UNIVERSITAT POLITÈCNICA DE CATALUNYA

Departament de Teoria del Senyal i Comunicacions

Ph.D. Dissertation

A Unified Framework for Communications through MIMO Channels

Author: **Daniel Pérez Palomar**
Advisor: **Miguel Angel Lagunas Hernández**

Barcelona, May 2003

This work has been awarded with the 2002/03 Rosina Ribalta first prize for the best Doctoral Thesis project within the areas of Information Technologies and Communications by the Epson Foundation.

Ph.D. committee:

Gregori Vázquez	Technical University of Catalonia - UPC
Ana Pérez-Neira	Technical University of Catalonia - UPC
Sergio Barbarossa	University of Rome “La Sapienza”
Mats Bengtsson	Royal Institute of Technology - KTH
Emilio Sanvicente	Technical University of Catalonia - UPC

Qualification: Cum Laude

Abstract

MULTIPLE-INPUT MULTIPLE-OUTPUT (MIMO) CHANNELS constitute a unified way of modeling a wide range of different physical communication channels, which can then be handled with a compact and elegant vector-matrix notation. The two paradigmatic examples are wireless multi-antenna channels and wireline Digital Subscriber Line (DSL) channels.

Research in antenna arrays (also known as smart antennas) dates back to the 1960's. However, the use of multiples antennas at both the transmitter and the receiver, which can be naturally modeled as a MIMO channel, has been recently shown to offer a significant potential increase in capacity. DSL has gained popularity as a broadband access technology capable of reliably delivering high data rates over telephone subscriber lines. A DSL system can be modeled as a communication through a MIMO channel by considering all the copper twisted pairs within a binder as a whole rather than treating each twisted pair independently.

This dissertation considers arbitrary MIMO channels regardless of the physical nature of the channels themselves; as a consequence, the obtained results apply to any communication system that can be modeled as such. After an extensive overview of MIMO channels, both fundamental limits and practical communication aspects of such channels are considered.

First, the fundamental limits of MIMO channels are studied. An information-theoretic approach is taken to obtain different notions of capacity as a function of the degree of channel knowledge for both single-user and multiuser scenarios. Specifically, a game-theoretic framework is adopted to obtain robust solutions under channel uncertainty.

Then, practical communication schemes for MIMO channels are derived for the single-user case or, more exactly, for point-to-point communications (either single-user or multiuser when coordination is possible at both sides of the link). In particular, a joint design of the transmit-receive linear processing (or beamforming) is obtained (assuming a perfect channel knowledge) for systems with either a power constraint or Quality of Service (QoS) constraints.

For power-constrained systems, a variety of measures of quality can be defined to optimize the performance. For this purpose, a novel unified framework that generalizes the existing results in the literature is developed based on majorization theory. In particular, the optimal structure of the transmitter and receiver is obtained for a wide family of objective functions that

can be used to measure the quality of a communication system. Using this unified framework, the original complicated nonconvex problem with matrix-valued variables simplifies into a much simpler convex problem with scalar variables. With such a simplification, the design problem can be then reformulated within the powerful framework of convex optimization theory, in which a great number of interesting design criteria can be easily accommodated and efficiently solved even though closed-form expressions may not exist. Among other results, a closed-form expression for optimum beamforming in terms of minimum average bit error rate (BER) is obtained. For other design criteria, either closed-form solutions are given or practical algorithms are derived within the framework of convex optimization theory.

For QoS-constrained systems, although the original problem is a complicated nonconvex problem with matrix-valued variables, with the aid of majorization theory, the problem is reformulated as a simple convex optimization problem with scalar variables. An efficient multi-level water-filling algorithm is given to optimally solve the problem in practice.

Finally, the more realistic situation of imperfect channel knowledge due to channel estimation errors is considered. The previous results on the joint transmit-receive design for MIMO channels are then extended in this sense to obtain robust designs.

A mis padres,

Acknowledgements

I FEEL EXTREMELY LUCKY to have met all the people I have met during the intriguing and appealing journey of my Ph.D.

Foremost, I would like to thank my advisor Prof. Miguel Angel Lagunas for his guidance and support over the past four years. His office door has always been open, both literally and metaphorically, and I could always talk to him no matter how busy he was. I can only feel flattered for the confidence he has always shown in me. Fortunately, I have benefited from his extraordinary motivation, great intuition, and technical insight. I just hope my thinking and working attitudes have been shaped according to such outstanding qualities.

I wish to thank my mentor at Stanford University, Prof. John M. Cioffi, for he welcomed me into his research group and treated me like another member of the group with no distinction, making my stay at Stanford a pleasant and rewarding experience. He has always shown confidence in me and has given me a continuous support which I appreciate. Very special and deep thanks go to Joice DeBolt for being such a great human being always ready to listen and give advice.

My thanks also extend to Prof. Javier R. Fonollosa, for he guided my first steps as a Ph.D. student and with whom I have had the pleasure of working in several European projects. I want to thank him for the support he has always given to me.

I will always be indebted to the people who have boldly reviewed this dissertation: Toni Pascual, Gonzalo Seco, Xavi Mestre, and Carlos Aldana. I appreciate very much the enormous help of the people who, at some time or another, have reviewed some of my papers (in chronological order): Dana Brooks, Laurent Schumacher, Carlos Aldana, Jérôme Louveaux, Wonjong Rhee, Wei Yu, Young-Han Kim, Jonathan Levin, Mats Bengtsson, George Ginis, and Jeannie Lee Fang. I would also like to thank the people with whom I have had the pleasure to collaborate: John M. Cioffi, Javier R. Fonollosa, Ana Pérez Neira, Montse Nájar, Toni Pascual, and Diego Bartolomé.

Special thanks go to some people who I have encountered during my Ph.D. journey: Gonzalo Seco for introducing me to the wonderful world of matrix theory at my early stage as a Ph.D. student, Wei Yu and Wonjong Rhee for the encouraging and enthusiastic “wireless and convex” discussions, and Mats Bengtsson for sharing with me the joy for convex optimization theory and for his always interesting comments on my work.

This long journey would have been nothing without all the people I have met. I want to thank all my colleagues at UPC: Pau Fernández, Rubén Villarino, Albert Guillén i Fàbregas, Christian Ibars, Xavi Mestre, Gonzalo Seco, Roger Gaspa, Toni Pascual, Diego Bartolomé, René Jativa, Alejandro Ramírez, Joel Solé, Carles Fernández, Jaume Padrell, Joan Bas, Toni Castro, Maribel Madueño, Helenca Duxans, Mónica Caballero, Javier Ruiz Hidalgo, Hugo Durney, Pedro Correa, Xavi Villares, Jose Antonio Lopez, Francesc Rey, Pere Pujol, Pablo Agüero, Ami Wiesel, Ali Nassar, Camilo Chang, Carine Simon, Alex Heldring, Eduard Úbeda, Juan Rendón, Pau Pachès, Gustavo Hernández Ábrego, Carles Antón, Lluis Garrido, Sílvia Pujalte, Ramon Morros, Alberto Sesma, David Font, David Urbano, and Carlos López. I also want to thank all the members and visitors of Cioffi's diverse and friendly group at Stanford University: Amal Ekbal, Ana García Armada, Ardavan Maleki-Tehrani, Atul Salvekar, Avneesh Agarwal, Carlos Aldana, Chaohuang (Steve) Zeng, Christian Ibars, Dimitrios-Alexandros Toumpakaris, George Ginis, Ghazi A. Al-rawi, Kazutomo Hasegawa (Hase), Jeannie Lee Fang, Jérôme Louveaux, José Tellado, Jungwon Lee, Kee-Bong Song, Marie-Laure Boucheret, Maxime Flament, Min Chuin Hoo (Louise), Norihito Mihota, Seong Taek Chung, Susan C. Lin, Wei Yu, Wonjong Rhee, and Prof. Yeheskel Bar-Ness. I could not possibly forget my good friends at Stanford: David Rebollo Monedero, Kevin Yarritu, Christian Ibars Casas, Ana García Armada, Gustavo Hernández Ábrego, and Isabelle da Piedade. I also want to thank all my colleagues at CTTC: Alex Graell, Ana Pérez Neira, Antonio Mollfulleda, Carles Antón, Carlos Bader, Carolina Pinart, Diego Bartolomé, Herman Ballarín, Jordi Cebrià, Jordi Mateu, Jordi Sorribes, José López, Ju Liu, Marc Realp, Marta Vallejo, Miquel Payaró, Mònica Navarro, Patricia Gallart, Raül Muñoz, Ricardo Martínez, Toni Morell, Toni Pascual, and Xavi Mestre. I could not forget to thank my good friends Xavi Martínez and Oscar González, as well as my biking partner Rodolfo Guichón. Please forgive me if your name should have been listed above and is missing.

Last, but not least, I would like to thank my parents, Laura, and my whole family for their unconditional love and support.

Daniel Pérez Palomar
April 2003

This work has been supported by the Catalan Government (DURSI) 1999FI 00588.

Contents

Notation

Boldface upper-case letters denote matrices, boldface lower-case letters denote column vectors, and italics denote scalars.

$\mathbf{X}^T,\ \mathbf{X}^*,\ \mathbf{X}^H$	Transpose, complex conjugate, and conjugate transpose (Hermitian) of matrix $\mathbf{X}$, respectively.
$(\cdot)^\star$	Optimal value.
$\mathbf{X}^{1/2}$	Hermitian square root of the Hermitian matrix $\mathbf{X}$, *i.e.*, $\mathbf{X}^{1/2}\mathbf{X}^{1/2} = \mathbf{X}$.
$\mathbf{P}_x$, $\mathbf{P}_x^\perp$	Projection matrix onto the subspace spanned by the columns of $\mathbf{X}$ and the orthogonal subspace, respectively.
$\mathrm{Tr}\left(\mathbf{X}\right)$	Trace of $\mathbf{X}$.
$\lvert\mathbf{X}\rvert$ or $\det\left(\mathbf{X}\right)$	Determinant of matrix $\mathbf{X}$.
$\lvert x\rvert$	Absolute value (modulus) of the scalar x.
$\lVert\cdot\rVert$	A norm.
$\lVert\mathbf{x}\rVert_2$	Euclidean norm of vector $\mathbf{x}$: $\lVert\mathbf{x}\rVert_2 = \sqrt{\mathbf{x}^H\mathbf{x}}$.
$\lVert\mathbf{X}\rVert_F$	Frobenius norm of matrix $\mathbf{X}$: $\lVert\mathbf{X}\rVert_F = \sqrt{\mathrm{Tr}\left(\mathbf{X}^H\mathbf{X}\right)}$.
$\mathbf{d}\left(\mathbf{X}\right)$	Vector of diagonal elements of matrix $\mathbf{X}$.
$\boldsymbol{\lambda}\left(\mathbf{X}\right)$	Vector of eigenvalues of matrix $\mathbf{X}$.
$\lambda_{\max}\left(\mathbf{X}\right)$, $\lambda_i\left(\mathbf{X}\right)$	Maximum eigenvalue and ith eigenvalue (in increasing or decreasing order), respectively, of matrix $\mathbf{X}$.
$\mathbf{u}_{\max}\left(\mathbf{X}\right)$, $\mathbf{u}_i\left(\mathbf{X}\right)$	Eigenvector associated to the maximum and to the ith eigenvalue (in increasing or decreasing order), respectively, of matrix $\mathbf{X}$.
$[\mathbf{X}]_{i,j}$ or $[\mathbf{X}]_{ij}$	The (ith,jth) element of matrix $\mathbf{X}$.
$[\mathbf{X}]_{:,j}$	The jth column of matrix $\mathbf{X}$.
$\mathrm{diag}\left(\{\mathbf{D}_k\}\right)$	Block-diagonal matrix with diagonal blocks given by the set $\{\mathbf{D}_k\}$. In particular, if the $\mathbf{D}_k$'s are scalars, it reduces to a diagonal matrix.

$\text{vec}(\cdot)$	Vec-operator: if $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_n]$, then $\text{vec}(\mathbf{X}) = \left[\mathbf{x}_1^T \cdots \mathbf{x}_n^T\right]^T$.
$\mathbf{I}$	Identity matrix. A subscript can be used to indicate the dimension.
$\mathbf{e}_i$	Canonical vector with all the elements being zero except the ith one which is equal to one.
$\mathbf{A} \geq \mathbf{B}$	$\mathbf{A} - \mathbf{B}$ is positive semidefinite.
$\mathbf{a} \geq \mathbf{b}$	Elementwise relation $a_i \geq b_i$.
$\mathbf{a} \succ \mathbf{b}$	$\mathbf{a}$ majorizes $\mathbf{b}$ or, equivalently, $\mathbf{b}$ is majorized by $\mathbf{a}$.
$\mathbf{a} \succ^w \mathbf{b}$	$\mathbf{a}$ weakly majorizes $\mathbf{b}$ or, equivalently, $\mathbf{b}$ is weakly majorized by $\mathbf{a}$.
$\triangleq$	Defined as.
$\propto$	Equal up to a scaling factor.
δ_{kl}	Kronecker delta: $\delta_{kl} = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$.
$[a, b]$, (a, b)	Closed interval ($a \leq x \leq b$) and open interval ($a < x < b$), respectively.
$[a, b)$, $(a, b]$	Half-closed (or half-open) intervals $a \leq x < b$ and $a < x \leq b$, respectively.
$\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_{++}$	The set of real, nonnegative real, and positive real numbers, respectively.
$\mathbb{R}^{n\times m}$, $\mathbb{C}^{n\times m}$	The set of $n \times m$ matrices with real- and complex-valued entries, respectively.
$\mathbf{S}^n$	The set of Hermitian $n \times n$ matrices $\mathbf{S}^n \triangleq \left\{\mathbf{X} \in \mathbb{C}^{n\times n} \mid \mathbf{X} = \mathbf{X}^H\right\}$.
$\mathbf{S}_+^n$	The set of Hermitian positive semidefinite $n \times n$ matrices $\mathbf{S}_+^n \triangleq \left\{\mathbf{X} \in \mathbb{C}^{n\times n} \mid \mathbf{X} = \mathbf{X}^H \geq \mathbf{0}\right\}$.
$\mathbf{S}_{++}^n$	The set of Hermitian positive definite $n \times n$ matrices $\mathbf{S}_{++}^n \triangleq \left\{\mathbf{X} \in \mathbb{C}^{n\times n} \mid \mathbf{X} = \mathbf{X}^H > \mathbf{0}\right\}$.
$\mathbb{E}[\cdot]$	Statistical expectation. A subscript can be used to indicate the random variable considered for the expectation.
$\sim$	Distributed according to.
$\mathcal{CN}(\mathbf{m}, \mathbf{C})$	Complex circularly symmetric Gaussian vector distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$.[1]
$\log(\cdot)$	Natural logarithm.
$\log_b(\cdot)$	Logarithm in base b.
$\text{Re}[\cdot]$, $\text{Im}[\cdot]$	Real and imaginary parts.

[1] A complex Gaussian random vector $\mathbf{z} = \mathbf{x} + j\mathbf{y}$ is circularly symmetric (also termed proper Gaussian) if $\mathbb{E}\left[\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \begin{bmatrix} \mathbf{x}^T & \mathbf{y}^T \end{bmatrix}\right] = \frac{1}{2}\begin{bmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix}$ so that $\mathbb{E}\left[\mathbf{z}\mathbf{z}^H\right] = \mathbf{A} + j\mathbf{B}$ [Nee93].

sup, inf	Supremum (lowest upper bound) and infimum (highest lower bound).
$\bigcap, \bigcup$	Intersection and union.
$(x)^{+}$	Positive part of x, *i.e.*, $\max(0, x)$. For matrices it is defined elementwise.
$g'(a)$	Derivative of function $g(x)$ evaluated at $x = a$.
$\nabla_{\mathbf{x}} f(\mathbf{x})$	Gradient of function $f(\mathbf{x})$ with respect to $\mathbf{x}$.[2]
$\text{dom}\, f$	Domain of function f.

[2] If $\mathbf{x}$ is a complex-valued vector and function $f(\mathbf{x})$ is not analytic, the well-known definition of the complex gradient operator is used, since it is very convenient, among other things, to determine the stationary points of a real-valued scalar function of a complex vector [Bra83].

Acronyms

1-D	One-dimensional.
3GPP	Third Generation Partnership Project.
A/D	Analog-to-Digital.
ADSL	Asymmetric DSL.
AWGN	Additive White Gaussian Noise.
BER	Bit Error Rate.
BLAST	Bell-labs LAyered Space-Time.
bps	Bits per second.
BPSK	Binary Phase Shift Keying.
CDMA	Code Division Multiple Access.
CO	Central Office.
COFDM	Coded Orthogonal Frequency Division Multiplexing.
CP	Cyclic Prefix.
CPE	Customer Premises Equipment.
CSI	Channel State Information.
CSIR	Channel State Information at the Receiver.
CSIT	Channel State Information at the Transmitter.
D/A	Digital-to-Analog.
DAB	Digital Audio Broadcasting.
DF	Decision-Feedback.
DFE	Decision-Feedback Equalizer.

DFT Discrete Fourier Transform.

DLST Diagonal Layered Space-Time.

DMT Discrete Multi-Tone.

DSL Digital Subscriber Line.

DVB Digital Video Broadcasting.

ETSI European Telecommunications Standards Institute.

EVD Eigenvalue Decomposition.

FDD Frequency Division Duplex.

FDMA Frequency Division Multiple Access.

FEXT Far-End Crosstalk.

FFT Fast Fourier Transform.

FIR Finite Impulse Response.

HLST Horizontal Layered Space-Time.

IBI Inter-Block Interference.

IDFT Inverse Discrete Fourier Transform.

IEEE Institute of Electrical and Electronical Engineers.

IFFT Inverse Fast Fourier Transform.

i.i.d. Independent and Identically Distributed.

ISI Inter-Symbol Interference.

KKT Karush-Kuhn-Tucker.

LHS Left-Hand Side.

LMMSE Linear Minimum Mean Square Error.

LP Linear Program.

LST Layered Space-Time.

LTI Linear Time-Invariant.

LTV Linear Time-Varying.

MAC Multiple-Access Channel.

MC-CDMA Multicarrier CDMA.

MIMO Multiple-Input Multiple-Output.

MISO Multiple-Input Single-Output.

ML Maximum Likelihood.

MLSE Maximum Likelihood Sequence Estimator.

MMSE Minimum Mean Square Error.

MSE Mean Square Error.

NEXT Near-End Crosstalk.

NLOS Non-Line-Of-Sight.

OFDM Orthogonal Frequency Division Multiplexing.

OFDMA Orthogonal Frequency Division Multiple Access.

P/S Parallel-to-Serial.

PAR Peak to Average Ratio.

pdf Probability Density Function.

PSD Power Spectral Density.

PSK Phase Shift Keying.

QAM Quadrature Amplitude Modulation.

QoS Quality of Service.

QP Quadratic Program.

QPSK Quadrature Phase Shift Keying.

RHS Right-Hand Side.

r.m.s. Root Mean Squared.

Rx Receiver.

S/P Serial-to-Parallel.

SIMO Single-Input Multiple-Output.

SINR Signal to Interference-plus-Noise Ratio.

SISO Single-Input Single-Output.

SNR Signal to Noise Ratio.

s.t. Subject To.

STBC	Space-Time Block Coding.
STC	Space-Time Coding.
STTC	Space-Time Trellis Coding.
SVD	Singular Value Decomposition.
TDD	Time Division Duplex.
TDMA	Time Division Multiple Access.
Tx	Transmitter.
UMTS	Universal Mobile Telecommunications System.
UTRA	UMTS Terrestrial Radio Access.
VDSL	Very-high-bit-rate DSL.
w/ and **w/o**	With and without, respectively.
WLAN	Wireless Local Area Network.
w.l.o.g.	Without Loss Of Generality.
ZF	Zero-Forcing.
ZP	Zero-Padding.

Chapter 1

Introduction

THE FOCUS OF THIS DISSERTATION is on communications through multiple-input multiple-output (MIMO) channels. The interest of studying MIMO channels is because many different types of real channels can be modeled as such. In other words, they represent a unified way to model a wide variety of scenarios. In addition, MIMO channels can be naturally handled with a convenient and elegant vector-matrix notation. The two paradigmatic examples are wireless multi-antenna channels and wireline channels, although many other typical scenarios are straightforwardly modeled as MIMO channels as well.

1.1 Motivation

1.1.1 Wireless Multi-Antenna Channels

The recent and anticipated growth of wireless communication systems has fueled research efforts investigating methods to increase system capacity. Doubtlessly, the rapid advance in technology, on the one hand, and the exploding demand for efficient high-quality services of digital wireless communications, on the other hand, play a dramatic role in this trend. The demand for these services is growing at an extremely rapid pace and these trends are likely to continue for several years.

The radio spectrum available for wireless services is extremely scarce. As a consequence, a prime issue in current wireless systems is the conflict between the increasing demand for wireless services and the scarce electromagnetic spectrum. Spectral efficiency is therefore of primary concern in the design of future wireless data communication systems with the omnipresent bandwidth constraint.

The current need for increased capacity and interference protection in wireless multiuser systems is at present treated through, among other techniques, limited microdiversity features

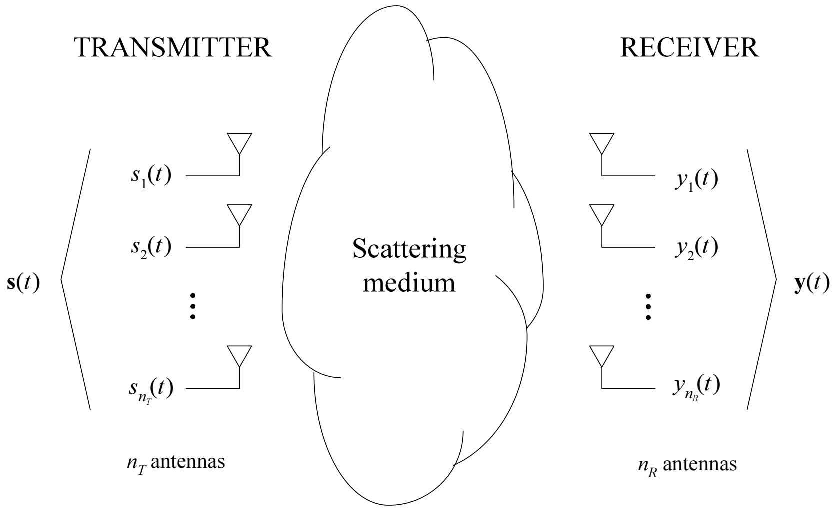


Figure 1.1: Wireless multi-antenna channel.

at cell sites, sectorization, and switched multibeam schemes. These techniques fall into the category of exploiting the spatial characteristics of wireless channels. Slowly but steadily, more sophisticated fully adaptive antenna arrays are being considered instead as a cost-effective higher-performance solution for the base station.

Research in antenna arrays (also known as *smart antennas*) as a means to generate spatial diversity, while dating back to the sixties, has seen a boom in activity in recent years due to digital processing and the real commercial need for higher capacity in wireless systems [Vee88, Kri96]. The use of multiple antennas at the receiver can significantly increase the channel capacity by exploiting the spatial diversity, for example, to combat fading and to perform interference cancellation. If simultaneous spatial diversity is employed both at the transmitter and the receiver (see Figure 1.1), then a MIMO channel naturally arises with the additional property that several substreams can be opened up for communication within the MIMO channel (this is the so-called multiplexing gain). This particular scenario has gained a significant popularity due to recent studies indicating a linear increase in capacity with the number of antennas [Tel95, Fos98]. As a consequence of the potential of multi-antenna wireless channels, an extraordinary number of publications have flourished in the open literature proposing a variety of efficient communication schemes such as spatio-temporal vector coding techniques [Ral98], space-time coding [Tar98], and layered architectures [Fos96].

1.1.2 Wireline DSL Channels

Digital Subscriber Line (DSL) has gained popularity as a broadband access technology capable of reliably delivering high data rates over telephone subscriber lines [Sta99]. It provides a solution to the *last-mile access* by utilizing the already existing copper twisted pair infrastructure, which

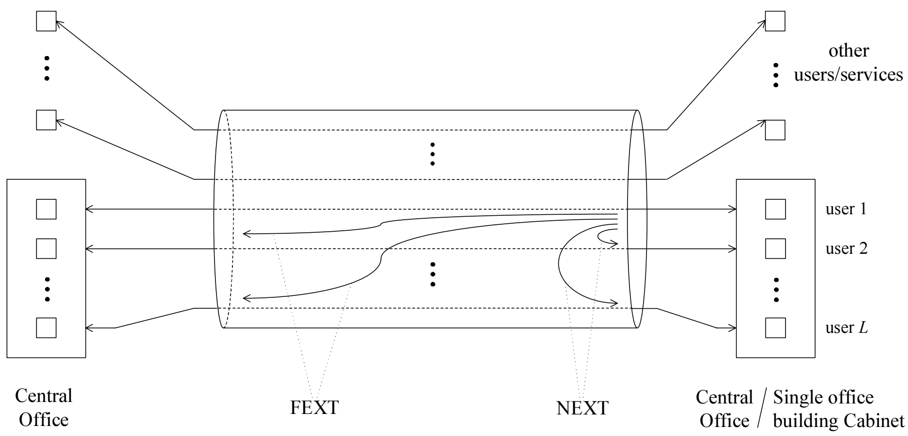


Figure 1.2: Wireline DSL channel (scheme of a bundle of copper twisted pairs).

was originally built with the purpose of providing telephone service. Asymmetric DSL (ADSL) systems have been successfully deployed, revealing the potential of this technology. Current efforts focus on very-high-bit-rate DSL (VDSL) which allows the use of a bandwidth up to 20 MHz.

The dominant impairment in DSL systems is crosstalk arising from electromagnetic coupling between neighboring twisted pairs. Near-end crosstalk (NEXT) comprises the signals originated in the same side of the received signal (due to the existence of downstream and upstream transmission) and far-end crosstalk (FEXT) includes the signal originated in the opposite side of the received signal. The impact of NEXT is generally suppressed by employing frequency division duplex (FDD) to separate downstream and upstream transmission. In addition to the crosstalk, DSL channels are highly frequency-selective; as a consequence, multicarrier transmission schemes are used in practice.

Modeling a DSL system as a MIMO channel presents many advantages with respect to treating each twisted pair independently [Hon90, Gin02]. In fact, modeling a wireline channel as a MIMO channel was done almost three decades ago [Lee76, Sal85]. A general scenario consists of a binder group composed of a set of intended users in the same physical location plus some other users that possibly belong to a different service provider and use different types of DSL systems (see Figure 1.2). The MIMO channel represents the communication of the intended users while the others are treated as interference.

In many situations, joint processing can be assumed at one side of the link, the Central Office (CO), whereas the other side, corresponding to the Customer Premises Equipment (CPE), must use independent processing per user since users are geographically distributed [Gin02]. In some cases of practical interest, however, both ends of the MIMO system are each terminated in a single physical location (see Figure 1.2), *e.g.*, links between CO's and Remote Terminals or links between CO's and private networks. This allows the utilization of joint processing at both sides of the link [Hon90].

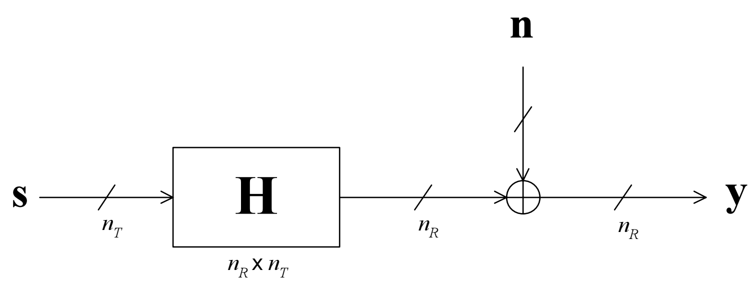


Figure 1.3: Generic MIMO channel that may represent, among others, a multi-antenna wireless channel and a wireline channel.

1.1.3 A Generic Approach: MIMO Channels

In addition to the two previous wireless multi-antenna and wireline DSL systems, there are many other common scenarios that can be naturally modeled as MIMO channels such as wireless single-antenna systems transmitting through time-dispersive channels (properly modeled on a block basis), multicarrier systems on frequency-selective channels, CDMA channels, or systems exploiting the polarization diversity. Recall that MIMO channels can be easily handled using a vector-matrix notation. In Figure 1.3, the communication process through a generic MIMO channel is depicted, where $\mathbf{H}$ represents the channel matrix that accepts multiple inputs and gives multiple outputs. Note that the structure of $\mathbf{H}$ will completely depend on the specific application at hand (in a time-dispersive channel, for example, $\mathbf{H}$ is a convolution matrix which has a very special structure).

Since MIMO channels are a unified way to represent many different real scenarios, one can focus on arbitrary MIMO channels regardless of the specific physical origin. In many cases, of course, it is instructive to step back from this generic formulation and study the details and particularities of the specific channel model under consideration.

There are two aspects that should be considered for investigation: fundamental limits and practical communication techniques. First, it is important to quantify the fundamental limits of MIMO channels. Then, practical communication schemes must be obtained aiming at achieving rates close to the limits of the channel. A plethora of communication techniques suitable for MIMO channels has flourished (and is still flourishing) in the literature. A variety of methods exists tailored to different types of channel characteristics and degree of knowledge of the actual channel realization (some of them are even being successfully tested in realistic environments). Nevertheless, new transmit-receive processing schemes improving upon the existing ones are still to come and, therefore, should be investigated.

1.2 Outline of Dissertation

In general terms, the focus of this dissertation is on the joint design of the transmitter and receiver for communications through arbitrary MIMO channels. The fundamental limits of MIMO channels are explored as well. The outline of each of the chapter is as follows.

Chapter 1, the present chapter, gives the motivation, outline, and contributions of this dissertation.

Chapter 2 overviews MIMO channels which arise in many different scenarios such as wireless multi-antenna systems or wireline DSL systems. In fact, there is a significant variety of situations that can be modeled as such. A MIMO channel is conveniently represented by a channel matrix which provides an elegant, compact, and unified way to deal with physical channels of completely different nature.

Chapter 3 introduces two important theories—convex optimization theory and majorization theory—on which many results of this dissertation are based.

Chapter 4 deals with the fundamental limits of communications through MIMO channels. An infomation-theoretic approach is taken to obtain different notions of capacity for different degrees of channel state information (CSI). Specifically, for the case of no CSI, a game-theoretic approach is taken to obtain robust solutions under channel uncertainty.

Chapter 5 considers communications through MIMO channels with a power constraint and designs transmit-receive beamforming (or linear processing) to optimize the performance of the system under a variety of design criteria. This chapter generalizes all the existing results in the literature by developing a novel unifying framework based on majorization theory that provides the optimal structure of the transmitter and receiver. With such a result, the original complicated nonconvex problem with matrix-valued variables simplifies into a much simpler convex problem with scalar variables. After such a simplification, the design problem can be then reformulated within the powerful framework of convex optimization theory, in which a great number of interesting design criteria can be easily accommodated and efficiently solved even though closed-form expressions may not exist. From this perspective, a wide range of design criteria is analyzed and, in particular, a closed-form expression for optimum beamforming in the sense of minimizing the average bit error rate (BER) is obtained. Efficient algorithms for practical implementation are given for the considered design criteria.

Chapter 6 considers communications through MIMO channels with a set of Quality of Service (QoS) requirements for the simultaneously established substreams. Linear transmit-receive processing (or beamforming) is designed to satisfy the QoS constraints with minimum transmitted power. Although the original problem is a complicated nonconvex problem with matrix-valued variables, with the aid of majorization theory, the problem is reformulated as a simple convex op-

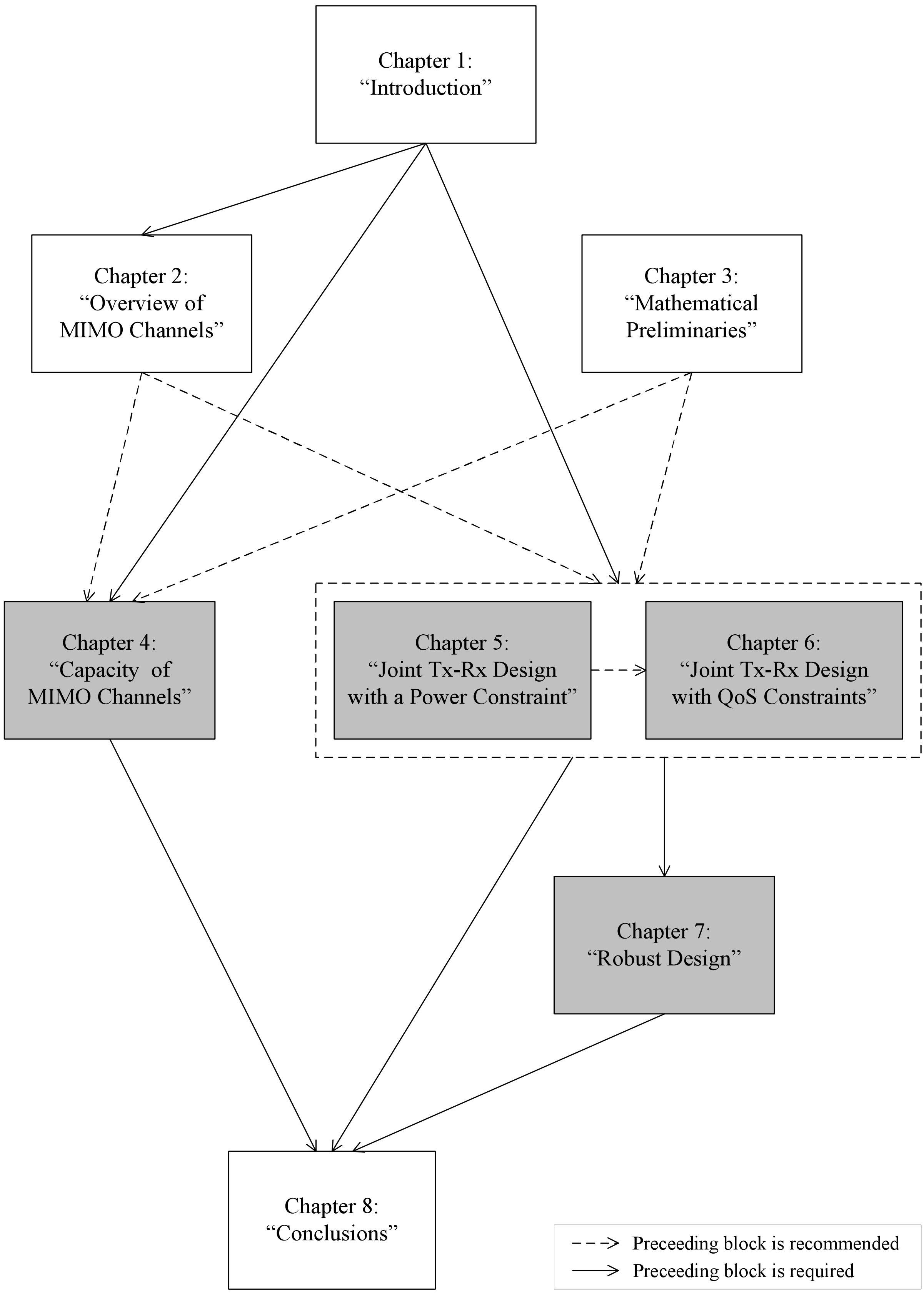


Figure 1.4: Dependence among chapters.

timization problem with scalar variables. An efficient multi-level water-filling algorithm is given to optimally solve the problem in practice.

Chapter 7 extends the results of Chapters 5 and 6, in which perfect CSI was assumed, to the more realistic situation of imperfect CSI accounting for channel estimation errors. Two completely different philosophies are used to obtain robust designs: worst-case robustness and stochastic (Bayesian) robustness.

Chapter 8 concludes the dissertation summarizing the main obtained results and enumerating future lines of work.

The dependence among the chapters is illustrated in Figure 1.4. For example, before reading Chapter 7, one should read first Chapters 5 and 6, for which one should first read Chapter 1 with the recommendation of reading Chapters 2 and 3 as well.

1.3 Research Contributions

The main contribution of this dissertation is the development of a general unified framework for the joint design of transmit-receive linear processing schemes for communications in MIMO channels. Details of the research contributions in each chapter are as follows.

Chapter 4

The main result in this chapter is regarding the game-theoretic formulation of the communication problem published in one journal paper and one conference paper:

- D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform Power Allocation in MIMO Channels: A Game-Theoretic Approach," *IEEE Trans. on Information Theory*, Vol. 49, No. 7, pp. 1707-1727, July 2003.
- D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform Power Allocation in MIMO Channels: A Game-Theoretic Approach," in *Proc. IEEE 2003 International Symposium on Information Theory (ISIT'03)*, p. 271, Pacifico, Yokohama, Japan, June 29-July 4, 2003.

Additional results have also been obtained for beamforming-constrained systems in realistic multi-antenna correlated channels with perfect CSI in one journal paper and four conference papers:

- D. P. Palomar and M. A. Lagunas, "Joint Transmit-Receive Space-Time Equalization in Spatially Correlated MIMO channels: A Beamforming Approach," *IEEE Journal on Selected Areas in Communications: Special Issue on MIMO Systems and Applications*, Vol. 21, No. 5, pp. 730-743, June 2003.

- D. P. Palomar, J. R. Fonollosa, and M. A. Lagunas, "Capacity results on frequency-selective Rayleigh MIMO channels," in *Proc. IST Mobile Communication Summit 2000*, pp. 491-496, Galway, Ireland, Oct. 1-4, 2000.
- D. P. Palomar, J. R. Fonollosa, and M. A. Lagunas, "Information-theoretic results for realistic UMTS MIMO channels," in *Proc. IST Mobile Communication Summit 2001*, Sitges, Barcelona, Spain, Sept. 9-12, 2001.
- D. P. Palomar, J. R. Fonollosa, and M. A. Lagunas, "Capacity results of spatially correlated frequency-selective MIMO channels in UMTS," in *Proc. IEEE Vehicular Technology Conf. Fall (VTC-Fall 2001)*, Atlantic City, NJ, Oct. 7-11, 2001.
- D. P. Palomar and M. A. Lagunas, "Capacity of spatially flattened frequency-selective MIMO channels using linear processing techniques in transmission," in *Proc. 35th IEEE Annual Conference on Information Sciences and Systems (CISS 2001)*, The John Hopkins University, Baltimore, MD, March 21-23, 2001.

Chapter 5

The main results in this chapter involve the joint optimization of the transmitter and receiver according to different criteria under a power constraint, for which a novel unified framework has been developed. The results have been published in two journal papers and three conference papers:

- D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx Beamforming Design for Multicarrier MIMO Channels: a Unified Framework for Convex Optimization," *IEEE Trans. on Signal Processing*, to appear in 2003 (submitted Feb. 2002, revised Dec. 2002).
- D. P. Palomar and M. A. Lagunas, "Joint Transmit-Receive Space-Time Equalization in Spatially Correlated MIMO channels: A Beamforming Approach," *IEEE Journal on Selected Areas in Communications: Special Issue on MIMO Systems and Applications*, Vol. 21, No. 5, pp. 730-743, June 2003.[1]
- D. P. Palomar, M. A. Lagunas, A. P. Iserte, and A. P. Neira "Practical implementation of jointly designed transmit-receive space-time IIR filters," in *Proc. 6th IEEE International Symposium on Signal Processing and its Applications (ISSPA-2001)*, pp. 521-524, Kuala-Lampur, Malaysia, Aug. 13-16, 2001.
- D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "On the Optimal Structure of Transmit-Receive Linear Processing for MIMO Channels," in *Proc. 40th Annual Allerton Conference*

[1] Note that this paper has been previously listed in the publications of the capacity results for beamforming-constrained systems corresponding to Chapter 4.

on Communication, Control, and Computing, pp. 683-692, Allerton House, Monticello, IL, Oct. 2-4, 2002.

- D. P. Palomar, J. M. Cioffi, M. A. Lagunas, and A. P. Iserte, "Convex Optimization Theory Applied to Joint Beamforming Design in Multicarrier MIMO Channels," in *Proc. IEEE 2003 International Conference on Communications (ICC'03)*, Anchorage, Alaska, USA, May 11-15, 2003.

Chapter 6

The main results in this chapter refer to the joint optimization of the transmitter and receiver to satisfy a set of QoS constraints with minimum transmitted power. The results have been published in one journal paper and one conference paper:

- D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum Linear Joint Transmit-Receive Processing for MIMO Channels with QoS Constraints," *IEEE Trans. on Signal Processing*, to appear in 2003 (submitted May 2002, revised Feb. 2003).

- D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum Joint Transmit-Receive Linear Processing for Vectored DSL Transmission with QoS Requirements," in *Proc. 36th Asilomar Conference on Signals, Systems & Computers*, pp. 388-392, Pacific Grove, CA, Nov. 3-6, 2002.

Chapter 7

The results in this chapter extend the previously obtained results for perfect CSI to the more realistic case of imperfect CSI. Partial results were presented in one journal paper and additional results are still to be submitted for publication:

- D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum Linear Joint Transmit-Receive Processing for MIMO Channels with QoS Constraints," *IEEE Trans. on Signal Processing*, to appear in 2003 (submitted May 2002, revised Feb. 2003).[2]

Other contributions not presented in this dissertation

During the first and a half years of the author's Ph.D. period, blind beamforming techniques were developed for spread spectrum systems with multiple receive antennas. The results were published in two journal papers and four conference papers:

[2] Note that this paper has been previously listed in the publications corresponding to Chapter 6.

- D. P. Palomar, M. Nájar, and M. A. Lagunas "Self-reference Spatial Diversity Processing for Spread Spectrum Communications," *AEÜ International Journal of Electronics and Communications*, Vol. 54, No. 5, pp. 267-276, Nov. 2000.

- D. P. Palomar and M. A. Lagunas "Temporal diversity on DS-CDMA communication systems for blind array signal processing," *EURASIP Signal Processing*, Vol. 81, No. 8, pp. 1625-1640, Aug. 2001.

- D. P. Palomar and M. A. Lagunas "Blind beamforming for DS-CDMA systems," in *Proc. of the Fifth Bayona Workshop on Emerging Technologies in Telecommunications*, pp. 83-87, Bayona, Spain, Sept. 6-8, 1999.

- D. P. Palomar, M. A. Lagunas, and M. Nájar, "Self-reference Spatial Diversity Processing for Spread Spectrum Communications," in *Proc. of International Symposium on Image/Video Communications over Fixed and Mobile Networks (ISIVC'2000)*, Invited Presentation, Vol. 1, pp. 81-96, Rabat, Morocco, April 17-20, 2000.

- D. P. Palomar and M. A. Lagunas "Self-reference beamforming for DS-CDMA communication systems," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2000)*, Vol. V, pp.3001-3004, Istanbul, Turkey, June 5-9, 2000.

- D. P. Palomar and M. A. Lagunas "Optimum Self-reference Spatial Diversity Processing for FDSS and FH communication systems," in *Proc. EUSIPCO 2000*, Vol. III, Tampere, Finland, Sept. 4-8, 2000.

Some work was also done in the area of multiuser dectection in CDMA systems with results published in one conference paper:

- D. P. Palomar, J. R. Fonollosa, and M. A. Lagunas, "MMSE Joint Detection in frequency-selective wireless communication channels for DS-CDMA systems," in *Proc. IEEE Sixth International Symposium on Spread Spectrum Techniques & Applications (ISSSTA 2000)*, Vol. 2, pp. 530-534, Parsippany, NJ, Sept. 6-8, 2000.

Chapter 2

MIMO Channels: An Overview

MULTIPLE-INPUT MULTIPLE-OUTPUT (MIMO) CHANNELS arise in many different scenarios such as wireline systems or multi-antenna wireless systems. There is a significant variety of situations that can be modeled as a MIMO system or as a communication through a MIMO channel. A MIMO channel can be represented by a channel matrix which provides an elegant, compact, and unified way to deal with physical channels of completely different nature.

This chapter is organized as follows. After describing the basic MIMO channel model in Section 2.1, a variety of illustrative examples of real communication systems that can be modeled as MIMO channels is given in Section 2.2. To gain insight into MIMO communication systems, Section 2.3 describes the basic characteristics and properties of MIMO channels. Section 2.4 gives an overview of the existing transmission techniques for MIMO channels and Section 2.5 focuses specifically on linear signal processing techniques which is the scope of this dissertation (see Chapters 5 and 6).

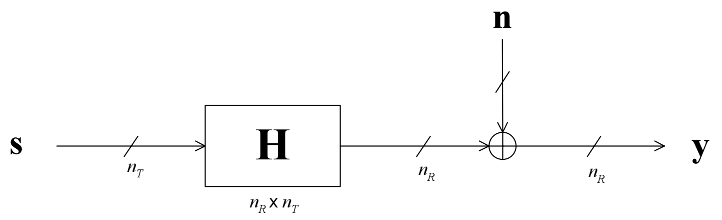
2.1 Basic MIMO Channel Model

The signal model for a MIMO channel with n_T transmit and n_R receive dimensions (see Figure 2.1(a)) is[1]

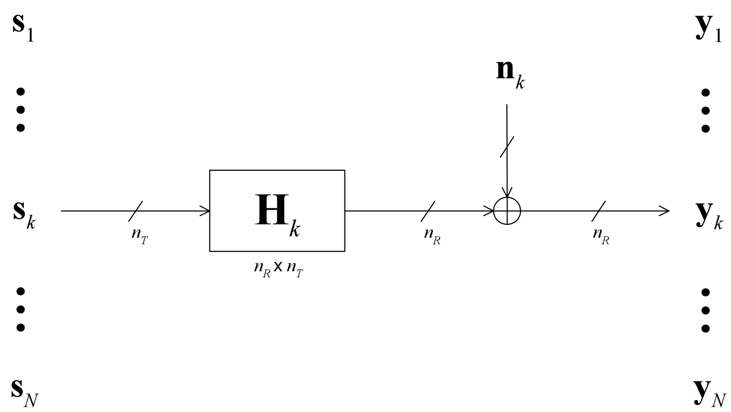
$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \tag{2.1}$$

where $\mathbf{s} \in \mathbb{C}^{n_T \times 1}$ is the transmitted vector, $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ is the channel matrix, $\mathbf{y} \in \mathbb{C}^{n_R \times 1}$ is the received vector, and $\mathbf{n} \in \mathbb{C}^{n_R \times 1}$ is a zero-mean circularly symmetric (also termed *proper* [Nee93]) complex Gaussian noise vector (which can also include other interference Gaussian signals) with arbitrary covariance matrix $\mathbf{R}_n$, *i.e.*, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_n)$. It is sometimes notationally convenient to

[1]Although a general MIMO channel may be nonlinear, we restrict to linear MIMO channels since the physical process of propagation can be accurately modeled as a linear transformation.



(a) Single MIMO channel



(b) Multiple MIMO channels

Figure 2.1: Scheme of a single MIMO channel and of a set of N parallel and independent MIMO channels.

utilize the whitened channel defined as

$$\tilde{\mathbf{H}} \triangleq \mathbf{R}_n^{-1/2}\mathbf{H}. \tag{2.2}$$

Note that the whitened channel is useful when the received signal vector $\mathbf{y}$ is pre-processed with the whitening matrix $\mathbf{R}_n^{-1/2}$ so that the pre-processed received signal vector $\mathbf{R}_n^{-1/2}\mathbf{y} = \tilde{\mathbf{H}}\mathbf{s} + \mathbf{w}$ has a white noise $\mathbf{w}$ with a unitary covariance matrix, *i.e.*, $\mathbb{E}[\mathbf{w}\mathbf{w}^H] = \mathbf{I}$.

The signal model of (2.1) represents a single transmission. A real communication is, of course, composed of multiple transmissions. It suffices to index the signals in (2.1) with a time-discrete index as $\mathbf{y}(n) = \mathbf{H}\mathbf{s}(n) + \mathbf{n}(n)$ (the channel can also be considered time-varying $\mathbf{H}(n)$). For the sake of notation, however, the discrete-time index is not used in the rest of the dissertation.

For the more general case of having a set of N parallel and independent MIMO channels (multiple MIMO channels) with n_T transmit and n_R receive dimensions each[2] (see Figure 2.1

[2]In general, one can consider that each MIMO channel has a different number of transmit and receive dimensions $n_{T,k}$ and $n_{R,k}$. For the sake of notation, however, we consider the same number of dimensions for all MIMO channels. This is without loss of generality since one can always take $n_T = \max_k \{n_{T,k}\}$ and $n_R = \max_k\{n_{R,k}\}$ and then fill in each MIMO channel with zero elements as necessary so that it has dimensions $n_R \times n_T$.

(b)), the signal model is

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{s}_k + \mathbf{n}_k \qquad 1 \leq k \leq N \tag{2.3}$$

where k denotes the channel index and $\mathbf{s}_k$, $\mathbf{H}_k$, $\mathbf{y}_k$, and $\mathbf{n}_k$ are defined as before for each MIMO channel k (noise vectors corresponding to different MIMO channels are considered independent). A natural example of this signal model is for vector transmission over frequency-selective channels using a multicarrier approach (assuming orthogonality among carriers) such as in a wireless multi-antenna system (see §2.2.3) or in a wireline system (see §2.2.4). Another example of multiple MIMO channels is when multiple binders in wireline communications are considered for transmission (assuming no crosstalk among binders).

The multiple MIMO channel model of (2.3) can be expressed as in (2.1) by defining the block-diagonal matrix $\mathbf{H} = \text{diag}(\{\mathbf{H}_k\})$ and stacking the vectors as $\mathbf{s} = \left[\mathbf{s}_1^T \cdots \mathbf{s}_N^T\right]^T$, $\mathbf{y} = \left[\mathbf{y}_1^T \cdots \mathbf{y}_N^T\right]^T$, and $\mathbf{n} = \left[\mathbf{n}_1^T \cdots \mathbf{n}_N^T\right]^T$. Clearly, the model in (2.1) is more general (it can model, for example, a multicarrier system with non-orthogonal carriers or intermodulation terms). However, the model in (2.3) proves useful when each MIMO channel is independently processed as opposed to (2.1) that treats all MIMO channels as a whole (c.f. §2.5.1.2).

When $n_T = 1$, the MIMO channel reduces to a single-input multiple-output (SIMO) channel (*e.g.*, when having multiple antennas only at the receiver). Similarly, when $n_R = 1$, the MIMO channel reduces to a multiple-input single-output (MISO) (*e.g.*, when having multiple antennas only at the transmitter). When both $n_T = 1$ and $n_R = 1$, the MIMO channel simplifies to a simple scalar or single-input single-output (SISO) channel.

2.2 Examples of MIMO Channels

This dissertation deals with MIMO channels as an abstract and convenient way to describe the communication process. In real systems, each particular scenario has a specific type of MIMO channel with a given structure. The results in this dissertation are completely general and do not depend on the specific scenario that is modeled as a MIMO channel (of course, for the numerical simulations of the proposed methods, a specific choice of the type of MIMO channel has to be made).

This section is devoted to show how different communication systems can be expressed as a communication over a MIMO channel. By doing this, a MIMO system is indeed seen as a unified way to represent many different types of channels. Some of the characteristics that define a channel are:

- the degree of frequency-selectivity: flat or narrowband channels vs. frequency-selective channels (also termed time-dispersive, broadband, or wideband channels).

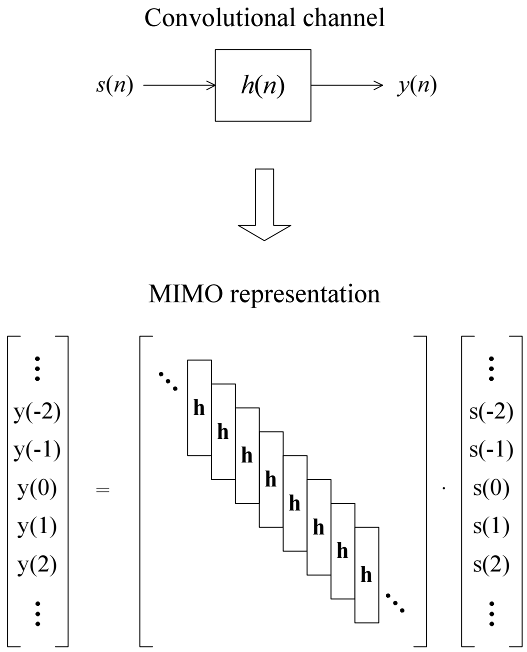


Figure 2.2: MIMO representation of a convolutional (frequency-selective) channel.

- the degree of time-selectivity [Bar01]: linear time-invariant (LTI) channels vs. linear time-varying (LTV) channels.
- number of space transmit and receive dimensions: either antennas in wireless systems or copper wires in wireline systems such as the popular digital subscriber line (DSL) systems.
- availability of other dimensions such as code and polarization axes.

2.2.1 Frequency-Selective Channel

Symbol-rate sampling of the output of the whitening matched filter (at the correct sampling phase) is optimum in the sense that yields a set of sufficient statistics for estimation of the transmitted sequence [GDF72, Ung74, Qur85]. Sampling at a rate higher than the symbol rate (fractional sampling) is considered in §2.2.5 and §2.2.6. Unless otherwise stated, the channel is considered to be LTI.

The considered discrete-time noiseless channel model after symbol-rate sampling is

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)\, s(n-k) = \sum_{k=0}^{L} h(k)\, s(n-k) \tag{2.4}$$

where $s(n)$ and $y(n)$ denote the transmitted symbols and the received samples, respectively, and $h(n)$ includes the physical channel, the shaping pulse at the transmitter, and the whitening matched filter at the receiver. The second expression in (2.4) is obtained assuming that the channel is causal and of finite order L.[3] Note that the main difficulty of dispersive channels is the inter-symbol interference (ISI) due to the convolution of the symbols with the channel impulse response.

The noiseless MIMO channel model corresponding to a block of N transmitted symbols $s(0), \cdots, s(N-1)$ is (see Figure 2.2)[4]

$$\begin{bmatrix} y(0) \\ \vdots \\ y(N-1) \\ y(N) \\ \vdots \\ y(N+L-1) \end{bmatrix} = \begin{bmatrix} h(L) & \cdots & h(0) & 0 & & \cdots & & 0 \\ 0 & \ddots & \vdots & \ddots & \ddots & & & \\ & \ddots & h(L) & & \ddots & 0 & & \vdots \\ \vdots & & 0 & \ddots & & h(0) & \ddots & \\ & & & \ddots & \ddots & \vdots & \ddots & 0 \\ 0 & & \cdots & & 0 & h(L) & \cdots & h(0) \end{bmatrix} \begin{bmatrix} s(-L) \\ \vdots \\ s(-1) \\ \boxed{\begin{matrix} s(0) \\ \vdots \\ s(N-1) \end{matrix}} \\ s(N) \\ \vdots \\ s(N+L-1) \end{bmatrix} \tag{2.5}$$

where the channel matrix has a very specific structure and is commonly termed convolution matrix. Note that a LTV channel can be straightforwardly accommodated in (2.5).

The main drawback of the block model in (2.5) is the so-called inter-block interference (IBI) caused by the preceding block (previously L transmitted symbols $s(-L), \cdots, s(-1)$) and the subsequent block (following L transmitted symbols $s(N), \cdots, s(N+L-1)$). This inconvenient can be easily circumvented by introducing a guard period between successive blocks, *i.e.*, by transmitting some known sequence of at least L symbols between blocks [Sca99a, Sca99b, Bar01]. The two simplest and most common ways of doing this in practice are the addition of zeros or zero-padding (ZP) and the introduction of a cyclic prefix (CP) which leads naturally to a multicarrier formulation (see [Sca99a, Sca99b] for a detailed treatment of both approaches). The introduction of the guard period has an implicit loss of spectral efficiency of $L/(N+L)$ which can be made as small as desired by choosing N large enough.

---

[3]A channel of order L means that the channel has length $L+1$ (the polynomial representation of the channel has order L).

[4]To define the signal model in (2.5) we have considered the received samples with contribution from the N transmitted symbols $s(0), \cdots, s(N-1)$. Another possibility would be to consider only a block of N receive samples.

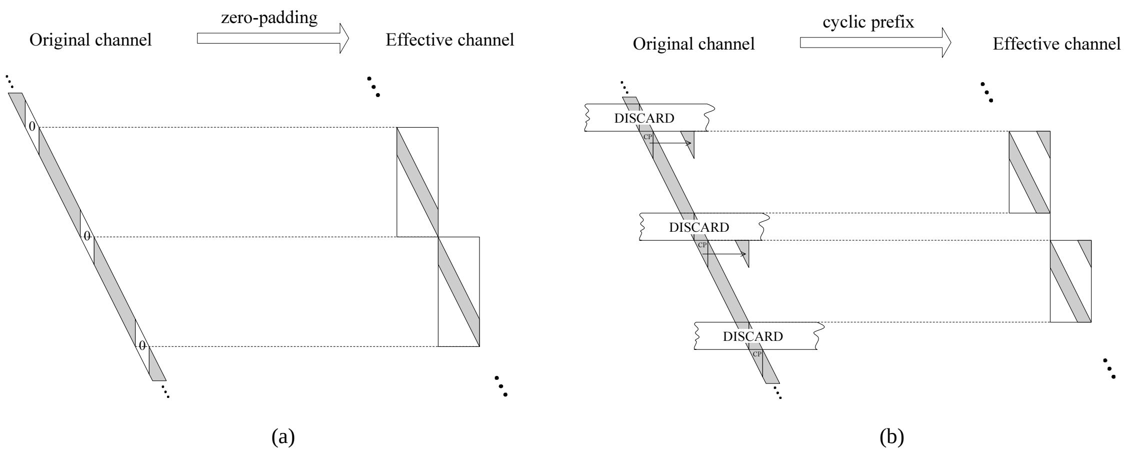


Figure 2.3: Simplification of the channel matrix (removal of IBI) after: (a) zero-padding and (b) the utilization of the cyclic prefix.

2.2.1.1 Zero-Padding

The channel model in (2.5) is greatly simplified if at least L zeros are padded between blocks, in which case the signal model simplifies to (see Figure 2.3(a))

$$\begin{bmatrix} y(0) \\ \vdots \\ y(N-1) \\ y(N) \\ \vdots \\ y(N+L-1) \end{bmatrix} = \begin{bmatrix} h(0) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h(L) & & \ddots & 0 \\ 0 & \ddots & & h(0) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h(L) \end{bmatrix} \begin{bmatrix} s(0) \\ \vdots \\ s(N-1) \end{bmatrix} \tag{2.6}$$

where the IBI has been effectively removed. In this case, a LTV channel can be readily incorporated in (2.6).

2.2.1.2 Cyclic Prefix

Another possibility to simplify the channel model in (2.5) is to introduce a cyclic prefix of at least L symbols consisting on the last L symbols of the block. To be more specific, the transmitter pre-appends the last L symbols of the block to finally transmit $s(N-L), \cdots, s(N-1), s(0), \cdots, s(N-1)$ and the receiver discards the last L received samples. The signal model in (2.5) simplifies then to (see Figure 2.3(b))

$$\begin{bmatrix} y(0) \\ \vdots \\ y(N-1) \end{bmatrix} = \begin{bmatrix} h(0) & 0 & \cdots & \cdots & 0 & h(L) & \cdots & h(1) \\ \vdots & \ddots & \ddots & & & 0 & \ddots & \vdots \\ \vdots & & h(0) & \ddots & & & \ddots & h(L) \\ h(L) & & \vdots & \ddots & 0 & & & 0 \\ 0 & \ddots & \vdots & & h(0) & \ddots & & \vdots \\ \vdots & \ddots & h(L) & & \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \vdots & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & h(L) & \cdots & \cdots & h(0) \end{bmatrix} \begin{bmatrix} s(0) \\ \vdots \\ s(N-1) \end{bmatrix} \quad (2.7)$$

where not only the IBI has been removed but also the channel matrix has now become a circulant matrix which has many interesting properties (c.f. §2.2.2). In other words, the linear convolution performed by the channel has been artificially converted into a circular convolution with the introduction of the cyclic prefix. It is important to point out that, for the model in (2.7) to be valid, the channel has to remain unchanged during the whole transmission of the block. Therefore, the maximum value of the block length N and consequently the spectral efficiency $N/(N+L)$ are in practice limited by the channel variability.

2.2.2 Multicarrier Channel

In a multicarrier communication system, the available bandwidth is partitioned into N subbands and then each subband is independently used for transmission [Kal89, Bin90]. Such an approach not only simplifies the communication process but it is also a capacity-achieving structure for a sufficiently high N [Gal68, Hir88, Ral98, Wan00]. In other words, a multicarrier structure can be adopted without loss of optimality.

Orthogonal frequency division multiplexing (OFDM) is a popular transmission scheme for wireless communications that belongs to the class of multicarrier transmission. OFDM has been adopted in many standards such as in digital audio/video broadcasting (DAB/DVB) standards in Europe [Wan00] and in wireless local area networks (WLAN) both in the European standard HIPERLAN/2 [ETS01] and in the US standard IEEE 802.11 [IEE99]. OFDM has also been proposed for digital cable television systems [Wan00]. Discrete multi-tone (DMT) is a more refined multicarrier transmission which has been applied to DSL modems over twisted pairs [Wan00]. DMT modulation makes use of the channel knowledge at the transmitter to properly distribute the bits over the carriers by using different constellations (to avoid, for example, faded frequencies) and to possibly use different coding schemes among the carriers as well [Rui92]. OFDM systems use the same constellation at each carrier (no matter the fading state) and

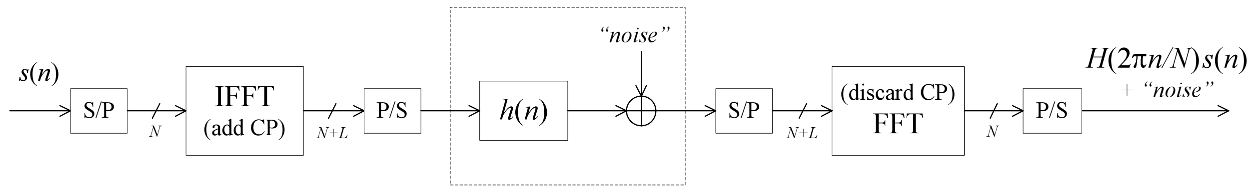


Figure 2.4: Classical scheme of an OFDM communication system.

therefore can be used for communications in which the transmitter does not know the channel (in practice, of course, OFDM is always combined with coding schemes to combat the detrimental effect of faded carriers on the performance of the system, yielding the so-called coded-OFDM (COFDM) [Wan00]). In the multicarrier literature, however, the difference between OFDM and DMT is sometimes overlooked and they are both interchangeably used to refer to the same scheme.

Multicarrier transmission is easily derived from the MIMO channel model of (2.7) obtained with the utilization of the cyclic prefix. Thanks to the introduction of the cyclic prefix at the transmitter and its removal at the receiver, the resulting channel matrix in (2.7) is a circulant matrix whose rows are composed of cyclically shifted versions of a sequence [Hor85]. Circulant matrices have a very interesting and useful property [Lan69, Gra72]: the eigenvectors are independent of the specific channel coefficients and are always given by complex exponentials. To be more precise, the eigenvalue decomposition (EVD) of the circulant channel matrix in (2.7) is

$$\begin{bmatrix} h(0) & 0 & \cdots & \cdots & 0 & h(L) & \cdots & h(1) \\ \vdots & \ddots & \ddots & & & 0 & \ddots & \vdots \\ \vdots & & h(0) & \ddots & & & \ddots & h(L) \\ h(L) & & \vdots & \ddots & 0 & & & 0 \\ 0 & \ddots & \vdots & & h(0) & \ddots & & \vdots \\ \vdots & \ddots & h(L) & & \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \vdots & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & h(L) & \cdots & \cdots & h(0) \end{bmatrix} = \mathbf{F}^H \mathbf{D}_H \mathbf{F} \tag{2.8}$$

where $\mathbf{F} = [\mathbf{f}_0, \cdots, \mathbf{f}_{N-1}]$ is the unitary discrete Fourier transform (DFT) (with $\mathbf{f}_k \triangleq \frac{1}{\sqrt{N}}[1, e^{-j\frac{2\pi}{N}k}, e^{-j\frac{2\pi}{N}2k}, \cdots, e^{-j\frac{2\pi}{N}(N-1)k}]^T$) and $\mathbf{D}_H = \text{diag}\left(\{H(2\pi k/N)\}_{k=0}^{N-1}\right)$ (with $H(w) = \sum_{n=0}^{L} h(n)e^{-jwn}$ being the channel transfer function) is a diagonal matrix with the DFT coefficients as diagonal elements [Lan69, Gra72]. Note that the eigenvectors of the circulant matrix are given by $\mathbf{f}_k^*$ for $0 \leq k \leq N-1$.

As a consequence of the simple EVD of a circulant matrix, such a channel can be easily diagonalized by performing the inverse DFT (IDFT) at the transmitter $\tilde{\mathbf{s}} = \mathbf{F}^H \mathbf{s}$ ($\tilde{\mathbf{s}}$ contains the temporal transmitted samples) and the DFT at the receiver $\mathbf{y} = \mathbf{F}\tilde{\mathbf{y}}$ ($\tilde{\mathbf{y}}$ contains the temporal received samples). In practice, the inverse fast Fourier transform (IFFT) and the fast Fourier transform (FFT) are used (see Figure 2.4 for a scheme of the whole communication process). As a consequence of the diagonalization of the channel matrix, the communication is effectively performed over a set of N parallel and independent subchannels with gains $\{H(2\pi k/N)\}_{k=0}^{N-1}$:

$$y(k) = H(2\pi k/N)\; s(k) + n(k) \qquad 0 \le k \le N-1 \tag{2.9}$$

or in matrix form

$$\mathbf{y} = \mathbf{D}_H \mathbf{s} + \mathbf{n} \tag{2.10}$$

where $\mathbf{y} = [y(0), \cdots, y(N-1)]^T$, $\mathbf{s} = [s(0), \cdots, s(N-1)]^T$, and $\mathbf{n} = [n(0), \cdots, n(N-1)]^T$.

It is remarkable that with a multicarrier approach, the original frequency-selective channel with ISI and IBI is transformed into a set of parallel flat subchannels that can be straightforwardly equalized. In addition, the different carriers can be used for multiplexing several users such as in Orthogonal Frequency Division Multiple Access (OFDMA), where different users are assigned different non-overlapping carriers, and in multicarrier CDMA (MC-CDMA), where each user spreads its signal over all the carriers by means of a spreading code [Wan00].

As a final comment, it is worth pointing out that a multicarrier system need not be implemented following the described block approach using the FFT and the IFFT. Instead, it can be implemented using banks of orthogonal filters [Aka98].

2.2.3 Multi-Antenna Wireless Channel

Among the scenarios that lead to MIMO representations, perhaps the most popular is that of wireless communications when multiple antennas are used at both the transmitter and the receiver (see Figure 2.5). The popularity of this particular scenario is mainly due to recent studies indicating a linear increase of capacity with the number of antennas [Tel95, Fos98] (see also [Shi98, Shi00, Chu02]).

To exploit antenna arrays in wireless communication, it is necessary to obtain an accurate, yet tractable, modeling of the MIMO channel. Existing models represent two extreme approaches. On the one hand is the statistical modeling which is a tractable and idealized abstraction of spatial propagation characteristics [Tel95, Fos98]. On the other hand are the parametric physical models which explicitly relate the scattering environment to the channel coefficients and dictate their statistics but are less tractable due to the nonlinearity in the spatial angles [Say02]. Some attempts at bridging the gap between the two modeling philosophies are [Say02, Ges02, Shi00, Ral98].

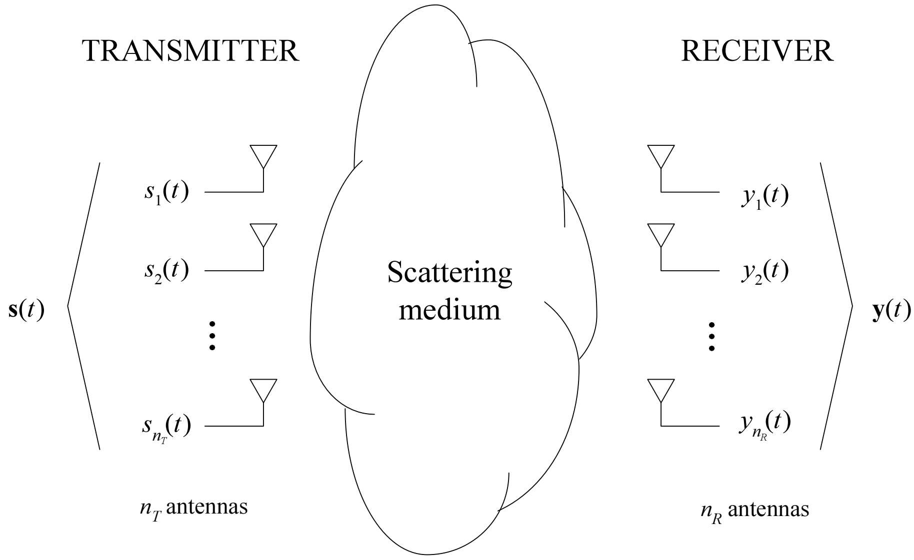


Figure 2.5: Example of a MIMO channel arising in wireless communications when multiple antennas are used at both the transmitter and the receiver.

2.2.3.1 Flat Multi-Antenna Wireless Channel

For the case of a flat channel, the MIMO channel model of (2.1) $\mathbf{y} = \mathbf{Hs} + \mathbf{n}$ is readily obtained defining the channel matrix as

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n_T} \\ h_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ h_{n_R1} & \cdots & \cdots & h_{n_Rn_T} \end{bmatrix} \tag{2.11}$$

where h_{ij} is the fading coefficient between the jth transmit antenna and the ith receive one. In this case, the channel matrix $\mathbf{H}$ is in general a full matrix with no structure. Note that the channel matrix may or may not vary at each transmission.

2.2.3.2 Frequency-Selective Multi-Antenna Wireless Channel

For the frequency-selective case, a discrete-time convolution with channel matrix coefficient is obtained (similarly to (2.4)):

$$\mathbf{y}(n) = \sum_{k=0}^{L} \mathbf{H}(k)\, \mathbf{s}(n-k) + \mathbf{n}(n) \tag{2.12}$$

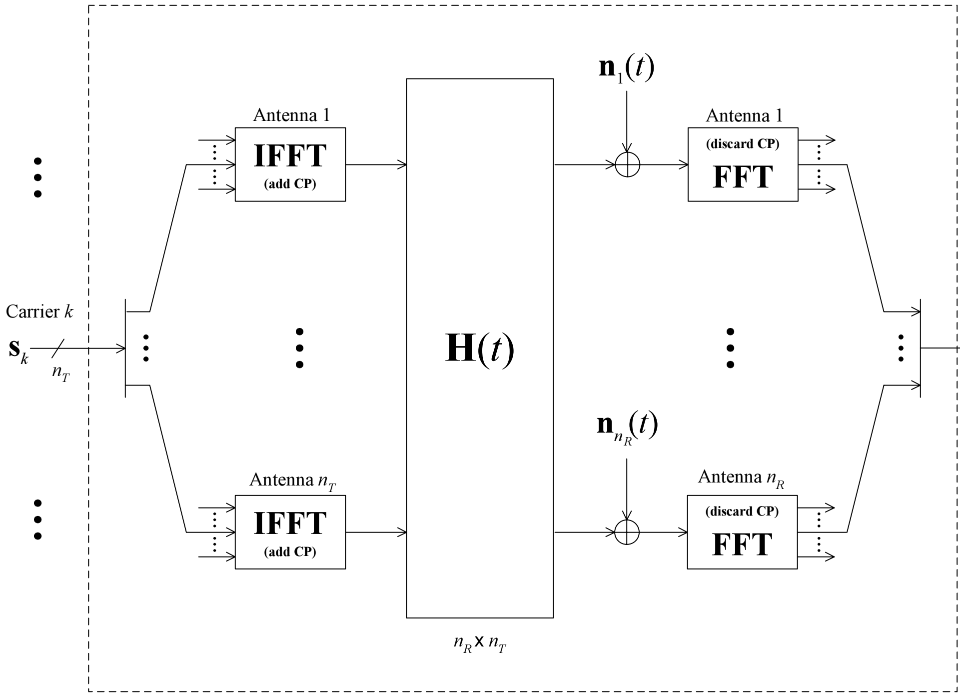


Figure 2.6: Scheme of a multi-antenna OFDM wireless communication system.

where $\mathbf{H}(n)$ is the discrete-time channel matrix defined as

$$\mathbf{H}(n) = \begin{bmatrix} h_{11}(n) & h_{12}(n) & \cdots & h_{1n_T}(n) \\ h_{21}(n) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ h_{n_R1}(n) & \cdots & \cdots & h_{n_Rn_T}(n) \end{bmatrix} \tag{2.13}$$

and $h_{ij}(n)$ is the discrete-time channel from the jth transmit antenna to the ith receive one. The channel order of the frequency-selective matrix channel $\mathbf{H}(n)$ is defined as the maximum of the orders among its elements $h_{ij}(n)$.

At this point, the frequency-selective channel of (2.12) can be manipulated as in §2.2.1 and §2.2.2. For example, the MIMO channel model obtained after zero-padding (similarly to (2.6)) is [Ral98]

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{n_R} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{1,1} & \cdots & \mathbf{H}_{1,n_T} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{n_R,1} & \cdots & \mathbf{H}_{n_R,n_T} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_{n_T} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_{n_R} \end{bmatrix} \tag{2.14}$$

where $\mathbf{y}_i = [y_i(0), \cdots, y_i(N-1), \cdots, y_i(N+L-1)]^T$ is the received signal and $\mathbf{n}_i = [n_i(0), \cdots, n_i(N-1), \cdots, n_i(N+L-1)]^T$ is the received noise at the ith receive antenna, $\mathbf{s}_j = [s_j(0), \cdots, s_j(N-1)]^T$ is the transmitted signal through the jth transmit antenna, and the convolutional channel matrix from the jth transmit antenna to the ith receive one is

$$\mathbf{H}_{ij} = \begin{bmatrix} h_{ij}(0) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h_{ij}(L) & & \ddots & 0 \\ 0 & \ddots & & h_{ij}(0) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{ij}(L) \end{bmatrix}. \tag{2.15}$$

If a cyclic prefix is used instead of a zero-padding and the IDFT/DFT operations are included at each antenna, a multi-antenna multicarrier MIMO channel model is obtained (similarly to 2.10) as

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{n_R} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{H_{1,1}} & \cdots & \mathbf{D}_{H_{1,n_T}} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_{H_{n_R,1}} & \cdots & \mathbf{D}_{H_{n_R,n_T}} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_{n_T} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_{n_R} \end{bmatrix} \tag{2.16}$$

where $\mathbf{y}_i$, $\mathbf{n}_i$, and $\mathbf{s}_j$ are defined as before and $\mathbf{D}_{H_{i,j}} = \text{diag}\left(\{H_{i,j}(2\pi k/N)\}_{k=0}^{N-1}\right)$. Reorganizing the elements of the transmitted and received vectors, the MIMO channel model can be written as (see Figure 2.6)

$$\begin{bmatrix} \mathbf{y}_0 \\ \vdots \\ \mathbf{y}_{N-1} \end{bmatrix} = \text{diag}\left(\{\mathbf{H}_k\}_{k=0}^{N-1}\right) \begin{bmatrix} \mathbf{s}_0 \\ \vdots \\ \mathbf{s}_{N-1} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_0 \\ \vdots \\ \mathbf{n}_{N-1} \end{bmatrix} \tag{2.17}$$

or, equivalently, following the multiple MIMO channel model of (2.3) as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{s}_k + \mathbf{n}_k \qquad 0 \le k \le N-1 \tag{2.18}$$

where $\mathbf{y}_k = [y_{k,1}, \cdots, y_{k,n_R}]^T$, $\mathbf{n}_k = [n_{k,1}, \cdots, n_{k,n_R}]^T$, $\mathbf{s}_k = [s_{k,1}, \cdots, s_{k,n_T}]^T$, and $[\mathbf{H}_k]_{i,j} = H_{i,j}(2\pi k/N)$ at the kth carrier.

2.2.4 Wireline DSL Channel

Digital Subscriber Line (DSL) technology has gained popularity as a broadband access technology capable of reliably delivering high data rates over telephone subscriber lines [Sta99]. Modeling a DSL system as a MIMO channel presents many advantages with respect to treating each twisted pair independently [Hon90, Gin02]. If fact, modeling a wireline channel as a MIMO channel was done almost three decades ago [Lee76, Sal85].

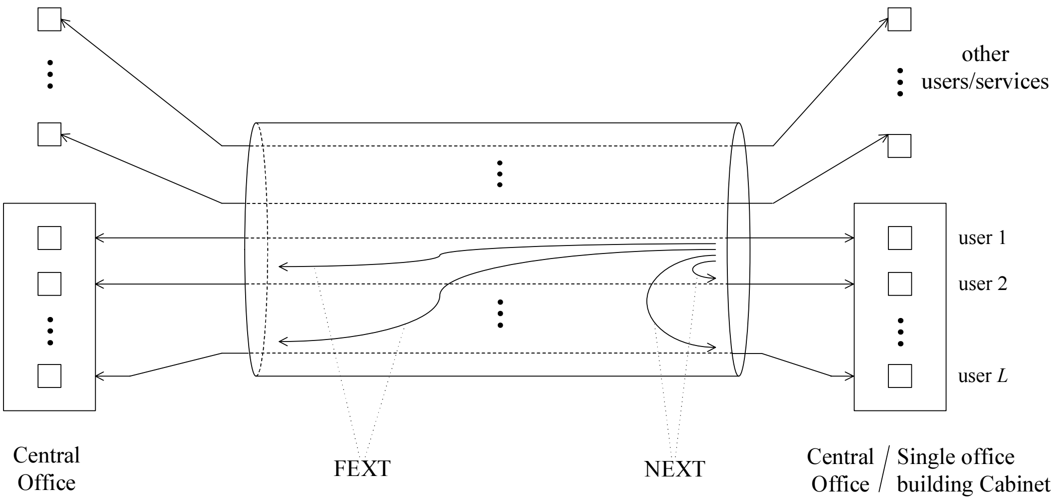


Figure 2.7: Scheme of a bundle of twisted pairs of a DSL system.

In many situations, joint processing can be assumed at one side of the link, the Central Office (CO), whereas the other side, corresponding to the Customer Premises Equipment (CPE), must use independent processing per user since users are geographically distributed [Gin02]. In some cases of practical interest, however, both ends of the MIMO system are each terminated in a single physical location, *e.g.*, links between CO's and Remote Terminals or links between CO's and private networks. This allows the utilization of joint processing at both sides of the link [Hon90].

The dominant impairment in DSL systems is crosstalk arising from electromagnetic coupling between neighboring twisted-pairs. Near-end crosstalk (NEXT) comprises the signals originated in the same side of the received signal (due to the existence of downstream and upstream transmission) and far-end crosstalk (FEXT) includes the signal originated in the opposite side of the received signal. The impact of NEXT is generally suppressed by employing frequency division duplex (FDD) to separate downstream and upstream transmission.

The general case under analysis consists of a binder group composed of L users in the same physical location plus some other users that possibly belong to a different service provider and use different types of DSL systems (see Figure 2.7). The MIMO channel represents the communication of the L intended users while the others are treated as interference.

DSL channels are highly frequency-selective; as a consequence, practical communication systems are based on DMT modulations, *i.e.*, multicarrier schemes. The channel model of such a system is exactly as the obtained for the frequency-selective multi-antenna system considered in §2.2.3 given by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{s}_k + \mathbf{n}_k \qquad 0 \leq k \leq N-1 \tag{2.19}$$

where $\mathbf{y}_k = [y_{k,1}, \cdots, y_{k,L}]^T$, $\mathbf{n}_k = [n_{k,1}, \cdots, n_{k,L}]^T$, $\mathbf{s}_k = [s_{k,1}, \cdots, s_{k,L}]^T$, and $\mathbf{H}_k$ is now an

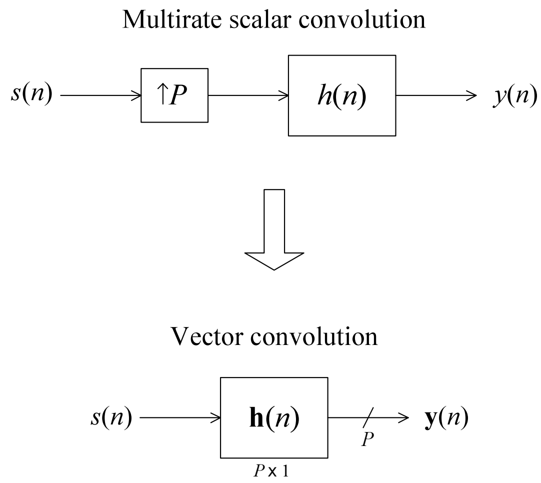


Figure 2.8: Equivalence between a multirate scalar convolution and a vector convolution.

$L \times L$ square matrix in which the diagonal elements represents the channel transfer function of each of the twisted pairs and the nondiagonal elements represent the FEXT.

2.2.5 Fractional Sampling Equivalent Channel

Excess-bandwidth systems (the majority of practical systems) utilize a transmit bandwidth higher than the minimum (Nyquist) bandwidth. Examples are systems using spreading codes (c.f. §2.2.6) and systems using a root-raised cosine transmit shaping pulse (with a nonzero rolloff factor) for which the baseband transmit spectrum is then band-limited to $|f| \leq (1+\beta)\, 1/2T$, where $\beta \in [0,1]$ is the rolloff factor which determines the excess bandwidth [Pro95]. Although in such cases the sampling theorem requires a sampling rate higher than the symbol rate for perfect reconstruction, symbol-rate sampling of the output of the whitening matched filter (at the correct sampling phase) is optimum in the sense that yields a set of sufficient statistics for estimation of the transmitted sequence [GDF72, Ung74, Qur85].

Fractional-rate sampling (sampling at a rate higher than the symbol rate), however, has significant practical advantages compared to symbol-rate sampling such as the insensitivity with respect to the sampling phase and the possibility to implement in discrete time many of the operations performed at the receiver such as the matched-filtering operation (as opposed to the analog matched-filtering required for the symbol-rate sampling), timing recovery, frequency recovery, and phase recovery (c.f. [Qur85]). Fractionally-sampled systems can be modeled as a multirate convolution which can be easily converted into the more convenient vector convolution as we now show.

The received continuous-time noiseless signal is

$$y_c(t) = \sum_{k=-\infty}^{\infty} s(k)\, h_c(t - kT) \tag{2.20}$$

where T is the symbol period, the subscript c is used to denote continuous time signals, and $h_c\left(t\right)$ includes the physical channel, the shaping pulse at the transmitter, and the bandpass filter at the receiver. Note that the discrete-time signal model after sampling at the symbol rate of (2.4) is readily obtained by defining $y\left(n\right) \triangleq y_c(nT)$ and $h\left(n\right) \triangleq h_c\left(nT\right)$.

The resulting discrete-time signal after fractional sampling every T/P seconds is given by the multirate convolution

$$y_c(nT/P) = \sum_{k=-\infty}^{\infty} s\left(k\right) h_c\left(\left(n-kP\right)T/P\right)$$

or, more compactly,

$$y\left(n\right) = \sum_{k=-\infty}^{\infty} s\left(k\right) h\left(n-kP\right) \tag{2.21}$$

where we have defined $y\left(n\right) \triangleq y_c(nT/P)$ and $h\left(n\right) \triangleq h_c\left(nT/P\right)$. Note that the multirate convolution of (2.21) can be seen as the classical convolution of the channel and the upsampled version of $s\left(n\right)$ (see Figure 2.8). As is usually done in multirate theory, by stacking $y\left(n\right)$ and $h\left(n\right)$ in vectors of length P as $\mathbf{y}(n) \triangleq \left[y\left(nP\right), y\left(nP+1\right), \cdots, y\left(nP+\left(P-1\right)\right)\right]^T$ and $\mathbf{h}(n) \triangleq \left[h\left(nP\right), h\left(nP+1\right), \cdots, h\left(nP+\left(P-1\right)\right)\right]^T$, the following vector discrete-time convolution signal model is obtained (see Figure 2.8):

$$\mathbf{y}\left(n\right) = \sum_{k=-\infty}^{\infty} s\left(k\right) \mathbf{h}\left(n-k\right).$$

Assuming now that the oversampled channel $h\left(n\right)$ has a finite support given by the interval $0 \le n < \left(L+1\right)P$, the final model is (similarly to (2.4))

$$\mathbf{y}\left(n\right) = \sum_{k=0}^{L} \mathbf{h}\left(k\right) s\left(n-k\right) \tag{2.22}$$

At this point, the MIMO representation can be easily obtained as was done for the scalar discrete-time convolution signal model in §2.2.1. For example, the MIMO channel model obtained after zero-padding is (similarly to (2.6))

$$\begin{bmatrix} \mathbf{y}(0) \\ \vdots \\ \mathbf{y}(N-1) \\ \mathbf{y}(N) \\ \vdots \\ \mathbf{y}(N+L-1) \end{bmatrix} = \begin{bmatrix} \mathbf{h}(0) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{h}(L) & & \ddots & 0 \\ 0 & \ddots & & \mathbf{h}(0) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{h}(L) \end{bmatrix} \begin{bmatrix} s(0) \\ \vdots \\ s(N-1) \end{bmatrix}. \tag{2.23}$$

2.2.6 CDMA Channel

In a code division multiple access (CDMA) system, multiple users transmit overlapping in time and frequency but using different signature waveforms or spreading codes (which are excess-bandwidth shaping pulses). The discrete-time model for such systems is commonly obtained following a matched filtering approach by sampling at the symbol rate the output of a bank of filters where each filter is matched to one of the signature waveforms [Ver98]. An alternative derivation of the discrete-time model for CDMA systems is based on a fractionally-sampled scheme [Tsa96] (see also [Kle96, Jun95]) as we now briefly describe.

The signal model obtained for the fractional-rate sampling in §2.2.5 is valid for each of the users of the CDMA system. In particular, the noiseless received continuous-time signal corresponding to just one user is directly given by (2.20). After fractional sampling at the proper sampling rate, the vector discrete-time convolution signal model of (2.22) is obtained. Adding up the effects of K users, the final noiseless discrete-time signal model is

$$\mathbf{y}(n) = \sum_{k=1}^{K} \sum_{l=0}^{L} \mathbf{h}_k(l)\, s_k(n-l) \tag{2.24}$$

where each user is indexed by the index k.

At this point, the MIMO representation can be easily obtained as was done for the scalar discrete-time convolution signal model in §2.2.1. For example, the MIMO channel model obtained after zero-padding is obtained similarly to (2.23) but adding up the effect of all K users.

2.3 Gains and Properties of MIMO Channels

MIMO channels have a number of advantages over traditional SISO channels such as the beamforming (or array) gain, the diversity gain, and the multiplexing gain. The beamforming and diversity gains are no exclusive of MIMO channels and also exist in SIMO and MISO channels. The multiplexing gain, however, is a unique characteristic of MIMO channels. Some gains can be simultaneously achieved while others compete and establish a tradeoff as is explained later. An excellent overview of the gains of MIMO channels is given in [Böl02].

In a nutshell, the use of multiple dimensions at both ends of a communication link offers significant improvements in terms of *spectral efficiency* and *link reliability*.

2.3.1 Beamforming Gain

Beamforming or array gain is the improvement in SINR obtained by coherently combining the signals on multiple transmit or multiple receive dimensions. If the BER of a communication system is plotted with respect to the transmitted power or the received power per antenna (using a logarithmic scale), the beamforming gain is easily characterized as a shift of the curve due to the gain in SINR.

Beamforming is a term traditionally associated with array processing or *smart antennas* in wireless communications where an array of antennas exists either at the transmitter or at the receiver [Vee88, Kri96]. We consider the concept of beamforming over any arbitrary dimension, generalizing the traditional meaning that refers only to the space dimension.

For illustration purposes, consider a SIMO channel with the received signal given by $\mathbf{y} = \mathbf{h}x + \mathbf{n}$ where $\mathbf{h}$ is the channel or signature of the desired signal and $\mathbf{n}$ is a white noise $\mathbb{E}\left[\mathbf{n}\mathbf{n}^H\right] = \sigma^2\mathbf{I}$. The receiver can now use a beamvector $\mathbf{w}$ to combine all the elements of $\mathbf{y}$ in a coherent way as $z = \mathbf{w}^H\mathbf{y}$. If the beamvector matches the channel, *i.e.*, if $\mathbf{w} = \mathbf{h}$, the SINR is maximized and given by $\text{SINR} = \|\mathbf{h}\|^2/\sigma^2$ which clearly shows the increase of SINR with respect to using a single receive dimension $\text{SINR} = |h|^2/\sigma^2$.

2.3.2 Diversity Gain

Diversity gain is the improvement in link reliability obtained by receiving replicas of the information signal through independently fading links, branches, or dimensions. This type of diversity is clearly related to the random nature of the channel and is closely connected to the specific channel statistics. If the BER of a communication system is plotted with respect to the transmitted power or the received power per antenna (using a logarithmic scale), the diversity gain is easily characterized as the increase of the slope of the curve in the low BER region.

The basic idea is that with high probability, at least one or more of these links will not be in a fade at any given instant. In other words, the use of multiple dimensions reduces the fluctuations of the received signal and eliminates the deep fades [Böl02]. Clearly, this concept is suited for wireless communications where fading exists due to multipath effects (this concept may not be useful for wireline communications where the fading effect does not exist).

The three main forms of diversity traditionally exploited in wireless communications systems are temporal diversity, frequency diversity, and spatial diversity [Böl02]. Other forms of diversity include, for example, code diversity and polarization diversity [Nab02].

Transmit diversity is more difficult to exploit than receive diversity since it requires special modulation and coding schemes, whereas receive diversity simply needs the multiple receive dimensions to fade independently without requiring any specific modulation or coding scheme.

2.3.3 Multiplexing Gain

Multiplexing gain is the increase of rate, at no additional power consumption, obtained through the use of multiple dimensions at both sides of the communication link [Böl02]. While the beamforming and the diversity gains can be obtained when multiple dimensions are present at either the transmit or the receive side, multiplexing gain requires multiple dimensions at both ends of the link.

The basic idea is to exploit the multiple dimensions to open up several parallel *subchannels* within the MIMO channel, also termed channel *eigenmodes*, which lead to a linear increase in capacity. The multiplexing property allows the transmission of several symbols simultaneously or, in other words, the establishment of several *substreams* for communication (c.f. Chapters 5 and 6).

Channel Eigenmodes, Parallel Subchannels, and Established Substreams

It is important to remark that, in principle, the concepts of parallel subchannels (or channel eigenmodes) and established substreams have a completely different meaning. Of course, they are closely related and in many cases their meaning coincide. By parallel subchannels or channel eigenmodes we refer to the parallel and orthogonal subchannels that the channel provides which can be obtained via the singular value decomposition (SVD) of the channel $\mathbf{H}$ or via the eigenvalue decomposition (EVD) of the squared channel $\mathbf{H}^H\mathbf{H}$. The amplitude gains of the subchannels are therefore given by the singular values of the channel and, equivalently, the power gains are given by the eigenvalues of the squared channel, hence the name channel eigenmodes. The number of available parallel subchannels is $\text{rank}\,(\mathbf{H})$ or, equivalently, $\text{rank}\left(\mathbf{H}^H\mathbf{H}\right)$.

By established substreams, we refer to the number of simultaneously transmitted symbols which may or may not interfere with each other depending on whether the channel matrix is diagonalized or not. If an orthogonality constraint is imposed over the established substreams to obtain a set of parallel non-interfering links (diagonalized channel matrix), then the substreams have to be established over the channel eigenmodes. In that case, both concepts become equivalent and, as a consequence, the maximum number of substreams is clearly limited by the channel rank. However, if we allow the established substreams not to be exactly orthogonal (not perfectly separable), both concepts are completely different; in fact, the maximum number of substreams is not directly limited anymore by the rank of the channel (although care has to be taken to guarantee a certain level of performance due to the non-perfect separation of the substreams). In Chapters 5 and 6, the possibility of establishing substreams over the channel eigenmodes both in parallel and non-parallel fashions are explored in depth.

2.3.4 Tradeoffs Between Gains

Beamforming and Diversity Gains

Beamforming gain is a concept that refers to the combination of multiple copies of the same signal for a specific channel realization regardless of the channel statistics. Diversity gain, however, is directly connected to the statistical behavior of the channel. With multiple receive dimensions, both gains can be simultaneously achieved by a coherent combination of the received signals and there is no tradeoff between them. With multiple transmit dimensions, beamforming gain requires channel knowledge at the transmitter whereas diversity gain can be achieved even when the channel is unknown.

Beamforming and Multiplexing Gains

Maximum beamforming gain on a MIMO channel implies that only the maximum singular value of the channel should be used [And00] (c.f. § 5.3.1). In terms of multiplexing gain, however, the optimum strategy is to use a subset of the channel singular values according to a water-filling strategy [Tel95] (c.f. §5.5.4). In other words, maximum beamforming gain requires establishing a single substream for communication, whereas maximum multiplexing gain requires, in general, establishing several simultaneous substreams (c.f. Chapter 5).

Diversity and Multiplexing Gains

Traditionally, the design of systems has been focused on either extracting maximum diversity gain or maximum multiplexing gain. Nevertheless, both gains can in fact be simultaneously achieved, but there is a fundamental tradeoff between how much of each type of gain any communication scheme can extract as was shown in the excellent groundbreaking paper [Zhe03]. Note that, since the diversity gain is related to the BER (it is the slope of the BER curve in the high SINR region) and the multiplexing gain is related to the achieved rate, the diversity-multiplexing tradeoff is essentially the tradeoff between the error probability and the data rate of a system.

2.4 State-of-the-Art of Transmission Techniques for MIMO Channels

A plethora of communication techniques exists for transmission over MIMO channels which we briefly review in this section. They basically depend on the degree of channel state information (CSI) available at the transmitter and at the receiver. Clearly, the more channel information, the better the performance of the system.

CSI at the receiver (CSIR) is traditionally acquired via the transmission of a training sequence (pilot symbols) that allows the estimation of the channel. It is also possible to use blind methods that do not require any training symbols but exploit knowledge of the structure of the transmitted signal or of the channel [Mou95, Liu96, Tsa97] (see also [Pal00, Pal01b] and references therein). CSI at the transmitter (CSIT) cannot be directly obtained as happened with CSIR. One possible way to achieve CSIT is to have a feedback channel from the receiver to the transmitter to send back the channel state as side information (this technique requires the channel to be sufficiently slowly varying and has a loss in spectral efficiency due to the utilization of part of the bandwidth to transmit the channel state). Another traditional way to acquire CSIT is to infer knowledge about the transmit channel from previous receive measurements [Ben01]. In general, the transmit and receive channels may be uncorrelated since they may be separated in time and/or frequency. In particular, in a time division duplex (TDD) system, the transmit and receive channels share the same frequency using different time slots. Similarly, in a frequency division duplex (FDD) system, the transmit and receive channels share the same time slot using different frequencies. Whenever the time/frequency separation is small compared to the time/frequency coherence of the channel, the reciprocity principle of electromagnetics implies that the instantaneous transmit and receive channels are identical. Otherwise, even when the channel reciprocity property does not hold, the transmit and receive channels still may show some statistical dependence which can be exploited [Ben01].

MIMO channel modeling has attracted a significant attention in two main practical scenarios: DSL channels and multi-antenna wireless channels. Before considering communication methods for MIMO channels, we give a glimpse of communication techniques for systems with multiple dimensions only at one side of the link, *i.e.*, multiple-input single-output (MISO) and single-input multiple-output (SIMO) channels. Transmitter diversity has traditionally been viewed more difficult to exploit than receiver diversity and, as a consequence, processing methods for reception clearly outnumber those for transmission. This is mainly because the transmitter is permitted to generate a different signal at each transmit dimension (with the consequent challenging signal design problem), whereas the receiver can only combine the copies of the received signal to achieve a performance gain. A review of space-time processing techniques up to 1997 can be found in [Pau97].

Regarding diversity only at the receiver, a wide variety of methods exists [Vee88, Kri96]. In [Bal92], for example, space-time filters were designed in the frequency domain using the minimum mean square error (MMSE) and the zero-forcing (ZF) criteria (a generalization of the classical temporal-only case [Qur85, GDF91]). In [Win94], receive diversity with optimum combining is shown to increase the capacity of a multiuser system. In [Cio95a, Cio95b], a comprehensive analysis of MMSE decision feedback (MMSE-DF) space-time filters for reception is given. A remarkable unified vision of design of space-time receive filters under the MMSE and ZF criteria can be found in [Ari99]. In [Lag00], an array combiner in conjunction with maximum likelihood sequence estimation (MLSE) was derived. The multiuser case has also been widely studied [Lup89, Mos96, Jun95].

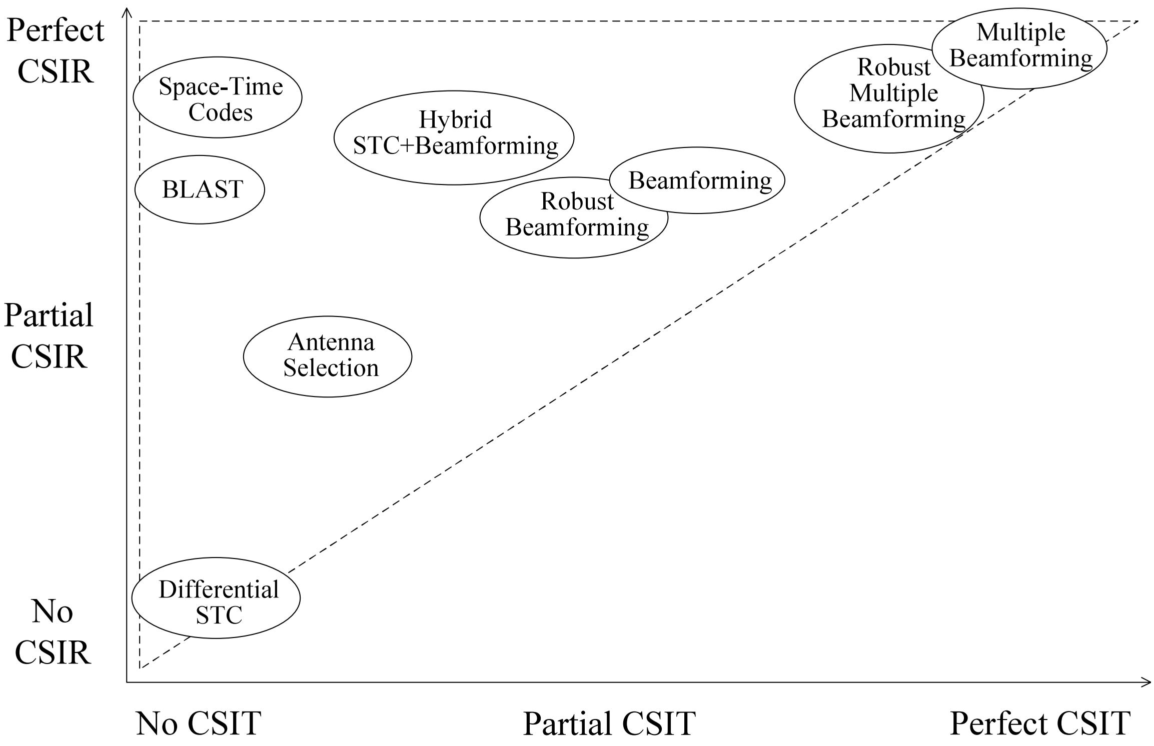


Figure 2.9: Classification of different communication methods as a function of the CSI at the transmitter and at the receiver.

Concerning methods with only transmit diversity, classical beamforming techniques can be used to obtain the array gain when perfect CSIT is available [Pau97]. With partial CSIT, beamforming techniques can also be used [Nar98]. For the case of no CSIT, the transmit diversity can be exploited either indirectly or directly. Indirect transmit diversity schemes convert spatial diversity into time or frequency diversity which can then be readily exploited by the receiver [Wit91, Ses94, Win98, Hir92]. Space-time coding (STC) techniques exploit the spatial diversity in a direct way [Ala98, Tar98].

The more general and appealing case of simultaneous transmit and receive diversity (*i.e.*, MIMO systems) is being extensively addressed since recent studies on wireless multi-antenna MIMO channels showed a linear increase of capacity with the number of antennas [Tel95, Fos98]. It is generally assumed that perfect CSI is available at the receiver. Nevertheless, the case of no CSI at any side of the link has also been considered using the so-called unitary and differential STC [Hoc00a, Hoc00b]. Regarding the CSIT, there are two main families of transmission methods that consider either no CSIT, for which STC techniques can be used, or perfect CSIT, for which linear precoders and beamforming techniques can be designed (c.f. §2.4.1 and §2.4.2, respectively). In practice, however, it may be more realistic to consider partial CSIT yielding robust designs (c.f. §2.4.3). In Figure 2.9, a variety of communication techniques for MIMO channels are positioned in a chart as a function of the required level of CSIT and CSIR.

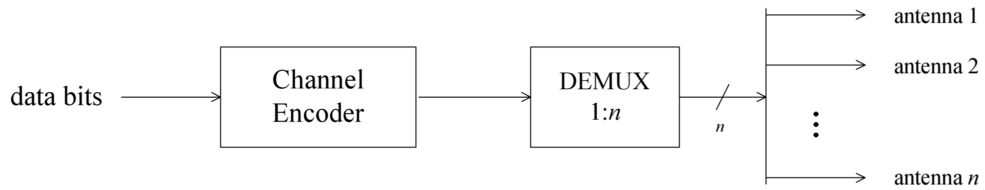


Figure 2.10: General scheme of space-time coding (STC).

2.4.1 Transmission Techniques with no CSIT

Transmission techniques that do not require CSIT can be encompassed within two main philosophies: space-time coding and layered architectures. Space-time coding generalizes the classical concept of coding on the temporal domain [Cal98] to coding on both spatial and temporal dimensions [Tar98]. Layered architectures refer to a particular case of a space-time coding when a separate coding scheme is used for each spatial branch [Fos96]. This constraint on the coding scheme represents a significant reduction of complexity at the receiver. Recently, hybrid schemes combining layered architectures with constituent space-time codes have been proposed as a reasonable tradeoff between performance and complexity [Ari00].

2.4.1.1 Space-Time Coding

The idea of space-time coding is that the transmitter does not have CSIT (see Figure 2.9) but introduces redundancy in the transmitted signal, both over space and time, that allows the receiver to recover the signal even in difficult propagation situations. To be more specific, the symbols are first encoded and then the encoded data is split into n substreams that are simultaneously transmitted using n transmit antennas (see Figure 2.10). The received signal is decoded using a maximum likelihood (ML) decoder. STC is indeed very effective as it combines the benefits of forward error correction coding and diversity transmission to provide considerable performance gains.

There are two main types of space-time coding techniques: space-time trellis coding (STTC) and space-time block coding (STBC). The key development of the STC concept was originally revealed in [Tar98] in the form of STTC. STTC may not be practical or cost-effective due to the high complexity of the ML detector. In an attempt to reduce complexity, orthogonal STBC that can be optimally decoded with a simple linear processing at the receiver were introduced and analyzed in [Ala98, Tar99a] (see also [Gan00b, Gan01a]). In fact, the orthogonal STBC designed for two transmit antennas in [Ala98] is so simple and useful that it has been included in the technical specifications for the UMTS Terrestrial Radio Access (UTRA) system of the third generation of mobile communication systems [3GP99].

It has to be noted that since 1998, an extraordinary number of publications (which we do not

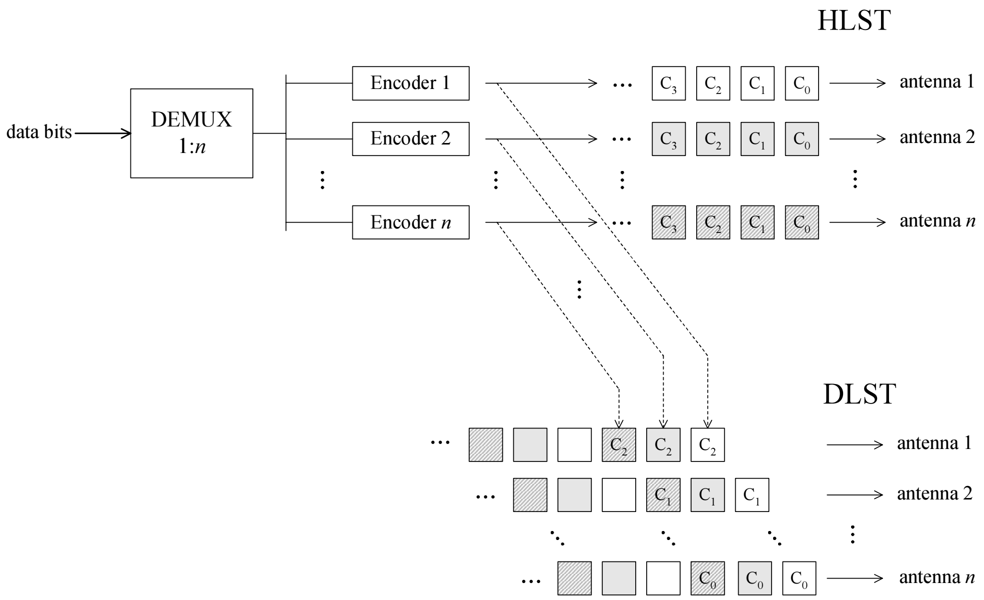


Figure 2.11: General scheme of the horizontal and diagonal layered space-time (HLST and DLST) architectures (also known as BLAST).

reference here) have flourished in the open literature dealing with the design of space-time codes under different performance criteria for different scenarios such as spatially correlated channels (see §2.4.3 for hybrids schemes combining STC with beamforming).

