}_{n_r} \otimes (\mathbf{C} - \mathbf{E}))\right)$$

$$= \prod_{i=1}^{r(\mathbf{C_R})} \lambda_i\left(\mathbf{U_C}\mathbf{\Lambda_C}\mathbf{U_Q}\mathbf{\Lambda_R}\mathbf{U}_\mathbf{Q}^H\mathbf{\Lambda}_\mathbf{C}^T\mathbf{U}_\mathbf{C}^H\right). \tag{6.37}$$

Because the code is rank-deficient, we may rewrite $\mathbf{\Lambda_C}$ as

$$\mathbf{\Lambda_C} = \begin{bmatrix} \mathbf{\Lambda}_\mathbf{C}' & \mathbf{0}_{n_r r(\tilde{\mathbf{E}}) \times n_r (n_t - r(\tilde{\mathbf{E}}))} \\ \mathbf{0}_{n_r(T - r(\tilde{\mathbf{E}})) \times n_r r(\tilde{\mathbf{E}})} & \mathbf{0}_{n_r(T - r(\tilde{\mathbf{E}})) \times n_r (n_t - r(\tilde{\mathbf{E}}))} \end{bmatrix} \tag{6.38}$$

such that

$$\mathbf{\Lambda_C}\mathbf{U_Q}\mathbf{\Lambda_R}\mathbf{U}_\mathbf{Q}^H\mathbf{\Lambda}_\mathbf{C}^T = \begin{bmatrix} \mathbf{\Lambda}_\mathbf{C}'\mathbf{Q}'\mathbf{\Lambda}_\mathbf{C}' & \mathbf{0}_{n_r r(\tilde{\mathbf{E}}) \times n_r (n_t - r(\tilde{\mathbf{E}}))} \\ \mathbf{0}_{n_r(T - r(\tilde{\mathbf{E}})) \times n_r r(\tilde{\mathbf{E}})} & \mathbf{0}_{n_r(T - r(\tilde{\mathbf{E}})) \times n_r (n_t - r(\tilde{\mathbf{E}}))} \end{bmatrix} \tag{6.39}$$

where $\mathbf{Q}'$ is a principal submatrix of $\mathbf{U_Q}\mathbf{\Lambda_R}\mathbf{U}_\mathbf{Q}^H$ of size $n_r r(\tilde{\mathbf{E}}) \times n_r r(\tilde{\mathbf{E}})$. Because the eigenvalues λ_i are not affected through left and right multiplication by unitary matrices $\mathbf{U_C}$, we have

$$\prod_{i=1}^{r(\mathbf{C_R})} \lambda_i(\mathbf{C_R}) = \prod_{i=1}^{n_r r(\tilde{\mathbf{E}})} \lambda_i\left(\mathbf{\Lambda}_\mathbf{C}'\mathbf{Q}'\mathbf{\Lambda}_\mathbf{C}'\right)$$

$$= \det\left(\mathbf{\Lambda}_\mathbf{C}'\mathbf{Q}'\mathbf{\Lambda}_\mathbf{C}'\right)$$

$$= \det\left(\mathbf{Q}'\right) \det\left(\mathbf{\Lambda}_\mathbf{C}'^2\right)$$

$$= \det\left(\mathbf{Q}'\right) \prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i^{n_r}\left(\tilde{\mathbf{E}}\right). \tag{6.40}$$

Hence, (6.27) eventually becomes

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r r(\tilde{\mathbf{E}})} (\det(\mathbf{Q}'))^{-1} \prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i^{-n_r}(\tilde{\mathbf{E}}) d\beta. \tag{6.41}$$

If we use the Kronecker representation, it is straightforward to write $\mathbf{Q}' = \mathbf{R}_r \otimes \mathbf{K}'$. The matrix $\mathbf{K}'$ is a principal submatrix of $\mathbf{K} = \mathbf{U}_\mathbf{E}^H \mathbf{R}_t \mathbf{U}_\mathbf{E}$ of size $r(\tilde{\mathbf{E}}) \times r(\tilde{\mathbf{E}})$ where $\mathbf{U}_\mathbf{E}$ is the left matrix of singular vectors of $\mathbf{C} - \mathbf{E}$. Therefore, the coding gain becomes

$$\det\left(\mathbf{Q}'\right) = (\det(\mathbf{R}_r))^{r(\tilde{\mathbf{E}})} \left(\det\left(\mathbf{K}'\right)\right)^{n_r}. \tag{6.42}$$

An upper bound on $P(\mathbf{C} \to \mathbf{E})$ is easily obtained by applying the inclusion principle (Appendix A) to matrix $\mathbf{Q}'$. Indeed, $\mathbf{Q}'$ is a principal submatrix of the Hermitian matrix $\mathbf{U}_\mathbf{Q} \boldsymbol{\Lambda}_\mathbf{R} \mathbf{U}_\mathbf{Q}^H$. Denoting the eigenvalues of $\mathbf{R}$ and $\mathbf{Q}'$ in increasing order respectively by $\lambda_i(\mathbf{R})$ and $\lambda_i(\mathbf{Q}')$, we are able to write the following inequalities

$$\lambda_i(\mathbf{R}) \leq \lambda_i\left(\mathbf{Q}'\right) \leq \lambda_{i+n_r n_t - n_r r(\tilde{\mathbf{E}})}(\mathbf{R}) \tag{6.43}$$

for $1 \leq i \leq n_r r(\tilde{\mathbf{E}})$. We finally obtain a lower bound for $\det\left(\mathbf{Q}'\right)$ as

$$\det\left(\mathbf{Q}'\right) = \prod_{i=1}^{n_r r(\tilde{\mathbf{E}})} \lambda_i(\mathbf{Q}') \geq \prod_{i=1}^{n_r r(\tilde{\mathbf{E}})} \lambda_i(\mathbf{R}) \tag{6.44}$$

and an upper bound for $P(\mathbf{C} \to \mathbf{E})$ as

$$P(\mathbf{C} \to \mathbf{E}) \leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r r(\tilde{\mathbf{E}})} \prod_{i=1}^{n_r r(\tilde{\mathbf{E}})} \lambda_i^{-1}(\mathbf{R}) \prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i^{-n_r}\left(\tilde{\mathbf{E}}\right) d\beta. \tag{6.45}$$

The diversity order achieved by the code is thus well equal to $n_r r(\tilde{\mathbf{E}})$ and is affected by spatial correlations under the hypothesis that the spatial correlation matrix is full rank. Analogous to full-rank codes, spatial correlations affect the coding gain, which is again made of two terms:

1. the term $\prod_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i^{n_r}\left(\tilde{\mathbf{E}}\right)$ is similar to the coding gain in i.i.d. Rayleigh slow fading channels for a codeword error matrix of rank $r(\tilde{\mathbf{E}})$
2. the term $\det(\mathbf{Q}')$ takes into account the interactions between the spatial correlation matrix and the code. Because $\mathbf{Q}'$ is a principal submatrix of $\mathbf{V}_\mathbf{C}^H \mathbf{R} \mathbf{V}_\mathbf{C}$, the performance highly depends on the projection of $\mathbf{V}_\mathbf{C}$ onto the spatial correlation matrix $\mathbf{R}$.

When the lower bound (6.44) is achieved, the performance may be severely degraded. Interestingly, in Kronecker-structured channels, the transmit correlation interacts with the code, whereas the receive correlation only decreases the coding gain analogous to the case of full-rank codes. Note also that (6.40) may be seen as the generalization of (6.30) to arbitrary space–time codes, where the special case of full-rank codes is covered by letting $\mathbf{Q}' = \mathbf{V}_\mathbf{C}^H \mathbf{R} \mathbf{V}_\mathbf{C}$ and $\det\left(\mathbf{Q}'\right) = \det(\mathbf{R})$. ∎

How realistic is the high SNR regime approximation?
In other words, in practical scenarios (i.e. at high but still realistic SNR), how representative of classical codes is the predicted behavior based on a high-SNR assumption? To address
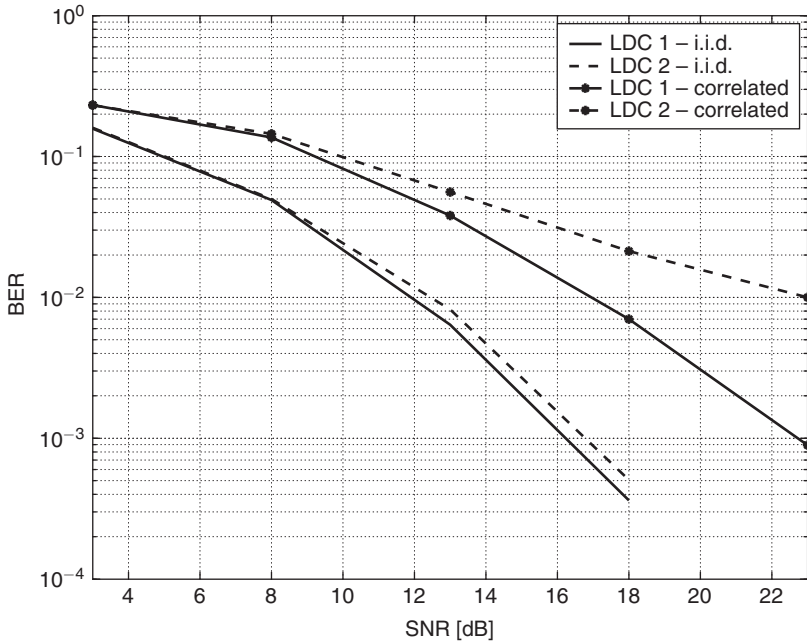
Figure 6.5   *Performance of full-rank LDCs in i.i.d. and correlated channels with $n_t = 2$ and $n_r = 2$*

this issue, let us analyze the performance of two linear dispersion codes for 2 transmit and 2 receive antennas in both i.i.d. and correlated slow fading Rayleigh channels. The correlated slow fading channel is generated following the geometry-based stochastic model of Section 2.3.3, with inline (with respect to the link axis) array configuration at the mobile terminal (antenna spacing of 0.5 λ) and broadside array configuration at the base station (antenna spacing of 20 λ). A Rayleigh channel with highly directional propagation along the link axis is considered, resulting in high spatial correlations at the transmitter.

Figure 6.5 displays the performance of two full-rank LDCs (denoted as "LDC 1" and "LDC 2") expanding on 2 symbol durations and transmitting 4 symbols per codeword, using QPSK constellations and ML decoding. Clearly, "LDC 1" and "LDC 2" have pretty much the same performance in i.i.d. channels. In correlated channels, we expect that the diversity be equal to 4 (as in i.i.d. channels) and the coding gain be decreased by det ($\mathbf{R}$), irrespective of the code: both codes should thus perform equally well in correlated channels, at least at high SNR. Yet, this is not what we observe, as the bit error rates in correlated channels are vastly different. How can we explain this ? The high regime is actually not achieved for "LDC 2", owing to destructive interactions between the channel and the code. This results in "LDC 2" being much more affected by spatial correlations than "LDC 1". While "LDC 1" approaches the gains foreseen by (6.30) and may be considered as evolving in the high SNR regime, "LDC 2" has a much limited diversity gain, even at SNR levels as high as 20 dB. This illustrates the need to investigate the performance of space–time

codes at realistic SNR when dealing with correlated channels. We will come back to this issue when we investigate the medium SNR regime later in this chapter.

Rayleigh block fading channels

In block fading channels, (6.26) is rewritten in the high SNR regime as

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-r(\mathbf{C_B})} \prod_{i=1}^{r(\mathbf{C_B})} \lambda_i^{-1}(\mathbf{C_B}) \, d\beta. \tag{6.46}$$

Denoting the blocks over which $\mathbf{E}_{(m)}$ is full rank as $\tau_{\mathcal{FR}} = \{m | r(\tilde{\mathbf{E}}_{(m)}) = n_t\}$ and the blocks over which $\mathbf{E}_{(m)}$ is rank-deficient as $\tau_{\mathcal{RD}} = \{m | r(\tilde{\mathbf{E}}_{(m)}) < n_t\}$, and assuming that the blocks fade independently, we simply obtain

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \left[\eta^{-n_t n_r \sharp \tau_{\mathcal{FR}}} \prod_{m \in \tau_{\mathcal{RD}}} \prod_{i=1}^{n_t n_r} \lambda_i^{-1}(\mathbf{C_{R}}_{(m)}) \right]$$
$$\left[\prod_{n \in \tau_{\mathcal{RD}}} \eta^{-r(\mathbf{C_{R}}_{(n)})} \prod_{i=1}^{r(\mathbf{C_{R}}_{(n)})} \lambda_i^{-1}(\mathbf{C_{R}}_{(n)}) \right] d\beta \tag{6.47}$$

where $\sharp \tau_{\mathcal{FR}}$ denotes the cardinality of the set $\tau_{\mathcal{FR}}$. Both terms in (6.47) can be further developed based on (6.29). Observations are thus similar to slow fading channels.

Rayleigh fast fading channels

In Rayleigh fast fading (or time-varying) channels, the high SNR regime implies that for each i, $\lambda_i(\mathbf{\Xi \Delta})\eta$ is much larger than one. In what follows, we assume that the channel varies sufficiently rapidly or that the interleaving depth is long, so that $r(\mathbf{\Xi}) = n_t n_r T$. From (6.24), the high-SNR PEP is given by

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-r(\mathbf{\Xi \Delta})} \prod_{i=1}^{r(\mathbf{\Xi \Delta})} \lambda_i^{-1}(\mathbf{\Xi \Delta}) d\beta \tag{6.48}$$

where $\mathbf{\Xi}$ denotes the spatio-temporal correlation matrix. Defining the set $\tau_{\mathbf{C,E}} = \{k | \mathbf{c}_k - \mathbf{e}_k \neq \mathbf{0}_{n_t \times 1}\}$, and its length $l_{\mathbf{C,E}} = \sharp \tau_{\mathbf{C,E}}$, we first construct the matrices $\mathbf{D}'$ by suppressing the columns of $\mathbf{D}$ that correspond to the time indexes k not contained in $\tau_{\mathbf{C,E}}$. Similarly, we construct the matrix $\mathbf{\Xi}'$ by suppressing the rows and columns that correspond to the zero-columns of $\mathbf{I}_{n_r} \otimes \mathbf{D}$. We also decompose $\mathbf{I}_{n_r} \otimes \mathbf{D}'^H$ as $\mathbf{U_{D'}} \mathbf{\Lambda_{D'}} \mathbf{V}_{\mathbf{D'}}^H$ where $\mathbf{\Lambda_{D'}}$ is made of a non-zero diagonal matrix $\mathbf{\Lambda}'_{\mathbf{D'}}$ of size $n_r l_{\mathbf{C,E}} \times n_r l_{\mathbf{C,E}}$. Equation (6.48) is then re-expressed as outlined in the following proposition.

Proposition 6.2 *At asymptotically high SNR in Rayleigh fast fading channels with full-rank spatio-temporal correlation matrix* $\mathbf{\Xi}$*, the average PEP reads as*

$$P(\mathbf{C} \rightarrow \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-r(\mathbf{\Xi}\mathbf{\Delta})} \prod_{i=1}^{r(\mathbf{\Xi}\mathbf{\Delta})} \lambda_i^{-1} \left(\mathbf{\Lambda}_{\mathbf{D}'}' \mathbf{O}' \mathbf{\Lambda}_{\mathbf{D}'}' \right) d\beta \tag{6.49}$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l_{\mathbf{C},\mathbf{E}}} \left(\det \left(\mathbf{O}' \right) \right)^{-1} \prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \| \mathbf{c}_k - \mathbf{e}_k \|^{-2n_r} d\beta \tag{6.50}$$

$$\leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l_{\mathbf{C},\mathbf{E}}} \prod_{i=1}^{n_r l_{\mathbf{C},\mathbf{E}}} \lambda_i^{-1}(\mathbf{\Xi}) \prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \| \mathbf{c}_k - \mathbf{e}_k \|^{-2n_r} d\beta. \tag{6.51}$$

where $\mathbf{O}'$ *is a principal submatrix of* $\mathbf{V}_{\mathbf{D}'}^H \mathbf{\Xi}' \mathbf{V}_{\mathbf{D}'}$ *of size* $n_r l_{\mathbf{C},\mathbf{E}} \times n_r l_{\mathbf{C},\mathbf{E}}$*. Note that in* (6.51)*, the eigenvalues of* $\mathbf{\Xi}$ *are ordered by increasing value.*

The above proposition has obvious similarities with the properties of rank-deficient codes in slow fading channels. Again, the diversity gain given by $n_r l_{\mathbf{C},\mathbf{E}}$ is not affected by spatio-temporal correlations. The coding gain is made of two terms:

1. the first term is given by the product distance of $\tilde{\mathbf{E}}$, and is equal to the coding gain achieved in i.i.d. Rayleigh fast fading channels
2. the second term $\det(\mathbf{O}')$ accounts for the interactions between the code and the channel through the projection of $\mathbf{V}_{\mathbf{D}'}$ onto the space spanned by $\mathbf{\Xi}$.

For all codes (both full-rank and rank-deficient codes), the coding gain is thus decreased by the interactions between the error matrices and the correlation matrix via the projection of each column of the error matrix onto the space spanned by the eigenvectors of $\mathbf{\Xi}$. Hence, two codes offering the same performance in i.i.d. Rayleigh channels may perform differently in spatio-temporally correlated Rayleigh channels.

Designing codes based on the distance-product criterion is thus not sufficient to guarantee a good performance in spatially correlated Rayleigh fast fading channels, irrespective of the rank of the code.

Example 6.2 *To illustrate the impact of transmit correlation on the error rate, we consider an arbitrary code (full-rank or rank-deficient) in a perfectly interleaved channel for which the Kronecker model may be applied. Hence,* $\mathbf{\Xi} = \mathbf{R}_r \otimes \mathbf{I}_T \otimes \mathbf{R}_t$*, and we easily*

obtain that

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l_{\mathbf{C},\mathbf{E}}} \left(\det\left(\mathbf{R}_r\right)\right)^{-l_{\mathbf{C},\mathbf{E}}}$$
$$\prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \left|(\mathbf{c}_k - \mathbf{e}_k)^H \mathbf{R}_t (\mathbf{c}_k - \mathbf{e}_k)\right|^{-n_r} d\beta \qquad (6.52)$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l_{\mathbf{C},\mathbf{E}}} (\det\left(\mathbf{R}_r\right))^{-l_{\mathbf{C},\mathbf{E}}} \prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{-2n_r}$$
$$\prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \left|\mathbf{U}_{\tilde{\mathbf{e}}_k}^H (1,:) \mathbf{R}_t \mathbf{U}_{\tilde{\mathbf{e}}_k}(:,1)\right|^{-n_r} d\beta \qquad (6.53)$$

where

$$\tilde{\mathbf{e}}_k = (\mathbf{c}_k - \mathbf{e}_k)(\mathbf{c}_k - \mathbf{e}_k)^H = \mathbf{U}_{\tilde{\mathbf{e}}_k} \mathbf{\Lambda}_{\tilde{\mathbf{e}}_k} \mathbf{U}_{\tilde{\mathbf{e}}_k}^H \qquad (6.54)$$

with $\mathbf{\Lambda}_{\tilde{\mathbf{e}}_k} = \operatorname{diag}\left\{\|\mathbf{c}_k - \mathbf{e}_k\|^2, 0, \ldots, 0\right\}$. This is very similar to (6.50). In ideally interleaved channels, the performance of a code is thus determined by projection of each column of the error matrix onto the space spanned by the eigenvectors of $\mathbf{R}_t$. From the point of view of the channel-code interactions, any code in ideally interleaved channel behaves similarly to Spatial Multiplexing (see Example 6.1).

PROOF OF PROPOSITION 6.2: Since $r(\mathbf{\Xi}) = n_t n_r T$, the total diversity achieved at high SNR by the pair of codewords $\{\mathbf{C}, \mathbf{E}\}$ is given by $r\left(\mathbf{\Xi}\left(\mathbf{I}_{n_r} \otimes \mathbf{D}\mathbf{D}^H\right)\right) = r\left(\mathbf{D}\mathbf{D}^H\right) n_r$ and the coding gain writes as

$$\prod_{i=1}^{r(\mathbf{\Xi}\mathbf{\Delta})} \lambda_i(\mathbf{\Xi}\mathbf{\Delta}) = \prod_{i=1}^{n_r r(\mathbf{D}\mathbf{D}^H)} \lambda_i\left(\left(\mathbf{I}_{n_r} \otimes \mathbf{D}^H\right) \mathbf{\Xi} \left(\mathbf{I}_{n_r} \otimes \mathbf{D}\right)\right). \qquad (6.55)$$

This coding gain is equivalently expressed as follows. Based on the following decompositions

$$\mathbf{I}_{n_r} \otimes \mathbf{D}'^H = \mathbf{U}_{\mathbf{D}'} \mathbf{\Lambda}_{\mathbf{D}'} \mathbf{V}_{\mathbf{D}'}^H,$$
$$\mathbf{\Xi}' = \mathbf{U}_{\mathbf{\Xi}'} \mathbf{\Lambda}_{\mathbf{\Xi}'} \mathbf{U}_{\mathbf{\Xi}'}^H,$$
$$\mathbf{U}_{\mathbf{O}} = \mathbf{V}_{\mathbf{D}'}^H \mathbf{U}_{\mathbf{\Xi}'}, \qquad (6.56)$$

we note that the non-zero eigenvalues of $\left(\mathbf{I}_{n_r} \otimes \mathbf{D}'^H\right) \mathbf{\Xi}' \left(\mathbf{I}_{n_r} \otimes \mathbf{D}'\right)$ and $\left(\mathbf{I}_{n_r} \otimes \mathbf{D}^H\right) \mathbf{\Xi} \left(\mathbf{I}_{n_r} \otimes \mathbf{D}\right)$ remain unchanged. Since it is always possible to express $\mathbf{\Lambda}_{\mathbf{D}'}$ as

$$\mathbf{\Lambda}_{\mathbf{D}'} = \left[\mathbf{\Lambda}'_{\mathbf{D}'} \; \mathbf{0}_{n_r l_{\mathbf{C},\mathbf{E}} \times n_r l_{\mathbf{C},\mathbf{E}}(n_t - 1)}\right], \qquad (6.57)$$

we have

$$\mathbf{\Lambda}_{\mathbf{D}'} \mathbf{U}_{\mathbf{O}} \mathbf{\Lambda}_{\mathbf{\Xi}'} \mathbf{U}_{\mathbf{O}}^H \mathbf{\Lambda}_{\mathbf{D}'}^T = \mathbf{\Lambda}'_{\mathbf{D}'} \mathbf{O}' \mathbf{\Lambda}'_{\mathbf{D}'} \qquad (6.58)$$

where $\mathbf{O}'$ is a principal submatrix of $\mathbf{U_O} \boldsymbol{\Lambda}_{\boldsymbol{\Xi}'} \mathbf{U_O}^H$ of size $n_r l_{\mathbf{C,E}} \times n_r l_{\mathbf{C,E}}$. The coding gain can thus be rewritten as

$$\prod_{i=1}^{r(\boldsymbol{\Xi}\boldsymbol{\Delta})} \lambda_i(\boldsymbol{\Xi}\boldsymbol{\Delta}) = \det(\mathbf{O}') \prod_{k\in\tau_{\mathbf{C,E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{2n_r}. \tag{6.59}$$

Equation (6.48) finally becomes

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l_{\mathbf{C,E}}} \left(\det(\mathbf{O}')\right)^{-1} \prod_{k\in\tau_{\mathbf{C,E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{-2n_r} d\beta. \tag{6.60}$$

Since $\mathbf{O}'$ is a principal submatrix of $\mathbf{U_O} \boldsymbol{\Lambda}_{\boldsymbol{\Xi}'} \mathbf{U_O}^H$ and $\boldsymbol{\Xi}'$ is a principal submatrix of $\boldsymbol{\Xi}$, we get, from the inclusion theorem (see Appendix A), the following lower bound

$$\det\left(\mathbf{O}'\right) = \prod_{i=1}^{n_r l_{\mathbf{C,E}}} \lambda_i(\mathbf{O}') \geq \prod_{i=1}^{n_r l_{\mathbf{C,E}}} \lambda_i\left(\boldsymbol{\Xi}'\right) \geq \prod_{i=1}^{n_r l_{\mathbf{C,E}}} \lambda_i(\boldsymbol{\Xi}) \tag{6.61}$$

where the eigenvalues of $\boldsymbol{\Xi}'$ and $\boldsymbol{\Xi}$ are in increasing order. The average PEP can thus be upper bounded by

$$P(\mathbf{C} \to \mathbf{E}) \leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l_{\mathbf{C,E}}} \prod_{i=1}^{n_r l_{\mathbf{C,E}}} \lambda_i^{-1}(\boldsymbol{\Xi}) \prod_{k\in\tau_{\mathbf{C,E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{-2n_r} d\beta. \tag{6.62}$$

The diversity gain is given by $n_r l_{\mathbf{C,E}}$ and is equal to the diversity gain in i.i.d. Rayleigh fast fading channels, i.e. it is function of the effective length of the error matrix only. Hence, a full-rank spatio-temporal correlation matrix does not affect the achievable diversity gain in the high SNR regime. However, the coding gain is affected by spatio-temporal correlations. Analogous to slow fading channels, the coding gain is expressed as the product of two terms:

1. the term $\prod_{k\in\tau_{\mathbf{C,E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{2n_r}$ is the product distance of the codeword error matrix
2. the term $\det(\mathbf{O}')$ accounts for the interaction between the code and the spatio-temporal correlation matrix and highly impacts the total coding gain. For full-rank codes, the situation is more critical in fast fading channels than in slow fading channels because the error rate is now affected by the projection of each error vector $\mathbf{c}_k - \mathbf{e}_k$ onto the space spanned by the transmit correlation matrix. Somehow, the robustness is similar to that of rank-deficient codes in slow fading channels. This part of the coding gain is always lower-bounded by the product of the $n_r l_{\mathbf{C,E}}$ smallest eigenvalues of $\boldsymbol{\Xi}$, but the bound is quite loose and not representative of the achieved performance. Therefore, a code design for correlated channels should take care that $\det(\mathbf{O}')$ is maximized. ∎

6.3.2 Medium SNR regime

The previous examples have illustrated that in correlated Rayleigh channels, space–time codes might not evolve in the high SNR regime for realistic SNR levels. To derive the error performance at more practical SNR, we define a so-called medium SNR regime, characterized by the fact that only a fraction of the non-zero eigenvalues of matrices $\mathbf{C_R}$, $\mathbf{\Xi\Delta}$ or $\mathbf{C_B}$ are able to contribute to the diversity and coding gains, whereas the remnant eigenvalues are much smaller than η^{-1}.

Rayleigh slow fading channels

Let us consider that α eigenvalues of $\mathbf{C_R}$ are such that $\lambda_i(\mathbf{C_R})\eta << 1$, $i = 1,\ldots,\alpha$ while $r(\mathbf{C_R}) - \alpha$ eigenvalues are in the high SNR regime. Based on those definitions, (6.25) may be rewritten as

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(r(\mathbf{C_R})-\alpha)} \prod_{i=1+\alpha}^{r(\mathbf{C_R})} \lambda_i^{-1}(\mathbf{C_R}) \prod_{j=1}^{\alpha} (1 + \lambda_j(\mathbf{C_R})\eta)^{-1} \, d\beta$$

$$\approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(r(\mathbf{C_R})-\alpha)} \prod_{i=1+\alpha}^{r(\mathbf{C_R})} \lambda_i^{-1}(\mathbf{C_R}) \, d\beta \tag{6.63}$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(r(\mathbf{C_R})-\alpha)} \prod_{i=1+\alpha}^{r(\mathbf{C_R})} \lambda_i^{-1} \left(\mathbf{\Lambda_C'} \mathbf{Q'} \mathbf{\Lambda_C'} \right) d\beta \tag{6.64}$$

where we assume that the eigenvalues λ_i are sorted by increasing order of magnitude, and (6.64) results from (6.28).

Let us now investigate why α non zero eigenvalues of $\mathbf{C_R}$ and $\mathbf{\Lambda_C'} \mathbf{Q'} \mathbf{\Lambda_C'}$ are possibly not able to contribute to diversity. To this end, we use the following decompositions

$$\mathbf{Q'} = \mathbf{Q'}^{1/2} \mathbf{Q'}^{H/2} = \mathbf{U_{Q'}} \mathbf{\Lambda_{Q'}^2} \mathbf{U_{Q'}^H} \tag{6.65}$$

$$\mathbf{\Lambda_C'} \mathbf{Q'} \mathbf{\Lambda_C'} = \mathbf{\Lambda_C'} \mathbf{U_{Q'}} \mathbf{\Lambda_{Q'}^2} \mathbf{U_{Q'}^H} \mathbf{\Lambda_C'}. \tag{6.66}$$

Clearly, the distribution of the eigenvalues of $\mathbf{\Lambda_C'} \mathbf{Q'} \mathbf{\Lambda_C'}$, as well as its condition number, depend on $\mathbf{U_{Q'}}$.

- If the eigenvalues of both $\mathbf{\Lambda_C'}$ and $\mathbf{\Lambda_{Q'}^2}$ are large, the code error matrices are characterized by small condition numbers and the propagation is described by a Rayleigh slow fading channel with low spatial correlation. Hence, the code evolves in the high SNR regime.
- If the eigenvalues of at least one of those matrices vanish, there is a transition from the high to medium SNR regime. As an example, if $\mathbf{U_{Q'}} = \mathbf{I}_{n_t}$, the minimum eigenvalue of $\mathbf{\Lambda_C'} \mathbf{Q'} \mathbf{\Lambda_C'}$, equal to the product of the minimum eigenvalues of $\mathbf{\Lambda_C'}$ and $\mathbf{\Lambda_{Q'}^2}$, would probably become much lower than η^{-1} and thereby hardly contribute to diversity at finite SNR.

A physical interpretation of this situation is easily found by assuming a Kronecker-structured channel, so that $\mathbf{C_R} = \mathbf{R}_r \otimes \mathbf{R}_t \tilde{\mathbf{E}}$. Hence, the decrease in diversity resulting from a transition from the high to the medium SNR regime may be caused

- by a high receive correlation, inducing low eigenvalues of $\mathbf{R}_r$
- by a high transmit correlation, inducing an interaction between the code (even if it is full rank!) and the transmit correlation matrix via low eigenvalues of $\mathbf{R}_t \tilde{\mathbf{E}}$.

Consequently, the performance in the medium SNR regime (i.e. at practical SNR levels) critically depends upon the distribution of eigenvalues of $\mathbf{Q}'$ and $\mathbf{\Lambda}'_{\mathbf{C}}$ and upon the shape of matrix $\mathbf{U_{Q'}}$, i.e. upon the projection of the error matrix onto the space of the spatial correlation matrix. The achievable error performance is summarized in the following proposition for two extreme cases:

1. in the first case, α non-zero eigenvalues of $\mathbf{Q}'$ are too small to contribute to diversity, while the $n_r r(\tilde{\mathbf{E}}) - \alpha$ remnant eigenvalues of $\mathbf{\Lambda_{Q'}}$ are contained in a principal diagonal submatrix $\mathbf{\Lambda}'_{\mathbf{Q'}}$
2. in the second case, α non-zero eigenvalues of the codeword error matrix $\Lambda'_{\mathbf{C}}$ are assumed to be small.

We stress that in real-world scenarios, both cases may occur simultaneously.

Proposition 6.3   *At finite SNR in Rayleigh slow fading channels with full-rank spatio-temporal correlation matrix $\mathbf{R}$, consider that α eigenvalues of $\mathbf{\Lambda}'_{\mathbf{C}} \mathbf{Q}' \mathbf{\Lambda}'_{\mathbf{C}}$ are much smaller than η^{-1}.*

1. *If high spatial correlations cause α non-zero eigenvalues of $\mathbf{Q}'$ to be negligible, the average PEP is expressed as*

$$P(\mathbf{C} \rightarrow \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r r(\tilde{\mathbf{E}}) - \alpha)} \left(\det\left(\mathbf{L}''\right)\right)^{-1} \left(\det\left(\mathbf{\Lambda}_{\mathbf{Q}}^{'2}\right)\right)^{-1} d\beta \tag{6.67}$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r r(\tilde{\mathbf{E}}) - \alpha)} \left(\det\left(\mathbf{L}''\right)\right)^{-1} \prod_{i=1+\alpha}^{n_t n_r} \lambda_i^{-1}(\mathbf{R}) \, d\beta \tag{6.68}$$

$$\leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r r(\tilde{\mathbf{E}}) - \alpha)} \prod_{i=1}^{r(\tilde{\mathbf{E}}) - \delta} \lambda_i^{-n_r}(\tilde{\mathbf{E}}) \prod_{i=\alpha+1}^{n_r r(\tilde{\mathbf{E}})} \lambda_i^{-1}(\mathbf{R}) \, d\beta \tag{6.69}$$

where L'' is a principal submatrix of $\mathbf{U}_{\mathbf{Q}'}^H \mathbf{\Lambda}_{\mathbf{C}}^{'2} \mathbf{U}_{Q'} \overset{(a)}{=} \mathbf{U}_{\mathbf{R}}^H \left(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}\right) \mathbf{U}_{\mathbf{R}}$ of size $n_r r(\tilde{\mathbf{E}}) - \alpha \times n_r r(\tilde{\mathbf{E}}) - \alpha$, and (6.68), (6.69) and equality (a) only hold true for full-rank codes.

2. *If $\alpha = \delta n_r$ non-zero eigenvalues of the codeword error matrix $\mathbf{\Lambda}'_{\mathbf{C}}$ are negligible, the average PEP takes the following form*

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r r(\tilde{\mathbf{E}}) - \alpha)} \left(\det\left(\mathbf{Q}''\right)\right)^{-1} \prod_{i=1+\delta}^{r(\tilde{\mathbf{E}})} \lambda_i^{-1}(\tilde{\mathbf{E}}) d\beta \tag{6.70}$$

$$\leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r r(\tilde{\mathbf{E}}) - \alpha)} \prod_{i=1}^{n_r r(\tilde{\mathbf{E}}) - \alpha} \lambda_i^{-1}(\mathbf{R}) \prod_{i=1+\delta}^{r(\tilde{\mathbf{E}})} \lambda_i^{-n_r}(\tilde{\mathbf{E}}) d\beta \tag{6.71}$$

where $\mathbf{Q}''$ is a principal submatrix of $\mathbf{Q}'$ of size $n_r r(\tilde{\mathbf{E}}) - \alpha \times n_r r(\tilde{\mathbf{E}}) - \alpha$.

All eigenvalues are ordered by increasing value.

Let us now analyze and understand the above proposition. By contrast to the high SNR regime, high spatial correlations decrease both the diversity and the coding gains. The diversity gain is given by $n_r r(\tilde{\mathbf{E}}) - \alpha$.

As far as the coding gain is concerned, all space–time codes (rank-deficient and full-rank codes) are affected in such a way that the channel and the code cannot be separated in the expression of the coding gain (whereas it was the case for full-rank codes in the high SNR regime). The coding gain is made of two terms: the first term is proportional to the i.i.d. coding gain and the second term accounts for the interactions between the spatial correlation matrix and the code error matrices, via the projection of the codewords onto the space spanned by the eigenvectors of $\mathbf{R}$ (i.e. the spatial distribution of the scatterers).

For rank-deficient codes, the medium SNR regime is not qualitatively different from the high SNR regime: the interaction between the codeword and spatial correlations is also observed at high SNR, it is simply accentuated at medium SNR through $\det(\mathbf{L}'')$. However, for full-rank codes, $\det(\mathbf{Q}'')$ significantly differs from $\det(\mathbf{R})$. The consequence is that, in the medium SNR regime, both the eigenvectors and the eigenvalues of the error matrices impact the full-rank code performance. We can now understand how a full-rank code behaves in the medium SNR regime as if it were rank-deficient.

The maximization of the coding gain in i.i.d. Rayleigh channels is therefore not a sufficient condition to guarantee the good performance of a code in correlated channels at finite SNR, even for full-rank codes.

The above discussion also suggests designing error matrices with large eigenvalues and small condition numbers, so that the code offers no weak direction. Orthogonal space–time block codes (Section 5.5.4) are particularly robust in that sense. Delay diversity (see Section 5.6.1) also presents a similar robustness. However, the error matrices of codes designed with the trace criterion (see Section 5.3.2) generally exhibit large condition numbers. These codes are thus far more sensitive to spatial correlations in Rayleigh slow fading channels.

Example 6.3   *To illustrate the decrease in diversity, let us come back to the SM scheme of Example 6.1. Assuming that there is no receive correlation matrix, the PEP at finite SNR is then rewritten as*

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \left(1 + \eta (\mathbf{C} - \mathbf{E})^H \mathbf{R}_t (\mathbf{C} - \mathbf{E}) \right)^{-n_r} d\beta$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \left(1 + \eta \left| \mathbf{U}_{\mathbf{E}}^H (1,:) \mathbf{R}_t \mathbf{U}_{\mathbf{E}}(:,1) \right| \left\| \mathbf{C} - \mathbf{E} \right\|^2 \right)^{-n_r} d\beta. \tag{6.72}$$

Depending on the value of $\left| \mathbf{U}_{\mathbf{E}}^H(1,:)\mathbf{R}_t\mathbf{U}_{\mathbf{E}}(:,1) \right|$, the high SNR assumption may be quite unrealistic. Indeed, if $\mathbf{U}_{\mathbf{E}}(:,1)$ is parallel to the space spanned by the eigenvector of $\mathbf{R}_t$ corresponding to its lowest eigenvalue, $\left| \mathbf{U}_{\mathbf{E}}^H(1,:)\mathbf{R}_t\mathbf{U}_{\mathbf{E}}(:,1) \right|$ will be very small, so that no diversity is achieved at finite SNR. Only at very high values of the SNR would the diversity gain reach n_r.

PROOF OF PROPOSITION 6.3:   Consider the first scenario. Since the non-zero eigenvalues of $\mathbf{\Lambda}_{\mathbf{C}}' \mathbf{Q}'^{1/2} \mathbf{Q}'^{H/2} \mathbf{\Lambda}_{\mathbf{C}}'$ are equal to the non-zero eigenvalues of $\mathbf{Q}'^{H/2} \mathbf{\Lambda}_{\mathbf{C}}'^2 \mathbf{Q}'^{1/2}$, the coding gain may be rewritten as

$$\prod_{i=1+\alpha}^{n_r r(\mathbf{C}-\mathbf{E})} \lambda_i^{-1} \left(\mathbf{\Lambda}_{\mathbf{C}}' \mathbf{Q}' \mathbf{\Lambda}_{\mathbf{C}}' \right) = \prod_{i=1+\alpha}^{n_r r(\tilde{\mathbf{E}})} \lambda_i^{-1} \left(\mathbf{\Lambda}_{\mathbf{Q}'} \mathbf{U}_{\mathbf{Q}'}^H \mathbf{\Lambda}_{\mathbf{C}}'^2 \mathbf{U}_{\mathbf{Q}'} \mathbf{\Lambda}_{\mathbf{Q}'} \right) \tag{6.73}$$

Since α eigenvalues of $\mathbf{Q}'$ are negligible, we set them to zero, so that $\mathbf{\Lambda}_{\mathbf{Q}'}$ is now made of a principal non-zero submatrix $\mathbf{\Lambda}_{\mathbf{Q}'}'$ of size $\left(n_r r(\tilde{\mathbf{E}}) - \alpha \right) \times \left(n_r r(\tilde{\mathbf{E}}) - \alpha \right)$

$$\mathbf{\Lambda}_{\mathbf{Q}'} \approx \begin{bmatrix} \mathbf{\Lambda}_{\mathbf{Q}'}' & \mathbf{0}_{n_r r(\tilde{\mathbf{E}})-\alpha \times \alpha} \\ \mathbf{0}_{\alpha \times n_r r(\tilde{\mathbf{E}})-\alpha} & \mathbf{0}_{\alpha \times \alpha} \end{bmatrix} \tag{6.74}$$

Thus, the coding gain becomes

$$\prod_{i=1+\alpha}^{n_r r(\tilde{\mathbf{E}})} \lambda_i \left(\mathbf{\Lambda}_{\mathbf{C}}' \mathbf{Q}' \mathbf{\Lambda}_{\mathbf{C}}' \right) \approx \det \left(\mathbf{\Lambda}_{\mathbf{Q}'}' \mathbf{L}'' \mathbf{\Lambda}_{\mathbf{Q}'}' \right)$$

$$= \det \left(\mathbf{L}'' \right) \det \left(\mathbf{\Lambda}_{\mathbf{Q}'}'^2 \right) \tag{6.75}$$

with $\mathbf{L}''$ a principal submatrix of $\mathbf{U}_{\mathbf{Q}'}^H \mathbf{\Lambda}_{\mathbf{C}}'^2 \mathbf{U}_{\mathbf{Q}'}$ of size $n_r r(\tilde{\mathbf{E}}) - \alpha \times n_r r(\tilde{\mathbf{E}}) - \alpha$. Assuming that all eigenvalues of $\mathbf{\Lambda}_{\mathbf{C}}'$ are large, the PEP is thus approximated by

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r r(\tilde{\mathbf{E}})-\alpha)} \left(\det \left(\mathbf{L}'' \right) \right)^{-1} \left(\det \left(\mathbf{\Lambda}_{\mathbf{Q}'}'^2 \right) \right)^{-1} d\beta. \tag{6.76}$$

Applying the inclusion theorem (see Appendix A), we may write

$$\det\left(\mathbf{L}''\right) \geq \prod_{i=1}^{n_r r(\tilde{\mathbf{E}})-\alpha} \lambda_i(\boldsymbol{\Lambda}_{\mathbf{C}}'^2) = \prod_{i=1}^{r(\tilde{\mathbf{E}})-\delta} \lambda_i^{n_r}(\tilde{\mathbf{E}})$$

$$\det\left(\boldsymbol{\Lambda}_{\mathbf{Q}'}'^2\right) = \prod_{i=\alpha+1}^{n_r r(\tilde{\mathbf{E}})} \lambda_i(\mathbf{Q}') \geq \prod_{i=\alpha+1}^{n_r r(\tilde{\mathbf{E}})} \lambda_i(\mathbf{R}) \tag{6.77}$$

where $\lambda_i(\boldsymbol{\Lambda}_{\mathbf{C}}'^2)$, $\lambda_i(\tilde{\mathbf{E}})$, $\lambda_i(\mathbf{Q}')$ and $\lambda_i(\mathbf{R})$ are the non-zero eigenvalues in increasing order and δ is chosen such that $\alpha = \delta n_r$. The PEP (6.76) is then upper-bounded as

$$P(\mathbf{C} \rightarrow \mathbf{E}) \leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r r(\tilde{\mathbf{E}})-\alpha)} \prod_{i=1}^{r(\tilde{\mathbf{E}})-\delta} \lambda_i^{-n_r}(\tilde{\mathbf{E}}) \prod_{i=\alpha+1}^{n_r r(\tilde{\mathbf{E}})} \lambda_i^{-1}(\mathbf{R}) d\beta. \tag{6.78}$$

The diversity gain is equal to $n_r r(\mathbf{C} - \mathbf{E}) - \alpha$, and the coding gain is made of two terms:

1. the term $\det\left(\boldsymbol{\Lambda}_{\mathbf{Q}'}'^2\right)$ can be thought of as the coding gain provided by an equivalent channel with $n_r r(\tilde{\mathbf{E}}) - \alpha$ modes
2. the term $\det\left(\mathbf{L}''\right)$ is the coding gain inherent to the code.

For full rank codes, $\boldsymbol{\Lambda}_{\mathbf{Q}'}^2 = \boldsymbol{\Lambda}_{\mathbf{R}}$ and $\boldsymbol{\Lambda}_{\mathbf{Q}'}'$ is obtained by taking the square root of the $n_t n_r - \alpha$ largest eigenvalues of $\boldsymbol{\Lambda}_{\mathbf{R}}$: the MIMO channel loses α degrees of freedom out of $n_t n_r$. The part of the coding gain provided by the channel is now given by $\det\left(\boldsymbol{\Lambda}_{\mathbf{Q}'}'^2\right) = \prod_{i=1+\alpha}^{n_t n_r} \lambda_i(\mathbf{R})$ (instead of $\det(\mathbf{R})$ in the high SNR regime). The part of the coding gain inherent to the code is given by $\det(\mathbf{L}'')$ with $\mathbf{L}''$ a principal submatrix of $\mathbf{U}_{\mathbf{R}}^H(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}})\mathbf{U}_{\mathbf{R}}$ and is a function of the mapping of the codeword onto the space spanned by the eigenvectors of the space correlation matrix. Hence, the PEP achieved by full rank codes simplifies into

$$P(\mathbf{C} \rightarrow \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r r(\tilde{\mathbf{E}})-\alpha)} \left(\det\left(\mathbf{L}''\right)\right)^{-1} \prod_{i=1+\alpha}^{n_t n_r} \lambda_i^{-1}(\mathbf{R}) d\beta. \tag{6.79}$$

Let us now assume that α non-zero eigenvalues of the codeword error matrix $\boldsymbol{\Lambda}_{\mathbf{C}}'$ are negligible and thereby set to zero:

$$\boldsymbol{\Lambda}_{\mathbf{C}}' \approx \begin{bmatrix} \boldsymbol{\Lambda}_{\mathbf{C}}'' & \mathbf{0}_{n_r r(\tilde{\mathbf{E}})-\alpha \times \alpha} \\ \mathbf{0}_{\alpha \times n_r r(\tilde{\mathbf{E}})-\alpha} & \mathbf{0}_{\alpha \times \alpha} \end{bmatrix} \tag{6.80}$$

Reasoning as above

$$\prod_{i=1+\alpha}^{n_r r(\tilde{\mathbf{E}})} \lambda_i\left(\boldsymbol{\Lambda}_{\mathbf{C}}'\mathbf{Q}'\boldsymbol{\Lambda}_{\mathbf{C}}'\right) \approx \prod_{i=1+\alpha}^{n_r r(\tilde{\mathbf{E}})} \lambda_i\left(\boldsymbol{\Lambda}_{\mathbf{C}}''\mathbf{Q}''\boldsymbol{\Lambda}_{\mathbf{C}}''\right)$$

$$= \det\left(\boldsymbol{\Lambda}_{\mathbf{C}}''\mathbf{Q}''\boldsymbol{\Lambda}_{\mathbf{C}}''\right)$$

$$= \det{(\mathbf{Q}'')}\det{(\mathbf{\Lambda}_{\mathbf{C}}''^2)}$$

$$= \det{(\mathbf{Q}'')}\prod_{i=1+\delta}^{r(\tilde{\mathbf{E}})} \lambda_i\big(\tilde{\mathbf{E}}\big) \tag{6.81}$$

where $\mathbf{Q}''$ is a principal submatrix of $\mathbf{Q}'$ of size $n_r r(\tilde{\mathbf{E}}) - \alpha \times n_r r(\tilde{\mathbf{E}}) - \alpha$ and δ is chosen as $\alpha = \delta n_r$ without loss of generality. The non-zero eigenvalues of $\tilde{\mathbf{E}}$, $\lambda_i\big(\tilde{\mathbf{E}}\big)$, are ordered in increasing order. If all eigenvalues of $\mathbf{Q}'$ are large, the PEP at medium SNR is thus approximated by

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi}\int_0^{\pi/2} \eta^{-(n_r r(\tilde{\mathbf{E}}) - \alpha)}\big(\det{(\mathbf{Q}'')}\big)^{-1}\prod_{i=1+\delta}^{r(\tilde{\mathbf{E}})} \lambda_i^{-1}\big(\tilde{\mathbf{E}}\big)d\beta. \tag{6.82}$$

Using the Kronecker model, $\mathbf{Q}''$ can be further decomposed as $\mathbf{R}_r \otimes \mathbf{K}''$ where $\mathbf{K}''$ is a principal submatrix of $\mathbf{K}'$. From the inclusion theorem of Appendix A, we have

$$\det{(\mathbf{Q}'')} \geq \prod_{i=1}^{n_r r(\tilde{\mathbf{E}}) - \alpha} \lambda_i(\mathbf{Q}') \geq \prod_{i=1}^{n_r r(\tilde{\mathbf{E}}) - \alpha} \lambda_i(\mathbf{R}) \tag{6.83}$$

yielding

$$P(\mathbf{C} \to \mathbf{E}) \leq \frac{1}{\pi}\int_0^{\pi/2} \eta^{-(n_r r(\tilde{\mathbf{E}}) - \alpha)}\prod_{i=1}^{n_r r(\tilde{\mathbf{E}}) - \alpha} \lambda_i^{-1}(\mathbf{R}) \prod_{i=1+\delta}^{r(\tilde{\mathbf{E}})} \lambda_i^{-n_r}\big(\tilde{\mathbf{E}}\big)d\beta \tag{6.84}$$

where the eigenvalues are ordered by increasing value.

The diversity gain is thus equal to $n_r r(\tilde{\mathbf{E}}) - \alpha$. The coding gain, analogous to the high SNR regime, is made of two terms:

1. the term $\prod_{i=1+\delta}^{r(\tilde{\mathbf{E}})} \lambda_i\big(\tilde{\mathbf{E}}\big)$ is similar to the i.i.d. coding gain, except that α eigenvalues are lost
2. the second term $\det{(\mathbf{Q}'')}$ accounts for the interaction between the codeword and the spatial correlations. ∎

Rayleigh block fading channels
Starting from (6.26), the expressions of the average PEP for block fading channels at medium SNR are straightforward, using the derivations carried out for slow fading channels.

Rayleigh fast fading channels
In fast fading channels at medium SNR, analogous to the slow fading case, we assume that α eigenvalues of $\mathbf{\Xi}\mathbf{\Delta}$ are too small to contribute to diversity. The PEP of (6.24) thereby

becomes

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \eta^{-r(\boldsymbol{\Xi}\boldsymbol{\Delta})-\alpha} \prod_{i=1+\alpha}^{r(\boldsymbol{\Xi}\boldsymbol{\Delta})} \lambda_i^{-1}(\boldsymbol{\Xi}\boldsymbol{\Delta}) \, d\beta$$

$$\overset{(a)}{=} \frac{1}{\pi} \int_0^{\pi/2} \eta^{-r(\boldsymbol{\Xi}\boldsymbol{\Delta})-\alpha} \prod_{i=1+\alpha}^{r(\boldsymbol{\Xi}\boldsymbol{\Delta})} \lambda_i^{-1} \left(\boldsymbol{\Lambda}'_{\mathbf{D}'} \mathbf{O}' \boldsymbol{\Lambda}'_{\mathbf{D}'} \right) d\beta \qquad (6.85)$$

where the non-zero eigenvalues λ_i are in increasing order, and (a) relies on (6.49). Again, the existence of small eigenvalues may result from two non-exclusive situations: α eigenvalues of $\boldsymbol{\Lambda}'_{\mathbf{D}'}$ or of $\mathbf{O}' = \mathbf{U}_{\mathbf{O}'} \boldsymbol{\Lambda}_{\mathbf{O}'}^2 \mathbf{U}_{\mathbf{O}'}^H$ are negligible. In the second situation, $\boldsymbol{\Lambda}_{\mathbf{O}'}$ is formed of a diagonal submatrix $\boldsymbol{\Lambda}'_{\mathbf{O}'}$ containing the $n_r l_{\mathbf{C},\mathbf{E}} - \alpha$ largest singular values.

Proposition 6.4 *Consider a finite SNR in Rayleigh fast fading channels with full-rank spatio-temporal correlation matrix* $\mathbf{R}$,

1. *if α eigenvalues of $\boldsymbol{\Lambda}'_{\mathbf{D}'}$ are negligible, the average PEP is expressed as*

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l'_{\mathbf{C},\mathbf{E}}} \left(\det \left(\mathbf{O}'' \right) \right)^{-1} \prod_{k \in \tau'_{\mathbf{C},\mathbf{E}}} \| \mathbf{c}_k - \mathbf{e}_k \|^{-2n_r} \, d\beta \qquad (6.86)$$

$$\leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l'_{\mathbf{C},\mathbf{E}}} \prod_{i=1}^{n_r l_{\mathbf{C},\mathbf{E}}-\alpha} \lambda_i^{-1}(\boldsymbol{\Xi}) \prod_{k \in \tau'_{\mathbf{C},\mathbf{E}}} \| \mathbf{c}_k - \mathbf{e}_k \|^{-2n_r} \, d\beta \qquad (6.87)$$

   where $\tau'_{\mathbf{C},\mathbf{E}} = \{k | \eta \| \mathbf{c}_k - \mathbf{e}_k \|^2 >> 1\}$ with $l'_{\mathbf{C},\mathbf{E}} = \sharp \tau'_{\mathbf{C},\mathbf{E}}$ and $\mathbf{O}''$ is a principal submatrix of $\mathbf{O}'$ of size $n_r l_{\mathbf{C},\mathbf{E}} - \alpha \times n_r l_{\mathbf{C},\mathbf{E}} - \alpha$.
2. *if α eigenvalues of $\mathbf{O}'$ are negligible, the average PEP is expressed as*

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r l_{\mathbf{C},\mathbf{E}}-\alpha)} \left(\det \left(\mathbf{M}'' \right) \right)^{-1} \left(\det \left(\boldsymbol{\Lambda}_{\mathbf{O}'}'^2 \right) \right)^{-1} d\beta \qquad (6.88)$$

$$\leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r l_{\mathbf{C},\mathbf{E}}-\alpha)} \prod_{i=1+\alpha}^{n_r l_{\mathbf{C},\mathbf{E}}} \lambda_i^{-1}(\boldsymbol{\Xi}) \prod_{k \in \tau''_{\mathbf{C},\mathbf{E}}} \| \mathbf{c}_k - \mathbf{e}_k \|^{-2n_r} \, d\beta \qquad (6.89)$$

   where $\mathbf{M}''$ is a principal submatrix of $\mathbf{U}_{\mathbf{O}'}^H \boldsymbol{\Lambda}_{\mathbf{D}'}'^2 \mathbf{U}_{\mathbf{O}'}$ of size $n_r l_{\mathbf{C},\mathbf{E}} - \alpha \times n_r l_{\mathbf{C},\mathbf{E}} - \alpha$. $\tau''_{\mathbf{C},\mathbf{E}}$ is the set containing the positions k that correspond to the $n_r l_{\mathbf{C},\mathbf{E}} - \alpha$ smallest $\| \mathbf{c}_k - \mathbf{e}_k \|^2$ in $\tau_{\mathbf{C},\mathbf{E}}$.

By contrast to the high SNR situation, both the diversity and the coding gains are affected by the interaction between the error matrices and the channel correlation matrix. They are a function of the projection of the columns of the error matrices onto the space spanned by eigenvectors of the channel correlation matrix. If those projections are small w.r.t. η^{-1}, the code acts as if its effective length was reduced. In the high SNR regime, the impact

on diversity is not present since the SNR is large enough to extract the effects induced by each non-zero column of the error matrix, irrespective of the projection on the channel correlation matrix.

Example 6.4   *Several reasons may account for* $\mathbf{O}'$ *to loose* α *eigenvalues at medium SNR. As an example, the temporal correlation may be so large that full time diversity is no longer achievable: the fast fading channel acts at medium SNR as if it were slow or block fading. In such case, (6.88) simplifies into the high-SNR expressions valid for slow or block fading.*

The loss of α *eigenvalues is also caused by the destructive interaction between codewords and spatial correlation. To illustrate this, let us assume that the Kronecker model is used to model spatial correlation. If the transmit correlation is large enough, there is a non-negligible probability for* $\mathbf{c}_k - \mathbf{e}_k$ *to lie in the space spanned by the eigenvector of* $\mathbf{R}_t$ *corresponding to the smallest eigenvalue, causing* $(\mathbf{c}_k - \mathbf{e}_k)^H \mathbf{R}_t(\mathbf{c}_k - \mathbf{e}_k)$ *to fall below* η^{-1}. *The position k will, therefore, not contribute to temporal diversity. Defining*

$$\tau_{medium} = \left\{ k \,|\, \eta| \,(\mathbf{c}_k - \mathbf{e}_k)^H \, \mathbf{R}_t \,(\mathbf{c}_k - \mathbf{e}_k)\,| \gg 1 \right\}, \tag{6.90}$$

the apparent effective length of the error matrix at medium SNR is now given by $\sharp\tau_{medium}$. *The diversity offered by the pair of codewords is equal to* $n_r\sharp\tau_{medium}$. *Eventually, the PEP (6.88) is rewritten as*

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r\sharp\tau_{medium}} \left(\det(\mathbf{R}_r)\right)^{-\sharp\tau_{medium}}$$

$$\times \prod_{k\in\tau_{medium}} |\,(\mathbf{c}_k - \mathbf{e}_k)^H \, \mathbf{R}_t \,(\mathbf{c}_k - \mathbf{e}_k)\,|^{-n_r} d\beta \tag{6.91}$$

which should be compared to (6.52) in the high SNR regime.

PROOF OF PROPOSITION 6.4:   We discuss the average PEP in the two aforementioned situations.

Assume that α eigenvalues of $\mathbf{\Lambda}'_{\mathbf{D}'}$ are significantly small and, without loss of generality, that $\alpha = \delta n_r$. Then, having small eigenvalues δ positions $k \in \tau_{\mathbf{C},\mathbf{E}}$ such that $\eta\|\mathbf{c}_k - \mathbf{e}_k\|^2$ is much smaller than one. We denote as $\tau'_{\mathbf{C},\mathbf{E}} = \{k\,|\,\eta\|\mathbf{c}_k - \mathbf{e}_k\|^2 \gg 1\}$ a new set of positions of size $l'_{\mathbf{C},\mathbf{E}} = \sharp\tau'_{\mathbf{C},\mathbf{E}}$. Matrix $\mathbf{\Lambda}'_{\mathbf{D}'}$ is thus made of a non-zero principal submatrix $\mathbf{\Lambda}''_{\mathbf{D}'}$ of size $n_r l_{\mathbf{C},\mathbf{E}} - \alpha \times n_r l_{\mathbf{C},\mathbf{E}} - \alpha$. Analogous to the previous sections, we get

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l'_{\mathbf{C},\mathbf{E}}} \left(\det(\mathbf{O}'')\right)^{-1} \prod_{k\in\tau'_{\mathbf{C},\mathbf{E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{-2n_r} d\beta \tag{6.92}$$

where $\mathbf{O}''$ is a principal submatrix of $\mathbf{O}'$ of size $n_r l_{\mathbf{C,E}} - \alpha \times n_r l_{\mathbf{C,E}} - \alpha$. Using the lower bound

$$\det\left(\mathbf{O}''\right) \geq \prod_{i=1}^{n_r l_{\mathbf{C,E}}-\alpha} \lambda_i(\boldsymbol{\Xi}) \tag{6.93}$$

with $\lambda_i(\boldsymbol{\Xi})$ in increasing order of magnitude, the PEP is finally upper-bounded as

$$P(\mathbf{C} \to \mathbf{E}) \leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l'_{\mathbf{C,E}}} \prod_{i=1}^{n_r l_{\mathbf{C,E}}-\alpha} \lambda_i^{-1}(\boldsymbol{\Xi}) \prod_{k \in \tau'_{\mathbf{C,E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{-2n_r} d\beta. \tag{6.94}$$

The diversity order is reduced by α. The coding gain is made of two terms:

1. a coding gain similar to the coding gain in i.i.d. Rayleigh channels, except that the effective length of the error matrix is reduced to $l'_{\mathbf{C,E}}$
2. a second term that accounts for the relative positions of the eigenvectors of the error matrix in the space spanned by the eigenvectors of the channel correlation matrix.

Assume now that $\mathbf{O}'$ has α negligible eigenvalues, and that we can decompose $\mathbf{O}' = \mathbf{O}'^{1/2}\mathbf{O}'^{H/2} = \mathbf{U}_{\mathbf{O}'}\boldsymbol{\Lambda}_{\mathbf{O}'}^2\mathbf{U}_{\mathbf{O}'}^H$. If α eigenvalues in $\boldsymbol{\Lambda}_{\mathbf{O}'}$ are negligible, we may set them to zero and denote the principal submatrix of $\boldsymbol{\Lambda}_{\mathbf{O}'}$ consisting in the non-zero eigenvalues as $\boldsymbol{\Lambda}'_{\mathbf{O}'}$. The PEP is then easily evaluated as

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r l_{\mathbf{C,E}}-\alpha)} \left(\det\left(\mathbf{M}''\right)\right)^{-1} \left(\det\left(\boldsymbol{\Lambda}_{\mathbf{O}'}'^2\right)\right)^{-1} d\beta \tag{6.95}$$

where $\mathbf{M}''$ is a principal submatrix of $\mathbf{U}_{\mathbf{O}'}^H \boldsymbol{\Lambda}_{\mathbf{D}'}'^2 \mathbf{U}_{\mathbf{O}'}$ of size $n_r l_{\mathbf{C,E}} - \alpha \times n_r l_{\mathbf{C,E}} - \alpha$. Furthermore, we may always express the following lower bounds

$$\det\left(\mathbf{M}''\right) \geq \prod_{i=1}^{n_r l_{\mathbf{C,E}}-\alpha} \lambda_i\left(\boldsymbol{\Lambda}_{\mathbf{D}'}'^2\right) = \prod_{k \in \tau''_{\mathbf{C,E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{2n_r}$$

$$\det\left(\boldsymbol{\Lambda}_{\mathbf{O}'}'^2\right) \geq \prod_{i=1+\alpha}^{n_r l_{\mathbf{C,E}}} \lambda_i\left(\mathbf{O}'\right) \geq \prod_{i=1+\alpha}^{n_r l_{\mathbf{C,E}}} \lambda_i\left(\boldsymbol{\Xi}'\right) \geq \prod_{i=1+\alpha}^{n_r l_{\mathbf{C,E}}} \lambda_i(\boldsymbol{\Xi}) \tag{6.96}$$

where $\tau''_{\mathbf{C,E}}$ is the set containing the positions k that correspond to the $n_r l_{\mathbf{C,E}} - \alpha$ smallest $\|\mathbf{c}_k - \mathbf{e}_k\|^2$ in $\tau_{\mathbf{C,E}}$ and the eigenvalues are ordered in increasing order. The PEP is finally upper-bounded as

$$P(\mathbf{C} \to \mathbf{E}) \leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-(n_r l_{\mathbf{C,E}}-\alpha)} \prod_{i=1+\alpha}^{n_r l_{\mathbf{C,E}}} \lambda_i^{-1}(\boldsymbol{\Xi}) \prod_{k \in \tau''_{\mathbf{C,E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{-2n_r} d\beta. \tag{6.97}$$

Note the difference with (6.94). The diversity order is still reduced by α and the two terms in the coding gain are now as follows:

1. the term $\det(\Lambda'^2_{\mathbf{O}'})$ can be thought of as the coding gain provided by an equivalent channel with $n_r l_{\mathbf{C,E}} - \alpha$ degrees of freedom
2. the term $\det(\mathbf{M}'')$ is the coding gain inherent to the code.

Both terms, however, account for the interaction between the channel and the code.   ∎

6.3.3   Low SNR regime

Proposition 6.5   *At low SNR, $\eta\lambda_i(\boldsymbol{\Xi}\boldsymbol{\Delta}) < 1$, and the average PEP is given by*

$$P(\mathbf{C} \rightarrow \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \left(1 + \eta n_r \operatorname{Tr}\{\mathbf{R}_t\tilde{\mathbf{E}}\}\right)^{-1} d\beta \tag{6.98}$$

$$\leq \frac{1}{\pi} \int_0^{\pi/2} \left(1 + \eta n_r \sum_{i=1}^{n_t} \lambda_i(\mathbf{R}_t)\lambda_i(\tilde{\mathbf{E}})\right)^{-1} d\beta, \tag{6.99}$$

with the eigenvalues of $\mathbf{R}_t$ ranked in increasing order, and those of $\tilde{\mathbf{E}}$ ranked in decreasing order.

In the low SNR regime, no diversity is achieved. The MIMO channel behaves as a SISO channel with equivalent transmitted codewords given by $\mathbf{R}_t^{1/2}\mathbf{C}$ such that $\operatorname{Tr}\{\mathbf{R}_t\tilde{\mathbf{E}}\}$ acts as the Euclidean distance seen by the receiver. In the low SNR regime, space–time codes interact with the channel through the projection of the eigenvectors of $\tilde{\mathbf{E}}$ onto the space spanned by the eigenvectors of the transmit correlation matrix. In this respect, (6.99) suggests that a code error matrices characterized by a large minimum eigenvalue and a small condition number should perform reasonably well in correlated channels, because such a code would not present any weak direction. Note, however, that (6.99) only gives an upper bound on the performance. By directly accounting for the destructive interactions between the error matrix and transmit correlation in the code design, the performance might be further improved.

PROOF OF PROPOSITION 6.4:   The proof of (6.98) is straightforward and left as an exercise to the reader. To prove the lower bound of (6.99), we decompose $\mathbf{R}_t = \mathbf{V}_{\mathbf{R}_t}\Lambda_{\mathbf{R}_t}\mathbf{V}_{\mathbf{R}_t}^H$ with the eigenvalues of $\Lambda_{\mathbf{R}_t}$ in increasing order. Similarly, $\tilde{\mathbf{E}} = \mathbf{V}_{\tilde{\mathbf{E}}}\Lambda_{\tilde{\mathbf{E}}}\mathbf{V}_{\tilde{\mathbf{E}}}^H$ but with the eigenvalues in $\Lambda_{\tilde{\mathbf{E}}}$ in decreasing order. We may then write

$$\operatorname{Tr}\{\mathbf{R}_t\tilde{\mathbf{E}}\} \geq \sum_{i=1}^{n_t} \lambda_i(\mathbf{R}_t)\,\lambda_i(\tilde{\mathbf{E}}). \tag{6.100}$$

Equality occurs when $\mathbf{V}_{\mathbf{R}_t}^H\mathbf{V}_{\tilde{\mathbf{E}}} = \mathbf{I}_{n_t}$.   ∎

6.3.4   Summary and examples

In Rayleigh fading channels in the **high SNR regime**,

- all space–time codes extract a diversity gain equal to the gain achieved in i.i.d. Rayleigh channels
- the coding gain is affected by spatio-temporal correlations:
  - for full-rank codes in slow fading conditions, the coding gain loss is independent of the code structure, hence the design of a full-rank code in spatially correlated Rayleigh slow fading channels can use criteria developed for i.i.d. Rayleigh slow fading channels
  - for rank-deficient codes or in fast fading channels, interactions between the spatio-temporal correlations and the code error matrices induce a code-dependent loss in the coding gain, and the performance of a code in correlated channels is not only related to its performance in i.i.d. channels.

In Rayleigh fading channels in the **medium SNR regime**, both the diversity and the coding gains are affected by spatio-temporal correlations, irrespective of the rank of the code:

- the total diversity gain is decreased by the number of negligible eigenvalues of $\mathbf{\Xi\,\Delta}$ (fast fading) or $\mathbf{C_R}$ (slow fading) with respect to the diversity in i.i.d. channels
- the coding gain strongly depends on the interactions between the code error matrices and the spatial correlations.

In Rayleigh fading channels in the **low SNR regime**,

- no diversity is extracted
- the coding gain is independent of the receive and temporal correlations, as only transmit correlations affect the error probability.

It is worth highlighting some obvious similarities between the pairwise error probability and the diversity-multiplexing trade-off in correlated Rayleigh slow fading channels at both infinite and finite SNR (see Section 4.7). In the high SNR regime, the diversity-multiplexing trade-off remains unaffected by spatial correlations, but not at finite SNR. Similarly, for full-rank codes, we have seen that the rank-determinant criterion is optimal at high SNR (in slow fading channels) while in the medium SNR regime, the design criterion should account for the spatial correlations. This is in contrast with the popular belief that full-rank codes are robust and that the design criteria in correlated channels are strictly identical to the rank-determinant and distance-product criteria. The above results clearly illustrate that this reasoning may be loose at practical SNR levels, because it neglects all interactions between the full-rank code error matrices and the spatio-temporal correlations. Using the i.i.d. designs at finite SNR might result in non-robust codes, for which the error matrix responsible for increasing the error probability depends on the spatial correlation and is thereby not necessarily the minimum distance code error matrix.
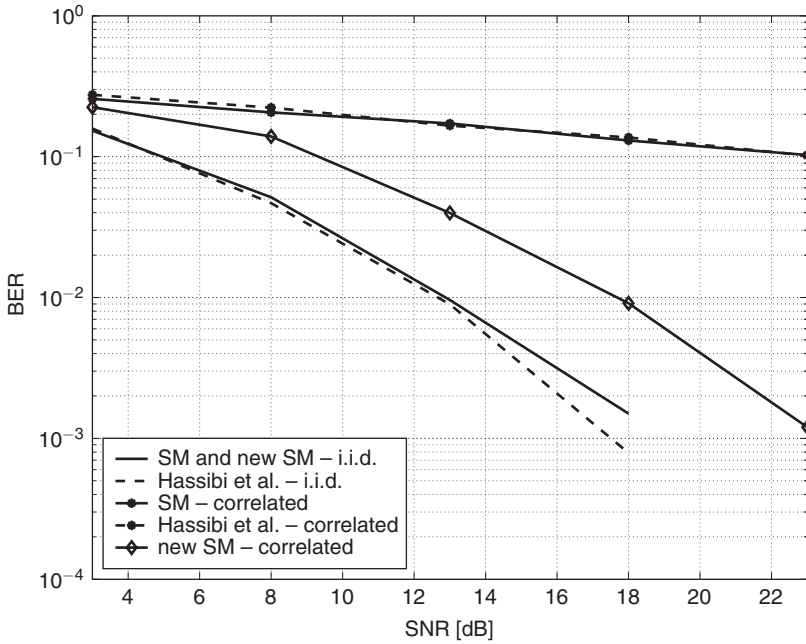
Figure 6.6   *Performance of rank-deficient LDCs on i.i.d. and correlated channels with $n_t = 2$ and $n_r = 2$*

To illustrate this important notion, let us analyze in Figure 6.6 the performance of

- a classical 2×2 Spatial Multiplexing (SM) scheme transmitting independent symbols on each transmit antenna
- a so-called robust Spatial Multiplexing code (obtained using the optimal criterion derived in Chapter 7) denoted as "new SM" and offering the same performance in i.i.d. channels as the classical SM scheme
- a rank-one linear dispersion code expanding on 2 symbol durations and transmitting 4 symbols per codeword (see Example 5.15), denoted as "Hassibi et al." [HH01]

in a correlated slow fading channel generated using the geometry-based stochastic model of Section 2.3.3, with inline (regarding the link axis) array configuration at the mobile terminal (antenna spacing of 0.5 λ) and broadside array configuration at the base station (antenna spacing of 20 λ). A Rayleigh scenario with highly directional propagation along the link axis is considered, resulting in high spatial correlations. Note also that all results are obtained with QPSK constellations and ML decoding. The strong interactions between the code error matrices and the spatial correlations cause the classical SM scheme and the "Hassibi et al." code to evolve in the low SNR regime: the diversity achieved by those codes in the correlated channel at 20 dB is very poor. Actually, the diversity foreseen by (6.29) would only be observed at extremely high SNR levels, far beyond SNR levels used in practice. By contrast, the robust SM scheme is evolving in the high SNR regime at
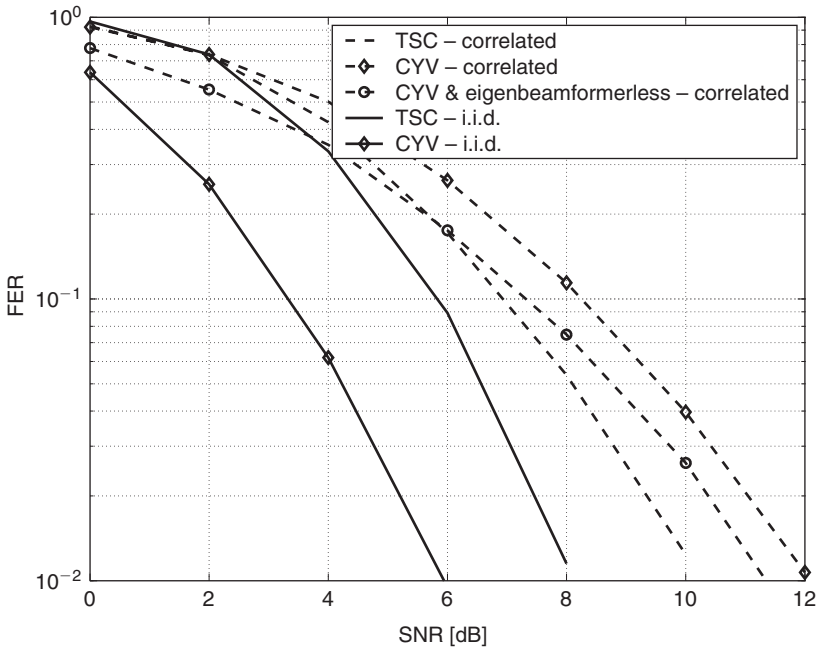
Figure 6.7 *Performance of STTCs in i.i.d. and correlated channels with $n_t = 2$ and $n_r = 4$*

15 dB SNR. Its diversity in correlated channels is the same as in i.i.d. channels, and only the coding gain is affected by the spatial correlations, resulting in a shift to the right.

In Figure 6.7, the frame error rates of 4-state space–time trellis codes in i.i.d. and correlated Rayleigh slow fading channels are displayed. Analogous to the LDCs, the performance of STTCs in correlated channels depends on the considered code: the "TSC" code [TSC98], while performing worse in i.i.d. channels than the "CYV" code [CYV01], is more robust in correlated channels. Once again, this illustrates that (6.30) gives a very poor estimation of the performance of space–time coding at finite SNR. Moreover, the combination of the "CYV" code with a supposedly optimal precoder (denoted as "CYV & eigenbeam-former"), designed with the knowledge of the transmit correlation (see Chapter 8), is only marginally improved. Why is that so? Actually, the optimal precoder might not be the one matching the minimum distance code error matrix to the transmit eigenvectors, as the above developments have shown that the minimum distance code error matrix is not necessarily the error matrix responsible for increasing the PEP in correlated channels at realistic SNR. But the precoder used here accounts for the transmit correlation by applying a solution aimed to work in the medium SNR regime to the minimum distance code error matrix, although the latter is responsible for the increase in PEP only in the high SNR regime. This might have succeeded for codes whose minimum distance code error matrix is independent of the channel correlations, e.g. for O-STBC and delay diversity schemes. However, for many other codes, this is not the best strategy, as illustrated in Figure 6.7. Note that the optimal design of precoders is further developed in Chapter 8.
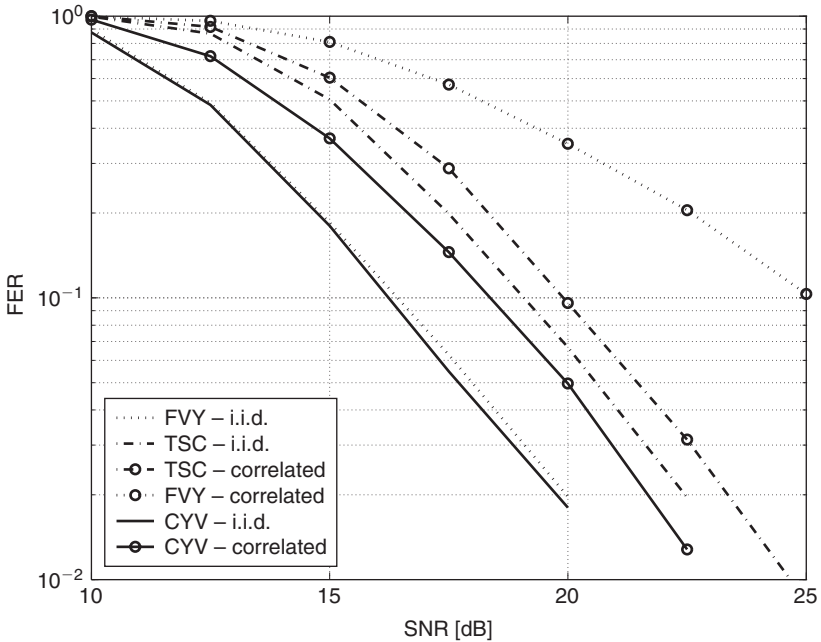
Figure 6.8   *Performance of STTCs in i.i.d. and correlated Rayleigh fast fading channels with $n_t = 2$ and $n_r = 1$*

In Figure 6.8, we investigate the performance of STTCs over spatially correlated and i.i.d. Rayleigh fast fading channels. The propagation channel and array configurations are analogous to the above slow fading scenario, except that the system is now ideally time interleaved. The "TSC" code [TSC98] has similar performance in i.i.d. and correlated channels in terms of diversity and coding gains. This suggests that this code evolves in the high SNR regime, and its performance is predicted by (6.50). This code can be qualified as robust (though its performance in i.i.d. channels could still be improved). By contrast, the performance of the "FVY" code [FVY01] is much better in i.i.d. channels than that of the "TSC" code, but its performance in correlated channels is extremely poor. Indeed, it looses both diversity and coding gains: this code evolves in the medium SNR regime in correlated channels, and its performance is better predicted by (6.82) and (6.88). Finally, the "CYV" code [CYV01] outperforms the two other codes in both i.i.d. and correlated Rayleigh fast fading channels. It can be considered as resulting from an 'optimal' design, as the high SNR regime is already achieved at realistic SNR levels.

6.4   Average pairwise error probability in Ricean fading channels

In Ricean fading channels, the analysis of the average PEP is much more difficult than in Rayleigh channels, as the performance is determined by the interactions between three terms: the spatio-temporal correlation matrix, the code error matrix and the coherent

component $\bar{\mathbf{H}}$. However, the intuitive impact of the coherent component can still be predicted in high, medium and low SNR regimes.

6.4.1 High SNR regime

Ricean slow fading channels

In the high SNR regime, $\eta\lambda_i(\mathbf{C}_\mathbf{R}) \gg 1$ for all values of i, the average PEP is highly dependent upon the rank of the error matrix $\tilde{\mathbf{E}}$, which we decompose as $\tilde{\mathbf{E}} = \mathbf{V}_{\tilde{\mathbf{E}}}\boldsymbol{\Lambda}_{\tilde{\mathbf{E}}}\mathbf{V}_{\tilde{\mathbf{E}}}^H$, with the eigenvalues in decreasing order.

Proposition 6.6 *At high SNR, the average PEP of a full-rank code in correlated Ricean slow fading channels is given by*

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_t n_r} (\det(\mathbf{R}))^{-1} (\det(\tilde{\mathbf{E}}))^{-n_r} e^{-K \operatorname{vec}(\bar{\mathbf{H}}^H)^H \mathbf{R}^{-1} \operatorname{vec}(\bar{\mathbf{H}}^H)} d\beta. \quad (6.101)$$

where $\eta = \rho/(4 \sin^2 \beta(1+K))$ is defined as the effective SNR.

The high-SNR PEP of a rank-deficient code in i.i.d. Ricean slow fading channels reads as

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r r(\tilde{\mathbf{E}})} \prod_{k=1}^{r(\tilde{\mathbf{E}})} \lambda_k^{-n_r}(\tilde{\mathbf{E}}) e^{-K \operatorname{Tr}\left\{\bar{\mathbf{H}}^H \bar{\mathbf{H}} \mathbf{V}_{\tilde{\mathbf{E}}} \underline{\boldsymbol{\Delta}} \mathbf{V}_{\tilde{\mathbf{E}}}^H\right\}} d\beta$$

$$\leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r r(\tilde{\mathbf{E}})} \prod_{k=1}^{r(\tilde{\mathbf{E}})} \lambda_k^{-n_r}(\tilde{\mathbf{E}}) e^{-K \sum_{i=1}^{r(\tilde{\mathbf{E}})} \lambda_i(\bar{\mathbf{H}}^H \bar{\mathbf{H}})} d\beta \quad (6.102)$$

where $\underline{\boldsymbol{\Lambda}} = \operatorname{diag}\{\mathbf{I}_{r(\tilde{\mathbf{E}})}, \mathbf{0}_{n_t - r(\tilde{\mathbf{E}})}\}$. The eigenvalues $\lambda_i(\bar{\mathbf{H}}^H \bar{\mathbf{H}})$ are ordered by increasing value.

What does reveal Proposition 6.6? First, we observe that the coherent component does not affect the achieved diversity gain equal to $n_r r(\tilde{\mathbf{E}})$, irrespective of the rank of the code. Note that this is similar to the high-SNR diversity-multiplexing trade-off in Ricean channels (see Chapter 4).

For full-rank codes, the coding gain is modified according to the projection of the coherent component $\bar{\mathbf{H}}$ onto the space spanned by the spatial correlation matrix $\mathbf{R}$, which is independent of the code itself. This has two consequences:

- the coherent component may be beneficial or detrimental to the PEP, depending on its projection on the space of $\mathbf{R}$; if the Rayleigh component is i.i.d. ($\mathbf{R} = \mathbf{I}_{n_t n_r}$), the impact of the coherent component is always beneficial;
- the presence of a dominant component does not impact the design criterion; hence, the design of a full-rank code in Ricean fading channels in the high SNR regime can be based on the rank-determinant criterion.