2.4.1.2 Layered Space-Time Architecture

Layered architectures, also known as layered space-time (LST) codes, are constructed by assembling one-dimensional (1-D) constituent codes (see Figure 2.11). The original data stream is demultiplexed into n data streams and then each data stream is encoded in such a way that the encoders can proceed without sharing any information with each other, *i.e.*, independent encoding. Note that CSIT is not required (see Figure 2.9). At the receiver, with the use of interference suppression and interference cancellation, these constituent codes can be separated and then decoded using conventional decoding algorithms developed for 1-D codes, leading to a much lower decoding complexity compared to ML decoding.

A diagonal LST (DLST) architecture was originally proposed by Foschini in 1996 [Fos96] (see Figure 2.11). In [Fos99], the less complex horizontal LST (HLST)[5] architecture was considered and the name BLAST was coined standing for Bell-labs LAyered Space-Time architecture (see Figure 2.11). In [Shi99], the layered architecture was further analyzed, giving error probability expressions and considering different types of constituent codes.

Hybrids strategies that combine layered architectures with space-time processing techniques

[5]The horizontal LST architecture is sometimes in the literature referred to as vertical LST architecture.

constitute a very promising solution in systems with many antennas to reduce complexity while keeping performance. In [Tar99b], space-time codes were used as constituent codes of a layered architecture for systems with a large number of transmit antennas to keep complexity at a reasonable level. In [Ari00], the hybrid scheme went one step further by encompassing turbo space-time processing within the layered architecture.

2.4.2 Transmission Techniques with Perfect CSIT

When perfect CSIT is available, the transmission can be adapted to each channel realization using signal processing techniques (c.f. §2.5.1). Historically speaking, there are two main scenarios that have motivated the development of communication methods for MIMO channels with CSIT: wireline channels and wireless channels.

The initial motivation to design techniques for communication over MIMO channels was originated in wireline systems by treating all the links within a bundle of cables as a whole [Lee76, Sal85, Hon90, Yan94b, Yan94a]. In [Kas90], a vector coding technique consisting on partitioning a frequency-selective SISO channel into an independent set of parallel ISI-free channels by means of the channel eigenvectors was developed. Optimal bit allocation and energy distribution over the set of parallel subchannels, where coset codes were applied, was also given. In [Rui92], the channel eigenvectors were replaced by the discrete multiple tones of the IDFT, leading to a lower complexity scheme. The idea of partitioning can be seen as a generalization and optimization of multitone methods traceable back to at least 1964 (Holsinger). In [AD96, Gin02], a decision-feedback equalizer (DFE) was considered at the receiver and also at the transmitter in the form of a precoder in [Gin02].

Another more recent source of motivation to design methods for communication over MIMO channels can be found in multi-antenna wireless systems [Ral98, And00, AD00, Sam01, Sca02]. In [Ral98], a spatio-temporal vector coding architecture was introduced for burst transmission over MIMO channels. To reduce complexity a simplified version called discrete matrix multitone space-frequency coding structure was also proposed. Transmit and receive linear processing (equivalently, multiple beamforming) was jointly designed in [Sam01, Sca02]. In [AD00], finite-length transmit and receive filters were derived assuming a DFE at the receiver. In [And00], the flat multi-antenna MIMO case was considered providing useful insights from the point of view of beamforming. For the case of single beamforming, a low-complexity transmission method was proposed in which knowledge of the whole channel matrix is not necessary at any side of the link (see Figure 2.9). In [Sca99a, Sca99b], a frequency-selective SISO channel was transformed into a MIMO channel by means of a multirate filterbank scheme.

2.4.3 Transmission Techniques with Partial CSIT

As previously mentioned, it is more reasonable in practice to consider partial CSIT. We now describe two existing techniques to cope with this situation.

2.4.3.1 Hybrid Schemes

The basic idea is to enhance the performance of a space-time code, for which no CSIT is required, by combining it with some type of beamforming or linear precoding, for which the existing (possibly partial) CSIT can be exploited (see Figure 2.9).

In [Neg99], the combination of beamforming and space-time codes in a system with multiple antennas at both sides of the link was considered aiming at maximizing the average SNR for scenarios with partial CSIT. The appealing combination of beamforming and space-time codes with partial CSIT was also taken in [Jön02] from a statistical viewpoint to minimize the pairwise error probability for different degrees of CSIT (from no CSIT to perfect CSIT). In [Hon03] a space-time code was combined with a linear precoding matrix to use the statistical knowledge on the channel. A virtual representation of the MIMO channel was used in [Liu02] to adapt space-time codes to the channel realization to cope with the channel correlation.

2.4.3.2 Robust Beamforming under Channel Estimation Errors

Since perfect CSIT is never available in practice, it is always necessary to quantify the degradation due to channel estimation errors of any system that assumes perfect CSIT (the degradation due to imperfect CSIR should also be quantified, *e.g.*, [Gan01b]). However, it is always better to obtain robust solutions by directly taking into account the existence of channel estimation errors when designing the system (see Figure 2.9). There have been some attempts to obtain robust solutions as we now mention (see also Chapter 7 to see how to take into account channel estimation errors in the beamforming schemes proposed in Chapters 5 and 6).

In [Cox87, Bel00, Vor03], robust designs were obtained in the context of classical receive beamforming. Robust designs of transmit beamforming with imperfect CSIT have been considered according to different criteria such as error probability [Wit95], SNR [Nar98], and mutual information [Nar98, Vis01]. In [Vis01], robust transmission schemes were obtained for imperfect CSIT in the form of either the mean or the covariance matrix of the channel distribution function (see also [Jaf01] for the case of multiple transmit and receive antennas). In [Nar98], imperfect CSIT was considered not only stemming from channel estimation errors but also from the effect of quantizing the channel estimate at the receiver to be fed back as side information.

Partial CSI for the design of single-antenna systems modeled as MIMO channels was considered in [Bar01], where a design maximizing the average SNR was adopted. A stochastic robust design for MIMO channels was obtained in [Mil00] to minimize the sum of MSE's with imperfect CSIT and perfect CSIR (a ZF receiver was assumed). In [Rey02], a robust solution to the minimization of the sum of the MSE's was derived for multi-antenna multicarrier systems with imperfect CSIT and CSIR. In [Ise03], a robust minimum BER solution was designed for single-antenna multicarrier systems with imperfect CSIT and perfect CSIR. In [Ben99, Ben01], a worst-case robust approach was taken in the context of multiuser communications between several mobiles and a multi-antenna base station.

2.4.4 The Multiuser Case

The more general case of multiuser communications has also been studied in the literature although to a lesser extent. Interestingly, beamforming happens to be optimal for multiuser MIMO systems provided that the number of users is sufficiently high [Rhe01b].

In [RF98a, RF98b, Vis99a], beamforming techniques were derived for uplink and downlink communications in multiuser systems with multiple antennas only at the base station. In [Ben99, Ben01], a more general scenario including robustness under channel estimation errors was considered.

In [Vis02], an interesting downlink transmission scheme was proposed based on opportunistic beamforming by transmitting at each time to a single user.

2.5 Linear Signal Processing for MIMO Channels

2.5.1 Linear Signal Processing

In a communication system, the optimal receiver is given by the Maximum Likelihood (ML) detector, also known as Maximum Likelihood Sequence Estimation (MLSE) [Pro95]. The ML detector is a nonlinear signal processing method and has an exponential complexity in the number of channel dimensions or transmitted symbols. If the structure of the channel matrix is adequate, *i.e.*, if it can be modeled as a hidden Markov model, the MLSE can be efficiently implemented with the Viterbi algorithm [Pro95]. To further reduce the complexity, other suboptimal nonlinear signal processing techniques can be used such as the popular Decision-Feedback (DF) scheme [Qur85, GDF91, Yan94a, Cio95a, Cio95b, Ari99] whose main drawback is the error propagation due to the feedback mainly at low SNR. A more pragmatic solution is the utilization of linear signal processing techniques which have low complexity and no error feedback problems at the expense of a worse performance. We now focus on linear transmit-receive processing schemes for MIMO systems. The meaning of the term beamforming, initially used for smart antennas[6]

[6]By smart antennas we refer to the classical scenario in which only one side of the communication link has multiple antennas.

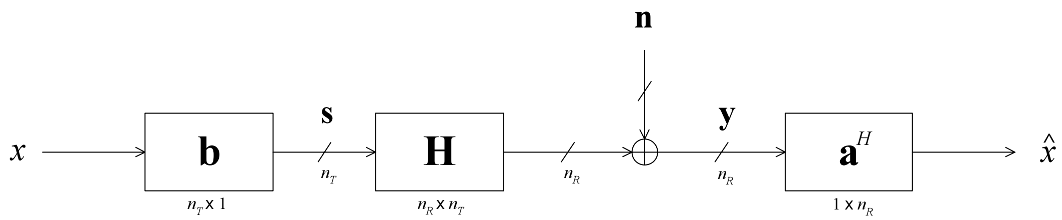


Figure 2.12: Scheme of a single beamforming strategy for a MIMO channel.

applications, is hereinafter generalized to refer to a linear processing scheme along an arbitrary dimension (not necessarily the spatial dimension), becoming then a *virtual beamforming.* Note that linear processing at the receiver and at the transmitter are commonly referred to as *linear equalization* and *linear precoding*, respectively.

2.5.1.1 Single Beamforming

In this subsection, we consider the simple case of single beamforming for MIMO channels, *i.e.*, the utilization of a single beamvector at the transmitter and a single beamvector at the receiver. Note that this concept is the natural extension of the classical beamforming approach used for smart antennas to the case of MIMO channels. In other words, single beamforming means that a single symbol is transmitted at each transmission through the MIMO channel. Therefore, only one data stream needs to be considered (as opposed to multiple beamforming approach as discussed next in §2.5.1.2) and coding and transmission can be done in a much easier manner (as in traditional single-antenna systems) [Nar99]. In addition, for the particular case of multi-antenna MIMO systems, beamforming becomes asymptotically optimal as the spatial channel fading correlation increases at least at one end of the communication link (see §4.2 and [Pal03e] for more details).

The transmitted vector when using a beamvector $\mathbf{b} \in \mathbb{C}^{n_T \times 1}$ at the transmitter (see Figure 2.12) is

$$\mathbf{s} = \mathbf{b}x \tag{2.25}$$

where $x \in \mathbb{C}$ is the scalar data symbol to be transmitted assumed zero-mean and with unit energy $\mathbb{E}\left[|x|^2\right] = 1$ without loss of generality (w.l.o.g.).[7] Assuming that the receiver uses a beamvector $\mathbf{a}^H \in \mathbb{C}^{1 \times n_R}$ (which can be assumed without loss of optimality in the sense of sufficient statistics, provided that the transmitter is also using a beamvector), the estimated data symbol is

$$\hat{x} = \mathbf{a}^H \mathbf{y}. \tag{2.26}$$

[7] The mean of the signal does not carry any information and can always be set to zero saving power at the transmitter. If the symbols do not have unit energy, they can always be normalized them and include the scaling factor in $\mathbf{b}$.

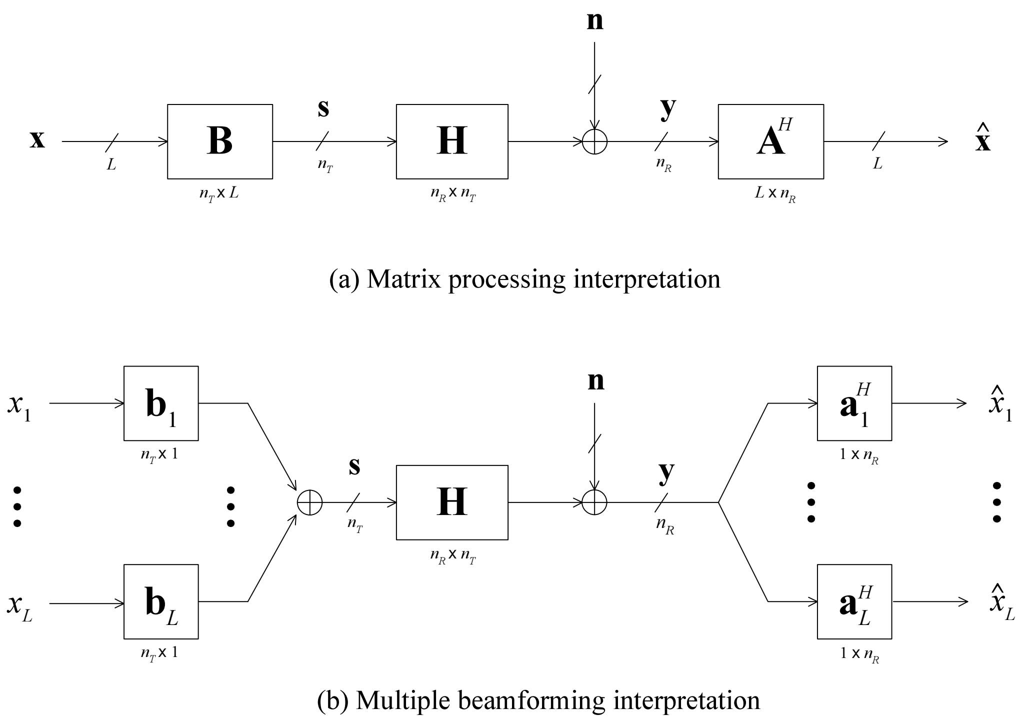


Figure 2.13: Scheme of a matrix-processing/multiple-beamforming strategy for a MIMO channel.

The signal model in (2.25) and (2.26) corresponds to a single transmission. Recall that a real communication system is composed of multiple transmissions that can be represented by explicitly using a discrete-time index n as $\mathbf{s}(n) = \mathbf{b}x(n)$ and $\hat{x}(n) = \mathbf{a}^H\mathbf{y}(n)$. For the sake of notation, however, we omit the utilization of an explicit indexing.

Similarly, for the more general case of having a set of N parallel and independent MIMO channels as in (2.3), the signal model at the kth MIMO channel is

$$\mathbf{s}_k = \mathbf{b}_k x_k \tag{2.27}$$

$$\hat{x}_k = \mathbf{a}_k^H \mathbf{y}_k \tag{2.28}$$

where zero-mean unit-energy symbols ($i.e.$, $\mathbb{E}[|x_k|^2] = 1$) are assumed w.l.o.g.

2.5.1.2 Multiple Beamforming

We now extend the concept of single beamforming of §2.5.1.1 to the more general case of multiple beamforming or matrix beamforming at both sides of the link. Multiple beamforming means that $L > 1$ symbols are simultaneously transmitted through the MIMO channel. Recall that a channel matrix with n_T transmit and n_R receive dimensions has $K \leq \min(n_T, n_R)$ channel eigenmodes (or, equivalently, nonvanishing singular values of the channel matrix) that can be used as a means of spatial multiplexing to transmit simultaneously L symbols by establishing L substreams (c.f.

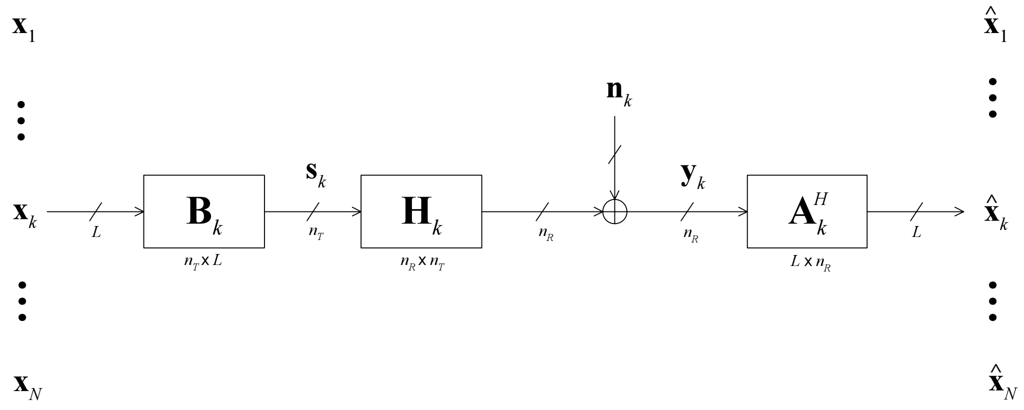


Figure 2.14: Scheme of a non-cooperative multiple beamforming strategy for a set of N parallel MIMO channels.

§2.3.3). For generality, we allow for arbitrary values of L; in a practical system we will typically have $L \leq K$ to have an acceptable performance.

The transmitted vector when using a transmit linear processing matrix $\mathbf{B} \in \mathbb{C}^{n_T \times L}$ (see Figure 2.13(a)) is

$$\mathbf{s} = \mathbf{B}\mathbf{x} = \sum_{i=1}^{L} \mathbf{b}_i x_i \tag{2.29}$$

where $\mathbf{x} = [x_1 \cdots x_L]^T \in \mathbb{C}^{L \times 1}$ is the vector of L data symbols assumed zero-mean, with unit energy, and uncorrelated (white) [8] $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}_L$ w.l.o.g. Vector $\mathbf{b}_i$ is the ith column of matrix $\mathbf{B}$ and can be regarded as the beamvector associated to the ith data symbol x_i, *i.e.*, a multiple beamforming architecture (see Figure 2.13(b)). Assuming that the receiver uses the receive linear processing matrix $\mathbf{A}^H \in \mathbb{C}^{L \times n_R}$, the estimated data vector is

$$\hat{\mathbf{x}} = \mathbf{A}^H \mathbf{y}. \tag{2.30}$$

Consider now the more general case of having a set of N parallel and independent MIMO channels as in (2.3). To take into account the possibility of transmitting a different number of symbols through each MIMO channel, we consider that L_k symbols are transmitted at the kth MIMO channel and, therefore, a total of $L_T = \sum_{k=1}^{N} L_k$ symbols are transmitted. Note that the total number of transmit and receive dimensions is now $n_T N$ and $n_R N$, respectively. Regarding the linear processing at the transmitter and at the receiver, two different cases are considered as we now describe [Pal03c]:

- *Noncooperative scheme*: this is when an independent signal processing per MIMO channel is used such as in a multi-antenna muticarrier system with independent processing at each

[8] White symbols account, for example, to having independent bit streams. In case of having colored symbols due, for example, to a coded transmission [GDF91], a prewhitening operation can be performed prior to precoding at the transmitter and the corresponding inverse operation can be performed after the equalizer at the receiver.

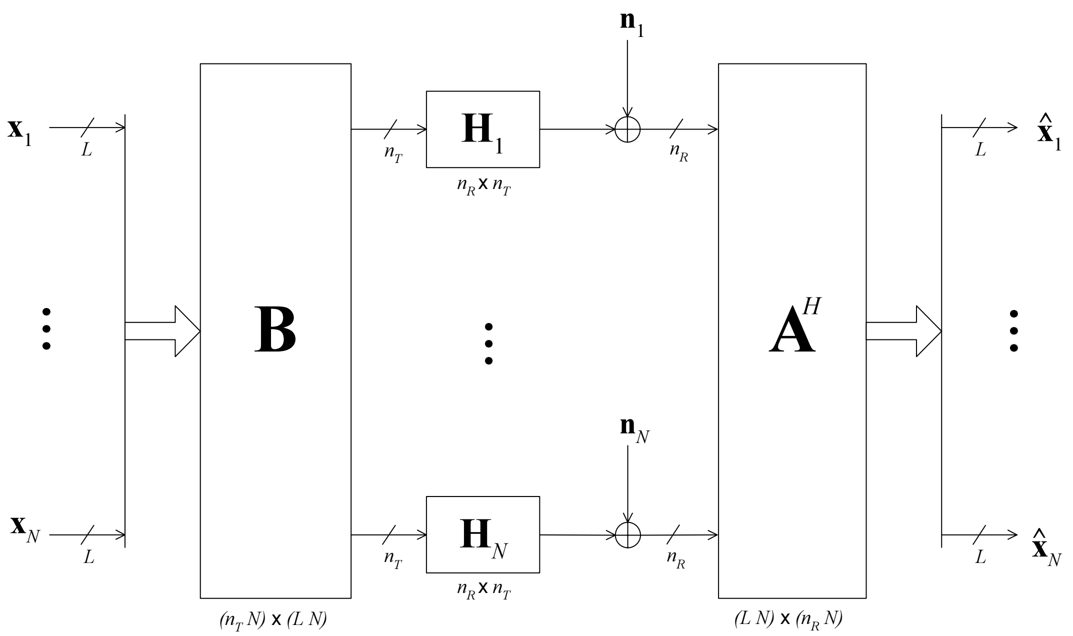


Figure 2.15: Scheme of a cooperative multiple beamforming strategy for a set of N parallel MIMO channels.

carrier (see Figure 2.14). The transmit and receive signal model at the kth MIMO channel considering the simultaneous transmission of L_k symbols by establishing L_k substreams is

$$\mathbf{s}_k = \mathbf{B}_k \mathbf{x}_k = \sum_{i=1}^{L_k} \mathbf{b}_{k,i} x_{k,i} \tag{2.31}$$

$$\hat{\mathbf{x}}_k = \mathbf{A}_k^H \mathbf{y}_k \tag{2.32}$$

where all the quantities are defined as for the single MIMO case and zero-mean unit-energy uncorrelated symbols (*i.e.*, $\mathbb{E}\left[\mathbf{x}_k \mathbf{x}_k^H\right] = \mathbf{I}_{L_k}$) are assumed w.l.o.g. The total average transmitted power is $P_T = \sum_k \mathrm{Tr}(\mathbf{B}_k \mathbf{B}_k^H)$.

- *Cooperative scheme*: this is a more general linear processing scheme that allows for co-operation or joint processing among the MIMO channels (see Figure 2.15). The signal model is obtained by stacking the vectors corresponding to all MIMO channels (*e.g.*, $\mathbf{x}^T = \left[\mathbf{x}_1^T, \cdots, \mathbf{x}_N^T\right]$), by considering global transmit and receive matrices $\mathbf{B} \in \mathbb{C}^{(n_T N) \times L_T}$ and $\mathbf{A}^H \in \mathbb{C}^{L_T \times (n_R N)}$, and by defining the global channel as $\mathbf{H} = \mathrm{diag}\left(\{\mathbf{H}_k\}\right) \in \mathbb{C}^{(n_R N) \times (n_T N)}$. This general block processing scheme was used in [Ral98] to obtain a capacity-achieving system.

It is important to realize that the noncooperative processing model of (2.31) and (2.32) can be obtained from the general model (2.29) and (2.30) by setting $\mathbf{B} = \mathrm{diag}\left(\{\mathbf{B}_k\}\right)$ and $\mathbf{A} = \mathrm{diag}\left(\{\mathbf{A}_k\}\right)$, *i.e.*, by imposing a block-diagonal structure on $\mathbf{B}$ and $\mathbf{A}$. This connection makes clear why the noncooperative scheme, due to the block-diagonal structure constraint in $\mathbf{B}$ and $\mathbf{A}$, cannot do better than the cooperative approach in which $\mathbf{B}$ and $\mathbf{A}$ are structurally unconstrained. Intuitively, the reason why the cooperative approach (see Figure 2.15) has potentially a better performance is that it can reallocate the symbols among the MIMO channels in an intelligent way (*e.g.*, if one MIMO channel is a bad channel, it will try to use other MIMO channels instead), whereas the noncooperative scheme (see Figure 2.14) will always transmit L_k symbols through the kth MIMO channel no matter the channel state.

2.5.2 Transmit Power Constraint

In this dissertation, we consider systems limited in the average total transmitted power. In particular, a constraint on the average power utilized to transmit the N-symbol (or block of symbols) in the single beamforming case is

$$\sum_{k=1}^{N} \mathbb{E}\left[\, \|\mathbf{s}_k\|_2^2 \,\right] = \sum_{k=1}^{N} \|\mathbf{b}_k\|_2^2 \leq P_T \tag{2.33}$$

where P_T is power in units of energy per transmission (the power in units of energy per second is $P_s = P_T/T_s$ where T_s is the duration of a transmission). Similarly, a constraint on the average power utilized to transmit the L_T-symbol (or block of symbols) in the multiple beamforming case is

$$\sum_{k=1}^{N} \mathbb{E}\left[\, \|\mathbf{s}_k\|_2^2 \,\right] = \sum_{k=1}^{N} \|\mathbf{B}_k\|_F^2 = \sum_{k=1}^{N} \mathrm{Tr}\left(\mathbf{B}_k \mathbf{B}_k^H\right) \leq P_T. \tag{2.34}$$

Note that the power constraint for the case of a single MIMO channel is readily obtained particularizing for $N = 1$.

Whenever the channel is random, the previous power constraint (either (2.33) or (2.34)) corresponds to a short-term power constraint in the sense that for each set of transmit matrices (in general designed as a function of the channel state), the power averaged over the possible transmitted symbols is constrained by P_T. On the other hand, a long-term power constraint can be considered by further averaging over all channel states. Such a constraint allows the transmitted power to exceed P_T for some channel states as long as it is compensated for some other channel states so that the power averaged over all channel states does not exceed P_T [Cai99b, Big01]. More generally, the power constraint can be averaged over N_T transmissions which contains as particular cases the short-term power constraint (for $N_T = 1$) and the long-term power constraint (for $N_T \to \infty$). The higher the value of N_T, the less restrictive is the power constraint (the long-term power constraint is clearly the least restrictive). A general case of $N_T > 1$, however, requires knowledge of some future realizations of the channel (or of the channel statistics for $N_T \to \infty$) [Bar01]. Note that the case of a finite $N_T > 1$ can be easily included in the short-term power constraint of (2.33)-(2.34) by defining a block-diagonal channel matrix that contains along the main diagonal the channel matrices corresponding to the N_T transmissions. In the limiting case of $N_T \to \infty$ a similar approach can be taken if the probability density function (pdf) of the channel eigenvalues is known [Gol97, Bar01].

For completeness, we now list other possible types of power constraints that may be interesting depending on the specific scenario:

- Power constraint per MIMO channel. This only applies to the case of multiple MIMO channels and consists on further constraining the power used in each MIMO channel: $\|\mathbf{B}_k\|_F^2 \leq P_k$.

- Power constraint per dimension. Similarly, the power used in the ith dimension can be constrained: $\sum_{k=1}^{N} \mathbb{E}\left[\,|s_{k,i}|^2\right] = \sum_{k=1}^{N} \|\,[\mathbf{B}_k]_{i,:}\,\|_2^2 \leq P_i$ (this constraint will be considered in Chapter 5 in the context of multi-antenna systems to control the dynamic range of the signal transmitted by each antenna).

- Peak power constraint per MIMO channel. Noting that $\max_i |s_{k,i}|^2 \leq \mathbf{s}_k^H \mathbf{s}_k = \mathbf{x}_k^H \mathbf{B}_k^H \mathbf{B}_k \mathbf{x}_k \leq \lambda_{\max}\left(\mathbf{B}_k^H \mathbf{B}_k\right) \|\mathbf{x}_k\|_2^2$, the average peak power can be controlled by constraining the maximum eigenvalue of the squared transmit matrix: $\mathbb{E}\left[\,\max_i |s_{k,i}|^2\right] \leq \lambda_{\max}\left(\mathbf{B}_k^H \mathbf{B}_k\right) L_k \leq P_k^{\text{peak}}$ [Gan00a, Sca02] (this constraint is briefly considered in Chapter 5).

- Peak power constraint. Similarly to the peak power constraint per MIMO channel, this one constrains the maximum eigenvalue of all MIMO channels: $\max_k \left\{\lambda_{\max}\left(\mathbf{B}_k^H \mathbf{B}_k\right) L_k\right\} \leq P_T^{\text{peak}}$.

- Peak average power constraint. This variation of the eigenvalue constraint is useful for multicarrier systems where the signals corresponding to different MIMO channel are added together: $\max_i \sum_{k=1}^{N} \mathbb{E}\left[\,|s_{k,i}|^2\right] \leq \sum_{k=1}^{N} \mathbb{E}\left[\mathbf{x}_k^H \mathbf{B}_k^H \mathbf{B}_k \mathbf{x}_k\right] \leq \sum_{k=1}^{N} \lambda_{\max}\left(\mathbf{B}_k^H \mathbf{B}_k\right) L_k \leq P_T^{\text{peak-ave}}$.

- Instantaneous power constraint. This is in fact a family of power constraints based on constraining the system for each realization of the data symbols $\mathbf{x}_k$ instead of averaging them. For example, an instantaneous total power constraint would be $\sum_{k=1}^{N} \|\mathbf{s}_k\|_2^2 = \sum_{k=1}^{N} \mathbf{x}_k^H \mathbf{B}_k^H \mathbf{B}_k \mathbf{x}_k \leq P_T^{\text{inst}}$. This type of constraints is very difficult to handle.

- Worst-case power constraint. The instantaneous power constraint can be simplified by considering that for each channel state, the transmitter has to be designed so that the constraint is satisfied for all possible $\mathbf{x}_k$'s: $\max_{\{\mathbf{x}_k\}} \sum_{k=1}^{N} \mathbf{x}_k^H \mathbf{B}_k^H \mathbf{B}_k \mathbf{x}_k \leq P_T^{\text{wc}}$.

2.5.3 Canonical Channel Model

The canonical channel model is a reduction of the original channel model in which extraneous dimensions are eliminated and which have other nice properties such as having a positive definite Hermitian channel matrix and a noise with covariance matrix identical to the canonical channel (see [Cio97] for a detailed treatment of the canonical channel model). The canonical channel

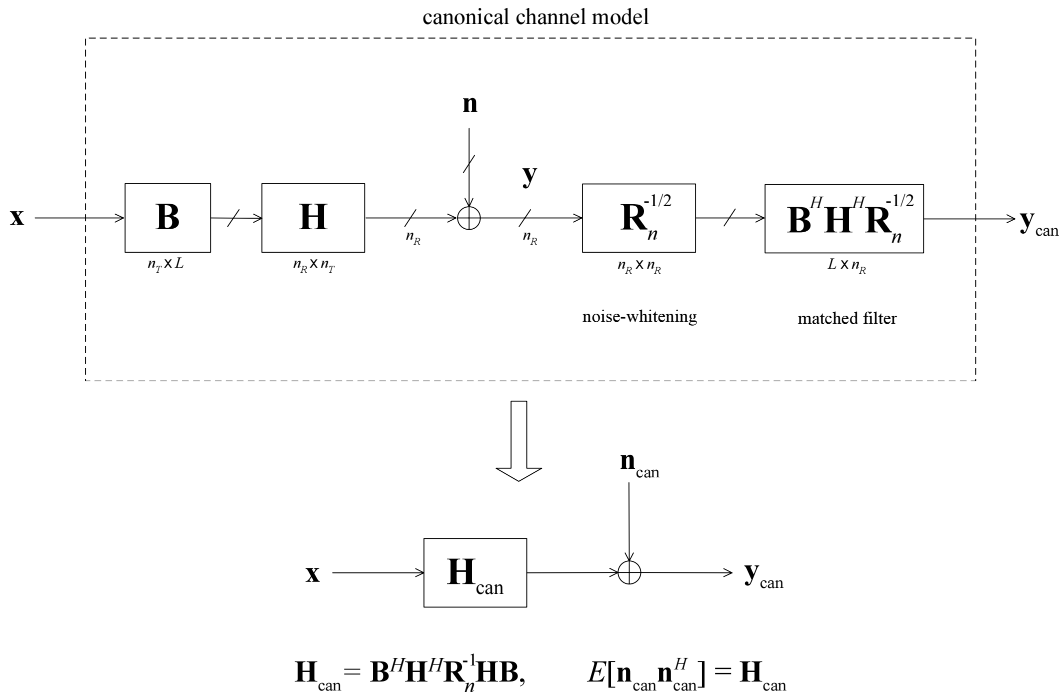


Figure 2.16: Scheme of the canonical channel model.

model is known to produce sufficient statistics for the estimation of the transmitted symbols [Sch91, Ch. 3][Cio97] and to be capacity-lossless from an information-theoretic point of view [Cio97].

It is obtained by a series of information-lossless linear transformations on the channel output $\mathbf{y}$. The first transformation consists of a noise-whitening stage $\mathbf{R}_n^{-1/2}$ producing the global channel $\mathbf{C} = \mathbf{R}_n^{-1/2}\mathbf{HB}$. Assume for the moment that $L \leq \text{rank}(\mathbf{C})$. The principle of the sufficiency of matched filtering may be now applied and the second stage is the matched filter $\mathbf{C}^H = \mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1/2}$. The canonical output signal after these two stages is

$$\mathbf{y}_{\text{can}} = \mathbf{H}_{\text{can}}\mathbf{x} + \mathbf{n}_{\text{can}} \tag{2.35}$$

where $\mathbf{H}_{\text{can}} = \mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{HB}$ is the canonical channel and $\mathbb{E}\left[\mathbf{n}_{\text{can}}\mathbf{n}_{\text{can}}^H\right] = \mathbf{H}_{\text{can}}$ is the covariance matrix of the canonical noise (see Figure 2.16).

For the particular case of single beamforming, the canonical channel reduces to a scalar channel $h_{\text{can}} = \mathbf{b}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{Hb}$ with a scalar noise n_{can} with covariance equal to the canonical channel $\mathbb{E}\left[|n_{\text{can}}|^2\right] = h_{\text{can}}$.

Consider now the case $L > \text{rank}(\mathbf{C})$ and define the effective number of transmit dimensions as $\check{L} = \min(L, \text{rank}(\mathbf{C}))$. It is clear that the component of $\mathbf{x}$ that lies in the right null space (kernel) of $\mathbf{C}$, *i.e.*, $\mathbf{P}_c^{\perp}\mathbf{x}$ where $\mathbf{P}_c^{\perp}$ is the projection onto the kernel, may be disregarded since it is undetectable. In other words, the output signal depends only on the effective input $\mathbf{x}_{\text{eff}} = \mathbf{P}_c\mathbf{x}$

and, therefore, only $\mathbf{x}_{\text{eff}}$ can convey information (note that this does not mean that more than $\check{L}$ symbols cannot be transmitted, although they will not be perfectly separable[9]). To be more exact, using the singular value decomposition (SVD) $\mathbf{C} = \mathbf{U\Sigma V}^H$, the output at the receiver depends only on $\mathbf{x}_{\text{eff}} = \mathbf{V}_1\mathbf{V}_1^H\mathbf{x}$ where $\mathbf{V}_1$ contains the right singular vectors of $\mathbf{C}$ associated with nonzero singular values.

2.5.4 Figures of Merit: MSE, SINR, BER, and Capacity

We now describe the four common measures of the performance of a communication system: MSE, SINR, BER, and capacity. The MSE, SINR, and BER arise naturally from a estimation framework. The notion of channel capacity, however, comes from a rather different information-theoretic perspective.

2.5.4.1 Average vs. Outage Quantities

Regardless of the specific measure that is used to quantify the performance of a system, when the channel is deterministic or when it changes at a very slow pace (such as in some wireline systems), the measurement is given by a single value. However, when the channel is random, due for example to fading in wireless channels [Pro95], the measurement becomes a random quantity. To properly characterize the system performance in such cases, the pdf of the measurement should be obtained. In practice, however, it is more convenient to characterize a system with a single value rather than with the whole pdf. The two most common ways of doing it is with the average value and with the outage value of the measurement whose choice depends on whether the scenario is ergodic or non-ergodic.

The ergodic scenario arises in communications without delay constraints in which the transmission duration is long enough to reveal the long-term ergodic properties of the fading channel (assuming that the channel is an ergodic process in time). In such cases, average values such as the average BER or the average capacity (also termed ergodic capacity) are useful measures.

The ergodicity assumption, however, is not necessarily satisfied in practical communication system with stringent delay constraints operating on fading channels, because no significant channel variability may occur during the whole transmission. In these circumstances, outage values are more appropriate and meaningful than average values which give more conservative

[9] As a simple illustrative example, consider the following transmission of 101 symbols through 100 parallel SISO channels with unit gain: transmit 100 symbols independently on each of the 100 SISO channels and the last symbol simultaneously through all SISO channel (scaled down by a properly chosen factor) in a CDMA fashion. The first 100 symbols will see a slight increase of noise and the last symbol using matched filter receiver in a CDMA fashion will benefit from the processing gain.

and unrealistic quantities. The outage value of a random variable is defined as the best value[10] that can be achieved with a high probability of $(1 - P_{\text{out}})$ or, in other words, the best value that is not achieved only with a small probability of P_{out} (termed outage probability). Parameterizing the outage value with respect to the outage probability yields a whole curve of performance vs. outage (the outage value at some given outage probability is just one point of this curve).

Practical wireless systems are in general non-ergodic scenarios and therefore outage values are more realistic than average ones.

2.5.4.2 MSE

The mean square error (MSE) corresponding to the estimation $\hat{x}$ of the transmitted symbol x is defined as

$$\text{MSE} \triangleq \mathbb{E}\left[|\hat{x} - x|^2\right] \tag{2.36}$$

and is bounded by $0 < \text{MSE} \leq 1$. The upper bound can always be achieved by setting $\hat{x} = 0$ (recall that $\mathbb{E}\left[|x|^2\right] = 1$) which means no effective communication.[11] The lower bound cannot be achieved unless the noise variance is zero (noiseless channel), which is never the case in any real system (note that a zero MSE implies that an infinite amount of information can be transmitted).

The smaller the MSE the better the system, since the estimation matches more closely the desired value. Hence, any reasonable system has to be designed to have a low MSE.

2.5.4.3 SINR

The signal to interference-plus-noise ratio (SINR) corresponding to the received signal $y = \alpha x + n$ is defined as the ratio between the desired component and the undesired component (assuming normalized symbols $\mathbb{E}\left[|x|^2\right] = 1$ and $\sigma_n^2 \triangleq \mathbb{E}\left[|n|^2\right]$):

$$\text{SINR} \triangleq \frac{|\alpha|^2}{\sigma_n^2} \tag{2.37}$$

and is bounded by $0 < \text{SINR} < \infty$. The upper bound cannot be achieved unless the channel is noiseless and, similarly, the lower bound cannot be achieved unless the signal component is zero which is never the case unless the transmitter stops the transmission. A closely related measure is the E_b/N_0 defined as the SINR per bit (assuming that N_0 includes all the interfering signals).

The higher the SINR the better the system, since it implies a higher ratio of the desired to undesired components. Hence, any reasonable system has to be designed to have a high SINR.

[10] Note that depending on the specific measure, the best value may be the maximum value (SINR and capacity) or the minimum value (MSE and BER).

[11] In principle, it is possible to obtain values of the MSE higher than one simply by using a terrible receiver (for example, by setting the estimation to $\hat{x} = 837$). However, we focus on reasonable systems.

2.5.4.4 BER

The ultimate performance of a digital communication system is given in terms of symbol error probability (fraction of symbols in error) or bit error probability (fraction of bits in error). We define the bit error rate (BER) as the bit error probability.[12]

Assuming that the interference-plus-noise component is Gaussian distributed,[13] the symbol error probability P_e can be analytically expressed as a function of the SINR [Pro95]:

$$P_e = \alpha\, \mathcal{Q}\left(\sqrt{\beta\ \mathrm{SINR}}\right) \tag{2.38}$$

where α and β are constants that depend on the signal constellation (see Appendix 2.A for more details) and $\mathcal{Q}$ is the $\mathcal{Q}$-function defined as $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\lambda^2/2} d\lambda$ [Pro95].[14]

It is sometimes convenient to use the Chernoff upper bound of the tail of the Gaussian distribution function $\mathcal{Q}(x) \leq \frac{1}{2} e^{-x^2/2}$ [Ver98] to approximate the symbol error probability (which becomes a reasonable approximation for high values of the SINR) as

$$P_e \approx \frac{1}{2} \alpha\, e^{-\beta/2\ \mathrm{SINR}}. \tag{2.39}$$

The BER can be approximately obtained from the symbol error probability (assuming that a Gray encoding is used to map the bits into the constellation points) as

$$\mathrm{BER} \approx P_e / k \tag{2.40}$$

where $k = \log_2 M$ is the number of bits per symbol and M is the constellation size. The BER is bounded by $0 < \mathrm{BER} \leq 0.5$. The lower bound cannot be achieved unless the channel is noiseless and the upper bound can always be achieved by randomly choosing the estimated bit which means no effective communication.[15] In [Cho02] a more exact approximation of the BER than that obtained combining (2.38) and (2.40) can be found.

The lower the BER the better the system, since it means that less errors are made when estimating the transmitted bits. Hence, any reasonable system has to be designed to have a low BER.

[12] When dealing with coded systems, the ultimate measure is the coded BER as opposed to the uncoded BER (obtained without using any coding). However, the coded BER is strongly related to the uncoded BER (in fact, for codes based on hard decisions, both quantities are strictly related). Therefore, it is in general sufficient to focus on the uncoded BER when optimizing the uncoded part of a communication system (keep in mind, however, that with soft-input decoding, it may not be good enough to deal just with the uncoded BER).

[13] The Gaussian assumption is a reasonable approximation even when each of the interfering signals is not Gaussian distributed as long as the number of interfering signals is sufficiently high (since the total interference contribution tends to have a Gaussian distribution as the number of interfering signals grows due to the central limit theorem).

[14] The $\mathcal{Q}$-function and the complementary error function erfc are related as $\mathrm{erfc}(x) = 2\, \mathcal{Q}\left(\sqrt{2}x\right)$.

[15] In principle, as happened with the upper bound of the MSE, it is possible to obtain values of the BER higher than 0.5 simply by using a terrible receiver (for example, by flipping the binary output of any properly designed system). However, we focus on reasonable systems.

2.5.4.5 Capacity

The notion of capacity comes from an information-theoretic point of view. The capacity of a channel is defined as the maximum rate of information that can be reliably transmitted [Sha48]. The adjective *reliably* means that the probability of error can be made as small as desired by using a sufficiently long transmission block.

To achieve the capacity of a channel, the transmitter has to be carefully designed by using an adequate power allocation and sufficiently long Gaussian-distributed codewords (c.f. Chapter 4). At the receiver, the optimal maximum likelihood (ML) estimation of the transmitted codewords is generally assumed. Interestingly, the receiver admits any transformation of the received signal without any loss of capacity as long as it is information-lossless [Cov91b].

Clearly, the higher the information rate the better the system, since it means that more information can be transmitted. Hence, any reasonable system has to be designed to have a high information rate (as close as possible to the channel capacity).

2.5.5 Optimum Linear Receiver: The Wiener Filter

We now obtain the optimum linear receiver to optimize the performance of the system in terms of either MSE, SINR, BER, and capacity (the design of the transmitter is far more involved and is the scope of Chapters 5 and 6).

It suffices to focus on the case of a single MIMO channel since for multiple MIMO channels the derivation can be done independently for each MIMO channel (recall that the MIMO channels do not interfere with each other).

2.5.5.1 Optimum Linear Receiver in Terms of MSE

Consider first the single beamforming signal model given by (2.1), (2.25), and (2.26), *i.e.*, $\hat{x} = \mathbf{a}^H (\mathbf{H}\mathbf{b}x + \mathbf{n})$. The MSE is

$$\begin{aligned} \mathrm{MSE}(\mathbf{b}, \mathbf{a}) &= |\mathbf{a}^H \mathbf{H}\mathbf{b} - 1|^2 + \mathbf{a}^H \mathbf{R}_n \mathbf{a} && (2.41) \\ &= \mathbf{a}^H \left(\mathbf{H}\mathbf{b}\mathbf{b}^H \mathbf{H}^H + \mathbf{R}_n\right) \mathbf{a} + 1 - \mathbf{a}^H \mathbf{H}\mathbf{b} - \mathbf{b}^H \mathbf{H}^H \mathbf{a}. && (2.42) \end{aligned}$$

Given $\mathbf{b}$, the optimum receive beamvector $\mathbf{a}$ is easily found by setting the gradient of the MSE to zero,[16] obtaining the solution

$$\mathbf{a}^{\star} = \left(\mathbf{H}\mathbf{b}\mathbf{b}^H\mathbf{H}^H + \mathbf{R}_n\right)^{-1}\mathbf{H}\mathbf{b} \tag{2.43}$$

$$= \mathbf{R}_n^{-1}\mathbf{H}\mathbf{b}\frac{1}{1+\mathbf{b}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{b}} \tag{2.44}$$

where the second expression is obtained after applying the matrix inversion lemma (see §3.3).[17] Using the optimum receive beamvector (2.43), the MSE expression (2.42) reduces to

$$\begin{aligned}\mathrm{MSE}(\mathbf{b}) &\triangleq \mathrm{MSE}(\mathbf{b},\mathbf{a}^{\star}) \\ &= \frac{1}{1+\mathbf{b}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{b}}.\end{aligned} \tag{2.45}$$

Now consider the multiple beamforming signal model given by (2.1), (2.29), and (2.30), *i.e.*, $\hat{\mathbf{x}} = \mathbf{A}^H(\mathbf{H}\mathbf{B}\mathbf{x}+\mathbf{n})$. The MSE matrix $\mathbf{E}$ is defined as the covariance matrix of the error vector (given by $\mathbf{e} \triangleq \hat{\mathbf{x}} - \mathbf{x}$):

$$\begin{aligned}\mathbf{E}(\mathbf{B},\mathbf{A}) &\triangleq \mathbb{E}\left[(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^H\right] \\ &= \left(\mathbf{A}^H\mathbf{H}\mathbf{B}-\mathbf{I}\right)\left(\mathbf{B}^H\mathbf{H}^H\mathbf{A}-\mathbf{I}\right)+\mathbf{A}^H\mathbf{R}_n\mathbf{A} \\ &= \mathbf{A}^H\left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H+\mathbf{R}_n\right)\mathbf{A}+\mathbf{I}-\mathbf{A}^H\mathbf{H}\mathbf{B}-\mathbf{B}^H\mathbf{H}^H\mathbf{A}\end{aligned} \tag{2.46}$$

from which the MSE of the ith substream is given by the ith diagonal element of matrix $\mathbf{E}$:

$$\begin{aligned}\mathrm{MSE}_i(\mathbf{B},\mathbf{a}_i) &= [\mathbf{E}]_{ii} \\ &= \mathbf{a}_i^H\left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H+\mathbf{R}_n\right)\mathbf{a}_i+1-\mathbf{a}_i^H\mathbf{H}\mathbf{b}_i-\mathbf{b}_i^H\mathbf{H}^H\mathbf{a}_i \\ &= \mathbf{a}_i^H\left(\mathbf{H}\mathbf{b}_i\mathbf{b}_i^H\mathbf{H}^H+\mathbf{R}_{n_i}\right)\mathbf{a}_i+1-\mathbf{a}_i^H\mathbf{H}\mathbf{b}_i-\mathbf{b}_i^H\mathbf{H}^H\mathbf{a}_i\end{aligned} \tag{2.47}$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ are the ith columns of $\mathbf{A}$ and $\mathbf{B}$, respectively, and $\mathbf{R}_{n_i} \triangleq \mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H + \mathbf{R}_n - \mathbf{H}\mathbf{b}_i\mathbf{b}_i^H\mathbf{H}^H$ is the noise covariance matrix seen by the ith substream. Note that (2.47) is identical in form to (2.42) in the single beamforming case.

The derivation of the receive beamvectors $\mathbf{a}_i$'s can be independently done for each of the substreams since the choice of the receive beamvector for one substream does not affect the others as can be seen from (2.47). The optimum receive beamvectors are given similarly to (2.43) by $\mathbf{a}_i^{\star} = \left(\mathbf{H}\mathbf{b}_i\mathbf{b}_i^H\mathbf{H}^H + \mathbf{R}_{n_i}\right)^{-1}\mathbf{H}\mathbf{b}_i$ which can be compactly written as the receive matrix

$$\mathbf{A}^{\star} = \left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H + \mathbf{R}_n\right)^{-1}\mathbf{H}\mathbf{B} \tag{2.48}$$

$$= \mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\left(\mathbf{I}+\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}. \tag{2.49}$$

[16]Since the considered function (2.42) is not analytic, we use the well-known definition of the complex gradient operator which is very convenient, among other things, to determine the stationary points of a real-valued scalar function of a complex vector/matrix [Bra83]. The same applies verbatim to all the non-analytic functions considered in this dissertation.

[17]In practice, it is simpler to implement (2.43) than (2.44), because (2.43) requires knowledge of the covariance matrix of the received signal (which can be easily estimated) rather than of the noise covariance matrix as in (2.44).

Expression (2.48) (as well as (2.43)) is the linear minimum MSE (LMMSE) filter or *Wiener filter* [Kay93] and minimizes simultaneously all the diagonal elements of the MSE matrix $\mathbf{E}$ as can be checked by "completing the squares" as follows (from (2.46)):

$$\begin{aligned}\mathbf{E}(\mathbf{B},\mathbf{A}) &= (\mathbf{A}-(\mathbf{HBB}^H\mathbf{H}^H+\mathbf{R}_n)^{-1}\mathbf{HB})^H(\mathbf{HBB}^H\mathbf{H}^H+\mathbf{R}_n)(\mathbf{A}-(\mathbf{HBB}^H\mathbf{H}^H+\mathbf{R}_n)^{-1}\mathbf{HB})\\ &\quad+\mathbf{I}-\mathbf{B}^H\mathbf{H}^H\left(\mathbf{HBB}^H\mathbf{H}^H+\mathbf{R}_n\right)^{-1}\mathbf{HB}\\ &\geq \mathbf{I}-\mathbf{B}^H\mathbf{H}^H\left(\mathbf{HBB}^H\mathbf{H}^H+\mathbf{R}_n\right)^{-1}\mathbf{HB}\end{aligned} \tag{2.50}$$

where we have used the fact that $\mathbf{X}+\mathbf{Y}\geq\mathbf{X}$ when $\mathbf{Y}$ is positive semidefinite. The lower bound is clearly achieved by (2.48). The concentrated MSE matrix is obtained by plugging (2.48) into (2.46) (or directly from (2.50)) as

$$\begin{aligned}\mathbf{E}(\mathbf{B}) &\triangleq \mathbf{E}(\mathbf{B},\mathbf{A}^\star)\\ &= \mathbf{I}-\mathbf{B}^H\mathbf{H}^H\left(\mathbf{HBB}^H\mathbf{H}^H+\mathbf{R}_n\right)^{-1}\mathbf{HB}\\ &= \mathbf{I}-\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{HB}\left(\mathbf{I}+\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{HB}\right)^{-1}\\ &= \left(\mathbf{I}+\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{HB}\right)^{-1}\end{aligned} \tag{2.51}$$

where we have used the matrix inversion lemma (see §3.3). From (2.51), the MSE at the ith substream is

$$\begin{aligned}\mathrm{MSE}_i(\mathbf{B}) &\triangleq \mathrm{MSE}_i(\mathbf{B},\mathbf{a}_i^\star)\\ &= 1-\mathbf{b}_i^H\mathbf{H}^H\left(\mathbf{Hb}_i\mathbf{b}_i^H\mathbf{H}^H+\mathbf{R}_{n_i}\right)^{-1}\mathbf{Hb}_i\\ &= \frac{1}{1+\mathbf{b}_i^H\mathbf{H}^H\mathbf{R}_{n_i}^{-1}\mathbf{Hb}_i}\end{aligned} \tag{2.52}$$

where the second expression is obtained after applying the matrix inversion lemma (see §3.3). Note that (2.52) is identical in form to the equivalent expression (2.45) in the single beamforming case. It is interesting to note from the fact that $\mathrm{MSE}_i(\mathbf{B})=[\mathbf{E}(\mathbf{B})]_{ii}$ (combining (2.51) and (2.52)) the following relation:

$$\left[\left(\mathbf{I}+\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{HB}\right)^{-1}\right]_{ii}=\frac{1}{1+\mathbf{b}_i^H\mathbf{H}^H\mathbf{R}_{n_i}^{-1}\mathbf{Hb}_i}.$$

As can be seen from the optimum receive matrix in (2.49) $\mathbf{A}=\mathbf{R}_n^{-1}\mathbf{HB}\left(\mathbf{I}+\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{HB}\right)^{-1}$ (and also from (2.44) in the single beamforming case), the receive processing $\mathbf{A}^H$ can be decomposed as the concatenation of a noise-whitening stage $\mathbf{R}_n^{-1/2}$, a matched filter stage $\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1/2}$, and an MSE stage $\mathbf{E}=\left(\mathbf{I}+\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{HB}\right)^{-1}$. The first two stages simplify the channel model significantly without any loss of performance into the canonical channel model[18]. Therefore, the global communication process (including the

[18]Strictly speaking this is only true if $\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{HB}$ is nonsingular (in particular, this implies $L\leq\mathrm{rank}(\mathbf{H})$).

pre- and post-processing) can be nicely related to the canonical channel combining (2.49) and (2.51) as

$$\hat{\mathbf{x}} = \mathbf{A}^H (\mathbf{HBx} + \mathbf{n}) \tag{2.53}$$

$$= \mathbf{E} (\mathbf{H}_{\text{can}}\mathbf{x} + \mathbf{n}_{\text{can}}) \tag{2.54}$$

where $\mathbf{H}_{\text{can}}$ and $\mathbf{n}_{\text{can}}$ are defined as in §2.5.3. Note that both $\mathbf{E}$ and $\mathbf{H}_{\text{can}}$ depend on the transmit matrix $\mathbf{B}$ which have to be properly designed (see Chapters 5 and 6).

2.5.5.2 Optimum Linear Receiver in Terms of SINR

Consider first the single beamforming signal model given by (2.1), (2.25), and (2.26), *i.e.*, $\hat{x} = \mathbf{a}^H (\mathbf{Hb}x + \mathbf{n})$. The SINR is

$$\text{SINR}(\mathbf{b}, \mathbf{a}) \triangleq \frac{\left|\mathbf{a}^H\mathbf{Hb}\right|^2}{\mathbf{a}^H\mathbf{R}_n\mathbf{a}} = \frac{\mathbf{a}^H \left(\mathbf{Hbb}^H\mathbf{H}^H\right)\mathbf{a}}{\mathbf{a}^H\mathbf{R}_n\mathbf{a}}. \tag{2.55}$$

We can now obtain the optimum receive beamvector that maximizes the SINR. Expression (2.55) is a generalized Rayleigh quotient that is maximized by the generalized eigenvector of the matrix pencil $\left(\mathbf{Hbb}^H\mathbf{H}^H, \mathbf{R}_n\right)$ corresponding to the maximum generalized eigenvalue of $\left(\mathbf{Hbb}^H\mathbf{H}^H\right)\mathbf{a} = \lambda\,\mathbf{R}_n\mathbf{a}$ [Gol96], *i.e.*,

$$\mathbf{a}^\star = \alpha\,\mathbf{R}_n^{-1}\mathbf{Hb} \tag{2.56}$$

where α is an arbitrary scaling factor that does not affect the SINR. Alternatively, the SINR in (2.55) can be upper-bounded as

$$\frac{\left|\mathbf{a}^H\mathbf{Hb}\right|^2}{\mathbf{a}^H\mathbf{R}_n\mathbf{a}} \leq \mathbf{b}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{Hb} \tag{2.57}$$

simply by using Cauchy-Schwarz's inequality (see §3.3) with vectors $(\mathbf{R}_n^{-1/2}\mathbf{Hb})$ and $(\mathbf{R}_n^{1/2}\mathbf{a})$:

$$|(\mathbf{a}^H\mathbf{R}_n^{1/2})\,(\mathbf{R}_n^{-1/2}\mathbf{Hb})|^2 \leq (\mathbf{a}^H\mathbf{R}_n\mathbf{a}^H)\,(\mathbf{b}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{Hb}).$$

The upper bound in (2.57) is achieved when $\mathbf{R}_n^{1/2}\mathbf{a} \propto \mathbf{R}_n^{-1/2}\mathbf{Hb}$ or, equivalently, by (2.56). It is important to remark that (2.56) is identical to (2.44) up to a scaling factor, *i.e.*, the Wiener filter is also optimal in terms of maximizing the SINR. Using the optimal receive beamvector (2.56), the SINR in (2.55) reduces to

$$\text{SINR}(\mathbf{b}) \triangleq \text{SINR}(\mathbf{b}, \mathbf{a}^\star) = \mathbf{b}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{Hb}. \tag{2.58}$$

For the multiple beamforming case, the SINR is similarly defined as

$$\text{SINR}_i(\mathbf{B}, \mathbf{a}_i) \triangleq \frac{\left|\mathbf{a}_i^H\mathbf{Hb}_i\right|^2}{\mathbf{a}_i^H\mathbf{R}_{n_i}\mathbf{a}_i} \tag{2.59}$$

where $\mathbf{R}_{n_i}$ is the noise covariance matrix seen by the ith substream. The derivation of the receive beamvectors $\mathbf{a}_i$'s can be independently done for each of the substreams since the choice of the receive beamvector for one substream does not affect the others. The optimal solution is again given by the Wiener filter (up to a scaling factor)

$$\mathbf{a}_i^\star = \alpha_i \, \mathbf{R}_{n_i}^{-1} \mathbf{H} \mathbf{b}_i \tag{2.60}$$

and the resulting SINR by

$$\mathrm{SINR}_i \left(\mathbf{B} \right) \triangleq \mathrm{SINR}_i \left(\mathbf{B}, \mathbf{a}_i^\star \right) = \mathbf{b}_i^H \mathbf{H}^H \mathbf{R}_{n_i}^{-1} \mathbf{H} \mathbf{b}_i. \tag{2.61}$$

Comparing now (2.61) with (2.52) (also (2.58) with (2.45)), it is clear that the SINR and the MSE are related by

$$\mathrm{SINR}_i = \frac{1}{\mathrm{MSE}_i} - 1. \tag{2.62}$$

Note that the SINR in (2.62) is a convex decreasing function of the MSE (clearly, minimizing the MSE is tantamount to maximizing the SINR).

2.5.5.3 Optimum Linear Receiver in Terms of BER

To design a system directly in terms of BER, it is convenient to relate the BER with the SINR and with the MSE.

BER vs. SINR

Both the exact BER function and the Chernoff upper bound are convex decreasing functions of the SINR as proved in Appendix 2.A. Since the BER is decreasing in the SINR, it follows that the Wiener filter minimizes the BER (in the multiple beamforming case, the BER of each substream depends only on its corresponding SINR).

BER vs. MSE

Using (2.62) in (2.38), the error probability can be alternatively expressed as a function of the MSE

$$P_e = \alpha \, \mathcal{Q}\left(\sqrt{\beta \left(\mathrm{MSE}^{-1} - 1 \right)} \right). \tag{2.63}$$

Both the exact BER function and the Chernoff upper bound happen to be convex increasing functions of the MSE for sufficiently small values of the argument (for BPSK and QPSK constellations, this is true for any value of the argument) as can be observed from Figure 2.17 (see Appendix 2.A for a formal proof). As a rule-of-thumb, the exact BER function and the Chernoff upper bound are indeed convex in the MSE for a BER less than 2×10^{-2} (recall that for BPSK and QPSK constellations, this is true for any value of the BER). Note that this is a mild assumption since any realistic system must have an uncoded BER[19] of less than 2×10^{-2}. Therefore,

[19] Given an uncoded bit error probability of at most 10^{-2} and using a proper coding scheme, coded bit error probabilities with acceptable low values such as 10^{-6} can be obtained.

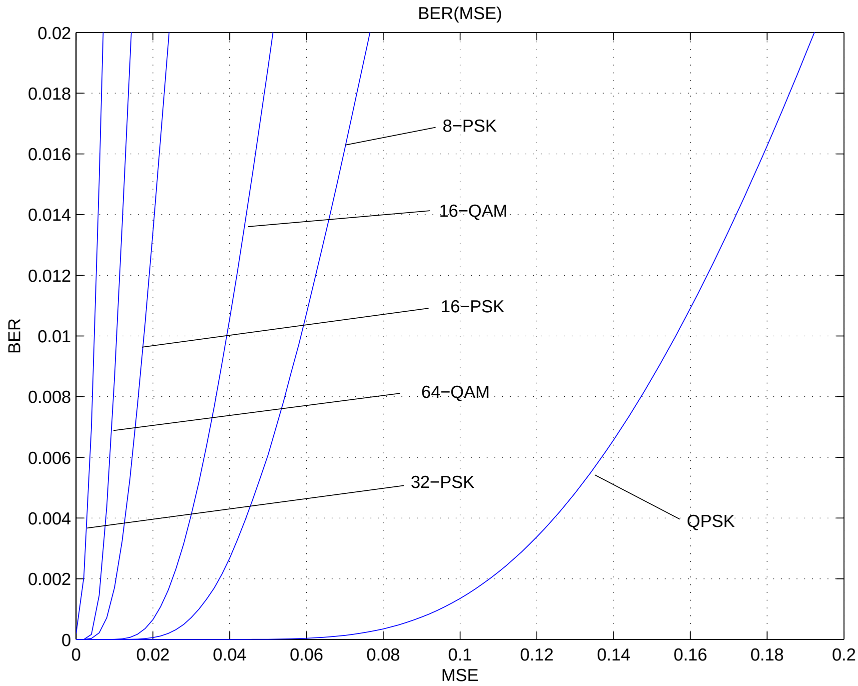


Figure 2.17: Convexity of the BER as a function of the MSE for the range of BER $\leq 2 \times 10^{-2}$.

for practical purposes, we can assume the exact BER and the Chernoff upper bound as convex functions of the MSE.

Since the BER is increasing in the MSE, minimizing the BER is tantamount to minimizing the MSE and it then follows that the Wiener filter minimizes the BER (in the multiple beamforming case, the BER of each substream depends only on its corresponding MSE).

2.5.5.4 Optimum Linear Receiver in Terms of Capacity

Recall that the optimum receiver in terms of capacity is the ML receiver. The Wiener filter, however, is an invertible transformation of the received signal and is, therefore, capacity-lossless [Cio97]. As a consequence, we can consider the Wiener filter as the front-end stage of the receiver without loss of generality.

2.5.5.5 A Summary of the Wiener filter

As has been shown in the previous subsections, the Wiener filter (equivalently, the LMMSE receiver) is the optimal linear receiver in the sense that each of the MSE's is minimized, each of the SINR's is maximized, and each of the BER's is minimized. In terms of capacity, the

Wiener filter is capacity-lossless and simplifies the signal model by possibly eliminating useless dimensions.

In addition, the Wiener filter is in fact the MMSE estimator (without imposing a linear structure) if the symbols to be estimated are Gaussian distributed (note that this is necessary to achieve capacity). In such a case, the linear structure of the receiver is without loss of optimality [Kay93].

2.5.5.6 Optimum Linear Receiver under a ZF Constraint

It is very common in the literature of equalization to include a zero-forcing (ZF) constraint in the design [Qur85, GDF91]. According to the channel model in (2.1), the ZF constraint (perfect equalization) is

$$\mathbf{A}^H\mathbf{H}\mathbf{B} = \mathbf{I}. \tag{2.64}$$

Note that for a solution to exist it must be that $L \leq \text{rank}(\mathbf{H})$.

The design of the transmit and receive matrices can now be based (as before) on the optimization of the MSE's, SINR's, or BER's but this time subject to the ZF constraint of (2.64). In that case, the MSE matrix in (2.46) reduces to $\mathbf{E}^{\text{ZF}} = \mathbf{A}^H\mathbf{R}_n\mathbf{A}$ (due to the perfect equalization of the channel) or, equivalently, $\text{MSE}_i^{\text{ZF}} = \mathbf{a}_i^H\mathbf{R}_n\mathbf{a}_i$. The SINR simplifies to $\text{SINR}_i^{\text{ZF}} = \frac{1}{\mathbf{a}_i^H\mathbf{R}_n\mathbf{a}_i}$ which implies the following relation

$$\text{SINR}_i^{\text{ZF}} = \frac{1}{\text{MSE}_i^{\text{ZF}}}. \tag{2.65}$$

As a consequence, the optimization of the MSE's, SINR's, or BER's subject to the ZF constraint is uniquely expressed as

$$\begin{array}{ll} \min\limits_{\mathbf{a}_i} & \mathbf{a}_i^H\mathbf{R}_n\mathbf{a}_i \\ \text{s.t.} & \mathbf{a}_i^H\mathbf{H}\mathbf{B} = \mathbf{e}_i^H. \end{array} \tag{2.66}$$

The associated Lagrangian is

$$L = \mathbf{a}_i^H\mathbf{R}_n\mathbf{a}_i + \text{Re}\left(\mathbf{a}_i^H\mathbf{H}\mathbf{B} - \mathbf{e}_i^H\right)\boldsymbol{\lambda}_R + \text{Im}\left(\mathbf{a}_i^H\mathbf{H}\mathbf{B} - \mathbf{e}_i^H\right)\boldsymbol{\lambda}_I. \tag{2.67}$$

Setting the gradient of the Lagrangian to zero and solving for the Lagrange multipliers, the solution $\mathbf{a}_i^{\text{ZF}} = \mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\left(\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}\mathbf{e}_i$ is obtained. More compactly, the optimal receiver matrix is

$$\mathbf{A}^{\text{ZF}} = \mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\left(\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}. \tag{2.68}$$

Using (2.68), the resulting MSE matrix, MSE's, and SINR's are given by

$$\mathbf{E}^{\mathrm{ZF}}(\mathbf{B}) = \left(\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}, \tag{2.69}$$

$$\mathrm{MSE}_i^{\mathrm{ZF}}(\mathbf{B}) = \left[\left(\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}\right]_{ii}, \text{ and} \tag{2.70}$$

$$\mathrm{SINR}_i^{\mathrm{ZF}}(\mathbf{B}) = \frac{1}{\left[\left(\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}\right]_{ii}}. \tag{2.71}$$

It may be useful for Chapters 5 and 6 to consider the following heuristic and intuitively appealing asymptotic derivation of the ZF solution. Define the following extended MSE matrix:

$$\mathbf{E}^{(\gamma)}(\mathbf{B},\mathbf{A}) \triangleq \gamma\left(\mathbf{A}^H\mathbf{H}\mathbf{B}-\mathbf{I}\right)\left(\mathbf{B}^H\mathbf{H}^H\mathbf{A}-\mathbf{I}\right) + \mathbf{A}^H\mathbf{R}_n\mathbf{A}. \tag{2.72}$$

Note that the only difference with respect to the regular MSE matrix (2.46) is the scaling factor γ. It is important to realize that, for $\gamma \to \infty$, the importance of the term $\left(\mathbf{A}^H\mathbf{H}\mathbf{B}-\mathbf{I}\right)\left(\mathbf{B}^H\mathbf{H}^H\mathbf{A}-\mathbf{I}\right)$ increases with respect to the term $\mathbf{A}^H\mathbf{R}_n\mathbf{A}$. As a consequence, a design based on the minimization of the diagonal elements of (2.72) for $\gamma \to \infty$ will produce $\left(\mathbf{A}^H\mathbf{H}\mathbf{B}-\mathbf{I}\right)\left(\mathbf{B}^H\mathbf{H}^H\mathbf{A}-\mathbf{I}\right)$ with zero diagonal elements or, equivalently, $\left(\mathbf{A}^H\mathbf{H}\mathbf{B}-\mathbf{I}\right) = \mathbf{0}$. Therefore, a design including the ZF constraint is easily obtained by using the extended MSE matrix of (2.72) and then letting $\gamma \to \infty$ in the solution.

Solving as for the unconstrained case but including the scaling factor γ we obtain (similarly to (2.48)-(2.49))

$$\mathbf{A}^{(\gamma)} = \left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H + \frac{1}{\gamma}\mathbf{R}_n\right)^{-1}\mathbf{H}\mathbf{B} \tag{2.73}$$

$$= \mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\left(\frac{1}{\gamma}\mathbf{I} + \mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}. \tag{2.74}$$

Note that expression (2.73) may not exist in the limit of $\gamma \to \infty$ since $\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H$ may be singular. Expression (2.74) always exists provided that $L \leq \mathrm{rank}(\mathbf{H})$, which is a necessary condition to guarantee the existence of the ZF constraint as previously discussed. The resulting MSE matrix is (similarly to (2.51))

$$\mathbf{E}^{(\gamma)}(\mathbf{B}) = \gamma\left(\mathbf{I} - \mathbf{B}^H\mathbf{H}^H\left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H + \frac{1}{\gamma}\mathbf{R}_n\right)^{-1}\mathbf{H}\mathbf{B}\right)$$

$$= \left(\frac{1}{\gamma}\mathbf{I} + \mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1} \tag{2.75}$$

and the MSE's (similarly to (2.52))

$$\mathrm{MSE}_i^{(\gamma)}(\mathbf{B}) = \left[\left(\frac{1}{\gamma}\mathbf{I} + \mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}\right]_{ii}$$

$$= \frac{1}{\frac{1}{\gamma} + \mathbf{b}_i^H\mathbf{H}^H\mathbf{R}_{n_i}^{(\gamma)\,-1}\mathbf{H}\mathbf{b}_i} \tag{2.76}$$

where $\mathbf{R}_{n_i}^{(\gamma)} \triangleq \mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H - \mathbf{H}\mathbf{b}_i\mathbf{b}_i^H\mathbf{H}^H + \frac{1}{\gamma}\mathbf{R}_n$ is an extended definition of the noise covariance matrix.

Similarly, an extended SINR can be defined as

$$\mathrm{SINR}_i^{(\gamma)} \triangleq \frac{\left|\mathbf{a}_i^H\mathbf{H}\mathbf{b}_i\right|^2}{\mathbf{a}_i^H\mathbf{R}_{n_i}^{(\gamma)}\mathbf{a}_i}.$$

Using the optimal receive matrix given by (2.68), the resulting SINR expression is

$$\mathrm{SINR}_i^{(\gamma)} = \mathbf{b}_i^H\mathbf{H}^H\mathbf{R}_{n_i}^{(\gamma)\,-1}\mathbf{H}\mathbf{b}_i. \tag{2.77}$$

The extended SINR can be related to the extended MSE as

$$\mathrm{SINR}_i^{(\gamma)} = \frac{1}{\mathrm{MSE}_i^{(\gamma)}} - \frac{1}{\gamma} \tag{2.78}$$

which includes as special cases the ZF-constrained case of (2.65) for $\gamma \to \infty$ and the unconstrained case of (2.62) for $\gamma = 1$.

The interest of the heuristic derivation based on the definition of the extended MSE matrix and SINR is that in the limit of $\gamma \to \infty$, the extended quantities tend to the ones obtained with the ZF constraint:

$$\begin{aligned} \mathbf{A}^{(\gamma)} &\longrightarrow \mathbf{A}^{\mathrm{ZF}}, \\ \mathbf{E}^{(\gamma)} &\longrightarrow \mathbf{E}^{\mathrm{ZF}}, \\ \mathrm{MSE}_i^{(\gamma)} &\longrightarrow \mathrm{MSE}_i^{\mathrm{ZF}}, \text{ and} \\ \mathrm{SINR}_i^{(\gamma)} &\longrightarrow \mathrm{SINR}_i^{\mathrm{ZF}}. \end{aligned}$$

Thus, in order to obtain the transmit matrix $\mathbf{B}$ under the ZF constraint, one can either use directly the ZF expressions (2.69)-(2.71) or use the extended expressions (2.75)-(2.77) and then let $\gamma \to \infty$.

Alternatively, the ZF constraint can be similarly imposed (with the same results) on different versions of the channel:

- the whitened channel $\tilde{\mathbf{H}} = \mathbf{R}_n^{-1/2}\mathbf{H}$: $\tilde{\mathbf{A}}^H\tilde{\mathbf{H}}\mathbf{B} = \mathbf{I}$ where $\mathbf{A} = \mathbf{R}_n^{-1/2}\tilde{\mathbf{A}}$
- the whitened and matched-filtered channel $\mathbf{H}_{\mathrm{wmf}} = \mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}$: $\tilde{\mathbf{A}}^H\mathbf{H}_{\mathrm{wmf}}\mathbf{B} = \mathbf{I}$ where $\mathbf{A} = \mathbf{R}_n^{-1}\mathbf{H}\tilde{\mathbf{A}}$
- the canonical channel $\mathbf{H}_{\mathrm{can}} = \mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}$: $\tilde{\mathbf{A}}^H\mathbf{H}_{\mathrm{can}}\mathbf{B} = \mathbf{I}$ where $\mathbf{A} = \mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\tilde{\mathbf{A}}$

2.6 Chapter Summary and Conclusions

In this chapter, we have given an overview of MIMO channels. After introducing the basic MIMO channel model, we have shown how many different real communication systems can be conveniently modeled as such. The basic gains of MIMO channels—beamforming, diversity, and multiplexing gains—have been described, followed by an overview of the existing communication methods that try to achieve such gains. In particular, linear processing methods have been considered in detail since they constitute the basis of Chapters 5-7.

No new material has been presented in this chapter, although the exposition of the subject has followed a personal perspective.

Appendix 2.A Analysis of the Error Probability Function

In this appendix, we prove that the BER function and also the corresponding Chernoff upper bound are convex decreasing functions of the SINR and convex increasing functions of the MSE (for sufficiently small values of the MSE).

Before proceeding, recall that the symbol error probability can be analytically expressed as a function of the SINR as $P_e = \alpha\, \mathcal{Q}\left(\sqrt{\beta\, \text{SINR}}\right)$. As an example, for M-ary PAM, M-ary QAM, and M-ary PSK constellations this relation is specifically given by

$$
\begin{aligned}
P_e^{\text{PAM}} &\cong 2\left(1-\frac{1}{M}\right)\mathcal{Q}\left(\sqrt{\frac{3}{M^2-1}\,\text{SINR}}\right),\\
P_e^{\text{QAM}} &\cong 4\left(1-\frac{1}{\sqrt{M}}\right)\mathcal{Q}\left(\sqrt{\frac{3}{M-1}\,\text{SINR}}\right), \text{ and}\\
P_e^{\text{PSK}} &\cong 2\,\mathcal{Q}\left(\sqrt{2\sin^2\left(\frac{\pi}{M}\right)\text{SINR}}\right) \quad \text{for } M \geq 4.
\end{aligned}
$$

Note that $\alpha \geq 1$ and $\beta \leq 1$. See Table 2.1 for specific values of the parameters (recall that M is the constellation size and $k = \log_2 M$ the number of bits per symbol).

BER as a function of the SINR

To prove that the BER function is convex decreasing in the SINR, it suffices to show that the first and second derivatives of $\mathcal{Q}\left(\sqrt{\beta\, x}\right)$ are negative and positive, respectively (note that a positive scaling factor preserves monotonicity and convexity):

$$
\begin{aligned}
\frac{\partial \mathcal{Q}\left(\sqrt{\beta\, x}\right)}{\partial x} &= -\sqrt{\frac{\beta}{8\pi}} e^{-\beta x/2} x^{-1/2} < 0 & 0 < x < \infty\\
\frac{\partial^2 \mathcal{Q}\left(\sqrt{\beta\, x}\right)}{\partial x^2} &= \frac{1}{2}\sqrt{\frac{\beta}{8\pi}} e^{-\beta x/2} x^{-1/2}\left(\frac{1}{x}+\beta\right) > 0 & 0 < x < \infty.
\end{aligned}
$$

Constellation	M	k	α	β ($\simeq$)	x_{z_1} ($\simeq$)	Convexity region of BER(MSE)
BPSK	2	1	1	1	0.5	$\text{BER} \leq 1.587 \times 10^{-1}$
4-PAM	4	2	1.5	0.2	6.834×10^{-2}	$\text{BER} \leq 3.701 \times 10^{-2}$
16-PAM	16	4	1.875	0.0118	3.927×10^{-3}	$\text{BER} \leq 1.971 \times 10^{-2}$
QPSK	4	2	2	1	0.5	$\text{BER} \leq 1.587 \times 10^{-1}$
16-QAM	16	4	3	0.2	6.834×10^{-2}	$\text{BER} \leq 3.701 \times 10^{-2}$
64-QAM	64	6	3.5	0.0476	1.596×10^{-2}	$\text{BER} \leq 2.526 \times 10^{-2}$
8-PSK	8	3	2	0.2929	10.15×10^{-2}	$\text{BER} \leq 3.576 \times 10^{-2}$
16-PSK	16	4	2	0.0761	2.559×10^{-2}	$\text{BER} \leq 2.218 \times 10^{-2}$
32-PSK	32	5	2	0.0192	6.414×10^{-3}	$\text{BER} \leq 1.692 \times 10^{-2}$

Table 2.1: Examples of parameters and convexity region of the BER for well-known constellations.

The same can be done for the Chernoff upper bound $e^{-\beta x/2}$:

$$\begin{aligned} \frac{\partial e^{-\beta x/2}}{\partial x} &= -\frac{\beta}{2} e^{-\beta x/2} < 0 && 0 < x < \infty \\ \frac{\partial^2 e^{-\beta x/2}}{\partial x^2} &= \left(\frac{\beta}{2}\right)^2 e^{-\beta x/2} > 0 && 0 < x < \infty. \end{aligned}$$

BER as a function of the MSE

To prove that the BER function is convex increasing in the MSE (assuming a MMSE receiver), it suffices to show that the first and second derivatives of $\mathcal{Q}\left(\sqrt{\beta\ (x^{-1}-1)}\right)$ are both positive (note that a positive scaling factor preserves monotonicity and convexity):

$$\begin{aligned} \frac{\partial \mathcal{Q}\left(\sqrt{\beta\ (x^{-1}-1)}\right)}{\partial x} &= \sqrt{\frac{\beta}{8\pi}} e^{-\beta\left(x^{-1}-1\right)/2} \left(x^3 - x^4\right)^{-1/2} \geq 0 && 0 < x \leq 1 \\ \frac{\partial^2 \mathcal{Q}\left(\sqrt{\beta\ (x^{-1}-1)}\right)}{\partial x^2} &= \frac{1}{2}\sqrt{\frac{\beta}{8\pi}} e^{-\beta\left(x^{-1}-1\right)/2} \left(x^3 - x^4\right)^{-1/2} \left(\frac{\beta}{x^2} - \frac{3-4x}{x-x^2}\right) \geq 0 && 0 < x \leq x_{z_1}, \\ & && x_{z_2} \leq x \leq 1. \end{aligned}$$

where the zeros are $x_{z_1} = \frac{(\beta+3)-\sqrt{\beta^2-10\beta+9}}{8}$ and $x_{z_2} = \frac{(\beta+3)+\sqrt{\beta^2-10\beta+9}}{8}$ (it has been tacitly assumed that $\beta \leq 1$). It is remarkable that for $\beta = 1$ both zeros coincide, which means that the BER function is convex for the whole range of MSE values. To be more specific, BPSK and QPSK constellations satisfy this condition and, consequently, their corresponding BER function is always convex in the MSE.

Consider now the Chernoff upper bound $e^{-\beta\left(x^{-1}-1\right)/2}$:

$$\begin{aligned} \frac{\partial e^{-\beta\left(x^{-1}-1\right)/2}}{\partial x} &= \frac{\beta}{2} e^{-\beta\left(x^{-1}-1\right)/2} x^{-2} > 0 && 0 < x < \infty \\ \frac{\partial^2 e^{-\beta\left(x^{-1}-1\right)/2}}{\partial x^2} &= \frac{\beta}{2} e^{-\beta\left(x^{-1}-1\right)/2} x^{-4} \left(\frac{\beta}{2} - 2x\right) \geq 0 && 0 < x \leq \frac{\beta}{4}. \end{aligned}$$

The Chernoff upper bound is then convex increasing in the MSE for MSE $\leq \beta/4$ (see Table 2.1 for specific values).

The same analysis can be performed for a ZF receiver, obtaining that the BER function $\mathcal{Q}\left(\sqrt{\beta \text{ MSE}^{-1}}\right)$ is convex increasing in the MSE for MSE $\leq \beta/3$ and that the Chernoff upper bound $e^{-\beta \text{ MSE}^{-1}/2}$ is convex increasing in the MSE for MSE $\leq \beta/4$ (exactly the same result as with the MMSE receiver).

Concluding, as a rule-of-thumb, both the exact BER function and the Chernoff upper bound are convex increasing functions of the MSE for a BER $\leq 2 \times 10^{-2}$ (see Table 2.1).

Chapter 3

Mathematical Preliminaries

AN OVERVIEW OF TWO IMPORTANT THEORIES —convex optimization theory and majorization theory—on which many results of this dissertation are based is given in this chapter.

3.1 Convex Optimization Theory

Most problems of practical interest can be appropriately formulated as constrained optimization problems. In some cases, possibly after some mathematical manipulations, the problems can be expressed in convex form. For these type of problems, there is a well developed body of theory and practice. In a nutshell, convex problems can be optimally solved very efficiently in practice. As a consequence, roughly speaking, one can say that once a problem has been expressed in convex form, it has been solved.

The two main mathematical references on the subject are [Lue69] and [Roc70]. An excellent reference from a practical implementation perspective with engineering applications is [Boy00].

3.1.1 Convex Problems

A *convex optimization problem* (*convex program*) is of the form [Lue69, Roc70, Boy00]:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0 \qquad 1 \leq i \leq m, \\ & h_i(\mathbf{x}) = 0 \qquad 1 \leq i \leq p, \end{aligned} \tag{3.1}$$

where $\mathbf{x} \in I\!R^n$ is the *optimization variable*, $f_0, \cdots, f_m$ are convex functions,[1] and $h_1, \cdots, h_p$ are linear functions (more exactly affine functions). The function f_0 is the *objective function* or *cost function*. The inequalities $f_i(\mathbf{x}) \leq 0$ are called *inequality constraints* and the equations $h_i(\mathbf{x}) = 0$ are called *equality constraints*. If there are no constraints, we say that the problem is *unconstrained*.

The set of points for which the objective and all constraint functions are defined, *i.e.*,

$$D = \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{i=1}^{p} \operatorname{dom} h_i$$

is called the *domain* of the optimization problem (3.1). A point $\mathbf{x} \in D$ is *feasible* if it satisfies all the constraints $f_i(\mathbf{x}) \leq 0$ and $h_i(\mathbf{x}) = 0$. Problem (3.1) is said to be feasible if there exists at least one feasible point and *infeasible* otherwise. The *optimal value* (minimal value) is denoted by $f^\star$ (if the problem is infeasible, it is commonly denoted by $f^\star = +\infty$) and is achieved at an optimal solution $\mathbf{x}^\star$, *i.e.*, $f^\star = f_0(\mathbf{x}^\star)$.

When the functions f_i and h_i in (3.1) are linear (affine), the problem is called a *linear program* (LP) and is much simpler to solve. If the objective function is quadratic and the constraint functions are linear (affine), the convex optimization problem (3.1) is called a *quadratic program* (QP).

Many analysis and design problems arising in engineering can be cast (or recast) in the form of a convex optimization problem. In general, some manipulations are required to convert the problem into a convex one (unfortunately, this is not always possible). The interest of expressing a problem in convex form is that, although an analytical solution may not exist and the problem may seem difficult to solve (it may have hundreds of variables and a nonlinear, nondifferentiable objective function), it can still be solved (numerically) very efficiently both in theory and practice [Boy00]. Another interesting feature of expressing the problem in convex form is that additional constraints can be straightforwardly added as long as they are convex.

Convex programming has been used in related areas such as FIR filter design [Wu96, Dav02], antenna array pattern synthesis [Leb97], power control for interference-limited wireless networks [Kan02], and beamforming design in a multiuser scenario with a multi-antenna base station [Ben99, Ben01].

3.1.2 Solving Convex Problems

In some cases, convex optimization problems can be analytically solved using duality theory and closed-form expressions can be obtained. In general, however, one must resort to iterative methods [Lue69, Boy00].

[1] A function $f : I\!R^n \longrightarrow I\!R$ is convex if, for all $x, y \in \operatorname{dom} f$ and $\theta \in [0,1]$, $\theta x + (1-\theta) y \in \operatorname{dom} f$ (*i.e.*, the domain is a convex set) and $f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$.

In the last ten years, there has been considerable progress and development of efficient algorithms for solving wide classes of convex optimization problems. Recently developed *interior-point methods* can be used to iteratively solve convex problems efficiently in practice. This was an important breakthrough achieved by Nesterov and Nemirovsky in 1988. They showed that interior-point methods (initially proposed only for linear programming by Karmarkar in 1984) can, in principle, be generalized to all convex optimization problems. In [Nes94], a very general framework was developed for solving convex optimization problems using interior-point methods. In addition, the difference between the objective value at each iteration and the optimum value can be upper-bounded using duality theory [Lue69, Boy00]. This allows the utilization of nonheuristic stopping criteria based on checking whether some prespecified resolution has been reached.

Interior-point methods solve constrained problems by solving a sequence of smooth (continuous second derivatives are assumed) unconstrained problems, usually using Newton's method [Lue69, Boy00]. The solutions at each iteration are all strictly feasible (they are in the interior of the domain), hence the name interior-point methods. They are also called *barrier methods* since at each iteration a barrier function is used to guarantee that the solution obtained is strictly feasible.

Alternatively, *cutting-plane methods* can be used [Boy00]. They use a completely different philosophy and do not require differentiability of the objective and constraint functions. They start with the feasible space and iteratively divide it into two halfspaces to reject the one that is known not to contain any optimal solution. *Ellipsoid methods* are related to cutting-plane methods in that they sequentially reduce an ellipsoid known to contain an optimal solution. In general, cutting-plane methods are less efficient for problems to which interior-point methods apply. See [Boy00] for details of implementation.

3.1.3 Duality Theory and KKT Conditions

The basic idea in Lagrangian duality is to take the constraints in (3.1) into account by augmenting the objective function with a weighted sum of the constraint functions. We define the *Lagrangian* associated with the problem (3.1) (not necessarily convex) as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \tag{3.2}$$

where λ_i is the *Lagrange multiplier* associated with the ith inequality constraint $f_i(\mathbf{x}) \leq 0$ and ν_i the Lagrange multiplier associated with the ith equality constraint $h_i(\mathbf{x}) = 0$.

The optimization variable $\mathbf{x}$ is called *primal variable* and the vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are called the *dual variables* or *Lagrange multiplier vectors* associated with the problem (3.1). The original objective function $f_0(\mathbf{x})$ is termed the *primal objective* or *primal function.* The *dual objective* or *dual function* $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is defined as the minimum value of the Lagrangian over $\mathbf{x}$, *i.e.*, for $\boldsymbol{\lambda} \in I\!R^m$ and $\boldsymbol{\nu} \in I\!R^p$,

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in D} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}), \tag{3.3}$$

which is concave even if the original problem is not convex because it is the pointwise infimum of a family of affine functions of $(\boldsymbol{\lambda}, \boldsymbol{\nu})$. We say that $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ are *dual feasible* if $\boldsymbol{\lambda} \geq \mathbf{0}$ and $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$.

The dual function $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is a lower bound on the optimal value $f^\star$ of the problem (3.1). For any feasible $(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$:

$$\begin{aligned} f_0(\mathbf{x}) &\geq f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) && (3.4) \\ &\geq \inf_{\mathbf{z} \in D} \left(f_0(\mathbf{z}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{z}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{z}) \right) && (3.5) \\ &= g(\boldsymbol{\lambda}, \boldsymbol{\nu}) && (3.6) \end{aligned}$$

where we have used the fact that $f_i(\mathbf{x}) \leq 0$ and $h_i(\mathbf{x}) = 0$ for any feasible $\mathbf{x}$ and $\lambda_i \geq 0$ for any feasible λ_i in the first inequality. Thus, for the set of feasible $(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$, it follows that

$$\min_{\mathbf{x}} f_0(\mathbf{x}) \geq \max_{\boldsymbol{\lambda}, \boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \tag{3.7}$$

which holds even if the original problem is not convex. The difference between the primal objective $f_0(\mathbf{x})$ and the dual objective $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is called the *duality gap*. If (3.7) is satisfied with strict inequality we say that weak duality holds; if (3.7) is achieved with equality we say that strong duality holds.

A central result in convex analysis [Lue69, Roc70, Boy00] is that when the problem is convex, under some technical conditions (called constraint qualifications), the duality gap reduces to zero at the optimal (strong duality holds), *i.e.*, (3.7) is achieved with equality for some $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$. Therefore, one way to solve the original problem is to solve its associated dual problem:

$$\begin{array}{ll} \max\limits_{\boldsymbol{\lambda}, \boldsymbol{\nu}} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{s.t.} & \boldsymbol{\lambda} \geq \mathbf{0}, \end{array} \tag{3.8}$$

which is always a convex optimization problem even if the original problem is not convex (the objective to be maximized $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is concave and the constraint is convex). It is interesting to note that a primal-dual feasible pair $(\mathbf{x}, (\boldsymbol{\lambda}, \boldsymbol{\nu}))$ localizes the optimal value of the primal (and dual) problem in an interval:

$$f^\star \in [g(\boldsymbol{\lambda}, \boldsymbol{\nu}), f_0(\mathbf{x})]$$

the width of which is the duality gap. If the duality gap is zero, *i.e.*, if $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x})$, then $\mathbf{x}$ is primal optimal and $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is dual optimal. These observations can be used in optimization algorithms to provide nonheuristic stopping criteria.

Let $\mathbf{x}^\star$ and $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ be the primal and dual variables at the optimum. If we substitute them in the chain of inequalities (3.4)-(3.6), we see that each of the inequalities must be satisfied with equality (assuming that strong duality holds). To have equality in (3.4), it must be that $\lambda_i f_i(\mathbf{x}) = 0$ (this is the so-called complementary slackness condition). Moreover, since the inequality in (3.5) must also be satisfied with equality, the infimum is achieved at $\mathbf{x}^\star$; in other words, the gradient of the Lagrangian with respect to $\mathbf{x}$ must be zero at $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$, *i.e.*, $\nabla_{\mathbf{x}} L(\mathbf{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = \mathbf{0}$. These two results, together with the constraints on the primal and dual variables, form the Karush-Kuhn-Tucker (KKT) conditions:

$$h_i(\mathbf{x}^\star) = 0, \qquad f_i(\mathbf{x}^\star) \le 0, \tag{3.9}$$

$$\lambda_i^\star \ge 0, \tag{3.10}$$

$$\nabla_{\mathbf{x}} f_0(\mathbf{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla_{\mathbf{x}} f_i(\mathbf{x}^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla_{\mathbf{x}} h_i(\mathbf{x}^\star) = \mathbf{0}, \tag{3.11}$$

$$\lambda_i^\star f_i(\mathbf{x}^\star) = 0. \tag{3.12}$$

Under some technical conditions (called constraint qualifications), the KKT conditions are necessary and sufficient for optimality. One simple version of the constraint qualifications is Slater's condition, which is satisfied when there exists $\mathbf{x}$ such that $f_i(\mathbf{x}) < 0$, $1 \le i \le m$ and $h_i(\mathbf{x}) = 0$, $1 \le i \le p$ (such a point is sometimes called *strictly feasible* since the inequality constraints hold with strict inequalities) [Roc70, Boy00].

In practice, the KKT conditions are very useful to obtain optimal solutions analytically (whenever this is possible).

3.1.4 Sensitivity Analysis

The optimal dual variables (Lagrange multipliers) of a convex optimization problem give useful information about the sensitivity of the optimal value with respect to perturbations of the constraints [Lue69, Boy00]. Consider the following perturbed version of the original convex problem (3.1):

$$\begin{array}{lll} \min\limits_{\mathbf{x}} & f_0(\mathbf{x}) & \\ \text{s.t.} & f_i(\mathbf{x}) \le u_i & 1 \le i \le m, \\ & h_i(\mathbf{x}) = v_i & 1 \le i \le p. \end{array} \tag{3.13}$$

This problem coincides with the original problem (3.1) if $u_i = 0$ and $v_i = 0$. When u_i is positive it means that we have relaxed the ith inequality constraint; when u_i is negative, it means that we have tightened the constraint. We define $f^\star(\mathbf{u}, \mathbf{v})$ as the optimal value of the perturbed problem with the perturbations $\mathbf{u}$ and $\mathbf{v}$. Note that $f^\star(\mathbf{0}, \mathbf{0}) = f^\star$.

Suppose that $f^\star(\mathbf{u}, \mathbf{v})$ is differentiable at $\mathbf{u} = \mathbf{0}$, $\mathbf{v} = \mathbf{0}$. Then, provided that strong duality holds, the optimal dual variables $\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star$ are related to the gradient of $f^\star(\mathbf{u}, \mathbf{v})$ at $\mathbf{u} = \mathbf{0}$, $\mathbf{v} = \mathbf{0}$

by [Boy00]

$$\lambda_i^\star = -\frac{\partial f^\star(\mathbf{0},\mathbf{0})}{\partial u_i}, \quad \nu_i^\star = -\frac{\partial f^\star(\mathbf{0},\mathbf{0})}{\partial v_i}. \tag{3.14}$$

This means that tightening the ith constraint a small amount (*i.e.*, taking u_i small and negative) yields an increase in $f^\star$ of approximately $-\lambda_i^\star u_i$ and, similarly, loosening the ith constraint a small amount (*i.e.*, taking u_i small and positive) yields a decrease in $f^\star$ of approximately $\lambda_i^\star u_i$.

The sensitivity result of (3.14) allows us to assign a numerical value to how *active* a constraint is at the optimum $\mathbf{x}^\star$. If $f_i(\mathbf{x}^\star) < 0$, then the constraint is inactive and it follows that the constraint can be tightened or loosened a small amount without affecting the optimal value (this agrees with the fact that $\lambda_i^\star$ must be zero by the complementary slackness condition). Suppose now that $f_i(\mathbf{x}^\star) = 0$, *i.e.*, the ith constraint is active at the optimum. The ith optimal Lagrange multiplier tells us how active the constraint is: if $\lambda_i^\star$ is small, it means that the constraint can be loosened or tightened a small amount without much effect on the optimal value; if $\lambda_i^\star$ is large, it means that if the constraint is loosened or tightened a small amount, the effect on the optimal value will be great.

3.2 Majorization Theory

Many of the problems addressed in this dissertation result in complicated nonconvex constrained optimization problems that involve matrix-valued variables. Majorization theory is a key tool that will allow us to convert these problems into simple convex problems with scalar variables that can be optimally solved.

In this section, we introduce the basic notion of majorization and state some basic results that will be needed in the sequel. A complete reference on the subject is [Mar79].

3.2.1 Basic Definitions

Majorization makes precise the vague notion that the components of a vector $\mathbf{x}$ are "less spread out" or "more nearly equal" than the components of a vector $\mathbf{y}$.

Definition 3.1 *[Mar79] For any* $\mathbf{x} \in \mathbb{R}^n$*, let*

$$x_{[1]} \geq \cdots \geq x_{[n]}$$

denote the components of vector $\mathbf{x}$ *in decreasing order, and let*

$$x_{(1)} \leq \cdots \leq x_{(n)}$$

denote the components of vector $\mathbf{x}$ *in increasing order.*

Definition 3.2 *[Mar79, 1.A.1] For any* $\mathbf{x}, \mathbf{y} \in I\!R^n$*,* $\mathbf{x}$ *is majorized by* $\mathbf{y}$ *(or* $\mathbf{y}$ *majorizes* $\mathbf{x}$*) if*

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]} \qquad 1 \leq k \leq n-1$$
$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}$$

and it is denoted by $\mathbf{x} \prec \mathbf{y}$ *or, equivalently, by* $\mathbf{y} \succ \mathbf{x}$*.*

Alternatively, the previous conditions can be rewritten as

$$\sum_{i=1}^{k} x_{(i)} \geq \sum_{i=1}^{k} y_{(i)} \qquad 1 \leq k \leq n-1$$
$$\sum_{i=1}^{n} x_{(i)} = \sum_{i=1}^{n} y_{(i)}.$$

Definition 3.3 *[Mar79, 1.A.2] For any* $\mathbf{x}, \mathbf{y} \in I\!R^n$*,* $\mathbf{x}$ *is weakly majorized by* $\mathbf{y}$ *(or* $\mathbf{y}$ *weakly majorizes* $\mathbf{x}$*) if*

$$\sum_{i=1}^{k} x_{(i)} \geq \sum_{i=1}^{k} y_{(i)} \qquad 1 \leq k \leq n$$

and it is denoted by $\mathbf{x} \prec^w \mathbf{y}$ *or, equivalently, by* $\mathbf{y} \succ^w \mathbf{x}$*.*[2]

Note that $\mathbf{x} \prec \mathbf{y}$ implies $\mathbf{x} \prec^w \mathbf{y}$; in other words, majorization is a more restrictive definition than weakly majorization.

Definition 3.4 *[Mar79, 3.A.1] A real-valued function* ϕ *defined on a set* $\mathcal{A} \subseteq I\!R^n$ *is said to be Schur-convex on* $\mathcal{A}$ *if*

$$\mathbf{x} \prec \mathbf{y} \quad \text{on } \mathcal{A} \quad \Rightarrow \quad \phi(\mathbf{x}) \leq \phi(\mathbf{y}).$$

Similarly, ϕ *is said to be Schur-concave on* $\mathcal{A}$ *if*

$$\mathbf{x} \prec \mathbf{y} \quad \text{on } \mathcal{A} \quad \Rightarrow \quad \phi(\mathbf{x}) \geq \phi(\mathbf{y}).$$

As a consequence, if ϕ *is Schur-convex on* $\mathcal{A}$ *then* $-\phi$ *is Schur-concave on* $\mathcal{A}$ *and vice-versa.*

It is important to remark that the sets of Schur-concave and Schur-convex functions do no form a partition of the set of all functions from $\mathcal{A} \subseteq I\!R^n$ to $I\!R$. In fact, neither are the two sets disjoint (the intersection is not empty), nor do they cover the entire set of all functions (see Figure 3.1). We now give some illustrative examples.

[2] More specifically, $\mathbf{x}$ is said to be weakly supermajorized by $\mathbf{y}$ (as opposed to the submajorization relation denoted by $\mathbf{x} \prec_w \mathbf{y}$ [Mar79, 1.A.2]).

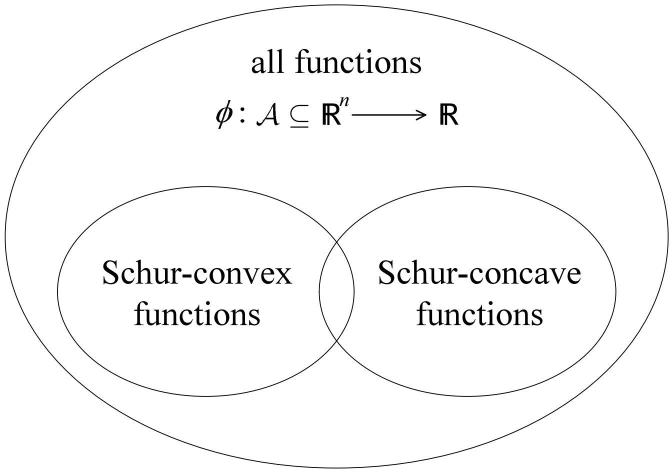


Figure 3.1: Illustration of the sets of Schur-convex and Schur-concave functions within the set of all functions $\phi : \mathcal{A} \subseteq I\!R^n \longrightarrow I\!R$.

Example 3.1 *The function* $\phi(\mathbf{x}) = \sum_{i=1}^{n} x_i$ *is both Schur-convex and Schur-concave since* $\phi(\mathbf{x}) = \phi(\mathbf{y})$ *for any* $\mathbf{x} \prec \mathbf{y}$.

Example 3.2 *The function* $\phi(\mathbf{x}) = c$ *is trivially both Schur-convex and Schur-concave.*

Example 3.3 *The function* $\phi(\mathbf{x}) = x_1 + 2x_2 + x_3$ *is neither Schur-convex nor Schur-concave as can be seen from the counterexample given by* $\mathbf{x} = [2,1,1]^T$, $\mathbf{y} = [2,2,0]^T$ *and* $\mathbf{z} = [4,0,0]^T$, *from which* $\mathbf{x} \prec \mathbf{y} \prec \mathbf{z}$ *but* $\phi(\mathbf{x}) < \phi(\mathbf{y}) > \phi(\mathbf{z})$.

Definition 3.5 *[Mar79, p. 21] A T-transform is a matrix of the form*

$$\mathbf{T} = \alpha \mathbf{I} + (1-\alpha)\mathbf{Q} \tag{3.15}$$

for some $\alpha \in [0,1]$ *and some permutation matrix* $\mathbf{Q} \in I\!R^{n\times n}$ *with* $n-2$ *diagonal entries equal to 1. Let* $[\mathbf{Q}]_{ij} = [\mathbf{Q}]_{ji} = 1$ *for some indices* $i < j$, *then*

$$\mathbf{Qy} = [y_1, \cdots, y_{i-1}, y_j, y_{i+1}, \cdots, y_{j-1}, y_i, y_{j+1}, \cdots, y_n]^T$$

and hence

$$\mathbf{Ty} = [y_1, \cdots, y_{i-1}, \alpha y_i + (1-\alpha) y_j, y_{i+1}, \cdots, y_{j-1}, \alpha y_j + (1-\alpha) y_i, y_{j+1}, \cdots, y_n]^T.$$

3.2.2 Basic Results

Lemma 3.1 *[Mar79, p. 7] For any* $\mathbf{x} \in I\!R^n$, *let* $\mathbf{1} \in I\!R^n$ *denote the constant vector with the ith element given by* $1_i \triangleq \sum_{j=1}^{n} x_j/n$, *then*

$$\mathbf{1} \prec \mathbf{x}.$$

Lemma 3.1 is simply stating the obvious fact that a vector of equal components has the "least spread out" or the "most equal" components.

Lemma 3.2 *[Mar79, 3.B.1] An increasing function of a Schur-convex (Schur-concave) function is Schur-convex (Schur-concave). Similarly, a decreasing function of a Schur-convex (Schur-concave) function is Schur-concave (Schur-convex).*

Lemma 3.3 *[Mar79, 3.H.2] Let* $\phi(\mathbf{x}) = \sum_i g_i(x_i)$ *where* $x_i \geq x_{i+1}$ *and each* g_i *is differentiable. Then* ϕ *is Schur-convex if and only if*

$$g_i'(a) \geq g_{i+1}'(b) \qquad \textit{whenever} \quad a \geq b, \quad i = 1, \ldots, n-1.$$

Corollary 3.1 *Let* $\phi(\mathbf{x}) = \sum_i g(x_i)$ *where* g *is convex. Then* ϕ *is Schur-convex.*

Lemma 3.4 *[Mar79, 5.A.9.a] For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *satisfying* $\mathbf{y} \succ^w \mathbf{x}$*, there exists a vector* $\mathbf{u}$ *such that*

$$\mathbf{u} \leq \mathbf{x} \quad \textit{and} \quad \mathbf{y} \succ \mathbf{u}.$$

The following lemma is a fundamental result of majorization theory.

Lemma 3.5 *[Mar79, 2.B.1] For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *satisfying* $\mathbf{x} \prec \mathbf{y}$*, there exists a sequence of T-transforms* $\mathbf{T}^{(1)}, \cdots, \mathbf{T}^{(K)}$ *such that* $\mathbf{x} = \mathbf{T}^{(K)} \cdots \mathbf{T}^{(1)}\mathbf{y}$ *and* $K < n$.

We now give an algorithm to obtain such a sequence of T-transforms from [Mar79, 2.B.1].

Algorithm 3.1 *[Mar79, 2.B.1] Algorithm to obtain a sequence of T-transforms such that* $\mathbf{x} = \mathbf{T}^{(K)} \cdots \mathbf{T}^{(1)}\mathbf{y}$.

Input: *Vectors* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *satisfying* $\mathbf{x} \prec \mathbf{y}$ *(it is assumed that the components of* $\mathbf{x}$ *and* $\mathbf{y}$ *are in decreasing order and that* $\mathbf{x} \neq \mathbf{y}$*).*

Output: *Set of T-transforms* $\mathbf{T}^{(1)}, \cdots, \mathbf{T}^{(K)}$.

0. *Let* $\mathbf{y}^{(0)} = \mathbf{y}$ *and* $k = 1$ *be the iteration index.*
1. *Find the largest index* i *such that* $y_i^{(k-1)} > x_i$ *and the smallest index* j *greater than* i *such that* $y_j^{(k-1)} < x_j$.
2. *Let* $\delta = \min\left(x_j - y_j^{(k-1)}, y_i^{(k-1)} - x_i\right)$ *and* $\alpha = 1 - \delta / \left(y_i^{(k-1)} - y_j^{(k-1)}\right)$.
3. *Use* α *to compute* $\mathbf{T}^{(k)}$ *as in (3.15) and let* $\mathbf{y}^{(k)} = \mathbf{T}^{(k)}\mathbf{y}^{(k-1)}$.
4. *If* $\mathbf{y}^{(k)} \neq \mathbf{x}$*, then set* $k = k + 1$ *and go to step 1; otherwise, finish.*

Lemma 3.6 *[Mar79, 9.B.1] Let* $\mathbf{R}$ *be an* $n \times n$ *Hermitian matrix with diagonal elements denoted by the vector* $\mathbf{d}$ *and eigenvalues denoted by the vector* $\boldsymbol{\lambda}$*, then*

$$\boldsymbol{\lambda} \succ \mathbf{d}.$$

Lemma 3.7 *[Mar79, 9.B.2] For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *satisfying* $\mathbf{x} \prec \mathbf{y}$*, there exists a real symmetric (and therefore Hermitian) matrix with diagonal elements given by* $\mathbf{x}$ *and eigenvalues given by* $\mathbf{y}$*.*

Lemma 3.7 is the converse of Lemma 3.6 (in fact it is stronger than the converse since it guarantees the existence of a real symmetric matrix instead of just a Hermitian matrix). A recursive algorithm to obtain a matrix with a given vector of eigenvalues and vector of diagonal elements is indicated in [Mar79, 9.B.2] and [Vis99b, Section IV-A]. We consider the practical and simple method obtained in [Vis99b, Section IV-A] and reproduce it here for completeness.