Furthermore, in the absence of receive correlation (taking a Kronecker-structured channel) and assuming that $\bar{\mathbf{H}}$ is rank-one, the average PEP can be upper-bounded by

$$P(\mathbf{C} \to \mathbf{E}) \leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_t n_r} \left(\det(\mathbf{R}_t)\right)^{-n_r} \left(\det(\tilde{\mathbf{E}})\right)^{-n_r} \exp\left(-K n_r\right) d\beta. \tag{6.103}$$

As it will appear in the proof, the coherent component is always beneficial to the PEP if $(1+K)^{n_t n_r} \exp\left(-K n_t n_r \lambda_{max}^{-1}(\mathbf{R}_t)\right) < 1$. Hence, if $\mathbf{R}_t$ is identity (so the Rayleigh component $\tilde{\mathbf{H}}$ is uncorrelated), the coherent component is always beneficial if $n_t \ln(1+K) < K$, irrespective of the value of the correlation matrix $\mathbf{R}_t$.

For rank-deficient codes, the impact of the coherent component on the coding gain depends on the code via the projection of the error matrix onto the space spanned by $\bar{\mathbf{H}}$. The coherent component is beneficial only if

$$(1+K)^{n_r r(\tilde{\mathbf{E}})} e^{-K \mathrm{Tr}\left\{\bar{\mathbf{H}}^H \bar{\mathbf{H}} \mathbf{V}_{\bar{\mathbf{E}}} \Delta \mathbf{V}_{\bar{\mathbf{E}}}^H\right\}} < 1.$$

Clearly, this depends on the code structure: for some codes, the coherent component may improve the error performance whereas for other codes, it may be detrimental. A good design should, therefore, prevent the eigenvectors of $\tilde{\mathbf{E}}$ corresponding to the largest eigenvalues to lie in the space spanned by the eigenvectors of $\bar{\mathbf{H}}$ corresponding to the smallest eigenvalues.

What is hidden in the above equations is that the effective SNR η is inversely proportional to the K-factor. This means that the high SNR regime is shifted to a larger SNR with respect to Rayleigh fading channels, and that the larger the K-factor, the larger this shift. Therefore, in high K-factor Ricean fading channels, the high-SNR assumption is even less realistic than in Rayleigh fading channels.

PROOF OF PROPOSITION 6.4:   For a full-rank error matrix, the proof of (6.101) is straightforward. The derivation of the upper bound relies on the Kronecker model $\mathbf{R} = \mathbf{R}_r \otimes \mathbf{R}_t$, which enables us to write (see Appendix A)

$$\mathrm{vec}\left(\bar{\mathbf{H}}^H\right)^H \mathbf{R}^{-1} \mathrm{vec}\left(\bar{\mathbf{H}}^H\right) = \mathrm{Tr}\left\{\bar{\mathbf{H}} \mathbf{R}_t^{-1} \bar{\mathbf{H}}^H \left(\mathbf{R}_r^{-1}\right)^T\right\}. \tag{6.104}$$

Assuming that $\mathbf{R}_r = \mathbf{I}_{n_r}$ and making the following decompositions $\bar{\mathbf{H}}^H \bar{\mathbf{H}} = \mathbf{V}_{\bar{\mathbf{H}}} \mathbf{\Lambda}_{\bar{\mathbf{H}}} \mathbf{V}_{\bar{\mathbf{H}}}^H$, $\mathbf{R}_t = \mathbf{V}_{\mathbf{R}_t} \mathbf{\Lambda}_{\mathbf{R}_t} \mathbf{V}_{\mathbf{R}_t}^H$, $\mathbf{V}_{\mathbf{Q}} = \mathbf{V}_{\bar{\mathbf{H}}}^H \mathbf{V}_{\mathbf{R}_t}$, we have the following lower bound

$$\mathrm{Tr}\left\{\bar{\mathbf{H}}^H \bar{\mathbf{H}} \mathbf{R}_t^{-1}\right\} = \mathrm{Tr}\left\{\mathbf{\Lambda}_{\bar{\mathbf{H}}} \mathbf{V}_{\mathbf{Q}} \mathbf{\Lambda}_{\mathbf{R}_t}^{-1} \mathbf{V}_{\mathbf{Q}}^H\right\} \tag{6.105}$$

$$= \sum_{k=1}^{n_t} \sum_{l=1}^{n_t} |\mathbf{V}_{\mathbf{Q}}(k,l)|^2 \lambda_l\left(\bar{\mathbf{H}}^H \bar{\mathbf{H}}\right) \lambda_k^{-1}\left(\mathbf{R}_t\right) \tag{6.106}$$

$$\geq \sum_{k=1}^{n_t} \lambda_k\left(\bar{\mathbf{H}}^H \bar{\mathbf{H}}\right) \lambda_k^{-1}\left(\mathbf{R}_t\right) \tag{6.107}$$

where the eigenvalues of $\bar{\mathbf{H}}^H\bar{\mathbf{H}}$, $\lambda_k(\bar{\mathbf{H}}^H\bar{\mathbf{H}})$ and $\mathbf{R}_t$, $\lambda_k(\mathbf{R}_t)$ are in increasing order of magnitude. Equality occurs when $\mathbf{V}_\mathbf{Q} = \mathbf{I}_{n_t}$, i.e. when $\mathbf{V}_{\bar{\mathbf{H}}} = \mathbf{V}_{\mathbf{R}_t}$. If the coherent component $\bar{\mathbf{H}}$ is rank-deficient, its only non-zero eigenvalue equals $n_t n_r$ and

$$\mathrm{Tr}\{\bar{\mathbf{H}}^H\bar{\mathbf{H}}\mathbf{R}_t^{-1}\} \geq n_t n_r \lambda_{max}^{-1}(\mathbf{R}_t) \geq n_r \qquad (6.108)$$

with $\lambda_{max}(\mathbf{R}_t)$ denoting the largest eigenvalue of $\mathbf{R}_t$. Since all but the largest eigenvalue of $\bar{\mathbf{H}}^H\bar{\mathbf{H}}$ are equal to zero, the lower bound in (6.108) is simply achieved if the main eigenvector of $\mathbf{R}_t$ lies in the space spanned by the main eigenvector of $\bar{\mathbf{H}}^H\bar{\mathbf{H}}$. This yields the upper bound of (6.103). Note that the coherent component is beneficial to the PEP if $(1+K)^{n_t n_r}\exp(-Kn_t n_r\lambda_{max}^{-1}(\mathbf{R}_t)) < 1$, i.e. if $\mathbf{R}_t$ is identity, or equally if the Rayleigh component $\tilde{\mathbf{H}}$ is uncorrelated.

For ***rank-deficient*** codes, the situation is sensibly different. For simplicity, let us assume that the Rayleigh component is i.i.d. Making use of the decomposition, $\tilde{\mathbf{E}} = \mathbf{V}_{\tilde{\mathbf{E}}}\mathbf{\Lambda}_{\tilde{\mathbf{E}}}\mathbf{V}_{\tilde{\mathbf{E}}}^H$, with the eigenvalues in $\mathbf{\Lambda}_{\tilde{\mathbf{E}}}$ ordered by decreasing value, we may write

$$\begin{aligned}
(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}})(\mathbf{I}_{n_r n_t} + \eta\mathbf{C}_\mathbf{R})^{-1} &= (\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}})(\mathbf{I}_{n_r n_t} + \eta(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}))^{-1} \\
&= \mathbf{I}_{n_r} \otimes \left[\mathbf{V}_{\tilde{\mathbf{E}}}\mathbf{\Lambda}_{\tilde{\mathbf{E}}}(\mathbf{I}_{n_r} + \eta\mathbf{\Lambda}_{\tilde{\mathbf{E}}})^{-1}\mathbf{V}_{\tilde{\mathbf{E}}}^H\right] \\
&\approx \eta^{-1}\left(\mathbf{I}_{n_r} \otimes \left[\mathbf{V}_{\tilde{\mathbf{E}}}\underline{\mathbf{\Lambda}}\mathbf{V}_{\tilde{\mathbf{E}}}^H\right]\right) \qquad (6.109)
\end{aligned}$$

with

$$\underline{\mathbf{\Lambda}} = \mathrm{diag}\{\underbrace{1,\ldots,1}_{r(\tilde{\mathbf{E}})},\underbrace{0,\ldots,0}_{n_t - r(\tilde{\mathbf{E}})}\} = \mathrm{diag}\left\{\mathbf{I}_{r(\tilde{\mathbf{E}})},\mathbf{0}_{n_t - r(\tilde{\mathbf{E}})}\right\}. \qquad (6.110)$$

The last equality of (6.109) uses the fact that, at high SNR, the non-zero eigenvalues of $\tilde{\mathbf{E}}$ are such that $\eta\lambda_i(\tilde{\mathbf{E}}) \gg 1$. Analogous to (6.105), we may further write

$$\mathrm{vec}(\bar{\mathbf{H}}^H)^H\left(\mathbf{I}_{n_r} \otimes \left[\mathbf{V}_{\tilde{\mathbf{E}}}\underline{\mathbf{\Lambda}}\mathbf{V}_{\tilde{\mathbf{E}}}^H\right]\right)\mathrm{vec}(\bar{\mathbf{H}}^H) = \mathrm{Tr}\left\{\bar{\mathbf{H}}^H\bar{\mathbf{H}}\mathbf{V}_{\tilde{\mathbf{E}}}\underline{\mathbf{\Lambda}}\mathbf{V}_{\tilde{\mathbf{E}}}^H\right\}$$

$$\geq \sum_{i=1}^{r(\tilde{\mathbf{E}})}\lambda_i(\bar{\mathbf{H}}^H\bar{\mathbf{H}}) \qquad (6.111)$$

with the eigenvalues of $\bar{\mathbf{H}}^H\bar{\mathbf{H}}$ in increasing order. Decomposing $\bar{\mathbf{H}}^H\bar{\mathbf{H}} = \mathbf{V}_{\bar{\mathbf{H}}}\mathbf{\Lambda}_{\bar{\mathbf{H}}}\mathbf{V}_{\bar{\mathbf{H}}}^H$, equality in (6.111) occurs if $\mathbf{V}_{\bar{\mathbf{H}}} = \mathbf{V}_{\tilde{\mathbf{E}}}$. Assuming that $\mathbf{R} = \mathbf{I}_{n_t n_r}$ finally yields the announced results. Note that if $\bar{\mathbf{H}}$ is (quasi) rank deficient with only one non-zero eigenvalue, (6.111) equals zero for rank-deficient codes. Note also that $\mathrm{Tr}\left\{\bar{\mathbf{H}}^H\bar{\mathbf{H}}\mathbf{V}_{\tilde{\mathbf{E}}}\underline{\mathbf{\Lambda}}\mathbf{V}_{\tilde{\mathbf{E}}}^H\right\}$ is always upper-bounded by $n_t n_r$. ∎

Ricean fast fading channels

Proposition 6.7 *In a perfectly time interleaved Ricean channel whose Rayleigh component is i.i.d., the average PEP reads as*

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l_{\mathbf{C},\mathbf{E}}} \prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{-2n_r} e^{-K \sum_{k \in \tau_{\mathbf{C},\mathbf{E}}} \left\| \bar{\mathbf{H}} \frac{\mathbf{c}_k - \mathbf{e}_k}{\|\mathbf{c}_k - \mathbf{e}_k\|} \right\|^2} d\beta \quad (6.112)$$

$$\leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l_{\mathbf{C},\mathbf{E}}} \prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{-2n_r} d\beta \quad (6.113)$$

The diversity gain is given by $n_r l_{\mathbf{C},\mathbf{E}}$, analogous to Rayleigh fast fading channels. Again, the coding gain is made of two terms:

- the first term is similar to gain obtained in i.i.d. Rayleigh channels and given by the product distance of the error matrix
- the second term accounts for the interactions between the channel and the coherent component via the projection of the principal eigenvector of $\mathbf{c}_k - \mathbf{e}_k$ onto the space spanned by the eigenvectors of $\bar{\mathbf{H}}^H \bar{\mathbf{H}}$.

In practical scenarios, if the principal eigenvector of $\mathbf{c}_k - \mathbf{e}_k$ is orthogonal to the principal eigenvector of $\bar{\mathbf{H}}^H \bar{\mathbf{H}}$, $\|\bar{\mathbf{H}} \frac{\mathbf{c}_k - \mathbf{e}_k}{\|\mathbf{c}_k - \mathbf{e}_k\|}\|^2 = 0$, and the code acts as if the effective length was reduced by one for its contribution to the argument of the exponential. Therefore, since the effective SNR is divided by $1 + K$, the error performance in Ricean fast fading channels is not necessarily better than in Rayleigh fast fading channels.

PROOF OF PROPOSITION 6.4:   Given our assumptions, $\mathbf{\Xi} = \mathbf{I}_{Tn_r n_t}$. Making the following decompositions $\mathbf{c}_k - \mathbf{e}_k = \mathbf{V}_k \mathbf{\Lambda}_k \mathbf{V}_k^H$ with $\mathbf{\Lambda}_k = \mathrm{diag}\{\|\mathbf{c}_k - \mathbf{e}_k\|^2, 0, \ldots, 0\}$ and $\mathbf{D} = \mathbf{V}_\mathbf{D} \mathbf{\Lambda}_\mathbf{D} \mathbf{V}_\mathbf{D}^H$ with $\mathbf{V}_\mathbf{D} = \mathrm{diag}\{\mathbf{V}_1, \ldots, \mathbf{V}_T\}$ and $\mathbf{\Lambda}_\mathbf{D} = \mathrm{diag}\{\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_T\}$ and reasoning similarly to (6.109), the average PEP (6.14) regime is approximated in the high SNR as

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r l_{\mathbf{C},\mathbf{E}}} \prod_{k \in \tau_{\mathbf{C},\mathbf{E}}} \|\mathbf{c}_k - \mathbf{e}_k\|^{-2n_r}$$

$$e^{-K \sum_{k \in \tau_{\mathbf{C},\mathbf{E}}} \mathrm{Tr}\left\{ \bar{\mathbf{H}}^H \bar{\mathbf{H}} \mathbf{V}_k \mathrm{diag}\{1,0,\ldots,0\} \mathbf{V}_k^H \right\}} d\beta \quad (6.114)$$

or equivalently as given by (6.112). ∎

6.4.2   Medium SNR regime

Ricean slow fading channels

For simplicity, we assume that the Rayleigh component is i.i.d. At medium SNR, $\eta \lambda_i(\tilde{\mathbf{E}}) < 1$ for $i = r(\tilde{\mathbf{E}}) - \delta + 1, \ldots, n_t$ with $\delta = \alpha/n_r$. Under those assumptions, the performance of a code is summarized in the following proposition.

Proposition 6.8 *In i.i.d. Ricean channels, the average PEP in the medium SNR regime reads as*

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r(r(\tilde{\mathbf{E}})-\delta)} \prod_{k=1}^{r(\tilde{\mathbf{E}})-\delta} \lambda_k^{-n_r}(\tilde{\mathbf{E}}) \, e^{-KTr\left\{\bar{\mathbf{H}}^H\bar{\mathbf{H}}\mathbf{V}_{\tilde{\mathbf{E}}}\underline{\boldsymbol{\Lambda}}'\mathbf{V}_{\tilde{\mathbf{E}}}^H\right\}} d\beta \qquad (6.115)$$

$$\leq \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r(r(\tilde{\mathbf{E}})-\delta)} \prod_{k=1}^{r(\tilde{\mathbf{E}})-\delta} \lambda_k^{-n_r}(\tilde{\mathbf{E}}) \, e^{-K\eta\lambda_{max}(\bar{\mathbf{H}}^H\bar{\mathbf{H}})\lambda_{\min}(\tilde{\mathbf{E}})} d\beta \qquad (6.116)$$

where $\underline{\boldsymbol{\Lambda}}' = \text{diag}\{\mathbf{I}_{r(\tilde{\mathbf{E}})-\delta}, \eta\lambda_{r(\tilde{\mathbf{E}})-\delta+1}(\tilde{\mathbf{E}}), \ldots, \eta\lambda_{n_t}(\tilde{\mathbf{E}})\}$, *and the upper bound is obtained assuming that* $\bar{\mathbf{H}}$ *is of rank one.*

We first note that the diversity gain is reduced by α. Then, the coding gain of any code is a function of a product of eigenvalues of the error matrix and of a term accounting for the interactions between the error matrix and the coherent component via the projection of the error matrix onto the space spanned by the coherent component. Equation (6.115) also suggests to maximize the minimum eigenvalue of the error matrices when designing a code. Yet, even two codes with the same set of eigenvalues may perform differently, because the eigenvectors of the error matrix also play a significant role. When $\mathbf{R} \neq \mathbf{I}_{n_t n_r}$, the derivation of the average PEP in the medium SNR regime is complicated and not dealt with in this book.

PROOF OF PROPOSITION 6.8: From (6.109), we may write in the medium SNR regime that

$$\left(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}\right)\left(\mathbf{I}_{n_r n_t} + \eta\mathbf{C}_{\mathbf{R}}\right)^{-1} \approx \eta^{-1}\left(\mathbf{I}_{n_r} \otimes \left[\mathbf{V}_{\tilde{\mathbf{E}}}\underline{\boldsymbol{\Lambda}}'\mathbf{V}_{\tilde{\mathbf{E}}}^H\right]\right) \qquad (6.117)$$

where

$$\underline{\boldsymbol{\Lambda}}' = \text{diag}\{\underbrace{1, \ldots, 1}_{r(\tilde{\mathbf{E}})-\delta}, \eta\lambda_{r(\tilde{\mathbf{E}})-\delta+1}(\tilde{\mathbf{E}}), \ldots, \eta\lambda_{n_t}(\tilde{\mathbf{E}})\}. \qquad (6.118)$$

$\mathbf{V}_{\tilde{\mathbf{E}}}\underline{\boldsymbol{\Lambda}}'\mathbf{V}_{\tilde{\mathbf{E}}}^H$ can be though of as an equivalent error matrix at medium SNR where the dominant eigenvalues are equal to 1 and the smallest eigenvalues are a function of the smallest eigenvalues of $\tilde{\mathbf{E}}$. The average PEP for an identity spatial correlation matrix at medium SNR is thus expressed as (6.115).

A lower bound on the performance can be obtained similarly to (6.111) as

$$\text{Tr}\left\{\bar{\mathbf{H}}^H\bar{\mathbf{H}}\mathbf{V}_{\tilde{\mathbf{E}}}\underline{\boldsymbol{\Lambda}}'\mathbf{V}_{\tilde{\mathbf{E}}}^H\right\} \geq \sum_{i=1}^{r(\tilde{\mathbf{E}})-\delta} \lambda_i(\bar{\mathbf{H}}^H\bar{\mathbf{H}}) + \eta \sum_{i=r(\tilde{\mathbf{E}})-\delta+1}^{n_t} \lambda_i(\bar{\mathbf{H}}^H\bar{\mathbf{H}})\lambda_i(\tilde{\mathbf{E}})$$

$$\geq \eta\lambda_{max}(\bar{\mathbf{H}}^H\bar{\mathbf{H}})\lambda_{min}(\tilde{\mathbf{E}}) \qquad (6.119)$$

where $\lambda_i\left(\bar{\mathbf{H}}^H\bar{\mathbf{H}}\right)$ are ordered in increasing order and $\lambda_i(\tilde{\mathbf{E}})$ in decreasing order, and (6.119) results from the assumption that $\bar{\mathbf{H}}$ is rank deficient with only one dominant eigenvalue, $\lambda_{max}\left(\bar{\mathbf{H}}^H\bar{\mathbf{H}}\right)$. ∎

Ricean fast fading channels

Assuming that the spatio-temporal correlation matrix $\boldsymbol{\Xi}$ is equal to $\mathbf{I}_{Tn_t n_r}$, the PEP in the medium SNR regime is similar to (6.112), except that $l_{\mathbf{C},\mathbf{E}}$ and $\tau_{\mathbf{C},\mathbf{E}}$ are replaced by $l'_{\mathbf{C},\mathbf{E}}$ and $\tau'_{\mathbf{C},\mathbf{E}}$ respectively, since δ positions $k \in \tau_{\mathbf{C},\mathbf{E}}$ are such that $\eta\|\mathbf{c}_k - \mathbf{e}_k\|^2 << 1$ at medium SNR. Hence, the PEP at medium SNR in perfect time interleaved i.i.d. Ricean fast fading channels is also predicted by Proposition 6.7.

6.4.3 Low SNR regime

Proposition 6.9   *In Ricean fading channels at low SNR, the average PEP behaves as*

$$P(\mathbf{C} \to \mathbf{E}) \approx \mathcal{Q}\left(\sqrt{\frac{\rho}{2}\frac{K}{1+K}\|\bar{\mathbf{H}}(\mathbf{C}-\mathbf{E})\|_F^2}\right) \tag{6.120}$$

$$\leq \mathcal{Q}\left(\sqrt{\frac{\rho}{2}\frac{K}{1+K}\lambda_{max}\left(\bar{\mathbf{H}}^H\bar{\mathbf{H}}\right)\lambda_{\min}(\tilde{\mathbf{E}})}\right). \tag{6.121}$$

Naturally, no diversity is achieved and the performance is exclusively a function of the projection of the error matrix onto the space spanned by the coherent component. The joint spatio-temporal correlation matrix does not have any impact on the PEP. Actually, the channel acts as an AWGN SISO channel where the performance is dictated by the Euclidean distance between codewords $\bar{\mathbf{H}}\mathbf{C}$ as seen by the receiver.

PROOF OF PROPOSITION 6.8:   As the Ricean K-factor increases, it is getting less realistic to assume a high effective SNR. When $\eta\lambda_i(\boldsymbol{\Xi}\boldsymbol{\Delta})$ is smaller than 1, we evolve into the low SNR regime and the PEP becomes

$$P(\mathbf{C} \to \mathbf{E}) \approx \frac{1}{\pi}\int_0^{\pi/2} \exp\left(-\eta K D^2\right) d\beta$$

$$= \mathcal{Q}\left(\sqrt{\frac{\rho}{2}\frac{K}{1+K}D^2}\right) \tag{6.122}$$

with

$$D^2 = \text{vec}\left(\left(\mathbf{1}_{1\times T} \otimes \bar{\mathbf{H}}\right)^H\right)^H \boldsymbol{\Delta}\,\text{vec}\left(\left(\mathbf{1}_{1\times T} \otimes \bar{\mathbf{H}}\right)^H\right)$$

$$= \sum_{k=0}^{T-1}\left\|\bar{\mathbf{H}}\left(\mathbf{c}_k - \mathbf{e}_k\right)\right\|_F^2 \tag{6.123}$$

$$\geq \sum_{i=1}^{n_t} \lambda_i(\tilde{\mathbf{E}})\lambda_i\left(\bar{\mathbf{H}}^H\bar{\mathbf{H}}\right) \tag{6.124}$$

$$\approx n_t n_r \lambda_{\min}(\tilde{\mathbf{E}}) \tag{6.125}$$

where in (6.124), $\lambda_i\left(\bar{\mathbf{H}}^H\bar{\mathbf{H}}\right)$ are ordered in increasing order of magnitude and $\lambda_i(\tilde{\mathbf{E}})$ in decreasing order. Equality occurs if $\mathbf{V}_{\bar{\mathbf{H}}} = \mathbf{V}_{\tilde{\mathbf{E}}}$. Equation (6.125) assumes that $\bar{\mathbf{H}}^H\bar{\mathbf{H}}$ has only one non-zero eigenvalue equal to $n_t n_r$. In this case, equality in (6.125) is simply achieved if the eigenvector of $\tilde{\mathbf{E}}$ corresponding to the smallest eigenvalue $\lambda_{\min}(\tilde{\mathbf{E}})$ lies in the space spanned by the eigenvector of $\bar{\mathbf{H}}^H\bar{\mathbf{H}}$ corresponding to the unique eigenvalue $n_t n_r$ [Gam02]. ∎

The performance is sensibly different from the high-SNR result. The code design should ensure that the lower bound (6.124) is not achieved, i.e. that the main eigenvectors of $\tilde{\mathbf{E}}$ do not lie in the space spanned by the eigenvectors of $\bar{\mathbf{H}}^H\bar{\mathbf{H}}$ corresponding to the zero eigenvalues. In any case, a code with large minimum determinant and thus with large $\lambda_{\min}(\tilde{\mathbf{E}})$ should perform well. But that is too conservative in the sense that $n_t n_r \lambda_{\min}(\tilde{\mathbf{E}})$ is only a lower bound on D^2.

6.4.4 Summary and examples

In Ricean fading channels in the **high SNR regime**,

- the diversity gain is not affected by the dominant component and is the same as the one achieved on Rayleigh fading
- the effective SNR is decreased by $1 + K$
- the coding gain is made of two terms: one provided by the Rayleigh component and similar to the coding gain in pure Rayleigh fading, and another term provided by the coherent component:
  - for full-rank codes in slow fading channels, the term induced by the coherent component in the coding gain is independent of the code, hence designing a full-rank code for spatially correlated Ricean slow fading channels results in designing a code for i.i.d. Rayleigh slow fading channels
  - for rank-deficient codes or in fast fading channels, there are interactions between the coherent component and the code, so that the coding gain provided by the coherent component is related to the considered code through the projection of the error matrix onto the space spanned by the coherent component.

In Ricean fading channels in the **medium SNR regime**, the coding gain is affected by the presence of the coherent component through its projection on the error matrix, irrespective of the rank of the code.
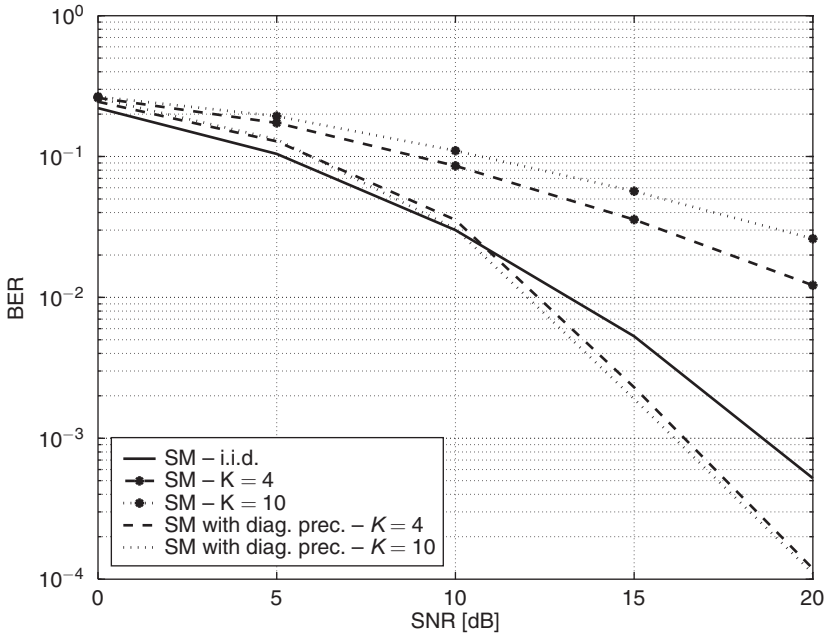
Figure 6.9   *Performance of rank-deficient Spatial Multiplexing schemes on* 2×2 *Ricean fading channels with* $K = 4$ *and* $K = 10$

In Ricean fading channels in the **low SNR regime**, the channel acts as an AWGN SISO channel where the performance is dictated by the Euclidean distance between codewords $\bar{\mathbf{H}}\mathbf{C}$ as seen by the receiver.

To illustrate these results, Figures 6.9 and 6.10 evaluate the impact of a coherent contribution on the performance of rank-deficient and full-rank codes, respectively. As rank-deficient codes, we use the classical SM and another scheme consisting of the classical SM combined with a unitary diagonal precoder designed in [HLHS03] (this precoder is discussed in greater detail in Chapter 7). The full-rank codes are the "Dayal" and the "Tilted-QAM" codes, presented in Section 5.5.7. Again, QPSK constellations and ML decoding are used. We assume the same array configurations as in our previous examples. However, the scattering around the mobile is now assumed to be very rich and a coherent LOS component with K-factor $K = 4$ and $K = 10$ is present along the link axis. For comparison, we also display the performance achieved in i.i.d. Rayleigh fading channels.

As highlighted in Section 6.4.1, the coherent component may be beneficial or detrimental to the performance, depending on the rank-deficient code considered. This is obvious in Figure 6.9 where the SM scheme with the diagonal precoder [HLHS03] beneficially exploits the Ricean fading while the classical SM is negatively affected by the coherent component. In such a configuration, its performance is worse in Ricean fading than in Rayleigh fading although the non-coherent (Rayleigh) component is close to i.i.d. thanks to the scattering richness.
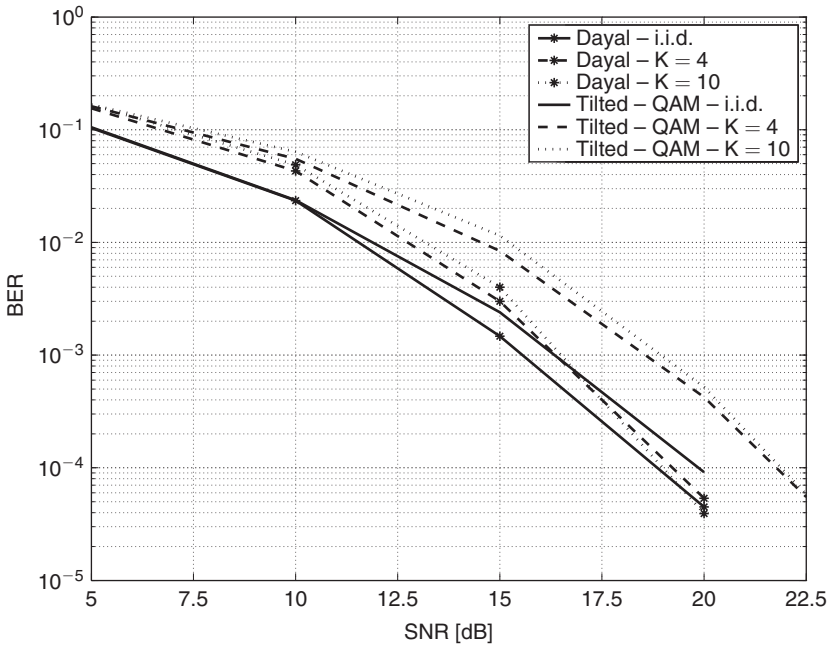
Figure 6.10 *Performance of approximately universal codes on 2×2 Ricean fading channels with $K = 4$ and $K = 10$*

At high SNR, theoretical derivations show that there is no interaction between the dominant component and the full-rank code. The coding gain offered by the coherent component is fixed and does not depend on the code but solely on the structure of the coherent component and the spatial correlation matrix. Moreover, in the absence of spatial correlation, the performance in Ricean channels should always be improved relative to Rayleigh fading channels. However, in Figure 6.10, the situation is different. First of all, we observe that, for a large range of SNR values, the coherent component is detrimental to the error rate, even at supposedly high SNR. Second, the performance in Ricean fading channels is not directly related to the performance in Rayleigh fading channels, as highlighted by (6.101), but the gap between both codes is much higher in Ricean fading than in Rayleigh fading. This shows that the high SNR regime is rarely reached in Ricean channels, and that the performance agrees more with Proposition 6.8.

6.5 Average pairwise error probability in dual-polarized channels

So far, we have analyzed the performance of Rayleigh and Ricean channels with identically distributed channel gains in mind (although the expressions are general). A typical counter-example is the case of dual-polarized schemes. Therefore, it seems interesting to analyze the impact of propagation conditions on the average PEP in 2×2 dual-polarized Rayleigh

and LOS Ricean fading channels, for which (6.25) and (6.120) can be used. Because dual-polarized channel models are complex, we simplify our study to two simple schemes:

- the Alamouti O-STBC
- the Spatial Multiplexing scheme.

When we consider Rayleigh slow fading channels with vertical and horizontal antennas at both ends, we assume that the antenna XPD is negligible ($\chi_a \cong 0$), so that

$$\tilde{\mathbf{H}}_{\times,a} = \tilde{\mathbf{H}}_\times = \tilde{\mathbf{X}} \odot \tilde{\mathbf{H}}, \tag{6.126}$$

where $\tilde{\mathbf{H}}$ is the uni-polarized channel matrix (i.e. for vertically polarized antennas at both ends), as proposed in Section 3.1.3, and

$$\text{vec}(\tilde{\mathbf{X}}^H) = \begin{bmatrix} 1 & 0 & 0 & \sqrt{\mu}\delta_1^* \\ 0 & \mu\chi & \sqrt{\mu}\chi\delta_2^* & 0 \\ 0 & \sqrt{\mu}\chi\delta_2 & \chi & 0 \\ \sqrt{\mu}\delta_1 & 0 & 0 & \mu \end{bmatrix} \text{vec}(\mathbf{X}_w^H). \tag{6.127}$$

Using the same models as those used when estimating the mutual information (see Section 4.5.1), we may write the dual-polarized correlation matrix as

$$\mathbf{R}_\times = \begin{bmatrix} 1 & 0 & 0 & s_1^*\sqrt{\mu} \\ 0 & \mu\chi & s_2^*\sqrt{\mu}\chi & 0 \\ 0 & s_2\sqrt{\mu}\chi & \chi & 0 \\ s_1\sqrt{\mu} & 0 & 0 & \mu \end{bmatrix} \tag{6.128}$$

if $\delta_1 = \delta_2 = 1$, and

$$\mathbf{R}_\times = \text{diag}\{1, \mu\chi, \chi, \mu\} \tag{6.129}$$

if $\delta_1 = \delta_2 = 0$. It is worth remembering that the uni-polarized correlation matrix is given by

$$\mathbf{R} = \begin{bmatrix} 1 & t^* & r^* & s_1^* \\ t & 1 & s_2^* & r^* \\ r & s_2 & 1 & t^* \\ s_1 & r & t & 1 \end{bmatrix}. \tag{6.130}$$

6.5.1   *Performance of orthogonal space–time block coding*

Restricting our analysis to the Alamouti scheme, the codeword matrix is given by

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{bmatrix} \tag{6.131}$$

where c_1 and c_2 are two symbols of a given constellation, and we denote the elements of the codeword error as $d_1 = (c_1 - e_1)/\sqrt{2}$ and $d_2 = (c_2 - e_2)/\sqrt{2}$.

Rayleigh fading channels

In Rayleigh slow fading channels, the matrix $\mathbf{C_R}$ is given by application of (6.16) as

$$\mathbf{C_R} = \mathbf{R}\left[|d_1|^2 + |d_2|^2\right] \tag{6.132}$$

with $\mathbf{R}$ being given by (6.128), (6.129) or (6.130). Therefore, the impact of the propagation channel on the error probability is directly given by the four eigenvalues of $\mathbf{R}$. Minimizing the average PEP is equivalent to maximizing

$$\det\left(\mathbf{I}_4 + \eta\mathbf{C_R}\right) = \prod_{i=1}^{r(\mathbf{C_R})} \left(1 + \eta\lambda_i\left(\mathbf{C_R}\right)\right)$$

or, using a Chernoff bound, to maximizing

$$\det\left(\mathbf{I}_4 + \frac{\rho}{4}\mathbf{C_R}\right) = \det\left(\mathbf{I}_4 + \rho_e\mathbf{R}\right)$$

where we define for better legibility an effective SNR as $\rho_e = \rho[|d_0|^2 + |d_1|^2]/4$. Dual-polarization is then preferred when

$$\det\left(\mathbf{I}_4 + \rho_e\mathbf{R}_\times\right) > \det\left(\mathbf{I}_4 + \rho_e\mathbf{R}\right). \tag{6.133}$$

In the high SNR regime, the condition simply becomes

$$\det\left(\mathbf{R}_\times\right) > \det\left(\mathbf{R}\right). \tag{6.134}$$

Note that the high SNR regime might only be reached for unrealistically high SNR when χ and/or μ are small, i.e. when some eigenvalues of $\mathbf{R}_\times$ are small. The above conditions are easily expressed in terms of the correlation coefficients, as well as μ and χ. As an example, let us consider a Kronecker-structured channel uncorrelated at the receiver (hence, $r = s_1 = s_2 = 0$). When $\mu = 1$ (i.e. when vertically and horizontally-polarized waves are equally attenuated), the above conditions are in turn rewritten as

$$\left(1 + \rho_e\right)^2\left(1 + \chi\rho_e\right)^2 > \left(1 + \rho_e\right)^4 + \rho_e^2|t|^2\left[\rho_e^2|t|^2 - 2\left(1 + \rho_e\right)^2\right] \tag{6.135}$$

and

$$|t| > \sqrt{1 - \chi^2}. \tag{6.136}$$

A similar condition on $|r|$ would be found for Kronecker-structured channels uncorrelated at the transmitter. This indicates that dual-polarization can offer increased performance at high spatial transmit/receive correlation levels despite the reduction of the average energy of the channel caused by χ. Naturally, the higher χ, the lower the transmit/receive correlation for which dual-polarized schemes outperform uni-polarized schemes. Note also that transmit and receive correlations equally affect the performance of O-STBCs.

It might also seem sensible to use the effective diversity order N_{div} as a natural metric to decide whether dual-polarized systems should be preferred to uni-polarized systems.

Applying directly the definition of (3.43), we find that the condition, under which dual-polarized transmissions (with the same polarization scheme at both Tx and Rx) perform better than uni-polarized transmissions, is given

- if $\delta_1 = \delta_2 = 1$ by

$$\frac{\left(1 + \mu + \mu\chi + \chi\right)^2}{1 + \mu^2 + \mu^2\chi^2 + \chi^2 + 2\mu\left(|s_1|^2 + \chi^2|s_2|^2\right)} > \frac{4}{1 + |t|^2 + |r|^2 + \frac{|s_1|^2 + |s_2|^2}{2}} \quad (6.137)$$

- if $\delta_1 = \delta_2 = 0$ by

$$\frac{\left(1 + \mu + \mu\chi + \chi\right)^2}{1 + \mu^2 + \mu^2\chi^2 + \chi^2} > \frac{4}{1 + |t|^2 + |r|^2 + \frac{|s_1|^2 + |s_2|^2}{2}}. \quad (6.138)$$

If we consider the same example as above ($r = s_1 = s_2 = 0$ and $\mu = 1$), (6.137) and (6.138) are both rewritten as

$$|t| > \frac{1 - \chi}{1 + \chi}. \quad (6.139)$$

This condition is actually looser than (6.136) with respect to favoring dual-polarized transmissions, as it only implies that the number of degrees of freedom offered by dual-polarized channels be higher than that offered by uni-polarized channels. It does not imply that the error probability is lower. When $\frac{1-\chi}{1+\chi} < |t| \leq \sqrt{1 - \chi^2}$, dual-polarized schemes exhibit a larger diversity order, but are still penalized by the energy loss since $\mathrm{Tr}\{\mathbf{R}_\times\} < \mathrm{Tr}\{\mathbf{R}\}$.

Ricean fading channels

We will analyze this case assuming that the Ricean component is the line-of-sight, and that the K-factor is high (so that the Rayleigh component can be neglected). We will then consider the impact of antenna XPD at one side only. Then, the performance solely depends on the Frobenius norm of the coherent component. Naturally, this norm is always smaller for dual-polarized schemes (since $\chi_a \leq 1$). The same conclusion is reached if we consider the effective diversity order (neglecting the Rayleigh component): $N_{div} = 2K$ for uni-polarized transmissions, while $N_{div} = (2 + \chi_a)K/2$ for dual-polarized systems. Therefore, the use of dual-polarized transmissions is not recommended for diversity schemes in Ricean channels.

6.5.2   *Performance of Spatial Multiplexing*

For a simple uncoded Spatial Multiplexing scheme, we have

$$(\mathbf{C} - \mathbf{E})(\mathbf{C} - \mathbf{E})^H = \begin{bmatrix} |d_1^2| & d_1 d_2^* \\ d_1^* d_2 & |d_2|^2 \end{bmatrix}. \quad (6.140)$$

Rayleigh fading channels

Given (6.140), the rank of both $\mathbf{C_R}$ and $\mathbf{C_{R_\times}}$ is two, with their two non-zero eigenvalues $\lambda_{1,2}$ reading as

$$\lambda_{1,2} = \frac{a + d \pm \sqrt{(a - d)^2 + 4|b|^2}}{2} \tag{6.141}$$

with

- for uni-polarized schemes

$$a = |d_1|^2 + |d_2|^2 + 2\Re[td_1d_2^*] \tag{6.142}$$
$$b = r\left(|d_1|^2 + |d_2|^2\right) + s_1\, d_1d_2^* + s_2\, d_1^*d_2 \tag{6.143}$$
$$d = a \tag{6.144}$$

- for dual-polarized schemes with $\delta_1 = \delta_2 = 1$

$$a = |d_1|^2 + \mu\chi|d_2|^2 \tag{6.145}$$
$$b = \sqrt{\mu}\left(s_1\, d_1d_2^* + s_2\, \chi\, d_1^*d_2\right) \tag{6.146}$$
$$d = \chi|d_1|^2 + \mu|d_2|^2 \tag{6.147}$$

- for dual-polarized schemes with $\delta_1 = \delta_2 = 0$

$$a = |d_1|^2 + \mu\chi|d_2|^2 \tag{6.148}$$
$$b = 0 \tag{6.149}$$
$$d = \chi|d_1|^2 + \mu|d_2|^2. \tag{6.150}$$

An estimate of the average symbol error rate is finally obtained by weighting the average PEP over the different possible symbols, using the union bound of (5.4). Various simulation results are illustrated in Figure 6.11 for a QPSK modulation. Considering first uni-polarized systems, we observe that transmit correlation is more harmful than receive correlation, as already seen in [NBE+02]. As for the use of dual-polarized arrays, it is clear that these are only beneficial when the transmit and/or receive correlations are higher than a certain level. This level is directly related to the co-polar gain imbalance and the scattering XPD. The smaller μ and/or χ, the larger this correlation level. Naturally, when both μ and χ are equal to one, dual-polarized systems always perform better than uni-polarized systems, as the decorrelation effect is not hampered by a decrease of power. For μ and/or χ smaller than one, the required correlation increases rapidly. Consequently, dual-polarized Spatial Multiplexing schemes perform well in Rayleigh fading channels when the uni-polarized spatial correlations are high enough.

Ricean fading channels

In LOS Ricean fading channels, the error probability is obtained through (6.15). Simulation results are displayed in Figure 6.12 for $\chi = 0.1$, $\mu = 1$ and $\delta_1 = \delta_2 = 0$ at 10 dB SNR. We
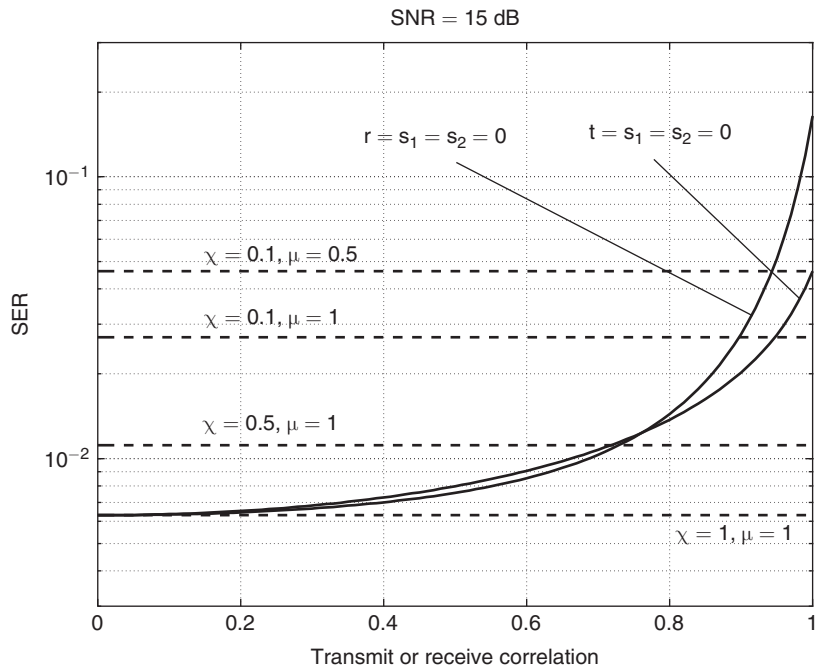
Figure 6.11   *Performance of QPSK Spatial Multiplexing for uni- and dual-polarized transmissions as a function of t or r*
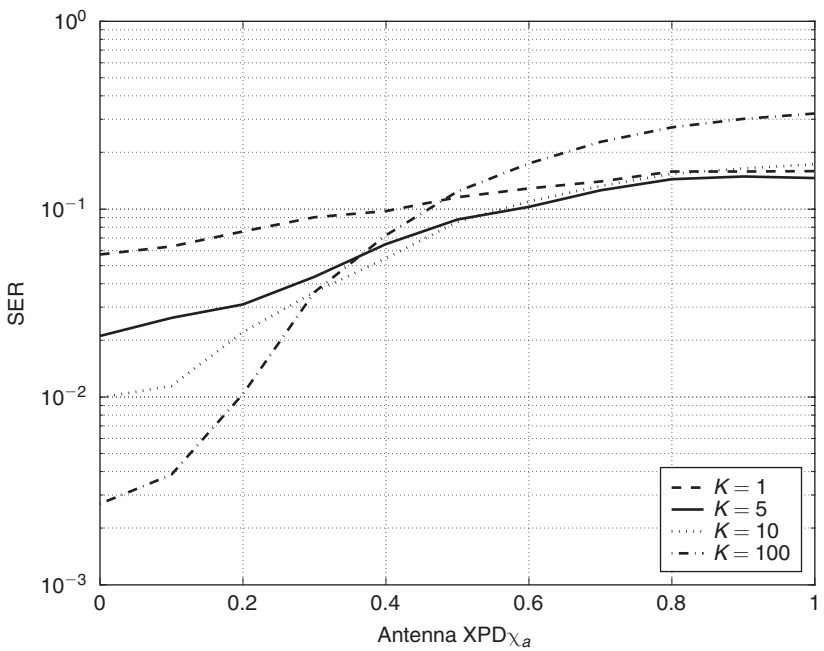
Figure 6.12   *Performance of QPSK Spatial Multiplexing in various Ricean channels for uni- and dual-polarized transmissions as a function of antenna XPD at 10 dB SNR*

observe that the use of dual-polarized arrays significantly improves the performance if antennas with cross-polar discrimination are used (the SER for uni-polarized links is given by $\chi_a = 1$, when K is high enough for the Rayleigh component to be neglected).

6.6 Perspectives on the space–time code design in realistic channels

Derivations obtained in this chapter provide some important rules for designing robust codes in correlated Ricean/Rayleigh channels. An important concept is the medium SNR regime, suggesting that a high-SNR assumption is not advised while predicting the performance of a code in correlated channels. In the medium SNR regime, interactions between the channel and the code highly affect the PEP, via the projection of the error matrix onto the space spanned by the spatio-temporal correlation matrix and/or the coherent component.

1. The diversity gain is decreased since the interactions induce the loss of several eigenvalues.
2. The coding gain counts an additional term, related to the mentioned interactions.

This suggests the following rule for designing efficient space–time codes in correlated Rayleigh/Ricean channels at useful SNR: choose a code offering a good performance in i.i.d. Rayleigh fading channels, which also accounts for the interaction between the propagation channel and the codewords. This is the goal of Chapter 7.

This page intentionally left blank

7

Space–time coding over real-world MIMO channels with no transmit channel knowledge

As highlighted in the previous chapters, the performance of space–time codes may be severely affected by the scattering richness, the array design or the presence of coherent components. Since a large variety of propagation conditions is likely to be encountered in practice, there is an obvious appeal to design codes that are robust with respect to all possible propagation conditions.

This is the objective of this chapter, assuming that the transmitter has no prior channel knowledge. In a sense, the criteria described hereafter may be thought of as being to correlated Rayleigh/Rician channels what the rank/determinant criteria of Chapter 5 are to i.i.d. Rayleigh fading channels. To this end, two design methodologies are covered. The first one takes on an information theory approach, while the second one yields design criteria motivated by a physical interpretation of the error probability.

7.1 Information theory motivated design methodology

Codes designed in Chapter 5 naturally perform very well in i.i.d. Rayleigh fading channels, but this might not be so in correlated Rayleigh or Rician channels. Hence, we shall derive a criterion to design codes in all possible slow fading channels, i.e. for all possible fading distributions.

The compound channel coding theorem [RV68] states that there exist codes with rate R bits/s/Hz that achieve reliable transmission over any slow fading channel realization which is not in outage, i.e. over any channel such that $\log_2 \det \left(\mathbf{I}_{n_t} + \frac{\rho}{n_t} \mathbf{H} \mathbf{H}^H \right) > R$. Those are the so-called universal codes. They achieve the outage capacity but require to code over infinite block lengths.

In Chapter 4, we have expressed that there is a fundamental trade-off between error probability and rate. In the high SNR regime, this trade-off has been relaxed and written in

terms of the SNR exponent of the error probability vs. the rate of transmission for an increasing SNR, i.e. the diversity-multiplexing trade-off. Chapter 5 detailed examples of codes that satisfy this trade-off for i.i.d. Rayleigh fading channels. Here, we may also express a trade-off between the error probability and the rate. However, with universal coding, we move one step further as we achieve the error probability vs. rate trade-off not only in i.i.d. Rayleigh fading channels (analogous to Chapters 4 and 5) but for all possible channel statistics. At high SNR, this trade-off between error probability and rate may also be written more easily as a trade-off between diversity and rate or diversity and multiplexing. Because it is much easier to obtain design criteria at high SNR, [TV06] introduces the notion of approximately universal property as an asymptotic (high-SNR) vision of the universality property. Being approximately universal is not sufficient for achieving outage capacity but is sufficient to achieve the diversity-multiplexing trade-off under all channel statistics. Moreover, as in the i.i.d. case, the trade-off is achievable with short block codes while outage capacity is only achievable with infinite block lengths. This approach has been originally investigated by [KW01, KW03] and recently extended in [TV06].

We must design codes that perform reliably when the channel is not in outage, but the channel fading distribution is not known at the transmitter. Hence, a sensible approach is to use the conditional PEP and find which channel realization, not in outage, induces the worst-case conditional PEP. By contrast to Chapter 5, averaging over a given fading distribution is not performed since the channel matrices are arbitrary. Let us recall that the conditional PEP in slow fading channels reads as

$$P(\mathbf{C} \to \mathbf{E}|\mathbf{H}) = \mathcal{Q}\left(\sqrt{\frac{\rho}{2}D^2}\right) \tag{7.1}$$

with $D^2 = \|\mathbf{H}(\mathbf{C}-\mathbf{E})\|_F^2$. Assuming an identity input covariance matrix, the set of channels not in outage is defined as

$$\mathcal{N}_{\mathcal{O}} = \left\{\mathbf{H} : \log_2 \det\left[\mathbf{I}_{n_t} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H\right] \geq R\right\}. \tag{7.2}$$

The philosophy behind the design criterion is thus outlined as follows.

Choose the code $\mathcal{C}$ that minimizes the worst-case conditional PEP over the set of channels that are not in outage $\mathcal{N}_{\mathcal{O}}$

$$\mathcal{C} = \arg\max \min_{\mathbf{H}\in\mathcal{N}_{\mathcal{O}}} D^2. \tag{7.3}$$

What conditions should D satisfy in order to be approximately universal and achieve the diversity-multiplexing trade-off over all channels not in outage? Reasoning as in the proof of Proposition 5.1, we know that the error probability is upper-bounded as

$$P_e(\rho) \leq P_{out}(R) + P(\text{error, no outage}). \tag{7.4}$$

To achieve the trade-off, we want the dominant errors to be caused by outage events. This requires [TV04] the second term of (7.4) to decay exponentially with the SNR ($= e^{-\rho^{\delta}}$ with $\delta > 0$) for all rates. To this end, the minimum distance between codewords must remain larger than the noise, so that codewords remain far apart from each other with regard to the noise power. This should hold true even if more constellations points are added when the SNR is increased. To better understand this concept, let us consider the following example [TV06].

Example 7.1   *We have seen in Chapter 5 that QAM constellations achieve the diversity-multiplexing trade-off in scalar Rayleigh slow fading channels. Let us now check that QAM constellations are also approximately universal for scalar slow fading channels, i.e. they achieve the diversity-multiplexing trade-off irrespective of the channel distribution. We may write the set of channel realizations not in outage as*

$$\mathcal{N_O} = \left\{ h : \log_2 \left(1 + \rho |h|^2 \right) \geq R \right\} \tag{7.5}$$

$$= \left\{ h : |h|^2 \geq \frac{2^R - 1}{\rho} \right\}. \tag{7.6}$$

The worst-case conditional PEP over all channel realizations not in outage is thus given by

$$P(c \rightarrow e|h) = \min_{h \in \mathcal{N_O}} \mathcal{Q} \left(\sqrt{\frac{\rho}{2} |h|^2 d^2} \right) \tag{7.7}$$

$$= \mathcal{Q} \left(\sqrt{\frac{2^R - 1}{2} d^2} \right) \tag{7.8}$$

with $d^2 = |c - e|^2$. If $2^R d^2 > 1$, the separation between constellation points is larger than the noise, and an error rarely occurs when the channel is not in outage. Since for QAM constellations, $d^2_{min} \approx 1/2^R$, these are approximately universal for scalar Rayleigh fading channels.

7.2   Information theory motivated code design in slow fading channels

7.2.1   *Universal code design criteria*

In order to obtain the general design criterion, we first evaluate the worst-case conditional PEP over all channels not in outage, which finds the channel realization **H** that minimizes

$$P(\mathbf{C} \rightarrow \mathbf{E}|\mathbf{H}) = \mathcal{Q} \left(\sqrt{\frac{\rho}{2} ||\mathbf{H}(\mathbf{C} - \mathbf{E})||_F^2} \right) \tag{7.9}$$

under the constraint that $\mathbf{H}$ is not in outage

$$\sum_{k=1}^{n} \log_2 \left(1 + \frac{\rho}{n_t} \lambda_k \left(\mathbf{H}^H \mathbf{H} \right) \right) \geq R. \tag{7.10}$$

where $\lambda_k \left(\mathbf{H}^H \mathbf{H} \right)$ with $k = 1, \ldots, n = \min \{n_t, n_r\}$ are the non-zero eigenvalues of $\mathbf{H}^H \mathbf{H}$. Clearly, the above constraint is only a function of the eigenvalues of the channel matrix, and not of the eigenvectors. However, as already explained in Chapter 6, the projection of $\mathbf{H}$ onto the error matrix $\mathbf{C} - \mathbf{E}$ significantly affects the squared distance D^2. Recalling that $\tilde{\mathbf{E}} = (\mathbf{C} - \mathbf{E})(\mathbf{C} - \mathbf{E})^H$, we use the following EVD decompositions

$$\tilde{\mathbf{E}} = \mathbf{U}_{\tilde{\mathbf{E}}} \mathbf{\Lambda}_{\tilde{\mathbf{E}}} \mathbf{U}_{\tilde{\mathbf{E}}}^H,$$

$$\mathbf{H}^H \mathbf{H} = \mathbf{U}_{\mathbf{H}} \mathbf{\Lambda}_{\mathbf{H}} \mathbf{U}_{\mathbf{H}}^H,$$

$$\mathbf{U}_{\mathbf{Q}} = \mathbf{U}_{\mathbf{H}}^H \mathbf{U}_{\tilde{\mathbf{E}}} \tag{7.11}$$

where $\mathbf{\Lambda}_{\tilde{\mathbf{E}}} = \mathrm{diag}\{\lambda_1(\tilde{\mathbf{E}}), \ldots, \lambda_{n_t}(\tilde{\mathbf{E}})\}$ contains the eigenvalues arranged by increasing order while $\mathbf{\Lambda}_{\mathbf{H}} = \mathrm{diag}\{\lambda_1(\mathbf{H}^H \mathbf{H}), \ldots, \lambda_{n_t}(\mathbf{H}^H \mathbf{H})\}$ contains the eigenvalues arranged by decreasing order.

The following proposition provides the channel matrix eigenvectors $\mathbf{U}_{\mathbf{H}}$ leading to the least favorable channel [KW01, KW03].

Proposition 7.1 *The squared distance D^2 can always be lower-bounded as*

$$D^2 = \|\mathbf{H}(\mathbf{C} - \mathbf{E})\|_F^2 \geq \sum_{k=1}^{n} \lambda_k(\tilde{\mathbf{E}}) \lambda_k(\mathbf{H}^H \mathbf{H}). \tag{7.12}$$

The lower bound is achieved if and only if $\mathbf{U}_{\mathbf{H}} = \mathbf{U}_{\tilde{\mathbf{E}}}$, i.e. when the kth eigenvector of $\mathbf{H}^H \mathbf{H}$ corresponding to the kth smallest eigenvalue lies in the space spanned by the kth eigenvector of $\tilde{\mathbf{E}}$ corresponding to the kth largest eigenvalue, and this for all $k = 1, \ldots, n$.

PROOF: Writing the squared distance D^2 as

$$D^2 = \|\mathbf{H}(\mathbf{C} - \mathbf{E})\|_F^2 = \mathrm{Tr}\{\tilde{\mathbf{E}} \mathbf{H}^H \mathbf{H}\} = \mathrm{Tr}\{\mathbf{U}_{\mathbf{Q}} \mathbf{\Lambda}_{\tilde{\mathbf{E}}} \mathbf{U}_{\mathbf{Q}}^H \mathbf{\Lambda}_{\mathbf{H}}\}$$

$$= \sum_{k=1}^{n_t} \sum_{l=1}^{n} \lambda_k(\tilde{\mathbf{E}}) \lambda_l(\mathbf{H}^H \mathbf{H}) |\mathbf{U}_{\mathbf{Q}}(l, k)|^2$$

$$\geq \sum_{k=1}^{n} \lambda_k(\tilde{\mathbf{E}}) \lambda_k(\mathbf{H}^H \mathbf{H}), \tag{7.13}$$

it is evident that equality occurs when $\mathbf{U}_{\mathbf{Q}} = \mathbf{I}_{n_t}$, i.e. when $\mathbf{U}_{\mathbf{H}} = \mathbf{U}_{\tilde{\mathbf{E}}}$. ∎

Still, we need to find the eigenvalues $\lambda_k \left(\mathbf{H}^H \mathbf{H}\right)$ minimizing $\frac{\rho}{2} \sum_{k=1}^{n} \lambda_k \left(\tilde{\mathbf{E}}\right) \lambda_k \left(\mathbf{H}^H \mathbf{H}\right)$ subject to the constraint (7.10). A Lagrangian optimization actually leads to the standard waterfilling solution

$$\lambda_k \left(\mathbf{H}^H \mathbf{H}\right) = \frac{1}{\rho} \left(\frac{1}{\mu \lambda_k \left(\tilde{\mathbf{E}}\right)} - n_t \right)^+ \tag{7.14}$$

where $(x)^+$ stands for $\max\{x, 0\}$ and μ is the Lagrange multiplier chosen such that

$$\sum_{k=1}^{n} \log_2 \left(1 + \left(\frac{1}{n_t \mu \lambda_k \left(\tilde{\mathbf{E}}\right)} - 1 \right)^+ \right) = R \tag{7.15}$$

or equivalently

$$\sum_{k=1}^{n} \left[\log_2 \left(\frac{1}{n_t \mu \lambda_k \left(\tilde{\mathbf{E}}\right)} \right) \right]^+ = R. \tag{7.16}$$

The worst-case conditional PEP is thus given by

$$\min_{\mathbf{H} \in \mathcal{N}_{\mathcal{O}}} P \left(\mathbf{C} \to \mathbf{E} | \mathbf{H}\right) = \mathcal{Q} \left(\sqrt{\frac{1}{2} \sum_{k=1}^{n} \left(\frac{1}{\mu} - n_t \lambda_k \left(\tilde{\mathbf{E}}\right) \right)^+} \right) \tag{7.17}$$

suggesting that codes should be designed according to the universal code design criterion.

Design criterion 7.1 *(Universal code design criterion for MIMO channels) Choose a code $\mathcal{C}$ that maximizes*

$$\sum_{k=1}^{n} \left(\frac{1}{\mu} - n_t \lambda_k \left(\tilde{\mathbf{E}}\right) \right)^+ . \tag{7.18}$$

The implementation of this criterion is quite complex. Therefore, it has been suggested [TV06] to focus on the high SNR regime in order to relax the previous design criterion.

Since $(x)^+$ is always larger than zero, a lower bound of (7.18) is obtained by suppressing the $()^+$ operator in (7.18) and (7.16) so that the Lagrange multiplier μ can be expressed as

$$\mu = 2^{-\frac{R}{n}} \frac{1}{n_t} \left(\prod_{k=1}^{n} \lambda_k^{-1} \left(\tilde{\mathbf{E}}\right) \right)^{\frac{1}{n}} \tag{7.19}$$

and

$$\sum_{k=1}^{n} \left(\frac{1}{\mu} - n_t \lambda_k \left(\tilde{\mathbf{E}}\right) \right)^+ \geq n_t n 2^{\frac{R}{n}} \left(\prod_{k=1}^{n} \lambda_k \left(\tilde{\mathbf{E}}\right) \right)^{\frac{1}{n}} - n_t \sum_{k=1}^{n} \lambda_k \left(\tilde{\mathbf{E}}\right). \tag{7.20}$$

Equality occurs when the rate in each subchannel is large. As the SNR increases, so does the rate, and the first term of the right-hand side of (7.20) becomes much larger than the second term.

Proposition 7.2   *A code is approximately universal if and only if it satisfies the approximately universal code design criterion.*

Design criterion 7.2   *(Approximately universal code design criterion for MIMO channels) Design the codebook $\mathcal{C}$ whose rate scales with the SNR in such a way that*

$$\min_{\substack{\mathbf{C},\mathbf{E}\in\mathcal{C}\\\mathbf{C}\neq\mathbf{E}}} 2^R \prod_{k=1}^{n} \lambda_k(\tilde{\mathbf{E}}) \dot{\geq} 1 \tag{7.21}$$

or equivalently

$$\min_{\substack{\mathbf{C},\mathbf{E}\in\mathcal{C}\\\mathbf{C}\neq\mathbf{E}}} \prod_{k=1}^{n} \lambda_k(\tilde{\mathbf{E}}) \dot{\geq} \rho^{-r} \tag{7.22}$$

where $\lambda_k(\tilde{\mathbf{E}})$, $k=1,\ldots,n$, are the n smallest eigenvalues of $\tilde{\mathbf{E}}$.

Proof: From (7.1), the proof consists in showing that P(error, no outage) decays exponentially with the SNR, so that the dominant errors are caused by outage events. When the channel is not in outage, this is guaranteed if the squared minimum distance between the codewords remains larger than the noise

$$\sum_{k=1}^{n} \left(\frac{1}{\mu} - n_t \lambda_k(\tilde{\mathbf{E}}) \right)^{+} \dot{\geq} 1 \tag{7.23}$$

and (7.21) simply results from the lower bound (7.20). A more rigorous proof is provided in [TV06].   ∎

For $n_r \geq n_t$, the approximately universal code design criterion (Criterion 7.2) is the same as the non-vanishing determinant criterion (Criterion 5.4). This is not surprising, since the proof of Proposition 5.2 already revealed that the non-vanishing determinant condition caused the distance between codewords to remain large compared to noise power. The minimum determinant has thus a broader meaning than the coding gain in i.i.d. Rayleigh slow fading channels. Furthermore, the results of Section 5.5.7 also yield the following property.

Proposition 7.3   *The Golden, tilted QAM and Dayal codes are approximately universal in slow fading MIMO channels.*

They achieve the optimal diversity-multiplexing trade-off at high SNR for every fading distribution, not only for the i.i.d. Rayleigh fading distribution.

We consider two particular interesting cases of MIMO channels: MISO channels and parallel channels.

7.2.2 MISO channels

For the particular case of MISO channels, $n_r = n = 1$ and the universal code design criterion (7.18) reduces to

Design criterion 7.3 *(Universal code design criterion for MISO channels) Choose the code $\mathcal{C}$ such that*

$$\mathcal{C} = \arg \max_{\substack{\mathbf{C}, \mathbf{E} \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{E}}} \min_{k=1,\ldots,n_t} \lambda_k(\tilde{\mathbf{E}}) \tag{7.24}$$

i.e. the code maximizing the smallest eigenvalue of $\tilde{\mathbf{E}} = (\mathbf{C} - \mathbf{E})(\mathbf{C} - \mathbf{E})^H$ evaluated over all pairs of codewords $\mathbf{C}$, $\mathbf{E}$ with $\mathbf{C} \neq \mathbf{E}$.

Proposition 7.4 *The Alamouti code with a QAM constellation is approximately universal in MISO channels.*

PROOF: For O-STBCs

$$\min_{\substack{\mathbf{C}, \mathbf{E} \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{E}}} \min_{k=1,\ldots,n_t} \lambda_k(\tilde{\mathbf{E}}) = \frac{T}{Qn_t} d_{min}^2$$

where d_{min} is the minimum Euclidean distance of the constellation. Since $d_{min}^2 \approx 1/2^{R/r_s}$ for QAM constellations

$$2^R \min_{\substack{\mathbf{C}, \mathbf{E} \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{E}}} \min_{k=1,\ldots,n_t} \lambda_k(\tilde{\mathbf{E}}) \overset{\cdot}{\geq} 1 \tag{7.25}$$

only if $r_s = 1$, which is satisfied by the Alamouti code. ∎

7.2.3 Parallel channels

Another particular class of channels consists in MIMO channels without interferences, also known as parallel channels. Interestingly, approximately universal codes in parallel channels can be used in D-BLAST schemes to obtain approximately universal codes in arbitrary MIMO channels. Considering one symbol period, we have on the $l^{\text{th}} = 1, \ldots, L$ subchannel

$$r^{(l)} = \sqrt{E_s} h^{(l)} c^{(l)} + n^{(l)}. \tag{7.26}$$

The conditional PEP is given by

$$P\left(\mathbf{C} \to \mathbf{E} | \{h^{(l)}\}_{l=1}^L\right) = \mathcal{Q}\left(\sqrt{\frac{\rho}{2} \sum_{l=1}^L |h^{(l)}|^2 |c^{(l)} - e^{(l)}|^2}\right) \tag{7.27}$$

while the set of channel realizations that are not in outage is expressed as

$$\sum_{l=1}^{L} \log_2 \left(1 + \frac{\rho}{L}|h^{(l)}|^2\right) \geq R. \tag{7.28}$$

The problem is similar to MIMO channels and the solution is again the standard waterfilling. The universal code design comes to maximize the following quantity

$$\sum_{l=1}^{L} \left(\frac{1}{L} - L|c^{(l)} - e^{(l)}|^2\right)^{+}. \tag{7.29}$$

In the high SNR regime, this criterion is simplified and the code is approximately universal if and only if

$$\min_{\substack{\mathbf{C},\mathbf{E} \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{E}}} 2^R \prod_{l=1}^{L} |c^{(l)} - e^{(l)}|^2 \overset{.}{\geq} 1. \tag{7.30}$$

In parallel channels, the approximately universal code design criterion reduces to the maximization of a minimum product distance.

Recently, [TV06] has proposed space-only codes that are approximately universal in parallel channels. Indeed, they achieve the optimal diversity-multiplexing trade-off at high SNR for all parallel channel distributions with block length $T = 1$. Those codes are also known as permutation codes. The basic idea is to send QAM symbols on each subchannel. Yet, instead of using repetition coding (that extracts the full diversity but lacks the multiplexing gain) and simply assigning the same symbol on each subchannel, permutation coding uses different mappings of the QAM points on each subchannel. The mappings are chosen so that if two codewords are spaced by the minimum distance in a QAM constellation on one subchannel, they are placed away from each other on the QAM constellation sent on the other subchannels, as shown by the following example.

Example 7.2   *Consider that $L = 2$. For a 16-QAM constellation, repetition coding consists in transmitting the same symbol on the two subchannels. This is illustrated in Figure 7.1 where the constellation sent on the two subchannels are displayed. A codeword is given by a pair of the same numbers. So if the pair (x, x) with $x = 1, \ldots, 16$ is transmitted, the same symbol is sent on the two subchannels. By contrast, Figure 7.2 illustrates the principle of permutation coding: two different symbols are sent when (x, x) is transmitted and they are sorted such that if x is close to y on the first constellation, x is far from y on the second constellation.*

For a total transmission rate scaling with SNR as $R = g_s \log_2(\rho)$, we have (assuming $L = 2$) that $d_{min}^2 = 1/2^R$ for a QAM constellation. Subsequently, the product distance (7.30) for repetition coding reads in this case as

$$\min_{\substack{\mathbf{C},\mathbf{E} \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{E}}} 2^R \prod_{l=1}^{2} |c^{(l)} - e^{(l)}|^2 = 2^R d_{min}^2 d_{min}^2 \approx \frac{1}{2^R} = \rho^{-g_s} \overset{.}{<} 1. \tag{7.31}$$

11	12	15	16		11	12	15	16
9	10	13	14		9	10	13	14
3	4	7	8		3	4	7	8
1	2	5	6		1	2	5	6

Figure 7.1   *Repetition coding for $L=2$ with rate $R=4$ bits/s/Hz*

11	12	15	16		7	3	8	4
9	10	13	14		15	11	16	12
3	4	7	8		5	1	6	2
1	2	5	6		13	9	14	10

Figure 7.2   *Permutation coding for $L=2$ with rate $R=4$ bits/s/Hz*

Clearly, repetition coding is not approximately universal in parallel channels, because a term ρ^{g_s} is missing in the product distance. This induces a large coding loss as the SNR increases. With permutation coding, the symbols mapped on the second constellation are separated by the length of a side of the QAM constellation $d_{max} \approx 1$. This compensates for the aforementioned ρ^{g_s} gap. The product distance with permutation is then written as

$$\min_{\substack{C,E \in \mathcal{C} \\ C \neq E}} 2^R \prod_{l=1}^{2} |c^{(l)} - e^{(l)}|^2 = 2^R d_{min}^2 d_{max}^2 \doteq 1. \tag{7.32}$$

That yields the following proposition.

Proposition 7.5   *Permutation codes are approximately universal in parallel channels.*

Permutation codes are particularly useful if they are used in conjunction with the D-BLAST transceiver architecture. Both V-BLAST and D-BLAST with MMSE-SIC receiver and ML decoder transform the MIMO channel into a parallel channel while still preserving the mutual information of the original MIMO channel, as highlighted by (5.126) in Section 5.5.2. However, unlike V-BLAST, D-BLAST also allows for coding across subchannels as outlined by (5.136).

Proposition 7.6   *Ignoring the initialization loss, using approximately universal codes (e.g. permutation codes) for each layer, the D-BLAST architecture with MMSE-SIC or ML decoding is approximately universal in MIMO channels with $n_r \geq n_t$.*

We now have a broader overview of the D-BLAST performance, as the optimality is not limited to i.i.d. Rayleigh fading channels: D-BLAST also achieves the optimal diversity-multiplexing trade-off over MIMO channels with arbitrary channel statistics with space-only codes.

The previous proposition can be further illustrated by verifying (7.21) for $n_r \geq n_t$, using (7.32). The error matrix of the D-BLAST scheme with 2 layers reads as

$$\mathbf{C} - \mathbf{E} = \begin{bmatrix} c_1^{(1)} - e_1^{(1)} & c_2^{(1)} - e_2^{(1)} & 0 \\ 0 & c_1^{(2)} - e_1^{(2)} & c_2^{(2)} - e_2^{(2)} \end{bmatrix}. \tag{7.33}$$

It is straightforward to note that with approximately universal codes for the parallel channel $2^R \det(\tilde{\mathbf{E}}) \overset{.}{\geq} 1$.

7.3 Error probability motivated design methodology

In Chapter 6, we have obtained the pairwise error probability averaged over the broad class of space–time correlated Ricean fading channels and discussed the impact of correlations and K-factor as a function of SNR. In this section, we use the so-derived expressions in order to extract practical design criteria:

- the design philosophy is briefly outlined in Section 7.3.1
- the interactions between the channel and the codewords are formalized in Section 7.3.2
- finally, Section 7.3.3 introduces the important notions of sum-wise catastrophic, product-wise catastrophic and robust codes and the main design criterion is obtained.

7.3.1 Designing robust codes

In Chapter 6, we have observed that the average PEP on correlated/Ricean fading channels generally reads as the product of two terms, one corresponding to i.i.d. Rayleigh fading and another that accounts for the interaction between the codewords and the non-idealities of the channel. This second term indicates that codes offering the same performance in i.i.d. Rayleigh channels may behave very differently if the channel becomes correlated or Ricean fading. Therefore, it is sensible to base a code design on choosing among the codes that perform optimally in i.i.d. Rayleigh fading channels those that also perform best in correlated Rayleigh/Ricean channels.

However, deriving such a code, performing well in correlated Rayleigh/Ricean channels, is a daunting task, as we should cover a very wide range of propagation conditions. To simplify the design, let us use the same rationale as in Section 6.1.1: the physical parameter representing the various scattering conditions and affecting the code performance is the transmit correlation, which is directly linked to the spatial distribution of the scatterers with regard to the transmit array or, in other words, to the direction power spectrum $\mathcal{A}_t(\mathbf{\Omega}_t)$. Therefore, designing a code for correlated Rayleigh/Ricean channels finds a code whose performance is roughly independent of $\mathcal{A}_t(\mathbf{\Omega}_t)$. As was highlighted in Chapter 6, the transmit correlation has a non-linear impact on the PEP. If the density of scatterers is sufficiently rich, the transmit correlation coefficient is small (typically, smaller than 0.6), and the performance is not significantly degraded. Hence, codes may be designed

assuming that the channel is i.i.d. Rayleigh distributed. However, if scatterers are only present in a few directions, the transmit correlation increases, and the error probability grows exponentially. The worst-case among those situations is the degenerate channel $(\mathcal{A}_t(\mathbf{\Omega}_t) \to \delta(\mathbf{\Omega}_t - \mathbf{\Omega}_{t,0}))$, where the scatterers as seen from the transmit array are located along a single direction $\mathbf{\Omega}_{t,0}$. In this scenario (assuming a 2-D propagation, reducing the direction to the azimuth θ_t), the error probability reads as

$$P\left(\mathbf{C} \to \mathbf{E}| \{\mathbf{H}_k\}_{k=0}^{T-1}\right) = \mathcal{Q}\left(\sqrt{\frac{\rho}{2}D_d^2}\right) \tag{7.34}$$

with $D_d^2 = \sum_{k=0}^{T-1} |(\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_{t,0})|^2 \|\mathbf{H}_k(1:n_r,1)\|^2$.

Naturally, $\theta_{t,0}$ is not known by the transmitter, hence the philosophy behind the PEP motivated design criterion may be expressed as follows.

> Within the space of codes providing a good performance in i.i.d. Rayleigh fading channels $\mathcal{C}_{i.i.d.}$, choose the code $\mathcal{C}$ that performs best in any channel contained in the set of degenerate channels $\mathcal{D}$
>
> $$\mathcal{C} = \arg \max_{\mathcal{C} \in \mathcal{C}_{i.i.d.}} \min_{\{\mathbf{H}_k\}_{k=0}^{T-1} \in \mathcal{D}} D_d^2. \tag{7.35}$$

We still need to define a measure of the performance in degenerate channels, which is the goal of the next sections [COV$^+$07 Cle05, CVJV05].

7.3.2 *Average pairwise error probability in degenerate channels*

The code performance in a degenerate channel may be formalized from the average PEP perspective developed in Section 6.2. To this end, we assume that the receive correlation matrix $\mathbf{R}_r$ is full rank, i.e. $r(\mathbf{R}_r) = n_r$, but that the transmit correlation matrix may tend toward rank deficiency.

Rayleigh fading channels
In Rayleigh fading channels ($K = 0$), if all scatterers are mainly located along one direction θ_{t_0}, the directional spread $\Lambda_t \to 0$, and the Rayleigh fading channel $\tilde{\mathbf{H}}$ becomes

$$\lim_{\Lambda_t \to 0} \text{vec}\left(\tilde{\mathbf{H}}^H\right) = \text{vec}\left(\tilde{\mathbf{H}}_{deg}^H\right) \otimes \mathbf{a}_t^*\left(\theta_{t_0}\right) \tag{7.36}$$

where $\tilde{\mathbf{H}}_{deg} = [\tilde{\mathbf{H}}_0(:,1) \ \tilde{\mathbf{H}}_1(:,1) \ \cdots \ \tilde{\mathbf{H}}_{T-1}(:,1)]$. The space–time correlation matrix $\mathbf{\Xi}$ reduces exactly to

$$\lim_{\Lambda_t \to 0} \mathbf{\Xi} = \mathbf{\Xi}_r \otimes \left(\mathbf{a}_t^*\left(\theta_{t_0}\right) \mathbf{a}_t^T\left(\theta_{t_0}\right)\right), \tag{7.37}$$

with $\mathbf{\Xi}_r$ the correlation matrix taking into account the joint receive-time correlation, given by

$$\mathbf{\Xi}_r = \mathcal{E} \left\{ \mathrm{vec} \left(\tilde{\mathbf{H}}^H_{deg} \right) \mathrm{vec} \left(\tilde{\mathbf{H}}^H_{deg} \right)^H \right\}. \tag{7.38}$$

Remember that the apparent Kronecker structure in (7.37) is exact in this low transmit angle spread configuration. Accordingly, the average PEP is a function of the direction of departure $\theta_{t,0}$. Indeed, since

$$\lim_{\Lambda_t \to 0} \left(\mathbf{I}_{n_r} \otimes \mathbf{D}^H \right) \mathbf{\Xi} \left(\mathbf{I}_{n_r} \otimes \mathbf{D} \right)$$
$$= \left(\mathbf{I}_{n_r} \otimes \left(\mathbf{I}_T \otimes \mathbf{a}^*_t(\theta_{t,0}) \right) \mathbf{D}^H \right) \mathbf{\Xi}_r \left(\mathbf{I}_{n_r} \otimes \left(\mathbf{I}_T \otimes \mathbf{a}^T_t(\theta_{t,0}) \right) \mathbf{D} \right), \tag{7.39}$$

the average PEP becomes

$$P\left(\mathbf{C} \to \mathbf{E}|\theta_t = \theta_{t,0}\right) = \frac{1}{\pi} \int_0^{\pi/2} \left(\det \left(\mathbf{I}_{Tn_r} + \eta \mathbf{\Xi}_r \left(\mathbf{I}_{n_r} \otimes \mathbf{D}_d \mathbf{D}^H_d \right) \right) \right)^{-1} d\beta$$
$$= \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{r(\mathbf{\Xi}_r(\mathbf{I}_{n_r} \otimes \mathbf{D}_d \mathbf{D}^H_d))} \left(1 + \eta \lambda_i \left(\mathbf{\Xi}_r \left(\mathbf{I}_{n_r} \otimes \mathbf{D}_d \mathbf{D}^H_d \right) \right) \right)^{-1} d\beta \tag{7.40}$$

where we define

$$\mathbf{D}_d = \left(\mathbf{I}_T \otimes \mathbf{a}^T_t(\theta_{t,0}) \right) \mathbf{D}$$
$$= \mathrm{diag} \left\{ (\mathbf{c}_0 - \mathbf{e}_0)^T \mathbf{a}_t(\theta_{t,0}), \ldots, (\mathbf{c}_{T-1} - \mathbf{e}_{T-1})^T \mathbf{a}_t(\theta_{t,0}) \right\}, \tag{7.41}$$

and $\eta = \rho / \left(4 \sin^2 \beta \right)$ denotes the effective SNR.

The impact of space–time correlation on the average PEP still depends on the SNR. In the low SNR regime, the determinant operator is approximated by the trace operator, hence

$$P\left(\mathbf{C} \to \mathbf{E}|\theta_t = \theta_{t,0}\right) \approx \frac{1}{\pi} \int_0^{\pi/2} \left(1 + \eta \mathrm{Tr}\{\mathbf{\Xi}_r \left(\mathbf{I}_{n_r} \otimes \mathbf{D}_d \mathbf{D}^H_d \right)\} \right)^{-1} d\beta$$
$$\approx \frac{1}{\pi} \int_0^{\pi/2} \left(1 + \eta n_r \left[\sum_{k=0}^{T-1} \left| (\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_{t,0}) \right|^2 \right] \right)^{-1} d\beta. \tag{7.42}$$

The receive-time correlation matrix $\mathbf{\Xi}_r$ does not affect the PEP, while the degenerate channel (i.e. the high transmit correlation) induces constructive or destructive interactions between the channel and the codewords through the term $\sum_{k=0}^{T-1} |(\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_{t,0})|^2$. Depending on the code, it may happen that

$$\sum_{k=0}^{T-1} \left| (\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_{t,0}) \right|^2 \geq \|\mathbf{C} - \mathbf{E}\|^2_F, \tag{7.43}$$

leading to a better PEP in transmit correlated Rayleigh fading channels than in i.i.d. Rayleigh fading channels.