Algorithm 3.2 *[Vis99b, Section IV-A] Algorithm to obtain a real symmetric matrix* $\mathbf{R}$ *with diagonal values given by* $\mathbf{x}$ *and eigenvalues given by* $\mathbf{y}$ *provided that* $\mathbf{x} \prec \mathbf{y}$*.*

Input: *Vectors* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *satisfying* $\mathbf{x} \prec \mathbf{y}$ *(it is assumed that the components of* $\mathbf{x}$ *and* $\mathbf{y}$ *are in decreasing order and that* $\mathbf{x} \neq \mathbf{y}$*).*

Output: *Matrix* $\mathbf{R}$*.*

1. *Using Algorithm 3.1, obtain a sequence of T-transforms such that* $\mathbf{x} = \mathbf{T}^{(K)} \cdots \mathbf{T}^{(1)} \mathbf{y}$*.*
2. *Define the Givens rotation* $\mathbf{U}^{(k)}$ *as* $\left[\mathbf{U}^{(k)}\right]_{ij} = \begin{cases} \sqrt{\left[\mathbf{T}^{(k)}\right]_{ij}} & \text{for } \ i < j \\ -\sqrt{\left[\mathbf{T}^{(k)}\right]_{ij}} & \text{otherwise} \end{cases}$*.*
3. *Let* $\mathbf{R}^{(0)} = \operatorname{diag}(\mathbf{y})$ *and* $\mathbf{R}^{(k)} = \mathbf{U}^{(k)T} \mathbf{R}^{(k-1)} \mathbf{U}^{(k)}$*. The desired matrix is given by* $\mathbf{R} = \mathbf{R}^{(K)}$*. Alternatively, define the unitary matrix* $\mathbf{Q} = \mathbf{U}^{(1)} \cdots \mathbf{U}^{(K)}$ *and the desired matrix is given by* $\mathbf{R} = \mathbf{Q}^T \operatorname{diag}(\mathbf{y}) \mathbf{Q}$*.*

Corollary 3.2 *For any* $\boldsymbol{\lambda} \in \mathbb{R}^n$*, there exists a real symmetric (and therefore Hermitian) matrix with equal diagonal elements and eigenvalues given by* $\boldsymbol{\lambda}$*.*

Proof. The proof is straightforward from Lemmas 3.1 and 3.7. ■

As before, such matrix can be obtained using Algorithm 3.2. However, for this particular case and allowing the desired matrix to be complex, it is easy to see that any unitary matrix $\mathbf{Q}$ satisfying the condition $|[\mathbf{Q}]_{ik}| = |[\mathbf{Q}]_{il}| \ \forall i, k, l$ provides a valid solution given by $\mathbf{R} = \mathbf{Q}^H \operatorname{diag}(\mathbf{y}) \mathbf{Q}$. As an example, the unitary Discrete Fourier Transform (DFT) matrix and the Hadamard matrix (when the dimensions are appropriate such as a power of two [Ver98, p.66][Pet78, Sec.5.6]) satisfy this condition. Nevertheless, Algorithm 3.2 has the nice property that the obtained matrix $\mathbf{Q}$ is real-valued and can be naturally decomposed (by construction) as the product of Givens rotations (where each term performs a single rotation [Söd89]). This simple structure plays a key role for practical implementation. Interestingly, an iterative approach to construct a matrix with equal diagonal elements and with a given set of eigenvalues was obtained in [Mul76], based also on a sequence of rotations.

3.3 Miscellaneous Algebra Results

In this section, we include for convenience a few basic algebra results that are standard fare in most textbooks (*e.g.*, [Hor85, Mag99, Söd89, Kay93, Sch91]) and that will be repeatedly used throughout this dissertation.

Basic results on the trace and determinant

The following relation, commonly referred to as the circularity of the trace, is widely used:

$$\mathrm{Tr}\left(\mathbf{AB}\right) = \mathrm{Tr}\left(\mathbf{BA}\right).$$

The following is a basic result on the determinant for conformable matrices [Mag99, Söd89]:

$$\left|\mathbf{I} + \mathbf{AB}\right| = \left|\mathbf{I} + \mathbf{BA}\right|.$$

Another useful result is

$$\mathbf{A} \geq \mathbf{B} \;\Rightarrow \left|\mathbf{A}\right| \geq \left|\mathbf{B}\right|.$$

Matrix Inversion Lemma

The general expression of the matrix inversion lemma is [Kay93, Sch91, Söd89, Hor85]

$$\left(\mathbf{A} + \mathbf{BCD}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1}\right)^{-1}\mathbf{DA}^{-1}.$$

A particular case is the Woodbury's Identity:

$$\left(\mathbf{R} + \gamma^2\mathbf{cc}^H\right)^{-1} = \mathbf{R}^{-1} - \frac{\gamma^2}{1 + \gamma^2\mathbf{c}^H\mathbf{R}^{-1}\mathbf{c}}\mathbf{R}^{-1}\mathbf{cc}^H\mathbf{R}^{-1}.$$

Cauchy-Schwarz's inequality

The Cauchy-Schwarz's inequality in vector form is [Mag99, Hor85]

$$\left|\mathbf{y}^H\mathbf{x}\right| \leq \left\|\mathbf{y}\right\|_2 \left\|\mathbf{x}\right\|_2$$

with equality if and only if $\mathbf{y} = \alpha\mathbf{x}$, *i.e.*, if $\mathbf{y}$ and $\mathbf{x}$ are linearly dependent.

Hadamard's inequality

Given an $n \times n$ positive semidefinite matrix $\mathbf{R}$, the following holds [Cov91b, Mag99, Hor85]:

$$|\mathbf{R}| \leq \prod_{i=1}^{n} [\mathbf{R}]_{ii}$$

with equality if and only if matrix $\mathbf{R}$ is diagonal (except in the trivial case in which $\mathbf{R}$ is singular).

Jensen's inequality [Boy00, Cov91b]

If f is a convex function, $x_1, \cdots, x_k \in \operatorname{dom} f$, and $\theta_1, \cdots, \theta_k \geq 0$ with $\theta_1 + \cdots + \theta_k = 1$, then

$$f\left(\theta_1 x_1 + \cdots + \theta_k x_k\right) \leq \theta_1 f\left(x_1\right) + \cdots + \theta_k f\left(x_k\right).$$

Moreover, if f is strictly convex, then equality implies that the x_i's for which $\theta_i > 0$ are equal.

The inequality extends to infinite sums, integrals, and expected values. For example, if x is a random variable such that $x \in \operatorname{dom} f$ with probability one and f is convex, then

$$f\left(\mathbb{E}\left[x\right]\right) \leq \mathbb{E}\left[f\left(x\right)\right],$$

provided the expectations exist. Moreover, if f is strictly convex, then equality implies that x is a constant.

Chapter 4

Capacity of MIMO Channels

THE CAPACITY OF A CHANNEL is a fundamental limit that characterizes the highest rate at which information can be reliably transmitted, *i.e.*, with arbitrarily low probability of error. The channel capacity theorem by Shannon back in 1948 [Sha48] is the central and most famous success of information theory. In this chapter, we explore the fundamental communication limits of MIMO channels for different degrees of channel state information (CSI). For the case of no CSI, novel results are obtained within the framework of game theory.

4.1 Introduction

The simplest communication scenario is that composed of a single transmitter and a single receiver, *i.e.*, a single-user communication system (see Figure 4.1). In this case, the message W, drawn from the index set $\{1, 2, \cdots, M\}$, results in the signal vector (codeword) $\mathbf{x}^n(W)$ of block-length n which is successively transmitted as $\mathbf{x}_1 \cdots \mathbf{x}_n$ in "n uses of the channel". The transmitted signal $\mathbf{x}^n(W)$ goes through the channel described by the transition probability $p(\mathbf{y}^n \mid \mathbf{x}^n)$ to be received as the random sequence $\mathbf{y}^n$, similarly composed of n received samples $\mathbf{y}_1 \cdots \mathbf{y}_n$. The receiver guesses the message W by an appropriate decoding rule $\hat{W}(\mathbf{y}^n)$. The channel is assumed to be memoryless such that $p(\mathbf{y}^n \mid \mathbf{x}^n) = \prod_{i=1}^{n} p(\mathbf{y}_i \mid \mathbf{x}_i)$. The whole communication process is illustrated in Figure 4.1.

A more general scenario is that consisting of m nodes trying to communicate with each other, *i.e.*, a multiuser communication system. In such a case, the more general notion of capacity

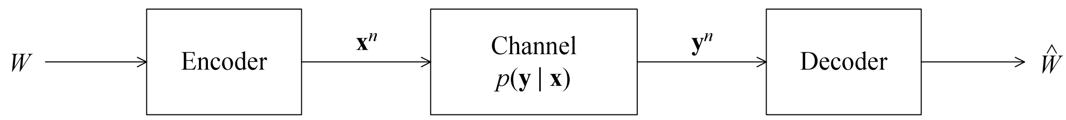
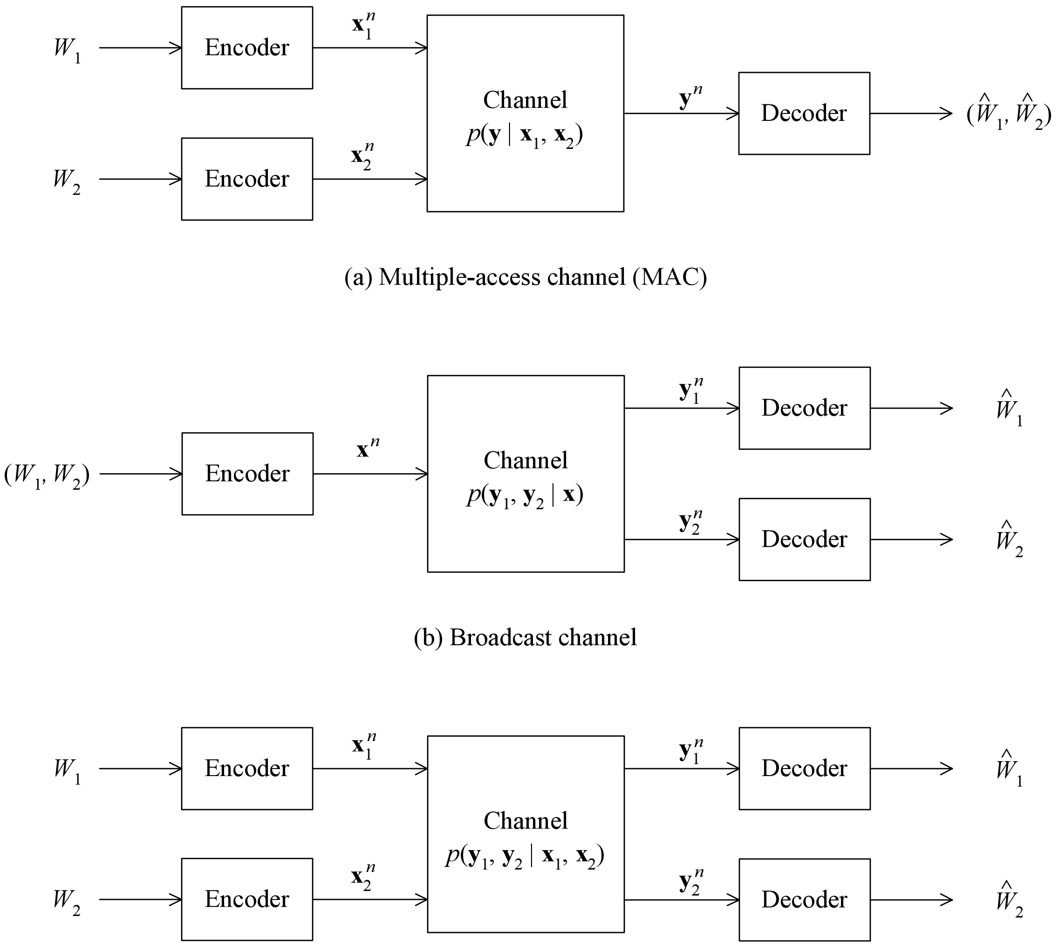


Figure 4.1: Scheme of a single-user communication system.



(a) Multiple-access channel (MAC)

(b) Broadcast channel

(c) Interference channel

Figure 4.2: Scheme of three different (and very common) particular cases of a multiuser communication system (for simplicity, only two users are considered).

region of the network is necessary to characterize the set of achievable rates. The channel is then described by the transition probability $p(\mathbf{y}_1, \cdots, \mathbf{y}_m \mid \mathbf{x}_1, \cdots, \mathbf{x}_m)$.[1] This general problem (m independent transmitters and m independent receivers) has not yet been solved and special cases have to be considered (see Figure 4.2) such as the multiple-access channel (MAC) (m independent transmitters and a single receiver), the broadcast channel (one transmitter and m independent receivers), and the interference channel (m independent pairs of transmitter-receiver that interfere with each other) [Cov91b]. Note that the difference between the different multiuser scenarios lies on whether cooperation is possible among the transmitters and/or among the receivers (see Figure 4.2).

In this chapter, we will only address the single-user channel and the MAC with Gaussian channel transition probabilities (they lead to simple analysis and real systems are in many cases

[1]In the most general case, all m nodes play the simultaneous role of transmitters and receivers.

accurately modeled with a Gaussian channel transition probability because the noise is Gaussian). For completeness, we recall the signal model for the single-user channel with n_T transmit and n_R receive (finite) dimensions

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{4.1}$$

where $\mathbf{x} \in \mathbb{C}^{n_T \times 1}$ is the transmitted vector, $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ is the channel matrix, $\mathbf{y} \in \mathbb{C}^{n_R \times 1}$ is the received vector, and $\mathbf{n} \in \mathbb{C}^{n_R \times 1}$ is a proper complex Gaussian noise vector $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_n)$.[2] For future reference, recall that $\tilde{\mathbf{H}} \triangleq \mathbf{R}_n^{-1/2}\mathbf{H}$ is the whitened channel matrix as was defined in (2.2). The covariance matrix of the transmitted vector signal is $\mathbf{Q} = \mathbb{E}[\mathbf{x}\mathbf{x}^H]$ and the transmitter is constrained in its average transmit power as $\mathrm{Tr}(\mathbf{Q}) \leq P_T$ (c.f. §2.5.2). By uniform power allocation we mean $\mathbf{Q} = P_T/n_T\, \mathbf{I}_{n_T}$, which also implies an independent signaling over the transmit dimensions if a Gaussian code is used. Another interesting constraint is the maximum eigenvalue constraint $\lambda_{\max}(\mathbf{Q}) \leq \alpha$ (c.f. §2.5.2) which will be revisited in Section 4.4. Note that $\lambda_{\max}(\mathbf{Q})$ is an upper-bound on the average transmitted power at each transmit dimension $\mathbb{E}[|x_i|^2] \leq \lambda_{\max}(\mathbf{Q})$.

Similarly, the signal model for the multiple-access channel with K users, each one transmitting over n_k dimensions with a (possibly different) power constraint $\mathrm{Tr}(\mathbf{Q}_k) \leq P_k$ and with channel $\mathbf{H}_k \in \mathbb{C}^{n_R \times n_k}$, is

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{x}_k + \mathbf{n}. \tag{4.2}$$

Different types of capacities are obtained depending on the characteristics of the communication system such as the deterministic/random nature of the channel, the degree of CSI at the transmitter (CSIT) and at the receiver (CSIR), and the delay constraints [Big98]. Perfect (instantaneous) CSIR is considered throughout this chapter under the reasonable assumption that the receiver may obtain sufficiently good channel estimates (c.f. §2.4). In fact, it is not necessary to assume an a priori perfect CSIR; provided that the channel remains fixed during a sufficiently long time, the capacity does not depend on whether the channel state is available or not at the receiver [Big98].[3] Regarding the transmitter, when perfect CSIT is available, the instantaneous capacity (also termed deterministic capacity) at each channel state is a meaningful measure. This situation may arise either when the channel is fixed and deterministic or when it is a fading channel with a sufficiently slow fading so that the channel can be estimated and fed back to the transmitter (c.f. §2.4). In many situations, however, the transmitter may not have an instantaneous CSI. In such cases, provided that the channel statistics are known at the transmitter (statistical CSIT),

[2] Recall that, for a given $\mathbf{H}$ and since the noise is Gaussian distributed, the channel transition probability is Gaussian: $p(\mathbf{y} \mid \mathbf{x}) = \frac{1}{\pi^{n_R}|\mathbf{R}_n|} e^{-(\mathbf{y}-\mathbf{H}\mathbf{x})^H \mathbf{R}_n^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x})}$.

[3] The intuitive explanation of this effect is that, since the channel state remains fixed for the transmission of the whole codeword, for sufficiently long codes, it can be estimated at the receiver by transmitting, for example, a training sequence with length proportional to $\sqrt{n}$ at no cost of rate as $n \to \infty$ [Wol78]. In fact, the channel state is not at all required at the receiver with the utilization of universal decoders [Lap98].

the notions of ergodic and outage capacities (depending on the delay constraints of the system) provide meaningful measures as will be further discussed. It may also be interesting to consider the most pessimistic scenario in which not even the statistics of the channel are known at the transmitter (no CSIT), implying the necessity of a robust communication scheme under channel uncertainty. In such a case, as will be analyzed in detail, the notion of worst-case capacity (also termed compound capacity) arises naturally and can be formulated as a game with two players: the transmitter and a malicious nature that controls the channel state and the noise statistics.

The main contribution of this chapter is the characterization of the worst-case capacity within the framework of game theory, obtaining the interesting novel result that a uniform power allocation (transmitting in all directions) is the best thing to do when nothing is known about the channel (under a mild isotropy condition). Interestingly, this result holds for the MAC as well.

This chapter is structured as follows. First, the instantaneous capacity as well as the ergodic/outage capacity are briefly described in Sections 4.2 and 4.3, respectively. After that, the worst-case capacity is extensively analyzed in Section 4.4, providing original results.

The results in this chapter regarding the instantaneous capacity in beamforming-constrained systems and the worst-case capacity under a game-theoretic framework have been published in [Pal01c, Pal03e] and [Pal03a, Pal03d], respectively.

4.2 Instantaneous Capacity

The notion of instantaneous capacity is meaningful only when instantaneous CSI is available at both sides of the link, which happens either when the channel remains basically fixed such as in DSL systems or when it changes sufficiently slow so that it can be considered fixed during the sufficient number of transmissions. In such a case, capacity is achieved by adapting the transmitted signal to the specific channel realization as we now review.

4.2.1 Capacity of the Single-User Channel

For the single-user channel, the optimum signaling to achieve rates up to the channel capacity is well known [Bra74, Cov91b, Tel95]. The capacity of a channel is the maximum mutual information between the transmitted and the received signals $\mathcal{I}(\mathbf{x};\mathbf{y})$ over all possible input distributions satisfying the power constraint [Bla87, Cov91b]:

$$\mathcal{C} = \max_{p(\mathbf{x}):\mathbb{E}\|\mathbf{x}\|^2 \leq P_T} \mathcal{I}(\mathbf{x};\mathbf{y}). \tag{4.3}$$

For the vector Gaussian channel under consideration, it is well known that the maximum mutual information is achieved with Gaussian inputs $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$ (*i.e.*, when a Gaussian code

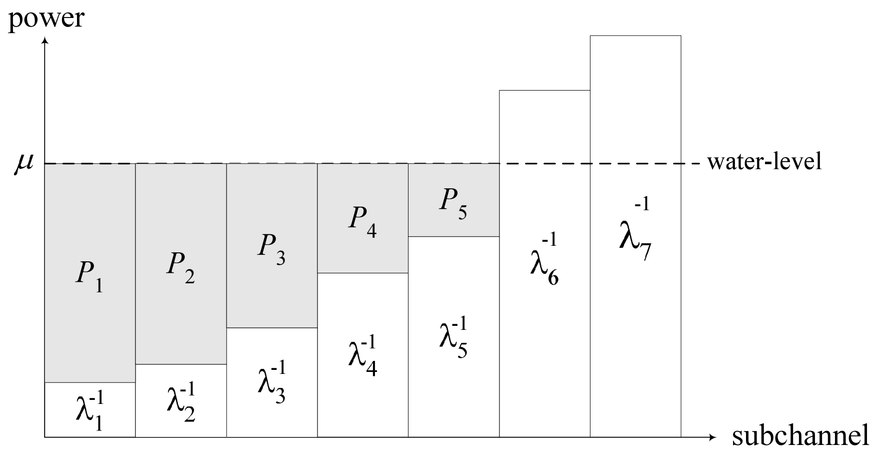


Figure 4.3: Illustration of the water-filling power allocation $P_i = \left(\mu - \lambda_i^{-1}\right)^+$.

is used for transmission), where $\mathbf{Q}$ is the covariance matrix of the transmitted vector $\mathbf{x}$, and the mutual information is then [Cov91b, Tel95]

$$\mathcal{I}\left(\mathbf{x};\mathbf{y}\right) = \log\det\left(\mathbf{I}_{n_R} + \mathbf{R}_n^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) \tag{4.4}$$

where the units are nats/transmission since the natural logarithm has been used (note that any other base can be chosen affecting only the units of the mutual information). The channel capacity for each channel state $\mathbf{H}$ is given by the maximum mutual information over all $\mathbf{Q}$ verifying the power constraint:

$$\mathcal{C}\left(\mathbf{H}\right) = \max_{\substack{\mathbf{Q}:\mathrm{Tr}(\mathbf{Q})\leq P_T,\\ \mathbf{Q}=\mathbf{Q}^H\geq \mathbf{0}}} \log\det\left(\mathbf{I}_{n_R} + \mathbf{R}_n^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) \tag{4.5}$$

where we have made explicit the dependence of the capacity on each particular channel state. This maximization problem has a well-known solution based on diagonalizing the whitened channel $\tilde{\mathbf{H}} = \mathbf{R}_n^{-1/2}\mathbf{H}$ and then distributing the available power over the channel eigenmodes in a water-filling fashion. To be more specific, the capacity-achieving solution is [Bra74, Cov91b, Ral98, Tel95, Sca99a]

$$\mathbf{Q} = \mathbf{U}_{R_H}\mathbf{D}_Q\mathbf{U}_{R_H}^H \tag{4.6}$$

where $\mathbf{U}_{R_H}$ contains the eigenvectors of $\mathbf{R}_H = \mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}$ (whose EVD is given by $\mathbf{R}_H = \mathbf{U}_{R_H}\mathbf{D}_{R_H}\mathbf{U}_{R_H}^H$) and $\mathbf{D}_Q$ is a diagonal matrix with diagonal elements $\{\lambda_{Q,i}\}$ given by the water-filling solution

$$\lambda_{Q,i} = \left(\mu - \lambda_{R_H,i}^{-1}\right)^+ \quad 1 \leq i \leq n_T \tag{4.7}$$

where $(x)^+ \triangleq \max(0,x)$ and μ is the water-level chosen to satisfy the power constraint with equality $\sum_i \lambda_{Q,i} = P_T$. This optimal power allocation has the appealing interpretation based on considering the available power as water which is poured over a surface inversely proportional to the channel gain, hence the name water-filling or water-pouring (see Figure 4.3) [Cov91b]. In particular, this implies that eigenmodes with higher gain receive higher power and vice-versa; in addition, there is a threshold under which no power is allocated to an eigenmode (for a given channel, the number of used eigenmodes depends on the available power at the transmitter P_T). The capacity can be then written as

$$\mathcal{C}(\mathbf{H}) = \sum_i \log\left(1 + \lambda_{Q,i}\,\lambda_{R_H,i}\right) \tag{4.8}$$

$$= \sum_i \left(\log\left(\mu\,\lambda_{R_H,i}\right)\right)^+ \tag{4.9}$$

from which the channel capacity is achieved by properly using the constituent channel eigenmodes.

It will be convenient in the sequel to express the mutual information between $\mathbf{x}$ and $\mathbf{y}$ explicitly as a function of $\mathbf{Q}$ and $\tilde{\mathbf{H}}$ (assuming a Gaussian code) as in [Tel95]:

$$\Psi\left(\mathbf{Q}, \tilde{\mathbf{H}}\right) = \log\det\left(\mathbf{I}_{n_R} + \tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^H\right) \tag{4.10}$$

$$= \log\det\left(\mathbf{I}_{n_T} + \mathbf{Q}\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right) \tag{4.11}$$

Beamforming-Constrained Systems

We now consider the capacity of a channel when the transmitter is constrained to use a beamvector (termed single beamforming in §2.5.1). Such a scheme is preferred in terms of complexity because only one data stream needs to be considered and coding and transmission can be done in a much easier manner (as in single antenna systems) [Nar99]. However, the utilization of a beamvector at the transmitter implies a rank-one transmit covariance matrix which may not be optimum. It is therefore necessary to analyze when such a communication structure does not incur any loss of optimality [Pal01c, Pal01a, Pal03e].

The capacity expression when the communication system is constrained to use a rank-one transmit covariance matrix (beamforming) is (similarly to (4.8)) given by [Pal01c, Pal03e]

$$\mathcal{C}^{\mathrm{bf}}(\mathbf{H}) = \log\left(1 + P_T\,\lambda_{R_H,\max}\right) \tag{4.12}$$

where only the best channel eigenmode is used for transmission. Such a beamforming scheme is, in principle, suboptimum because it uses only one channel eigenmode (compare (4.12) with (4.8)). To be exact, beamforming is optimum if and only if $P_T \leq \lambda_{R_H,2}^{-1} - \lambda_{R_H,1}^{-1}$ where the $\lambda_{R_H,i}$'s are assumed in decreasing order. This condition is satisfied if a sufficiently small value of transmit power is used (provided that $\lambda_{R_H,1} \neq \lambda_{R_H,2}$).

In wireless multi-antenna MIMO systems, for example, the optimality conditions are satisfied for some situations with high spatial correlation. In particular, beamforming is asymptotically optimum (in the sense of achieving capacity) as the spatial channel fading correlation increases at least at one end of the link because the channel matrix becomes rank-one: $\mathbf{H} \rightarrow \mathbf{h}_R\mathbf{h}_T^H$ [Pal03e].

In [Jaf01, Sim03], the optimality of beamforming was analyzed for different degrees of channel feedback quality. In [Ivr03], different transmission strategies were considered (including a

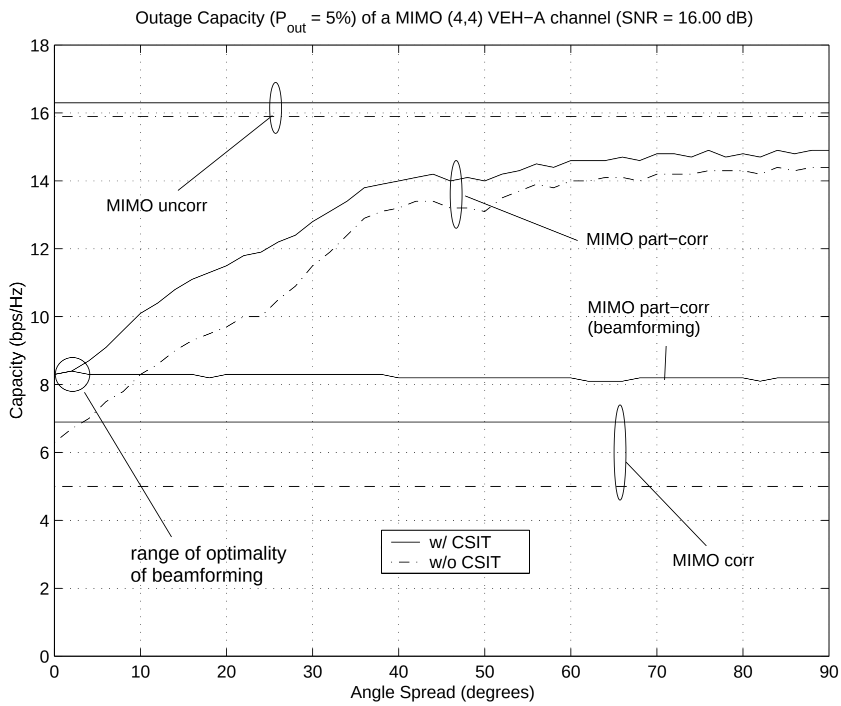


Figure 4.4: Outage capacity (at an outage probability of 5%) of a 4 × 4 MIMO Vehicular-A channel (SNR=16dB) as a function of the angle spread (part-corr) with CSIT (w/ CSIT), without CSIT (w/o CSIT), and when beamforming is used. In addition, the capacities with and without CSIT corresponding to fully uncorrelated (uncorr) and completely correlated (corr) MIMO channels are also plotted for comparison.

beamforming approach) depending on the degree of channel knowledge at the transmitter and the spatial correlation.

Numerical Example

We now evaluate numerically the range of optimality of beamforming in a typical wireless multi-antenna MIMO channel (in particular, we consider 4 transmit and 4 receive antennas).

The transmitter (mobile unit) is assumed to be immersed in a rich scattering environment and has an almost uncorrelated fading among the antennas. The receiver (base station), however, is assumed to be on top of a building, receiving the signal from a mean direction of arrival with a given angle spread. The spatial fading correlation was obtained using the closed-form expression derived in [Sal94] (where a uniform-shaped angular distribution was assumed[4]) along with the procedure explained in [Ped00] to include the corresponding spatial correlation into the generated MIMO channel. The generated channel is frequency-selective following a Vehicular power delay profile as specified by ETSI [ETS98b]. The noise covariance matrix was assumed white and the signal was received with an average SNR of 16dB. For a more detailed description of the

[4]Depending on the exact shape of the considered angular ray distribution, different spatial correlation are obtained. In general, however, this is just a minor detail and the conclusions derived under different ray distributions should be roughly the same.

simulation scenario and for additional plots, the reader is referred to [Pal03e].

In Figure 4.4, the capacity[5] is plotted as a function of the angle spread (partially correlated channel) for the following situations: with instantaneous CSIT (w/ CSIT) according to the water-filling solution, without CSIT (w/o CSIT) using a uniform power allocation among all the transmit dimensions, and when beamforming is used (for which instantaneous CSIT is required). In addition, the capacities with and without CSIT corresponding to fully uncorrelated (uncorr) and completely correlated (corr) MIMO channels are also plotted for comparison. For the particular scenario under consideration, we can infer that for an angle spread on the order of 4-8 degrees, the loss in capacity of the beamforming scheme is negligible.[6]

4.2.2 Capacity Region of the Multiple-Access Channel (MAC)

In the multiple-access channel, we do not deal anymore with a single capacity measure but with a capacity region, which is the set of achievable rates $R_1, \cdots, R_K$ (recall that, in the MAC, the users transmit independently, whereas the receiver can perform a joint detection). For the vector Gaussian MAC, it is well known that all the boundaries defining the achievable region are maximized when Gaussian codes are used for transmission (in other words, the rate region corresponding to a Gaussian signaling contains the rate region of other signaling distributions). The achievable rate region corresponding to a set of channel states $\{\mathbf{H}_k\}$ and a set of covariance matrices $\{\mathbf{Q}_k\}$ (assuming Gaussian codes) is [Cov91b, Ver89]

$$\mathcal{R}\left(\{\mathbf{Q}_k\}, \{\mathbf{H}_k\}\right) = \Bigg\{ (R_1, \cdots, R_K) : \\ 0 \leq \sum_{k \in \mathcal{S}} R_k \leq \log\det\left(\mathbf{I}_{n_R} + \sum_{k \in \mathcal{S}} \mathbf{R}_n^{-1} \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^H\right), \forall \mathcal{S} \subseteq \{1, \cdots, K\} \Bigg\}. \tag{4.13}$$

Recall that a joint detection is necessary in general to obtain all the rates within the region; otherwise, if an independent detection is used for all the users, the interference channel model as previously described (see Figure 4.2) would be more appropriate.

Assuming that the transmit covariance matrices are constrained in their average transmit

[5] To be more exact, we plot the capacity at an outage probability of 5% (*i.e.*, the capacity that is achieved in 95% of the cases). In this case, however, the transmitter has instantaneous CSIT which is a fundamental difference with respect to the outage capacity considered in §4.3 where only statistical CSIT is assumed.

[6] Note that the curve for partial correlation in Figure 4.4 is not monotonic on the angle spread. This is due to the specific model of the ray distribution (uniform-shaped angular distribution) used to compute the fading correlation [Sal94]. For other distribution models, slightly different curves are obtained but the underlying trend is always the same.

power, the capacity region is [Ver89, Yu01b]

$$\mathcal{C}\left(\{\mathbf{H}_k\}\right) = \bigcup_{\substack{\mathrm{Tr}(\mathbf{Q}_k)\leq P_k,\\ \mathbf{Q}_k=\mathbf{Q}_k^H\geq 0}} \mathcal{R}\left(\{\mathbf{Q}_k\},\{\mathbf{H}_k\}\right). \tag{4.14}$$

Note that the convex closure operation usually needed [Cov91b] is unnecessary in this case because the region is already closed and convex as shown in [Ver89, Yu01b].

Although the capacity region has been fully characterized in (4.14), it is not obvious how to obtain in practice the boundary points defining the region. Since (4.14) is a convex region, its boundary points can be found by maximizing the weighted sum of the data rates (different set of weights $\{\mu_k\}$ give different points of the boundary)

$$\begin{array}{ll} \max\limits_{\{R_k\}} & \sum_k \mu_k R_k \\ \text{s.t.} & \{R_k\} \in \mathcal{C}\left(\{\mathbf{H}_k\}\right) \end{array} \tag{4.15}$$

which is a convex optimization problem and therefore can be efficiently solved in practice (see §3.1). This problem was extensively treated in [Yu01b].

A simple measure of particular interest in multiple-access channels is the sum capacity, which is a single performance parameter as opposed to the capacity region. The sum capacity represents the maximum total rate $\sum_k R_k$ that can be achieved by the system and is obtained as the solution to the following convex optimization problem:

$$\begin{array}{lll} \max\limits_{\{\mathbf{Q}_k\}} & \log\det\left(\mathbf{I}_{n_R} + \sum\limits_{k=1}^{K} \mathbf{R}_n^{-1}\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H\right) & \\ \text{s.t.} & \mathrm{Tr}\left(\mathbf{Q}_k\right) \leq P_k, & 1 \leq k \leq K \\ & \mathbf{Q}_k = \mathbf{Q}_k^H \geq 0. & \end{array} \tag{4.16}$$

This problem turns out to be easily solved by the iterative water-filling algorithm proposed in [Yu01b], which solves, for each of the users and in a sequential manner, the single-user water-filling solution when the rest of the users are considered as noise, *i.e.*, when each user k considers the interference-plus-noise covariance matrix $\mathbf{R}_{n,k} = \sum_{l\neq k} \mathbf{H}_l\mathbf{Q}_l\mathbf{H}_l^H + \mathbf{R}_n$ [Yu01b]. As was found in [Yu01b], at an optimal point, each user must water-fill the rest of the users considered as noise.

Beamforming-Constrained Systems

We can also consider a beamforming-constrained system for the vector Gaussian MAC where each user transmit using beamforming. The optimality conditions obtained for the single-user case (sufficiently low transmit power and/or sufficiently high spatial correlation in a multi-antenna system) are also valid for the MAC. In this scenario, however, the beamforming approach turns out to be asymptotically optimal with the number of users regardless of the channel correlation and of the transmit power as shown in [Rhe01b]. To be more exact, the number of users must be sufficiently large compared to the number of receive dimensions (or antennas in a multi-antenna system) [Rhe01b]. The underlying idea is that the multiplexing gain of a MIMO channel is fully utilized on a multiuser level (by having many users doing beamforming) rather than on the individual level.

Utilizing beamforming implies rank-one transmit covariance matrices, *i.e.*, $\text{rank}(\mathbf{Q}_k) = 1$. The definition of the capacity region corresponding to such a constrained system is straightforward by including the rank-one constraint in the region defined in (4.14). The computation of such a capacity region, however, turns out to be a very complicated problem because the rank-one constraint is nonconvex and so is the whole problem (even the computation of the sum capacity is a nonconvex problem). It is useful to parameterize the transmit covariance matrices as $\mathbf{Q}_k = \mathbf{b}_k \mathbf{b}_k^H$ where $\mathbf{b}_k$ is the beamvector of the kth user (this parameterization is also very appropriate to describe CDMA systems where the beamvector denotes the code in time [Ulu01]). The sum capacity is then characterized as the following nonconvex problem:

$$\begin{array}{lll} \displaystyle\max_{\{\mathbf{b}_k\}} & \log\det\left(\mathbf{I}_{n_R} + \sum_{k=1}^{K} \mathbf{R}_n^{-1}\mathbf{H}_k \mathbf{b}_k \mathbf{b}_k^H \mathbf{H}_k^H\right) & \\ \text{s.t.} & \mathbf{b}_k^H \mathbf{b}_k \leq P_k, & 1 \leq k \leq K. \end{array} \tag{4.17}$$

As in the rank-unconstrained case, we can try to solve this maximization problem using an iterative algorithm such that, at each iteration, the kth user's beamvector is updated as $\mathbf{b}_k = \sqrt{P_k}\,\mathbf{u}_{\max}(\mathbf{H}_k^H \mathbf{R}_{n,k}^{-1} \mathbf{H}_k)$ where $\mathbf{R}_{n,k} = \sum_{l \neq k} \mathbf{H}_l \mathbf{b}_l \mathbf{b}_l^H \mathbf{H}_l^H + \mathbf{R}_n$. This iterative algorithm, however, is not guaranteed to converge to a global optimal solution due to the nonconvexity of the problem (see [Ulu01] and references therein where this problem was considered in the CDMA context). For illustrative purposes, we give a very simple counter-example which shows that such an iterative approach does not converge.

Counter-Example[7]

Consider the following two-user channel

$$\mathbf{y} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}_1 + \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \mathbf{x}_2 + \mathbf{n}$$

where $\mathbf{R}_n = \mathbf{I}$ and the power constraint is given by $P_1 = P_2 = 10$. The optimal solution with the rank-one constraint is given using the best of the two subchannels for each user, *i.e.*, by $\mathbf{b}_1 = \left[\sqrt{10},\ 0\right]^T$ and $\mathbf{b}_2 = \left[0,\ \sqrt{10}\right]^T$ (in fact, this is the optimal solution even if the rank-one constraint is removed, since each user water-fills the other user treated as noise). However, if the starting point is $\mathbf{b}_1 = \left[0,\ \sqrt{10}\right]^T$ and $\mathbf{b}_2 = \left[\sqrt{10},\ 0\right]^T$, then the iterative algorithm cannot escape from such a suboptimal point and fails to converge to the optimum.

[7] To the author's knowledge, this counter-example is due to Prof. Zhi-Quan (Tom) Luo.

4.3 Ergodic and Outage Capacities

As we have seen, with perfect CSIT, it is possible to adapt $\mathbf{Q}$ to each channel state to achieve the instantaneous capacity. However, obtaining CSI at the transmitter requires either a feedback channel or the application of the channel reciprocity property when transmission and reception operate at the same carrier frequency and the time variation of the channel is sufficiently slow (c.f. §2.4). In many cases, the channel estimate at the transmitter may become significantly inaccurate, mainly due to a fast time-varying nature of the channel. In fact, many practical communication systems assume no CSI at the transmitter (see §2.4). For these situations, it becomes necessary to utilize transmission techniques (and in particular a transmit power allocation) independent of the current channel realization. The channel is therefore treated as a random quantity drawn according to some known probability density function (pdf) $p_{\mathbf{H}}(\mathbf{H})$. We now introduce the notions of ergodic and outage capacities for the single-user case (for the multiuser case, the same ideas apply).

The natural extension of the capacity (provided that there are no delay constraints in the communication system) when the channel fading state is a random quantity is given by the ergodic capacity obtained as the maximum mutual information averaged over all channel states [Tel95]. The ergodic capacity is a useful measure of the achievable bit rate when the transmission duration is so long as to reveal the long-term ergodic properties of the fading process, which is assumed to be an ergodic process in time [Tel95, Big98]. The mutual information for a given transmit covariance matrix $\mathbf{Q}$ is

$$\mathcal{I}^{\text{erg}}(\mathbf{Q}) = \mathbb{E}_{\mathbf{H}} \log\det\left(\mathbf{I}_{n_R} + \mathbf{R}_n^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right). \tag{4.18}$$

The ergodicity assumption, however, is not necessarily satisfied in practical communication systems operating on fading channels because no significant channel variability may occur during the whole transmission for applications with stringent delay constraints such as speech transmission. In these cases, the Shannon capacity may be zero [Big98], *i.e.*, there may be a non-negligible probability that the value of the actual transmitted rate (no matter how small) exceeds the instantaneous mutual information provided by the channel. It is then useful to associate an outage probability to any given rate, *i.e.*, the probability that the channel cannot support the rate [Oza94, Big98]. An appropriate measure is then the outage capacity defined as the rate that cannot be satisfied only with a small outage probability ϵ (also known as ϵ-achievable rate [Cai99b, Big01]) [Oza94, Big98, Tel95, Cai99b]. The mutual information with outage probability ϵ for a given transmit covariance matrix $\mathbf{Q}$ is

$$\mathcal{I}_{\epsilon}^{\text{out}}(\mathbf{Q}) = \sup_{R}\left\{R : \Pr\left\{\log\det\left(\mathbf{I}_{n_R} + \mathbf{R}_n^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) \leq R\right\} \leq \epsilon\right\}. \tag{4.19}$$

Since the pdf or the statistics of the channel are a priori known, an optimal fixed power allocation (independent of the actual channel realization) can be precomputed to maximize either

$\mathcal{I}^{\text{erg}}(\mathbf{Q})$ or $\mathcal{I}^{\text{out}}_{\epsilon}(\mathbf{Q})$ over the set of $\mathbf{Q}$ satisfying the power constraint to obtain the ergodic capacity $\mathcal{C}^{\text{erg}}$ or the outage capacity $\mathcal{C}^{\text{out}}_{\epsilon}$, respectively. The same ideas are easily extended to the multiuser case.

In the following, we will mainly treat the ergodic capacity since it is simpler to analyze than the outage capacity, because it results in a convex optimization problem unlike the outage capacity formulation (which will be briefly considered as well in the single-user case).

4.3.1 Capacity of the Single-User Channel

First of all, it is instructive to realize that the formulation of the ergodic capacity results in a convex optimization problem (by the concavity of the logdet function and using the fact that the expectation operator is linear):

$$\begin{aligned} \max_{\mathbf{Q}} \quad & \mathbb{E}_{\mathbf{H}} \log\det\left(\mathbf{I}_{n_R} + \mathbf{R}_n^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) \\ \text{s.t.} \quad & \operatorname{Tr}(\mathbf{Q}) \leq P_T, \\ & \mathbf{Q} = \mathbf{Q}^H \geq 0. \end{aligned} \tag{4.20}$$

In [Tel95], the ergodic capacity problem was solved for the particular case of a Gaussian-distributed channel with i.i.d. entries, obtaining the uniform power allocation $\mathbf{Q} = P_T/n_T\,\mathbf{I}_{n_T}$ as the optimal solution. The proof follows easily from the concavity of the logdet function and the invariance under rotation (isotropy property) of the Gaussian random channel $\mathbf{H} \sim \mathbf{HU}$ where $\mathbf{U}$ is a fixed unitary matrix [Tel95]. In fact, the optimality of the uniform power allocation holds not just for the Gaussian-distributed channel matrix $\mathbf{H}$ but for any channel matrix satisfying the isotropy property $\mathbf{H} \sim \mathbf{HU}$. The uniform power distribution has also been shown optimum for some particular cases of interest such as frequency-selective SISO channels [Oza94, Cai99a] and the dual case of flat time-varying SISO channels [Cai99b]. (In case that the channel matrix entries are correlated, it is possible to improve upon the uniform power allocation by using some statistical knowledge of the channel, *e.g.*, using a stochastic water-filling solution as proposed in [Shi98].) We now extend the results to the more general case of arbitrary channel distributions (not necessarily Gaussian or satisfying the isotropic condition $\mathbf{H} \sim \mathbf{HU}$) having i.i.d. entries with symmetric pdf.

Proposition 4.1 *The ergodic capacity of the vector Gaussian memoryless channel (4.1) where the channel state $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ is a random matrix with i.i.d. entries drawn from a symmetric pdf, $p_{h_{ij}}(\nu) = p_{h_{ij}}(-\nu)$, subject to a transmit power constraint P_T is (assuming instantaneous CSIR and statistical CSIT)*

$$\mathcal{C}^{erg} = \mathbb{E}_{\mathbf{H}} \log\det\left(\mathbf{I}_{n_R} + P_T/n_T\,\mathbf{R}_n^{-1}\mathbf{H}\mathbf{H}^H\right). \tag{4.21}$$

The capacity-achieving solution is given by a Gaussian code with a uniform power allocation $\mathbf{Q} = P_T/n_T\,\mathbf{I}_{n_T}$*, i.e., by* $\mathbf{x} \sim \mathcal{CN}\left(\mathbf{0}, P_T/n_T\,\mathbf{I}_{n_T}\right)$*.*

Proof. See Appendix 4.A. ■

The characterization of the outage capacity is a significantly more difficult problem since its formulation results in a nonconvex (combinatorial) optimization problem. In [Fos98], the capacity of a MIMO channel with no CSIT was obtained assuming a uniform power allocation over the transmit antennas. The choice of the uniform distribution was based on the symmetry of the problem, *i.e.*, the fact that the fading between each transmit-receive pair of antennas is identically distributed and uncorrelated with the fading between any other pair of antennas (spatially uncorrelated channel). In [Tel95], it was conjectured that, for the Gaussian-distributed channel $\mathbf{H}$ with i.i.d. entries, the optimal solution is a uniform power allocation only over a subset of the transmit dimensions. The problem can be formulated from two different points of view (although equivalent in essence): given an outage probability obtain the corresponding rate or given a desired rate obtain the corresponding outage probability.[8] Consider the latter formulation for a desired rate R:

$$\begin{array}{ll} \min\limits_{\mathbf{Q}} & \Pr\left\{\log\det\left(\mathbf{I}_{n_R} + \mathbf{R}_n^{-1}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) \leq R\right\} \\ \text{s.t.} & \operatorname{Tr}\left(\mathbf{Q}\right) \leq P_T, \\ & \mathbf{Q} = \mathbf{Q}^H \geq 0. \end{array} \tag{4.22}$$

Unfortunately, the problem is nonconvex due to the nonconvexity of the outage probability as a function of $\mathbf{Q}$. In fact, this is a well-known open problem. Even in the simple case of a Gaussian-distributed channel $\mathbf{H}$ with i.i.d. entries, the problem remains unsolved [Tel95]. We will now show how this problem can be relaxed to make it convex (solving the relaxed problem will yield upper and lower bounds on the optimal solution).

For practical purposes, we can approximate the set of channels by a finite set with an approximation error as small as desired. In other words, the set of channels can be quantized with the sufficient number of points. Denoting the finite set of N channels by $\{\mathbf{H}_i\}$ and the probability of the ith channel $\mathbf{H}_i$ by p_i, the problem can be rewritten as

$$\begin{array}{lll} \min\limits_{\mathbf{Q},\{t_i\}} & \sum_{i=1}^{N} p_i\, t_i & \\ \text{s.t.} & \log\det\left(\mathbf{I}_{n_R} + \mathbf{R}_n^{-1}\mathbf{H}_i\mathbf{Q}\mathbf{H}_i^H\right) \geq R\left(1 - t_i\right) & 1 \leq i \leq N, \\ & t_i \in \{0,1\}, & \\ & \operatorname{Tr}\left(\mathbf{Q}\right) \leq P_T, & \\ & \mathbf{Q} = \mathbf{Q}^H \geq 0, & \end{array} \tag{4.23}$$

where the variables t_i's are indicators of whether a channel $\mathbf{H}_i$ generates an outage event ($t_i = 1$) or not ($t_i = 0$) for a given $\mathbf{Q}$. The problem is not convex due to the constraint $t_i \in \{0,1\}$ which

[8] In fact, the full characterization of the outage capacity is the whole curve of outage probability vs. rate.

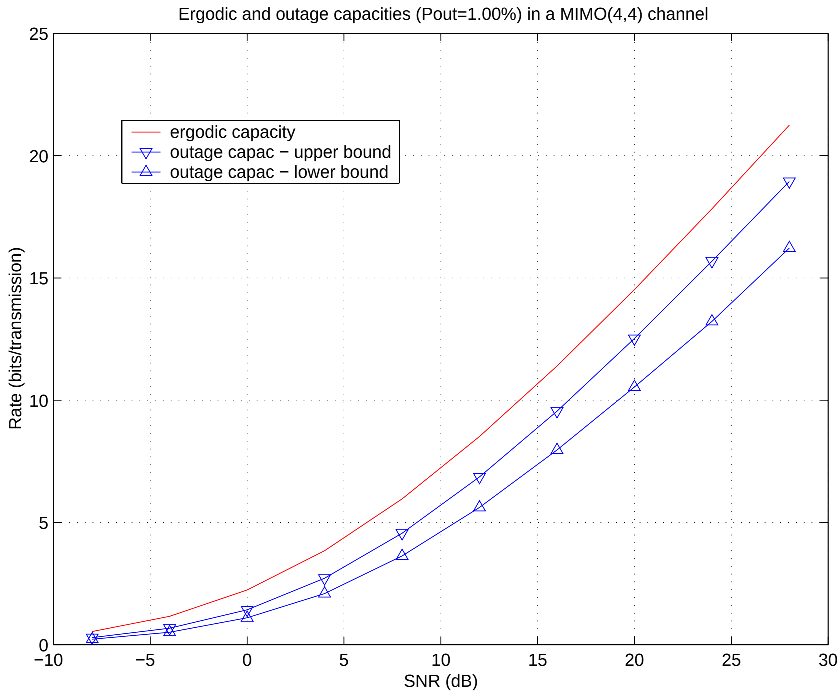


Figure 4.5: Lower and upper bounds of the outage capacity (at an outage probability of 1%) along with the ergodic capacity.

makes the problem a combinatorial one. Note that if we fix the set of channels corresponding to an outage event, *i.e.*, if we fix the set $\{t_i\}$, then the problem is convex. It is possible, however, to convexify problem (4.23) by relaxing the "hard" constraints $t_i \in \{0,1\}$ to $t_i \in [0,1]$ (*i.e.*, $0 \leq t_i \leq 1$):

$$\begin{array}{lll} \min\limits_{\mathbf{Q}} & \sum_{i=1}^{N} p_i \, t_i & \\ \text{s.t.} & \log\det\left(\mathbf{I}_{n_R} + \mathbf{R}_n^{-1}\mathbf{H}_i\mathbf{Q}\mathbf{H}_i^H\right) \geq R\,(1-t_i) & 1 \leq i \leq N, \\ & 0 \leq t_i \leq 1, & \\ & \mathrm{Tr}\,(\mathbf{Q}) \leq P_T, & \\ & \mathbf{Q} = \mathbf{Q}^H \geq 0. & \end{array} \tag{4.24}$$

The relaxed problem allows to obtain upper and lower bounds on the outage probability (equivalently, on the outage capacity) and can be solved in practice since it is a convex problem (c.f. §3.1). The minimum objective value of the relaxed problem (4.24) gives a lower bound on the outage probability of the original (discretized) problem (4.23) and the upper bound can be readily obtained by plugging the optimum transmit covariance matrix of the relaxed problem (4.24) into (4.23) and solving for the t_i's.

As a final comment, it is important to remark that the tradeoff between rate and outage probability is closely related to the tradeoff between multiplexing and diversity gains which has been recently characterized in [Zhe03].[9] In terms of diversity, the uniform power allocation has been shown to be asymptotically optimal for a sufficiently high SNR [Zhe03].

Numerical Example

For illustration purposes, we plot in Figure 4.5 the upper and lower bounds obtained from solving numerically the relaxed problem (4.24) using an interior-point method (see §3.1) for the specific case of a real Gaussian 4×4 MIMO channel matrix with i.i.d. unit-variance entries.

4.3.2 Capacity Region of the Multiple-Access Channel (MAC)

The ergodic capacity region of the multiple-access channel has a simple characterization similar to that in (4.13)-(4.14). The achievable rate region corresponding to a set of transmit covariance matrices $\{\mathbf{Q}_k\}$ (assuming that a Gaussian signaling is used) is [Tel01, Rhe01a]

$$\mathcal{R}^{\text{erg}}\left(\{\mathbf{Q}_k\}\right) = \Bigg\{ (R_1, \cdots, R_K) : \\ 0 \leq \sum_{k \in \mathcal{S}} R_k \leq \mathbb{E}_{\mathbf{H}} \log\det\left(\mathbf{I}_{n_R} + \sum_{k \in \mathcal{S}} \mathbf{R}_n^{-1}\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H\right), \forall \mathcal{S} \subseteq \{1, \cdots, K\} \Bigg\}. \tag{4.25}$$

The capacity region when the transmit covariance matrices are constrained in their average transmit power is [Tel01, Rhe01a]

$$\mathcal{C}^{\text{erg}} = \bigcup_{\substack{\text{Tr}(\mathbf{Q}_k) \leq P_k, \\ \mathbf{Q}_k = \mathbf{Q}_k^H \geq 0}} \mathcal{R}^{\text{erg}}\left(\{\mathbf{Q}_k\}\right).$$

Note that the concept of ergodic capacity when the transmitter has instantaneous CSI can also be considered [Tse98].

For the specific case of Gaussian-distributed channel matrices with i.i.d. entries, the uniform power allocation for each of the users, $\mathbf{Q}_k = P_k/n_k \, \mathbf{I}_{n_k} \ \forall k$, was proved to be optimal in terms of ergodic capacity in [Tel01, Rhe01a] (the proof is the natural extension of that of the single-user case given in [Tel95] based on the concavity of the logdet function applied to each of the constraints defining the rate region). The ergodic capacity region is then

$$\mathcal{C}^{\text{erg}} = \Bigg\{ (R_1, \cdots, R_K) : \\ 0 \leq \sum_{k \in \mathcal{S}} R_k \leq \mathbb{E}_{\mathbf{H}} \log\det\left(\mathbf{I}_{n_R} + \sum_{k \in \mathcal{S}} P_k/n_k \, \mathbf{R}_n^{-1}\mathbf{H}_k\mathbf{H}_k^H\right), \forall \mathcal{S} \subseteq \{1, \cdots, K\} \Bigg\}. \tag{4.26}$$

[9]As defined in [Zhe03], the diversity gain is basically the exponent of the outage probability or, in other words, the slope of the curve of outage probability vs. SNR for high SNR in a log-log scale. The multiplexing gain is defined in [Zhe03] as the slope of the curve of rate vs. SNR in dB for high SNR.

This result also holds for more general channel distributions (not necessarily Gaussian or satisfying the isotropic condition $\mathbf{H}_k \sim \mathbf{H}_k \mathbf{U}_k$) that have i.i.d. entries with symmetric pdf as obtained in Proposition 4.1 for the single-user case (the proof for the MAC is straightforward using the result of Proposition 4.1 and is therefore not included).

4.4 Worst-Case Capacity: A Game-Theoretic Approach

This section considers the case in which not even the channel statistics are known at the transmitter, obtaining therefore a robust power allocation under channel uncertainty. We formulate the problem within a game-theoretic framework [Osb94, Fud92], in which the payoff function of the game is the mutual information and the players are the transmitter and a malicious nature. The formulation of the problem as a game allows a better characterization of the problem (different types of games are considered such as a strategic game, a Stackelberg game, and a mixed-strategy strategic game). Mathematically, this is formulated as a maximin problem that is known to lead to robust solutions [Kas85]. Well-known examples of robust maximin and minimax formulations are universal source coding and universal portfolio [Cov91b, Cov91a]. The problem turns out to be the characterization of the capacity of a compound vector Gaussian channel [Lap98, Wol78]. Interestingly, the uniform power allocation is obtained as the solution of the game in terms of capacity (under the mild condition that the set of channels is isotropically unconstrained, meaning that the transmission "directions" are unconstrained). Note that well-known communication schemes, such as space-time codes and layered architectures (*e.g.*, BLAST) (c.f. §2.4), use indeed a uniform power allocation. The results are easily extended to ergodic and outage capacities. The loss in terms of capacity of the robust power allocation with respect to the optimal one (adapted to the specific channel realization) is analyzed using the concept of duality gap arising in convex optimization theory [Lue69, Boy00]. The robustness of the uniform power allocation from a maximin viewpoint also holds for the more interesting and general case of a multiple-access channel. In particular, the worst-case rate region corresponding to the uniform power distribution is shown to contain the worst-case rate region of any other possible power allocation strategy. In other words, the capacity region of the compound vector MAC is achieved when each of the users is using a uniform power allocation.

Channel Model

In wireless communications, the channel may undergo slow and/or fast fading due to shadowing and Doppler effects. Essentially, matrix $\mathbf{H}$ is not fixed and changes in time. One possible way to deal with this is by considering the channel as a random variable with a known pdf $p_{\mathbf{H}}(\mathbf{H})$ which naturally leads to the notions of ergodic capacity and outage capacity as considered in §4.3. In this section, we are interested in a robust design obtained by including uncertainty about the channel at both the transmitter and the receiver. There is a significant variety of channel models that can be used to model channel uncertainty (see [Lap98] for a great overview of reliable communication under channel uncertainty). If the fading is sufficiently slow (the channel coherence time is much higher than the duration of a transmission), the system can be modeled as a compound channel, where the channel state remains unchanged during the course of a transmission and it is assumed to belong to a set of possible channel states but otherwise unknown [Wol78, Csi81, Lap98, Big98] (the capacity of the compound vector Gaussian channel was obtained in [Roo68]). For fast fading, however, the compound channel is no longer appropriate and other models such as a compound finite-state channel (FSC) [Lap98] or an arbitrarily varying channel (AVC) [Wol78, Csi81, Lap98] may be necessary. In the AVC, the channel state can arbitrarily change from symbol to symbol during the course of a transmission (see [Hug88] for results on the vector Gaussian AVC). Recall that, in situations where the unknown channel remains unchanged over multiple transmissions, the utilization of a training sequence to estimate the channel at the receiver is particularly attractive. The reader is referred to [Lap98] for a detailed discussion on the applicability of each model.

We consider that the fading is slow enough so that the compound channel model is valid (see [Roo68], for example, where the compound channel was used to model a wireless MIMO system). In other words, we assume that the transmission duration is sufficiently long so that the information-theoretic coding arguments are valid and sufficiently short so that the channel remains effectively unchanged during a transmission (c.f. [Fos98, Big01]). This type of channel is usually referred to as block-fading channel [Big98, Big01].

4.4.1 Capacity of the Single-User Channel

In this subsection, the problem of obtaining a robust transmit power allocation when the transmitter does not even know the channel statistics is formulated within the framework of game theory [Osb94, Fud92]. The idea of robustness implies being able to function in all possible scenarios and, in particular, the worst-case scenario. This concept fits naturally into the context of game theory.

4.4.1.1 Game-Theoretic Formulation

We will consider a game in which the payoff function (by which the result of the game is measured) is the mutual information and the players are: the transmitter that selects the best signaling scheme $p(\mathbf{x})$ and a malicious nature that chooses the worst communication conditions or channel transition probability $p(\mathbf{y} \mid \mathbf{x})$. It is interesting to note that the formulation of the communication process explicitly as a game was first proposed more than 40 years ago by Blachman [Bla57] using a mutual information payoff. We constrain our search to Gaussian-distributed signal and noise since it is well known that they constitute a robust solution (a saddle point) to a mutual

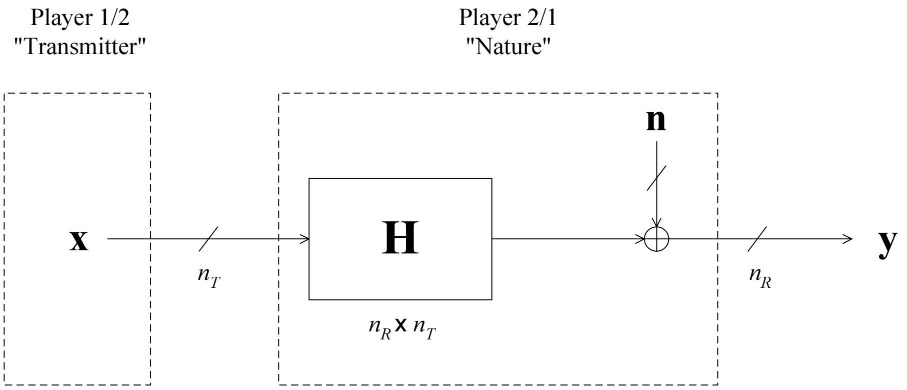


Figure 4.6: Communication interpreted as a two-player game.

information game for the memoryless vector channel [Bor85, Dig01].[10] In this case, $p(\mathbf{y} \mid \mathbf{x})$ is a vector Gaussian distribution parameterized with the channel state $(\mathbf{H}, \mathbf{R}_n)$. In the sequel, by "channel" we will simply refer to the whitened channel state $\tilde{\mathbf{H}} = \mathbf{R}_n^{-1/2}\mathbf{H}$ and not to the channel transition probability $p(\mathbf{y} \mid \mathbf{x})$. The two-player game is illustrated in Figure 4.6.

With the previous considerations, the unknowns of the game are the transmit covariance matrix $\mathbf{Q}$ and the whitened channel $\tilde{\mathbf{H}}$ (which implicitly includes the noise covariance matrix $\mathbf{R}_n$ and the original channel $\mathbf{H}$). The payoff function of the game is then the mutual information given by $\Psi(\mathbf{Q}, \tilde{\mathbf{H}})$ in (4.10). The game would be meaningless and trivial unless we placed restrictions on the players. Therefore, we suppose that the channel $\tilde{\mathbf{H}}$ must belong to a set of possible channels $\tilde{\mathcal{H}}$ and, similarly, $\mathbf{Q}$ must belong to a set of possible covariance matrices $\mathcal{Q}$. It is important to bear in mind that, for simplicity of notation, we write $\tilde{\mathbf{H}} \in \tilde{\mathcal{H}}$ instead of $(\mathbf{H}, \mathbf{R}_n) \in \mathcal{H} \times \mathcal{R}_n$ with no loss of generality (one can always define $\tilde{\mathcal{H}}$ as the set of matrices $\tilde{\mathbf{H}}$ that can be parameterized as $\tilde{\mathbf{H}} = \mathbf{R}_n^{-1/2}\mathbf{H}$ for some $(\mathbf{H}, \mathbf{R}_n) \in \mathcal{H} \times \mathcal{R}_n$). The set $\mathcal{Q}$ considered in this section is defined by the average transmit power constraint

$$\mathcal{Q} \triangleq \left\{ \mathbf{Q} : \operatorname{Tr}(\mathbf{Q}) \leq P_T,\ \mathbf{Q} = \mathbf{Q}^H \geq \mathbf{0} \right\}. \tag{4.27}$$

We remark that the results of this section still hold if the eigenvalue constraint $\lambda_{\max}(\mathbf{Q}) \leq \alpha$ is utilized instead to define $\mathcal{Q}$. Regarding the set $\tilde{\mathcal{H}}$, since we are interested in finding a robust $\mathbf{Q}$ for all possible channels, we would like not to impose any constraint on the allowable set of channels. However, this would be a poor choice because the trivial solution $\tilde{\mathbf{H}} = \mathbf{0}$ would be obtained. To avoid this effect, we are forced to introduce some artificial constraints (unlike the constraint used to define $\mathcal{Q}$ which is very natural). But this may have the side effect that the solution to the game formulation may depend on the particular constraints chosen. Fortunately, as proved in §4.4.1.2, the solution to the game formulation is independent of the particular channel constraints under the mild condition that the constraints guarantee an isotropy property in $\tilde{\mathcal{H}}$ or $\mathcal{H}$ (c.f. §4.4.1.2).

[10]For complex-valued signals, the saddle-point property holds for proper complex Gaussian distributions [Nee93].

As has been previously argued, to take into account the effect of channel uncertainty, we consider that the channel is known to belong to a set of possible channels $\tilde{\mathcal{H}}$ but otherwise unknown. The worst-case channel for a given $\mathbf{Q}$ is given by the minimizing solution to $\inf_{\tilde{\mathbf{H}}\in\tilde{\mathcal{H}}} \Psi(\mathbf{Q}, \tilde{\mathbf{H}})$. The transmitter will maximize the worst-case mutual information over the set $\mathcal{Q}$, yielding the following maximin formulation of the problem:[11]

$$\sup_{\mathbf{Q}\in\mathcal{Q}} \inf_{\tilde{\mathbf{H}}\in\tilde{\mathcal{H}}} \Psi(\mathbf{Q}, \tilde{\mathbf{H}}). \tag{4.28}$$

At this point, it is interesting to recall that a compound channel is precisely a channel that is known to belong to a set of possible channels (unchanged during the course of a transmission) but otherwise unknown [Wol78, Csi81, Lap98]. As discussed at the beginning of this section, this type of channel may be useful to model communication under channel uncertainty for sufficiently slow fading. The capacity of the compound channel (the capacity that can be guaranteed for the set of possible channels $\tilde{\mathcal{H}}$) was extensively treated in [Wol78] where an expression similar to (4.28) was shown to be the capacity of the compound discrete memoryless channel. In [Roo68], the vector Gaussian channel was specifically considered and (4.28) was indeed shown to be the capacity of the compound vector Gaussian channel when the actual channel state is unknown at both the transmitter and the receiver (under the mild assumption that $\tilde{\mathcal{H}}$ is bounded). Note that knowledge of the channel state at the receiver does not increase the compound channel capacity [Wol78],[3] although the receiver may be simpler to implement with this knowledge. Clearly, the capacity of the compound channel cannot exceed the capacity of any channel in the family. In principle, it may not even be equal to the infimum of the capacities of the individual channels in the family (this is because codes and their decoding sets must be found, not just to give small error probability in the worst channel, but uniformly across the class of channels, which is a more stringent condition) [Roo68, Lap98].

Alternatively, we can consider the compound channel when the transmitter knows the channel state (as in the previous case, it is indifferent whether the receiver knows the channel state or not [Wol78]). In this case, in principle, a different coding-decoding strategy can be used for each channel realization and the capacity of such a compound channel is given by the following minimax formulation:

$$\inf_{\tilde{\mathbf{H}}\in\tilde{\mathcal{H}}} \sup_{\mathbf{Q}\in\mathcal{Q}} \Psi(\mathbf{Q}, \tilde{\mathbf{H}}) \tag{4.29}$$

i.e., the infimum of the capacities of the family of channels $\tilde{\mathcal{H}}$.

From a game-theoretic perspective, the problem can be viewed as a two-player zero-sum (players with diametrically opposed preferences) game, also known as strictly competitive game (the transmitter is the maximizing player and nature is the minimizing player) [Osb94] (see Figure

[11]For the particular sets $\mathcal{Q}$ and $\mathcal{H}$ considered in this chapter, the formulation sup-inf reduces to max-min. For the sake of generality, however, we stick to the sup-inf notation throughout the chapter.

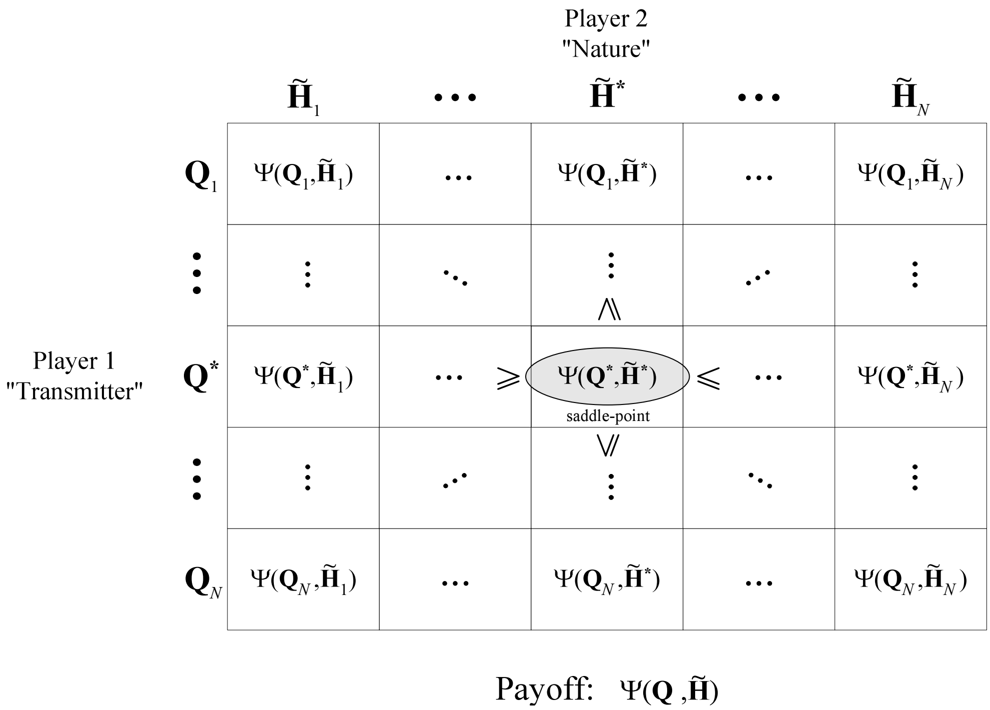


Figure 4.7: Two-player zero-sum strategic game in which player 1 (the transmitter) and player 2 (nature) move simultaneously. The optimal power allocation is found as a saddle-point (Nash equilibrium). (Note that for illustration purposes the sets $\mathcal{Q}$ and $\widetilde{\mathcal{H}}$ have been considered finite.)

4.6). In the following, we cast the problem in three different types of games: a strategic game both with pure strategies and with mixed strategies and a Stackelberg game.

Strategic Game with Pure Strategies

The simplest formulation (from a game-theoretic standpoint) is that of a strategic game, in which the players select their strategies without knowing the other players' choices, *i.e.*, they "move" simultaneously (see Figure 4.7). In such cases, there may exist a set of equilibrium points called Nash equilibria characterized for being robust or locally optimal in the sense that no player wants to deviate from such points. In our case (a two-player zero-sum game), a Nash equilibrium is also termed saddle point $\left(\mathbf{Q}^{\star}, \widetilde{\mathbf{H}}^{\star}\right)$ and it is a simultaneously optimal point for both players (see Figure 4.7):

$$\Psi\left(\mathbf{Q}, \widetilde{\mathbf{H}}^{\star}\right) \leq \Psi\left(\mathbf{Q}^{\star}, \widetilde{\mathbf{H}}^{\star}\right) \leq \Psi\left(\mathbf{Q}^{\star}, \widetilde{\mathbf{H}}\right) \tag{4.30}$$

where $\Psi\left(\mathbf{Q}^{\star}, \widetilde{\mathbf{H}}^{\star}\right)$ is called the value of the game (whenever it exists) and is equal to the maximin and minimax solutions of (4.28) and (4.29) [Osb94], *i.e.*,

$$\Psi\left(\mathbf{Q}^{\star}, \widetilde{\mathbf{H}}^{\star}\right) = \sup_{\mathbf{Q}\in\mathcal{Q}} \inf_{\widetilde{\mathbf{H}}\in\widetilde{\mathcal{H}}} \Psi\left(\mathbf{Q}, \widetilde{\mathbf{H}}\right) = \inf_{\widetilde{\mathbf{H}}\in\widetilde{\mathcal{H}}} \sup_{\mathbf{Q}\in\mathcal{Q}} \Psi\left(\mathbf{Q}, \widetilde{\mathbf{H}}\right). \tag{4.31}$$

Note that one of the major techniques for designing systems that are robust with respect to modeling uncertainties is the minimax approach, in which the goal is the optimization of the worst-case performance [Ver84, Kas85]. Interesting examples of minimax design in information theory are the problem of source coding or data compression when the data distribution is completely unknown and the problem of portfolio investment when nothing is known about the stock market [Cov91b, Cov91a]. Both problems can be formulated as a game in which two players compete: the source encoding scheme vs. the data distribution and the portfolio investor vs. the market.

Strategic Game with Mixed Strategies

The function $\Psi\big(\mathbf{Q}, \tilde{\mathbf{H}}\big)$ may or may not have any saddle point depending on the particular set $\tilde{\mathcal{H}}$ (c.f. §4.4.1.2). However, so far we have only considered pure strategies, *i.e.*, strategies given by a single fixed (deterministic) pair $\big(\mathbf{Q}, \tilde{\mathbf{H}}\big)$. The game can be extended to include mixed strategies, *i.e.*, the possibility of choosing a randomization over a set of pure strategies (the randomizations of the different players is independent) [Osb94]. In this case, the payoff is the average of $\Psi\big(\mathbf{Q}, \tilde{\mathbf{H}}\big)$ over the mixed strategies, $\mathbb{E}_{p_{\mathbf{Q}} p_{\tilde{\mathbf{H}}}} \Psi\big(\mathbf{Q}, \tilde{\mathbf{H}}\big)$, and the saddle point is similarly defined as

$$\mathbb{E}_{p_{\mathbf{Q}} p^{\star}_{\tilde{\mathbf{H}}}} \Psi\big(\mathbf{Q}, \tilde{\mathbf{H}}\big) \leq \mathbb{E}_{p^{\star}_{\mathbf{Q}} p^{\star}_{\tilde{\mathbf{H}}}} \Psi\big(\mathbf{Q}, \tilde{\mathbf{H}}\big) \leq \mathbb{E}_{p^{\star}_{\mathbf{Q}} p_{\tilde{\mathbf{H}}}} \Psi\big(\mathbf{Q}, \tilde{\mathbf{H}}\big). \tag{4.32}$$

It is well known that a strategic game always has a mixed strategy Nash equilibrium under the assumption that each set of pure strategies is closed, bounded, and convex [Osb94]. In fact, for our specific problem, even if we allow more general sets (which need not be closed, bounded, and convex) such as the set $\tilde{\mathcal{H}}$ defined by $\lambda_{\max}\big(\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}\big) \geq \beta$ (which is nonconvex and unbounded), it can be shown that the problem always has an infinite set of Nash equilibria (c.f. §4.4.1.2). One can interpret mixed strategies in different ways. In this problem, perhaps, the most relevant interpretation is to consider the mixed strategy Nash equilibrium as a steady state of an environment in which players act repeatedly, learning other players' mixed strategies (see [Osb94] for other interpretations).

Stackelberg Game

Alternatively, instead of modeling our problem as a strategic game (which, in general, does not have a pure strategy Nash equilibrium), we can formulate it in a more general way as an extensive game[12] in which the selected strategy of a user may depend on the previously selected strategy of another user[13] (as opposed to the previous strategic interpretation in which both players move simultaneously) [Osb94, Fud92]. For the specific case of a two-player zero-sum game, in the parlance of game theory, such an extensive game is called Stackelberg game [Osb94, Fud92]. Consider the case in which the transmitter moves first and then nature moves aware of the transmitter's move (see Figure 4.8). In such a case, the maximin solution of (4.28) is always a pure strategy Nash equilibrium. In fact, such a solution is a subgame perfect equilibrium (called in this case Stackelberg equilibrium) which is a more refined definition of equilibrium[14]

[12] An extensive game is an explicit description of the sequential structure of the decision problems encountered by the players in a strategic situation [Osb94].

[13] By extensive game we always refer to those with perfect information (imperfect information can also be considered) [Osb94].

[14] The solution concept of Nash equilibrium is unsatisfactory in extensive games since it ignores the sequential

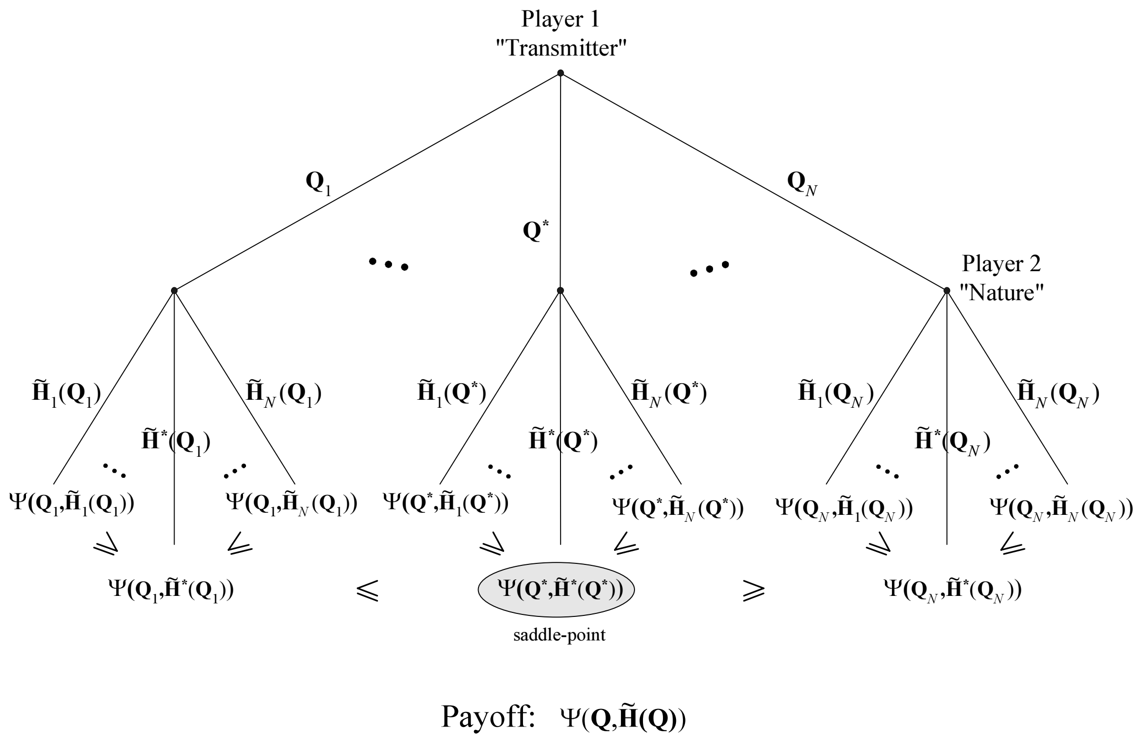


Figure 4.8: Two-player zero-sum extensive game in which player 1 (the transmitter) moves first and then player 2 (nature) moves aware of player 1's move, *i.e.*, Stackelberg game. The optimal power allocation is found as a saddle-point (subgame perfect equilibrium which is also a Nash equilibrium). (Note that for illustration purposes the sets $\mathcal{Q}$ and $\tilde{\mathcal{H}}$ have been considered finite.)

[Osb94, Fud92]. In this case, a saddle point $\left(\mathbf{Q}^\star, \tilde{\mathbf{H}}^\star\left(\mathbf{Q}^\star\right)\right)$ is characterized by

$$\Psi\left(\mathbf{Q}, \tilde{\mathbf{H}}^\star\left(\mathbf{Q}\right)\right) \leq \Psi\left(\mathbf{Q}^\star, \tilde{\mathbf{H}}^\star\left(\mathbf{Q}^\star\right)\right) \leq \Psi\left(\mathbf{Q}^\star, \tilde{\mathbf{H}}\left(\mathbf{Q}^\star\right)\right). \tag{4.33}$$

Similarly, we can also consider the opposite formulation of the Stackelberg game in which nature moves first and then the transmitter moves aware of nature's move with saddle point given by

$$\Psi\left(\mathbf{Q}(\tilde{\mathbf{H}}^\star), \tilde{\mathbf{H}}^\star\right) \leq \Psi\left(\mathbf{Q}^\star(\tilde{\mathbf{H}}^\star), \tilde{\mathbf{H}}^\star\right) \leq \Psi\left(\mathbf{Q}^\star(\tilde{\mathbf{H}}), \tilde{\mathbf{H}}\right). \tag{4.34}$$

Note that the saddle points of (4.33) and (4.34) are always satisfied by the solutions to problems (4.28) and (4.29), respectively.

Existing Results

A significant part of the literature that has modeled communication as a game has dealt with the characterization of saddle points satisfying (4.30), *i.e.*, implicitly adopting a formulation of the problem as a strategic game. Reference [Bla57] is one of the earliest papers dealing with such a problem using a mutual information payoff. (Note that other payoff functions have also been considered, such as the mean square error in [Bas83] to deal with communication over a channel

structure of the decision problem; as a consequence, more refined definitions of equilibrium have been proposed [Osb94].

with an intelligent jammer.) A two-player zero-sum game was explicitly adopted in [Bor85] obtaining the Gaussian distribution as a saddle point. In [Sta88], m-dimensional strategies were considered in a game-theoretic formulation of communication over channels with block memory, where it was found that memoryless jamming and memoryless coding constitute a saddle point. In [Yan93], a two-player zero-sum game was explicitly formulated for communication in the presence of jamming using a power constraint for both players. In [Dig01], communication under the worst additive noise under a covariance constraint was analyzed (the Gaussian distribution was obtained as a saddle-point solution) with emphasis on covariances satisfying correlation constraints at different lags. The vector Gaussian AVC was considered in [Hug88] obtaining a saddle point given by a water-filling solution for the jammer and for the coder. In [Chi01], the maximin and minimax problems of (4.28) and (4.29) in a multi-antenna wireless scenario were solved for a specific set of channels $\tilde{\mathcal{H}}$ defined by $\mathrm{Tr}\left(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right) \geq \beta$, *i.e.*, the two Stackelberg games previously formulated were implicitly considered.

The rest of the section focuses mainly on finding a robust power allocation when the channel is unknown, *i.e.*, in solving the maximin problem of (4.28). Such a solution has many interpretations. Under some conditions (obtained in §4.4.1.2), it constitutes a saddle point of the strategic game formulation of (4.30) with the inherent properties of robustness. In any case, if mixed strategies are allowed in the strategic game, the solution to (4.28) always forms a saddle point defined by (4.32) (c.f. §4.4.1.2). Finally, even if we restrict the game to pure strategies, the solution to (4.28) always constitutes a saddle point as defined in (4.33) corresponding to a Stackelberg game. (The opposite minimax problem formulation of (4.29) is briefly considered in §4.4.1.2.3 as well.)

4.4.1.2 Worst-Case Capacity and Robust Power Allocation

The main purpose of this subsection is to solve the maximin formulation of (4.28) and to characterize the conditions under which the solution forms a saddle point in the strategic formulation of the game (with pure strategies and mixed strategies).

As pointed out in §4.4.1.1, we have to define some artificial constraint on the channel to avoid the trivial solution. Noting from (4.10)-(4.11) that the payoff function $\Psi\big(\mathbf{Q},\tilde{\mathbf{H}}\big)$ depends on $\tilde{\mathbf{H}}$ through $\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}$ (the left singular vectors of $\tilde{\mathbf{H}}$ are irrelevant), it is convenient to define $\tilde{\mathcal{H}}$ as

$$\tilde{\mathcal{H}} \triangleq \left\{\tilde{\mathbf{H}} : \mathbf{R}_H \triangleq \tilde{\mathbf{H}}^H\tilde{\mathbf{H}} = \mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H} \in \mathcal{R}_H\right\}. \tag{4.35}$$

To define the set $\mathcal{R}_H$ we consider any kind of spectral (eigenvalue) constraint given by

$$\mathcal{R}_H \triangleq \{\mathbf{R}_H\colon \{\lambda_i\left(\mathbf{R}_H\right)\} \in \mathcal{L}_{R_H}\} \tag{4.36}$$

where $\mathcal{L}_{R_H}$ denotes an arbitrary eigenvalue constraint (in §4.4.1.2.1 some specific eigenvalue constraints are considered). (Clearly, the set $\mathcal{L}_{R_H}$ cannot contain the all-zero vector that would correspond to $\tilde{\mathbf{H}} = \mathbf{0}$.) In defining the set $\mathcal{R}_H$ as in (4.36), we are deliberately leaving the eigenvectors of $\mathbf{R}_H$ (equivalently, the right singular vectors of $\tilde{\mathbf{H}}$) totally unconstrained. This is so that no preference is given to any signaling direction,[15] *i.e.*, to guarantee the isotropy of $\mathcal{R}_H$ (any direction is possible).

Definition 4.1 *A set of matrices $\mathcal{H}$ is isotropically unconstrained if the right singular vectors of the elements of the set are unconstrained, i.e., if for each $\mathbf{H} \in \mathcal{H}$ then $\mathbf{HU} \in \mathcal{H}$ for any unitary matrix $\mathbf{U}$.*

Clearly, the set $\tilde{\mathcal{H}}$ (defined according to (4.35) and (4.36)) is isotropically unconstrained. We remark that the results in this subsection are valid regardless of the particular eigenvalue constraint chosen to define the set $\mathcal{L}_{R_H}$.

We now obtain the uniform power allocation as the maximin solution of (4.28), *i.e.*, as the capacity-achieving solution of the compound vector Gaussian channel. Note that this could be proved in a shorter way by contradiction, *i.e.*, by showing that, for any given power allocation, we can always find some channel that yields a lower capacity than the minimum capacity corresponding to the uniform power allocation (indeed, this is the technique used in §4.4.2 for the multiple-access channel). Nevertheless, we obtain a more complete proof by characterizing the "shape" of the worst channel for any given power allocation and then we give some examples in order to gain insight into the problem.

Before proceeding to the main result, recall that the capacity of the compound vector Gaussian memoryless channel when the channel state is unknown was obtained in [Roo68][16] as

$$\mathcal{C}(\tilde{\mathcal{H}}) = \sup_{\mathbf{Q}\in\mathcal{Q}} \inf_{\tilde{\mathbf{H}}\in\tilde{\mathcal{H}}} \log\det\left(\mathbf{I}_{n_R} + \tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^H\right) \tag{4.37}$$

under the mild assumption that $\tilde{\mathcal{H}}$ is bounded (if not, we can simply bound $\tilde{\mathcal{H}}$ by adding the constraint $\lambda_{\max}\left(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right) \leq c$ for a sufficiently large value of c, which can be done without loss of generality based on physical interpretations of the channel $\tilde{\mathbf{H}}$). The achievability was proved in [Roo68] by showing the existence of a code (along with the decoding sets). Therefore, in theory, one can always find a code to achieve rates arbitrarily close to capacity and then use a universal decoder that decodes the received word according to the decoding set it belongs to (note that no knowledge of the channel state is required).

Theorem 4.1 *The capacity of the compound vector Gaussian memoryless channel with power constraint P_T, n_T transmit dimensions, and n_R receive dimensions (with no CSI) is*

$$\mathcal{C}(\tilde{\mathcal{H}}) = \inf_{\tilde{\mathbf{H}}\in\tilde{\mathcal{H}}} \log\det\left(\mathbf{I}_{n_R} + P_T/n_T\,\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right) \tag{4.38}$$

[15] For a flat multi-antenna system, the term "direction" means literally spatial direction.

[16] The extension to the complex-valued case is straightforward using the results of [Nee93].

where the class of channels $\tilde{\mathcal{H}}$ is an isotropically unconstrained set defined by (4.35)-(4.36) (unconstrained right singular vectors). The capacity-achieving solution of (4.38) is given by a Gaussian code with a uniform power allocation

$$\mathbf{Q}^{\star} = P_T/n_T \, \mathbf{I}_{n_T} \tag{4.39}$$

which implies an independent signaling over the transmit dimensions.

Proof. Intuitively speaking, due to the symmetry of the problem, if the transmitter does not use a uniform power distribution, the channel will do an "inverse water-filling", *i.e.*, it will redistribute its singular values so that the highest ones align with the lowest eigenvalues of $\mathbf{Q}$ (see Lemma 4.3 in Appendix 4.B). Therefore, maximizing the lowest eigenvalues of $\mathbf{Q}$ seems to be appropriate to avoid such a behavior. Indeed, this is achieved by the uniform power allocation.

See Appendix 4.B for a formal proof. ■

Note that the worst-case capacity expression (4.38) obtained in Theorem 4.1 can be simplified as

$$\inf_{\tilde{\mathbf{H}} \in \tilde{\mathcal{H}}} \log\det\left(\mathbf{I}_{n_R} + P_T/n_T \, \tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\right) = \inf_{\{\lambda_{R_H,i}\} \in \mathcal{L}_{R_H}} \sum_{i=1}^{n_T} \log\left(1 + P_T/n_T \, \lambda_{R_H,i}\right). \tag{4.40}$$

Theorem 4.1 is basically saying that when the channel state is unknown but known to belong to a set of possible channels $\tilde{\mathcal{H}}$, the optimum solution in the sense of providing the best worst-case performance is given by the uniform power allocation of (4.39). In other words, it is the solution to the problem formulation as a Stackelberg game in which the transmitter moves first as depicted in Figure 4.8. Note that if we had used instead the eigenvalue constraint $\lambda_{Q,\max} \leq \alpha$ to define the set $\mathcal{Q}$, it would have immediately followed $\lambda^{\star}_{Q,i} = \alpha \; \forall i$, *i.e.*, a uniform solution as well.

The uniform power allocation and the corresponding minimizing channel always constitute a saddle point of the Stackelberg game as defined in (4.33). Depending on the specific definition of the set of channels $\tilde{\mathcal{H}}$, they will also form a saddle point of the strategic game as given in (4.30). The following corollary gives the exact conditions.

Corollary 4.1 *The uniform power allocation $\mathbf{Q}^{\star} = P_T/n_T \, \mathbf{I}_{n_T}$ obtained in Theorem 4.1 and the corresponding minimizing channel form a saddle point of the strategic game given by (4.30) if and only if the minimizing channel satisfies $\lambda^{\star}_{R_H,i} = \beta \; \forall i$ (in particular this implies $n_R \geq n_T$).*

Proof. Since the right inequality of (4.30) is satisfied by any solution to (4.28), it suffices to find the conditions under which the left inequality is satisfied, *i.e.*, $\Psi\left(\mathbf{Q}, \tilde{\mathbf{H}}^{\star}\right) \leq \Psi\left(P_T/n_T \, \mathbf{I}, \tilde{\mathbf{H}}^{\star}\right)$ where $\tilde{\mathbf{H}}^{\star} \triangleq \tilde{\mathbf{H}}^{\star}\left(P_T/n_T \, \mathbf{I}\right)$ is the minimizing channel of Theorem 4.1 corresponding to the uniform power allocation. Recalling that $\Psi\left(\mathbf{Q}, \tilde{\mathbf{H}}\right)$ is maximized when the eigenvectors of $\mathbf{Q}$ align with the right singular vectors of $\tilde{\mathbf{H}}$ and when the eigenvalues of $\mathbf{Q}$ water-fill the eigenvalues of $\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}$ [Tel95], it must be that $\tilde{\mathbf{H}}^{\star H}\tilde{\mathbf{H}}^{\star}$ is a diagonal matrix and has equal eigenvalues. Thus, it must be that $\tilde{\mathbf{H}}^{\star H}\tilde{\mathbf{H}}^{\star} = \beta \, \mathbf{I}$ or, equivalently, $\lambda^{\star}_{R_H,i} = \beta \; \forall i$ for some β. ■

In the next subsection, specific definitions of $\tilde{\mathcal{H}}$ are considered and Corollary 4.1 will be invoked to show in which cases the uniform power allocation constitutes a saddle point of the strategic game.

In [Dig01], the existence of a saddle point as defined in (4.30) was proved for any set of channels $\tilde{\mathcal{H}}$ such that $\mathbf{H} = \mathbf{I}$ and $\mathbf{R}_n \in \mathcal{R}_n$ where $\mathcal{R}_n$ is closed, bounded, and convex. With the additional constraint that $\mathcal{R}_n$ be isotropically unconstrained, the existence result of [Dig01] can be combined with Theorem 4.1 and Corollary 4.1 to conclude that in such a case, the uniform solution for both the transmitter and the noise always constitute a saddle point of the strategic game as given in (4.30). We state this in the following corollary for further reference.

Corollary 4.2 *Consider the set of channels $\tilde{\mathcal{H}}$ defined such that $\mathbf{H} = \mathbf{I}$ ($n_T = n_R$) and $\mathbf{R}_n \in \mathcal{R}_n$ where $\mathcal{R}_n$ is closed, bounded, convex, and isotropically unconstrained (i.e., unconstrained eigenvectors). It then follows that the uniform power allocation $\mathbf{Q}^\star = P_T/n_T\,\mathbf{I}$ and the noise $\mathbf{R}_n = \sigma_n^2\,\mathbf{I}$ always form a saddle point of the strategic game given by (4.30).*

Many papers have obtained the uniform solution for both the transmitter and the noise (or jammer) as mutual information saddle points for the set of noise covariances with power constraint given by $\mathrm{Tr}\left(\mathbf{R}_n\right) \leq P_n$, *e.g.*, [Bla57, Sta88, Yan93] (also [Hug88][17] for the particular case in which the background noise is removed). Corollary 4.2 generalizes such a result to an arbitrary set of noise covariances $\mathcal{R}_n$ (provided it is closed, bounded, convex, and isotropically unconstrained). Note that a constraint on the channel eigenvalues $\{\lambda_{R_H,i}\}$ can be alternatively expressed (whenever $n_T = n_R$) as a constraint of the form $\mathbf{H} = \mathbf{I}$ and $\mathbf{R}_n \in \mathcal{R}_n$ as considered in Corollary 4.2 since we can write $\lambda_i\left(\mathbf{R}_n\right) = \lambda_{n_T-i+1}\left(\mathbf{R}_H\right)^{-1}$.

As mentioned in §4.4.1.1, even when the strategic game does not have a saddle point or Nash equilibrium, if mixed strategies are allowed the game has then an infinite set of saddle points or Nash equilibria as defined in (4.32) (see Appendix 4.C). In particular, as proved in Appendix 4.C, the mixed-strategy Nash equilibria are given by a pure strategy for the transmitter $\mathbf{Q}^\star = P_T/n_T\,\mathbf{I}_{n_T}$ (uniform power allocation) and a mixed strategy for nature that, for example, puts equal probability on each element of the set $\left\{\tilde{\mathbf{H}} = \mathbf{U}_{\tilde{H}}\boldsymbol{\Sigma}^\star_{\tilde{H}}\mathbf{P}\mathbf{V}^H_{\tilde{H}} : \mathbf{P} \in \boldsymbol{\Pi}\right\}$ where $\boldsymbol{\Sigma}^\star_{\tilde{H}}$ contains in the main diagonal the optimum (worst-case) singular values corresponding to $\mathbf{Q}^\star = P_T/n_T\,\mathbf{I}_{n_T}$ (as in Theorem 4.1), $\mathbf{U}_{\tilde{H}}$ and $\mathbf{V}_{\tilde{H}}$ are two arbitrary unitary matrices, and $\boldsymbol{\Pi}$ is the set of the $n_T!$ different permutation matrices of size $n_T \times n_T$ (see Appendix 4.C for a proof).

4.4.1.2.1 Examples of Channel Constraints

To gain further insight into the problem, we now analyze in detail some particular constraints to define the set of channels $\tilde{\mathcal{H}}$. In principle, for each of the different constraints, it is possible to directly solve the corresponding maximin problem of (4.28). Using the result obtained in

[17] Although [Hug88] deals with the vector Gaussian AVC, the final problem formulation is also given by maximin and minimax mathematical problems.

Theorem 4.1, however, we already know (provided that the set $\tilde{\mathcal{H}}$ is isotropically unconstrained) that the optimal solution is the uniform power allocation $\mathbf{Q}^{\star} = P_T/n_T\,\mathbf{I}_{n_T}$ and that the worst-case channel is given by the minimizing solution to (4.40). It is important to remark that, to find the worst-case channel, it is not necessary to solve the minimization for an arbitrary set $\{\lambda_{Q,i}\}$ to obtain $\{\lambda^{\star}_{R_H,i}(\{\lambda_{Q,i}\})\}$; it suffices to consider directly the uniform solution $\lambda_{Q,i} = P_T/n_T\ \forall i$ as in (4.40) and obtain $\{\lambda^{\star}_{R_H,i}(\{P_T/n_T\})\}$, which is a great simplification.

General Individual Channel Eigenvalue Constraint: $\{\lambda_{R_H,i} \geq \beta_i\}$

Consider a general and individual constraint on each channel eigenvalue $\{\lambda_{R_H,i} \geq \beta_i\}$, where it is assumed that $\beta_i \geq \beta_{i+1} \geq 0$ and that all eigenvalues have a corresponding β_i without loss of generality (if not, one can always set $\beta_i = \beta_{i-1}$ or $\beta_i = 0$ as appropriate). The minimizing channel of (4.40) is easily obtained by minimizing each of the terms of the RHS of (4.40) as

$$\lambda^{\star}_{R_H,i} = \beta_i \qquad 1 \leq i \leq n_T. \tag{4.41}$$

This solution is in general non-uniform and, by Corollary 4.1, is not a saddle point of the strategic game as given in (4.30). Note that Corollary 4.2 cannot be invoked to prove the existence of a saddle point (for $\mathbf{H} = \mathbf{I}$) since the constraints expressed in terms of noise eigenvalues $\lambda_{n,i} = \lambda^{-1}_{R_H,n_T-i+1} \leq \beta^{-1}_{n_T-i+1}$ in general define a nonconvex and unbounded region for the set $\mathcal{R}_n$.

Consider now a constraint just on the maximum channel eigenvalue $\lambda_{R_H,\max} \geq \beta$. This specific constraint has a special interest since the maximum eigenvalue of $\mathbf{R}_H = \tilde{\mathbf{H}}^H\tilde{\mathbf{H}}$ is an upper-bound on the elements of $\mathbf{R}_H$ and, in particular, on the received power corresponding to the ith transmit dimension, $[\mathbf{R}_H]_{ii} = \|\tilde{\mathbf{h}}_i\|^2$, where $\tilde{\mathbf{h}}_i$ is the ith column of the channel matrix $\tilde{\mathbf{H}}$. The minimizing channel is given by

$$\begin{cases} \lambda^{\star}_{R_H,\max} = \beta \\ \quad \lambda^{\star}_{R_H,i} = 0 \qquad 2 \leq i \leq n_T. \end{cases} \tag{4.42}$$

It is also of interest to consider a constraint just on the minimum channel eigenvalue $\lambda_{R_H,\min} \geq \beta$. The minimizing channel is now

$$\lambda^{\star}_{R_H,i} = \beta \qquad 1 \leq i \leq n_T. \tag{4.43}$$

For this particular case, the minimizing channel is uniform and then, by Corollary 4.1, the uniform power allocation forms a saddle point of the strategic game as given in (4.30). Alternatively, Corollary 4.2 could have been invoked to show the existence of a saddle point (for $\mathbf{H} = \mathbf{I}$) since the constraints expressed in terms of noise eigenvalues $\lambda_{n,\max} \leq \beta^{-1}$ form a closed, bounded, convex, and isotropically unconstrained set $\mathcal{R}_n$.

Channel Trace Constraint: $\mathrm{Tr}\,(\mathbf{R}_H) = \sum_i \lambda_{R_H,i} \geq \beta$

The channel trace constraint is probably the most reasonable constraint from a physical standpoint since it represents the total channel energy $\|\tilde{\mathbf{H}}\|_F^2 = \mathrm{Tr}\,(\mathbf{R}_H)$. In [Chi01], this channel constraint was considered, obtaining the same results.

Since function $f(\mathbf{x}) = \sum_{i=1}^{n} \log(1 + x_i\,\alpha)$ is Schur-concave ($-f(\mathbf{x})$ is Schur-convex) [Mar79, 3.H.2] and any eigenvalue distribution is majorized by $(\sum_i \lambda_{R_H,i}, 0, \cdots, 0)$ [Mar79, p.7] (see the proof of Lemma 4.4 for a similar reasoning), it follows that its minimum value is achieved by

$$\begin{cases} \lambda^{\star}_{R_H,\max} = \beta \\ \quad \lambda^{\star}_{R_H,i} = 0 \qquad 2 \leq i \leq n_T. \end{cases} \tag{4.44}$$

This solution is clearly non-uniform and, by Corollary 4.1, does not constitute a saddle point of the strategic game as given in (4.30). Note that Corollary 4.2 cannot be invoked either to prove the existence of a saddle point (for $\mathbf{H} = \mathbf{I}$) since the constraint expressed in terms of noise eigenvalues $\sum_i \lambda_{n,i}^{-1} \geq \beta$ defines a nonconvex and unbounded region for the set $\mathcal{R}_n$.

Maximum Noise Eigenvalue Constraint: $\lambda_{n,\max} \leq \sigma^2$

This constraint is identical to the minimum channel eigenvalue constraint with solution given by (4.43).

Noise Trace Constraint: $\mathrm{Tr}(\mathbf{R}_n) = \sum_i \lambda_{n,i} \leq \sigma^2$

This is the constraint considered in most publications since it is a very natural constraint when the noise is interpreted as a jammer constrained in its average transmit power (as is the intended transmitter). See, for example, [Bla57, Sta88, Yan93] and also [Hug88] for the particular case in which the background noise is removed.

For this particular constraint, we can directly invoke Corollary 4.2 to show that the worst-case noise is given by

$$\lambda^{\star}_{n,i} = \sigma^2 / n_R \qquad 1 \leq i \leq n_R \tag{4.45}$$

and that the uniform power allocation constitutes a saddle point of the strategic game as given in (4.30).

Banded Noise Covariance Constraint

In [Dig01, Sec. III], a banded noise covariance constraint (a noise with correlation constraints at different lags) was analyzed in detail. Such a constraint is not isotropically unconstrained and, consequently, the results of this paper do not apply. Therefore, we cannot conclude that the uniform power allocation is the maximin solution to the mutual information game. In fact, the saddle-point solution was obtained in [Dig01, Sec. III] to be given by the maximum-entropy extension for the noise and by a water-filling solution for the transmitter which in general is non-uniform.

4.4.1.2.2 On the Specific Choice of the Channel Constraints

As has been shown in Theorem 4.1, the uniform power allocation is the solution to the maximin problem of (4.28). In other words, it is a robust solution under channel uncertainty.

Other aspects and observations of the solution, such as whether it is better to have many antennas or just a few in a multi-antenna system, depend on the particular choice of constraints that define the set of channels $\tilde{\mathcal{H}}$ which have to be tailored to each specific application. To illustrate this effect, we now consider some heuristic choices as examples (for simplicity, we assume $\mathbf{R}_n = \mathbf{I}$).

Inspired by a communication system with multiple transmit and receive antennas with a unit-energy channel in the sense of expected value $\mathbb{E}\left[\left|\,[\mathbf{H}]_{i,j}\right|^2\right] = 1$ (which implies $\mathbb{E}\left[\mathrm{Tr}\left(\mathbf{H}^H\mathbf{H}\right)\right] = n_T\, n_R$), we can similarly consider a worst-case problem formulation with the trace constraint defined as

$$\mathrm{Tr}\left(\mathbf{H}^H\mathbf{H}\right) \geq \alpha\, n_T\, n_R$$

where α is a scaling factor that, for example, guarantees that the constraint is satisfied with a certain probability (if the constraint is not satisfied, an outage event is declared). In this case, using the results of §4.4.1.2.1, the worst-case capacity is given by

$$\log\left(1 + \alpha\, P_T\, n_R\right)$$

from which we can conclude that, while adding transmit antennas does not increase the worst-case capacity when the channel state is unknown, adding receive antennas is always beneficial.

Inspired by a set of parallel subchannels, each with unit gain, we can instead define the trace constraint as (assuming $n_T \leq n_R$)

$$\mathrm{Tr}\left(\mathbf{H}^H\mathbf{H}\right) \geq \alpha\, n_T.$$

In this case, the worst-case capacity is given by

$$\log\left(1 + \alpha\, P_T\right)$$

from which we can conclude that the worst-case performance is independent of the number of transmit and receive antennas when the channel state is unknown. However, for this scenario corresponding to a set of parallel subchannels, it may be more appropriate to consider the minimum channel eigenvalue constraint (assuming $n_T \leq n_R$)

$$\lambda_{\min}\left(\mathbf{H}^H\mathbf{H}\right) \geq \alpha,$$

obtaining a worst-case capacity (using the result in §4.4.1.2.1) given by the increasing function of n_T

$$n_T \log\left(1 + \alpha\, P_T / n_T\right) \underset{n_T \to \infty}{\longrightarrow} \alpha\, P_T$$

from which it is always beneficial to add transmit and also receive antennas.

A Numerical Example

In Figure 4.9, the capacity of the uniform power allocation is compared to that of a non-uniform allocation (simply chosen according to the distribution $\boldsymbol{\lambda}_Q = [0.6,\, 0.2,\, 0.1,\, 0.1]^T\, P_T$)

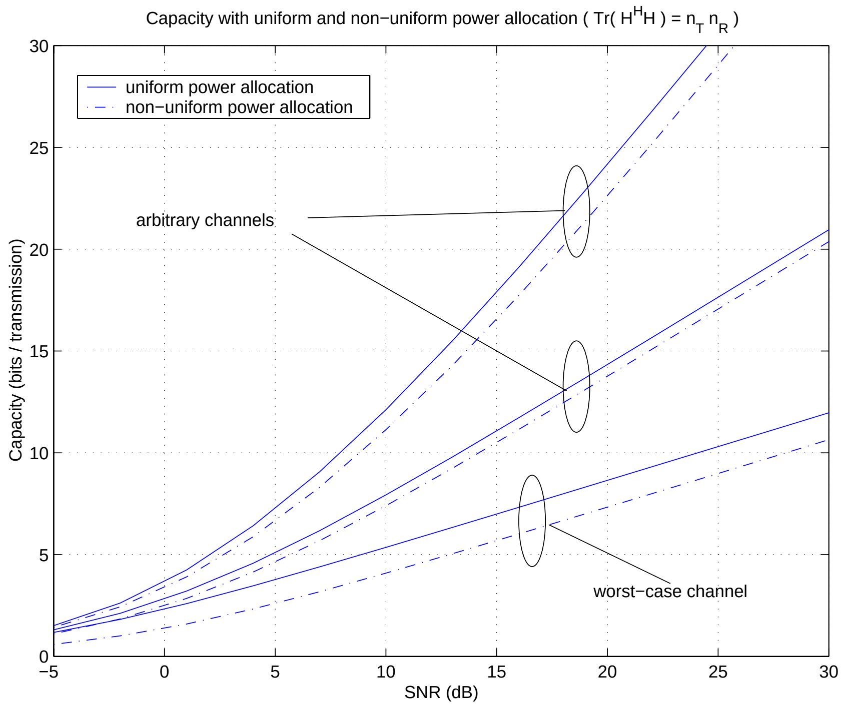


Figure 4.9: Capacity of the uniform and a non-uniform (according to the distribution $\boldsymbol{\lambda}_Q = [0.6, 0.2, 0.1, 0.1]^T P_T$) power allocations vs. the SNR for two arbitrary channels and for the worst channel of the set defined by $\mathrm{Tr}\left(\mathbf{H}^H\mathbf{H}\right) = n_T\, n_R$ and $\mathbf{R}_n = \sigma_n^2\, \mathbf{I}$.

as a function of the SNR defined as $\mathrm{Tr}\left(\mathbf{Q}\right)/\sigma_n^2$, where the noise covariance matrix was fixed to $\mathbf{R}_n = \sigma_n^2\,\mathbf{I}$ and the set of channels $\mathcal{H}$ was constrained using the channel trace constraint $\mathrm{Tr}\left(\mathbf{H}^H\mathbf{H}\right) = n_T\, n_R$ for $n_T = n_R = 4$ (equivalently, $\tilde{\mathcal{H}}$ is defined by $\mathrm{Tr}\left(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right) = n_T\, n_R/\sigma_n^2$). The capacities corresponding to two arbitrary channels and to the worst channel adapted to each power distribution are plotted. As expected, the capacity of the uniform distribution is always the best for the worst-case channel (note that in general, for an arbitrary channel, this may or may not be the case).