As the SNR increases, the observed diversity on degenerate channels depends on the temporal correlation.

- For large temporal correlations owing to poor interleaving or slow fading conditions, $\mathbf{\Xi}_r$ reads as $\mathbf{R}_r \otimes \mathbf{1}_{T \times T}$, and

$$\lim_{\substack{\mathbf{\Xi}_r \to \\ \mathbf{R}_r \otimes \mathbf{1}_{T \times T}}} r\left(\mathbf{\Xi}_r \left(\mathbf{I}_{n_r} \otimes \mathbf{D}_d \mathbf{D}_d^H\right)\right) = n_r,$$

$$\lim_{\substack{\mathbf{\Xi}_r \to \\ \mathbf{R}_r \otimes \mathbf{1}_{T \times T}}} \left\{\lambda_i \left(\mathbf{\Xi}_r \left(\mathbf{I}_{n_r} \otimes \mathbf{D}_d \mathbf{D}_d^H\right)\right)\right\}_{i=1}^{n_r} = \left\{\lambda_i \left(\mathbf{R}_r\right) \sum_{k=0}^{T-1} \left|(\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_{t,0})\right|^2\right\}_{i=1}^{n_r},$$
(7.44)

$$\lim_{\substack{\mathbf{\Xi}_r \to \\ \mathbf{R}_r \otimes \mathbf{1}_{T \times T}}} \left\{\lambda_i \left(\mathbf{\Xi}_r \left(\mathbf{I}_{n_r} \otimes \mathbf{D}_d \mathbf{D}_d^H\right)\right)\right\}_{i=n_r+1}^{r\left(\mathbf{\Xi}_r\left(\mathbf{I}_{n_r} \otimes \mathbf{D}_d \mathbf{D}_d^H\right)\right)} = 0.$$

The PEP becomes

$$P\left(\mathbf{C} \to \mathbf{E} | \theta_t = \theta_{t,0}\right)$$

$$\approx \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{n_r} \left(1 + \eta \left[\sum_{k=0}^{T-1} \left|(\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_{t,0})\right|^2\right] \lambda_i(\mathbf{R}_r)\right)^{-1} d\beta \qquad (7.45)$$

The receive spatial correlation reduces the coding gain, while no transmit diversity is exploited by the code. Subsequently, the PEP is only determined by the coding gain, i.e., $\|(\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_{t,0})\|^2$. The total diversity achieved by the pair of codeword $\{\mathbf{C}, \mathbf{E}\}$ is equal to n_r at high SNR[1] if $\|(\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_{t,0})\|^2 \neq 0$ and equal to 0 otherwise.

- For low temporal correlations (the channel is fast varying or the interleaving depth is very long), $r(\mathbf{\Xi}_r) = n_r T$ and the total diversity achieved at high SNR[2] by the pair of codewords $\{\mathbf{C}, \mathbf{E}\}$ is given by $r\left(\mathbf{\Xi}_r \left(\mathbf{I}_{n_r} \otimes \mathbf{D}_d \mathbf{D}_d^H\right)\right) = r\left(\mathbf{D}_d \mathbf{D}_d^H\right) n_r$. The corresponding coding gain is given by

$$\prod_{i=1}^{n_r r\left(\mathbf{D}_d \mathbf{D}_d^H\right)} \lambda_i \left((\mathbf{I}_{n_r} \otimes \mathbf{D}_d) \mathbf{\Xi}_r (\mathbf{I}_{n_r} \otimes \mathbf{D}_d^H)\right). \qquad (7.46)$$

This coding gain can equivalently be expressed as follows. Defining a function $\tau_{deg}(\theta_t)$ by

$$\tau_{deg}(\theta_t) = \{k | (\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_t) \neq 0\}, \qquad (7.47)$$

---

[1] i.e. such that $\left\{\|(\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_{t,0})\|^2 \lambda_i(\mathbf{R}_r)\frac{\rho}{4} \gg 1\right\}_{i=1}^{n_r}$

[2] i.e. such that $\left\{\lambda_i \left((\mathbf{I}_{n_r} \otimes \mathbf{D}_d) \mathbf{\Xi}_r (\mathbf{I}_{n_r} \otimes \mathbf{D}_d^H)\right)\frac{\rho}{4} \gg 1\right\}_{i=1}^{r\left(\mathbf{D}_d \mathbf{D}_d^H\right) n_r}$

we may construct the matrices $\mathbf{D}'_d$ by suppressing the rows and columns that correspond to the time indexes k that are not contained in $\tau_{deg}(\theta_t)$. Similarly, we construct the matrix $\mathbf{\Xi}'_r$ by suppressing the rows and columns that correspond to the zero-rows and zero-columns of $\mathbf{I}_{n_r} \otimes \mathbf{D}_d$. Consequently, the non-zero eigenvalues of $\left(\mathbf{I}_{n_r} \otimes \mathbf{D}'_d\right) \mathbf{\Xi}'_r \left(\mathbf{I}_{n_r} \otimes \mathbf{D}'^H_d\right)$ and $\left(\mathbf{I}_{n_r} \otimes \mathbf{D}_d\right) \mathbf{\Xi}_r \left(\mathbf{I}_{n_r} \otimes \mathbf{D}^H_d\right)$ remain identical and we may rewrite the coding gain as

$$
\prod_{i=1}^{n_r r\left(\mathbf{D}_d \mathbf{D}^H_d\right)} \lambda_i \left(\left(\mathbf{I}_{n_r} \otimes \mathbf{D}_d\right) \mathbf{\Xi}_r \left(\mathbf{I}_{n_r} \otimes \mathbf{D}^H_d\right)\right)
$$

$$
= \det\left(\mathbf{\Xi}'_r \left(\mathbf{I}_{n_r} \otimes \mathbf{D}'_d \mathbf{D}'^H_d\right)\right)
$$

$$
= \det\left(\mathbf{\Xi}'_r\right) \left(\det\left(\mathbf{D}'_d \mathbf{D}'^H_d\right)\right)^{n_r}
$$

$$
= \det\left(\mathbf{\Xi}'_r\right) \prod_{k \in \tau_{deg(\theta_{t,0})}} \left|(\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_{t,0})\right|^{2n_r}
$$

$$
\leq \prod_{k \in \tau_{deg(\theta_{t,0})}} \left|(\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_{t,0})\right|^{2n_r}. \tag{7.48}
$$

In the high SNR regime, the PEP is thus approximated as

$$
P\left(\mathbf{C} \to \mathbf{E} | \theta_t = \theta_{t,0}\right) \approx \frac{1}{\pi} \int_0^{\pi/2} \eta^{-n_r \sharp \tau_{deg(\theta_{t,0})}} \left(\det\left(\mathbf{\Xi}'_r\right)\right)^{-1}
$$

$$
\prod_{k \in \tau_{deg(\theta_{t,0})}} \left|(\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_{t,0})\right|^{-2n_r} d\beta \tag{7.49}
$$

where we recall that $\sharp \tau_{deg(\theta_{t,0})}$ denotes the cardinality of the set $\tau_{deg(\theta_{t,0})}$. On the one hand, the receive-time correlation thereby induces a loss in the coding gain but there is no interaction between the channel and the codewords. The loss is smaller as the receive correlation decreases and/or the temporal correlation decreases. On the other hand, the transmit correlation may reduce the temporal diversity, since the diversity achieved by a pair of codewords $\{\mathbf{C}, \mathbf{E}\}$ is equal to $\sharp \tau_{deg}(\theta_{t,0})$ in channels degenerate along the direction $\theta_{t,0}$. Such diversity order may be sensibly different from the diversity extracted in i.i.d. channels, equal to the length of error event path $\tau_{\mathbf{C},\mathbf{E}} = \{k | \mathbf{c}_k - \mathbf{e}_k \neq 0\}$.

Ricean fading channels
In mobile scenarios, the transmit angle spread of the Ricean component often tends toward zero, i.e. the Ricean component is essentially composed of a single coherent component of direction of departure $\theta_{t,0}$

$$
\text{vec}\left(\left(\mathbf{1}_{1 \times T} \otimes \bar{\mathbf{H}}\right)^H\right)^H = \left(\bar{\mathbf{H}}(:,1)\right)^T \otimes \mathbf{1}_{1 \times T} \otimes \mathbf{a}_t^T(\theta_{t,0}). \tag{7.50}
$$

As the Ricean K-factor increases and/or the SNR decreases such that $\eta = \rho/(4 \sin^2 \beta (1 + K)) << 1$, the relationship between the PEP and the direction of departure $\theta_{t,0}$ is

outlined by

$$P\left(\mathbf{C} \to \mathbf{E} | \theta_t = \theta_{t,0}\right) \approx \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\eta K \|(\mathbf{C}-\mathbf{E})^T \mathbf{a}_t(\theta_{t,0})\|^2 \|\bar{\mathbf{H}}(:,1)\|^2\right) d\beta$$

$$\approx \mathcal{Q}\left(\sqrt{\frac{\rho}{2}\frac{K}{1+K}\|(\mathbf{C}-\mathbf{E})^T \mathbf{a}_t(\theta_{t,0})\|^2 \|\bar{\mathbf{H}}(:,1)\|^2}\right). \qquad (7.51)$$

Analogous to slow fading degenerate Rayleigh channels, the PEP depends on the Euclidean distance between codewords $\mathbf{a}_t^T(\theta_{t,0})\,\mathbf{C}$ and $\mathbf{a}_t^T(\theta_{t,0})\,\mathbf{E}$.

7.3.3 Catastrophic codes and general design criteria

As with MIMO i.i.d. Rayleigh fading channels, we define a minimum squared Euclidean distance $G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$ and a minimum product distance $G_{product}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$ in the case of MIMO degenerate channels as follows [Cle05, COV$^+$on, CVJV05].

Definition 7.1 *The squared minimum Euclidean distance* $G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$ *of a code characterized by the codebook* $\mathcal{C}$ *is defined as*

$$G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t)) \triangleq \min_{\substack{\mathbf{C},\mathbf{E}\in\mathcal{C} \\ \mathbf{C}\neq\mathbf{E}}} \sum_{k=0}^{T-1} \left|(\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_t)\right|^2. \qquad (7.52)$$

Definition 7.2 *The minimum product distance* $G_{product}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$ *of a code characterized by the codebook* $\mathcal{C}$ *is defined as*

$$G_{product}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$$
$$\triangleq \begin{cases} \min_{\substack{\mathbf{C},\mathbf{E}\in\mathcal{C} \\ \mathbf{C}\neq\mathbf{E}}} \prod_{\substack{k\in\tau_{deg}(\theta_t) \\ \sharp\tau_{deg}(\theta_t)=l_{min}(\theta_t)}} |(\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_t)|^2 & \text{if } l_{min}(\theta_t) = L_{min} \\ 0 & \text{if } l_{min}(\theta_t) < L_{min} \end{cases} \qquad (7.53)$$

where $l_{min}(\theta_t) = \min_{\substack{\mathbf{C},\mathbf{E}\in\mathcal{C} \\ \mathbf{C}\neq\mathbf{E}}} \sharp\tau_{deg}(\theta_t)$ *and* L_{min} *is the minimum effective length of the code in i.i.d. channels.*

Over the space of candidate codes $\mathcal{C}_{i.i.d.}$ of interest, we are now able to define sum-wise catastrophic, product-wise catastrophic and robust space–time codes.

Definition 7.3 *A sum-wise catastrophic space–time code* $\mathcal{C} \in \mathcal{C}_{i.i.d.}$ *in the direction of departure* θ_t *satisfies*

$$G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t)) = 0. \qquad (7.54)$$

Definition 7.4   *A product-wise catastrophic space–time code $\mathcal{C} \in \mathcal{C}_{i.i.d.}$ in the direction of departure θ_t satisfies*

$$G_{product}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t)) = 0. \tag{7.55}$$

Definition 7.5   *A robust space–time code $\mathcal{C} \in \mathcal{C}_{i.i.d.}$ satisfies*

$$\min_{\theta_t} G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t)) \geq \epsilon_1$$

$$\min_{\theta_t} G_{product}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t)) \geq \epsilon_2 \tag{7.56}$$

for some $\epsilon_1, \epsilon_2 > 0$.

Focusing on the minimization of the maximum PEP (i.e. on the optimization of the worst case), discussions in the previous section lead to the following design criterion.

Design criterion 7.4   *Over the space of codes that perform well on i.i.d. Rayleigh channels, denoted as $\mathcal{C}_{i.i.d.}$,*

- *in slow fading channels, choose a robust code $\in \mathcal{C}_{i.i.d.}$ with large G_{sum} over all directions of departure*
- *in fast fading channels, choose a robust code $\in \mathcal{C}_{i.i.d.}$ with large $G_{product}$ over all directions of departure.*

It is important to realize that the above code design criterion requires the knowledge of the signal constellation and of the geometry of the transmit antenna array.

Furthermore, in fast fading channels, it is sensible to choose among codes with large $G_{product}$ those with large G_{sum}. Indeed, in the low SNR regime, a coherent component makes the channel behave as if it were slow fading, even if the Rayleigh component is fast fading. Obviously, it might not be possible to simultaneously maximize $\min_{\theta_t} G_{sum}$ and $\min_{\theta_t} G_{product}$.

Since the MIMO channel is converted to a SIMO channel in channels degenerate along any direction of departure θ_t, a simple SIMO transmission based on a code $\mathcal{C}_{SIMO}$ may perform better than the MIMO transmission. Assuming that we focus only on the maximum PEP, it is sufficient[3] that

$$G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t)) < d_e^2 \tag{7.57}$$

in slow fading channels, and that

$$l_{min}(\theta_t) < L_{min, \mathcal{C}_{SIMO}} \tag{7.58}$$

---

[3] To be more rigorous, the distance spectrum should be taken into account.

or that

$$G_{product}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t)) < d_p^2 \quad \text{if } l_{min}(\theta_t) = L_{min,\mathcal{C}_{\text{SIMO}}} \tag{7.59}$$

in fast fading channels, where d_e^2 and d_p^2 are, respectively, the minimum squared Euclidean distance and[4] the minimum product distance over the error event path of minimum length, $L_{min,\mathcal{C}_{\text{SIMO}}}$ of $\mathcal{C}_{\text{SIMO}}$.

Equation (7.40) also provides a revealing insight into the diversity achieved by sum-wise and product-wise catastrophic codes.

Proposition 7.7 *In the high SNR regime[5], a non sum-wise catastrophic code achieves a transmit diversity order ≥ 1 in any slow Rayleigh fading MIMO channel. A sum-wise catastrophic code achieves a diversity order equal to zero in degenerate slow fading channels for some directions of departure. At high SNR[6], a non product-wise catastrophic code achieves a transmit diversity order $= L_{min}$ on any fast fading MIMO channel (i.e. channels with full rank Ξ_r). A product-wise catastrophic code achieves a diversity order $l_{min}(\theta_t) \leq L_{min}$ in degenerate fast fading channels for some directions of departure.*

Corollary 7.1 *In a slow fading channel degenerate in the direction of departure $\theta_{t,0}$, a space–time code sum-wise catastrophic along $\theta_{t,0}$ is always outperformed by a SIMO transmission, irrespective of its rate.*

It is worth mentioning that in slow fading scenarios, the channel degeneration decreases by at least $n_t - 1$ the diversity order achievable by the code. By contrast, the impact of degeneration in fast fading channels is sensibly different. In this case, the transmit diversity achieved by the code is not affected by the degeneration at high SNR if the code is non product-wise catastrophic.

Note also that the G_{sum} criterion is not limited to Rayleigh or Ricean distributed degenerate channels but to any channel distribution, as highlighted by (6.9).

It is important to emphasize that degenerate channels rarely exist, as they are equivalent to a zero angle-spread. Yet, many channels may be considered as quasi-degenerate at realistic SNR as soon as the transmit correlation is high enough. In such conditions, (7.45) and (7.49) may also be written. The larger G_{sum} and $G_{product}$, the smaller the SNR range for which the code appears as quasi sum-wise or product-wise catastrophic in some directions. Furthermore, even if the channel is not degenerate but presents some transmit correlation (i.e. the transmit angle spread is non-zero but limited), the finite SNR regime still implies that some eigenvalues of the channel correlation matrix do not contribute to diversity (see Chapter 6 for more information). What the above comment means is that the design criteria

---

[4] with the same total transmit power
[5] i.e. such that $\|(\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_t)\|^2 \frac{\rho}{4} >> 1$
[6] i.e. such that $\|(\mathbf{c}_k - \mathbf{e}_k)^T \mathbf{a}_t(\theta_t)\|^2 \frac{\rho}{4} >> 1 \quad \forall k \in \tau_{deg}(\theta_t)$ with $\sharp\tau_{deg}(\theta_t) = l_{min}(\theta_t)$

based on G_{sum} and $G_{product}$ naturally account for spatio-temporal correlations at finite SNR. Indeed, Chapter 6 has highlighted the importance of having error matrices whose dominant eigenvectors (i.e. corresponding to the largest eigenvalues) lie in the space spanned by the dominant eigenvector of the transmit correlation matrix. In that sense, the vector $\mathbf{a}_t(\theta_t)$ can be more generally thought of as an eigenvector of the transmit correlation matrix and not only as the transmit array response. Actually, $\mathbf{a}_t(\theta_t)$ is exactly the dominant eigenvector of the transmit correlation matrix if the channel is degenerate. This indicates that the G_{sum} and $G_{product}$ design criteria ensure that the dominant eigenvectors of the error matrices never lie in the space spanned by the weakest eigenvectors of the transmit correlation matrix (i.e. those corresponding to the smallest eigenvalues). This observation really allows us to appreciate the usefulness of G_{sum} and $G_{product}$ and clarifies the relationships between the developments of Chapter 6 and the design criteria of the present chapter.

Full rank vs. rank-deficient codes
Let us first define $\check{\mathbf{E}}$ as

$$\check{\mathbf{E}} = (\mathbf{C} - \mathbf{E})^*(\mathbf{C} - \mathbf{E})^T. \tag{7.60}$$

Note that $\check{\mathbf{E}} = \tilde{\mathbf{E}}^*$.

Lemma 7.1 *The minimum of $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ over all directions of departure θ_t is lower bounded by $\min_{\theta_t} \|\mathbf{a}_t(\theta_t)\|^2 \lambda_{min}$*

$$\min_{\theta_t} G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t)) \geq \min_{\theta_t} \|\mathbf{a}_t(\theta_t)\|^2 \lambda_{min} \tag{7.61}$$

with λ_{min} the smallest eigenvalue of $\tilde{\mathbf{E}}$ evaluated over all pairs of codewords $\mathbf{C}$, $\mathbf{E}$ with $\mathbf{C} \neq \mathbf{E}$. The lower bound is achieved iff, for a direction of departure $\bar{\theta}_t = \arg\min_{\theta_t} \|\mathbf{a}_t(\theta_t)\|^2$, $\mathbf{a}_t(\bar{\theta}_t)$ lies in the space spanned by the eigenvectors of $\check{\mathbf{E}}$ corresponding to λ_{min}.

PROOF: Taking the EVD decompositions of $\check{\mathbf{E}}$ and $\mathbf{A}$

$$\check{\mathbf{E}} = \mathbf{U}_{\check{\mathbf{E}}} \mathbf{\Lambda}_{\check{\mathbf{E}}} \mathbf{U}_{\check{\mathbf{E}}}^H,$$

$$\mathbf{A} = \mathbf{a}_t(\theta_t)\mathbf{a}_t^H(\theta_t) = \mathbf{U}_{\mathbf{A}} \mathbf{\Lambda}_{\mathbf{A}} \mathbf{U}_{\mathbf{A}}^H,$$

$$\mathbf{U}_{\mathbf{Q}} = \mathbf{U}_{\mathbf{A}}^H \mathbf{U}_{\check{\mathbf{E}}}, \tag{7.62}$$

and arranging the eigenvalues in $\mathbf{\Lambda}_{\check{\mathbf{E}}} = \mathrm{diag}\{\lambda_1(\check{\mathbf{E}}), \ldots, \lambda_{n_t}(\check{\mathbf{E}})\}$ and $\mathbf{\Lambda}_{\mathbf{A}} = \mathrm{diag}\{\lambda_1(\mathbf{A}), 0, \ldots, 0\}$ in decreasing order, $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ is rewritten as

$$\|(\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_t)\|^2 = \mathrm{Tr}\{\check{\mathbf{E}}\mathbf{A}\} = \mathrm{Tr}\{\mathbf{U}_{\mathbf{Q}} \mathbf{\Lambda}_{\check{\mathbf{E}}} \mathbf{U}_{\mathbf{Q}}^H \mathbf{\Lambda}_{\mathbf{A}}\}$$

$$= \lambda_1(\mathbf{A}) \sum_{i=1}^{n_t} |\mathbf{U}_{\mathbf{Q}}(1, i)|^2 \lambda_i(\check{\mathbf{E}})$$

$$\geq \min_{\theta_t} \|\mathbf{a}_t(\theta_t)\|^2 \min_{i=1,\ldots,n_t} \lambda_i(\check{\mathbf{E}}). \tag{7.63}$$

Since $\mathbf{U_Q}(1,:) = \mathbf{U}_\mathbf{A}^H(1,:)\mathbf{U}_{\check{\mathbf{E}}} = \frac{1}{\|\mathbf{a}_t(\theta_t)\|}\mathbf{a}_t^H(\theta_t)\mathbf{U}_{\check{\mathbf{E}}}$, equality in (7.63) occurs when, for a direction of departure $\bar{\theta}_t$ such that $\bar{\theta}_t = \arg\min_{\theta_t}\|\mathbf{a}_t(\theta_t)\|^2$, $\mathbf{a}_t(\bar{\theta}_t)$ lies in the space spanned by $\{\mathbf{U}_{\check{\mathbf{E}}}(:,i)\}_{i\in\mathbf{J}}$ where $\mathbf{J} = \{i \in \{1,\ldots,n_t\} : \lambda_i(\check{\mathbf{E}}) = \min_{i=1,\ldots,n_t}\lambda_i(\check{\mathbf{E}})\}$. Since eigenvalues of $\check{\mathbf{E}}$ and $\tilde{\mathbf{E}}$ are the same, the minimization over all pairs of codewords $\mathbf{C}, \mathbf{E}$ with $\mathbf{C} \neq \mathbf{E}$ yields the expected result. ∎

It is worth noting the similarities with Proposition 7.1.

Lemma 7.1 leads to the following properties for full-rank and rank-deficient codes.

Proposition 7.8 *A full diversity code is never sum-wise catastrophic. A rank-deficient code is sum-wise catastrophic in the direction of departure θ_t iff $\mathbf{a}_t(\theta_t)$ lies in the null space of $\check{\mathbf{E}}$.*

PROOF: It comes directly from the application of Lemma 7.1 to full-rank and rank-deficient codes. ∎

About the robustness of classical Spatial Multiplexing
In order to better understand the notion of catastrophic codes, let us consider the classical Spatial Multiplexing.

Proposition 7.9 *The PSK/QAM based classical Spatial Multiplexing is catastrophic[7] for a balanced transmit antenna array.*

PROOF: Let us first focus on the particular case of two transmit antennas. For an arbitrary transmit steering vector expressed as $\mathbf{a}_t(\theta_t) = [1 \quad \beta(\theta_t)e^{j\varphi(\theta_t)}]^T$, where $\beta(\theta_t)$ is the antenna gain imbalance and $\varphi(\theta_t)$ is a function of the direction of departure θ_t, the space orthogonal to $\mathbf{a}_t(\theta_t)$ is spanned by the vector $\mathbf{a}_t^\perp(\theta_t) = [\beta(\theta_t) \quad -e^{j\varphi(\theta_t)}]^T$. Taking the decomposition $\check{\mathbf{E}} = \mathbf{U}_{\check{\mathbf{E}}}\mathbf{\Lambda}_{\check{\mathbf{E}}}\mathbf{U}_{\check{\mathbf{E}}}^H$ and arranging the eigenvalues in decreasing order from top left to bottom right, Proposition 7.8 indicates that the code is not catastrophic if the eigenvector of $\check{\mathbf{E}}$ corresponding to the non-zero eigenvalue (rank 1 code) $\mathbf{U}_{\check{\mathbf{E}}}(:,1)$ does not belong to the space spanned by $\mathbf{a}_t^\perp(\theta_t)$. So this eigenvector $\mathbf{U}_{\check{\mathbf{E}}}(:,1)$ must be such that $\mathbf{U}_{\check{\mathbf{E}}}(:,1) \neq \alpha\,\mathbf{a}_t^\perp(\theta_t)$ with α a complex scalar. This must be valid for all $\mathbf{a}_t^\perp(\theta_t)$, i.e. for all possible[8] φ. For a balanced array, this requires that $\mathbf{U}_{\check{\mathbf{E}}}(:,1)$ be such that $|\mathbf{U}_{\check{\mathbf{E}}}(1,1)| \neq |\mathbf{U}_{\check{\mathbf{E}}}(2,1)|$. In the case of SM, $(\mathbf{C}-\mathbf{E}) = [c_1 - e_1 \quad c_2 - e_2]^T$ is a column vector, and the singular value decomposition of $\check{\mathbf{E}}$ yields that $\mathbf{U}_{\check{\mathbf{E}}}(:,1)$ is equal to $(\mathbf{C}-\mathbf{E})^*$ multiplied by a normalization factor to guarantee the unitarity. So a non-catastrophic SM would require that $|c_0 - e_0| \neq |c_1 - e_1|$ for all possible $c_1 - e_1$ and $c_2 - e_2$. However, for classical SM, this condition is never fulfilled: or any PSK/QAM constellation, taking $c_1 = c_2$ and $e_1 = e_2$ with $c_1 - e_1 \neq 0$ leads to $|c_1 - e_1| = |c_2 - e_2|$.

---

[7] Sum-wise catastrophic is equivalent to product-wise catastrophic for SM schemes, the denomination catastrophic denoting thus either sum-wise catastrophic either product-wise catastrophic.

[8] We will assume that φ in (6.4) may vary over 2π. For a linear array, this requires that $d/\lambda \geq 0.5$.
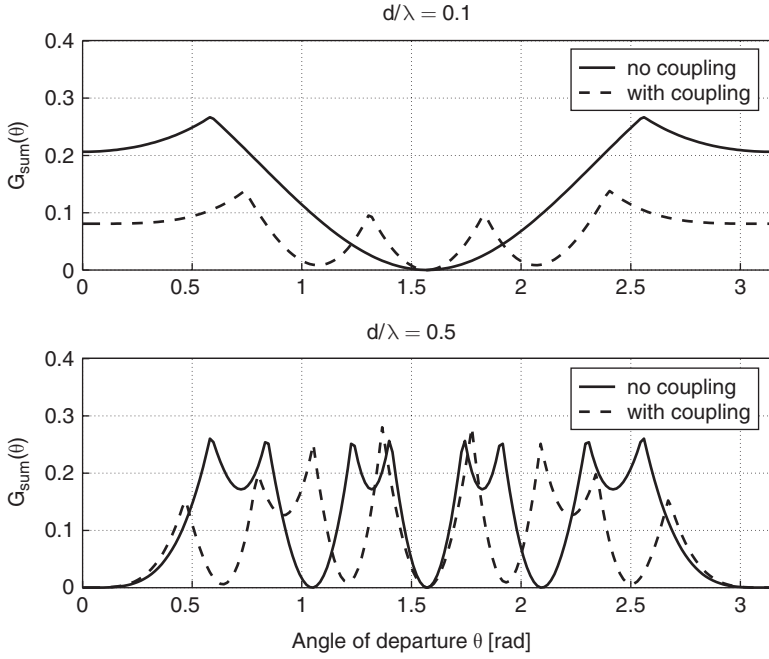
Figure 7.3   *Mutual coupling effects on the $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ of a SM scheme with QPSK*

For $n_t \geq 2$, having $|(\mathbf{C} - \mathbf{E})(1, 1)| = |(\mathbf{C} - \mathbf{E})(2, 1)|$ and $|(\mathbf{C} - \mathbf{E})(3 : n_t, 1)| = 0$ is a sufficient condition to show that classical SM is catastrophic for the PSK and QAM constellations irrespective of the array configuration.   ∎

Note that in the case of an unbalanced array, classical SM is not necessarily catastrophic. In degenerate channels, the classical SM may thus perform better in the presence of an unbalanced array than in the presence of a balanced array. This is also the case of any catastrophic code using PSK/QAM constellations.

Mutual coupling: a natural precoder

We have just seen that classical SM is catastrophic if the transmit array is balanced. Let us now illustrate that the mutual coupling between antennas may improve the situation, considering two transmit antennas. The proof of Proposition 7.9 indeed tells us that SM with mutual coupling is not robust if $|a(c_1 - e_1) + b(c_2 - e_2)| = |b(c_1 - e_1) + a(c_2 - e_2)|$ for some c_1, e_1, c_2, e_2, a and b denoting the elements of the coupling matrix. This situation may take place, irrespective of a and b, e.g. if $c_1 - e_1 = c_2 - e_2 \neq 0$ (this is common with the classical PSK and QAM constellations). Consequently, SM is also catastrophic in the presence of antenna coupling in some directions. However, mutual coupling may improve the performance in certain cases. The squared minimum distances $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ using QPSK constellation are displayed in Figure 7.3 for antenna spacings of 0.1λ and 0.5λ. The resulting symbol error rates corresponding to absolute transmit correlations $|t|$ of 0.95 and
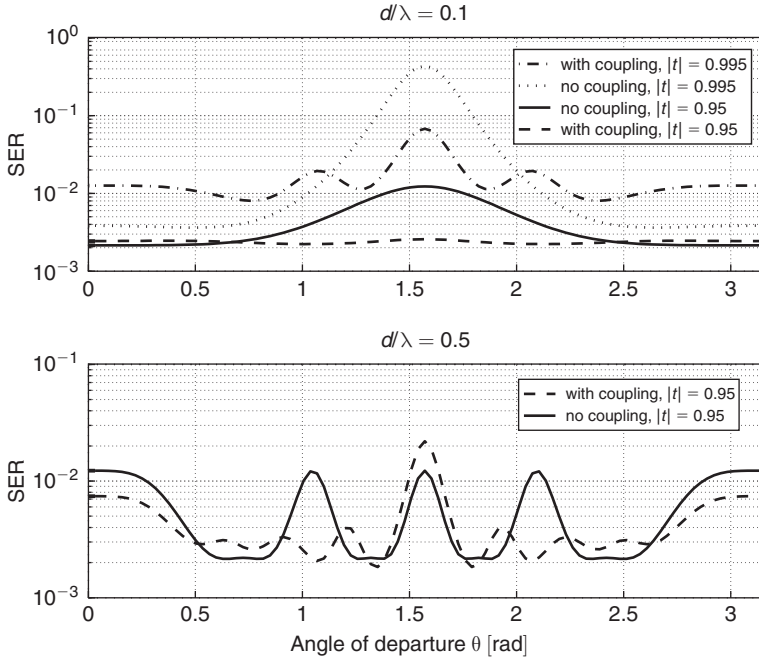
Figure 7.4 *Mutual coupling effects on the SER of a SM scheme with QPSK*

0.995 (the receive correlation r is taken as zero) are displayed in Figure 7.4 as a function of the direction of departure at a SNR of 20 dB.

Let us first focus on the smaller antenna spacing ($d/\lambda = 0.1$). When mutual coupling is neglected, a large null appears around $\pi/2$ in the plot of $G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$. In Figure 7.4, this null corresponds to significant degradations. When mutual coupling is taken into account, the null remains (SM with coupling is still catastrophic) but becomes nevertheless narrower. From a SER perspective,

- for $|t| = 0.95$, the performance remains flat around $\theta_t = \pi/2$, as if the null of $G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$ was invisible
- for $|t| = 0.995$, a bump appears in that same direction, as expected.

What does this mean? We know that if $G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$ presents a null along a specific angle of departure, the code performance may be degraded if the channel is degenerated along this direction. The severity of the degradation is strongly related to the relative width of the null with respect to the channel transmit directional spread, or equivalently, with respect to the absolute transmit correlation. As the dip becomes narrower, a smaller directional spread (i.e. a higher transmit correlation) is required for this dip to impact the SER.

Consequently, for small antenna spacings, mutual coupling may improve the performance of SM in certain directions by reducing the width of the nulls in $G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$. Yet,

for angles of departure smaller than 0.75 or larger than 2.25, the performance without coupling is at least equal or better than that with coupling because $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ is larger without coupling than with coupling in those directions (and this impact is visible because the SNR is finite).

For $d/\lambda = 0.5$, the same discussion holds true. Yet, mutual coupling also increases the width of the nulls (e.g. around $\pi/2$), which leads to a performance degradation.

As a more general conclusion, this example illustrates that codes with nulls in $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ should be avoided. If it is not possible, the nulls should be as narrow as possible.

7.4 Error probability motivated code design in slow fading channels

7.4.1 Full rank codes

From Lemma 7.1, the following design criterion guarantees a large $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ for full-rank codes.

Design criterion 7.5 *Choose the code $\mathcal{C} \in \mathcal{C}_{i.i.d.}$ such that*

$$\mathcal{C} = \arg \max_{\mathcal{C}_{i.i.d.}} \min_{\substack{\mathbf{C}, \mathbf{E} \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{E}}} \min_{k=1,\dots,n_t} \lambda_k\left(\breve{\mathbf{E}}\right) \tag{7.64}$$

i.e. choose the code $\mathcal{C}$ that maximizes the smallest eigenvalue of $\breve{\mathbf{E}} = (\mathbf{C} - \mathbf{E})^*(\mathbf{C} - \mathbf{E})^T$ evaluated over all pairs of codewords $\mathbf{C}$, $\mathbf{E}$ with $\mathbf{C} \neq \mathbf{E}$.

This design criterion is quite similar to the universal code design criterion for MISO channels. Unlike Criterion 7.4, the above criterion designs robust full-rank space-time codes independently of the transmit antenna array configuration. However, it has to be noted that (7.64) is only a sufficient condition. As will be shown in the design examples, codes not fulfilling (7.64), but satisfying Criterion 7.4 for a given transmit array configuration, may outperform codes designed via (7.64). Criterion 7.5 has also been derived in [Gam02] for the case of line-of-sight channels.

7.4.2 Linear space–time block codes

As seen in (5.51), a linear STBC is expressed in its general form as

$$\mathbf{C} = \sum_{q=1}^{Q} \mathbf{\Phi}_q \Re[c_q] + \mathbf{\Phi}_{q+Q} \Im[c_q] \tag{7.65}$$

where $\mathbf{\Phi}_q$ are complex basis matrices of size $(n_t \times T)$ with T being the codeword duration relative to the symbol duration, c_q stands for the complex information symbol taken from

PSK or QAM constellations, Q is the number of complex symbols c_q transmitted over a codeword, $\Re$ and $\Im$ stand for the real and imaginary parts.

We have already discussed the case of Spatial Multiplexing. The following proposition deals with the ultimate performance of a linear dispersion code based on wide ($T \geq n_t$) unitary basis matrices such that $\boldsymbol{\Phi}_q \boldsymbol{\Phi}_q^H = \frac{T}{Qn_t} \mathbf{I}_{n_t}\ \forall q = 1, \ldots, 2Q$.

Proposition 7.10 *For a direction of departure θ_t, a PSK/QAM based linear dispersion code consisting of wide unitary basis matrices achieves the largest $G_{sum}\left(\theta_t | \mathcal{C},\ \mathbf{a}_t(\theta_t)\right)$ equal to*

$$\frac{T}{Qn_t} \|\mathbf{a}_t(\theta_t)\|^2 d_{min}^2 \tag{7.66}$$

if the matrices $\left\{\boldsymbol{\Phi}_q\right\}_{q=1}^{2Q}$ satisfy the conditions

$$\mathrm{Tr}\left\{\mathbf{a}_t^*(\theta_t)\mathbf{a}_t^T(\theta_t)\left(\boldsymbol{\Phi}_q \boldsymbol{\Phi}_l^H + \boldsymbol{\Phi}_l \boldsymbol{\Phi}_q^H\right)\right\} = 0 \quad q \neq l \tag{7.67}$$

where d_{min}^2 is the minimum squared Euclidean distance of the constellation used.

Proof: Denoting $d_q = c_q - e_q$ with e_q the complex symbols of the codeword $\mathbf{E}$, observe that

$$\min_{\substack{\mathbf{C},\mathbf{E}\in\mathcal{C} \\ \mathbf{C}\neq\mathbf{E}}} \|(\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_t)\|^2$$

$$\leq \min_{q=1,\ldots,Q} \min_{d_q} \left[\sum_{q=1}^{Q} \|\mathbf{a}_t^T(\theta_t)\left(\boldsymbol{\Phi}_q \Re[d_q] + \boldsymbol{\Phi}_{q+Q}\Im[d_q]\right)\|\right]^2$$

$$\overset{(a)}{\leq} \min_{q=1,\ldots,Q} \min_{d_q} \|\mathbf{a}_t(\theta_t)\|^2 \frac{T}{Qn_t}\left[|\Re[d_q]|^2 + |\Im[d_q]|^2\right]$$

$$+ \mathrm{Tr}\left\{\mathbf{a}_t^*(\theta_t)\mathbf{a}_t^T(\theta_t)\left(\boldsymbol{\Phi}_q \boldsymbol{\Phi}_{q+Q}^H + \boldsymbol{\Phi}_{q+Q}\boldsymbol{\Phi}_q^H\right)\right\}\Re[d_q]\Im[d_q]$$

$$\overset{(b)}{\leq} \frac{T}{Qn_t}\|\mathbf{a}_t(\theta_t)\|^2 \min_{d_q}\left[|\Re[d_q]|^2 + |\Im[d_q]|^2\right]$$

$$= \frac{T}{Qn_t}\|\mathbf{a}_t(\theta_t)\|^2 d_{min}^2 \tag{7.68}$$

where in (a) we exploit the unitarity of the basis matrices, and in (b) we use the fact that for PSK or QAM constellations, $\Re[d_q]$ and $\Im[d_q]$ may be positive or negative irrespective of

$$\mathrm{Tr}\left\{\mathbf{a}_t^*(\theta_t)\mathbf{a}_t^T(\theta_t)\left(\boldsymbol{\Phi}_q \boldsymbol{\Phi}_{q+Q}^H + \boldsymbol{\Phi}_{q+Q}\boldsymbol{\Phi}_q^H\right)\right\}. \tag{7.69}$$

Taking the minimum over $\mathbf{C}, \mathbf{E}$ with $\mathbf{C} \neq \mathbf{E}$ of

$$
\|(\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_t)\|^2
$$

$$
= \sum_{q=1}^{Q} \|\mathbf{a}_t^T(\theta_t)\mathbf{\Phi}_q\|^2 |\Re[d_q]|^2 + \|\mathbf{a}_t^T(\theta_t)\mathbf{\Phi}_{q+Q}\|^2 |\Im[d_q]|^2
$$

$$
+ \mathrm{Tr}\{\mathbf{a}_t^*(\theta_t)\mathbf{a}_t^T(\theta_t)(\mathbf{\Phi}_q\mathbf{\Phi}_{q+Q}^H + \mathbf{\Phi}_{q+Q}\mathbf{\Phi}_q^H)\}\Re[d_q]\Im[d_q]
$$

$$
+ \sum_{q=1}^{Q}\sum_{l<q}^{Q} \mathrm{Tr}\{\mathbf{a}_t^*(\theta_t)\mathbf{a}_t^T(\theta_t)(\mathbf{\Phi}_q\mathbf{\Phi}_l^H + \mathbf{\Phi}_l\mathbf{\Phi}_q^H)\}\Re[d_q]\Re[\epsilon_l]
$$

$$
+ \mathrm{Tr}\{\mathbf{a}_t^*(\theta_t)\mathbf{a}_t^T(\theta_t)(\mathbf{\Phi}_{q+Q}\mathbf{\Phi}_{l+Q}^H + \mathbf{\Phi}_{l+Q}\mathbf{\Phi}_{q+Q}^H)\}\Im[d_q]\Im[\epsilon_l]
$$

$$
+ \mathrm{Tr}\{\mathbf{a}_t^*(\theta_t)\mathbf{a}_t^T(\theta_t)(\mathbf{\Phi}_q\mathbf{\Phi}_{l+Q}^H + \mathbf{\Phi}_{l+Q}\mathbf{\Phi}_q^H)\}\Re[d_q]\Im[\epsilon_l]
$$

$$
+ \mathrm{Tr}\{\mathbf{a}_t^*(\theta_t)\mathbf{a}_t^T(\theta_t)(\mathbf{\Phi}_{q+Q}\mathbf{\Phi}_l^H + \mathbf{\Phi}_l\mathbf{\Phi}_{q+Q}^H)\}\Re[\epsilon_l]\Im[d_q]. \tag{7.70}
$$

and equaling it to (7.68) leads to the sufficient condition in (7.66)   ∎

Proposition 7.11   *For a given slow fading channel with channel matrix $\mathbf{H}$, a PSK/QAM based linear dispersion code consisting of wide unitary basis matrices achieves the minimum conditional maximum PEP if the matrices $\{\mathbf{\Phi}_q\}_{q=1}^{2Q}$ satisfy the conditions*

$$
\mathrm{Tr}\{\mathbf{H}^H\mathbf{H}(\mathbf{\Phi}_q\mathbf{\Phi}_l^H + \mathbf{\Phi}_l\mathbf{\Phi}_q^H)\} = 0 \quad q \neq l. \tag{7.71}
$$

PROOF: It follows exactly from a proof to that of Proposition 7.10.   ∎

Corollary 7.2   *In order to fulfill Propositions 7.10 and 7.11, it is sufficient that the unitary basis matrices $\{\mathbf{\Phi}_q\}_{q=1}^{2Q}$ be pairwise skew-Hermitian*

$$
\mathbf{\Phi}_q\mathbf{\Phi}_l^H + \mathbf{\Phi}_l\mathbf{\Phi}_q^H = 0 \quad q \neq l. \tag{7.72}
$$

Orthogonal codes are the most robust linear codes based on unitary matrices and achieve the minimum conditional maximum PEP.

PROOF: This results directly from Propositions 7.10 and 7.11 and the properties of orthogonal codes.   ∎

It should be noted that the skew-hermitianity condition of Corollary 7.2 is sufficient but not necessary. The skew-hermitianity is necessary to maximize $\|(\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_t)\|^2$, but is

only sufficient to maximize $\min_{\mathbf{C},\mathbf{E}} \|(\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_t)\|^2$. The skew-Hermitian properties are thus necessary to optimize the G_{sum} distance spectrum. Depending on the structure of $\mathbf{H}$, it should be possible to almost satisfy (7.71) with non-skew-Hermitian basis matrices.

Defining the following quantity

$$\mathcal{Z}_q = \text{vec}\left(\begin{bmatrix} \Re\left(\mathbf{a}_t^T(\theta_t)\mathbf{\Phi}_q\right) \\ \Im\left(\mathbf{a}_t^T(\theta_t)\mathbf{\Phi}_q\right) \end{bmatrix}\right) \quad 1 \leq q \leq 2Q \tag{7.73}$$

and making use of the following relationship

$$\text{Tr}\{\mathbf{X}\mathbf{Y}^H + \mathbf{Y}\mathbf{X}^H\} = 2 \ \text{vec}\left(\begin{bmatrix} \Re(\mathbf{X}) \\ \Im(\mathbf{X}) \end{bmatrix}\right)^T \text{vec}\left(\begin{bmatrix} \Re(\mathbf{Y}) \\ \Im(\mathbf{Y}) \end{bmatrix}\right), \tag{7.74}$$

Proposition 7.10 leads to the following design criterion for linear codes based on unitary basis matrices.

Design criterion 7.6 *Choose the linear code $\in \mathcal{C}_{i.i.d.}$ consisting of unitary basis matrices $\{\mathbf{\Phi}_j\}_{j=1}^{2Q}$ such that*

$$\{\mathbf{\Phi}_q\}_{q=1}^{2Q} = \arg\min_{\mathcal{C}_{i.i.d.}} \max_{\theta_t} \max_{\substack{q=1,\dots,2Q \\ l<q}} \left|\mathcal{Z}_q^T \mathcal{Z}_l\right|. \tag{7.75}$$

Unlike design criteria 7.4 and 7.5, Criterion 7.6 does not need to know the signal constellation used. However, it requires to know the shape of the transmit array and the inter-element spacing. This might be useful if the constellation is not known or if the constellation size is so large that finding good codes based on criteria 7.4 or 7.5 is computationally cumbersome.

Corollary 7.2 leads to the following design criterion.

Design criterion 7.7 *Choose the linear code $\in \mathcal{C}_{i.i.d.}$ consisting of unitary basis matrices $\{\mathbf{\Phi}_q\}_{q=1}^{2Q}$ such that*

$$\{\mathbf{\Phi}_q\}_{q=1}^{2Q} = \arg\min_{\mathcal{C}_{i.i.d.}} \max_{\substack{q=1,\dots,2Q \\ l<q}} \left\|\mathbf{\Phi}_q\mathbf{\Phi}_l^H + \mathbf{\Phi}_l\mathbf{\Phi}_q^H\right\|_F^2. \tag{7.76}$$

This criterion does not require any information about the array or the constellations. Orthogonal codes are the most robust linear codes, irrespective of the structure of the steering vector $\mathbf{a}_t(\theta_t)$ and the signal constellation. This was intuitively predictable. Indeed, the structure of orthogonal codes is such that $\tilde{\mathbf{E}} = \delta\mathbf{I}_{n_t}$ with δ a scalar, indicating that the code excites all directions equally if the transmit antenna array is balanced. In this case,

$\|(\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_t)\|^2 = n_t \delta$ and is independent of $\mathbf{a}_t(\theta_t)$. However, the skew-hermitianity condition of Corollary 7.2 is sufficient. Depending on the knowledge of the inter-element spacing, the signal constellation and/or the shape of the transmit antenna array, the robustness of orthogonal codes is approached by applying criteria 7.4, 7.5, 7.6 or 7.7. It is trivial that the deeper the knowledge about the transmit array and/or the constellation used, the more robust the code. So, if possible, Criterion 7.4 should always be considered first.

From Proposition 7.10, one also obtains the following results.

Corollary 7.3   *In slow fading degenerate channels, a linear dispersion code consisting of unitary basis matrices and based on a constellation with a minimum squared Euclidean distance d_{min}^2 is always outperformed by an uncoded SIMO transmission based on a constellation with a minimum squared Euclidean distance larger than T/Qd_{min}^2.*

This indicates that most LDCs are always outperformed by a SIMO transmission in degenerate slow fading channels.

7.4.3   Virtual channel representation based design criterion

The virtual channel model presented in Section 3.2.2 may also be used to design robust precoders for Spatial Multiplexing schemes over Rayleigh fading channels with uniform linear arrays [HLHS03].

Recalling, from Section 3.2.2, that the correlation matrix may be written as

$$\mathbf{R} = \left[\hat{\mathbf{A}}_r^* \otimes \hat{\mathbf{A}}_t^*\right] \mathbf{R}_v \left[\hat{\mathbf{A}}_r^T \otimes \hat{\mathbf{A}}_t^T\right] \tag{7.77}$$

with $\hat{\mathbf{A}}_t = \left[\hat{\mathbf{a}}_t(\hat{\theta}_{t,1}) \ \ldots \ \hat{\mathbf{a}}_t(\hat{\theta}_{t,n_t})\right]$ and $\hat{\mathbf{A}}_r = \left[\hat{\mathbf{a}}_r(\hat{\theta}_{r,1}) \ \ldots \ \hat{\mathbf{a}}_r(\hat{\theta}_{r,n_r})\right]$, the average PEP in (6.25) now reads as

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \left(\det\left(\mathbf{I}_{n_r n_t} + \eta \mathbf{R}_v \left(\mathbf{I}_{n_r} \otimes \hat{\mathbf{A}}_t^T \tilde{\mathbf{E}} \hat{\mathbf{A}}_t^*\right)\right)\right)^{-1} d\beta \tag{7.78}$$

with $\tilde{\mathbf{E}} = (\mathbf{C} - \mathbf{E})(\mathbf{C} - \mathbf{E})^H$. For the Spatial Multiplexing scheme, codewords $\mathbf{C}$ are $n_t \times 1$ vectors.

The virtual channel representation assumes that $\mathbf{R}_v$ is diagonal, hence the average PEP (7.78) becomes

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \prod_{q=1}^{n_r} \left(1 + \eta \sum_{p=1}^{n_t} \mathbf{R}_v(x,x) \|(\mathbf{C} - \mathbf{E})^T \hat{\mathbf{a}}_t(\hat{\theta}_{t,p})\|^2\right)^{-1} d\beta \tag{7.79}$$

with $x = n_r(q-1) + p$. Since $\mathbf{R}_v$ is unknown to the transmitter, the maximum diversity order is extracted if $\|(\mathbf{C} - \mathbf{E})^T \hat{\mathbf{a}}_t(\hat{\theta}_{t,p})\|^2 > 0$ for $p = 1, \ldots, n_t$. Moreover, maximizing this

quantity is always advised since it contributes to a larger coding gain. Focusing on the minimization of the worst-case PEP, the design criterion based on the virtual channel model for Spatial Multiplexing in Rayleigh fading channels with uniform linear arrays relies on a unitary precoder $\mathbf{P}$ used in front of the original SM codewords $\mathbf{C} = 1/\sqrt{n_t}\begin{bmatrix} c_1 & \dots & c_{n_t} \end{bmatrix}^T \in \mathcal{C}$ in (5.80), such that the new transmitted codewords $\mathbf{C}' \in \mathcal{C}'$ are given by $\mathbf{C}' = \mathbf{P}\mathbf{C}$.

Design criterion 7.8 *Over the space of unitary precoder $\mathcal{P}$, choose the precoder $\mathbf{P} \in \mathcal{P}$ such that*

$$\mathbf{P} = \arg\max_{\mathcal{P}} \min_{\{\hat{\theta}_{tp}\}_{p=1}^{n_t}} \min_{\substack{\mathbf{C}, \mathbf{E} \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{E}}} \|(\mathbf{C} - \mathbf{E})^T \mathbf{P}^T \hat{\mathbf{a}}_t(\hat{\theta}_t)\|^2 \tag{7.80}$$

or equivalently

$$\mathbf{P} = \arg\max_{\mathcal{P}} \min_{\{\hat{\theta}_{tp}\}_{p=1}^{n_t}} G_{sum}(\hat{\theta}_t | \mathcal{C}', \hat{\mathbf{a}}_t(\hat{\theta}_t)) \tag{7.81}$$

i.e. choose the precoder $\mathbf{P} \in \mathcal{P}$ that maximizes the minimum of $G_{sum}(\hat{\theta}_t | \mathcal{C}', \hat{\mathbf{a}}_t(\hat{\theta}_t))$ over the n_t virtual directions of departure.

The precoder is chosen as $\mathbf{P} = \mathbf{W}\hat{\mathbf{A}}_t^*$ with $\mathbf{W}$ a diagonal or general unitary precoder. The effect of such a precoder is quite clear. Writing the received vector[9] as $\mathbf{r} = \sqrt{E_s}\mathbf{H}\mathbf{P}\mathbf{C} + \mathbf{n} = \sqrt{E_s}\hat{\mathbf{A}}_r\mathbf{H}_v\hat{\mathbf{A}}_t^T\mathbf{W}\hat{\mathbf{A}}_t^*\mathbf{C} + \mathbf{n}$, we observe that the product $\hat{\mathbf{A}}_t^T\mathbf{W}\hat{\mathbf{A}}_t^*$ acts as a filter so that the codewords $\mathbf{C}$ interact with the channel $\mathbf{H}_v$ in a constructive fashion.

Although there are connections with Criterion 7.4, the present criterion should be used with care when designing codes. Indeed, the number of directions of departure over which G_{sum} is maximized is now finite as the spatial domain is sampled into n_t virtual directions, $\hat{\theta}_{t,k}$, $k = 1, \dots, n_t$. By contrast, in Section 7.3.3, G_{sum} is a continuous function of θ_t. As observed in Chapter 3, taking $\mathbf{R}_v$ as diagonal may lead to incorrect channel representations and to overestimating the mutual information. As will be observed later, this assumption has also a significant impact when designing codes based on Criterion 7.8: the so-designed codes are not necessarily robust in all directions of departure, but only in the n_t virtual directions. Therefore, if the transmit correlation is high and if scatterers are distributed along directions not contained in the set of virtual directions of departure, the error performance might be highly degraded because the assumption of $\mathbf{R}_v$ being diagonal does not hold.

7.4.4   *Relationship with information theory motivated design*

Reasoning analogous to Section 7.1, the design of universal codes for degenerate channels comes to find codes that maximize

$$\frac{\rho}{2}D_d^2 = \frac{\rho}{2}\|(\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_t)\|^2 \|\mathbf{H}(1 : n_r, 1)\|^2 \tag{7.82}$$

---

[9] We suppress the time subscript k for simplicity.

under the constraint that

$$\log_2\left(1 + \frac{\rho}{n_t}\|\mathbf{H}(1:n_r,1)\|^2\|\mathbf{a}_t(\theta_t)\|^2\right) \geq R. \tag{7.83}$$

It follows that the worst-case PEP over all degenerate channels not in outage is given by

$$\min_{\substack{\mathbf{H}\in\mathcal{D} \\ \mathbf{H}\in\mathcal{N}_{\mathcal{O}}}} P(\mathbf{C}\rightarrow\mathbf{E}|\mathbf{H}) = \mathcal{Q}\left(\sqrt{\frac{\|(\mathbf{C}-\mathbf{E})^T\mathbf{a}_t(\theta_t)\|^2(2^R-1)n_t}{2\|\mathbf{a}_t(\theta_t)\|^2}}\right). \tag{7.84}$$

Being approximately universal for the degenerate channel is equivalent to guarantee

$$\min_{\substack{\mathbf{C},\mathbf{E}\in\mathcal{C} \\ \mathbf{C}\neq\mathbf{E}}} 2^R\|(\mathbf{C}-\mathbf{E})^T\mathbf{a}_t(\theta_t)\|^2 \overset{.}{\geq} 1. \tag{7.85}$$

Clearly, any approximately universal code in MIMO channels is approximately universal in degenerate channels since G_{sum} is always lower bounded by λ_{min} (see Lemma 7.1). A code with a non-vanishing determinant has also a non-vanishing minimum eigenvalue λ_{min}.

The Alamouti scheme is also approximately universal in degenerate channels. Indeed, O-STBCs achieve

$$G_{sum} = \frac{T}{Qn_t}\|\mathbf{a}_t(\theta_t)\|^2 d_{min}^2. \tag{7.86}$$

For QAM constellations, $d_{min}^2 \approx 1/2^{R/r_s}$, hence (7.85) is satisfied iff $r_s = 1$, i.e. by the Alamouti code.

Although information theory and error probability motivated approaches are different, their combination provides a much broader view of the performance of space–time codes in correlated channels. On the one hand, the information theory approach obtains a necessary condition to guarantee the approximate universality property, i.e. to achieve the high SNR trade-off over any slow fading channel distribution. This condition maximizes the minimum determinant and is similar to the code design obtained for i.i.d. Rayleigh slow fading channels. This is in agreement with the results of Chapter 6, where the performance of full-rank codes were shown to be only determined by the minimum determinant in the presence of Rayleigh and Ricean fading channels. Recall that the high-SNR diversity-multiplexing trade-offs in correlated Rayleigh and Ricean channels was also shown in Chapter 4 to be identical to the one obtained in i.i.d. Rayleigh fading channels. On the other hand, the error probability motivated approach focuses on a finite SNR behavior by accounting for the presence of fading correlations and/or a coherent component via their interactions with the codewords in the particular sets of Rayleigh and Ricean fading channels. However, further research is necessary to investigate how codes designed using both approaches perform in terms of the diversity-multiplexing trade-off at finite SNR.

7.4.5 Practical code designs in slow fading channels

There are several practical methods to design a robust code in slow fading channels. A first solution consists in optimizing a unitary precoder maximizing the global squared minimum Euclidean distance G_{sum}. Indeed, applying a unitary precoder $\mathbf{P}$ in front of a space–time code whose codebook is denoted as $\mathcal{C}$ does not modify its performance in i.i.d. Rayleigh fading channels. Hence, the precoder is chosen in the set of unitary precoders $\mathcal{P}$ such that

$$\mathbf{P} = \arg \max_{\mathbf{P}\in\mathcal{P}} \min_{\theta} \min_{\substack{\mathbf{C},\mathbf{E}\in\mathcal{C} \\ \mathbf{C}\neq\mathbf{E}}} \|(\mathbf{C} - \mathbf{E})^T \mathbf{P}^T \mathbf{a}_t(\theta_t)\|_F^2. \tag{7.87}$$

Another solution is to choose among all codes performing equally well in i.i.d. Rayleigh fading channels one that also performs best in correlated channels.

We have mentioned that Criterion 7.4 requires specification of the array configuration and the constellation. In the following, we consider an horizontal uniform linear array with 2 transmit antennas and various PSK modulations. We also assume that the transmit antenna spacing is larger than half wavelength, so that the minimization of Criterion 7.4 is carried out for φ varying over 2π.

Linear space–time block codes
The first method relies on the optimization of a unitary precoder, using a code performing well in i.i.d. Rayleigh fading channels. Let us apply this method in the case of $n_r \times 2$ Spatial Multiplexing. The error matrix $(\mathbf{C} - \mathbf{E}) = 1/\sqrt{2}[c_0 - e_0 \quad c_1 - e_1]^T$ is a column vector where c_0, e_0, c_1, e_1 are chosen from a given signal constellation. The unitary precoder

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{7.88}$$

must satisfy

$$|p_{11}(c_0 - e_0) + p_{12}(c_1 - e_1)| \neq |p_{21}(c_0 - e_0) + p_{22}(c_1 - e_1)| \tag{7.89}$$

for all $c_0 - e_0$ and $c_1 - e_1$ simultaneously different of zero. Expressing the precoder as

$$\mathbf{P} = \begin{bmatrix} -\cos\beta & \sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} e^{j\alpha} & 0 \\ 0 & 1 \end{bmatrix}, \tag{7.90}$$

the optimization must be carried out over two parameters β and α. For 4, 8 and 16 PSK constellations, the optimal values of β and α are summarized in Table 7.1.

In Figure 7.5, the squared minimum Euclidean distance $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ is displayed as a function of the angle of departure θ_t for various SM schemes: non-precoded classical SM, precoded classical SM (using QPSK and Table 7.1), and two precoded schemes, designed using the virtual channel motivated criterion (Criterion 7.8), with a diagonal
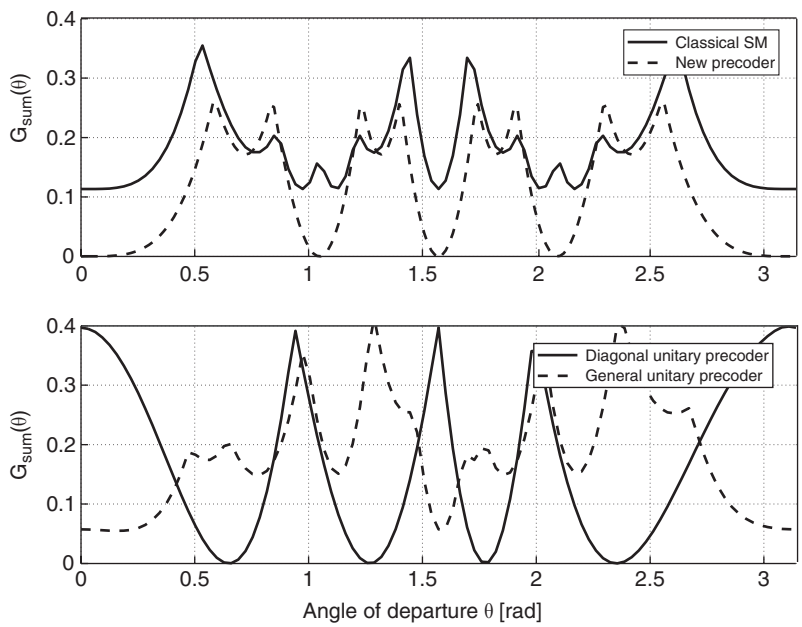
Figure 7.5 $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ *of several SM schemes as a function of the angle of departure θ_t [rad]*

Table 7.1   *Optimal precoder* **P** *for 4, 8 and 16 PSK SM transmission with $n_t = 2$*

	4-PSK	8-PSK	16-PSK
α	$\pi/4$	$\pi/8$	$\pi/16$
β	1.2631	1.1121	0.9716

unitary precoder and a general unitary precoder [HLHS03]. The classical SM scheme, as well as the diagonal unitary precoded scheme, are catastrophic in some directions, and their performance should significantly degrade if the azimuth power spectrum is confined in one of those directions. For comparison, the QPSK-based Alamouti scheme is characterized by a $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ of 2.

The symbol error rate (SER) achieved by the classical and precoded 2×2 SM schemes are displayed in Figure 7.6 for uncorrelated receive antennas over a Rayleigh fading channel, at a SNR of 20 dB and using a QPSK modulation. In the first graph, the SER is represented as a function of the phase of the transmit correlation t ($|t|$ being fixed to 0.95). In the second graph, the SER is represented as a function of $|t|$ (while the phase of t is equal to 0). For the sake of comparison, an alternative scheme, proposed in [NBP01], is also considered. However, this scheme assumes that the statistical properties of the channel are known by the transmitter (i.e. the transmitter is aware of the value of t), as illustrated in
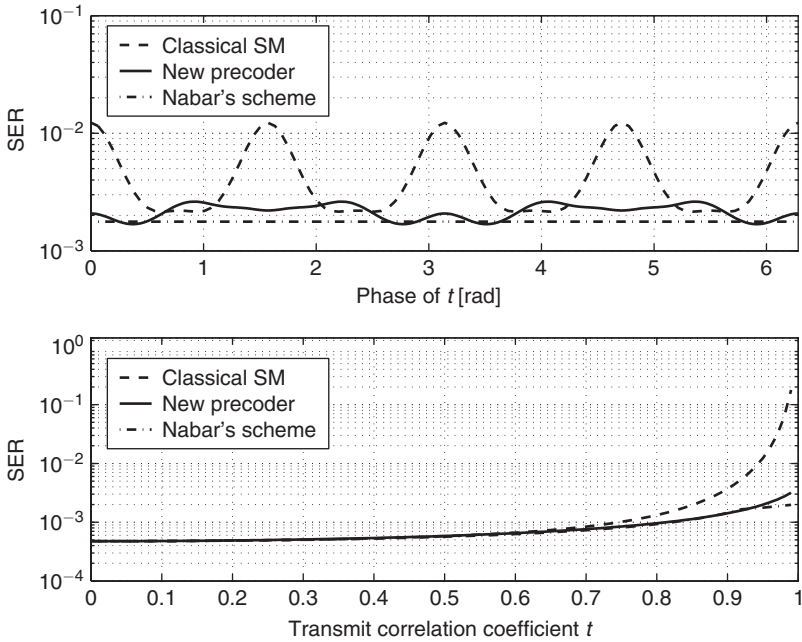
Figure 7.6   *SER as a function of the phase of t for |t| = 0.95 (up) and SER as a function of |t| with the phase of t equal to 0 (down)*

Chapter 8 (see Section 8.4). The precoded scheme performs satisfactorily for all phases of the transmit correlation coefficient, and its SER is actually quite close to that obtained by the scheme of [NBP01], although it does not require any channel information to be fed back to the transmitter.

To simulate the error probability of these schemes in real-world propagation scenarios, an uplink correlated Rayleigh fading channel is generated following the model developed in Section 2.3.3 in combination with the tap-delay line of SUI model 6. Highly directional scattering is considered (the local scattering ratio is fixed to zero, i.e. there is no local ring), and the receive antennas (at the base station) are spaced by several tens of wavelengths. This configuration should yield a high transmit correlation, whose phase is related to the transmit array orientation with respect to the transmit-to-receive axis. Two array orientations are considered so that scatterers are mostly located around directions $\theta_t = 0$ or direction $\theta_t = 0.63$, corresponding to directions for which the classical SM and diagonal unitary precoded schemes are respectively sum-wise catastrophic.

In Figure 7.7, we observe that sum-wise catastrophic codes may achieve a very poor diversity, in agreement with the G_{sum} behavior. By contrast, the newly designed precoded SM scheme offers the same diversity in both uncorrelated and correlated channels, irrespective of the main direction of departure. It also outperforms both codes designed using the virtual channel motivated criterion [HLHS03] (see Section 7.4.3).
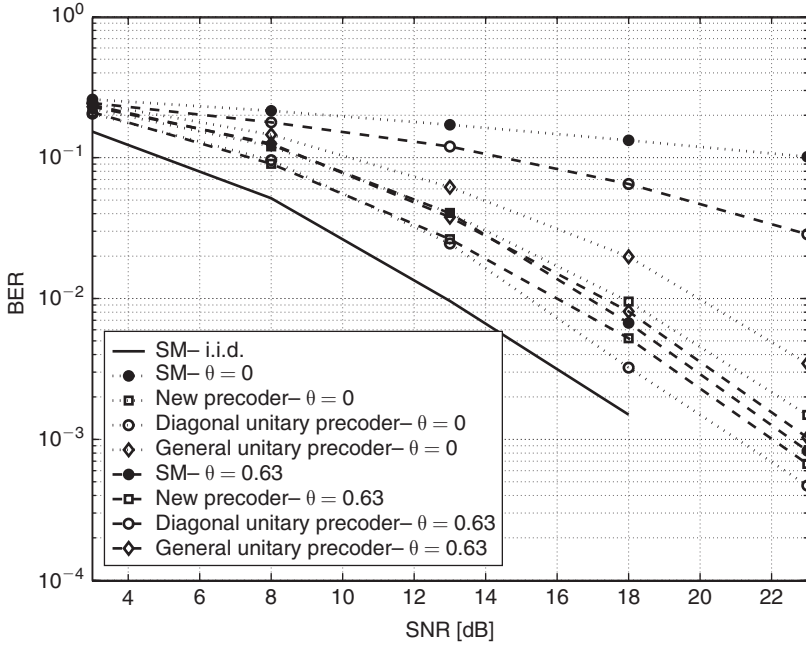
Figure 7.7   *Bit error rate of several 2×2 SM schemes in i.i.d. and correlated Rayleigh fading channels (with two antenna orientations $\theta_t = 0$ and $\theta_t = 0.63$)*

The second design method consists in choosing the best code in terms of G_{sum} among a number of codes performing equally well in i.i.d. Rayleigh fading channels. From Proposition 5.4, we know that a linear STBC minimizes the worst-case average PEP in uncorrelated Rayleigh channels at low SNR or for a large number of receive antennas if the following condition is met:

$$\mathcal{X}^T \mathcal{X} = \frac{T}{Q} \mathbf{I}_{2Q} \tag{7.91}$$

where

$$\mathcal{X} = \left[\text{vec}\left(\begin{bmatrix} \Re(\mathbf{\Phi}_1) \\ \Im(\mathbf{\Phi}_1) \end{bmatrix} \right) \cdots \text{vec}\left(\begin{bmatrix} \Re(\mathbf{\Phi}_{2Q}) \\ \Im(\mathbf{\Phi}_{2Q}) \end{bmatrix} \right) \right] \tag{7.92}$$

The approximation of a large number of receive antennas is mostly adequate for high-rate codes for a realistic number of receive antennas.

Among codes satisfying this property, we may thereby choose those that are robust. As an example, for linear dispersion codes with $T = 2$ and $K = 8$, a large number of matrices χ satisfying (7.91) have been randomly generated, and the code performing best in terms of
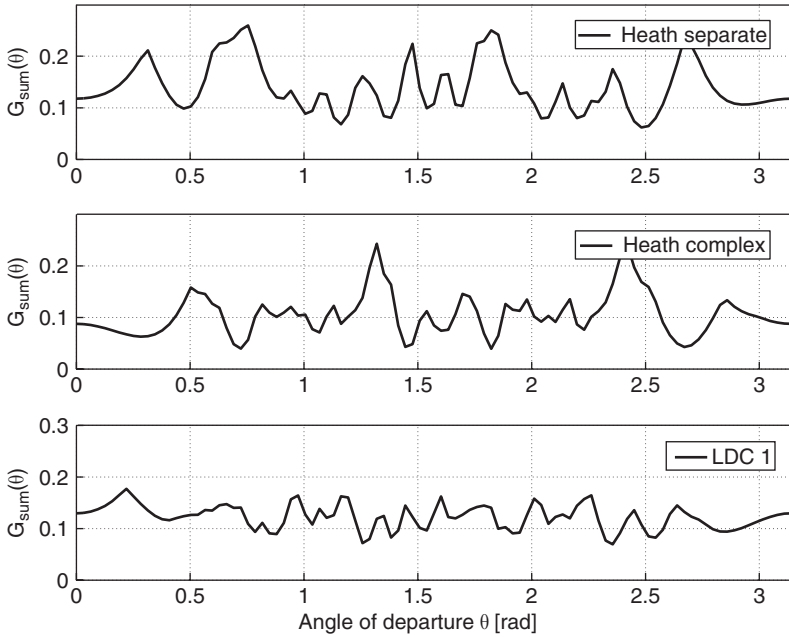
Figure 7.8 $G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$ *of several LDCs as a function of the angle of departure* θ_t *[rad]*

maximizing $G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$ has been selected, and denoted as 'LDC 1'. It is outlined by

$$
\mathcal{X} = \begin{pmatrix}
-0.2640 & 0.3792 & -0.2880 & -0.0563 & -0.1637 & 0.3719 & 0.1481 & 0.1158 \\
0.1335 & -0.2087 & -0.3019 & 0.3454 & 0.3609 & 0.1649 & 0.2608 & 0.0520 \\
0.3760 & -0.2501 & -0.2006 & -0.3419 & -0.2928 & 0.0767 & 0.1540 & 0.1536 \\
-0.1453 & 0.0117 & 0.1888 & 0.1031 & -0.0849 & -0.2804 & 0.3692 & 0.4586 \\
-0.1865 & -0.1891 & 0.2090 & -0.3588 & 0.3579 & 0.2784 & -0.0539 & 0.2202 \\
-0.4110 & -0.2576 & -0.3423 & -0.1805 & 0.0048 & -0.2907 & 0.0571 & -0.1650 \\
-0.0643 & -0.0892 & 0.2883 & -0.0340 & -0.0969 & 0.1683 & 0.4488 & -0.4056 \\
-0.2053 & -0.3741 & 0.0779 & 0.2958 & -0.3354 & 0.2442 & -0.2059 & 0.0989
\end{pmatrix} .
\tag{7.93}
$$

In Figures 7.8 and 7.9, we compare $G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$ of various full-rate LDCs which are all capacity-efficient in i.i.d. Rayleigh fading channels: a robust LDC ('LDC 1'), an arbitrary LDC satisfying (7.91) ('LDC 2') and LDCs presented in Chapter 5 (Hassibi [HH01], Sandhu [San02] and Heath [Hea01]).

On the one hand, the Hassibi and Sandhu codes, as well as 'LDC 2' present the drawback that, in some directions of departure, G_{sum} is quite low (some of them are even catastrophic). This should lead to large error rates if the azimuth power spectrum is concentrated along those directions. On the other hand, 'LDC 1' and both Heath codes should be more robust.
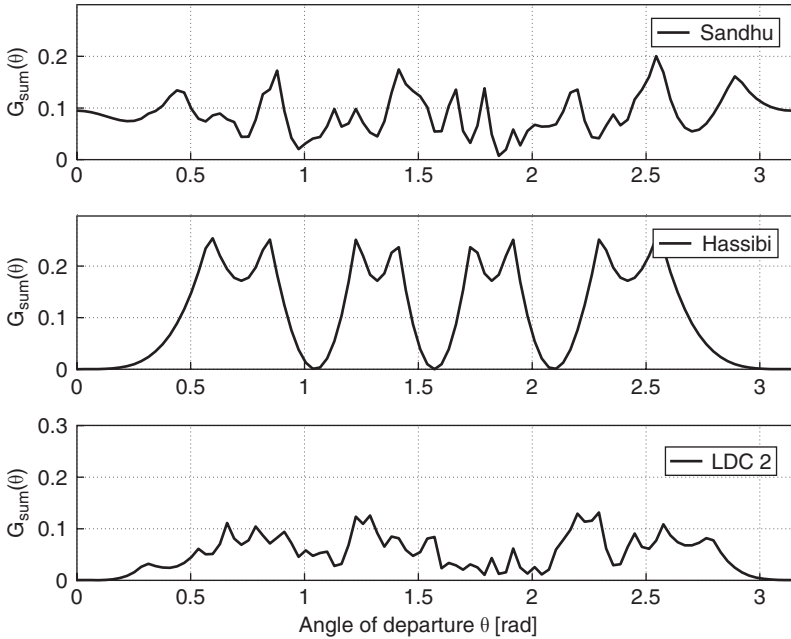
Figure 7.9 $G_{sum}(\theta_t | \mathcal{C}, \mathbf{a}_t(\theta_t))$ *of several LDCs as a function of the angle of departure* θ_t *[rad]*

Comparisons between bit error rates achieved at 20 dB by the classical SM, precoded SM, 'LDC 1', 'LDC 2' and Hassibi code in i.i.d. and correlated Rayleigh fading channels have been depicted in Figures 6.5 and 6.6 considering two receive antennas, and an azimuth power spectrum peaking around zero degree. Once again, the performance matches perfectly the G_{sum} behavior.

By Corollary 7.3, it should also be noted that all previous SM schemes and LDCs perform worse in highly correlated channels than an uncoded SIMO transmission based on a 16-QAM constellation. Indeed, the squared minimum Euclidean distance of a 16-QAM constellation is 1, while for the previous schemes, $G_{sum} << 1$.

Finally, in Figure 7.10, we investigate 3 full rate algebraic codes presented in Section 5.5.7: Yao, Dayal and $B_{2,\phi}$ codes. The Yao and Dayal codes achieve the diversity multiplexing trade-off of Section 5.4.2 (they satisfy the non-vanishing determinant criterion), and are also approximately universal. In terms of Euclidean distance, the Dayal code achieves the largest G_{sum}.

In Figures 6.9 and 6.10, the impact of a coherent component on the performance of rank-deficient codes and full-rank codes has been illustrated, with K-factors of $K = 4$ and $K = 10$, and a peaky azimuth power spectrum aligned with the array baseline. As with Rayleigh fading, the performance perfectly matches the G_{sum} behavior. The SM scheme with the unitary diagonal precoder presents a large G_{sum} in the direction of departure $\theta_t = 0$, which explains why it is able to exploit the coherent component. In the same direction, the
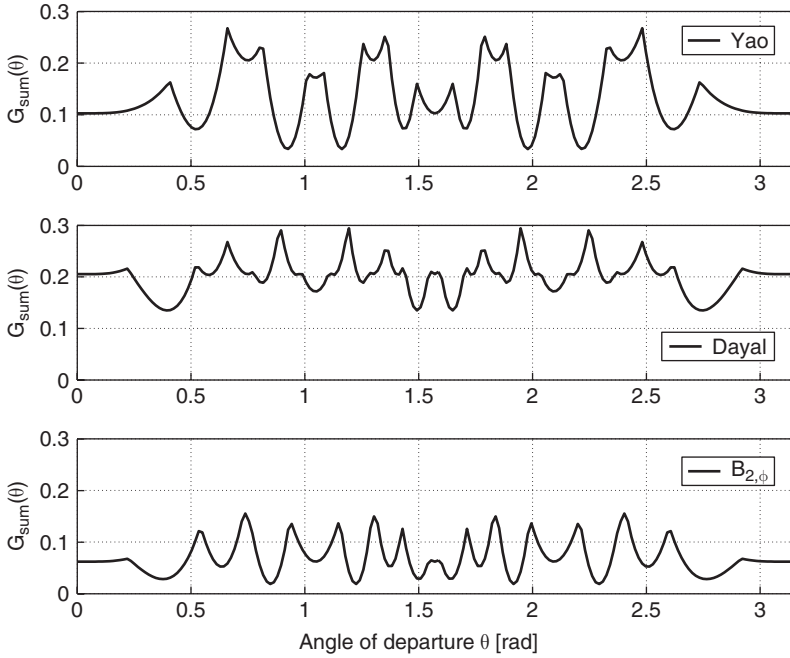
Figure 7.10 $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ *of several LDCs achieving the multiplexing diversity trade-off* [*ZT03*] *as a function of the angle of departure* θ_t [*rad*]

classical SM scheme is sum-wise catastrophic. Regarding approximately universal codes, the Dayal code presents a larger coding gain than the Yao code, thanks to a larger value of G_{sum} in $\theta_t = 0$.

Space–time trellis codes

In i.i.d. Rayleigh fading channels, the assumption of large receive arrays leads to the rank-trace design criterion (see Section 5.3.2). Codes satisfying this criterion have been shown to outperform previously developed codes based on the rank-determinant criterion (see Section 5.3.2). Hence, codes satisfying the trace criterion may be selected and combined with a unitary precoder (denoted as 'New & Prec') to maximize $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$.

Figure 7.11 represents $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ for three QPSK 4-state codes: 'TSC' [TSC98], 'CYV' [CYV01], as well as the combination of 'CYV' with an optimized unitary precoder. We also present four QPSK 8-state codes: 'TSC' [TSC98], 'CYV' [CYV01], an arbitrary trellis code performing similarly to 'CYV' in uncorrelated channels (denoted as 'New') and the combination of the latter with an optimized unitary precoder (denoted as 'New & Prec'). The trellis representations of the 4-state and 8-state codes have been depicted in Figures 5.13 and 5.14 respectively. The optimized precoder for both the 4- and 8-state codes is outlined by

$$\mathbf{P} = \begin{pmatrix} -0.3736 - 0.5965j & -0.6039 + 0.3741j \\ -0.6598 + 0.2630j & -0.2575 - 0.6551j \end{pmatrix}. \tag{7.94}$$
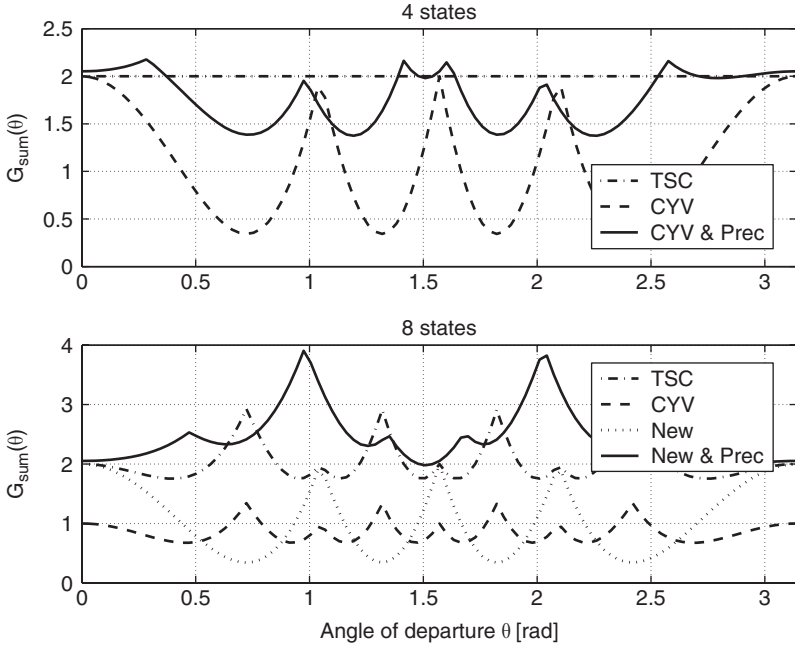
Figure 7.11 $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ *of several 4- and 8-state STTCs as a function of the angle of departure* θ_t

The 4- and 8-state 'TSC' codes offer a large value for G_{sum} in all directions of departure, unlike the 4- and 8-state 'CYV' in some directions. The combined codes using the 'CYV' or the new trellis codes with the unitary precoder present a much larger $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ than the 'CYV' code alone.

For generating Figures 7.12 and 7.13, 4×2 systems are considered in quasi-degenerate channels along directions of departure equal to $\pi/4$ (Figure 7.12) and 2.7π (Figure 7.13). Those two directions correspond to those for which the considered codes present their worse performance. We note that the 'CYV' code, while performing well in uncorrelated channels, critically suffers from transmit correlation. By contrast, the degradation experienced by the 'TSC' code is limited. To better understand this difference, let us take a more intuitive approach.