4.4.1.2.3 Opposite Problem Formulation: Nature Moves First

For completeness, we now briefly consider the opposite problem formulation, *i.e.*, the minimax problem of (4.29). A solution to (4.29) will always be a saddle point as defined in (4.34) corresponding to the Stackelberg game in which nature moves first and then the transmitter moves aware of natures's move. In some cases, it will also form a saddle point of the strategic game as defined in (4.30).

It is well known that $\Psi\left(\mathbf{Q}, \tilde{\mathbf{H}}\right)$ is maximized when the eigenvectors of $\mathbf{Q}$ align with the right singular vectors of $\tilde{\mathbf{H}}$ and when the eigenvalues of $\mathbf{Q}$ water-fill the eigenvalues of $\mathbf{R}_H = \tilde{\mathbf{H}}^H\tilde{\mathbf{H}}$

[Tel95]. The minimax problem of (4.29) reduces then to

$$\begin{array}{ll} \min_{\{\lambda_{R_H,i}\}} & \sum_{i=1}^{n_T} \log\left(1+\lambda_{Q,i}^{\star}\left(\{\lambda_{R_H,i}\}\right)\lambda_{R_H,i}\right) \\ \text{s.t.} & \{\lambda_{R_H,i}\} \in \mathcal{L}_{R_H} \end{array} \tag{4.46}$$

where $\lambda_{Q,i}^{\star}\left(\{\lambda_{R_H,i}\}\right) = \left(\nu - \lambda_{R_H,i}^{-1}\right)^{+}$ is the water-filling solution and ν is the water-level chosen to satisfy the power constraint of (4.27) with equality. Clearly, we can relabel the $\lambda_{R_H,i}$'s so that they are in decreasing order without loss of generality and, as a consequence of the water-filling solution, the $\lambda_{Q,i}^{\star}$'s will also be in decreasing order.

For the cases considered in §4.4.1.2.1 where saddle points were obtained (minimum channel eigenvalue constraint, maximum noise eigenvalue constraint, noise trace constraint, banded noise covariance constraint), we already know that the same solutions are obtained when nature moves first simply by the definition of saddle point in (4.30).

For the case of a general individual channel eigenvalue constraint $\{\lambda_{R_H,i} \geq \beta_i\}$, the worst-case channel is simply obtained as in (4.41) (although in this case, the eigenvalues of $\mathbf{Q}$ water-fill those of $\mathbf{R}_H$).

The channel trace constraint was considered in [Chi01], where it was found that for low values of the SNR defined as $P_T\,\beta/\sigma_n^2$ (the noise covariance matrix was assumed fixed and given by $\mathbf{R}_n = \sigma_n^2\,\mathbf{I}$) the worst channel is given by $\lambda_{R_H,i}^{\star} = \beta/\min\left(n_T,\, n_R\right)\ \forall i$ and for high values of the SNR the worst channel is similarly given except that a dominant eigenvalue arises.

4.4.1.2.4 Extension to Ergodic and Outage Capacities

In addition to analyzing robustness in terms of instantaneous capacity (which implies a fixed channel state $\tilde{\mathbf{H}}$), it is also interesting to consider other statistics of the capacity such as average and outage values as described in §4.3. Whereas channel pdf was assumed known and fixed in §4.3, we now consider the case in which it is known to belong to a class of pdf's but otherwise unknown.

In this sense, the maximin formulation is as in (4.28) but now the payoff function is given either by $\mathcal{I}^{\text{erg}}\left(\mathbf{Q}\right)$ or $\mathcal{I}_{\epsilon}^{\text{out}}\left(\mathbf{Q}\right)$ (see (4.18) and (4.19)) and the minimization is over the set of possible channel pdf's $p_{\tilde{\mathbf{H}}} \in \mathcal{P}_{\tilde{H}}$ in which the channel singular vectors are unconstrained (isotropic property), *e.g.*, $\mathbb{E}\left[\lambda_{\max}\left(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right)\right] \geq \beta$. Without going into details, we justify why the uniform power allocation is also obtained as a robust solution in terms of ergodic and outage capacities. Simply note that $p_{\tilde{\mathbf{H}}}$ can always be chosen as a function of the utilized $\mathbf{Q}$ to put positive probability only on channel states $\tilde{\mathbf{H}}$ with the singular vectors chosen to perform an "inverse water-filling" on $\mathbf{Q}$ (c.f. Theorem 4.1), against which the best solution for the transmitter is a uniform power allocation. Therefore, it is an optimal solution for every choice of $\tilde{\mathbf{H}}$ and hence for other capacity statistics such as the average and the outage values.

It is interesting to point out that if $p_{\tilde{\mathbf{H}}}$ does a randomization over a set of channel states in $\tilde{\mathcal{H}}$ as defined in the previous subsections (this need not be in a general case), the ergodic capacity problem then results in a mixed-strategy formulation of a game in which the pure strategies are defined by $\tilde{\mathcal{H}}$ and, therefore, the previously obtained results on mixed strategy Nash equilibria apply (c.f. Appendix 4.C).

It is important to bear in mind that the optimality of the uniform power allocation in terms of ergodic and outage capacities is in the worst-case sense, *i.e.*, when $p_{\tilde{\mathbf{H}}}$ is known to belong to a set $\mathcal{P}_{\tilde{H}}$ but otherwise unknown. Therefore, it cannot be concluded from the obtained results that the uniform power allocation is optimum in terms of outage capacity for the case, for example, of a random $\tilde{\mathbf{H}}$ with i.i.d. Gaussian entries which is a well-known open problem as discussed in [Tel95] (where it was conjectured that the uniform power allocation could be the optimal solution, but only over a certain number of transmit dimensions). It is interesting, however, to remark that, by definition, the worst-case instantaneous capacity for a set of channels $\mathcal{H}$ as previously considered happens to be the zero-outage capacity (also termed delay-limited capacity [Big98, Cai99b, Big01]) for any $p_{\mathbf{H}}$ that puts non-zero probability on each member of $\mathcal{H}$. Unfortunately, this result is not very useful for the case of a random channel $\mathbf{H}$ with i.i.d. $\mathcal{CN}(0,1)$ entries, since it has a zero worst-case capacity [Big01].[18]

Note that if one considers that nature can only control the channel eigenvalues but not the eigenvectors, then the optimality of the uniform power allocation need not hold.

4.4.1.2.5 Cost of Robustness

Robustness is a desirable property that comes with a price. For a given channel state $\tilde{\mathbf{H}}$, one can explicitly compute the loss in performance of the robust uniform power distribution with respect to the optimum allocation (obtained with a perfect instantaneous knowledge of the channel state). However, it is also interesting to know the worst-case loss of performance for a given class of channels $\tilde{\mathcal{H}}$. In this section, the cost of robustness for a family of channels $\tilde{\mathcal{H}}$ (or, equivalently, $\mathcal{R}_H$) is analyzed using the concept of duality gap arising in convex optimization theory (see §3.1 and [Lue69, Roc70, Boy00]) following the approach proposed in [Yu01a].

Assuming for the moment a fixed channel state given by $\{\lambda_{R_H,i}\}$, the maximization of the mutual information can be expressed in convex form (we use in this section logarithms in base 2 and natural logarithms denoted by $\log_2$ and ln, respectively) as

$$\begin{array}{ll} \min & f_0(\mathbf{x}) = -\sum_{i=1}^{n} \log_2\left(1 + x_i\,\lambda_{R_H,i}\right) \\ \text{s.t.} & \sum_i x_i \le P, \\ & x_i \ge 0 \qquad 1 \le i \le n. \end{array} \tag{4.47}$$

[18] The worst-case capacity studied in this paper is equivalent to the delay-limited capacity considered in [Big01] under a short-term power constraint (in [Big01], however, perfect CSI was assumed).

(note that Slater's condition is satisfied and, therefore, strong duality holds) and the Lagrangian is

$$L(\mathbf{x}, (\lambda, \boldsymbol{\mu})) = -\sum_{i=1}^{n} \log_2 (1 + x_i\, \lambda_{R_H,i}) + \lambda \left(\sum_i x_i - P \right) - \sum_{i=1}^{n} \mu_i\, x_i. \tag{4.48}$$

The dual objective, $g(\lambda, \boldsymbol{\mu})$, is obtained by setting $\frac{\partial L}{\partial x_i} = 0$, which gives the water-filling solution

$$x_i + \frac{1}{\lambda_{R_H,i}} = \frac{1}{(\lambda - \mu_i)} \frac{1}{\log 2}. \tag{4.49}$$

If we now evaluate the duality gap (see §3.1) denoted by $\Gamma(\mathbf{x}, (\lambda, \boldsymbol{\mu}))$ at any $\mathbf{x}$ and with $\boldsymbol{\mu}$ chosen so that the water-filling condition (4.49) is satisfied, then the duality gap is

$$\Gamma(\mathbf{x}, \lambda) = -\frac{1}{\log 2} \sum_{i=1}^{n} \frac{x_i\, \lambda_{R_H,i}}{1 + x_i\, \lambda_{R_H,i}} + \lambda P. \tag{4.50}$$

(Note that a better choice of the Lagrange multipliers to obtain a smaller gap could be made, however this choice produces a simple closed-form expression.) Using the smallest possible value for λ (such that all the Lagrange multipliers λ and μ_i's are non-negative)

$$\lambda = \frac{1}{\log 2} \max_i \left(\frac{\lambda_{R_H,i}}{1 + x_i\, \lambda_{R_H,i}} \right) \tag{4.51}$$

and assuming that the power constraint is satisfied with equality $\sum_i x_i = P$, we can write the duality gap as

$$\Gamma(\mathbf{x}) = \frac{1}{\log 2} \sum_{i=1}^{n} x_i \left(\max_j \left(\frac{\lambda_{R_H,j}}{1 + x_j\, \lambda_{R_H,j}} \right) - \frac{\lambda_{R_H,i}}{1 + x_i\, \lambda_{R_H,i}} \right). \tag{4.52}$$

Finally, evaluating the gap for a uniform power allocation $x_i = P/n$, we obtain

$$\Gamma^{\text{uni}}(\boldsymbol{\lambda}_{R_H}) \triangleq \Gamma(\{x_i = P/n\}) = \frac{1}{\log 2} \sum_{i=2}^{n} \left(\frac{P/n\, \lambda_{R_H,\max}}{1 + P/n\, \lambda_{R_H,\max}} - \frac{P/n\, \lambda_{R_H,i}}{1 + P/n\, \lambda_{R_H,i}} \right) \tag{4.53}$$

where we have made explicit the dependence of the gap on the channel eigenvalues $\{\lambda_{R_H,i}\}$ which are assumed in decreasing order. For a channel with equal eigenvalues, $\lambda_{R_H,i} = \kappa$, the uniform power allocation is optimum and the gap becomes zero as expected. Note that for $P \rightarrow \infty$ (with positive $\lambda_{R_H,i}$'s) the gap also tends to zero, *i.e.*, for high SNR the uniform distribution tends to be optimal (this observation was empirically made in [Cho93] and further analyzed in [Yu01a]).

Now we can use the closed-form expression in (4.53) to easily obtain an upper bound on the worst-case loss of performance for the class of channels $\mathcal{R}_H$. For example, if we consider a maximum channel eigenvalue constraint $\lambda_{R_H,\max} \geq \beta$, the gap is

$$\Gamma^{\text{uni}} = \frac{n-1}{\log 2} \frac{P/n\, \lambda_{R_H,\max}}{1 + P/n\, \lambda_{R_H,\max}} \underset{\lambda_{R_H,\max} \rightarrow \infty}{\longrightarrow} \frac{n-1}{\log 2}. \tag{4.54}$$

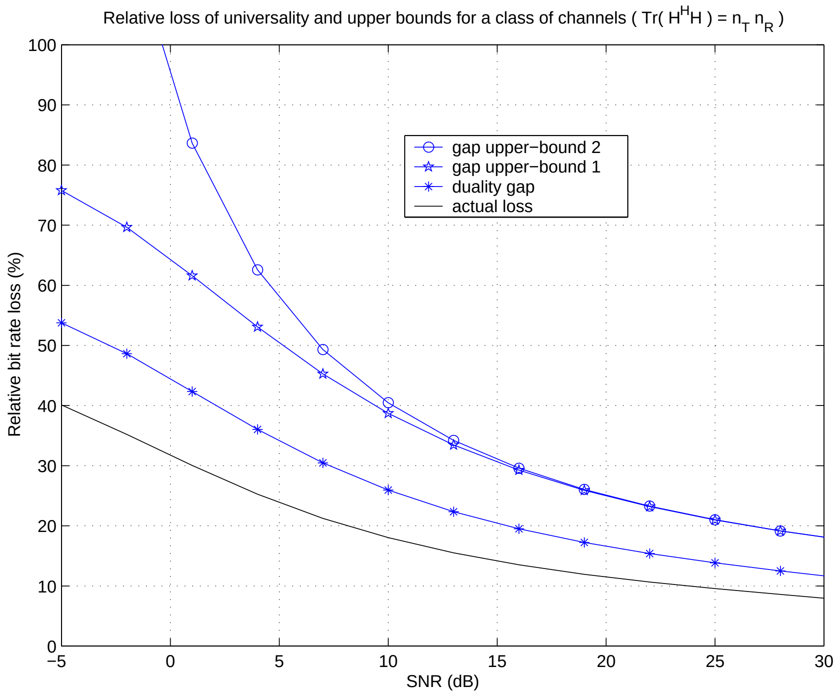


Figure 4.10: Relative bit-rate loss and duality gap of the uniform power allocation, along with two upper bounds, vs. the SNR for a channel realization ($\boldsymbol{\lambda}_{R_H} = [56\%,\ \sim 44\%,\ 10^{-3}\%,\ 5 \times 10^{-4}\%]^T$) corresponding to the class of channels defined by $\mathrm{Tr}\left(\mathbf{H}^H\mathbf{H}\right) = n_T\, n_R$ and $\mathbf{R}_n = \sigma_n^2\, \mathbf{I}$.

Note that for a channel trace constraint $\sum_i \lambda_{R_H,i} \geq \beta$, the same gap is obtained. If instead we consider a minimum channel eigenvalue constraint $\lambda_{R_H,\min} \geq \beta$, the gap is

$$\Gamma^{\mathrm{uni}} = \frac{n-1}{\log 2}\left(\frac{P/n\,\lambda_{R_H,\max}}{1+P/n\,\lambda_{R_H,\max}} - \frac{P/n\,\beta}{1+P/n\,\beta}\right) \underset{\lambda_{R_H,\max}\to\infty}{\longrightarrow} \frac{n-1}{\log 2}\,\frac{1}{1+P/n\,\beta}. \tag{4.55}$$

In any case, the gap in (4.53) is always upper-bounded as

$$\Gamma^{\mathrm{uni}}\left(\boldsymbol{\lambda}_{R_H}\right) \leq \frac{n-1}{\log 2} \tag{4.56}$$

which in turn is upper-bounded by $n/\log 2$ bits/transmission or, equivalently, by 1.4427 bits/transmission/dimension as was found in [Yu01a].

Example

As an illustrative example, we consider a channel trace constraint given by $\mathrm{Tr}\left(\mathbf{H}^H\mathbf{H}\right) = n_T\, n_R$ for $n_T = n_R = 4$ (the noise covariance matrix was fixed to $\mathbf{R}_n = \sigma_n^2\, \mathbf{I}$) and plot in Figure 4.10 the actual relative bit-rate loss and duality gap as given in (4.53) for a channel realization, along with the gap upper-bounds of (4.54) (both the asymptotic and the non-asymptotic versions), as a function of the SNR defined as $\mathrm{Tr}\left(\mathbf{Q}\right)/\sigma_n^2$.

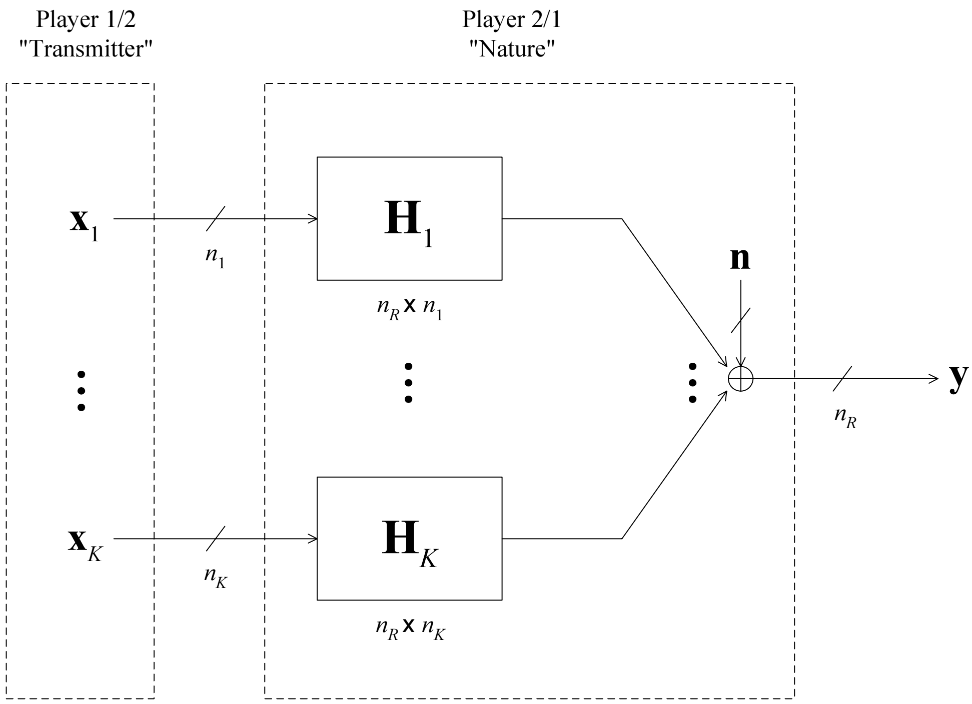


Figure 4.11: Communication interpreted as a two-player game for the multiple-access channel.

4.4.2 Capacity Region of the Multiple-Access Channel (MAC)

In this section, we extend the previous single-user results and prove the optimality of the uniform power allocation in terms of robustness for the multiple-access channel (MAC). In particular, we show that all rates inside the capacity region of the compound vector MAC are achieved when each user uses a uniform power allocation.

4.4.2.1 Game-Theoretic Formulation

As in the single-user case of §4.4.1, we constrain our search to Gaussian-distributed signals and noise, since they constitute a robust solution (a saddle point) to the mutual information game for the memoryless vector MAC (this follows by applying the results of [Bor85, Dig01] to each of the constraints that define the capacity region).

We impose some constraints on the set of possible channels $\{\tilde{\mathbf{H}}_k\} \in \tilde{\mathcal{H}} \triangleq \tilde{\mathcal{H}}_1 \times \cdots \times \tilde{\mathcal{H}}_K$ to avoid the trivial solution (note that the class of channels seen by each user may be different). We assume that each set $\tilde{\mathcal{H}}_k$ is isotropically unconstrained, *i.e.*, with unconstrained right singular vectors (see Definition 4.1 in §4.4.1).

From the perspective of robustness under channel uncertainty, we are interested in the worst-case capacity region, *i.e.*, in the set of rates that can be achieved regardless of the set of channel states chosen from the set of possible channels $\{\tilde{\mathbf{H}}_k\} \in \tilde{\mathcal{H}}$. This can be formulated as a game (see Figure 4.11) where the first player is the transmitter and the second player, who controls the whole set of channels $\{\tilde{\mathbf{H}}_k\}$ and is aware of the first player's move, is nature.

4.4.2.2 Worst-Case Capacity Region and Robust Power Allocation

The worst-case capacity region is in fact the notion of capacity region of the compound MAC [Ver89, Lap98] (see also [Csi81, p. 288]). Mathematically, the worst-case region of the set of achievable rates for a fixed set of transmit covariance matrices $\{\mathbf{Q}_k\}$ is expressed (similarly to the maximin formulation of the single-user case in (4.28)) as the following intersection:

$$\mathcal{R}\big(\{\mathbf{Q}_k\}, \tilde{\mathcal{H}}\big) = \bigcap_{\{\tilde{\mathbf{H}}_k\} \in \tilde{\mathcal{H}}} \mathcal{R}\big(\{\mathbf{Q}_k\}, \{\tilde{\mathbf{H}}_k\}\big) \tag{4.57}$$

which is closed and convex because it is the intersection of closed and convex sets. Recall that $\mathcal{R}\big(\{\mathbf{Q}_k\}, \{\tilde{\mathbf{H}}_k\}\big)$ is the set of achievable sets for the set of channel states $\{\tilde{\mathbf{H}}_k\}$ as given in (4.13). Assuming that the transmit covariance matrices are constrained in their average transmit power, the worst-case capacity region (capacity region of the compound vector Gaussian MAC) is [Ver89][19]

$$\mathcal{C}(\tilde{\mathcal{H}}) = \bigcup_{\substack{\mathrm{Tr}(\mathbf{Q}_k) \leq P_k, \\ \mathbf{Q}_k = \mathbf{Q}_k^H \geq 0}} \mathcal{R}\big(\{\mathbf{Q}_k\}, \tilde{\mathcal{H}}\big) \tag{4.58}$$

which also happens to be closed and convex as shown in Theorem 4.2. In [Han98], an expression similar to (4.58) was obtained as the delay-limited MAC capacity region (although the case of perfect CSIT was considered therein). The worst-case capacity region is formally characterized in the following theorem.

Theorem 4.2 *The capacity region of the compound vector Gaussian memoryless multiple-access channel (MAC) composed of K users with power constraints $\{P_k\}$, number of transmit dimensions $\{n_k\}$, and n_R receive dimensions (with no CSI) is*

$$\mathcal{C}(\tilde{\mathcal{H}}) = \Bigg\{ (R_1, \cdots, R_K) : \\ 0 \leq \sum_{k \in \mathcal{S}} R_k \leq \inf_{\{\tilde{\mathbf{H}}_k\} \in \tilde{\mathcal{H}}} \log\det\Bigg(\mathbf{I}_{n_R} + \sum_{k \in \mathcal{S}} P_k/n_k \, \tilde{\mathbf{H}}_k \tilde{\mathbf{H}}_k^H\Bigg), \forall \mathcal{S} \subseteq \{1, \cdots, K\} \Bigg\} \tag{4.59}$$

where the class of channels $\tilde{\mathcal{H}}$ is an isotropically unconstrained set (unconstrained right singular vectors). All set of rates within the region of (4.59) are achieved when each user utilizes a

[19] As argued in [Ver89], achievability follows easily using randomized codes and the converse is established since reliable communication has to be guaranteed no matter what channel state is in effect. Similarly to the single-user case, the capacity region remains the same if the receiver is uninformed of the channel state [Csi81, p. 293].

Gaussian code with a uniform power allocation

$$\mathbf{Q}_k^{\star} = P_k/n_k \, \mathbf{I}_{n_k} \qquad 1 \leq k \leq K \tag{4.60}$$

which implies an independent signaling over the transmit dimensions for each user.

Proof. See Appendix 4.D. ■

It is important to remark that all points inside the worst-case capacity region are achieved by the same structure of transmit covariance matrices $\{\mathbf{Q}_k\}$, *i.e.*, by a uniform power allocation $\{P_k/n_k \, \mathbf{I}_{n_k}\}$. This is a significant difference with respect to the case with CSIT obtained from (4.14) in which each point of the region requires, in general, a different structure for the transmit covariance matrices [Yu01b].

It is possible to further simplify the expression for each of the boundaries of the worst-case capacity region (4.59) obtained in Theorem 4.2, provided that the left singular vectors of the class of channels are unconstrained as well (this means unconstrained receive as well as transmit directions). In other words, only the singular values of the channels are constrained and, therefore, $\tilde{\mathbf{H}}_k \in \tilde{\mathcal{H}}_k$ if and only if $\left\{\lambda_i\big(\tilde{\mathbf{H}}_k\tilde{\mathbf{H}}_k^H\big)\right\} \in \mathcal{L}_{H_k}$ (we similarly define $\mathcal{L}_H \triangleq \mathcal{L}_{H_1} \times \cdots \times \mathcal{L}_{H_K}$). We first state a lemma and then proceed to simplify the boundaries of the worst-case capacity region in (4.59).

Lemma 4.1 *Let $\{\mathbf{R}_k\}$ be a set of J $n \times n$ Hermitian matrices. Then, the following inequality is verified*

$$\det\left(\mathbf{R}_1 + \cdots + \mathbf{R}_J\right) \geq \prod_i \left(\lambda_i\left(\mathbf{R}_1\right) + \cdots + \lambda_i\left(\mathbf{R}_J\right)\right) \tag{4.61}$$

where $\lambda_i(\cdot)$ denotes the ith ordered eigenvalue in decreasing order and equality is achieved when all the $\mathbf{R}_i$'s have the same eigenvectors with eigenvalues in the same order.

Proof. This result is a generalization of the particular case $J = 2$ considered in [Mar79, 9.G.3.a] and is proved in Appendix 4.E. ■

Since the left singular vectors of the channels $\{\tilde{\mathbf{H}}_k\}$ are unconstrained, we can invoke Lemma 4.1 to obtain

$$\inf_{\{\tilde{\mathbf{H}}_k\}\in\tilde{\mathcal{H}}} \log\det\left(\mathbf{I}_{n_R} + \sum_{k\in\mathcal{S}} P_k/n_k \, \tilde{\mathbf{H}}_k\tilde{\mathbf{H}}_k^H\right) = \inf_{\{\lambda_i(\tilde{\mathbf{H}}_k\tilde{\mathbf{H}}_k^H)\}\in\mathcal{L}_H} \sum_{i=1}^{n_R} \log\left(1 + \sum_{k\in\mathcal{S}} P_k/n_k \, \lambda_i\big(\tilde{\mathbf{H}}_k\tilde{\mathbf{H}}_k^H\big)\right). \tag{4.62}$$

This implies that the worst-case is obtained by choosing the same left singular vectors for each $\tilde{\mathbf{H}}_k$ (the right singular vectors are irrelevant) such that the eigenvalues of $\tilde{\mathbf{H}}_k\tilde{\mathbf{H}}_k^H$ are ordered in the same way for all k.

Example

In Figure 4.12, the worst-case capacity region of a two-user system is plotted for a class of 3×3 matrix channels with eigenvalues constrained to be exactly $\boldsymbol{\lambda}(\tilde{\mathbf{H}}_1\tilde{\mathbf{H}}_1^H) = [9.80, 9.24, 4.59]^T$ and

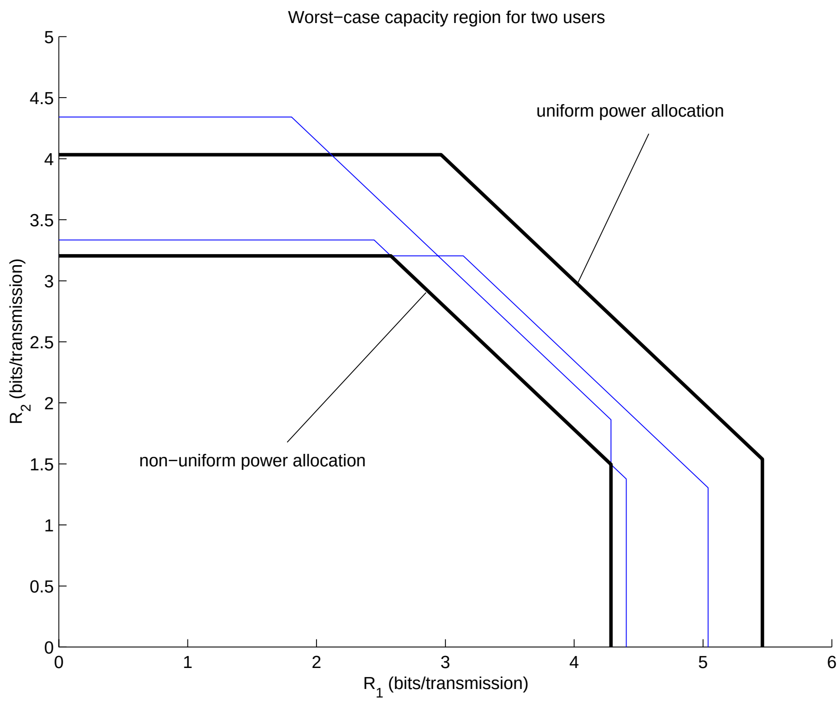


Figure 4.12: Worst-case capacity region corresponding to the channel eigenvalues $\boldsymbol{\lambda}(\tilde{\mathbf{H}}_1\tilde{\mathbf{H}}_1^H) = [9.80, 9.24, 4.59]^T$ and $\boldsymbol{\lambda}(\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_2^H) = [9.19, 5.45, 1.29]^T$, when using a uniform and non-uniform (according to $\boldsymbol{\lambda}_{Q_1} = [0.64, 0.34, 0.02]^T$ and $\boldsymbol{\lambda}_{Q_2} = [0.52, 0.40, 0.08]^T$) power allocation. The latter is obtained as the intersection of the three capacity regions plotted in thin lines.

$\boldsymbol{\lambda}(\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_2^H) = [9.19, 5.45, 1.29]^T$. The three inequalities defining the capacity region corresponding to the uniform power allocation are simultaneously minimized by the same worst-case set of channels according to (4.62). We also plot the worst-case capacity region corresponding to a non-uniform power allocation, in particular for $\boldsymbol{\lambda}_{Q_1} = [0.64, 0.34, 0.02]^T$ and $\boldsymbol{\lambda}_{Q_2} = [0.52, 0.40, 0.08]^T$. In this case, however, the three inequalities are not simultaneously minimized by the same choice of $\tilde{\mathbf{H}}_1$ and $\tilde{\mathbf{H}}_2$. To obtain the capacity region, therefore, we have to obtain the three capacity regions in which each one of the three inequalities is minimized[20] (plotted in thin lines) and then compute the intersection. This is due to the fact that, in general, there are no channels $\tilde{\mathbf{H}}_1$ and $\tilde{\mathbf{H}}_2$ that simultaneously minimize all inequalities, unlike in the uniform case.

4.5 Chapter Summary and Conclusions

In this chapter, we have analyzed the capacity of MIMO channels for different degrees of CSI. First of all, the instantaneous capacity (instantaneous CSIT) has been overviewed, with the well-known

[20] The worst-case capacity region corresponding to the non-uniform power allocation in Figure 4.12 has been computed by choosing the channels with left singular vectors as dictated by Lemma 4.1 and by arbitrarily choosing the right singular vectors to diagonalize the transmit covariance matrices and then optimizing over the permutations only. The ultimate worst-case capacity region by properly optimizing the right singular vectors may be even smaller.

solution based on diagonalizing the channel matrix and distributing the power over the channel eigenmodes in a water-filling fashion. The particular case of practical interest corresponding to beamforming-constrained systems has been specifically addressed. Next, the ergodic and outage capacities (statistical CSIT) have been reviewed. Finally and more extensively, the case in which not even the channel statistics are known (no CSIT) has been investigated in great detail to obtain robust solutions under channel uncertainty.

To be specific, the problem with no CSIT has been formulated within the framework of game theory in which the payoff function of the game is the mutual information and the players are the transmitter and a malicious nature. Mathematically, this has been expressed as a maximin problem, obtaining a robust power allocation under channel uncertainty. This problem characterizes the capacity of the compound vector Gaussian channel. Interestingly, the uniform power allocation has been obtained as a robust solution to the game for the class of isotropically unconstrained channels (unconstrained "directions"). The loss of capacity when using the uniform power allocation has been analytically bounded, showing that for high SNR the loss is small.

For the more interesting and general case of a multiple-access channel (MAC), a uniform power allocation for each of the users also constitutes a robust solution. To be more specific, the worst-case rate region corresponding to the uniform power distribution has been shown to contain the worst-case rate region of any other possible power allocation strategy. In other words, the capacity region of the compound vector Gaussian MAC is achieved when each of the users is using a uniform power allocation.

The main contribution of this chapter is the characterization of the worst-case capacity of MIMO channels, in which both the channel and the noise covariance matrix are unknown, and its formulation as different types of games within the framework of game theory (Section 4.4), obtaining the interesting result that a uniform power allocation (transmitting in all directions) is the best solution when nothing is known about the channel. Of course, the extension to the MAC is even more interesting than the single-user case. Other small contributions of this chapter include: the explicit consideration of beamforming-constrained systems (both in the single-user and multiuser cases), the characterization of the optimum transmit strategy in terms of ergodic capacity for a random channel matrix with i.i.d. entries drawn from a symmetric pdf (not necessarily Gaussian or isotropic), the explicit formulation of the outage capacity problem as a nonconvex problem and its relaxation to obtain upper and lower bounds, and the personal treatment and exposition of the subject.

Appendix 4.A Proof of Proposition 4.1

We first obtain a lemma and then proceed to prove Proposition 4.1.

Lemma 4.2 *If the entries of the complex random matrix* $\mathbf{H}$ *are i.i.d. with a symmetric pdf along the origin (i.e.,* $p_\nu(\nu) = p_\nu(-\nu)$*), then* $\mathbf{H} \sim \mathbf{HPS}$*, where* $\mathbf{P}$ *is an arbitrary permutation matrix and* $\mathbf{S}$ *is a diagonal matrix with diagonal elements in the set* $\{-1, +1\}$*. This result also applies to distributions with circular symmetry,* $p_\nu(\nu) = p_\nu(\nu e^{j\phi})$ $\forall\phi$*, since it is a particular case of the symmetry along the origin.*

Proof. Since the elements of $\mathbf{H}$ are i.i.d., its pdf is $p_{\mathbf{H}}(\mathbf{H}) = \prod_{i,j} p_h(h_{ij})$, from which it is clear that a permutation of the factors induced by matrix $\mathbf{P}$ does not alter the final product. Furthermore, since the entries are symmetric, $p_\nu(\nu) = p_\nu(-\nu)$, any change of sign induced by matrix $\mathbf{S}$ does not affect the final product either. ■

Proof of Proposition 4.1. Using the mutual information function as defined in (4.10) (which implies a Gaussian code) and invoking Lemma 4.2, it follows that

$$\Psi(\mathbf{Q}, \tilde{\mathbf{H}}) \sim \Psi(\mathbf{Q}, \tilde{\mathbf{H}}\mathbf{PS})$$

or, equivalently,

$$\Psi(\mathbf{Q}, \tilde{\mathbf{H}}) \sim \Psi(\mathbf{PSQS}^H\mathbf{P}^H, \tilde{\mathbf{H}}).$$

In particular, this implies that $\mathbb{E}_{\mathbf{H}}\,\Psi(\mathbf{Q}, \tilde{\mathbf{H}}) = \mathbb{E}_{\mathbf{H}}\,\Psi(\mathbf{PSQS}^H\mathbf{P}^H, \tilde{\mathbf{H}})$.

Considering now all possible $n_T!\,2^{n_T}$ different combinations of permutations and changes of sign of $\mathbf{P}$ and $\mathbf{S}$, we have

$$\begin{aligned}
\mathbb{E}_{\mathbf{H}}\,\Psi(\mathbf{Q}, \tilde{\mathbf{H}}) &= \frac{1}{n_T!2^{n_T}} \sum_{\substack{\mathbf{P}\in\Pi \\ \mathbf{S}\in\Omega}} \mathbb{E}_{\mathbf{H}}\,\Psi(\mathbf{PSQS}^H\mathbf{P}^H, \tilde{\mathbf{H}}) \\
&\leq \mathbb{E}_{\mathbf{H}}\,\Psi\left(\frac{1}{n_T!2^{n_T}} \sum_{\substack{\mathbf{P}\in\Pi \\ \mathbf{S}\in\Omega}} \mathbf{PSQS}^H\mathbf{P}^H, \tilde{\mathbf{H}}\right)
\end{aligned}$$

where Π and Ω represent the set of $n_T!$ different permutation matrices and 2^{n_T} sign matrices, respectively, and we have used Jensen's inequality (see §3.3).

Therefore, the covariance matrix $\tilde{\mathbf{Q}} \triangleq \frac{1}{n_T!2^{n_T}} \sum_{\mathbf{P},\mathbf{S}} \mathbf{PSQS}^H\mathbf{P}^H$ obtains a higher mutual information (note that the trace constraint is also verified by $\tilde{\mathbf{Q}}$). By inspection, it can be seen that $\tilde{\mathbf{Q}}$ corresponds to a uniform power allocation, *i.e.*, it is proportional to the identity matrix,

$$\begin{aligned}
[\tilde{\mathbf{Q}}]_{ij} &= \kappa_1 \sum_{\substack{\mathbf{P}\in\Pi \\ \mathbf{S}\in\Omega}} \mathbf{e}_i^H \mathbf{PSQS}^H\mathbf{P}^H \mathbf{e}_j \\
&= \kappa_2 \sum_{\substack{(\pi_i,\pi_j) \\ \mathbf{S}\in\Omega}} \mathbf{e}_{\pi_i}^H \mathbf{SQS}^H \mathbf{e}_{\pi_j} \\
&= \kappa_2 \sum_{\substack{(\pi_i,\pi_j) \\ \mathbf{S}\in\Omega}} [\mathbf{S}]_{\pi_i,\pi_i} [\mathbf{S}]_{\pi_j,\pi_j} [\mathbf{Q}]_{\pi_i,\pi_j} \\
&= \kappa_3 \delta_{i,k}
\end{aligned}$$

where $\mathbf{e}_i$ is the ith unit vector (all zeros except a one in the ith position). The domain of the permutation indexes (π_i, π_j) is given by $\pi_i = \pi_j \in \{1, \cdots, n_T\}$ for $i = j$, and $\pi_i \in \{1, \cdots, n_T\}, \pi_j \in \{1, \cdots, n_T\} - \pi_i$ for $i \neq j$. The last equality is clear for $i = j$, because $\left| [\mathbf{S}]_{\pi_i,\pi_i} \right|^2 = 1$. For the case $i \neq j$, it follows from the fact that for each pair (π_i, π_j) (note that $\pi_i \neq \pi_j$), half of the terms of the summation over $\mathbf{S} \in \Omega$ have a positive sign and the other half a negative sign.

Thus, the uniform power allocation $\mathbf{Q} = P_T/n_T\, \mathbf{I}_{n_T}$ always give an upper bound on the ergodic mutual information and the ergodic capacity is then

$$\mathbb{E}_{\mathbf{H}}\, \Psi\big(P_t/n_T\, \mathbf{I}_{n_T}, \tilde{\mathbf{H}}\big) = \mathbb{E}_{\mathbf{H}} \log\det\big(\mathbf{I}_{n_R} + P_t/n_T\, \mathbf{R}_n^{-1}\mathbf{H}\mathbf{H}^H\big)\,.$$

■

Appendix 4.B Proof of Theorem 4.1

We first present a couple of lemmas and then proceed to prove Theorem 4.1.

Lemma 4.3 *Given two positive semidefinite $n \times n$ Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$, the following holds*

$$\log\det(\mathbf{I} + \mathbf{AB}) \geq \sum_{i=1}^{n} \log\left(1 + \lambda_{A,i}\, \lambda_{B,n-i+1}\right), \tag{4.63}$$

where $\lambda_{A,i}$ and $\lambda_{B,i}$ denote the eigenvalues in decreasing order ($\lambda_i \geq \lambda_{i+1}$) of $\mathbf{A}$ and $\mathbf{B}$, respectively. Equality in (4.63) is achieved for $\mathbf{U}_A = \mathbf{U}_B\mathbf{J}$, where $\mathbf{U}_A$ and $\mathbf{U}_B$ contain the eigenvectors corresponding to the eigenvalues in decreasing order of $\mathbf{A}$ and $\mathbf{B}$, respectively, and $\mathbf{J}$ is the "backward identity" permutation matrix [Hor85] defined as

$$\mathbf{J} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & \ddots & & 1 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 1 & & \ddots & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Proof. Consider the EVD $\mathbf{A} = \mathbf{U}_A\mathbf{D}_A\mathbf{U}_A^H$ and $\mathbf{B} = \mathbf{U}_B\mathbf{D}_B\mathbf{U}_B^H$, where $\mathbf{D}_A = \mathrm{diag}(\{\lambda_{A,i}\})$ and $\mathbf{D}_B = \mathrm{diag}(\{\lambda_{B,i}\})$ (we assume eigenvalues in decreasing order). It follows that

$$\det(\mathbf{I} + \mathbf{AB}) = \det\big(\mathbf{I} + \mathbf{D}_A\tilde{\mathbf{U}}^H\mathbf{D}_B\tilde{\mathbf{U}}\big)$$

where $\tilde{\mathbf{U}} = \mathbf{U}_B^H\mathbf{U}_A$. If $\mathbf{A}$ has $n - k$ zero eigenvalues, we can write

$$\mathbf{D}_A = \begin{bmatrix} \mathbf{D}_{A,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{D}_{A,1} \in \mathbb{C}^{k\times k} \quad \text{(nonsingular)}$$

$$\tilde{\mathbf{U}} = \begin{bmatrix} \tilde{\mathbf{U}}_1 & \tilde{\mathbf{U}}_2 \end{bmatrix}, \qquad \tilde{\mathbf{U}}_1 \in \mathbb{C}^{n\times k}$$

and then

$$\begin{aligned}
\det\left(\mathbf{I}+\mathbf{D}_A\tilde{\mathbf{U}}^H\mathbf{D}_B\tilde{\mathbf{U}}\right) &= \det\left(\mathbf{I}+\begin{bmatrix}\mathbf{D}_{A,1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{U}}_1^H\\ \tilde{\mathbf{U}}_2^H\end{bmatrix}\mathbf{D}_B\begin{bmatrix}\tilde{\mathbf{U}}_1 & \tilde{\mathbf{U}}_2\end{bmatrix}\right)\\
&= \det\left(\mathbf{I}+\begin{bmatrix}\mathbf{D}_{A,1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{U}}_1^H\\ \mathbf{0}\end{bmatrix}\mathbf{D}_B\begin{bmatrix}\tilde{\mathbf{U}}_1 & \mathbf{0}\end{bmatrix}\right)\\
&= \det\left(\mathbf{I}+\mathbf{D}_{A,1}\tilde{\mathbf{U}}_1^H\mathbf{D}_B\tilde{\mathbf{U}}_1\right)\\
&= \det\left(\mathbf{D}_{A,1}\right)\,\det\left(\mathbf{D}_{A,1}^{-1}+\tilde{\mathbf{U}}_1^H\mathbf{D}_B\tilde{\mathbf{U}}_1\right)\\
&\geq \left(\prod_{i=1}^{k}\lambda_{A,i}\right)\left(\prod_{i=1}^{k}\left(\lambda_{A,k-i+1}^{-1}+\lambda_i\big(\tilde{\mathbf{U}}_1^H\mathbf{D}_B\tilde{\mathbf{U}}_1\big)\right)\right)\\
&\geq \prod_{i=1}^{k}\left(1+\lambda_{A,k-i+1}\,\lambda_{B,n-k+i}\right)\\
&= \prod_{i=1}^{n}\left(1+\lambda_{A,n-i+1}\,\lambda_{B,i}\right)
\end{aligned}$$

where $\lambda_i(\cdot)$ denotes the ith eigenvalue in decreasing order. In the first inequality, we have used the inequality [Mar79, 9.G.3.a]

$$\det(\mathbf{A}+\mathbf{B}) \geq \prod_{i=1}^{n}\left(\lambda_i(\mathbf{A})+\lambda_i(\mathbf{B})\right)$$

with equality verified for $\tilde{\mathbf{U}}_1^H\mathbf{D}_B\tilde{\mathbf{U}}_1$ diagonal, *i.e.*, when $\tilde{\mathbf{U}}_1$ is a permutation matrix. In the second inequality, we have used the Poincaré Separation theorem [Mag99, p. 209]

$$\lambda_{B,i} \geq \lambda_i\big(\tilde{\mathbf{U}}_1^H\mathbf{D}_B\tilde{\mathbf{U}}_1\big) \geq \lambda_{B,i+n-k} \qquad 1 \leq i \leq k$$

with equality verified when $\tilde{\mathbf{U}}_1$ (note that $\tilde{\mathbf{U}}_1^H\tilde{\mathbf{U}}_1=\mathbf{I}_k$) selects the k smallest diagonal elements of $\mathbf{D}_B$. Since the logarithm is a monotonic increasing function, taking the logarithm on both sides completes the proof. Equality is achieved for $\tilde{\mathbf{U}}$ being a permutation matrix that sorts the diagonal elements of $\mathbf{D}_B$ in increasing order, *i.e.*, $\tilde{\mathbf{U}}=\mathbf{J}$. Note, however, that if $\mathbf{A}$ has zero eigenvalues, then $\tilde{\mathbf{U}}_2$ can be freely chosen as long as $\tilde{\mathbf{U}}$ remains unitary. ■

Lemma 4.4 *The global optimal solution to the following convex optimization problem*

$$\begin{array}{lll}\min\limits_{\mathbf{x}} & f(\mathbf{x})=-\sum_{i=1}^{n}\log\left(1+x_i\,\alpha_i\right) & \textit{with}\quad 0\leq\alpha_i\leq\alpha_{i+1}\\ s.t. & \sum_{i=1}^{n}x_i\leq P, & \\ & x_i\geq x_{i+1}\geq 0, \qquad 1\leq i\leq n-1 & \end{array} \tag{4.64}$$

is given by the uniform solution

$$x_i^\star = P/n, \qquad 1\leq i\leq n. \tag{4.65}$$

Proof. From an intuitive viewpoint, we can see that without the constraint $x_i \geq x_{i+1}$, the solution would be a water-filling, which would imply $x_i \leq x_{i+1}$. With the additional constraint, however, the solution will try to water-fill but always verifying the constraint $x_i \geq x_{i+1}$, resulting in $x_i = x_{i+1}$.

This result can be straightforwardly proved in a formal way using majorization theory [Mar79]. First, rewrite the objective function as $f(\mathbf{x}) = \sum_{i=1}^{n} g_i(x_i)$ where $g_i(x) = -\log(1 + x\,\alpha_i)$. Since $g_i'(a) \geq g_{i+1}'(b)$ whenever $a \geq b$, function $f(\mathbf{x})$ is Schur-convex [Mar79, 3.H.2]. Now, from the definition of Schur-convexity [Mar79, 3.A.1] and using the fact that the uniform solution is majorized by any other solution [Mar79, p.7], it follows that the minimum of $f(\mathbf{x})$ is attained by the uniform solution of (4.65). This result can be alternatively proved using convex optimization theory [Lue69, Boy00]. ■

Proof of Theorem 4.1. We use the relation (4.10)-(4.11) and the fact that the eigenvectors of $\mathbf{R}_H = \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H}$ are unconstrained (see (4.36)) to simplify the inner minimization of (4.37) for a given $\mathbf{Q}$:

$$\begin{aligned}\inf_{\mathbf{R}_H \in \mathcal{R}_H} \log\det(\mathbf{I} + \mathbf{Q}\mathbf{R}_H) &= \inf_{\{\lambda_{R_H,i}\} \in \mathcal{L}_{R_H}} \sum_{i=1}^{n_T} \log(1 + \lambda_{Q,i}\,\lambda_{R_H,n_T-i+1}) \\ &= \sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\,\lambda^\star_{R_H,n_T-i+1}(\{\lambda_{Q,i}\})\right)\end{aligned}$$

where Lemma 4.3 has been used (the minimizing eigenvectors are chosen according to $\mathbf{U}_{R_H} = \mathbf{U}_Q \mathbf{J}$) and $\{\lambda^\star_{R_H,i}(\{\lambda_{Q,i}\})\}$ denote the minimizing eigenvalues of $\mathbf{R}_H$ as a function of $\{\lambda_{Q,i}\}$, which depend on the particular constraint used to define the set $\mathcal{L}_{R_H}$ (in §4.4.1.2.1, some specific examples of $\mathcal{L}_{R_H}$ are considered).

The outer maximization of (4.37) can be now written as

$$\begin{array}{ll}\max\limits_{\{\lambda_{Q,i}\}} & \sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\,\lambda^\star_{R_H,n_T-i+1}(\{\lambda_{Q,i}\})\right) \\ \text{s.t.} & \sum_i \lambda_{Q,i} \leq P_T, \\ & \lambda_{Q,i} \geq \lambda_{Q,i+1} \geq 0 \qquad 1 \leq i \leq n_T - 1\end{array}$$

with solution given by $\lambda^\star_{Q,i} = P_T/n_T\ \forall i$. To show this, we just have to apply Lemma 4.4:

$$\sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\,\lambda^\star_{R_H,n_T-i+1}(\{P_T/n_T\})\right) \leq \sum_{i=1}^{n_T} \log\left(1 + P_T/n_T\,\lambda^\star_{R_H,n_T-i+1}(\{P_T/n_T\})\right)$$

and then the obvious relation

$$\inf_{\{\lambda_{R_H,i}\} \in \mathcal{L}_{R_H}} \sum_{i=1}^{n_T} \log(1 + \lambda_{Q,i}\,\lambda_{R_H,n_T-i+1}) \leq \sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\,\lambda^\star_{R_H,n_T-i+1}(\{P_T/n_T\})\right)$$

to finally obtain

$$\sum_{i=1}^{n_T} \log\left(1 + \lambda_{Q,i}\, \lambda^{\star}_{R_H, n_T - i + 1}\left(\{\lambda_{Q,i}\}\right)\right) \leq \sum_{i=1}^{n_T} \log\left(1 + P_T/n_T\, \lambda^{\star}_{R_H, n_T - i + 1}\left(\{P_T/n_T\}\right)\right).$$

Thus, the maximizing solution is given by $\lambda^{\star}_{Q,i} = P_T/n_T\ \forall i$, *i.e.*, a uniform power allocation $\mathbf{Q}^{\star} = P_T/n_T\, \mathbf{I}_{n_T}$. ■

Appendix 4.C Mixed Strategy Nash Equilibria

In this appendix, we characterize the solutions to the mixed-strategy saddle point given by (4.32).

By the saddle-point property of (4.32), it must be that

$$\mathbb{E}_{p^{\star}_{\tilde{\mathbf{H}}}} \mathbb{E}_{p^{\star}_{\mathbf{Q}}} \Psi\big(\mathbf{Q}, \tilde{\mathbf{H}}\big) \geq \mathbb{E}_{p^{\star}_{\tilde{\mathbf{H}}}} \Psi\big(\mathbb{E}_{p^{\star}_{\mathbf{Q}}} \mathbf{Q}, \tilde{\mathbf{H}}\big).$$

However, by the concavity of the logdet function [Hor85], it holds that $\mathbb{E}_{p_{\mathbf{Q}}} \Psi\big(\mathbf{Q}, \tilde{\mathbf{H}}\big) \leq \Psi\big(\mathbb{E}_{p_{\mathbf{Q}}} \mathbf{Q}, \tilde{\mathbf{H}}\big)$. Therefore, it must be the case that

$$\mathbb{E}_{p^{\star}_{\mathbf{Q}}} \Psi\big(\mathbf{Q}, \tilde{\mathbf{H}}\big) = \Psi\big(\mathbb{E}_{p^{\star}_{\mathbf{Q}}} \mathbf{Q}, \tilde{\mathbf{H}}\big) \qquad \forall\, \tilde{\mathbf{H}} : p^{\star}_{\tilde{\mathbf{H}}}\big(\tilde{\mathbf{H}}\big) > 0$$

which, by the strict concavity of the logdet function [Hor85] and (4.10), implies that

$$\tilde{\mathbf{H}}\mathbf{Q}_1\tilde{\mathbf{H}}^H = \tilde{\mathbf{H}}\mathbf{Q}_2\tilde{\mathbf{H}}^H \qquad \forall\, \mathbf{Q}_1, \mathbf{Q}_2, \tilde{\mathbf{H}} : p^{\star}_{\mathbf{Q}}\left(\mathbf{Q}_1\right) > 0,\ p^{\star}_{\mathbf{Q}}\left(\mathbf{Q}_2\right) > 0,\ p^{\star}_{\tilde{\mathbf{H}}}\big(\tilde{\mathbf{H}}\big) > 0.$$

Thus, we can conclude that $\mathbf{Q}_1 = \mathbf{Q}_2$ for $\forall\ \mathbf{Q}_1, \mathbf{Q}_2 : p^{\star}_{\mathbf{Q}}\left(\mathbf{Q}_1\right) > 0,\ p^{\star}_{\mathbf{Q}}\left(\mathbf{Q}_2\right) > 0$ (note that if the set of used $\tilde{\mathbf{H}}$'s have a common null space, by the nature of the saddle point in (4.32), all the used $\mathbf{Q}$'s will be orthogonal to that subspace). In other words, the optimal mixed strategy $p^{\star}_{\mathbf{Q}}$ reduces to a pure strategy $\mathbf{Q}^{\star}$. We can now invoke Theorem 4.1: if $\mathbf{Q}^{\star}$ was not the uniform power allocation, the set of optimal $\tilde{\mathbf{H}}$'s would align their largest singular values with the smallest eigenvalues of $\mathbf{Q}^{\star}$, and the best solution is then given by the uniform power allocation $\mathbf{Q}^{\star} = P_T/n_T\, \mathbf{I}$.

The problem now is to find a mixed strategy $p^{\star}_{\tilde{\mathbf{H}}}$ so that the saddle-point conditions are satisfied:

$$\mathbb{E}_{p_{\mathbf{Q}} p^{\star}_{\tilde{\mathbf{H}}}} \Psi\big(\mathbf{Q}, \tilde{\mathbf{H}}\big) \leq \mathbb{E}_{p^{\star}_{\tilde{\mathbf{H}}}} \Psi\big(P_T/n_T\, \mathbf{I}, \tilde{\mathbf{H}}\big) \leq \mathbb{E}_{p_{\tilde{\mathbf{H}}}} \Psi\big(P_T/n_T\, \mathbf{I}, \tilde{\mathbf{H}}\big).$$

Recall that the mixed strategy $p^{\star}_{\tilde{\mathbf{H}}}$ must satisfy $\Psi\big(P_T/n_T\, \mathbf{I}, \tilde{\mathbf{H}}_1\big) = \Psi\big(P_T/n_T\, \mathbf{I}, \tilde{\mathbf{H}}_2\big)$ for $\forall\, \tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2 : p^{\star}_{\tilde{\mathbf{H}}}\big(\tilde{\mathbf{H}}_1\big) > 0,\ p^{\star}_{\tilde{\mathbf{H}}}\big(\tilde{\mathbf{H}}_2\big) > 0$ [Osb94]. Function $\Psi\big(P_T/n_T\, \mathbf{I}, \tilde{\mathbf{H}}\big)$ only depends on $\tilde{\mathbf{H}}$ through its singular values and it is minimized by some optimal set $\{\sigma^{\star}_{\tilde{H},i}\}$. Therefore, any $p_{\tilde{\mathbf{H}}}$ that puts positive probability on channels $\tilde{\mathbf{H}}$'s with singular values given by $\{\sigma^{\star}_{\tilde{H},i}\}$ and arbitrary right and left singular vectors satisfies the right inequality of the saddle point. We just have to find the appropriate $p_{\tilde{\mathbf{H}}}$ such that the left inequality of the saddle point is also satisfied. An example of such an optimal mixed strategy $p^{\star}_{\tilde{\mathbf{H}}}$ is one that puts equal probability on each element of the set $\left\{\tilde{\mathbf{H}} = \mathbf{U}_{\tilde{H}}\boldsymbol{\Sigma}^{\star}_{\tilde{H}}\mathbf{P}\mathbf{V}^{H}_{\tilde{H}} : \mathbf{P} \in \Pi\right\}$ where $\boldsymbol{\Sigma}^{\star}_{\tilde{H}}$ contains in the main diagonal the optimum singular values $\{\sigma^{\star}_{\tilde{H},i}\}$, $\mathbf{U}_{\tilde{H}}$ and $\mathbf{V}_{\tilde{H}}$ are two arbitrary unitary matrices, and Π is the set of the $n_T!$ different permutation matrices of size $n_T \times n_T$. To check that the left inequality of the saddle point is verified just note that

$$\begin{aligned}
\mathbb{E}_{p^{\star}_{\tilde{\mathbf{H}}}}\Psi\big(\mathbf{Q},\tilde{\mathbf{H}}\big) &= \frac{1}{n_T!}\sum_{\mathbf{P}\in\Pi}\log\det\left(\mathbf{I}+\boldsymbol{\Sigma}^{\star H}_{\tilde{H}}\boldsymbol{\Sigma}^{\star}_{\tilde{H}}\mathbf{P}\mathbf{V}^{H}_{\tilde{H}}\mathbf{Q}\mathbf{V}_{\tilde{H}}\mathbf{P}^{H}\right)\\
&\le \log\det\left(\mathbf{I}+\boldsymbol{\Sigma}^{\star H}_{\tilde{H}}\boldsymbol{\Sigma}^{\star}_{\tilde{H}}\left(\frac{1}{n_T!}\sum_{\mathbf{P}\in\Pi}\left(\mathbf{P}\tilde{\mathbf{Q}}\mathbf{P}^{H}\right)\right)\right)\\
&\le \log\det\left(\mathbf{I}+\boldsymbol{\Sigma}^{\star H}_{\tilde{H}}\boldsymbol{\Sigma}^{\star}_{\tilde{H}}\,\mathrm{diag}\left(\frac{1}{n_T!}\sum_{\mathbf{P}\in\Pi}\left(\mathbf{P}\tilde{\mathbf{Q}}\mathbf{P}^{H}\right)\right)\right)\\
&\le \log\det\left(\mathbf{I}+P_T/n_T\,\boldsymbol{\Sigma}^{\star H}_{\tilde{H}}\boldsymbol{\Sigma}^{\star}_{\tilde{H}}\right)\\
&= \mathbb{E}_{p^{\star}_{\tilde{\mathbf{H}}}}\Psi\big(P_T/n_T\,\mathbf{I},\tilde{\mathbf{H}}\big)
\end{aligned}$$

where $\tilde{\mathbf{Q}} \triangleq \mathbf{V}^{H}_{H}\mathbf{Q}\mathbf{V}_{H}$ and $\mathrm{diag}\,(\mathbf{X})$ denotes a diagonal matrix with the diagonal elements of $\mathbf{X}$. The first inequality comes from the concavity of the logdet function, the second from Hadamard's inequality (see §3.3), and the third from the fact that the diagonal elements of $\frac{1}{n_T!}\sum_{\mathbf{P}\in\Pi}\left(\mathbf{P}\tilde{\mathbf{Q}}\mathbf{P}^{H}\right)$ equal $\frac{1}{n_T}\mathrm{Tr}(\tilde{\mathbf{Q}}) = \frac{1}{n_T}\mathrm{Tr}\,(\mathbf{Q}) \le P_T/n_T$. It then follows that $\mathbb{E}_{p_{\mathbf{Q}}}\mathbb{E}_{p^{\star}_{\tilde{\mathbf{H}}}}\Psi\big(\mathbf{Q},\tilde{\mathbf{H}}\big) \le \mathbb{E}_{p^{\star}_{\tilde{\mathbf{H}}}}\Psi\big(P_T/n_T\,\mathbf{I},\tilde{\mathbf{H}}\big)$. Thus, we have characterized the uniform power allocation $\mathbf{Q}^{\star} = P_T/n_T\,\mathbf{I}$ as a mixed strategy saddle point of the strategic game.

Appendix 4.D Proof of Theorem 4.2

The rate region given by (4.57) is the intersection of a set of regions each of which is in turn defined by the intersection of $2^K - 1$ non-trivial inequalities as in (4.13). We can therefore rewrite the rate region of (4.57) as the region defined by the more restrictive of each one of the $2^K - 1$ inequalities over the set of possible channels (as was done in [Ver89]):

$$\mathcal{R}\left(\{\mathbf{Q}_k\},\mathcal{H}\right) = \left\{ (R_1,\cdots,R_K) : . \right.$$
$$\left. 0 \le \sum_{k\in\mathcal{S}} R_k \le \inf_{\{\mathbf{H}_k\}\in\mathcal{H}} \log\det\left(\mathbf{I}_{n_R} + \sum_{k\in\mathcal{S}}\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H\right), \forall\mathcal{S}\subseteq\{1,\cdots,K\} \right\} \tag{4.66}$$

Note that the capacity region of the compound vector Gaussian MAC as given by (4.58) and (4.66) is the natural counterpart of the capacity of the single-user compound vector Gaussian channel of (4.37). Similarly, expression (4.59) is the natural counterpart of (4.38).

We have to show now that the inequalities defining the rate region in (4.66) corresponding to non-uniform power distributions are always more restrictive than for the uniform power distribution, *i.e.*,

$$\inf_{\{\mathbf{H}_k\}\in\mathcal{H}} \log\det\left(\mathbf{I}_{n_R} + \sum_{k\in\mathcal{S}} \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H\right) \leq \inf_{\{\mathbf{H}_k\}\in\mathcal{H}} \log\det\left(\mathbf{I}_{n_R} + \sum_{k\in\mathcal{S}} P_k/n_k\,\mathbf{H}_k\mathbf{H}_k^H\right) \tag{4.67}$$
$$\forall\mathcal{S}\subseteq\{1,\cdots,K\}$$
$$\forall\mathbf{Q}_k : \mathrm{Tr}\left(\mathbf{Q}_k\right)\leq P_k, \mathbf{Q}_k=\mathbf{Q}_k^H\geq 0 \quad 1\leq k\leq K.$$

This has the important consequence that the worst-case rate region of the uniform power allocation contains the worst-case rate region corresponding to any other power allocation strategy, *i.e.*,

$$\mathcal{R}\left(\left\{\mathbf{Q}_k\right\},\mathcal{H}\right)\subseteq\mathcal{R}\left(\left\{P_k/n_k\,\mathbf{I}_{n_k}\right\},\mathcal{H}\right) \qquad \forall\mathbf{Q}_k : \mathrm{Tr}\left(\mathbf{Q}_k\right)\leq P_k, \mathbf{Q}_k=\mathbf{Q}_k^H\geq 0 \quad 1\leq k\leq K.$$

Therefore, the expression of the worst-case capacity region in (4.58) reduces to

$$\mathcal{C}\left(\mathcal{H}\right)=\mathcal{R}\left(\left\{P_k/n_k\,\mathbf{I}_{n_k}\right\},\mathcal{H}\right)$$

which, together with (4.66), gives the desired result of (4.59). Now that (4.58) has been rewritten as (4.59), it is clear that it is a closed and convex region.

We now focus on proving the inequalities of (4.67). We first consider a single user in the set S and show that with a uniform power distribution the boundary can never decrease. Then, we apply the same idea for the rest of the users in S. Consider the minimization of the boundary with respect to the channel $\mathbf{H}_l$ of the lth user in $\mathcal{S}$; for any given set of channels $\{\mathbf{H}_k\}_{k\neq l}$, we have

$$\begin{aligned}
\inf_{\mathbf{H}_l\in\mathcal{H}_l} \log\det\left(\mathbf{I}_{n_R} + \sum_{k\in\mathcal{S}} \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H\right) &= \inf_{\mathbf{H}_l\in\mathcal{H}_l} \log\det\left(\mathbf{R}_{n_l} + \mathbf{H}_l\mathbf{Q}_l\mathbf{H}_l^H\right)\\
&= \inf_{\mathbf{H}_l\in\mathcal{H}_l} \log\det\left(\mathbf{I}_{n_T} + \mathbf{Q}_l\mathbf{H}_l^H\mathbf{R}_{n_l}^{-1}\mathbf{H}_l\right) + \log\det\left(\mathbf{R}_{n_l}\right)\\
&\leq \inf_{\mathbf{H}_l\in\mathcal{H}_l} \log\det\left(\mathbf{I}_{n_T} + P_l/n_l\,\mathbf{H}_l^H\mathbf{R}_{n_l}^{-1}\mathbf{H}_l\right) + \log\det\left(\mathbf{R}_{n_l}\right)\\
&= \inf_{\mathbf{H}_l\in\mathcal{H}_l} \log\det\left(\mathbf{I}_{n_R} + P_l/n_l\,\mathbf{H}_l\mathbf{H}_l^H + \sum_{\substack{k\in\mathcal{S}\\ k\neq l}} \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H\right)
\end{aligned}$$

where $\mathbf{R}_{n_l} \triangleq \left(\mathbf{I}_{n_R} + \sum_{\substack{k\in\mathcal{S}\\ k\neq l}} \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H\right)$ is the interference-plus-noise covariance matrix seen by the lth user and the inequality comes from invoking Lemmas 4.3 and 4.4 as was done when proving Theorem 4.1 in Appendix 4.B for the single-user case. The previous reasoning can be sequentially applied to each of the users in the set $\mathcal{S}$ to finally obtain

$$\inf_{\{\mathbf{H}_k\}\in\mathcal{H}} \log\det\left(\mathbf{I}_{n_R} + \sum_{k\in\mathcal{S}} \mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^H\right) \leq \inf_{\{\mathbf{H}_k\}\in\mathcal{H}} \log\det\left(\mathbf{I}_{n_R} + \sum_{k\in\mathcal{S}} P_k/n_k\,\mathbf{H}_k\mathbf{H}_k^H\right).$$

Therefore, a non-uniform power allocation always has a lower (or at most equal) worst-case boundary for all inequalities defining the capacity region in (4.67). This concludes the proof. ■

Appendix 4.E Proof of Lemma 4.1

In this proof, we make use of majorization theory [Mar79]. For definitions and further details, the interested reader is referred to §3.2.

Using the following consequence of Poincaré separation theorem [Mag99, p. 211]:

$$\max_{\mathbf{X}^H\mathbf{X}=\mathbf{I}_k} \operatorname{Tr}\left(\mathbf{X}^H\mathbf{A}\mathbf{X}\right) = \sum_{i=1}^{k} \lambda_i\left(\mathbf{A}\right)$$

where $\mathbf{A}$ is a $n \times n$ Hermitian matrix, $\mathbf{X} \in \mathbb{C}^{n\times k}$ with $k \leq n$, and $\lambda_i\left(\cdot\right)$ denotes the ith eigenvalue in decreasing order, we obtain

$$\begin{aligned}
\sum_{i=1}^{k} \lambda_i \left(\sum_j \mathbf{R}_j\right) &= \max_{\mathbf{X}^H\mathbf{X}=\mathbf{I}_k} \operatorname{Tr}\left(\mathbf{X}^H\left(\sum_j \mathbf{R}_j\right)\mathbf{X}\right) \\
&\leq \sum_j \max_{\mathbf{X}^H\mathbf{X}=\mathbf{I}_k} \operatorname{Tr}\left(\mathbf{X}^H\mathbf{R}_j\mathbf{X}\right) \\
&= \sum_{i=1}^{k}\sum_j \lambda_i\left(\mathbf{R}_j\right).
\end{aligned}$$

In addition, for $k = n$, we have

$$\sum_{i=1}^{n} \lambda_i \left(\sum_j \mathbf{R}_j\right) = \operatorname{Tr}\left(\sum_j \mathbf{R}_j\right) = \sum_j \operatorname{Tr}\left(\mathbf{R}_j\right) = \sum_{i=1}^{n}\sum_j \lambda_i\left(\mathbf{R}_j\right).$$

Therefore, we have proved that the sum of the eigenvalues majorizes the eigenvalues of the sum:

$$\left(\lambda_1\left(\sum_j \mathbf{R}_j\right), \cdots, \lambda_n\left(\sum_j \mathbf{R}_j\right)\right) \prec \left(\sum_j \lambda_1\left(\mathbf{R}_j\right), \cdots, \sum_j \lambda_n\left(\mathbf{R}_j\right)\right).$$

We can now proceed as in [Mar79, 9.G.3.a] for $J = 2$. Using [Mar79, 5.A.2.c], we have

$$\left(\log\lambda_1\left(\sum_j \mathbf{R}_j\right), \cdots, \log\lambda_n\left(\sum_j \mathbf{R}_j\right)\right) \prec^w \left(\log\sum_j \lambda_1\left(\mathbf{R}_j\right), \cdots, \log\sum_j \lambda_n\left(\mathbf{R}_j\right)\right)$$

or, equivalently,

$$\begin{aligned}
&\sum_{i=k}^{n} \log\lambda_i\left(\sum_j \mathbf{R}_j\right) \geq \sum_{i=k}^{n}\log\sum_j \lambda_i\left(\mathbf{R}_j\right) \\
\iff &\log\prod_{i=k}^{n}\lambda_i\left(\sum_j \mathbf{R}_j\right) \geq \log\prod_{i=k}^{n}\sum_j \lambda_i\left(\mathbf{R}_j\right).
\end{aligned}$$

In particular, for $k = 1$,

$$\det\left(\mathbf{R}_1 + \cdots + \mathbf{R}_J\right) \geq \prod_{i=1}^{n} \left(\lambda_i\left(\mathbf{R}_1\right) + \cdots + \lambda_i\left(\mathbf{R}_J\right)\right).$$

■

Chapter 5

Joint Design of Tx-Rx Linear Processing for MIMO Channels with a Power Constraint: A Unified Framework

THE DESIGN OF A COMMUNICATION SYSTEM requires the definition of an objective function to measure the performance of the system. This chapter considers communication through MIMO channels and designs transmit-receive beamforming or linear processing (also termed linear precoder at the transmitter and linear equalizer at the receiver) to optimize the performance of the system under a variety of design criteria subject to a power constraint. A variety of interesting design criteria can be adopted depending on the specific application of interest. The common way of proceeding is to choose a specific design criterion and then design the system accordingly. This chapter generalizes all the existing results in the literature by developing a novel unifying framework that provides the optimal structure of the transmitter and receiver. With such a result, the original complicated nonconvex problem with matrix-valued variables simplifies and then the design problem can be reformulated within the powerful framework of convex optimization theory, in which a great number of interesting design criteria can be easily accommodated and efficiently solved even though closed-form expressions may not exist. From this perspective, a wide range of design criteria is analyzed and, in particular, optimum beamforming in the sense of minimizing the average bit error rate (BER) is obtained. Efficient algorithms for practical implementation are given for the considered design criteria. Numerical results from simulations are given to support the mathematical development of the problem.

5.1 Introduction

Communications over multiple-input multiple-output (MIMO) channels have recently gained considerable attention [Hon90, Yan94b, Fos96, Ral98, Sca99b]. They arise in many different scenarios such as when a bundle of twisted pairs in digital subscriber lines (DSL) is treated as a whole [Hon90], when multiple antennas are used at both sides of a wireless link [Fos96, Ral98], or simply when a frequency-selective channel is properly modeled by using, for example, transmit and receive filterbanks [Sca99b] (c.f. §2.2). In particular, MIMO channels arising from the use of multiple antennas at both the transmitter and the receiver have attracted a significant interest because they provide an important increase in capacity over single-input single-output (SISO) channels under some uncorrelation conditions [Tel95, Fos98]. The use of multiple antennas at both ends of a wireless link provides significant improvement not only in terms of spectral efficiency (multiplexing gain) [Tel95, Fos98] but also in terms of link reliability (beamforming and diversity gains) [Böl02] (c.f. §2.3).

Communication techniques for MIMO systems fall into two main categories depending on whether channel state information (CSI) is available at the transmitter (CSIT) (note that CSI at the receiver (CSIR) is in general assumed) (c.f. §2.4). With no CSIT, popular space-time coding techniques independent of the channel realization can be readily used [Fos96, Ala98, Tar98]. With perfect CSIT, the transmission can be adapted to each channel realization using signal processing techniques [Yan94b, Ral98, Sca99b]. In this chapter, we focus on the latter (*i.e.*, when CSI is available at both sides of the link) and in particular when linear processing is utilized for the sake of complexity. Note that it is also possible to consider intermediate situations such as combining space-time codes and signal processing when having partial CSI [Neg99, Jön02] (see Chapter 7 to see how to take into account channel estimation errors). For the sake of complexity and perhaps with some loss of optimality, we focus on linear processing techniques as opposed to nonlinear techniques such as maximum likelihood (ML) detection[1] or decision feedback (DF) schemes (see §2.5.1 for more details).

In order to design a communication system, it is necessary to have an objective function to measure the performance (c.f. §2.5.4). The system can be then optimized in the sense of improving the performance as given by the objective function while the transmitter is assumed to be constrained in its average transmitted power to limit the interference level of the system. A variety of criteria can be used to design transmit and receive signal processing techniques for MIMO channels. Alternatively, it is possible to formulate the problem from the opposite point of view, *i.e.*, to minimize the transmitted power subject to some Quality of Service (QoS) constraints to guarantee a certain level of performance in the communication process (this approach is treated in detail in Chapter 6).

[1] In some very specific cases, the ML detector may happen to be linear such as in orthogonal space-time block coding [Ala98, Tar99a, Gan01a].

The capacity of a channel is the fundamental bound on the maximum reliably achievable data rate. Therefore, in terms of spectral efficiency, a system should be designed to approach the capacity of the channel [Cov91b, Ral98, Sca99a]. A capacity-achieving design dictates that the channel matrix must be diagonalized and then a *water-filling* power allocation must be used on the channel eigenmodes [Cov91b, Ral98, Sca99a]. In theory, this solution has the implication that an ideal Gaussian code should be used on each channel eigenmode according to its allocated power [Cov91b]. In practice, however, each Gaussian code is substituted by a simple (and suboptimal) signal constellation and a practical (and suboptimal) coding scheme (if any). Hence, the resulting system may be far from optimum. The complexity of such a solution is still significative since each channel eigenmode requires a different combination of signal constellation and code depending on the allocated power. To further reduce the complexity, the system can be constrained to use the same constellation and code in all channel eigenmodes (possibly optimizing the set of used eigenmodes to transmit only over those with a sufficiently high gain), *i.e.*, an equal-rate transmission. Examples of this pragmatic and simple solution are found in the European standard HIPERLAN/2 [ETS01] and in the US standard IEEE 802.11 [IEE99] for wireless local area networks (WLAN) that use a multicarrier approach with the same constellation on each carrier.

Consider now that the specific signal constellations and coding schemes for all the established substreams have been selected either after some bit distribution method or simply by taking a uniform bit distribution. It is then possible to further optimize the system to improve the quality of each of the established substreams. In particular, we consider the joint design of linear processing at both ends of the link (commonly referred to as linear precoding at the transmitter and equalization at the receiver) according to a variety of criteria as we now review. In [Lee76, Sal85, Yan94b, Sca99b], the sum of the mean square error (MSE) of all channel substreams (the trace of the MSE matrix) was used as the objective to minimize. This criterion was generalized by using a weighted sum (weighted trace) in [Sam01]. In [Sca99b], a maximum signal to interference-plus-noise ratio (SINR) criterion with a zero-forcing (ZF) constraint was also considered. For these criteria, the original complicated design problem is greatly simplified because the channel turns out to be diagonalized by the transmit-receive processing. In [Yan94a], the determinant of the MSE matrix was minimized and the diagonal structure was found to be optimal as well. In [Sca02], the results were extended to the case of a peak power constraint (maximum eigenvalue constraint) with similar results.

At this point, it is important to step back for a second and realize that the problem of designing jointly the transmit and receive linear processing is an old one; in fact, it dates back to the sixties. In [Ber67] (and references therein), transmit-receive filters for frequency-selective SISO channels were jointly designed to minimize the MSE, where an iterative water-filling algorithm was found for optimum energy distribution. In [Ami84], the solution was extended to 2×2 MIMO channels. Decision feedback schemes were considered in [Kav85]. The first generalization to $N \times N$ matrix channels was obtained in [Sal85]. In [Hon92], the case was generalized to an arbitrary $M \times N$

matrix channel with correlated data symbols, colored noise, both near- and far-end crosstalk, and excess bandwidth (although a closed-form expression was not provided, an iterative solution was presented). In [Yan94b, Yan94a], joint transmit-receive filters were derived using an elegant notation for a general framework including excess bandwidth and decision feedback systems. Remarkably, the joint transmit-receive design for MIMO systems was already solved in 1976 for flat channels [Lee76]. In [And00], the flat multi-antenna MIMO case was considered providing useful insights from the point of view of beamforming.

Going back to the design of transmit and receive processing under different criteria, we remark that the channel-diagonalizing property is of paramount importance in order to be able to solve the problem.

The main interest of the diagonalizing structure is that it allows a *scalarization* of the problem (meaning that all matrix equations are substituted with scalar ones) with the consequent great simplification. In light of the optimality of the channel-diagonalizing structure in all the aforementioned examples (including the capacity-achieving solution), one may wonder whether the same holds for other criteria. Examples of other reasonable criteria to design a communication system are the minimization of the maximum bit error rate (BER) of the substreams, the minimization of the average BER, or the maximization of the minimum SINR of the substreams. In these cases, it is not clear whether one can assume a diagonal structure as was obtained in the previous cases. In fact, as will be shown, the diagonal structure is not optimal for these and other criteria. Of course, a simple solution is to impose such a structure to simplify the design, but doing so may be far from optimum such as in [Ong03], where the average BER (and also of the Chernoff upper bound) was minimized imposing a diagonal structure, and in [Sam01], where the minimum of the SINR's was maximized imposing a diagonal structure.

In this chapter, we consider different design criteria based on optimizing the MSE's, the SINR's, and also the BER's directly. Instead of considering each design criterion separately, we develop a unifying framework and generalize the existing results by considering two families of objective functions that embrace most reasonable criteria to design a communication system: Schur-concave and Schur-convex functions (these types of functions arise in majorization theory [Mar79]). For Schur-concave objective functions, the channel-diagonalizing structure is always optimal, whereas for Schur-convex functions, an optimal solution diagonalizes the channel only after a very specific rotation of the transmitted symbols. Once the optimal structure of the transmit-receive processing is known, the design problem simplifies and can be formulated within the powerful framework of convex optimization theory, in which a great number of interesting design criteria can be easily accommodated and efficiently solved even though closed-form expressions may not exist (see §3.1 for details on convex optimization theory). We analyze a variety of criteria and, in particular, we derive optimal beamvectors in the sense of having minimum average BER. Additional constraints on the Peak-to-Average Ratio (PAR) or on the signal dynamic range of the transmitted signal are easily included in the design within the convex optimization framework. We propose two multi-level water-filling practical solutions that perform very close to the optimal in terms of average BER with a low implementation complexity. Interestingly, the optimal solution in the sense of minimum average BER can be obtained in closed-form.

This chapter is organized as follows. Section 5.3 considers the case of single beamforming, which refers to the transmission of a single symbol through the MIMO channel by using transmit and receive beamvectors. Section 5.4 extends the results to the more general case of multiple beamforming or matrix beamforming, which refers to the simultaneous transmission of L symbols through the MIMO channel by using transmit and receive multiple beamvectors or, equivalently, transmit and receive beam-matrices. It is in this case of designing transmit and receive matrices where the main result is obtained, *i.e.*, the optimal transmit-receive structure for Schur-concave and Schur-convex objective functions. In both sections, the transmission over a single MIMO channel is considered first and then extended to the case of having multiple MIMO channels which can model, for example, a multicarrier system. Once the optimal structure of the solution is found, the problem simplifies and a variety of design criteria are considered and formulated in Section 5.5 under the powerful framework of convex optimization theory. In Section 5.6, additional constraints to control the dynamic range of the power amplifier and the PAR of the transmitted signal are included in the design. In Section 5.7, numerical results for the proposed methods are obtained from simulations using realistic channel models. Finally, in Section 5.8, a summary of the chapter is given along with the final conclusions.

The problem formulation under the framework of convex optimization theory for the single beamforming and multiple beamforming cases have been published in [Pal03b] and [Pal02a, Pal03c], respectively.

5.2 Design Criterion

In order to design the system, we consider a general design criterion based on an arbitrary objective function of any the three basic figures of merit that measure the performance of the system as described in §2.5.4: the MSE, the SINR, and the BER.

The objective function is an indicator of how well the system performs. As an example, if two MIMO systems are identical except in one of the substreams for which one of the systems outperforms the other, any reasonable function should properly reflect this difference. Therefore, it suffices to consider only these *reasonable functions*.[2] Mathematically, this is equivalent to saying that the objective function must be monotonic in each of its arguments while having the rest fixed. To be more specific, a reasonable function of the MSE's or of the BER's must be

[2]Given an *unreasonable* objective function, it is always possible to redefine it in a *reasonable* way so that it better reflects the system performance.

increasing in each variable and a reasonable function of the SINR's must be decreasing in each variable.

The optimum linear signal processing at the receiver was obtained in §2.5.5 as the classical LMMSE receiver or Wiener filter given by $\mathbf{a}_k = \left(\mathbf{H}_k\mathbf{b}_k\mathbf{b}_k^H\mathbf{H}_k^H + \mathbf{R}_{n_k}\right)^{-1}\mathbf{H}_k\mathbf{b}_k$ in the single beamforming case and by $\mathbf{A}_k = \left(\mathbf{H}_k\mathbf{B}_k\mathbf{B}_k^H\mathbf{H}_k^H + \mathbf{R}_{n_k}\right)^{-1}\mathbf{H}_k\mathbf{B}_k$ in the multiple beamforming case (see (2.43) and (2.48)). As shown in §2.5.5, the Wiener filter is optimal in the sense that each of the MSE's is minimized, each of the SINR's is maximized, and each of the BER's is minimized. Therefore, it only remains to obtain the optimum linear signal processing at the transmitter.

It is extremely important to remark that any objective function of the SINR's can be alternatively expressed as a function of the MSE's by means of the relation in (2.62) $\text{SINR}_{k,i} = \text{MSE}_{k,i}^{-1} - 1$. Similarly, any objective function of the BER's can be expressed as a function of the SINR's by using the $\mathcal{Q}$-function as in (2.38) and (2.40). Thus, it suffices to consider objective functions of the MSE's $f_0\left(\{\text{MSE}_{k,i}\}\right)$ increasing in each variable without loss of generality.

For the sake of notation, we define the squared whitened channel matrix $\mathbf{R}_H \triangleq \mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}$ (note that the eigenvectors and eigenvalues of $\mathbf{R}_H$ are the right singular vectors and the squared singular values, respectively, of the whitened channel $\mathbf{R}_n^{-1/2}\mathbf{H}$) with maximum eigenvalue and corresponding eigenvector given by $\lambda_{H,\max} \triangleq \lambda_{\max}\left(\mathbf{R}_H\right)$ and $\mathbf{u}_{H,\max} \triangleq \mathbf{u}_{\max}\left(\mathbf{R}_H\right)$, respectively. For multiple MIMO channels, we similarly define $\mathbf{R}_{H_k} \triangleq \mathbf{H}_k^H\mathbf{R}_{n_k}^{-1}\mathbf{H}_k$, $\lambda_{H_k,\max} \triangleq \lambda_{\max}\left(\mathbf{R}_{H_k}\right)$, and $\mathbf{u}_{H_k,\max} \triangleq \mathbf{u}_{\max}\left(\mathbf{R}_{H_k}\right)$.

5.3 Single Beamforming

In this section, the simple case of single beamforming for MIMO channels as formulated in §2.5.1.1 is considered. First, the simple case of a single MIMO channel is analyzed in §6.3.1 and then the result is extended in § 6.3.2 to the case of multiple MIMO channels (typical of multicarrier systems).

In [Ise01, Ise02, Pal03b], a variety of design criteria were considered for single beamforming transmission in multicarrier applications. A convex optimization approach was explicitly adopted in [Pal03b].

5.3.1 Single MIMO Channel

Single beamforming on a single MIMO channel is a trivial case and has a simple solution. It will serve as a reference when dealing with the more general case of having a set of parallel MIMO channels.

Consider the single MIMO channel model of (2.1) and the single beamforming approach of (2.25)-(2.26) given by $\hat{x} = \mathbf{a}^H (\mathbf{Hb}x + \mathbf{n})$. As obtained in §2.5.5, the optimal receive beamvector is the Wiener filter $\mathbf{a} = \left(\mathbf{Hbb}^H\mathbf{H}^H + \mathbf{R}_n\right)^{-1}\mathbf{Hb}$ and the resulting MSE is $\text{MSE} = \frac{1}{1+\mathbf{b}^H\mathbf{R}_H\mathbf{b}}$. The problem reduces then to obtain the optimal transmit beamvector $\mathbf{b}$ to minimize the MSE subject to the power constraint at the transmitter:

$$\begin{array}{ll} \min\limits_{\mathbf{b}} & \frac{1}{1+\mathbf{b}^H\mathbf{R}_H\mathbf{b}} \\ \text{s.t.} & \mathbf{b}^H\mathbf{b} \leq P_T. \end{array} \tag{5.1}$$

The optimal solution to this (nonconvex) optimization problem is trivially given by

$$\mathbf{b} = \sqrt{P_T}\,\mathbf{u}_{H,\max}. \tag{5.2}$$

which satisfies the power constraint with equality and has the direction of the eigenvector associated to the maximum eigenvalue of matrix $\mathbf{R}_H$ (with arbitrary phase). The resulting MSE is

$$\text{MSE} = \frac{1}{1 + P_T\,\lambda_{H,\max}} \tag{5.3}$$

and the SINR is (using (2.62))

$$\text{SINR} = P_T\,\lambda_{H,\max}. \tag{5.4}$$

This solution fully agrees with intuition. It simply means that when transmitting a single symbol through a MIMO channel, the eigenmode with highest gain should be used. In [And00], an iterative solution to obtain the transmit and receive beamvectors for a flat multi-antenna MIMO channel without full channel matrix knowledge was proposed (the main idea is based on the well-known power iteration method that iteratively computes the eigenvector associated to the maximum eigenvalue of a matrix [Gol96]).

5.3.2 Multiple MIMO Channels

The single beamforming approach for multiple MIMO channels, *i.e.*, for a set of N parallel and independent MIMO channels, is similar in essence to the single MIMO channel previously considered, although the power allocation of the optimum solution depends heavily on the specific design criterion utilized.

Consider the multiple MIMO channel model of (2.3) and the single beamforming approach of (2.27)-(2.28) given by $\hat{x}_k = \mathbf{a}_k^H (\mathbf{H}_k\mathbf{b}_k x_k + \mathbf{n}_k)$. As obtained in §2.5.5, the optimal receive beamvectors are the Wiener filters $\mathbf{a}_k = \left(\mathbf{H}_k\mathbf{b}_k\mathbf{b}_k^H\mathbf{H}_k^H + \mathbf{R}_{n_k}\right)^{-1}\mathbf{H}_k\mathbf{b}_k$ and the resulting MSE is $\text{MSE}_k = \frac{1}{1+\mathbf{b}_k^H\mathbf{R}_{H_k}\mathbf{b}_k}$. The problem then reduces to obtain the optimal transmit beamvectors $\mathbf{b}_k$'s to minimize some function f_0 of the MSE's subject to the power constraint at the transmitter:

$$\begin{array}{ll} \min\limits_{\{\mathbf{b}_k\}} & f_0\left(\left\{\frac{1}{1+\mathbf{b}_k^H\mathbf{R}_{H_k}\mathbf{b}_k}\right\}_{k=1}^N\right) \\ \text{s.t.} & \sum_{k=1}^N \mathbf{b}_k^H\mathbf{b}_k \leq P_T. \end{array} \tag{5.5}$$

In this case, the optimal solution is again obtained when each $\mathbf{b}_k$ has the direction of the eigenvector associated to the maximum eigenvalue of matrix $\mathbf{R}_{H_k}$ (with arbitrary phase). However, the scaling factor associated to each $\mathbf{b}_k$ will depend on the particular objective function utilized (this is the subject of §5.5). The solution can be written as

$$\mathbf{b}_k = \sqrt{z_k}\ \mathbf{u}_{H_k,\max} \qquad 1 \leq k \leq N \tag{5.6}$$

where $z_k = \mathbf{b}_k^H\mathbf{b}_k$ is the power allocated to the kth MIMO channel still to be determined and constrained by $\sum_k z_k \leq P_T$ (recall that the optimal z_k's depend on the specific objective function and is the subject of §5.5). The resulting MSE's are

$$\text{MSE}_k = \frac{1}{1 + z_k\ \lambda_{H_k,\max}} \qquad 1 \leq k \leq N \tag{5.7}$$

and the SINR's are (using (2.62))

$$\text{SINR}_k = z_k\ \lambda_{H_k,\max} \qquad 1 \leq k \leq N. \tag{5.8}$$

Note that each MSE_k is a convex decreasing function of z_k and that each SINR_k is a linear function of z_k.

5.4 Multiple Beamforming

This section extends the results of §5.3 to the more general case of multiple beamforming or matrix beamforming for MIMO channels as formulated in §2.5.1.2.

The joint transmit-receive matrix design is in general a complicated and nonconvex problem. As previously mentioned, for some specific design criteria the original complicated problem is greatly simplified because the channel turns out to be diagonalized by the transmit-receive processing, which allows a *scalarization* of the problem (meaning that all matrix equations are substituted with scalar ones). Examples are the minimization of the (weighted) sum of the MSE's of all channel spatial substreams [Yan94b, Sca99b, Sam01], the minimization of the determinant of the MSE matrix [Yan94a], and the maximization of the mutual information [Cov91b, Ral98, Sca99a]. For other design criteria (such as the minimization of the average/maximum BER or the maximization of the minimum SINR), however, it is not known *a priori* whether the channel-diagonalizing structure is optimal as well. In the following, we generalize these results by developing a unified framework based on considering two families of functions: Schur-concave and Schur-convex functions.

It is important to recall here the difference between established substreams and channel eigenmodes (c.f. §2.3.3). The established substreams are the scalar channels that are generated over the matrix channel which may or may not be parallel and orthogonal; in other words, it refers to the number of simultaneous symbols transmitted. The channel eigenmodes are the parallel and orthogonal scalar channels existing within the matrix channel with gain given by the channel eigenvalues. In general, the number of substreams is different to the number of channel eigenmodes (preferably smaller, although not necessarily). When a diagonal structure is used for transmission, each substream is established through a channel eigenmode and both concepts coincide.

First, in §5.4.1, we deal with the case of a single MIMO channel, for which the main result of this chapter is obtained. In particular, Theorem 5.1 develops a unified framework and obtains the optimal transmit-receive structure for any Schur-concave or Schur-convex objective function, generalizing the existing results in the literature. In §5.4.2, we then extend the results to the case of multiple MIMO channels (typical of multicarrier systems). The results in this section were obtained in [Pal03c] and [Pal02a].

5.4.1 Single MIMO Channel

Consider the single MIMO channel model of (2.1) and the matrix processing model of (2.29)-(2.30) given by $\hat{\mathbf{x}} = \mathbf{A}^H (\mathbf{HBx} + \mathbf{n})$. As obtained in §2.5.5, the optimal receive matrix is the Wiener filter $\mathbf{A} = \left(\mathbf{HBB}^H\mathbf{H}^H + \mathbf{R}_n\right)^{-1}\mathbf{HB}$ and the resulting MSE matrix is $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}$. The problem reduces then to obtain the optimal transmit matrix $\mathbf{B}$ to minimize some function f_0 of the MSE's (diagonal elements of the MSE matrix) subject to the power constraint at the transmitter:

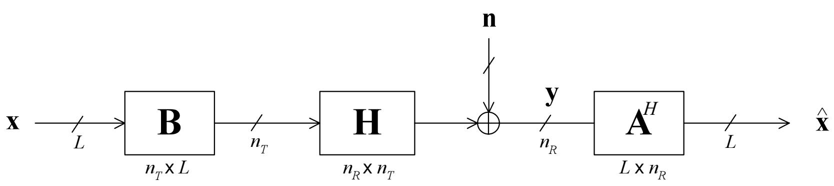
$$\begin{array}{ll} \min\limits_{\mathbf{B}} & f_0\left(\left\{\left[\left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right]_{ii}\right\}_{i=1}^{L}\right) \\ \text{s.t.} & \operatorname{Tr}\left(\mathbf{BB}^H\right) \leq P_T. \end{array} \tag{5.9}$$

Such a constrained nonconvex optimization problem is nontrivial and requires previous simplification.

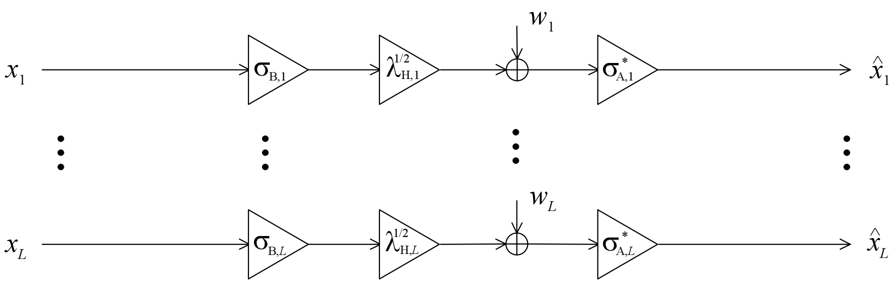
We first solve the problem in an optimal way in §5.4.1.1 by developing a unified framework, which constitutes the main result of the chapter, and then consider a simple suboptimal approach in §5.4.1.2 based on imposing a diagonality constraint on the MSE matrix.

5.4.1.1 Optimum Solution

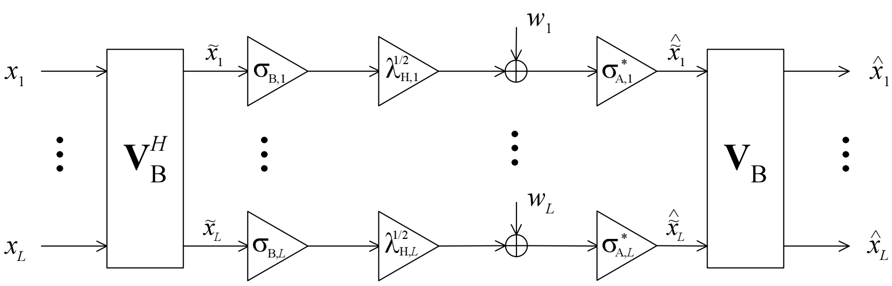
To simplify the design of the system, we now show that for Schur-concave and Schur-convex objective functions [Mar79] (see §3.2), the problem can be *scalarized*, meaning that the complicated



(a) Original system



(b) Fully diagonalized system



(c) Diagonalized (up to a rotation) system

Figure 5.1: Original matrix system, fully diagonalized system, and diagonalized (up to a rotation) system.

matrix function $\left[\left(\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right]_{ii}$ can be simplified to a set of simple scalar expressions without matrices involved. In particular, for Schur-concave functions the system is fully diagonalized, whereas for Schur-convex functions it is diagonalized up to a rotation matrix (see Figure 5.1). This is the main result of this chapter and is formally stated in the following theorem.

Theorem 5.1 *Consider the following constrained optimization problem:*

$$\begin{array}{ll} \min\limits_{\mathbf{B}} & f_0\left(\mathbf{d}(\mathbf{E}\left(\mathbf{B}\right))\right) \\ s.t. & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T \end{array}$$

where matrix $\mathbf{B} \in \mathbb{C}^{n_T \times L}$ *is the optimization variable,* $\mathbf{d}(\mathbf{E}\left(\mathbf{B}\right))$ *is the vector of diagonal elements* *of the MSE matrix* $\mathbf{E}(\mathbf{B}) = \left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ *(the diagonal elements of* $\mathbf{E}(\mathbf{B})$ *are assumed in decreasing order w.l.o.g.),* $\mathbf{R}_H \in \mathbb{C}^{n_T \times n_T}$ *is a positive semidefinite Hermitian matrix, and* $f_0 : \mathbb{R}^L \longrightarrow \mathbb{R}$ *is an arbitrary objective function (increasing in each variable). It then follows that there is an optimal solution* $\mathbf{B}$ *of at most rank* $\check{L} \triangleq \min(L, \text{rank}(\mathbf{R}_H))$ *with the following structure:*

- *If* f_0 *is Schur-concave, then*
$$\mathbf{B} = \mathbf{U}_{H,1} \mathbf{\Sigma}_{B,1} \tag{5.10}$$
where $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times \check{L}}$ *has as columns the eigenvectors of* $\mathbf{R}_H$ *corresponding to the* $\check{L}$ *largest eigenvalues in increasing order and* $\mathbf{\Sigma}_{B,1} = [\mathbf{0} \; \text{diag}(\{\sigma_{B,i}\})] \in \mathbb{C}^{\check{L} \times L}$ *has zero elements except along the rightmost main diagonal (which can be assumed real w.l.o.g.).*

- *If* f_0 *is Schur-convex, then*
$$\mathbf{B} = \mathbf{U}_{H,1} \mathbf{\Sigma}_{B,1} \mathbf{V}_B^H \tag{5.11}$$
where $\mathbf{U}_{H,1}$ *and* $\mathbf{\Sigma}_{B,1}$ *are defined as before, and* $\mathbf{V}_B \in \mathbb{C}^{L \times L}$ *is a unitary matrix (rotation) such that* $\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ *has identical diagonal elements. This rotation can be computed using Algorithm 3.2 (reproduced from [Vis99b, Section IV-A]) or with any rotation matrix* $\mathbf{Q}$ *that satisfies* $|[\mathbf{Q}]_{ik}| = |[\mathbf{Q}]_{il}| \; \forall i,k,l$ *such as the Discrete Fourier Transform (DFT) matrix or the Hadamard matrix when the dimensions are appropriate such as a power of two (see §3.2 for more details).*

Proof. See Appendix 5.A. ■

For the single beamforming case $L = 1$, Theorem 5.1 simplifies and the diagonal structure simply means that the channel eigenmode with highest gain is used (this is indeed the result obtained in §5.3).

We now give some interesting remarks and corollaries of Theorem 5.1.

Remark 5.1 *For Schur-concave objective functions, the global communication process including pre- and post-processing* $\mathbf{A}^H \mathbf{H} \mathbf{B}$ *is fully diagonalized (see Figure 5.1(b)) as well as the MSE matrix* $\mathbf{E}$*. Note that the canonical channel* $\mathbf{H}_{can} = \mathbf{B}^H \mathbf{R}_H \mathbf{B}$ *(as defined in §2.5.3) is also fully diagonalized.*
Among the L established substreams, only $\check{L}$ *are associated to nonzero channel eigenvalues whereas the remainder* $L_0 = L - \check{L}$ *are associated to zero eigenvalues. The global communication process is*[3]
$$\hat{\mathbf{x}} = \left(\mathbf{I} + \mathbf{\Sigma}_{B,1}^H \mathbf{D}_{H,1} \mathbf{\Sigma}_{B,1}\right)^{-1} \mathbf{\Sigma}_{B,1}^H \mathbf{D}_{H,1}^{1/2} \left(\mathbf{D}_{H,1}^{1/2} \mathbf{\Sigma}_{B,1} \mathbf{x} + \mathbf{w}\right)$$
or, equivalently,
$$\hat{x}_i = \begin{cases} 0 & 1 \leq i \leq L_0 \\ \frac{\sigma_{B,(i-L_0)}^2 \lambda_{H,(i-L_0)}}{1+\sigma_{B,(i-L_0)}^2 \lambda_{H,(i-L_0)}} x_i + \frac{\sigma_{B,(i-L_0)} \lambda_{H,(i-L_0)}^{1/2}}{1+\sigma_{B,(i-L_0)}^2 \lambda_{H,(i-L_0)}} w_i & L_0 < i \leq L \end{cases}$$

[3] Note that $\mathbf{A} = \left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H + \mathbf{R}_n\right)^{-1} \mathbf{H}\mathbf{B} = \mathbf{R}_n^{-1} \mathbf{H}\mathbf{B} \left(\mathbf{I} + \mathbf{B}^H \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H}\mathbf{B}\right)^{-1}$.

where $\mathbf{D}_{H,1} = \mathrm{diag}\left(\{\lambda_{H,i}\}_{k=1}^{\check{L}}\right)$, *the* $\lambda_{H,i}$*'s are the* $\check{L}$ *largest eigenvalues of* $\mathbf{R}_H$ *in increasing order,* $\mathbf{w}$ *is a normalized equivalent white noise, and* $\sigma_{B,i}$ *is the* i*th diagonal element of the rightmost main diagonal of* $\mathbf{\Sigma}_{B,1}$*. The MSE matrix is* $\mathbf{E} = \left(\mathbf{I} + \mathbf{\Sigma}_{B,1}^H \mathbf{D}_{H,1} \mathbf{\Sigma}_{B,1}\right)^{-1}$ *and the corresponding MSE's are given by*

$$\mathrm{MSE}_i = \begin{cases} 1 & 1 \leq i \leq L_0 \\ \frac{1}{1+\sigma_{B,(i-L_0)}^2 \lambda_{H,(i-L_0)}} & L_0 < i \leq L. \end{cases} \tag{5.12}$$

Similarly, the SINR's are given using (2.62) by

$$\mathrm{SINR}_i = \begin{cases} 0 & 1 \leq i \leq L_0 \\ \sigma_{B,(i-L_0)}^2 \lambda_{H,(i-L_0)} & L_0 < i \leq L. \end{cases} \tag{5.13}$$

It is clear from (5.12) or from (5.13), that if we try to establish more substreams (equivalently, transmit more symbols) than the rank of the channel or the number of nonvanishing channel eigenmodes (i.e., if $L > \mathrm{rank}\,(\mathbf{R}_H)$ *or, equivalently,* $L_0 > 0$*), then the system fails since some substreams will have an MSE equal to* 1 *or, equivalently, a BER equal to* 0.5*. Therefore, for Schur-concave objective functions, a communication system should be designed such that* $L \leq \mathrm{rank}\,(\mathbf{R}_H)$ *in order to have an acceptable performance. In case that* $L > \mathrm{rank}\,(\mathbf{R}_H)$*, the substreams associated to zero eigenvalues can be simply ignored in the optimization methods of §5.5.*

Remark 5.2 *For Schur-convex objective functions, the global communication process including pre- and post-processing* $\mathbf{A}^H\mathbf{H}\mathbf{B}$ *is diagonalized only up to a very specific rotation of the data symbols (see Figure 5.1(c)) and the MSE matrix* $\mathbf{E}$ *is nondiagonal with equal diagonal elements (equal MSE's).*

In particular, assuming a pre-rotation of the data symbols at the transmitter $\tilde{\mathbf{x}} = \mathbf{V}_B^H \mathbf{x}$ *and a post-rotation of the estimates at the receiver* $\tilde{\hat{\mathbf{x}}} = \mathbf{V}_B^H \hat{\mathbf{x}}$ *the same diagonalizing results of Schur-concave functions apply (see Figure 5.1(c)). Since the diagonal elements of the MSE matrix* $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ *are equal whenever the appropriate rotation is included, the MSE's are identical and given by*

$$\mathrm{MSE}_i = \frac{1}{L}\,\mathrm{Tr}\,(\mathbf{E}) = \frac{1}{L}\left(L_0 + \sum_{j=1}^{\check{L}} \frac{1}{1+\sigma_{B,j}^2 \lambda_{H,j}}\right) \qquad 1 \leq i \leq L. \tag{5.14}$$

Similarly, the SINR's are given using (2.62) by

$$\mathrm{SINR}_i = \frac{L}{L_0 + \sum_{j=1}^{\check{L}} \frac{1}{1+\sigma_{B,j}^2 \lambda_{H,j}}} - 1 \qquad 1 \leq i \leq L. \tag{5.15}$$

Note that, during the design process, the rotation matrix can be initially ignored since the minimization can be based directly on the MSE expression in (5.14). The rotation can be computed at a later stage of the design as explained in Theorem 5.1. Observe that for Schur-convex functions *(unlike for Schur-concave ones), it is possible to have* $L > \operatorname{rank}(\mathbf{R}_H)$ *(equivalently,* $L_0 > 0$*) and still obtain an acceptable performance. This is because the* L *symbols are transmitted over the* $\check{L}$ *nonzero eigenvalues in a distributed way (as opposed to the parallel and independent transmission of the symbols for fully diagonalized systems); in other words, each substream is established using several channel eigenmodes.*

In both cases of Schur-concave and Schur-convex objective functions, the expressions of the MSE's of (5.12) and (5.14) have been *scalarized* in the sense that the original complicated matrix expressions have been reduced to simple scalar expressions. For Schur-concave functions, the specific power distribution among the channel eigenmodes will depend on the particular objective function f_0. Interestingly, for Schur-convex functions, the power distribution is independent of the specific choice of f_0 as formally stated in the following corollary.

Corollary 5.1 *For Schur-convex objective functions, the optimal solution of the constrained optimization problem stated in Theorem 5.1 happens to be independent of the particular objective function* f_0 *chosen and can always be found by first minimizing* $\operatorname{Tr}(\mathbf{E})$ *and then obtaining the appropriate rotation matrix as indicated in Theorem 5.1.*

Proof. The proof is straightforward since for Schur-convex functions, the MSE expression to be minimized given by (5.14) and the rotation matrix to make the diagonal elements of the MSE matrix equal are completely independent of the particular choice of f_0. ■

In other words, among the infinite solutions that minimize $\operatorname{Tr}(\mathbf{E})$, only that which yields equal diagonal elements in $\mathbf{E}$ is the optimal solution for a Schur-convex objective function.

Corollary 5.2 *If a function* f_0 *is both Schur-concave and Schur-convex (c.f. §3.2), then it is invariant with respect to post-rotations of* $\mathbf{B}$ *(i.e., it admits an arbitrary post-rotation in the solution) and vice-versa.*

Proof. See Appendix 5.B. ■

As an example, the function $\operatorname{Tr}(\mathbf{E})$ is clearly both Schur-concave and Schur-convex and consequently it is invariant with respect to post-rotations of $\mathbf{B}$, which on the other hand is obvious from the circularity of the trace (see §3.3).

Remark 5.3 *Theorem 5.1 still holds if a peak power constraint or maximum eigenvalue constraint* $\lambda_{\max}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_{peak}$ *(c.f. [Sca02]) is used instead of the average power constraint or trace constraint* $\operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T$*. Furthermore, the optimal power allocation is trivially given by* $\mathbf{\Sigma}_{B,1} = \sqrt{P_{peak}}\left[\mathbf{0}_{\check{L}\times(L-\check{L})}\ \mathbf{I}_{\check{L}\times\check{L}}\right] \in \mathbb{C}^{\check{L}\times L}$ *regardless of the specific choice of* f_0*, i.e., all the* $\check{L}$ *eigenmodes are used with the maximum peak power. The proof of Theorem 5.1 is still valid except when invoking Lemma 5.11 (which is not valid anymore). Instead, it is not difficult to show that given that* $\mathbf{B}^H\mathbf{R}_H\mathbf{B}$ *is diagonal, it is never worse to use* $\mathbf{B} = \sqrt{P_{peak}}\,\mathbf{U}_{H,1}\left[\mathbf{0}\ \mathbf{I}\right]$.

5.4.1.2 Suboptimum Solution: a Simple Approach Imposing Diagonality

At this point, it is interesting to consider a suboptimal but very simple solution to the considered problem. The simplicity of the solution comes from imposing a diagonality constraint in the MSE matrix, *i.e.*, from forcing $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ to have a diagonal structure. This in turn implies the diagonality of the canonical channel $\mathbf{H}_{\text{can}} = \mathbf{B}^H \mathbf{R}_H \mathbf{B}$ as defined in §2.5.3 and of the global channel including the pre- and post-processing $\mathbf{A}^H \mathbf{H} \mathbf{B}$ (see (2.53) and (2.54)). Imposing such a structure implies that the transmission is performed in a parallel fashion through the channel eigenmodes. In Lemma 5.1, we formally state such a simple solution.

Lemma 5.1 *Consider the following constrained optimization problem:*

$$\begin{array}{lll} \min_{\mathbf{B}} & f_0\left(\mathbf{d}\left(\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}\right)\right) & \\ s.t. & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T & \\ & \mathbf{B}^H \mathbf{R}_H \mathbf{B} & \textit{diagonal} \end{array}$$

where matrix $\mathbf{B} \in \mathbb{C}^{n_T \times L}$ *is the optimization variable,* $\mathbf{R}_H \in \mathbb{C}^{n_T \times n_T}$ *is a positive semidefinite Hermitian matrix,* $\mathbf{d}(\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1})$ *is the vector of diagonal elements of the MSE matrix* $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ *(the diagonal elements of* $\mathbf{E}$ *are assumed in decreasing order w.l.o.g.), and* $f_0 : \mathbb{R}^L \longrightarrow \mathbb{R}$ *is an arbitrary objective function (increasing in each variable). The optimal solution is of the form* $\mathbf{B} = \mathbf{U}_{H,1} \mathbf{\Sigma}_{B,1}$ *and has at most rank* $\check{L} \triangleq \min\left(L, \mathrm{rank}\left(\mathbf{R}_H\right)\right)$ *where* $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times \check{L}}$ *has as columns the eigenvectors of* $\mathbf{R}_H$ *corresponding to the* $\check{L}$ *largest eigenvalues in increasing order and* $\mathbf{\Sigma}_{B,1} = \left[\mathbf{0} \;\; \mathrm{diag}\left(\{\sigma_{B,i}\}\right)\right] \in \mathbb{C}^{\check{L} \times L}$ *has zero elements except along the rightmost main diagonal (which can be assumed real w.l.o.g.).*

Proof. See Appendix 5.C. ■

Bear in mind that the optimal solution obtained in Lemma 5.1 under the diagonality constraint need not be the optimal solution of the original problem considered in Theorem 5.1 without the diagonality constraint. For Schur-concave functions, however, the solution of Lemma 5.1 happens to be the optimal solution of the original problem in Theorem 5.1 since the diagonal structure is without loss of optimality.

As happened with Schur-concave objective functions, the global communication process including pre- and post-processing $\mathbf{A}^H \mathbf{H} \mathbf{B}$ is fully diagonalized as well as the MSE matrix $\mathbf{E}$. Therefore, the system should be designed such that $L \leq \mathrm{rank}\left(\mathbf{R}_H\right)$ in order to have an acceptable performance. The global communication process is then (assuming $L \leq \mathrm{rank}\left(\mathbf{R}_H\right)$)

$$\hat{x}_i = \frac{\sigma_{B,i}^2 \, \lambda_{H,i}}{1 + \sigma_{B,i}^2 \, \lambda_{H,i}} \, x_i + \frac{\sigma_{B,i} \, \lambda_{H,i}^{1/2}}{1 + \sigma_{B,i}^2 \, \lambda_{H,i}} \, w_i \qquad 1 \leq i \leq L.$$

The corresponding MSE's are

$$\mathrm{MSE}_i = \frac{1}{1 + \sigma_{B,i}^2 \, \lambda_{H,i}} \qquad 1 \leq i \leq L \tag{5.16}$$

and the SINR's are

$$\text{SINR}_i = \sigma_{B,i}^2 \lambda_{H,i} \qquad 1 \leq i \leq L. \tag{5.17}$$

5.4.2 Multiple MIMO Channels

This section extends the results of multiple beamforming in a single MIMO channel of §5.4.1 to the case of a set of N parallel MIMO channels. The extension is straightforward as we show next.

Consider the multiple MIMO channel model of (2.3) and the matrix processing model of (2.31)-(2.32) given by $\hat{\mathbf{x}}_k = \mathbf{A}_k^H (\mathbf{H}_k \mathbf{B}_k \mathbf{x}_k + \mathbf{n}_k)$ where L_k symbols are transmitted through the kth MIMO channel (L_k substreams). As obtained in §2.5.5, the optimal receive matrices are the Wiener filters $\mathbf{A}_k = \left(\mathbf{H}_k \mathbf{B}_k \mathbf{B}_k^H \mathbf{H}_k^H + \mathbf{R}_{n_k}\right)^{-1} \mathbf{H}_k \mathbf{B}_k$ and the resulting MSE matrices are $\mathbf{E}_k = \left(\mathbf{I} + \mathbf{B}_k^H \mathbf{R}_{H_k} \mathbf{B}_k\right)^{-1}$. The problem reduces then to obtain the optimal transmit matrices $\mathbf{B}_k$'s to minimize some function f_0 of the MSE's subject to the power constraint at the transmitter:

$$\begin{array}{ll} \displaystyle\min_{\{\mathbf{B}_k\}} & f_0\left(\left\{\left\{\left[\left(\mathbf{I} + \mathbf{B}_k^H \mathbf{R}_{H_k} \mathbf{B}_k\right)^{-1}\right]_{ii}\right\}_{i=1}^{L}\right\}_{k=1}^{N}\right) \\ \text{s.t.} & \sum_{k=1}^N \text{Tr}\left(\mathbf{B}_k \mathbf{B}_k^H\right) \leq P_T \end{array} \tag{5.18}$$

Similarly to the case of a single MIMO channel, we define $\check{L}_k \triangleq \min\left(L_k, \text{rank}\left(\mathbf{R}_{H_k}\right)\right)$ as the number of available nonzero eigenmodes to support the L_k substreams at the kth MIMO channel.

Theorem 5.1 is easily extended to the case of multiple MIMO channels as follows. For any MIMO channel k, consider the matrices corresponding to the rest of the MIMO channels $\{\mathbf{B}_l\}_{l \neq k}$ fixed, and Theorem 5.1 can be directly applied to the kth MIMO channel to show the optimal structure for $\mathbf{B}_k$. Note that it is required that f_0 is either Schur-concave or Schur-convex for each of the MIMO channels. For example, if f_0 is Schur-concave for some k then the MSE's corresponding to the kth MIMO channel will be equal but different in general to the MSE's the other MIMO channels.

5.5 Analysis of Different Design Criteria: A Convex Optimization Approach

In this section, using the optimal receive matrix given by (2.48) and the unified framework based on Schur-concave and Schur-convex functions obtained in Theorem 5.1, we systematically consider a variety of design criteria. The potential of the proposed unified framework is made evident by showing that a great number of interesting and appealing objective functions are either Schur-concave or Schur-convex and thus Theorem 5.1 can be applied to scalarize and simplify the design. The aim of this section is to express each problem in convex form, so that the well-developed body of literature on convex optimization theory [Lue69, Roc70, Boy00] can be used to obtain optimal solutions very efficiently in practice using, for example, interior-point methods (c.f. §3.1). In fact, it is possible in many cases to obtain simple closed-form solutions by means of the Karush-Kuhn-Tucker (KKT) optimality conditions that can be easily implemented in practice with low-complexity algorithms.

The following results were obtained in [Pal03c, Pal02a] for the multiple beamforming case and in [Pal03b] for the single beamforming case.

For simplicity of notation we define $z_{k,i} \triangleq \sigma^2_{B_k,i}$ and $\lambda_{k,i} \triangleq \lambda_{H_k,i}$ (k indexes the MIMO channels and i the channel eigenmodes). Recall that whenever the fully diagonal structure is used (optimal solution for Schur-concave functions and suboptimal solution for Schur-convex functions), in case that $L_k > \text{rank}\left(\mathbf{R}_{H_k}\right)$, then the $L_k - \check{L}_k$ substreams associated to zero eigenvalues are simply ignored in the optimization process.

5.5.1 Minimization of the ARITH-MSE

The minimization of the (weighted) arithmetic mean of the MSE's (ARITH-MSE) was considered in [Yan94b, Sca99b, Sam01]. We deal with the weighted version as was extended in [Sam01] under the unified framework of Theorem 5.1. The objective function is

$$f_0\left(\{\text{MSE}_{k,i}\}\right) = \sum_{k,i} \left(w_{k,i}\ \text{MSE}_{k,i}\right). \tag{5.19}$$

Lemma 5.2 *The function $f_0\left(\{x_i\}\right) = \sum_i \left(w_i\, x_i\right)$ (assuming $x_i \geq x_{i+1}$) is minimized when the weights are in increasing order $w_i \leq w_{i+1}$ and it is then a Schur-concave function.*

Proof. See Appendix 5.D. ■

By Lemma 5.2, the objective function (5.19) is Schur-concave on each MIMO channel k. Therefore, by Theorem 5.1, the diagonal structure is optimal and the MSE's are given by (5.12). The problem in convex form (the objective is convex and the constraints linear) is[4]

$$\begin{array}{ll} \min\limits_{\{z_{k,i}\}} & \sum_{k,i} w_{k,i} \frac{1}{1+\lambda_{k,i} z_{k,i}} \\ \text{s.t.} & \sum_{k,i} z_{k,i} \leq P_T, \\ & z_{k,i} \geq 0 \qquad 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k. \end{array} \tag{5.20}$$

[4]Note that it is not necessary to explicitly include the constraints corresponding to $x_i \geq x_{i+1}$ of Lemma 5.2 in the convex problem formulation of (5.20) since an optimal solution always satisfies these constraints (if $x_i < x_j$ for $i < j$, then those terms could be swapped so that $x_i \geq x_j$ is satisfied, but this would imply $\lambda_{k,i} > \lambda_{k,j}$ which cannot be optimal by Lemma 5.11).

As formally stated in Proposition 5.1, this problem can be solved very efficiently in practice with Algorithm 5.1 that obtains the water-filling solution[5]

$$z_{k,i} = \left(\mu^{-1/2}\, w_{k,i}^{1/2}\, \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}\right)^{+} \tag{5.21}$$

where $(x)^{+} \triangleq \max(0, x)$ and $\mu^{-1/2}$ is the *water-level* chosen to satisfy the power constraint with equality. Since the optimal solution (5.21) does not distinguish between the two types of indexes k and i, Algorithm 5.1 uses a single index for the sake of notation w.l.o.g.

Proposition 5.1 *The following convex optimization problem:*

$$\begin{array}{ll} \min\limits_{\{z_k\}} & \sum_{i=1}^{\check{L}} w_i \frac{1}{1+\lambda_i z_i} \\ s.t. & \sum_{i=1}^{\check{L}} z_i \leq P_T, \\ & z_i \geq 0 \qquad\qquad 1 \leq i \leq \check{L}, \end{array}$$

is optimally solved by the water-filling solution $z_i = \left(\mu^{-1/2}\, w_i^{1/2}\, \lambda_i^{-1/2} - \lambda_i^{-1}\right)^{+}$ *(it is tacitly assumed that all the* λ_i*'s are strictly positive) where the water-level* $\mu^{-1/2}$ *is chosen such that* $\sum_{i=1}^{\check{L}} z_i = P_T$.

Furthermore, the optimal water-filling solution can be efficiently obtained in practice with Algorithm 5.1 in no more than $\check{L}$ *iterations (worst-case complexity).*

Proof. See Appendix 5.F. ■

Algorithm 5.1 *Modified weighted water-filling algorithm (see Proposition 5.1).*

Input: *Number of available positive eigenvalues* $\check{L}$*, set of weights* $\{w_i\}$*, set of eigenvalues* $\{\lambda_i\}$*, and maximum power* P_T*.*

Output: *Set of allocated powers* $\{z_i\}$ *and water-level* $\mu^{-1/2}$*.*

0. *Reorder the set of pairs of weights and eigenvalues* $\{(w_i, \lambda_i)\}$ *so that the terms* $(w_i\, \lambda_i)$ *are in decreasing order (define* $w_{\check{L}+1}\lambda_{\check{L}+1} \triangleq 0$*). Set* $\tilde{L} = \check{L}$*.*
1. *Set* $\mu = w_{\tilde{L}}\, \lambda_{\tilde{L}}$ *(if* $w_{\tilde{L}}\lambda_{\tilde{L}} = w_{\tilde{L}+1}\lambda_{\tilde{L}+1}$*, then set* $\tilde{L} = \tilde{L} - 1$ *and go to step 1).*
2. *If* $\mu^{-1/2} \geq \frac{P_T + \sum_{i=1}^{\tilde{L}} \lambda_i^{-1}}{\sum_{i=1}^{\tilde{L}} w_i^{1/2} \lambda_i^{-1/2}}$*, then set* $\tilde{L} = \tilde{L} - 1$ *and go to step 1.*
   Otherwise, obtain the definitive water-level $\mu^{-1/2}$ *and allocated powers as*
   $$\mu^{-1/2} = \frac{P_T + \sum_{i=1}^{\tilde{L}} \lambda_i^{-1}}{\sum_{i=1}^{\tilde{L}} w_i^{1/2} \lambda_i^{-1/2}} \quad \text{and}$$
   $$z_i = \left(\mu^{-1/2}\, w_i^{1/2}\, \lambda_i^{-1/2} - \lambda_i^{-1}\right)^{+},$$
   undo the reordering done at step 0, and finish.

[5] To be exact, this is a modified water-filling solution as opposed to the classical water-filling solution of the form $z_k = \left(\mu^{-1} - \lambda_k^{-1}\right)^{+}$.

5.5.2 Minimization of the GEOM-MSE

The objective function corresponding to the minimization of the weighted geometric mean of the MSE's (GEOM-MSE) is

$$f_0\left(\{\text{MSE}_{k,i}\}\right) = \prod_{k,i} \left(\text{MSE}_{k,i}\right)^{w_{k,i}}. \tag{5.22}$$

Lemma 5.3 *The function $f_0\left(\{x_i\}\right) = \prod_i x_i^{w_i}$ (assuming $x_i \geq x_{i+1} > 0$) is minimized when the weights are in increasing order $w_i \leq w_{i+1}$ and it is then a Schur-concave function.*

Proof. See Appendix 5.D. ■

By Lemma 5.3, the objective function (5.22) is Schur-concave on each MIMO channel k. Therefore, by Theorem 5.1, the diagonal structure is optimal and the MSE's are given by (5.12). The problem in convex form (since the objective is log-convex, it is also convex [Boy00]) is[6]

$$\begin{array}{ll} \underset{\{z_{k,i}\}}{\min} & \prod_{k,i} \left(\frac{1}{1+\lambda_{k,i} z_{k,i}}\right)^{w_{k,i}} \\ \text{s.t.} & \sum_{k,i} z_{k,i} \leq P_T, \\ & z_{k,i} \geq 0 \qquad 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k. \end{array} \tag{5.23}$$

This problem, however, is more tractable if the logarithm of the objective is used as objective.

As formally stated in Proposition 5.2, this problem can be solved very efficiently in practice with Algorithm 5.2 that obtains the water-filling solution

$$z_{k,i} = \left(\mu^{-1}\, w_{k,i} - \lambda_{k,i}^{-1}\right)^+ \tag{5.24}$$

where μ^{-1} is the water-level chosen to satisfy the power constraint with equality. Since the optimal solution (5.24) does not distinguish between the two types of indexes k and i, Algorithm 5.2 uses a single index for simplicity w.l.o.g. Note that for the unweighted case $w_{k,i} = 1$, (5.24) becomes the classical capacity-achieving water-filling solution[7] [Cov91b, Ral98].

Proposition 5.2 *The following convex optimization problem:*

$$\begin{array}{ll} \underset{\{z_i\}}{\min} & \sum_{i=1}^{\check{L}} w_i \log\left(\frac{1}{1+\lambda_i z_i}\right) \\ s.t. & \sum_{i=1}^{\check{L}} z_i \leq P_T, \\ & z_i \geq 0 \qquad\qquad 1 \leq i \leq \check{L}, \end{array}$$

is optimally solved by the water-filling solution $z_i = \left(\mu^{-1}\, w_i - \lambda_i^{-1}\right)^+$ (it is tacitly assumed that all the λ_i's are strictly positive) where the water-level μ^{-1} is chosen such that $\sum_{i=1}^{\check{L}} z_i = P_T$.

[6] Again, it is not necessary to explicitly include the constraints corresponding to $x_i \geq x_{i+1}$ of Lemma 5.3 in the convex problem formulation of (5.23), since an optimal solution always satisfies these constraints.

[7] Under the constraint of using $\check{L}_k$ eigenmodes on the kth MIMO channel.

Furthermore, the optimal water-filling solution can be efficiently obtained in practice with Algorithm 5.2 in no more than $\check{L}$ iterations (worst-case complexity).

Proof. See Appendix 5.F. ■

Algorithm 5.2 *Classical weighted water-filling algorithm (see Proposition 5.2).*

Input: *Number of available positive eigenvalues $\check{L}$, set of weights $\{w_i\}$, set of eigenvalues $\{\lambda_i\}$, and maximum power P_T.*

Output: *Set of allocated powers $\{z_i\}$ and water-level μ^{-1}.*

0. *Reorder the set of pairs of weights and eigenmode gains $\{(w_i, \lambda_i)\}$ so that the terms $(w_i \lambda_i)$ are in decreasing order (define $w_{\check{L}+1}\lambda_{\check{L}+1} \triangleq 0$). Set $\tilde{L} = \check{L}$.*

1. *Set $\mu = w_{\tilde{L}} \lambda_{\tilde{L}}$ (if $w_{\tilde{L}}\lambda_{\tilde{L}} = w_{\tilde{L}+1}\lambda_{\tilde{L}+1}$, then set $\tilde{L} = \tilde{L} - 1$ and go to step 1).*

2. *If $\mu^{-1} \geq \frac{P_T + \sum_{i=1}^{\tilde{L}} \lambda_i^{-1}}{\sum_{i=1}^{\tilde{L}} w_i}$, then set $\tilde{L} = \tilde{L} - 1$ and go to step 1.*
   Otherwise, obtain the definitive water-level μ^{-1} and allocated powers as

$$\mu^{-1} = \frac{P_T + \sum_{i=1}^{\tilde{L}} \lambda_i^{-1}}{\sum_{i=1}^{\tilde{L}} w_i} \quad \text{and}$$
$$z_i = \left(\mu^{-1} w_i - \lambda_i^{-1}\right)^+ ,$$

   undo the reordering of step 0, and finish.

5.5.3 Minimization of the Determinant of the MSE Matrix

The minimization of the determinant of the MSE matrix was considered in [Yan94a]. We now show how this particular criterion is easily accommodated in our framework as a Schur-concave function of the diagonal elements of the MSE matrix $\mathbf{E}$. (For multiple MIMO channels with a noncooperative scheme, the same reasoning applies for each MIMO channel.)

Using the fact that $\mathbf{X} \geq \mathbf{Y} \Rightarrow |\mathbf{X}| \geq |\mathbf{Y}|$, it follows that the $|\mathbf{E}|$ is minimized for the choice of the receive matrix given by (2.48). From the expression of the $\mathbf{E}$ in (2.51), it is clear that $|\mathbf{E}|$ does not change if the transmit matrix $\mathbf{B}$ is post-multiplied by a unitary matrix (a rotation). Therefore, we can always choose a rotation matrix so that $\mathbf{E}$ is diagonal without loss of optimality (as we already knew from [Yan94a]) and then

$$|\mathbf{E}| = \prod_j \lambda_j (\mathbf{E}) = \prod_j [\mathbf{E}]_{jj} . \tag{5.25}$$

Therefore, the minimization of $|\mathbf{E}|$ is equivalent to the minimization of the (unweighted) product of the MSE's treated in §5.5.2.

5.5.4 Maximization of Mutual Information

The maximization of the mutual information can be used to obtain a capacity-achieving solution [Cov91b]:

$$\max_{\mathbf{Q}} \quad I = \log \left| \mathbf{I} + \mathbf{R}_n^{-1} \mathbf{H} \mathbf{Q} \mathbf{H}^H \right| \tag{5.26}$$

where $\mathbf{Q}$ is the transmit covariance matrix. Using the fact that $|\mathbf{I} + \mathbf{XY}| = |\mathbf{I} + \mathbf{YX}|$ and that $\mathbf{Q} = \mathbf{B}\mathbf{B}^H$ (from (2.29)), the mutual information can be written as $I = \log \left| \mathbf{I} + \mathbf{B}^H \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H} \mathbf{B} \right|$. Comparing this with (2.51), it follows that the mutual information can be expressed (see [Cio97] for detailed connections between the mutual information and the MSE matrix) as

$$I = -\log |\mathbf{E}| \tag{5.27}$$

and, therefore, the maximization of I is equivalent to the minimization of $|\mathbf{E}|$ treated in §5.5.3.

Hence, the minimization of the unweighted product of the MSE's, the minimization of the determinant of the MSE matrix, and the maximization of the mutual information are all equivalent criteria with solution given by a channel-diagonalizing structure and the classical capacity-achieving water-filling for the power allocation:

$$z_{k,i} = \left(\mu^{-1} - \lambda_{k,i}^{-1} \right)^+ . \tag{5.28}$$

5.5.5 Minimization of the MAX-MSE

In general, the overall performance (average BER) is dominated by the substream with highest MSE. It makes sense then to minimize the maximum of the MSE's (MAX-MSE) [Pal02a]. In fact, as will be seen in the numerical results of §5.7, this criterion happens to perform very well in practice in terms of average BER. The objective function is

$$f_0 \left(\{\mathrm{MSE}_{k,i}\} \right) = \max_{k,i} \left\{ \mathrm{MSE}_{k,i} \right\} . \tag{5.29}$$

Lemma 5.4 *The function $f_0 (\{x_i\}) = \max_i \{x_i\}$ is a Schur-convex function.*

Proof. See Appendix 5.D. ■

By Lemma 5.4, the objective function (5.29) is Schur-convex on each MIMO channel k. Therefore, by Theorem 5.1, the optimal solution has a nondiagonal MSE matrix $\mathbf{E}_k$.

5.5.5.1 Suboptimum Solution: A Simple Approach Imposing Diagonality

Although suboptimal, it is always interesting to solve the problem imposing a diagonal structure by using the transmit matrix $\mathbf{B}_k = \mathbf{U}_{H_k,1} \mathbf{\Sigma}_{B_k,1}$. The MSE's are given by (5.16) and the problem in convex form (the objective is linear and the constraints are all convex) is

$$\begin{array}{lll} \underset{t,\{z_{k,i}\}}{\min} & t & \\ \text{s.t.} & t \geq \frac{1}{1+\lambda_{k,i} z_{k,i}} & 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k \\ & \sum_{k,i} z_{k,i} \leq P_T, & \\ & z_{k,i} \geq 0 & 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k. \end{array} \tag{5.30}$$

It is straightforward to see that an optimal solution must satisfy all constraints on t with equality $t^{-1} = 1 + \lambda_{k,i} z_{k,i}$ $\forall k, i$ (if the constraint was not satisfied with equality for some indexes k and i, then $z_{k,i}$ could be decreased saving some power that could be redistributed among all eigenmodes, which means that t could be further reduced and therefore it was not an optimal solution). The solution has a simple closed-form expression given by

$$z_{k,i} = \lambda_{k,i}^{-1} \frac{P_T}{\sum_{l,j} \lambda_{l,j}^{-1}}. \tag{5.31}$$

The same solution is obtained if one minimizes the maximum eigenvalue of the MSE matrix (or, equivalently, if one maximizes the minimum eigenvalue of the SINR matrix defined as $\mathbf{\Gamma} \triangleq \mathbf{E}^{-1} - \mathbf{I}$) [Sca02].

5.5.5.2 Optimum Solution

Consider now the optimal solution with a nondiagonal MSE matrix $\mathbf{E}_k$ with equal diagonal elements. The MSE's are given by (5.14) which have to be minimized (scalarized problem). Recall that after minimizing the MSE's, it remains to obtain the optimal rotation matrix at each MIMO channel such that the diagonal elements of each MSE matrix $\mathbf{E}_k$ are identical. The scalarized problem in convex form (the objective is linear and the constraints are all convex) is

$$\begin{array}{lll} \underset{t,\{z_{k,i}\}}{\min} & t & \\ \text{s.t.} & t \geq \frac{1}{L_k} \left(\left(L_k - \check{L}_k\right) + \sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}} \right) & 1 \leq k \leq N, \\ & \sum_{k,i} z_{k,i} \leq P_T, & \\ & z_{k,i} \geq 0 & 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k. \end{array} \tag{5.32}$$

It is also possible to consider the weighted version of the problem by including some weights in (5.29). In such a case, although more involved, the problem can still be reformulated in convex form and, therefore, optimally solved. Before attempting to solve the nontrivial problem (5.32), let us consider two important particular cases.

For the case of single beamforming [Ise02, Pal03b] (*i.e.*, $L_k = 1$) problem (5.32) simplifies to one similar to (5.30) with solution (as (5.31))

$$z_k = \lambda_k^{-1} \frac{P_T}{\sum_l \lambda_l^{-1}}. \tag{5.33}$$

For the case of a single MIMO channel (or multiple MIMO channels with a cooperative approach), problem (5.32) simplifies to the minimization of the unweighted ARITH-MSE considered in §5.5.1 (see also Proposition 5.1) as expected from the result in Corollary 5.1.

In the general case, as formally stated in Proposition 5.3, the problem can be solved very efficiently in practice with Algorithm 5.3 that obtains the multi-level water-filling solution

$$z_{k,i} = \left(\bar{\mu}_k^{1/2}\,\lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}\right)^+ \tag{5.34}$$

where $\{\bar{\mu}_k^{1/2}\}$ are multiple water-levels chosen to satisfy the constraints on t and the power constraint all with equality.

Proposition 5.3 *The following convex optimization problem:*

$$\begin{array}{lll} \underset{t,\{z_{k,i}\}}{\min} & t & \\ \text{s.t.} & t \geq \frac{1}{L_k}\left(\left(L_k - \check{L}_k\right) + \sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}}\right) & 1 \leq k \leq N, \\ & \sum_{k,i} z_{k,i} \leq P_T, & \\ & z_{k,i} \geq 0 & 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k, \end{array}$$

is optimally solved by the multi-level water-filling solution (it is tacitly assumed that all the λ_i's are strictly positive)

$$z_{k,i} = \left(\bar{\mu}_k^{1/2}\,\lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}\right)^+$$

where the multiple water-levels $\{\bar{\mu}_k^{1/2}\}$ are chosen positive such that the constraints on t and the power constraint are satisfied with equality:

$$\begin{array}{ll} t = \frac{1}{L_k}\left(\left(L_k - \check{L}_k\right) + \sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}}\right) & 1 \leq k \leq N, \\ \sum_{k,i} z_{k,i} = P_T. & \end{array}$$

Furthermore, the optimal multi-level water-filling solution can be efficiently obtained in practice with Algorithm 5.3 in no more than $\check{L}_T \triangleq \sum_{k=1}^{N} \check{L}_k$ iterations (worst-case complexity).

Proof. See Appendix 5.F. ■

Algorithm 5.3 *Multi-level water-filling algorithm for the MAX-MSE criterion (see Proposition 5.3).*

Input: *Number of channels N, number of substreams per channel $\{L_k\}$, number of available positive eigenvalues per channel $\{\check{L}_k\}$, set of eigenvalues $\{\lambda_{k,i}\}$, and maximum power P_T.*

Output: *Set of allocated powers $\{z_{k,i}\}$ and set of water-levels $\{\bar{\mu}_k^{1/2}\}$.*

0. *Reorder the set $\{\lambda_{k,i}\}_{i=1}^{\check{L}_k}$ in decreasing order (define $\lambda_{k,\check{L}_k+1} \triangleq 0$) and set $\tilde{L}_k = \check{L}_k$ for $1 \leq k \leq N$.*

1. *Set* $t^{\max\text{-}\mathrm{lb}} = \max_{1\leq k\leq N}\left\{\frac{1}{L_k}\left(\left(L_k-\tilde{L}_k\right)+\lambda_{k,\tilde{L}_k+1}^{1/2}\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)\right\}$ *and*
   $t^{\min\text{-}\mathrm{ub}} = \min_{1\leq k\leq N}\left\{\frac{1}{L_k}\left(\left(L_k-\tilde{L}_k\right)+\lambda_{k,\tilde{L}_k}^{1/2}\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)\right\}$
   (denote the minimizing k *by* $k_{\min}$*).*

2. *If* $t^{\max\text{-}\mathrm{lb}} < t^{\min\text{-}\mathrm{ub}}$ *and* $\sum_{k=1}^{N}\frac{\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)^2}{L_k t^{\min\text{-}\mathrm{ub}}-\left(L_k-\tilde{L}_k\right)} < P_T+\sum_{k=1}^{N}\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1}$, *then accept the hypothesis and go to step 3.*
   Otherwise reject the hypothesis, set $\tilde{L}_{k_{\min}} = \tilde{L}_{k_{\min}}-1$, *and go to step 1.*

3. *Obtain the definitive* t, *water-levels, and allocated powers as*

$$t : \sum_{k=1}^{N}\frac{\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)^2}{L_k t-\left(L_k-\tilde{L}_k\right)} = P_T+\sum_{k=1}^{N}\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1},$$

(recall that $t\in\left[t^{\max\text{-}\mathrm{lb}},\ t^{\min\text{-}\mathrm{ub}}\right)$*)*

$$\bar{\mu}_k^{1/2} = \frac{\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}}{L_k t-\left(L_k-\tilde{L}_k\right)}, \text{ and}$$

$$z_{k,i} = \left(\bar{\mu}_k^{1/2}\,\lambda_{k,i}^{-1/2}-\lambda_{k,i}^{-1}\right)^+,$$

undo the reordering done at step 0, and finish.

5.5.6 Maximization of the ARITH-SINR

The objective function to be minimized for the maximization of the (weighted) arithmetic mean of the SINR's (ARITH-SINR) is

$$\tilde{f}_0\left(\{\mathrm{SINR}_{k,i}\}\right) = -\sum_{k,i}\left(w_{k,i}\ \mathrm{SINR}_{k,i}\right) \tag{5.35}$$

which can be expressed as a function of the MSE's using (2.62) as

$$f_0\left(\{\mathrm{MSE}_{k,i}\}\right) = \tilde{f}_0(\{\mathrm{MSE}_{i,k}^{-1}-1\}) = -\sum_{k,i} w_{k,i}\,(\mathrm{MSE}_{i,k}^{-1}-1). \tag{5.36}$$

Lemma 5.5 *The function* $f_0\left(\{x_i\}\right) = -\sum_i w_i\left(x_i^{-1}-1\right)$ *(assuming* $x_i\geq x_{i+1}>0$*) is minimized when the weights are in increasing order* $w_i\leq w_{i+1}$ *and it is then a Schur-concave function.*

Proof. See Appendix 5.D. ■

By Lemma 5.5, the objective function (5.36) is Schur-concave on each MIMO channel k. Therefore, by Theorem 5.1, the diagonal structure is optimal and the SINR's are given by (5.13).

The problem expressed in convex form (it is actually an LP since the objective and the constraints are all linear) is[8]

$$\begin{array}{ll} \max\limits_{\{z_{k,i}\}} & \sum_{k,i} \left(w_{k,i}\, \lambda_{k,i}\right)\; z_{k,i} \\ \text{s.t.} & \sum_{k,i} z_{k,i} \leq P_T, \\ & z_{k,i} \geq 0 \qquad 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k. \end{array} \tag{5.37}$$

The optimal solution is to allocate all the available power to the eigenmode with maximum weighted gain $(w_{k,i}\, \lambda_{k,i})$ (otherwise the objective value could be increased by transferring power from other eigenmodes to this eigenmode). Although this solution maximizes indeed the weighted sum of the SINR's, it is a terrible solution in practice due to the extremely poor spectral efficiency (only one substream would be conveying information). This criterion gives a pathological solution and should not be used.

5.5.7 Maximization of the GEOM-SINR

The objective function to be minimized for the maximization of the (weighted) geometric mean of the SINR's (GEOM-SINR) is

$$\tilde{f}_0\left(\{\text{SINR}_{k,i}\}\right) = -\prod_{k,i} \left(\text{SINR}_{k,i}\right)^{w_{k,i}} \tag{5.38}$$

which can be expressed as a function of the MSE's using (2.62) as

$$f_0\left(\{\text{MSE}_{k,i}\}\right) = \tilde{f}_0(\{\text{MSE}_{k,i}^{-1} - 1\}) = -\prod_{k,i} (\text{MSE}_{k,i}^{-1} - 1)^{w_{k,i}}. \tag{5.39}$$

Note that the maximization of the product of the SINR's is equivalent to the maximization of the sum of the SINR's expressed in dB.

Lemma 5.6 *The function* $f_0\left(\{x_i\}\right) = -\prod_i \left(x_i^{-1} - 1\right)^{w_i}$ *(assuming* $0.5 \geq x_i \geq x_{i+1} > 0$*) is minimized when the weights are in increasing order* $w_i \leq w_{i+1}$ *and it is then a Schur-concave function.*

Proof. See Appendix 5.D. ■

By Lemma 5.6, the objective function (5.39) is Schur-concave on each MIMO channel k provided that $\text{MSE}_{k,i} \leq 0.5\ \forall k, i$ (this is a mild assumption since a MSE greater than 0.5 is unreasonable for a practical communication system). Therefore, by Theorem 5.1, the diagonal

[8] Note that it is not necessary to explicitly include the constraints corresponding to $\text{SINR}_{k,i} \leq \text{SINR}_{k,i+1}$ in the convex problem formulation ($x_i \geq x_{i+1}$ in Lemma 5.5) since an optimal solution always satisfies them.

.

structure is optimal and the SINR's are given by (5.13). The problem expressed in convex form (the weighted geometric mean is a concave function[9] [Roc70, Boy00]) is[10]

$$\begin{array}{ll} \max\limits_{\{z_{k,i}\}} & \prod_{k,i} (\lambda_{k,i}\, z_{k,i})^{\tilde{w}_{k,i}} \\ \text{s.t.} & \sum_{k,i} z_{k,i} \leq P_T, \\ & z_{k,i} \geq 0, \qquad 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k, \end{array} \tag{5.40}$$

where $\tilde{w}_{k,i} = w_{k,i} / \left(\sum_{l,j} w_{l,j} \right)$ (recall that $\lambda_{k,i} > 0\ \forall k, i$, otherwise the problem has trivial solution $z_{k,i} = 0\ \forall k, i$).

For the unweighted case $w_{k,i} = 1$, the problem can be rewritten as the maximization of the geometric mean subject to the arithmetic mean:

$$\begin{array}{ll} \max\limits_{\{z_{k,i}\}} & \prod_{k,i} z_{k,i}^{1/\check{L}_T} \\ \text{s.t.} & 1/\check{L}_T \sum_{k,i} z_{k,i} \leq P_T/\check{L}_T, \\ & z_{k,i} \geq 0, \end{array} \tag{5.41}$$

where $\check{L}_T \triangleq \sum_{k=1}^{N} \check{L}_k$. From the arithmetic-geometric mean inequality $\left(\prod_k x_k\right)^{1/N} \leq \frac{1}{N} \sum_k x_k$ (with equality if and only if $x_k = x_l\ \forall k, l$) [Mag99, p.202][Hor85], it follows that the optimal solution is the uniform power allocation

$$z_{k,i} = P_T/\check{L}_T. \tag{5.42}$$

Note that the uniform power distribution is commonly used due to its simplicity, *e.g.*, [Won01].

For the general case, problem (5.40) can be easily solved by forming the Lagrangian and then solving the KKT conditions (see §3.1) obtaining the optimal solution

$$z_{k,i} = \tilde{w}_{k,i}\, P_T. \tag{5.43}$$

5.5.8 Maximization of the HARM-SINR

The maximization of the harmonic mean of the SINR's (HARM-SINR) was considered in [Ise02] for the case of single beamforming. Using the unified framework of Theorem 5.1, we can extend this result to the case of multiple beamforming. The objective function to be minimized is

$$\tilde{f}_0\left(\{\text{SINR}_{k,i}\}\right) = \sum_{k,i} \frac{1}{\text{SINR}_{k,i}} \tag{5.44}$$

[9] The concavity of the geometric mean is easily verified by showing that the Hessian matrix is positive semidefinite for positive values of the arguments. The extension to include boundary points (points with zero-valued arguments) is straightforward either by using a continuity argument to show that $f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \geq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})$ for $0 \leq \theta \leq 1$ or by considering the epigraph of the function and using [Lue69, Prop. 2.7.4].

[10] Again, it is not necessary to explicitly include the constraints corresponding to $\text{SINR}_{k,i} \leq \text{SINR}_{k,i+1}$ ($x_i \geq x_{i+1}$ in Lemma 5.6).

which can be expressed as a function of the MSE's using (2.62) as

$$f_0\left(\{\text{MSE}_{k,i}\}\right) = \sum_{k,i} \frac{\text{MSE}_{k,i}}{1 - \text{MSE}_{k,i}}. \tag{5.45}$$

Lemma 5.7 *The function $f_0\left(\{x_i\}\right) = \sum_i \frac{x_i}{1-x_i}$ (for $0 \le x_i < 1$) is a Schur-convex function.*

Proof. See Appendix 5.D. ■

By Lemma 5.7, the objective function (5.45) is Schur-convex on each MIMO channel k. Therefore, by Theorem 5.1, the optimal solution has a nondiagonal MSE matrix $\mathbf{E}_k$.

5.5.8.1 Suboptimum Solution: A Simple Approach Imposing Diagonality

Although suboptimal, let us consider the problem imposing a diagonal structure by using the transmit matrix $\mathbf{B}_k = \mathbf{U}_{H_k,1}\mathbf{\Sigma}_{B_k,1}$. The SINR's are given by (5.13) and the problem in convex form is

$$\begin{array}{lll} \underset{\{z_{k,i}\}}{\min} & \sum_{k,i} \frac{1}{\lambda_{k,i} z_{k,i}} & \\ \text{s.t.} & \sum_{k,i} z_{k,i} \le P_T, & \\ & z_{k,i} \ge 0 \qquad 1 \le k \le N, \quad 1 \le i \le \check{L}_k. & \end{array} \tag{5.46}$$

Interestingly, this problem happens to be identical in form to the optimal solution of the ARITH-MSE-ZF criterion in (5.66) for $w_{k,i} = 1$ with solution

$$z_{k,i} = \lambda_{k,i}^{-1/2} \frac{P_T}{\sum_{l,j} \lambda_{l,j}^{-1/2}}. \tag{5.47}$$

5.5.8.2 Optimum Solution

Consider now the optimal solution with a nondiagonal MSE matrix $\mathbf{E}_k$ with equal diagonal elements. The MSE's are given by (5.14) which have to be minimized (scalarized problem). Recall that, after minimizing the MSE's, it remains to obtain the optimal rotation matrix at each MIMO channel such that the diagonal elements of each error covariance matrix $\mathbf{E}_k$ are identical. The scalarized problem in convex form is

$$\begin{array}{lll} \underset{\{t_k\},\{z_{k,i}\}}{\min} & \sum_k \frac{t_k}{L_k - t_k} & \\ \text{s.t.} & L_k > t_k \ge \left(L_k - \check{L}_k\right) + \sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}} & 1 \le k \le N, \\ & \sum_{k,i} z_{k,i} \le P_T, & \\ & z_{k,i} \ge 0 & 1 \le k \le N, \quad 1 \le i \le \check{L}_k. \end{array} \tag{5.48}$$

Before attempting to solve problem (5.32), let us consider two important particular cases.

For the case of single beamforming [Ise02] (*i.e.*, $L_k = 1$), the problem simplifies to (similarly to (5.46))

$$\begin{array}{ll} \min\limits_{\{z_k\}} & \sum_k \frac{1}{\lambda_k z_k} \\ \text{s.t.} & \sum_k z_k \leq P_T, \\ & z_k \geq 0 \qquad 1 \leq k \leq N \end{array} \tag{5.49}$$

with solution

$$z_k = \lambda_k^{-1/2} \frac{P_T}{\sum_l \lambda_l^{-1/2}}. \tag{5.50}$$

For the case of a single MIMO channel (or multiple MIMO channels with a cooperative approach), problem (5.48) simplifies to the minimization of the ARITH-MSE considered in §5.5.1 (see also Proposition 5.1) as expected from the result in Corollary 5.1.

In the general case, as formally stated in Proposition 5.4, the problem can be solved very efficiently in practice with Algorithm 5.4 that obtains the multi-level water-filling solution

$$z_{k,i} = \left(\bar{\mu}_k^{1/2}\, \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}\right)^+ \tag{5.51}$$

where $\{\bar{\mu}_k^{1/2}\}$ are multiple water-levels.

Proposition 5.4 *The following convex optimization problem:*

$$\begin{array}{lll} \min\limits_{\{t_k\},\{z_{k,i}\}} & \sum_{k=1}^N \frac{t_k}{L_k - t_k} & \\ s.t. & L_k > t_k \geq \left(L_k - \check{L}_k\right) + \sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}} & 1 \leq k \leq N, \\ & \sum_{k,i} z_{k,i} \leq P_T, & \\ & z_{k,i} \geq 0 & 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k \end{array}$$

is optimally solved by the multi-level water-filling solution (it is tacitly assumed that all the λ_i's are strictly positive)

$$z_{k,i} = \left(\bar{\mu}_k^{1/2}\, \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}\right)^+$$

where the multiple water-levels $\{\bar{\mu}_k^{1/2}\}$ are chosen positive such that (ν is a positive parameter)

$$\begin{array}{ll} t_k = \left(L_k - \check{L}_k\right) + \sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}} & 1 \leq k \leq N, \\ \bar{\mu}_k^{1/2} = \nu \frac{L_k^{1/2}}{L_k - t_k}, & \\ \sum_{k,i} z_{k,i} = P_T. & \end{array}$$

Furthermore, the optimal multi-level water-filling solution can be efficiently obtained in practice with Algorithm 5.4 in no more than $\check{L}_T \triangleq \sum_{k=1}^N \check{L}_k$ iterations (worst-case complexity).

Proof. See Appendix 5.F. ■

Algorithm 5.4 *Multi-level water-filling algorithm for the HARM-SINR criterion (see Proposition 5.4).*

Input: *Number of channels N, number of substreams per channel $\{L_k\}$, number of available positive eigenvalues $\{\check{L}_k\}$, set of eigenvalues $\{\lambda_{k,i}\}$, and maximum power P_T.*

Output: *Set of allocated powers $\{z_{k,i}\}$ and set of water-levels $\{\bar{\mu}_k^{1/2}\}$.*

0. *Reorder the set $\{\lambda_{k,i}\}_{i=1}^{\check{L}_k}$ in decreasing order (define $\lambda_{k,\check{L}_k+1} \triangleq 0$) and set $\tilde{L}_k = \check{L}_k$ for $1 \leq k \leq N$.*

1. *Set* $$\nu^{\text{min-ub}} = \min_{1\leq k\leq N}\left\{\frac{1}{L_k^{1/2}}\left(\tilde{L}_k\,\lambda_{k,\tilde{L}_k+1}^{-1/2} - \sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)\right\}$$ *and* $$\nu^{\text{max-lb}} = \max_{1\leq k\leq N}\left\{\frac{1}{L_k^{1/2}}\left(\tilde{L}_k\,\lambda_{k,\tilde{L}_k}^{-1/2} - \sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)\right\}$$ *(denote the maximizing k by $k_{\max}$).*

2. *If $\nu^{\text{max-lb}} < \nu^{\text{min-ub}}$ and $\nu^{\text{max-lb}} < \frac{P_T+\sum_{k=1}^{N}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1}-\frac{1}{\tilde{L}_k}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)^2\right)}{\sum_{k=1}^{N}\frac{L_k^{1/2}}{\tilde{L}_k}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)}$, then accept the hypothesis and go to step 3.*

   Otherwise reject the hypothesis, set $\tilde{L}_{k_{\max}} = \tilde{L}_{k_{\max}} - 1$, and go to step 1.

3. *Obtain the definitive t, water-levels, and allocated powers as*

$$\nu = \frac{P_T + \sum_{k=1}^{N}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1} - \frac{1}{\tilde{L}_k}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)^2\right)}{\sum_{k=1}^{N}\frac{L_k^{1/2}}{\tilde{L}_k}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)}$$

$$\bar{\mu}_k^{1/2} = \frac{1}{\tilde{L}_k}\left(\nu\,L_k^{1/2} + \sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right), \text{ and}$$

$$z_{k,i} = \left(\bar{\mu}_k^{1/2}\,\lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}\right)^+,$$

   undo the reordering done at step 0, and finish.

5.5.9 Maximization of the PROD-(1+SINR)

Consider for a moment the following maximization:

$$\max \quad \prod_{k,i}\left(1+\text{SINR}_{k,i}\right). \tag{5.52}$$

Using the relation between the MSE and the SINR in (2.62), this maximization can be equivalently expressed as the following minimization:

$$\min \quad \prod_{k,i} \text{MSE}_{k,i} \tag{5.53}$$

which is equivalent to the unweighted MAX-MSE criterion of (5.22), to the minimization of the determinant of the MSE matrix in §5.5.3, and to the maximization of the mutual information in §5.5.4 with solution given by the capacity-achieving expression (5.28). This result is completely natural since maximizing the logarithm of (5.52) is tantamount to maximizing the mutual information $I = \sum_{k,i} \log\left(1 + \text{SINR}_{k,i}\right)$.

5.5.10 Maximization of the MIN-SINR

The objective function to be minimized for the maximization of the minimum of the SINR's (MIN-SINR) is

$$\tilde{f}_0\left(\{\text{SINR}_{k,i}\}\right) = -\min_{k,i}\left\{\text{SINR}_{k,i}\right\} \tag{5.54}$$

which can be expressed as a function of the MSE's using (2.62) as

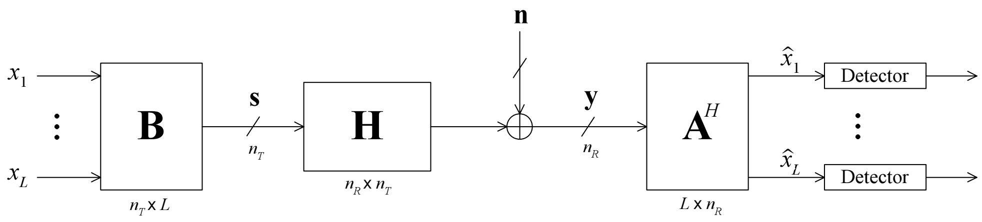
$$f_0\left(\{\text{MSE}_{k,i}\}\right) = \tilde{f}_0(\{\text{MSE}_{k,i}^{-1} - 1\}) = -\min_{k,i}\{\text{MSE}_{k,i}^{-1} - 1\}. \tag{5.55}$$

Since the SINR is a monotonic decreasing function of the MSE, the minimization of (5.55) is equivalent to the minimization of the maximum of the MSE's $\max_{k,i}\left\{\text{MSE}_{k,i}\right\}$ which was treated with detail in §5.5.5. In [Sam01], the same criterion was used imposing a channel diagonal structure.

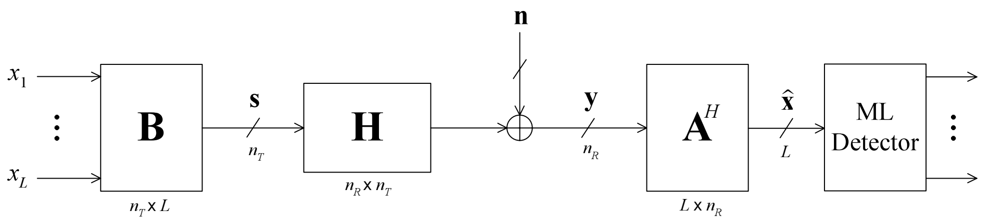
5.5.11 Minimization of the ARITH-BER

The ultimate measure of a digital communication system is the BER (c.f. § 2.5.4). In practice, MIMO communication systems use some type of coding over all the dimensions and possibly over multiple transmissions to reduce the BER (usually some orders of magnitude). The ultimate measure is then the coded BER as opposed to the uncoded BER (obtained without using any coding). However, the coded BER is strongly related to the uncoded BER (in fact, for codes based on hard decisions, both quantities are strictly related). Therefore, it suffices to focus on the uncoded BER when designing the uncoded part of a communication system.

Under the Gaussian assumption, the uncoded BER can be analytically expressed as a function of the SINR using the $\mathcal{Q}$-function (c.f. §2.5.4.4). Recall that the Gaussian assumption, which refers to the distribution of the global interference-plus-noise component, does not necessarily imply that each interfering signal is Gaussian distributed. In fact, as long as the number of



(a) Independent (parallel) detection



(b) Joint ML detection

Figure 5.2: Scheme of a vector transmission using a transmit-receive linear processing with either an independent or a joint detection of the established substreams.

interfering signals is sufficiently high, it is a reasonable assumption even when each of them is not Gaussian distributed (since the total interference contribution tends to have a Gaussian distribution as the number of interfering signals grows due to the central limit theorem).

Assuming that after the linear processing at the receiver each substream is detected independently (see Figure 5.2(a)), the BER of the MIMO system is given by the average BER of the established substreams or, equivalently, by the arithmetic mean of the BER's (ARITH-BER). In principle, a joint detection of all the substreams using an ML detector (see Figure 5.2(b)) is the optimum receiver and outperforms the previously mentioned independent detection, for which the minimization of the maximum eigenvalue of the MSE matrix[11] may be an interesting criterion [Sca02]. For complexity reasons, however, we assume an independent detection in the following.

The problem of minimizing the average BER in MIMO systems has recently been receiving a considerable attention. In [Ong03], the problem was treated in detail imposing a diagonal structure in the transmission (the approximation of the BER by the Chernoff upper bound was also considered). The minimum BER solution without imposing any structural constraint has been independently obtained in [Din02, Cha02a, Din03a] and in [Pal02a, Pal03c], where it has been shown that the optimal solution consists of a nondiagonal transmission (nondiagonal MSE matrix). In [Din03b], the minimum BER solution was used to improve the performance of a DMT transmission using different constellations.

[11] Minimizing the maximum eigenvalue of the MSE matrix $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H} \mathbf{B}\right)^{-1}$ is equivalent to maximizing the minimum eigenvalue of the SINR matrix defined as $\mathbf{\Gamma} \triangleq \mathbf{B}^H \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H} \mathbf{B}$ [Sca02].

The objective function to minimize is

$$\tilde{f}_0\left(\{\mathrm{BER}_{k,i}\}\right) = \sum_{k,i} \mathrm{BER}_{k,i} \tag{5.56}$$

which can be expressed as a function of the MSE's using (2.62) and (2.38)-(2.40) (or the Chernoff upper bound (2.39)) as

$$f_0\left(\{\mathrm{MSE}_{k,i}\}\right) = \sum_{k,i} \mathrm{BER}\left(\mathrm{MSE}_{k,i}^{-1} - 1\right). \tag{5.57}$$

Lemma 5.8 *The function* $f_0\left(\{x_i\}\right) = \sum_i \mathrm{BER}\left(x_i^{-1} - 1\right)$ *(assuming* $\theta \geq x_i > 0$*, for sufficiently small* θ *such that* $\mathrm{BER}\left(x_i^{-1} - 1\right) \leq 2 \times 10^{-2}$ $\forall i$*) is a Schur-convex function.*

Proof. See Appendix 5.D. ■

By Lemma 5.8, the objective function (5.57) is Schur-convex on each MIMO channel k (assuming the same constellation/coding on all substreams of channel k) provided that $\mathrm{BER}_{k,i} \leq 2 \times 10^{-2}$ (for BPSK and QPSK constellations, this is true for any value of the BER). Note that this is a mild assumption since a practical system generally has an uncoded BER[12] less than 2×10^{-2}. Therefore, by Theorem 5.1, we can assume that for practical purposes the optimal solution has a nondiagonal MSE matrix $\mathbf{E}_k$.

5.5.11.1 Suboptimum Solution: A Simple Approach Imposing Diagonality

Consider first a suboptimal solution obtained by imposing a diagonal structure with $\mathbf{B}_k = \mathbf{U}_{H_k,1} \mathbf{\Sigma}_{B_k,1}$ (this case was extensively treated in [Ong03]). The SINR's are given by (5.13) and the problem in convex form is (recall that the BER function is convex decreasing on the SINR as shown in §2.5.4.4) is

$$\begin{array}{ll} \min\limits_{\{z_{k,i}\}} & \sum_{k,i} \alpha_{k,i} \mathcal{Q}\left(\sqrt{\beta_{k,i} \lambda_{k,i} z_{k,i}}\right) \\ \text{s.t.} & \sum_{k,i} z_{k,i} \leq P_T, \\ & z_{k,i} \geq 0 \qquad 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k. \end{array} \tag{5.58}$$

This problem does not have a closed-form solution and one has to resort to general purpose methods such as interior-point methods (see §3.1). For completeness we give the gradient and the Hessian of the BER function and of the Chernoff approximation in Appendix 5.E.

[12] Given an uncoded BER of at most 10^{-2} and using a proper coding scheme, a coded BER with acceptable low values such as 10^{-6} can be obtained. Note, however, that it is possible to design a system with an uncoded BER higher than 10^{-2} for example with the utilization of low-density parity-check codes.

5.5.11.2 Optimum Solution

Consider now the optimal solution with a nondiagonal MSE matrix $\mathbf{E}_k$ with equal diagonal elements. The MSE's are given by (5.14) which have to be minimized (scalarized problem). Recall that after minimizing the MSE's, it remains to obtain the optimal rotation matrix at each MIMO channel such that the diagonal elements of each MSE matrix $\mathbf{E}_k$ are identical. The scalarized problem in convex form is[13]

$$\begin{array}{lll} \underset{\{t_k\},\{z_{k,i}\}}{\min} & \sum_k \alpha_k \mathcal{Q}\left(\sqrt{\beta_k\left(t_k^{-1}-1\right)}\right) & \\ \text{s.t.} & \theta \geq t_k \geq \frac{1}{L_k}\left(\left(L_k - \check{L}_k\right) + \sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}}\right) & 1 \leq k \leq N, \\ & \sum_{k,i} z_{k,i} \leq P_T, & \\ & z_{k,i} \geq 0, & 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k. \end{array} \tag{5.59}$$

Note that we have explicitly included the upper bound θ on the MSE's to guarantee the convexity of the BER function and therefore of the whole problem.

The problem in convex form when using the Chernoff approximation is similarly given by

$$\begin{array}{lll} \underset{\{t_k\},\{z_{k,i}\}}{\min} & \sum_k \alpha_k e^{\beta_k\left(t_k^{-1}-1\right)/2} & \\ \text{s.t.} & \theta \geq t_k \geq \frac{1}{L_k}\left(\left(L_k - \check{L}_k\right) + \sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}}\right) & 1 \leq k \leq N, \\ & \sum_{k,i} z_{k,i} \leq P_T, & \\ & z_{k,i} \geq 0, & 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k. \end{array} \tag{5.60}$$

For a general case with $N > 1$ and $L_k > 1$, problems (5.59) and (5.60) do not have a simple closed-form solution and one has to resort to general purpose methods such as interior-point methods (see §3.1). For completeness, we give in Appendix 5.E the gradients and the Hessians of the objective functions of (5.59) and (5.60) and of the log-barrier functions corresponding to the constraints on the t_k's: $t_k \geq \frac{1}{L_k}\left(\left(L_k - \check{L}_k\right) + \sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}}\right)$.

For the case of single beamforming (*i.e.*, $L_k = 1$), problem (5.59) simplifies to that in (5.58). For the case of a single MIMO channel (or multiple MIMO channels with a cooperative approach), problems (5.59) and (5.60) have a closed-form solution as is stated below.

Remark 5.4 *The minimization of the average BER (assuming the same constellation on all substreams) of a vector transmission through a single MIMO channel (this also includes multiple MIMO channels with a cooperative approach) can be obtained by first minimizing* $\mathrm{Tr}\,(\mathbf{E})$ *(method ARITH-MSE considered in §5.5.1) and then including the appropriate rotation matrix*

[13] We are implicitly assuming the same constellation and code on the substreams corresponding to each MIMO channel k.

$\mathbf{Q}$ as indicated in Theorem 5.1 (such as the unitary DFT matrix) to make the diagonal elements of the MSE matrix equal. The solution can be then written in closed form, following the notation of Theorem 5.1, as $\mathbf{B} = \mathbf{U}_{H,1}\mathbf{\Sigma}_{B,1}\mathbf{Q}$, where the squared-diagonal elements of $\mathbf{\Sigma}_{B,1}$, denoted by $\{z_i\}$, are given by $z_i = P_T \, \lambda_i^{-1/2} / \sum_j \lambda_j^{-1/2}$ if a ZF receiver is used (see (5.68)) and by $z_i = \left(\mu^{-1/2} \, \lambda_i^{-1/2} - \lambda_i^{-1}\right)^+$ if a MMSE receiver is used (see (5.21), for which Algorithm 5.1 can be conveniently used in practice).

The previous remark is just a particularization of Corollary 5.1.

5.5.12 Minimization of the GEOM-BER

Although in terms of BER the minimization of the arithmetic mean is the soundest criterion from a coding perspective, let us briefly consider the minimization of the geometric mean of the BER's (GEOM-BER) for completeness. The objective function is

$$\tilde{f}_0\left(\{\mathrm{BER}_{k,i}\}\right) = \prod_{k,i} \mathrm{BER}_{k,i} \tag{5.61}$$

which can be expressed as a function of the MSE's using (2.62) and (2.38)-(2.40) (or the Chernoff upper bound (2.39)) as

$$f_0\left(\{\mathrm{MSE}_{k,i}\}\right) = \prod_{k,i} \mathrm{BER}\left(\mathrm{MSE}_{k,i}^{-1} - 1\right). \tag{5.62}$$

Lemma 5.9 *The function $f_0\left(\{x_i\}\right) = \prod_i \mathrm{BER}\left(x_i^{-1} - 1\right)$ is a Schur-concave function for $\theta \geq x_i > 0$ with θ such that $\left(\frac{\partial \, \mathrm{BER}\left(x^{-1}-1\right)}{\partial x}\right)^2 \geq \mathrm{BER}\left(x^{-1} - 1\right) \frac{\partial^2 \, \mathrm{BER}\left(x^{-1}-1\right)}{\partial x^2}$ for $\theta \geq x > 0$.*

Proof. See Appendix 5.D. ■

By Lemma 5.9, the objective function (5.62) is Schur-concave on each MIMO channel k (assuming the same constellation/coding on all substreams of channel k) provided that $\mathrm{MSE}_{k,i} \leq \theta$ $\forall k, i$. This condition is trivially satisfied by the Chernoff approximation for any value of θ and by the exact BER function for high values of θ in general around 0.8 (this is a mild assumption since a MSE greater than 0.5 is unreasonable for a practical communication system). Therefore, by Theorem 5.1, we can assume that for practical purposes that the diagonal structure is optimal and the MSE's are given by (5.12). For simplicity, we consider the formulation of the problem in convex form using the Chernoff bound (it is actually an LP since the objective and the constraints are all linear):

$$\begin{array}{ll} \max\limits_{\{z_{k,i}\}} & \sum_{k,i} \left(\beta_k \lambda_{k,i}\right) z_{k,i} \\ \text{s.t.} & \sum_{k,i} z_{k,i} \leq P_T, \\ & z_{k,i} \geq 0, \qquad 1 \leq k \leq N, \quad 1 \leq i \leq \check{L}_k. \end{array} \tag{5.63}$$

Problem (5.63) is exactly identical to the one appearing in the maximization of the ARITH-SINR with solution given by allocating all the available power to the eigenmode with maximum gain ($\beta_k \lambda_{k,i}$). Although this solution maximizes indeed the product of the BER's, it is a terrible solution in practice due to the extremely poor spectral efficiency (only one substream would be conveying information). This criterion gives a pathological solution and should not be used.

5.5.13 Minimization of the MAX-BER

The objective function for the minimization of the maximum of the BER's (MAX-BER) is

$$\tilde{f}_0\left(\{\mathrm{BER}_{k,i}\}\right) = \max_{k,i} \{\mathrm{BER}_{k,i}\} \tag{5.64}$$

which can be expressed as a function of the MSE's using (2.62), (2.38) and (2.40) as

$$f_0\left(\{\mathrm{MSE}_{k,i}\}\right) = \max_{k,i} \left\{\mathrm{BER}\left(\mathrm{MSE}_{k,i}^{-1} - 1\right)\right\}. \tag{5.65}$$

The minimization of (5.65) is equivalent to the minimization of the maximum of the MSE's[14] $\max_{k,i}\{\mathrm{MSE}_{k,i}\}$ which was treated with detail in §5.5.5.

5.5.14 Including a ZF Constraint

It is interesting at this point to see how easily the ZF constraint can be imposed in the design of the transmitter. Two different approaches can be taken. One is based on using the extended MSE matrix in (2.75) to obtain the solution and then taking the asymptotic result for $\gamma \to \infty$. The other approach is based on using directly the limiting MSE matrix in (2.69). In any case, Theorem 5.1 still holds. We consider three illustrative examples, but the same approach can be applied to any other criteria.

5.5.14.1 ARITH-MSE-ZF

Consider the criterion ARITH-MSE treated in §5.5.1, but using instead the extended MSE matrix defined in (2.75) (or, equivalently, the MSE's given in (2.76)). The problem in convex form is

$$\begin{array}{ll} \displaystyle\min_{\{z_{k,i}\}} & \sum_{k,i} w_{k,i} \frac{1}{\frac{1}{\gamma} + \lambda_{k,i} z_{k,i}} \\ \text{s.t.} & \sum_{k,i} z_{k,i} \le P_T, \\ & z_{k,i} \ge 0 \qquad 1 \le k \le N, \quad 1 \le i \le \check{L}_k. \end{array} \tag{5.66}$$

with solution

$$z_{k,i} = \left(\mu^{-1/2}\, w_{k,i}^{1/2}\, \lambda_{k,i}^{-1/2} - \frac{1}{\gamma}\lambda_{k,i}^{-1}\right)^{+} \longrightarrow \mu^{-1/2}\, w_{k,i}^{1/2}\, \lambda_{k,i}^{-1/2} \tag{5.67}$$

[14] We are implicitly assuming the same constellation and code on all the substreams.

where $\mu^{-1/2} = \frac{P_T}{\sum_{k,i} w_{k,i}^{1/2} \lambda_{k,i}^{-1/2}}$. In other words, the asymptotic solution is

$$z_{k,i}^{\mathrm{ZF}} = w_{k,i}^{1/2}\ \lambda_{k,i}^{-1/2} \frac{P_T}{\sum_{l,j} w_{l,j}^{1/2} \lambda_{l,j}^{-1/2}} \tag{5.68}$$

which coincides with the suboptimal solution to the HARM-SINR criterion (5.47) in §5.5.8. Note that this criterion for the unweighted case was considered in [Bar01] and also in [Sca99b] under the name MAX-SNR/ZF.

5.5.14.2 GEOM-MSE-ZF

Consider now the criterion GEOM-MSE treated in §5.5.2, but using instead the extended MSE matrix defined in (2.75). The solution is

$$z_{k,i} = \left(\mu^{-1}\ w_{k,i} - \frac{1}{\gamma}\lambda_{k,i}^{-1}\right)^{+} \longrightarrow \mu^{-1}\ w_{k,i} \tag{5.69}$$

where $\mu^{-1} = \frac{P_T}{\sum_{k,i} w_{k,i}}$. The asymptotic solution can be rewritten as

$$z_{k,i}^{\mathrm{ZF}} = w_{k,i} \frac{P_T}{\sum_{l,j} w_{l,j}} \tag{5.70}$$

which coincides with the solution to the GEOM-SINR criterion (5.43) in §5.5.7.

5.5.14.3 MAX-MSE-ZF

Finally, consider the MAX-MSE criterion treated in §5.5.5 but using instead the extended MSE matrix defined in (2.75). The suboptimal solution is not affected by the inclusion of the ZF constraint and it is given by (5.31). The optimal solution is given by

$$z_{k,i} = \left(\bar{\mu}_k^{1/2}\ \lambda_{k,i}^{-1/2} - \frac{1}{\gamma}\lambda_{k,i}^{-1}\right)^{+} \longrightarrow \bar{\mu}_k^{1/2}\ \lambda_{k,i}^{-1/2} \tag{5.71}$$

where $\bar{\mu}_k^{1/2} = t^{-1} \sum_i \lambda_{k,i}^{-1/2}$ and $t^{-1} = \frac{P_T}{\sum_k \left(\sum_i \lambda_{k,i}^{-1/2}\right)^2}$. The asymptotic solution can be rewritten as

$$z_{k,i}^{\mathrm{ZF}} = \lambda_{k,i}^{-1/2} \sum_j \lambda_{k,j}^{-1/2} \frac{P_T}{\sum_l \left(\sum_j \lambda_{l,j}^{-1/2}\right)^2}. \tag{5.72}$$

5.6 Introducing Additional Constraints

As explained in §3.1, one of the nice properties of expressing a problem in convex form is that additional constraints can be added as long as they are convex without affecting the solvability of the problem. As an example, we now consider two additional constraints that may be interesting to impose in a multicarrier multi-antenna communication system with N carriers and n_T transmit antennas. Of course, with the additional constraints, the closed-form solutions previously obtained in §5.5 are not valid any more and have to be properly modified.

Recall that the transmitted vector signals in a multi-antenna multicarrier system are given by $\mathbf{s}_k = \mathbf{B}_k \mathbf{x}_k$ $1 \leq k \leq N$ and the transmitted sequence by the ith antenna is then $s_i(n) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} s_{k,i} \, e^{-j \frac{2\pi}{N}(k-1)(n-1)}$ $1 \leq n \leq N$. Therefore, the average transmitted power by the ith antenna is $P_i \triangleq \mathbb{E}\left[|s_i(n)|^2\right] = \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}\left[|s_{k,i}|^2\right]$.

Dynamic Range of Power Amplifier

We can easily add constraints on the dynamic range of the power amplifier at each transmit antenna element as was done in [Ben01]. Consider a Schur-concave objective function and assume for simplicity $L_k = \check{L}_k = L \ \forall k$. From the optimal structure in (5.10) $\mathbf{B}_k = \mathbf{U}_{H_k,1} \mathbf{\Sigma}_{B_k,1}$, the total average transmitted power (in units of energy per symbol period) by the ith antenna is

$$P_i = \frac{1}{N} \sum_{k=1}^{N} \sum_{l=1}^{L} \left| [\mathbf{B}_k]_{i,l} \right|^2 = \frac{1}{N} \sum_{k=1}^{N} \sum_{l=1}^{L} \sigma_{B_k,l}^2 \left| [\mathbf{U}_{H_k,1}]_{i,l} \right|^2 \tag{5.73}$$

which is linear in the variables $\{\sigma_{B_k,i}^2\}$. (For the carrier-cooperative scheme, $P_i = \frac{1}{N} \sum_{k=1}^{N} \sum_{l=1}^{NL} \left| [\mathbf{B}]_{i+(k-1)n_T,l} \right|^2$.) Therefore, the following constraints are linear

$$\alpha_i^L \leq P_i \leq \alpha_i^U \qquad 1 \leq i \leq n_T$$

where α_i^L and α_i^U are the lower and upper bounds for the ith antenna. Similarly, it is straightforward to set limits on the relative dynamic range of a single element in comparison to the total power for the whole array [Ben01]:

$$\rho_i^L \, P_{\text{array}} \leq P_i \leq \rho_i^U \, P_{\text{array}} \qquad 1 \leq i \leq n_T$$

where ρ_i^L and ρ_i^U are the relative bounds, and $P_{\text{array}} = \sum_{i=1}^{n_T} P_i$ is the total power also linear in $\{\sigma_{B_k,i}^2\}$.

Peak to Average Ratio (PAR)

One of the main practical problems that OFDM systems face is the PAR. Indeed, multicarrier signals exhibit Gaussian-like time-domain waveforms with relatively high PAR, *i.e.*, they exhibit large amplitude spikes when several frequency components add in-phase. These spikes may have a serious impact on the design complexity and feasibility of the transceiver's analog front-end (*i.e.*, high resolution of D/A-A/D[15] converters and power amplifiers with a linear behavior over a large dynamic range). In practice, the transmitted signal has to be clipped when it exceeds a certain threshold which has detrimental effects on the BER. A variety of techniques have been devised to deal with the PAR [Mes96, Tel98, Tel00]. In this section we show how the PAR can be taken

[15] D/A: Digital to Analog. A/D: Analog to Digital.

into account into the design of the beamvectors using a convex optimization framework. Note that the already existing techniques to cope with the PAR and this approach are not exclusive and can be simultaneously used.

The PAR is defined as

$$\text{PAR} \triangleq \max_{0 \le t \le T_s} \frac{A^2(t)}{\sigma^2} \tag{5.74}$$

where T_s is the symbol period, $A(t)$ is the zero-mean transmitted signal, and $\sigma^2 = \mathbb{E}\left[A^2(t)\right]$. Since the number of carriers is usually large ($N \geq 64$), $A(t)$ can be accurately modeled as a Gaussian random process (central-limit theorem) with zero mean and variance σ^2 [Mes96]. Using this assumption, the probability that the PAR exceeds certain threshold or, equivalently, the probability that the instantaneous amplitude exceeds a clipping value A_{clip} is

$$\Pr\left\{|A(t)| > A_{\text{clip}}\right\} = 2\ \mathcal{Q}\left(\frac{A_{\text{clip}}}{\sigma}\right) \tag{5.75}$$

where $\mathcal{Q}$ is the $\mathcal{Q}$-function defined as $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\lambda^2/2} d\lambda$ [Pro95]. The clipping probability of an OFDM symbol is then [Mes96]

$$P_{\text{clip}}(\sigma) = 1 - \left(1 - 2\ \mathcal{Q}\left(\frac{A_{\text{clip}}}{\sigma}\right)\right)^{2N}. \tag{5.76}$$

In other words, in order to have a clipping probability lower than P with respect to the maximum instantaneous amplitude A_{clip}, the average signal power must satisfy

$$\sigma \leq \sigma_{\text{clip}}(P) = \frac{A_{\text{clip}}}{\mathcal{Q}^{-1}\left(\frac{1-(1-P)^{1/(2N)}}{2}\right)}. \tag{5.77}$$

When using multiple antennas for transmission, the previous equation has to be satisfied for all transmit antennas. Those constraints can be easily incorporated in any of the convex designs derived in §5.5 with a Schur-concave objective function. Using (5.73) the constraint is

$$\frac{1}{N} \sum_{k=1}^{N} \sum_{l=1}^{L} \sigma_{B_k,l}^2 \left| \left[\mathbf{U}_{H_k,1}\right]_{i,l} \right|^2 \leq \sigma_{\text{clip}}^2 \qquad 1 \leq i \leq n_T \tag{5.78}$$

which is linear in the optimization variables $\{\sigma_{B_k,i}^2\}$. Such a constraint has two effects in the solution: (i) the power distribution over the carriers changes with respect to the distribution without the constraint, and (ii) the total transmitted power drops as necessary.

5.7 Simulation Results

For the numerical results, we consider a multi-antenna wireless channel. In particular, we have chosen the European standard HIPERLAN/2 for WLAN [ETS01]. It is based on the multi-carrier modulation OFDM (64 carriers are used in the simulations). We consider multiple antennas at both the transmitter and the receiver, obtaining therefore the MIMO model used

Delay (ns)	Average relative power (dB)
0	-3.3
10	-3.6
20	-3.9
30	-4.2
50	0.0
80	-0.9
110	-1.7
140	-2.6
180	-1.5
230	-3.0
280	-4.4
330	-5.9
400	-5.3
490	-7.9
600	-9.4
730	-13.2
880	-16.3
1050	-21.2

(a)

$$\mathbf{R}_{\mathrm{BS}} = \begin{bmatrix} 1.0000 & 0.4154 & 0.2057 & 0.1997 \\ 0.4154 & 1.0000 & 0.3336 & 0.3453 \\ 0.2057 & 0.3336 & 1.0000 & 0.5226 \\ 0.1997 & 0.3453 & 0.5226 & 1.0000 \end{bmatrix}$$

$$\mathbf{R}_{\mathrm{MS}} = \begin{bmatrix} 1.0000 & 0.3644 & 0.0685 & 0.3566 \\ 0.3644 & 1.0000 & 0.3245 & 0.1848 \\ 0.0685 & 0.3245 & 1.0000 & 0.3093 \\ 0.3566 & 0.1848 & 0.3093 & 1.0000 \end{bmatrix}$$

(b)

Figure 5.3: (a) Power delay profile type C for HIPERLAN/2. (b) Envelope correlation matrices at the base station (BS) and at the mobile station (MS) in the *Nokia* environment.

throughout the chapter (either multiple MIMO channels or a single MIMO channel depending on whether a carrier-noncooperative approach or a carrier-cooperative approach is used as described in §2.5.1.2). Perfect CSI is assumed at both sides of the communication link (channel estimation errors are considered in Chapter 7).

The frequency selectivity of the channel is modeled using the power delay profile type C for HIPERLAN/2 as specified in [ETS98a] (see Figure 5.3(a)), which corresponds to a typical large open space indoor environment for NLOS conditions with 150ns average r.m.s. delay spread and 1050ns maximum delay (the sampling period is 50ns) [ETS01]. The spatial correlation of the MIMO channel is modeled according to the *Nokia* model defined in [Sch01] (which corresponds to a reception hall) specified by the correlation matrices of the envelope of the channel fading at the transmit and receive side given in Figure 5.3(b) where the base station is the receiver (uplink) (see [Sch01] for details of the model). It models a large open indoor environment with two floors, which could easily illustrate a conference hall or a shopping galleria scenario. The matrix channel generated was normalized so that $\sum_n \mathbb{E}\left[| \left[\mathbf{H}(n)\right]_{ij} | \right]^2 = 1$. The SNR is defined as the transmitted power normalized with the noise variance.

For the numerical simulations, the following design criteria have been considered: ARITH-MSE, GEOM-MSE, MAX-MSE (equivalently, MIN-SINR or MAX-BER), GEOM-SINR, HARM-SINR, and ARITH-BER (benchmark). The utilization of the Chernoff upper bound instead of the exact BER function gives indistinguishable results and is therefore not presented in the simulation results. The constellations used for transmission are fixed and are not involved in the optimization

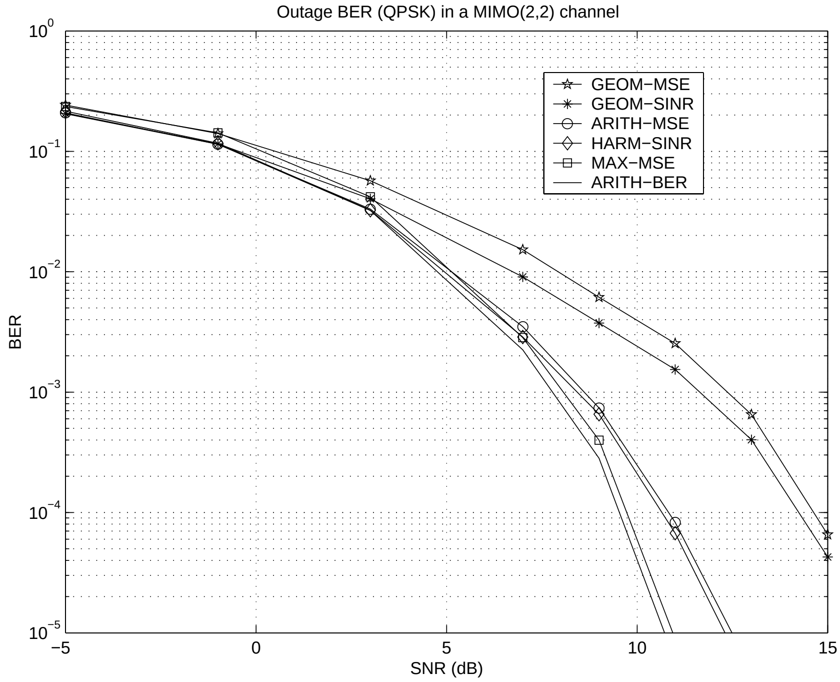


Figure 5.4: BER (at an outage probability of 5%) vs. SNR when using QPSK in a 2×2 MIMO channel with $L = 1$ for the GEOM-MSE, GEOM-SINR, ARITH-MSE, HARM-SINR, MAX-MSE, and ARITH-BER criteria (without carrier cooperation).

process, regardless of whether some channel eigenmodes are not allocated any power (due to the water-filling nature of many of the solutions) for fair comparisons. Unless otherwise specified, a carrier-noncooperative approach is considered. The performance is given in terms of outage BER (averaged over the channel substreams), *i.e.*, the BER that can be guaranteed with some probability or, equivalently, the BER that is not achieved with some small outage probability. In particular, we consider the BER with an outage probability of 5%. Note that for typical systems with delay constraints, the outage BER is a more realistic measure than the commonly used mean BER that only makes sense when the transmission coding block is long enough to reveal the long-term ergodic properties of the fading process (no delay constraints).

Single Beamforming

First we show some results when using a single channel spatial substream ($L_k = 1 \ \forall k$). In Figure 5.4, the BER is plotted vs. the SNR for a 2×2 MIMO channel using QPSK constellations. Clearly, the ARITH-BER criterion has the lowest BER because it was designed for that. The MAX-MSE criterion performs really close to the ARITH-BER and can be considered the second best criterion. The HARM-SINR and also the ARITH-MSE perform reasonably well (in fact, for values of the BER higher than 10^{-2} they outperform the MAX-MSE). The GEOM-MSE and the GEOM-SINR criteria perform really bad in terms of BER and should not be used. In Figure 5.5, the same results are shown for a 4×2 MIMO channel using 16-QAM constellations and the same observations hold.

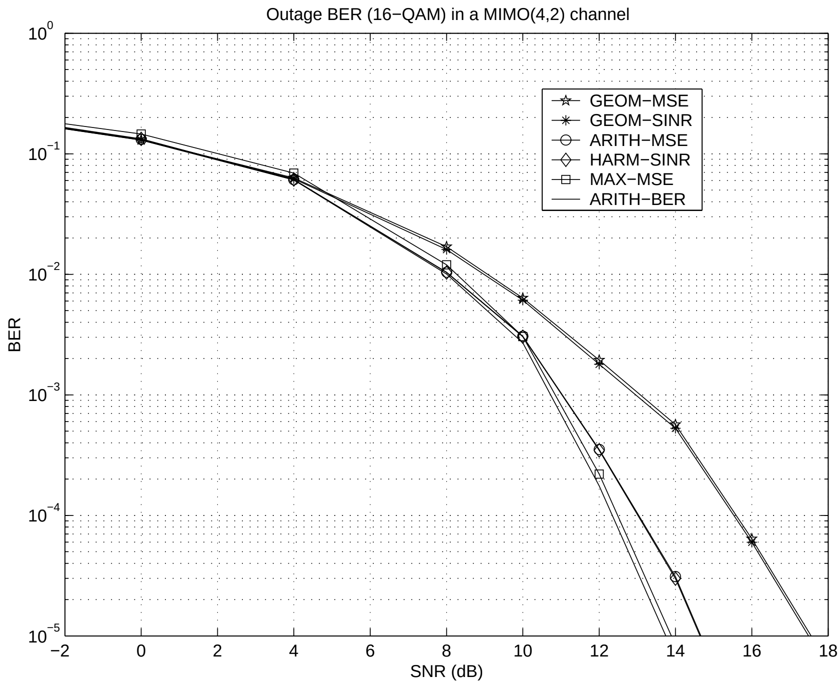


Figure 5.5: BER (at an outage probability of 5%) vs. SNR when using 16-QAM in a 4×2 MIMO channel (2 transmit and 4 receive antennas) with $L = 1$ for the GEOM-MSE, GEOM-SINR, ARITH-MSE, HARM-SINR, MAX-MSE, and ARITH-BER criteria (without carrier cooperation).

Therefore, the best criteria are (in order): ARITH-BER, MAX-MSE, HARM-SINR, and ARITH-MSE.

Including PAR Constraints

We now consider the introduction of PAR constraints as described in §5.6. We parameterize the clipping amplitude with respect to μ as $A_{\text{clip}} = \mu\sqrt{\frac{P_T}{n_T}}$ to make the results independent of the total transmitted power P_T. In Figure 5.6, the probability of clipping along with the BER (when using QPSK constellations) is shown for the ARITH-MSE criterion in a 2×2 MIMO channel both with PAR constraints ($P_{\text{clip}} = 10^{-2}$) and without them. In 5.6(a), the results are shown as a function of μ. It can be observed how the design with the PAR constraints always has a clipping probability no greater than the prespecified value 10^{-2} as expected. The BER, however, can be severely affected if a very low clipping probability is imposed due to power back-offs. From Figure 5.6(a), a choice of $\mu = 4$ seems reasonable. In Figure 5.6(b), the results are shown as a function of the SNR for $\mu = 4$. For the design with PAR constraints, the BER is slightly higher due to the additional constraint. However, the system is guaranteed to have a clipping probability of at most 10^{-2} unlike in the unconstrained case where nothing can be guaranteed. Recall that in a practical system, the final BER increases due to the clipping.

Including a ZF Constraint

We now consider the effect of introducing a ZF constraint. Clearly, since the design has additional constraints, the performance will decrease with respect to the design without the ZF

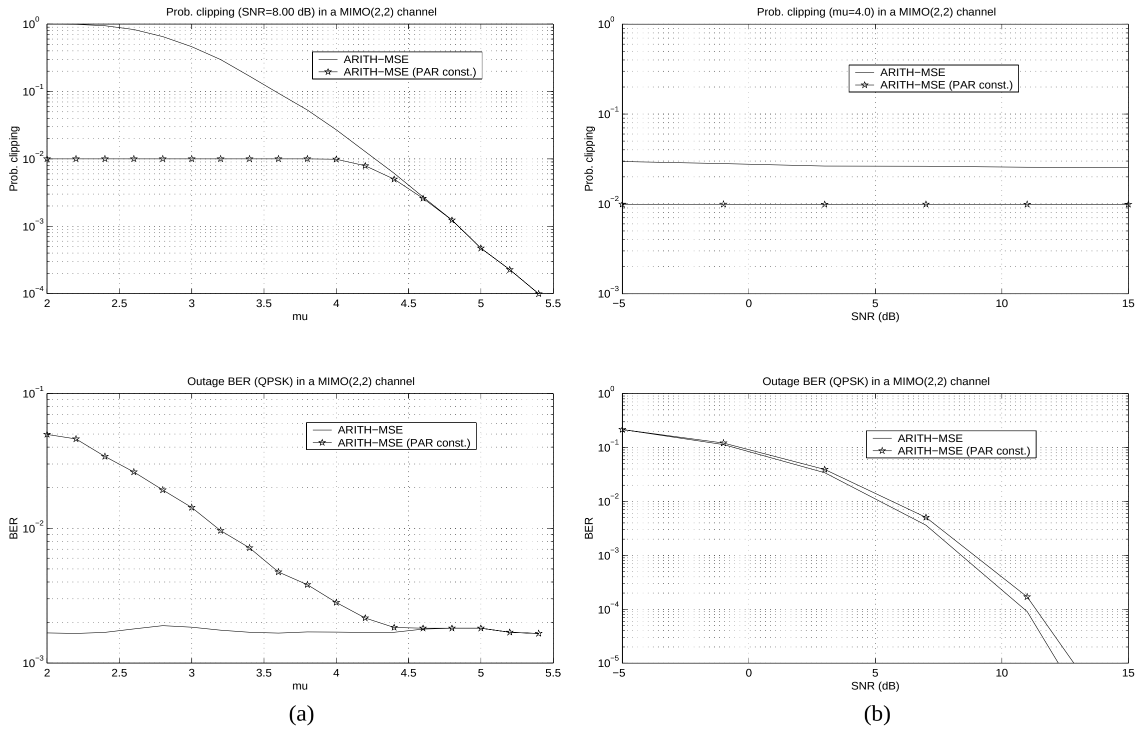


Figure 5.6: Probability of clipping and BER (at an outage probability of 5%) when using QPSK in a 2×2 MIMO channel with $L = 1$ for the ARITH-MSE criterion with and without PAR constraints (without carrier cooperation): (a) as a function of μ (for SNR=8 dB and $P_{\text{clip}} \leq 10^{-2}$), and (b) as a function of the SNR (for $\mu = 4$ and $P_{\text{clip}} \leq 10^{-2}$).

constraint. In Figure 5.7, the BER is plotted vs. the SNR for a 2×2 MIMO channel using QPSK constellations with and without the ZF constraint for the ARITH-MSE and MAX-MSE criteria. The ZF constraint introduces a penalty of about 1-2 dB for low SNR whereas for high SNR the performance with the ZF constraint converges to that without the ZF constraint as expected.

Therefore, the ZF constraint should not be used unless strictly necessary.

Multiple Beamforming

We now consider the simultaneous transmission of more than one symbol per carrier, *i.e.*, multiple beamforming (we consider $L_k = L \; \forall k$).

In Figure 5.8, the BER is plotted vs. the SNR for a 4×4 MIMO channel with $L = 2$ using QPSK constellations. In general, similar observations hold as for the single beamforming case. However, it is worth pointing out that in this case the HARM-SINR method performs much closer to the benchmark than the ARITH-MSE method.

In Figure 5.9, the BER is plotted vs. the SNR for a 4×4 MIMO channel with $L = 4$ (fully loaded system) using QPSK constellations. In this extreme situation, it can be observed the superiority of Schur-convex criteria with respect to the Schur-concave methods that have a

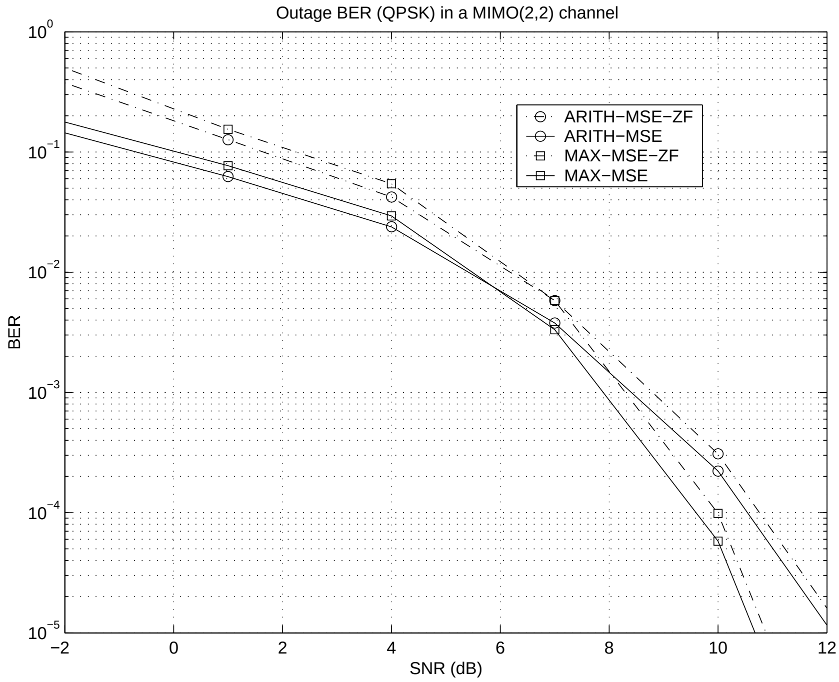


Figure 5.7: BER (at an outage probability of 5%) vs. SNR when using QPSK in a 2×2 MIMO channel with $L = 1$ for the ARITH-MSE and MAX-MSE criteria with and without the ZF constraint (without carrier cooperation).

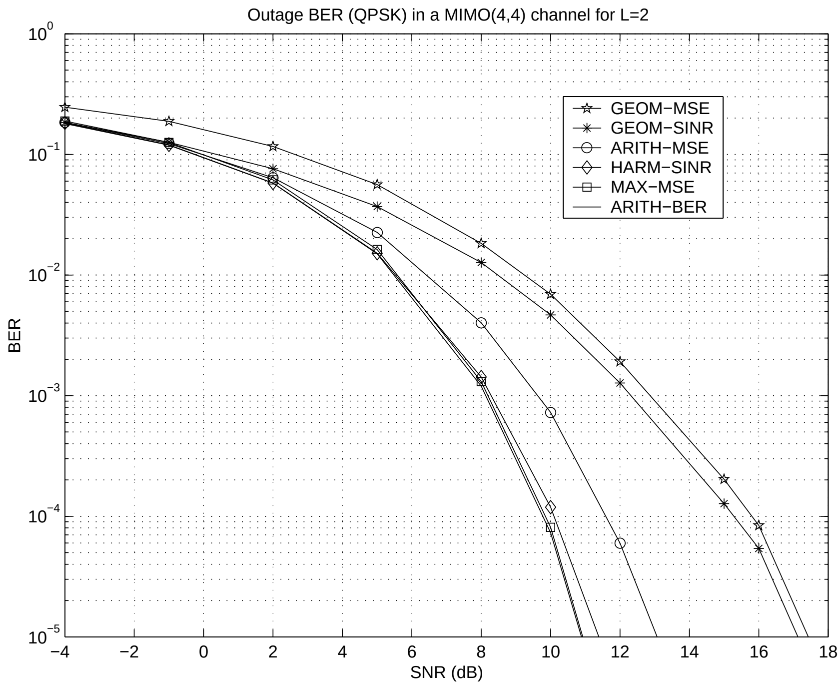


Figure 5.8: BER (at an outage probability of 5%) vs. SNR when using QPSK in a 4×4 MIMO channel with $L = 2$ for the GEOM-MSE, GEOM-SINR, ARITH-MSE, HARM-SINR, MAX-MSE, and ARITH-BER criteria (without carrier cooperation).

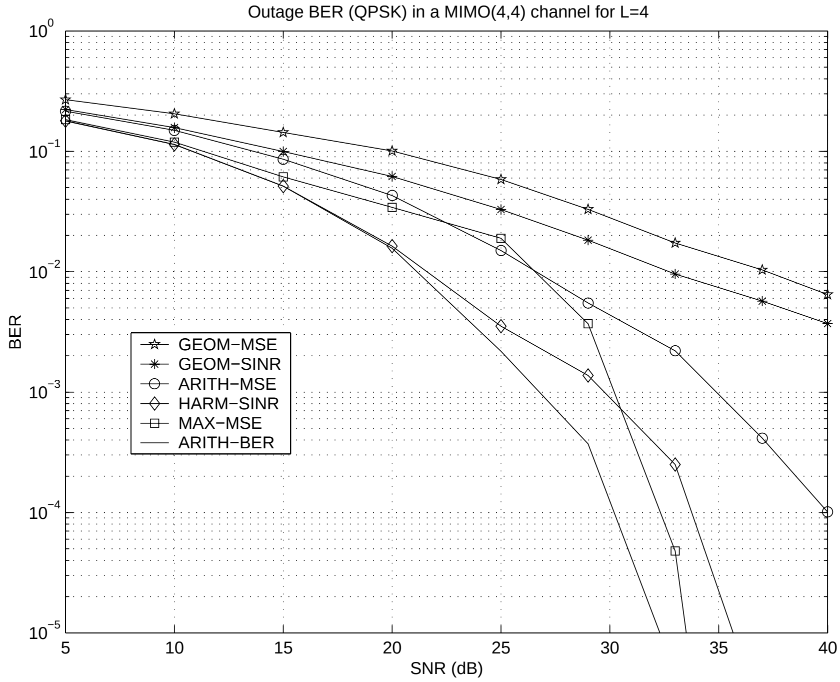


Figure 5.9: BER (at an outage probability of 5%) vs. SNR when using QPSK in a 4×4 MIMO channel with $L = 4$ (fully loaded system) for the GEOM-MSE, GEOM-SINR, ARITH-MSE, HARM-SINR, MAX-MSE, and ARITH-BER criteria (without carrier cooperation).

channel-diagonalizing structure. This is due to the fact that Schur-concave functions transmit each symbol through each channel eigenmode with the consequent waste of power if some of the eigenmodes has a low gain. On the other hand, Schur-convex functions transmit all the symbols in a distributed way and therefore have more flexibility to properly use the channel eigenmodes. It is also worth pointing out that, as can be clearly observed in Figure 5.9, the different BER vs. SNR curves have different slopes for the high SNR regime; this can be interpreted as different diversity orders corresponding to different design criteria.

Single vs. Multiple MIMO Channel Modeling (Carrier Cooperation)

We now analyze the improvement in performance when using cooperation among the MIMO channels as described in §2.5.1.2 (in this case, cooperation among the carriers) for the best methods: ARITH-MSE, HARM-SINR, MAX-MSE, and ARITH-BER. Recall that with carrier cooperation, the HARM-SINR, MAX-MSE and ARITH-BER criteria provide the same solution since they are all Schur-convex functions (see Corollary 5.1).

In Figure 5.10, the BER is plotted vs. the SNR with and without carrier cooperation for a 2×2 MIMO channel with $L = 1$ using QPSK constellations. Cooperation in this case gives an improvement of about 0.5-1dB. In Figure 5.11, the BER is plotted vs. SNR with and without carrier cooperation for a 2×2 MIMO channel with $L = 2$ (fully loaded system) using QPSK constellations. In this fully loaded system ($L = \text{rank}(\mathbf{H})$), the value of cooperation is much more significant with an improvement of about 5-10 dB (note that since the ARITH-MSE method has a

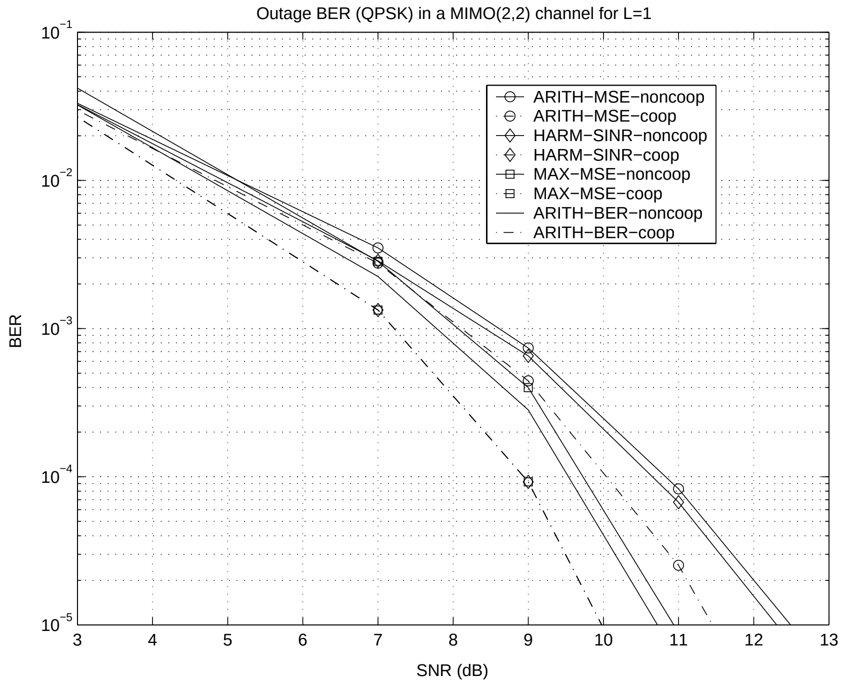


Figure 5.10: BER (at an outage probability of 5%) vs. SNR when using QPSK in a 2×2 MIMO channel with $L = 1$ for the ARITH-MSE, HARM-SINR, MAX-MSE, and ARITH-BER criteria with and without carrier cooperation.

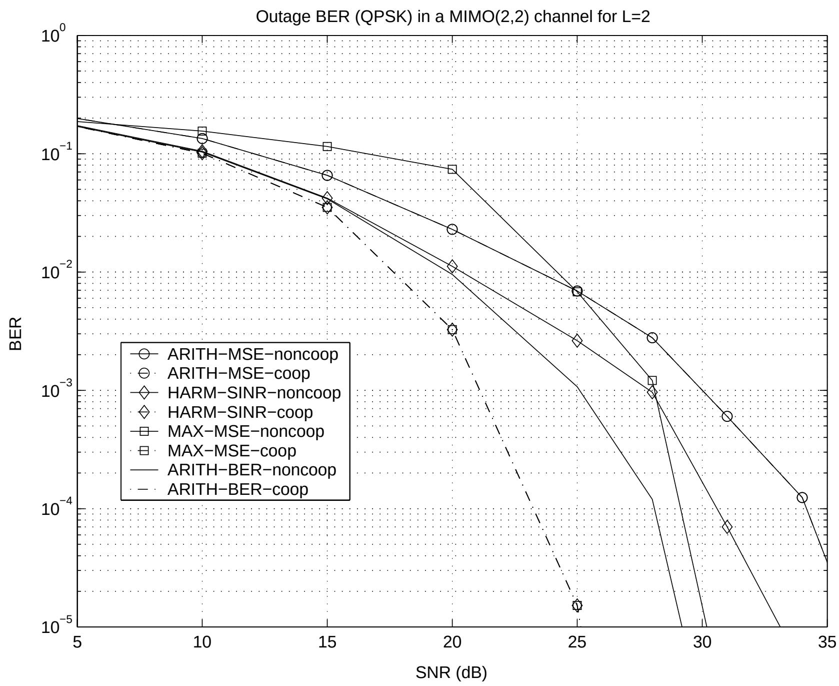


Figure 5.11: BER (at an outage probability of 5%) vs. SNR when using QPSK in a 2×2 MIMO channel with $L = 2$ (fully loaded system) for the ARITH-MSE, HARM-SINR, MAX-MSE, and ARITH-BER criteria with and without carrier cooperation.

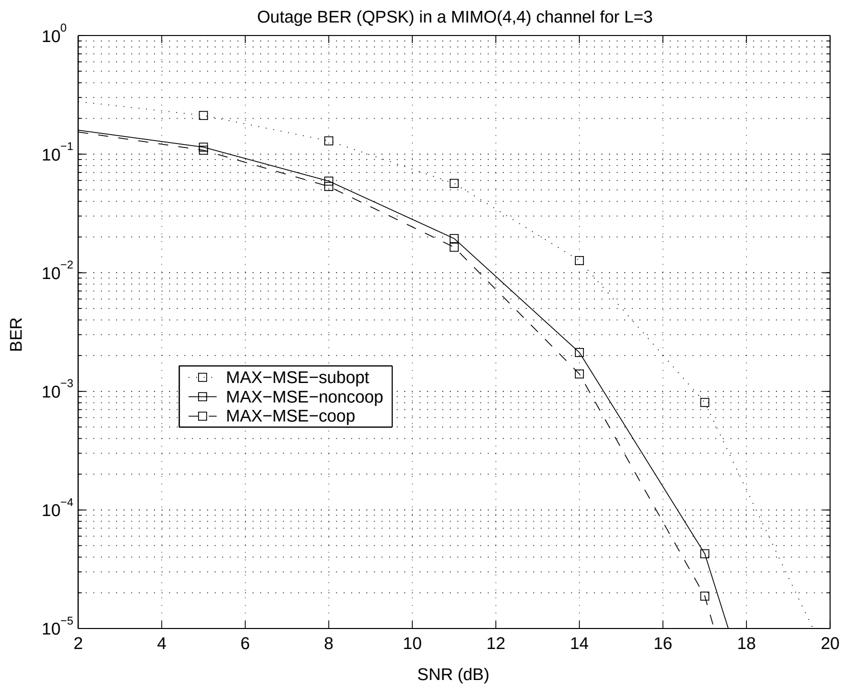


Figure 5.12: BER (at an outage probability of 5%) vs. SNR when using QPSK in a 4×4 MIMO channel with $L = 3$ for the MAX-MSE criterion (optimal solution with and without carrier cooperation and suboptimal solution).

channel-diagonalizing structure and all channel eigenmodes are used, both the carrier-cooperative and the carrier-noncooperative approaches are equivalent).

Therefore, whereas cooperation among carriers may give a small improvement in underloaded systems, for high loaded systems the difference becomes important and cooperation should be used.

Optimum vs. Suboptimum (Imposing Diagonality) Design for Schur-Convex Criteria

We now compare the difference in performance between the optimal and suboptimal solutions of a Schur-convex function: the MAX-MSE criterion. In Figure 5.12, the BER is plotted vs. the SNR for a 4×4 MIMO channel with $L = 3$ using QPSK constellations. The optimal cooperative method only shows a small improvement of about 0.5 dB with respect the optimal noncooperative approach. The suboptimal approach (imposing the diagonal structure) has a significant performance degradation of about 3 dB. In Figure 5.13, the BER is plotted vs. the SNR for a 4×4 MIMO channel for $L = 4$. In this case, the system is fully loaded and the difference of performance is much more significant. The optimal cooperative method is about 7 dB better than the optimal noncooperative approach and the suboptimal solution has an additional degradation of another 7 dB. In fact, as previously seen in Figure 5.9, methods with a diagonal structure fail for fully loaded systems.

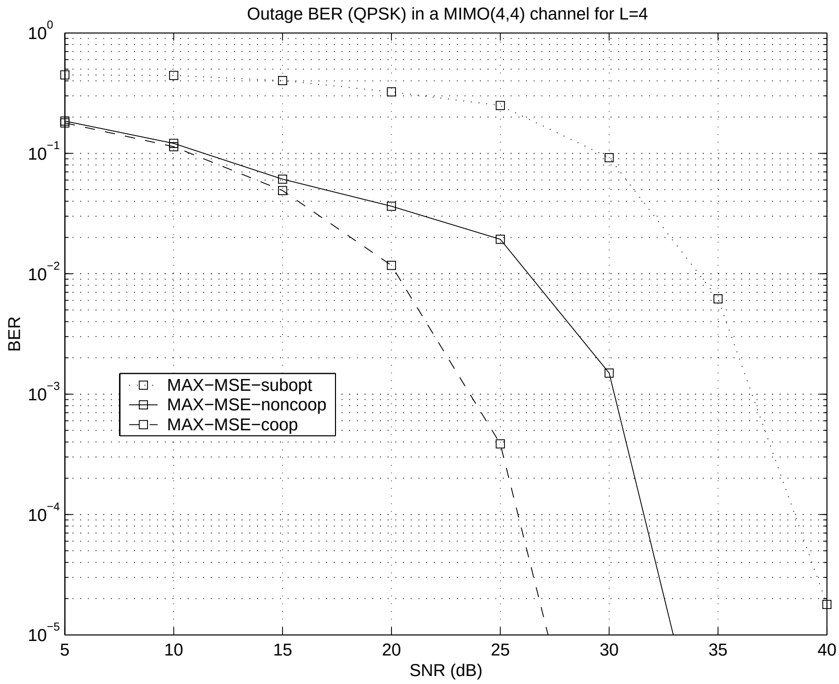


Figure 5.13: BER (at an outage probability of 5%) vs. SNR when using QPSK in a 4×4 MIMO channel with $L = 4$ (fully loaded system) for the MAX-MSE criterion (optimal solution with and without carrier cooperation and suboptimal solution).