For delay-diversity schemes, two different codewords originating from the same state and merging in the same state are expressed as

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} d & \dots & f \\ e & \dots & d \end{bmatrix}$$

$$\mathbf{E} = \frac{1}{\sqrt{2}} \begin{bmatrix} d & \dots & h \\ g & \dots & d \end{bmatrix} \tag{7.95}$$
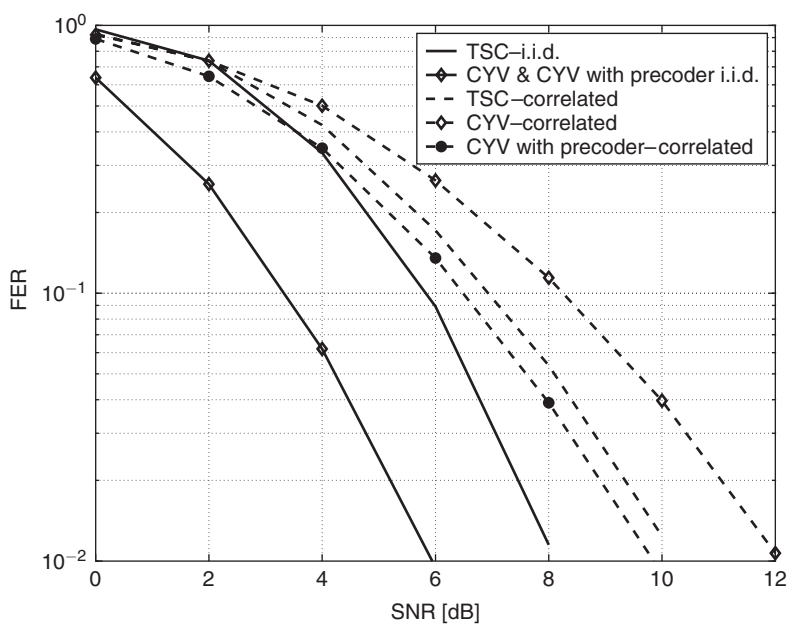
Figure 7.12 *Frame error rate of several 4-state STTCs in i.i.d. and correlated Rayleigh fading channels with $n_t = 2$ and $n_r = 4$*
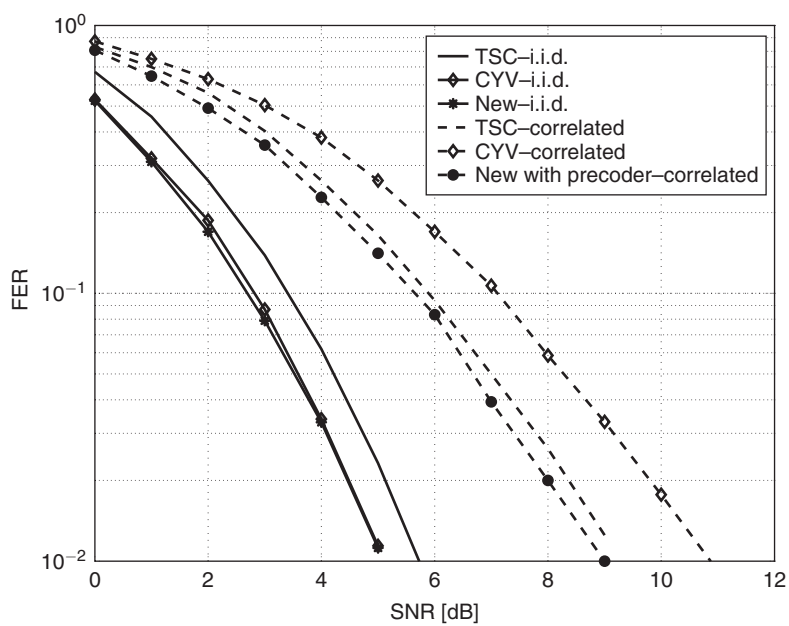
Figure 7.13 *Frame error rate of several 8-state STTCs in i.i.d. and correlated Rayleigh fading channels with $n_t = 2$ and $n_r = 4$*

where d, e, f, g and h are points in the constellation. The error matrix $\mathbf{C} - \mathbf{E}$ is then given by

$$\mathbf{C} - \mathbf{E} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & \cdots & f - h \\ e - g & \cdots & 0 \end{bmatrix} \tag{7.96}$$

with $e \neq g$ and $f \neq h$. Because the following inequality holds true

$$
\begin{aligned}
\|\mathbf{a}_t^T(\theta_t)(\mathbf{C} - \mathbf{E})\|_F^2 &= \sum_{k=0}^{T-1} |\mathbf{a}_t^T(\theta_t)(\mathbf{c}_k - \mathbf{e}_k)|^2 \\
&\geq \sum_{\substack{k \in \tau_{deg}(\theta_t) \\ \sharp \tau_{deg}(\theta_t) = l_{min}(\theta_t)}} |\mathbf{a}_t^T(\theta_t)(\mathbf{c}_k - \mathbf{e}_k)|^2,
\end{aligned}
\tag{7.97}
$$

it is obvious that with delay-diversity for two transmit antennas $l_{min}(\theta_t) = 2 \ \forall \theta_t$ and that

$$\min_{\substack{\mathbf{C}, \mathbf{E} \in \mathcal{C} \\ \mathbf{C} \neq \mathbf{E}}} \|\mathbf{a}_t^T(\theta_t)(\mathbf{C} - \mathbf{E})\|_F^2 = d_{\min}^2 \quad \forall \theta_t \tag{7.98}$$

where d_{min}^2 is the squared minimum Euclidean distance of the constellation. The delay diversity code has the advantage that over a path of error length equal to two, the quantity $\|(\mathbf{c}_k - \mathbf{e}_k)(1)| - |(\mathbf{c}_k - \mathbf{e}_k)(2)\|$ is large, providing a code which is far from being sum-wise catastrophic.

The previous discussion may suggest that the design criteria for i.i.d. and correlated Rayleigh fading channels are contradictory. Indeed, in correlated channels, maximizing the quantity $\min_{\theta_t} \|\mathbf{a}_t^T(\theta_t)(\mathbf{C} - \mathbf{E})\|_F^2$ requires codewords to be of the delay-diversity code type, i.e.

$$\mathbf{C} - \mathbf{E} = \begin{bmatrix} \blacklozenge & \blacklozenge & & \\ & & \blacklozenge & \blacklozenge \end{bmatrix} \tag{7.99}$$

However, the trace or determinant criterion yields codewords of the following form

$$\mathbf{C} - \mathbf{E} = \begin{bmatrix} \blacklozenge & \blacklozenge & \blacklozenge & \blacklozenge \\ \blacklozenge & \blacklozenge & \blacklozenge & \blacklozenge \end{bmatrix} \tag{7.100}$$

suggesting that interactions between codewords transmitted on different antennas always occur.

The precoded 4-state 'CYV' code and the newly developed precoded 8-state code offer the best performance in i.i.d. and correlated channels. Again, this motivates the design of space–time codes using two criteria: one for uncorrelated channels and one for correlated channels.

It is worth noting that the 4- and 8-state 'CYV' codes as well as the 8-state 'new' trellis code achieve the lower bound (7.61) for some directions of departure. While this lower bound

is not changed by a precoder, G_{sum} is highly affected. This explains why even a code with a small lower bound (7.61) is able to perform satisfactorily in correlated channels. The 8-state 'TSC' has a much larger lower bound than the 8-state 'new' code, but its performance in correlated channels is worse.

A QPSK-based uncoded SIMO transmission presenting a minimum squared Euclidean distance equal to 2 is only outperformed by the precoded 8-state 'new' code, whereas all the other codes offer a squared minimum Euclidean distance G_{sum} equal or lower than 2.

7.5 Error probability motivated code design in fast fading channels

7.5.1 Product-wise catastrophic codes

While the previous discussion has mainly been concerned by sum-wise catastrophic codes, let us now derive the conditions a code has to fulfill in order to be product-wise catastrophic. Therefore, we will assume in what follows that the channel is varying sufficiently fast, i.e. the receive-time correlation Ξ_r is full rank. Recalling that $\tau_{\mathbf{C},\mathbf{E}} = \{k | \mathbf{c}_k - \mathbf{e}_k \neq 0\}$, we have the following results.

Lemma 7.2 *In a fast fading channel degenerate in a direction of departure θ_t, the diversity achieved by a pair of codewords $\{\mathbf{C}, \mathbf{E}\}$ differs from the diversity achieved by the same pair of codewords in i.i.d. channels iff $\exists k \in \tau_{\mathbf{C},\mathbf{E}}$ such that $\mathbf{a}_t(\theta_t)$ lies in the null space of $\mathbf{c}_k - \mathbf{e}_k$.*

PROOF: It results from Lemma 7.1 applied to $\mathbf{c}_k - \mathbf{e}_k$. ∎

The following property of product-wise catastrophic codes is thus derived.

Proposition 7.12 *A code is product-wise catastrophic in the direction of departure θ_t iff $\exists n > \sharp\tau_{\mathbf{C},\mathbf{E}} - L_{min}$ positions $k_1, \ldots, k_n$ such that $\mathbf{a}_t(\theta_t)$ lies in the null space of each $\mathbf{c}_k - \mathbf{e}_k$, $k = k_1, \ldots, k_n$.*

The next corollary is useful in order to easily identify product-wise catastrophic codes in the case of two transmit antennas in a balanced array.

Corollary 7.4 *For a balanced two transmit antenna array[10], a code is product-wise catastrophic in a direction $0 \leq \theta_t \leq 2\pi$ if $\exists k \in \tau_{\mathbf{C},\mathbf{E}}$ with $\sharp\tau_{\mathbf{C},\mathbf{E}} = L_{min}$ such that $|(\mathbf{c}_k - \mathbf{e}_k)(1)| = |(\mathbf{c}_k - \mathbf{e}_k)(2)|$.*

---

[10] We will assume that φ in (6.4) may vary over 2π. For a linear array, this requires that $d/\lambda \geq 0.5$.

PROOF: The condition $|(\mathbf{c}_k - \mathbf{e}_k)(1)| = |(\mathbf{c}_k - \mathbf{e}_k)(2)|$ is just another expression of the fact that $\mathbf{a}_t(\theta_t)$ lies in the null space of $\mathbf{c}_k - \mathbf{e}_k$ for a balanced transmit array when φ may vary over 2π. It directly results from the proof of Proposition 7.9.   ∎

7.5.2   Practical code designs in fast fading channels

If the environment is fast fading, a well-designed code should ideally exploit the temporal diversity offered by the channel. Hence, we choose among the set of codes $\mathcal{C}_{i.i.d.}$ performing well in i.i.d. fast fading channels those that are also robust in correlated channels. As above, we resort to the optimization of a unitary precoder maintaining $l_{min}(\theta_t) = L_{min}$ for all possible θ_t and so maximizing $\min_{\theta_t} G_{product}$ (Criterion 7.4). As already mentioned, it is also preferable to ensure that $\min_{\theta_t} G_{sum}$ is large in order to guarantee a low error rate in Ricean fading channels.

Regarding the code design, we focus again on a horizontal uniform linear array using two transmit antennas and QPSK modulation, with a transmit antenna spacing larger than $0.5\,\lambda$ (so that the minimization is carried out for φ varying over 2π).

We consider in the sequel various QPSK 4-state space–time trellis codes in i.i.d. and spatially correlated channels: 'TSC' [TSC98], 'BBH' [BBH00], 'CYV' [CYV01] and 'FVY' [FVY01]. The trellis representations have been depicted in Figure 5.13. For simplicity, let us consider perfectly time interleaved fading channels.

In Figures 7.14 and 7.15, $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ and $G_{product}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ are both represented for the aforementioned 4-state codes, as well as for two new precoded schemes. The first one ('CYV & Prec slow') consists of the 'CYV' code combined with the unitary precoder designed for slow fading channels in (7.94) (i.e. optimized only to maximize $\min_{\theta_t} G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$). The second scheme ('CYV & Prec fast') combines the 'CYV' code with a unitary precoder designed for fast fading channels (i.e. optimized following Criterion 7.4). This precoder is given by

$$\mathbf{P} = \begin{pmatrix} -0.1436 - 0.8340j & 0.5233 - 0.0997j \\ -0.5262 - 0.0833j & -0.1174 + 0.8381j \end{pmatrix}. \qquad (7.101)$$

The reason why the precoder is optimized for the 'CYV' code is that the latter achieves excellent performance in i.i.d. Rayleigh slow and fast fading channels. While this is obvious in slow fading channels (since it was designed for slow fading channels [CYV01]), it can be shown that it also performs in fast fading channels as well as the 'FVY' code [FVY01] (specifically designed for fast fading channels). The 'FVY' and 'CYV' codes have indeed the same minimum product distance in i.i.d. fast fading Rayleigh channels.

Figures 7.14 and 7.15 illustrate that the 'TSC' code presents an excellent robustness in degenerate channels in both slow and fast fading channels since both G_{sum} and $G_{product}$ are large for all directions of departure. Reasoning as in (7.95) and (7.96), since $l_{min}(\theta_t) = 2$,
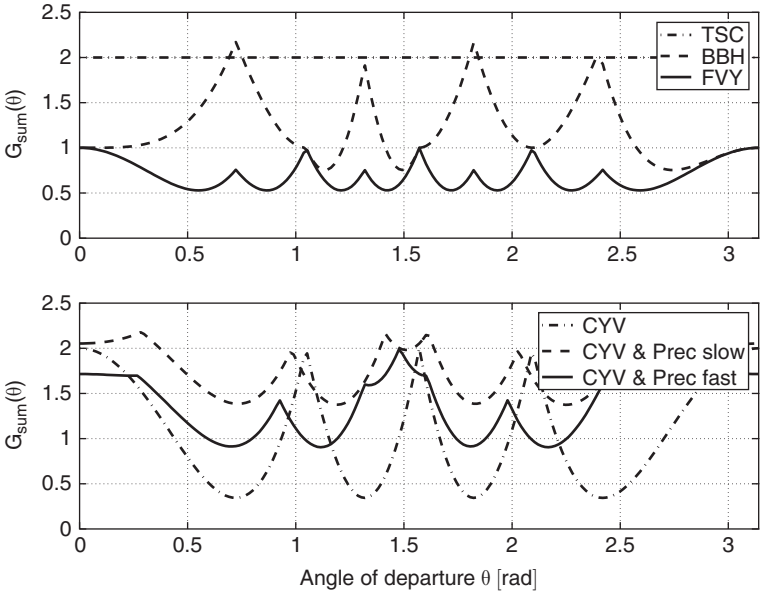
Figure 7.14 $G_{sum}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ of several 4-state STTCs as a function of the angle of departure θ_t
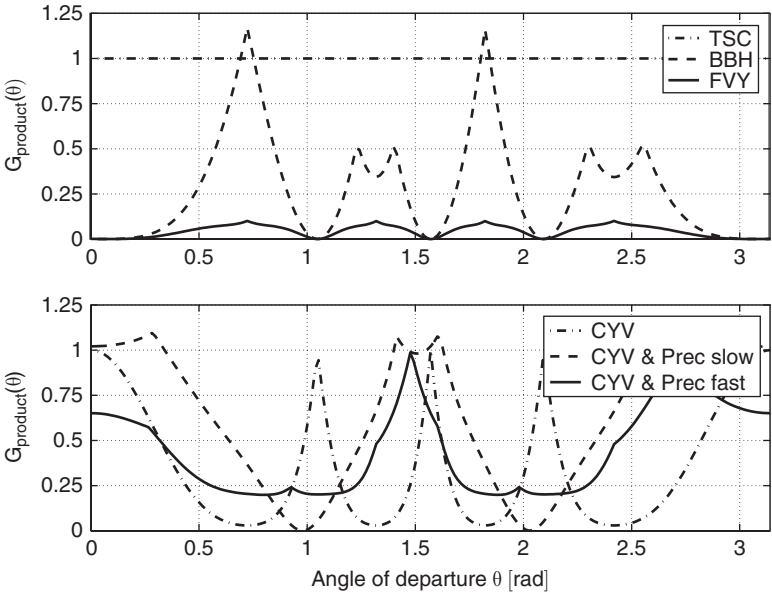
Figure 7.15 $G_{product}(\theta_t|\mathcal{C}, \mathbf{a}_t(\theta_t))$ of several 4-state STTCs as a function of the angle of departure θ_t
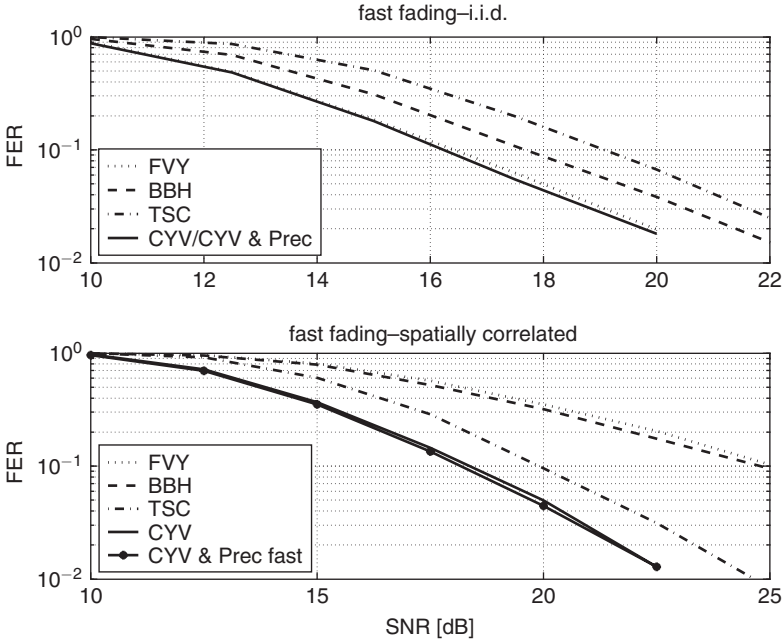
Figure 7.16 *Frame error rate of several 4-state STTCs in i.i.d. and correlated Rayleigh fast fading channels with $n_t = 2$ and $n_r = 1$*

we have $\forall \theta_t$

$$\min_{\substack{\mathbf{C},\mathbf{E}\in\mathcal{C} \\ \mathbf{C}\neq\mathbf{E}}} \prod_{\substack{k\in\mathcal{V}_{deg}(\theta_t) \\ \sharp\mathcal{V}_{deg}(\theta_t)=l_{min}(\theta_t)}} |(\mathbf{c}_k - \mathbf{e}_k)^T\mathbf{a}_t(\theta_t)|^2 = \frac{1}{4}d_{min}^4, \quad \forall\theta_t \qquad (7.102)$$

where d_{min}^2 is the minimum squared Euclidean distance of the constellation. Both G_{sum} and $G_{product}$ are thus flat as a function of θ_t. The delay diversity code has the advantage that, over a path of error length equal to 2, $\|(\mathbf{c}_k - \mathbf{e}_k)(1)| - |(\mathbf{c}_k - \mathbf{e}_k)(2)\|$ is large, leading to a code that is far from being product-wise catastrophic.

By contrast, for the 'BBH', 'CYV' and 'FVY' codes, G_{sum} and/or $G_{product}$ are low. The 'BBH' and 'FVY' codes are even product-wise catastrophic in some directions. This can also be verified from Corollary 7.4, since there exist some error matrices that are such that $|(\mathbf{c}_k - \mathbf{e}_k)(1)| = |(\mathbf{c}_k - \mathbf{e}_k)(2)|$. The 'CYV & Prec slow' code (designed for slow fading scenarios) presents a larger G_{sum} than the 'CYV' code, the values of $G_{product}$ are still very low, leading to large error rates in degenerate fast fading channels. Finally, the 'CYV & Prec fast' code is neither sum-wise nor product-wise catastrophic, and should thereby be robust against fading correlations in both slow and fast fading scenarios.

In Figure 7.16, the frame error rates (FERs) achieved by the various trellis codes in 1×2 i.i.d. and correlated channels are represented (each frame consists of 128 symbols out of

each transmit antenna). The azimuth power spectrum is close to a delta function around 0 degree. On the one hand, the performance of the 'BBH' and 'FVY' codes is severely degraded. As predicted by Proposition 7.7, the diversity achieved by these codes is equal to $l_{min}(0) = 1$. On the other hand, the degradation observed for the 'TSC' code is practically invisible. This could be predicted from the value of $G_{product}$, which equals the minimum product distance achieved in i.i.d. Rayleigh fading channel. The reason for the slight loss is caused by a larger number of code error matrices with a path length of two in the correlated case, resulting in a modified distance spectrum (components $\mathbf{c}_k - \mathbf{e}_k$ for some $k \in \tau_{\mathbf{C},\mathbf{E}}$ are such that $|(\mathbf{c}_k - \mathbf{e}_k)(1)| = |(\mathbf{c}_k - \mathbf{e}_k)(2)|$ for error events presenting a length $\sharp \tau_{\mathbf{C},\mathbf{E}} > 2$). As expected, the precoded 4-state 'CYV & Prec fast' code is robust.

Note that the 'TSC' code has a lower performance than the 'CYV' code in the direction 0 whereas the values of $G_{product}$ are similar. This is due to the fact that the minimum product distance of the 'CYV' code is larger than the minimum product distance of the 'TSC' code. Indeed, we know from Chapter 6 that the performance of a code is a function of the product distance and of the interactions between the channel and the codewords. Those interactions are totally accounted for through $G_{product}$. When the channel becomes close to degenerate, the main factor affecting the performance is $G_{product}$. By contrast, when the scattering richness is high, the relevant parameter is the product distance. Therefore, it is much better for a code to present a large product distance and a large value of $G_{product}$. Since the 'CYV' code has a larger product distance and a similar value for $G_{product}$ around 0 degree, its performance is improved.

In practice, it may not always be possible to optimize a unitary precoder for a given space–time code, especially if the number of states or the constellation are large. In these cases, a product-wise catastrophic code may be converted into a non product-wise catastrophic code by means of the unitary precoder optimized for a Spatial Multiplexing scheme using the same constellations (see Table 7.1). While not optimal, this method may significantly improve the code performance in degenerate fast fading channels.

This page intentionally left blank

8

Space–time coding with partial transmit channel knowledge

In Chapter 1, we have outlined that MIMO systems essentially deliver four gains: diversity gain, spatial multiplexing gain, coding gain and array gain. So far, we have extensively covered the first three gains.

- In Chapters 1 and 5, we have shown how space–time coding architectures could exploit spatial multiplexing and diversity gains in i.i.d. Rayleigh fading channels.
- In Chapter 4, we have illustrated via the mutual information and the diversity-multiplexing trade-off that those gains are significantly affected by the channel correlations and the K-factor under a uniform input power allocation.
- In Chapter 6, we have formalized the impact of channel correlations and K-factor on the error performance of codes obtained in Chapter 5 for i.i.d. Rayleigh fading channels, with the result that codes maximizing the spatial multiplexing gain are more affected than codes maximizing diversity.
- In Chapter 7, design criteria have been developed to construct robust codes with respect to propagation conditions, without requiring any transmit channel knowledge.

Yet, improving the exploitation of the array gain performance is still possible if the transmitter has some channel knowledge. As already evoked in Chapter 1, channel knowledge is the key to array gain. At the receiver, where perfect channel knowledge is assumed throughout this book, Chapter 1 has detailed the calculation of the array gain for several schemes. At the transmitter, the array gain is a further function of the amount of channel knowledge available. This channel knowledge may be of various kinds. For instance, we may have no channel knowledge at all. This problem has been dealt with in Chapters 5 and 7. In contrast, full (perfect) channel knowledge at the transmitter is naturally the best solution. In Chapters 1 and 4, we have evaluated what could be gained by designing precoders based on perfect CSIT to maximize the mutual information or the received SNR. Such precoders yield a larger capacity, extract the full diversity and achieve large array gains when the total transmit power is constrained. Moreover, the overall system complexity may be reduced by using linear precoding/decoding and by the decoupling properties of certain codes providing parallel virtual transmissions. Note that CSIT based precoders may also

rely on MMSE criteria, see for example [SSB$^+$02, SSP01]. Naturally, exploiting perfect CSIT might sometimes be hardly achievable, especially when the channel varies rapidly, because it requires a high-rate feed back link between the receiver and the transmitter. As an intermediate solution, we find schemes that are based on channel statistics only or that exploit a limited amount of feed back at the transmitter (e.g. a quantized version of the channel). These schemes are covered in this chapter.

Beyond the type of channel knowledge, other factors impact the design of schemes using partial channel knowledge, namely the transmit signaling scheme, the receiver detector, the performance criterion and the transmit power constraint. For both types of channel knowledge (channel statistics and limited feed back), we discuss in this chapter the design of precoders for dominant eigenmode transmission, spatial multiplexing and O-STBC, using either ML or suboptimal detectors. As far as criteria are concerned, both information theory and error probability motivated performance criteria are addressed. Finally, two transmit power constraints are considered. The first is the classical total transmit power constraint (also known as the sum power constraint). An alternative constraint consists in limiting the peak-to-average power ratio. While the former has been used extensively throughout this book, the latter is useful in practical scenarios where keeping the operation point in the linearity range of the amplifiers is especially important.

Exploiting channel statistics at the transmitter

The advantage of exploiting the channel statistics at the transmitter instead of the full CSI is that it does only require a low-rate feed back link to keep the transmitter informed. Indeed, the statistical properties of the channel (correlations, K-factor) vary at a much slower rate than the fading channel itself. The receiver may thus estimate the channel stochastic properties and then send them back to the transmitter 'once in a while'. As an example, we have illustrated in Chapter 3 that, for indoor downlink scenarios, the transmit correlation matrix remains essentially unchanged when the receiver moves over a few meters. It is also important to realize that, for a critical antenna spacing, the phase of a high transmit correlation is directly linked to the direction where the transmit direction spectrum is maximum, hence the knowledge that the transmit correlation enables the transmitter to somehow stir a beam into this direction and so to maximize the power transmission through the channel.

It is actually not the first time that we deal with stochastic (or partial) transmit channel knowledge: in Chapter 4, we have derived the capacity-wise optimal covariance matrix and we have shown that the capacity could be substantially improved (depending on the SNR and the propagation conditions) through a precoding along preferred directions in combination with appropriate power allocation, i.e. via matrix $\mathbf{Q}$. Ultimately, we have observed that the optimal allocation is sometimes to transmit the total available power into a unique direction. This technique is known as beamforming and exploits the transmit array gain (since the average received power is increased). In this chapter, we further investigate how to exploit the transmit array gain by designing practical precoders minimizing the error probability.

A brief word about stationarity. Sending back the channel statistical properties to the transmitter relies on the channel stationarity. While strict stationarity (i.e. the fact that the statistical properties of the channel do not change with time) is not encountered in real-world scenarios, the wide sense stationarity assumption is much more practical (see Chapter 2): in this case, the correlation function of the time-variant channel response is only the function of the time difference. If this only holds true over limited time intervals, the channel is referred to as quasi wide sense stationary.

Exploiting a limited amount of feed back at the transmitter

There are essentially two approaches to design schemes using a limited feed back. The first method consists in directly quantizing the channel and feeding back the quantized version of the channel to the transmitter. The quantized channel structure can be exploited in schemes designed for perfect CSIT. The second method relies on a codebook of precoding matrices, i.e. a finite set of precoders, designed off-line and known to both the transmitter and receiver. The receiver estimates the best precoder as a function of the current channel and feeds back only the index of this best precoder in the codebook. Intuitively, we can consider each precoder as being more adapted to a particular set of channel realizations. The first method usually leads to a larger amount of feed back (each complex entry of the channel structure must be quantized) and the performance is not necessarily better than with the second method. Moreover, it is quite sensitive to quantization errors, which is not the case with the latter. Indeed, since the receiver chooses the best precoder based on the perfect channel state information, the eigenstructure of the channel is not compromised during the quantization, making the second method more robust [LHS03, LH05a, LH05b].

8.1 Introduction to channel statistics based precoding techniques

8.1.1 A general framework

Precoding techniques are particularly attractive as they keep the complexity low. That explains why they are currently integrated in many standards, such as IEEE 802.16e and 802.11n.

Analogous to Chapter 7, we consider that the transmitter results from the concatenation of an encoder, whose output is a codeword $\mathbf{C}$, and of a precoder $\mathbf{P}$ that processes the codeword $\mathbf{C}$ and transmits $\mathbf{C}' = \mathbf{PC}$. Various criteria can be used to design $\mathbf{P}$, but similarly to previous chapters, we focus here on information theory and error probability motivated design methodologies. A general precoding structure is depicted in Figure 8.1. The linear precoder $\mathbf{P} = \mathbf{U_P \Sigma_P}$ essentially consists in two steps:

- a multi-mode beamformer spreads $\mathbf{C}''$ in orthogonal directions related to the channel state or distribution, via a unitary matrix $\mathbf{U_P}$, so that $\mathbf{C}' = \mathbf{U_P C''}$
- a constellation shaper transforms codeword $\mathbf{C}$ into codeword $\mathbf{C}''$ by allocating higher power and rate to selected modes of $\mathbf{U_P}$ so that $\mathbf{C}'' = \mathbf{\Sigma_P C}$ (if $\mathbf{\Sigma_P}$ is real-valued
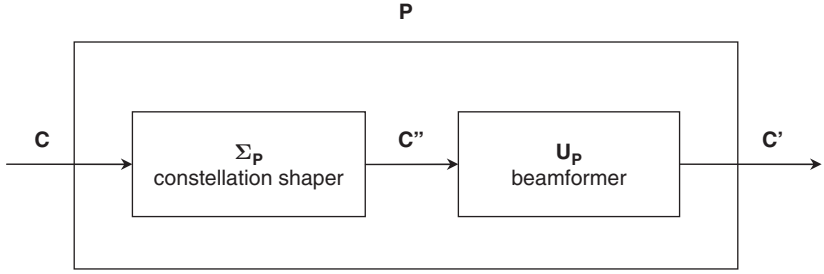
Figure 8.1 *Overview of precoder* **P***:* **U$_P$** *acts as a multi-mode beamformer,* **C″** *is the codeword being shaped by* **Σ$_P$**

and diagonal, $\Sigma_{\mathbf{P}}^2 = \boldsymbol{\Lambda}_{\mathbf{P}}$ can be thought of as the power allocation scheme across the modes).

The precoder is always normalized in order to satisfy a power constraint. The classical sum power constraint translates into $\mathcal{E}\{\mathrm{Tr}\{\mathbf{C}'\mathbf{C}'^H\}\} = T$. Assuming that the encoded codewords **C** are normalized as $\mathcal{E}\{\mathrm{Tr}\{\mathbf{C}\mathbf{C}^H\}\} = T$ (see Chapter 5), the precoder must be normalized as

$$\mathrm{Tr}\{\mathbf{P}\mathbf{P}^H\} = n_t. \tag{8.1}$$

8.1.2 Information theory motivated design methodologies

From an information theory point of view, we know from Chapter 4 that an identity input covariance matrix $\mathbf{Q} = \mathbf{I}_{n_t}$ maximizes the average mutual information in i.i.d. Rayleigh fading channels. In correlated channels, the optimal **Q**, based on partial transmit channel knowledge (CDIT), is generally colored. Such an input covariance matrix is easily realized without loss in capacity by splitting the full space–time encoder with CSIT/CDIT into a space–time code itself and a precoder which is a function of the feed back data [SJ03]. The codebook of the space–time code is drawn from a Gaussian distribution with i.i.d. components, hence the precoder colors the input covariance matrix of the space–time code to obtain the capacity-achieving input covariance matrix **Q**. If the codewords **C** are i.i.d. Gaussian distributed $\left(\mathcal{E}\{\mathbf{C}\mathbf{C}^H\} = \mathbf{I}_{n_t}/n_t\right)$, the precoder is given by $\frac{1}{n_t}\mathbf{P}\mathbf{P}^H = \mathbf{Q} = \mathbf{U}_{\mathbf{Q}}\boldsymbol{\Lambda}_{\mathbf{Q}}\mathbf{U}_{\mathbf{Q}}^H$. Without loss of generality, we may write $\mathbf{U}_{\mathbf{P}} = \mathbf{U}_{\mathbf{Q}}$ and $\boldsymbol{\Lambda}_{\mathbf{P}} = n_t\boldsymbol{\Lambda}_{\mathbf{Q}}$. Note that the constellation shaper is defined up to a unitary matrix, which we take here as the identity. This unitary matrix has no impact if the code is i.i.d. Gaussian distributed. However, if the code is drawn from a finite alphabet, this unitary matrix may significantly affect the error performance (typically, if the precoder is designed according to a PEP criterion).

Let us now summarize what are the optimal input covariance matrices derived in Chapter 4 for several propagation scenarios.

- In semi-correlated Rayleigh channels with transmit correlation only, $\mathbf{Q} = \mathbf{U}_{\mathbf{R}_t}\boldsymbol{\Lambda}_{\mathbf{Q}}\mathbf{U}_{\mathbf{R}_t}^H$, hence we may consider that the beamformer is $\mathbf{U}_{\mathbf{P}} = \mathbf{U}_{\mathbf{R}_t}$ and the constellation shaper $\boldsymbol{\Sigma}_{\mathbf{P}} = \boldsymbol{\Lambda}_{\mathbf{P}}^{1/2}\sqrt{n_t}\boldsymbol{\Lambda}_{\mathbf{Q}}^{1/2}$ is diagonal and acts as a power allocation scheme, favoring the

directions of $\mathbf{R}_t$ corresponding to large eigenvalues. This optimal power allocation requires a numerical optimization. For example, if such a precoder is used in front of an Alamouti scheme, previous results and discussions in Chapter 5 indicate that this precoding scheme will not induce any capacity loss if $n_r = 1$. An approximate power allocation obtained thanks to Jensen's inequality is given by a so-called stochastic waterfilling

$$\lambda_k(\mathbf{Q}) = \left[\mu - \frac{1}{\rho n_r \lambda_k(\mathbf{R}_t)} \right]^+. \tag{8.2}$$

- In correlated channels with separable Rx and Tx correlations, beamforming is also performed in the direction given by the eigenvectors of $\mathbf{R}_t$, and the power allocation is also determined through numerical computation (though there is no intuitive approximation).
- In Ricean channels for which the coherent component is of rank one and the Rayleigh component is i.i.d. distributed, the dominant eigenvector of $\mathbf{Q}$ is equal to the transposed conjugate of the first row of $\bar{\mathbf{H}}$ normalized to a unit norm. The other columns of $\mathbf{Q}$ are chosen such that $\mathbf{U}_\mathbf{Q}$ is unitary. The dominant eigenvalue of $\mathbf{Q}$, $\lambda_1(\mathbf{Q})$ is determined numerically, while the remnant eigenvalues are equal and given by $(1 - \lambda_1(\mathbf{Q}))/(n_t - 1)$.
- In Ricean channels for which the rank of the coherent contribution is larger than one or the Rayleigh contribution is correlated, no analytical solution exists for $\mathbf{Q}$. However, an approximation derived from Jensen's inequality yields that $\mathbf{U}_\mathbf{Q}$ is chosen as the eigenvectors of $\mathcal{E}\{\mathbf{H}^H\mathbf{H}\}$ and $\mathbf{\Lambda}_\mathbf{Q}$ is simply obtained by the water-filling solution applied on the eigenvalues of $\mathcal{E}\{\mathbf{H}^H\mathbf{H}\}$.

8.1.3 Error probability motivated design methodologies

As an alternative to information theory motivated approaches, the minimization of the maximum average PEP or the average SER may also be an objective to design precoders.

The average PEP in correlated Ricean slow fading channels in the presence of a precoder $\mathbf{P}$ is easily obtained from (6.15) by expressing the error matrix as $\mathbf{P}\tilde{\mathbf{E}}\mathbf{P}^H$

$$P(\mathbf{C} \rightarrow \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} [\det(\mathbf{I}_{n_r n_t} + \eta \mathbf{C}_\mathbf{R})]^{-1}$$
$$\exp\left[-\eta K \text{vec}(\bar{\mathbf{H}}^H)^H (\mathbf{I}_{n_r} \otimes \mathbf{P}\tilde{\mathbf{E}}\mathbf{P}^H)(\mathbf{I}_{n_r n_t} + \eta \mathbf{C}_\mathbf{R})^{-1} \text{vec}(\bar{\mathbf{H}}^H) \right] d\beta \tag{8.3}$$

with $\mathbf{C}_\mathbf{R}$ now given by $\mathbf{C}_\mathbf{R} = \mathbf{R}(\mathbf{I}_{n_r} \otimes \mathbf{P}\tilde{\mathbf{E}}\mathbf{P}^H)$ and $\eta = \rho/(4(1+K)\sin^2 \beta)$. The most classical precoder design relies on minimizing the maximum PEP, which finds the optimal precoder $\mathbf{P}^\star$ satisfying

$$\mathbf{P}^\star = \arg \min_\mathbf{P} \max_{\tilde{\mathbf{E}} \neq \mathbf{0}} P(\mathbf{C} \rightarrow \mathbf{E}). \tag{8.4}$$

Proposing a unified theory of precoder design applicable to arbitrary space–time codes is a grueling task, mostly because the precoder design is related to the error matrix given by $\max_{\tilde{\mathbf{E}} \neq \mathbf{0}} P(\mathbf{C} \rightarrow \mathbf{E})$, itself depending on the shape of $\mathbf{P}$. Therefore, both $\max_{\tilde{\mathbf{E}} \neq \mathbf{0}} P(\mathbf{C} \rightarrow \mathbf{E})$

and **P** are mutually coupled. As a consequence, the next sections deal with the design of precoders for specific space–time codes such as orthogonal space–time block codes and Spatial Multiplexing schemes. Furthermore, the following designs are also often limited to Kronecker Rayleigh fading channels, although a few results for non-Kronecker or non-Rayleigh fading channels will be reviewed whenever possible.

8.2 Channel statistics based precoding for orthogonal space–time block coding

The design of precoders for O-STBCs has been first studied by [JSO02], assuming imperfect CSIT at the transmitter. Around the same time, several results regarding the CDIT-design of precoders have been reported in [SP02, ZG03, FADT02, HGA06, HG07b], using the channel mean or the channel correlations as partial transmit channel knowledge. However, the various designs differ in the considered channel models and criteria (worst-case PEP, Chernoff bound on PEP, exact SER). More recently, those results have been generalized to the precoding design using joint channel mean and channel correlation feed back [VP06, HG07a].

In Section 5.5.4, we have observed that the error matrices of O-STBCs may be written as $\tilde{\mathbf{E}} = Td^2/(Qn_t)\mathbf{I}_{n_t}$ where $d^2 = \sum_{q=1}^{Q} |c_q - e_q|^2$ is an Euclidean distance and Q/T is the spatial multiplexing rate of the code. From (8.3), the average PEP of precoded O-STBCs in Ricean fading channels reads as

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \mathcal{O}\left(-\frac{d^2}{\sin^2 \beta}\right) d\beta \qquad (8.5)$$

where

$$\mathcal{O}\left(-\frac{d^2}{\sin^2 \beta}\right) = [\det(\mathbf{I}_{n_r n_t} + \zeta\mathbf{C_R})]^{-1}$$
$$\exp\left[-\zeta K \text{vec}(\bar{\mathbf{H}}^H)^H (\mathbf{I}_{n_r} \otimes \mathbf{P}\mathbf{P}^H)(\mathbf{I}_{n_r n_t} + \zeta\mathbf{C_R})^{-1}\text{vec}(\bar{\mathbf{H}}^H)\right] \qquad (8.6)$$

with $\mathbf{C_R} = \mathbf{R}(\mathbf{I}_{n_r} \otimes \mathbf{P}\mathbf{P}^H)$ and $\zeta \triangleq \eta Td^2/(Qn_t)$.

Because $\tilde{\mathbf{E}}$ is proportional to the identity, the design of optimal precoders for O-STBCs becomes very simple. Indeed for O-STBCs, $\max_{\tilde{\mathbf{E}}\neq\mathbf{0}} P(\mathbf{C} \to \mathbf{E})$ is reached as soon as $d = d_{min}$ (the minimum squared Euclidean distance of the constellation), irrespective of **P**, the SNR, the channel correlation matrix **R** and the coherent component $\bar{\mathbf{H}}$. Hence, the optimal precoder minimizing the maximum PEP is now simply given by

$$\mathbf{P}^\star = \arg\min_{\mathbf{P}} P(\mathbf{C} \to \mathbf{E}), \qquad (8.7)$$

where d is fixed to d_{min} in the expression (8.5) of the PEP.

Furthermore, orthogonal space–time block coding decouples the signals at the receiver, thereby allowing for the direct evaluation of the SER in i.i.d. channels (see Chapter 5). In correlated Rayleigh channels, it is also possible to obtain the expressions of the average SER for M-PSK (denoted as $\bar{P}_{PSK}$), M-PAM (denoted as $\bar{P}_{PAM}$) and M-QAM (denoted as $\bar{P}_{QAM}$) [SA00, HG07a]:

$$\bar{P}_{PSK} = \frac{1}{\pi} \int_0^{\frac{M-1}{M}\pi} \mathcal{O}\left(-\frac{d^2_{min,PSK}}{\sin^2\beta}\right) d\beta \tag{8.8}$$

$$\bar{P}_{PAM} = \frac{2}{\pi}\frac{M-1}{M} \int_0^{\frac{\pi}{2}} \mathcal{O}\left(-\frac{d^2_{min,PAM}}{\sin^2\beta}\right) d\beta \tag{8.9}$$

$$\bar{P}_{QAM} = \frac{4}{\pi}\frac{\sqrt{M}-1}{\sqrt{M}} \left[\frac{1}{\sqrt{M}} \int_0^{\frac{\pi}{4}} \mathcal{O}\left(-\frac{d^2_{min,QAM}}{\sin^2\beta}\right) d\beta + \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \mathcal{O}\left(-\frac{d^2_{min,QAM}}{\sin^2\beta}\right) d\beta\right] \tag{8.10}$$

where $d^2_{min,PSK} = 4\sin^2\left(\frac{\pi}{M}\right)$, $d^2_{min,PAM} = \frac{12}{M^2-1}$ and $d^2_{min,QAM} = \frac{6}{M-1}$ are the minimum squared Euclidean distance of M-PSK, M-PAM and M-QAM, respectively. The above equations are very similar to (8.5), suggesting that designs relying on the PEP or the SER may result from similar approaches.

8.2.1   Optimal precoding in Kronecker Rayleigh fading channels

Beamforming

In Rayleigh fading channels, (8.5) and (8.8) to (8.10) are functions of terms reading as

$$\mathcal{G}(\zeta) = \int_{\beta_1}^{\beta_2} \left[\det\left(\mathbf{I}_{n_r n_t} + \zeta\mathbf{R}\left(\mathbf{I}_{n_r} \otimes \mathbf{P}\mathbf{P}^H\right)\right)\right]^{-1} d\beta \tag{8.11}$$

where $\zeta = \eta T \delta^2/(Qn_t)$ and $\delta = d_{min}$, $d_{min,PSK}$, $d_{min,PAM}$ or $d_{min,QAM}$. The structure of the optimal precoder in Kronecker Rayleigh fading channels ($\mathbf{R} = \mathbf{R}_r \otimes \mathbf{R}_t$) is then outlined by the following proposition [SP02, VP06, HG07b].

Proposition 8.1 *In Kronecker Rayleigh fading channels, the optimal precoder minimizing the average PEP/SER is given by $\mathbf{P} = \mathbf{U}_{\mathbf{R}_t}\boldsymbol{\Sigma}_{\mathbf{P}}$ where*

- $\mathbf{U}_{bf\mathbf{R}_t}$ *is the eigenvector matrix of $\mathbf{R}_t$, i.e. $\mathbf{R}_t = \mathbf{U}_{\mathbf{R}_t}\boldsymbol{\Lambda}_{\mathbf{R}_t}\mathbf{U}_{\mathbf{R}_t}^H$,*
- $\boldsymbol{\Sigma}_{\mathbf{P}} = \mathbf{D}_{\mathbf{P}}\mathbf{V}_{\mathbf{P}}^H$, $\mathbf{D}_{\mathbf{P}}$ *being a real-valued diagonal matrix accounting for the power allocation and $\mathbf{V}_{\mathbf{P}}$, an arbitrary unitary matrix.*

PROOF: The proof consists in showing that the optimal precoder achieves the lower bound on $\mathcal{G}(\zeta)$. From Hadamard's inequality, the determinant of a non-singular matrix is upper-bounded by the product of its diagonal elements (see Appendix A). Equality naturally

occurs if and only if the matrix is diagonal. We assume that the channel correlation is Kronecker-separable and apply Hadamard's inequality to $\mathcal{G}(\zeta)$. Since the inequality is valid for all β within the integration interval $[\beta_1, \beta_2]$, the inequality also holds true at the integral level, so that

$$\mathcal{G}(\zeta) \geq \int_{\beta_1}^{\beta_2} \prod_{k=1}^{n_r} \Big[\det \Big(\mathrm{diag}\Big(\mathbf{I}_{n_t} + \zeta\lambda_k(\mathbf{R}_r)\mathbf{\Lambda}_{\mathbf{R}_t}^{1/2}\mathbf{U}_{\mathbf{R}_t}^H \mathbf{P}\mathbf{P}^H \mathbf{U}_{\mathbf{R}_t} \mathbf{\Lambda}_{\mathbf{R}_t}^{1/2}\Big)\Big)\Big]^{-1} d\beta. \quad (8.12)$$

Equality occurs if and only if the argument of the determinant operator is diagonal for all $\beta \in [\beta_1, \beta_2]$, i.e. if and only if $\mathbf{U}_{\mathbf{R}_t}^H \mathbf{P}\mathbf{P}^H \mathbf{U}_{\mathbf{R}_t}$ is diagonal. Choosing $\mathbf{P}\mathbf{P}^H = \mathbf{U}_{\mathbf{R}_t} \mathbf{\Lambda}_{\mathbf{P}} \mathbf{U}_{\mathbf{R}_t}^H$ and $\mathbf{\Sigma}_{\mathbf{P}} = \mathbf{D}_{\mathbf{P}} \mathbf{V}_{\mathbf{P}}^H$, $\mathbf{V}_{\mathbf{P}}$ being an arbitrary unitary matrix, the proposition is satisfied.

As a consequence, the optimal beamforming directions in Kronecker Rayleigh channels are given by the eigenvectors of the transmit correlation matrix $\mathbf{R}_t$, irrespective of the receive correlation profile. In Figure 8.1, the multi-mode beamformer $\mathbf{U}_{\mathbf{P}}$ is thus equal to $\mathbf{U}_{\mathbf{R}_t}$. Moreover, the constellation shaper $\mathbf{\Sigma}_{\mathbf{P}}$ is a real diagonal matrix $\mathbf{D}_{\mathbf{P}}$, up to an arbitrary unitary matrix $\mathbf{V}_{\mathbf{P}}^H$. The latter does not impact the performance and may be chosen as the identity for the sake of simplicity. The optimal constellation shaper $\mathbf{\Sigma}_{\mathbf{P}}$ thus determines the power allocated in each of the beamforming directions of $\mathbf{U}_{\mathbf{P}}$. Interestingly, there is a striking similarity with the capacity-achieving input covariance matrix $\mathbf{Q}$ using CDIT (see Section 4.3.2), although we will soon find that the optimal power allocation is different.

Example 8.1   *Let us consider the Alamouti O-STBC with two transmit antennas. Denoting $\mathbf{\Sigma}_{\mathbf{P}} = \mathrm{diag}\{\sigma_1(\mathbf{P}), \sigma_2(\mathbf{P})\} = \mathrm{diag}\{\sqrt{\lambda_1(\mathbf{P})}, \sqrt{\lambda_2(\mathbf{P})}\}$, the transmitted codewords are proportional, at the first time instant, to*

$$\frac{1}{\sqrt{2}} \mathbf{U}_{\mathbf{R}_t} \mathbf{\Sigma}_{\mathbf{P}} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \mathbf{U}_{\mathbf{R}_t}(:,1)\sigma_1(\mathbf{P})c_1 + \frac{1}{\sqrt{2}} \mathbf{U}_{\mathbf{R}_t}(:,2)\sigma_2(\mathbf{P})c_2 \quad (8.13)$$

and, at the second time instant, to

$$\frac{1}{\sqrt{2}} \mathbf{U}_{\mathbf{R}_t} \mathbf{\Sigma}_{\mathbf{P}} \begin{bmatrix} -c_2^* \\ c_1^* \end{bmatrix} = -\frac{1}{\sqrt{2}} \mathbf{U}_{\mathbf{R}_t}(:,1)\sigma_1(\mathbf{P})c_2^* + \frac{1}{\sqrt{2}} \mathbf{U}_{\mathbf{R}_t}(:,2)\sigma_2(\mathbf{P})c_1^*. \quad (8.14)$$

Clearly, the precoder sends symbols c_1 and c_2 on two orthogonal directions given by the eigenvectors of the transmit correlation matrix, also defined as transmit eigenbeams in Chapter 3. Depending on the channel correlation and the SNR, the power allocated to each eigenbeam will vary, allowing for a smooth transition between the classical transmit diversity Alamouti scheme ($\lambda_1(\mathbf{P}) = \lambda_2(\mathbf{P}) = 1$) and a beamforming in the dominant eigenbeam ($\lambda_1(\mathbf{P}) = 2$, $\lambda_2(\mathbf{P}) = 0$), also known as eigenbeamforming. The physical interpretation is rather straightforward: the precoder allocates more power to angular directions corresponding to the peaks of the transmit direction power spectrum $\mathcal{A}_t(\mathbf{\Omega}_t)$. If $\mathcal{A}_t(\mathbf{\Omega}_t) \to \delta(\mathbf{\Omega}_t - \mathbf{\Omega}_{t,0})$, the transmit correlation increases, reducing the effective diversity: it is therefore more efficient (in terms of error rate) to focus the transmit power in the

particular direction $\mathbf{\Omega}_{t,0}$ in order to improve the received SNR. If the scatterers are more uniformly distributed around the transmit array, the power allocation similarly follows a uniform distribution to exploit the full diversity.


Power allocation

As already mentioned, the constellation shaper reduces for O-STBCs to a simple power allocation scheme. The above example is a good introduction to the derivation of the optimal power allocation $\mathbf{\Lambda_P}$. Let us now formalize it to a general O-STBC [SP02, ZG03].

To simplify the analysis, we consider a Kronecker Rayleigh channel with no receive correlation, i.e. $\mathbf{R} = \mathbf{I}_{n_r} \otimes \mathbf{R}_t$. The integral evaluation is further alleviated by using the Chernoff upper-bound on the error probability. For the optimal precoder of Proposition 8.1, $\mathcal{G}(\zeta)$ is then upper-bounded by $[\det(\mathbf{I}_{n_t} + \zeta \mathbf{\Lambda_{R_t}} \mathbf{\Lambda_P})]^{-n_r}$ where $\zeta = \frac{\rho}{4} \frac{T}{Qn_t} \delta^2$. Denoting the diagonal elements of $\mathbf{\Lambda_P}$ as $\lambda_k(\mathbf{P})$, the power allocation minimizing this upper-bound is the allocation that maximizes $\sum_{k=1}^{n_t} \log(1 + \zeta \lambda_k(\mathbf{R}_t) \lambda_k(\mathbf{P}))$ under the conditions that $\mathrm{Tr}\{\mathbf{\Lambda_P}\} = n_t$ and that $\mathbf{\Lambda_P}$ is positive semidefinite. The optimum power allocation $\lambda_k^\star(\mathbf{P})$ is given by the water-filling solution of Proposition 4.1

$$\lambda_k^\star(\mathbf{P}) = \left[\frac{1}{\mu} - \frac{1}{\zeta \lambda_k(\mathbf{R}_t)}\right]^+ = \left[\frac{1}{\mu} - \frac{4Qn_t}{\rho T \delta^2 \lambda_k(\mathbf{R}_t)}\right]^+ \tag{8.15}$$

where μ is chosen to meet the power constraint and δ is generally taken as the minimum Euclidean distance of the constellation in order to minimize the PEP. To better apprehend the physical meaning of (8.15), Figures 8.2 and 8.3 illustrate the performance of the optimally precoded Alamouti scheme, together with a few other schemes, in two 2×2 Kronecker Rayleigh channels with transmit correlations $t = 0.7$ or $t = 0.95$, respectively (remember that the receive correlation is set to zero). We observe that, as a function of $\mathbf{R}_t$ and the SNR ρ, the optimal scheme evolves between dominant-mode eigenbeamforming, also known as single-mode eigenbeamforming (the total power is transmitted on a single mode corresponding to the strongest eigenmode of $\mathbf{R}_t$) and a classical O-STBC with all transmit eigenbeams receiving the same amount of power. Between these two extremes, a so-called multi-mode eigenbeamforming is performed. For a given transmit correlation matrix (i.e. t is fixed but assumed strictly smaller than 1), it is also clear that single-mode eigenbeamforming is only optimal at low SNR, as it lacks in exploiting transmit diversity. As the SNR increases, the power is spread over an increasing number of modes, reaching a uniform power allocation in the high SNR regime (only if $t < 1$). Moreover, for a given SNR, if the condition number of $\mathbf{R}_t$ increases (i.e. if t increases), the power is preferably transmitted on the dominant mode(s) only, canceling the power transmitted over the weakest modes. Actually, for $t = 0.95$, the power is transmitted along the dominant eigenbeam and the transitions to multi-mode eigenbeamforming and eventually to classical O-STBC only appear at larger SNR levels than those considered in Figure 8.3. With the single-mode
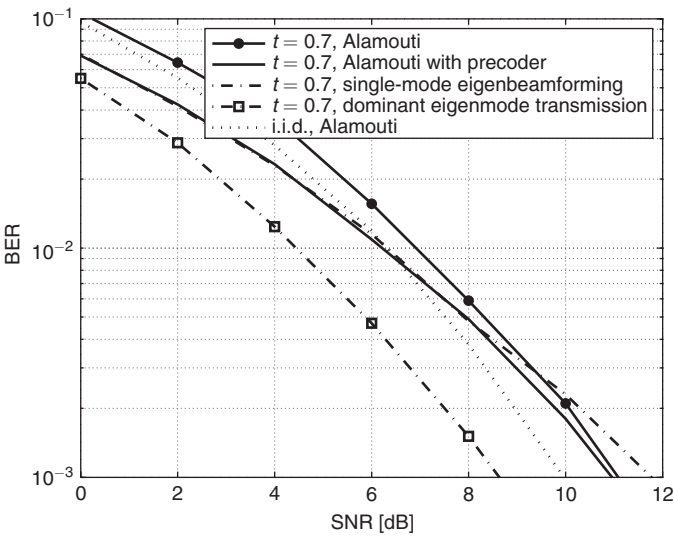
Figure 8.2 *Performance of a transmit correlation based precoded Alamouti scheme in 2×2 transmit correlated ($t = 0.7$) Rayleigh channels*
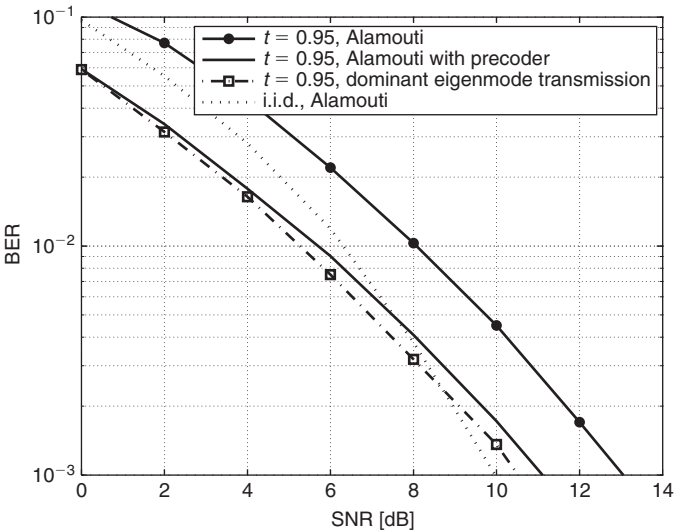
Figure 8.3 *Performance of a transmit correlation based precoded Alamouti scheme in 2×2 transmit correlated ($t = 0.95$) Rayleigh channels*

eigenbeamformer, we also note that the maximum array gain is obtained as the transmit correlation matrix approaches the identity, hence $\max_k \lambda_k(\mathbf{R}_t) = n_t$, yielding $g_a = n_t$. In Figure 8.3, an array gain close to 3 dB is indeed obtained w.r.t. the non-precoded Alamouti scheme, while this gain is much lower in Figure 8.2.

The transition between beamforming and uniform power allocation is very similar to the capacity-achieving power allocation. Comparing $\mathbf{P}$ and $\mathbf{Q}$, we observe that the beam-forming directions are identical but that the optimal power allocation is slightly different owing to the presence of δ^2 in (8.15). Finally, we might remember that CDIT-based pre-coding asymptotically approaches the ergodic capacity obtained via perfect CSIT (with no bias if $n_r \geq n_t$). This originally justified the use of CDIT instead of CSIT in practical applications.

The single-mode eigenbeamforming scheme of the above precoder may also be put in relationship with the dominant eigenmode beamformer using perfect CSIT (see Chapter 1). The former is designed based on the statistical properties of the channel matrix $\mathbf{H}$, while the latter is designed using the instantaneous knowledge of $\mathbf{H}$. Both schemes increase the average received SNR thanks to the exploitation of the transmit array gain (the array gain of dominant eigenmode transmission is equal to $\lambda_{max} = \max_k \lambda_k(\mathbf{H})$, as outlined in Chapter 1). However, there is a fundamental difference between both schemes: the single-mode eigenbeamformer never provides any transmit diversity gain while the dominant eigenmode transmission delivers the full transmit diversity (as long as the channel is almost uncorrelated). Yet, as t increases, single-mode eigenbeamforming performs almost as well as the dominant eigenmode transmission based on perfect CSIT, since the channel looses diversity.

Finally, the graphical results also confirm that O-STBCs are particularly robust to channel correlations. Even for $t = 0.7$, the performance is not so much degraded (although it should be kept in mind that the degradations caused by spatial correlations increase as the number of antennas increases).

If the receive correlation matrix is not equal to the identity matrix, the power allocation depends on the receive and transmit correlation profile, as it must maximize $\sum_{l=1}^{n_r} \sum_{k=1}^{n_t} \log\left(1 + \zeta\lambda_l(\mathbf{R}_r)\lambda_k(\mathbf{R}_t)\lambda_k(\mathbf{P})\right)$ using the Chernoff bound and the same power constraints as above. The Lagrangian optimization leads to the following power allocation strategy

$$\sum_{l=1}^{n_r} \frac{\zeta\lambda_l(\mathbf{R}_r)\lambda_k(\mathbf{R}_t)}{1 + \zeta\lambda_l(\mathbf{R}_r)\lambda_k(\mathbf{R}_t)\lambda_k(\mathbf{P})} = \mu \qquad (8.16)$$

where the Lagrange multiplier μ is chosen to satisfy the power constraint. For $n_r = 2$, two solutions are obtained. However, only one meets the positive semidefinite constraint, and is given by

$$\lambda_k^\star(\mathbf{P}) = \left[\frac{1}{\mu} - \frac{1}{2}(a_1 + a_2) + \sqrt{\frac{(a_1 - a_2)^2}{4} + \frac{1}{\mu^2}}\right]^+ \qquad (8.17)$$

with $a_l^{-1} = \zeta\lambda_l(\mathbf{R}_r)\lambda_k(\mathbf{R}_t)$. For $n_r > 2$, the optimal solution has to be numerically computed.

8.2.2   *Optimal precoding in non-Kronecker Rayleigh channels*

For a non-separable correlation matrix $\mathbf{R}$, obtaining the optimal precoder is more challenging, as a numerical optimization is usually required [HGA06]. Moreover, the optimal beamformer $\mathbf{U_P}$ is now a function of the full correlation matrix $\mathbf{R}$, hence the physical intuition is missing. If we use the Chernoff upper-bound of the PEP criterion and neglect the positive semidefinite constraint for simplicity, the unconstrained Lagrange function reads as

$$\mathcal{L}(\mathbf{P}) = \det\left(\mathbf{I}_{n_r n_t} + \zeta\mathbf{R}\left(\mathbf{I}_{n_r} \otimes \mathbf{PP}^H\right)\right) - \mu\left(\mathrm{Tr}\{\mathbf{PP}^H\} - n_t\right). \tag{8.18}$$

The optimal precoder is derived by equaling $\frac{\partial\mathcal{L}(\mathbf{P})}{\partial\mathrm{vec}(\mathbf{P}^\star)}$ to the zero vector. Since

$$\frac{\partial\mathrm{Tr}\{\mathbf{PP}^H\}}{\partial\mathrm{vec}(\mathbf{P}^\star)} = \mathrm{vec}(\mathbf{P}), \tag{8.19}$$

the optimal precoder results from the application of fixed point iterations to

$$\mathrm{vec}(\mathbf{P}) = \frac{1}{\mu}\frac{\partial \det\left(\mathbf{I}_{n_r n_t} + \zeta\mathbf{R}\left(\mathbf{I}_{n_r} \otimes \mathbf{PP}^H\right)\right)}{\partial\mathrm{vec}(\mathbf{P}^\star)}. \tag{8.20}$$

The derivation of the right term of (8.20) is detailed in [HGA06], but we restrict our analysis to a quick sketch of the algorithm. Choosing an initial value for $\mathbf{P}$ and calculating the right side of (8.20), the new estimate of $\mathbf{P}$ is given by the left term of (8.20). The result is then normalized to ensure that $\mathrm{Tr}\{\mathbf{PP}^H\} = n_t$. Re-iterating with the new estimate of $\mathbf{P}$ until the difference between the previous and new estimates is less than a given value, the so-derived precoder should finally be checked *a posteriori* against the positive semidefinite constraint.

If the optimization criterion is the exact error rate rather than its Chernoff upper-bound, a similar algorithm may be used [HG07b], as the unconstrained Lagrange function becomes

$$\mathcal{L}(\mathbf{P}) = P - \mu\left(\mathrm{Tr}\{\mathbf{PP}^H\} - n_t\right) \tag{8.21}$$

with P given by the exact PEP or SER. Fixed point iterations are then applied to

$$\mathrm{vec}(\mathbf{P}) = \frac{1}{\mu}\frac{\partial P}{\partial\mathrm{vec}(\mathbf{P}^\star)}. \tag{8.22}$$

Expressions of the partial derivatives are found in [HG07b], which also shows that precoders designed upon the exact SER or upon the Chernoff upper-bound offer similar performance for all SNR values.

8.2.3 Optimal precoding in Ricean fading channels

In Ricean fading channels with a Kronecker Rayleigh contribution and transmit correlation only ($\mathbf{R} = \mathbf{I}_{n_r} \otimes \mathbf{R}_t$), we may rewrite (8.6) by observing [VP06] that

$$\text{vec}(\bar{\mathbf{H}}^H)^H (\mathbf{I}_{n_r} \otimes \mathbf{PP}^H)(\mathbf{I}_{n_r n_t} + \zeta \mathbf{C_R})^{-1} \text{vec}(\bar{\mathbf{H}}^H)$$

$$= \text{Tr}\left\{ \bar{\mathbf{H}} \mathbf{PP}^H (\mathbf{I}_{n_t} + \zeta \mathbf{R}_t \mathbf{PP}^H)^{-1} \bar{\mathbf{H}}^H \right\} \tag{8.23}$$

using the classical relationships between the trace and vec operators (see Appendix A). Since $\mathbf{R}_t$ is non singular, we multiply the numerator and the denominator of (8.6) by $[\det(\mathbf{R}_t)]^{n_r} \exp\left(\text{Tr}\{K\bar{\mathbf{H}}\mathbf{R}_t^{-1}\bar{\mathbf{H}}^H\}\right)$ and use the matrix inversion lemma (see Appendix A) in order to express

$$\mathbf{PP}^H (\mathbf{I}_{n_t} + \zeta \mathbf{R}_t \mathbf{PP}^H)^{-1} - \frac{1}{\zeta}\mathbf{R}_t^{-1} = -\frac{1}{\zeta}\mathbf{W}^{-1} \tag{8.24}$$

with $\mathbf{W} = \mathbf{R}_t + \zeta \mathbf{R}_t \mathbf{PP}^H \mathbf{R}_t$. The average PEP is eventually rewritten as

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(K\text{Tr}\{\bar{\mathbf{H}}\mathbf{W}^{-1}\bar{\mathbf{H}}^H\}\right)[\det(\mathbf{R}_t)]^{n_r}$$

$$[\det(\mathbf{W})]^{-n_r} \exp\left(-\text{Tr}\{K\bar{\mathbf{H}}\mathbf{R}_t^{-1}\bar{\mathbf{H}}^H\}\right)d\beta. \tag{8.25}$$

To simplify the derivation, let us use the minimization of the Chernoff bound of (8.25) as a design criterion. Hence, we may alleviate the integration problem and take $\zeta = \frac{\rho}{4(1+K)}\frac{T}{Qn_t}\delta^2$. The optimization is also equivalent to minimizing the logarithm of the Chernoff bound of (8.25). Ignoring the terms that do not depend on $\mathbf{W}$, the objective function reads as

$$K\text{Tr}\{\bar{\mathbf{H}}\mathbf{W}^{-1}\bar{\mathbf{H}}^H\} - n_r \log\det(\mathbf{W}),$$

subject to the power constraint $\text{Tr}\{\mathbf{PP}^H\} = \sum_{k=1}^{n_t} \lambda_k(\mathbf{P}) = n_t$ and the condition that $\lambda_k(\mathbf{P}) \geq 0$, $k = 1, \ldots, n_t$. Therefore, the optimization problem may be summarized as follows [VP06]

$$\min_{\mathbf{W}} \quad K\text{Tr}\{\bar{\mathbf{H}}\mathbf{W}^{-1}\bar{\mathbf{H}}^H\} - n_r \log\det(\mathbf{W})$$

$$\text{subject to } \text{Tr}\{\mathbf{R}_t^{-1}\mathbf{W}\mathbf{R}_t^{-1} - \mathbf{R}_t^{-1}\} = n_t \zeta \text{ and } \mathbf{R}_t^{-1}\mathbf{W}\mathbf{R}_t^{-1} - \mathbf{R}_t^{-1} \succcurlyeq 0 \tag{8.26}$$

where the last inequality results from the positive semidefinite property of $\mathbf{PP}^H$, i.e. the fact that $\lambda_k(\mathbf{P}) \geq 0$, $k = 1, \ldots, n_t$.

Analogous to the water-filling solution, the Lagrangian optimization is first performed neglecting the positive semidefinite condition on $\mathbf{PP}^H$. Using various properties of derivatives outlined in [HGonc] and summarized in Appendix A, we obtain

$$\mathbf{W} = \frac{1}{2\mu}\mathbf{R}_t\left(n_r\mathbf{I}_{n_t} + \mathbf{\Psi}^{1/2}\right)\mathbf{R}_t \tag{8.27}$$

with $\boldsymbol{\Psi} = n_r^2 \mathbf{I}_{n_t} + 4\mu K \mathbf{R}_t^{-1} \bar{\mathbf{H}}^H \bar{\mathbf{H}} \mathbf{R}_t^{-1}$ and μ is the Lagrange multiplier attached to the power constraint. Inserting (8.27) into (8.26) leads to the following equation on μ

$$n_t n_r + \sum_{k=1}^{n_t} \sqrt{n_r^2 + 4\mu\lambda_k \left(\mathbf{R}_t^{-1} \bar{\mathbf{H}}^H \bar{\mathbf{H}} \mathbf{R}_t^{-1} \right)} - 2\left[\mathrm{Tr}\{\mathbf{R}_t^{-1}\} + n_t \zeta \right] \mu = 0 \qquad (8.28)$$

which requires a numerical search when $n_t > 1$. Once μ is found

$$\mathbf{PP}^H = \frac{1}{\zeta} \left[\frac{1}{2\mu} \left(n_r \mathbf{I}_{n_t} + \boldsymbol{\Psi}^{1/2} \right) - \mathbf{R}_t^{-1} \right] \qquad (8.29)$$

is evaluated and checked against the positive semidefinite property of $\mathbf{PP}^H$. If satisfied, the EVD $\mathbf{PP}^H = \mathbf{U_P} \boldsymbol{\Lambda_P} \mathbf{U_P}^H$ yields a precoder reading as

$$\mathbf{P} = \mathbf{U_P} \boldsymbol{\Sigma_P} \mathbf{V_P}^H$$

where $\mathbf{V_P}^H$ is an arbitrary unitary matrix (in practice, $\mathbf{V_P} = \mathbf{I}_{n_t}$). If $\mathbf{PP}^H$ is not positive semidefinite, the smallest eigenvalue of $\mathbf{PP}^H$ is set to zero and the total power is redistributed over the $n_t - 1$ largest eigenvalues, analogous to the iterative water-filling algorithm described in Chapter 4. By contrast to Rayleigh fading channels, the power on each mode and also the beamforming directions vary at each iteration, according to the modification of $\boldsymbol{\Psi}$.

If the particular case of uncorrelated Rayleigh contribution ($\mathbf{R}_r = \mathbf{I}_{n_r}$ and $\mathbf{R}_t = \mathbf{I}_{n_t}$), (8.29) yields

$$\mathbf{PP}^H = \underbrace{\mathbf{V}_{\bar{\mathbf{H}}}}_{\mathbf{U_P}} \underbrace{\frac{1}{\zeta} \left[\frac{1}{2\mu} \left(n_r^2 \mathbf{I}_{n_t} + 4\mu K \boldsymbol{\Lambda}_{\bar{\mathbf{H}}} \right)^{1/2} - \left(1 - \frac{n_r}{2\mu} \right) \mathbf{I}_{n_t} \right]}_{\boldsymbol{\Lambda_P}} \underbrace{\mathbf{V}_{\bar{\mathbf{H}}}^H}_{\mathbf{U_P}^H} \qquad (8.30)$$

where $\mathbf{U}_{\bar{\mathbf{H}}} \boldsymbol{\Sigma}_{\bar{\mathbf{H}}} \mathbf{V}_{\bar{\mathbf{H}}}^H$ is the SVD of $\bar{\mathbf{H}}$. Since $\mathbf{U_P} = \mathbf{V}_{\bar{\mathbf{H}}}$, a multi-mode beamforming is performed along the right singular vectors of $\bar{\mathbf{H}}$. The constellation shaper acts as a power allocation scheme whose optimal value [JSO02, VP06] is given by

$$\lambda_k^\star(\mathbf{P}) = \left[\frac{n_r + \sqrt{n_r^2 + 4\mu K \lambda_k(\bar{\mathbf{H}})}}{2\mu\zeta} - \frac{1}{\zeta} \right]^+ \qquad (8.31)$$

where $\lambda_k(\bar{\mathbf{H}})$ are the diagonal elements of $\boldsymbol{\Lambda}_{\bar{\mathbf{H}}} = \boldsymbol{\Sigma}_{\bar{\mathbf{H}}}^2$.

If $\mathbf{R}$ is not separable, there is no semi-closed form solution. Numerical evaluation is required [HG07a], relying on an algorithm similar to (8.22) with P being the exact or the Chernoff bound of the PEP/SER (8.5)–(8.10). Again, fixed point iterations solve the problem.

It is worth analyzing two asymptotic cases [VP06]. First, consider that the K-factor is infinite whereas the SNR remains finite. This assumption allows us to rewrite (8.5) considering that

$$\mathcal{O}\left(-\frac{\delta^2}{\sin^2 \beta}\right) = \exp\left(-\zeta K \text{Tr}\{\bar{\mathbf{H}}^H \bar{\mathbf{H}} \mathbf{P} \mathbf{P}^H\}\right). \tag{8.32}$$

We have

$$\text{Tr}\{\bar{\mathbf{H}}^H \bar{\mathbf{H}} \mathbf{P} \mathbf{P}^H\} \leq \max_{k=1,\dots,n_t} \lambda_k(\bar{\mathbf{H}}) \max_{k=1,\dots,n_t} \lambda_k(\mathbf{P}) \tag{8.33}$$

where equality occurs when the dominant eigenvector (i.e. the one corresponding to the largest eigenvalue) of $\mathbf{P}\mathbf{P}^H$ lies in the space spanned by the dominant right singular vector of $\bar{\mathbf{H}}$. Hence, the optimal precoding scheme comes to transmit the total power in a single direction corresponding to the dominant singular vector of $\bar{\mathbf{H}}$ [VP06]. Naturally, the channel correlation matrix should not be accounted for in the precoder design as the K-factor is very high. This is nothing else than conventional matched beamforming. In MISO systems, this is also capacity-achieving for infinite K-factor (see Section 4.4.2). However, in MIMO channels, the power allocation schemes obtained using PEP or information theory motivated criteria substantially differ. Indeed, the results of Section 4.4.2 call for a multi-mode beamformer even when K approaches infinity.

The second asymptotic case is an infinite SNR while the K-factor remains finite. Since the SNR is very large, the total power is spread over n_t directions and the precoder $\mathbf{P}$ is full rank. From (8.6), the precoder only affects the Rayleigh component and can be designed in the same way as for Rayleigh fading channels. Practically speaking, this further reveals that for a finite K-factor in the high SNR regime, the coherent component has a lower impact on the precoder design than the correlation matrix $\mathbf{R}$.

8.3 Channel statistics based precoding for codes with non-identity error matrices

The design of precoders for O-STBCs is relatively simple because error matrices are all similarly proportional to the identity matrix. Yet, the vast majority of space–time codes do not share this convenient property (see Chapter 5 for examples), thereby complicating the precoder design. Because each error matrix may have a different form, the precoder design (8.4) does not simply reduce to $\mathbf{P}^\star = \arg\min_{\mathbf{P}} P(\mathbf{C}^. \to \mathbf{E}^.)$, where $P(\mathbf{C}^. \to \mathbf{E}^.)$ is the average PEP evaluated for a particular error matrix $\tilde{\mathbf{E}}^.$ (for O-STBCs, $\tilde{\mathbf{E}}^. = T d_{min}^2/(Q n_t)\mathbf{I}_{n_t}$). Indeed, there is absolutely no evidence that $\tilde{\mathbf{E}}^.$ precisely leads to the maximum PEP for any precoder structure, SNR level, channel correlation matrix $\mathbf{R}$ and coherent component $\bar{\mathbf{H}}$.

Precoder design for a particular error matrix
Still, precoders have been sometimes successfully designed for a particular error matrix $\tilde{\mathbf{E}}^.$ [SP02, VP06]. In [SP02, FADT02], $\tilde{\mathbf{E}}^.$ is chosen as the minimum distance error matrix,

i.e. $\tilde{\mathbf{E}}^{\cdot} = \arg\min_{\tilde{\mathbf{E}} \neq 0} \det(\tilde{\mathbf{E}})$. In Kronecker Rayleigh fading channels with transmit correlation only, the precoder minimizing the worst-case PEP consists in a beamformer given by

$$\mathbf{U_P} = \mathbf{U_{R_t}}, \tag{8.34}$$

and a constellation shaper outlined by

$$\mathbf{\Sigma_P} = \mathbf{D_P V_P^H}, \tag{8.35}$$

where $\mathbf{D_P}$ is a real diagonal matrix realizing a power allocation and $\mathbf{V_P} = \mathbf{U_{\tilde{E}}}$, i.e. $\mathbf{V_P}$ is the eigenvector matrix of $\tilde{\mathbf{E}}^{\cdot} = \mathbf{U_{\tilde{E}}} \mathbf{\Lambda_{\tilde{E}}} \mathbf{U_{\tilde{E}}^H}$. The precoder is thus written as $\mathbf{P} = \mathbf{U_{R_t} D_P U_{\tilde{E}}^H}$. This is quite different from the O-STBC case where $\mathbf{V_P}$ could be arbitrarily chosen. Here, the choice of $\mathbf{V_P}$ (i.e. the choice of $\tilde{\mathbf{E}}^{\cdot}$) directly impacts the performance.

Using the Chernoff bound of the average PEP, the optimal power allocation must maximize

$$\sum_{k=1}^{n_t} \left(1 + \eta \lambda_k(\mathbf{R}_t) \lambda_k(\mathbf{P}) \lambda_k(\tilde{\mathbf{E}}^{\cdot}) \right), \tag{8.36}$$

under the power constraint $\sum_{k=1}^{n_t} \lambda_k(\mathbf{P}) = n_t$. The optimum power allocation therefore results from the well-known water-filling solution

$$\lambda_k^{\star}(\mathbf{P}) = \left[\frac{1}{\mu} - \frac{1}{\eta \lambda_k(\mathbf{R}_t) \lambda_k(\tilde{\mathbf{E}}^{\cdot})} \right]^{+} \tag{8.37}$$

with μ chosen to satisfy the power constraint.

In summary, the effect of the precoder $\mathbf{P}$ for non-identity error matrices is to remap the signal directions from $\mathbf{U_{\tilde{E}}}$ onto $\mathbf{U_{R_t}}$ and to redistribute the power across the modes according to $\mathbf{\Lambda_{\tilde{E}}}$ and $\mathbf{\Lambda_{R_t}}$ in such a way that the effective error matrix now reads as $\mathbf{P\tilde{E}^{\cdot}P^H} = \mathbf{U_{R_t} D_P \Lambda_{\tilde{E}} D_P U_{R_t}^H}$.

In Figure 6.7 (see Chapter 6), such a precoder has been applied under the denomination 'CYV & eigenbeamformer' to a 4-state STTC in a 4×2 Rayleigh fading channel with high transmit correlation (implying a power transmission in the dominant mode of $\mathbf{R}_t$ for a wide range of SNR). The matrix $\tilde{\mathbf{E}}^{\cdot}$ has been chosen as the minimum distance error matrix. In Figure 8.4, we show that this choice might not have been the best one as the gain w.r.t. the non-precoded scheme is rather weak. We are now able to explain this disappointing performance. Indeed, $\tilde{\mathbf{E}}^{\cdot} = \arg\min_{\tilde{\mathbf{E}} \neq 0} \det(\tilde{\mathbf{E}})$ is only responsible for the worst-case PEP at very high SNR. By contrast, a precoder aims primarily at improving the performance in the low and medium SNR regimes. Interestingly, if $\tilde{\mathbf{E}}^{\cdot}$ is chosen as the identity matrix with $\mathbf{U_{\tilde{E}}} = \mathbf{I}_{n_t}$, the performance of the 4-state code in the same propagation channel is substantially improved.
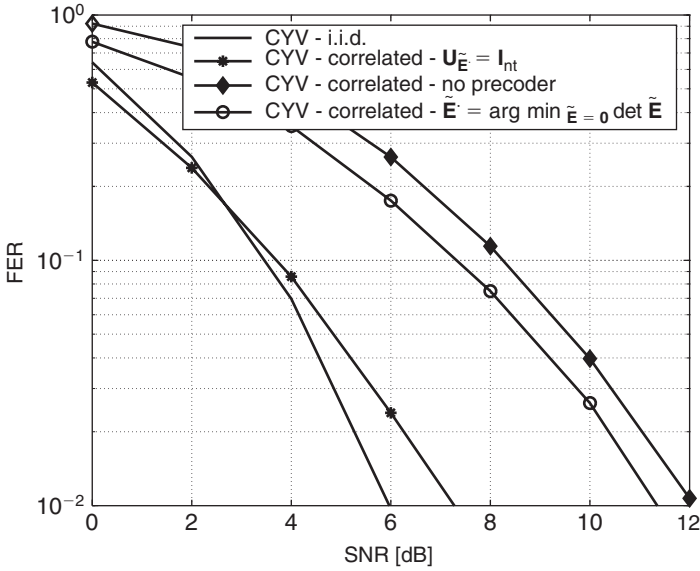
Figure 8.4 *Performance of a 4-state STTC 'CYV' in a correlated Rayleigh fading channel ($n_t = 2$, $n_r = 4$), with a high transmit correlation using two precoding schemes: $\mathbf{U}_{\tilde{\mathbf{E}}} = \mathbf{I}_{n_t}$ and $\tilde{\mathbf{E}} = \arg\min_{\tilde{\mathbf{E}} \neq 0} \det(\tilde{\mathbf{E}})$ (proposed in [SP02])*

Quite evidently, the choice of $\tilde{\mathbf{E}}$ has a significant impact on the performance. That can be further investigated using the Chernoff bound. The average PEP in transmit-correlated Kronecker Rayleigh fading channels may indeed be expressed as

$$P(\mathbf{C} \to \mathbf{E}) \leq \left[\det\left(\mathbf{I}_{n_t} + \eta \mathbf{\Lambda}_{\mathbf{R}_t} \mathbf{D}_{\mathbf{P}} \mathbf{U}_{\tilde{\mathbf{E}}}^H \tilde{\mathbf{E}} \mathbf{U}_{\tilde{\mathbf{E}}} \mathbf{D}_{\mathbf{P}} \right) \right]^{-n_r}. \tag{8.38}$$

If we assume that α eigenvalues of $\mathbf{R}_t$ are negligible, the power is spread over the $n_t - \alpha$ dominant eigenvectors of $\mathbf{R}_t$. This implies that α eigenvalues in (8.37) are equal to zero. Denoting by $\mathbf{D}'_{\mathbf{P}}$ the submatrix of size $(n_t - \alpha) \times (n_t - \alpha)$ containing the non-zero rows and the non-zero columns of $\mathbf{D}_{\mathbf{P}}$ and by $\mathbf{\Lambda}'_{\mathbf{R}_t}$ the diagonal submatrix of $\mathbf{\Lambda}_{\mathbf{R}_t}$ containing the $n_t - \alpha$ non-negligible eigenvalues, the average PEP is rewritten as

$$P(\mathbf{C} \to \mathbf{E}) \leq \left[\det\left(\mathbf{I}_{n_t - \alpha} + \eta \mathbf{D}'_{\mathbf{P}} \mathbf{\Lambda}'_{\mathbf{R}_t} \mathbf{D}'_{\mathbf{P}} \mathbf{\Lambda}'_{\tilde{\mathbf{E}}} \right) \right]^{-n_r} \tag{8.39}$$

where $\mathbf{\Lambda}'_{\tilde{\mathbf{E}}}$ $(n_t - \alpha) \times (n_t - \alpha)$ is the principal submatrix of $\mathbf{U}_{\tilde{\mathbf{E}}}^H \tilde{\mathbf{E}} \mathbf{U}_{\tilde{\mathbf{E}}}$. From the inclusion theorem (see Appendix A), the determinant in (8.39) is highly affected by the choice of $\mathbf{U}_{\tilde{\mathbf{E}}}$.

Let us now come back to Figure 8.4. We assume that the transmit correlation is high enough to cancel $\alpha = n_t - 1$ modes. Hence, the full power is transmitted in the dominant mode. From (8.39), the performance is then dictated by $\mathbf{\Lambda}'_{\tilde{\mathbf{E}}}(1,1)$. Choosing $\tilde{\mathbf{E}}$ to maximize

$\min_{\tilde{\mathbf{E}} \neq 0} \mathbf{\Lambda}'_{\tilde{\mathbf{E}}}(1, 1)$ should improve the performance. We may thus compare $\mathbf{\Lambda}'_{\tilde{\mathbf{E}}}(1, 1)$ for both simulated precoders:

- the choice $\mathbf{U}_{\tilde{\mathbf{E}}} = \mathbf{I}_{n_t}$ yields $\mathbf{\Lambda}'_{\tilde{\mathbf{E}}}(1, 1) = 2$
- using the minimum distance error matrix yields $\mathbf{\Lambda}'_{\tilde{\mathbf{E}}}(1, 1) = 0.34$.

For low-rate full-rank codes, the first solution generally provides $\mathbf{\Lambda}'_{\tilde{\mathbf{E}}}(1, 1) = \min_{\tilde{\mathbf{E}} \neq 0} \tilde{\mathbf{E}}(1, 1)$. This should be large, since the values of the diagonal elements of $\tilde{\mathbf{E}}$ are usually large for all $\tilde{\mathbf{E}} \neq \mathbf{0}$. By contrast, for Spatial Multiplexing, Hassibi's linear dispersion code or more general high-rate codes, this is not the case. Finding a good $\tilde{\mathbf{E}}$ becomes a really complex task and empirical rules appear as necessary. Naturally, these rules do not guarantee any optimality. The following proposition formalizes the above discussion.

Proposition 8.2   *In a Kronecker Rayleigh fading channel with transmit correlation only* $(\mathbf{R} = \mathbf{I}_{n_r} \otimes \mathbf{R}_t)$, *and for a PSK/QAM based linear STBC consisting of wide unitary basis matrices* $\mathbf{\Phi}_k$, *a precoder* $\mathbf{P} = \mathbf{U}_{\mathbf{R}_t} \mathbf{D}_{\mathbf{P}} \mathbf{V}_{\mathbf{P}}^H$, *with* $\mathbf{D}_{\mathbf{P}}$ *a real diagonal matrix and* $\mathbf{V}_{\mathbf{P}}$ *a unitary matrix, minimizes the worst-case average PEP if the wide unitary matrices* $\{\mathbf{\Phi}_k\}_{q=1}^{2Q}$ *are pairwise skew-hermitian* $\mathbf{\Phi}_q \mathbf{\Phi}_p^H + \mathbf{\Phi}_p \mathbf{\Phi}_q^H = \mathbf{0}_{n_t}$ *for* $q \neq p$. *This condition is necessary and sufficient to minimize the union bound on the average PEP.*

Proof: Applying Hadamard's inequality and reasoning in the same way as in the proof of Proposition 5.3, we have

$$\min_{q=1\ldots Q} \min_{d_q} \int_0^{\pi/2} \det\left(\mathbf{I}_{n_t} + \eta \mathbf{R}_t \mathbf{P} \tilde{\mathbf{E}} \mathbf{P}^H\right) d\beta$$
$$\leq \int_0^{\pi/2} \det\left(\mathbf{I}_{n_t} + \eta \mathbf{\Lambda}_{\mathbf{R}_t} \mathbf{D}_{\mathbf{P}} \frac{T}{Qn_t} \min_{d_q}\left[|\Re[d_q]|^2 + |\Im[d_q]|^2\right]\right) d\beta$$
$$= \int_0^{\pi/2} \det\left(\mathbf{I}_{n_t} + \zeta \mathbf{\Lambda}_{\mathbf{R}_t} \mathbf{D}_{\mathbf{P}}\right) d\beta \tag{8.40}$$

with $\zeta = \eta T d_{min}^2/(Qn_t)$. The equality in (8.40) is obtained if $\mathbf{\Phi}_q \mathbf{\Phi}_p^H + \mathbf{\Phi}_p \mathbf{\Phi}_q^H = \mathbf{0}_{n_t}$, $q \neq p$. Necessary and sufficient conditions result from [San02]. ∎

Rigorously, O-STBCs are the only codes satisfying the pairwise skew-hermitian condition. However, a precoder $\mathbf{P} = \mathbf{U}_{\mathbf{R}_t} \mathbf{D}_{\mathbf{P}} \mathbf{U}_{\tilde{\mathbf{E}}}^H$ might still perform well for non-orthogonal codes whose basis matrices are almost pairwise skew-hermitian. This is simply tested by evaluating $\max_{q \neq p} \left\| \mathbf{\Phi}_q^H \mathbf{\Phi}_p + \mathbf{\Phi}_p^H \mathbf{\Phi}_q \right\|_F^2$.

We stress that the above proposition only claims that the pairwise skew-hermitian property is sufficient to minimize the worst-case PEP. Therefore, it is possible to find

non-skew-hermitian codes for which the precoder can be safely used. For example, consider the delay-diversity scheme for $n_t = 2$. From (7.96), we know that this code presents diagonal error matrices when the error length is equal to two, corresponding to the minimum error length of the code. In particular, the error matrix equal to $d_{min}^2 \mathbf{I}_{n_t}$ causes the worst-case PEP in both i.i.d. and correlated Rayleigh fading channels analogous to O-STBCs. A precoder designed for the Alamouti scheme could thus be reused for the delay-diversity code with $n_t = 2$. From [VP06], this could also be applied to QO-STBCs.

8.4 Channel statistics based precoding for Spatial Multiplexing

An important scheme for which the precoders designed in the previous sections do not provide any robustness improvement is Spatial Multiplexing. Therefore, ad-hoc methodologies should be developed. For simplicity, we restrict our analysis to a Spatial Multiplexing scheme with 2 transmit and n_r receive antennas in Rayleigh flat fading channels, but the general approach is theoretically valid for more than two transmit antennas.

The physically transmitted codewords of size $n_t \times 1$ by $\mathbf{C}' = [c_1' \; c_2']^T$ are related to the original SM codewords[1] by $\mathbf{C} = [c_1 \; c_2]^T$ via $\mathbf{C}' = \mathbf{PC}$. Denoting as $\mathbf{E}' = [e_2' \; e_2']^T = \mathbf{PE}$ another physically transmitted codeword, and defining $\tilde{\mathbf{E}}' = (\mathbf{C}' - \mathbf{E}')(\mathbf{C}' - \mathbf{E}')^H$, the linear precoder yields

$$\tilde{\mathbf{E}}' = \mathbf{P}\tilde{\mathbf{E}}\mathbf{P}^H, \tag{8.41}$$

where we recall that $\tilde{\mathbf{E}} \triangleq (\mathbf{C} - \mathbf{E})(\mathbf{C} - \mathbf{E})^H$. From (6.25), the average PEP in Kronecker Rayleigh fading channels is expressed as

$$P(\mathbf{C} \rightarrow \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \prod_{k=1}^{n_r} \left(1 + \eta \lambda_k(\mathbf{R}_r) D^2\right)^{-1} d\beta \tag{8.42}$$

where D is the equivalent Euclidean distance seen by the receiver

$$D^2 = \text{Tr}\{\mathbf{R}_t \tilde{\mathbf{E}}'\} = |c_1' - e_1'|^2 + |c_2' - e_2'|^2 + 2\Re\{t(c_1' - e_1')(c_2' - e_2')^*\}, \tag{8.43}$$

and t denotes the transmit correlation coefficient.

An estimate of the average symbol error probability [NBP01, NBE+02, BP01] is obtained via a weighted average of the Chernoff bound on the PEP (8.42) over all possible pairs of codeword vectors $\mathbf{C}$ and $\mathbf{E}$. The weights, denoted as s, result from the fact that different codeword vectors $\mathbf{C}$ and $\mathbf{E}$ may cause a different number of symbol errors, i.e.

$$s(c_1, e_1, c_2, e_2) = \begin{cases} 2, & c_1 - e_1 \neq 0 \text{ and } c_2 - e_2 \neq 0 \\ 1, & c_1 - e_1 = 0 \text{ or } c_2 - e_2 = 0 \\ 0, & c_1 - e_1 = 0 \text{ and } c_2 - e_2 = 0. \end{cases} \tag{8.44}$$

---

[1] Without loss of generality, we have included the codeword normalization factor of $1/\sqrt{2}$ within the symbol.

If the symbols are picked up in a M-point constellation $\mathcal{B}$, the average symbol error rate is approximated as

$$\bar{P} \approx \frac{1}{M^2} \sum_{c_1 \in \mathcal{B}} \sum_{e_1 \in \mathcal{B}} \sum_{c_2 \in \mathcal{B}} \sum_{e_2 \in \mathcal{B}} s(c_1, e_1, c_2, e_2) \prod_{k=1}^{n_r} \left(1 + \frac{\rho}{4} \lambda_k(\mathbf{R}_r) D^2\right)^{-1}. \tag{8.45}$$

Note that analogous to all union bound expressions, the above approximation is usually not very accurate at low SNR.

Before studying different beamformers and constellation shapers, we first investigate whether a precoder obtained for O-STBC may also be used for SM. Such a precoder is given by $\mathbf{P} = \mathbf{U}_{\mathbf{R}_t} \text{diag}\{\sqrt{\lambda_1^\star(\mathbf{P})}, \sqrt{\lambda_2^\star(\mathbf{P})}\}$ so that D^2 becomes

$$D^2 = (1 + |t|)\lambda_1^\star(\mathbf{P})|c_1 - e_1|^2 + (1 - |t|)\lambda_2^\star(\mathbf{P})|c_2 - e_2|^2. \tag{8.46}$$

Clearly, this precoding structure does not improve the performance in correlated channels. For instance, if $|t|$ is large enough so that the power is allocated mainly over the dominant mode (i.e. $\lambda_1^\star(\mathbf{P}) >> \lambda_2^\star(\mathbf{P})$), taking $c_1 - e_1 = 0$ and $c_2 - e_2 \neq 0$ yields an extremely high PEP.

8.4.1 Beamforming

Remember that the beamforming step is outlined by $\mathbf{C}' = \mathbf{U_P}\mathbf{C}''$. For SM, we choose $\mathbf{U_P}$ as a diagonal matrix $\mathbf{U_P} = \text{diag}\{1, e^{j\psi}\}$, and $\mathbf{C}'' = \begin{bmatrix} c_1'' & c_2'' \end{bmatrix}^T$ is a codeword from an intermediate constellation yet to be defined through the linear relationship $\mathbf{C}'' = \mathbf{\Sigma_P}\mathbf{C}$. The chosen $\mathbf{U_P}$ is clearly different from the capacity-achieving and O-STBC precoders. While there is no proof that this is optimal, ψ is generally chosen as the phase of the transmit correlation coefficient t, hence

$$\mathbf{U_P} = \text{diag}\{1, t/|t|\}. \tag{8.47}$$

This can be intuitively understood as an eigenbeamformer in the direction of the dominant eigenvector of $\mathbf{R}_t$. The exact physical meaning and the difference with O-STBC precoding will be made clearer later. Yet, one advantage of (8.47) is to separate the effect of the phase and the magnitude of t. Indeed, the formalism of (8.47) results in that $\mathbf{C}''$ can be designed irrespective of the phase correlation coefficient because D^2 is only a function of $|t|$ and $\tilde{\mathbf{E}}'' \triangleq (\mathbf{C}'' - \mathbf{E}'')(\mathbf{C}'' - \mathbf{E}'')^H$, as outlined by

$$D^2 = \text{Tr}\left\{\begin{bmatrix} 1 & |t| \\ |t| & 1 \end{bmatrix} \tilde{\mathbf{E}}''\right\} = (1 - |t|)\left\|\mathbf{C}'' - \mathbf{E}''\right\|^2 + |t|\left\|[1\ \ 1](\mathbf{C}'' - \mathbf{E}'')\right\|^2$$

$$= (1 - |t|)\left[|c_1'' - e_1''|^2 + |c_2'' - e_2''|^2\right] + |t||(c_1'' + c_2'') - (e_1'' + e_2'')|^2. \tag{8.48}$$

The squared Euclidean distance seen by the receiver is expressed as a linear combination of two terms: the former is the squared Euclidean distance between the codewords $\mathbf{C}''$ and $\mathbf{E}''$, and the latter is the squared Euclidean distance between the constellation points $c_1'' + c_2''$

and $e_1'' + e_2''$. In terms of minimum distance, it can be considered as the G_{sum} parameter (see Section 7.3.3) of SM codewords $\{\mathbf{C}''\}$ evaluated in a direction of departure corresponding to a phase of t equal to zero.

For low transmit correlation, only the first term affects the performance. Hence, the minimization of the SER/PEP is obtained through a maximization of the Euclidean distance between the symbols of the constellations, analogous to classical constellations. Therefore, the unitary precoder $\mathbf{P}$ is optimal.

As the transmit correlation t increases, the second term should be accounted for. As a result, the structure of $\mathbf{\Sigma_P}$ should be chosen in such a way that the equivalent constellation formed by the points $c_1'' + c_2''$ presents large Euclidean distances. Ultimately, for $|t| = 1$, $\mathbf{\Sigma_P}$ acts similarly to the robust precoders designed in Chapter 7. However, the difference lies in the fact that we do not need to maximize the minimum of G_{sum} over all directions of departure but only in the direction corresponding to a phase of t equal to zero. Indeed, the rest is taken care of by $\mathbf{U_P} = \text{diag}\{1, t/|t|\}$ which acts as a beamformer in the preferred direction outlined by $t/|t|$.

8.4.2 Constellation shaping

Three different constellation shapers $\mathbf{\Sigma_P}$ are investigated:

- precoder I: $\mathbf{\Sigma_P}^{(I)} = \sqrt{2}\begin{bmatrix} \sqrt{p} & 0 \\ 0 & \sqrt{1-p} \end{bmatrix}$

- precoder II: $\mathbf{\Sigma_P}^{(II)} = \begin{bmatrix} 1 & 0 \\ \sqrt{1-p} & \sqrt{p} \end{bmatrix}$

- precoder III: $\mathbf{\Sigma_P}^{(III)} = \begin{bmatrix} \sqrt{p}e^{j\theta} & \sqrt{1-p} \\ \sqrt{p'}e^{j\theta} & \sqrt{1-p'} \end{bmatrix}$.

The idea behind precoder I is to prevent error vectors from lying in the null space of the channel. This is somewhat similar to the robust code designs of Chapter 7, except that t is now a known parameter. Both streams remain totally independent analogous to non-precoded SM. Precoder II combines the inputs c_1 and c_2 in such a way that c_2'' is no longer independent of c_1'', unlike the cases of classical Spatial Multiplexing and precoder I. The constellation transmitted on the second antenna thereby depends on the symbol $c_1'' = c_1$ transmitted on the first branch. Finally, precoder III generalizes the previous precoder but also offers the opportunity to spread symbols on both antennas.

The main problem with SM precoders is that the error matrix $\tilde{\mathbf{E}}$ may take many different forms. This is in strong contrast to O-STBC for which the error matrix is always proportional to $d_{min}^2 \mathbf{I}_{n_t}$ irrespective of the precoder and the channel correlations. Defining $d_1 = c_1 - e_1$ and $d_2 = c_2 - e_2$, the equivalent squared Euclidean distance D^2 may be

rewritten as

$$D^2 = \text{Tr} \left\{ \boldsymbol{\Sigma}_{\mathbf{P}}^H \begin{bmatrix} 1 & |t| \\ |t| & 1 \end{bmatrix} \boldsymbol{\Sigma}_{\mathbf{P}} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \begin{bmatrix} d_1^* & d_2^* \end{bmatrix} \right\}. \tag{8.49}$$

Focusing on the minimization of the maximum PEP, the optimal constellation shaper is given by

$$\boldsymbol{\Sigma}_{\mathbf{P}}^\star = \arg \max_{\boldsymbol{\Sigma}_{\mathbf{P}}} \min_{\substack{\mathbf{C},\mathbf{E} \\ \mathbf{C} \neq \mathbf{E}}} D^2. \tag{8.50}$$

In order to obtain a closed-form solution for a specific precoder, we need to find which error events are responsible for the minimum Euclidean distance D_{min}. In low transmit correlation scenarios, the precoder is practically equal to an identity matrix, hence $\{d_1', d_2'\} \cong \{d_1, d_2\}$, and these events are outlined by $\{|d_1|^2, |d_2|^2\} = \{d_{min}^2, 0\}$ and $\{|d_1|^2, |d_2|^2\} = \{0, d_{min}^2\}$, d_{min}^2 being the squared minimum Euclidean distance of the constellation. As the transmit correlation increases, the error events causing a minimum Euclidean distance D_{min} are not limited to those two events. Using the following EVD

$$\begin{bmatrix} 1 & |t| \\ |t| & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1+|t| & 0 \\ 0 & 1-|t| \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \frac{1}{\sqrt{2}}, \tag{8.51}$$

we observe that the error events, which we are interested in, are those lying within the subspace spanned by $[1 \ -1]^T$, i.e. by the eigenvector of (8.51) corresponding to the lowest eigenvalue $1 - |t|$. Expressing the $(i,j)^{\text{th}}$ element of $\boldsymbol{\Sigma}_{\mathbf{P}}$ as $\sigma_{ij}(\mathbf{P})$, the error events $\{d_1', d_2'\} = \{\sigma_{11}(\mathbf{P})d_1 + \sigma_{12}(\mathbf{P})d_2, \sigma_{21}(\mathbf{P})d_1 + \sigma_{22}(\mathbf{P})d_2\}$ that lie within this subspace have a similar magnitude, i.e. $\min_{d_1,d_2} ||d_1'| - |d_2'|| \to 0$, and an opposite phase. When all problematic error events are identified, the precoder is designed in order to maximize the minimum distance D among those corresponding to these events. Once again, this shows that a good precoder design for SM cannot rely on one particular error matrix but should consider a set of several problematic error matrices.

Constellation shaping using precoder I

We consider that $\mathbf{U}_{\mathbf{P}} = \text{diag}\{1, t/|t|\}$ and that $\boldsymbol{\Sigma}_{\mathbf{P}} = \boldsymbol{\Sigma}_{\mathbf{P}}^{(I)}$. Minimizing the maximum PEP requires to find $p^\star = \arg \max_p D_{min}^2$ with

$$D_{min}^2 = \min_{d_1,d_2} D^2 = \min_{d_1,d_2} \left\{ p|d_1|^2 + (1-p)|d_2|^2 + 2\sqrt{p}\sqrt{1-p}|t|\Re[d_2 d_1^*] \right\}. \tag{8.52}$$

Let us deal with a QPSK constellation for simplicity. Reasoning as explained above regarding $\{d_1', d_2'\}$, D_{min}^2 is only achieved when:

- $\{|d_1|^2, |d_2|^2\} = \{d_{min}^2, 0\}$
- $\{|d_1|^2, |d_2|^2\} = \{0, d_{min}^2\}$
- $\{d_1, d_2\} = \{d_{min}e^{j\varphi_{d_1}}, -d_{min}e^{j\varphi_{d_1}}\}$.

Evaluating D^2 in each of these cases, the optimal value of p should satisfy

$$p^\star = \arg\max_p \min \left\{ p,\ 1-p,\ 1 - 2\sqrt{p}\sqrt{1-p}|t| \right\} \tag{8.53}$$

whose solution is given by

$$p^\star = \begin{cases} 0.5 & |t| \le 0.5 \\ \frac{1}{1+4|t|^2} \quad \text{or} \quad \frac{4|t|^2}{1+4|t|^2} & |t| \ge 0.5 \end{cases}. \tag{8.54}$$

In the extreme case $|t| = 1$, the optimal constellation shaper outputs a constellation whose symbols read as $\sqrt{2}(\sqrt{0.8}c_1 + \sqrt{0.2}c_2)$, i.e. exactly as 16-QAM symbols.

Another approach relying on precoder I uses the virtual channel representation (see Chapter 3). In Section 7.4.3, a design criterion for Spatial Multiplexing based on the virtual channel representation was already discussed, but this design is easily extended to take into account that the transmitter knows $\mathbf{R}_v$. In [HLHS03], a precoder $\mathbf{P} = \boldsymbol{\Sigma}_{\mathbf{P}}^{(I)}\mathbf{W}\hat{\mathbf{A}}_t^*$ is used, with $\mathbf{W}$ being a unitary matrix and $\hat{\mathbf{A}}_t$, the virtual steering matrix at the transmitter. In order to extract the best performance out of this precoder, both $\boldsymbol{\Sigma}_{\mathbf{P}}^{(I)}$ and $\mathbf{W}$ should be jointly optimized. However, a suboptimal solution consists in taking $\mathbf{W}$ as the diagonal or general unitary precoder designed in Section 7.8. No beamformer $\mathbf{U}_{\mathbf{P}}$ is used, and the power allocation $\boldsymbol{\Sigma}_{\mathbf{P}}^{(I)}$ is optimized in order to minimize the maximum PEP based on the Chernoff upper-bound of (7.78), i.e.

$$p^\star = \arg\min_p \max_{\substack{\mathbf{C},\mathbf{E} \\ \mathbf{C} \ne \mathbf{E}}} \left(\det \left(\mathbf{I}_{n_r n_t} + \eta \mathbf{R}_v \left(\mathbf{I}_{n_r} \otimes \hat{\mathbf{A}}_t^T \tilde{\mathbf{P}} \mathbf{E} \mathbf{P}^H \hat{\mathbf{A}}_t^* \right) \right) \right)^{-1}. \tag{8.55}$$

Precoders minimizing the SER rather than the worst-case PEP may also be built on precoder I. In [NBP01], a precoder with $\mathbf{U}_{\mathbf{P}} = \text{diag}\{1, e^{j\psi}\}$ is used, relaxing the assumption that ψ equals the phase of t. Both parameters ψ and p must then be optimized to minimize the average SER

$$\{p^\star, \psi^\star\} = \arg\min_{p,\psi} \bar{P}. \tag{8.56}$$

The optimization is carried out through a numerical search in the domains $0 < p \le 0.5$ and $0 \le \psi \le \frac{\pi}{2}$. As shown later in the simulation results, the effect of this precoder is mostly to prevent error events lying in the null space of the channel analogous to the precoders designed in Chapter 7. No beamforming effect is really provided by this scheme, leading to a performance that is only slightly better than robust precoding based on G_{sum}.

Constellation shaping using precoder II
We now choose $\mathbf{U}_{\mathbf{P}} = \text{diag}\{1, t/|t|\}$ and $\boldsymbol{\Sigma}_{\mathbf{P}} = \boldsymbol{\Sigma}_{\mathbf{P}}^{(II)}$, so that $d_1' = d_1$ and $d_2' = \sqrt{1-p}d_1 + \sqrt{p}d_2$, and D^2 is given by

$$D^2 = \left[2 + 2\sqrt{1-p}|t| - p \right] |d_1|^2 + p|d_2|^2 + 2\left[\sqrt{p}|t| + \sqrt{p}\sqrt{1-p} \right] \Re\,[d_2 d_1^*]. \tag{8.57}$$

For a QPSK constellation, the problematic error events are such that d_1 and d_2 are out of phase and of equal magnitude d_{min}. As a consequence, D_{min}^2 is achieved when:

- $\{|d_1|^2, |d_2|^2\} = \{d_{min}^2, 0\}$
- $\{|d_1|^2, |d_2|^2\} = \{0, d_{min}^2\}$
- $\{d_1, d_2\} = \{d_{min}e^{j\varphi_{d_1}}, -d_{min}e^{j\varphi_{d_1}}\}$.

The optimal p is then given by

$$p^\star = \arg \max_p \min \left\{ p, \, 2 - 2\sqrt{1-p}\sqrt{p} + 2|t|\left[\sqrt{1-p} - \sqrt{p}\right] \right\}. \tag{8.58}$$

On the one hand, $p^\star = 1$ if $|t| \leq 0.64$. On the other hand, when $|t| > 0.64$, $p^\star$ is the unique value of p between 0 and 1 satisfying $p + 2\sqrt{1-p}\sqrt{p} - 2|t|(\sqrt{1-p} - \sqrt{p}) - 2 = 0$. For $|t| = 1$, it is straightforward to see that $p^\star = 0.64$. Although it is possible to solve the equation numerically, a good approximation of the optimal power allocation reads as

$$p = \begin{cases} 1 & |t| \leq 0.64 \\ 0.7 - \frac{1}{6}(|t| - 0.64) & |t| > 0.64 \end{cases}. \tag{8.59}$$

Note that the solution does not vary much with $|t|$ for $|t| > 0.64$. Therefore, a practical implementation would consist in

$$p = \begin{cases} 1 & |t| \leq 0.64 \\ 0.65 & |t| > 0.64 \end{cases}. \tag{8.60}$$

The effect of both $\mathbf{U_P}$ and $\mathbf{\Sigma_P^{(II)}}$ are clearly distinct. The latter shapes the constellation transmitted on the second antenna, since it shrinks the original QPSK transmitted on the second antenna by a factor $\sqrt{p}$ and translates it around symbol $\sqrt{1-p}c_1$. The former then rotates c_2'' similarly to an eigenbeamformer. When the channel is highly correlated in transmission, the distance between c_1'' and c_2'' becomes smaller and most of the energy is transmitted in the strongest eigenbeam. As the transmit correlation decreases, the distance between c_1'' and c_2'' increases and the energy transmitted in the direction of the direction orthogonal to the main eigenvector of $\mathbf{R}_t$ (i.e. the second eigenbeam, corresponding to the lowest eigenvalue $1 - |t|$) increases. If $t = 0$, the power is transmitted uniformly in all directions.

Interestingly, the equivalent symbols seen at the receiver for $|t| = 1$ are $1.6c_1 + 0.8c_2 = \sqrt{1.6}\sqrt{2}(\sqrt{0.8}c_1 + \sqrt{0.2}c_2)$, analogous to precoder I except that the average power is 1.6 times larger. This 2 dB gain comes from the partial exploitation of the transmit array gain. By contrast, the form of precoder I maximizes the equivalent minimum distance by avoiding destructive interactions between the codewords and the transmit correlation, but is unable to extract any array gain potentially offered by the knowledge of transmit correlation.

Constellation shaping using precoder III

When $\mathbf{U_P} = \text{diag}\{1, t/|t|\}$ and $\mathbf{\Sigma_P} = \mathbf{\Sigma_P}^{(III)}$, D^2 takes a more complicated form

$$D^2 = \left[p + p' + 2|t|\sqrt{p}\sqrt{p'} \right] |d_1|^2 + \left[2 - p - p' + 2|t|\sqrt{1-p}\sqrt{1-p'} \right] |d_2|^2$$
$$+ 2 \left[\sqrt{1-p}\sqrt{p} + |t|\sqrt{1-p}\sqrt{p'} + |t|\sqrt{1-p'}\sqrt{p} + \sqrt{1-p'}\sqrt{p'} \right] \Re \left[e^{-j\theta} d_2 d_1^* \right].$$

(8.61)

The optimization becomes very complex owing to the large number of parameters. Yet, it has been shown that $p' = 1 - p$ offers good results as long as the transmit correlation remains small. For such a precoder, $d_1' = \sqrt{p}e^{j\theta}d_1 + \sqrt{1-p}d_2$ and $d_2' = \sqrt{1-p}e^{j\theta}d_1 + \sqrt{p}d_2$. It is observed that d_1' and d_2' are out of phase when $e^{j\theta}d_1$ and d_2 have opposite phases. Therefore, the use of $e^{j\theta}$ reduces the contribution of the vector $\left[e^{j\theta}d_1 \ d_2 \right]$ lying in the space spanned by $\left[1 \ -1 \right]$. For QPSK constellation, $\theta = \pi/4$ seems optimal. Again, we can focus on three error events:

- $\{|d_1|^2, |d_2|^2\} = \{d_{min}^2, 0\}$
- $\{|d_1|^2, |d_2|^2\} = \{0, d_{min}^2\}$
- $\{d_1, d_2\} = \{d_{min}e^{j\varphi_{d_1}}, -d_{min}e^{j\varphi_{d_1}}\}$.

Evaluating D^2 achieved by each of these events, the optimal p for low transmit correlation must satisfy

$$p^\star = \arg \max_p \min \left\{ 1 + 2|t|\sqrt{1-p}\sqrt{p}, \quad 2 + 4\sqrt{1-p}\sqrt{p}(|t| + A) + 2A|t| \right\} \quad (8.62)$$

where $A = -\sqrt{2}/2$. This can be solved to yield

$$p^\star = \frac{|t| + 2A \pm \sqrt{(1 - 4A^2)(|t|^2 - 1)}}{2|t| + 4A}. \quad (8.63)$$

For high transmit correlation, it is best to choose $p' = p$ and $\theta = 0$. This is equivalent to classical beamforming, since the same apparent symbol is transmitted on both antennas. As a result, $d_1' = \sqrt{p}d_1 + \sqrt{1-p}d_2$ and $d_2' = \sqrt{p}d_1 + \sqrt{1-p}d_2$. The vector $\left[d_1' \ d_2' \right]^T$ always lies in the space spanned by $\left[1 \ 1 \right]^T$ and excites thus the largest eigenvalue $1 + |t|$. The power allocation problem is similar to precoder I, since $D^2 = 2(1 + |t|)\left| \sqrt{p}d_1 + \sqrt{1-p}d_2 \right|^2$. This reveals that the optimal value of p is such that

$$p^\star = \arg \max_p \min \left| \sqrt{p}d_1 + \sqrt{1-p}d_2 \right|^2, \quad (8.64)$$

which is the optimization solved for precoder I, with $|t| = 1$ (see (8.52)). The optimal power allocation $p_{opt} = 0.2$ or 0.8. Compared with the previous precoders, this comes to transmit the same 16-QAM symbol on each antenna when the transmit antennas are fully

correlated. By contrast to precoders I and II, the array gain at the transmitter is now fully exploited. Indeed, D_{min}^2 is exactly twice as large as the squared minimum distance D_{min}^2 of precoder I.

It is also worthwhile to establish a relationship between precoder III or the capacity-achieving or O-STBC precoders. To this end, let us express the beamformer $\mathbf{U_P}$ as

$$\mathbf{U_P} = \begin{bmatrix} 1 & 0 \\ 0 & t/|t| \end{bmatrix} = \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ t/|t| & -t/|t| \end{bmatrix}}_{\mathbf{U_{R_t}}} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{8.65}$$

For high $|t|$, precoder III acts as follows

$$\mathbf{U_P P}''_{(3)} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \mathbf{U_{R_t}} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{0.8} & \sqrt{0.2} \\ \sqrt{0.8} & \sqrt{0.2} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \tag{8.66}$$

$$= \mathbf{U_{R_t}} \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{0.8}c_1 + \sqrt{0.2}c_2 \\ 0 \end{bmatrix}. \tag{8.67}$$

This last equation reveals that when $|t| = 1$, the total power is transmitted along the dominant eigenbeam. Instead of transmitting two QPSK-based independent layers, only one 16-QAM-based layer is transmitted to conserve the total rate. This can be put in relationship with the capacity-achieving precoder: when $|t| = 1$, the total power is also transmitted only on the dominant eigenmode of $\mathbf{R}_t$, and adaptive modulation/coding are performed on only one layer. For the Alamouti scheme, since the rate of the code is one, there is no need to transmit higher order modulation to keep the rate constant. However, keeping the same total transmission rate, a 16-QAM-based precoded Alamouti scheme and a QPSK-based precoded SM with $\mathbf{\Sigma_P}^{(III)}$ provides for $|t| = 1$ the same scheme, consisting in the transmission of a 16-QAM symbol in the dominant eigenvector of $\mathbf{R}_t$. Yet, the precoder in the Alamouti scheme aims to trades the spatial multiplexing gain (of the Alamouti scheme) with the array gain (of the beamformer), whereas the SM precoder trades the spatial multiplexing gain with the beamforming array gain.

In summary, precoder III may take the following forms

- $|t| < 0.78$: $\mathbf{\Sigma_P}^{(III)} = \begin{bmatrix} \sqrt{p}e^{j\pi/4} & \sqrt{1-p} \\ \sqrt{1-p}e^{j\pi/4} & \sqrt{p} \end{bmatrix}$ with $p = \frac{|t| + 2A \pm \sqrt{(1-4A^2)(|t|^2 - 1)}}{2|t| + 4A}$

- $|t| \geq 0.78$: $\mathbf{\Sigma_P}^{(III)} = \begin{bmatrix} \sqrt{p} & \sqrt{1-p} \\ \sqrt{p} & \sqrt{1-p} \end{bmatrix}$ with $p = 0.5 \pm 0.3$

where $A = -\sqrt{2}/2$.

Comparing precoders I, II and III

It has been seen that precoders I to III shape the original QPSK constellations so that in highly transmit-correlated channels, the receiver sees an equivalent 16-QAM constellation.
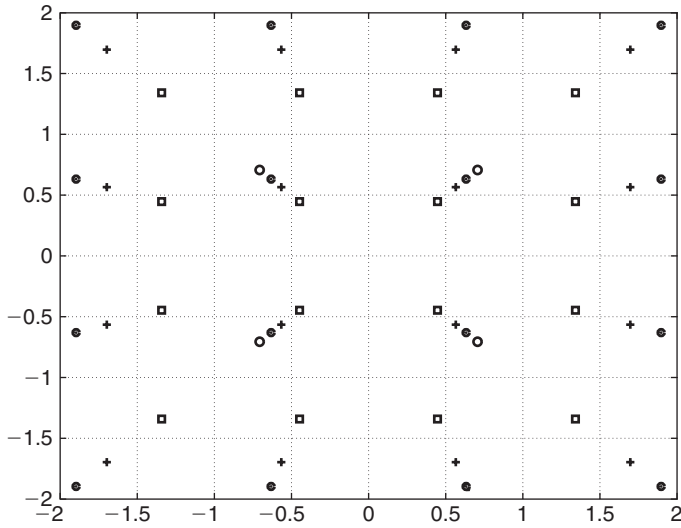
Figure 8.5 *QPSK (○), constellation perceived at the receiver when* $|t| = 1$ *and QPSK used with precoder I (□), precoder II (+) and precoder III (●)*

This is illustrated in Figure 8.5. We may also note that the difference between all three precoders is the average power of the so-induced 16-QAM constellations: the 16-QAM constellations of precoders II and III have a respective average power 1.6 and 2 times larger than the 16-QAM constellation of precoder I. Hence, precoders II and III have a respective coding gain of 2 dB and 3 dB higher than the coding gain of precoder I, suggesting that precoders II and III are much more efficient in exploiting the transmit array gain. Figure 8.6 also compares the squared minimum distance achieved by each precoder as a function of the absolute transmit correlation $|t|$. Precoder III performs best, as it has the interesting property that for small transmit correlation $(0 \leq |t| \leq 0.7)$, D^2_{min} is larger than the minimum Euclidean distance of the QPSK constellation $(d^2_{min} = 2)$.

Figures 8.7 and 8.8 further depict how the SM schemes, designed based on the maximum PEP criterion, perform in terms of error rate/required SNR in different propagation conditions. In Figure 8.7, the performance analysis of several schemes as a function of the transmit correlation in a 2×2 Kronecker Rayleigh fading channel assumes that the receive correlation is equal to zero.

On the one hand, we observe that the robust precoder (designed and denominated as 'new precoder' in Chapter 7) achieves the same performance as precoder I over a large range of t, being only outperformed for very high t. This suggests that precoder I should not be used in practice since a robust precoder may achieve similar performance without requiring the knowledge of t at the transmitter. On the other hand, precoders II and III achieve a performance that is almost independent of t, which shows that the transmit knowledge of t maintains the performance achieved in i.i.d. Rayleigh fading channels.
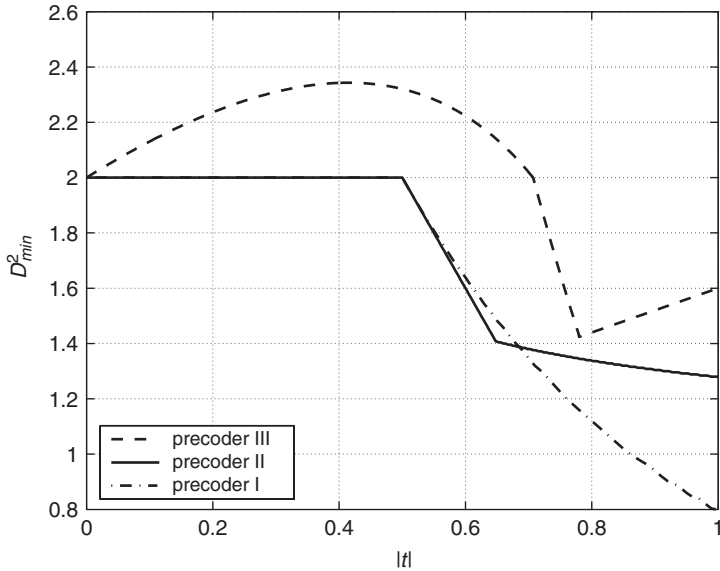
Figure 8.6 D_{min}^2 *achieved by the three precoders with QPSK constellation as a function of the magnitude of the transmit correlation coefficient*
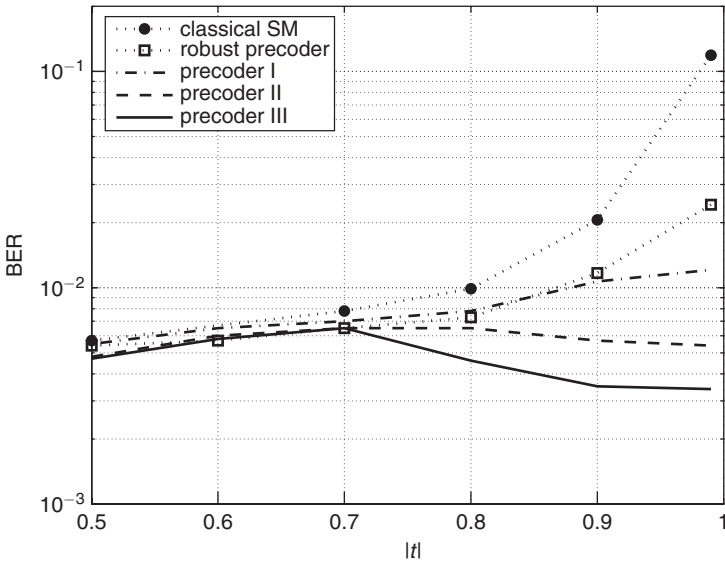
Figure 8.7   *Bit error rate of Spatial Multiplexing as a function of the transmit correlation coefficient* t *in* 2×2 *correlated MIMO channels with and without precoding (SNR = 15 dB)*

To produce the results of Figure 8.8, we have assumed that most scatterers and the transmit array are aligned with the link axis, so that the transmit azimuth power spectrum $\mathcal{A}_t(\theta_t) \cong \delta(\theta_t)$, and $t \cong e^{j\pi}$ for half-wavelength spacing at the transmitter. The gains offered by precoders II and III over precoder I match the theoretical gains of 2 and 3 dB. The
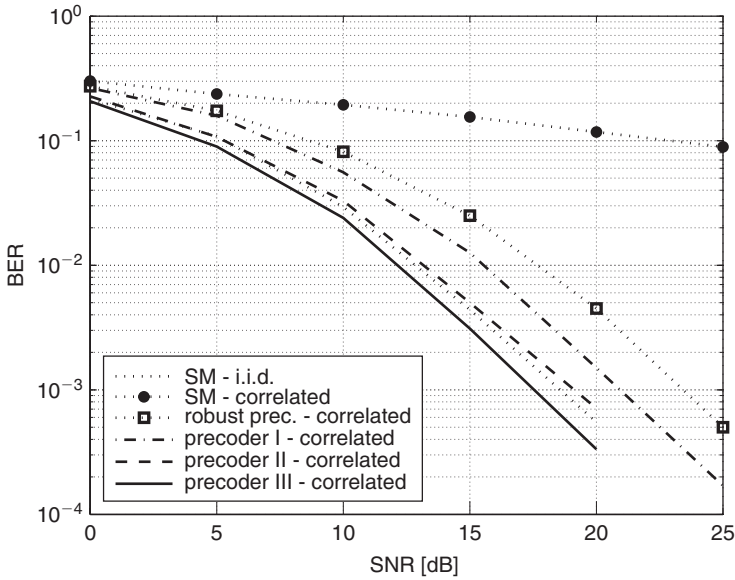
Figure 8.8 *Bit error rate of Spatial Multiplexing in correlated channels with and without precoding: precoders I, II and III exploit the knowledge of* t, *while the robust precoder has been designed following the* G_{sum} *criterion in Chapter 7*

performance of the classical SM scheme and the robust SM obtained by precoding the classical SM with a robust precoder designed following the G_{sum} criterion (see Chapter 7) is also proposed for comparison. We find that for high transmit correlation, precoder I has about 2–2.5 dB gain over the robust precoder (which does not require the knowledge of t). Precoders II and III offer 2 to 3 dB gain over precoder I. For comparison, the performance of SM in i.i.d. Rayleigh fading channels is also displayed. We observe that even if the channel is highly correlated, the exploitation of channel statistics at the transmitter results in error rates similar to (with precoder I, even better than) those obtained in uncorrelated Rayleigh channels with classical SM.

8.4.3   A non-linear approach to constellation shaping

Using linear constellation shapers is the common method to design precoders for Spatial Multiplexing schemes. However, it is also possible to use a non-linear approach consisting in the numerical optimization of the constellations $\mathbf{C}''$ [CVVJO04]. This results in signal constellations that are very similar to those obtained with the second linear precoder (precoder II).

In this non-linear method, the symbols c_1'' transmitted on the first antenna are selected from a constellation $\mathbf{S}$ (containing the symbols S_m, with $m = 1, \ldots, M_0$). The symbols c_2'' transmitted on the second antenna are selected from a constellation $\mathbf{K}_m$ (containing the

symbols K_{mn}, with $n = 1, \ldots, M_1$) which depends on the symbol S_m (with $m = 1, \ldots, M_0$) transmitted on the first antenna

$$\mathbf{C}''^T = \begin{bmatrix} S_m & K_{mn} \end{bmatrix}. \tag{8.68}$$

Analogous to precoder II, c_2'' is no longer independent of c_1'', as the constellation $\mathbf{K}_m$ depends on S_m. The difference is that the relationship between $\mathbf{C}$ and $\mathbf{C}''$ is no longer constrained by a linear shaper $\mathbf{\Sigma}_{\mathbf{P}}^{(II)}$.

Making use of the precoder $\mathbf{U_P} = \text{diag}\{1, t/|t|\}$, the physically transmitted symbols are chosen in the constellations $\mathbf{S}$ and $\{\mathbf{Q}_m\}$ (containing the symbols Q_{mn}, with $m = 1, \ldots, M_0$ and $n = 1, \ldots, M_1$) obtained after rotation of the constellations $\{\mathbf{K}_m\}$ as $\mathbf{Q}_m = e^{j\phi_t}\mathbf{K}_m$ where $t = |t|e^{j\phi_t}$. We thus have $\mathbf{C}'^T = [S_m \ Q_{mn}]$.

For simplicity, we will further assume that $\mathbf{R}_r = \mathbf{I}_2$. Following (8.45), the average symbol error probability can then be expressed as a function of $\mathbf{S}$, $\{\mathbf{K}_m\}$ and the magnitude of t

$$\bar{P} \approx \frac{1}{M_0 M_1} \sum_{m=1}^{M_0} \sum_{n=1}^{M_1} \sum_{u=1}^{M_0} \sum_{v=1}^{M_1} s(S_m, K_{mn}, S_u, K_{uv}) \left(1 + \left[|S_m - S_u|^2 \right. \right.$$

$$\left. \left. + |K_{mn} - K_{uv}|^2 + 2\Re \left\{ |t| \, (S_m - S_u) \, (K_{mn} - K_{uv})^* \right\} \right] \frac{\rho}{4} \right)^{-n_r}. \tag{8.69}$$

The selection of the signal constellations $\mathbf{S}$ and $\{\mathbf{K}_m\}_{m=1}^{M_0}$ should minimize the average symbol error probability $\bar{P}$ in (8.69) under an average power constraint [FGW74]. Let $\mathbf{f}$ be the vector defined as

$$\mathbf{f} = \begin{bmatrix} S_1 & \ldots & S_{M_0} & K_{11} & \ldots & K_{1M_1} & \ldots & K_{M_01} & \ldots & K_{M_0 M_1} \end{bmatrix}^T \tag{8.70}$$

and $\mathbf{f}^k$ the vector $\mathbf{f}$ at the kth step of the algorithm. The unconstrained gradient algorithm is described by $\mathbf{f}^{k+1} = \mathbf{f}^k - \alpha \nabla \bar{P} \left(\mathbf{f}^k \right)$, where α is the step size and $\nabla \bar{P} \left(\mathbf{f}^k \right)$ is the gradient of $\bar{P}$ with respect to $\mathbf{f}^k$. The signal vector $\mathbf{f}^k$ is normalized at each step so that

$$\mathbf{f}^k(1 : M_0) = \sqrt{M_0} \frac{\mathbf{f}^k(1 : M_0)}{\|\mathbf{f}^k(1 : M_0)\|}$$

$$\mathbf{f}^k(M_0 + 1 : M_0 + M_0 M_1) = \sqrt{M_0 M_1} \frac{\mathbf{f}^k(M_0 + 1 : M_0 + M_0 M_1)}{\|\mathbf{f}^k(M_0 + 1 : M_0 + M_0 M_1)\|} \tag{8.71}$$

in order to fix the average signal power to one.

The optimal constellations $\mathbf{S}$ and $\{\mathbf{K}_m\}$ are given in Figure 8.9 for a 4-bit/s/Hz system. They have been optimized [CVVJO04] with $t = 0.95$ while ρ is chosen to obtain a symbol error rate of 10^{-4}. Figure 8.9 displays the four symbols S_m ($m = 1, \ldots, 4$) which can be transmitted on the first antenna as well as the four symbols K_{1n} ($n = 1, \ldots, 4$) which can
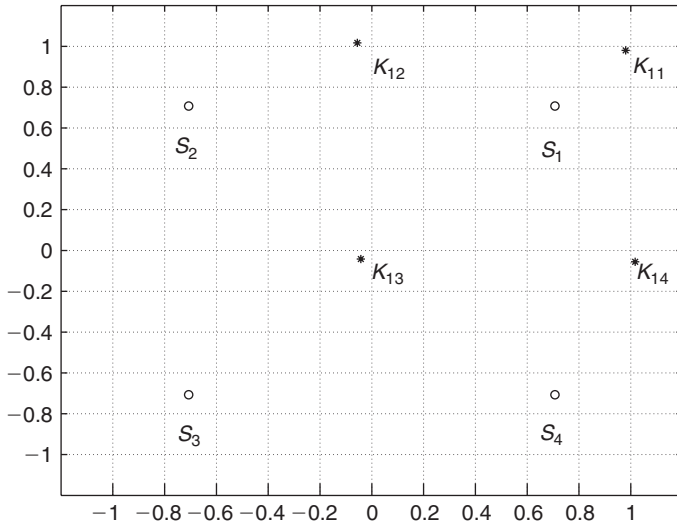
Figure 8.9 *Signal constellations* **S** *and* **K**$_1$ *for a 4-bit/s/Hz system using two transmit antennas*

be sent (in the form of Q_{1n}) on the second antenna if symbol S_1 is transmitted on the first one. The remaining symbols are not displayed to keep the figure legible. The full numerical description of the constellation is detailed in [CVVJO04].

As expected, symbols K_{mn} in Figure 8.9 are located somewhere around symbols S_m. Note that these constellations are very similar to the constellations of precoder II. The equivalent constellation seen by the receiver and given by the points $S_m + K_{mn}$ is also a 16-QAM constellation with the same average power as precoder II. However, for higher order QAM, the constellations obtained through this non-linear approach do not simply follow a shrink-and-translate effect. The interested reader is referred to [CVVJO04] for examples of 8-QAM constellations.

Figure 8.10 compares the performance achieved by the schemes designed based on the SER criterion for precoders I and II (entitled, respectively, 'Nabar's precoder' and 'non-linear constellations'), and the scheme designed following the virtual channel representation (entitled 'Hong's precoder'). The performance is evaluated in terms of the SNR required to achieve a target symbol error rate of 10^{-2} in two extreme scattering distributions (in both cases, a large inter-element distance is taken at the receiver, and the transmit array is oriented perpendicularly to the transmit-to-receive axis):

- most scatterers are uniformly distributed on a ring around the transmitter, corresponding to a large local scattering ratio (LSR $= 1$, see Section 2.3.3), so that the transmit correlation is well approximated by $t \cong J_0(2\pi d/\lambda)$
- most scatterers are aligned with the transmit-to-receive axis (small local scattering ratio, LSR $= 0$) so that the transmit azimuth power spectrum $\mathcal{A}_t(\theta_t)$ is close to $\delta(\theta_t)$.
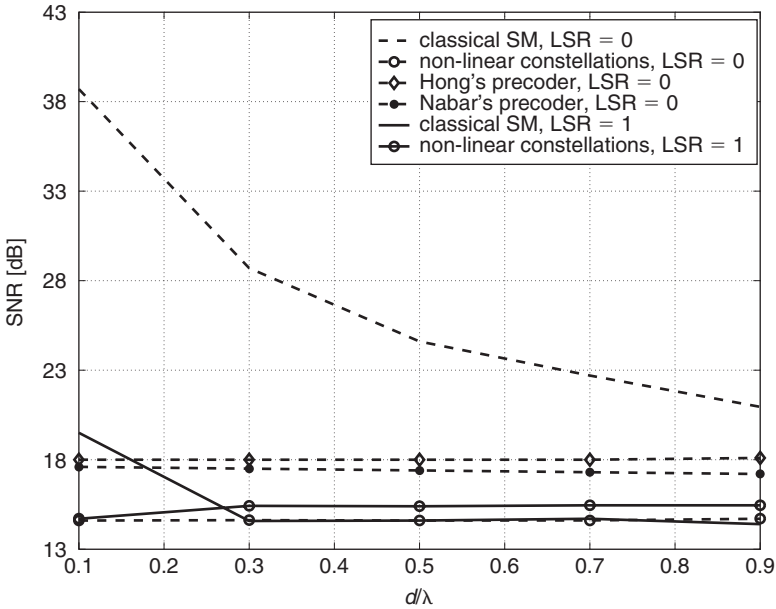
Figure 8.10 *Required SNR as a function of the inter-element distance over a Rayleigh channel for the classical and the new constellations in a 2×2 4-bit/s/Hz transmission in a broadside configuration*

The non-linear constellation outperforms Nabar's precoder by 2–2.5 dB, which corresponds to the gain of precoder II with respect to precoder I. Indeed, while those two schemes are designed based on a SER optimization, they perform almost similarly to precoders designed with a PEP criterion. Also, the non-linear constellations outperform Hong's precoder by 3 to 4 dB. This is not only because Hong's precoder is based on the virtual channel model (whose limitations have been highlighted in Chapters 3 and 7), but also because this precoder suffers from the same weakness as precoder I. Moreover, no beamformer is used.

Interestingly, although the non-linear constellations have been originally optimized for $|t| = 0.95$, they still perform very well for other values of t [CVVJO04]. This robustness directly results from the fact that the distance between symbols S_m and K_{mn} is small enough to convey energy in adequate directions in poor scattering conditions, and yet, sufficiently large as to excite many directions in rich scattering conditions. That explains why $\mathbf{\Sigma}_{\mathbf{P}}^{(II)}$ or the non-linear constellations, though designed with poor scattering scenarios in mind (i.e. for $t = 0.95$) can be computed off-line and stored permanently at the transmitter and the receiver. We only have to dynamically adapt the precoder $\mathbf{U_P}$ as a function of the phase of the transmit correlation coefficient. This makes those designs particularly suited for practical applications. For instance, it has been shown in [CVVJO04] that the non-linear constellations are sufficiently robust to exploit the channel estimation of the reverse link even in a FDD mode.

Remember that the precoders and non-linear constellations have been obtained for Rayleigh fading channels. In Ricean channels, they should theoretically be optimized depending on the Ricean K-factor, the transmit correlation and the structure of the coherent component, analogous to what is implemented for O-STBCs. However, if the angle of departure/arrival of the coherent component is centered with respect to the angles of departure/arrival of the variable components, it is possible to rely on the above design by using the phase of the transmit correlation (and not the covariance) in Ricean fading channels. Note that this becomes optimal as the number of receive antenna or the K-factor increase.

8.4.4 Precoder design for suboptimal receivers

So far, the derivations assume that ML detection is used at the receiver. However, ML decoding is quite complex, so that suboptimal receivers as those presented in Chapter 5 may be of interest. While designed for ML decoding, the aforementioned precoders/constellations are easily used with a linear decoder [AR05, GA05].

Assuming that $\mathbf{R}_r = \mathbf{I}_{n_r}$, a Kronecker Rayleigh channel may only be ill-conditioned due to high transmit correlations. In such case, the receiver cannot invert the channel matrix $\mathbf{H} = \mathbf{H}_w \mathbf{R}_t^{1/2}$. Yet, it may invert the i.i.d. component $\mathbf{H}_w$, e.g. thanks to a ZF filter. After ZF filtering, we may write ($\dagger$ denotes pseudo-inverse)

$$\mathbf{H}_w^\dagger \mathbf{r}_k = \sqrt{E_s} \underbrace{\begin{bmatrix} 1 & t^* \\ t & 1 \end{bmatrix}^{1/2} \overbrace{\begin{bmatrix} 1 & 0 \\ 0 & t/|t| \end{bmatrix}}^{\mathbf{U_P}}}_{\mathbf{H}_{eq}} \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} c_1'' \\ c_2'' \end{bmatrix}}_{\mathbf{C}''} + \mathbf{H}_w^\dagger \mathbf{n}_k. \tag{8.72}$$

Matched filtering with respect to $\mathbf{H}_{eq}$ yields

$$\mathbf{H}_{eq}^H \mathbf{H}_w^\dagger \mathbf{r}_k = \sqrt{\frac{E_s}{2}} \begin{bmatrix} 1 & |t| \\ |t| & 1 \end{bmatrix} \begin{bmatrix} c_1'' \\ c_2'' \end{bmatrix} + \mathbf{H}_{eq}^H \mathbf{H}_w^\dagger \mathbf{n}_k. \tag{8.73}$$

The estimation of $\{c_1'', c_2''\}$ is then based on a SIC-like decoder using detection/subtraction steps. Using precoders I or II for simplicity, the detection of c_1'' occurs first based on the first line of (8.73). Since c_2'' has a much larger average power than c_1'' (precoder I) or since c_2'' depends on c_1'' (precoder II), the symbol c_2'' is perceived as an interference in this first step. The subtraction of the estimated c_1'' from the second line of (8.73) allows for the detection of c_2''. When $|t| = 0$, this reduces to a standard MIMO ZF receiver. Analogous to ML decoding, the transmit array gain is efficiently exploited.

For high transmit correlation, the term $c_1'' + |t|c_2''$ can be seen as the equivalent constellation perceived by the receiver after ZF and matched filtering. If QPSK constellations are used, this equivalent constellation is a 16-QAM constellation with an increased average power compared to the initial unit-energy QPSK constellations. When a SIC-like decoder is used, the detection of symbol c_1'' first detects one of the quadrant of the 16-QAM, and the second

symbol is detected among the four symbols contained in the detected quadrant. Note the similarity with superposition coding.

In [AR05, GA05], a precoder based on precoder I is designed for suboptimal receivers following the BER balancing criterion (BBC), which states that all components of the SM codeword **C** should be detected with a similar target error rate. While it is not directly related to the PEP/SER criterion used in the derivation of the aforementioned precoder, the power allocation obtained for $|t| = 1$ with precoder I and the BBC is the same as the one obtained with the optimal PEP/SER criterion. The advantage of this criterion is that the analytical derivation of the precoder is straightforward. Therefore, the power allocation for larger constellations or a higher number of transmit antennas is obtained much more easily than with the PEP/SER criterion.

8.5   Introduction to quantized precoding and antenna selection techniques

We have observed in Chapter 4, as well as in the previous sections, that channel statistics (CDIT) based precoders may significantly improve the error rates w.r.t. to open-loop schemes not requiring any channel knowledge at the transmitter. Yet, the achieved gains are smaller than the potential gains achievable by schemes using perfect CSIT such as those mentioned in Chapter 1. On the other hand, the complexities of both approaches are not comparable in terms of feed back bits. While CDIT based precoders require very low-rate feed back link, perfect CSIT based techniques require high-rate feed backs, especially in mobile scenarios (see Section 3.3.3).

One particular precoding technique that may help reduce the feed back rate is the antenna selection at the receiver and/or the transmitter already discussed in Chapters 1, 4 and 5. We have mentioned that antenna selection allows for reducing the number of RF chains, making it highly attractive from an implementation cost viewpoint. Furthermore, it is also attractive in terms of feed back because the receiver may only feed back the index of the selected antennas to the transmitter, thereby limiting the feed back rate. However, remember that if antenna selection techniques extract the full diversity, they lack in providing any transmit array gain, unlike full-CSIT or channel statistics based techniques.

To limit the feed back rate, an approach that subsumes antenna selection as a particular case consists in exploiting a limited amount of feed back at the transmitter. In this context, many results have applied this methodology to the design of beamforming [NLTW98, MSEA03, LHS03, LH05d, LH06, ZWG05, XG04], O-STBCs [LGSW02, LH05a, JS04] and Spatial Multiplexing schemes [LH05b, LH05c]. As already outlined in the introduction of this chapter, the use of a codebook of precoding matrices leads to lower feed back and improved performance. The idea behind quantized precoding is that precoders are chosen within a finite codebook known by both the transmitter and the receiver. The receiver simply selects the optimal precoder based on the current CSIT and feeds back the index

of this precoder to the transmitter. We assume throughout this section that the feed back channel is zero-delay and error-free.

For three different schemes (beamforming, O-STBC and Spatial Multiplexing), we provide in the sequel the selection and codebook design criteria in order to minimize the error probability. In the last section, we also deal with a codebook design maximizing the channel capacity.

8.6 Quantized precoding and antenna selection for dominant eigenmode transmissions

This transmission scheme is analyzed in [LHS03, LH05d, LH06]. As outlined by (1.58) in Chapter 1, a dominant eigenmode transmission is described by the following system model

$$\mathbf{y} = \sqrt{E_s}\mathbf{H}\mathbf{w}_t c + \mathbf{n}, \tag{8.74}$$

$$z = \mathbf{w}_r^H \mathbf{y},$$

$$= \sqrt{E_s}\mathbf{w}_r^H \mathbf{H}\mathbf{w}_t c + \mathbf{w}_r^H \mathbf{n} \tag{8.75}$$

where $\mathbf{w}_r$ and $\mathbf{w}_t$ are respectively $n_r \times 1$ and $n_t \times 1$ vectors. Performing MRC at the receiver in order to maximize the SNR, we choose $\mathbf{w}_r = (\mathbf{H}\mathbf{w}_t)^H$ and the optimal beamforming vector $\mathbf{w}_t$ that maximizes the SNR is given by

$$\mathbf{w}_t^\star = \arg \max_{\mathbf{w} \in \mathcal{C}_w} \|\mathbf{H}\mathbf{w}\|^2 \tag{8.76}$$

where the maximization is performed over all possible precoders $\mathbf{w}$ contained in the set of unit-norm vectors $\mathcal{C}_w$. The unit-norm constraint is a direct consequence of the total transmit power constraint $\|\mathbf{w}\|^2 = 1$. If $\mathcal{C}_w$ is taken as the entire set of unit vectors in the n_t-dimensional complex vector space $\mathbb{C}^{n_t}$, the solution has been shown in Chapter 1 to be given by the dominant right singular vector of $\mathbf{H}$ (i.e. corresponding to the largest singular value of $\mathbf{H}$). If such a choice is made, $\mathbf{w}_r = (\mathbf{H}\mathbf{w}_t)^H$ is simply given by the dominant left singular vector of the channel matrix.

Yet, the goal is now to reduce the number of feed back bits. To this end, we need to introduce a quantization method. This limits the space over which the precoders can be chosen to a set of finite cardinality n_p, called codebook, that is known to both the transmitter and the receiver. The receiver evaluates the best precoder $\mathbf{w}_t$ in the codebook, following the selection procedure of (8.76), but only feed back its index in the codebook. The number of required feed back bits n_b is a direct function of the size of the codebook, $n_b = \lceil \log_2 (n_p) \rceil$.

8.6.1   Selection criterion and codebook design in I.I.D. Rayleigh fading channels

Let us assume in this section that the channel is i.i.d. Rayleigh distributed ($\mathbf{H} = \mathbf{H}_w$). We resort to limit the space $\mathcal{C}_w$ over which $\mathbf{w}$ can be chosen to a codebook called $\mathcal{W}$. We also denote by $\mathbf{M}_w = [\mathbf{w}_1 \ldots \mathbf{w}_{n_p}]$ the codebook matrix obtained by stacking all precoders within one matrix. The receiver evaluates the best precoder $\mathbf{w}_t$ among all precoders $\mathbf{w}_i \in \mathcal{W}$ (with $i = 1, \ldots, n_p$) such that

$$\mathbf{w}_t = \arg \max_{\substack{1 \leq i \leq n_p \\ \mathbf{w}_i \in \mathcal{W}}} \|\mathbf{H}\mathbf{w}_i\|^2. \tag{8.77}$$

In order to design the codebook, a distortion function is introduced. If non-quantized dominant eigenmode transmission were performed, an array gain of $\|\mathbf{H}\mathbf{w}_t\|^2 = \lambda_{max}$ (with λ_{max} the largest eigenvalue value of $\mathbf{H}\mathbf{H}^H$) would be obtained. Hence, the chosen distortion function d_f is a measure of the average (over all channel realizations) array gain loss induced by the quantization process

$$d_f = \mathcal{E}_{\mathbf{H}}\Big\{\lambda_{max} - \|\mathbf{H}\mathbf{w}_t\|^2\Big\}. \tag{8.78}$$

Using the SVD decomposition of $\mathbf{H}$, d_f is easily upper-bounded by

$$d_f \leq \mathcal{E}_{\mathbf{H}}\{\lambda_{max}\}\mathcal{E}_{\mathbf{H}}\Big\{1 - |\mathbf{v}_{max}^H \mathbf{w}_t|^2\Big\} \tag{8.79}$$

where $\mathbf{v}_{max}$ is the dominant right singular vector of $\mathbf{H}$ (i.e. the right singular vector of $\mathbf{H}$ that corresponds to the largest singular value $\sqrt{\lambda_{max}}$). We note that the distortion measure is upper-bounded by the product of two terms. The first term $\mathcal{E}_{\mathbf{H}}\{\lambda_{max}\}$ indicates the quality of the channel. The second term $\mathcal{E}_{\mathbf{H}}\big\{1 - |\mathbf{v}_{max}^H \mathbf{w}_t|^2\big\}$ is an indication of the codebook quality. We also observe that if the condition number of the channel matrix increases and λ_{max} becomes largely dominant over the other eigenvalues (typically, in correlated Rayleigh or Ricean channels), this upper-bound on d_f is maintained only by improving the codebook quality.

It is further shown in [LHS03] that minimizing the upper-bound of (8.79) is equivalent to maximizing the minimum distance

$$\delta_{line}(\mathcal{W}) = \min_{1 \leq k < l \leq n_p} \sqrt{1 - |\mathbf{w}_k^H \mathbf{w}_l|^2}. \tag{8.80}$$

The above discussion results, therefore, in the following codebook design criterion for quantized dominant eigenmode transmission.

Design criterion 8.1   *Choose a codebook $\mathcal{W}$ made of n_p unit-norm vectors $\mathbf{w}_i$ ($i = 1, \ldots, n_p$) such that the minimum distance $\delta_{line}(\mathcal{W})$ is maximized.*

The quantity in (8.80) is a measure of the minimum distance between the spaces spanned by two vectors of the codebook $\mathcal{W}$. Since $\mathbf{w}_i$ is a unit-norm vector, the space spanned by

$\mathbf{w}_i$ is usually referred to as a line. A line passes through the origin of $\mathbb{C}^{n_t}$ and the distance $\delta_{line}(\mathcal{W})$ between two vectors in (8.80) can be thought of as the sine of the angle between the so-corresponding lines.

Hence, the goal of Design criterion 8.1 is to pack n_p lines in $\mathbb{C}^{n_t}$ in such a way that the minimum distance between any pair of lines is maximized. This problem is commonly known as the Grassmannian line packing problem. Naturally, its solution is trivial when the number of lines is smaller than the dimension of the space, i.e. when $n_p \leq n_t$. In this case, it is indeed sufficient to take n_p vectors out of the n_t columns of any $n_t \times n_t$ unitary matrix.

Note that additional constraints on the precoder structure may be added in practice. For example, it might be useful to constraint the precoders entries to have an equal magnitude $1/\sqrt{n_t}$. This can be thought of as a quantized equal gain transmission applied to MIMO systems. Because Criterion 8.1 is general, it can be used to design equal gain precoders [LHS03, LH03], which perform equally well as more general precoders based on quantized dominant eigenmode transmission.

### 8.6.2   Antenna selection and achievable diversity gain

We know from Chapter 1 that perfect dominant eigenmode transmission achieves the full diversity. Therefore, it is of particular importance to investigate how large the codebook should be in order to guarantee full diversity with the quantized dominant eigenmode transmission. To this end, let us consider some results from antenna selection. Antenna selection is indeed a particular case of a quantized precoding whose codebook is chosen as the columns of the identity matrix $\mathbf{I}_{n_t}$. It is shown in [LHS03] that a codebook larger than the number of transmit antennas (i.e. $n_p \geq n_t$) with a full-rank codebook matrix $\mathbf{M}_w$ (i.e. $r(\mathbf{M}_w) = n_t$) satisfies the following relationship

$$\max_{1 \leq i \leq n_p} \|\mathbf{H}\mathbf{w}_i\|^2 \geq \frac{\sigma_{n_t}^2(\mathbf{M}_w)}{n_p n_r} \max_{k,l} |\mathbf{H}(k,l)|^2 \tag{8.81}$$

with $\sigma_{n_t}(\mathbf{M}_w)$ the n_t^{th} largest singular value of $\mathbf{M}_w$ (which is also the minimum singular value). Interestingly, $\max_{k,l} |\mathbf{H}(k,l)|^2$ is the equivalent channel gain of a system with $n_t n_r$ antennas performing antenna selection. Considering that antenna selection delivers the full diversity and that a codebook constructed following Criterion 8.1 has a full-rank codebook matrix when $n_p \geq n_t$, we can conclude that this criterion also provides codebooks extracting the full diversity. Furthermore, the term $\sigma_{n_t}(\mathbf{M}_w)/(n_p n_r)$ can be thought of as a lower-bound on the array gain, suggesting that codebook matrices with small singular values are not advised.

To sum up, choosing the vectors $\mathbf{w}_i$ as the columns of any $n_t \times n_t$ unitary matrix is the most simple way to extract a full diversity. Yet, by the unitary invariance of the Gaussian
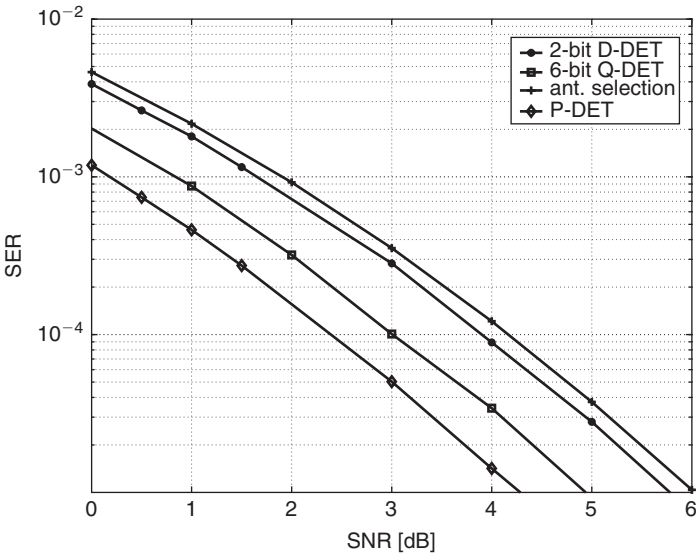
Figure 8.11   *Symbol error rate (SER) of a 3×3 MIMO system using 2-bit and 6-bit quantized BPSK-based dominant eigenmode transmissions (courtesy of D. Love [LHS03])*

Table 8.1   *Codebook for quantized dominant eigenmode transmission for $n_t = 3$ and $n_p = 4$*

$$\mathcal{W} = \left\{ \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} \frac{j}{\sqrt{3}} \\ \frac{-1}{\sqrt{3}} \\ \frac{-j}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} \frac{-1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} \frac{-j}{\sqrt{3}} \\ \frac{-1}{\sqrt{3}} \\ \frac{j}{\sqrt{3}} \end{bmatrix} \right\} \qquad (8.82)$$

distribution, we know that the performance is not affected by the multiplication of each codebook element by a unitary precoder. This means that choosing the vectors $\mathbf{w}_i$ as columns of a unitary matrix is essentially equivalent to choosing those vectors as columns of $\mathbf{I}_{n_t}$, which simply reduces to selecting the transmit antenna that maximizes the received SNR. By adjusting the size of the codebook, we are also able to improve the array gain, achieving more flexibility than a scheme based only on antenna selection.

Figure 8.11 depicts the performance of 2-bit and 6-bit quantized BPSK dominant eigenmode transmissions (respectively denoted as '2-bit Q-DET' and '6-bit Q-DET') over a 3×3 i.i.d. Rayleigh fading channel. The codebooks have been designed following Criterion 8.1 and the optimal constructions available in [HMR+00]. An example of codebook for the 2-bit transmission in given in Table 8.1. Comparisons with perfect dominant eigenmode transmission ('P-DET') and conventional antenna selection are also given. A gain of 0.2 dB is achieved by '2-bit Q-DET' over the antenna selection for the same amount of

feed back. Using a 6-bit quantization provides an extra 0.9 dB gain but remains outperformed by the 'P-DET' by about 0.6 dB.

8.6.3 How many feed back bits are required?

Naturally, the use of quantized dominant eigenmode transmission is associated with a capacity loss w.r.t. to perfect dominant eigenmode transmission. To compare the ergodic capacities achieved by perfect and quantized dominant eigenmode transmissions, respectively denoted as $\bar{C}_{perfect}$ and $\bar{C}_{quant}$, [LHS03] identifies the minimum codebook size n_p required to guarantee a normalized capacity loss of $\bar{C}_{loss} = \left(\bar{C}_{perfect} - \bar{C}_{quant}\right)/\bar{C}_{perfect}$ as

$$n_p \gtrsim \left(1 - \bar{C}_{loss}\right) \left(\frac{4n_t}{n_t - 1}\right)^{n_t - 1}. \tag{8.83}$$

Similarly, in order to guarantee a normalized SNR loss $\rho_{loss} = d_f / \mathcal{E}_{\mathbf{H}}\{\lambda_{max}\}$, the codebook size should satisfy

$$n_p \gtrsim \left(1 - \rho_{loss}\right) \left(\frac{4n_t}{n_t - 1}\right)^{n_t - 1} \left(1 - \frac{n_t - 1}{4n_t}\right)^{-1}. \tag{8.84}$$

8.6.4 Selection criterion and codebook design in spatially correlated Rayleigh fading channels

Analogous to [LH06], let us now investigate how to modify the codebook design when the channel is spatially correlated Rayleigh distributed and the spatial correlation matrix is known to the transmitter. For simplicity, we assume a MISO transmission, so that the channel matrix is written as $\mathbf{h} = \mathbf{h}_w \mathbf{R}_t^{1/2}$, with $\mathbf{h}$ and $\mathbf{h}_w$ being $1 \times n_t$ vectors. As already mentioned in Chapter 1, the dominant eigenmode precoder for $n_r = 1$ is the matched beamformer (also known as maximum ratio transmission in analogy with maximum ratio combining)

$$\mathbf{w}_t = \frac{\mathbf{R}_t^{H/2} \mathbf{w}_w^H}{\left\| \mathbf{R}_t^{1/2} \mathbf{w}_w \right\|} \tag{8.85}$$

with $\mathbf{w}_w = \mathbf{h}_w / \|\mathbf{h}_w\|$. In i.i.d. Rayleigh fading channels, the optimal precoder would have been given by $\mathbf{w}_w^H$. In correlated Rayleigh channels, the matched beamforming is also a function of $\mathbf{R}_t^{1/2}$, modifying the transmit power. Hence, a normalization by $\left\| \mathbf{R}_t^{1/2} \mathbf{w}_w \right\|$ is necessary.

To design the quantized precoder, we may account for the fact that $\mathbf{R}_t$ is fixed and known to the transmitter. Subsequently, the precoding vectors only depend on $\mathbf{w}_w$, and the codebook
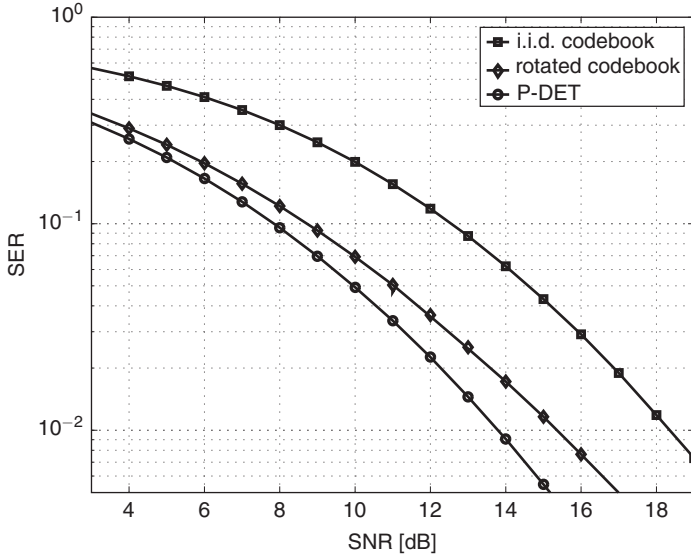
Figure 8.12   *Symbol error rate (SER) of a 8×1 MISO system using 6-bit quantized i.i.d. and rotated dominant eigenmode transmissions in correlated Rayleigh channels (courtesy of D. Love [LH06])*

$\mathcal{W}_c$ in correlated channels may be written as

$$\mathcal{W}_c = \left\{ \frac{\mathbf{R}_t^{1/2}\mathbf{w}_1}{\left\|\mathbf{R}_t^{1/2}\mathbf{w}_1\right\|}, \ldots, \frac{\mathbf{R}_t^{1/2}\mathbf{w}_{n_p}}{\left\|\mathbf{R}_t^{1/2}\mathbf{w}_{n_p}\right\|} \right\} \tag{8.86}$$

where each $\mathbf{w}_i$ is a unit-norm vector. In practice, it is even better to choose the codebook $\mathcal{W}$ such that the first element of $\mathcal{W}_c$ is equal to the dominant eigenvector of $\mathbf{R}_t$. Therefore single mode eigenbeam forming is included in the codebook. Defining a distortion function similarly to (8.78), it can be concluded [LH06] that the set of vectors $\{\mathbf{w}_i\}$ may also be designed following Criterion 8.1. Hence, the codebook $\mathcal{W}_c$ in correlated channels is simply constructed by rotating the codebook $\mathcal{W}$ derived for i.i.d. Rayleigh channels and by normalizing each element according to (8.86). The gain offered by the rotation is illustrated in Figure 8.12 where a 8×1 MISO channel is considered using the 802.11 TGn model (see Section 2.5.1 in Chapter 2) assuming three clusters and a 5-degree azimuth-spread. Comparisons between the i.i.d. codebook, the rotated codebook and perfect maximum ratio transmission ('P-DET') are illustrated. We note that the rotation provides a gain of about 3 dB.

8.7   Quantized precoding and antenna selection for orthogonal space–time block coding

In Section 8.2, we have shown how to design precoders for O-STBC when channel statistics are available at the transmitter. We now investigate the case of limited feed back

using a SNR-based criterion [LH05a]. This will result in a design criterion very similar to Criterion 8.1.

8.7.1 Selection criterion and codebook design

The problem is as follows. We want to use an orthogonal code designed for n_e transmit antennas in combination with a precoder[2] $\mathbf{W}$ of size $n_t \times n_e$ (n_t can be larger than n_e). Thanks to the orthogonality of the code, the pairwise error probability simply reads as

$$P(\mathbf{C} \to \mathbf{E}|\mathbf{H}) = \mathcal{Q}\left(\sqrt{\frac{\rho}{2}\|\mathbf{HW}Td^2/(Qn_e)\|_F^2}\right) \tag{8.87}$$

where d^2 is defined as in (8.5). To minimize the PEP, the optimal precoder $\mathbf{W}_t \in \mathcal{C}_{\mathbf{W}}$ should satisfy

$$\mathbf{W}_t = \arg\max_{\mathbf{W}\in\mathcal{C}_{\mathbf{W}}} \|\mathbf{HW}\|_F^2. \tag{8.88}$$

Note the similitude with (8.76). If $\mathcal{C}_{\mathbf{W}}$ is chosen as $\mathbb{C}^{n_t \times n_e}$, the optimal precoder $\mathbf{W}$ consists in transmitting only in the dominant eigenvector of $\mathbf{H}$, i.e. in performing dominant eigenmode transmission for which quantized precoders have been proposed in Section 8.6.

However, if we impose a peak power constraint imposing that the largest singular value of $\mathbf{W}$ remains smaller than one, the situation is radically different. Let us assume for simplicity that the singular values of $\mathbf{W}$ and $\mathbf{H}$ are ranked in decreasing order. With this notation, the peak power constraint is simply written as

$$\sigma_1(\mathbf{W}) \leq 1. \tag{8.89}$$

Under this constraint, it is clear that $\|\mathbf{HW}\|_F \leq \sum_{i=1}^{n_e} \sigma_i(\mathbf{H})$, and equality occurs when $\mathbf{W}$ is made of the first n_e columns of the right singular matrix of $\mathbf{H}$. Denoting the SVD of $\mathbf{H}$ as $\mathbf{U_H}\mathbf{\Sigma_H}\mathbf{V_H^H}$, $\mathbf{W}$ should thereby be chosen as the first n_e columns of $\mathbf{V_H}$, denoted as $\hat{\mathbf{V}}_{\mathbf{H}}$. More generally, the precoders in the codebook should be chosen as tall unitary matrices (i.e. its columns are orthogonal). Indeed since the peak power constraint imposes the singular values of the precoder to be smaller than 1, we should take all singular values as large as possible (i.e. having them all equal to 1) in order to achieve larger $\|\mathbf{HW}\|_F$.