Overloaded Systems for Schur-Convex Criteria ($L > \text{rank}(\mathbf{H})$)

We consider now an overloaded system, *i.e.* $L > \text{rank}(\mathbf{H})$ or equivalently $L_0 > 0$ for Schur-convex criteria (it does not make any sense to consider an overloaded system for a Schur-concave criterion, since the substreams associated to zero eigenmodes would be always in error). Such systems are interference-limited and consequently the BER cannot be made as small as desired. In Figure 5.14, the BER is plotted vs. the SNR for a 4×4 MIMO channel with $L = 5$ using QPSK constellations for the MAX-MSE criterion. As can be observed, the suboptimal solution (imposing the channel-diagonalizing structure) has a BER about one order of magnitude higher than the optimal solutions with and without carrier cooperation.

For systems designed with a channel-diagonalized structure (either Schur-concave functions of Schur-convex criteria suboptimally solved), the performance for a sufficiently high SNR will inevitably tend to $\text{BER} = 0.5 \times L_0/L$ (considering that the $\check{L}$ substreams associated to nonzero eigenmodes have a negligible BER and that the L_0 substreams associated to zero eigenmodes have a BER of 0.5). Systems designed with a channel-nondiagonalized structure behave in a CDMA-fashion: it is like having L users using optimally designed codes with a spreading factor of $\check{L}$.

Summary of Observations

It is important to remark that all the simulations results have been obtained in terms of BER assuming that the set of constellations had already been determined. In this sense, it is not

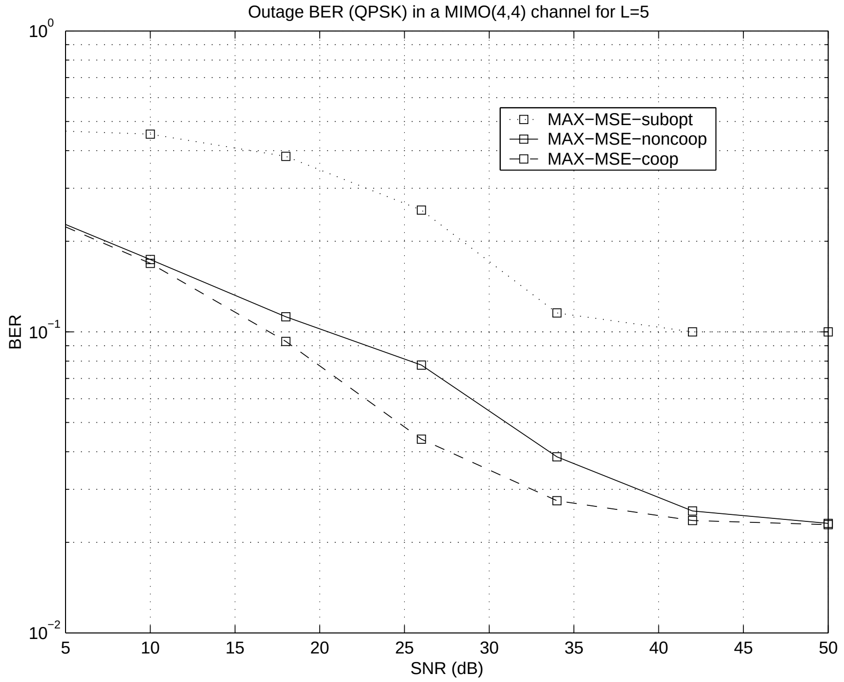


Figure 5.14: BER (at an outage probability of 5%) vs. SNR when using QPSK in a 4×4 MIMO channel with $L = 5$ (overloaded system) for the MAX-MSE criterion (optimal solution with and without carrier cooperation and suboptimal solution).

surprising that the maximization of the mutual information performs very poorly as a means to improve the BER performance for a given set of practical constellations (such a criterion only makes sense when optimizing the constellations).

We now summarize the observations made from the numerical simulations:

- Schur-convex criteria are in general superior to Schur-concave criteria in terms of BER performance (the superiority is evident for highly loaded systems). In fact, the weakness of Schur-concave functions is the diagonal structure that forces the transmission of each symbol through each channel eigenmode with the consequent waste of power if some of the eigenmodes has a low gain. On the other hand, Schur-convex functions transmit all the symbols in a distributed way and therefore have more flexibility to properly use the channel eigenmodes.

- Schur-convex criteria optimally solved are significantly better than when suboptimally solved imposing a diagonal structure (the difference is higher for highly loaded systems). The reason again is the lack of robustness inherent to the diagonal structure.

- Carrier cooperation always improves the performance. The difference becomes significative for highly loaded systems.

- It is very common in the literature of equalization to include a ZF constraint in the design.

Such a constraint is easily introduced in the unified framework but the performance is degraded due to the additional constraint in the design.

- The best criteria (in order) are: ARITH-BER (benchmark), MAX-MSE, HARM-SINR, and ARITH-MSE.

5.8 Chapter Summary and Conclusions

In this chapter, we have formulated and solved the joint design of transmit and receive multiple beamvectors or beam-matrices (also known as linear precoder at the transmitter and linear equalizer at the receiver) for a set of MIMO channels under a variety of design criteria. Instead of considering each design criterion separately, we have developed a unifying framework that generalizes the existing results by considering two families of objective functions that embrace most reasonable criteria to design a communication system: Schur-concave and Schur-convex functions. For Schur-concave objective functions (ARITH-MSE, GEOM-MSE, ARITH-SINR, GEOM-SINR, GEOM-BER), the channel-diagonalizing structure is always optimal, whereas for Schur-convex functions (MAX-MSE, HARM-SINR, MIN-SINR, ARITH-BER, MAX-BER), an optimal solution diagonalizes the channel only after a very specific rotation of the transmitted symbols.

Knowing the optimal structure of the communication process, the design problem has been formulated within the powerful framework of convex optimization theory, in which a great number of interesting design criteria can be easily accommodated and efficiently solved even though closed-form expressions may not exist. From this perspective, a variety of design criteria have been analyzed and, in particular, optimal beamvectors have been derived in the sense of having minimum average BER. It has been shown how to include additional constraints on the design to control the dynamic range of the transmitted signal and the PAR. We have also considered the more general scheme in which cooperation among the processing at different MIMO channels is allowed to improve performance.

Some important points that should be remarked regarding the different design criteria are:

- Most of the presented methods under the framework of convex optimization theory have nice closed-form solutions which can be easily implemented in practice using simple and efficient algorithms.

- Method ARITH-BER is clearly the best in terms of average BER. For the noncooperative scheme, there is no closed-form solution and an iterative approach is necessary such as an interior-point method. Interestingly, for the case of a single MIMO channel or multiple MIMO channels using a cooperative approach, the solution can be obtained in closed-form (see Remark 5.4). In any case, it always serves as a benchmark for comparison.

- The utilization of the Chernoff approximation instead of the exact BER function gives indistinguishable results.
- Methods ARITH-MSE, HARM-SINR, and MAX-MSE have very simple solutions and perform really close to the benchmark given by ARITH-BER. These methods should therefore be considered for practical purposes.
- Two novel multi-level water-filling solutions (with the corresponding practical implementation algorithms) have been obtained for the MAX-MSE and the HARM-SINR criteria.
- Cooperation among different MIMO channels improves performance without significant increase on the complexity (each MIMO channel can be diagonalized independently and then the largest eigenmodes are selected).
- A striking result is that, for single MIMO channels or multiple MIMO channels with cooperation, all criteria with Schur-convex objective functions (*e.g.*, MAX-MSE, HARM-SINR, MIN-SINR, ARITH-BER, and MAX-BER) have the same optimal solution. Hence, the best performance in terms of average BER (given by the ARITH-BER criterion with cooperation among the MIMO channels) can be achieved in practice with low complexity using the simple water-filling solution of the ARITH-MSE criterion plus the computation of the rotation matrix (see Remark 5.4).

All the material presented in this chapter, which is strongly based on majorization theory and convex optimization theory, is a novel contribution of this dissertation.

Appendix 5.A Proof of Theorem 5.1

We first present a couple of lemmas and then proceed to the proof of Theorem 5.1.

Lemma 5.10 *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are* $n \times n$ *positive semidefinite Hermitian matrices, then*

$$\mathrm{Tr}\left(\mathbf{AB}\right) \geq \sum_{i=1}^{n} \lambda_{A,i}\, \lambda_{B,n-i+1}$$

where $\lambda_{A,i}$ *and* $\lambda_{B,i}$ *are the eigenvalues of* $\mathbf{A}$ *and* $\mathbf{B}$*, respectively, in decreasing order.*

Proof. The proof can be found in [Mar79, 9.H.1.h]. However, due to its simplicity using basic results of majorization theory, we rewrite it here. Given the EVD of $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} = \mathbf{U}_A \mathbf{D}_A \mathbf{U}_A^H$ and $\mathbf{B} = \mathbf{U}_B \mathbf{D}_B \mathbf{U}_B^H$, we have

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{AB}\right) &= \mathrm{Tr}\left(\mathbf{D}_B \mathbf{U}_B^H \mathbf{A} \mathbf{U}_B\right) \\ &= \sum_{i=1}^{n} \lambda_{B,i} \left[\mathbf{U}_B^H \mathbf{A} \mathbf{U}_B\right]_{ii}.\end{aligned}$$

Since the function $\sum_i w_i \, x_i$ (with x_i in decreasing order) is minimized when w_i are in increasing order and it is then a Schur-concave function (see Lemma 5.2), the minimum is achieved by an extreme point (*i.e.*, a vector that majorizes any other). Recalling that the vector of eigenvalues majorizes the vector of diagonal elements (Lemma 3.6), it follows that $\mathrm{Tr}\,(\mathbf{AB}) \geq \sum_{i=1}^{n} \lambda_{B,n-i+1} \, \lambda_{A,i}$ where $\lambda_{A,i}$ and $\lambda_{B,i}$ are in decreasing order. ∎

Lemma 5.11 *Given a matrix* $\mathbf{B} \in \mathbb{C}^{n_T \times L}$ *and a positive semidefinite Hermitian matrix* $\mathbf{R}_H \in \mathbb{C}^{n_T \times n_T}$ *such that* $\mathbf{B}^H \mathbf{R}_H \mathbf{B}$ *is a diagonal matrix with diagonal elements in increasing order (possibly with some zero diagonal elements), it is always possible to find another matrix of the form* $\tilde{\mathbf{B}} = \mathbf{U}_{H,1} \mathbf{\Sigma}_{B,1}$ *of at most rank* $\check{L} \triangleq \min\left(L, \mathrm{rank}\left(\mathbf{R}_H\right)\right)$ *that satisfies* $\tilde{\mathbf{B}}^H \mathbf{R}_H \tilde{\mathbf{B}} = \mathbf{B}^H \mathbf{R}_H \mathbf{B}$ *with* $\mathrm{Tr}\left(\tilde{\mathbf{B}} \tilde{\mathbf{B}}^H\right) \leq \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$*, where* $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times \check{L}}$ *has as columns the eigenvectors of* $\mathbf{R}_H$ *corresponding to the* $\check{L}$ *largest eigenvalues in increasing order and* $\mathbf{\Sigma}_{B,1} = \left[\mathbf{0}_{\check{L} \times (L-\check{L})} \;\; \mathrm{diag}\left(\{\sigma_{B,i}\}\right)_{\check{L} \times \check{L}}\right] \in \mathbb{C}^{\check{L} \times L}$ *has zero elements except along the rightmost main diagonal (which can be assumed real w.l.o.g.).*

Proof. Although the basic idea follows easily from the application of Lemma 5.10, the formal proof for arbitrary values of n_T, L, and $\mathrm{rank}\left(\mathbf{R}_H\right)$ becomes notationally involved.

Since $\mathbf{B}^H \mathbf{R}_H \mathbf{B} \in \mathbb{C}^{L \times L}$ (recall that $\mathbf{R}_H \in \mathbb{C}^{n_T \times n_T}$) is diagonal with diagonal elements in increasing order, we can write

$$\mathbf{B}^H \mathbf{R}_H \mathbf{B} = \tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}$$

where $\mathbf{D}$ is a square diagonal matrix (with real diagonal elements) of dimension $\check{L} \triangleq \min\left(L, \mathrm{rank}\left(\mathbf{R}_H\right)\right)$. Using the SVD [Hor85], we can then write $\mathbf{R}_H^{1/2} \mathbf{B} = \mathbf{Q}\mathbf{\Sigma}$ where $\mathbf{Q} \in \mathbb{C}^{n_T \times n_T}$ is a unitary matrix whose columns are the left singular vectors, the right singular vectors (eigenvectors of $\tilde{\mathbf{D}}$) are the canonical vectors, and matrix $\mathbf{\Sigma} \in \mathbb{C}^{n_T \times L}$ is composed of zero elements and contains $\mathbf{D}^{1/2}$ in its top-right block so that $\mathbf{\Sigma}^H \mathbf{\Sigma} = \tilde{\mathbf{D}}$. In particular, if $n_T \geq L$, then $\tilde{\mathbf{D}} = \mathbf{D}$ and $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{D}^{1/2} \\ \mathbf{0} \end{bmatrix}$, and if $n_T < L$, then $\mathbf{\Sigma} = \left[\mathbf{0} \; \mathbf{D}^{1/2}\right]$ (note that the position of matrix $\mathbf{D}^{1/2}$ within $\mathbf{\Sigma}$ is different to the classical definition of the SVD [Hor85] because the diagonal elements of $\mathbf{B}^H \mathbf{R}_H \mathbf{B}$ are assumed in increasing order).

Assuming that matrix $\mathbf{R}_H$ is nonsingular with EVD given by $\mathbf{R}_H = \mathbf{U}_H \mathbf{D}_H \mathbf{U}_H^H$, we can write

$$\mathbf{B} = \mathbf{R}_H^{-1/2} \mathbf{Q}\mathbf{\Sigma} = \mathbf{U}_H \mathbf{D}_H^{-1/2} \mathbf{U}_H^H \mathbf{Q}\mathbf{\Sigma}. \tag{5.79}$$

In case that $\mathbf{R}_H$ is singular, clearly $\mathbf{B}$ must be orthogonal to the null space of $\mathbf{R}_H$, otherwise this component could be made zero without changing the value of $\mathbf{B}^H \mathbf{R}_H \mathbf{B}$ and decreasing $\mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$. Knowing that $\mathbf{B}$ must be orthogonal to the null space of $\mathbf{R}_H$, expression (5.79) is still valid using the pseudo-inverse of $\mathbf{R}_H$ instead of the inverse.

The idea now is to find another matrix $\mathbf{B}$ by changing the unitary matrix $\mathbf{Q}$ in (5.79) with the lowest possible value of $\mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$ (note that any matrix $\mathbf{B}$ obtained from (5.79) satisfies by definition the desired constraint $\mathbf{B}^H\mathbf{R}_H\mathbf{B} = \tilde{\mathbf{D}}$). Using Lemma 5.10, $\mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$ can be lower-bounded as follows:

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) &= \mathrm{Tr}\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^H\tilde{\mathbf{U}}^H\mathbf{D}_H^{-1}\tilde{\mathbf{U}}\right)\\ &\geq \sum_{i=1}^{\check{L}} d_i\ \lambda_{H,i}^{-1}\end{aligned}$$

where $\tilde{\mathbf{U}} \triangleq \mathbf{U}_H^H\mathbf{Q}$, d_i is the ith diagonal element of $\mathbf{D}$ in increasing order and $\{\lambda_{H,i}\}_{i=1}^{\check{L}}$ are the $\check{L}$ largest eigenvalues of $\mathbf{R}_H$ in increasing order. If the d_i's are different, the lower bound is achieved by matrix $\tilde{\mathbf{U}}$ being a permutation matrix. For subsets of equal d_i's, the corresponding subblock in $\tilde{\mathbf{U}}$ can be any rotation matrix. Therefore, if $\tilde{\mathbf{U}}$ is chosen as a permutation matrix $\mathbf{P}$ that selects the $\check{L}$ largest eigenvalues of $\mathbf{R}_H$ in the same ordering as the d_i's, the lower bound is achieved. From (5.79), we obtain that the optimal $\mathbf{B}$ (in the sense of minimizing the value of $\mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$) has at most rank $\check{L}$ and is of the form $\mathbf{B} = \mathbf{U}_H\mathbf{D}_H^{-1/2}\mathbf{P}\boldsymbol{\Sigma} = \mathbf{U}_{H,1}\boldsymbol{\Sigma}_{B,1}$ where $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T\times\check{L}}$ has as columns the eigenvectors of $\mathbf{R}_H$ corresponding to the $\check{L}$ largest eigenvalues in increasing order and $\boldsymbol{\Sigma}_{B,1} = \left[\mathbf{0}_{\check{L}\times(L-\check{L})}\ \ \mathrm{diag}\left(\{\sigma_{B,i}\}\right)_{\check{L}\times\check{L}}\right] \in \mathbb{C}^{\check{L}\times L}$ has zero elements except along the rightmost main diagonal (which can be assumed real w.l.o.g.). ■

Proof of Theorem 5.1. The constrained optimization problem to be solved is

$$\begin{aligned}\min_{\mathbf{B}} \quad & f_0\left(\mathbf{d}(\mathbf{E}\left(\mathbf{B}\right))\right)\\ \text{s.t.} \quad & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T\end{aligned}$$

where $\mathbf{d}(\mathbf{E})$ is the vector of diagonal elements of the MSE matrix $\mathbf{E}\left(\mathbf{B}\right) = \left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}$. It is mathematically convenient to assume the diagonal elements of $\mathbf{E}\left(\mathbf{B}\right)$ in decreasing order, *i.e.*, $d_i(\mathbf{E}) \geq d_{i+1}(\mathbf{E})$, without loss of generality. In fact, most reasonable objective functions (in particular, all the objective functions considered in §5.5) have a fixed preferred ordering of the arguments, *i.e.*, the value of the function is minimized with a very specific ordering of the arguments. In such cases, it suffices to relabel the arguments so that the preferred ordering is decreasing. In a more general case, however, a function may not have a fixed preferred ordering since it may depend on the specific value of the arguments. In such a case, since we are interested in minimizing the objective function, we can define and use instead the function $\tilde{f}_0\left(\mathbf{x}\right) = \min_{\mathbf{P}\in\Pi} f_0\left(\mathbf{P}\mathbf{x}\right)$, where $\mathbf{P}$ is a permutation matrix and Π is the set of the $L!$ different permutation matrices. The original minimization of f_0 without the ordering constraint is equivalent to the minimization of $\tilde{f}_0$ with the ordering constraint. Therefore, we can always assume that the function to be minimized has been properly defined so that the ordering constraint can be included without loss of generality (c.f. §5.5).

If f_0 is Schur-concave, using the definition of Schur-concavity (Definition 3.4) and the relation $\mathbf{d} \prec \boldsymbol{\lambda}$ (Lemma 3.6), it follows that $f_0(\boldsymbol{\lambda}(\mathbf{E})) \leq f_0(\mathbf{d}(\mathbf{E}))$ where $\boldsymbol{\lambda}(\mathbf{E})$ is the vector of eigenvalues of $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ in decreasing order. The lower bound $f_0(\boldsymbol{\lambda}(\mathbf{E}))$ is achieved if matrix $\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ is diagonal with diagonal elements in decreasing order or, equivalently, if $\mathbf{B}^H \mathbf{R}_H \mathbf{B}$ is diagonal with diagonal elements in increasing order. Furthermore, for any given $\mathbf{B}$, one can always find a rotation matrix $\mathbf{Q}$ so that $\mathbf{Q}^H \left(\mathbf{B}^H \mathbf{R}_H \mathbf{B}\right) \mathbf{Q}$ becomes diagonal and use instead the transmit matrix $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{Q}$ (the sum of diagonal elements of $\mathbf{E}$ remains the same regardless of $\mathbf{Q}$) improving the performance (by this rotation the utilized power remains the same $\mathrm{Tr}\left(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^H\right) = \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$). This implies that for Schur-concave functions, there is an optimal $\mathbf{B}$ with a structure such that $\mathbf{B}^H \mathbf{R}_H \mathbf{B}$ is diagonal with diagonal elements in increasing order.

If f_0 is Schur-convex, the opposite happens. From the definition of Schur-convexity (Definition 3.4) and the relation $\mathbf{1} \prec \mathbf{d}$ (Lemma 3.1), it follows that $f_0(\mathbf{d}(\mathbf{E}))$ is minimized when $\mathbf{E}$ has equal diagonal elements. Furthermore, for any given $\mathbf{B}$, one can always find (see Corollary 3.2) a rotation matrix $\mathbf{Q}$ so that $\mathbf{Q}^H \left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1} \mathbf{Q}$ has identical diagonal elements and use instead the transmit matrix $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{Q}$ (the sum of diagonal elements of $\mathbf{E}$ remains the same regardless of $\mathbf{Q}$) improving the performance (the transmit power remains the same). Therefore, for an optimal $\mathbf{B}$ we have that $\left[\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}\right]_{ii} = \frac{1}{L} \mathrm{Tr}\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$. Interestingly, regardless of the specific function f_0, the optimal $\mathbf{B}$ can be found by first minimizing $\mathrm{Tr}\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ (without imposing the constraint that the diagonal elements be equal) and then including the rotation to make the diagonal elements identical. The rotation can be found using Algorithm 3.2 (reproduced from [Vis99b, Section IV-A]) or with any rotation matrix that satisfies $|[\mathbf{Q}]_{ik}| = |[\mathbf{Q}]_{il}| \ \forall i, k, l$ such as the Discrete Fourier Transform (DFT) matrix or the Hadamard matrix when the dimensions are appropriate such as a power of two (see §3.2 for more details). Regarding the minimization of $\mathrm{Tr}\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$, since it is a Schur-concave function of the diagonal elements ($f_0(\mathbf{d}) = \sum_{i=1}^{L} d_i$), the previous result can be applied to show that there is an optimal $\mathbf{B}$ (excluding for the moment the rotation) such that $\mathbf{B}^H \mathbf{R}_H \mathbf{B}$ is diagonal with diagonal elements in increasing order.

Given that $\mathbf{B}^H \mathbf{R}_H \mathbf{B}$ is diagonal, it follows from Lemma 5.11 that $\mathbf{B}$ has at most rank $\check{L} \triangleq \min(L, \mathrm{rank}(\mathbf{R}_H))$ and can be written as $\mathbf{B} = \mathbf{U}_{H,1} \boldsymbol{\Sigma}_{B,1}$ where $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times \check{L}}$ has as columns the eigenvectors of $\mathbf{R}_H$ corresponding to the $\check{L}$ largest eigenvalues in increasing order and $\boldsymbol{\Sigma}_{B,1} = \left[\mathbf{0} \ \ \mathrm{diag}\left(\{\sigma_{B,i}\}\right)\right] \in \mathbb{C}^{\check{L} \times L}$ has zero elements except along the rightmost main diagonal.

Hence, we can write an optimal $\mathbf{B}$ as

$$\mathbf{B} = \begin{cases} \mathbf{U}_{H,1} \boldsymbol{\Sigma}_{B,1} & \text{for } f_0 \text{ Schur-concave} \\ \mathbf{U}_{H,1} \boldsymbol{\Sigma}_{B,1} \mathbf{V}_B^H & \text{for } f_0 \text{ Schur-convex} \end{cases}$$

where $\mathbf{U}_{H,1}$ and $\boldsymbol{\Sigma}_{B,1}$ are defined as before and $\mathbf{V}_B \in \mathbb{C}^{L \times L}$ is the rotation to make the diagonal elements of $\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ identical. ■

Appendix 5.B Proof of Lemma 5.1

The proof is straightforward from Lemma 5.11. Since the diagonal elements of $\left(\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}$ can be assumed in decreasing ordering w.l.o.g. (c.f. proof of Theorem 5.1 in Appendix 5.A) and $\mathbf{B}^H\mathbf{R}_H\mathbf{B}$ is diagonal, it follows from Lemma 5.11 that an optimal solution can be expressed as $\mathbf{B}=\mathbf{U}_{H,1}\mathbf{\Sigma}_{B,1}$ where $\mathbf{U}_{H,1}\in\mathbb{C}^{n_T\times\check{L}}$ has as columns the eigenvectors of $\mathbf{R}_H$ corresponding to the $\check{L}\triangleq\min\left(L,\operatorname{rank}\left(\mathbf{R}_H\right)\right)$ largest eigenvalues in increasing order and $\mathbf{\Sigma}_{B,1}=\left[\mathbf{0}\ \operatorname{diag}\left(\{\sigma_{B,i}\}\right)\right]\in\mathbb{C}^{\check{L}\times L}$ has zero elements except along the rightmost main diagonal (assumed real w.l.o.g.). ■

Appendix 5.C Proof of Corollary 5.2

The proof in one direction is very similar to that of Theorem 5.1. First, note that the vector of eigenvalues $\boldsymbol{\lambda}(\mathbf{E}(\mathbf{B}))$ (obtained when the channel is fully diagonalized with $\mathbf{V}_B=\mathbf{I}$) majorizes any vector of diagonal elements $\mathbf{d}(\mathbf{E}(\mathbf{B}))$ (obtained when an arbitrary rotation $\mathbf{V}_B$ is used) which in turn majorizes the vector of equal elements $\mathbf{1}(\mathbf{E}(\mathbf{B}))$ (obtained when a very specific rotation $\mathbf{V}_B$ is used as described in Theorem 5.1 for Schur-convex functions). Since the function is Schur-concave, it must be that $f_0(\boldsymbol{\lambda})\leq f_0(\mathbf{d})\leq f_0(\mathbf{1})$. At the same time, since the function is Schur-convex, it must be that $f_0(\boldsymbol{\lambda})\geq f_0(\mathbf{d})\geq f_0(\mathbf{1})$. Therefore, it follows that $f_0(\boldsymbol{\lambda})=f_0(\mathbf{d})=f_0(\mathbf{1})$. The opposite direction is trivial since if the function is invariant with respect post-rotations of $\mathbf{B}$, it then follows that in particular $f_0(\mathbf{d}_1)=f_0(\mathbf{d}_2)$ for $\mathbf{d}_1\succ\mathbf{d}_2$. ■

Appendix 5.D Proof of Schur-Convexity/Concavity Lemmas

Proof of Lemma 5.2 $\left(f_0(\mathbf{x})=\sum_i(w_i\,x_i)\right)$

Since the x_i's are in decreasing order $x_i\geq x_{i+1}$, the function $f_0(\mathbf{x})=\sum_i(w_i\,x_i)$ is minimized with the weights in increasing order $w_i\leq w_{i+1}$. To show this assume for a moment that for $i<j$ $(x_i\geq x_j)$ the weights are such that $w_i>w_j$. It follows that the term $(w_i\,x_i+w_j\,x_j)$ can be minimized by simply swapping the weights:

$$\begin{aligned} & w_i\ (x_i-x_j)\geq w_j\ (x_i-x_j) \\ \Longleftrightarrow\qquad & w_i\,x_i+w_j\,x_j\geq w_i\,x_j+w_j\,x_i. \end{aligned}$$

To prove that the function f_0 is Schur-concave (see Definition 3.4), define $\phi(\mathbf{x})\triangleq -f_0(\mathbf{x})=\sum_i g_i(x_i)$ where $g_i(x)=-w_i\,x$. Function ϕ is Schur-convex because $g_i'(a)\geq g_{i+1}'(b)$ whenever $a\geq b$ (see Lemma 3.3) and, therefore, f_0 is Schur-concave (see Definition 3.4). ■

Proof of Lemma 5.3 **($f_0(\mathbf{x}) = \prod_i x_i^{w_i}$)**

Since the x_i's are strictly positive and in decreasing order $x_i \geq x_{i+1} > 0$, the function $f_0(\mathbf{x}) = \prod_i x_i^{w_i}$ is minimized with the weights in increasing order $w_i \leq w_{i+1}$. To show this assume for a moment that for $i < j$ $(x_i \geq x_j)$ the weights are so that $w_i > w_j$. It follows that the term $(x_i^{w_i} x_j^{w_j})$ can be minimized by simply swapping the weights:

$$\begin{aligned} & w_i \log\left(\frac{x_i}{x_j}\right) \geq w_j \log\left(\frac{x_i}{x_j}\right) \\ \iff \quad & \left(\frac{x_i}{x_j}\right)^{w_i} \geq \left(\frac{x_i}{x_j}\right)^{w_j} \\ \iff \quad & x_i^{w_i} x_j^{w_j} \geq x_i^{w_j} x_j^{w_i}. \end{aligned}$$

To prove that the function f_0 is Schur-concave (see Definition 3.4), define $\phi(\mathbf{x}) \triangleq -\log f_0(\mathbf{x}) = \sum_i g_i(x_i)$ where $g_i(x) = -w_i \log x$. Function ϕ is Schur-convex because $g_i'(a) \geq g_{i+1}'(b)$ whenever $a \geq b$ (see Lemma 3.3). Since $f_0(\mathbf{x}) = e^{-\phi(\mathbf{x})}$ and function e^{-x} is decreasing in x, f_0 is Schur-concave by Lemma 3.2. ■

Proof of Lemma 5.4 **($f_0(\mathbf{x}) = \max_i\{x_i\}$)**

From Definition 3.1, it follows that $f_0(\mathbf{x}) = \max_i\{x_i\} = x_{[1]}$. If $\mathbf{x} \prec \mathbf{y}$ it must be that $x_{[1]} \leq y_{[1]}$ (from Definition 3.2) and, therefore, $f_0(\mathbf{x}) \leq f_0(\mathbf{y})$. This means that f_0 is Schur-convex by Definition 3.4. ■

Proof of Lemma 5.5 **($f_0(\mathbf{x}) = -\sum_i \left(w_i \left(x_i^{-1} - 1\right)\right)$)**

Since the x_i's are strictly positive and in decreasing order $x_i \geq x_{i+1} > 0$, the function $f_0(\mathbf{x}) = -\sum_i \left(w_i \left(x_i^{-1} - 1\right)\right)$ is minimized with the weights in increasing order $w_i \leq w_{i+1}$ (this can be similarly proved as was done in the proof of Lemma 5.2).

To prove that the function f_0 is Schur-concave (see Definition 3.4), define $\phi(\mathbf{x}) \triangleq -f_0(\mathbf{x}) = \sum_i g_i(x_i)$ where $g_i(x) = w_i \left(x^{-1} - 1\right)$. Function ϕ is Schur-convex because $g_i'(a) \geq g_{i+1}'(b)$ whenever $a \geq b$ (see Lemma 3.3) and, therefore, f_0 is Schur-concave by Definition 3.4. ■

Proof of Lemma 5.6 **($f_0(\mathbf{x}) = -\prod_i \left(x_i^{-1} - 1\right)^{w_i}$)**

Since the x_i's are strictly positive and in decreasing order $x_i \geq x_{i+1} > 0$, it follows that $f_0(\mathbf{x}) = -\prod_i \left(x_i^{-1} - 1\right)^{w_i}$ is minimized with the weights in increasing order $w_i \leq w_{i+1}$ (this can be similarly proved as was done in the proof of Lemma 5.3).

To prove that the function f_0 is Schur-concave for $x_i \leq 0.5$ (see Definition 3.4), define $\phi(\mathbf{x}) \triangleq \log(-f_0(\mathbf{x})) = \sum_i g_i(x_i)$ where $g_i(x) = w_i \log\left(x^{-1} - 1\right)$. Function ϕ is Schur-convex because $g_i'(a) \geq g_{i+1}'(b)$ whenever $0.5 \geq a \geq b$ [16] (see Lemma 3.3). Since $f_0(\mathbf{x}) = -e^{\phi(\mathbf{x})}$ and function $-e^{x}$ is decreasing in x, f_0 is Schur-concave by Lemma 3.2. ■

[16] Function $(1-x)\,x$ is increasing in x for $0 \leq x \leq 0.5$.

Proof of Lemma 5.7 ($f_0(\mathbf{x}) = \sum_i \frac{x_i}{1-x_i}$)

To prove that the function $f_0(\mathbf{x}) = \sum_i \frac{x_i}{1-x_i}$ is Schur-convex, rewrite it as $f_0(\mathbf{x}) = \sum_i g(x_i)$ where $g(x) = \frac{x}{1-x}$. Since function g is convex, it follows that f_0 is Schur-convex by Corollary 3.3. ■

Proof of Lemma 5.8 ($f_0(\mathbf{x}) = \sum_i \mathrm{BER}\left(x_i^{-1} - 1\right)$)

To prove that the function $f_0(\mathbf{x}) = \sum_i \mathrm{BER}\left(x_i^{-1} - 1\right)$ is Schur-convex for $\theta \geq x_i > 0$ (for sufficiently small θ such that $\mathrm{BER}\left(x_i^{-1} - 1\right) \leq 10^{-2}\ \forall i$), write $f_0(\mathbf{x}) = \sum_i g(x_i)$ where $g(x) = \mathrm{BER}\left(x^{-1} - 1\right)$. Since function g is convex within the range $(0, \theta]$ (see §2.5.4.4), it follows that f_0 is Schur-convex by Corollary 3.1. ■

Proof of Lemma 5.9 ($f_0(\mathbf{x}) = \prod_i \mathrm{BER}\left(x_i^{-1} - 1\right)$)

To prove that the function $f_0(\mathbf{x}) = \prod_i \mathrm{BER}\left(x_i^{-1} - 1\right)$ is Schur-concave for $\theta \geq x_i > 0$ with θ such that $\left(\frac{\partial\, \mathrm{BER}\left(x^{-1}-1\right)}{\partial x}\right)^2 \geq \mathrm{BER}\left(x^{-1} - 1\right) \frac{\partial^2\, \mathrm{BER}\left(x^{-1}-1\right)}{\partial x^2}$ for $\theta \geq x > 0$, define $\phi(\mathbf{x}) \triangleq -\log f_0(\mathbf{x}) = \sum_i g(x_i)$ where $g(x) = -\log \mathrm{BER}\left(x^{-1} - 1\right)$. Function g is convex for $\theta \geq x > 0$ because $\frac{\partial^2 g(x)}{\partial x^2} = \frac{1}{\left(\mathrm{BER}\left(x^{-1}-1\right)\right)^2}\left(\left(\frac{\partial\, \mathrm{BER}\left(x^{-1}-1\right)}{\partial x}\right)^2 - \mathrm{BER}\left(x^{-1} - 1\right) \frac{\partial^2\, \mathrm{BER}\left(x^{-1}-1\right)}{\partial x^2}\right) \geq 0$ and ϕ is Schur-convex by Corollary 3.1. Since $f_0(\mathbf{x}) = e^{-\phi(\mathbf{x})}$ and function e^{-x} is decreasing in x, f_0 is Schur-concave by Lemma 3.2. ■

Appendix 5.E Gradients and Hessians for the ARITH-BER

Suboptimal Problem Formulation of (5.58)

- Gradient and the Hessian of the exact BER function $f(\mathbf{x}) = \sum_{k=1}^{N} \alpha_k \mathcal{Q}\left(\sqrt{\tilde{\beta}_k x_k}\right)$ where $\tilde{\beta}_k = \beta_k \lambda_k$ (see Appendix 5.A):

$$\nabla f(\mathbf{x}) = -\sqrt{\frac{1}{8\pi}} \begin{bmatrix} \alpha_1 \tilde{\beta}_1^{1/2} e^{-\tilde{\beta}_1 x_1/2} x_1^{-1/2} \\ \vdots \\ \alpha_N \tilde{\beta}_N^{1/2} e^{-\tilde{\beta}_N x_N/2} x_N^{-1/2} \end{bmatrix}$$

$$Hf(\mathbf{x}) = \frac{1}{2}\sqrt{\frac{1}{8\pi}}\, \mathrm{diag}\left(\left\{\alpha_k \tilde{\beta}_k^{1/2} e^{-\tilde{\beta}_k x_k/2} x_k^{-1/2}\left(x_k^{-1} + \tilde{\beta}_k\right)\right\}_{k=1}^{N}\right),$$

  and of the Chernoff approximation $f(\mathbf{x}) = \sum_{k=1}^{N} \frac{1}{2} \alpha_k e^{-\tilde{\beta}_k x_k/2}$ (see Appendix 5.A):

$$\nabla f(\mathbf{x}) = -\frac{1}{4} \begin{bmatrix} \alpha_1 \tilde{\beta}_1 e^{-\tilde{\beta}_1 x_1/2} \\ \vdots \\ \alpha_N \tilde{\beta}_N e^{-\tilde{\beta}_N x_N/2} \end{bmatrix}$$

$$Hf(\mathbf{x}) = \frac{1}{8}\, \mathrm{diag}\left(\left\{\alpha_k \tilde{\beta}_k^2 e^{-\tilde{\beta}_k x_k/2}\right\}\right).$$

Optimal Problem Formulation of (5.59)

- Gradient and the Hessian of the exact BER function $f(\mathbf{x}) = \sum_{k=1}^{N} \alpha_k \mathcal{Q}\left(\sqrt{\beta_k \left(x_k^{-1} - 1\right)}\right)$ (see Appendix 5.A):

$$\nabla f(\mathbf{x}) = \sqrt{\frac{1}{8\pi}} \begin{bmatrix} \alpha_1 \beta_1^{1/2} e^{-\beta_1\left(x_1^{-1}-1\right)/2} \left(x_1^3 - x_1^4\right)^{-1/2} \\ \vdots \\ \alpha_N \beta_N^{1/2} e^{-\beta_N\left(x_N^{-1}-1\right)/2} \left(x_N^3 - x_N^4\right)^{-1/2} \end{bmatrix}$$

$$Hf(\mathbf{x}) = \frac{1}{2}\sqrt{\frac{1}{8\pi}} \operatorname{diag}\left(\left\{\alpha_k \beta_k^{1/2} e^{-\beta_k\left(x_k^{-1}-1\right)/2} \left(x_k^3 - x_k^4\right)^{-1/2} \left(\frac{\beta_k}{x_k^2} - \frac{3-4x_k}{x_k - x_k^2}\right)\right\}_{k=1}^{N}\right),$$

and of the Chernoff approximation $f(\mathbf{x}) = \sum_{k=1}^{N} \frac{1}{2}\alpha_k e^{-\beta_k\left(x_k^{-1}-1\right)/2}$ (see Appendix 5.A):

$$\nabla f(\mathbf{x}) = \frac{1}{4} \begin{bmatrix} \alpha_1 \beta_1 e^{-\beta_1\left(x_1^{-1}-1\right)/2} x_1^{-2} \\ \vdots \\ \alpha_N \beta_N e^{-\beta_N\left(x_N^{-1}-1\right)/2} x_N^{-2} \end{bmatrix}$$

$$Hf(\mathbf{x}) = \frac{1}{4} \operatorname{diag}\left(\left\{\alpha_k \beta_k e^{-\beta_k\left(x_k^{-1}-1\right)/2} x_k^{-4} \left(\beta_k/2 - 2x_k\right)\right\}_{k=1}^{N}\right).$$

- Gradient and the Hessian of the log-barrier function $f(\mathbf{x}) = -\mathbf{1}^T \log(\mathbf{b} - \mathbf{A}\mathbf{x})$ corresponding to the linear constraints $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ [Boy00]:

$$\nabla f(\mathbf{x}) = \mathbf{A}^T \operatorname{diag}\left(1/\left(\mathbf{b} - \mathbf{A}\mathbf{x}\right)\right)\mathbf{1}$$
$$Hf(\mathbf{x}) = \mathbf{A}^T \operatorname{diag}\left(1/\left(\mathbf{A}\mathbf{x} - \mathbf{b}\right)^2\right)\mathbf{A}.$$

- Gradient and the Hessian of the log-barrier function $f(\mathbf{x}) = -\log\left(t - \frac{1}{L}\left(L_0 + \sum_{i=1}^{\check{L}} \frac{1}{1+\lambda_i x_i}\right)\right)$ (we define $\mathbf{x} \triangleq [t, x_1, \cdots, x_{\check{L}}]^T$) corresponding to the constraints on the t_k's $t_k \geq \frac{1}{L_k}\left(\left(L_k - \check{L}_k\right) + \sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} x_{k,i}}\right)$:

$$\nabla f(\mathbf{x}) = -\frac{1}{D} \begin{bmatrix} 1 \\ \frac{1}{L}\frac{\lambda_1}{(1+\lambda_1 x_1)^2} \\ \vdots \\ \frac{1}{L}\frac{\lambda_{\check{L}}}{(1+\lambda_{\check{L}} x_{\check{L}})^2} \end{bmatrix}$$

$$Hf(\mathbf{x}) = \begin{bmatrix} 1 \\ \frac{1}{L}\frac{\lambda_1}{(1+\lambda_1 x_1)^2} \\ \vdots \\ \frac{1}{L}\frac{\lambda_{\check{L}}}{(1+\lambda_{\check{L}} x_{\check{L}})^2} \end{bmatrix} \frac{1}{D^2} \begin{bmatrix} 1 & \frac{1}{L}\frac{\lambda_1}{(1+\lambda_1 x_1)^2} & \cdots & \frac{1}{L}\frac{\lambda_{\check{L}}}{(1+\lambda_{\check{L}} x_{\check{L}})^2} \end{bmatrix}$$
$$+ \frac{2}{D} \operatorname{diag}\left(\left\{0, \left\{\frac{1}{L}\frac{\lambda_i^2}{(1+\lambda_i x_i)^3}\right\}_{i=1}^{\check{L}}\right\}\right)$$

where $D \triangleq t - \frac{1}{L}\left(L_0 + \sum_{i=1}^{\check{L}} \frac{1}{1+\lambda_i x_i}\right)$.

Appendix 5.F Proof of Water-Filling Results

Proof of Proposition 5.1 (Modified weighted water-filling.)

Since the problem is convex, an optimal solution can be obtained based on convex optimization theory (the reader is referred to §3.1 and [Lue69, Roc70, Boy00] for details and definitions). We first obtain the closed-form optimal solution to the problem and then proceed to prove the optimality of Algorithm 5.1.

Optimal Solution. The Lagrangian corresponding to the constrained convex problem is

$$L = \sum_{i=1}^{\check{L}} w_i \frac{1}{1+\lambda_i z_i} + \mu\left(\sum_{i=1}^{\check{L}} z_i - P_T\right) - \sum_{i=1}^{\check{L}} \gamma_i z_i \tag{5.80}$$

where μ and $\{\gamma_i\}$ are the dual variables or Lagrange multipliers. The water-filling solution is easily found from the sufficient and necessary KKT optimality conditions (the problem satisfies the Slater's condition and therefore strong duality holds):

$$\begin{aligned}
&\sum_{i=1}^{\check{L}} z_i \le P_T, && z_i \ge 0,\\
&\mu \ge 0, && \gamma_i \ge 0,\\
&\mu = w_i \tfrac{\lambda_i}{(1+\lambda_i z_i)^2} + \gamma_i, &&\\
&\mu\left(\sum_{i=1}^{\check{L}} z_i - P_T\right) = 0, && \gamma_i z_i = 0.
\end{aligned} \tag{5.81}$$

It is assumed that there is at least one $\lambda_i > 0$; otherwise the problem has trivial solution given by $z_i = 0 \ \forall i$. If $\mu = 0$, then $\gamma_i = -w_i \frac{\lambda_i}{(1+\lambda_i z_i)^2}$, which cannot be (since $\gamma_i \ge 0$). Therefore, at an optimal solution $\mu > 0$ and consequently (by the complementary slackness condition) the power constraint must be satisfied with equality. Note that if $\lambda_i = 0$ for some i, then $\gamma_i = \mu > 0 \Rightarrow z_i = 0$, *i.e.*, no power is allocated to eigenmodes with zero gain (as expected). From now on, we consider that all zero-valued λ_i's have been removed.

If $z_i > 0$, then $\gamma_i = 0$ (by the complementary slackness condition), $\mu = w_i \frac{\lambda_i}{(1+\lambda_i z_i)^2}$ (note that $\mu < w_i \lambda_i$), and $z_i = \mu^{-1/2}\, w_i^{1/2}\, \lambda_i^{-1/2} - \lambda_i^{-1}$ (note that if $\lambda_i = 0$, then $\gamma_i = \mu > 0$ and $z_i = 0$ as expected). If $z_i = 0$, then $\mu = w_i \lambda_i + \gamma_i$ (note that $\mu \ge w_i \lambda_i$). Equivalently,

$$z_i = \begin{cases} \mu^{-1/2}\, w_i^{1/2}\, \lambda_i^{-1/2} - \lambda_i^{-1} > 0 & \text{if} \quad \mu < w_i \lambda_i \\ 0 & \text{if} \quad \mu \ge w_i \lambda_i \end{cases}$$

or, more compactly,

$$z_i = \left(\mu^{-1/2}\, w_i^{1/2}\, \lambda_i^{-1/2} - \lambda_i^{-1}\right)^+ \tag{5.82}$$

where $\mu^{-1/2}$ is the water-level chosen so that $\sum_{i=1}^{\check{L}} z_i = P_T$. Note that this solution satisfies all KKT conditions and is therefore optimal.

Optimal Algorithm. Algorithm 5.1 is based on hypothesis testing. It first makes the assumption that all $\check{L}$ eigenmodes are active ($z_i > 0$ for $1 \leq i \leq \check{L}$) and then checks whether the maximum power is exceeded, in which case the hypothesis is rejected, a new hypothesis with one less active eigenmode is made, and so forth.

In more detail, Algorithm 5.1 first reorders the positive eigenmodes so that the terms $(w_i \lambda_i)$ are in decreasing order. With this ordering, since $\lambda_i\, z_i = \left(\mu^{-1/2}\ (w_i\, \lambda_i)^{1/2} - 1\right)^+$, a hypothesis is completely described by the set of active eigenmodes $\tilde{L}$ (such that $z_i > 0$ for $1 \leq i \leq \tilde{L}$ and zero otherwise). This allows a reduction of the total number of hypotheses from $2^{\tilde{L}}$ to $\tilde{L}$. The initial hypothesis chooses the highest number of active eigenmodes $\tilde{L} = \check{L}$.

For each hypothesis, the water-level $\mu^{1/2}$ must be such that the considered $\tilde{L}$ eigenmodes are indeed active while the rest remain inactive:

$$\begin{cases} \mu^{-1/2}\, w_i^{1/2}\, \lambda_i^{-1/2} - \lambda_i^{-1} > 0 & 1 \leq i \leq \tilde{L} \\ \mu^{-1/2}\, w_i^{1/2}\, \lambda_i^{-1/2} - \lambda_i^{-1} \leq 0 & \tilde{L} < i \leq \check{L} \end{cases}$$

or, more compactly,

$$\left(w_{\tilde{L}} \lambda_{\tilde{L}}\right)^{-1/2} < \mu^{-1/2} \leq \left(w_{\tilde{L}+1} \lambda_{\tilde{L}+1}\right)^{-1/2}$$

where we define $w_{\check{L}+1}\lambda_{\check{L}+1} \triangleq 0$ for simplicity of notation. Assuming that $w_{\tilde{L}}\lambda_{\tilde{L}} \neq w_{\tilde{L}+1}\lambda_{\tilde{L}+1}$ (otherwise, the hypothesis is clearly rejected since the set of possible water-levels is empty), the algorithm checks whether the power constraint cannot be satisfied with the minimum water-level in the limiting case, *i.e.*, $\mu = w_{\tilde{L}}\lambda_{\tilde{L}}$. Note that for this limiting value, the eigenmode $\tilde{L}$ becomes inactive which corresponds to a different hypothesis. Therefore, the current hypothesis is rejected when $\sum_{i=1}^{\tilde{L}} z_i \geq P_T$ for $\mu = w_{\tilde{L}}\lambda_{\tilde{L}}$ or, equivalently, when $\mu^{-1/2} \geq \frac{P_T + \sum_{i=1}^{\tilde{L}} \lambda_i^{-1}}{\sum_{i=1}^{\tilde{L}} w_i^{1/2} \lambda_i^{-1/2}}$. In such a case, the z_i's must be decreased or, equivalently, the water-level $\mu^{-1/2}$ must be decreased and the eigenmode $\tilde{L}$ becomes inactive which corresponds to the new hypothesis. Otherwise, the current hypothesis (the current set of active eigenmodes $1 \leq i \leq \tilde{L}$) is accepted since removing more active eigenmodes (decreasing the water-level) would further increase the objective function and the addition of more active eigenmodes has already been tested and rejected (for exceeding the maximum power). This reasoning can be applied as many times as needed for each remaining set of active eigenmodes. Once the optimal set of active eigenmodes is known, the value of the active z_i's is then increased to satisfy the power constraint with equality. The definitive water-level is then

$$\mu^{-1/2} = \frac{P_T + \sum_{i=1}^{\tilde{L}} \lambda_i^{-1}}{\sum_{i=1}^{\tilde{L}} w_i^{1/2} \lambda_i^{-1/2}}. \tag{5.83}$$

By the nature of the algorithm, the maximum number of iterations (worst-case complexity) is $\check{L}$. ■

Proof of Proposition 5.2 (Classical weighted water-filling.)

This proof very similar to that of Proposition 5.1 and, therefore, we simply give a sketch of the proof.

Optimal Solution. The Lagrangian corresponding to the constrained convex problem is

$$L = \sum_{i=1}^{\check{L}} -w_i \log(1 + \lambda_i z_i) + \mu \left(\sum_{i=1}^{\check{L}} z_i - P_T \right) - \sum_{i=1}^{\check{L}} \gamma_i z_i \tag{5.84}$$

where μ and $\{\gamma_i\}$ are the dual variables or Lagrange multipliers. The water-filling solution is easily found from the sufficient and necessary KKT optimality conditions (the problem satisfies the Slater's condition and therefore strong duality holds):

$$\begin{array}{ll} \sum_{i=1}^{\check{L}} z_i \le P_T, & z_i \ge 0, \\ \mu \ge 0, & \gamma_i \ge 0, \\ \mu = w_i \frac{\lambda_i}{1+\lambda_i z_i} + \gamma_i, & \\ \mu \left(\sum_{i=1}^{\check{L}} z_i - P_T \right) = 0, & \gamma_i z_i = 0. \end{array} \tag{5.85}$$

It is assumed that there is at least one $\lambda_i > 0$; otherwise the problem has trivial solution given by $z_i = 0 \ \forall i$. As in the proof of Proposition 5.1, it follows that $\mu > 0$ and, consequently, the power constraint must be satisfied with equality. It is simple to obtain the water-filling solution

$$z_i = \left(\mu^{-1} w_i - \lambda_i^{-1} \right)^+ \tag{5.86}$$

where μ^{-1} is the water-level chosen so that $\sum_{i=1}^{\check{L}} z_i = P_T$.

Optimal Algorithm. The optimality of Algorithm 5.1 is proved as in the proof of Proposition 5.1. In this case, however, the comparison $\sum_{i=1}^{\check{L}} z_i \ge P_T$ is given by $\mu^{-1} \ge \frac{P_T + \sum_{i=1}^{\check{L}} \lambda_i^{-1}}{\sum_{i=1}^{\check{L}} w_i}$. ■

Proof of Proposition 5.3 (Multi-level water-filling for the MAX-MSE criterion.)

Since the problem is convex, the optimal solution can be found using convex optimization theory (the reader is referred to §3.1 and [Lue69, Roc70, Boy00] for details and definitions). We first obtain the closed-form optimal solution to the problem and then proceed to prove the optimality of Algorithm 5.3.

Optimal Solution. The Lagrangian corresponding to the constrained convex problem is

$$\begin{aligned} L = t + \sum_{k=1}^{N} \mu_k \left(\frac{1}{L_k} \left((L_k - \check{L}_k) + \sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}} \right) - t \right) \\ + \mu_0 \left(\sum_{k=1}^{N} \sum_{i=1}^{\check{L}_k} z_{k,i} - P_T \right) - \sum_{k=1}^{N} \sum_{i=1}^{\check{L}_k} \gamma_{k,i} z_{k,i} \end{aligned} \tag{5.87}$$

where μ_0, $\{\mu_k\}$, and $\{\gamma_{k,i}\}$ are the dual variables or Lagrange multipliers. The multi-level water-filling solution is found from the sufficient and necessary KKT optimality conditions (the problem satisfies the Slater's condition and therefore strong duality holds):

$$\begin{array}{lll} t \geq \frac{1}{L_k} \left((L_k - \check{L}_k) + \sum_i \frac{1}{1+\lambda_{k,i} z_{k,i}} \right), & \sum_{k,i} z_{k,i} \leq P_T, & z_{k,i} \geq 0, \\ \mu_k \geq 0, & \mu_0 \geq 0, & \gamma_{k,i} \geq 0, \\ \mu_0 = \mu_k / L_k \frac{\lambda_{k,i}}{(1+\lambda_{k,i} z_{k,i})^2} + \gamma_{k,i}, & & \sum_k \mu_k = 1, \\ \mu_k \left(\frac{1}{L_k} \left((L_k - \check{L}_k) + \sum_i \frac{1}{1+\lambda_{k,i} z_{k,i}} \right) - t \right) = 0, & \mu_0 \left(\sum_{k,i} z_{k,i} - P_T \right) = 0, & \gamma_{k,i} z_{k,i} = 0. \end{array} \tag{5.88}$$

It is assumed that for each k there is at least one eigenmode with nonzero gain $\lambda_{k,i} > 0$ (otherwise the problem has trivial solution given by $t = 1$ and $z_{k,i} = 0 \ \forall k, i$).

First of all, we show that μ_0 and $\{\mu_k\}$ must be positive values and consequently (by the complementary slackness conditions) the constraints on t and the power constraint must be satisfied with equality (this can also be concluded from a direct inspection of the problem formulation). If $\mu_0 = 0$, then $\gamma_{k,i} = -\mu_k / L_k \frac{\lambda_{k,i}}{(1+\lambda_{k,i} z_{k,i})^2}$. This clearly implies $\gamma_{k,i} = 0 \ \forall k, i$ and since for each k it is assumed that $\lambda_{k,i} > 0$ for some i, it also implies $\mu_k = 0 \ \forall k$, which cannot possibly be because of the constraint $\sum_k \mu_k = 1$ and consequently $\mu_0 > 0$. Note that in case $\lambda_{k,i} = 0$ for some k and i, then $\gamma_{k,i} = \mu_0 > 0 \Rightarrow z_{k,i} = 0$, *i.e.*, no power is allocated to eigenmodes with zero gain (as expected). From now on, we consider that all zero-valued $\lambda_{k,i}$'s have been removed. Now, if $\mu_k = 0$ for some k, then $\gamma_{k,i} = \mu_0 > 0 \ \forall i \Rightarrow z_{k,i} = 0 \ \forall i$, and $t \geq \frac{1}{L_k} \left((L_k - \check{L}_k) + \sum_i \frac{1}{1+\lambda_{k,i} z_{k,i}} \right) = 1$. Consider now that $\mu_l > 0$ for some $l \neq k$, this implies that $t = \frac{1}{L_l} \left((L_l - \check{L}_l) + \sum_i \frac{1}{1+\lambda_{l,i} z_{l,i}} \right)$ (by the complementary slackness condition) which can be if only if $t = 1$ and $z_{l,i} = 0 \ \forall i$. As a consequence, if $\mu_k = 0$ for some k, it would then follow that $z_{k,i} = 0 \ \forall k, i$ which does not agree with constraint $\sum_{k,i} z_{k,i} = P_T$. Therefore, it must be that $\mu_k > 0 \ \forall k$.

If $z_{k,i} > 0$, then $\gamma_{k,i} = 0$ (by the complementary slackness condition), $\mu_0 = \mu_k / L_k \frac{\lambda_{k,i}}{(1+\lambda_{k,i} z_{k,i})^2}$ (note that $\mu_0 < \mu_k / L_k \lambda_{k,i}$), and $z_{k,i} = (\mu_k / (L_k \mu_0))^{1/2} \ \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}$. If $z_{k,i} = 0$, then $\mu_0 =$

$\mu_k/L_k\lambda_{k,i} + \gamma_{k,i}$ (note that $\mu_0 \geq \mu_k/L_k\lambda_{k,i}$). Equivalently,

$$z_{k,i} = \begin{cases} (\mu_k/(L_k\mu_0))^{1/2}\, \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1} & \text{if} \quad \mu_0 < \mu_k/L_k\lambda_{k,i} \\ 0 & \text{if} \quad \mu_0 \geq \mu_k/L_k\lambda_{k,i} \end{cases}$$

or, more compactly,

$$z_{k,i} = \left((\mu_k/(L_k\mu_0))^{1/2}\, \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1} \right)^+$$

where μ_0 and $\{\mu_k\}$ are positive values so that the constraints on t and the power constraint are satisfied with equality and $\sum_k \mu_k = 1$.

Defining $\bar{\mu}_k \triangleq \mu_k/(L_k\mu_0)$ (note that given $\{\bar{\mu}_k\}$ one can always find μ_0 and $\{\mu_k\}$ as $\mu_k = L_k\mu_0\, \bar{\mu}_k$ and $\mu_0 = 1/\sum_k (L_k\bar{\mu}_k)$), the optimal solution can be finally expressed as

$$z_{k,i} = \left(\bar{\mu}_k^{1/2}\, \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1} \right)^+ \tag{5.89}$$

where $\{\bar{\mu}_k^{1/2}\}$ are positive water-levels chosen to satisfy

$$\begin{aligned} & t = \tfrac{1}{L_k} \left(\left(L_k - \check{L}_k \right) + \textstyle\sum_{i=1}^{\check{L}_k} \tfrac{1}{1+\lambda_{k,i} z_{k,i}} \right) \qquad 1 \leq k \leq N, \\ & \textstyle\sum_{k,i} z_{k,i} = P_T. \end{aligned} \tag{5.90}$$

Note that this solution satisfies all KKT conditions and is therefore optimal.

Optimal Algorithm. Algorithm 5.3 is based on hypothesis testing. It first makes the hypothesis that all $\check{L}_T = \sum_k \check{L}_k$ eigenmodes are active ($z_{k,i} > 0\ \forall k, i$), imposes the equality constraints on t, and then checks whether the maximum power is exceeded, in which case the hypothesis is rejected, a new hypothesis is made with one less active eigenmode, and so forth.

To formally derive Algorithm 5.3, we first simplify the problem by properly defining the hypotheses. If the eigenmodes are ordered in decreasing order for each k (*i.e.*, $\lambda_{k,i} \geq \lambda_{k,i+1}$) and since $\lambda_{k,i}\, z_{k,i} = \left(\bar{\mu}_k^{1/2}\, \lambda_{k,i}^{1/2} - 1 \right)^+$, a hypothesis is completely described by the set of active eigenmodes $\{\tilde{L}_k\}$ (such that $z_{k,i} > 0$ for $1 \leq k \leq N,\ 1 \leq i \leq \tilde{L}_k$ and zero otherwise). This ordering of the eigenmodes, which we assume in the derivation of the algorithm, allows a reduction of the total number of hypotheses from $\prod_k 2^{\check{L}_k} = 2^{\sum_k \check{L}_k}$ to $\prod_k \check{L}_k$. Using the fact that $\sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}} = \left(\check{L}_k - \tilde{L}_k \right) + \bar{\mu}_k^{-1/2} \sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1/2}$, the constraints on t can be rewritten as

$$t = \frac{1}{L_k} \left(\left(L_k - \tilde{L}_k \right) + \bar{\mu}_k^{-1/2} \sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1/2} \right) \qquad 1 \leq k \leq N \tag{5.91}$$

and the constraints of the original problem (5.90) can be rewritten by parameterizing the water-levels $\{\bar{\mu}_k^{1/2}\}$ with respect to t as

$$\begin{aligned} & \bar{\mu}_k^{1/2} = \tfrac{\sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1/2}}{tL_k - (L_k - \tilde{L}_k)} \qquad 1 \leq k \leq N, \\ & \textstyle\sum_{k,i} z_{k,i} = P_T. \end{aligned} \tag{5.92}$$

Note that the first constraint of (5.92) is nontrivial since each water-level $\bar{\mu}_k^{1/2}$ is obtained as a function of the parameter t and also of $\tilde{L}_k$, but at the same time $\tilde{L}_k$ depends on the water-level $\bar{\mu}_k^{1/2}$.

Consider a hypothesis given by $\{\tilde{L}_k\}$. The water-levels for such a hypothesis must be such that the considered $\tilde{L}_k$ eigenmodes are indeed active while the rest remain inactive:

$$\begin{cases} \bar{\mu}_k^{1/2}\,\lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1} > 0 & 1 \le k \le N,\ 1 \le i \le \tilde{L}_k \\ \bar{\mu}_k^{1/2}\,\lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1} \le 0 & 1 \le k \le N,\ \tilde{L}_k < i \le \check{L}_k \end{cases}$$

$$\iff \qquad \lambda_{k,\tilde{L}_k+1}^{1/2} \le \bar{\mu}_k^{-1/2} < \lambda_{k,\tilde{L}_k}^{1/2} \qquad 1 \le k \le N \tag{5.93}$$

where we define $\lambda_{k,\check{L}_k+1} \triangleq 0\ \forall k$ for simplicity of notation. These constraints on the water-levels implicitly constrain the parameter t. Combining (5.93) and (5.91), the range of possible values for t is given by

$$t_k^{\mathrm{lb}}\big(\tilde{L}_k\big) \le t < t_k^{\mathrm{ub}}\big(\tilde{L}_k\big) \qquad 1 \le k \le N \tag{5.94}$$

or, equivalently,

$$t \in \bigcap_{k=1}^{N} \left[t_k^{\mathrm{lb}}\big(\tilde{L}_k\big),\ t_k^{\mathrm{ub}}\big(\tilde{L}_k\big)\right) \tag{5.95}$$

where $t_k^{\mathrm{lb}}\big(\tilde{L}_k\big) = \frac{1}{L_k}\left(\big(L_k - \tilde{L}_k\big) + \lambda_{k,\tilde{L}_k+1}^{1/2}\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)$ and $t_k^{\mathrm{ub}}\big(\tilde{L}_k\big) = \frac{1}{L_k}\left(\big(L_k - \tilde{L}_k\big) + \lambda_{k,\tilde{L}_k}^{1/2}\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)$. More compactly, we can write

$$t \in \left[t^{\mathrm{max\text{-}lb}}\big(\{\tilde{L}_k\}\big),\ t^{\mathrm{min\text{-}ub}}\big(\{\tilde{L}_k\}\big)\right) \tag{5.96}$$

where $t^{\mathrm{max\text{-}lb}}\big(\{\tilde{L}_k\}\big) = \max_{1\le k\le N}\left\{t_k^{\mathrm{lb}}\big(\tilde{L}_k\big)\right\}$ and $t^{\mathrm{min\text{-}ub}}\big(\{\tilde{L}_k\}\big) = \min_{1\le k\le N}\left\{t_k^{\mathrm{ub}}\big(\tilde{L}_k\big)\right\}$. Denote the maximizing k by $k_{\max}$ and the minimizing k by $k_{\min}$. It is also important to notice that $t_k^{\mathrm{ub}}\big(\tilde{L}_k + 1\big) = t_k^{\mathrm{lb}}\big(\tilde{L}_k\big)$ which implies that the different hypotheses partition the whole range of possible values of t for a given k:

$$\begin{array}{l} \bigcup_{\tilde{L}_k=1}^{\check{L}_k}\left[t_k^{\mathrm{lb}}\big(\tilde{L}_k\big),\ t_k^{\mathrm{ub}}\big(\tilde{L}_k\big)\right) = \left[t_k^{\mathrm{lb}}\big(\check{L}_k\big),\ t_k^{\mathrm{ub}}(1)\right) = \left[\big(1 - \check{L}_k/L_k\big),\ 1\right) \\ \left[t_k^{\mathrm{lb}}\big(\tilde{L}_k^{(1)}\big),\ t_k^{\mathrm{ub}}\big(\tilde{L}_k^{(1)}\big)\right) \bigcap \left[t_k^{\mathrm{lb}}\big(\tilde{L}_k^{(2)}\big),\ t_k^{\mathrm{ub}}\big(\tilde{L}_k^{(2)}\big)\right) = \varnothing \qquad \tilde{L}_k^{(1)} \ne \tilde{L}_k^{(2)}. \end{array}$$

In other words, the partitioning of the interval $\left[\big(1 - \check{L}_k/L_k\big),\ 1\right)$ is given by (see Figure 5.15)

$$\ldots\left[t_k^{\mathrm{lb}}\big(\tilde{L}_k+1\big),\ t_k^{\mathrm{ub}}\big(\tilde{L}_k+1\big)\right),\ \left[t_k^{\mathrm{lb}}\big(\tilde{L}_k\big),\ t_k^{\mathrm{ub}}\big(\tilde{L}_k\big)\right),\ \left[t_k^{\mathrm{lb}}\big(\tilde{L}_k-1\big),\ t_k^{\mathrm{ub}}\big(\tilde{L}_k-1\big)\right),\ \ldots \tag{5.97}$$

We are now ready to derive an efficient optimal algorithm which starts with the smallest possible value of t (highest values of the $\tilde{L}_k$'s) and then increases it until the power constraint is satisfied. In more detail, Algorithm 5.3 first reorders the positive eigenmodes in decreasing order for each k by setting $\tilde{L}_k = \check{L}_k$ for $1 \le k \le N$ to obtain the initial hypothesis.

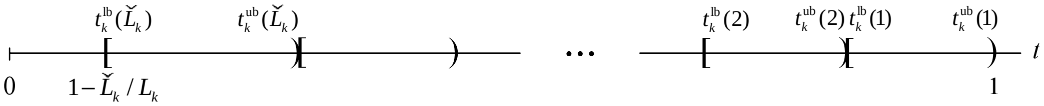


Figure 5.15: Illustrative scheme of the partition of the domain of t for a given index k.

At this point, if $t^{\text{max-lb}}\big(\{\tilde{L}_k\}\big) \geq t^{\text{min-ub}}\big(\{\tilde{L}_k\}\big)$, then the considered hypothesis cannot satisfy the optimality conditions by (5.96) and must be rejected. Otherwise, if $t^{\text{max-lb}}\big(\{\tilde{L}_k\}\big) < t^{\text{min-ub}}\big(\{\tilde{L}_k\}\big)$, then we still have to check whether there is some value of $t \in \Big[t^{\text{max-lb}}\big(\{\tilde{L}_k\}\big),\, t^{\text{min-ub}}\big(\{\tilde{L}_k\}\big)\Big)$ such that the power constraint is satisfied, which can be done simply by checking if with $t^{\text{min-ub}}\big(\{\tilde{L}_k\}\big)$ the power constraint is strictly satisfied (since higher values of t require less power). From $\bar{\mu}_k^{1/2} = \frac{\sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1/2}}{tL_k-(L_k-\tilde{L}_k)}$ and $z_{k,i} = \left(\bar{\mu}_k^{1/2}\, \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}\right)^+$, the strict power constraint condition $\sum_{k=1}^{N} \sum_{i=1}^{\tilde{L}_k} z_{k,i} < P_T$ can be written as

$$\sum_{k=1}^{N} \frac{\left(\sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1/2}\right)^2}{tL_k - \big(L_k - \tilde{L}_k\big)} < P_T + \sum_{k=1}^{N} \sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1}. \tag{5.98}$$

If the power constraint is not satisfied for $t = t^{\text{min-ub}}\big(\{\tilde{L}_k\}\big)$, the hypothesis cannot satisfy the optimality conditions and must be rejected. Otherwise, the hypothesis is accepted (as is later shown, such a hypothesis is in fact the optimal) and the water-levels are recomputed to satisfy the power constraint with equality, *i.e.*,

$$t : \sum_{k=1}^{N} \frac{\left(\sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1/2}\right)^2}{L_k t - \big(L_k - \tilde{L}_k\big)} = P_T + \sum_{k=1}^{N} \sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1}, \tag{5.99}$$

$$\left(\text{recall that } t \in \left[t^{\text{max-lb}}\big(\{\tilde{L}_k\}\big),\, t^{\text{min-ub}}\big(\{\tilde{L}_k\}\big)\right]\right)$$

$$\bar{\mu}_k^{1/2} = \frac{\sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1/2}}{L_k t - \big(L_k - \tilde{L}_k\big)}, \text{ and} \tag{5.100}$$

$$z_{k,i} = \left(\bar{\mu}_k^{1/2}\, \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}\right)^+. \tag{5.101}$$

After a hypothesis has been rejected, a new one has to be made. We now show by induction how to do it in an efficient way such that the total number of hypotheses is reduced from $\prod_k \check{L}_k$ to $\sum_k \check{L}_k$ (recall that the initial number of hypotheses was $2^{\sum_k \check{L}_k}$). Assume that hypothesis $\{\tilde{L}_k\}$ has been rejected and that the optimal hypothesis (the one that contains the optimal t) is known not to contain higher values for any of the $\tilde{L}_k$'s (this clearly holds for the initial hypothesis since the values of the $\tilde{L}_k$'s are chosen as the highest possible). As a consequence, a new hypothesis can be made only by decreasing some of the $\tilde{L}_k$'s. Noting that if any of the $\tilde{L}_k$'s is decreased to some value $\tilde{L}_k^{\text{new}}$ then $t_k^{\text{lb}}\big(\tilde{L}_k^{\text{new}}\big) \geq t_k^{\text{ub}}\big(\tilde{L}_k\big)$ (from (5.97)) and also that $t_k^{\text{ub}}\big(\tilde{L}_k\big) \geq t_{k_{\min}}^{\text{ub}}\big(\tilde{L}_{k_{\min}}\big)$ (by definition of $k_{\min}$), it follows that $t_k^{\text{lb}}\big(\tilde{L}_k^{\text{new}}\big) \geq t_{k_{\min}}^{\text{ub}}\big(\tilde{L}_{k_{\min}}\big)$. This means that if a new hypothesis is made by decreasing some of the $\tilde{L}_k$'s for $k \neq k_{\min}$, then the range of possible values for t will be empty since $\left[t_{k_{\min}}^{\mathrm{lb}}\left(\tilde{L}_{k_{\min}}\right),\, t_{k_{\min}}^{\mathrm{ub}}\left(\tilde{L}_{k_{\min}}\right)\right) \bigcap \left[t_{k}^{\mathrm{lb}}\left(\tilde{L}_{k}^{\mathrm{new}}\right),\, t_{k}^{\mathrm{ub}}\left(\tilde{L}_{k}^{\mathrm{new}}\right)\right) = \varnothing$. Thus, we can guarantee that the optimal hypothesis must have a lower value for $\tilde{L}_{k_{\min}}$ of at most $\tilde{L}_{k_{\min}} - 1$, which is taken as the next hypothesis to evaluate. Since it was assumed that the optimal hypothesis was known not to contain higher values for any of the $\tilde{L}_k$'s of the original hypothesis, the new hypothesis also satisfies this condition (since it only differs in $\tilde{L}_{k_{\min}}$ and we have just shown that the optimal hypothesis has a value of $\tilde{L}_{k_{\min}}$ of at most $\tilde{L}_{k_{\min}} - 1$). Therefore, by induction (recall that the initial hypothesis also satisfies this condition), the previous mechanism to generate new hypotheses can be repeatedly applied.

It is simple to see that the accepted hypothesis must be the optimal one since removing more active eigenmodes (decreasing water-levels and increasing t) would further increase t which is the objective to minimize and adding more active eigenmodes has already been checked and rejected.

By the nature of the algorithm, the maximum number of iterations (worst-case complexity) is $\sum_k \check{L}_k$. ∎

Proof of Proposition 5.4 (Multi-level water-filling for the HARM-SINR criterion.)

Since the problem is convex, the optimal solution can be found using convex optimization theory (the reader is referred to §3.1 and [Lue69, Roc70, Boy00] for details and definitions). We first obtain the closed-form optimal solution to the problem and then proceed to prove the optimality of Algorithm 5.4.

Optimal Solution. The Lagrangian corresponding to the constrained convex problem is

$$\begin{aligned} L = & \sum_{k=1}^{N} \frac{t_k}{L_k - t_k} + \sum_{k=1}^{N} \mu_k \left(\left(L_k - \check{L}_k\right) + \sum_{i=1}^{\check{L}_k} \frac{1}{1 + \lambda_{k,i} z_{k,i}} - t_k \right) + \sum_{k=1}^{N} \alpha_k \left(t_k - L_k\right) \\ & + \mu_0 \left(\sum_{k=1}^{N} \sum_{i=1}^{\check{L}_k} z_{k,i} - P_T \right) - \sum_{k=1}^{N} \sum_{i=1}^{\check{L}_k} \gamma_{k,i} z_{k,i} \end{aligned} \tag{5.102}$$

where μ_0, $\{\mu_k\}$, $\{\alpha_k\}$, and $\{\gamma_{k,i}\}$ are the dual variables or Lagrange multipliers. The multi-level water-filling solution is found from the sufficient and necessary KKT optimality conditions (the problem satisfies the Slater's condition and therefore strong duality holds):

$$
\begin{aligned}
& t_k \geq \left(L_k - \check{L}_k\right) + \sum_i \tfrac{1}{1+\lambda_{k,i} z_{k,i}}, && t_k \leq L_k, && \sum_{k,i} z_{k,i} \leq P_T, && z_{k,i} \geq 0, \\
& \mu_k \geq 0, && \alpha_k \geq 0, && \mu_0 \geq 0, && \gamma_{k,i} \geq 0, \\
& \mu_0 = \mu_k \tfrac{\lambda_{k,i}}{\left(1+\lambda_{k,i} z_{k,i}\right)^2} + \gamma_{k,i}, && && \mu_k = \tfrac{L_k}{\left(L_k - t_k\right)^2} + \alpha_k, \\
& \mu_k \left(\left(L_k - \check{L}_k\right) + \sum_i \tfrac{1}{1+\lambda_{k,i} z_{k,i}} - t_k \right) = 0, && \alpha_k \left(t_k - L_k\right) = 0, \\
& && \mu_0 \left(\sum_{k,i} z_{k,i} - P_T \right) = 0, && \gamma_{k,i} z_{k,i} = 0.
\end{aligned}
\tag{5.103}
$$

Note that although the problem formulation allows $t_k = L_k$, this will never happen at an optimal solution since the objective would become infinite. Therefore, $t_k < L_k \ \forall k$ (and consequently $\alpha_k = 0 \ \forall k$). It is assumed that for each k there is at least one eigenmode with a nonzero gain $\check{L}_k > 0$ (otherwise the problem has trivial solution given by $t_k = L_k$ and $z_{k,i} = 0 \ \forall k, i$).

First, we show that μ_0 and $\{\mu_k\}$ must be positive values and consequently the lower constraints on the t_k's and the power constraint must be satisfied with equality. From the KKT conditions, $\mu_k = \frac{L_k}{(L_k - t_k)^2} + \alpha_k \geq \frac{L_k}{(L_k - t_k)^2} \geq \frac{1}{L_k} > 0 \ \forall k$ and $\mu_0 = \mu_k \frac{\lambda_{k,i}}{\left(1+\lambda_{k,i} z_{k,i}\right)^2} + \gamma_{k,i} \geq \mu_k \frac{\lambda_{k,i}}{\left(1+\lambda_{k,i} z_{k,i}\right)^2} > 0$ at least for some (k, i) for which $\lambda_{k,i} > 0$. Note that if $\lambda_{k,i} = 0$ for some k and i, then $\gamma_{k,i} = \mu_0 > 0 \Rightarrow z_{k,i} = 0$, *i.e.*, no power is allocated to eigenmodes with zero gain (as expected). From now on, we consider that all zero-valued $\lambda_{k,i}$'s have been removed.

If $z_{k,i} > 0$, then $\gamma_{k,i} = 0$ (by the complementary slackness condition), $\mu_0 = \mu_k \frac{\lambda_{k,i}}{\left(1+\lambda_{k,i} z_{k,i}\right)^2}$ (note that $\mu_0 < \mu_k \lambda_{k,i}$), and $z_{k,i} = (\mu_k/\mu_0)^{1/2} \ \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}$. If $z_{k,i} = 0$, then $\mu_0 = \mu_k \lambda_{k,i} + \gamma_{k,i}$ (note that $\mu_0 \geq \mu_k \lambda_{k,i}$). Equivalently,

$$
z_{k,i} = \begin{cases} (\mu_k/\mu_0)^{1/2} \ \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1} & \text{if} \quad \mu_0 < \mu_k \lambda_{k,i} \\ 0 & \text{if} \quad \mu_0 \geq \mu_k \lambda_{k,i} \end{cases}
$$

or, more compactly

$$
z_{k,i} = \left((\mu_k/\mu_0)^{1/2} \ \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1} \right)^+
$$

where μ_0 and $\{\mu_k\}$ are positive values so that the lower constraints on the t_k's and the power constraint are satisfied with equality, and $\mu_k = \frac{L_k}{(L_k - t_k)^2}$.

Defining $\bar{\mu}_k \triangleq \mu_k/\mu_0$ and $\nu \triangleq \mu_0^{-1/2}$, the optimal solution can be finally expressed as

$$
z_{k,i} = \left(\bar{\mu}_k^{1/2} \ \lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1} \right)^+ \tag{5.104}
$$

where ν is a positive parameter and $\{\bar{\mu}_k^{1/2}\}$ are positive water-levels chosen to satisfy:

$$
\begin{aligned}
& t_k = \left(L_k - \check{L}_k\right) + \textstyle\sum_{i=1}^{\check{L}_k} \frac{1}{1+\lambda_{k,i} z_{k,i}} \qquad 1 \leq k \leq N, \\
& \bar{\mu}_k^{1/2} = \nu \, \frac{L_k^{1/2}}{L_k - t_k}, \\
& \textstyle\sum_{k,i} z_{k,i} = P_T.
\end{aligned}
\tag{5.105}
$$

Note that this solution satisfy all KKT conditions and is therefore optimal.

Optimal Algorithm. Algorithm 5.4 is based on hypothesis testing. It first makes the assumption that all $\check{L}_T = \sum_k \check{L}_k$ eigenmodes are active ($z_{k,i} > 0\ \forall k, i$), imposes the two first constraints of (5.105), and then checks whether the maximum power is exceeded, in which case the hypothesis is rejected, a new hypothesis is made with one less active eigenmode, and so forth. The derivation of Algorithm 5.4 is very similar to that of Algorithm 5.3 obtained in the proof of Proposition 5.3.

To formally derive Algorithm 5.4, we first simplify the problem by properly defining the hypotheses. If the eigenmodes are ordered in decreasing order for each k (*i.e.*, $\lambda_{k,i} \geq \lambda_{k,i+1}$) and since $\lambda_{k,i}\, z_{k,i} = \left(\bar{\mu}_k^{1/2}\, \lambda_{k,i}^{1/2} - 1\right)^+$, a hypothesis is completely described by the set of active eigenmodes $\{\tilde{L}_k\}$ (such that $z_{k,i} > 0$ for $1 \leq k \leq N,\ 1 \leq i \leq \tilde{L}_k$ and zero otherwise). This ordering of the eigenmodes, which we assume in the derivation of the algorithm, allows a reduction of the total number of hypotheses from $\prod_k 2^{\check{L}_k} = 2^{\sum_k \check{L}_k}$ to $\prod_k \check{L}_k$. The two first constraints of (5.105) can be rewritten as

$$\left.\begin{aligned} t_k &= \left(L_k - \tilde{L}_k\right) + \bar{\mu}_k^{-1/2} \textstyle\sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1/2} \\ t_k &= L_k - \nu\, L_k^{1/2}\, \bar{\mu}_k^{-1/2} \end{aligned}\right. \qquad 1 \leq k \leq N$$

from which the following constraints on ν are obtained:

$$\nu = \frac{1}{L_k^{1/2}} \left(\tilde{L}_k\, \bar{\mu}_k^{1/2} - \sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1/2} \right) \qquad 1 \leq k \leq N. \tag{5.106}$$

The constraints of the original problem (5.105) can be rewritten in a simpler way by parameterizing the water-levels $\{\bar{\mu}_k^{1/2}\}$ with respect to ν as

$$\begin{aligned} &\bar{\mu}_k^{1/2} = \tfrac{1}{\tilde{L}_k} \left(\nu\, L_k^{1/2} + \textstyle\sum_{i=1}^{\tilde{L}_k} \lambda_{k,i}^{-1/2}\right) \qquad 1 \leq k \leq N, \\ &\textstyle\sum_{k,i} z_{k,i} = P_T. \end{aligned} \tag{5.107}$$

Note that the first constraint of (5.107) is nontrivial since each water-level $\bar{\mu}_k^{1/2}$ is obtained as a function of the parameter ν and also of $\tilde{L}_k$, but at the same time $\tilde{L}_k$ depends on the water-level $\bar{\mu}_k^{1/2}$. It is important to notice that any solution satisfying the simplified problem given by (5.107) and (5.104) also satisfies the original problem given by (5.105) and (5.104). Consequently, any solution satisfying (5.107) and (5.104) is optimal since satisfies all KKT conditions (to see this simply undo the steps taken to go from (5.105) to (5.107)).

Consider a hypothesis given by $\{\tilde{L}_k\}$. The water-levels for such a hypothesis must be such that the considered $\tilde{L}_k$ eigenmodes are indeed active while the rest remain inactive (see (5.93)):

$$\lambda_{k,\tilde{L}_k+1}^{-1/2} \geq \bar{\mu}_k^{1/2} > \lambda_{k,\tilde{L}_k}^{-1/2} \qquad 1 \leq k \leq N \tag{5.108}$$

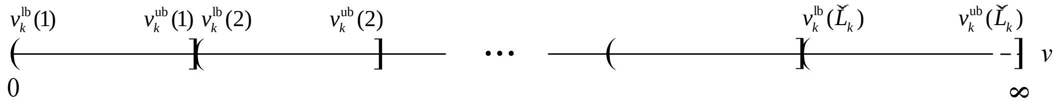


Figure 5.16: Illustrative scheme of the partition of the domain of ν for a given index k.

where we define $\lambda_{k,\check{L}_k+1} \triangleq 0 \ \forall k$ for simplicity of notation. These constraints on the water-levels implicitly constrain the parameter ν. Combining (5.108) and (5.106), the range of possible values for ν is given by

$$\nu_k^{\mathrm{lb}}\big(\tilde{L}_k\big) < \nu \le \nu_k^{\mathrm{ub}}\big(\tilde{L}_k\big) \qquad 1 \le k \le N \tag{5.109}$$

where the lower bound is $\nu_k^{\mathrm{lb}}\big(\tilde{L}_k\big) = \frac{1}{L_k^{1/2}}\left(\tilde{L}_k\,\lambda_{k,\tilde{L}_k}^{-1/2} - \sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)$ and the upper bound $\nu_k^{\mathrm{ub}}\big(\tilde{L}_k\big) = \frac{1}{L_k^{1/2}}\left(\tilde{L}_k\,\lambda_{k,\tilde{L}_k+1}^{-1/2} - \sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)$. More compactly, we can write

$$\nu \in \left(\nu^{\max\text{-}\mathrm{lb}}\big(\{\tilde{L}_k\}\big),\ \nu^{\min\text{-}\mathrm{ub}}\big(\{\tilde{L}_k\}\big)\right] \tag{5.110}$$

where $\nu^{\max\text{-}\mathrm{lb}}\big(\{\tilde{L}_k\}\big) = \max_{1\le k\le N}\left\{\nu_k^{\mathrm{lb}}\big(\tilde{L}_k\big)\right\}$ and $\nu^{\min\text{-}\mathrm{ub}}\big(\{\tilde{L}_k\}\big) = \min_{1\le k\le N}\left\{\nu_k^{\mathrm{ub}}\big(\tilde{L}_k\big)\right\}$. Denote the maximizing k by $k_{\max}$ and the minimizing k by $k_{\min}$. It is also important to notice that $\nu_k^{\mathrm{ub}}\big(\tilde{L}_k\big) = \nu_k^{\mathrm{lb}}\big(\tilde{L}_k+1\big)$ which implies that the different hypotheses partition the whole range of possible values of ν for a given k (similarly to (5.97)):

$$\begin{array}{l} \bigcup_{\tilde{L}_k=1}^{\check{L}_k}\left(\nu_k^{\mathrm{lb}}\big(\tilde{L}_k\big),\ \nu_k^{\mathrm{ub}}\big(\tilde{L}_k\big)\right] = \left(\nu_k^{\mathrm{lb}}(1),\ \nu_k^{\mathrm{ub}}\big(\check{L}_k\big)\right] = (0,\ \infty] \\ \left(\nu_k^{\mathrm{lb}}\big(\tilde{L}_k^{(1)}\big),\ \nu_k^{\mathrm{ub}}\big(\tilde{L}_k^{(1)}\big)\right] \bigcap \left(\nu_k^{\mathrm{lb}}\big(\tilde{L}_k^{(2)}\big),\ \nu_k^{\mathrm{ub}}\big(\tilde{L}_k^{(2)}\big)\right] = \varnothing \qquad \tilde{L}_k^{(1)} \ne \tilde{L}_k^{(2)}. \end{array}$$

In other words, the partitioning of the interval $(0,\ \infty]$ is given by (see Figure 5.16)

$$\ldots\left(\nu_k^{\mathrm{lb}}\big(\tilde{L}_k-1\big),\ \nu_k^{\mathrm{ub}}\big(\tilde{L}_k-1\big)\right],\ \left(\nu_k^{\mathrm{lb}}\big(\tilde{L}_k\big),\ \nu_k^{\mathrm{ub}}\big(\tilde{L}_k\big)\right],\ \left(\nu_k^{\mathrm{lb}}\big(\tilde{L}_k+1\big),\ \nu_k^{\mathrm{ub}}\big(\tilde{L}_k+1\big)\right],\ \ldots \tag{5.111}$$

We are now ready to derive an efficient optimal algorithm which starts with the highest possible value of ν (highest values of the $\tilde{L}_k$'s) and then decreases it until the power constraint is satisfied. In more detail, Algorithm 5.4 first reorders the positive eigenmodes in decreasing order for each k by setting $\tilde{L}_k = \check{L}_k$ for $1 \le k \le N$ to obtain the initial hypothesis.

At this point, if $\nu^{\max\text{-}\mathrm{lb}}\big(\{\tilde{L}_k\}\big) \ge \nu^{\min\text{-}\mathrm{ub}}\big(\{\tilde{L}_k\}\big)$, then the considered hypothesis cannot satisfy the optimality conditions by (5.110) and must be rejected. Otherwise, if $\nu^{\max\text{-}\mathrm{lb}}\big(\{\tilde{L}_k\}\big) < \nu^{\min\text{-}\mathrm{ub}}\big(\{\tilde{L}_k\}\big)$, then we still have to check whether there is some value of $\nu \in \left(\nu^{\max\text{-}\mathrm{lb}}\big(\{\tilde{L}_k\}\big),\ \nu^{\min\text{-}\mathrm{ub}}\big(\{\tilde{L}_k\}\big)\right]$ such that the power constraint is satisfied, which can be done simply by checking if with $\nu^{\max\text{-}\mathrm{lb}}\big(\{\tilde{L}_k\}\big)$ the power constraint is strictly satisfied (since lower values of ν means lower water-levels which require less power). From $\bar{\mu}_k^{1/2} = \frac{1}{\tilde{L}_k}\left(\nu\,L_k^{1/2} + \sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)$ and $z_{k,i} = \left(\bar{\mu}_k^{1/2}\,\lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}\right)^{+}$, the strict power constraint condition $\sum_{k=1}^{N}\sum_{i=1}^{\tilde{L}_k} z_{k,i} < P_T$ can be written as

$$\nu < \frac{P_T + \sum_{k=1}^{N}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1} - \frac{1}{\tilde{L}_k}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)^2\right)}{\sum_{k=1}^{N}\frac{L_k^{1/2}}{\tilde{L}_k}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)}. \tag{5.112}$$

If the power constraint is not satisfied for $\nu = \nu^{\text{max-lb}}\left(\{\tilde{L}_k\}\right)$, the hypothesis cannot satisfy the optimality conditions and must be rejected. Otherwise, the hypothesis is accepted (as is later shown, such a hypothesis is in fact the optimal) and the water-levels are recomputed to satisfy the power constraint with equality, *i.e.*,

$$\nu = \frac{P_T + \sum_{k=1}^{N}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1} - \frac{1}{\tilde{L}_k}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)^2\right)}{\sum_{k=1}^{N}\frac{L_k^{1/2}}{\tilde{L}_k}\left(\sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right)} \tag{5.113}$$

$$\bar{\mu}_k^{1/2} = \frac{1}{\tilde{L}_k}\left(\nu\, L_k^{1/2} + \sum_{i=1}^{\tilde{L}_k}\lambda_{k,i}^{-1/2}\right), \text{ and} \tag{5.114}$$

$$z_{k,i} = \left(\bar{\mu}_k^{1/2}\,\lambda_{k,i}^{-1/2} - \lambda_{k,i}^{-1}\right)^{+}. \tag{5.115}$$

After a hypothesis has been rejected, a new one has to be made. We now show by induction how to do it in an efficient way so that the total number of hypotheses is reduced from $\prod_k \check{L}_k$ to $\sum_k \check{L}_k$ (recall that the initial number of hypotheses was $2^{\sum_k \check{L}_k}$). Assume that hypothesis $\{\tilde{L}_k\}$ has been rejected and that the optimal hypothesis (the one that contains the optimal ν) is known not to contain higher values for any of the $\tilde{L}_k$'s (this clearly holds for the initial hypothesis since the values of the $\tilde{L}_k$'s are chosen as the highest possible). As a consequence, a new hypothesis can be made only by decreasing some of the $\tilde{L}_k$'s. Noting that if any of the $\tilde{L}_k$'s is decreased to some value $\tilde{L}_k^{\text{new}}$ then $\nu_k^{\text{ub}}\left(\tilde{L}_k^{\text{new}}\right) \le \nu_k^{\text{lb}}\left(\tilde{L}_k\right)$ (from (5.111)) and also that $\nu_k^{\text{lb}}\left(\tilde{L}_k\right) \le \nu_{k_{\max}}^{\text{lb}}\left(\tilde{L}_{k_{\max}}\right)$ (by definition of $k_{\max}$), it follows that $\nu_k^{\text{ub}}\left(\tilde{L}_k^{\text{new}}\right) \le \nu_{k_{\max}}^{\text{lb}}\left(\tilde{L}_{k_{\max}}\right)$. This means that if a new hypothesis is made by decreasing some of the $\tilde{L}_k$'s for $k \neq k_{\max}$, then the range of possible values for ν will be empty since $\left[\nu_k^{\text{lb}}\left(\tilde{L}_k^{\text{new}}\right), \nu_k^{\text{ub}}\left(\tilde{L}_k^{\text{new}}\right)\right) \bigcap \left[\nu_{k_{\max}}^{\text{lb}}\left(\tilde{L}_{k_{\max}}\right), \nu_{k_{\max}}^{\text{ub}}\left(\tilde{L}_{k_{\max}}\right)\right) = \emptyset$. Thus, we can guarantee that the optimal hypothesis must have a lower value for $\tilde{L}_{k_{\max}}$ of at most $\tilde{L}_{k_{\max}}-1$, which is taken as the next hypothesis to evaluate. Since it was assumed that the optimal hypothesis was known not to contain higher values for any of the $\tilde{L}_k$'s of the original hypothesis, the new hypothesis also satisfies this condition (since it only differs in $\tilde{L}_{k_{\max}}$ and we have just shown that the optimal hypothesis has a value of $\tilde{L}_{k_{\max}}$ of at most $\tilde{L}_{k_{\max}} - 1$). Therefore, by induction (recall that the initial hypothesis also satisfies this condition), the previous mechanism to generate new hypotheses can be repeatedly applied.

It is simple to see that the accepted hypothesis must be the optimal one since removing more active eigenmodes would further increase the objective to minimize (removing active eigenmodes implies decreasing water-levels, increasing the t_k's, and increasing the terms $\frac{t_k}{L_k - t_k}$) and adding more active eigenmodes has already been checked and rejected.

By the nature of the algorithm, the maximum number of iterations (worst-case complexity) is $\sum_k \check{L}_k$. ■

Chapter 6

Joint Design of Tx-Rx Linear Processing for MIMO Channels with QoS Constraints

THIS CHAPTER CONSIDERS COMMUNICATIONS THROUGH MIMO channels with a set of Quality of Service (QoS) requirements for the simultaneously established substreams in terms of MSE, SINR, or BER. Linear transmit-receive processing (also termed linear precoder at the transmitter and linear equalizer at the receiver) is designed to satisfy the QoS constraints with minimum transmitted power (the exact conditions under which the problem becomes unfeasible are given). Although the original problem is a complicated nonconvex problem with matrix-valued variables, with the aid of majorization theory, we reformulate it as a simple convex optimization problem with scalar variables. We then propose a practical and efficient multi-level water-filling algorithm to optimally solve the problem for the general case of different QoS requirements. The optimal transmit-receive processing is shown to diagonalize the channel matrix only after a very specific pre-rotation of the data symbols. For situations in which the resulting transmit power is too large, we give the precise way to relax the QoS constraints in order to reduce the required power based on a perturbation analysis. Numerical results from simulations are also given to support the mathematical development of the problem.

6.1 Introduction

Communications over multiple-input multiple-output (MIMO) channels have recently gained considerable attention [Hon90, Yan94b, Fos96, Ral98, Sca99b]. They arise in many different scenarios such as when a bundle of twisted pair copper wires in digital subscriber lines (DSL) is treated as a whole [Hon90], when multiple antennas are used at both sides of a wireless link [Fos96, Ral98], or simply when a time-dispersive or frequency-selective channel is properly modeled for block transmission by using, for example, transmit and receive filterbanks [Sca99b] (c.f. §2.2). In particular, MIMO channels arising from the use of multiple antennas at both the transmitter and the receiver have recently attracted a significant interest because they provide an important increase in capacity over single-input single-output (SISO) channels under some uncorrelation conditions [Tel95, Fos98].

The transmitter and the receiver may or may not have channel state information (CSI) (c.f. §2.4). For slowly varying channels, it is generally assumed perfect CSI at the receiver (CSIR). Regarding CSI at the transmitter (CSIT), a significant part of the publications deal with the case of no CSIT such as the popular space-time coding techniques [Fos96, Ala98, Tar98], whereas another great block of research has been devoted to the situation with perfect CSIT [Yan94b, Ral98, Sca99b, Sam01]. We focus on the latter and, in particular, when linear processing is utilized for the sake of complexity.

In many situations, the transmitter is constrained on its average transmit power to limit the interference level of the system and the transmit-receive processing is designed to maximize the quality of the communication according to some criterion as was considered in Chapter 5.

In other situations, however, the approach of maximizing the quality subject to a transmit power constraint may not be the desired objective as argued next. From a system level point of view, it may be interesting to consider the opposite formulation of the problem. Given that several substreams are to be established through the MIMO channel and that each substream requires a (possibly different) Quality of Service (QoS), the communication system wishes to satisfy these QoS constraints with minimum transmitted power. This type of design has been considered in the literature mainly for multiuser scenarios, in which multiple distributed users coexist and joint multiuser processing cannot be assumed at one side of the link since the users are geographically distributed. In [RF98a], a multiuser scenario with a multi-element base station was considered and optimal beamvectors and power allocation were obtained for the uplink and downlink under QoS constraints in terms of SINR. In [Ben99, Ben01], the same problem was solved under the powerful framework of convex optimization theory which allows the introduction of additional constraints (for example to control the dynamic range of the transmitted signal or to increase the robustness against channel estimation errors). In [Cha02b], the problem was generalized to the case of having multiple antennas at both sides of the link, although a global optimal solution was not found due to the nonconvexity of the problem (a suboptimal iterative optimization approach was taken). A similar signal model arising from a single-antenna multiuser multicarrier CDMA system was treated in [Lok00] (with QoS constraints in term of SINR as well), in which a suboptimal (due to the nonconvexity of the problem) gradient-type algorithm was used. In [Vis99c], a single-antenna multiuser CDMA system was characterized in terms of user capacity by optimally designing the CDMA codes of the users with SINR requirements.

The approach in this chapter is similar to the aforementioned examples in that we deal with the optimization of a system subject to QoS requirements and different in that it is assumed that joint processing is possible at both the transmitter and the receiver. Of course, the previously considered multiuser scenario in which the users are geographically distributed is not valid any-more[1] (note that the considered model does not correspond to a multiple-access channel since both sides of the link are allowed to cooperate). Hence, the considered scenario is just a point-to-point communication system[2] where more than one substreams are simultaneously established with (possibly different) QoS requirements.[3] In fact, this situation happens naturally in spectrally efficient systems which are designed to approach the capacity of the MIMO channel as we now describe.

The capacity-achieving solution dictates that the channel matrix has to be diagonalized and that the power at the transmitter has to be allocated following a water-filling distribution on the channel eigenmodes [Cov91b, Ral98, Sca99a]. In theory, this solution has the implication that an ideal Gaussian code should be used on each channel eigenmode according to its allocated power [Cov91b]. In practice, however, Gaussian codes are substituted with simple (and suboptimal) signal constellations and practical (and suboptimal) coding schemes (if any). Therefore, the uncoded part of a practical system basically transmits a set of (possibly different) constellations simultaneously through the MIMO channel. In light of these observations, an interesting way to design the uncoded part of a communication system is based on the gap approximation [Sta99, p. 206], which gives the optimal bit distribution (following a water-filling allocation similar to that of the capacity-achieving solution) under the assumption that practical constellations such as QAM of different sizes are used. Of course, in order to reduce the complexity of a system employing different constellations and codes, it can be constrained to use the same constellation and code in all channel eigenmodes (possibly optimizing the utilized bandwidth to transmit only over those eigenmodes with a sufficiently high gain), *i.e.*, an equal-rate transmission. Examples of this pragmatic and simple solution are found in the European standard HIPERLAN/2 [ETS01] and in the US standard IEEE 802.11 [IEE99] for wireless local area networks (WLAN). In any case, once the constellations to be used at each of the substreams are known, the system can be further optimized such that each established substream satisfies, for example, a given BER.

Hence, this chapter considers the transmission of a vector of data symbols through a channel

[1] Some very specific multiuser scenarios allow for cooperation at both sides of the link such as in DSL systems where both ends of the MIMO system are each terminated in a single physical location, *e.g.*, links between central offices and remote terminals (and also private networks) [Hon90] (see also [Pal02b]).

[2] For the more general case of a multiuser scenario, each user with multiple transmit and receive dimensions, the results of this chapter are not optimal, although they can still be applied iteratively for each of the users, obtaining good solutions in practice [Ben02].

[3] A very simple example of a single-user communication with several established substreams, each with a different QoS requirement, arises when the user wants to transmit simultaneously different services such as audio and video (since video typically requires a higher SINR than audio).

matrix subject to (possibly different) QoS constraints given in terms of MSE, SINR, or BER. The coding and modulation schemes used on the different substreams are assumed given and are not involved in the optimization process (therefore, different services can employ different signal constellations and different error control coding schemes yielding a general multirate communication system). Linear transmit-receive processing is designed to satisfy the QoS constraints with minimum transmitted power (the exact conditions under which the problem becomes unfeasible are given). The original formulation of the problem is a complicated nonconvex optimization problem with matrix-valued variables. With the aid of majorization theory, however, the problem can be reformulated as a simple convex problem with scalar variables. We propose then a practical and efficient multi-level water-filling algorithm that obtains an optimal solution for the general case of different QoS requirements among the established substreams. The optimal solution is shown to diagonalize the channel matrix only after a very specific pre-rotation of the data symbols. In some situations, when the transmit power required to satisfy the QoS constraints results too large, it may be desirable to relax some QoS requirements. By using a sensitivity analysis of the perturbed system, we obtain the precise way to relax the QoS constraints in order to reduce the power needed.

This chapter is structured as follows. Section 6.3 considers the case of single beamforming, which refers to the transmission of a single symbol through the MIMO channel by using transmit and receive beamvectors. Section 6.4 extends the results to the more general case of multiple beamforming, which refers to the simultaneous transmission of L symbols through the MIMO channel by using transmit and receive multiple beamvectors or, equivalently, transmit and receive matrices (or beam-matrices). In this section, the main result of the paper is obtained, *i.e.*, the optimal transmit-receive linear processing for a communication with equal and different MSE-based QoS requirements (as well as a simple and suboptimal transmit-receive scheme of interest that will serve as a benchmark for comparison). In Section 6.5, the relaxation of the QoS requirements is considered with application in practical systems. Numerical results obtained from simulations are given in Section 6.6. Finally, in Section 6.7, a summary of the chapter is given along with the final conclusions.

The results in this chapter have been published in [Pal02b, Pal03f].

6.2 QoS Requirements

The problem formulation of Chapter 5 was based on the optimization of some global measure of the system quality subject to a power constraint. The opposite formulation of the problem is the minimization of the transmitted power subject to a constraint on the system quality. The analysis of such a formulation is straightforward using the results of Chapter 5. In particular, the main result stated in Theorem 5.1 can be directly invoked and the practical algorithms obtained in §5.5 for different measures of quality can be easily modified (in fact, the algorithms of §5.5 can be directly used by iteratively fixing some transmit power, optimizing the quality with the corresponding algorithm of §5.5, and then adjusting the power so that the achieved quality approaches the desired value on an iterative fashion). In this chapter, however, we consider a different problem in that we fix the quality of each of the substreams independently rather than a single global measure of quality of the whole system. It is important to point out that the particular case of constraining the BER averaged over a set of substreams using the same constellation is equivalent to constraining each substream independently to the same BER (c.f. §5.5.11 and [Pal03c]).

To design the system, we consider that each of the established substreams has a (possibly different) QoS requirement expressed in terms of either the MSE's

$$\mathrm{MSE}_{k,i} \triangleq \mathbb{E}[\,|\hat{x}_{k,i} - x_{k,i}|^2] \leq \rho_{k,i} \qquad 1 \leq i \leq L,\ 1 \leq k \leq N, \tag{6.1}$$

the SINR's

$$\mathrm{SINR}_{k,i} \triangleq \frac{\left|\mathbf{a}_{k,i}^H \mathbf{H}_k \mathbf{b}_{k,i}\right|^2}{\mathbf{a}_{k,i}^H \mathbf{R}_{n_{k,i}} \mathbf{a}_{k,i}} \geq \gamma_{k,i} \qquad 1 \leq i \leq L,\ 1 \leq k \leq N, \tag{6.2}$$

or the BER's

$$\mathrm{BER}_{k,i} \triangleq \mathrm{BER}\left(\mathrm{SINR}_{k,i}\right) \leq p_{k,i} \qquad 1 \leq i \leq L,\ 1 \leq k \leq N, \tag{6.3}$$

where BER (SINR) is the BER function assuming a Gaussian-distributed interference-plus-noise component (c.f. §2.5.4.4). As in Chapter 5, we only consider the uncoded part of the communication system and, therefore, the BER always refers to the uncoded BER (recall that, in practice, an outer code should always be used on top of the uncoded part).

The optimum signal processing at the receiver was obtained in §2.5.5 as the classical LMMSE filter or Wiener filter given by $\mathbf{a}_k = \left(\mathbf{H}_k \mathbf{b}_k \mathbf{b}_k^H \mathbf{H}_k^H + \mathbf{R}_{n_k}\right)^{-1} \mathbf{H}_k \mathbf{b}_k$ in the single beamforming case and by $\mathbf{A}_k = \left(\mathbf{H}_k \mathbf{B}_k \mathbf{B}_k^H \mathbf{H}_k^H + \mathbf{R}_{n_k}\right)^{-1} \mathbf{H}_k \mathbf{B}_k$ in the multiple beamforming case (see (2.43) and (2.48)). As shown in §2.5.5, the Wiener filter is optimal in the sense that each of the MSE's is minimized, each of the SINR's is maximized, and each of the BER's is minimized. Therefore, for any given feasible transmit matrix $\mathbf{B}_k$ (a $\mathbf{B}_k$ such that there exists some $\mathbf{A}_k$ with which the QoS constraints can be satisfied), the Wiener filter will always give a feasible solution (clearly, if the Wiener filter does not satisfy some of the QoS constraints, no other $\mathbf{A}_k$ will).[4]

The rest of the chapter is devoted to obtaining the optimal transmitter in terms of minimum power that satisfies the QoS requirements (the exact conditions under which the problem becomes unfeasible are derived). No matter whether the QoS requirements of the system are specified in terms of MSE, SINR, or BER, the problem can always be formulated in terms of MSE constraints

[4]To be exact, if some QoS constraints are satisfied with strict inequality for a given transmit matrix $\mathbf{B}_k$ and the corresponding Wiener filter at the receiver, there must exist some other feasible receive matrix $\mathbf{A}_k$. However, we stick to the Wiener filter since it guarantees that for any feasible $\mathbf{B}_k$ it will always yield a feasible solution.

as we consider in the rest of the paper without loss of generality (this is straightforward using the relation between the SINR and the MSE in (2.62) and the relation between the BER and the SINR in (2.38)-(2.40)).

For the sake of notation, we define the squared whitened channel matrix $\mathbf{R}_H \triangleq \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H}$ (note that the eigenvectors and eigenvalues of $\mathbf{R}_H$ are the right singular vectors and the squared singular values, respectively, of the whitened channel $\mathbf{R}_n^{-1/2}\mathbf{H}$). For multiple MIMO channels, we similarly define $\mathbf{R}_{H_k} \triangleq \mathbf{H}_k^H \mathbf{R}_{n_k}^{-1} \mathbf{H}_k$.

The results in this section were obtained in [Pal02b, Pal03f].

6.3 Single Beamforming

In this section, the simple case of single beamforming for MIMO channels as formulated in §2.5.1.1 is considered. First, the simple case of a single MIMO channel is analyzed in §6.3.1 and then the result is extended in §6.3.2 to the case of multiple MIMO channels (typical of multicarrier systems).

6.3.1 Single MIMO Channel

Single beamforming on a single MIMO channel is a trivial case and has a simple solution. It will serve as a reference when dealing with the more general case of having a set of parallel MIMO channels.