Let us now design the codebook $\mathcal{W}$, containing the precoders $\mathbf{W}_i$, $i = 1, \ldots, n_p$, which are tall unitary matrices. The receiver computes the best precoder $\mathbf{W}_t$ according to the following SNR maximization

$$\mathbf{W}_t = \arg\max_{\mathbf{W}_i \in \mathcal{W}} \|\mathbf{HW}_i\|_F^2 \tag{8.90}$$

---

[2] $\mathbf{W}$ is used instead of $\mathbf{P}$ to differentiate CDIT based and quantized precoders.

and feeds its index back to the transmitter. Analogous to quantized dominant eigenmode transmission, a distortion measure is defined and upper-bounded as

$$d_f = \mathcal{E}_{\mathbf{H}} \left\{ \min_{\mathbf{W}_i \in \mathcal{W}} \left[\|\mathbf{H}\hat{\mathbf{V}}_{\mathbf{H}}\|_F^2 - \|\mathbf{H}\mathbf{W}_i\|_F^2 \right] \right\} \tag{8.91}$$

$$\leq \frac{1}{2} \mathcal{E}_{\mathbf{H}}\{\lambda_1(\mathbf{H})\} \mathcal{E}_{\mathbf{H}} \left\{ \min_{\mathbf{W}_i \in \mathcal{W}} \left\| \hat{\mathbf{V}}_{\mathbf{H}} \hat{\mathbf{V}}_{\mathbf{H}}^H - \mathbf{W}_i \mathbf{W}_i^H \right\|_F^2 \right\}. \tag{8.92}$$

We can further bound the term related to the codebook quality, since minimizing this bound comes to maximize the distance between the column spaces spanned by each precoder matrix [LH05a]. This distance is commonly called the chordal distance, and its minimum value over all precoders is given by

$$\delta_{ch}(\mathcal{W}) = \min_{1 \leq k < l \leq n_p} \frac{1}{\sqrt{2}} \left\| \mathbf{W}_k \mathbf{W}_k^H - \mathbf{W}_l \mathbf{W}_l^H \right\|_F. \tag{8.93}$$

By contrast to the dominant eigenmode transmission, the problem consists now in finding n_p subspaces in the set of all column spaces of tall unitary ($n_t \times n_e$) matrices with the largest minimum chordal distance. This is commonly referred to as the Grassmannian subspace packing problem. The subsequent design criterion simply reads as follows.

Design criterion 8.2   *Choose a codebook $\mathcal{W}$ made of n_p tall unitary matrices $\mathbf{W}_i$ ($i = 1, \ldots, n_p$) of size $n_t \times n_e$ such that the minimum chordal distance $\delta_{ch}(\mathcal{W})$ is maximized.*

In [LH05a], it is proposed to use the non-coherent space–time modulation designs of [HMR+00] to construct the codebook. The proposed codebook $\mathcal{W}$ has the following form:

$$\mathcal{W} = \{\mathcal{D}, \mathbf{\Theta}\mathcal{D}, \ldots, \mathbf{\Theta}^{n_p-1}\mathcal{D}\} \tag{8.94}$$

where the $n_t \times n_e$ matrix $\mathcal{D}$ has a structure similar to a DFT matrix with its $(k, l)^{\text{th}}$ entry given by $1/\sqrt{n_t}e^{j(2\pi/n_t)kl}$. The diagonal matrix $\mathbf{\Theta}$ reads as

$$\mathbf{\Theta} = \text{diag}\left\{ e^{j\frac{2\pi}{n_p}u_1}, \ldots, e^{j\frac{2\pi}{n_p}u_{n_t}} \right\}. \tag{8.95}$$

The value of each u_i is selected in the finite set of integers $\{0, \ldots, n_p - 1\}$. Over the $n_p^{n_t}$ possible combinations for $\mathbf{\Theta}$, the best $\mathbf{\Theta}$ is selected in order to maximize the minimum chordal distance between any pair of precoders in (8.94).

8.7.2   Antenna subset selection and achievable diversity gain

What is the diversity order achieved by quantized precoders for O-STBC? We know already that antenna subset selection is a form of quantized precoding. Indeed, remember that

antenna subset selection chooses from the n_t columns of $\mathbf{I}_{n_t}$ the n_e columns that minimize the probability of error, which is equivalent to maximize the received SNR in the O-STBC context. In [GP02], such a selection scheme is examined in combination with antenna selection at the receiver (n_r' antennas are selected out of n_r). The number of possible selections is then given by $\binom{n_t}{n_e}\binom{n_r}{n_r'}$. Let us denote as $\mathbf{H}'$ the matrix constructed from the remnant rows and columns of $\mathbf{H}$ after antenna selection at both the transmitter and receiver. Reasoning along the lines of [PNG03], we easily write the inequality

$$\frac{\|\mathbf{H}\|_F^2}{n_e n_r'} \geq \frac{\|\mathbf{H}'\|_F^2}{n_e n_r'} \geq \frac{\|\mathbf{H}\|_F^2}{n_t n_r}, \tag{8.96}$$

suggesting that O-STBC combined with antenna subset selection is able to exploit all degrees of freedom of the original channel $\mathbf{H}$ and therefore achieve the full diversity gain of $n_t n_r$. Let us now extend this result to the case where a general quantized precoder is used. It is shown in [LH05a] that

$$\max_{\mathbf{W}_i \in \mathcal{W}} \|\mathbf{H}\mathbf{W}_i\|_F^2 \geq \frac{\lambda_{n_t}(\mathbf{M}_W)}{n_p n_e n_t^2 n_r} \|\mathbf{H}\|_F^2 \tag{8.97}$$

where $\lambda_{n_t}(\mathbf{M}_W)$ is the n_t^{th} largest eigenvalue of the codebook matrix $\mathbf{M}_W = [\mathbf{W}_1 \dots \mathbf{W}_{n_p}]$. By taking $n_p \geq n_t/n_e$ and ensuring that the columns of $\mathbf{M}_W$ span $\mathbb{C}^{n_t}$ (i.e. $r(\mathbf{M}_W) = n_t$), $\sigma_{n_t}(\mathbf{M}_W)$ becomes the minimum non-zero singular value of $\mathbf{M}_W$. The scalar $\|\mathbf{H}\|_F^2$ is the equivalent channel resulting in an O-STBC transmission in a $n_r \times n_t$ MIMO channel. Therefore, analogous to antenna subset selection, the full diversity gain of $n_t n_r$ is achieved if $n_p \geq n_t/n_e$ and the columns of the precoders span $\mathbb{C}^{n_t}$. Furthermore, $\lambda_{n_t}(\mathbf{M}_W)/(n_p n_e n_t^2 n_r)$ is a lower-bound on the achievable array gain, suggesting to design the codebook matrix with large singular values.

In brief, quantized precoding can be used with O-STBC over any number of transmit antennas and can still achieve the full diversity. Moreover, it may offer more flexibility than antenna subset selection [GP02].

The performance of quantized precoder for O-STBC is shown in Figure 8.13. We consider a 2×4 system using a precoded QPSK Alamouti scheme in i.i.d. Rayleigh fading channels. The large gain offered by antenna selection w.r.t. the unprecoded Alamouti scheme originates the achieved diversity gain, equal to 4 for antenna selection but only 2 for the Alamouti scheme. The 3-bit precoded Alamouti scheme presents a slight gain over antenna selection for the same complexity. Indeed, antenna selection selects 2 antennas out of 4, requiring also $n_b = \lceil \log_2 \binom{4}{2} \rceil = 3$ bits of feed back. The 6-bit precoded Alamouti scheme offers an additional gain over the 3-bit precoder. In practical scenarios, it is advisable to verify the exact gain resulting from increasing n_b, as the relative gain becomes negligible for large n_b.

Note that antenna subset selection may also rely on channel statistics rather than on the full or quantized CSIT. As an example, in [GP02], the selection is based on the transmit
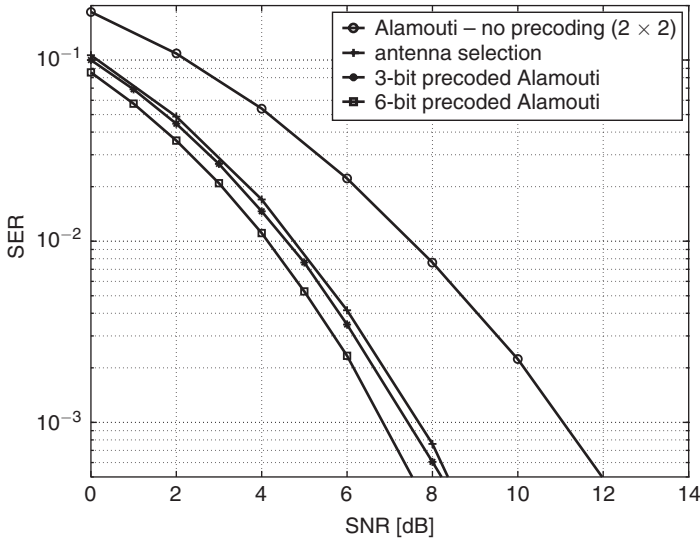
Figure 8.13   *Symbol error rate (SER) of 3-bit and 6-bit precoded Alamouti schemes in 2×4 i.i.d. Rayleigh fading channels (courtesy of D. Love [LH05a])*

and receive correlation matrices $\mathbf{R}_t$ and $\mathbf{R}_r$ and uses an average PEP based design for full-rank codes at high SNR. The limitations of the high SNR assumption have been detailed in Chapter 6. However, we might remember that the high SNR regime is a realistic regime for O-STBC (this might not be so for other codes). From (6.30) and (6.35), a simple selection algorithm that minimizes the average error probability over all channel realizations consists in choosing the subset of transmit and receive antennas that maximize $\det\left(\mathbf{R}'_t\right)$ and $\det\left(\mathbf{R}'_r\right)$. $\mathbf{R}'_t\left[n_e \times n_e\right]$ and $\mathbf{R}'_r\left[n'_r \times n'_r\right]$ are the transmit and receive correlation matrices obtained after subset selection. This selection algorithm attempts to maximize the coding gain by selecting the subset of antennas that present low antenna correlation. At high SNR, this selection procedure does not have any impact on the diversity gain, but the maximum diversity gain is only $n_e n'_r$, unlike CSIT based antenna selection.

8.8   Quantized precoding and antenna selection for Spatial Multiplexing

Consider now that a Spatial Multiplexing scheme is used in place of O-STBC. Therefore, we transmit n_e independent symbols that are precoded by a $n_t \times n_e$ matrix $\mathbf{W}$ spreading the n_e streams over the n_t physical antennas. We assume that the channel is i.i.d. Rayleigh distributed and that n_r is larger than n_e to enable suboptimal decoding. Two classes of receivers are investigated, based on ML decoding or suboptimal ZF decoding. To derive the selection and codebook design criteria minimizing the probability of error [LH05b], we again account for a peak power constraint, so that the largest singular value of each precoder is upper-bounded by one (as already observed, this designs the precoders as tall unitary matrices).

8.8.1 Selection criterion and codebook design

Unlike O-STBC, the performance of SM is not exclusively a function of the Frobenius norm of the channel. Recalling from Chapter 5 that the conditional PEP with ML decoding is directly related to the minimum distance at the receiver, the optimal precoder $\mathbf{W}_t \in \mathcal{C}_\mathbf{W}$ should satisfy

$$\mathbf{W}_t = \arg \max_{\mathbf{W} \in \mathcal{C}_\mathbf{W}} D^2 \tag{8.98}$$

with $D^2 = \min_{\mathbf{C} \neq \mathbf{E}} \|\mathbf{HW}(\mathbf{C} - \mathbf{E})\|^2$ where $\mathbf{C}$ and $\mathbf{E}$ are two different SM codewords (see Appendix A (5.80)). A useful lower-bound on D^2 is easily derived as (see Appendix A)

$$D^2 \geq \min_{\mathbf{C} \neq \mathbf{E}} \|\mathbf{C} - \mathbf{E}\|^2 \sigma^2_{min}(\mathbf{HW}) \tag{8.99}$$

with $\sigma^2_{min}(\mathbf{HW})$ denoting the minimum non-zero singular value of $\mathbf{HW}$.

For ZF filtering, in Chapter 5 we have seen that the performance is mainly a function of the worst SNR among the n_e streams given by (5.105)

$$\rho_{min} = \min_{1 \leq q \leq n_e} \frac{\rho}{n_e} \frac{1}{(\mathbf{W}^H \mathbf{H}^H \mathbf{HW})^{-1}(q, q)}. \tag{8.100}$$

While it is possible to use the above expression of ρ_{min} to select the precoders, it has been proposed in [LH05b] to use a lower-bound of ρ_{min}

$$\rho_{min} \geq \frac{\rho}{n_e} \frac{1}{\lambda_{max}((\mathbf{W}^H \mathbf{H}^H \mathbf{HW})^{-1})} = \frac{\rho}{n_e}[\sigma_{min}(\mathbf{HW})]^2 \tag{8.101}$$

with $\lambda_{max}((\mathbf{W}^H \mathbf{H}^H \mathbf{HW})^{-1})$ the maximum eigenvalue of $(\mathbf{W}^H \mathbf{H}^H \mathbf{HW})^{-1}$. Note that this selection criterion has originally been proposed by [HSP01] for the particular case of antenna subset selection for Spatial Multiplexing. In this case, the precoders can be chosen as matrices whose columns are n_e columns of $\mathbf{I}_{n_t}$. The selection criterion for ZF filtering eventually consists in choosing $\mathbf{W}_t \in \mathcal{C}_\mathbf{W}$ as

$$\mathbf{W}_t = \arg \max_{\mathbf{W} \in \mathcal{C}_\mathbf{W}} \sigma_{min}(\mathbf{HW}). \tag{8.102}$$

Interestingly, we observe that even if the selection criteria are different for ML and ZF decoding, they are both related to $\sigma_{min}(\mathbf{HW})$. This suggests designing the codebook in such a way that $\sigma_{min}(\mathbf{HW})$ is maximized. Under the peak power constraint, if $\mathcal{C}_\mathbf{W}$ is chosen as the set of tall unitary matrices, the application of the inclusion principle (see Appendix A) leads to

$$\sigma_{min}(\mathbf{HW}_t) = \sigma_{n_e}(\mathbf{HW}_t) \leq \sigma_{n_e}(\mathbf{H}), \tag{8.103}$$

where singular values are ordered by decreasing amplitude. The upper-bound is achieved if $\mathbf{W}_t = \hat{\mathbf{V}}_\mathbf{H}$. For a finite codebook $\mathcal{W}$, the distortion function to be minimized may thus be written as

$$d_f = \mathcal{E}_\mathbf{H} \left\{ \lambda_{min}(\mathbf{H}\hat{\mathbf{V}}_\mathbf{H}) - \max_{\mathbf{W}_i \in \mathcal{W}} \lambda_{min}(\mathbf{HW}_i) \right\}, \tag{8.104}$$

and $\mathbf{W}_t$ is chosen according to (8.98) or (8.102) in the finite set $\mathcal{C}_{\mathbf{W}} = \mathcal{W}$. This minimization is shown in [LH05b] to be related to the maximization of the minimum of the projection two-norm distance between any pair of precoder column spaces

$$\delta_p(\mathcal{W}) = \min_{1 \leq k < l \leq n_p} \sqrt{1 - \lambda_{min}\big(\mathbf{W}_k^H \mathbf{W}_l\big)}. \tag{8.105}$$

Hence, the design criterion reads as follows.

Design criterion 8.3 *Choose a codebook $\mathcal{W}$ made of n_p tall unitary matrices $\mathbf{W}_i$ ($i = 1, \ldots, n_p$) of size $n_t \times n_e$ such that the minimum projection two-norm distance $\delta_p(\mathcal{W})$ is maximized.*

Those codebooks can be designed using the non-coherent constellations in [HMR+00]. Note that if $n_e = 1$, the projection two-norm distance reduces to (8.80), since the error probability is then a function of the received SNR. We should further mention that [LH05b] also proposes selection and codebook design criteria for MMSE receivers.

8.8.2 *Impact of decoding strategy on error probability*

In Figure 8.14, we illustrate the performance of a 2-stream 6-bit limited feed back precoding for Spatial Multiplexing in 2×4 i.i.d. Rayleigh fading channels. The error rates obtained by the minimum singular value selection criterion of (8.102) with ZF decoding ('quantized precoding (MSV-ZF)') and the ML selection criterion in (8.98) with ML decoding ('quantized precoding (ML)') are displayed. For the sake of comparison, the error rates achieved without precoding and with an optimal MMSE based precoding, using a perfect channel knowledge and a total power constraint [SSB+02], are also depicted. Clearly, the quantized precoder significantly decreases the error rate compared to non-precoded transmissions. The non-quantized MMSE precoder with total power constraint has only a 1.5 dB gain over the ZF quantized precoding (MSV-ZF) with peak power constraint (as the total power constraint fully exploits the transmit array gain), but is outperformed by the quantized ML precoder thanks to the larger diversity gain of the latter (at the price of a much larger complexity). Finally, it is important to stress that a suboptimal decoder combined with quantized precoding outperforms non-precoded ML-decoded schemes. This shows that it is possible to achieve a better performance than ML decoding with a lower complexity at the expense of a low-rate feed back.

8.8.3 *Extension to multi-mode precoding*

We have discussed the codebook design for quantized precoded Spatial Multiplexing when the number of streams n_e is fixed. This is also known as a single-mode precoding. Non-negligible improvements can be achieved by considering multi-mode precoding, i.e. letting n_e vary as a function of the channel state under the constraint of a fixed data rate. This requires to adaptively modify the rate of each stream as n_e varies. Hence, rather than
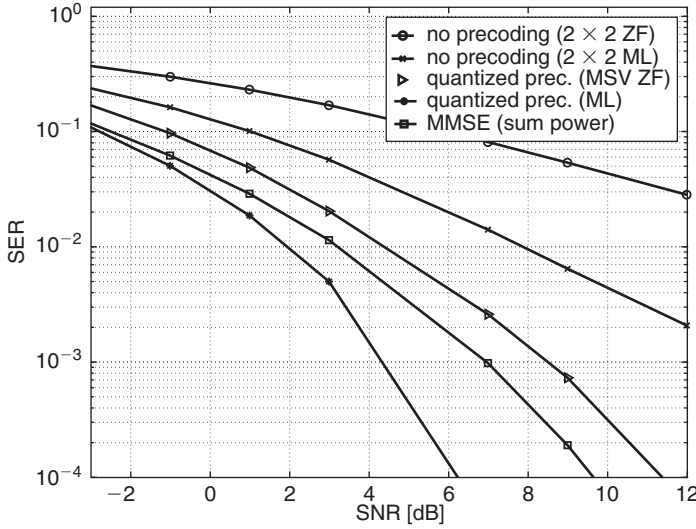
Figure 8.14 *Symbol vector error rate (SVER) of a 6-bit precoded BPSK-based SM scheme in i.i.d. Rayleigh fading channels with $n_t = 4$ and $n_r = 2$ (courtesy of D. Love [LH05b])*

having a single codebook designed for a particular value of n_e, the codebook $\mathcal{W}_{n_e}$ is made of precoders $\mathbf{W}_i^{(n_e)}$ for each supported mode value n_e. The set of supported modes is denoted as $\mathcal{N}_e$ and the size of $\mathcal{W}_{n_e}$ is denoted as $n_p^{(n_e)}$. The selection and codebook design criteria are extensively covered in [LH05c].

The selection procedure valid for both ML and ZF receivers is related to the maximization of the minimum singular value of the effective channel, analogous to (8.102). The first step consists in selecting for each mode n_e the precoder $\mathbf{W}_t^{(n_e)}$ in the codebook $\mathcal{W}_{n_e}$ that maximizes the minimum non-zero singular value of $\mathbf{HW}_i^{(n_e)}$ over all $\mathbf{W}_i^{(n_e)} \in \mathcal{W}_{n_e}$. In otherwords

$$\mathbf{W}_t^{(n_e)} = \arg \max_{\mathbf{W}_i^{(n_e)} \in \mathcal{W}_{n_e}} \sigma_{min}\left(\mathbf{HW}_i^{(n_e)}\right), \ \forall n_e \in \mathcal{N}_e. \tag{8.106}$$

This is similar to single-mode precoding.

The second step is to select among the selected precoders the best mode and the corresponding precoder as the precoder with the largest minimum singular value, taking into account that the constellation used for each mode is different. This is expressed as follows

$$\mathbf{W}_t = \arg \max_{n_e \in \mathcal{N}_e} \frac{\lambda_{min}\left(\mathbf{HW}_t^{(n_e)}\right)}{n_e} d_{min}^2(n_e) \tag{8.107}$$

where $d_{min}^2(n_e)$ is the minimum squared Euclidean distance of the constellation used for mode n_e. In order to keep the data rate constant, larger constellations are used as n_e

decreases. The receiver then feeds back the best mode number and the index of the selected precoder in the corresponding codebook.

Regarding the codebook design, the difficulty of multi-mode precoding is to calculate how to spread efficiently the total number of feed back bits among the different codebooks, i.e. how to distribute the total number of precoding matrices among the total number of supported modes. Since the dominant eigenmode transmission is a particular case of this scheme for $n_e = 1$, we can intuitively expect that multi-mode precoding extracts the full diversity for a fixed total rate if $n_p^{(1)} \geq n_t$. This is actually the case and fixes the required size of the codebook for $n_e = 1$. A more rigorous proof is provided in [LH05c]. Requirements for the case $n_e = n_t$ may also be obtained by assuming that the rate is allowed to grow with the SNR (analogous to the diversity-multiplexing trade-off). As far as the codebook design criterion is concerned, minimizing the error probability with ML or ZF decoding designs each codebook based on the projection two-norm distance of (8.105). Hence, each codebook $\mathcal{W}_{n_e}$ can be designed based on Criterion 8.3. The interested reader is referred to [LH05c] for more information.

An antenna selection technique very close to multi-mode precoding has been proposed in [Nar03]. The idea is to select a subset of antennas over which Spatial Multiplexing is performed. As with multi-mode precoding, the number and subset of transmit antennas together with the symbol constellations are allowed to vary over time under the constraints that equal power is allocated to each antenna and that the data rate is fixed. However, by contrast to multi-mode precoding, the optimization is performed based on the channel statistics (CDIT) rather than on the current channel state information (CSIT). As with statistical antenna selection techniques presented in Section 8.7.2 for O-STBC, this scheme requires a very low feed back rate as the channel statistics vary much slower than the channel itself.

We assume a ZF receiver and use the lower-bound of (8.101). Since antenna selection is applied at the transmitter, the equivalent channel $\mathbf{HW}$ can simply be written as the matrix $\mathbf{H}^{(n_e)}$, which is constructed based on the n_e selected columns of $\mathbf{H}$. In Kronecker-structured Rayleigh fading channels, it is shown in [Nar03] that the lower-bound (8.101) can be further developed as

$$\rho_{min} \; \geq \; \frac{\rho}{n_e} \lambda_{min}\left(\mathbf{H}^{(n_e)}\right) \tag{8.108}$$

$$\geq \; \frac{\rho}{n_e} \lambda_{min}(\mathbf{R}_r)\lambda_{min}\left(\mathbf{R}_t^{(n_e)}\right)\lambda_{min}\left(\mathbf{H}_w^{(n_e)}\right) \tag{8.109}$$

where $\mathbf{R}_t^{(n_e)}$ is the transmit correlation matrix obtained after selection of n_e transmit antennas, and $\mathbf{H}_w^{(n_e)}$ is defined analogously. Making use of this lower-bound, the selection criterion then consists in selecting the number and optimal set of transmit antennas to maximize the minimum SNR margin [Nar03]. However, since $\mathbf{H}_w$ is unknown to the transmitter and because the optimization is only performed based on the transmit correlation matrix (receive antenna selection is not performed, hence the selection criterion does not depend on the receive correlation matrix), the selection procedure relies on the average

value of $\lambda_{min}\left(\mathbf{H}_w^{(n_e)}\right)$ and not on its instantaneous value. The same spectral efficiency is allocated to each antenna and the constellation size is adapted as a function of the number of selected antennas.

8.9 Information theory motivated quantized precoding

Naturally, quantized precoding may be designed in view of maximizing the capacity. Assuming uncorrelated Gaussian signaling on each stream, the receiver chooses the precoder that maximizes the mutual information of the equivalent channel $\mathbf{HW}$

$$\mathcal{I}_e(\mathbf{H}, \mathbf{W}) = \log_2 \det \left[\mathbf{I}_{n_e} + \frac{\rho}{n_e} \mathbf{W}^H \mathbf{H}^H \mathbf{H} \mathbf{W} \right]. \tag{8.110}$$

The optimal precoder is therefore given by

$$\mathbf{W}_t = \arg \max_{\mathbf{W} \in \mathcal{C}_\mathbf{W}} \mathcal{I}_e(\mathbf{H}, \mathbf{W}). \tag{8.111}$$

Under a peak power constraint, the optimal precoder is shown in [LH05b] to be $\mathbf{W}_t = \hat{\mathbf{V}}_\mathbf{H}$. Note that this power constraint eliminates the power allocation optimization since all singular values are fixed to one for unitary precoders. This is radically different from the derivations of Chapter 4, which rely on a sum power constraint. In that case, the optimal scheme consists also in transmitting along the right singular vectors of $\mathbf{H}$ but the total power must be distributed over all streams based on a water-filling solution. Note that if $N_T \geq \mathcal{W}_t$, $\hat{\mathbf{V}}_\mathbf{H}$ is square unitary, and precoding does not increase the capacity unless a mode selection is performed.

To measure the capacity loss resulting from the choice of a precoder in a finite set $\mathcal{W}$ instead of the optimal precoder $\mathbf{W}_t = \hat{\mathbf{V}}_\mathbf{H}$, [LH05b] defines the average distortion function as

$$d_f = \mathcal{E}_\mathbf{H} \left\{ \det \left[\mathbf{I}_{n_e} + \frac{\rho}{n_e} \hat{\mathbf{\Lambda}}_\mathbf{H}^T \hat{\mathbf{\Lambda}}_\mathbf{H} \right] \right\} \left(1 - \mathcal{E}_\mathbf{H} \left\{ \max_{\mathbf{W}_i \in \mathcal{W}} \left| \det \left(\hat{\mathbf{V}}_\mathbf{H}^H \mathbf{W}_i \right) \right|^2 \right\} \right) \tag{8.112}$$

whose interpretation is similar to (8.79). The main diagonal of $\hat{\mathbf{\Lambda}}_\mathbf{H}$ of size $n_r \times n_e$ is made of the n_e largest singular values of $\mathbf{H}$ in decreasing order. Hence, $\hat{\mathbf{\Lambda}}_\mathbf{H}^T \hat{\mathbf{\Lambda}}_\mathbf{H}$ is diagonal with its entries given by the n_e largest eigenvalues of $\mathbf{H}$ in decreasing order. The distortion function is again expressed as a decreasing function of the distance between two precoder column spaces, known as the Fubini-Study distance. Its minimum value over any pair of precoder column spaces is given by

$$\delta_{FS}(\mathcal{W}) = \min_{1 \leq k < l \leq n_p} \arccos \left| \det \left(\mathbf{W}_k^H \mathbf{W}_l \right) \right|. \tag{8.113}$$

Hence, the design criterion reads as

Design criterion 8.4 *Choose a codebook $\mathcal{W}$ made of n_p tall unitary matrices $\mathbf{W}_i$ ($i = 1, \ldots, n_p$) of size $n_t \times n_e$ such that the minimum Fubini-Study distance $\delta_{FS}(\mathcal{W})$ is maximized.*

Those codebooks can be designed using the non-coherent constellations in [HMR$^+$00]. Note that for $n_e = 1$, the Fubini-Study distance reduces to (8.80), because the error probability and capacity based criteria are exclusively related to the received SNR when only one stream is transmitted.

As previously, we may allow the number of streams to vary as a function of the channel state. Naturally, this improves the performance as we authorize switching on/off the power on a variable number of streams depending on the propagation condition. Recall that if full CSIT is provided at the transmitter, the number of effective streams does also vary with the channel conditions as the power is adaptively allocated on each stream based on the water-filling algorithm. Yet, in the quantized version, the power on each stream only takes two values (on or off). More details can be found in [LH05c]. The design is very similar to multi-mode precoding. Instead of a SNR maximization, the selection and codebook design criteria are based on the capacity maximization criteria analogous to the single-mode precoding of (8.111) and Criterion 8.4.

9

Space–time coding for frequency selective channels

Chapters 4 to 8 have dealt extensively with frequency flat fading channels. This last chapter is dedicated to signaling techniques in MIMO frequency selective channels, covering both single-carrier and multi-carrier transmissions. The same aspects as in the previous chapters will be addressed: the mutual information, the error probability and the code design.

In Chapters 2 and 3, we have introduced the frequency selective MIMO channel matrix in the continuous time-delay domain as

$$\mathbf{H}_t[\tau] \triangleq \mathbf{H}(t, \tau) = \sum_{k=0}^{n_s-1} \mathbf{H}(t, \tau_k) \, \delta(\tau - \tau_k). \tag{9.1}$$

In what follows, we consider slow fading channels (the channel is constant over each frame) and the delay-sampled frequency selective representation, which is the combination of L flat fading taps (as already mentioned, L is generally different from n_s)

$$\mathbf{H}_k = \sum_{l=0}^{L-1} \mathbf{H}[l] \, \delta(k - l). \tag{9.2}$$

The lth tap $\mathbf{H}[l]$ is characterized on its own by a statistical (flat fading) model

$$\mathbf{H}[l] = \sqrt{\frac{K_l}{1 + K_l}} \, \bar{\mathbf{H}}[l] + \sqrt{\frac{1}{1 + K_l}} \, \tilde{\mathbf{H}}[l] \tag{9.3}$$

where K_l denotes the Ricean K-factor, $\sqrt{K_l/(1 + K_l)} \, \bar{\mathbf{H}}[l]$ is the coherent component (constant from frame to frame) and $\sqrt{1/(1 + K_l)}\tilde{\mathbf{H}}[l]$ is the variable (Rayleigh) component whose entries are spatially correlated circularly symmetric complex Gaussian random variables. It is also generally assumed that taps fade independently (see Chapter 2).

9.1   Single-carrier vs. multi-carrier transmissions

There are basically two different approaches to transmit information over frequency selective MIMO channels. The first one consists in modulating a single carrier over the full bandwidth B (MIMO-SC), whereas the second approach converts the frequency selective channel into a set of multiple parallel flat fading channels in the frequency domain. This is realized by means of orthogonal frequency division multiplexing (OFDM) modulation and is then referred to as MIMO-OFDM.

9.1.1   Single-carrier transmissions

Analogous to flat fading channels, single-carrier (SC) modulation considers the direct transmission of a codeword $\mathbf{C} = [\mathbf{c}_0 \ldots \mathbf{c}_{T-1}]$ (of size $n_t \times T$) over the channel. The difference with the flat fading case lies in the presence of L resolvable taps, which are responsible for inter-symbol interference (ISI). Yet, under the assumption that the taps fade independently, they provide the receiver with L replicas of the same codeword, i.e. they offer a Lth order diversity. The main challenge is then to design codewords such that a ML decoder is able to efficiently exploit the frequency and spatial diversity without suffering from the ISI. As a result, the frequency selectivity generally considered as a curse may thus become a blessing in fading channels if codewords are adequately designed.

At the kth time instant, the transmitted and received signals are linked by the following relationship

$$\mathbf{y}_k = \sqrt{E_s} \sum_{l=0}^{L-1} \mathbf{H}[l] \mathbf{c}_{k-l} + \mathbf{n}_k \tag{9.4}$$

where $\mathbf{y}_k$, $\mathbf{n}_k$, E_s are defined analogous to frequency flat fading channels. Interestingly, the L taps may be thought of as virtual transmit antennas. Re-expressing (9.4) as

$$\mathbf{y}_k = \sqrt{E_s} \big[\mathbf{H}[0] \ldots \mathbf{H}[L-1] \big] \big[\mathbf{c}_k^T \ldots \mathbf{c}_{k-L+1}^T \big]^T + \mathbf{n}_k \tag{9.5}$$

we obtain a virtual transmit array of $n_t L$ antennas. The $n_r \times n_t L$ virtual channel matrix reads as

$$\underline{\mathbf{H}} \triangleq \big[\mathbf{H}[0] \ldots \mathbf{H}[L-1] \big]. \tag{9.6}$$

In this virtual array formalism, an equivalent codeword may be expressed as

$$\underline{\mathbf{C}} = \begin{bmatrix} \mathbf{c}_0 & \mathbf{c}_1 & \ldots & \mathbf{c}_{T-1} \\ \mathbf{c}_{-1} & \mathbf{c}_0 & \ldots & \mathbf{c}_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_{1-L} & \mathbf{c}_{2-L} & \ldots & \mathbf{c}_{T-L} \end{bmatrix}. \tag{9.7}$$

This actually corresponds to the transmission over $T + L - 1$ symbol durations of codeword vectors $\mathbf{c}_k$, $k = 1 - L, \ldots, T - 1$, or equivalently to the transmission of a codeword $\mathbf{C}' = [\mathbf{c}_{1-L} \ldots \mathbf{c}_0 \ldots \mathbf{c}_{T-1}]$ of size $n_t \times (T + L - 1)$. As a result, a maximum diversity gain of $n_r n_t L$ could theoretically be extracted.

### 9.1.2 Multi-carrier transmissions: MIMO-OFDM

Let us first detail the principle of OFDM modulation before describing various MIMO coding schemes based on OFDM.

Orthogonal frequency division multiplexing (OFDM)

Unlike single-carrier transmissions, MIMO-OFDM can be considered as a frequency domain approach to exploit the frequency diversity, as the OFDM front-end converts the frequency selective channel into a number of parallel flat fading channels. The basic idea consists in turning the channel matrix into a circulant matrix via the addition of a cyclic prefix to the transmitted sequence. A circulant matrix presents the interesting property that its left and right singular vector matrices are respectively DFT (discrete Fourier transform) and IDFT (inverse discrete Fourier transform) matrices. Consequently, the multiplication by an IDFT matrix at the transceiver and by a DFT matrix at the receiver transforms the frequency selective channel into a diagonal matrix, whose elements are the singular values of the circulant matrix. That explains how the original frequency selective channel in the time domain becomes a set of parallel flat fading channels in the frequency domain. This construction allows for a considerable reduction of complexity in terms of equalization and demodulation.

Let us now formalize the OFDM concept. The basic operations are illustrated in Figure 9.1. We detail all steps in the sequel. First applying an IDFT to the codeword $\mathbf{C}$, we obtain as output at the nth time interval ($n = 0, \ldots, T - 1$)

$$\mathbf{x}_n = \frac{1}{\sqrt{T}} \sum_{k=0}^{T-1} \mathbf{c}_k e^{j \frac{2\pi}{T} kn}, \tag{9.8}$$
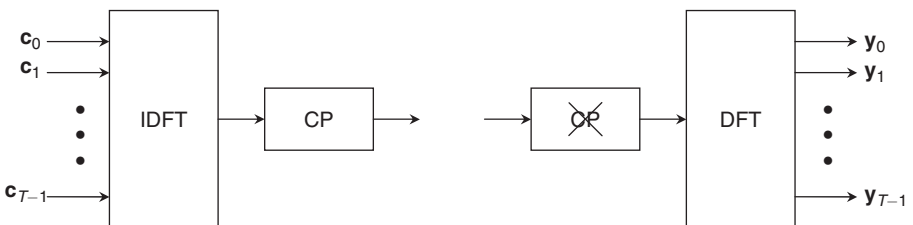
Figure 9.1 *OFDM modulator and demodulator*

or equivalently in a matrix form

$$\begin{bmatrix} \mathbf{x}_0 & \ldots & \mathbf{x}_{T-1} \end{bmatrix}^T = \mathcal{D}^H \begin{bmatrix} \mathbf{c}_0 & \ldots & \mathbf{c}_{T-1} \end{bmatrix}^T. \tag{9.9}$$

The $T \times T$ matrix $\mathcal{D}^H$ realizes the IDFT operation. Hence, $\mathcal{D}$ is a DFT matrix reading as

$$\mathcal{D} = \frac{1}{\sqrt{T}} \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & e^{-j\frac{2\pi}{T}} & e^{-j\frac{2\pi}{T}2} & \ldots & e^{-j\frac{2\pi}{T}(T-1)} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & e^{-j\frac{2\pi}{T}(T-2)} & e^{-j\frac{2\pi}{T}(T-2)2} & \ldots & e^{-j\frac{2\pi}{T}(T-2)(T-1)} \\ 1 & e^{-j\frac{2\pi}{T}(T-1)} & e^{-j\frac{2\pi}{T}(T-1)2} & \ldots & e^{-j\frac{2\pi}{T}(T-1)(T-1)} \end{bmatrix}. \tag{9.10}$$

To avoid intersymbol interference, a guard interval vector of length $L - 1$ denoted as $\mathbf{X}_g = \begin{bmatrix} \mathbf{x}_{-(L-1)} & \ldots & \mathbf{x}_{-1} \end{bmatrix}$ is added in front of the codeword $\mathbf{X} = \begin{bmatrix} \mathbf{x}_0 & \ldots & \mathbf{x}_{T-1} \end{bmatrix}$, so that the OFDM symbol $\mathbf{X}' = \begin{bmatrix} \mathbf{X}_g & \mathbf{X} \end{bmatrix}$ of size $n_t \times (T + L - 1)$ is transmitted. At the receiver, the guard interval is first removed and T output samples are gathered as

$$\begin{bmatrix} \mathbf{r}_0 & \ldots & \mathbf{r}_{T-1} \end{bmatrix}^T = \mathbf{H}_g \begin{bmatrix} \mathbf{x}_{-(L-1)} & \ldots & \mathbf{x}_{T-1} \end{bmatrix}^T + \begin{bmatrix} \mathbf{n}_0 & \ldots & \mathbf{n}_{T-1} \end{bmatrix}^T \tag{9.11}$$

where

$$\mathbf{H}_g = \begin{bmatrix} \mathbf{H}[L-1] & \ldots & \mathbf{H}[1] & \mathbf{H}[0] & \mathbf{0}_{n_r \times n_t} & \ldots & \mathbf{0}_{n_r \times n_t} \\ \mathbf{0}_{n_r \times n_t} & \mathbf{H}[L-1] & \ddots & \mathbf{H}[1] & \mathbf{H}[0] & \ldots & \mathbf{0}_{n_r \times n_t} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{n_r \times n_t} & \ldots & \mathbf{0}_{n_r \times n_t} & \mathbf{H}[L-1] & \mathbf{H}[L-2] & \ldots & \mathbf{H}[0] \end{bmatrix} \tag{9.12}$$

is a $T \times (T + L - 1)$ matrix representing the channel seen by the OFDM symbol.

Usually, the guard interval vector $\mathbf{X}_g$ is chosen in such a way that $\mathbf{x}_{-n} = \mathbf{x}_{T-n}$, for $n = 1, \ldots, L - 1$. Hence, the guard interval vector becomes $\mathbf{X}_g = \begin{bmatrix} \mathbf{x}_{T-(L-1)} \ldots \mathbf{x}_{T-1} \end{bmatrix}$, which is commonly known as the cyclic prefix.

The system model of (9.11) can now be rewritten as

$$\begin{bmatrix} \mathbf{r}_0 & \ldots & \mathbf{r}_{T-1} \end{bmatrix}^T = \mathbf{H}_{cp} \begin{bmatrix} \mathbf{x}_0 & \ldots & \mathbf{x}_{T-1} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_0 & \ldots & \mathbf{n}_{T-1} \end{bmatrix}^T \tag{9.13}$$

with

$$\mathbf{H}_{cp} = \begin{bmatrix} \mathbf{H}[0] & \mathbf{0}_{n_r \times n_t} & \cdots & \mathbf{0}_{n_r \times n_t} & \mathbf{H}[L-1] & \cdots & \mathbf{H}[1] \\ \mathbf{H}[1] & \mathbf{H}[0] & \mathbf{0}_{n_r \times n_t} & \cdots & \mathbf{0}_{n_r \times n_t} & \cdots & \mathbf{H}[2] \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{H}[L-2] & \cdots & \mathbf{H}[0] & \mathbf{0}_{n_r \times n_t} & \cdots & \mathbf{0}_{n_r \times n_t} & \mathbf{H}[L-1] \\ \mathbf{H}[L-1] & \cdots & \mathbf{H}[1] & \mathbf{H}[0] & \mathbf{0}_{n_r \times n_t} & \cdots & \mathbf{0}_{n_r \times n_t} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{n_r \times n_t} & \cdots & \mathbf{H}[L-1] & \mathbf{H}[L-2] & \cdots & \mathbf{H}[0] & \mathbf{0}_{n_r \times n_t} \\ \mathbf{0}_{n_r \times n_t} & \cdots & \mathbf{0}_{n_r \times n_t} & \mathbf{H}[L-1] & \cdots & \mathbf{H}[1] & \mathbf{H}[0] \end{bmatrix}. \tag{9.14}$$

We observe that $\mathbf{H}_{cp}$ is a blockwise circulant matrix of size $T \times T$. As a result, its SVD decomposition $\mathbf{H}_{cp} = \mathcal{D}^H \mathbf{\Lambda}_{cp} \mathcal{D}$ is such that $\mathbf{\Lambda}_{cp}$ is a diagonal matrix whose elements are obtained by a blockwise DFT of $\begin{bmatrix} \mathbf{H}[0] & \mathbf{H}[1] & \ldots & \mathbf{H}[L-1] \end{bmatrix}$, i.e.

$$\mathbf{\Lambda}_{cp}(k,k) = \sum_{l=0}^{L-1} \mathbf{H}[l] \, e^{-j\frac{2\pi}{T}kl}, \tag{9.15}$$

irrespective of the channel matrix. The full information about the channel is accounted for in $\mathbf{\Lambda}_{cp}$, and the eigenvectors of $\mathbf{H}_{cp}$ are independent of the channel matrices $\mathbf{H}[l]$. The use of IDFT matrices at the transmitter is now clear as (9.13) can be rewritten as

$$\begin{bmatrix} \mathbf{r}_0 & \ldots & \mathbf{r}_{T-1} \end{bmatrix}^T = \mathcal{D}^H \mathbf{\Lambda}_{cp} \begin{bmatrix} \mathbf{c}_0 & \ldots & \mathbf{c}_{T-1} \end{bmatrix}^T + \begin{bmatrix} \mathbf{n}_0 & \ldots & \mathbf{n}_{T-1} \end{bmatrix}^T. \tag{9.16}$$

Applying a DFT operation to the received vector, we finally obtain

$$\begin{aligned} \begin{bmatrix} \mathbf{y}_0 & \ldots & \mathbf{y}_{T-1} \end{bmatrix}^T &= \mathcal{D} \begin{bmatrix} \mathbf{r}_0 & \ldots & \mathbf{r}_{T-1} \end{bmatrix}^T \\ &= \mathbf{\Lambda}_{cp} \begin{bmatrix} \mathbf{c}_0 & \ldots & \mathbf{c}_{T-1} \end{bmatrix}^T + \mathcal{D} \begin{bmatrix} \mathbf{n}_0 & \ldots & \mathbf{n}_{T-1} \end{bmatrix}^T. \end{aligned} \tag{9.17}$$

As expected, the original frequency selective channel has been converted into a set of T parallel flat fading channels in the frequency domain, the channel gains being given by the diagonal elements of $\mathbf{\Lambda}_{cp}$. Because the DFT matrix $\mathcal{D}$ does not affect the noise statistics, the input-output relationship on each parallel channel $k = 0, \ldots, T-1$ may be expressed without loss of generality as

$$\mathbf{y}_k = \sqrt{E_s} \mathbf{H}_{(k)} \mathbf{c}_k + \mathbf{n}_k \tag{9.18}$$

with

$$\mathbf{H}_{(k)} = \sum_{l=0}^{L-1} \mathbf{H}[l] \, e^{-j\frac{2\pi}{T}kl}. \tag{9.19}$$

The $n_r \times 1$ vector $\mathbf{y}_k$ is the received signal to be decoded, and $\mathbf{n}_k$ is a $n_r \times 1$ zero mean complex additive white Gaussian noise (AWGN) vector with $\mathcal{E}\{\mathbf{n}_k \mathbf{n}_{k'}^H\} = \sigma_n^2 \mathbf{I}_{n_r} \delta[k-k']$.
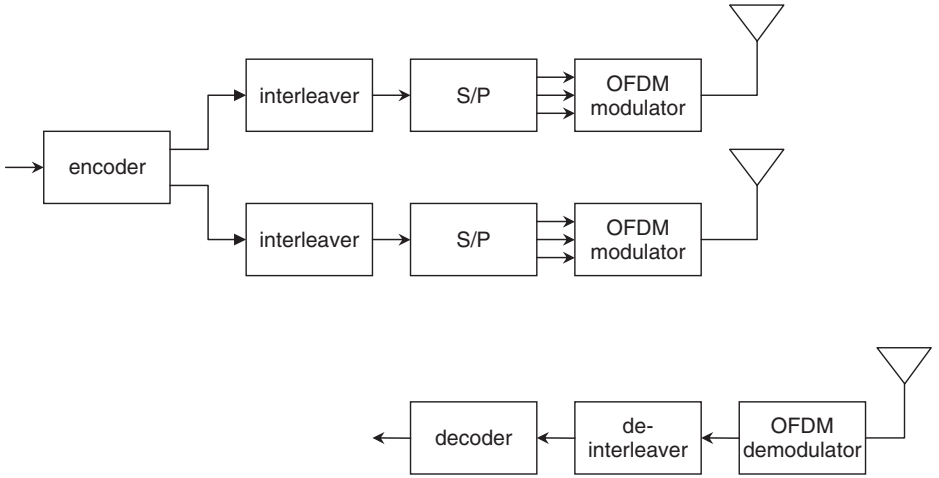
Figure 9.2   *Block diagram of a MIMO-OFDM system*

If ML decoding is applied, the decoder computes an estimate of the transmitted codeword according to

$$\hat{\mathbf{C}} = \arg \min_{\mathbf{C}} \sum_{k=0}^{T-1} \left\| \mathbf{y}_k - \sqrt{E_s} \mathbf{H}_{(k)} \mathbf{c}_k \right\|^2. \tag{9.20}$$

The system block diagram of a MIMO-OFDM system is depicted in Figure 9.2. We will discuss the design of the encoder and the importance of the interleaver in further sections.

As for now, we want to highlight the strong analogy between the input-output relationship over a flat fading MIMO channel in (5.1) and the system model of (9.18). In the latter, the temporal dimension of the former is replaced by the frequency dimension, i.e. the kth time instant in MIMO flat fading corresponds here to the kth frequency subband (or tone) $\mathbf{H}_{(k)}$. Note that we use an index (k) in order to make a distinction with the index k denoting the time instant in the flat fading case.

Space-frequency coded MIMO-OFDM
The most popular coding technique in MIMO-OFDM exploits the independent fading of the parallel channels $\mathbf{H}_{(k)}$ across the frequency tones. By coding across space and frequency, both space and frequency diversity gains can be extracted. Space-frequency coded MIMO-OFDM is the direct implementation of the transmission scheme described above. Intuitively, we understand that if the coherence bandwidth of the channel is small (i.e. if the delay-spread and the number of taps are large), the channel gains $\mathbf{H}_{(k)}$ vary significantly from tone to tone. The channel in the frequency domain can then be considered as fast fading. Swapping frequency with time, one expects that a code designed for flat fast fading channels will perform well if used as a space-frequency code in a highly frequency selective channel. This also suggests that certain codes designed for slow fading channels

might not perform well if they are used as space-frequency codes in highly selective channels. For instance, O-STBCs require the channel to remain static over the codeword duration. As an example, if the Alamouti scheme is used across two tones, one should make sure that the channel coherence bandwidth is larger than the width of these tones. Otherwise, the matched filtering operation will not decouple the streams at the receiver, causing a performance degradation. By contrast, the space-frequency version of a trellis code optimally designed for fast fading channels may perform equally well in frequency selective channels.

We have shown that a single-carrier transmission over a frequency selective MIMO channel could be seen as a transmission over an equivalent $n_r \times L n_t$ MIMO channel. This is also the case here. For space-frequency coded MIMO-OFDM, let us express the virtual array model by rewriting (9.18) as follows

$$\mathbf{y}_k = \sqrt{E_s}\underline{\mathbf{H}}\left[\mathbf{c}_k^T \ \cdots \ e^{-j\frac{2\pi}{T}kl}\mathbf{c}_k^T \ \cdots \ e^{-j\frac{2\pi}{T}k(L-1)}\mathbf{c}_k^T \right]^T + \mathbf{n}_k, \qquad (9.21)$$

where $\underline{\mathbf{H}}$ is defined analogous to single-carrier transmissions. Nevertheless, the equivalent codeword defined as

$$\underline{\mathbf{C}} = \begin{bmatrix} \mathbf{c}_0 \cdots & & \mathbf{c}_k & \cdots & & \mathbf{c}_{T-1} \\ \vdots & \ddots & \vdots & & \vdots & \vdots \\ & & e^{-j\frac{2\pi}{T}kl}\mathbf{c}_k & & & \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \mathbf{c}_0 & \cdots & e^{-j\frac{2\pi}{T}k(L-1)}\mathbf{c}_k & \cdots & e^{-j\frac{2\pi}{T}(T-1)(L-1)}\mathbf{c}_{T-1} \end{bmatrix} \qquad (9.22)$$

differs from the SC case, as readily observed by comparing (9.5) and (9.21).

Analogous to SC transmissions, a maximum diversity gain of $n_r n_t L$ could theoretically be extracted. Indeed, MIMO-OFDM does not reduce the number of degrees of freedom. Rigorously, a slight loss in spectral efficiency is caused by the cyclic prefix, but for long frame lengths T this loss is negligible.

By contrast to SC transmissions, OFDM modulation generally requires a better synchronization and is more susceptible to phase noise. Furthermore, it is subject to large peak-to-average power ratios. Despite those drawbacks, SF-MIMO OFDM is often preferred to MIMO-SC and is now part of an increasing number of standards.

Space–time and space–time-frequency coded MIMO-OFDM
Space–time coded MIMO-OFDM spreads information symbols in space and time analogous to space–time coding for flat fading channels. Although this scheme can also be represented along the formalism of (9.18), the codewords are defined in a different way,

as they are sent on different OFDM symbols on a per tone basis. Each tone simply acts as one parallel channel. Therefore, space–time codes designed for flat fading channels can be directly used for space–time coded MIMO-OFDM.

Example 9.1   *If the channel remains static over two OFDM symbols, space–time coded MIMO-OFDM may use the Alamouti scheme by transmitting $\begin{bmatrix} c_1 & c_2 \end{bmatrix}^T$ over a given tone of the first OFDM symbol and $\begin{bmatrix} -c_2^* & c_1^* \end{bmatrix}^T$ over the same tone of the second OFDM symbol. All other tones are similarly occupied by Alamouti codewords matrices.*

However, this scheme is not very practical for large codeword lengths T, as it requires the channel to remain constant over T OFDM symbols, i.e. over a usually large period. Furthermore, no coding is performed across the frequency tones, implying that no frequency diversity is extracted.

Space–time-frequency coded MIMO-OFDM is a combination of the above schemes, where frequency diversity is exploited by coding across frequency tones of one OFDM symbol, time diversity is extracted by coding across multiple OFDM symbols and space diversity is provided thanks to coding across antennas. Space–time and space–time-frequency coded MIMO-OFDM schemes are not covered in this chapter.

9.1.3   *A unified representation for single and multi-carrier transmissions*

As already outlined, both SC and MIMO-OFDM transmission models can be represented by means of a unique virtual $n_r \times Ln_t$ MIMO channel

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_0 & \cdots & \mathbf{y}_{T-1} \end{bmatrix} = \sqrt{E_s}\underline{\mathbf{H}}\,\mathbf{C} + \begin{bmatrix} \mathbf{n}_0 & \cdots & \mathbf{n}_{T-1} \end{bmatrix} \tag{9.23}$$

where the virtual channel matrix $\underline{\mathbf{H}}$ of dimension $n_r \times Ln_t$ is given by (9.6). Analogous to (9.3), it can be decomposed into two components

$$\underline{\mathbf{H}} = \bar{\underline{\mathbf{H}}} + \tilde{\underline{\mathbf{H}}} \tag{9.24}$$

where $\bar{\underline{\mathbf{H}}}$ and $\tilde{\underline{\mathbf{H}}}$ are naturally defined by

$$\bar{\underline{\mathbf{H}}} = \begin{bmatrix} \sqrt{\frac{K_0}{1+K_0}}\,\bar{\mathbf{H}}[0] & \sqrt{\frac{K_1}{1+K_1}}\,\bar{\mathbf{H}}[1] & \cdots & \sqrt{\frac{K_{L-1}}{1+K_{L-1}}}\,\bar{\mathbf{H}}[L-1] \end{bmatrix}$$
$$\tilde{\underline{\mathbf{H}}} = \begin{bmatrix} \sqrt{\frac{1}{1+K_0}}\,\tilde{\mathbf{H}}[0] & \sqrt{\frac{1}{1+K_1}}\,\tilde{\mathbf{H}}[1] & \cdots & \sqrt{\frac{1}{1+K_{L-1}}}\,\tilde{\mathbf{H}}[L-1] \end{bmatrix}. \tag{9.25}$$

The equivalent transmitted codewords in the virtual $n_r \times Ln_t$ MIMO representation are outlined by

$$\underline{\mathbf{C}} = \begin{bmatrix} \mathbf{C}_{(0)} \\ \vdots \\ \mathbf{C}_{(L-1)} \end{bmatrix}. \tag{9.26}$$

For space-frequency coding, rewriting (9.22) further yields

$$\begin{bmatrix} \mathbf{C}_{(0)} \\ \vdots \\ \mathbf{C}_{(L-1)} \end{bmatrix} = \begin{bmatrix} \mathbf{CD}_{(0)} \\ \vdots \\ \mathbf{CD}_{(L-1)} \end{bmatrix} = [\mathbf{I}_L \otimes \mathbf{C}] \underbrace{\begin{bmatrix} \mathbf{D}_{(0)} \\ \vdots \\ \mathbf{D}_{(L-1)} \end{bmatrix}}_{\mathbf{D}} \tag{9.27}$$

with $\mathbf{D}_{(l)} = \operatorname{diag}\left\{1, \ldots, e^{-j\frac{2\pi}{T}kl}, \ldots, e^{-j\frac{2\pi}{T}(T-1)l}\right\}$. For single-carrier transmissions

$$\mathbf{C}_{(l)}(m,k) = \mathbf{c}_{k-l}(m,1), \quad k = 0, \ldots, T-1, \quad m = 1, \ldots, n_t. \tag{9.28}$$

In what follows, we assume that the receiver knows perfectly the instantaneous channel realizations while the transmitter does not possess any information about the channel. Maximum-likelihood (ML) decoding is performed, so that the estimate of the transmitted codeword is given by

$$\hat{\mathbf{C}} = \arg\min_{\mathbf{C}} \left\| \mathbf{Y} - \sqrt{E_s}\underline{\mathbf{H}}\,\underline{\mathbf{C}} \right\|^2 \tag{9.29}$$

where the minimization is performed over all possible codeword vectors $\mathbf{C}$ (recall that there is a one-to-one relationship between $\underline{\mathbf{C}}$ and $\mathbf{C}$, for both SC and MIMO-OFDM).

Linear block codes

The virtual array model can be further particularized to the case of linear block codes. In Chapter 5, linear block codewords $\mathbf{C}$ read as

$$\mathbf{C} = \sum_{q=1}^{Q} \mathbf{\Phi}_q \Re[c_q] + \mathbf{\Phi}_{q+Q}\Im[c_q] \tag{9.30}$$

where $\mathbf{\Phi}_q$ are complex basis matrices of size $(n_t \times T)$, c_q stands for the complex information symbol taken from PSK or QAM constellations and Q is the number of complex symbols c_k transmitted over a codeword. In frequency selective channels, the transmitted codewords in the virtual array formalism are given by

$$\underline{\mathbf{C}} = \sum_{q=1}^{Q} \underline{\mathbf{\Phi}}_q \Re[c_q] + \underline{\mathbf{\Phi}}_{q+Q}\Im[c_q] \tag{9.31}$$

with

$$
\underline{\mathbf{\Phi}}_q = \begin{bmatrix} \mathbf{\Phi}_q^{(0)} \\ \vdots \\ \mathbf{\Phi}_q^{(L-1)} \end{bmatrix}
\tag{9.32}
$$

denoting the equivalent basis matrices in the virtual array representation.

For space-frequency coding, those equivalent basis matrices are further rewritten as

$$
\begin{bmatrix} \mathbf{\Phi}_q^{(0)} \\ \vdots \\ \mathbf{\Phi}_q^{(L-1)} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_q \mathbf{D}_{(0)} \\ \vdots \\ \mathbf{\Phi}_q \mathbf{D}_{(L-1)} \end{bmatrix} = [\mathbf{I}_L \otimes \mathbf{\Phi}_q]\underline{\mathbf{D}}
\tag{9.33}
$$

whereas in single-carrier transmissions, they are such that

$$
\mathbf{\Phi}_q^{(l)}(m,k) = \mathbf{\Phi}_q(m,k-l), \;\; k = 0,\ldots,T-1, \; m = 1,\ldots,n_t.
\tag{9.34}
$$

Finally, we define the notion of wide unitary basis matrices.

Definition 9.1 *Wide $(T \geq Ln_t)$ unitary basis matrices are such that $\underline{\mathbf{\Phi}}_q \underline{\mathbf{\Phi}}_q^{H} = \frac{T}{LQn_t}\mathbf{I}_{Ln_t} \; \forall q = 1,\ldots,2Q$.*

9.2   Information theoretic aspects for frequency selective MIMO channels

9.2.1   *Capacity considerations*

For a given channel realization of the frequency selective channel, the mutual information is directly derived from the flat fading case by an integration over the frequency band of interest B

$$
\mathcal{I}_{FS}\left(\{\mathbf{H}(f)\}_f, \{\mathbf{Q}(f)\}_f\right) = \frac{1}{B}\int_B \log_2 \det\left[\mathbf{I}_{n_r} + \rho(f)\mathbf{H}(f)\mathbf{Q}(f)\mathbf{H}(f)^H\right] df
\tag{9.35}
$$

where $\mathrm{Tr}\{\mathbf{Q}(f)\} = P_B$ is the total average transmit power over the frequency band B. The quantities $\mathbf{H}(f)$, $\mathbf{Q}(f)$ and $\rho(f)$ are the channel transfer function, entry covariance matrix and SNR at frequency f, respectively.

The capacity of the frequency selective MIMO channel with perfect transmit channel knowledge (CSIT) is then defined as

$$
C_{CSIT,FS} = \max_{\int_B \mathrm{Tr}\{\mathbf{Q}(f)\}=P_B} \mathcal{I}_{FS}\left(\{\mathbf{H}(f)\}_f, \{\mathbf{Q}(f)\}_f\right)
\tag{9.36}
$$

By analogy with the MIMO-OFDM transmission, let us assume that the frequency bandwidth B is divided into T frequency flat subbands with identical spectral noise density.

The input-output relationship in each of these subbands is then given by (9.18). Hence, neglecting the loss in spectral efficiency due to the cyclic prefix, the mutual information for a given frequency selective MIMO channel realization is the sum of the mutual information $\mathcal{I}_{(k)}$ achieved over each subband, divided by the number of frequency subbands [BGP02] . Hence, we have

$$\mathcal{I}_{FS}\left(\{\mathbf{H}_{(k)}\}_k, \{\mathbf{Q}_{(k)}\}_k\right) = \frac{1}{T}\sum_{k=0}^{T-1}\mathcal{I}_{(k)}$$

$$= \frac{1}{T}\sum_{k=1}^{T}\log_2\det\left[\mathbf{I}_{n_r} + \rho\mathbf{H}_{(k)}\mathbf{Q}_{(k)}\mathbf{H}_{(k)}^H\right] \qquad (9.37)$$

where $\mathbf{Q}_{(k)}$ is the covariance matrix of the Gaussian vector $\mathbf{c}_k$. Normalizing the total transmit power analogous to the flat fading case, the transmit power constraint can be rewritten as $\sum_{k=0}^{T-1}\text{Tr}\{\mathbf{Q}_{(k)}\} = T$. The capacity is then outlined by

$$C_{CSIT,FS} = \max_{\sum_{k=0}^{T-1}\text{Tr}\{\mathbf{Q}_{(k)}\}=T}\mathcal{I}_{FS}\left(\{\mathbf{H}_{(k)}\}_k, \{\mathbf{Q}_{(k)}\}_k\right) \qquad (9.38)$$

$$= \frac{1}{T}\max_{\sum_{k=0}^{T}\text{Tr}\{\mathbf{Q}_{(k)}\}=T}\sum_{k=1}^{T}\log_2\det\left[\mathbf{I}_{n_r} + \rho\mathbf{H}_{(k)}\mathbf{Q}_{(k)}\mathbf{H}_{(k)}^H\right]. \qquad (9.39)$$

Remember that in flat fading channels, when the channel realizations are known at the transmitter, the capacity is obtained by performing spatial water-filling. In frequency selective channels, the power allocation is now performed not only in the spatial domain but also in the frequency domain, and is known as space-frequency water-filling [RC98].

We may also introduce the notions of ergodic capacity $\bar{C}_{CSIT,FS} = \mathcal{E}\{C_{CSIT,FS}\}$ and outage capacity. As before, the ergodic capacity is relevant when coding is performed across frequency tones of a large number of OFDM symbols experiencing many different channel realizations.

9.2.2 Mutual information with equal power allocation

When the power is uniformly allocated over the frequency subbands (or tones, using an OFDM formalism), the entry covariance matrices $\mathbf{Q}_{(k)}$ are identity matrices and the mutual information simply reads as

$$\mathcal{I}_{e,FS} = \frac{1}{T}\sum_{k=1}^{T}\log_2\det\left[\mathbf{I}_{n_r} + \rho\mathbf{H}_{(k)}\mathbf{H}_{(k)}^H\right]. \qquad (9.40)$$

If the channel is flat fading over the full bandwidth B, $\mathbf{H}_{(k)} = \mathbf{H}$ and (9.40) reduces to the mutual information with equal power allocation of the flat fading channel. If the channel is frequency-selective, $\mathbf{H}_{(k)}$ varies from tone to tone and the mutual information $\mathcal{I}_{e,FS}$ is

simply an average of $\mathcal{I}_{(k)}$ over the tones. This suggests that if the channel is frequency selective enough to decorrelate the identically distributed tones, the mutual information converges to a fixed quantity $\bar{\mathcal{I}}_{e,FS} = \mathcal{E}\{\mathcal{I}_{e,FS}\}$ as the number of tones approaches infinity. This shows that as the frequency selectivity increases, the distribution of the mutual information becomes narrower and the outage performance is improved.

9.2.3 Diversity-multiplexing trade-off

We now discuss the diversity-multiplexing trade-off in frequency selective channels. In this context, [GT04, Gro05] reveal that the trade-off is actually the same as if all paths were individually resolved, thereby achieving the same performance as the matched filter bound (MFB). In other words, this means that the interference created by the different echoes does not degrade the achievable performance.

Proposition 9.1 *The asymptotic diversity-multiplexing trade-off $g_d^*(g_s, \infty)$ of the frequency selective SISO fading channel consisting of L i.i.d. Rayleigh distributed taps is given by*

$$g_d^*(g_s, \infty) = L(1 - g_s), \ g_s \in [0, 1]. \tag{9.41}$$

PROOF: The proof relies on the assumption that each path can be resolved individually so that matched filtering is performed on each tap. Let us denote the impulse response of the channel as the vector $\underline{\mathbf{h}} = \left[h[0], \ldots, h[L-1] \right]$. For a total transmission rate that scales with the SNR as $R = g_s \log_2(\rho)$, the outage probability simply reads as

$$P_{out}(R) = P\left(\log_2\left(1 + \rho\|\underline{\mathbf{h}}\|^2\right) \leq R \right) \tag{9.42}$$

$$\dot{=} P\left(\rho\|\underline{\mathbf{h}}\|^2 \leq \rho^{g_s}\right) \tag{9.43}$$

$$\dot{=} \rho^{-L(1-g_s)} \tag{9.44}$$

for $0 \leq g_s \leq 1$ (the operator $\dot{=}$ has been introduced in Chapter 4). ∎

As expected, a frequency selective channel provides an Lth order diversity for a fixed rate code. It is interesting to point out the similarity with the trade-off obtained with SIMO, MISO or O-STBC-based MIMO channels (see Example 4.2). The L taps act as L virtual antennas providing the same trade-off as a SIMO or MISO system with L physical antennas.

Ignoring the loss in efficiency related to the cyclic prefix, the asymptotic optimal diversity-multiplexing trade-off of an OFDM-based transmission in frequency selective channels is also given by (9.41). Theoretically, it should be possible to derive the trade-off by considering the OFDM transmission as a set of parallel channels and by applying the results of Chapter 4. However, the OFDM parallel channels are not independent owing to the limited number of taps in the channel, making the derivation much more complicated.

9.3 Average pairwise error probability

For a given channel realization $\underline{\mathbf{H}}$, the conditional pairwise error probability (PEP), i.e. the probability of decoding a codeword $\underline{\mathbf{E}}$ when $\underline{\mathbf{C}}$ is transmitted, is generalized to the frequency selective case as

$$P(\underline{\mathbf{C}} \rightarrow \underline{\mathbf{E}}|\underline{\mathbf{H}}) = \mathcal{Q}\left(\sqrt{\frac{E_s}{2\sigma_n^2}\|\underline{\mathbf{H}}(\underline{\mathbf{C}} - \underline{\mathbf{E}})\|_F^2}\right). \tag{9.45}$$

Defining $\underline{\boldsymbol{\Delta}} \triangleq \mathbf{I}_{n_r} \otimes \tilde{\underline{\mathbf{E}}} \triangleq \mathbf{I}_{n_r} \otimes (\underline{\mathbf{C}} - \underline{\mathbf{E}})(\underline{\mathbf{C}} - \underline{\mathbf{E}})^H$, we are able to write

$$\|\underline{\mathbf{H}}(\underline{\mathbf{C}} - \underline{\mathbf{E}})\|_F^2 = \text{vec}(\underline{\mathbf{H}}^H)^H \underline{\boldsymbol{\Delta}}\text{vec}(\underline{\mathbf{H}}^H). \tag{9.46}$$

Since this last equation turns out to be an Hermitian quadratic form of complex Gaussian random variables, the rationale used in Chapter 6 is also valid. Consequently, the PEP averaged over the spatially correlated Ricean fading channel with correlated taps is given by

$$P(\underline{\mathbf{C}} \rightarrow \underline{\mathbf{E}}) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(-\eta\text{vec}(\bar{\underline{\mathbf{H}}}^H)^H \underline{\boldsymbol{\Delta}}(\mathbf{I}_{Ln_rn_t} + \eta\underline{\mathbf{R}}\underline{\boldsymbol{\Delta}})^{-1}\text{vec}(\bar{\underline{\mathbf{H}}}^H)\right)$$
$$(\det(\mathbf{I}_{Ln_rn_t} + \eta\underline{\mathbf{R}}\underline{\boldsymbol{\Delta}}))^{-1}d\beta \tag{9.47}$$

where $\eta = \rho/(4\sin^2\beta)$ and $\underline{\mathbf{R}}$ denotes the so-called space-tap correlation matrix defined as

$$\underline{\mathbf{R}} = \mathcal{E}\left\{\text{vec}(\tilde{\underline{\mathbf{H}}}^H)\text{vec}(\tilde{\underline{\mathbf{H}}}^H)^H\right\}. \tag{9.48}$$

The space-tap correlation matrix accounts for spatial correlations within each narrowband MIMO tap as well as for the correlations between taps.

In Rayleigh slow fading channels, $\{K_l = 0\}_{l=0}^{L-1}$ for all taps and the average PEP (9.47) reduces to

$$P(\underline{\mathbf{C}} \rightarrow \underline{\mathbf{E}}) = \frac{1}{\pi}\int_0^{\pi/2}\left(\det\left(\mathbf{I}_{Tn_r} + \eta\underline{\mathbf{C}}_{\underline{\mathbf{R}}}\right)\right)^{-1}d\beta \tag{9.49}$$

where we define $\underline{\mathbf{C}}_{\underline{\mathbf{R}}} = \left(\mathbf{I}_{n_r} \otimes (\underline{\mathbf{C}} - \underline{\mathbf{E}})^H\right)\mathbf{R}(\mathbf{I}_{n_r} \otimes (\underline{\mathbf{C}} - \underline{\mathbf{E}}))$.

To simplify the notations, we decompose the space-tap correlation matrix $\underline{\mathbf{R}}$ into n_r^2 blocks $\mathbf{R}^{(pq)}$ of size $Ln_t \times Ln_t$

$$\underline{\mathbf{R}} = \begin{bmatrix} \mathbf{R}^{(11)} & \cdots & \mathbf{R}^{(1n_r)} \\ \vdots & \ddots & \vdots \\ \mathbf{R}^{(n_r1)} & \cdots & \mathbf{R}^{(n_rn_r)} \end{bmatrix} \tag{9.50}$$

where each block $\mathbf{R}_{ij}^{(pq)}$ is further decomposed into L^2 $n_t \times n_t$ sub-blocks $\mathbf{R}_{lj}^{(pq)}$ as

$$\mathbf{R}^{(pq)} = \begin{bmatrix} \mathbf{R}_{00}^{(pq)} & \cdots & \mathbf{R}_{0(L-1)}^{(pq)} \\ \vdots & \ddots & \vdots \\ \mathbf{R}_{(L-1)0}^{(pq)} & \cdots & \mathbf{R}_{(L-1)(L-1)}^{(pq)} \end{bmatrix}. \tag{9.51}$$

Hence, $\underline{\mathbf{C}_{\mathbf{R}}}$ can be expressed in terms of its internal blocks

$$\underline{\mathbf{C}_{\mathbf{R}}}^{(pq)} = (\underline{\mathbf{C}} - \underline{\mathbf{E}})^H \mathbf{R}^{(pq)} (\underline{\mathbf{C}} - \underline{\mathbf{E}}) \tag{9.52}$$

for $p, q = 1, \ldots, n_r$.

If each tap is spatially i.i.d. Rayleigh distributed with an average power β_l and if there is no correlation between taps, the space-tap correlation matrix reduces to

$$\underline{\mathbf{R}} = \mathbf{I}_{n_r} \otimes \mathrm{diag}\{\beta_0, \ldots, \beta_{L-1}\} \otimes \mathbf{I}_{n_t} \tag{9.53}$$

and the average PEP simply becomes

$$P(\underline{\mathbf{C}} \to \underline{\mathbf{E}}) = \frac{1}{\pi} \int_0^{\pi/2} \left[\det \left(\mathbf{I}_{Ln_t} + \eta \, \mathrm{diag}\{\beta_0, \ldots, \beta_{L-1}\} \otimes \mathbf{I}_{n_t} \tilde{\underline{\mathbf{E}}} \right) \right]^{-n_r} d\beta. \tag{9.54}$$

9.4   Code design criteria for single-carrier transmissions in Rayleigh fading channels

We assume that $\mathbf{R}$ is full rank. To achieve the full diversity at high SNR, $r(\underline{\mathbf{C}_{\mathbf{R}}})$ should equal $n_r n_t L$. This means that the equivalent (or virtual) codewords must be designed in such a way that $r(\tilde{\underline{\mathbf{E}}}) = n_t L$, with $\tilde{\underline{\mathbf{E}}} \triangleq (\underline{\mathbf{C}} - \underline{\mathbf{E}})(\underline{\mathbf{C}} - \underline{\mathbf{E}})^H$.

Space–time code design for single-carrier transmissions has been the subject of much research [ZG01, LFT01, LP00, GHL+03, GSP02]. Yet, owing to the large receiver complexity, this type of transmission has often not been considered in practical applications. Nevertheless, we will review below a few well-known schemes that exploit the full diversity in SC-MIMO transmissions.

9.4.1   Generalized delay-diversity

In the classical delay-diversity scheme over a flat fading channel (see Chapter 5), the stream transmitted on the mth antenna is the stream transmitted on the first antenna delayed by $m - 1$ symbol durations. We have explained that this delay makes the channel appear as frequency selective at the receiver. When the channel is naturally frequency selective, the generalized delay-diversity (GDD) scheme [GSP02] further increases the channel selectivity. Practically, on the mth antenna, the transmitter delays the stream transmitted on the first

antenna by $(m-1)L$ symbol durations. The equivalent codeword $\underline{\mathbf{C}}$ simply corresponds to the codeword $\mathbf{C}$ obtained with a classical delay-diversity scheme over $n_t L$ transmit antennas. Note that if the delay is smaller than $(m-1)L$, the equivalent error matrix is not full rank and the full diversity cannot be achieved. This is illustrated in the following example.

Example 9.2   *We want to send the stream of information symbols $c_1, \ldots, c_T$ over two antennas in a 2-tap Rayleigh fading channel $(L=2)$. With the classical delay-diversity scheme, the transmitted codeword $\mathbf{C}$ is given by*

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & c_2 & \ldots & c_T & 0 & 0 \\ 0 & c_1 & c_2 & \ldots & c_T & 0 \end{bmatrix}, \tag{9.55}$$

where the zeros at the end of the frame avoid the ISI with the next frame. In a 2-tap channel, the equivalent codeword $\underline{\mathbf{C}}$ of the classical scheme then becomes

$$\underline{\mathbf{C}} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & c_2 & \ldots & c_T & 0 & 0 \\ 0 & c_1 & c_2 & \ldots & c_T & 0 \\ 0 & c_1 & c_2 & \ldots & c_T & 0 \\ 0 & 0 & c_1 & c_2 & \ldots & c_T \end{bmatrix}. \tag{9.56}$$

Clearly, this code cannot achieve the full diversity but only provides a total diversity gain of $3n_r$ since the second and third rows of $\underline{\mathbf{C}}$ are identical.

With the generalized delay-diversity scheme, the delay on the second antenna is equal to two, hence the transmitted codeword is

$$\mathbf{C} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & c_2 & \ldots & c_T & 0 & 0 & 0 \\ 0 & 0 & c_1 & c_2 & \ldots & c_T & 0 \end{bmatrix}, \tag{9.57}$$

so that the equivalent codeword $\underline{\mathbf{C}}$ reads as

$$\underline{\mathbf{C}} = \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 & c_2 & \ldots & c_T & 0 & 0 & 0 \\ 0 & 0 & c_1 & c_2 & \ldots & c_T & 0 \\ 0 & c_1 & c_2 & \ldots & c_T & 0 & 0 \\ 0 & 0 & 0 & c_1 & c_2 & \ldots & c_T \end{bmatrix}. \tag{9.58}$$

This scheme achieves the full transmit diversity of 4, hence a total diversity gain of $4n_r$. Naturally, the trellis representation of this scheme can also be constructed similarly to the flat fading case.

However, note that by increasing the length of the channel (i.e. the number of taps L) and the number of antennas, the efficiency of the transmission decreases as more and more zeros have to be added in order to avoid inter symbol interferences. For the same reason,

we have seen in Chapter 5 that standard delay-diversity achieves the optimal diversity-multiplexing trade-off for $n_r = 1$ only when T is infinite. For any finite T, the trade-off is affected by this efficiency loss. This is a small drawback of delay-diversity schemes when performed in the temporal domain. Later we will find that this can be improved when a similar scheme is used in the frequency domain.

9.4.2 Lindskog-Paulraj scheme

The Lindskog-Paulraj (LP) scheme [LP00] is a direct extension of the Alamouti scheme to single-carrier transmissions in frequency selective channels. This scheme is limited to two transmit antennas. For simplicity, we also assume here that $n_r = 1$. Extensions to a larger number of antennas and additional details may be found in [SL01, LSLL02].

Instead of transmitting two complex symbols c_1, c_2 like the Alamouti code, we transmit two streams of symbols of length T, $c_1[k]$ and $c_2[k]$. Instead of employing usual matrix multiplications, we use a FIR filter representation of the frequency selective channel to simplify the notations. Denoting by q^{-1} the unit delay operator, the impulse response of the channel from the mth transmit antenna to the single receive antenna can be written as[1]

$$h_m(q^{-1}) = \sum_{l=0}^{L-1} \mathbf{H}[l](1,m)q^{-l} \tag{9.59}$$

whose conjugate is defined as

$$\left(h_m(q^{-1})\right)^* = h_m^*(q) = \sum_{l=0}^{L-1} \mathbf{H}[l]^*(1,m)q^{l}. \tag{9.60}$$

The operation $h_m(q^{-1}) c_m[k]$ is simply defined as

$$h_m(q^{-1}) c_m[k] = \sum_{l=0}^{L-1} \mathbf{H}[l](1,m)c_m[k-l]. \tag{9.61}$$

The LP scheme works as follows. During the first burst of T symbol durations, the streams $c_1[k]$ and $c_2[k]$ are transmitted respectively from the first and second antennas. The received stream $y_1[k]$ during this first burst is then expressed as

$$y_1[k] = \sqrt{\frac{E_s}{2}} \left[h_1(q^{-1}) \ h_2(q^{-1}) \right] \begin{bmatrix} c_1[k] \\ c_2[k] \end{bmatrix} + n_1[k]. \tag{9.62}$$

This is then followed by a guard interval of L symbol durations to avoid ISI with the second burst. During the second burst, the time reversed and complex conjugated versions of $c_1[k]$

---

[1] $\mathbf{H}[l](m,n)$ denotes the $(m,n)^{\text{th}}$ element of $\mathbf{H}[l]$ and $\mathbf{H}[l]^*(m,n)$ denotes the $(m,n)^{\text{th}}$ element of $(\mathbf{H}[l])^*$.

and $c_2[k]$, i.e. $-c_2^*[T - k + 1]$ and $c_1^*[T - k + 1]$ are transmitted from the first and second antennas, respectively. Note the analogy with the vector $\begin{bmatrix} -c_2^* & c_1^* \end{bmatrix}^T$ transmitted by the classical Alamouti scheme. The received stream $y_2[k]$ during the second burst thereby reads as

$$y_2[k] = \sqrt{\frac{E_s}{2}} \begin{bmatrix} h_1(q^{-1}) & h_2(q^{-1}) \end{bmatrix} \begin{bmatrix} -c_2^*[T - k + 1] \\ c_1^*[T - k + 1] \end{bmatrix} + n_2'[k]. \qquad (9.63)$$

By time-reversing and complex conjugating $y_2[k]$, we get

$$y_2^*[T - k + 1] = \sqrt{\frac{E_s}{2}} \begin{bmatrix} h_2^*(q) & -h_1^*(q) \end{bmatrix} \begin{bmatrix} c_1[k] \\ c_2[k] \end{bmatrix} + n_2[k] \qquad (9.64)$$

where $n_2[k] = n_2'^*[T - k + 1]$. Combining $y_1[k]$ and $y_2^*[T - k + 1]$, we eventually obtain

$$\begin{bmatrix} y_1[k] \\ y_2^*[T - k + 1] \end{bmatrix} = \sqrt{\frac{E_s}{2}} \underbrace{\begin{bmatrix} h_1(q^{-1}) & h_2(q^{-1}) \\ h_2^*(q) & -h_1^*(q) \end{bmatrix}}_{\mathbf{H}_{eff}} \begin{bmatrix} c_1[k] \\ c_2[k] \end{bmatrix} + \begin{bmatrix} n_1[k] \\ n_2[k] \end{bmatrix}. \qquad (9.65)$$

After matched filtering, the final received signal is given by

$$\begin{bmatrix} z_1[k] \\ z_2[k] \end{bmatrix} = \mathbf{H}_{eff}^H \begin{bmatrix} y_1[k] \\ y_2^*[T - k + 1] \end{bmatrix}$$

$$= \begin{bmatrix} h_1^*(q)h_1(q^{-1}) + h_2^*(q)h_2(q^{-1}) \end{bmatrix} \begin{bmatrix} c_1[k] \\ c_2[k] \end{bmatrix} + \mathbf{H}_{eff}^H \begin{bmatrix} n_1[k] \\ n_2[k] \end{bmatrix}, \qquad (9.66)$$

revealing that both streams are completely decoupled, which allows for the use of two independent MLSE decoders.

9.4.3 Alternative constructions

A few alternative code constructions have been proposed in the literature to achieve the full diversity and maximize the coding gain in frequency selective channels, mostly generalizing classical space–time trellis codes [GHL+03]. The drawback of such codes lies in the very high receiver complexity. Indeed, the inner trellis combined with the impulse response of the channel acts as a super trellis whose number of states is proportional to the product of the number of states of the inner code by the number of channel taps. MLSE decoding with so many states leads to very complex receivers. This is another reason why space-frequency MIMO-OFDM has emerged as a more practical approach to tackle frequency selectivity.

9.5   Code design criteria for space-frequency coded MIMO-OFDM transmissions in Rayleigh fading channels

The literature regarding space-frequency coded MIMO-OFDM is extremely large [ATNS98, LW00a, BLWY01, GL02, HH02b, LW00b, BGP02, BP00b, BP01, BBP03, SSOL03, GL01, LXG02]. In this section, we restrict our analysis to the diversity and coding gains achieved by space-frequency codes and we give some insight into how to design efficient codes.