Consider the single MIMO channel model of (2.1) and the single beamforming approach of (2.25)-(2.26) given by $\hat{x} = \mathbf{a}^H (\mathbf{H}\mathbf{b}x + \mathbf{n})$. As obtained in §2.5.5, the optimal receive beamvector is the Wiener filter $\mathbf{a} = \left(\mathbf{H}\mathbf{b}\mathbf{b}^H\mathbf{H}^H + \mathbf{R}_n\right)^{-1} \mathbf{H}\mathbf{b}$ and the resulting MSE is $\text{MSE} = \frac{1}{1+\mathbf{b}^H \mathbf{R}_H \mathbf{b}}$. The problem then reduces to obtain the optimal transmit beamvector $\mathbf{b}$ to minimize the transmitted power subject to a MSE QoS constraint:

$$\begin{array}{ll} \min\limits_{\mathbf{b}} & \mathbf{b}^H \mathbf{b} \\ \text{s.t.} & \frac{1}{1+\mathbf{b}^H \mathbf{R}_H \mathbf{b}} \leq \rho. \end{array} \tag{6.4}$$

The optimal solution to this (nonconvex) optimization problem (similarly to that obtained in §5.3.1) is trivially given by

$$\mathbf{b} = \sqrt{\lambda_{H,\max}^{-1} \left(\rho^{-1} - 1\right)}\, \mathbf{u}_{H,\max} \tag{6.5}$$

where $\lambda_{H,\max} \triangleq \lambda_{\max}(\mathbf{R}_H)$ and $\mathbf{u}_{H,\max} \triangleq \mathbf{u}_{\max}(\mathbf{R}_H)$. The optimal $\mathbf{b}$ satisfies the MSE constraint with equality and has the direction of the eigenvector associated to the maximum eigenvalue of matrix $\mathbf{R}_H$ (with arbitrary phase). The minimum required power at the transmitter is

$$P_T = \lambda_{H,\max}^{-1} \left(\rho^{-1} - 1\right). \tag{6.6}$$

6.3.2 Multiple MIMO Channels

The single beamforming approach for multiple MIMO channels, *i.e.*, for a set of N parallel and independent MIMO channels, is exactly the same as that previously obtained for the single MIMO channel.

Consider the multiple MIMO channel model of (2.3) and the single beamforming approach of (2.27)-(2.28) given by $\hat{x}_k = \mathbf{a}_k^H (\mathbf{H}_k \mathbf{b}_k x_k + \mathbf{n}_k)$. As obtained in §2.5.5, the optimal receive beamvectors are the Wiener filters $\mathbf{a}_k = \left(\mathbf{H}_k \mathbf{b}_k \mathbf{b}_k^H \mathbf{H}_k^H + \mathbf{R}_{n_k}\right)^{-1} \mathbf{H}_k \mathbf{b}_k$ and the resulting MSE is $\text{MSE}_k = \frac{1}{1+\mathbf{b}_k^H \mathbf{R}_{H_k} \mathbf{b}_k}$. The problem then reduces to obtain the optimal transmit beamvectors $\mathbf{b}_k$'s to minimize the transmitted power subject to MSE QoS constraints:

$$\begin{array}{ll} \min\limits_{\{\mathbf{b}_k\}} & \sum_{k=1}^N \mathbf{b}_k^H \mathbf{b}_k \\ \text{s.t.} & \frac{1}{1+\mathbf{b}_k^H \mathbf{R}_{H_k} \mathbf{b}_k} \leq \rho_k \qquad 1 \leq k \leq N. \end{array} \tag{6.7}$$

In this case, the optimal solution is again obtained as

$$\mathbf{b}_k = \sqrt{\lambda_{H_k,\max}^{-1} \left(\rho_k^{-1} - 1\right)}\, \mathbf{u}_{H_k,\max} \tag{6.8}$$

where $\lambda_{H_k,\max} \triangleq \lambda_{\max}(\mathbf{R}_{H_k})$ and $\mathbf{u}_{H_k,\max} \triangleq \mathbf{u}_{\max}(\mathbf{R}_{H_k})$. As before, each optimal $\mathbf{b}_k$ has the direction of the eigenvector associated to the maximum eigenvalue of matrix $\mathbf{R}_{H_k}$ with norm such the MSE constraint is satisfied with equality (with arbitrary phase). The minimum required power at the transmitter is

$$P_T = \sum_{k=1}^N \lambda_{H_k,\max}^{-1} \left(\rho_k^{-1} - 1\right). \tag{6.9}$$

6.4 Multiple Beamforming

This section extends the results of §6.3 to the more general case of multiple beamforming or matrix beamforming for MIMO channels as formulated in §2.5.1.2.

The original formulation of the problem is a complicated nonconvex optimization problem with matrix-valued variables. With the aid of majorization theory (see §3.2), however, the problem can be reformulated as a simple convex problem with scalar variables.

First, in §6.4.1, we deal with the case of a single MIMO channel, for which the main result of this chapter is obtained. In particular, the optimum solution in terms of minimum transmitted power is obtained to satisfy the given set of QoS constraints. In §6.4.2, we then extend the results to the case of multiple MIMO channels (typical of multicarrier systems). The results in this section were obtained in [Pal02b, Pal03f].

6.4.1 Single MIMO Channel

Consider the single MIMO channel model of (2.1) and the matrix processing model of (2.29)-(2.30) given by $\hat{\mathbf{x}} = \mathbf{A}^H (\mathbf{HBx} + \mathbf{n})$. As obtained in §2.5.5, the optimal receive matrix is the Wiener filter $\mathbf{A} = \left(\mathbf{HBB}^H\mathbf{H}^H + \mathbf{R}_n\right)^{-1}\mathbf{HB}$ and the resulting MSE matrix is $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}$. The problem reduces then to obtain the optimal transmit matrix $\mathbf{B}$ to minimize the transmitted power subject to a set of MSE QoS constraints:

$$\begin{array}{ll} \min\limits_{\mathbf{B}} & \mathrm{Tr}\left(\mathbf{BB}^H\right) \\ \text{s.t.} & \left[\left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right]_{ii} \leq \rho_i \qquad 1 \leq i \leq L. \end{array} \tag{6.10}$$

Such a constrained optimization problem is nonconvex and requires previous manipulations (even in the simple scalar and real case, $\frac{1}{1+b^2r} \leq \rho$ is a nonconvex region in b). Although (6.10) seems at first a formidable problem to solve, with the aid of majorization theory (see §3.2) it can be reformulated as a simple convex problem with scalar variables that can be optimally solved.

We first consider the simpler case with equal MSE QoS constraints in §6.4.1.1 and then extend the solution to the more general (and complicated) case with different MSE QoS constraints in §6.4.1.2 (in both cases, efficient algorithms are given for practical implementations). We then consider a simple and suboptimum approach in §6.4.1.3 based on imposing a diagonality constraint on the MSE matrix (such a constraint implies that the channel matrix is diagonalized and then each symbol is transmitted through a different channel eigenmode).

6.4.1.1 Optimum Solution for Equal MSE QoS Constraints

The optimal solution for equal MSE QoS constraints $\rho_i = \rho \; \forall i$ (along with the feasibility condition) is formally stated in the following theorem.

Theorem 6.1 *The following nonconvex optimization problem subject to equal MSE QoS constraints:*

$$\begin{array}{ll} \min\limits_{\mathbf{B}} & \mathrm{Tr}\left(\mathbf{BB}^H\right) \\ s.t. & \left[\left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right]_{ii} \leq \rho \qquad 1 \leq i \leq L, \end{array} \tag{6.11}$$

can be optimally solved by first solving the simple convex optimization problem:

$$\begin{array}{lll} \min\limits_{\{z_i\}} & \sum_{i=1}^{\check{L}} z_i & \\ s.t. & \frac{1}{L}\left(L_0 + \sum_{i=1}^{\check{L}} \frac{1}{1+z_i\lambda_{H,i}}\right) \leq \rho, & \\ & z_i \geq 0, & 1 \leq i \leq \check{L} \end{array} \tag{6.12}$$

where L is the number of established links, $\check{L} \triangleq \min(L, \text{rank}(\mathbf{R}_H))$ is the number of effective channel eigenvalues used, $L_0 \triangleq L - \check{L}$ is the number of links associated to zero eigenvalues, and the set $\{\lambda_{H,i}\}_{i=1}^{\check{L}}$ contains the $\check{L}$ largest eigenvalues of $\mathbf{R}_H$ in increasing order.

The optimal solution to (6.11) satisfies all QoS constraints with equality and is given by $\mathbf{B} = \mathbf{U}_{H,1}\mathbf{\Sigma}_{B,1}\mathbf{Q}$ where $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times \check{L}}$ has as columns the eigenvectors of $\mathbf{R}_H$ corresponding to the $\check{L}$ largest eigenvalues in increasing order, $\mathbf{\Sigma}_{B,1} = [\mathbf{0} \ \text{diag}(\{\sigma_{B,i}\})] \in \mathbb{C}^{\check{L} \times L}$ has zero elements except along the rightmost main diagonal (assumed real w.l.o.g.) which are given by $\sigma_{B,i}^2 = z_i$, $1 \leq i \leq \check{L}$ (the z_i's are the solution to the convex problem (6.12)), and $\mathbf{Q}$ is a unitary matrix such that the diagonal elements of $\left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}$ are equal. This rotation matrix can be computed using Algorithm 3.2 (reproduced from [Vis99b, Section IV-A]) or with any rotation matrix $\mathbf{Q}$ that satisfies $|[\mathbf{Q}]_{ik}| = |[\mathbf{Q}]_{il}|$ $\forall i,k,l$ such as the Discrete Fourier Transform (DFT) matrix or the Hadamard matrix when the dimensions are appropriate (see §3.2 for more details). The problem is feasible if and only if $\rho > L_0/L$.

Proof. See Appendix 6.A. ■

Now that the original nonconvex problem has been reformulated as a simple convex problem, we know that a global optimal solution can be obtained in practice by using, for example, interior-point methods (see §3.1). Nevertheless, the particular convex problem (6.12) obtained in Theorem 6.1 can be optimally solved by a simple water-filling algorithm (Algorithm 6.1) as stated in the following proposition (by setting $\tilde{\rho} = \rho L - L_0$).

Proposition 6.1 *The optimal solution to the following convex optimization problem:*

$$\begin{array}{lll} \min\limits_{\{z_i\}} & \sum_{i=1}^{\check{L}} z_i & \\ s.t. & \sum_{i=1}^{\check{L}} \frac{1}{1+z_i\lambda_i} \leq \tilde{\rho}, & \\ & z_i \geq 0, & 1 \leq i \leq \check{L}, \end{array}$$

is given (if feasible) by the water-filling solution $z_i = \left(\mu^{1/2}\, \lambda_i^{-1/2} - \lambda_i^{-1}\right)^+$ (it is tacitly assumed that all the λ_i's are strictly positive) where $\mu^{1/2}$ is the water-level chosen such that the MSE constraint is satisfied with equality: $\sum_{i=1}^{\check{L}} \frac{1}{1+z_i\lambda_i} = \tilde{\rho}$.

Furthermore, the optimal water-filling solution can be efficiently obtained in practice with Algorithm 6.1 in no more than $\check{L}$ iterations (worst-case complexity). (It is assumed that $\tilde{\rho} < \check{L}$, otherwise the optimal solution is trivially given by $z_i = 0$ $\forall i$, i.e., by not transmitting anything.)

Proof. See Appendix 6.B. ■

Algorithm 6.1 *Practical water-filling algorithm to solve the convex problem corresponding to the design with equal MSE QoS requirements of Theorem 6.1.*

Input: *Number of available positive eigenvalues* $\check{L}$, *set of eigenvalues* $\{\lambda_i\}_{i=1}^{\check{L}}$, *and MSE constraint* $\tilde{\rho}$.

Output: *Set of allocated powers* $\{z_i\}_{i=1}^{\check{L}}$ *and water-level* $\mu^{1/2}$.

0. *Reorder the* λ_i*'s in decreasing order (define* $\lambda_{\check{L}+1} \triangleq 0$*). Set* $\tilde{L} = \check{L}$.

1. *Set* $\mu = \lambda_{\tilde{L}}^{-1}$ *(if* $\lambda_{\tilde{L}} = \lambda_{\tilde{L}+1}$, *then set* $\tilde{L} = \tilde{L} - 1$ *and go to step 1).*

2. *If* $\mu^{1/2} \geq \frac{\sum_{i=1}^{\tilde{L}} \lambda_i^{-1/2}}{\tilde{\rho} - (\check{L} - \tilde{L})}$, *then set* $\tilde{L} = \tilde{L} - 1$ *and go to step 1.*
Otherwise, obtain the definitive water-level $\mu^{1/2}$ *and allocated powers as*

$$\mu^{1/2} = \frac{\sum_{i=1}^{\tilde{L}} \lambda_i^{-1/2}}{\tilde{\rho} - \left(\check{L} - \tilde{L}\right)} \quad \textit{and}$$
$$z_i = \left(\mu^{1/2}\, \lambda_i^{-1/2} - \lambda_i^{-1}\right)^+ ,$$

undo the reordering done at step 0, and finish.

Thus, the problem of finding a transmit matrix $\mathbf{B}$ that minimizes the required transmit power subject to equal MSE QoS requirements has been completely solved in a practical and optimal way. It is remarkable that, in general, the optimal solution does not consist on transmitting each symbol through a channel eigenmode in a parallel fashion (diagonal transmission); instead, the symbols are transmitted in a distributed way over all the channel eigenmodes.

6.4.1.2 Optimum Solution for Different MSE QoS Constraints

This section generalizes the results obtained in §6.4.1.1 by allowing different QoS constraints in terms of MSE. This problem is more general and far more complicated than the one with equal MSE constraints. Nevertheless, using majorization theory (see §3.2) we can still reformulate the original complicated nonconvex problem as a simple convex optimization problem that can be optimally solved as is formally stated in the following theorem (the feasibility condition of the problem is also given).

Theorem 6.2 *The following nonconvex optimization problem subject to different MSE QoS constraints (assumed in decreasing order* $\rho_i \geq \rho_{i+1}$ *w.l.o.g.):*

$$\begin{array}{ll} \min\limits_{\mathbf{B}} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \\ s.t. & \left[\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}\right]_{ii} \leq \rho_i \qquad 1 \leq i \leq L, \end{array} \tag{6.13}$$

can be optimally solved by first solving the simple convex optimization problem:

$$
\begin{array}{lll}
\min\limits_{\{z_i\}} & \sum_{i=1}^{\check{L}} z_i & \\
s.t. & \sum_{i=k}^{\check{L}} \frac{1}{1+z_i\lambda_{H,i}} \leq \sum_{i=k+L_0}^{L} \rho_i & 1 \leq k \leq \check{L}, \\
 & \sum_{i=1}^{\check{L}} \frac{1}{1+z_i\lambda_{H,i}} \leq \sum_{i=1}^{L} \rho_i - L_0, & \\
 & z_k \geq 0, & 1 \leq k \leq \check{L}
\end{array}
\tag{6.14}
$$

where L is the number of established links, $\check{L} \triangleq \min(L, \operatorname{rank}(\mathbf{R}_H))$ is the number of effective channel eigenvalues used, $L_0 \triangleq L - \check{L}$ is the number of links associated to zero eigenvalues, and the set $\{\lambda_{H,i}\}_{i=1}^{\check{L}}$ contains the $\check{L}$ largest eigenvalues of $\mathbf{R}_H$ in increasing order.

The optimal solution to (6.13) satisfies all QoS constraints with equality and is given by $\mathbf{B} = \mathbf{U}_{H,1}\boldsymbol{\Sigma}_{B,1}\mathbf{Q}$ where $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times \check{L}}$ has as columns the eigenvectors of $\mathbf{R}_H$ corresponding to the $\check{L}$ largest eigenvalues in increasing order, $\boldsymbol{\Sigma}_{B,1} = [\mathbf{0}\ \operatorname{diag}(\{\sigma_{B,i}\})] \in \mathbb{C}^{\check{L} \times L}$ has zero elements except along the rightmost main diagonal (assumed real w.l.o.g.) which are given by $\sigma_{B,i}^2 = z_i$, $1 \leq i \leq \check{L}$ (the z_i's are the solution to the convex problem (6.14)), and $\mathbf{Q}$ is a unitary matrix such that $\left[\left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right]_{ii} = \rho_i$, $1 \leq k \leq L$ which can be computed using Algorithm 3.2 (reproduced from [Vis99b, Section IV-A]). The problem is feasible if and only if $\sum_{i=1}^{L} \rho_i > L_0$ (a simple sufficient condition for the feasibility of the problem is the rule of thumb: $L \leq \operatorname{rank}(\mathbf{H})$ [5]).

Proof. See Appendix 6.C. ■

Noting that $\sum_{i=1}^{L} \rho_i - L_0 < \sum_{i=1+L_0}^{L} \rho_i$ (recall that $\rho_i < 1$), problem (6.14) can be rewritten more compactly as

$$
\begin{array}{lll}
\min\limits_{\{z_i\}} & \sum_{i=1}^{\check{L}} z_i & \\
\text{s.t.} & \sum_{i=k}^{\check{L}} \frac{1}{1+z_i\lambda_{H,i}} \leq \sum_{i=k}^{\check{L}} \tilde{\rho}_i, & 1 \leq k \leq \check{L}, \\
 & z_k \geq 0, &
\end{array}
\tag{6.15}
$$

where $\tilde{\rho}_i \triangleq \begin{cases} \sum_{k=1}^{L_0+1} \rho_k - L_0 & \text{for } i = 1 \\ \rho_{i+L_0} & \text{for } 1 < i \leq \check{L} \end{cases}$ (note that the resulting $\tilde{\rho}_i$'s need not be in decreasing ordering as the ρ_i's).

By Theorem 6.2, the original formidable problem has been reformulated as a convex problem which can always be solved in practice using, for example, interior-point methods (see §3.1). However, as happened with the case of equal MSE QoS requirements in §6.4.1.1, this problem can be solved with a multi-level water-filling algorithm (Algorithms 6.2 and 6.3) as shown next.

[5] In a practical situation, a threshold should be considered in the rank determination to avoid extremelly small eigenvalues to artificially increase the rank.

Proposition 6.2 *The optimal solution to the following convex optimization problem:*

$$\begin{array}{lll} \min\limits_{\{z_i\}} & \sum_{i=1}^{\check{L}} z_i & \\ s.t. & \sum_{i=k}^{\check{L}} \frac{1}{1+z_i\lambda_i} \leq \sum_{i=k}^{\check{L}} \tilde{\rho}_i, & 1 \leq k \leq \check{L}, \\ & z_k \geq 0, & \end{array}$$

is given (if feasible) by the multi-level water-filling solution $z_i = \left(\tilde{\mu}_i^{1/2}\,\lambda_i^{-1/2} - \lambda_i^{-1}\right)^+$ *(it is tacitly assumed that all the* λ_i*'s are strictly positive) where the multiple water-levels* $\tilde{\mu}_i^{1/2}$*'s are chosen to satisfy:*

$$\begin{array}{ll} \sum_{i=k}^{\check{L}} \frac{1}{1+z_i\lambda_i} \leq \sum_{i=k}^{\check{L}} \tilde{\rho}_i & 1 < k \leq \check{L} \\ \sum_{i=1}^{\check{L}} \frac{1}{1+z_i\lambda_i} = \sum_{i=1}^{\check{L}} \tilde{\rho}_i & \\ \tilde{\mu}_k \geq \tilde{\mu}_{k-1} \quad (\tilde{\mu}_0 \triangleq 0) & \\ (\tilde{\mu}_k - \tilde{\mu}_{k-1}) \left(\sum_{i=k}^{\check{L}} \frac{1}{1+z_i\lambda_i} - \sum_{i=k}^{\check{L}} \tilde{\rho}_i\right) = 0. & \end{array}$$

Furthermore, the optimal multi-level water-filling solution can be efficiently obtained in practice with Algorithm 6.2 (or with the equivalent Algorithm 6.3) in no more (worst-case complexity) than $\check{L}\left(\check{L}+1\right)/2$ *inner iterations (simple water-fillings) or, more exactly,* $\check{L}^2\left(\check{L}+1\right)/6$ *basic iterations (iterations within each simple water-filling). (It is assumed that* $\tilde{\rho}_i < 1$*, otherwise the optimal solution is trivially given by* $z_i = 0\ \forall i$*, i.e., by not transmitting anything.)*

Proof. See Appendix 6.D. ■

Algorithm 6.2 *Practical multi-level water-filling algorithm to solve the convex problem corresponding to the design with different MSE QoS requirements of Theorem 6.2. Version 1. (See Figure 6.1 for an illustrative example of the application of the algorithm.)*

Input: *Number of available positive eigenvalues* $\check{L}$*, set of eigenvalues* $\{\lambda_i\}_{i=1}^{\check{L}}$*, and set of MSE constraints* $\{\tilde{\rho}_i\}_{i=1}^{\check{L}}$ *(note that the appropriate ordering of the* λ_i*'s and of the* $\tilde{\rho}_i$*'s is independent of this algorithm).*

Output: *Set of allocated powers* $\{z_i\}_{i=1}^{\check{L}}$ *and set of water-levels* $\{\tilde{\mu}_i^{1/2}\}_{i=1}^{\check{L}}$*.*

0. *Set* $\tilde{L} = \check{L}$*.*

1. *Perform an outer iteration.*

2. *If* $k_0 = 1$*, then finish. Otherwise set* $\tilde{L} = k_0 - 1$ *and go to step 1.*

Outer iteration:

0. *Set* $k_0 = 1$*.*

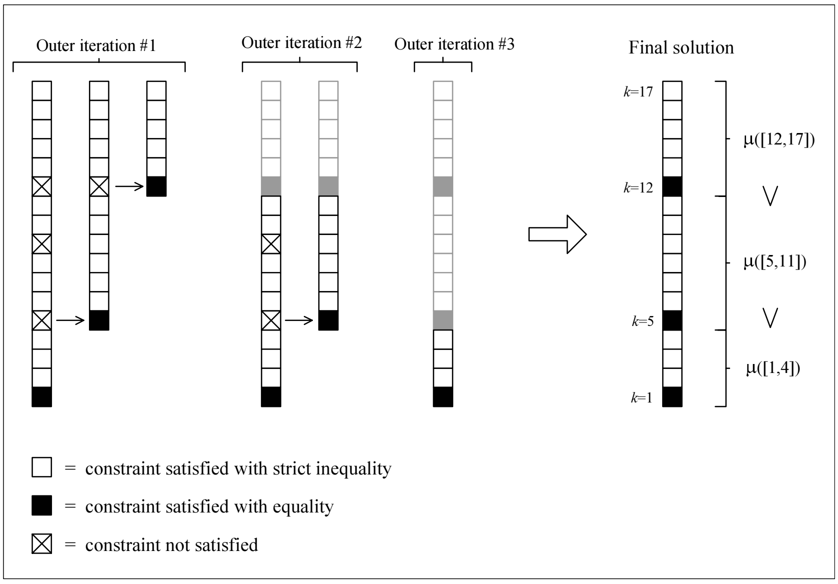


Figure 6.1: Example of the execution of Algorithm 6.2 to solve the problem with different QoS requirements.

1. *Perform an inner iteration: solve the equal MSE QoS constrained problem in* $\left[k_0, \tilde{L}\right]$ *using Algorithm 6.1 with the set of* $\tilde{L} - k_0 + 1$ *eigenvalues* $\{\lambda_i\}_{i=k_0}^{\tilde{L}}$ *and with the MSE constraint given by* $\tilde{\rho} = \sum_{i=k_0}^{\tilde{L}} \tilde{\rho}_i$.
2. *If all intermediate constraints are also satisfied (i.e., if* $\sum_{i=k}^{\tilde{L}} \frac{1}{1+z_i\lambda_i} \leq \sum_{i=k}^{\tilde{L}} \tilde{\rho}_i$, $k_0 < k \leq \tilde{L}$*), then finish.*
   Otherwise, set k_0 *equal to the smallest index whose constraint is not satisfied and go to step 1.*

Note that each outer iteration implicitly computes the water-level for the set $\left[k_0, \tilde{L}\right]$ denoted by $\mu^{1/2}\left(\left[k_0, \tilde{L}\right]\right)$; in other words, $\tilde{\mu}_k = \mu\left(\left[k_0, \tilde{L}\right]\right),\ k_0 \leq k \leq \tilde{L}$.

Algorithm 6.2 was conveniently written to prove its optimality. For a practical implementation, however, it can be rewritten in a much simpler way as in Algorithm 6.3.

Algorithm 6.3 *Practical multi-level water-filling algorithm to solve the convex problem corresponding to the design with different MSE QoS requirements of Theorem 6.2. Version 2.*

Input: *Number of available positive eigenvalues* $\check{L}$, *set of eigenvalues* $\{\lambda_i\}_{i=1}^{\check{L}}$, *and set of MSE constraints* $\{\tilde{\rho}_i\}_{i=1}^{\check{L}}$ *(note that the appropriate ordering of the* λ_i*'s and of the* $\tilde{\rho}_i$*'s is independent of this algorithm).*

Output: *Set of allocated powers* $\{z_i\}_{i=1}^{\check{L}}$ *and set of water-levels* $\{\tilde{\mu}_i^{1/2}\}_{i=1}^{\check{L}}$.

0. *Set* $k_0 = 1$ *and* $\tilde{L} = \check{L}$.
1. *Solve the equal MSE QoS constrained problem in* $\left[k_0, \tilde{L}\right]$ *using Algorithm 6.1 with the set of* $\tilde{L} - k_0 + 1$ *eigenvalues* $\{\lambda_i\}_{i=k_0}^{\tilde{L}}$ *and with the MSE constraint given by* $\tilde{\rho} = \sum_{i=k_0}^{\tilde{L}} \tilde{\rho}_i$.

2. *If any intermediate constraint* ($\sum_{i=k}^{\tilde{L}} \frac{1}{1+z_i\lambda_i} \leq \sum_{i=k}^{\tilde{L}} \tilde{\rho}_i,\ k_0 < k \leq \tilde{L}$) *is not satisfied, then set* k_0 *equal to the smallest index whose constraint is not satisfied and go to step 1. Otherwise, if* $k_0 = 1$ *finish and if* $k_0 > 1$ *set* $\tilde{L} = k_0 - 1$, $k_0 = 1$, *and go to step 1.*

This section has obtained the main result of the chapter: an efficient and optimal way to solve in practice the problem of finding a transmit matrix $\mathbf{B}$ that minimizes the required transmit power while satisfying different MSE QoS requirements. It suffices to use the multi-level water-filling algorithm (Algorithm 6.3) and then to find the proper rotation matrix $\mathbf{Q}$ as described in Theorem 6.2.

As happened in the case of equal MSE QoS constraints, the optimal solution does not generally consist on transmitting each symbol through a channel eigenmode in a parallel fashion (diagonal transmission); instead, the symbols are transmitted in a distributed way over all the channel eigenmodes.

6.4.1.3 Suboptimum Solution: A Simple Approach Imposing Diagonality

At this point, it is interesting to consider a suboptimum but very simple solution to the considered problem. The simplicity of the solution comes from imposing a diagonality constraint in the MSE matrix, *i.e.*, from forcing $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ to have a diagonal structure. Imposing such a structure implies that the transmission is performed in a parallel fashion through the channel eigenmodes. In Lemma 6.1, we obtain such a simple solution and give the feasibility condition. Then, in Lemma 6.2, we state the exact conditions under which such a constrained solution happens to be the optimum solution to the original problem considered in Theorem 6.2.

Lemma 6.1 *The following nonconvex optimization problem subject to different MSE QoS constraints (assumed in decreasing order* $\rho_i \geq \rho_{i+1}$ *w.l.o.g. and bounded by* $0 < \rho_i < 1$*):*

$$\begin{array}{ll} \min\limits_{\mathbf{B}} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \\ s.t. & \left[\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}\right]_{ii} \leq \rho_i \qquad 1 \leq i \leq L, \\ & \mathbf{B}^H \mathbf{R}_H \mathbf{B} \qquad \textit{diagonal} \end{array}$$

is feasible if and only if the number of established links L *satisfy* $L \leq \mathrm{rank}\left(\mathbf{R}_H\right)$ *and the optimal solution is then given by* $\mathbf{B} = \mathbf{U}_{H,1}\mathbf{\Sigma}_{B,1}$, *where* $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times L}$ *has as columns the eigenvectors of* $\mathbf{R}_H$ *corresponding to the* L *largest eigenvalues in increasing order, denoted by* $\{\lambda_{H,i}\}$, *and* $\mathbf{\Sigma}_{B,1} \in \mathbb{C}^{L \times L}$ *is a diagonal matrix with squared-diagonal elements given by*

$$z_i = \lambda_{H,i}^{-1}\left(\rho_i^{-1} - 1\right) \qquad 1 \leq i \leq L.$$

Proof. See Appendix 6.E. ■

Lemma 6.2 *The optimal solution obtained in Lemma 6.1 under the diagonality constraint of the MSE matrix* $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ *is the optimal solution to the problem considered in Theorem 6.2 without the diagonality constraint on* $\mathbf{E}$ *if and only if*

$$\lambda_{H,i}\, \rho_i^2 \geq \lambda_{H,i+1}\, \rho_{i+1}^2 \qquad 1 \leq i < L \tag{6.16}$$

where the ρ_i*'s are in decreasing order and the* $\lambda_{H,i}$*'s are the* L *largest eigenvalues of* $\mathbf{R}_H$ *in increasing order. Note that this condition implies the feasibility condition* $L \leq \text{rank}\left(\mathbf{R}_H\right)$*.*

The conditions in (6.16) can be stated in words by saying that the singular values of the whitened channel $\lambda_{H,i}^{1/2}$ *have to increase at a slower rate than the decrease of the MSE constraints* ρ_i*.*

Proof. See Appendix 6.F. ■

As an example, the conditions of Lemma 6.2 are always satisfied for channels with equal singular values (this corresponds to a diagonal squared channel matrix $\mathbf{R}_H$) and, consequently, a parallel transmission is always the optimum structure for such channels.

Another interesting example arises for systems with equal MSE constraints as treated in Theorem 6.1, for which the conditions in (6.16) are never satisfied (unless the squared-channel has equal eigenvalues).

6.4.2 Multiple MIMO Channels

This section extends the results of multiple beamforming in a single MIMO channel of §6.4.1 to the case of a set of N parallel MIMO channels. The extension is straightforward as we show next.

Consider the multiple MIMO channel model of (2.3) and the matrix processing model of (2.31)-(2.32) given by $\hat{\mathbf{x}}_k = \mathbf{A}_k^H \left(\mathbf{H}_k \mathbf{B}_k \mathbf{x}_k + \mathbf{n}_k\right)$ where L_k symbols are transmitted through the kth MIMO channel (L_k substreams). As obtained in §2.5.5, the optimal receive matrices are the Wiener filters $\mathbf{A}_k = \left(\mathbf{H}_k \mathbf{B}_k \mathbf{B}_k^H \mathbf{H}_k^H + \mathbf{R}_{n_k}\right)^{-1} \mathbf{H}_k \mathbf{B}_k$ and the resulting MSE matrices are $\mathbf{E}_k = \left(\mathbf{I} + \mathbf{B}_k^H \mathbf{R}_{H_k} \mathbf{B}_k\right)^{-1}$. The problem reduces then to obtain the optimal transmit matrices $\mathbf{B}_k$'s to minimize the transmitted power subject to a set of MSE QoS constraints:

$$\begin{array}{ll} \displaystyle\min_{\{\mathbf{B}_k\}} & \sum_{k=1}^N \text{Tr}\left(\mathbf{B}_k \mathbf{B}_k^H\right) \\ \text{s.t.} & \left[\left(\mathbf{I} + \mathbf{B}_k^H \mathbf{R}_{H_k} \mathbf{B}_k\right)^{-1}\right]_{ii} \leq \rho_{k,i} \qquad 1 \leq i \leq L,\ 1 \leq k \leq N. \end{array} \tag{6.17}$$

This problem is clearly separable in each of the N MIMO channels. Therefore, the results previously obtained in §6.4.1 are directly applicable to each of the MIMO channels.

6.5 Relaxation of the QoS Requirements

In situations where the problem is feasible but the required transmit power exceeds some pre-specified maximum level, the system may be forced to relax some QoS requirements so that the required power is reduced. For that purpose, we identify which QoS constraints produce the largest reduction in transmit power when relaxed by means of a perturbation analysis (see §3.1.4). The questions of whether and when these relaxations are necessary is a high-level decision[6] that may depend on factors as disparate as the energy left on the batteries at the transmitter or the number of services/users requesting a link. Such high-level decisions are out of the scope of this chapter.

It is well known from convex optimization theory (see §3.1) that the optimal dual variables (Lagrange multipliers) of a convex optimization problem give useful information about the sensitivity of the optimal objective value with respect to perturbations of the constraints. Consider the following relaxation of the original QoS constraints of (6.1):

$$\mathrm{MSE}_i \leq \rho_i + u_i(\delta_i) \tag{6.18}$$

where $u_i(\delta_i) : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is a positive differentiable function parameterized with δ_i such that $u_i(\delta_i) \underset{\delta_i \to 0^+}{\longrightarrow} 0$. By Theorem 6.2, the relaxed problem in convex form is (for simplicity of exposition and w.l.o.g., we focus on the multiple beamforming architecture in a single MIMO channel and consider $L \leq \mathrm{rank}(\mathbf{H})$ in (6.14)):

$$\begin{array}{lll} \min\limits_{\{z_i\}} & \sum_{i=1}^{L} z_i & \\ \text{s.t.} & \sum_{i=k}^{L} \frac{1}{1+z_i\lambda_{H,i}} \leq \sum_{i=k}^{L} \rho_i + \check{u}_k(\boldsymbol{\delta}) & 1 \leq k \leq L, \\ & z_k \geq 0. & \end{array}$$

where $\check{u}_k(\boldsymbol{\delta}) \triangleq \sum_{i=k}^{L} u_i(\delta_i)$ and $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_L]^T$. Defining $p^\star(\check{\mathbf{u}})$ as the optimal objective value of the relaxed problem as a function of $\check{\mathbf{u}} = [\check{u}_1, \cdots, \check{u}_L]^T$, the following local sensitivity result given in (3.14) holds (note that the problem satisfies the Slater's condition and therefore strong duality holds, c.f. §3.1):

$$\mu_k^\star = -\left.\frac{\partial p^\star}{\partial \check{u}_k}\right|_{\boldsymbol{\delta}=\mathbf{0}} \tag{6.19}$$

where $\mu_k^\star$ is the Lagrange multiplier of the Lagrangian of the relaxed problem (similar to (6.26)) at an optimal point. Using the chain rule for differentiation $\frac{\partial p^\star}{\partial \delta_i} = \sum_{k=1}^{L} \frac{\partial p^\star}{\partial \check{u}_k} \frac{\partial \check{u}_k}{\partial \delta_i}$, noting that $\frac{\partial \check{u}_k(\boldsymbol{\delta})}{\partial \delta_i} = \begin{cases} u_i(\delta_i) & i \geq k \\ 0 & \text{otherwise} \end{cases}$, and using (6.19), it follows that

$$\left.\frac{\partial p^\star}{\partial \delta_i}\right|_{\boldsymbol{\delta}=\mathbf{0}} = -\tilde{\mu}_i^\star \left.\frac{\partial u_i(\delta_i)}{\partial \delta_i}\right|_{\boldsymbol{\delta}=\mathbf{0}} \tag{6.20}$$

[6] Advanced communication systems are envisioned to exploit cross-layer signaling to further control the performance of the whole system.

where we have defined $\tilde{\mu}_i^\star \triangleq \sum_{k=1}^{i} \mu_k^\star$.

The largest value of $-\left.\frac{\partial p^\star}{\partial \delta_i}\right|_{\boldsymbol{\delta}=\mathbf{0}}$ for $1 \leq i \leq L$ indicates the QoS constraint that should be relaxed in order to get the largest reduction of the required transmitted power. Note that the optimal $\tilde{\mu}_i$'s used in (6.20) are readily given by the water-levels implicitly obtained in Algorithms 6.2 and 6.3: for each subblock $[k_1, k_2]$ of the partition on $[1, L]$, choose $\tilde{\mu}_k = \mu\left([k_1, k_2]\right)\ k_1 \leq k \leq k_2$. The term $\left.\frac{\partial u_i(\delta_i)}{\partial \delta_i}\right|_{\boldsymbol{\delta}=\mathbf{0}}$ in (6.20) depends on the particular cost function that relates the QoS in terms of MSE as used in the problem formulation and the QoS as seen by the service/user which can be in terms of MSE, SINR, or BER. We consider now a few examples.

Example 1: $\mathrm{MSE}_i \leq \rho_i + \delta_i$

In this case, it follows that $\frac{\partial u_i(\delta_i)}{\partial \delta_i} = 1$ and therefore $\left.\frac{\partial p^\star}{\partial \delta_i}\right|_{\boldsymbol{\delta}=\mathbf{0}} = -\tilde{\mu}_i^\star$. The largest $\tilde{\mu}_i^\star$ is given by $\tilde{\mu}_L^\star$ (or any other $\tilde{\mu}_i^\star$ belonging to the same water-filling subblock as obtained in Appendix 6.D). This means that the best way to relax the constraints in this case is by relaxing the tightest constraint.

Example 2: $\mathrm{MSE}_i \leq \mathrm{BER}^{-1}\left(p_i\left(1+\delta_i\right)\right)$

In this example, the QoS are given in terms of BER. The relaxation can be expressed as $\mathrm{MSE}_i \leq \rho_i + u_i\left(\delta_i\right)$ by defining $\rho_i \triangleq \mathrm{BER}^{-1}\left(p_i\right)$ and $u_i\left(\delta_i\right) \triangleq \mathrm{BER}^{-1}\left(p_i\left(1+\delta_i\right)\right) - \mathrm{BER}^{-1}\left(p_i\right)$. It follows that $\frac{\partial u_i(\delta_i)}{\partial \delta_i} = \frac{\partial\, \mathrm{BER}^{-1}(p_i(1+\delta_i))}{\partial \delta_i}$.

As a final remark, note that the previous sensitivity analysis is local and, therefore, the relaxation of the QoS constraints has to be performed in sufficiently small steps (recall that the decreasing order of the ρ_i's must be kept at each step).

6.6 Simulation Results

We now present numerical results obtained from simulations using realistic channel models for a wireless multi-antenna communication system and for a wireline (DSL) communication system.

6.6.1 Wireless Multi-Antenna Communication System

We consider a wireless communication system with multiple antennas at both sides of the link as was used for the numerical simulations in Chapter 5 (in particular 4 transmit and 4 receive antennas are considered). Perfect CSI is assumed at both sides of the communication system (channel estimation errors are considered in Chapter 7). The channel model includes the frequency-selectivity and the spatial correlation as measured in real scenarios. The MIMO channels were generated using the parameters of the WLAN European standard HIPERLAN/2 [ETS01], which is based on the multicarrier modulation OFDM (64 carriers were used in the simulations). The frequency selectivity of the channel was modeled using the power delay profile type

C for HIPERLAN/2 as specified in [ETS98a] (see Figure 5.3(a)), which corresponds to a typical large open space indoor environment for NLOS conditions with 150ns average r.m.s. delay spread and 1050ns maximum delay (the sampling period is 50ns [ETS01]). The spatial correlation of the MIMO channel was modeled according to the *Nokia* model defined in [Sch01] (for the uplink) as specified by the correlation matrices of the envelope of the channel fading at the transmit and receive side (see Figure 5.3(b)). It models a large open indoor environment with two floors, which could easily illustrate a conference hall or a shopping galleria scenario (see [Sch01] for details of the model). The matrix channel generated was normalized so that $\sum_n \mathbb{E}\left[\left| \left[\mathbf{H}(n)\right]_{ij} \right|^2 \right] = 1$.

The results are given in terms of required transmit power at some outage probability P_{out}, *i.e.*, the transmit power that will not suffice to satisfy the QoS constraints with a small probability P_{out} (it will suffice for $(1 - P_{\text{out}})$ of the time). In particular, we consider an outage probability of 5%. The reason of using the outage power instead of the average power is that, for typical wireless systems with delay constraints, the former is more realistic than the latter which only makes sense when the transmission coding block is long enough to reveal the long-term ergodic properties of the fading process (no delay constraints). Instead of plotting absolute values of the required transmit power, we plot relative values of the transmitted power normalized with the noise spectral density N_0. We call this normalized transmitted power SNR which is defined as $\text{SNR} = \text{Tr}\left(\mathbf{B}\mathbf{B}^H\right) / (N N_0)$ [7] where N is the number of carriers. Note that the plots are valid only for the channel normalization used.

In the following, numerical results for the proposed methods obtained in this chapter are given (Algorithm 6.1 for equal MSE QoS requirements and Algorithm 6.3 for different MSE QoS requirements). As a means of comparison, the simple benchmark obtained by imposing the diagonality of the MSE matrix as obtained in §6.4.1.3 is also considered. We consider two multicarrier approaches as described in §2.5.1.2: the carrier-noncooperative scheme (optimizing each carrier independently) and the carrier-cooperative scheme (treating all carriers as a single MIMO channel). Note that the total number of established substreams is LN where L denotes the number of spatial dimensions used per carrier.

Equal QoS Constraints

We now consider equal QoS requirements in terms of BER which correspond to equal QoS requirements in terms of MSE because the same constellation (QPSK) is used on all the substreams.

In Figure 6.2, the SNR per spatial dimension SNR/L is plotted as a function of L subject to equal QoS requirements given by $\text{BER} = 10^{-3}$. For $L = 3$, the gain over the benchmark is about 2dB. For $L = 4$, the gain is of 7dB for the carrier-noncooperative scheme and of 12dB for the carrier-cooperative scheme. The required power for $L = 4$ increases significantly with respect to

[7] Note that such an SNR definition is a measure of the total transmitted power per symbol (normalized with N_0) and does not correspond to the SNR at each receive antenna.

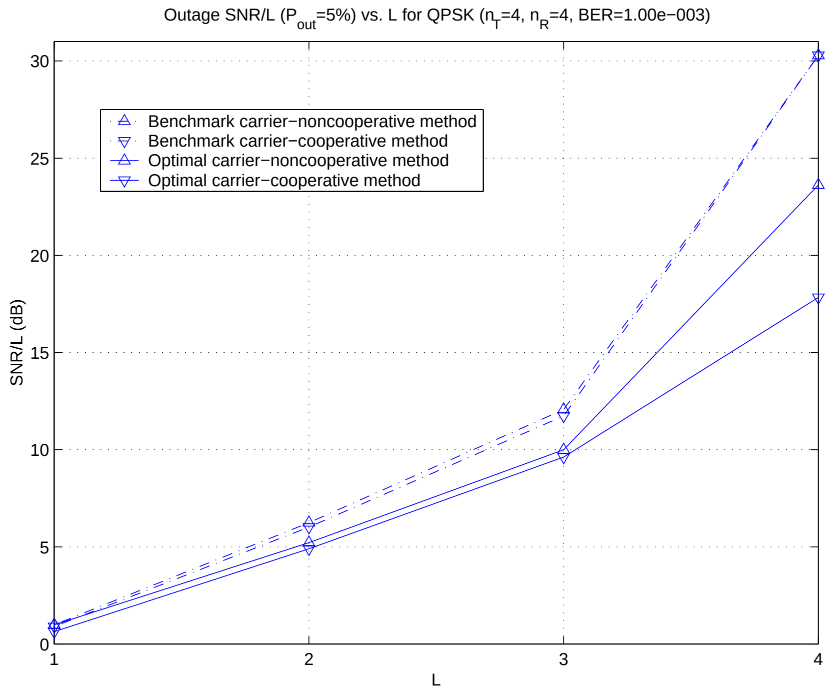


Figure 6.2: Outage SNR per spatial dimension vs. the number of spatial dimensions utilized L when using QPSK in a multicarrier 4×4 MIMO channel with an equal QoS for all substreams given by BER$\leq 10^{-3}$.

$L = 3$ and therefore we choose the latter for the rest of the simulations.

In Figure 6.3, the SNR is given as a function of the (equal) QoS constraints in terms of BER for $L = 3$. It can be observed that the gain over the benchmark is about 2-3dB and constant for all range of the BER. Cooperation among carriers improves the performance in no more than 0.5dB. Given that the carrier-noncooperative scheme has an attractive parallel implementation (since each carrier is independently processed), it may be an interesting solution for a practical and efficient implementation.

Different QoS Constraints

We now generalize the set-up by considering different QoS constraints both in terms of BER and MSE (since the same constellations are used on all the substreams).

In Figure 6.4, the SNR is plotted as a function of the nominal QoS corresponding to the first substream (the QoS constraints for the others substreams are obtained by scaling this nominal QoS constraint with the factors 0.5 and 0.1). Similar observations to those corresponding to Figure 6.3 hold in this case.

Relaxation of the QoS Requirements

In Figure 6.5, an example of a relaxed system is given (as explained in §6.5). The relaxed-I case consists of 10 relaxations per carrier with $\delta = 0.1$ of the form $\text{MSE}_i \leq \text{BER}^{-1}\left(p_i \, 10^{\delta_i}\right)$ (note that an extreme case of this series of relaxations amounts to increasing the BER of a single

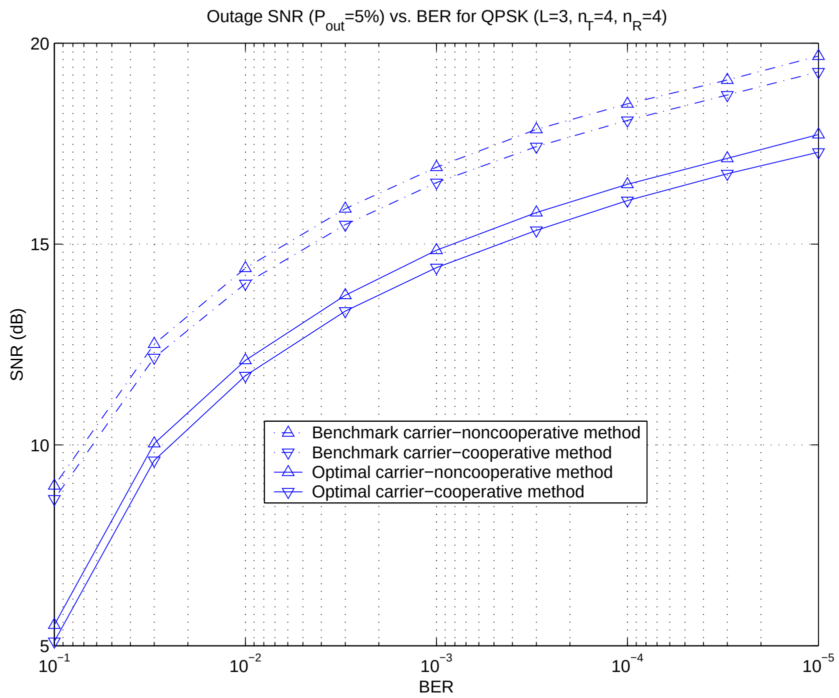


Figure 6.3: Outage SNR vs. the equal QoS constraints given in terms of BER when using QPSK in a multicarrier 4×4 MIMO channel with $L = 3$.

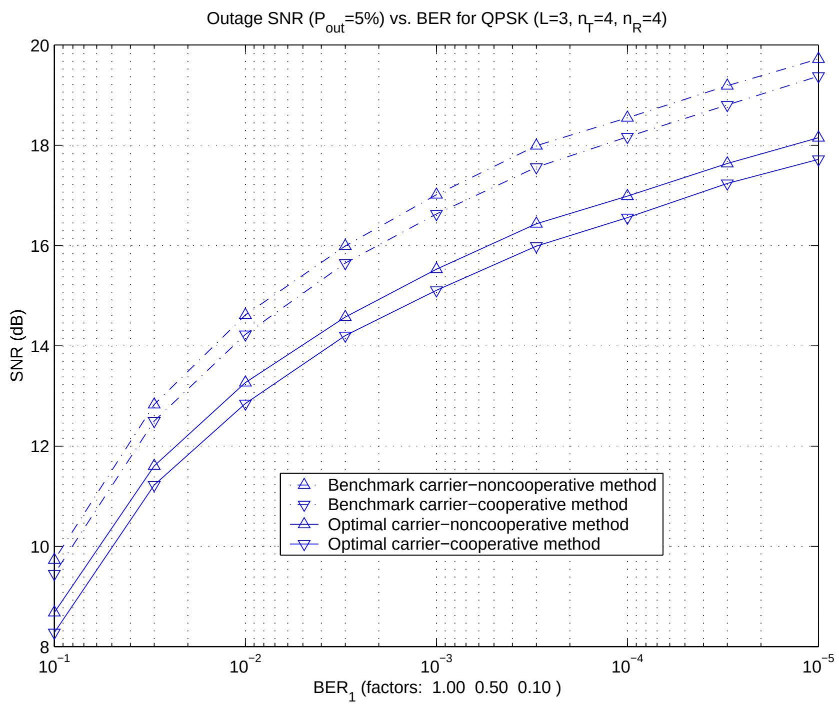


Figure 6.4: Outage SNR vs. the different QoS constraints given in terms of BER when using QPSK in a multicarrier 4×4 MIMO channel with $L = 3$. (The BER of the first substream is along the x-axis and the BER of the second and third substreams are given by scaling with the factors 0.5 and 0.1.)

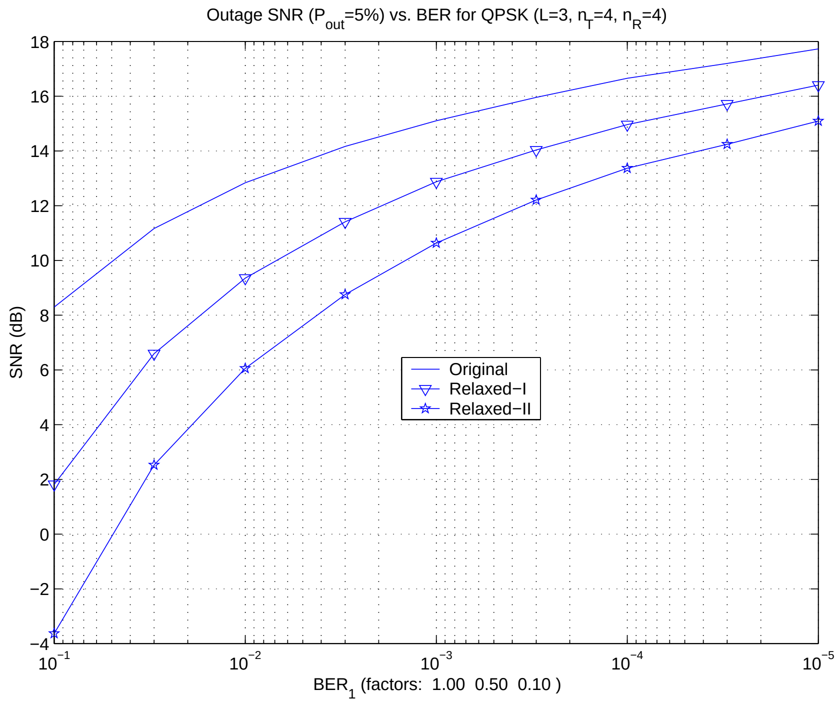


Figure 6.5: Outage SNR vs. the different QoS constraints given in terms of BER for the original carrier-cooperative method and for two succesive series of relaxations when using QPSK in a multicarrier 4×4 MIMO channel with $L = 3$. (The BER of the first substream is along the x-axis and the BER of the second and third substreams are given by scaling with the factors 0.5 and 0.1.)

substream one order of magnitude) and the relaxed-II case consists of 10 additional relaxations of the same type. As is observed, a significant reduction of the required transmit power can be achieved at the expense of the relaxation of some QoS requirements (recall that these relaxations are done in an optimal way).

6.6.2 Wireline (DSL) Communication System

We consider a VDSL system with typical parameters and DMT specifications: 4096 carriers on a bandwidth of 17.664 MHz, 26-gauge cable (AWG 26), background noise with a flat PSD at -140 dBm/Hz. We simulate a bundle with a total of 25 copper pairs, 20 of which are the intended users that can perform joint processing and the other 5 are considered as interference (no alien interference was included).

The MIMO DSL channel includes wide-scale frequency variations with statistics determined from measured FEXT transfer functions (see [Fan02] for details on the model). For illustrative purposes, a few realizations of the crosstalk are plotted in Figure 6.6.

The results are given in terms of required transmit power (since the channel is basically deterministic it is not necessary to take outage values as in the wireless case).

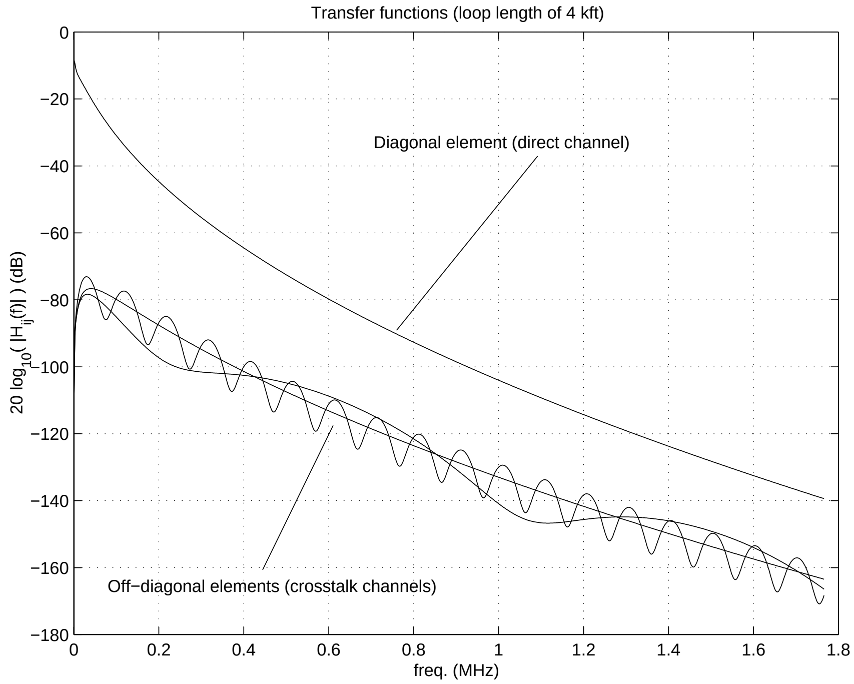


Figure 6.6: Example of a diagonal element and of some off-diagonal elements of the DSL channel matrix for a loop length of 4 kft.

For the simulations, we consider the obtained solutions to the problem of minimum transmitted power subject to a set of QoS requirements as given by Algorithm 6.1 for equal MSE QoS constraints and by Algorithm 6.3 for different MSE QoS constraints. As a means of comparison, the simple benchmark obtained by imposing the diagonality of the MSE matrix as obtained in §6.4.1.3 is also considered. We consider two multicarrier approaches as described in §2.5.1.2: the carrier-cooperative scheme (optimizing each carrier independently) and the carrier-noncooperative scheme (treating all carriers as a single MIMO channel).

Prior to using the benchmark and the proposed methods, a bit distribution is performed[8] using the gap approximation [Sta99, p. 206] (with a bit cap[9] of 12 bits) to obtain a probability of symbol error of $P_e = 10^{-7}$. Note that the benchmark method simply computes the minimum required transmit power for the given bit distribution with a diagonalized channel.

Transmitted Power vs. Loop Length

In Figure 6.7, the required transmit power to achieve $P_e = 10^{-7}$ on all the substreams[10] is plotted as a function of the loop length (for a nominal transmit power budget of 14.5dBm[11]). For

[8]The exact algorithm used for the bit distribution is not important for the results of the paper. A simple suboptimal bit distribution was used in the simulations.

[9]A bit cap is simply a constraint on the maximum number of transmitted bits per dimension.

[10]Note that the same BER constraints for different constellations turn into different MSE constraints and Algorithm 6.3 must be used.

[11]Note that the minimum required Tx-power corresponding to the benchmark method is never exactly equal to

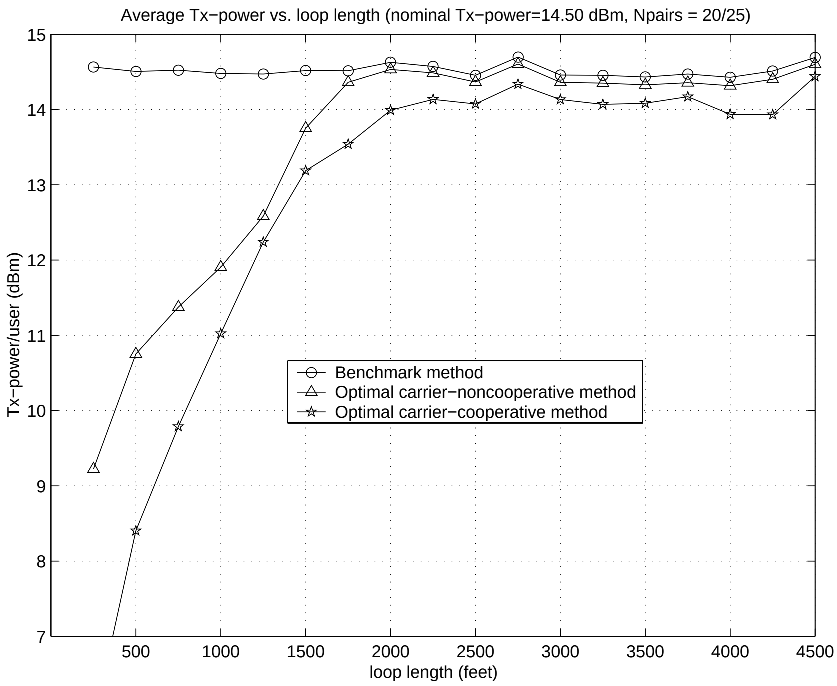


Figure 6.7: Required Tx-power to achieve $P_e = 10^{-7}$ as a function of the loop length (for a nominal Tx-power of 14.5dBm).

short loop lengths, the proposed method gives a saving in transmit power of 1-5dB for the carrier-noncooperative approach and of 2-8dB for the carrier-cooperative approach (this improvement comes with a slightly higher complexity). The improvement depends highly on the condition number of the whitened channel matrix. This is why for long loop lengths, where the interference is attenuated below the background noise level and the whitened channel matrix becomes almost a scaled identity, the improvement decreases significantly.

Transmitted Power vs. Nominal Power Budget

In Figure 6.8, the required transmit power to achieve $P_e = 10^{-7}$ on all the substreams is plotted as a function of the nominal transmit power budget (for a loop length of 1500 feet). It can be observed that the improvement increases for higher values of the transmit power.

6.7 Chapter Summary and Conclusions

In this chapter, we have formulated and solved the joint design of transmit and receive multiple beamvectors or beam-matrices (also known as linear precoder at the transmitter and linear equalizer at the receiver) for MIMO channels. In particular, the minimization of the transmitted power has been considered subject to (possibly different) QoS requirements for each of the

the nominal Tx-power due to the discrete nature of the bit distribution.

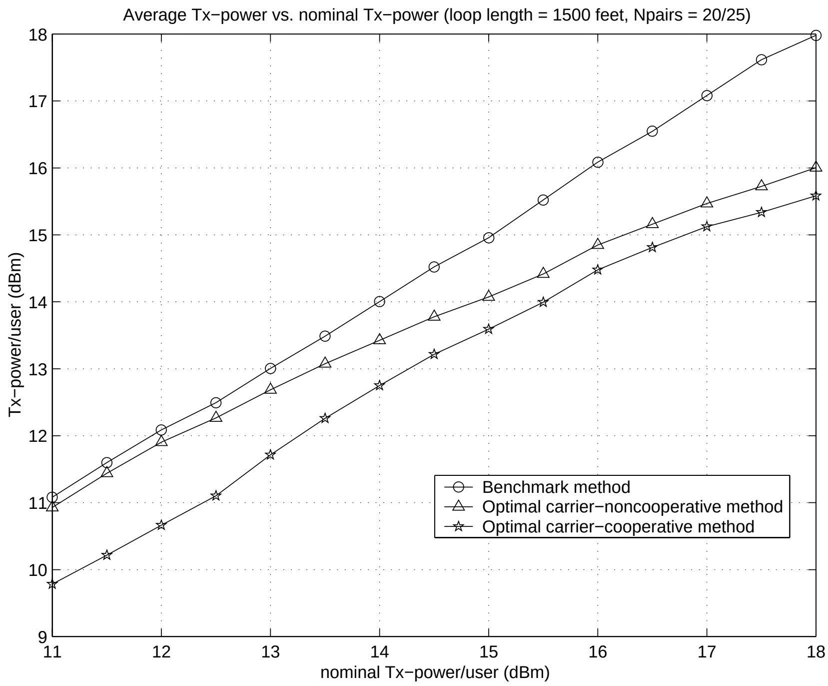


Figure 6.8: Required Tx-power to achieve $P_e = 10^{-7}$ as a function of the nominal Tx-power (for a loop length of 1500 feet).

established substreams in terms of MSE, SINR, and BER. Although the original problem formulation is a complicated nonconvex problem with matrix-valued variables, by using majorization theory we have been able to reformulate it as a simple convex optimization problem with scalar variables. To optimally solve the convex problem in practice, we have proposed a practical and efficient multi-level water-filling algorithm. For situations in which the required power results too large, a perturbation analysis has been conducted to obtain the optimal way in which the QoS requirements should be relaxed in order to reduce the power needed.

All the material presented in this chapter, which is strongly based on majorization theory and convex optimization theory, is a novel contribution of this dissertation.

Appendix 6.A Proof of Theorem 6.1

First rewrite the original problem as

$$\begin{array}{ll} \min\limits_{\mathbf{B}} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \\ \text{s.t.} & \max\limits_{i} \left[\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}\right]_{ii} \leq \rho. \end{array}$$

Note that this problem is exactly the opposite formulation of one of the design criteria considered in §5.5.5 (the MAX-MSE criterion) where $\max_i \left[\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}\right]_{ii}$ was minimized subject to a power constraint $\text{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T$.[12] We claim that matrix $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}$ must have equal diagonal elements at an optimal point and that the QoS constraints must be all satisfied with equality. Otherwise, by Corollary 3.2, we could find a unitary matrix $\mathbf{Q}$ with $\mathbf{Q}^H\mathbf{E}\mathbf{Q}$ having identical diagonal elements equal to $\frac{1}{L}\text{Tr}\left(\mathbf{E}\right)$ (this amounts to using $\mathbf{BQ}$ as transmit matrix instead of $\mathbf{B}$). Since $\frac{1}{L}\text{Tr}\left(\mathbf{E}\right) \leq \max_i \left[\mathbf{E}\right]_{ii}$ with equality if and only if all diagonal elements of $\mathbf{E}$ are equal, using $\mathbf{BQ}$ would satisfy all QoS constraints with strict inequality and the objective value could be further minimized by scaling down the whole transmit matrix. Therefore, the problem can be rewritten as

$$\begin{array}{ll} \min\limits_{\mathbf{B}} & \text{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \\ \text{s.t.} & \left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1} \quad \text{with equal diag. elements} \\ & \frac{1}{L}\text{Tr}\left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1} \leq \rho. \end{array}$$

By Corollary 3.2, it follows that for any given $\mathbf{B}$ we can always find a unitary matrix $\mathbf{Q}$ such that $\mathbf{Q}^H\left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\mathbf{Q}$ has equal diagonal elements. Therefore, we can simplify the problem by decomposing $\mathbf{B}$ as $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{Q}$ and imposing $\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}$ to be diagonal w.l.o.g. It is important to remark that imposing a diagonal structure on $\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}$ does not affect the problem since the objective value remains the same $\text{Tr}(\mathbf{B}\mathbf{B}^H) = \text{Tr}\left(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^H\right)$ and the QoS constraint also remains unchanged $\text{Tr}\left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1} = \text{Tr}\left(\mathbf{I}+\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}\right)^{-1}$. In addition, since $\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}$ is diagonal, the unitary matrix $\mathbf{Q}$ such that $\mathbf{Q}^H\left(\mathbf{I}+\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}\right)^{-1}\mathbf{Q}$ has equal diagonal elements can be found using Algorithm 3.2 (reproduced from [Vis99b, Section IV-A]) or with any rotation matrix that satisfies $\left|[\mathbf{Q}]_{ik}\right| = \left|[\mathbf{Q}]_{il}\right| \ \forall i,k,l$ such as the DFT matrix or the Hadamard matrix when the dimensions are appropriate (see §3.2 for more details).

The problem can be finally written as

$$\begin{array}{lll} \min\limits_{\tilde{\mathbf{B}}} & \text{Tr}\left(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^H\right) & \\ \text{s.t.} & \tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}} & \text{diagonal} \\ & \frac{1}{L}\text{Tr}\left(\mathbf{I}+\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}\right)^{-1} \leq \rho. & \end{array}$$

Since $\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}$ is diagonal, $\tilde{\mathbf{B}}$ can be assumed without loss of optimality (by Lemma 5.11) of the form $\tilde{\mathbf{B}} = \mathbf{U}_{H,1}\mathbf{\Sigma}_{B,1}$ where $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times \check{L}}$ has as columns the eigenvectors of $\mathbf{R}_H$ corresponding to the $\check{L} \triangleq \min\left(L, \text{rank}\left(\mathbf{R}_H\right)\right)$ largest eigenvalues in increasing order and $\mathbf{\Sigma}_{B,1} = \left[\mathbf{0} \ \text{diag}\left(\{\sigma_{B,i}\}\right)\right] \in \mathbb{C}^{\check{L} \times L}$ has zero elements except along the rightmost main diagonal (assumed real w.l.o.g.). The problem formulation of (6.12) follows by defining $z_i \triangleq \sigma_{B,i}^2$ and denoting with the set $\{\lambda_{H,i}\}_{i=1}^{\check{L}}$ the $\check{L}$ largest eigenvalues of $\mathbf{R}_H$ in increasing order. Note that the term $L_0 \triangleq L - \check{L}$ in (6.12) arises from the zero diagonal elements of $\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}$.

[12] Both problems are equivalent in the sense that they both describe the same curve of required power P_T for a given MSE constraint ρ (in the considered case the curve is parameterized with respect to ρ, $P_T(\rho)$, and in [Pal03c] the curve is parameterized with respect to P_T, $\rho(P_T)$).

Rewriting the MSE constraint in (6.12) as $\sum_{i=1}^{\check{L}} \frac{1}{1+z_i\lambda_{H,i}} \leq \rho L - L_0$, it becomes clear that it can be satisfied for sufficiently large values of the z_i's (equivalently, the problem is feasible) if and only if $\rho > L_0/L$. ■

Appendix 6.B Proof of Proposition 6.1

We first obtain the closed-form solution to the problem using convex optimization theory (see §3.1) and then proceed to prove the optimality of Algorithm 6.1.

Optimal Solution. The Lagrangian corresponding to the constrained convex problem is

$$\mathcal{L} = \sum_{i=1}^{\check{L}} z_i + \mu \left(\sum_{i=1}^{\check{L}} \frac{1}{1+z_i\lambda_i} - \tilde{\rho} \right) - \sum_{i=1}^{\check{L}} \gamma_i \, z_i,$$

where μ and the γ_i's are the dual variables or Lagrange multipliers. The water-filling solution is easily found from the sufficient and necessary KKT optimality conditions (the problem satisfies the Slater's condition and therefore strong duality holds):

$$\sum\nolimits_{i=1}^{\check{L}} \frac{1}{1+z_i\lambda_i} \leq \tilde{\rho}, \qquad z_i \geq 0 \tag{6.21}$$

$$\mu \geq 0, \qquad \gamma_i \geq 0 \tag{6.22}$$

$$\mu \frac{\lambda_i}{(1+z_i\lambda_i)^2} + \gamma_i = 1 \tag{6.23}$$

$$\mu \left(\sum\nolimits_{i=1}^{\check{L}} \frac{1}{1+z_i\lambda_i} - \tilde{\rho} \right) = 0, \quad \gamma_i z_i = 0. \tag{6.24}$$

Note that if $\mu = 0$, then $\gamma_i = 1 \; \forall i$ and $z_i = 0 \; \forall i$, which cannot be since the MSE constraint would not be satisfied because it was assumed that $\tilde{\rho} < \check{L}$. If $z_i > 0$, then $\gamma_i = 0$ (by the complementary slackness condition $\gamma_i z_i = 0$), $\mu \frac{\lambda_i}{(1+z_i\lambda_i)^2} = 1$ (note that $\mu \lambda_i > 1$), and $z_i = \mu^{1/2} \lambda_i^{-1/2} - \lambda_i^{-1}$. If $z_i = 0$, then $\mu \lambda_i + \gamma_i = 1$ (note that $\mu \lambda_i \leq 1$). Equivalently,

$$z_i = \begin{cases} \mu^{1/2} \lambda_i^{-1/2} - \lambda_i^{-1} & \text{if} \quad \mu\lambda_i > 1 \\ 0 & \text{if} \quad \mu\lambda_i \leq 1 \end{cases}$$

or, more compactly,

$$z_i = \left(\mu^{1/2} \lambda_i^{-1/2} - \lambda_i^{-1} \right)^+$$

where $\mu^{1/2}$ is the water-level chosen such that $\sum_{i=1}^{\check{L}} \frac{1}{1+z_i \lambda_i} = \tilde{\rho}$. This solution satisfies all KKT conditions and is therefore optimal.

Optimal Algorithm. Algorithm 6.1 is based on hypothesis testing. It first makes the assumption that all $\check{L}$ eigenmodes are active ($z_i > 0$ for $1 \leq i \leq \check{L}$) and then checks whether the MSE constraint could be satisfied with less power, in which case the current hypothesis is rejected, a new hypothesis with one less active eigenmode is made, and so forth.

In more detail, Algorithm 6.1 first reorders the eigenvalues in decreasing order. With this ordering, since $\lambda_i z_i = \left(\mu^{1/2} \lambda_i^{1/2} - 1\right)^+$, a hypothesis is completely described by the set of active eigenmodes $\tilde{L}$ (such that $z_i > 0$ for $1 \leq i \leq \tilde{L}$ and zero otherwise). This allows a reduction of the total number of hypotheses from $2^{\check{L}}$ to $\check{L}$. The initial hypothesis chooses the highest number of active eigenmodes $\tilde{L} = \check{L}$.

For each hypothesis, the water-level $\mu^{1/2}$ must be such that the considered $\tilde{L}$ eigenmodes are indeed active while the rest remain inactive:

$$\begin{cases} \mu^{1/2} \lambda_i^{-1/2} - \lambda_i^{-1} > 0 & 1 \leq i \leq \tilde{L} \\ \mu^{1/2} \lambda_i^{-1/2} - \lambda_i^{-1} \leq 0 & \tilde{L} < i \leq \check{L} \end{cases}$$

or, more compactly,

$$\lambda_{\tilde{L}}^{-1/2} < \mu^{1/2} \leq \lambda_{\tilde{L}+1}^{-1/2}$$

where we define $\lambda_{\check{L}+1} \triangleq 0$ for simplicity of notation. Assuming that $\lambda_{\tilde{L}} \neq \lambda_{\tilde{L}+1}$ (otherwise, the hypothesis is clearly rejected since the set of possible water-levels is empty), the algorithm checks whether the QoS constraint can be satisfied with the highest water-level of the subsequent hypothesis $\mu^{1/2} = \lambda_{\tilde{L}}^{-1/2}$, in which case the current hypothesis is rejected since the QoS constraint can be satisfied with a lower water-level and, equivalently, with lower z_i's (a reduced power). To be more precise, the algorithm checks whether $\sum_{i=1}^{\check{L}} \frac{1}{1+z_i \lambda_i} = \left(\check{L} - \tilde{L}\right) + \lambda_{\tilde{L}}^{1/2} \sum_{i=1}^{\tilde{L}} \lambda_i^{-1/2} \leq \tilde{\rho}$ or, equivalently, whether $\lambda_{\tilde{L}}^{-1/2} \geq \frac{\sum_{i=1}^{\tilde{L}} \lambda_i^{-1/2}}{\tilde{\rho} - (\check{L} - \tilde{L})}$.

If the current hypothesis is rejected, the algorithm forms a new hypothesis by decreasing $\tilde{L}$ to account for the decrease of the water-level. Otherwise, the current hypothesis is accepted since it contains the optimum water-level that satisfies the QoS constraint with equality (removing more active eigenmodes would keep the QoS constraint away of being satisfied and the addition of more active eigenmodes has already been tested and rejected for requiring a higher power to satisfy the QoS constraint). This reasoning can be applied as many times as needed for each remaining set of active eigenmodes. Once the optimal set of active eigenmodes is known, the active z_i's are obtained such that the QoS constraint is satisfied with equality and the definitive water-level is then

$$\mu^{1/2} = \frac{\sum_{i=1}^{\tilde{L}} \lambda_i^{-1/2}}{\tilde{\rho} - \left(\check{L} - \tilde{L}\right)}.$$

By the nature of the algorithm, the maximum number of iterations (worst-case complexity) is $\check{L}$. ■

Appendix 6.C Proof of Theorem 6.2

We prove the theorem in two steps. First, we show the equivalence of the original complicated problem and a simpler problem and, then, we solve the simple problem.

The original problem in (6.13) (problem P1) is equivalent to the following problem (problem P2):

$$\begin{array}{ll} \min_{\tilde{\mathbf{B}}} & \text{Tr}\left(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^H\right) \\ \text{s.t.} & \tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}} \qquad \text{diagonal} \\ & \mathbf{d}\left(\left(\mathbf{I}+\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}\right)^{-1}\right) \succ^w \boldsymbol{\rho}. \end{array}$$

Intuitively, the second constraint will guarantee the existence of a unitary matrix $\mathbf{Q}$ such that $\mathbf{d}\left(\mathbf{Q}^H\left(\mathbf{I}+\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}\right)^{-1}\mathbf{Q}\right) \leq \boldsymbol{\rho}$ (note the difference with respect to the case with equal MSE constraints for which there always exists $\mathbf{Q}$ such that $\mathbf{E}$ has equal diagonal elements by Corollary 3.2). To prove the equivalence of both problems, it suffices to show that for any feasible point $\mathbf{B}$ of problem P1 (*i.e.*, a point that satisfies the constraints of the problem) there is a corresponding feasible point $\tilde{\mathbf{B}}$ in problem P2 with the same objective value, *i.e.*, $\text{Tr}\left(\mathbf{B}\mathbf{B}^H\right) = \text{Tr}\left(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^H\right)$, and vice-versa. Therefore, solving one problem is tantamount to solving the other problem.

We prove first one direction. Let $\mathbf{B}$ be a feasible point of problem P1 with objective value $\text{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$. Define $\boldsymbol{\lambda}_B \triangleq \boldsymbol{\lambda}\left(\left(\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right)$ and $\mathbf{d}_B \triangleq \mathbf{d}\left(\left(\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right)$. Since $\mathbf{B}$ is feasible, $\mathbf{d}_B \leq \boldsymbol{\rho}$ and, therefore, $\mathbf{d}_B \succ^w \boldsymbol{\rho}$ (see Definition 3.3). It then follows by Lemma 3.6 that $\boldsymbol{\lambda}_B \succ^w \boldsymbol{\rho}$ ($\boldsymbol{\lambda}_B \succ \mathbf{d}_B \Rightarrow \boldsymbol{\lambda}_B \succ^w \mathbf{d}_B$). Find a unitary matrix $\mathbf{Q}$ that diagonalizes $\mathbf{B}^H\mathbf{R}_H\mathbf{B}$ and define $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{Q}$. It is straightforward to check that $\tilde{\mathbf{B}}$ is a feasible point of problem P2 (clearly $\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}$ is diagonal by selection of $\mathbf{Q}$ and, therefore, $\mathbf{d}_{\tilde{B}} = \boldsymbol{\lambda}_{\tilde{B}} = \boldsymbol{\lambda}_B \succ^w \boldsymbol{\rho}$) with the same objective value ($\text{Tr}\left(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^H\right) = \text{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$).

We prove now the other direction. Let $\tilde{\mathbf{B}}$ be a feasible point of problem P2. Since $\boldsymbol{\lambda}_{\tilde{B}} = \mathbf{d}_{\tilde{B}} \succ^w \boldsymbol{\rho}$, by Lemma 3.4, there exists a vector $\check{\boldsymbol{\rho}}$ such that $\check{\boldsymbol{\rho}} \leq \boldsymbol{\rho}$ and $\boldsymbol{\lambda}_{\tilde{B}} \succ \check{\boldsymbol{\rho}}$. We can now invoke Lemma 3.7 to show that there exists a unitary matrix $\mathbf{Q}$ such that $\mathbf{d}\left(\mathbf{Q}^H\left(\mathbf{I}+\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}\right)^{-1}\mathbf{Q}\right) = \check{\boldsymbol{\rho}} \leq \boldsymbol{\rho}$. Defining $\mathbf{B} = \tilde{\mathbf{B}}\mathbf{Q}$ (choosing $\mathbf{Q}$ such that the diagonal elements of $\left(\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}$ are in decreasing order), we have that $\mathbf{B}$ is a feasible point of problem P1 ($\mathbf{d}\left(\left(\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right) \leq \boldsymbol{\rho}$ or, equivalently, $\left[\left(\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right]_{ii} \leq \rho_i$). Note that if $\tilde{\mathbf{B}}$ is such that $\mathbf{d}_{\tilde{B}} \succ \boldsymbol{\rho}$, then $\mathbf{B}$ will satisfy the constraints of problem P1 with equality and vice-versa.

Now that problems P1 and P2 have been shown to be equivalent, we focus on solving problem P2 which is much simpler than problem P1. Since in problem P2 matrix $\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}$ is diagonal with diagonal elements in increasing order (recall that the diagonal elements of $\left(\mathbf{I}+\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}\right)^{-1}$ are considered in decreasing order because the ρ_i's are in decreasing order by definition), Lemma 5.11 can be invoked to show that $\tilde{\mathbf{B}}$ can be assumed without loss of optimality of the form

$\tilde{\mathbf{B}} = \mathbf{U}_{H,1}\mathbf{\Sigma}_{B,1}$ where $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times \check{L}}$ has as columns the eigenvectors of $\mathbf{R}_H$ corresponding to the $\check{L} \triangleq \min(L, \text{rank}(\mathbf{R}_H))$ largest eigenvalues in increasing order and $\mathbf{\Sigma}_{B,1} = [\mathbf{0} \ \text{diag}(\{\sigma_{B,i}\})] \in \mathbb{C}^{\check{L} \times L}$ has zero elements except along the rightmost main diagonal (assumed real w.l.o.g.). Writing the weakly majorization constraint of problem P2 explicitly according to Definition 3.3 (note that $\sum_{i=1}^{k} \rho_{(i)} = \sum_{i=L-k+1}^{L} \rho_i$ because the ρ_i's and the $\rho_{(i)}$'s are in decreasing and increasing ordering, respectively, and the same applies to the diagonal elements of $\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$), defining $z_i \triangleq \sigma_{B,i}^2$, and denoting with the set $\{\lambda_{H,i}\}_{i=1}^{\check{L}}$ the $\check{L}$ largest eigenvalues of $\mathbf{R}_H$ in increasing order, the problem reduces to

$$\begin{array}{lll}
\min\limits_{\{z_i\}} & \sum_{i=1}^{\check{L}} z_i & \\
\text{s.t.} & \sum_{i=k-L_0}^{\check{L}} \frac{1}{1+z_i \lambda_{H,i}} \leq \sum_{i=k}^{L} \rho_i & L_0 < k \leq L, \\
& (L_0 - k + 1) + \sum_{i=1}^{\check{L}} \frac{1}{1+z_i \lambda_{H,i}} \leq \sum_{i=k}^{L} \rho_i & 1 \leq k \leq L_0, \\
& z_k \geq 0, &
\end{array}$$

Note that the term $L_0 \triangleq L - \check{L}$ for the range $1 \leq k \leq L_0$ arises from the zero diagonal elements of $\tilde{\mathbf{B}}^H \mathbf{R}_H \tilde{\mathbf{B}}$.

To be exact, the explicit weakly majorization constraint should also include the ordering constraints:

$$\frac{1}{1 + z_k \lambda_{H,k}} \geq \frac{1}{1 + z_{k+1} \lambda_{H,k+1}} \qquad 1 \leq k < \check{L}. \tag{6.25}$$

Note that the remaining ordering constraints are trivially verified since $1 \geq \frac{1}{1+z_k \lambda_{H,k}}$. However, it is not necessary to include such ordering constraints since an optimal solution always satisfies them. Otherwise, we could reorder the terms $(z_i \lambda_{H,i})$'s to satisfy (6.25) and the solution obtained this way would still satisfy the other constraints of the convex problem with the same objective value. At this point, however, the $\lambda_{H,i}$'s would not be in increasing order and by, Lemma 5.11, this is not an optimal solution since the terms $(z_i \lambda_{H,i})$'s could be put back in increasing order with a lower objective value.

The problem formulation of (6.14) follows by noting that, since $\rho_i < 1$, the constraints for $1 \leq k \leq L_0$ imply and are implied by the constraint for $k = 1$: $\sum_{i=1}^{\check{L}} \frac{1}{1+z_i \lambda_{H,i}} \leq \sum_{i=1}^{L} \rho_i - L_0$. This constraint can be satisfied for sufficiently large values of the z_i's (equivalently, the problem is feasible) if and only if $\sum_{i=1}^{L} \rho_i > L_0$ (the constraints for $L_0 < k \leq L$ can always be satisfied).

It is straightforward to see that $\sum_{i=1}^{\check{L}} \frac{1}{1+z_i \lambda_{H,i}} \leq \sum_{i=1}^{L} \rho_i - L_0$ must be satisfied with equality at an optimal point. Otherwise, z_1 could be decreased until it is satisfied with equality or z_1 becomes zero (in which case, the same reasoning applies to z_2 and so forth). This means that an optimal solution to problem P2 must satisfy $\mathbf{d}\left(\left(\mathbf{I}+\tilde{\mathbf{B}}^H \mathbf{R}_H \tilde{\mathbf{B}}\right)^{-1}\right) \succ \boldsymbol{\rho}$, which in turn implies that the QoS constraints in problem P1 must be satisfied with equality: $\left[\left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}\right]_{ii} = \rho_i$. ■

Appendix 6.D Proof of Proposition 6.2

We first present a lemma and then proceed to prove Proposition 6.2.

Lemma 6.3 *In Algorithm 6.2 (for different QoS constraints), the water-level of each outer iteration (if more than one) is strictly lower than that of the previous iteration.*

Proof. Let $\mu([k_1, k_2])$ denote the squared water-level when applying the single water-filling of Algorithm 6.1 on $[k_1, k_2]$.

For any outer iteration that has more than one inner iteration (by inner iteration we refer to one execution of steps 1 and 2 of the outer iteration), after the first inner iteration we obtain k_0 which is the smallest index whose constraint $\sum_{i=k_0}^{\tilde{L}} \frac{1}{1+z_i\lambda_i} \leq \sum_{i=k_0}^{\tilde{L}} \tilde{\rho}_i$ is not satisfied. It follows that $\mu([1,\tilde{L}]) < \mu([k_0,\tilde{L}])$ because the squared water-level obtained in the first inner iteration $\mu([1,\tilde{L}])$ was not high enough to satisfy the constraint at k_0 and has to be strictly increased. It also follows that $\mu([k,\tilde{L}]) \leq \mu([1,\tilde{L}])$ for $1 < k \leq k_0$ since the water-filling over $[1,\tilde{L}]$ also satisfies the constraints for $1 < k < k_0$. Therefore, we have that $\mu([k,\tilde{L}]) < \mu([k_0,\tilde{L}])$ for $1 \leq k < k_0$. The same reasoning applies to all subsequent inner iterations. Thus, after each outer iteration on $[1,\tilde{L}]$, we have that $\mu([k,\tilde{L}]) < \mu([k_0,\tilde{L}])$ for $1 \leq k < k_0$.

The following outer iteration (if any) is on $[1, k_0 - 1]$. Any water-filling performed in this outer iteration verifies $\mu([k, k_0 - 1]) < \mu([k,\tilde{L}])$ for $1 \leq k \leq k_0 - 1$ as we now show. Using $\mu([k,\tilde{L}])$ on $[k,\tilde{L}]$ implies that the constraint is satisfied with equality $\sum_{i=k}^{\tilde{L}} \frac{1}{1+z_i\lambda_i} = \sum_{i=k}^{\tilde{L}} \tilde{\rho}_i$. We have shown before that $\mu([k_0,\tilde{L}]) > \mu([k,\tilde{L}])$ for $1 \leq k < k_0$; therefore, using $\mu([k,\tilde{L}])$ only on $[k, k_0 - 1]$ and $\mu([k_0,\tilde{L}])$ on $[k_0,\tilde{L}]$ the constraint is satisfied with strict inequality, $\sum_{i=k}^{\tilde{L}} \frac{1}{1+z_i\lambda_i} < \sum_{i=k}^{\tilde{L}} \tilde{\rho}_i$ which can be rewritten as $\sum_{i=k}^{k_0-1} \frac{1}{1+z_i\lambda_i} < \sum_{i=k}^{k_0-1} \tilde{\rho}_i$ (recalling that $\sum_{i=k_0}^{\tilde{L}} \frac{1}{1+z_i\lambda_i} = \sum_{i=k_0}^{\tilde{L}} \tilde{\rho}_i$ since we are using $\mu([k_0,\tilde{L}])$ in that range). Since using $\mu([k,\tilde{L}])$ only on $[k, k_0 - 1]$ implies $\sum_{i=k}^{k_0-1} \frac{1}{1+z_i\lambda_i} < \sum_{i=k}^{k_0-1} \tilde{\rho}_i$ and using $\mu([k, k_0 - 1])$ implies $\sum_{i=k}^{k_0-1} \frac{1}{1+z_i\lambda_i} = \sum_{i=k}^{k_0-1} \tilde{\rho}_i$ (by definition), it follows that $\mu([k,\tilde{L}]) > \mu([k, k_0 - 1])$.

Finally, since $\mu([k, k_0 - 1]) < \mu([k,\tilde{L}])$ and $\mu([k,\tilde{L}]) < \mu([k_0,\tilde{L}])$, it follows that $\mu([k, k_0 - 1]) < \mu([k_0,\tilde{L}])$ for $1 \leq k < k_0$. In other words, the water-level of each outer iteration is strictly lower than that of the previous one. ■

Proof of Proposition 6.2. We, first, obtain the closed-form solution to the problem using convex optimization theory (see §3.1) and, then, proceed to prove the optimality of Algorithm 6.2.

Optimal Solution. The Lagrangian corresponding to the constrained convex problem is

$$\mathcal{L} = \sum_{k=1}^{\check{L}} z_k + \sum_{k=1}^{\check{L}} \mu_k \left(\sum_{i=k}^{\check{L}} \frac{1}{1 + z_i\lambda_i} - \sum_{i=k}^{\check{L}} \tilde{\rho}_i \right) - \sum_{k=1}^{\check{L}} \gamma_k \, z_k, \tag{6.26}$$

where the μ_k's and the γ_k's are the dual variables or Lagrange multipliers. The water-filling solution is easily found from the sufficient and necessary KKT optimality conditions (the problem satisfies the Slater's condition and therefore strong duality holds):

$$\sum_{i=k}^{\check{L}} \frac{1}{1+z_i\lambda_i} \le \sum_{i=k}^{\check{L}} \tilde{\rho}_i, \qquad z_k \ge 0 \tag{6.27}$$

$$\mu_k \ge 0, \qquad \gamma_k \ge 0 \tag{6.28}$$

$$\left(\sum_{i=1}^{k} \mu_i\right) \frac{\lambda_k}{(1+z_k\lambda_k)^2} + \gamma_k = 1 \tag{6.29}$$

$$\mu_k \left(\sum_{i=k}^{\check{L}} \frac{1}{1+z_i\lambda_i} - \sum_{i=k}^{\check{L}} \tilde{\rho}_i\right) = 0, \qquad \gamma_k\, z_k = 0. \tag{6.30}$$

It is important to point out here that μ_1 cannot be zero at an optimal solution as we now shown. If $\mu_1 = 0$, then $\gamma_1 = 1$ and $z_1 = 0$. It follows then, from the inequality $\sum_{i=1}^{\check{L}} \frac{1}{1+z_i\lambda_i} \le \sum_{i=1}^{\check{L}} \tilde{\rho}_i$ and using the trivial assumption that $\tilde{\rho}_1 < 1$, that $\sum_{i=2}^{\check{L}} \frac{1}{1+z_i\lambda_i} < \sum_{i=2}^{\check{L}} \tilde{\rho}_i$ (note the strict inequality), which in turn implies that $\mu_2 = 0$. This reasoning can be repeatedly applied for $k = 2, \cdots, \check{L}$ to show that if $\mu_1 = 0$ then $\mu_k = 0 \ \forall k$, but this cannot be since it would imply that $z_k = 0 \ \forall k$ and then the constraints $\sum_{i=k}^{\check{L}} \frac{1}{1+z_i\lambda_i} \le \sum_{i=k}^{\check{L}} \tilde{\rho}_i \quad 1 \le k \le \check{L}$ would not be satisfied. Thus, it must be that $\mu_1 > 0$ which implies that $\sum_{i=1}^{\check{L}} \frac{1}{1+z_i\lambda_i} = \sum_{i=1}^{\check{L}} \tilde{\rho}_i$.

By defining $\tilde{\mu}_k \triangleq \sum_{i=1}^{k} \mu_i$, the KKT conditions involving the μ_k's can be more compactly rewritten as (define $\tilde{\mu}_0 \triangleq 0$):

$$\begin{aligned} &\tilde{\mu}_k \ge \tilde{\mu}_{k-1} && 1 \le k \le L \\ &\tilde{\mu}_k \tfrac{\lambda_k}{(1+z_k\lambda_k)^2} + \gamma_k = 1 && \\ &(\tilde{\mu}_k - \tilde{\mu}_{k-1}) \left(\textstyle\sum_{i=k}^{\check{L}} \tfrac{1}{1+z_i\lambda_i} - \sum_{i=k}^{\check{L}} \tilde{\rho}_i\right) = 0. && \end{aligned} \tag{6.31}$$

If $z_k > 0$, then $\gamma_k = 0$ (by the complementary slackness condition $\gamma_k\, z_k = 0$), $\tilde{\mu}_k \frac{\lambda_k}{(1+z_k\lambda_k)^2} = 1$ (note that $\tilde{\mu}_k\,\lambda_k > 1$), and $z_k = \tilde{\mu}_k^{1/2}\,\lambda_k^{-1/2} - \lambda_k^{-1}$. If $z_k = 0$, then $\tilde{\mu}_k\,\lambda_k + \gamma_k = 1$ (note that $\tilde{\mu}_k\,\lambda_k \le 1$). Equivalently,

$$z_k = \begin{cases} \tilde{\mu}_k^{1/2}\,\lambda_k^{-1/2} - \lambda_k^{-1} & \text{if} \quad \tilde{\mu}_k\lambda_k > 1 \\ 0 & \text{if} \quad \tilde{\mu}_k\lambda_k \le 1 \end{cases}$$

or, more compactly,

$$z_k = \left(\tilde{\mu}_k^{1/2}\,\lambda_k^{-1/2} - \lambda_k^{-1}\right)^+$$

where the water-levels $\tilde{\mu}_k^{1/2}$'s are chosen to satisfy the remaining KKT conditions:

$$\begin{aligned} &\textstyle\sum_{i=k}^{\check{L}} \tfrac{1}{1+z_i\lambda_i} \le \sum_{i=k}^{\check{L}} \tilde{\rho}_i && 1 < k \le \check{L} \\ &\textstyle\sum_{i=1}^{\check{L}} \tfrac{1}{1+z_i\lambda_i} = \sum_{i=1}^{\check{L}} \tilde{\rho}_i && \\ &\tilde{\mu}_k \ge \tilde{\mu}_{k-1} \quad (\tilde{\mu}_0 \triangleq 0) && \\ &(\tilde{\mu}_k - \tilde{\mu}_{k-1}) \left(\textstyle\sum_{i=k}^{\check{L}} \tfrac{1}{1+z_i\lambda_i} - \sum_{i=k}^{\check{L}} \tilde{\rho}_i\right) = 0. && \end{aligned}$$

This solution satisfies all KKT conditions and is therefore optimal.

Optimal Algorithm. We now prove the optimality of the solution given by Algorithm 6.2 (and also of the equivalent Algorithm 6.3) by showing that the solution it gives satisfies the KKT conditions (6.27)-(6.30). Note that after running Algorithm 6.2, the set $\left[1, \check{L}\right]$ is partitioned into subsets, each one solved by a single water-filling given by Algorithm 6.1. By the construction of the algorithm, the constraints $\sum_{i=k}^{\check{L}} \frac{1}{1+z_i\lambda_i} \leq \sum_{i=k}^{\check{L}} \rho_i$ are clearly satisfied. Since the algorithm produces a partition on the set $\left[1, \check{L}\right]$ (each subset solved by a single water-filling) the following conditions are necessarily satisfied (from (6.21), (6.22) and (6.24)):

$$z_k \geq 0, \quad \gamma_k \geq 0, \quad \text{and} \quad \gamma_k z_k = 0.$$

The remaining conditions are given by (6.31). If for each subblock of the partition on $\left[1, \check{L}\right]$, $[k_1, k_2]$, we choose $\tilde{\mu}_k = \mu\left([k_1, k_2]\right)$ $k_1 \leq k \leq k_2$, it can be readily checked that they are satisfied. From Lemma 6.3, it follows that $\tilde{\mu}_k \geq \tilde{\mu}_{k-1}$ is verified. Also, $\tilde{\mu}_k \frac{\lambda_k}{(1+z_k\lambda_k)^2} + \gamma_k = 1$ is satisfied by the nature of the single water-filling solution (see (6.23)). Finally, since on each subblock $[k_1, k_2]$ the water-level is fixed, it follows that $(\tilde{\mu}_k - \tilde{\mu}_{k-1}) = 0$ for $k_1 < k \leq k_2$ and that $\sum_{i=k_1}^{k_2} \frac{1}{1+z_i\lambda_i} = \sum_{i=k_1}^{k_2} \tilde{\rho}_i$; hence, condition $(\tilde{\mu}_k - \tilde{\mu}_{k-1}) \left(\sum_{i=k}^{\check{L}} \frac{1}{1+z_i\lambda_i} - \sum_{i=k}^{\check{L}} \tilde{\rho}_i\right) = 0$ is satisfied for $k_1 \leq k \leq k_2$.

The worst-case number of outer iterations in Algorithm 6.2 is $\check{L}$ and the worst-case number of inner iterations (simple water-fillings) for an outer iteration on $\left[1, \tilde{L}\right]$ is $\tilde{L}$; consequently, the worst-case number of total inner iterations is $\check{L}\left(\check{L}+1\right)/2$. If, instead, we evaluate the complexity in terms of basic iterations (iterations within each simple water-filling), the worst-case number of these basic iterations is approximately equal to $\check{L}^2\left(\check{L}+1\right)/6$. ■

Appendix 6.E Proof of Lemma 6.1

Since the MSE matrix $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}$ is constrained to be diagonal (recall that by definition the ρ_i's are in decreasing order and the diagonal elements of $\mathbf{E}$ can be assumed in decreasing order w.l.o.g.), it follows from Lemma 5.11 that an optimal solution can be expressed as $\mathbf{B} = \mathbf{U}_{H,1}\mathbf{\Sigma}_{B,1}$ where $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times \check{L}}$ has as columns the eigenvectors of $\mathbf{R}_H$ corresponding to the $\check{L} \triangleq \min\left(L, \text{rank}\left(\mathbf{R}_H\right)\right)$ largest eigenvalues in increasing order ($L_0 \triangleq L - \check{L}$ is the number of zero eigenvalues used) and $\mathbf{\Sigma}_{B,1} = \left[\mathbf{0} \ \text{diag}\left(\{\sigma_{B,i}\}\right)\right] \in \mathbb{C}^{\check{L} \times L}$ has zero elements except along the rightmost main diagonal (assumed real w.l.o.g.). Defining $z_i \triangleq \sigma_{B,i}^2$ and denoting with the set $\{\lambda_{H,i}\}_{i=1}^{\check{L}}$ the $\check{L}$ largest eigenvalues of $\mathbf{R}_H$ in increasing order, the original problem is then simplified to the convex problem

$$\begin{array}{lll} \min\limits_{\{z_i\}} & \sum_{i=1}^{\check{L}} z_i & \\ \text{s.t.} & \frac{1}{1+z_i\lambda_{H,i}} \leq \rho_{i+L_0} & 1 \leq i \leq \check{L}, \\ & 1 \leq \rho_i & \check{L} < i \leq L, \\ & z_i \geq 0. & \end{array}$$

The problem is clearly feasible if and only if $\rho_i \geq 1$ for $\check{L} < i \leq L$, but this cannot be since by definition we know that $\rho_i < 1$. Therefore, the problem is feasible if and only if $\check{L} = L$ or, equivalently, $L \leq \text{rank}(\mathbf{R}_H)$. In such a case, the optimal solution to the problem is trivially given by

$$z_i = \lambda_{H,i}^{-1} \left(\rho_i^{-1} - 1\right).$$

■

Appendix 6.F Proof of Lemma 6.2

To study the optimality of the solution obtained in Lemma 6.1 (under the diagonality constraint of the MSE matrix) with respect the original problem in Theorem 6.2 (without the diagonality constraint), it suffices to check under which conditions the solution obtained in Lemma 6.1 satisfies the KKT conditions obtained in the proof of Proposition 6.2 (which solves the convex problem obtained in Theorem 6.2).

Since $z_i > 0$ for $1 \leq i \leq L$, it must be that $\gamma_i = 0$ for $1 \leq i \leq L$ and, therefore, $\tilde{\mu}_i \frac{\lambda_i}{(1+z_i\lambda_i)^2} = 1 \Longrightarrow \tilde{\mu}_i = \frac{(1+z_i\lambda_i)^2}{\lambda_i} = \frac{1}{\lambda_i \rho_i^2}$ (recall that $\lambda_i \triangleq \lambda_{H,i}$). At this point, all KKT conditions (6.27)-(6.30) are clearly satisfied except $\tilde{\mu}_i \geq \tilde{\mu}_{i-1} \quad 1 < i \leq L$ which is satisfied if and only if:

$$\lambda_i \rho_i^2 \geq \lambda_{i+1} \rho_{i+1}^2 \qquad 1 \leq i < L$$

which clearly cannot be satisfied if $L > \text{rank}(\mathbf{R}_H)$ since the λ_i's are in increasing order. ■

Chapter 7

Robust Design against Channel Estimation Errors

THIS CHAPTER PROPOSES SEVERAL WAYS TO INCLUDE ROBUSTNESS in the solutions previously obtained in Chapters 5 and 6, where perfect Channel State Information (CSI) was assumed. Two different philosophies are considered for robust designs: the worst-case and the stochastic (Bayesian) approaches. The ultimate goal of this chapter is to obtain simple robust solutions that can be readily implemented in practice. The problem of robust design, however, deserves a more in-depth treatment (rather than merely a chapter) and several lines for future research are pointed out.

7.1 Introduction

A common problem in practical communication systems arises from channel estimation errors (see [Cox87] and references therein) which cause imperfect Channel State Information (CSI). For a given communication channel, the best spectral efficiency is obviously achieved when perfect CSI is available at both sides of the link. In a realistic wireless environment, however, CSI has to be periodically estimated due to the random nature of the channel and, because of channel estimation errors, only non-perfect CSI can be obtained [Ben01] (see Figure 7.1). This effect is of paramount importance in practical implementations and should be taken into account when designing a system.

CSI at the receiver (CSIR) is traditionally acquired via the transmission of a training sequence (pilot symbols) that allows the estimation of the channel (it is also possible to use blind methods that exploit the structure of the transmitted signal or of the channel (c.f. §2.4)). CSI at the transmitter (CSIT) can be obtained either by means of a feedback channel from the receiver to the transmitter or from previous receive measurements thanks to the channel reciprocity property

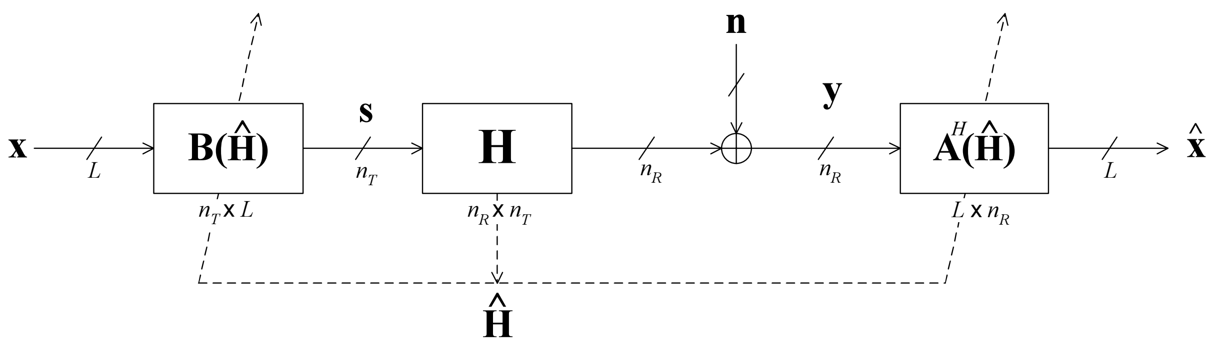


Figure 7.1: Scheme of a MIMO communication system in which the transmitter **B** and receiver **A** are explicitly designed as a function of the channel estimation $\hat{\mathbf{H}}$ rather than the real channel **H**.

(c.f. §2.4). See [Ben01] for a description of channel estimation strategies in real systems. Whereas a sufficiently accurate CSIR can be assumed in many cases, CSIT will be far from perfect in any realistic situation. Hence, as a first approximation, one can assume perfect CSIR and imperfect CSIT, although it would be more exact to consider imperfect CSIR as well.

To design properly the transmit and receive processing matrices **B** and **A**, it is necessary to obtain CSI, which not only includes knowledge of the propagation channel **H**, but also of the receive noise covariance matrix $\mathbf{R}_n$. If, however, the receiver has imperfect CSI given by $\hat{\mathbf{H}}$ and $\hat{\mathbf{R}}_n$, then a robust design should be used for the derivation of the transmit and/or the receive matrices. There are two basic schemes for the joint design of the transmitter and the receiver as we now describe (see Figure 7.2 for a classification).

On the one hand, the receiver can acquire CSIR, design both **A** and **B**, and then feed **B** back to the transmitter. In this case, there are two basic possibilities to analyze (see Figure 7.2):

- perfect CSI (treated in Chapters 5 and 6), and
- imperfect CSI which affects the design of both **A** and **B**.

On the other hand, the receiver may acquire CSIR and design only **A**, whereas the transmitter should obtain CSIT (either by using a feedback channel from the receiver or using the channel reciprocity property[1]) and design **B**. In this case, there are three basic possibilities to study (see Figure 7.2):[2]

- perfect CSIR and perfect CSIT (treated in Chapters 5 and 6),
- perfect CSIR and imperfect CSIT (recall that this is a first approximation to the problem of channel estimation errors, since we can assume that the receiver can estimate the channel

[1] In this case, the noise covariance matrix still has to be fed back to the transmitter or assumed white.

[2] From a physical standpoint, CSIR is generally better than CSIT. We do not consider the case of perfect CSIT and imperfect CSIR due to its relatively rare occurrence in the real world [Big98].

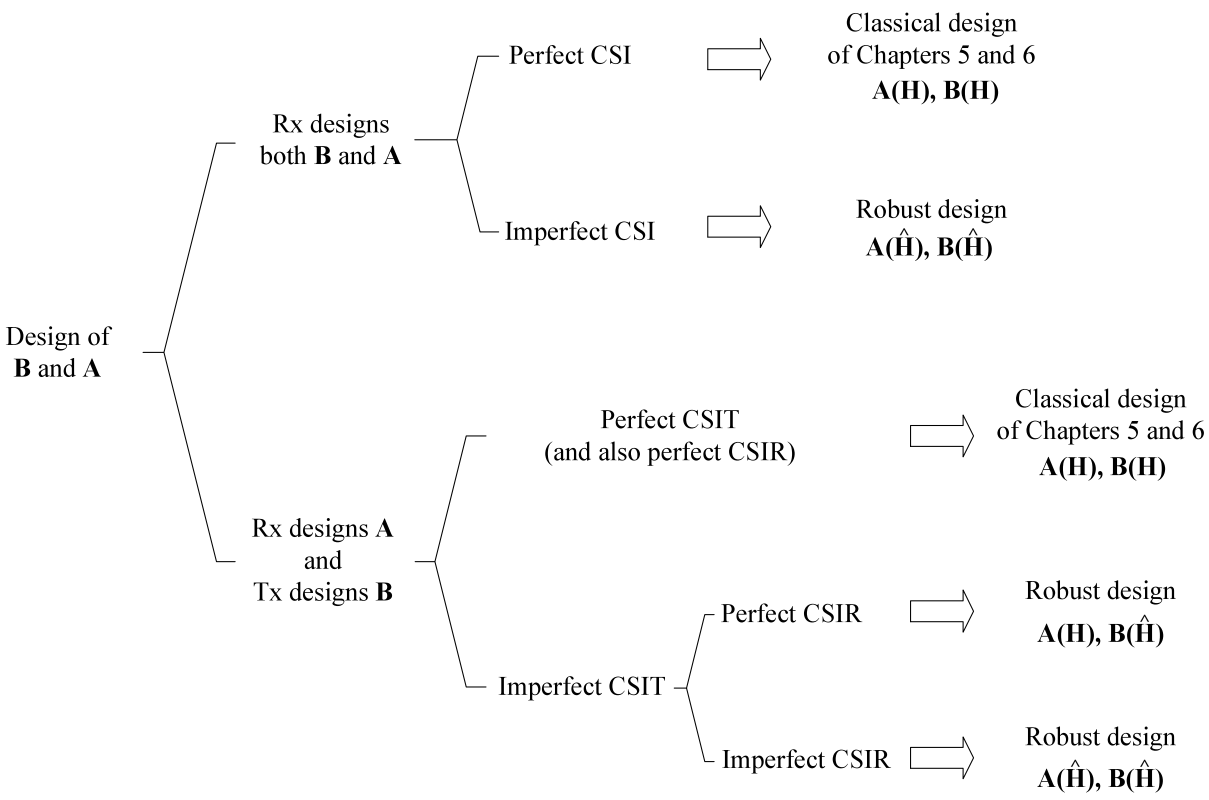


Figure 7.2: Classification of the classical (naive) vs. robust approaches for: i) the case in which the receiver estimates the channel, designs both **A** and **B**, and feeds **B** back to the transmitter, and ii) the case in which the receiver estimates the channel and designs **A** and, independently, the transmitter obtains a channel estimate and designs **B**.

with a sufficient accuracy whereas the transmitter will always have some error in the channel estimation due to the time-varying nature of the channel), and

- imperfect CSIR and imperfect CSIT.