9.5.1   Diversity gain analysis

Similarly to the flat fading case, let us denote the effective length of the mth row ($m = 1, \ldots, n_t$) of the space-frequency codeword error $\mathbf{C} - \mathbf{E}$ by $l_{\mathbf{C},\mathbf{E}}(l) = \sharp \tau_{\mathbf{C},\mathbf{E}}(l)$ with

$$\tau_{\mathbf{C},\mathbf{E}}(l) = \{k \mid \mathbf{c}_k(l) - \mathbf{e}_k(l) \neq 0\}.$$

The effective length of the pair of codewords $\{\mathbf{C}, \mathbf{E}\}$ is denoted as $l_{\mathbf{C},\mathbf{E}} = \sharp \tau_{\mathbf{C},\mathbf{E}}$ with $\tau_{\mathbf{C},\mathbf{E}} = \{k \mid \mathbf{c}_k - \mathbf{e}_k \neq 0\}$. The rank of the codewords error matrix $\tilde{\mathbf{E}} \triangleq (\mathbf{C} - \mathbf{E})(\mathbf{C} - \mathbf{E})^H$ is finally denoted as usual by $r(\tilde{\mathbf{E}})$. The diversity achievable by a pair of codewords $\{\mathbf{C}, \mathbf{E}\}$ over a full-rank correlated channel is outlined by the following proposition [WSZ04, Cle05].

Proposition 9.2   *On a L-tap MIMO channel with full-rank correlation matrix $\underline{\mathbf{R}}$, the diversity gain achieved by a pair of space-frequency codewords $\{\mathbf{C}, \mathbf{E}\}$ with effective lengths $\{l_{\mathbf{C},\mathbf{E}}(l)\}_{l=1}^{n_t}$ and a rank $r(\tilde{\mathbf{E}})$ is*

$$\max_{\{\tau'_{\mathbf{C},\mathbf{E}}(l)\}_{l \in \mathcal{V}_{ind}}} \sharp \left\{ \bigcup_{l \in \mathcal{V}_{ind}} \tau'_{\mathbf{C},\mathbf{E}}(l) \right\} n_r \tag{9.67}$$

where $\mathcal{V}_{ind}$ is the set of the indexes of the rows that form the basis of $\mathbf{C} - \mathbf{E}$ and $\tau'_{\mathbf{C},\mathbf{E}}(l) \subset \tau_{\mathbf{C},\mathbf{E}}(l)$ with $\sharp \tau'_{\mathbf{C},\mathbf{E}}(l) = \min\{l_{\mathbf{C},\mathbf{E}}(l), L\}$. Maximization is carried out over all possible $\{\tau'_{\mathbf{C},\mathbf{E}}(l)\}_{l \in \mathcal{V}_{ind}}$.

PROOF: Since the rank of a matrix is unchanged by left or right multiplication by a non-singular matrix [HJ95].

$$r(\mathbf{C}_{\underline{\mathbf{R}}}) = r(\mathbf{I}_{n_r} \otimes (\underline{\mathbf{C}} - \underline{\mathbf{E}})^H) = n_r r((\underline{\mathbf{C}} - \underline{\mathbf{E}})^H), \tag{9.68}$$

the proof comes to find the rank of matrix

$$\underline{\mathbf{D}}^H (\mathbf{I}_L \otimes (\mathbf{C} - \mathbf{E})^H) = \left[((\mathbf{C} - \mathbf{E})\mathbf{D}_{(0)})^H \ \ldots \ ((\mathbf{C} - \mathbf{E})\mathbf{D}_{(L-1)})^H \right]. \tag{9.69}$$

It then consists in reducing the size of this matrix until it becomes full rank (i.e. its rank is equal to the minimum between its number of rows and columns). Since permuting the

columns of this matrix does not change its rank, we may use the following matrix

$$\Big[\ \mathbf{D}^H_{(0)}((\mathbf{C} - \mathbf{E})(1,:\,))^H\ \ldots\ \mathbf{D}^H_{(L-1)}((\mathbf{C} - \mathbf{E})(1,:\,))^H$$

$$\ldots\ \mathbf{D}^H_{(0)}((\mathbf{C} - \mathbf{E})(n_t,:\,))^H\ \ldots\ \mathbf{D}^H_{(L-1)}((\mathbf{C} - \mathbf{E})(n_t,:\,))^H\ \Big]$$

$$= \Bigg[\ \text{diag}\{(\mathbf{C} - \mathbf{E})^*(1,:\,)\} \begin{bmatrix} \mathbf{d}_{(0)} \\ \vdots \\ \mathbf{d}_{(L-1)} \end{bmatrix}^H$$

$$\ldots\ \text{diag}\{(\mathbf{C} - \mathbf{E})^*(n_t,:\,)\} \begin{bmatrix} \mathbf{d}_{(0)} \\ \vdots \\ \mathbf{d}_{(L-1)} \end{bmatrix}^H\ \Bigg] \tag{9.70}$$

where

$$\mathbf{d}_{(l)} = \begin{bmatrix} 1 & \ldots & e^{-j2\pi kl/T} & \ldots & e^{-j2\pi(T-1)l/T} \end{bmatrix} \tag{9.71}$$

is formed by the diagonal elements of $\mathbf{D}_{(l)}$. In (9.70), all zero rows may be removed, yielding a matrix of size $l_{\mathbf{C},\mathbf{E}} \times n_t L$. Then, since $r(\tilde{\mathbf{E}})$ is the dimension of the subspace spanned by the rows of $\mathbf{C} - \mathbf{E}$, there are $r(\tilde{\mathbf{E}})$ independent rows in $\mathbf{C} - \mathbf{E}$. Let us denote by $\mathcal{V}_{ind}$ the set of the indexes of those rows. Assuming without loss of generality that the basis of the space spanned by the rows of $\mathbf{C} - \mathbf{E}$ is made of $r(\tilde{\mathbf{E}})$ rows of $\mathbf{C} - \mathbf{E}$, $\mathcal{V}_{ind}$ is then given by

$$\mathcal{V}_{ind} = \left\{ l\ \middle|\ \begin{array}{l} \text{the } l^{\text{th}} \text{ row of } \mathbf{C} - \mathbf{E} \text{ is a vector of the basis} \\ \text{of the space spanned by the rows of } \mathbf{C} - \mathbf{E} \end{array} \right\}. \tag{9.72}$$

Naturally, $\sharp\mathcal{V}_{ind} = r(\tilde{\mathbf{E}})$. We may thereby remove the block matrices

$$\mathbf{B}_l = \text{diag}\{(\mathbf{C} - \mathbf{E})^*(l,:\,)\} \begin{bmatrix} \mathbf{d}_{(0)} \\ \vdots \\ \mathbf{d}_{(L-1)} \end{bmatrix}^H \tag{9.73}$$

for $l \notin \mathcal{V}_{ind}$; the resulting matrix being of size $l_{\mathbf{C},\mathbf{E}} \times r(\tilde{\mathbf{E}})L$. We further reduce the dimension of this matrix by suppressing in the remnant block matrices $\{\mathbf{B}_l\}_{l \in \mathcal{V}_{ind}}$ the columns that are linear combinations of other columns of the same block matrices. Since each block matrix $\mathbf{B}_l$ is a Vandermonde matrix left multiplied by a diagonal matrix, its rank is equal to the minimum between the number of columns and rows. There are $l_{\mathbf{C},\mathbf{E}}(l)$ non-zero rows in the block matrix $\mathbf{B}_l$. The rank of $\mathbf{B}_l$ is thus equal to $\min\{l_{\mathbf{C},\mathbf{E}}(l), L\}$. Let us denote by $\mathbf{B}'_l$ the matrix obtained by keeping $\min\{l_{\mathbf{C},\mathbf{E}}(l), L\}$ columns and rows of $\mathbf{B}_l$. The matrix $\mathbf{B}'_l$ is thus full rank and the resulting number of columns is thus equal to $\sum_{l \in \mathcal{V}_{ind}} \min\{l_{\mathbf{C},\mathbf{E}}(l), L\}$.

The elements of $\{(\mathbf{C} - \mathbf{E})(l,:\,)\}_{l \in \mathcal{V}_{ind}}$ are independent. Furthermore, each block matrix $\mathbf{B}'_l$ is full rank. As a result, the rank of the total matrix is a function of the location of the block

matrices within the matrix of interest. Mathematically, this can be expressed as follows. Denoting as $\tau_{C,E}(l) = \{k \mid c_k(l) - e_k(l) \neq 0\}$ with $l \in \mathcal{V}_{ind}$ and as $\tau'_{C,E}(l)$ a subset of $\tau_{C,E}(l)$ of size $\sharp \tau'_{C,E}(l) = \min\{l_{C,E}(l), L\}$, the rank of $\underline{\mathbf{D}}^H\left(\mathbf{I}_L \otimes (\mathbf{C} - \mathbf{E})^H\right)$ is such that

$$r\left(\underline{\mathbf{D}}^H\left(\mathbf{I}_L \otimes (\mathbf{C} - \mathbf{E})^H\right)\right) = \max_{\{\tau'_{C,E}(l)\}_{l \in \mathcal{V}_{ind}}} \sharp\left\{\bigcup_{l \in \mathcal{V}_{ind}} \tau'_{C,E}(l)\right\}. \tag{9.74}$$

For full-rank space-frequency codewords, Proposition 9.2 simplifies into   ∎

Corollary 9.1 *For full-rank space-tap correlation matrix $\underline{\mathbf{R}}$, the diversity gain achieved by a pair of full-rank $\left(r\left(\tilde{\mathbf{E}}\right) = n_t\right)$ space-frequency codewords $\{\mathbf{C}, \mathbf{E}\}$ with effective lengths $\{l_{C,E}(l)\}_{l=1}^{n_t}$ is*

$$\max_{\{\tau'_{C,E}(l)\}_{l=1}^{n_t}} \sharp\left\{\bigcup_{1 \leq l \leq n_t} \tau'_{C,E}(l)\right\} n_r \tag{9.75}$$

where $\tau'_{C,E}(l) \subset \tau_{C,E}(l) = \{k \mid c_k(l) - e_k(l) \neq 0\}$ with $\sharp \tau'_{C,E}(l) = \min\{(l)_{C,E}(l), L\}$. Maximization is carried out over all possible $\{\tau'_{C,E}(l)\}_{(l)=1}^{n_t}$.

This diversity gain is always upper-bounded as obtained in the following proposition.

Proposition 9.3 *For full-rank space-tap correlation matrix $\underline{\mathbf{R}}$, the diversity gain achievable by a pair of space-frequency codewords $\{\mathbf{C}, \mathbf{E}\}$ with an effective length $l_{C,E}$, effective lengths $\{l_{C,E}(l)\}_{l=1}^{n_t}$ and a rank $r(\tilde{\mathbf{E}})$ is upper-bounded by $\min\{l_{C,E}, r(\tilde{\mathbf{E}})L\}n_r$. Equality occurs if $l_{C,E}(l) \geq L, \forall l = 1, \ldots, n_t$.*

PROOF: Assuming that $\underline{\mathbf{R}}$ is full rank, the rank of $\mathbf{C}_{\mathbf{R}}$ is upper-bounded by

$$n_r \min r\left(\underline{\mathbf{D}}^H\left(\mathbf{I}_L \otimes (\mathbf{C} - \mathbf{E})^H\right)\right). \tag{9.76}$$

Without loss of generality, we can assume that in $\underline{\mathbf{D}}$, we have removed the columns and rows whose positions correspond to the positions of the zero columns in $\mathbf{C} - \mathbf{E}$. It is straightforward to find that $r(\underline{\mathbf{D}}) = l_{C,E}$. Moreover, since $r(\mathbf{I}_L \otimes (\mathbf{C} - \mathbf{E})) = Lr(\mathbf{C} - \mathbf{E})$, the rank inequality on the product of matrices implies that the maximum achievable diversity is equal to $n_r \min\{l_{C,E}, r(\tilde{\mathbf{E}})L\}$. The equality results from the application of Proposition 9.2 to the case where $l_{C,E}(l) \geq L, \forall l = 1, \ldots, n_t$.   ∎

Corollary 9.2 *For full-rank space-tap correlation matrix $\underline{\mathbf{R}}$, a pair of space-frequency codewords $\{\mathbf{C}, \mathbf{E}\}$ with an effective length $l_{C,E}$, effective lengths $\{l_{C,E}(l)\}_{l=1}^{n_t}$ and a rank $r(\tilde{\mathbf{E}})$ achieves the full diversity $n_t n_r L$ if*

$$r(\tilde{\mathbf{E}}) = n_t,$$

$$l_{C,E}(l) \geq L, \ \forall l = 1, \ldots, n_t,$$

$$l_{C,E} \geq n_t L.$$

Example 9.3 *Let us consider the pair of space-frequency codewords* $\mathbf{C}, \mathbf{E}$ *is given by the following examples (a diamond* ♦ *denotes a non-zero value).*

- $r(\tilde{\mathbf{E}}) = 2$ *and* $L = 3$ *with*

$$
\mathbf{C} - \mathbf{E} = \begin{bmatrix} \blacklozenge\ \blacklozenge & & \\ \blacklozenge\ \blacklozenge & & \\ & \blacklozenge\ \blacklozenge\ \blacklozenge\ \blacklozenge\ \blacklozenge & \end{bmatrix} \tag{9.77}
$$

Since this is a rank-2 error matrix, we may choose e.g. $\mathcal{V}_{ind}$ *as* $\mathcal{V}_{ind} = \{2, 3\}$ *and write for row 2,* $l_{\mathbf{C},\mathbf{E}}(2) = 2$, $\sharp\tau'_{\mathbf{C},\mathbf{E}}(2) = 2$ *and for row 3,* $l_{\mathbf{C},\mathbf{E}}(3) = 5$, $\sharp\tau'_{\mathbf{C},\mathbf{E}}(3) = 3$. *The subset* $\tau'_{\mathbf{C},\mathbf{E}}(2)$ *is easily chosen as* $\tau'_{\mathbf{C},\mathbf{E}}(2) = \{1, 2\}$ *while for* $\tau'_{\mathbf{C},\mathbf{E}}(3)$, *several options are possible. A first option is* $\tau'_{\mathbf{C},\mathbf{E}}(3) = \{2, 3, 4\}$ *so that* $\tau'_{\mathbf{C},\mathbf{E}}(2) \cup \tau'_{\mathbf{C},\mathbf{E}}(3) = \{1, 2, 3, 4\}$ *and the cardinality of this set is 4. Another option is* $\tau'_{\mathbf{C},\mathbf{E}}(3) = \{3, 4, 5\}$ *such that* $\tau'_{\mathbf{C},\mathbf{E}}(2) \cup \tau'_{\mathbf{C},\mathbf{E}}(3) = \{1, 2, 3, 4, 5\}$ *and the cardinality of this set is 5. Since the achieved diversity is the maximal cardinality over all possible subsets, the transmit diversity achieved by the present error matrix is 5.*

- $r(\tilde{\mathbf{E}}) = 1$ *and* $L = 2$ *with* $\mathbf{C} - \mathbf{E}$ *such that*

$$
\mathbf{C} - \mathbf{E} = \begin{bmatrix} \blacklozenge\ \blacklozenge\ \blacklozenge\ \blacklozenge \\ \blacklozenge\ \blacklozenge\ \blacklozenge\ \blacklozenge \end{bmatrix} \tag{9.78}
$$

The transmit diversity is now equal to 2 and achieves the upper-bound of Proposition 9.3.

- $r(\tilde{\mathbf{E}}) = 2$ *and* $L = 2$ *with* $\mathbf{C} - \mathbf{E}$ *such that*

$$
\mathbf{C} - \mathbf{E} = \begin{bmatrix} \blacklozenge\ \blacklozenge\ \blacklozenge\ \blacklozenge \\ \blacklozenge\ \blacklozenge\ \blacklozenge\ \blacklozenge \end{bmatrix} \tag{9.79}
$$

The transmit diversity is equal to the full diversity of 4 (see Corollary 9.2).

Therefore, the achievable diversity is not a function of the space-tap correlation matrix (as long as the correlation matrix is full rank) but of the code properties. However, analogous to the flat fading case, space-tap correlations may decrease the coding gain. Hence, even if the asymptotic diversity (at high SNR) is not affected by space-tap correlations, the apparent finite-SNR diversity might be severely affected when the channel is highly correlated at the transmitter. In the rest of Section 9.5, we therefore assume that the transmit correlation is zero and consider joint receive-tap semi-correlated channels. By contrast, Section 9.5 deals with space-tap correlated channels with high transmit correlation.

9.5.2 Coding gain analysis

From space-tap correlation to space-frequency correlation

We consider that the inter-element spacing and the scattering richness at the transmit side are large enough to neglect the transmit correlation and represent the space-tap correlation

matrix as $\underline{\mathbf{R}} = \mathbf{R}_{r,L} \otimes \mathbf{I}_{n_t}$. The joint receive-tap correlation matrix is defined as

$$\mathbf{R}_{r,L} = \mathcal{E}\left\{\text{vec}\left(\tilde{\mathbf{H}}_{r,L}^H\right)\text{vec}\left(\tilde{\mathbf{H}}_{r,L}^H\right)^H\right\} \tag{9.80}$$

with

$$\tilde{\underline{\mathbf{H}}}_{r,L} = \left[\; \tilde{\mathbf{H}}[0](:,1)\; \tilde{\mathbf{H}}[1](:,1)\; \cdots\; \tilde{\mathbf{H}}[L-1](:,1)\;\right]. \tag{9.81}$$

As with (9.50), we decompose $\mathbf{R}_{r,L}$ into n_r^2 blocks $\mathbf{R}_{r,L}^{(pq)}$ for $p = 1, \ldots, n_r$ and $q = 1, \ldots, n_r$. Each block $\mathbf{R}_{lj}^{(pq)}$ is then modeled as $\mathbf{R}_{r,L}^{(pq)}(l,j)\mathbf{I}_{n_t}$. In this case, $\underline{\mathbf{C}_{\mathbf{R}}}^{(pq)} = (\underline{\mathbf{C}}-\underline{\mathbf{E}})^H \mathbf{R}^{(pq)}(\underline{\mathbf{C}}-\underline{\mathbf{E}})$ can be reformulated as

$$\begin{aligned}
(\underline{\mathbf{C}}-\underline{\mathbf{E}})^H \mathbf{R}^{(pq)}(\underline{\mathbf{C}}-\underline{\mathbf{E}}) &= \sum_{j=0}^{L-1}\sum_{l=0}^{L-1}(\mathbf{C}-\mathbf{E})_{(l)}^H \mathbf{R}_{lj}^{(pq)}(\mathbf{C}-\mathbf{E})_{(j)} \\
&= \sum_{j=0}^{L-1}\sum_{l=0}^{L-1}\mathbf{R}_{r,L}^{(pq)}(l,j)(\mathbf{C}-\mathbf{E})_{(l)}^H(\mathbf{C}-\mathbf{E})_{(j)} \\
&= \mathbf{R}_f^{(pq)} \odot \tilde{\mathbf{E}}^H
\end{aligned} \tag{9.82}$$

where $\mathbf{R}_f^{(pq)}$ is defined as

$$\mathbf{R}_f^{(pq)} = \sum_{j=0}^{L-1}\sum_{l=0}^{L-1}\mathbf{R}_{r,L}^{(pq)}(l,j)\mathbf{d}_{(l)}^H\mathbf{d}_{(j)}. \tag{9.83}$$

We recall that $\mathbf{d}_{(l)}$ was defined in (9.71). We may construct a space-frequency correlation matrix $\mathbf{R}_f$ from the blocks $\mathbf{R}_f^{(pq)}$ as

$$\mathbf{R}_f = \begin{bmatrix} \mathbf{R}_f^{(11)} & \cdots & \mathbf{R}_f^{(1n_r)} \\ \vdots & \ddots & \vdots \\ \mathbf{R}_f^{(n_r 1)} & \cdots & \mathbf{R}_f^{(n_r n_r)} \end{bmatrix}. \tag{9.84}$$

The matrix $\underline{\mathbf{C}_{\mathbf{R}}}$ then becomes

$$\underline{\mathbf{C}_{\mathbf{R}}} = \mathbf{R}_f \odot \left[\mathbf{1}_{n_r \times n_r} \otimes \tilde{\mathbf{E}}^H\right]. \tag{9.85}$$

The space-frequency correlation matrix $\mathbf{R}_f$ has the same meaning as the correlation function of the aperture-resolved transfer function of (2.9) and takes into account the correlation across both the receive discrete spatial domain and the discrete frequency domains (i.e. the tones). Note that if each tap l is i.i.d. Rayleigh distributed with an average power β_l and if there is no correlation between taps, the space-frequency correlation matrix reads as

$$\mathbf{R}_f = \mathbf{I}_{n_r} \otimes \left[\sum_{l=0}^{L-1}\beta_l \mathbf{d}_{(l)}^H\mathbf{d}_{(l)}\right]. \tag{9.86}$$

Impact of frequency correlation on the coding gain

The determinant operator in the average PEP of (9.49) can now be expressed as

$$
\begin{aligned}
\det{}&(\mathbf{I}_{Tn_r} + \eta\mathbf{C_R}) \\
&= \det\left(\mathbf{I}_{Tn_r} + \eta\mathbf{R}_f \odot \left[\mathbf{1}_{n_r \times n_r} \otimes \tilde{\mathbf{E}}^H\right]\right) \\
&\leq \det\left(\mathbf{I}_{Tn_r} + \eta\mathrm{diag}\left\{\mathbf{R}_f^{11} \ldots \mathbf{R}_f^{n_r n_r}\right\} \odot \left[\mathbf{1}_{n_r \times n_r} \otimes \tilde{\mathbf{E}}^H\right]\right) \\
&= \prod_{n=1}^{n_r} \det\left(\mathbf{I}_{Tn_r} + \eta\mathbf{I}_{n_r} \otimes \left[\mathbf{R}_f^{nn} \odot \tilde{\mathbf{E}}^H\right]\right) \\
&\leq \prod_{k=0}^{T-1} \left(1 + \eta\left[\sum_{l=0}^{L-1} \beta_l\right] \|\mathbf{c}_k - \mathbf{e}_k\|^2\right)^{n_r}.
\end{aligned}
\tag{9.87}
$$

Clearly, the performance of MIMO-OFDM approaches the performance of narrowband MIMO transmissions over i.i.d. fast fading Rayleigh channels if $\mathbf{R}_f$ is diagonal, indicating that the presence of receive-frequency correlation reduces the achievable coding gain. Yet, it is possible to reduce the frequency correlation between adjacent tones by means of an interleaver. Indeed, interleaving optimizes the symbol positions in which the pair of codewords differ in order to minimize the off-diagonal elements of $\mathbf{R}_f$. The previous discussion also suggests using codes with large effective length and product distance for OFDM transmissions as in fast fading channels.

If a random interleaver is classically used [ATNS98, LW00a, BLWY01, GL02, HH02b, LW00b, SSOL03, GL01, LXG02], a block interleaver whose depth is optimized as a function of L and $\{\beta_l\}_{l=0,\ldots,L-1}$ may provide an increased performance. Indeed, it is able to control the frequency correlation between adjacent tones in a much more efficient manner [GHL+03, WSZ04]. The block interleaver is designed so that pairs of codewords differ at positions corresponding to a low frequency correlation. When nothing is known at the transmitter about the number of taps and the power-delay profile, there is no closed form solution for the interleaving depth. However, it is likely that the dominant error events are those with the smallest error lengths. Hence, as a rule of thumb, the interleaver should guarantee that the pairs of codewords with the smallest effective lengths differ at positions where the frequency correlation is low.

In order to illustrate the impact of the interleaver and the effect of the number of taps on the performance, the frame error rate (FER) of the 16-state 'FVY' code [FVY01] in 1×2 i.i.d. Rayleigh channels with uniform tap power distribution ($\beta_l = \beta$, $l = 0, \ldots, L - 1$) is displayed in Figure 9.3 (128 tones are used). The minimum effective length of this code is 3, suggesting that the largest achievable diversity gain is also equal to 3. The graph shows that the performance improves as L increases when the interleaver is used, the degrees of freedom in the channel being then more efficiently exploited. Moreover, the frequency correlation between adjacent tap also decreases as L increases. Both the diversity and the coding gains are therefore improved.
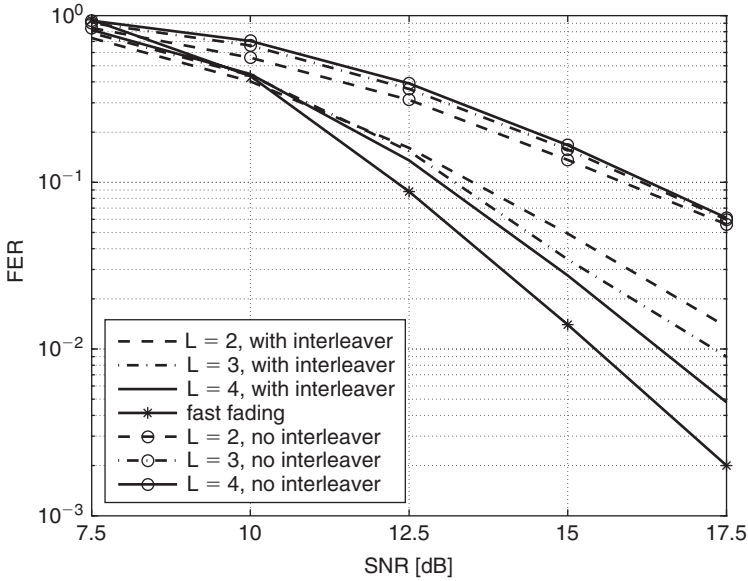
Figure 9.3   *FER of the 16-state 'FVY' code for $L = 2$, 3 and 4 in uniformly distributed i.i.d. Rayleigh channels with and without interleaver*

9.5.3   Space-frequency linear block coding

As far as linear space-frequency block codes are concerned, the following proposition provides some optimality conditions to minimize the maximum average PEP when the channel is spatially i.i.d. Rayleigh distributed with independent non-uniformly distributed taps. In such case, we know that the average PEP is given by (9.54).

Proposition 9.4   *A PSK/QAM-based linear space-frequency block code consisting of wide unitary basis matrices $\left\{\underline{\Phi}_q\right\}_{q=1}^{2Q}$ minimizes the worst-case PEP averaged over i.i.d. Rayleigh slow fading channels with independent non-uniformly distributed taps if (sufficient condition) the matrices $\left\{\underline{\Phi}_q\right\}_{q=1}^{2Q}$ are pairwise skew-hermitian, i.e. they satisfy the conditions*

$$\underline{\Phi}_q\underline{\Phi}_p^H + \underline{\Phi}_p\underline{\Phi}_q^H = 0 \ \ q \neq p. \tag{9.88}$$

Those conditions are sufficient and necessary in order to minimize the average Chernoff upper-bound of the union bound (5.4).

PROOF: The proof is similar to the proof of Proposition 5.3.   ∎

Orthogonal space-frequency block coding
Orthogonal space-frequency block codes (O-SFBCs) satisfy the pairwise skew-hermitian condition and may be thought of as a generalization of O-STBCs to frequency selective

channels. To design these codes, let us decompose $\underline{\Phi}_q$ into L blocks, the lth block being equal to $\Phi_q \mathbf{D}_{(l)}$. The $(m, n)^{\text{th}}$ element of the $(u, v)^{\text{th}}$ block of $\underline{\Phi}_q \underline{\Phi}_p^H + \underline{\Phi}_p \underline{\Phi}_q^H$, denoted as $\left(\underline{\Phi}_q \underline{\Phi}_p^H + \underline{\Phi}_p \underline{\Phi}_q^H \right)_{u,v}(m, n)$, is given by

$$\left(\underline{\Phi}_q \underline{\Phi}_p^H + \underline{\Phi}_p \underline{\Phi}_q^H \right)_{u,v}(m, n)$$
$$= \sum_{k=0}^{T-1} e^{-j\frac{2\pi}{T} k(u-v)} \left(\Phi_q(m, k)\Phi_p^*(n, k) + \Phi_p(n, k)\Phi_q^*(m, k) \right). \quad (9.89)$$

The conditions of (9.88) are equivalent to

$$\left(\underline{\Phi}_q \underline{\Phi}_p^H + \underline{\Phi}_p \underline{\Phi}_q^H \right)_{u,v}(m, n) = \begin{cases} 2\frac{T}{LQn_t} & q = p \text{ and } u = v \text{ and } m = n. \\ 0 & \text{otherwise} \end{cases} \quad (9.90)$$

When $u = v$, those conditions are only satisfied by an orthogonal code or repetition code whose inner code is orthogonal. Since the former is only able to exploit the transmit diversity, we focus on the latter. The repetition rate is denoted as Σ and the frequency separation is denoted as Δ in terms of tones between the repeated orthogonal codes. It is quite clear from (9.89) that the conditions of (9.90) for $u \neq v$ are only satisfied if $\Sigma = L$ and $\Delta = T/L$. The repetition rate of L ensures that the full diversity is achieved while Δ plays a role similar to the aforementioned block interleaver. Hence, we maximize the coding gain and achieve the full diversity. This code is also very efficient in terms of decoding complexity since the received streams are decoupled after matched filtering (similarly to O-STBC). Naturally, the price to pay is a reduction in spectral efficiency caused by the repetition. Hence, usually $\Sigma = 1$ is chosen to exploit full transmit diversity and an outer code is used to exploit the remnant frequency diversity.

Example 9.4 *Let us express the equivalent channel (after matched filtering) of an Alamouti based space-frequency repetition code in 1×2 frequency selective Rayleigh channels of various lengths ($L = 1, 2$ or 3). Since $n_r = 1$, the channel matrices are denoted in lower case* $\mathbf{h}$*. Assume that we design the code for a two-tap channel ($L = 2$). Hence, we must take $\Sigma = 2$ and $\Delta = T/2$. In other words, if the first space-frequency codeword is transmitted on tones v and $v + w$, the second codeword is thus transmitted on tones $v + T/2$ and $v + w + T/2$. The transfer function between transmit and receive antennas on tone v is written as $\mathbf{h}_{(v)} = \begin{bmatrix} \mathbf{h}_{(v)}(1) & \mathbf{h}_{(v)}(2) \end{bmatrix}$. The received signals on the tones v, $v + w$, $v + T/2$ and $v + T/2 + w$ are as follows*

$$y_v = \mathbf{h}_{(v)}(1)c_1 + \mathbf{h}_{(v)}(2)c_2 + n_v$$
$$y_{v+w} = -\mathbf{h}_{(v+w)}(1)c_2^* + \mathbf{h}_{(v+w)}(2)c_1^* + n_{v+w}r$$
$$y_{v+T/2} = \mathbf{h}_{(v+T/2)}(1)c_1 + \mathbf{h}_{(v+T/2)}(2)c_2 + n_{v+T/2}$$
$$y_{v+T/2+w} = -\mathbf{h}_{(v+T/2+w)}(1)c_2^* + \mathbf{h}_{(v+T/2+w)}(2)c_2^* + n_{v+T/2+w}. \quad (9.91)$$

This can be reformulated as follows

$$
\underbrace{\begin{bmatrix} y_v \\ y_{v+w}^* \\ y_{v+T/2} \\ y_{v+T/2+w}^* \end{bmatrix}}_{\mathbf{y}} = \mathbf{H}_{eff} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \underbrace{\begin{bmatrix} n_v \\ n_{v+w}^* \\ n_{v+T/2} \\ n_{v+T/2+w}^* \end{bmatrix}}_{\mathbf{n}}
\tag{9.92}
$$

where

$$
\mathbf{H}_{eff} = \begin{bmatrix} \mathbf{h}_{(v)}(1) & \mathbf{h}_{(v)}(2) \\ \mathbf{h}_{(v+w)}^*(2) & -\mathbf{h}_{(v+w)}^*(1) \\ \mathbf{h}_{(v+T/2)}(1) & \mathbf{h}_{(v+T/2)}(2) \\ \mathbf{h}_{(v+T/2+w)}^*(2) & -\mathbf{h}_{(v+T/2+w)}^*(1) \end{bmatrix}.
\tag{9.93}
$$

Let us examine the received signal after matched filtering for the three following channels.

- $L = 1$. *The channel is equivalent to a frequency flat channel, hence no frequency selectivity can be extracted. The channel gains[2] are* $\mathbf{h}_{(z)}(1) = \mathbf{h}[0](1)$ *and* $\mathbf{h}_{(z)}(2) = \mathbf{h}[0](2)$ $\forall z = 0, \ldots, T-1$. *After matched filtering* $(\mathbf{H}_{eff}^H)$, *we have*

$$
\mathbf{H}_{eff}^H \mathbf{y} = 2\left[|\mathbf{h}[0](1)|^2 + |\mathbf{h}[0](2)|^2\right] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} + \mathbf{H}_{eff}^H \mathbf{n}.
\tag{9.94}
$$

  The decoupling is achieved but no frequency selectivity is exploited.
- $L = 2$. *The channel gains are* $\mathbf{h}_{(z)}(1) = \mathbf{h}[0](1) + \mathbf{h}[1](1)e^{-j2\pi z/T}$ *and* $\mathbf{h}_{(z)}(2) = \mathbf{h}[0](2) + \mathbf{h}[1](2)e^{-j2\pi z/T}$, $\forall z = 0, \ldots, T-1$. *After matched filtering, we have*

$$
\mathbf{H}_{eff}^H \mathbf{y} = 2\|\underline{\mathbf{h}}\|_F^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} + \mathbf{H}_{eff}^H \mathbf{n}
\tag{9.95}
$$

  with $\|\underline{\mathbf{h}}\|_F^2 = \left[|\mathbf{h}[0](1)|^2 + |\mathbf{h}[0](2)|^2 + |\mathbf{h}[1](1)|^2 + |\mathbf{h}[1](2)|^2\right]$. *As expected, the streams are decoupled and the frequency selectivity is fully exploited (diversity gain of 4) irrespective of w.*
- $L = 3$. *The channel is longer than the repetition rate of the code. The channel gains are*

$$
\mathbf{h}_{(z)}(1) = \mathbf{h}[0](1) + \mathbf{h}[1](1)e^{-j2\pi z/T} + \mathbf{h}[2](1, 1)e^{-j2\pi 2z/T}
\tag{9.96}
$$

$$
\mathbf{h}_{(z)}(2) = \mathbf{h}[0](2) + \mathbf{h}[1](2)e^{-j2\pi z/T} + (\mathbf{h}[2])(2)e^{-j2\pi 2z/T}
\tag{9.97}
$$

$\forall z = 0, \ldots, T-1$. *After matched filtering, we have*

$$
\mathbf{H}_{eff}^H \mathbf{y} = \begin{bmatrix} b & a \\ a^* & c \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} + \mathbf{H}_{eff}^H \mathbf{n}
\tag{9.98}
$$

---

[2] $\mathbf{h}[l](m)$ denotes the mth element of $\mathbf{h}[l]$.

where

$$b = 2\left[|\mathbf{h}[0](1) + \mathbf{h}[2](1)e^{-j2\pi 2v/T}|^2 + |\mathbf{h}[1](1)|^2 \right.$$
$$\left. + |\mathbf{h}[1](2)|^2 + |\mathbf{h}[0](2) + \mathbf{h}[2](2)e^{-j2\pi 2(v+w)/T}|^2 \right]$$

$$c = 2\left[|\mathbf{h}[0](1) + \mathbf{h}[2](1)e^{-j2\pi 2(v+w)/T}|^2 + |\mathbf{h}[1](1)|^2 \right.$$
$$\left. + |\mathbf{h}[1](2)|^2 + |\mathbf{h}[0](2) + \mathbf{h}[2](2)e^{-j2\pi 2v/T}|^2 \right]$$

$$a = 2\left[\mathbf{h}[0]^*(1,1)\mathbf{h}[2](2)e^{-j2\pi 2v/T} - \mathbf{h}[2](2)\mathbf{h}[0]^*(1)e^{-j2\pi 2(v+w)/T} \right.$$
$$\left. + \mathbf{h}[2]^*(1)\mathbf{h}[0](2)e^{j2\pi 2v/T} - \mathbf{h}[0](2)\mathbf{h}[2]^*(1)e^{j2\pi 2(v+w)/T} \right].$$

A diversity gain of 4 is achieved. Moreover, it is recommended to choose w as small as possible (e.g. w = 1) in order to exploit the decoupling properties if the number of tones T is large. Indeed if w/T << 1, then a ≈ 0.

9.5.4 Cyclic delay-diversity

Cyclic delay-diversity (CDD) [DK01, GSP02], also known as cyclic shift diversity, is an adaptation of the generalized delay-diversity (GDD) scheme (see Section 9.4.1) to OFDM systems. The idea is to send on each antenna a circularly shifted version of the same OFDM symbol in the time domain. Hence, the temporal delay introduced on each antenna in the GDD scheme is transformed into a cyclic delay in the CDD scheme. Analogous to GDD, CDD turns the MIMO channel into a SIMO channel with enhanced frequency selectivity, and the subsequent frequency diversity may then be extracted by appropriate outer codes.

An advantage of the CDD scheme is the reduced guard interval with regard to GDD. In the latter, a non-negligible delay proportional to the number of transmit antennas and the length of the channel must be introduced to exploit the full diversity and avoid ISI. In CDD, the cyclic character of the delay reduces the guard interval and thereby improves the spectral efficiency. The guard interval in CDD is only proportional to the channel length L. Another advantage of CDD over O-SFBC is the increased flexibility and the scalability to any number of transmit antennas. In addition, there is no rate loss if more than two transmit antennas are used, and the channel is not required to remain constant over several tones, unlike in O-SFBC and O-STBC. Finally, the receiver is essentially the same as a classical SIMO receiver. However, the number of states of the outer code necessary to exploit the full diversity is much larger with CDD than with O-SFBC.

As shown in Figure 9.4, a CDD transmission works as follows. A sequence $\mathbf{c}$ of symbols c_k with $k = 0, \ldots, T-1$ (that have been preliminary error-control encoded) is serially to parallel converted and OFDM modulated. The output sequence is denoted as $\mathbf{x}$ (as in (9.8)), whose entries are denoted as x_n, $n = 0, \ldots, T-1$. This sequence is transmitted on
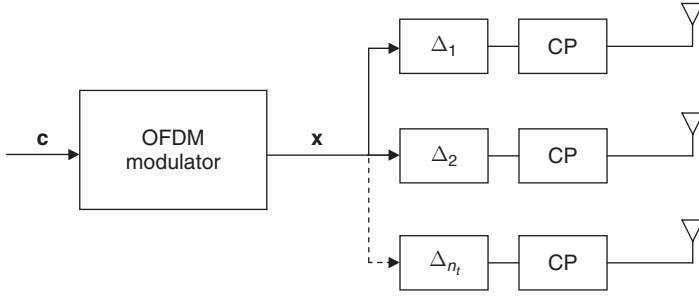
Figure 9.4   *Cyclic delay diversity in MIMO-OFDM*

each antenna with a cyclic delay Δ_m, $m = 1, \ldots, n_t$ so that the output symbol on antenna m ($m = 1, \ldots, n_t$) at time n ($n = 0, \ldots, T-1$) is given by $x_{(n-\Delta_m) \bmod T}$. Finally, a cyclic prefix is added on each antenna, analogous to conventional OFDM transmissions. At the receiver, the cyclic prefix is removed, and OFDM demodulation and decoding are performed.

As previously mentioned, CDD transforms the MIMO transmission into an equivalent SIMO transmission with enhanced frequency selectivity. Indeed, a cyclic shift in the time domain corresponds to the multiplication by a phase shift in the frequency domain. Therefore, the received signal in the frequency domain may be rewritten as

$$\mathbf{y}_k = \sqrt{\frac{E_s}{n_t}} \mathbf{h}_{eq,(k)} c_k + \mathbf{n}_k \tag{9.99}$$

where the equivalent SIMO channel matrix on the kth tone, denoted as $\mathbf{h}_{eq,(k)}$, is given by

$$\mathbf{h}_{eq,(k)} = \sum_{m=1}^{n_t} \mathbf{H}_{(k)}(:,m) \, e^{-j\frac{2\pi}{T}k\Delta_m} \tag{9.100}$$

and $\mathbf{H}_{(k)}$ is the DFT of the impulse response evaluated on the kth subcarrier and outlined by (9.19). Note that we can also express the equivalent channel induced by CDD as

$$\mathbf{h}_{eq,(k)} = \mathbf{H}_{(k)} \left[e^{-j\frac{2\pi}{T}k\Delta_1} \, \cdots \, e^{-j\frac{2\pi}{T}k\Delta_{n_t}} \right]^T, \tag{9.101}$$

suggesting that the transmitted codeword c_k transmitted on tone k (as defined by (9.18)) can be seen as

$$c_k = \frac{1}{\sqrt{n_t}} \left[e^{-j\frac{2\pi}{T}k\Delta_1} \, \cdots \, e^{-j\frac{2\pi}{T}k\Delta_{n_t}} \right]^T c_k. \tag{9.102}$$

If the channel is flat fading ($L = 0$), the equivalent SIMO channel matrix is simply given by

$$\mathbf{h}_{eq,(k)} = \sum_{m=1}^{n_t} \mathbf{H}(:,m) \, e^{-j\frac{2\pi}{T}k\Delta_m}. \tag{9.103}$$

Consequently, the equivalent channel is virtually made of Ln_t taps, and an outer code with a large number of states is able to exploit a diversity gain of Ln_t. However, the

effective number of taps (i.e. those increasing the diversity) highly depends on Δ_m, which should be optimized in order to exploit the full diversity and provide a large array gain. If $\Delta_m = (m-1)\Delta$, Δ should be chosen as at least equal or larger than L. Taking Δ smaller than L reduces the number of effective taps since some of them may experience the same phase shift. This issue is very similar to the choice of the delay on the different antennas in GDD. The discussion of Section 9.5.2 about the frequency correlation matrix $\mathbf{R}_f$ also comes to mind, as optimizing Δ is nothing more than minimizing the frequency correlation between adjacent tones. As expected from Section 9.5.2, the choice of the interleaver and the outer code (with large minimum effective length) for CDD is critical to fully extract frequency diversity. More information regarding the optimization of the cyclic delay, the interleaver and the outer code for CDD is found in [TS04, BM06]. We limit ourselves to the following example.

Example 9.5 *Let us transmit the sequence* $\mathbf{c} = [\, c_0 \,\ldots\, c_{T-1} \,]^T$ *applying CDD in a L-tap channel with $n_t = 2$ and $n_r = 1$. The signal after IDFT writes as $\mathbf{x} = \mathcal{D}^H \mathbf{c}$. Consider the two following cases: $L = 1$ and $L = 2$.*

- $L = 1$. *The codeword sent on second antenna is cyclically shifted with regard to the codeword sent on the first antenna. If the cyclic shift corresponds to one symbol duration (analogous to standard delay-diversity), $\Delta_m = m - 1$ and the codeword matrix before the CP extension reads as*

$$\mathbf{X} = \frac{1}{\sqrt{2}} \begin{bmatrix} x_0 & x_1 & \ldots & x_{T-2} & x_{T-1} \\ x_1 & x_2 & \ldots & x_{T-1} & x_0 \end{bmatrix}. \tag{9.104}$$

Since the channel is flat fading, no CP is actually required. The channel matrix is simply given by $\mathbf{h} = [\, h_1 \; h_2 \,]$ and the received signal reads as

$$\mathbf{z} = \sqrt{E_s}\mathbf{h}\mathbf{X} + \mathbf{n} \tag{9.105}$$

which can be re-formulated as

$$\mathbf{z}^T = \sqrt{\frac{E_s}{2}} \begin{bmatrix} h_1 & h_2 & 0 & \ldots & 0 \\ 0 & h_1 & h_2 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ h_2 & \ldots & 0 & \ldots & h_1 \end{bmatrix} \underbrace{\begin{bmatrix} x_0 \\ \vdots \\ x_{T-1} \end{bmatrix}}_{\mathbf{x}} + \mathbf{n}^T. \tag{9.106}$$

We note that CDD results in a circulant matrix. Hence, applying the DFT operator on $\mathbf{z}^T$ we get

$$\mathbf{y}^T = \mathcal{D}\mathbf{z}^T = \sqrt{\frac{E_s}{2}} \operatorname{diag}\left\{\mathbf{h}_{eq,(0)}, \ldots, \mathbf{h}_{eq,(T-1)}\right\}\mathbf{c} + \mathcal{D}\mathbf{n}^T \tag{9.107}$$

where $\mathbf{h}_{eq,(k)}$ are the DFT values of the vector channel $\mathbf{h}$

$$\mathbf{h}_{eq,(k)} = \sum_{m=1}^{n_t} h_m \, e^{-j\frac{2\pi}{T}k(m-1)}. \tag{9.108}$$

Note that it is possible to improve the frequency selectivity if Δ_m is further optimized.

* *$L = 2$. We shift now the codeword on the second antenna by two symbol durations, $\Delta_m = 2(m-1)$. The codeword matrix before the CP extension is then given by*

$$\mathbf{X} = \frac{1}{\sqrt{2}} \begin{bmatrix} x_0 & x_1 & x_2 & \dots & x_{T-2} & x_{T-1} \\ x_2 & x_3 & \dots & x_{T-1} & x_0 & x_1 \end{bmatrix}. \tag{9.109}$$

Since the channel length is equal to 2, a CP of length 1 is required. Let us transmit x_{T-1} and x_1 as the CP on antennas 1 and 2, respectively.

$$\mathbf{X}' = \frac{1}{\sqrt{2}} \begin{bmatrix} x_{T-1} & x_0 & x_1 & x_2 & \dots & x_{T-2} & x_{T-1} \\ x_1 & x_2 & x_3 & \dots & x_{T-1} & x_0 & x_1 \end{bmatrix}. \tag{9.110}$$

Denoting the virtual channel matrix as $\underline{\mathbf{h}} = \begin{bmatrix} \mathbf{h}[0] & \mathbf{h}[1] \end{bmatrix}$, the received signal after removing the CP becomes

$$\mathbf{z} = \sqrt{\frac{E_s}{2}} \underline{\mathbf{h}} \, \underline{\mathbf{X}} + \mathbf{n} \tag{9.111}$$

with

$$\underline{\mathbf{X}} = \begin{bmatrix} x_0 & x_1 & \dots & x_{T-1} \\ x_2 & x_3 & \dots & x_1 \\ x_{T-1} & x_0 & \dots & x_{T-2} \\ x_1 & x_2 & \dots & x_0 \end{bmatrix}. \tag{9.112}$$

Equation (9.111) can be rewritten as

$$\mathbf{z}^T = \sqrt{\frac{E_s}{2}} \mathbf{H}_{cp}\mathbf{x} + \mathbf{n}^T \tag{9.113}$$

with

$$\mathbf{H}_{cp} = \begin{bmatrix} h[0](1) & h[1](2) & h[0](2) & 0 & \dots & 0 & h[1](1) \\ h[1](1) & h[0](1) & h[1](2) & h[0](2) & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ h[1](2) & h[0](2) & 0 & \dots & 0 & h[1](1) & h[0](1) \end{bmatrix}. \tag{9.114}$$

Since $\mathbf{H}_{cp}$ is circulant, the application of the DFT operator $\mathcal{D}$ on $\mathbf{z}^T$ transforms (9.113) into a set of parallel channels $\{\mathbf{h}_{eq,(k)}\}$ whose gains are given by the DFT values of the vector channel

$$\mathbf{h}_{eq,(k)} = \sum_{m=1}^{n_t} \sum_{l=0}^{L-1} \mathbf{h}[l](m) \, e^{-j\frac{2\pi}{T}k(l+2(m-1))}. \tag{9.115}$$

It is worth noting that $\underline{\mathbf{X}}$ is reminiscent of the virtual codeword $\underline{\mathbf{C}}$ introduced in Section 9.1.3. Therefore, it is easy to conclude that this scheme delivers the full diversity gain since $\underline{\mathbf{X}}$ is circulant and thereby full rank [GSP02].

9.6 On the robustness of codes in spatially correlated frequency selective channels

We have shown that the high SNR diversity gain is not affected by full-rank correlation matrices, but that this might not be so at finite SNR when the transmit correlation is high. To better quantify the achievable diversity gain in this case, we introduce the concept of degenerate taps. The motivation for such a concept is analogous to the flat fading case and results from discussions detailed in Chapters 6 and 7.

9.6.1 Degenerate taps

To simplify the analysis, we consider that each tap is semi-correlated Rayleigh fading with no correlation at the receiver. In this case, the space-tap correlation matrix may be written as $\underline{\mathbf{R}} = \mathbf{I}_{n_r} \otimes \mathbf{R}_{t,L}$, where the joint transmit-tap correlation matrix reads as

$$\mathbf{R}_{t,L} = \mathcal{E}\left\{ \text{vec}(\tilde{\underline{\mathbf{H}}}_{t,L}^H)\text{vec}(\tilde{\underline{\mathbf{H}}}_{t,L}^H)^H \right\} \tag{9.116}$$

with

$$\tilde{\underline{\mathbf{H}}}_{t,L} = \left[\tilde{\mathbf{H}}[0](1,:) \cdots \tilde{\mathbf{H}}[L-1](1,:) \right]. \tag{9.117}$$

Analogous to eqn (9.50), we decompose $\mathbf{R}_{t,L}$ into L^2 blocks $\mathbf{R}_{t,L}^{(lj)}$ for $l = 0, \ldots, L-1$ and $j = 0, \ldots, L-1$. The average PEP reduces in this case to

$$P(\underline{\mathbf{C}} \rightarrow \underline{\mathbf{E}}) = \frac{1}{\pi} \int_0^{\pi/2} \left[\det\left(\mathbf{I}_T + \eta \sum_{j=0}^{L-1}\sum_{l=0}^{L-1} (\mathbf{C} - \mathbf{E})_{(l)}^H \mathbf{R}_{t,L}^{(lj)}(\mathbf{C} - \mathbf{E})_{(j)} \right) \right]^{-n_r} d\beta. \tag{9.118}$$

We now introduce the notion of degenerate taps.

Definition 9.2 *The lth tap is said to be degenerate in the direction of departure $\theta_{t,l}$ if all scatterers seen by the transmitter whose time of arrival falls within in the lth tap are located along the same direction $\theta_{t,l}$.*

We further assume that the correlation between taps is zero and define $\mathcal{V}_{deg}$ as the set of taps that are spatially correlated, all other taps being spatially uncorrelated. Then

$$\sum_{j=0}^{L-1}\sum_{l=0}^{L-1}(\mathbf{C}-\mathbf{E})^H_{(l)}\mathbf{R}^{(lj)}_{t,L}(\mathbf{C}-\mathbf{E})_{(j)} \qquad (9.119)$$

$$= \underbrace{(\underline{\mathbf{C}}-\underline{\mathbf{E}})^H_{(l\in\mathcal{V}_{deg})}\mathbf{R}_{deg}(\underline{\mathbf{C}}-\underline{\mathbf{E}})_{(l\in\mathcal{V}_{deg})}}_{\mathbf{C}_{\mathbf{R}_{deg}}} + \underbrace{\sum_{l\notin\mathcal{V}_{deg}}(\mathbf{C}-\mathbf{E})^H_{(l)}(\mathbf{C}-\mathbf{E})_{(j)}}_{\mathbf{C}_{\mathbf{R}_{i.i.d.}}}$$

where $(\underline{\mathbf{C}}-\underline{\mathbf{E}})_{(l\in\mathcal{V}_{deg})}$ is obtained by stacking up the blocks $(\mathbf{C}-\mathbf{E})_{(l)}$ with $l\in\mathcal{V}_{deg}$

$$(\underline{\mathbf{C}}-\underline{\mathbf{E}})_{(l\in\mathcal{V}_{deg})} = \begin{bmatrix} \vdots \\ (\mathbf{C}-\mathbf{E})_{(l)} \\ \vdots \end{bmatrix}_{l\in\mathcal{V}_{deg}} \qquad (9.120)$$

and $\mathbf{R}_{deg}$ is made of the blocks $\mathbf{R}^{(lj)}_{t,L}$ with $l,j\in\mathcal{V}_{deg}$.

As the transmit angle-spreads $\Lambda^{(l)}_t$ of the taps $\in\mathcal{V}_{deg}$ reduce each to zero. The transmit direction power spectrum of each tap tends to a delta function around $\theta_{t,l}$, and we easily derive that

$$\lim_{\{\Lambda^{(l)}_t\}_{l\in\mathcal{V}_{deg}}\to 0}\mathbf{C}_{\mathbf{R}_{deg}} = \sum_{l\in\mathcal{V}_{deg}}\sigma^2_l\underbrace{(\mathbf{C}-\mathbf{E})^H_{(l)}\mathbf{a}^*_t(\theta_{t,l})\mathbf{a}^T_t(\theta_{t,l})(\mathbf{C}-\mathbf{E})_{(l)}}_{\mathbf{C}_{\mathbf{R}_l}} \qquad (9.121)$$

and $\sharp\mathcal{V}_{deg}\big(r(\mathbf{C}_{\mathbf{R}_l})-1\big)$ eigenvalues of $\mathbf{C}_{\mathbf{R}_{deg}}$ become equal to zero.

Depending on the last eigenvalues $\lambda(\mathbf{C}_{\mathbf{R}_l}) = \big\|(\mathbf{C}-\mathbf{E})^T_{(l)}\mathbf{a}_t(\theta_{t,l})\big\|^2$, $\mathbf{C}_{\mathbf{R}_l}$ is of rank zero or one and the effective number of virtual antennas corresponding to each degenerate tap is reduced by n_t or n_t-1, respectively. Since the code cannot exploit the transmit diversity offered by the set $\mathcal{V}_{deg}$ of highly transmit-correlated taps, it may only exploit the additional frequency selectivity offered by those taps. The contribution of any degenerate tap $l\in\mathcal{V}_{deg}$ to the total performance is directly related to $\big\|(\mathbf{C}-\mathbf{E})^T_{(l)}\mathbf{a}_t(\theta_{t,l})\big\|^2$. From eqn (9.118), this term can be viewed as a coding gain, since the larger this quantity, the smaller the SNR required to observe the frequency diversity of this tap. Physically, this indicates that taps with small angle-spreads degenerate into SIMO channels where the $1\times T$ transmitted codeword seen by the lth tap $\in\mathcal{V}_{deg}$ is given by $\mathbf{a}^T_t(\theta_{t,l})\mathbf{C}_{(l)}$.

Note that for both single-carrier and MIMO-OFDM transmissions

$$\big\|(\mathbf{C}-\mathbf{E})^T_{(l)}\mathbf{a}_t(\theta_{t,l})\big\|^2 = \big\|(\mathbf{C}-\mathbf{E})^T\mathbf{a}_t(\theta_{t,l})\big\|^2. \qquad (9.122)$$

This implies that many derivations and concepts valid in flat fading channels (see Chapter 7) are also applicable to the frequency-selective case. For example, the G_{sum} concept suggests

that the Lindskog-Paulraj scheme and O-SFBC are the most robust schemes for single-carrier and MIMO-OFDM transmissions, respectively. It also suggests that CDD may be highly sensitive to transmit correlations. Indeed from eqn (9.102), the distance G_{sum} reads as

$$\left\| (\mathbf{C} - \mathbf{E})^T \mathbf{a}_t(\theta_{t,l}) \right\|^2 = \frac{1}{n_t} \sum_{k=0}^{T-1} |c_k - e_k|^2 \left| \left[e^{-j\frac{2\pi}{T}k\Delta_1} \; \ldots \; e^{-j\frac{2\pi}{T}k\Delta_{n_t}} \right] \mathbf{a}_t(\theta_{t,l}) \right|^2. \quad (9.123)$$

Typically, if the delay is chosen such that $\Delta_m = (m-1)\Delta$ and a linear array is used, G_{sum} may be low for some directions of departure. From Chapter 7, we know that the use of a precoder designed to maximize G_{sum} may be a low complexity solution to substantially decrease the error rate in correlated channels.

9.6.2 Application to space-frequency MIMO-OFDM

For space-frequency coding, eqn (9.119) can be reformulated similarly to eqn (9.82) as

$$\mathbf{C}_{\mathbf{R}_{deg}} + \mathbf{C}_{\mathbf{R}_{i.i.d.}}$$

$$= \sum_{l \in \mathcal{V}_{deg}} \mathbf{R}_{f,l} \odot \left[(\mathbf{C} - \mathbf{E})^H \mathbf{a}_t^*(\theta_{t,l}) \mathbf{a}_t^T(\theta_{t,l})(\mathbf{C} - \mathbf{E}) \right] + \mathbf{R}_{f,i.i.d.} \odot \tilde{\mathbf{E}}^H \quad (9.124)$$

where $\mathbf{R}_{f,i.i.d.} = \sum_{l \notin \mathcal{V}_{deg}} \beta_l \mathbf{d}_{(l)}^H \mathbf{d}_{(l)}$ denotes the frequency correlation matrix induced by all spatially uncorrelated taps, while $\mathbf{R}_{f,l}$ denotes the frequency correlation matrix corresponding to the lth degenerate tap. We have mentioned that in uncorrelated Rayleigh channels, the behavior of a space-frequency code in a highly frequency-selective channel is somehow similar to the behavior of a space–time code in fast fading channels. This implies that the performance is strongly related to the effective length and the product distance of the code. However, in correlated channels, the previous discussion illustrates that the behavior of space-frequency coding with degenerate taps differs from the behavior of space–time coding over correlated fast fading channels with high transmit correlation. Indeed, a space-frequency code is affected by a degenerate tap in the same way that a space–time code is affected by a degenerate slow fading channel, so that it is G_{sum} that dictates the performance. The product distance $G_{product}$ is only relevant in the highly improbable situation where all taps are degenerate and aligned in the same direction.

Quantifying the exact diversity achieved by a space-frequency code in the presence of degenerate taps is not an easy task. However, we can very simply quantify an upper bound of the diversity gain.

Proposition 9.5 *On a L-tap MIMO channel with $\sharp \mathcal{V}_{deg}$ degenerate taps, the diversity gain achievable by a pair of space-frequency codewords $\{\mathbf{C}, \mathbf{E}\}$ with an effective length*

$l_{\mathbf{C},\mathbf{E}}$, *effective lengths* $\{l_{\mathbf{C},\mathbf{E}}(l)\}_{l=1}^{n_t}$ *and a rank* $r(\tilde{\mathbf{E}})$ *is upper-bounded by*

$$\min\left\{l_{\mathbf{C},\mathbf{E}}, r(\tilde{\mathbf{E}})(L - \sharp\mathcal{V}_{deg}) + \sum_{l\in\mathcal{V}_{deg}}\left\lceil\left\|(\mathbf{C} - \mathbf{E})^T\mathbf{a}_t(\theta_l)\right\|^2\right\rceil\right\}n_r \qquad (9.125)$$

where $\lceil x\rceil = \begin{cases} 1 \text{ if } x > 0 \\ 0 \text{ if } x = 0 \end{cases}$.

PROOF: It follows from a development similar to the proof of Proposition 9.3.   ■
As a design methodology, it is thus advisable to use space-frequency codes with large
effective lengths and minimum product distance, but also with large values of G_{sum} and
$\sharp\tau_{\deg.}(\theta_{t,l})$ for all possible $\theta_{t,l}$.

It is important to note that even if degenerate taps do not exist in practice, taps may appear
as degenerate if the transmit correlation is sufficiently high and/or the SNR is sufficiently
low so that only the largest eigenvalue of $\mathbf{C}_{\mathbf{R}_{deg}}$ contributes non-negligibly to the average
PEP. The derivations and interpretations are similar to those outlined for the flat fading
case in Chapters 6 and 7.

Appendix A
Useful mathematical and matrix properties

Some mathematical and matrix properties [HJ95] used throughout this book are summarized here.

(1) $\|\mathbf{A}\|_F^2 = \text{Tr}\{\mathbf{A}\mathbf{A}^H\} = \text{Tr}\{\mathbf{A}^H\mathbf{A}\}$

(2) $\text{Tr}\{\mathbf{A}\mathbf{B}\} = \text{Tr}\{\mathbf{B}\mathbf{A}\}$

(3) $\det(\mathbf{I} + \mathbf{A}\mathbf{B}) = \det(\mathbf{I} + \mathbf{B}\mathbf{A})$

(4) *Cauchy-Schwartz's inequality*: $|\text{Tr}\{\mathbf{A}\mathbf{B}\}|^2 \leq \|\mathbf{A}\|_F^2 \|\mathbf{B}\|_F^2$

(5) *Hadamard's inequality*: $\det(\mathbf{A}) \leq \prod_{k=1}^{n} \mathbf{A}(k,k)$ if $\mathbf{A} > 0$ of size $n \times n$

(6) *Fischer's inequality*: $\det\left(\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^H & \mathbf{C} \end{bmatrix}\right) \leq \det(\mathbf{A})\det(\mathbf{C})$ if $\mathbf{A}, \mathbf{C}$ square; equality occurs if $\mathbf{B} = \mathbf{0}$

(7) *Weyl's inequality*: for $\mathbf{A}$ and $\mathbf{B}$ $n \times n$ hermitian matrices whose eigenvalues $\{\lambda_k(\mathbf{A})\}$ and $\{\lambda_k(\mathbf{B})\}$ are sorted by decreasing order of magnitude

$$\lambda_k(\mathbf{A}) + \lambda_n(\mathbf{B}) \leq \lambda_k(\mathbf{A} + \mathbf{B}) \leq \lambda_k(\mathbf{A}) + \lambda_1(\mathbf{B}),$$

for $k = 1, \ldots, n$

(8) $\|\mathbf{A} - \mathbf{B}\mathbf{X}\|_F^2 \geq \|(\mathbf{I} - \mathbf{B}(\mathbf{B}^H\mathbf{B})^{-1}\mathbf{B}^H)\mathbf{A}\|_F^2$ for $\mathbf{B}$ full column rank

(9) $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$

(10) $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$

(11) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D})$

(12) $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ if $\mathbf{A}, \mathbf{B}$ square and non singular

(13) $\det(\mathbf{A}_{m \times m} \otimes \mathbf{B}_{n \times n}) = \det(\mathbf{A})^n \det(\mathbf{B})^m$

(14) $\text{Tr}\{\mathbf{A} \otimes \mathbf{B}\} = \text{Tr}\{\mathbf{A}\}\text{Tr}\{\mathbf{B}\}$

(15) $\text{Tr}\{\mathbf{A}\mathbf{B}\} \geq \text{Tr}\{\mathbf{A}\}\sigma_{min}^2(\mathbf{B})$ with $\sigma_{min}(\mathbf{B})$ the smallest singular value of $\mathbf{B}$

(16) $\text{vec}(\mathbf{A}\mathbf{B}\mathbf{C}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$

(17) $\det(\mathbf{I} + \epsilon\mathbf{A}) = 1 + \epsilon\text{Tr}\{\mathbf{A}\}$ if $\epsilon \ll 1$

(18) *matrix inversion lemma*: $(\mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{D})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{C}(\mathbf{B} - \mathbf{D}\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{D}\mathbf{A}^{-1}$

(19) *inclusion principle*: if $\mathbf{A}$ is a $n \times n$ hermitian matrix and $\mathbf{A}'$ is a $r \times r$ principal submatrix of $\mathbf{A}$ (obtained by deleting $n - r$ rows and the corresponding columns

403

from **A**), then for any k such that $1 \le k \le r$

$$\lambda_k(\mathbf{A}) \le \lambda_k(\mathbf{A}') \le \lambda_{k+n-r}(\mathbf{A})$$

where the eigenvalues are denoted by increasing order

(20) for $\mathbf{A}, \mathbf{B}, \mathbf{Z} \in \mathbb{C}^{n \times m}$

$$\frac{\partial \mathrm{Tr}\{\mathbf{AZ}\}}{\partial \mathbf{Z}} = \mathbf{A}^T$$

$$\frac{\partial \mathrm{Tr}\left\{\mathbf{Z}^H \mathbf{A}\right\}}{\partial \mathbf{Z}} = \mathbf{0}_{n \times m}$$

$$\frac{\partial \mathrm{Tr}\left\{\mathbf{ZAZ}^H \mathbf{B}\right\}}{\partial \mathbf{Z}} = \mathbf{B}^T \mathbf{Z}^* \mathbf{A}^T$$

$$\frac{\partial \mathrm{Tr}\left\{\mathbf{AZ}^{-1}\right\}}{\partial \mathbf{Z}} = -\left(\mathbf{Z}^T\right)^{-1} \mathbf{A}^T \left(\mathbf{Z}^T\right)^{-1}$$

$$\frac{\partial \det(\mathbf{AZB})}{\partial \mathbf{Z}} = \det(\mathbf{AZB}) \mathbf{A}^T \left(\mathbf{B}^T \mathbf{Z}^T \mathbf{A}^T\right)^{-1} \mathbf{B}^T$$

(the interested reader is referred to [HGonc] for more properties)

(21) *Jensen's inequality*: $\mathcal{E}_x\{\mathcal{F}(x)\} \le \mathcal{F}(\mathcal{E}_x\{x\})$ if $\mathcal{F}$ concave

(22) *Chernoff bound*: $\mathcal{Q}(x) \le \exp\left(-x^2/2\right)$

Appendix B

Complex Gaussian random variables and matrices

B.1 Some useful probability distributions

Consider that h is a complex Gaussian variable, i.e. a circularly complex variable with zero-mean and variance σ^2. This means that both the real and imaginary parts of h are zero-mean Gaussian variables of variance σ^2. In this case, $s \triangleq |h|$ follows a Rayleigh distribution

$$p_s(s) = \frac{s}{\sigma^2} \exp\left(-\frac{s^2}{2\sigma^2}\right), \tag{B.1}$$

with the following properties:

$$\mathcal{E}\{s\} = \sigma\sqrt{\frac{\pi}{2}}, \tag{B.2}$$

$$\mathcal{E}\{s^2\} = 2\sigma^2. \tag{B.3}$$

The CDF of s is given by

$$P[s < S] = 1 - e^{-S^2/2\sigma^2}. \tag{B.4}$$

The random variable $y \triangleq |h|^2 = s^2$ follows a χ^2 distribution (with two degrees of freedom)

$$p_y(y) = \frac{1}{2\sigma^2} \exp\left(-\frac{y}{2\sigma^2}\right). \tag{B.5}$$

Note that for small ε, we have that $P[y < \varepsilon] \approx \varepsilon$.

Let us now consider n i.i.d. zero-mean complex Gaussian variables $h_1, \ldots, h_n$ with variance σ^2. Defining $u = \sum_{k=1}^{n} |h_k|^2$, the moment generating function of u is given by

$$\mathcal{M}_u(\tau) = \mathcal{E}\{e^{\tau u}\} = \prod_{k=1}^{n} \frac{1}{1 - \tau} = \left[\frac{1}{1 - \tau}\right]^n, \tag{B.6}$$

and the distribution of u reads as

$$p_u(u) = \frac{1}{\sigma^{2n} 2^n \Gamma(n)} u^{n-1} e^{-u} \exp\left(-\frac{u}{2\sigma^2}\right). \tag{B.7}$$

This distribution is known as the χ^2 distribution with $2n$ degrees of freedom (the case $n=1$ reduces to (B.5)). The corresponding CDF is given by

$$P[u < U] = 1 - e^{-U/2\sigma^2} \sum_{k=1}^{n-1} \frac{1}{k!} \left(\frac{U}{2\sigma^2}\right)^k, \tag{B.8}$$

$$= 1 - \frac{\Gamma(n, U/2\sigma^2)}{(n-1)!}, \tag{B.9}$$

$$= \frac{1}{n!} \left(\frac{U}{2\sigma^2}\right)^n \left[1 + o\left(\frac{U}{2\sigma^2}\right)\right]. \tag{B.10}$$

Finally, let us now analyze the case when $h_1, \ldots, h_n$ are non-zero mean. Assume that the real and imaginary parts of h_k are Gaussian variables of mean μ_k and variance σ^2. In this situation, $w = \sum_{k=1}^n |h_k|^2$ follows a non-central χ^2 distribution with $2n$ degrees of freedom

$$p_w(w) = \frac{1}{2\sigma^2} \left(\frac{w}{\beta}\right)^{\frac{n-1}{2}} \exp\left(-\frac{w+\beta}{2\sigma^2}\right) I_{n-1}\left(\frac{\sqrt{w\beta}}{\sigma^2}\right), \tag{B.11}$$

where $\beta = \sum_{k=1}^n 2\mu_k^2$ is called the non-centrality parameter of the distribution.

B.2   Determinant and product eigenvalues of Wishart matrices

In this section, we formalize a few properties of complex Wishart matrices. Indeed, for i.i.d. Rayleigh channels represented by a $n_r \times n_t$ complex Gaussian matrix $\mathbf{H}_w$, $\mathbf{W} = \mathbf{H}_w \mathbf{H}_w^H$ (for $n_t > n_r$) and $\mathbf{W} = \mathbf{H}_w^H \mathbf{H}_w$ (for $n_t < n_r$) are central Wishart matrices with parameters n and N, which we denote as $\mathbf{W}(n, N)$. In our notations, $N = \max\{n_t, n_r\}$ and $n = \min\{n_t, n_r\}$.

Lemma B.1   *The joint probability density [Ede89] of the elements of a matrix $\mathbf{W}(n, N)$ is*

$$\frac{1}{2^{Nn} \tilde{\Gamma}_n(N)} e^{\text{Tr}\{-\frac{1}{2}\mathbf{W}\}} [\det(\mathbf{W})]^{N-n}, \tag{B.12}$$

where $\tilde{\Gamma}_n(N) = \pi^{n(n-1)/2} \prod_{k=1}^n \Gamma(N - k + 1)$ is the multivariate gamma function.

B.2.1   Determinant and product of eigenvalues of Wishart matrices

Using a bi-diagonalization of $\mathbf{H}_w$, it is shown in [Ede89] that Wishart matrices have the following property.

Theorem B.1. *If* $\mathbf{W}$ *has the central Wishart distribution with parameters n and N, its determinant has the distribution* $\prod_{k=1}^{n} \chi^2_{2(N-n+k)}$ *(the notation refers to the distribution of the product of n random variables with the indicated* χ^2 *distribution). As the determinant is also the product of the non-zero eigenvalues, this product has the same distribution.*

PROOF: The proof relies on the fact that a complex Gaussian matrix with independent entries is unitarily similar to a bi-diagonal matrix whose entries are independent χ^2 variables [Ede89]. ∎

Corollary B.1 *The expected value of the determinant of* $\mathbf{W}$ *is equal to* $\frac{N!}{(N-n)!}$.

It is important to note that the above theorem does not imply that ordered or randomly selected eigenvalues of $\mathbf{W}$ are χ^2 distributed, as the theorem only concerns their product.

B.2.2 Distribution of ordered eigenvalues

The ordered eigenvalues of $\mathbf{W}$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, have the following distribution:

$$p_{\lambda_1,\ldots,\lambda_n}(\lambda_1,\ldots,\lambda_n) = \frac{2^{-Nn}\pi^{n(n-1)}}{\tilde{\Gamma}_n(N)\tilde{\Gamma}_n(n)} e^{-\sum_{k=1}^{n}\lambda_k} \prod_{k=1}^{n} \lambda_k^{N-n} \prod_{k<l} (\lambda_k - \lambda_l)^2. \qquad (\text{B.13})$$

Asymptotic results regarding the smallest and the largest eigenvalues of $\mathbf{W}$ are derived in [Ede89] and summarized as follows.

Proposition B.1 *If n and N tend to infinity in a constant ratio n/N ≤ 1, the smallest and the largest eigenvalues of* $\mathbf{W}$ *respectively satisfy*

$$\frac{1}{n}\lambda_{min} \overset{a.s.}{\to} 2\left(1 - \sqrt{n/N}\right)^2 \qquad (\text{B.14})$$

$$\frac{1}{N}\lambda_{max} \overset{a.s.}{\to} \left(1 + \sqrt{n/N}\right)^2. \qquad (\text{B.15})$$

This proposition is particularly helpful as it also provides directly the mean values of λ_{min} and λ_{max}.

B.2.3 Distribution of non-ordered eigenvalues

Theorem B.2. *The distribution* $p_\lambda(\lambda)$ *of a randomly selected (non-ordered) eigenvalue of* $\mathbf{W}$ *is*

$$p_\lambda(\lambda) = \frac{1}{n} \sum_{k=0}^{n-1} \frac{k!\lambda^{N-n}e^{-\lambda}}{(k+N-n)!} L_k^{N-n}(\lambda), \qquad (\text{B.16})$$
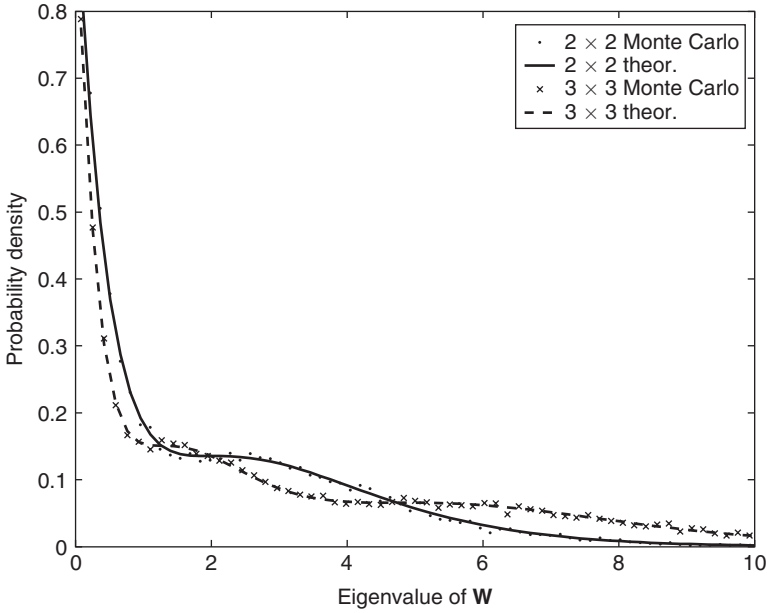
Figure B.1   *Distribution of a randomly selected eigenvalue of* **W** *in several scenarios*

and $L_k^p(x)$ is the Laguerre polynomial of order k

$$L_k^p(x) = \sum_{m=0}^{k} (-1)^m \binom{k+p}{k-m} \frac{x^m}{m!}. \tag{B.17}$$

Interestingly, all non-ordered eigenvalues are independent and have the same distribution. As an example, Figure B.1 illustrates the theoretical distribution together with the distribution obtained from Monte Carlo simulations for 2×2 and 3×3 i.i.d. Rayleigh channels.

Appendix C
Stanford University Interim channel models

This appendix describes the six SUI tap-delay line models, which are also used by the IEEE 802.16 channel models. Models 1 and 2 correspond to a mostly flat terrain with light tree density; models 5 and 6 correspond to a hilly terrain with moderate to heavy tree density; models 3 and 4 are used for intermediate propagation conditions.

Table C.1 *SUI-1 channel parameters (omnidirectional antennas)*

	Tap 1	Tap 2	Tap 3
Delay [μs]	0.0	0.4	0.9
Power [dB]	0	−15	−20
90% K-factor	4	0	0
Doppler v_m [Hz]	0.40	0.30	0.50
Envelope antenna correlation	0.7	0.7	0.7

Table C.2 *SUI-2 channel parameters (omnidirectional antennas)*

	Tap 1	Tap 2	Tap 3
Delay [μs]	0.0	0.4	1.1
Power [dB]	0	−12	−15
90% K-factor	2	0	0
Doppler v_m [Hz]	0.20	0.15	0.25
Envelope antenna correlation	0.5	0.5	0.5

Table C.3   *SUI-3 channel parameters (omnidirectional antennas)*

	Tap 1	Tap 2	Tap 3
Delay [μs]	0.0	0.4	0.9
Power [dB]	0	-5	-10
90% K-factor	1	0	0
Doppler v_m [Hz]	0.40	0.30	0.50
Envelope antenna correlation	0.4	0.4	0.4

Table C.4   *SUI-4 channel parameters (omnidirectional antennas)*

	Tap 1	Tap 2	Tap 3
Delay [μs]	0.0	1.5	4.0
Power [dB]	0	-4	-8
90% K-factor	0	0	0
Doppler v_m [Hz]	0.20	0.15	0.25
Envelope antenna correlation	0.3	0.3	0.3

Table C.5   *SUI-5 channel parameters (omnidirectional antennas)*

	Tap 1	Tap 2	Tap 3
Delay [μs]	0.0	4.0	10.0
Power [dB]	0	-5	-10
90% K-factor	0	0	0
Doppler v_m [Hz]	2.00	1.50	2.50
Envelope antenna correlation	0.3	0.3	0.3

Table C.6   *SUI-6 channel parameters (omnidirectional antennas)*

	Tap 1	Tap 2	Tap 3
Delay [μs]	0.0	14.0	20.0
Power [dB]	0	-10	-14
90% K-factor	0	0	0
Doppler v_m [Hz]	0.40	0.30	0.50
Envelope antenna correlation	0.3	0.3	0.3

Appendix D
Antenna coupling model

In this appendix, the antenna coupling model presented in Section 2.6.2 is rigorously derived. We explain why the coupling matrix may only be derived if the antennas are minimum scatterers. We distinguish the cases of antennas being minimum scatterers with regard to impedance parameters, and minimum scatterers with regard to admittance parameters.

D.1 Minimum scatterers with regard to impedance parameters

As explained in [RP94], antennas that are minimum scatterers with regard to impedance parameters are such that an open-circuited antenna behaves almost exactly as if it were not present at all. A typical example of such an antenna is the half-wave length dipole. Indeed, open-circuited half-wave length dipole elements become quarter-wave length wires, which have very small scattering cross sections [Han98].

The use of such antennas implies the following simplifications:

(1) for receive antennas, the open circuit voltages do not depend on the presence of other elements.
(2) the self impedances do not depend on the presence of other open-circuited antennas.
(3) the element pattern in an open-circuited environment is identical to the isolated element pattern.

D.1.1 Circuit representation

Consider two coupled antennas modeled as a two-port network. Each antenna is excited by a source represented by a Thevenin equivalent source with two series impedances Z_1 and Z_2. Denoting the currents at the antenna ports as I_1 and I_2 (see Figure D.1), we express [GK83] the voltages on the terminations, V_1 and V_2, as

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = - \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} + \begin{bmatrix} V_{01} \\ V_{02} \end{bmatrix} \tag{D.1}$$
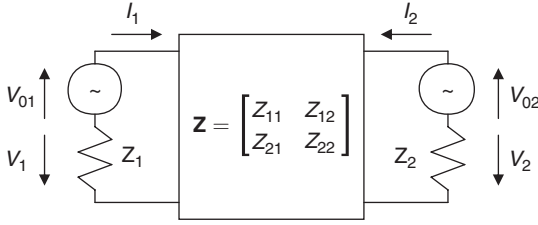
Figure D.1 *Equivalent circuit of the two coupled antennas*

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \qquad (D.2)$$

where V_{01} and V_{02} are the source voltages. Under simplification (1), V_{01} and V_{02} are constant and do not depend on I_1 and I_2. The elements Z_{ij} are the elements of the network impedance matrix **Z**. Under simplification (2), $Z_{11} = Z_{A1}$ and $Z_{22} = Z_{A2}$ with Z_{A1} and Z_{A2} being the antenna impedances of isolated antennas 1 and 2, respectively.

It is worth noting that this representation is valid for transmit and receive arrays. Indeed, the source voltages respectively correspond to the generator and the open circuit voltages while the termination impedances correspond to the generator and the load impedances.

At the receive side, we solve the previous set of equations in order to express the voltages on the loads as a function of the open circuit voltages

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \left[\mathbf{I}_2 + \mathbf{Z} \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}^{-1} \right]^{-1} \begin{bmatrix} V_{01} \\ V_{02} \end{bmatrix}. \qquad (D.3)$$

Assuming that $Z_1 = Z_2 = Z_T$, we can rewrite eqn (D.3) equivalently as

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \frac{Z_T}{Z_T + Z_A} \underbrace{(Z_T + Z_A)\left[\mathbf{I}_2 Z_T + \mathbf{Z}\right]^{-1}}_{\mathbf{M}_r} \begin{bmatrix} V_{01} \\ V_{02} \end{bmatrix} \qquad (D.4)$$

where we define the receive coupling matrix $\mathbf{M}_r$. For large inter-element spacings, mutual coupling is negligible and $\mathbf{M}_r$ becomes an identity matrix.

In order to derive the transmit coupling matrix, we first express the currents on each antenna as a function of the source voltages (i.e. the open circuit voltages) and then use the superposition principle in order to estimate, thanks to simplification (2) and (3), the voltages effectively induced on each antenna by its current. From (D.1) and (D.2), we have

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \left[\begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} + \mathbf{Z} \right]^{-1} \begin{bmatrix} V_{01} \\ V_{02} \end{bmatrix}. \qquad (D.5)$$

In order to estimate the radiated field, we may consider those currents as perfect current sources and superpose the fields radiated by each current. When we focus on one of those

current sources, the other currents are set to zero (the antennas are open-circuited). Since, for minimum scatterers with respect to impedance parameters, an open-circuited antenna is invisible from assumption (3), we focus on each antenna excited by its own current and write the voltages on the antennas V_{A1} and V_{A2} as

$$\begin{bmatrix} V_{A1} \\ V_{A2} \end{bmatrix} = \begin{bmatrix} Z_{A1} & 0 \\ 0 & Z_{A2} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}. \tag{D.6}$$

Assuming $Z_1 = Z_2 = Z_T$ and $Z_{A1} = Z_{A2} = Z_A$, combining eqns (D.6) and (D.5) leads to

$$\begin{bmatrix} V_{A1} \\ V_{A2} \end{bmatrix} = \frac{Z_A}{Z_T + Z_A} \underbrace{(Z_T + Z_A) \left[\mathbf{I}_2 Z_T + \mathbf{Z} \right]^{-1}}_{\mathbf{M}_t} \begin{bmatrix} V_{01} \\ V_{02} \end{bmatrix} \tag{D.7}$$

where we define the transmit coupling matrix $\mathbf{M}_t$. Comparing with eqn (D.4), we observe that the coupling matrices have the same form at the transmitter and the receiver.

D.1.2 Radiation patterns

We may also consider mutual coupling in terms of radiation patterns. The array pattern is evaluated as follows [KS93]

$$E(\theta, \psi) = \sum_{p=1}^{N} V_{0p} g_p (\theta, \psi) e^{jk \, \hat{\mathbf{r}} \bullet_a \, \mathbf{r}_p} \tag{D.8}$$

where $g_p(\theta, \psi)$ is the phase-adjusted unit-excitation active element pattern, k is the free-space propagation constant, $\hat{\mathbf{r}}$ is the unit radial vector from the coordinate origin in the observation direction (θ, ψ), $\mathbf{r}_p$ is a position vector from the origin to the center of the pth element, and $\bullet_a$ designates the scalar product in the array coordinate system. The pattern $g_p(\theta, \psi)$ represents the pattern radiated by the array when the pth element is excited by a unit voltage with its associated generator impedance Z_p, and the other elements are loaded by their respective generator impedance $\{Z_p\}$. Such a pattern is expressed in the coordinate system of antenna p. The term $e^{jk \, \hat{\mathbf{r}} \bullet_a \, \mathbf{r}_p}$ translates the origin of the coordinate system of antenna p such that $g_p(\theta, \psi) e^{jk \, \hat{\mathbf{r}} \bullet_a \, \mathbf{r}_p}$ expresses the same pattern but in the array coordinate system (this is called the unit-excitation active element pattern). By assumption (3), the element pattern $g_p(\theta, \psi)$ can be evaluated as the sum over the array elements with the currents as coefficients [Han98], all multiplied by the isolated element pattern

$$g_p(\theta, \psi) = g_{isol}(\theta, \psi) \sum_{q=1}^{N} I_q^{(p)} e^{jk \, \hat{\mathbf{r}} \bullet_p \, \mathbf{r}_q} \tag{D.9}$$

where $\bullet_p$ designates the scalar product in the pth antenna coordinate system, and $I_q^{(p)}$ is the feeding current evaluated on antenna q when antenna p is excited by a unit voltage with its associated generator impedance Z_p, while the other elements are loaded by their

respective generator impedance $\{Z_q\}$. The currents $I_q^{(p)}$ with $q \neq p$ originate from the antenna coupling. $g_{isol}(\theta, \psi)$ is the radiation pattern of an isolated array element. Those currents can be evaluated based on the aforementioned equivalent circuit representation of the coupled antenna network. Let us deal with two antennas. We look at $I_q^{(p)}$. Denoting by the superscript$^{(p)}$ that a unit voltage is applied to antenna p and the other antennas are not excited, we can express

$$\begin{bmatrix} V_1^{(p)} \\ V_2^{(p)} \end{bmatrix} = -\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}\begin{bmatrix} I_1^{(p)} \\ I_2^{(p)} \end{bmatrix} + \begin{bmatrix} V_{01}^{(p)} \\ V_{02}^{(p)} \end{bmatrix} \tag{D.10}$$

$$\begin{bmatrix} V_1^{(p)} \\ V_2^{(p)} \end{bmatrix} = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}\begin{bmatrix} I_1^{(p)} \\ I_2^{(p)} \end{bmatrix} \tag{D.11}$$

with

$$V_{0q}^{(p)} = \begin{cases} 0 & \text{if } p \neq q \\ V_{0p} = 1 & \text{if } p = q \end{cases} \tag{D.12}$$

and we can finally express the currents $I_q^{(p)}$ as

$$\begin{bmatrix} I_1^{(p)} \\ I_2^{(p)} \end{bmatrix} = \left[\begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} + \mathbf{Z}\right]^{-1}\begin{bmatrix} V_{01}^{(p)} \\ V_{02}^{(p)} \end{bmatrix}. \tag{D.13}$$

Focusing only on θ plane (see Figure D.1), the phase-adjusted unit-excitation active element patterns in eqn (D.9) g_1 and g_2 can now be evaluated by

$$g_1(\theta, \psi) = g_{isol}(\theta, \psi)\left[I_1^{(1)} + I_2^{(1)}e^{-jkd\cos(\theta)}\right] \tag{D.14}$$

$$g_2(\theta, \psi) = g_{isol}(\theta, \psi)\left[I_1^{(2)}e^{jkd\cos(\theta)} + I_2^{(2)}\right]. \tag{D.15}$$

Using the same array coordinate system as antenna 1, the array radiation pattern $E(\theta, \psi)$ reads as

$$\begin{aligned} E(\theta, \psi) &= V_{01}\, g_1(\theta, \psi) + V_{02}\, g_2(\theta, \psi)e^{-jkd\cos(\theta)} \\ &= g_{isol}(\theta, \psi)\begin{bmatrix} 1 & e^{-jkd\cos(\theta)} \end{bmatrix}\begin{bmatrix} I_1^{(1)} & I_1^{(2)} \\ I_2^{(1)} & I_2^{(2)} \end{bmatrix}\begin{bmatrix} V_{01} \\ V_{02} \end{bmatrix} \\ &= g_{isol}(\theta, \psi)\begin{bmatrix} 1 & e^{-jkd\cos(\theta)} \end{bmatrix}\left[\begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} + \mathbf{Z}\right]^{-1}\begin{bmatrix} V_{01} \\ V_{02} \end{bmatrix}. \end{aligned} \tag{D.16}$$

If $Z_1 = Z_2 = Z_T$, then all mutual coupling effects on array pattern are taken into account in the matrix $[Z_T\mathbf{I}_2 + \mathbf{Z}]^{-1}$. For simplicity, we normalize this matrix in such a way that in the absence of coupling, the coupling matrix becomes identity, so

$$\mathbf{M} = (Z_T + Z_A)\,[Z_T\mathbf{I}_2 + \mathbf{Z}]^{-1}. \tag{D.17}$$

Thanks to the introduction of the coupling matrices, the evaluation of radiation patterns in the presence of coupling is straightforward, based on (D.14), (D.15) and (D.16).

D.2   Minimum scatterers with regard to admittance parameters

Antennas that are minimum scatterers with regard to admittance parameters are such that a short-circuited antenna behaves almost exactly as if the antenna were not present at all [RP94]. A typical example of such an antenna is the slot type antenna.

The use of such antennas implies the following simplifications:

(1) the short circuit currents do not depend on the presence of other elements.
(2) the self admittances do not depend on the presence of other short-circuited antennas.
(3) the element pattern in a short-circuited environment is identical to the isolated element pattern.

Consider two coupled antennas, modeled as a two-port network. Each antenna is excited by a source modeled as a Norton equivalent current source with two parallel termination impedances Z_1 and Z_2, respectively. Denoting the voltages on the terminations as V_1 and V_2 (and at the entrance of the antennas also), we express the currents in the terminations, I_1 and I_2, in the following ways

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = - \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + \begin{bmatrix} I_{01} \\ I_{02} \end{bmatrix} \tag{D.18}$$

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \tag{D.19}$$

with I_{01} and I_{02} the source currents (if all antennas are short-circuited, $V_1 = V_2 = 0$ and I_1, I_2 are equal to I_{01} and I_{02}, respectively). Under simplification (1), I_{01} and I_{02} are constant and do not depend on V_1 and V_2. The elements Y_{ij} are the elements of the network admittance matrix $\mathbf{Y} = \mathbf{Z}^{-1}$. Under simplification (2), $Y_{11} = Y_{A1}$ and $Y_{22} = Y_{A2}$ with Y_{A1} and Y_{A2} the antenna admittance of antenna 1 and 2, respectively. Again, this representation is valid for transmit and receive arrays.

At the *receive side*, we can solve the previous set of equations in order to express the currents in the loads as a function of the short circuit currents

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \left[\mathbf{I}_2 + \mathbf{Z}^{-1} \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \right]^{-1} \begin{bmatrix} I_{01} \\ I_{02} \end{bmatrix}. \tag{D.20}$$

Assuming $Z_1 = Z_2 = Z_T$, we can rewrite eqn (D.20) equivalently as

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \frac{Z_A}{Z_T + Z_A} \underbrace{\frac{(Z_T + Z_A)}{Z_A} \mathbf{Z} \left[\mathbf{I}_2 Z_T + \mathbf{Z} \right]^{-1}}_{\mathbf{M}_r} \begin{bmatrix} I_{01} \\ I_{02} \end{bmatrix} \tag{D.21}$$

where we introduce the coupling matrix $\mathbf{M}_r$ at the receive side. All mutual coupling effects are taken into account in this matrix. For large inter-element spacings, mutual coupling is negligible and $\mathbf{M}_r$ becomes an identity matrix.

To derive the coupling matrix at the transmit side, we can first express the voltages on each antenna as a function of the short circuit currents and then use the superposition principle in order to estimate, thanks to simplifications (2) and (3), the currents effectively induced on each antenna by its voltage. From (D.18) and (D.19), we have

$$
\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \left[\begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}^{-1} + \mathbf{Z}^{-1} \right]^{-1} \begin{bmatrix} I_{01} \\ I_{02} \end{bmatrix}.
\tag{D.22}
$$

The radiated field is estimated considering those voltages as perfect voltage sources and superpose the fields radiated by each voltage source. When we focus on one of those voltage sources, the other voltages are set to zero (the antennas are short-circuited). Since for minimum scatterers with respect to admittance parameters, a short-circuited antenna is invisible via assumption (3), we can focus on each antenna excited by its own voltage and write for each antenna the currents in the antennas I_{A1} and I_{A2} as

$$
\begin{bmatrix} I_{A1} \\ I_{A2} \end{bmatrix} = \begin{bmatrix} Z_{A1}^{-1} & 0 \\ 0 & Z_{A2}^{-1} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}
\tag{D.23}
$$

where we used the fact that the self impedance is the same as the antenna impedance by assumption (2). Assuming $Z_1 = Z_2 = Z_T$ and $Z_{A1} = Z_{A2} = Z_A$, combining (D.23) and (D.22) leads to

$$
\begin{bmatrix} I_{A1} \\ I_{A2} \end{bmatrix} = \frac{Z_T}{Z_T + Z_A} \underbrace{\frac{(Z_T + Z_A)}{Z_A} \mathbf{Z} \left[\mathbf{I}_2 Z_T + \mathbf{Z} \right]^{-1}}_{\mathbf{M}_t} \begin{bmatrix} I_{01} \\ I_{02} \end{bmatrix}
\tag{D.24}
$$

where we introduce the coupling matrix $\mathbf{M}_t$ at the transmit side. Comparing with eqn (D.21), we observe that the coupling matrices have the same form at the transmitter and the receiver. However, they are different from those derived for minimum scatterers with regard to impedance parameters in (D.4) and (D.7). Interpretations in terms of radiation patterns are naturally similar to minimum scatterers with regard to impedance parameters.

Appendix E

Derivation of the average pairwise error probability

There are several methods to evaluate the average PEP of space–time codes. In the sequel, we present a method based on Craig's formula that has been extensively used in [SA00, Sim01] because of the simplicity of its derivation. Other methods can be found in [UG01, UG04, TB02].

When a codeword $\mathbf{C}$ is transmitted through n_t antennas, we are interested in the probability that a ML decoder decodes the codeword $\mathbf{E} = [\mathbf{e}_1 \dots \mathbf{e}_T]$ instead of codeword $\mathbf{C}$. This probability is known as the pairwise error probability (PEP). When the PEP is conditioned on the channel realizations $\{\mathbf{H}_k\}_{k=1}^T$, it is defined as the conditional PEP

$$P\big(\mathbf{C} \to \mathbf{E} | \{\mathbf{H}_k\}_{k=1}^T\big) = \mathcal{Q}\left(\sqrt{\frac{\rho}{2} \sum_{k=1}^T \|\mathbf{H}_k\,(\mathbf{c}_k - \mathbf{e}_k)\|_F^2}\right) \tag{E.1}$$

where $\mathcal{Q}(x)$ is the Gaussian $\mathcal{Q}$-function defined as

$$\mathcal{Q}(x) \overset{\Delta}{=} P(x \geq y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty \exp\left(-\frac{y^2}{2}\right) dy. \tag{E.2}$$

The average PEP, denoted as $P(\mathbf{C} \to \mathbf{E})$, is finally obtained by averaging the conditional PEP of eqn (E.1) over the probability distribution of the channel gains, i.e.

$$P(\mathbf{C} \to \mathbf{E}) = \mathcal{E}_{\mathbf{H}_k}\left\{P\big(\mathbf{C} \to \mathbf{E} | \{\mathbf{H}_k\}_{k=1}^T\big)\right\}. \tag{E.3}$$

This requires to integrate eqn (E.1) weighted by the probability density function of the path gains.

This integration is sometimes difficult to calculate. Therefore, alternative forms of the Gaussian $\mathcal{Q}$-function are used. One such form is generally referred to as Craig's formula [Sim01]

$$\mathcal{Q}(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2(\beta)}\right) d\beta. \tag{E.4}$$

However, while the averaging can be performed with Craig's formula, the integration over parameter β is generally difficult and does not always lead to intuitive results. Hence, an upper-bound on the average PEP based on the Chernoff bound is generally used. The Gaussian $\mathcal{Q}$-function is then upper-bounded thanks to the Chernoff bound as

$$\mathcal{Q}(x) \le \exp\left(-\frac{x^2}{2}\right). \tag{E.5}$$

Hence, the Chernoff bound is obtained by taking $\sin^2(\beta) = 1$ in the integrand of Craig's formula. In the sequel, we evaluate the average PEP in Ricean channels with joint space–time correlation. The PEP is also particularized to i.i.d. Rayleigh slow and fast fading channels. We derive an upper-bound on the average PEP based on the Chernoff bound, and the exact average PEP based on Craig's formula.

We can derive the average PEP as follows

$$P(\mathbf{C} \to \mathbf{E}) = \mathcal{E}_{\mathbf{H}_k}\left\{ P\left(\mathbf{C} \to \mathbf{E} | \{\mathbf{H}_k\}_{k=1}^T\right)\right\}$$

$$= \frac{1}{\pi} \int_0^{\pi/2} M_\Gamma\left(-\frac{1}{2\sin^2(\beta)}\right) d\beta \tag{E.6}$$

$$\le M_\Gamma\left(-\frac{1}{2}\right) \tag{E.7}$$

where in (E.6) we use Craig's formula and in (E.7) the Chernoff bound. The quantity

$$M_\Gamma(\gamma) \triangleq \int_0^\infty \exp(\gamma\Gamma) p_\Gamma(\Gamma) \, d\Gamma \tag{E.8}$$

is the moment generating function (MGF) of $\Gamma = \frac{\rho}{2} \sum_{k=1}^T \|\mathbf{H}_k(\mathbf{c}_k - \mathbf{e}_k)\|_F^2$. The next step consists in expressing Γ as a Hermitian quadratic form of complex Gaussian random variables and then make use of the following theorem [SBS66, Tur60].

Theorem E.1. *The moment generating function of a Hermitian quadratic form in complex Gaussian random variable $y = \mathbf{z}\mathbf{F}\mathbf{z}^H$, where $\mathbf{z}$ is a circularly symmetric complex Gaussian vector with mean $\bar{\mathbf{z}}$ and a covariance matrix $\mathbf{R_z}$ and $\mathbf{F}$ a Hermitian matrix, is given by*

$$M_y(s) \triangleq \int_0^\infty \exp(sy) p_y(y) dy = \frac{\exp\left(s\bar{\mathbf{z}}\mathbf{F}(\mathbf{I} - s\mathbf{R_z}\mathbf{F})^{-1}\bar{\mathbf{z}}^H\right)}{\det(\mathbf{I} - s\mathbf{R_z}\mathbf{F})}. \tag{E.9}$$

The exact PEP or its Chernoff upper-bound can then be easily derived from (E.6) and (E.7). We apply this procedure to derive the expression of the average PEP in various kinds of channels in the sequel.

E.1 Joint space–time correlated Ricean fading channels

Defining

$$\mathbf{H} = \sqrt{\frac{K}{1+K}}\left(\mathbf{1}_{1\times T} \otimes \bar{\mathbf{H}}\right) + \sqrt{\frac{1}{1+K}}\left[\tilde{\mathbf{H}}_1 \ \tilde{\mathbf{H}}_2 \ \cdots \ \tilde{\mathbf{H}}_T\right]$$
$$\mathbf{D} = \mathrm{diag}\{\mathbf{c}_1 - \mathbf{e}_1, \mathbf{c}_2 - \mathbf{e}_2, \ldots, \mathbf{c}_T - \mathbf{e}_T\}, \tag{E.10}$$

we may write

$$\sum_{k=1}^{T} \|\mathbf{H}_k(\mathbf{c}_k - \mathbf{e}_k)\|_F^2 = \|\mathbf{H}\mathbf{D}\|_F^2 = \mathrm{Tr}\{\mathbf{H}\mathbf{D}\mathbf{D}^H\mathbf{H}^H\}$$
$$= \mathrm{vec}(\mathbf{H}^H)^H \mathbf{\Delta} \ \mathrm{vec}(\mathbf{H}^H) \tag{E.11}$$

where $\mathbf{\Delta} = \mathbf{I}_{n_r} \otimes \mathbf{D}\mathbf{D}^H$. Since this last equation turns out to be a Hermitian quadratic form of complex Gaussian random variables, the MGF of Γ can be evaluated thanks to (E.9) and the PEP averaged over the space–time correlated Ricean fading channel is finally given by

$$P(\mathbf{C} \rightarrow \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\eta K \bar{\mathcal{H}}^H \mathbf{\Delta}(\mathbf{I}_{Tn_r n_t} + \eta \mathbf{\Xi}\mathbf{\Delta})^{-1}\bar{\mathcal{H}}\right)$$
$$(\det(\mathbf{I}_{Tn_r n_t} + \eta\mathbf{\Xi}\mathbf{\Delta}))^{-1} d\beta \tag{E.12}$$

where the effective SNR η is defined as $\eta = \rho/(4\sin^2(\beta)(1+K))$. $\mathbf{\Xi} = \mathcal{E}\{\tilde{\mathcal{H}}\tilde{\mathcal{H}}^H\}$ denotes the space–time correlation matrix with

$$\tilde{\mathcal{H}} = \mathrm{vec}\left(\left[\ \tilde{\mathbf{H}}_1 \ \tilde{\mathbf{H}}_2 \ \cdots \ \tilde{\mathbf{H}}_T \ \right]^H\right)$$
$$\bar{\mathcal{H}} = \mathrm{vec}\left(\left(\mathbf{1}_{1\times T} \otimes \bar{\mathbf{H}}\right)^H\right). \tag{E.13}$$

Since $\det(\mathbf{I} + \mathbf{A}) = \prod_{k=1}^{r(\mathbf{A})}(1 + \lambda_k(\mathbf{A}))$, such that the PEP can equivalently be expressed as

$$P(\mathbf{C} \rightarrow \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\eta K \bar{\mathcal{H}}^H \mathbf{\Delta}(\mathbf{I}_{Tn_r n_t} + \eta \mathbf{\Xi}\mathbf{\Delta})^{-1}\bar{\mathcal{H}}\right)$$
$$\prod_{i=1}^{r(\mathbf{\Xi}\mathbf{\Delta})} (1 + \eta\lambda_i(\mathbf{\Xi}\mathbf{\Delta}))^{-1} d\beta \tag{E.14}$$

where $r(\mathbf{\Xi}\mathbf{\Delta})$ stands for the rank of $\mathbf{\Xi}\mathbf{\Delta}$, $\{\lambda_i(\mathbf{\Xi}\mathbf{\Delta})\}_{i=1}^{r(\mathbf{\Xi}\mathbf{\Delta})}$ are the non-zero eigenvalues of $\mathbf{\Xi}\mathbf{\Delta}$. This expression is totally general and expresses the average PEP over a Ricean fading channel with any space–time correlated Rayleigh component. No assumption is made about the channel variation or its spatial structure.

If the channel is Rayleigh distributed, $K = 0$ and the average PEP simplifies into

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} (\det(\mathbf{I}_{Tn_r n_t} + \eta \boldsymbol{\Xi} \boldsymbol{\Delta}))^{-1} d\beta \tag{E.15}$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{r(\boldsymbol{\Xi} \boldsymbol{\Delta})} (1 + \eta \lambda_i(\boldsymbol{\Xi} \boldsymbol{\Delta}))^{-1} d\beta. \tag{E.16}$$

Note that the effective SNR now becomes $\eta = \rho/(4 \sin^2(\beta))$.

Integration over β is generally quite difficult to calculate. Therefore, the Chernoff bound is used to upper-bound the average PEP. As explained above, the bound is simply obtained by removing the $\frac{1}{\pi} \int_0^{\pi/2} d\beta$ operator and taking $\sin^2(\beta) = 1$. For example, the average PEP in (E.15) can simply be upper-bounded as

$$P(\mathbf{C} \to \mathbf{E}) \leq \frac{1}{\pi} \int_0^{\pi/2} \left(\det \left(\mathbf{I}_{Tn_r n_t} + \frac{\rho}{4} \boldsymbol{\Xi} \boldsymbol{\Delta} \right) \right)^{-1} d\beta \tag{E.17}$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{r(\boldsymbol{\Xi} \boldsymbol{\Delta})} \left(1 + \frac{\rho}{4} \lambda_i(\boldsymbol{\Xi} \boldsymbol{\Delta}) \right)^{-1} d\beta. \tag{E.18}$$

In the sequel, we particularize (E.14) to slow and block fading channels. Sometimes, the Hermitian quadratic form differs from (E.11), which may lead to more intuitive expressions.

E.2   Space correlated Ricean slow fading channels

In the case of slow fading, $\widetilde{\mathbf{H}}_k = \widetilde{\mathbf{H}}$ for all k. This implies that $\mathbf{H}_k = \mathbf{H}$, hence we obtain the quadratic form

$$\sum_{k=1}^{T} \|\mathbf{H}_k(\mathbf{c}_k - \mathbf{e}_k)\|_F^2 = \|\mathbf{H}(\mathbf{C} - \mathbf{E})\|_F^2 = \mathrm{Tr}\{\mathbf{H} \tilde{\mathbf{E}} \mathbf{H}^H\} \tag{E.19}$$
$$= \mathrm{vec}(\mathbf{H}^H)^H (\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}) \mathrm{vec}(\mathbf{H}^H)$$

with $\tilde{\mathbf{E}} = (\mathbf{C} - \mathbf{E})(\mathbf{C} - \mathbf{E})^H$. The average PEP becomes

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \prod_{i=1}^{r(\mathbf{C_R})} (1 + \eta \lambda_i(\mathbf{C_R}))^{-1}$$

$$\exp\left(-\eta K \mathrm{vec}(\bar{\mathbf{H}}^H)^H (\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}})(\mathbf{I}_{n_r n_t} + \eta \mathbf{C_R})^{-1} \mathrm{vec}(\bar{\mathbf{H}}^H) \right) d\beta \tag{E.20}$$

where $\mathbf{C_R} = \mathbf{R}(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}})$ with $\mathbf{R} = \mathcal{E}\{\mathrm{vec}(\tilde{\mathbf{H}}^H)\mathrm{vec}(\tilde{\mathbf{H}}^H)^H\}$ the space correlation matrix. Equally, we could have derived (E.20) from eqn (E.14) by evaluating $\boldsymbol{\Xi}$ for slow fading channels.

E.3 Joint space–time correlated Ricean block fading channels

Another interesting class of channels is the block fading channels. In the case of block fading, the codeword length T extends over M channel blocks of length $N = T/M$. The channel gains are assumed to be independent from block to block and constant within a block. We denote the matrix of the mth channel block as $\mathbf{H}_{(m)}$ and define the quantities $\tilde{\mathbf{E}}_{(m)} = (\mathbf{C} - \mathbf{E})_{(m)}(\mathbf{C} - \mathbf{E})_{(m)}^H$ with $(\mathbf{C} - \mathbf{E})_{(m)} = \big[\mathbf{c}_{(m-1)N+1} - \mathbf{e}_{(m-1)N+1} \quad \ldots \quad \mathbf{c}_{mN} - \mathbf{e}_{mN}\big]$. While equivalent, we take here a different approach to express $\sum_{k=1}^{T} \|\mathbf{H}_k(\mathbf{c}_k - \mathbf{e}_k)\|_F^2$

$$\sum_{k=1}^{T} \|\mathbf{H}_k(\mathbf{c}_k - \mathbf{e}_k)\|_F^2 = \sum_{m=1}^{M} \|\mathbf{H}_{(m)}(\mathbf{C} - \mathbf{E})_{(m)}\|_F^2$$

$$= \sum_{m=1}^{M} \text{vec}\big(\mathbf{H}_{(m)}^H\big)^H \big(\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}_{(m)}\big) \text{vec}\big(\mathbf{H}_{(m)}^H\big)$$

$$= \mathcal{H}_b^H \, \tilde{\mathbf{E}}_{\mathbf{B}} \, \mathcal{H}_b \tag{E.21}$$

where

$$\mathcal{H}_b^H = \Big[\text{vec}\big(\mathbf{H}_{(1)}^H\big)^H \quad \ldots \quad \text{vec}\big(\mathbf{H}_{(M)}^H\big)^H\Big], \tag{E.22}$$

$$\tilde{\mathbf{E}}_{\mathbf{B}} = \text{diag}\big\{\mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}_{(1)}, \ldots, \mathbf{I}_{n_r} \otimes \tilde{\mathbf{E}}_{(M)}\big\}. \tag{E.23}$$

The average PEP then becomes

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \exp\Big(-\eta K \bar{\mathcal{H}}_b^H \tilde{\mathbf{E}}_{\mathbf{B}} (\mathbf{I}_{n_r n_t M} + \eta \mathbf{C}_{\mathbf{B}})^{-1} \bar{\mathcal{H}}_b\Big)$$

$$\prod_{i=1}^{r(\mathbf{C}_{\mathbf{B}})} (1 + \eta \lambda_i(\mathbf{C}_{\mathbf{B}}))^{-1} d\beta \tag{E.24}$$

where $\mathbf{C}_{\mathbf{B}} = \mathbf{B}\tilde{\mathbf{E}}_{\mathbf{B}}$ with $\mathbf{B} = \mathcal{E}\big\{\tilde{\mathcal{H}}_b \tilde{\mathcal{H}}_b^H\big\}$ the space–time correlation matrix of the block fading channel

$$\tilde{\mathcal{H}}_b^H = \Big[\text{vec}\big(\tilde{\mathbf{H}}_{(1)}^H\big)^H \quad \ldots \quad \text{vec}\big(\tilde{\mathbf{H}}_{(M)}^H\big)^H\Big] \tag{E.25}$$

$$\bar{\mathcal{H}}_b^H = \Big[\mathbf{1}_{1\times M} \otimes \text{vec}\big(\bar{\mathbf{H}}^H\big)^H\Big]. \tag{E.26}$$

If the blocks are independent, as is usually assumed, $\mathbf{B} = \text{diag}\big\{\mathbf{R}_{(1)}, \ldots, \mathbf{R}_{(M)}\big\}$ with $\mathbf{R}_{(m)} = \mathcal{E}\big\{\text{vec}\big(\tilde{\mathbf{H}}_{(m)}^H\big) \text{vec}\big(\tilde{\mathbf{H}}_{(m)}^H\big)^H\big\}$.

E.4   I.i.d. Rayleigh slow and fast fading channels

In the particular case of i.i.d. Rayleigh ($K = 0$) *slow fading* channels, $\Xi = \mathbf{I}_{n_r} \otimes \mathbf{1}_{T \times T} \otimes \mathbf{I}_{n_t}$ or equivalently $\mathbf{R} = \mathbf{I}_{n_r} \otimes \mathbf{I}_{n_t}$ and the average PEP eqn (E.12) (or equivalently eqn (E.20)) reduces to

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \left[\det\left(\mathbf{I}_{n_t} + \eta \tilde{\mathbf{E}}\right) \right]^{-n_r} d\beta \tag{E.27}$$

$$\leq \left[\det\left(\mathbf{I}_{n_t} + \frac{\rho}{4} \tilde{\mathbf{E}}\right) \right]^{-n_r} \tag{E.28}$$

where in (E.28) we use the Chernoff bound to upper-bound the average PEP.

In the case of i.i.d. Rayleigh ($K = 0$) *fast fading* channels, $\Xi = \mathbf{I}_{n_r n_t T}$ or equivalently $\mathbf{R} = \mathbf{I}_{n_r} \otimes \mathbf{I}_{n_t}$ and the average PEP eqns (E.12) or (E.24) reduces to

$$P(\mathbf{C} \to \mathbf{E}) = \frac{1}{\pi} \int_0^{\pi/2} \prod_{k=1}^T \left(1 + \eta \| \mathbf{c}_k - \mathbf{e}_k \|^2 \right)^{-n_r} d\beta \tag{E.29}$$

$$\leq \prod_{k=1}^T \left(1 + \frac{\rho}{4} \| \mathbf{c}_k - \mathbf{e}_k \|^2 \right)^{-n_r}. \tag{E.30}$$

Bibliography

[AG05]   J. Akhtar and D. Gesbert. Spatial Multiplexing over correlated MIMO channels with a closed-form precoder. *IEEE Trans. Wireless Commun.*, 4(5):2400–2409, September 2005.

[AK02]   A. Abdi and M. Kaveh. A space–time correlation model for multielement antenna systems in mobile fading channels. *IEEE J. Select. Areas Commun.*, 20(4):550–560, April 2002.

[Ala98]   S.M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE J. Select. Areas Commun.*, 16(10):1451–1458, October 1998.

[AMSM02]   H. Asplund, A.F. Molisch, M Steinbauer, and N.B. Mehta. Clustering of scatterers in mobile radio channels-evaluation and modeling in the COST 259 directional channel model. In *Proc. ICC 2002 – IEEE Int. Conf. Commun.*, volume 2, pages 901–905, New York City, NY, May 2002.

[AN00]   G.E. Athanasiadou and A.R. Nix. Investigation into the sensitivity of the power predictions of a microcellular ray-tracing propagation model. *IEEE Trans. Veh. Technol.*, 49(4):1140–1151, July 2000.

[ANM00]   G.E. Athanasiadou, A.R. Nix, and J.P. McGeehan. A microcellular ray-tracing propagation model and evaluation of its narrow-band and wide-band predictions. *IEEE J. Select. Areas Commun.*, 18(3):322–335, April 2000.

[APM02]   A. Algans, K.I. Pedersen, and P.E. Mogensen. Experimental analysis of the joint statistical properties of azimuth spread, delay spread, and shadow fading. *IEEE J. Select. Areas Commun.*, 20(3):523–531, March 2002.

[AS00]   M.S. Alouini and M. K. Simon. An MGF-based performance analysis of generalized selection combining over Rayleigh fading channels. *IEEE Trans. Commun.*, 48(3):401–415, March 2000.

[ATM03]   P. Almers, F. Tufvesson, and A.F. Molisch. Keyhole effects in MIMO wireless channels – measurements and theory. In *Proc. Globecom 2003 – IEEE Global Telecommunications Conf.*, volume 3, pages 1781–1785, San Francisco, CA, USA, December 2003.

[ATNS98]   D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri. Space–time coded OFDM for high data-rate wireless communication over wideband channels. In *Proc. VTC – IEEE Vehicular Technology Conf.*, volume 3, pages 2232–2236, 1998.

[Bal05]   C.A. Balanis. *Antenna theory: analysis and design*. Wiley, New York, NY, 2005.

[BBH00]   S. Baro, G. Bauch, and A. Hansmann. Improved codes for space–time trellis-coded modulation. *IEEE Commun. Lett.*, 4(1):20–22, January 2000.

[BBP03]   H. Boelcskei, M. Borgmann, and A.J. Paulraj. Space-frequency coded MIMO-OFDM with variable multiplexing-diversity tradeoff. In *Proc. ICC 2003 – IEEE Int. Conf. Commun.*, volume 4, pages 2837–2841, Anchorage, AK, May 2003.

[BDA03]   I. Bahceci, T.M. Duman, and Y. Altunbasak. Antenna selection for multiple-antenna transmission systems: performance analysis and code construction. *IEEE Trans. Inform. Theory*, 49(10):2669–2681, October 2003.

[BDMS91]   E. Biglieri, D. Divsalar, P.J. McLane, and K.K. Simon. *Introduction to TCM with applications*. Macmillan, New York, NY, 1991.

[BGN+00]   D.S. Baum, D.A. Gore, R.U. Nabar, S. Panchanathan, K.V.S. Hari, V. Erceg, and A.J. Paulraj. Measurement and characterization of broadband MIMO fixed wireless channels at 2.5 GHz. In *Proc. Int. Conf. on Personal Wireless Commun. (ICPWC2000)*, Hyderabad, India, December 2000.

[BGP02]   H. Boelcskei, D. Gesbert, and A.J. Paulraj. On the capacity of OFDM-based spatial multiplexing systems. *IEEE Trans. Commun.*, 50(2):225–234, February 2002.

[BGPvdV06]   H. Boelcskei, D. Gesbert, C.B. Papadias, and A.J. van der Veen. *Space–Time Wireless Systems: from array processing to MIMO communications*. Cambridge University Press, Cambridge, UK, 2006.

[BLWY01]   R. Blum, Y. Li, J. Winters, and Q. Yan. Improved space–time coding for MIMO-OFDM wireless communications. *IEEE Trans. Commun.*, 49(11):1873–1878, November 2001.

[BM06]   G. Bauch and J.S. Malik. Cyclic delay diversity with bit-interleaved coded modulation in orthogonal frequency division multiple access. *IEEE Trans. Wireless Commun.*, 5(8):2092–2100, August 2006.

[BOH+03]   E. Bonek, H. Oezcelik, M. Herdin, W. Weichselberger, and J. Wallace. Deficiencies of the kronecker MIMO radio channel model. In *Proc.*

WPMC 2003 – Wireless Pers. Multimedia Commun., Yokosuka, Japan, October 2003.

[BP00a] H. Boelcskei and A.J. Paulraj. Performances analysis of space–time codes in correlated Rayleigh fading environment. In *Proc. Asilomar Conf. On Signals, Systems, and Computers*, Pacific Grove, CA, November 2000.

[BP00b] H. Boelcskei and A.J. Paulraj. Space-frequency coded broadband OFDM systems. In *Proc. IEEE Wireless Comm. and Networking Conf.*, volume 1, pages 1–6, 2000.

[BP01] H. Boelcskei and A.J. Paulraj. Space-frequency codes for broadband fading channels. In *Proc. ISIT 2001 – IEEE Int. Symp. Information Theory*, page 219, June 2001.

[BR03] J.C. Belfiore and G. Rekaya. Quaternionic lattices for space–time coding. In *Proc. IEEE Inform. Theory Workshop*, Paris, France, April 2003.

[BRV05] J.C. Belfiore, G. Rekaya, and E. Viterbo. The golden code: a 2×2 full-rate space–time code with non-vanishing determinants. *IEEE Trans. Inform. Theory*, 51(4):1432–1436, April 2005.

[BTT02] E. Biglieri, G. Taricco, and A. Tulino. Performance of space–time codes for a large number of antennas. *IEEE Trans. Inform. Theory*, 48(7):1794–1803, July 2002.

[Bur03] A.G. Burr. Capacity bounds and estimates for the finite scatterers MIMO wireless channel. *IEEE J. Select. Areas Commun.*, 21(5):812–818, June 2003.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

[CCG$^+$on] G. Calcev, D. Chizhik, B. Goransson, S. Howard, H. Huang, A. Kogiantis, A. Molisch, A. Moustakas, D. Reed, and H. Xu. A MIMO channel model for third-generation cellular systems. *IEEE Trans. Wireless Commun.*, accepted for publication.

[CKT98] C.N. Chuah, J.M. Kahn, and D. Tse. Capacity of multi-antenna array systems in indoor wireless environment. In *Proc. Globecom 1998 – IEEE Global Telecommunications Conf.*, volume 4, pages 1894–1899, Sydney, Australia, 1998.

[Cle05] B. Clerckx. *Space–Time Signaling for Real-World MIMO Channels*. PhD thesis, Université catholique de Louvain, September 2005.

[CLW$^+$03] D. Chizhik, J. Ling, P.W. Wolniansky, R.A. Valenzuela, N. Costa, and K. Huber. Multiple-input – multiple-output measurements and modeling in Manhattan. *IEEE J. Select. Areas Commun.*, 21(3):321–331, April 2003.

[Cor01]   L.M. Correia. *COST 259 – Wireless flexible personalized communica-tions*. Wiley, London, UK, 2001.

[Cor06]   L.M. Correia. *COST 273 – Towards mobile broadband multimedia networks*. Elsevier, London, UK, 2006.

[COV⁺07]   B. Clerckx, C. Oestges, L. Vandendorpe, D. Vanhoenacker-Janvier, and A.J. Paulraj. Design and performance of space–time codes for spa-tially correlated MIMO channels. *IEEE Trans. Commun.*, 55(1):64–68 January 2007.

[CT91]   T. Cover and T. Thomas. *Elements of Information Theory*. Wiley, New York, NY, 1991.

[CTKV02]   C.N. Chuah, D. Tse, J.M. Kahn, and R.A. Valenzuela. Capacity scaling in MIMO wireless systems under correlated fading. *IEEE Trans. Inform. Theory*, 48(3):637–650, March 2002.

[CVJV05]   B. Clerckx, D. Vanhoenacker-Janvier, and L. Vandendorpe. Robustness of space–time coding in spatially correlated fast fading MIMO channels. In *Proc. VTC 2005 Fall – IEEE 62ᵗʰ Vehicular Technology Conf.*, Dallas, TX, September 2005.

[CVVJO04]   B. Clerckx, L. Vandendorpe, D. Vanhoenacker-Janvier, and C. Oest-ges. Optimization of non-linear signal constellations for real-world MIMO channels. *IEEE Trans. Signal Processing*, 52(4):894–902, April 2004.

[CYV01]   Z. Chen, J. Yuan, and B. Vucetic. Improved space–time trellis coded modulation scheme on slow Rayleigh fading channels. *Elect. Lett.*, 37(7):440–441, March 2001.

[DCB00]   O. Damen, A. Chkeif, and J.-C. Belfiore. Lattice code decoder for space–time codes. *IEEE Commun. Lett.*, 4(5):161–163, May 2000.

[DK01]   A. Dammann and S. Kaiser. Standard conformable antenna diversity tech-niques for OFDM systems and its application to the DVB-T system. In *Proc. Globecom 2001 – IEEE Global Telecommunications Conf.*, pages 3100–3105, San Antonio, TX, November 2001.

[DLLD05]   J. Dumont, P. Loubaton, S. Lasaulce, and M. Debbah. On the asymptotic performance of MIMO correlated Ricean channels. In *Proc. ICASSP 2005 – IEEE Int. Conf. Acoust. Speech and Signal Processing*, pages 813–816, Philadelphia, PA, March 2005.

[DM05]   M. Debbah and R. Muller. MIMO channel modelling and the principle of maximum entropy. *IEEE Trans. Inform. Theory*, 51(5):1667–1690, May 2005.

[DR98]   G. Durgin and T.S. Rappaport. Basic relationship between multipath angular spread and narrowband fading in wireless channels. *Elect. Lett.*, 34(25):2431–2432, December 1998.

[DTB02]   M.O. Damen, A. Tewfik, and J.-C. Belfiore. A construction of a space–time code based on number theory. *IEEE Trans. Inform. Theory*, 48(3):753–760, March 2002.

[DV05]   P. Dayal and M.K. Varanasi. An optimal two transmit antenna space–time code and its stacked extensions. *IEEE Trans. Inform. Theory*, 51(12):4348–4355, December 2005.

[ECS+98]   R.B. Ertel, P. Cardieri, K.W. Sowerby, T.S. Rappaport, and J.H. Reed. Space–time coding with maximum diversity gains over frequency selective fading channels. *IEEE Personal Commun. Mag.*, pages 10–21, February 1998.

[Ede89]   A. Edelman. *Eigenvalues and condition numbers of random matrices*. PhD thesis, Massachusetts Institute of Technology, May 1989.

[Eea01]   V. Erceg *et al*. IEEE p802.16 – channel models for fixed wireless applications (ieee802.16.3c-01/29r4), 2001.

[Eea04]   V. Erceg *et al*. IEEE p802.11 – wireless LANs: TGn channel models, 2004.

[ER99]   R.B. Ertel and J.H. Reed. Angle and time of arrival statistics for circular and elliptical scattering models. *IEEE J. Select. Areas Commun.*, 17(11):1829–1840, November 1999.

[ERKP+06]   P. Elia, K. Raj Kumar, S.A. Pawar, P. Vijay Kumar, and H-F. Lu. Explicit space–time codes achieving the diversity-multiplexing gain tradeoff. *IEEE Trans. Inform. Theory*, 52(9): 3869–3884, September 2006.

[FADT02]   C. Fragouli, N. Al-Dhahir, and W. Turin. Effect of spatio-temporal channel correlation on the performance of space–time codes. In *Proc. ICC 2002 – IEEE Int. Conf. Commun.*, volume 2, pages 826–830, New York City, NY, May 2002.

[FG98]   G.J. Foschini and M.J. Gans. On the limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6:311–335, March 1998.

[FGW74]   G.J. Foschini, R.D. Gitlin, and S.B. Weinstein. Optimization of two-dimensional signal constellations in the presence of gaussian noise. *IEEE Trans. Commun.*, 22(1):28–38, January 1974.

[Fle00]   B.H. Fleury. First- and second-order characterization of direction dispersion and space selectivity in the radio channel. *IEEE Trans. Inform. Theory*, 46(6):2027–2044, June 2000.

[FMK+06]   T. Fuegen, J. Maurer, T Kayser, and W. Wiesbeck. Verification of 3D ray-tracing with non-directional and directional measurements in urban macrocellular environments. In *Proc. VTC 2006 Spring – IEEE 61st Vehicular Technology Conf.*, Melbourne, Australia, May 2006.

[Fos96]   G.J. Foschini. Layered space–time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Tech. J.*, pages 41–59, Autumn 1996.

[FP85]   U. Fincke and M. Pohst. Improved methods for calculating vectors of short length in a lattice, including a complexity analysis. *Math. Comput.*, 44:463–471, April 1985.

[FVY01]   W. Firmanto, B. Vucetic, and J. Yuan. Space–time TCM with improved performance on fast fading channels. *IEEE Commun. Lett.*, 5(4):154–156, April 2001.

[GA05]   D. Gesbert and J. Akhtar. Transmitting over ill-conditioned MIMO channels: From spatial to constellation multiplexing. *Smart Antennas in Europe – State-of-the-Art (EURASIP Book Series),* T. Kaiser *et al.*, Ed. Sylvania, OH: Hindawi Publishing Co., 2005.

[Gam02]   H.E. Gamal. On the robustness of space–time coding. *IEEE Trans. Signal Processing*, 50(10):2417–2428, October 2002.

[GBGP02]   D. Gesbert, H. Boelcskei, D.A. Gore, and A.J. Paulraj. Outdoor MIMO wireless channels: models and performance prediction. *IEEE Trans. Commun.*, 50(12):1926–1934, December 2002.

[GCD04]   H.E. Gamal, G. Caire, and M.O. Damen. Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels. *IEEE Trans. Inform. Theory*, 50(6):968–985, June 2004.

[GD03]   H.E. Gamal and M.O. Damen. Universal space–time coding. *IEEE Trans. Inform. Theory*, 49(5):1097–1119, May 2003.

[GFVW99]   G.D. Golden, G.J. Foschini, R.A. Valenzuela, and P.W. Wolniansky. Detection algorithm and initial laboratory results using the V-BLAST space–time communication architecture. *Elect. Lett.*, 35(1):14–15, January 1999.

[GGP03a]   A. Gorokhov, D. Gore, and A.J. Paulraj. Performance bounds for antenna selection in MIMO systems. In *Proc. ICC 2003 – IEEE Int. Conf. Commun.*, volume 4, pages 3021–3025, Anchorage, AK, May 2003.

[GGP03b]   A. Gorokhov, D. Gore, and A.J. Paulraj. Receive antenna selection for MIMO flat-fading channels: theory and algorithms. *IEEE Trans. Inform. Theory*, 49(10):2687–2696, October 2003.

[GHL⁺03]   H.E. Gamal, A.R. Hammons, Y. Liu, M.P. Fitz, and O.Y. Takeshita. On the design of space–time and space-frequency codes for MIMO frequency-selective fading channels. *IEEE Trans. Inform. Theory*, 49(9):2277–2292, September 2003.

[GK83]   I.J. Gupta and A.K. Ksienski. Effect of mutual coupling on the performance of adaptive arrays. *IEEE Trans. Antennas Propagat.*, 31(5):785–791, May 1983.

[GL01]   Y. Gong and K.B. Letaief. Space-frequency-time coded OFDM for broadband wireless communications. In *Proc. Globecom 2001 – IEEE Global Telecommunications Conf.*, San Antonio, TX, November 2001.

[GL02]   Y. Gong and K.B. Letaief. An efficient space-frequency coded wideband OFDM system for wireless communications. In *Proc. ICC 2002 – IEEE Int. Conf. Commun.*, volume 1, pages 475–479, New York City, NY, May 2002.

[God97]   L.C. Godara. Applications of antenna arrays to mobile communications, part II: Beamforming and direction-of-arrival considerations. *Proceedings IEEE*, 85(8):1195–1245, August 1997.

[GP02]   D.A. Gore and A.J. Paulraj. MIMO antenna subset selection with space–time coding. *IEEE Trans. Signal Processing*, 50(10):2580–2588, October 2002.

[Gro05]   L. Grokop. Diversity multiplexing tradeoff in ISI channels. Master's thesis, University of California-Berkeley, May 2005.

[GSP02]   D. Gore, S. Sandhu, and A. Paulraj. Delay diversity codes for frequency selective channels. In *Proc. ICC 2002 – IEEE Int. Conf. Commun.*, pages 1949–1953, New York, May 2002.

[GSS⁺03]   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.*, 21(3):281–302, April 2003.

[GT04]   L. Grokop and D. Tse. Diversity/multiplexing tradeoff in ISI channels. In *Proc. ISIT 2004 – IEEE Int. Symp. Information Theory*, page 96, Chicago, June 2004.

[Gui05]   M. Guillaud. *Transmission and Channel Modeling Techniques for Multiple-Antenna Communication Systems*. PhD thesis, Ecole Nationale Superieure des Telecommunications, July 2005.

[HAN92]   A. Hiroike, F. Adachi, and N. Nakajima. Combined effects of phase sweeping transmitter diversity and channel coding. *IEEE Trans. Veh. Technol.*, 41(2):170–176, May 1992.

[Han98]   R.C. Hansen. *Phased-array antennas*. Wiley, New York, NY, 1998.

[Hea01]   R. Heath. *Space–Time Signaling in Multi-Antenna Systems*. PhD thesis, Stanford University, November 2001.

[Her04]   M. Herdin. *Non-stationary indoor MIMO radio channels*. PhD thesis, Technische Universitat Wien, August 2004.

[HGA06]   A. Hjorungnes, D. Gesbert, and J. Akhtar. Precoding of space–time block coded signals for joint transmit-receive correlated mimo channels. *IEEE Trans. Wireless Commun.*, 5(3):492–497, March 2006.

[HG07a]   A. Hjorungnes and D. Gesbert. Precoded orthogonal space–time block codes over correlated Ricean MIMO channels. *IEEE Trans. Signal Processing*, 55(2):779–783, February 2007.

[HG07b]   A. Hjorungnes and D. Gesbert. Precoding of orthogonal space–time block codes in arbitrarily correlated MIMO channels: iterative and closed-form solutions. *IEEE Trans. Wireless Commun.*, to appear in 7(2) March 2007.

[HGonc]   A. Hjorungnes and D. Gesbert. Complex-valued matrix differentiation: techniques and key results. *IEEE Trans. Signal Processing*, accepted for publication.

[HH01]   B. Hassibi and B. Hochwald. High-rate linear space–time codes. In *Proc. ICASSP 2001 – IEEE Int. Conf. Acoust. Speech and Signal Processing*, volume 4, pages 2461–2464, Salt Lake City, UT, May 2001.

[HH02a]   B. Hassibi and B.M. Hochwald. High-rate codes that are linear in space and time. *IEEE Trans. Inform. Theory*, 48(7):1804–1824, July 2002.

[HH02b]   Z. Hong and B. Hughes. Robust space–time codes for broadband OFDM systems. In *Proc. IEEE Wireless Comm. and Networking Conf.*, volume 1, pages 105–108, 2002.

[HJ95]   R.A. Horn and C.R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, UK, 1995.

[HL05]   R.W. Heath and D.J. Love. Multimode antenna selection for spatial multiplexing systems with linear receivers. *IEEE Trans. Signal Processing*, 53(6):3042–3056, August 2005.

[HLHS03]   Z. Hong, K. Liu, R.W. Heath, and A. Sayeed. Spatial multiplexing in correlated fading via the virtual channel representation. *IEEE J. Select. Areas Commun.*, 21(5):856–866, June 2003.

[HMR$^+$00] B.M. Hochwald, T.L. Marzetta, T.J. Richardson, W. Sweldens, and R. Urbanke. Systematic design of unitary space–time constellations. *IEEE Trans. Inform. Theory*, 46(9):1962–1973, September 2000.

[HP02]   R. Heath and A.J. Paulraj. Linear dispersion codes for MIMO systems based on frame theory. *IEEE Trans. Signal Processing*, 50(10):2429–2441, October 2002.

[HSP01]   R.W. Heath, S. Sandhu, and A.J. Paulraj. Antenna selection for spatial multiplexing systems with linear receivers. *IEEE Commun. Lett.*, 5(4):142–144, April 2001.

[HVBon]   Y. Hong, E. Viterbo, and J.-C. Belfiore. Golden space–time trellis coded modulation. *IEEE Trans. Inform. Theory*, submitted for publication.

[IMYL01]   M. Ionescu, K.K. Mukkavilli, Z. Yan, and J. Lilleberg. Improved 8- and 16-state space–time codes for 4PSK with two transmit antennas. *IEEE Commun. Lett.*, 5(7):301–305, July 2001.

[IN03]   M.T. Ivrlac and J.A. Nossek. Quantifying diversity and correlation of Rayleigh fading MIMO channels. In *Proc. 3rd IEEE Int. Symp. on Signal Processing and Information Technology, (ISSPIT2003)*, pages 158–161, Darmstadt, Germany, December 2003.

[Ish98]   A. Ishimaru. *Wave propagation and scattering in random media (vol. 2)*. Academic Press, New York, NY, 1998.

[Jaf01]   H. Jafarkhani. A quasi orthogonal space–time block code. *IEEE Trans. Commun.*, 49(1):1–4, January 2001.

[Jan00]   R. Janaswamy. *Radiowave propagation and smart antenna for wireless communications*. Kluwer Academic Publishers, Boston, MA, 2000.

[JB04]   E. Jorswieck and H. Boche. Optimal transmission strategies and impact of correlation in multi-antenna systems with different types of channel state information. *IEEE Trans. Signal Processing*, 52(12):3440–3453, December 2004.

[JG01]   S.A. Jafar and A. Goldsmith. On optimality of beamforming for multiple antenna systems with imperfect feedback. In *Proc. ISIT 2001 – IEEE Int. Symp. Information Theory*, Washington, DC, June 2001.

[JG04]   S.A. Jafar and A. Goldsmith. Transmitter optimization and optimality of beamforming for multiple antenna systems with imperfect feedback. *IEEE Trans. Wireless Commun.*, 3(4):1165–1175, July 2004.

[JH05]   H. Jafarkhani and N. Hassanpour. Super-quasi-orthogonal space–time trellis codes for four transmit antennas. *IEEE Trans. Wireless Commun.*, 4(1):215–227, January 2005.

[JS03]   H. Jafarkhani and N. Seshadri. Super-orthogonal space–time trellis codes. *IEEE Trans. Inform. Theory*, 49(4):937–950, April 2003.

[JS04]   G. Jongren and M. Skoglund. Quantized feedback information in orthogonal space–time block coding. *IEEE Trans. Inform. Theory*, 50(10):2473–2486, October 2004.

[JSO02]   G. Jongren, M. Skoglund, and B. Ottersten. Combining beamforming and orthogonal space–time block coding. *IEEE Trans. Inform. Theory*, 48(3):611–627, March 2002.

[KA06a]   M. Kang and M.S. Alouini. Capacity of correlated MIMO Rayleigh channels. *IEEE Trans. Wireless Commun.*, 54(1):143–155, January 2006.

[KA06b]   M. Kang and M.S. Alouini. Capacity of MIMO Ricean channels. *IEEE Trans. Wireless Commun.*, 54(1):112–122, January 2006.

[Kat02]   R. Kattenbach. Statistical modeling of small-scale fading in directional channels. *IEEE J. Select. Areas Commun.*, 20(3):584–592, April 2002.

[Kel62]   J.B. Keller. Geometrical theory of diffraction. *J. Opt. Soc. Of America*, 52(2):116–130, February 1962.

[KP74]   R.G. Kouyoumjian and P.H. Pathak. A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface. *Proceedings of the IEEE*, 62(11):1448–1461, November 1974.

[KS93]   D.F. Kelley and W.L. Stutzman. Array antenna pattern modelling methods that include mutual coupling effects. *IEEE Trans. Antennas Propagat.*, 41(12):1625–1632, December 1993.

[KSP$^+$02] J.P. Kermoal, L. Schumacher, K.I. Pedersen, P. Mogensen, and F. Frederiksen. A stochastic MIMO radio channel model with experimental validation. *IEEE J. Select. Areas Commun.*, 20(6):1211–1226, June 2002.

[KSR05]   T. Kiran and B. Sundar Rajan. STBC-schemes with non-vanishing determinant for certain number of transmit antennas. *IEEE Trans. Inform. Theory*, 51(8):2984–2992, August 2005.

[KST06]   W.A.T. Kotterman, G. Sommerkorn, and R.S. Thoma. Cross-correlation values for dual-polarised indoor MIMO links and realistic antenna elements. In *Proc. 3rd Int. Symp. Wireless Communication Systems*, Valencia, Spain, September 2006.

[KW01]   C. Köse and R.D. Wesel. Code design metrics for space–time systems under arbitrary fading. In *Proc. ICC 2004 – IEEE Int. Conf. Commun.*, Helsinki, Finland, June 2001.

[KW03]   C. Köse and R.D. Wesel. Universal space–time trellis codes. *IEEE Trans. Inform. Theory*, 40(10):2717–2727, October 2003.

[Lee97]   W.C.Y. Lee. *Mobile communications engineering*. 2nd ed., McGraw-Hill, New York, NY, 1997.

[LFT01]   Y. Liu, M.P. Fitz, and O.Y. Takeshita. Space–time codes performance criteria and design for frequency selective fading channels. In *Proc. ICC 2005 – IEEE Int. Conf. Commun.*, volume 9, pages 2800–2804, Helsinki, Finland, June 2001.

[LGSW02]   E. Larsson, G. Ganesan, P. Stoica, and W.H. Wong. On the performance of orthogonal space–time block coding with quantized feedback. *IEEE Commun. Lett.*, 11(11):487–489, November 2002.

[LH03]   D.J. Love and R.W. Heath. Equal gain transmission in multiple-input multiple-output wireless systems. *IEEE Trans. Commun.*, 51(7):1102–1110, July 2003.

[LH05a]   D.J. Love and R.W. Heath. Limited feedback unitary precoding for orthogonal space–time block codes. *IEEE Trans. Signal Processing*, 53(1):64–73, January 2005.

[LH05b]   D.J. Love and R.W. Heath. Limited feedback unitary precoding for spatial multiplexing systems. *IEEE Trans. Inform. Theory*, 51(8):2967–2976, August 2005.

[LH05c]   D.J. Love and R.W. Heath. Multimode precoding for MIMO wireless systems. *IEEE Trans. Signal Processing*, 53(10):3674–3687, October 2005.

[LH05d]   D.J. Love and R.W. Heath. Necessary and sufficient conditions for full diversity order in correlated Rayleigh fading beamforming and combining systems. *IEEE Trans. Wireless Commun.*, 4(1):20–23, January 2005.

[LH06]   D.J. Love and R.W. Heath. Limited feedback diversity techniques for correlated channels. *IEEE Trans. Veh. Technol.*, 55(2):718–722, March 2006.

[LHS03]   D.J. Love, R.W. Heath, and T. Strohmer. Grassmannian beamforming for multiple-input multiple-output wireless systems. *IEEE Trans. Inform. Theory*, 49(10):2735–2747, October 2003.

[LP00]   E. Lindskog and A.J. Paulraj. A transmit diversity scheme for channels with intersymbol interference. In *Proc. ICC 2000 – IEEE Int. Conf. Commun.*, volume 1, pages 307–311, New Orleans, June 2000.

[LR96]   J.C. Liberti and T.S. Rappaport. A geometrically based model for line-of-sight multipath radio channels. In *Proc. VTC 1996 Spring – IEEE 46th Vehicular Technology Conf.*, volume 2, pages 844–848, Atlanta, GA, May 1996.

[LSLL02]   E.G. Larsson, P. Stoica, E. Lindskog, and J. Li. Space–time block coding for frequency-selective channels. In *Proc. ICASSP 2002 – IEEE Int. Conf. Acoust. Speech and Signal Processing*, pages 2405–2408, Orlando, FL, May 2002.

[LTV03]   A. Lozano, A.M. Tulino, and S. Verdu. Multiple-antenna capacity in the low-power regime. *IEEE Trans. Inform. Theory*, 49(10):2527–2544, October 2003.

[LTV05]   A. Lozano, A.M. Tulino, and S. Verdu. High-SNR power offest in multi-antenna Ricean channels. In *Proc. ICC 2001 – IEEE Int. Conf. Commun.*, volume 1, pages 683–687, Seoul, South Korea, May 2005.

[Lue84]   R.J. Luebbers. Finite conductivity uniform GTD versus knife edge diffraction in prediction of propagation path-loss. *IEEE Trans. Antennas Propagat.*, 32(1):70–75, January 1984.

[LW00a]   K. Lee and D. Williams. A space-frequency transmitter diversity technique for OFDM systems. In *Proc. Globecom 2000 – IEEE Global Telecommunications Conf.*, volume 3, pages 1473–1477, San Francisco, CA, December 2000.

[LW00b]   B. Lu and X. Wang. Space–time code design in OFDM systems. In *Proc. Globecom 2000 – IEEE Global Telecommunications Conf.*, volume 2, pages 1000–1004, San Francisco, CA, December 2000.

[LXG02]   Z. Liu, Y. Xin, and G. Giannakis. Space–time-frequency coded OFDM over frequency selective fading channels. *IEEE Trans. Signal Processing*, 50(10):2465–2476, October 2002.

[MBF02]   D.P. McNamara, M.A. Beach, and P.N. Fletcher. Spatial correlation in indoor MIMO channels. In *Proc. PIMRC 2002 – IEEE 13th Int. Symp. on Pers., Indoor and Mobile Radio Commun.*, volume 1, pages 290–294, Lisbon, Portugal, September 2002.

[MP67]   V.A. Marčenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR – Sbornik*, 1(4):457–483, 1967.

[MPM90]   D. A. McNamara, C.W.I. Pistorius, and J.A.G. Malherbe. *Introduction to the uniform geometrical theory of diffraction*. Artech House, London, UK, 1990.

[MSEA03]   K.K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang. On beamforming with finite rate feedback in multiple-antenna systems. *IEEE Trans. Inform. Theory*, 49(10):2562–2579, October 2003.

[MW04]   A.F. Molisch and M.Z. Win. MIMO systems with antenna selection. *IEEE Microw. Mag.*, 5(1):46–56, March 2004.

[MWCW05]   A.F. Molisch, M.Z. Win, Y.S. Choi, and J.H. Winters. Capacity of MIMO systems with antenna selection. *IEEE Trans. Wireless Commun.*, 4(4):1759–1772, July 2005.

[MWW01]   A.F. Molisch, M.Z. Win, and J.H. Winters. Capacity of MIMO systems with antenna selection. In *Proc. ICC 2001 – IEEE Int. Conf. Commun.*, volume 2, pages 570–574, Helsinki, Finland, June 2001.

[Nab03]   R.U. Nabar. *Performance Analysis and Transmit Optimization for General MIMO Channels*. PhD thesis, Stanford University, February 2003.

[Nar03]   R. Narasimhan. Spatial multiplexing with transmit antenna and constellation selection for correlated MIMO fading channels. *IEEE Trans. Signal Processing*, 51(11):2829–2838, November 2003.

[Nar05]   R. Narasimhan. Finite-snr diversity performance of rate-adaptive MIMO systems. In *Proc. Globecom 2005 – IEEE Global Telecommunications Conf.*, St. Louis, MO, December 2005.

[Nar06]   R. Narasimhan. Finite-snr diversity-multiplexing tradeoff for correlated Rayleigh and Rician MIMO channels. *IEEE Trans. Inform. Theory*, 52(9):3995–3979, September 2006.

[NBE$^+$02]   R.U. Nabar, H. Boelcskei, V. Erceg, D. Gesbert, and A.J. Paulraj. Performance of multiantenna signaling techniques in the presence of polarization diversity. *IEEE Trans. Signal Processing*, 50(10):2553–2562, October 2002.

[NBP01]   R.U. Nabar, H. Boelcskei, and A.J. Paulraj. Transmit optimization for spatial multiplexing in the presence of spatial fading correlation. In *Proc. Globecom 2001 – IEEE Global Telecommunications Conf.*, volume 1, pages 131–135, San Antonio, TX, November 2001.

[NBP05]   R.U. Nabar, H. Boelcskei, and A.J. Paulraj. Diversity and outage performance in space–time block coded Ricean MIMO channels. *IEEE Trans. Wireless Commun.*, 4(5):2519–2532, September 2005.

[NLTW98]   A. Narula, M.J. Lopez, M.D. Trott, and G.W. Wornell. Efficient use of side information in multiple-antenna data transmission over fading channels. *IEEE J. Select. Areas Commun.*, 16(10):1423–1436, October 1998.

[OCB05]   H. Oezcelik, N. Czink, and E. Bonek. What makes a good MIMO channel model? In *Proc. VTC 2005 Spring–IEEE 61st Vehicular Technology Conf.*, volume 1, pages 156–160, Stockholm, Sweden, May 2005.

[OCRVJ02]   C. Oestges, B. Clerckx, L. Raynaud, and D. Vanhoenacker-Janvier. Deterministic channel modeling and performance simulation of microcellular wideband communication systems. *IEEE Trans. Veh. Technol.*, 51(6):1422–1430, June 2002.

[OEP03]   C. Oestges, V. Erceg, and A.J. Paulraj. A physical scattering model for MIMO macrocellular broadband wireless channels. *IEEE J. Select. Areas Commun.*, 21(5):721–729, June 2003.

[OEP04]   C. Oestges, V. Erceg, and A.J Paulraj. Propagation modeling of multi-polarized MIMO fixed wireless channels. *IEEE Trans. Veh. Technol.*, 53(3):644–654, May 2004.

[Oes06]   C. Oestges. Mutual information of non-Kronecker structured dual-polarized 2×2 channels. *IEEE Trans. Veh. Technol.*, 56(1):410–413, January 2006.

[Oez04]   H. Oezcelik. *Indoor MIMO channel models*. PhD thesis, Technische Universitat Wien, December 2004.

[ONBP03]   O. Oyman, R.U. Nabar, H. Boelcskei, and A.J. Paulraj. Characterizing the statistical properties of mutual information in MIMO channels. *IEEE Trans. Signal Processing*, 51(11):2784–2795, November 2003.

[OO05]   H. Oezcelik and C. Oestges. Some remarkable properties of diagonally correlated MIMO channels. *IEEE Trans. Veh. Technol.*, 54(6):2143–2145, November 2005.

[OP04]   C. Oestges and A.J. Paulraj. Beneficial impact of channel correlations on MIMO capacity. *Elect. Lett.*, 40(10):606–607, May 2004.

[ORBV06]   F. Oggier, G. Rekaya, J.C. Belfiore, and E. Viterbo. Perfect space–time block codes. *IEEE Trans. Inform. Theory*, 52(9):3885–3902, September 2006.

[OV04]   F.E. Oggier and E. Viterbo. *Algebraic number theory and code design for Rayleigh fading channels*. Foundations and Trends in Communications and Information Theory, 2004.

[OVJC04]   C. Oestges, D. Vanhoenacker-Janvier, and B. Clerckx. SIMO measurements and directional characterization of outdoor wireless channels at 1.9 GHz. *IEEE Trans. Veh. Technol.*, 53(4):1190–1202, July 2004.

[Par00]   J.D. Parsons. *The mobile radio propagation channel*. 2nd ed., Wiley, London, UK, 2000.

[PBN06]   A. Pal, M. Beach, and A. Nix. A novel quantification of 3D directional spread from small-scale fading analysis. In *Proc. VTC 2006 Spring – IEEE 61st Vehicular Technology Conf.*, Melbourne, Australia, May 2006.

[PF03]   C.B. Papadias and G.J. Foschini. Capacity-approaching space–time codes for systems employing four transmitter antennas. *IEEE Trans. Inform. Theory*, 49(3):726–732, March 2003.

[PK94]   A.J. Paulraj and T. Kailath. Increasing capacity in wireless broadcast systems using distributed transmission/directional reception. *U.S. Patent, no. 5,345,599*, 1994.

[PNG03]   A. Paulraj, R. Nabar, and D. Gore. *Introduction to Space–Time Wireless Communications*. Cambridge University Press, Cambridge, UK, 2003.

[Poh81]   M. Pohst. On the computation of lattice vectors of minimal length, successive minima and reduced bases with applications. *ACM SIGSAM Bull.*, 15:37–44, February 1981.

[Poz98]   D.M. Pozar. *Microwave Engineering*. Wiley, New York, NY, 1998.

[Pro01]   J.G. Proakis. *Digital communications*. 4th ed., McGraw-Hill, New York, NY, 2001.

[PRR02]   P. Petrus, J.H. Reed, and T.S. Rappaport. Geometrical-based statistical macrocell channel model for mobile environments. *IEEE Trans. Commun.*, 50(3):495–502, March 2002.

[RBV04]   G. Rekaya, J.C. Belfiore, and E. Viterbo. Algebraic 3×3, 4×4 and 6×6 space–time codes with non-vanishing determinants. In *Int. Symp. Inform. Theory and its Applications*, pages 325–329, Parma, Italy, October 2004.

[RC98]   G. Raleigh and J. Cioffi. Spatio-temporal coding for wireless communication. *IEEE Trans. Commun.*, 46(3):357–366, March 1998.

[RFLFV00] F. Rashid-Farrokhi, A. Lozano, G. Foschini, and R.A. Valenzuela. Spectral efficiency of wireless systems with multiple transmit and receive antennas. In *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC)*, volume 1, pages 373–377, 2000.

[RK05]   K. Rosengren and P.-S. Kildal. The effect of source and load impedance on radiation efficiency and diversity gain of two parallel dipoles. In *Proc. IEEE Antennas and Propagation Society Symposium*, pages 1422–1430, Washington, D.C., July 2005.

[RP94]   A.J. Roscoe and R.A. Perrott. Large finite array analysis using infinite array data. *IEEE Trans. Antennas Propagat.*, 42(7):983–992, July 1994.

[RV68]   W.L. Root and P.P. Varaiya. Capacity of classes of Gaussian channels. *SIAM Journal of Applied Mathematics*, 16(6):1350–1393, November 1968.

[SA00]   M.K. Simon and M.-S. Alouini. *Digital communications over fading channels: a unified approach to performance analysis*. Wiley, New York, NY, 2000.

[San02]   S. Sandhu. *Signal Design for Multiple-Input Multiple-Output Wireless: A Unified Perspective*. PhD thesis, Stanford University, August 2002.

[Say02]   A.M. Sayeed. Deconstructing multiantenna fading channels. *IEEE Trans. Signal Processing*, 50(10):2563–2579, October 2002.

[SBS66]   M. Schwartz, W.R. Bennett, and S. Stein. *Communication systems and techniques*. Inter-University Electronics/McGraw-Hill, New York, NY, 1966.

[SF02a]   S. Siwamogsatham and M.P. Fitz. Improved high-rate space–time codes via concatenation of expanded orthogonal block code and M-TCM. In *Proc. ICC 2002 – IEEE Int. Conf. Commun.*, volume 1, pages 636–640, April 2002.

[SF02b]   S. Siwamogsatham and M.P. Fitz. Robust space–time codes for correlated Rayleigh fading channels. *IEEE Trans. Signal Processing*, 50(10): 2408–2416, October 2002.

[SFGK00]   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.*, 48(3):502–513, March 2000.

[SHP01]   S. Sandhu, R. Heath, and A. Paulraj. Space–time block codes versus space–time trellis codes. In *Proc. ICC 2001 – IEEE Int. Conf. Commun.*, volume 4, pages 1132–1136, Helsinki, Finland, June 2001.

[Sim01]   M.K. Simon. Evaluation of average bit error probability for space–time coding based on a simpler exact evaluation of pairwise error probability. *Journal of Communications and Networks*, 3(3):257–264, September 2001.

[SJ03]   M. Skoglund and G. Jongren. On the capacity of a multiple-antenna communication link with channel side information. *IEEE J. Select. Areas Commun.*, 21(3):395–405, April 2003.

[SJJS00]   Q.H. Spencer, B.D. Jeffs, M.A. Jensen, and A.L. Swindlehurst. Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel. *IEEE J. Select. Areas Commun.*, 18(3):347–360, March 2000.

[SL01]   P. Stoica and E. Lindskog. Space–time block coding for channels with intersymbol interference. In *Proc. Asilomar Conf. On Signals, Systems, and Computers*, volume 1, pages 252–256, Pacific Grove, CA, November 2001.

[SL03]   H. Shin and J.H. Lee. Capacity of multiple-antenna channels: Spatial fading correlation, double scattering and keyhole. *IEEE Trans. Inform. Theory*, 49(10):2636–2647, October 2003.

[SM03]   S. Simon and A. Moustakas. Optimizing MIMO antenna systems with channel covariance feedback. *IEEE J. Select. Areas Commun.*, 21(3):406–417, April 2003.

[SMB01]   M. Steinbauer, A.F. Molisch, and E. Bonek. The double-directional radio channel. *IEEE Antennas Propagat. Mag.*, 43(4):51–63, August 2001.

[SP02]   H. Sampath and A.J. Paulraj. Linear precoding for space–time coded systems with known fading correlations. *IEEE Commun. Lett.*, 6(6): 239–241, June 2002.

[SP03]   N. Sharma and C.B. Papadias. Improved quasi-orthogonal codes through constellation rotation. *IEEE Trans. Commun.*, 51(3):332–335, March 2003.

[SSB$^+$02]   A. Scaglione, P. Stoica, S. Barbarossa, G.B. Giannakis, and H. Sampath. Optimal designs for space–time linear precoders and decoders. *IEEE Trans. Signal Processing*, 50(5):1051–1064, May 2002.

[SSL04]   W. Su, Z. Safar, and K.J.R. Liu. Diversity analysis of space–time modulation over time-correlated Rayleigh fading channels. *IEEE Trans. Inform. Theory*, 50(8):1832–1840, August 2004.

[SSOL03]   W. Su, Z. Safar, M. Olfat, and K.J.R. Liu. Obtaining full-diversity space-frequency codes from space–time codes via mapping. *IEEE Trans. Signal Processing*, 51(11):2905–2916, November 2003.

[SSP01]   H. Sampath, P. Stoica, and A. Paulraj. Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion. *IEEE Trans. Commun.*, 49(12):2198–2206, December 2001.

[SSRS03]   B.A. Sethuraman, B. Sundar Rajan, and V. Shashidhar. Full-diversity, high-rate, space–time block codes from division algebras. *IEEE Trans. Inform. Theory*, 49(10):2596–2616, October 2003.

[SV87]   A.M. Saleh and R.A. Valenzuela. A statistical model for indoor multipath propagation. *IEEE J. Select. Areas Commun.*, 5(2):128–137, March 1987.

[SW94]   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. *Int. J. Wireless Information Networks*, 1:49–60, 1994.

[SX04]   W. Su and X. Xia. Signal constellations for quasi-orthogonal space–time block codes with full diversity. *IEEE Trans. Inform. Theory*, 50(10):2331–2347, October 2004.

[SZM+06]   M. Shafi, M. Zhang, A.L. Moustakas, P.J. Smith, A.F. Molisch, F. Tufvesson, and S.H. Simon. Polarized MIMO channels in 3-D: models, measurements and mutual information. *IEEE J. Select. Areas Commun.*, 24(3):514–527, March 2006.

[TB02]   G. Taricco and E. Biglieri. Exact pairwise error probability of space–time codes. *IEEE Trans. Inform. Theory*, 48(2):510–513, February 2002.

[TBH00]   O. Tirkkonen, A. Boariu, and A. Hottinen. Minimal non-orthogonality rate 1 space–time block code for 3+ Tx antennas. In *IEEE 6th Int. Symp. on Spread-Spectrum Tech. and Appl. (ISSSTA 2000)*, pages 429–432, September 2000.

[Tel95]   E. Telatar. Capacity of multiantenna Gaussian channels. *Tech. Rep., AT&T Bell Labs.*, 1995.

[Tel99]   E. Telatar. Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecomm.*, 10(6):585–596, November 1999.

[TJC99]   V. Tarokh, H. Jafarkhani, and A.R. Calderbank. Space–time block codes from orthogonal designs. *IEEE Trans. Inform. Theory*, 45(7):1456–1467, July 1999.

[TS04]   J. Tan and G.L. Stuber. Multicarrier delay diversity modulation for MIMO systems. *IEEE Trans. Wireless Commun.*, 3(5):1756–1763, September 2004.

[TSC98]   V. Tarokh, N. Seshadri, and A.R. Calderbank. Space–time codes for high data rate wireless communication: Performance criterion and code construction. *IEEE Trans. Inform. Theory*, 44(3):744–765, March 1998.

[Tur60]   G.L. Turin. The characteristic function of Hermitian quadratic forms in complex normal random variables. *Biometrika*, pages 199–201, June 1960.

[TV04]   S. Tavildar and P. Viswanath. Permutation codes: achieving the diversity-multiplexing tradeoff. In *Proc. ISIT 2004 – IEEE Int. Symp. Information Theory*, page 98, Chicago, June 2004.

[TV05]   D. Tse and P. Viswanath. *Fundamentals of wireless communication*. Cambridge University Press, Cambridge, UK, 2005.

[TV06]   S. Tavildar and P. Viswanath. Approximately universal codes over slow fading channels. *IEEE Trans. Inform. Theory*, 52(7):3233–3258, July 2006.

[UG01]   M. Uysal and C.N. Georghiades. Effect of spatial fading correlation on performance of space–time codes. *Elect. Lett.*, 37(3):181–183, February 2001.

[UG04]   M. Uysal and C.N. Georghiades. On the error performance analysis of space–time trellis codes. *IEEE Trans. Wireless Commun.*, 3(4):1118–1123, July 2004.

[VB99]   E. Viterbo and J. Boutros. A universal lattice code decoder for fading channels. *IEEE Trans. Inform. Theory*, 45(5):1639–1642, July 1999.

[VG97]   M.K. Varanasi and T. Guess. Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel. In *Proc. Asilomar Conf. On Signals, Systems, and Computers*, Monterey, CA, November 1997.

[VHU02]   I. Viering, H. Hofstetter, and W. Utschick. Validity of spatial covariance matrices over time and frequency. In *Proc. Globecom 2002 – IEEE Global Telecommunications Conf.*, 2002.

[VP05]   M. Vu and A.J. Paulraj. Capacity optimization for Rician correlated MIMO wireless channels. In *Proc. Asilomar Conf. On Signals, Systems, and Computers*, Pacific Grove, CA, November 2005.

[VP06]   M. Vu and A. Paulraj. Optimal linear precoders for mimo wireless correlated channels with nonzero mean in space–time coded systems. *IEEE Trans. Signal Processing*, 54(6):2318–2332, June 2006.

[Vu06]   M. Vu. *Exploiting transmit channel side information in MIMO wireless systems*. PhD thesis, Stanford University, 2006.

[WBS+03]   M.Z. Win, N.C. Beaulieu, L.A. Shepp, B.F. Logan, and J.H. Winters. On the SNR penalty of MPSK with hybrid selection/maximal ratio combining over i.i.d. Rayleigh fading channels. *IEEE Trans. Commun.*, 51(6):1012–1023, June 2003.

[Wei03]   W. Weichselberger. *Spatial Structure of Multiple Antenna Radio Channels*. PhD thesis, Technische Universitat Wien, December 2003.

[WHOB06]   W. Weichselberger, M. Herdin, H. Oezcelik, and E. Bonek. A stochastic MIMO channel model with joint correlation of both link ends. *IEEE Trans. Wireless Commun.*, 5(1): 90–100, January 2006.

[Win84]   J.H. Winters. Optimum combining in digital mobile radio with cochannel interference. *IEEE J. Select. Areas Commun.*, 2:528–539, August 1984.

[Wit93]   A. Wittneben. A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation. In *Proc. ICC 1993 – IEEE Int. Conf. Commun.*, pages 1630–1633, 1993.

[WJ02]   J.W. Wallace and M.A. Jensen. Modeling the indoor MIMO wireless channel. *IEEE Trans. Antennas Propagat.*, 50(5):591–599, May 2002.

[WJ04]   J.W. Wallace and M.A. Jensen. Mutual coupling in MIMO wireless systems: A rigorous network theory analysis. *IEEE Trans. Wireless Commun.*, 3(4):1317–1325, July 2004.

[WSZ04]   X. Wang, Y.R. Shayan, and M. Zeng. On the code and interleaver design of broadband OFDM systems. *IEEE Commun. Lett.*, 8(11):653–655, November 2004.

[WW99]   M.Z. Win and J.H. Winters. Analysis of hybrid selection/maximal ratio combining in Rayleigh fading. *IEEE Trans. Commun.*, 47(12):1773–1776, December 1999.

[WX05]   D. Wang and X.G. Xia. Optimal diversity product rotations for quasi-orthogonal STBC with MPSK symbols. *IEEE Commun. Lett.*, 9(5):420–422, May 2005.

[XG04]   P. Xia and G.B. Giannakis. Design and analysis of transmit-beamforming based on limited-rate feedback. In *Proc. VTC 2004 Fall – IEEE 54th Vehicular Technology Conf.*, volume 3, pages 1653–1657, Los Angeles, CA, September 2004.

[XL05]   L. Xian and H. Liu. Optimal rotation angles for quasi-orthogonal space–time codes with PSK modulation. *IEEE Commun. Lett.*, 9(8):676–678, August 2005.

[Yac93]   M.D. Yacoub. *Foundation of mobile radio engineering*. CRC Press, Boca Raton, FL, 1993.

[YBO+01]   K. Yu, M. Bengtsson, B. Ottersten, D.P. McNamara, P. Karlsson, and M.A. Beach. Second order statistics of NLOS indoor MIMO channels based on 5.2 GHz measurements. In *Proc. Globecom 2001 – IEEE Global Telecommunications Conf.*, volume 1, pages 156–160, San Antonio, TX, USA, November 2001.

[YBO+04]   K. Yu, M. Bengtsson, B. Ottersten, D.P. McNamara, P. Karlsson, and M.A. Beach. Modeling of wideband MIMO radio channels based on NLOS indoor measurements. *IEEE Trans. Veh. Technol.*, 53(3):655–665, May 2004.

[YW03]   H. Yao and G.W. Wornell. Structured space–time block codes with optimal diversity-multiplexing tradeoff and minimum delay. In *Proc. Globecom 2003 – IEEE Global Telecommunications Conf.*, volume 4, pages 1941–1945, San Francisco, CA, December 2003.

[ZG01]   S. Zhou and G. Giannakis. Space–time coding with maximum diversity gains over frequency selective fading channels. *IEEE Signal Processing Lett.*, 8(10):269–272, October 2001.

[ZG03]   S. Zhou and G.B. Giannakis. Optimal transmitter eigen-beamforming and space–time block coding based on channel correlations. *IEEE Trans. Inform. Theory*, 49(7):1673–1690, July 2003.

[ZT03]   L. Zheng and D. Tse. Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels. *IEEE Trans. Inform. Theory*, 49(5):1073–1096, May 2003.

[ZWG05]   S. Zhou, Z. Wang, and G.B. Giannakis. Quantifying the power loss when transmit beamforming relies on finite-rate feedback. *IEEE Trans. Wireless Commun.*, 4(4):1948–1957, July 2005.

This page intentionally left blank

Index