In this chapter, for the sake of notation, we consider $L \leq \text{rank}\,(\mathbf{H})$[3] (although the results are valid for other values of L as well) and we formulate the robust design using the signal model for the case of multiple beamforming (single beamforming is just a particular case) in a single MIMO channel (the case of multiple MIMO channels can be readily obtained).

Part of the results of this chapter were obtained in [Pal03f].

7.2 Worst-Case vs. Stochastic Robust Designs

There are two main approaches to deal with imperfect CSI: the worst-case and the stochastic (Bayesian) viewpoints [Boy00].

[3]Recall that L is the number of transmitted symbols or, equivalently, the number of established substreams.

On the one hand, the worst-case design guarantees a certain system performance for any possible channel sufficiently close to the estimated one, provided that the estimation error is bounded as expected (in a real situation, this assumption will be satisfied with some high probability, declaring an outage otherwise). This approach is extremely pessimistic (in most cases unnecessarily), which translates into a significant increase of the required transmit power.

In [Cox87], a worst-case robust approach was taken in the context of classical beamforming design at the receiver, obtaining the nice result that the heuristic approach of adding a small scaled identity matrix to the noise covariance matrix is optimum when a white noise gain constraint is imposed (see also [Vor03]). In [Ben99, Ben01], a worst-case robust approach was considered in the context of multiuser communications between several mobile users and a multi-antenna base station.

On the other hand, the stochastic design only guarantees a certain system performance averaged over the channels that could have caused the current estimated channel. It follows a Bayesian philosophy by considering the distribution function of the actual channel conditioned on the obtained estimation. In other words, it optimizes the average performance given the estimation of the channel. In this sense, nothing can be said for a given realization of the actual and estimated channels. The stochastic approach avoids the high pessimism inherent to the worst-case design with the consequent saving in transmit power.

Regarding the stochastic robustness, the main concern has been with respect to imperfect CSIT while assuming perfect CSIR (although the classical beamforming scenario with multiples antennas only at the receiver has also been analyzed from a Bayesian perspective [Bel00]). Robust designs of transmit beamforming with imperfect CSIT have been considered according to different criteria such as error probability [Wit95], SNR [Nar98], and mutual information [Nar98, Vis01]. In [Vis01], robust transmission schemes were obtained for imperfect CSIT in the form of either the mean or the covariance matrix of the channel distribution function (see also [Jaf01] for the case of multiple transmit and receive antennas). In [Nar98], imperfect CSIT was considered not only stemming from channel estimation errors but also from the effect of quantizing the channel estimate at the receiver to be fed back as side information. Partial CSI for the design of single-antenna systems modeled as MIMO channels was considered in [Bar01], where a design maximizing the average SNR was adopted. A stochastic robust design for MIMO channels was obtained in [Mil00] to minimize the sum of MSE's with imperfect CSIT and perfect CSIR (a ZF receiver was assumed). In [Rey02], a robust solution to the minimization of the sum of the MSE's was derived for multi-antenna multicarrier systems with imperfect CSIT and CSIR. Robust minimum BER solutions for multicarrier systems with imperfect CSIT and perfect CSIR were designed in [Ise03] and [Rey03] for the single-antenna and multi-antenna cases, respectively.

In [Neg99], the combination of beamforming and space-time codes in a system with multiple antennas at both sides of the link was considered aiming at maximizing the average SNR with partial CSIT (modeled from a physical perspective). The appealing combination of beamforming and space-time codes with partial CSIT was also taken in [Jön02] from a statistical viewpoint to minimize the pairwise error probability for different degrees of CSIT (from no CSIT to perfect CSIT).

In the rest of the chapter, we formulate the worst-case and the stochastic robust designs of the transmitter and receiver for MIMO channels under the general framework considered in Chapters 5 and 6.

7.3 Worst-Case Robust Design

The underlying idea of a worst-case robust design against parameter estimation errors is to assume that the actual parameter can be expressed as the estimated value plus some error bounded with some norm. The design is then based on the worst error that satisfies the bound, hence the name worst-case robust design. In the framework of communication systems, such a robust approach is in general pessimistic, but guarantees the system performance under all possible channels under the previous assumption of a bounded error.

We consider the case of imperfect CSIT and perfect CSIR and also the more realistic case of imperfect CSIT and imperfect CSIR. Part of the results in this section were obtained in [Pal03f].

7.3.1 Error Modeling and Problem Formulation

Channel and Noise Estimation Errors

Instead of taking the channel estimate $\hat{\mathbf{H}}$ as perfect (naive approach), the worst-case robust approach assumes that the actual channel $\mathbf{H}$ can be written as

$$\mathbf{H} = \hat{\mathbf{H}} + \mathbf{H}_{\Delta} \tag{7.1}$$

where $\mathbf{H}_{\Delta}$ represents the channel estimation error bounded as $\|\mathbf{H}_{\Delta}\| \leq \epsilon_H$ [Ben99, Ben01] or as $\|\mathbf{H}_{\Delta}\| / \|\hat{\mathbf{H}}\| \leq \epsilon$ for some small ϵ_H or ϵ (both constraints are equivalent if $\epsilon_H \triangleq \epsilon \|\hat{\mathbf{H}}\|$).

The same idea applies to the estimated noise covariance matrix $\hat{\mathbf{R}}_n$, *i.e.*, we assume that the actual $\mathbf{R}_n$ can be written as

$$\mathbf{R}_n = \hat{\mathbf{R}}_n + \mathbf{R}_{n,\Delta} \tag{7.2}$$

where $\mathbf{R}_{n,\Delta}$ represents the estimation error bounded as $\|\mathbf{R}_{n,\Delta}\| \leq \epsilon_n$.

We now state a lemma that will prove very useful in the sequel.

Lemma 7.1 *Given the matrices $\mathbf{X} \in \mathbb{C}^{n\times m}$ and $\mathbf{Y} \in \mathbb{C}^{n\times m}$, it follows that*

$$\mathbf{X}\mathbf{Y}^H + \mathbf{Y}\mathbf{X}^H \leq 2\sigma_{\max}(\mathbf{X})\,\sigma_{\max}(\mathbf{Y})\,\mathbf{I}_n$$

and, in particular, that

$$\mathbf{X}\mathbf{X}^H \leq \lambda_{\max}\left(\mathbf{X}\mathbf{X}^H\right)\mathbf{I}_n.$$

Proof. See Appendix 7.A. ■

Problem Formulation with Perfect CSIR and Imperfect CSIT

Under the assumption of perfect CSIR, the optimum receive matrix $\mathbf{A}$ for a given transmit matrix $\mathbf{B}$ is given by the Wiener filter $\mathbf{A} = \left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H + \mathbf{R}_n\right)^{-1}\mathbf{H}\mathbf{B}$ as obtained in §2.5.5 and the instantaneous MSE matrix is then $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}$ as in (2.51).

Both the power-constrained problem formulation of Chapter 5 and the QoS-constrained problem formulation of Chapter 6 depend on the channel $\mathbf{H}$ and the noise covariance matrix $\mathbf{R}_n$ through the squared whitened channel matrix $\mathbf{R}_H = \mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}$ (see, for example, (5.9) and (6.10)). Therefore, we need to obtain the worst-case error induced in $\mathbf{R}_H$.

We can easily deal with the estimation error of the noise covariance matrix by considering the norm on $\mathbf{S}_{++}^n$ (where $\mathbf{S}_{++}^n$ is the set of Hermitian positive definite $n \times n$ matrices) induced by the Euclidean norm on $\mathbb{C}^n$, *i.e.*, the spectral norm, leading to $\lambda_{\max}\left(\mathbf{R}_{n,\Delta}\right) \leq \epsilon_n$. The spectral norm implies that $\mathbf{R}_{n,\Delta} \leq \epsilon_n \mathbf{I}$. It then follows from (7.2) that $\mathbf{R}_n \leq \widehat{\mathbf{R}}_n \triangleq \hat{\mathbf{R}}_n + \epsilon_n \mathbf{I}$ and, using the fact that $\mathbf{X} \geq \mathbf{Y} \Rightarrow \mathbf{X}^{-1} \leq \mathbf{Y}^{-1}$ (for invertible matrices), we finally have that

$$\mathbf{R}_H \geq \mathbf{H}^H \widehat{\mathbf{R}}_n^{-1} \mathbf{H}. \tag{7.3}$$

To deal with the estimation error of the channel, consider again the norm on $\mathbb{C}^{n\times m}$ induced by the Euclidean norm on $\mathbb{C}^n$, *i.e.*, the spectral norm, leading to $\sigma_{\max}\left(\mathbf{H}_\Delta\right) \leq \epsilon_H$. Assuming for the moment only channel estimation errors, we can decompose $\mathbf{R}_H$ as $\mathbf{R}_H = \hat{\mathbf{R}}_H + \mathbf{R}_{H,\Delta}$, where $\hat{\mathbf{R}}_H \triangleq \hat{\mathbf{H}}^H\mathbf{R}_n^{-1}\hat{\mathbf{H}}$ and $\mathbf{R}_{H,\Delta} \triangleq \hat{\mathbf{H}}^H\mathbf{R}_n^{-1}\mathbf{H}_\Delta + \mathbf{H}_\Delta^H\mathbf{R}_n^{-1}\hat{\mathbf{H}} + \mathbf{H}_\Delta^H\mathbf{R}_n^{-1}\mathbf{H}_\Delta$. It follows (using Lemma 7.1) that $\mathbf{R}_{H,\Delta}$ is bounded as

$$\left|\lambda_i\left(\mathbf{R}_{H,\Delta}\right) - \frac{\epsilon_H^2}{\lambda_{\min}(\mathbf{R}_n)}\right| \leq 2\frac{\epsilon_H\,\sigma_{\max}(\hat{\mathbf{H}})}{\lambda_{\min}(\mathbf{R}_n)}.$$

Since $\epsilon_H \ll \sigma_{\max}(\hat{\mathbf{H}})$, we can ignore the quadratic term and finally write:

$$\mathbf{R}_H = \hat{\mathbf{R}}_H + \mathbf{R}_{H,\Delta} \qquad \begin{array}{l} \left|\lambda_i\left(\mathbf{R}_{H,\Delta}\right)\right| \leq \epsilon_R, \\ \mathbf{R}_H = \mathbf{R}_H^H \geq \mathbf{0}, \end{array} \tag{7.4}$$

where $\epsilon_R \triangleq 2\frac{\epsilon_H\,\sigma_{\max}(\hat{\mathbf{H}})}{\lambda_{\min}(\mathbf{R}_n)}$ and we have made explicit the inherent constraints on $\mathbf{R}_{H,\Delta}$ to guarantee the positive semidefiniteness of $\mathbf{R}_H$. (If the Frobenius norm is used instead of the maximum-singular-value norm, similar bounds on $\lambda_i\left(\mathbf{R}_{H,\Delta}\right)$ are obtained.) From (7.4) and using the eigen-decomposition $\hat{\mathbf{R}}_H = \hat{\mathbf{U}}_H\hat{\mathbf{D}}_H\hat{\mathbf{U}}_H^H$, the eigenvalues of $\mathbf{R}_H$ can be shown to be lower-bounded as

$$\lambda_{\min}\left(\hat{\mathbf{R}}_H + \mathbf{R}_{H,\Delta}\right) \geq \left(\lambda_{\min}(\hat{\mathbf{R}}_H) - \epsilon_R\right)^+$$
$$\vdots$$
$$\lambda_i\left(\hat{\mathbf{R}}_H + \mathbf{R}_{H,\Delta}\right) \geq \left(\lambda_i(\hat{\mathbf{R}}_H) - \epsilon_R\right)^+$$
$$\vdots$$
$$\lambda_1\left(\hat{\mathbf{R}}_H + \mathbf{R}_{H,\Delta}\right) \geq \left(\lambda_1(\hat{\mathbf{R}}_H) - \epsilon_R\right)^+$$

where $\lambda_i(\cdot)$ is the ith eigenvalue in decreasing order. The lower bounds $\lambda_i(\mathbf{R}_H) \geq \left(\lambda_i(\hat{\mathbf{R}}_H) - \epsilon_R\right)^+$ can be compactly expressed as

$$\mathbf{R}_H \geq \hat{\mathbf{U}}_H\left(\hat{\mathbf{D}}_H - \epsilon_R \mathbf{I}\right)^+ \hat{\mathbf{U}}_H^H. \tag{7.5}$$

Combining now the lower bounds on $\mathbf{R}_H$ due to the noise covariance error (7.3) (obtained by using the upper bound $\widehat{\mathbf{R}}_n = \hat{\mathbf{R}}_n + \epsilon_n \mathbf{I}$ as the noise covariance matrix) and due to the channel error (7.5), we can finally bound the squared whitened channel matrix as

$$\mathbf{R}_H \geq \breve{\mathbf{R}}_H \triangleq \hat{\mathbf{U}}_{H,n}\left(\hat{\mathbf{D}}_{H,n} - \epsilon_{R,n}\mathbf{I}\right)^+ \hat{\mathbf{U}}_{H,n}^H \tag{7.6}$$

where $\hat{\mathbf{H}}^H \widehat{\mathbf{R}}_n^{-1} \hat{\mathbf{H}} = \hat{\mathbf{U}}_{H,n}\hat{\mathbf{D}}_{H,n}\hat{\mathbf{U}}_{H,n}^H$ is the eigendecomposition of $\hat{\mathbf{R}}_H$ when using the upper bound $\widehat{\mathbf{R}}_n$ instead of $\hat{\mathbf{R}}_n$ and $\epsilon_{R,n} \triangleq 2\frac{\epsilon_H \sigma_{\max}(\hat{\mathbf{H}})}{\lambda_{\min}(\hat{\mathbf{R}}_n)+\epsilon_n}$. Note that the lower bound on $\mathbf{R}_H$ (7.6) implies the following upper bound on the MSE matrix:

$$\mathbf{E} \leq \widehat{\mathbf{E}} \triangleq \left(\mathbf{I} + \mathbf{B}^H \breve{\mathbf{R}}_H \mathbf{B}\right)^{-1} \tag{7.7}$$

where $\widehat{\mathbf{E}}$ is the worst-case MSE matrix.

Thus, a worst-case robust design with imperfect CSIT and perfect CSIR can be obtained in practice simply by using the lower bound $\breve{\mathbf{R}}_H$ in lieu of $\mathbf{R}_H$ (c.f. §7.3.2 and §7.3.3). This reduces to adding a small scaled identity matrix to the noise covariance matrix (commonly termed "diagonal loading" [Cox87, Car88, Vor03]) and then slightly decreasing the eigenvalues of the squared whitened channel matrix.

Problem Formulation with Imperfect CSIR and Imperfect CSIT

In case of imperfect CSIR, the receiver has to be designed in a robust way as well. The MSE matrix (2.46) is

$$\mathbf{E} = \left(\mathbf{A}^H\mathbf{H}\mathbf{B} - \mathbf{I}\right)\left(\mathbf{B}^H\mathbf{H}^H\mathbf{A} - \mathbf{I}\right) + \mathbf{A}^H\mathbf{R}_n\mathbf{A}$$

where $\mathbf{H}$ and $\mathbf{R}_n$ have an estimation error as in (7.1) and (7.2), respectively. The noise component is easily bounded as before (using the spectral norm) $\mathbf{A}^H\mathbf{R}_n\mathbf{A} \leq \mathbf{A}^H\widehat{\mathbf{R}}_n\mathbf{A}$ where $\widehat{\mathbf{R}}_n \triangleq \hat{\mathbf{R}}_n + \epsilon_n\mathbf{I}$. The remaining component can be bounded as

$$\begin{aligned}\left(\mathbf{A}^H\mathbf{H}\mathbf{B} - \mathbf{I}\right)\left(\mathbf{B}^H\mathbf{H}^H\mathbf{A} - \mathbf{I}\right) &= \left(\left(\mathbf{A}^H\hat{\mathbf{H}}\mathbf{B} - \mathbf{I}\right) + \mathbf{A}^H\mathbf{H}_\Delta\mathbf{B}\right)\left(\left(\mathbf{B}^H\hat{\mathbf{H}}^H\mathbf{A} - \mathbf{I}\right) + \mathbf{B}^H\mathbf{H}_\Delta^H\mathbf{A}\right) \\ &\leq \left(\mathbf{A}^H\hat{\mathbf{H}}\mathbf{B} - \mathbf{I}\right)\left(\mathbf{B}^H\hat{\mathbf{H}}^H\mathbf{A} - \mathbf{I}\right) + \epsilon_{AB}\mathbf{A}^H\mathbf{A}\end{aligned}$$

where $\epsilon_{AB} = \epsilon_H^2 \lambda_{\max}\left(\mathbf{B}\mathbf{B}^H\right) + 2\epsilon_H \sigma_{\max}(\hat{\mathbf{H}}) \lambda_{\max}\left(\mathbf{B}\mathbf{B}^H\right) + 2\epsilon_H \sigma_{\max}(\mathbf{B})/\sigma_{\min}(\mathbf{A})$ (note that the first term can be ignored since $\epsilon_H \ll \sigma_{\max}(\hat{\mathbf{H}})$) and we have used the following bounds (obtained using Lemma 7.1 and assuming that $\mathbf{A}$, which is a tall matrix, is full rank):

$$\begin{aligned} \mathbf{A}^H \mathbf{H}_\Delta \mathbf{B}\mathbf{B}^H \mathbf{H}_\Delta^H \mathbf{A} &\leq \epsilon_H^2 \lambda_{\max}\left(\mathbf{B}\mathbf{B}^H\right) \mathbf{A}^H \mathbf{A} \\ \mathbf{A}^H \hat{\mathbf{H}} \mathbf{B}\mathbf{B}^H \mathbf{H}_\Delta^H \mathbf{A} + \mathbf{A}^H \mathbf{H}_\Delta \mathbf{B}\mathbf{B}^H \hat{\mathbf{H}}^H \mathbf{A} &\leq 2\epsilon_H \sigma_{\max}(\hat{\mathbf{H}}) \lambda_{\max}\left(\mathbf{B}\mathbf{B}^H\right) \mathbf{A}^H \mathbf{A} \\ \mathbf{A}^H \mathbf{H}_\Delta \mathbf{B} + \mathbf{B}^H \mathbf{H}_\Delta^H \mathbf{A} &\geq -2\epsilon_H \frac{\sigma_{\max}(\mathbf{B})}{\sigma_{\min}(\mathbf{A})} \mathbf{A}^H \mathbf{A}. \end{aligned}$$

The MSE matrix can be finally upper-bounded as

$$\mathbf{E} \leq \widehat{\mathbf{E}} \triangleq \mathbf{A}^H \left(\hat{\mathbf{H}} \mathbf{B}\mathbf{B}^H \hat{\mathbf{H}}^H + \hat{\mathbf{R}}_n + \xi \mathbf{I}\right) \mathbf{A} + \mathbf{I} - \mathbf{A}^H \hat{\mathbf{H}} \mathbf{B} - \mathbf{B}^H \hat{\mathbf{H}}^H \mathbf{A} \tag{7.8}$$

where $\xi = \epsilon_n + 2\epsilon_H \left(\sigma_{\max}(\hat{\mathbf{H}}) \lambda_{\max}\left(\mathbf{B}\mathbf{B}^H\right) + \lambda_{\max}^{1/2}\left(\mathbf{B}\mathbf{B}^H\right) / \lambda_{\min}^{1/2}\left(\mathbf{A}^H \mathbf{A}\right)\right)$[4] and $\widehat{\mathbf{E}}$ is the worst-case MSE matrix. Although the scalar ξ depends on $\mathbf{B}$ and $\mathbf{A}$, this can be circumvented by approximating the values of $\lambda_{\max}\left(\mathbf{B}\mathbf{B}^H\right)$ and $\lambda_{\min}\left(\mathbf{A}^H \mathbf{A}\right)$ with the average value of past realizations. In practice, however, the exact value of ξ should be obtained from numerical simulations to avoid the high pessimism inherent in each of the bounds used (c.f. §7.5). The optimum receive matrix that minimizes the diagonal elements of the worst-case MSE matrix $\widehat{\mathbf{E}}$ is easily obtained (similarly to the Wiener filter in §2.5.5) as

$$\mathbf{A}_\xi = \left(\hat{\mathbf{H}} \mathbf{B}\mathbf{B}^H \hat{\mathbf{H}}^H + \hat{\mathbf{R}}_n + \xi \mathbf{I}\right)^{-1} \hat{\mathbf{H}} \mathbf{B} \tag{7.9}$$

and the resulting MSE matrix is then

$$\widehat{\mathbf{E}} = \left(\mathbf{I} + \mathbf{B}^H \hat{\mathbf{H}}^H (\hat{\mathbf{R}}_n + \xi \mathbf{I})^{-1} \hat{\mathbf{H}} \mathbf{B}\right)^{-1}. \tag{7.10}$$

Thus, a worst-case robust design with imperfect CSIT and CSIR can be obtained in practice simply by diagonal loading, *i.e.*, by adding a small scaled identity matrix to the noise covariance matrix [Cox87, Car88, Vor03].

In the following subsections, the robust design is formulated for power-constrained and QoS-constrained systems (extending the results of Chapters 5 and 6, respectively). With imperfect CSIR, the receiver can only use the fixed receive matrix of (7.9) for the given estimated channel and noise covariance matrix. With perfect CSIR, however, the receive matrix is assumed to be optimally designed for each possible channel and noise covariance matrix.

[4]Since $\mathbf{A}$ is a tall matrix, it follows that $\sigma_{\min}(\mathbf{A}) = \lambda_{\min}^{1/2}\left(\mathbf{A}^H \mathbf{A}\right)$.

7.3.2 Power-Constrained Systems

Perfect CSIR and Imperfect CSIT

The worst-case power-constrained design, with perfect CSIR and imperfect CSIT, can be formulated (similarly to (5.9)) as

$$\begin{array}{lll} \min_{\mathbf{B}} & \max_{\mathbf{H},\mathbf{R}_n} f_0\left(\left\{\left[\left(\mathbf{I}+\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}\right]_{ii}\right\}_{i=1}^{L}\right) & \\ \text{s.t.} & \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T, & \\ & \mathbf{H} = \hat{\mathbf{H}} + \mathbf{H}_\Delta & \mathbf{H}_\Delta : \|\mathbf{H}_\Delta\| \leq \epsilon_H, \\ & \mathbf{R}_n = \hat{\mathbf{R}}_n + \mathbf{R}_{n,\Delta} & \mathbf{R}_{n,\Delta} : \ \|\mathbf{R}_{n,\Delta}\| \leq \epsilon_n, \\ & & \mathbf{R}_n = \mathbf{R}_n^H > \mathbf{0}, \end{array} \tag{7.11}$$

or, more compactly, as

$$\begin{array}{lll} \min_{\mathbf{B}} & \max_{\mathbf{R}_H} f_0\left(\left\{\left[\left(\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right]_{ii}\right\}_{i=1}^{L}\right) & \\ \text{s.t.} & \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T, & \\ & \mathbf{R}_H = \hat{\mathbf{R}}_H + \mathbf{R}_{H,\Delta} & \mathbf{R}_{H,\Delta} : \ \left|\lambda_i\left(\mathbf{R}_{H,\Delta}\right)\right| \leq \epsilon_{R,n}, \\ & & \mathbf{R}_H = \mathbf{R}_H^H \geq \mathbf{0}. \end{array} \tag{7.12}$$

Now, since $f_0\left(\{x_i\}\right)$ is nondecreasing in each x_i by definition (see §5.2 for details), we can use the upper bound on the MSE matrix of (7.7) to finally write the problem as

$$\begin{array}{ll} \min_{\mathbf{B}} & f_0\left(\left\{\left[\left(\mathbf{I}+\mathbf{B}^H\breve{\mathbf{R}}_H\mathbf{B}\right)^{-1}\right]_{ii}\right\}_{i=1}^{L}\right) \\ \text{s.t.} & \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T, \end{array} \tag{7.13}$$

where the only difference with respect to the non-robust design of (5.9) is the use of $\breve{\mathbf{R}}_H$ instead of $\mathbf{R}_H$.

Imperfect CSIR and Imperfect CSIT

For the case of imperfect CSIR and CSIT, the worst-case power-constrained design can be formulated (similarly to (5.9)) as

$$\begin{array}{lll} \min_{\mathbf{B},\mathbf{A}} & \max_{\mathbf{H},\mathbf{R}_n} f_0\left(\left\{\left[\mathbf{A}^H\left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H+\mathbf{R}_n\right)\mathbf{A}+\mathbf{I}-\mathbf{A}^H\mathbf{H}\mathbf{B}-\mathbf{B}^H\mathbf{H}^H\mathbf{A}\right]_{ii}\right\}_{i=1}^{L}\right) & \\ \text{s.t.} & \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T, & \\ & \mathbf{H} = \hat{\mathbf{H}} + \mathbf{H}_\Delta & \mathbf{H}_\Delta : \|\mathbf{H}_\Delta\| \leq \epsilon_H, \\ & \mathbf{R}_n = \hat{\mathbf{R}}_n + \mathbf{R}_{n,\Delta} & \mathbf{R}_{n,\Delta} : \ \|\mathbf{R}_{n,\Delta}\| \leq \epsilon_n, \\ & & \mathbf{R}_n = \mathbf{R}_n^H > \mathbf{0}. \end{array} \tag{7.14}$$

Since $f_0(\{x_i\})$ is nondecreasing in each x_i by definition (c.f. §5.2), we can use the upper bound on the MSE matrix of (7.8) and the problem reduces then to

$$\begin{array}{ll} \min\limits_{\mathbf{B},\mathbf{A}} & f_0\left(\left\{\left[\mathbf{A}^H(\hat{\mathbf{H}}\mathbf{B}\mathbf{B}^H\hat{\mathbf{H}}^H + \hat{\mathbf{R}}_n + \xi\mathbf{I})\mathbf{A} + \mathbf{I} - \mathbf{A}^H\hat{\mathbf{H}}\mathbf{B} - \mathbf{B}^H\hat{\mathbf{H}}^H\mathbf{A}\right]_{ii}\right\}_{i=1}^{L}\right) \\ \text{s.t.} & \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T, \end{array} \quad (7.15)$$

from which the optimum $\mathbf{A}$ is given by (7.9) and the upper bound on the MSE matrix reduces to (7.10). The problem can be finally written as

$$\begin{array}{ll} \min\limits_{\mathbf{B}} & f_0\left(\left\{\left[\left(\mathbf{I} + \mathbf{B}^H\hat{\mathbf{H}}^H(\hat{\mathbf{R}}_n + \xi\mathbf{I})^{-1}\hat{\mathbf{H}}\mathbf{B}\right)^{-1}\right]_{ii}\right\}_{i=1}^{L}\right) \\ \text{s.t.} & \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \leq P_T, \end{array} \quad (7.16)$$

where the only difference with respect to the non-robust design of (5.9) is the use of $\hat{\mathbf{R}}_n + \xi\mathbf{I}$ instead of $\mathbf{R}_n$.

7.3.3 QoS-Constrained Systems

Perfect CSIR and Imperfect CSIT

The worst-case QoS-constrained design, with perfect CSIR and imperfect CSIT, can be formulated (similarly to (6.10)) as

$$\begin{array}{lll} \min\limits_{\mathbf{B}} & \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) & \\ \text{s.t.} & \max\limits_{\mathbf{H},\mathbf{R}_n}\left[\left(\mathbf{I} + \mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}\right]_{ii} \leq \rho_i, & 1 \leq i \leq L, \\ & \mathbf{H} = \hat{\mathbf{H}} + \mathbf{H}_\Delta & \mathbf{H}_\Delta : \|\mathbf{H}_\Delta\| \leq \epsilon_H, \\ & \mathbf{R}_n = \hat{\mathbf{R}}_n + \mathbf{R}_{n,\Delta} & \mathbf{R}_{n,\Delta} : \ \|\mathbf{R}_{n,\Delta}\| \leq \epsilon_n, \\ & & \qquad \mathbf{R}_n = \mathbf{R}_n^H > \mathbf{0}, \end{array} \quad (7.17)$$

or, more compactly, as

$$\begin{array}{lll} \min\limits_{\mathbf{B}} & \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) & \\ \text{s.t.} & \max\limits_{\mathbf{R}_H}\left[\left(\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right]_{ii} \leq \rho_i, & 1 \leq i \leq L, \\ & \mathbf{R}_H = \hat{\mathbf{R}}_H + \mathbf{R}_{H,\Delta} & \mathbf{R}_{H,\Delta} : \ |\lambda_i\left(\mathbf{R}_{H,\Delta}\right)| \leq \epsilon_{R,n}, \\ & & \qquad \mathbf{R}_H = \mathbf{R}_H^H \geq \mathbf{0}. \end{array} \quad (7.18)$$

We can now use the upper bound on the MSE matrix of (7.7) to finally write the problem as

$$\begin{array}{ll} \min\limits_{\mathbf{B}} & \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \\ \text{s.t.} & \left[\left(\mathbf{I} + \mathbf{B}^H\check{\mathbf{R}}_H\mathbf{B}\right)^{-1}\right]_{ii} \leq \rho_i, \qquad 1 \leq i \leq L, \end{array} \quad (7.19)$$

where, again, the only difference with respect to the non-robust design of (6.10) is the use of $\breve{\mathbf{R}}_H$ instead of $\mathbf{R}_H$.

Imperfect CSIR and Imperfect CSIT

For the case of imperfect CSIR and CSIT, the worst-case QoS-constrained design can be formulated (similarly to (6.10)) as

$$\begin{array}{lll} \min\limits_{\mathbf{B},\mathbf{A}} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) & \\ \text{s.t.} & \max\limits_{\mathbf{H},\mathbf{R}_n} \left[\mathbf{A}^H\left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H+\mathbf{R}_n\right)\mathbf{A}+\mathbf{I}-\mathbf{A}^H\mathbf{H}\mathbf{B}-\mathbf{B}^H\mathbf{H}^H\mathbf{A}\right]_{ii} \leq \rho_i, & 1 \leq i \leq L, \\ & \mathbf{H} = \hat{\mathbf{H}} + \mathbf{H}_\Delta \qquad\qquad \mathbf{H}_\Delta : \|\mathbf{H}_\Delta\| \leq \epsilon_H, & \\ & \mathbf{R}_n = \hat{\mathbf{R}}_n + \mathbf{R}_{n,\Delta} \qquad \mathbf{R}_{n,\Delta} : \ \|\mathbf{R}_{n,\Delta}\| \leq \epsilon_n, & \\ & \qquad\qquad\qquad\qquad \mathbf{R}_n = \mathbf{R}_n^H > \mathbf{0}. & \end{array} \tag{7.20}$$

We can now use the upper bound on the MSE matrix of (7.8) and the problem reduces then to

$$\begin{array}{lll} \min\limits_{\mathbf{B},\mathbf{A}} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) & \\ \text{s.t.} & \left[\mathbf{A}^H\left(\hat{\mathbf{H}}\mathbf{B}\mathbf{B}^H\hat{\mathbf{H}}^H+\hat{\mathbf{R}}_n+\xi\mathbf{I}\right)\mathbf{A}+\mathbf{I}-\mathbf{A}^H\hat{\mathbf{H}}\mathbf{B}-\mathbf{B}^H\hat{\mathbf{H}}^H\mathbf{A}\right]_{ii} \leq \rho_i, & 1 \leq i \leq L, \end{array} \tag{7.21}$$

from which the optimum $\mathbf{A}$ is obtained as in (7.9). The problem can be finally written as

$$\begin{array}{lll} \min\limits_{\mathbf{B},\mathbf{A}} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) & \\ \text{s.t.} & \left[\left(\mathbf{I}+\mathbf{B}^H\hat{\mathbf{H}}^H(\hat{\mathbf{R}}_n+\xi\mathbf{I})^{-1}\hat{\mathbf{H}}\mathbf{B}\right)^{-1}\right]_{ii} \leq \rho_i, & 1 \leq i \leq L, \end{array} \tag{7.22}$$

where, again, the only difference with respect to the non-robust design of (6.10) is the use of $\hat{\mathbf{R}}_n + \xi\mathbf{I}$ instead of $\mathbf{R}_n$.

7.4 Stochastic Robust Design

An alternative to the pessimistic worst-case robust design is the stochastic approach that has a completely different philosophy. The underlying idea is to model the pdf of the actual parameter given the estimation and then consider the average performance over all possible parameters that could have caused the estimated value (see for example [Goe99, Jön02]).

We consider in this section the case of imperfect CSIT and perfect CSIR and also the more realistic case of imperfect CSIT and imperfect CSIR. The robust design problem is formulated in a rather general framework. However, simple results are obtained only for some particular cases.

7.4.1 Error Modeling and Problem Formulation

Instead of utilizing the instantaneous value of the MSE matrix $\mathbf{E}$, we have to consider averaged values of functions of $\mathbf{E}$ with respect to the conditional pdf $p_{\mathbf{H},\mathbf{R}_n|\hat{\mathbf{H}},\hat{\mathbf{R}}_n}(\mathbf{H},\mathbf{R}_n \mid \hat{\mathbf{H}},\hat{\mathbf{R}}_n)$, *i.e.*, the pdf of the actual channel and noise covariance matrix conditioned on the estimated values.

Assuming that the errors in the estimation of the channel and the noise covariance matrix are independent, the pdf can be factored as $p_{\mathbf{H}|\hat{\mathbf{H}}}(\mathbf{H} \mid \hat{\mathbf{H}})\, p_{\mathbf{R}_n|\hat{\mathbf{R}}_n}(\mathbf{R}_n \mid \hat{\mathbf{R}}_n)$. In terms of expected values, this reduces to $\mathbb{E}_{\mathbf{H},\mathbf{R}_n|\hat{\mathbf{H}},\hat{\mathbf{R}}_n}[\cdot] = \mathbb{E}_{\mathbf{R}_n|\hat{\mathbf{R}}_n}\mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}}[\cdot]$.

Channel Estimation Errors

First, we consider channel estimation errors. Given the actual channel $\mathbf{H}$, we model the estimated channel as

$$\hat{\mathbf{H}} = \mathbf{H} + \mathbf{H}_{\Delta} \tag{7.23}$$

where the actual channel and the estimation error (equivalently, the actual and estimated channels) are jointly Gaussian distributed [Jön02]. More specifically, defining $\mathbf{h} \triangleq \mathrm{vec}(\mathbf{H})$ and $\hat{\mathbf{h}} \triangleq \mathrm{vec}(\hat{\mathbf{H}})$, we have that

$$\begin{bmatrix} \mathbf{h} \\ \hat{\mathbf{h}} \end{bmatrix} \sim \mathcal{CN}\left(\begin{bmatrix} \mathbf{m}_h \\ \mathbf{m}_{\hat{h}} \end{bmatrix}, \begin{bmatrix} \mathbf{C}_{hh} & \mathbf{C}_{h\hat{h}} \\ \mathbf{C}_{\hat{h}h} & \mathbf{C}_{\hat{h}\hat{h}} \end{bmatrix} \right)$$

where $\mathbf{m}_h$ and $\mathbf{m}_{\hat{h}}$ are the means of $\mathbf{h}$ and $\hat{\mathbf{h}}$, respectively, $\mathbf{C}_{hh}$ and $\mathbf{C}_{\hat{h}\hat{h}}$ are the covariance matrices of $\mathbf{h}$ and $\hat{\mathbf{h}}$, respectively, and $\mathbf{C}_{h\hat{h}}$ is the cross-covariance matrix of $\mathbf{h}$ and $\hat{\mathbf{h}}$. It is well known [Kay93] that the distribution of the actual channel $\mathbf{h}$ conditioned on the estimated one $\hat{\mathbf{h}}$ is also Gaussian distributed $\mathbf{h} \mid \hat{\mathbf{h}} \sim \mathcal{CN}(\mathbf{m}_{h|\hat{h}}, \mathbf{C}_{h|\hat{h}})$, *i.e.*,

$$p_{h|\hat{h}}(\mathbf{h} \mid \hat{\mathbf{h}}) = \frac{1}{\pi^{(n_T n_R)}\left|\mathbf{C}_{h|\hat{h}}\right|} e^{-(\mathbf{h}-\mathbf{m}_{h|\hat{h}})^H \mathbf{C}_{h|\hat{h}}^{-1}(\mathbf{h}-\mathbf{m}_{h|\hat{h}})}$$

where $\mathbf{m}_{h|\hat{h}}$ and $\mathbf{C}_{h|\hat{h}}$ are the mean and covariance matrices of the actual channel $\mathbf{h}$ conditioned on the estimation $\hat{\mathbf{h}}$. Assuming that the estimation error $\mathbf{H}_{\Delta}$ is independent of the channel realization $\mathbf{H}$ ($\mathbf{C}_{hh_{\Delta}} = \mathbf{0}$), it follows that[5]

$$\begin{aligned} \mathbf{m}_{h|\hat{h}} &= \mathbf{m}_h + \mathbf{C}_{hh}\left(\mathbf{C}_{hh} + \mathbf{C}_{h_\Delta h_\Delta}\right)^{-1}\left(\hat{\mathbf{h}} - \mathbf{m}_{\hat{h}}\right) \\ \mathbf{C}_{h|\hat{h}} &= \left(\mathbf{C}_{hh}^{-1} + \mathbf{C}_{h_\Delta h_\Delta}^{-1}\right)^{-1}. \end{aligned}$$

If we further assume that the actual and estimated channels have zero means, $\mathbf{m}_h = \mathbf{0}$ and $\mathbf{m}_{\hat{h}} = \mathbf{0}$, and that $\mathbf{C}_{h_\Delta h_\Delta} \ll \mathbf{C}_{hh}$, we can approximate $\mathbf{m}_{h|\hat{h}}$ and $\mathbf{C}_{h|\hat{h}}$ as

$$\mathbf{m}_{h|\hat{h}} \simeq \hat{\mathbf{h}} \tag{7.24}$$

$$\mathbf{C}_{h|\hat{h}} \simeq \mathbf{C}_{h_\Delta h_\Delta}. \tag{7.25}$$

[5] Using $\mathbf{m}_{h|\hat{h}} = \mathbf{m}_h + \mathbf{C}_{h\hat{h}}\mathbf{C}_{\hat{h}\hat{h}}^{-1}(\hat{\mathbf{h}} - \mathbf{m}_{\hat{h}})$, $\mathbf{C}_{h|\hat{h}} = \mathbf{C}_{hh} - \mathbf{C}_{h\hat{h}}\mathbf{C}_{\hat{h}\hat{h}}^{-1}\mathbf{C}_{\hat{h}h}$, and applying the matrix inversion lemma (see §3.3) in the latter.

The expressions for the conditional mean and covariance matrix in (7.24)-(7.25) have the appealing intuitive interpretation that the actual channel $\mathbf{h}$ given the estimated channel $\hat{\mathbf{h}}$ is on average the estimated value $\hat{\mathbf{h}}$ plus a zero-mean Gaussian error with covariance matrix given by $\mathbf{C}_{h_\Delta h_\Delta}$. A consequence of (7.24)-(7.25) is that $\mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} f(\mathbf{H}) = \mathbb{E}_{\mathbf{H}_\Delta} f(\hat{\mathbf{H}} + \mathbf{H}_\Delta)$ for any function f.

At this point it would be possible to include some structure in the covariance matrix of the estimation error $\mathbf{C}_{h_\Delta h_\Delta}$ based on physical interpretations. This way, the error could be modeled more accurately and the system would be more robust to that particular error model. For simplicity, however, we assume an unstructured error and consider $\mathbf{C}_{h_\Delta h_\Delta} = \sigma_H^2 \mathbf{I}_{(n_T n_R)}$.

It is worth mentioning that another interesting model of the channel estimation errors arises when the channel singular vectors are perfectly known and the uncertainty is only with respect to the singular values [Bar01]. This situation typically happens in time-invariant frequency-selective channels with the utilization of the cyclic prefix (c.f. §2.2.1), because the singular vectors are complex exponentials.[6]

Noise Estimation Errors

The noise covariance matrix $\mathbf{R}_n = \mathbb{E}\left[\mathbf{n}\mathbf{n}^H\right]$ can be readily estimated using the sample covariance matrix (zero mean is assumed) from a collection of N samples of the noise as $\frac{1}{N}\sum_{k=1}^{N} \mathbf{n}_k \mathbf{n}_k^H$. Assuming that the number of samples is high compared to the observation dimension, *i.e.*, $N \gg n_R$, the strong law of the large numbers can be invoked to show that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} \mathbf{n}_k \mathbf{n}_k^H = \mathbf{R}_n$$

with probability one (simply note that each of the elements is a sum of random variables with the corresponding element of $\mathbf{R}_n$ as the mean and bounded variance). Therefore, for a sufficiently large N, the noise covariance matrix is well approximated by the sample mean:

$$\mathbf{R}_n \simeq \frac{1}{N}\sum_{k=1}^{N} \mathbf{n}_k \mathbf{n}_k^H.$$

To be exact, the sample covariance matrix (without the normalization factor) is a random matrix that follows a central complex Wishart distribution (assuming that the noise samples $\mathbf{n}_k$ are i.i.d.) [Mui82].

In practice, however, the receiver may only be able to obtain a corrupted version of the noise samples. For example, the receiver may estimate the channel using a training sequence and then subtract the component of the received signal corresponding to the training sequence to obtain an estimation of the noise samples. In such a case, we can model an estimated noise sample $\hat{\mathbf{n}}$ as

$$\hat{\mathbf{n}} = \mathbf{n} + \mathbf{n}_\Delta \tag{7.26}$$

[6]The same happens in time-varying flat channels, because the channel singular vectors are the canonical vectors (in such a case, the channel matrix is a diagonal matrix with diagonal elements given by the channel gain at each instant).

where $\mathbf{n}_\Delta$ is the estimation error. The estimated noise covariance matrix is then

$$\hat{\mathbf{R}}_n = \frac{1}{N}\sum_{k=1}^{N} \hat{\mathbf{n}}_k \hat{\mathbf{n}}_k^H.$$

Proceeding now in the same way as in the previous analysis of channel estimation errors, we assume that the noise realization $\mathbf{n}$ and the noise estimation $\hat{\mathbf{n}}$ are jointly Gaussian distributed. The actual noise $\mathbf{n}$ conditioned on the estimated noise $\hat{\mathbf{n}}$ follows a Gaussian distribution $\mathbf{n} \mid \hat{\mathbf{n}} \sim \mathcal{CN}\left(\mathbf{m}_{n|\hat{n}}, \mathbf{C}_{n|\hat{n}}\right)$ where $\mathbf{m}_{n|\hat{n}} \simeq \hat{\mathbf{n}}$ and $\mathbf{C}_{n|\hat{n}} \simeq \mathbf{C}_{n_\Delta n_\Delta}$. Assuming an unstructured noise error covariance matrix $\mathbf{C}_{n_\Delta n_\Delta} = \sigma_n^2 \mathbf{I}_{n_R}$, it is straightforward to obtain the following:

$$\begin{aligned}
\mathbb{E}_{n|\hat{n}}\left[\mathbf{R}_n\right] &\simeq \mathbb{E}_{n|\hat{n}}\left[\frac{1}{N}\sum_{k=1}^{N} \mathbf{n}_k \mathbf{n}_k^H\right] \\
&= \mathbb{E}_{n|\hat{n}}\left[\frac{1}{N}\sum_{k=1}^{N} \hat{\mathbf{n}}_k \hat{\mathbf{n}}_k^H\right] + \mathbb{E}_{n|\hat{n}}\left[\frac{1}{N}\sum_{k=1}^{N} \left(\hat{\mathbf{n}}_k - \mathbf{n}_k\right)\left(\hat{\mathbf{n}}_k - \mathbf{n}_k\right)^H\right] \\
&= \hat{\mathbf{R}}_n + \sigma_n^2 \mathbf{I}.
\end{aligned}$$

With some abuse of notation, we denote $\mathbb{E}_{n|\hat{n}}\left[\cdot\right]$ by $\mathbb{E}_{\mathbf{R}_n|\hat{\mathbf{R}}_n}\left[\cdot\right]$ and, then, we can finally write

$$\mathbb{E}_{\mathbf{R}_n|\hat{\mathbf{R}}_n}\left[\mathbf{R}_n\right] \simeq \hat{\mathbf{R}}_n + \sigma_n^2 \mathbf{I}. \tag{7.27}$$

Problem Formulation with Perfect CSIR and Imperfect CSIT

In the case of perfect CSIR, for each of the possible channels $\mathbf{H}$ that could have caused the estimation $\hat{\mathbf{H}}$, the receiver can use the optimum receive matrix given by the Wiener filter $\mathbf{A} = \left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H + \mathbf{R}_n\right)^{-1}\mathbf{H}\mathbf{B}$ (as obtained in §2.5.5) and the instantaneous MSE matrix is then $\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}$as in (2.51).

Problem Formulation with Imperfect CSIR and Imperfect CSIT

In the case of imperfect CSIR, the receiver cannot use the optimum receive matrix for each of the possible channels $\mathbf{H}$ that could have caused the estimation $\hat{\mathbf{H}}$ and the only possibility is to use a fixed receive matrix for each channel estimation $\hat{\mathbf{H}}$. In the same way that the optimal receive matrix was obtained in §2.5.5 such that each MSE (diagonal element of $\mathbf{E}$) was independently minimized, we take here the same approach but considering the minimization of the averaged MSE with respect to the pdf $p_{\mathbf{H},\mathbf{R}_n|\hat{\mathbf{H}},\hat{\mathbf{R}}_n}(\mathbf{H}, \mathbf{R}_n \mid \hat{\mathbf{H}}, \hat{\mathbf{R}}_n)$. We first state the following lemma.

Lemma 7.2 *Let* $\mathbf{h} \triangleq \text{vec}\left(\mathbf{H}\right)$ *be a Gaussian-distributed random vector with mean and covariance matrix conditioned on* $\hat{\mathbf{h}} \triangleq \text{vec}(\hat{\mathbf{H}})$ *given by* $\mathbf{m}_{h|\hat{h}} = \hat{\mathbf{h}}$ *and* $\mathbf{C}_{h|\hat{h}} = \sigma_H^2 \mathbf{I}$*, respectively. It then follows that*

$$\mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}}\left[\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H\right] = \hat{\mathbf{H}}\mathbf{B}\mathbf{B}^H\hat{\mathbf{H}}^H + \sigma_H^2 \,\text{Tr}\left(\mathbf{B}\mathbf{B}^H\right)\mathbf{I}.$$

Proof. See Appendix 7.B. ■

The conditional averaged MSE matrix is[7]

$$\bar{\mathbf{E}} = \mathbb{E}_{\mathbf{R}_n|\hat{\mathbf{R}}_n} \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} [\mathbf{E}] . \tag{7.28}$$

Consider first the expectation with respect to the channel matrix:

$$\begin{aligned} \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} [\mathbf{E}] &= \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \left[\mathbf{A}^H \left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H + \mathbf{R}_n\right) \mathbf{A} + \mathbf{I} - \mathbf{A}^H\mathbf{H}\mathbf{B} - \mathbf{B}^H\mathbf{H}^H\mathbf{A}\right] \\ &= \mathbf{A}^H \left(\hat{\mathbf{H}}\mathbf{B}\mathbf{B}^H\hat{\mathbf{H}}^H + \mathbf{R}_n + \sigma_H^2 \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \mathbf{I}\right) \mathbf{A} + \mathbf{I} - \mathbf{A}^H\hat{\mathbf{H}}\mathbf{B} - \mathbf{B}^H\hat{\mathbf{H}}^H\mathbf{A} \end{aligned} \tag{7.29}$$

where we have used $\mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} [\mathbf{H}] = \hat{\mathbf{H}}$ (from (7.24)) and Lemma 7.2. Including now the expectation with respect to the noise covariance matrix, we finally obtain

$$\bar{\mathbf{E}} = \mathbb{E}_{\mathbf{R}_n|\hat{\mathbf{R}}_n} \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} [\mathbf{E}] = \mathbf{A}^H(\hat{\mathbf{H}}\mathbf{B}\mathbf{B}^H\hat{\mathbf{H}}^H + \hat{\mathbf{R}}_n + \xi\mathbf{I})\mathbf{A} + \mathbf{I} - \mathbf{A}^H\hat{\mathbf{H}}\mathbf{B} - \mathbf{B}^H\hat{\mathbf{H}}^H\mathbf{A} \tag{7.30}$$

where $\xi = \sigma_H^2 \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) + \sigma_n^2$. Note that the scalar ξ depends on $\operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$. For power-constrained systems, we have that $\operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right) = P_T$ and, therefore, $\xi = \sigma_H^2 P_T + \sigma_n^2$. For QoS-constrained system, ξ depends on the objective value which still has to be found. To avoid such a dependence, $\operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$ can be approximated, for example, by its expected value or by its averaged value over the past realizations (recall that the values of σ_H^2 and σ_n^2 have to be approximated as well and, therefore, they can absorb any error in the *a priori* approximated value of $\operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$). As a consequence, for practical purposes, ξ can be taken as a constant value.

The optimum receive matrix that minimizes the diagonal elements of the averaged MSE matrix in (7.30) is easily obtained (similarly to the Wiener filter in §2.5.5) as

$$\mathbf{A}_\xi = \left(\hat{\mathbf{H}}\mathbf{B}\mathbf{B}^H\hat{\mathbf{H}}^H + \hat{\mathbf{R}}_n + \xi\mathbf{I}\right)^{-1}\hat{\mathbf{H}}\mathbf{B}. \tag{7.31}$$

Thus, the optimum robust receive matrix (7.31) is easily obtained in practice simply by adding a small scaled identity matrix to the noise covariance matrix. This result has been obtained elsewhere [Cox87, Car88, Vor03] and is in fact a well-known trick commonly referred to as "diagonal loading".

In the following subsections, the robust design is formulated for power-constrained and QoS-constrained systems (extending the results of Chapters 5 and 6, respectively) in a rather general way. Specific solutions, however, are only obtained for some particular cases. With imperfect CSIR, the receiver can only use the fixed receive matrix of (7.31) for the given estimated channel and noise covariance matrix. With perfect CSIR, however, the receive matrix is assumed to be optimally designed for each channel and noise covariance matrix that could have caused the estimated ones.

[7] Note that minimizing the conditional averaged MSE matrix as in (7.28) is tantamount to minimizing the joint average $\mathbb{E}_{\mathbf{H},\hat{\mathbf{H}},\mathbf{R}_n,\hat{\mathbf{R}}_n} [\mathbf{E}]$.

7.4.2 Power-Constrained Systems

The stochastic power-constrained design can be formulated (similarly to (5.9)) for the case of perfect CSIR and imperfect CSIT as

$$\begin{array}{ll} \min_{\mathbf{B}} & \mathbb{E}_{\mathbf{H},\mathbf{R}_n|\hat{\mathbf{H}},\hat{\mathbf{R}}_n} f_0\left(\left\{\left[\left(\mathbf{I}+\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}\right]_{ii}\right\}_{i=1}^{L}\right) \\ \text{s.t.} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \le P_T, \end{array} \tag{7.32}$$

and for the case of imperfect CSIR and imperfect CSIT as

$$\begin{array}{ll} \min_{\mathbf{B}} & \mathbb{E}_{\mathbf{H},\mathbf{R}_n|\hat{\mathbf{H}},\hat{\mathbf{R}}_n} f_0\left(\left\{\left[\mathbf{A}_\xi^H\left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H+\mathbf{R}_n\right)\mathbf{A}_\xi+\mathbf{I}-\mathbf{A}_\xi^H\mathbf{H}\mathbf{B}-\mathbf{B}^H\mathbf{H}^H\mathbf{A}_\xi\right]_{ii}\right\}_{i=1}^{L}\right) \\ \text{s.t.} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \le P_T. \end{array} \tag{7.33}$$

Example: Minimization of $\mathrm{Tr}\left(\mathbf{E}\right)$ *with Imperfect CSIR and CSIT*

Since both the trace and the expectation operators are linear, they can be interchanged without affecting the problem. Using the conditional expectation of the MSE matrix $\bar{\mathbf{E}}$ of (7.30) and plugging in the expression of the optimum robust receive matrix $\mathbf{A}_\xi$ of (7.31), the problem formulation reduces to

$$\begin{array}{ll} \min_{\mathbf{B}} & \mathrm{Tr}\left(\mathbf{I}+\mathbf{B}^H\hat{\mathbf{H}}^H\left(\hat{\mathbf{R}}_n+\xi\mathbf{I}\right)^{-1}\hat{\mathbf{H}}\mathbf{B}\right)^{-1} \\ \text{s.t.} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \le P_T, \end{array} \tag{7.34}$$

which falls exactly within the framework developed in Chapter 5 but using $\hat{\mathbf{H}}$ and $\hat{\mathbf{R}}_n+\xi\mathbf{I}$ instead of $\mathbf{H}$ and $\mathbf{R}_n$, respectively. Interestingly, to obtain the optimum robust transmit matrix $\mathbf{B}$, it suffices to add a small scaled identity matrix to the noise covariance matrix (*i.e.*, diagonal loading) as obtained elsewhere [Cox87, Car88, Vor03] (although a different result would be obtained if some structure was introduced in the estimation errors by means of $\mathbf{C}_{h_\Delta h_\Delta}$ and $\mathbf{C}_{n_\Delta n_\Delta}$).

7.4.3 QoS-Constrained Systems

The stochastic QoS-constrained design can be formulated (similarly to (6.10)) for the case of perfect CSIR and imperfect CSIT as

$$\begin{array}{ll} \min_{\mathbf{B}} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \\ \text{s.t.} & \mathbb{E}_{\mathbf{H},\mathbf{R}_n|\hat{\mathbf{H}},\hat{\mathbf{R}}_n} f_i\left(\left[\left(\mathbf{I}+\mathbf{B}^H\mathbf{H}^H\mathbf{R}_n^{-1}\mathbf{H}\mathbf{B}\right)^{-1}\right]_{ii}\right) \le f_i\left(\rho_i\right), \qquad 1 \le i \le L, \end{array} \tag{7.35}$$

and for the case of imperfect CSIR and imperfect CSIT as

$$\begin{array}{ll} \min_{\mathbf{B}} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \\ \text{s.t.} & \mathbb{E}_{\mathbf{H},\mathbf{R}_n|\hat{\mathbf{H}},\hat{\mathbf{R}}_n} f_i\left(\left[\mathbf{A}_\xi^H\left(\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H+\mathbf{R}_n\right)\mathbf{A}_\xi+\mathbf{I}-\mathbf{A}_\xi^H\mathbf{H}\mathbf{B}-\mathbf{B}^H\mathbf{H}^H\mathbf{A}_\xi\right]_{ii}\right) \le f_i\left(\rho_i\right), \\ & \qquad 1 \le i \le L. \end{array} \tag{7.36}$$

Example: MSE-QoS Constraints with Imperfect CSIR and CSIT

MSE-QoS constraints translate into the functions $f_i(x) = x$. Since the f_i's and the expectation operator are linear, they can be interchanged without affecting the problem. Using the conditional expectation of the MSE matrix $\bar{\mathbf{E}}$ of (7.30) and plugging in the expression of the optimum robust receive matrix $\mathbf{A}_\xi$ of (7.31), the problem formulation reduces to

$$\begin{array}{ll} \min_{\mathbf{B}} & \mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right) \\ \text{s.t.} & \left[\left(\mathbf{I} + \mathbf{B}^H\hat{\mathbf{H}}^H(\hat{\mathbf{R}}_n + \xi\mathbf{I})^{-1}\hat{\mathbf{H}}\mathbf{B}\right)^{-1}\right]_{ii} \leq \rho_i, \qquad 1 \leq i \leq L, \end{array} \tag{7.37}$$

which fits the case considered in Chapter 6 but using $\hat{\mathbf{H}}$ and $\hat{\mathbf{R}}_n + \xi\mathbf{I}$ instead of $\mathbf{H}$ and $\mathbf{R}_n$. As happened in the minimization of $\mathrm{Tr}(\mathbf{E})$ with imperfect CSIR and CSIT, to obtain the optimum robust transmit matrix $\mathbf{B}$, it suffices to add a small scaled identity matrix to the noise covariance matrix (*i.e.*, diagonal loading) as obtained elsewhere [Cox87, Car88, Vor03] (recall that this result has been obtained assuming unstructured estimation errors, *i.e.*, $\mathbf{C}_{h_\Delta h_\Delta} = \mathbf{I}$ and $\mathbf{C}_{n_\Delta n_\Delta} = \mathbf{I}$).

7.5 Simulation Results

Illustrative Example

In this illustrative example, we compare in a qualitative way the structural difference between the simple robust solutions based on decreasing the channel eigenvalues and diagonal loading of the noise covariance matrix. The former is obtained in (7.5) for the worst-case design with perfect CSIR and imperfect CSIT. The latter is obtained in (7.8) for the worst-case design with imperfect CSIR and CSIT and in (7.30) for the stochastic design with imperfect CSIR and CSIT. For simplicity, we only consider channel estimation errors and the interference-plus-noise covariance matrix is assumed perfectly known.

In Figure 7.3, the response of different transmit beamvectors is depicted for a simple scenario with 6 transmit and 6 receive antennas with a QoS constraint. In particular, we consider a phased array design (which does not take into account the interference structure), a naive design (which assumes the estimated channel as perfect), a worst-case robust design based on decreasing the channel eigenvalues (c.f. §7.3), and a stochastic robust design based on diagonal loading (c.f. §7.4). The channel has two propagation rays with departure angles of 0° and 60° and arrival angles of 0° and 60° (with gains of 0dB and -0.9dB, respectively). The noise covariance matrix contains a strong directive interference signal with an arrival angle of 5° plus white noise (with an interference to noise ratio of 26dB).

As can be observed from Figure 7.3, the phased array design ignores the directional interference and points towards the two transmit directions of the channel rays at 0° and 60°. The naive design

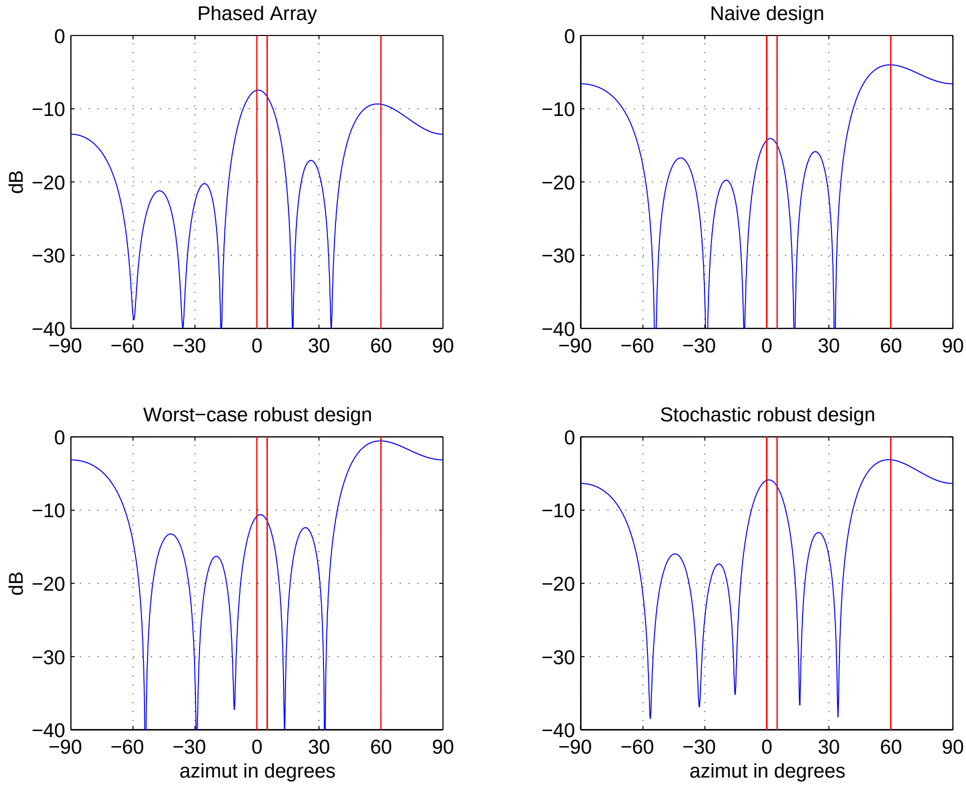


Figure 7.3: Qualitative comparison of the transmit beamformers for a 6×6 MIMO channel composed of two rays with departure angles of $0°$ and $60°$ and arrival angles of $0°$ and $60°$ (with a gain of 0dB and -0.9dB respectively) and a strong directive interference signal with an arrival angle of $5°$ plus white noise (with an interference to noise ratio of 26dB).

considers the directional interference and, consequently, weights less the transmit ray at $0°$ since it arrives very close to the directional interference. The worst-case robust design (based on decreasing the channel eigenvalues) has the same shape as the naive design but with an increased gain (*i.e.*, an increase in transmit power) to account for possible errors. The stochastic robust design (based on diagonal loading) not only has an increased gain with respect to the naive design but also has a different shape which is between that of the phased array and that of the naive design (this is because the diagonal loading has the effect of whitening the directional interference signals). Recall that, on average, the stochastic robust design is the best thing to do.

Realistic Simulations

For the realistic numerical simulations, we consider the same wireless scenario used in Chapters 5 and 6, which we briefly review for completeness.

We model a wireless communication system with multiple antennas at both sides of the link (in particular 4 transmit and 4 receive antennas). The MIMO channels were generated using the parameters of the European standard HIPERLAN/2 [ETS01] for wireless local area networks (WLAN), which is based on the multicarrier modulation OFDM (64 carriers are used in the simulations). The channel model includes the frequency-selectivity and the spatial correlation as measured in real scenarios. In particular, the frequency selectivity of the channel is modeled using the power delay profile type C for HIPERLAN/2 as specified in [ETS98a] and the spatial correlation of the MIMO channel is modeled according to the *Nokia* model defined in [Sch01]. The matrix channel generated was normalized so that $\sum_n \mathbb{E}\left[\left|\,[\mathbf{H}(n)]_{ij}\right|^2\right] = 1$. The SNR is defined as the transmitted power normalized with the noise spectral density N_0.

The estimated channel matrix contains a random estimation error drawn according to a Gaussian pdf with zero mean and uncorrelated elements with variance $\sigma_H^2 = 0.01$ (note that the real channel matrix at each carrier follows a Gaussian pdf with zero-mean elements with unit variance). The noise covariance matrix is assumed white and perfectly known.

For the worst-case robust design, the upper bound on the norm of the channel estimation error (see (7.1)) should be chosen in a real system such that it is satisfied with high probability; otherwise an outage event is declared.[8] For an outage probability of 5%, the upper bound on the norm should be chosen as $\epsilon_H \simeq \sqrt{14.6645 \times \sigma_H^2}$ for the case of 4 transmit and 4 receive antennas. In practice, however, this value results unnecessarily high due to the pessimism inherent in all bounds used in §7.3 to derive the worst-case design. Therefore, to overcome such excessive pessimism, it is necessary to include in ϵ_H an additional factor α_H which has to be found by numerical simulations (its value is in general on the order of $10^{-3} \sim 10^{-2}$).

Whenever the methods are evaluated for several independent realizations of the channel matrix, the results are given in terms of outage values as in Chapters 5 and 6 (c.f. §2.5.4). To be more specific, outage values of the BER refer to the BER that is guaranteed only for $(1 - P_{\text{out}})$ of the time and outage values of the SNR refer to the SNR required in $(1 - P_{\text{out}})$ of the time (*i.e.*, the SNR that will not suffice with probability P_{out}).

Worst-Case Robust Design with Perfect CSIR and Imperfect CSIT

In this example, we design a robust system with minimum transmit power subject to a set of different QoS constraints in terms of BER (a carrier-noncooperative approach is taken to deal with the multicarrier channel model). We make the approximation that perfect CSIR is available and design a worst-case robust solution to cope with imperfect CSIT as obtained in §7.3.3. To be specific, we consider the problem formulation of (7.19) in which a lower bound of $\mathbf{R}_H$ (obtained by decreasing its eigenvalues) is utilized.

In Figure 7.4, the required SNR and the obtained BER are plotted as a function of the desired BER given as a QoS constraint for three different cases: exact design with perfect CSIT, naive design with imperfect CSIT, and worst-case robust design against imperfect CSIT. As previously

[8] The worst-case performance is measured only when an outage event is not declared.

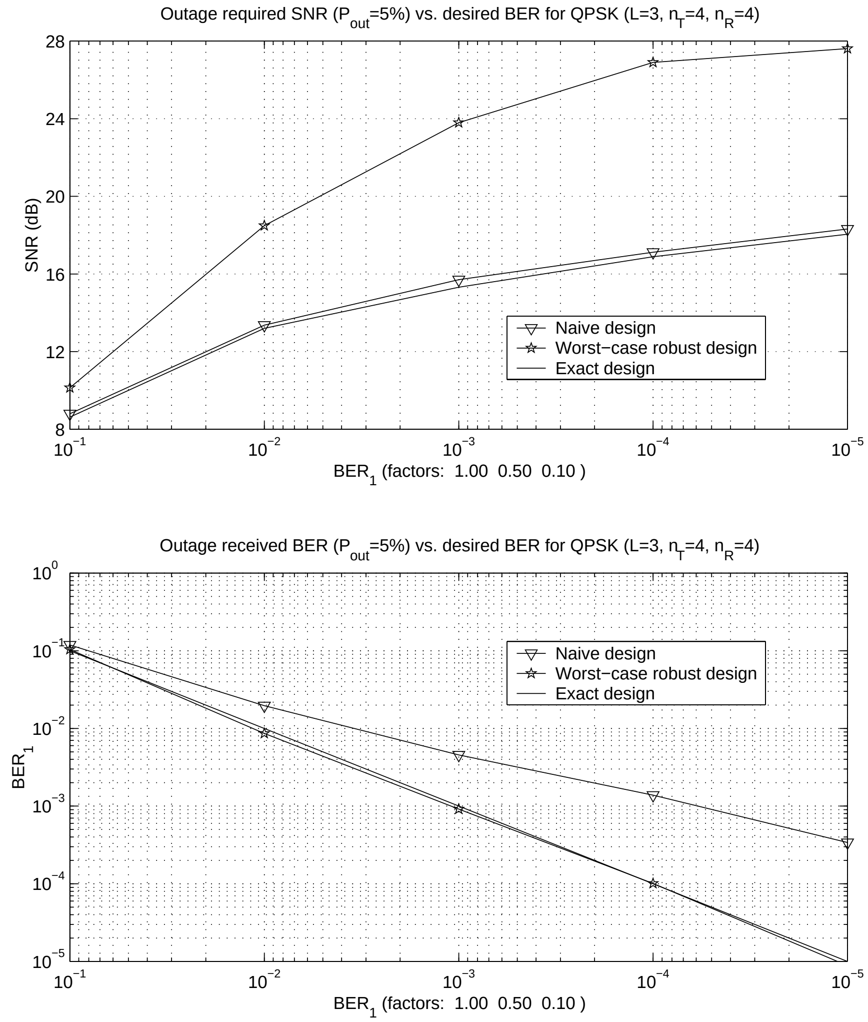
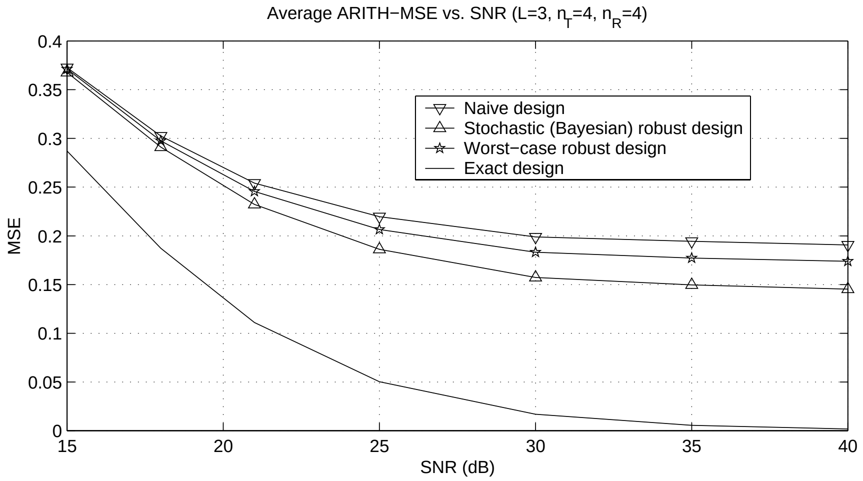


Figure 7.4: Outage required SNR and obtained BER vs. the required BER for the (carrier-noncooperative) exact, naive, and worst-case ($\alpha_H = 10^{-3}$) robust methods (with perfect CSIR and imperfect CSIT) when using QPSK in a multicarrier 4×4 MIMO channel with $L = 3$. (The BER of the first substream is along the x-axis and the BER of the second and third substreams are given by scaling with the factors 0.5 and 0.1.)

mentioned in §7.2, the worst-case robust design requires a significant increase of power with respect to the ideal design with perfect CSIT. The performance of the robust design behaves as predicted by theory, *i.e.*, the worst-case obtained BER (excluding outage events) satisfies the required BER. The performance of the naive design is completely unacceptable since it does not satisfy the required BER by about one order of magnitude.

Worst-Case and Stochastic Robust Designs with Imperfect CSIR and CSIT

We now consider the more realistic case of imperfect CSIR and CSIT to design a robust communication system (a carrier-noncooperative approach is taken to deal with the multicarrier channel model). For clarity of presentation, we plot simulation results for a single realization of the channel matrix $\mathbf{H}$ and many realizations of the channel estimation error $\mathbf{H}_{\Delta}$. If we also



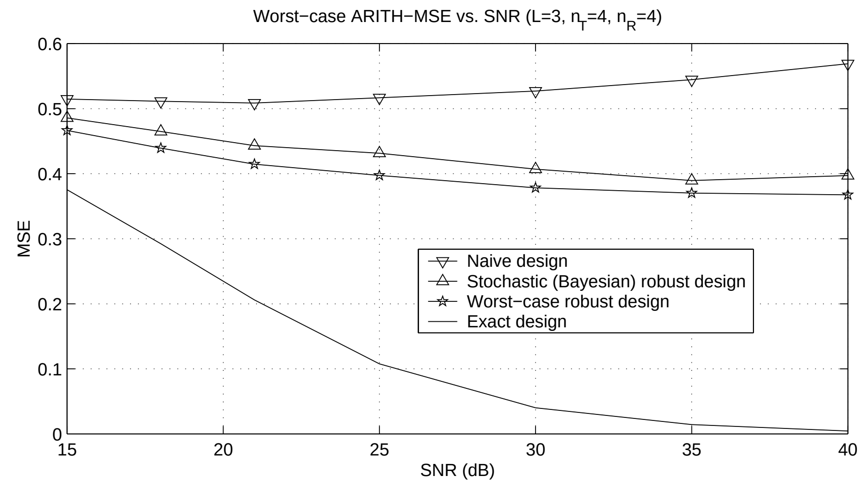


Figure 7.5: Average and worst-case obtained ARITH-MSE vs. the desired MSE for the (carrier-noncooperative) exact, naive, worst-case robust ($\alpha_H = 10^{-2}$), and stochastic robust methods (with imperfect CSIR and CSIT) in a multicarrier 4×4 MIMO channel with $L = 3$.

consider many realizations of $\mathbf{H}$ and plot the results in terms of outage values, very similar curves are obtained. Note that, for both the worst-case and the stochastic robust designs, the solution is obtained by diagonal loading; the difference lies on the determination of the scaling factor ξ.[9]

We first consider a power-constrained system in which the objective function to minimize is the MSE averaged over the carriers and substreams (ARITH-MSE method in §5.5.1). We derive worst-case and stochastic robust solutions according to the problem formulation in (7.16) and (7.34), respectively.

In Figure 7.5, the average and outage values of the ARITH-MSE are plotted as a function

[9]In both the worst-case and stochastic robust solutions, the value of ξ depends on $\mathbf{B}$ and $\mathbf{A}$ (in particular, it depends on $\sigma_{\max}(\mathbf{B})$ and $\sigma_{\min}(\mathbf{A})$ in the worst-case design and on $\mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right)$ in the stochastic design). This dependence can be avoided in practice by approximating ξ using the solutions $\mathbf{B}$ and $\mathbf{A}$ obtained in the previous block.

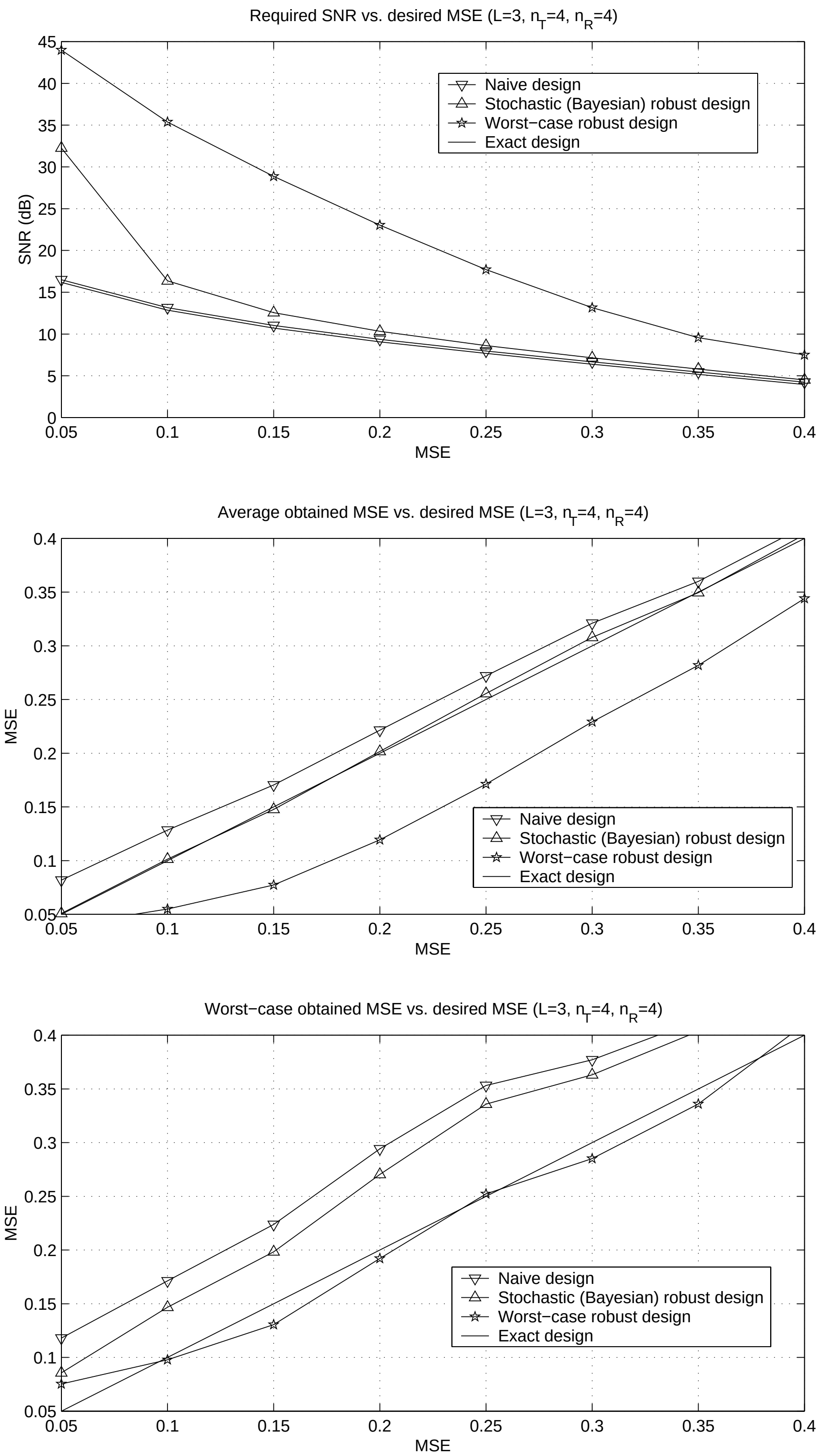


Figure 7.6: Required SNR, average obtained MSE, and worst-case obtained MSE vs. the desired MSE for the (carrier-noncooperative) exact, naive, worst-case robust ($\alpha_H = 10^{-2}$), and stochastic robust methods (with imperfect CSIR and CSIT) in a multicarrier 4×4 MIMO channel with $L = 3$.

of the SNR for four different cases: exact design, naive design, worst-case robust design, and stochastic robust design. We can observe that the stochastic robust design has the best average performance, as expected, among the three methods with imperfect CSIT and CSIR. The best worst-case performance, as predicted by theory, is obtained by the worst-case robust design. However, the worst-case robust approach requires the determination of the additional factor α_H for a fine tuning of the scaling factor ξ used in the diagonal loading. If ξ is not well chosen, the performance degrades significantly. Hence, the worst-case approach is not recommended in practice.

It is also worth pointing out that all methods with imperfect CSI suffer from an MSE floor as can be observed in Figure 7.5, unlike the exact design with perfect CSI whose performance can be as good as desired by increasing the transmit power. For the stochastic robust design, it is straightforward to obtain the value of the average MSE floor as $\bar{\mathbf{E}} \to \bar{\mathbf{E}}_{\text{fl}} \triangleq \left(\mathbf{I}+\frac{1}{\sigma_H^2}\bar{\mathbf{B}}^H\hat{\mathbf{H}}^H\hat{\mathbf{H}}\bar{\mathbf{B}}\right)^{-1}$ as $P_T \to \infty$, where $\bar{\mathbf{B}}$ is the normalized transmit matrix defined as $\bar{\mathbf{B}} = \mathbf{B}/\sqrt{\text{Tr}\left(\mathbf{B}\mathbf{B}^H\right)}$. The optimal $\bar{\mathbf{B}}$ is obtained as the solution to the minimization of $\text{Tr}\left(\bar{\mathbf{E}}_{\text{fl}}\right)$ subject to $\text{Tr}\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^H\right) \leq 1$.

We then consider a QoS-constrained system with minimum transmit power subject to a set of QoS requirements in terms of MSE. We derive worst-case and stochastic robust solutions according to the problem formulation in (7.22) and (7.37), respectively.

In Figure 7.6, the required SNR, the obtained average MSE, and the obtained worst-case MSE are plotted as a function of the desired MSE for four different cases: exact design, naive design, worst-case robust design, and stochastic robust design. We can observe that the stochastic robust design needs, in general, less than 3dB more transmit power than the exact design with perfect CSI. The worst-case robust design, however, requires significantly much more power. The MSE achieved by the stochastic robust design is, on average, the desired MSE as expected from the analytical derivations; however, nothing can be guaranteed for the worst-case MSE. As predicted by theory, the worst-case MSE is satisfied by the worst-case robust design (provided that the scaling factor ξ used in the diagonal loading is properly chosen by numerical simulations). The worst-case approach, however, is not recommended for practical purposes due to its sensitivity to the exact value of ξ and to the excessive increase of required transmit power.

7.6 Chapter Summary and Conclusions

This chapter has considered the omnipresent problem of channel estimation errors in practical communication systems. The problem formulation of Chapters 5 and 6, where perfect CSI was assumed, has been extended to account for imperfect CSI, obtaining robust designs. Two different philosophies have been considered: the worst-case and the stochastic (Bayesian) robust designs. In particular, a robust formulation of the joint transmit-receive design for power-constrained and QoS-constrained systems has been given under both philosophies. For the worst-case robust approach, the problem has been easily formulated by using some bounds on the channel and the noise covariance matrices. For the stochastic robust design, the general problem has been formulated in a rather general framework, although simple results have been obtained only for some particular cases such as when the trace of the MSE matrix is the objective to minimize or when QoS constraints in terms of MSE are imposed. In both cases, the stochastic robust design translates into the well-known trick of adding a small scaled identity matrix to the estimated noise covariance matrix (although a different result would be obtained if a structured estimation error was considered).

The analytical derivations have been checked by numerical simulations using realistic channel models. The robust designs manage to perform well under channel estimation errors due to their robustness (although they also show a limited performance). For power-constrained systems, the robust solutions perform better than the naive solution with the same transmit power. For QoS-constrained systems, the robust solutions satisfy the QoS requirements at the expense of an increased transmit power with respect to the exact design with perfect CSI, which is the price to be paid for robustness. In general terms, the worst-case philosophy has been found impractical due to its sensitivity with respect to exact choice of its parameters and, therefore, the stochastic philosophy is considered as the most appropriate solution in practice.

The contribution of this chapter has been the application of the well-known worst-case and stochastic philosophies to obtain robust designs within the framework used in Chapters 5 and 6.

Appendix 7.A Proof of Lemma 7.1

The inequality $\mathbf{X}\mathbf{X}^H \leq \lambda_{\max}\left(\mathbf{X}\mathbf{X}^H\right)\mathbf{I}_n$ is straightforward from the definition of maximum eigenvalue as we now show. Consider that $\mathbf{c}^H\mathbf{X}\mathbf{X}^H\mathbf{c} > \lambda_{\max}\left(\mathbf{X}\mathbf{X}^H\right)\mathbf{c}^H\mathbf{c}$ for some $\mathbf{c}$. This would imply that $\lambda_{\max}\left(\mathbf{X}\mathbf{X}^H\right) < \mathbf{c}^H\mathbf{X}\mathbf{X}^H\mathbf{c}/\mathbf{c}^H\mathbf{c}$ which cannot be since the maximum eigenvalue is defined as $\lambda_{\max}\left(\mathbf{X}\mathbf{X}^H\right) = \max_{\mathbf{c}} \mathbf{c}^H\mathbf{X}\mathbf{X}^H\mathbf{c}/\mathbf{c}^H\mathbf{c}$.

The inequality $\mathbf{X}\mathbf{Y}^H + \mathbf{Y}\mathbf{X}^H \leq 2\sigma_{\max}\left(\mathbf{X}\right)\sigma_{\max}\left(\mathbf{Y}\right)\mathbf{I}_n$ can be proved by showing that

$$\mathbf{c}^H\left(\mathbf{X}\mathbf{Y}^H + \mathbf{Y}\mathbf{X}^H\right)\mathbf{c} \leq 2\sigma_{\max}\left(\mathbf{X}\right)\sigma_{\max}\left(\mathbf{Y}\right)\mathbf{c}^H\mathbf{c} \qquad \forall\mathbf{c}.$$

We can now establish the following chain of inequalities:

$$\begin{aligned}\mathbf{c}^H\left(\mathbf{X}\mathbf{Y}^H+\mathbf{Y}\mathbf{X}^H\right)\mathbf{c} &= 2\,\mathrm{Re}\left[\mathbf{c}^H\mathbf{X}\mathbf{Y}^H\mathbf{c}\right]\\ &\leq 2\left|\mathbf{c}^H\mathbf{X}\mathbf{Y}^H\mathbf{c}\right|\\ &\leq 2\left\|\mathbf{X}^H\mathbf{c}\right\|_2\left\|\mathbf{Y}^H\mathbf{c}\right\|_2\\ &= 2\sqrt{\mathbf{c}^H\mathbf{X}\mathbf{X}^H\mathbf{c}}\sqrt{\mathbf{c}^H\mathbf{Y}\mathbf{Y}^H\mathbf{c}}\\ &\leq 2\,\lambda_{\max}^{1/2}\left(\mathbf{X}\mathbf{X}^H\right)\lambda_{\max}^{1/2}\left(\mathbf{Y}\mathbf{Y}^H\right)\,\mathbf{c}^H\mathbf{c}\\ &= 2\,\sigma_{\max}\left(\mathbf{X}\right)\sigma_{\max}\left(\mathbf{Y}\right)\,\mathbf{c}^H\mathbf{c}\end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz's inequality (see §3.3) and the third inequality from $\mathbf{X}\mathbf{X}^H \leq \lambda_{\max}\left(\mathbf{X}\mathbf{X}^H\right)\mathbf{I}_n$ and $\mathbf{Y}\mathbf{Y}^H \leq \lambda_{\max}\left(\mathbf{Y}\mathbf{Y}^H\right)\mathbf{I}_n$. ∎

Appendix 7.B Proof of Lemma 7.2

We first obtain a lemma and then proceed to prove Lemma 7.2.

Lemma 7.3 *Let $\mathbf{H}\in\mathbb{C}^{n\times m}$ be a random matrix with i.i.d. proper Gaussian-distributed entries with zero mean and variance σ^2, i.e., $[\mathbf{H}]_{ij}\sim\mathcal{CN}\left(0,\sigma^2\right)$. Then, for any fixed unitary matrix $\mathbf{U}\in\mathbb{C}^{m\times m}$, the distribution of $\mathbf{HU}$ is the same as the distribution of $\mathbf{H}$, i.e., $\mathbf{HU}\sim\mathbf{H}$.*

Proof. Let $\mathbf{G}=\mathbf{HU}$. Since the rows of $\mathbf{H}$ are independent, the rows of $\mathbf{G}$ are also independent. It remains to check that each row of $\mathbf{G}$ has the same distribution as that of $\mathbf{H}$. Since the rows of $\mathbf{H}$ are proper complex Gaussian vectors, so are those of $\mathbf{G}$. Therefore, only the first and second moments have to be compared. Let $\mathbf{g}_i^H$ and $\mathbf{h}_i^H$ be the ith row of $\mathbf{G}$ and $\mathbf{H}$, respectively ($\mathbf{g}_i^H=\mathbf{h}_i^H\mathbf{U}$). It is clear that the first moment is equal: $\mathbb{E}\left[\mathbf{g}_i\right]=\mathbf{U}^H\mathbb{E}\left[\mathbf{h}_i\right]=\mathbf{0}$. The second moment follows easily as well: $\mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i^H\right]=\mathbf{U}^H\mathbb{E}\left[\mathbf{h}_i\mathbf{h}_i^H\right]\mathbf{U}=\mathbf{U}^H\mathbf{U}=\mathbf{I}$. ∎

Proof of Lemma 7.2. First of all, note that

$$\begin{aligned}\mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}}\left[\mathbf{H}\mathbf{B}\mathbf{B}^H\mathbf{H}^H\right] &= \mathbb{E}_{\mathbf{H}_\Delta}[(\hat{\mathbf{H}}+\mathbf{H}_\Delta)\mathbf{B}\mathbf{B}^H(\hat{\mathbf{H}}+\mathbf{H}_\Delta)^H]\\ &= \hat{\mathbf{H}}\mathbf{B}\mathbf{B}^H\hat{\mathbf{H}}^H+\mathbb{E}_{\mathbf{H}_\Delta}\left[\mathbf{H}_\Delta\mathbf{B}\mathbf{B}^H\mathbf{H}_\Delta^H\right]\end{aligned}$$

where $\mathbf{h}_\Delta\triangleq\mathrm{vec}\left(\mathbf{H}_\Delta\right)$ has zero mean $\mathbb{E}\left[\mathbf{h}_\Delta\right]=\mathbf{0}$ and covariance matrix $\mathbb{E}\left[\mathbf{h}_\Delta\mathbf{h}_\Delta^H\right]=\sigma_H^2\,\mathbf{I}$. It remains to show that $\mathbb{E}_{\mathbf{H}_\Delta}\left[\mathbf{H}_\Delta\mathbf{B}\mathbf{B}^H\mathbf{H}_\Delta^H\right]=\sigma_H^2\,\mathrm{Tr}\left(\mathbf{B}\mathbf{B}^H\right)\mathbf{I}$.

From Lemma 7.3 and using the eigendecomposition $\mathbf{B}\mathbf{B}^H = \mathbf{U}\mathbf{D}_B\mathbf{U}^H$, it follows that $\mathbb{E}_{\mathbf{H}_\Delta}\left[\mathbf{H}_\Delta\mathbf{B}\mathbf{B}^H\mathbf{H}_\Delta^H\right] = \mathbb{E}_{\mathbf{H}_\Delta}\left[\mathbf{H}_\Delta\mathbf{D}_B\mathbf{H}_\Delta^H\right]$. We then have that

$$\begin{aligned}
\left[\mathbb{E}_{\mathbf{H}_\Delta}\left[\mathbf{H}_\Delta\mathbf{D}_B\mathbf{H}_\Delta^H\right]\right]_{ij} &= \sum_k \lambda_{B,k}\,\mathbb{E}_{\mathbf{H}_\Delta}\left[[\mathbf{H}_\Delta]_{i,k}\left[\mathbf{H}_\Delta^H\right]_{k,j}\right] = 0 \qquad \text{for } i \neq j \\
\left[\mathbb{E}_{\mathbf{H}_\Delta}\left[\mathbf{H}_\Delta\mathbf{D}_B\mathbf{H}_\Delta^H\right]\right]_{ii} &= \sum_k \lambda_{B,k}\,\mathbb{E}_{\mathbf{H}_\Delta}\left[\left|[\mathbf{H}_\Delta]_{i,k}\right|^2\right] \\
&= \sigma_H^2 \sum_k \lambda_{B,k} \\
&= \sigma_H^2 \operatorname{Tr}\left(\mathbf{B}\mathbf{B}^H\right).
\end{aligned}$$

■

Chapter 8

Conclusions and Future Work

THIS DISSERTATION HAS CONSIDERED communications through MIMO channels which includes many specific scenarios such as wireless multi-antenna systems and wireline DSL systems. Both information-theoretic and practical communications aspects of MIMO channels have been analyzed. For the former, the fundamental limits of MIMO channels with different degrees of CSI have been studied; specifically, for the case of no CSI, a game-theoretic approach was adopted to obtain robust solutions under channel uncertainty. For the latter, an optimal joint design of the transmitter and receiver has been analyzed in detail. First, power-constrained systems were considered and a unified framework was developed to obtain the optimal transmit-receive structure for a wide family of design criteria. Then, QoS-constrained systems were studied and the optimal design with minimum transmitted power was obtained. These results, which were initially obtained for the case of perfect CSI, were later extended to account for imperfect CSI due, for example, to channel estimation errors.

8.1 Conclusions

After giving the motivation of the dissertation in Chapter 1, an overview of MIMO channels has been given in Chapter 2 and two important theories on which many results of this dissertation are based, namely, convex optimization theory and majorization theory, have been briefly described in Chapter 3.

Chapter 4 has focused on the information-theoretic analysis of MIMO channels with different degrees of CSI for the single-user and mutiuser (multiple-access channel) cases. After reviewing the well-known cases of instantaneous and statistic CSI (beamforming-constrained systems were specifically addressed), the case in which not even the channel statistics are known was investigated in great detail to obtain robust solutions under channel uncertainty. To be specific, the problem was formulated within the framework of game theory in which the payoff function of the game is the mutual information and the players are the transmitter and a malicious nature. This problem characterizes the capacity of the compound vector Gaussian channel. The uniform power allocation was obtained as a robust solution to the game in terms of capacity for the class of isotropically unconstrained channels (unconstrained transmit "directions"). Interestingly, for the multiple-access channel, the uniform power allocation for each of the users also constitutes a robust solution (the worst-case rate region corresponding to the uniform power distribution was shown to contain the worst-case rate region of any other possible power allocation strategy). In other words, the capacity region of the compound vector Gaussian MAC is achieved when each of the users is using a uniform power allocation.

Chapter 5 has formulated and solved the joint design of transmit and receive linear processing (also termed multiple beamvectors or beam-matrices) for MIMO channels (only point-to-point communications were considered) under a variety of design criteria subject to a power constraint. Instead of considering each design criterion separately, a unifying framework that generalizes the existing results in the literature was developed by considering two families of objective functions that embrace most reasonable criteria to design a communication system: Schur-concave and Schur-convex functions. Under such a framework, the optimal structure of the transmitter and receiver is readily obtained and the problem can be then formulated within the powerful framework of convex optimization theory, in which solutions can be efficiently obtained even though closed-form expressions may not exist. From this perspective, different design criteria were analyzed and, in particular, a closed-form optimal solution in terms of minimum BER was obtained. Efficient algorithms for practical implementation were given for the considered design criteria.

Chapter 6 has formulated and solved the joint design of transmit and receive linear processing for MIMO channels (only point-to-point communications were considered) in terms of minimum transmitted power subject to QoS constraints for each of the established substreams. Although the original problem formulation was a complicated nonconvex problem with matrix-valued variables, by using majorization theory, the problem was reformulated as a simple convex optimization problem with scalar variables. To solve optimally the convex problem in practice, a practical and efficient multi-level water-filling algorithm was obtained.

Chapter 7 has extended the results of Chapters 5 and 6, in which perfect CSI was assumed, to the more realistic situation of imperfect CSI due to channel estimation errors. Two different philosophies were used to obtain robust designs: worst-case robustness and stochastic (Bayesian) robustness.

Summarizing, the main results obtained in this dissertation are:

- the game-theoretic formulation of the communication process, obtaining the uniform power allocation as a robust solution under channel uncertainty,
- the unifying framework for the joint transmit-receive design that generalizes the existing re-

sults in the literature by considering two families of functions that embrace most reasonable design criteria (and, in particular, the closed-form solution in terms of minimum BER), and

- the optimum solution in terms of minimum transmitted power subject to a set of QoS constraints.

8.2 Future Work

As a result of the work undertaken during the elaboration of the present dissertation, several lines for future research have been singled out to extend the obtained results as we now mention.

Regarding the capacity results of Chapter 4, some open problems are:

- Obtaining an iterative method (similar to the iterative water-filling algorithm [Yu01b]) to solve the maximization of the sum rate of the beamforming-constrained MAC of (4.17).
- Obtaining a simple iterative method to maximize an arbitrary weighted sum of the rates (generalizing the iterative water-filling algorithm that maximizes the sum rate of (4.16) as obtained in [Yu01b]).
- Characterizing the optimal transmit covariance matrix in terms of outage capacity for the simple case of a Gaussian-distributed channel matrix with i.i.d. elements.

With respect to the joint transmit-receive design for MIMO channels, some interesting problems are:

- Extending the ARITH-BER method of §5.5.11 to include different constellations.
- Generalizing the unified framework of Theorem 5.1 to include a DF scheme and/or a peak-power constraint.
- Performing an analytic characterization of the performance of the different criteria of Chapter 5 (rather than using numerical simulations as a means of assessment and comparison).
- Further investigation along the line of Chapter 7 to obtain robust designs; in particular, an analytical characterization of the performance under channel estimation errors.
- Extension of the obtained results in the joint transmit-receive design to the multiuser case. The downlink of a multiuser wireless system (broadcast channel) not only requires the design of the transmit and receive matrices for all the users but also has to allocate the total available power among the users. This is currently an open problem in the literature since it is a highly nonconvex and complicated problem. The uplink of a multiuser wireless system

(multiple-access channel) may be solved in a centralized way as the downlink. It seems more natural, however, to consider distributed solutions that do not require a significant user cooperation. Such an approach fits naturally within the framework of game theory in which each user could be considered as a player in a game with multiple players. In any case, both centralized and distributed problem formulations are currently open problems.

Bibliography

[3GP99] 3GPP, "Physical layer procedures (TDD)," *3G TS 25.224 (v3.1.1), TSGR, http://www.3gpp.org*, Dec. 1999.

[AD96] N. Al-Dhahir and J. M. Cioffi, "Block transmission over dispersive channels: Transmit filter optimization and realization, and MMSE-DFE receiver performance," *IEEE Trans. on Information Theory*, vol. 42, no. 1, pp. 137–160, Jan. 1996.

[AD00] N. Al-Dhahir and A. H. Sayed, "The finite-length multi-input multi-output MMSE-DFE," *IEEE Trans. on Signal Processing*, vol. 48, no. 10, pp. 2921–2936, Oct. 2000.

[Aka98] A. N. Akansu, P. Duhamel, X. Lin, and M. de Courville, "Orthogonal transmultiplexers in communication: A review," *IEEE Trans. on Signal Processing*, vol. 46, no. 4, pp. 979–995, April 1998.

[Ala98] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[Ami84] N. Amitay and J. Salz, "Linear equalization theory in digital data transmission over dually polarized fading radio channels," *At&T Bell Labs. Technical Journal*, vol. 63, no. 10, pp. 2215–2259, Dec. 1984.

[And00] J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 11, pp. 2172–2178, Nov. 2000.

[Ari99] S. L. Ariyavisitakul, J. H. Winters, and I. Lee, "Optimum space-time processors with dispersive interference: Unified analysis and required filter span," *IEEE Trans. on Communications*, vol. 47, no. 7, pp. 1073–1083, July 1999.

[Ari00] S. L. Ariyavisitakul, "Turbo space-time processing to improve wireless channel capacity," *IEEE Trans. on Communications*, vol. 48, no. 8, pp. 1347–1359, Aug. 2000.

[Bal92] P. Balaban and J. Salz, "Optimum diversity combining and equalization in digital data transmission with applications to cellular mobile radio. Part I: Theoretical considerations, and Part II: Numerical results," *IEEE Trans. on Communications*, vol. 40, no. 5, pp. 885–894, 895–907, May 1992.

[Bar01] S. Barbarossa and A. Scaglione, "Time-varying fading channels," in *Signal Processing Advances in Wireless and Mobile Communications - Volume II: Trends in Single- and Multi-User Systems*, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds., chapter 1. Prentice Hall, Inc., Upper Saddle River, NJ, 2001.

[Bas83] T. Basar, "The Gaussian test channel with an intelligent jammer," *IEEE Trans. on Information Theory*, vol. IT-29, no. 1, pp. 152–157, Jan. 1983.

[Bel00] K. L. Bell, Y. Ephraim, and H. L. V. Trees, "A Bayesian approach to robust adaptive beamforming," *IEEE Trans. on Signal Processing*, vol. 48, no. 2, pp. 386–398, Feb. 2000.

[Ben99] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. 37th Annual Allerton Conference on Communication, Control, and Computing*, pp. 987–996, Allerton House, Monticello, IL, Sept. 22-24, 1999.

[Ben01] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications*, L. C. Godara, Ed. CRC Press, 2001.

[Ben02] M. Bengtsson, "A pragmatic approach to multi-user spatial multiplexing," in *Proc. 2nd IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM-2002)*, Rosslyn, VA, Aug. 4-6, 2002.

[Ber67] T. Berger and D. W. Tufts, "Optimum pulse amplitude modulation. Part I: Transmitter-receiver design and bounds from information theory," *IEEE Trans. on Information Theory*, vol. IT-13, no. 2, pp. 196–208, April 1967.

[Big98] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2619–2691, Oct. 1998.

[Big01] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. on Information Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.

[Bin90] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Communication Magazine*, vol. 28, no. 5, pp. 5–14, May 1990.

[Böl02] H. Bölcskei and A. J. Paulraj, "Multiple-input multiple-output (MIMO) wireless systems," in *The Communications Handbook*, J. Gibson, Ed., pp. 90.1–90.14. CRC Press, 2nd edition, 2002.

[Bla57] N. M. Blachman, "Communication as a game," in *Proc. IRE WESCON Conference*, pp. 61–66, Aug. 1957.

[Bla87] R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, Reading, MA, 1987.

[Bor85] J. M. Borden, D. M. Mason, and R. J. McEliece, "Some information theoretic saddlepoints," *SIAM J. Control and Optimization*, vol. 23, no. 1, pp. 129–143, Jan. 1985.

[Boy00] S. Boyd and L. Vandenberghe, *Introduction to Convex Optimization with Engineering Applications*, Course Notes (available at http://www.stanford.edu/class/ee364). Stanford University, 2000.

[Bra74] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: The multivariate case," *The Bell System Technical Journal*, vol. 53, no. 5, pp. 745–778, May-June 1974.

[Bra83] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc.*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

[Cai99a] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. on Information Theory*, vol. 45, no. 6, pp. 2007–2019, Sept. 1999.

[Cai99b] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. on Information Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.

[Cal98] A. R. Calderbank, "The art of signaling: Fifty years of coding theory," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2561–2595, Oct. 1998.

[Car88] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 397–401, July 1988.

[Cha02a] S. S. Chan, T. N. Davidson, and K. M. Wong, "Asymptotically minimum bit error rate block precoders for minimum mean square error equalization," in *Proc. 2nd IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM-2002)*, pp. 140–144, Rosslyn, VA, Aug. 4-6, 2002.

[Cha02b] J.-H. Chang, L. Tassiulas, and F. Rashid-Farrokhi, "Joint transmitter receiver diversity for efficient space division multiaccess," *IEEE Trans. on Wireless Communications*, vol. 1, no. 1, pp. 16–27, Jan. 2002.

[Chi01] N. Chiurtu, B. Rimoldi, and E. Telatar, "On the capacity of multi-antenna Gaussian channels," in *Proc. IEEE International Symposium on Information Theory (ISIT 2001)*, p. 53, Washington, DC, June 24-29, 2001.

[Cho93] P. S. Chow, "Bandwidth optimized digital transmission techniques for spectrally shaped channels with impulse noise," Ph.D. dissertation, Stanford University, Stanford, CA, May 1993.

[Cho02] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. on Communications*, vol. 50, no. 7, pp. 1074–1080, July 2002.

[Chu02] C.-N. Chuah, D. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in dual-antenna-array wireless systems," *IEEE Trans. on Information Theory*, vol. 48, pp. 637–650, March 2002.

[Cio95a] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding - Part I: Equalization results," *IEEE Trans. on Communications*, vol. 43, no. 10, pp. 2582–2594, Oct. 1995.

[Cio95b] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding - Part II: Coding results," *IEEE Trans. on Communications*, vol. 43, no. 10, pp. 2595–2604, Oct. 1995.

[Cio97] J. M. Cioffi and G. D. Forney, "Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise," in *Communications, Computation, Control and Signal Processing*, A. Paulraj, V. Roychowdhury, and C. D. Schaper, Eds., chapter 4. Kluwer Academic Publishers, Boston, MA, 1997.

[Cov91a] T. Cover, "Universal portfolios," *Mathematical Finance*, vol. 1, no. 1, pp. 1–29, Jan. 1991.

[Cov91b] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[Cox87] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 10, pp. 1365–1376, Oct. 1987.

[Csi81] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.

[Dav02] T. N. Davidson, Z.-Q. Luo, and J. F. Sturm, "Linear matrix inequality formulation of spectral mask constraints with applications to FIR filter design," *IEEE Trans. on Signal Processing*, vol. 50, no. 11, pp. 2702–2715, Nov. 2002.

[Dig01] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. on Information Theory*, vol. 47, no. 7, pp. 3072–3081, Nov. 2001.

[Din02] Y. Ding, T. N. Davidson, J.-K. Zhang, Z.-Q. Luo, and K. M. Wong, "Minimum BER precoders for zero-forcing equalization," in *Proc. 2002 IEEE Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP'02)*, Orlando, Florida, May 2002.

[Din03a] Y. Ding, T. N. Davidson, Z.-Q. Luo, and K. M. Wong, "Minimum BER block precoders for zero-forcing equalization," *IEEE Trans. on Signal Processing*, to appear 2003.

[Din03b] Y. Ding, T. N. Davidson, and K. M. Wong, "On improving the BER performance of rate-adaptive block-by-block transceivers with applications to DMT," in *Proc. IEEE 2003 Global Communications Conference (GLOBECOM-2003)*, San Francisco, CA, submitted, Dec. 1-5, 2003.

[ETS98a] ETSI, "Channel models for HIPERLAN/2 in different indoor scenarios," *ETSI EP BRAN 3ERI085B*, pp. 1–8, March 1998.

[ETS98b] ETSI, "Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS (UMTS 30.03)," *Tech. Rep. TR101 112, v3.2.0,ETSI*, 1998.

[ETS01] ETSI, "Broadband radio access networks (BRAN); HIPERLAN type 2; physical (PHY) layer," *ETSI TS 101 475 V1.2.2*, pp. 1–41, Feb. 2001.

[Fan02] J. L. Fang, "Modeling and characterization of copper access systems," Ph.D. dissertation, Stanford University, Stanford, CA, Dec. 2002.

[Fos96] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, Autumn 1996.

[Fos98] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[Fos99] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE Journal on Selected Areas in Comm.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.

[Fud92] D. Fudenberg and J. Tirole, *Game Theory*, MIT Press, Cambridge, MA, 1992.

[Gal68] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc., USA, 1968.

[Gan00a] G. Ganesan and P. Stoica, "Space-time diversity," in *Signal Processing Advances in Communication*, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds., pp. 79–127. Prentice Hall, Upper Saddle River, NJ, 2000.

[Gan00b] G. Ganesan and P. Stoica, "Space-time diversity using orthogonal and amicable orthogonal designs," in *Proc. 2000 IEEE Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP'00)*, vol. 5, pp. 2561–2564, Istanbul, Turkey, June 5-9, 2000.

[Gan01a] G. Ganesan and P. Stoica, "Space-time block codes: A maximum SNR approach," *IEEE Trans. on Information Theory*, vol. 47, no. 4, pp. 1650–1656, May 2001.

[Gan01b] G. Ganesan and P. Stoica, "Utilizing space-time diversity for wireless communications," *Wireless Personal Communications*, vol. 18, pp. 149–163, 2001.

[GDF72] J. G. David Forney, "Maximum-likelihood sequence estimation of digital sequence in the presence of intersymbol interference," *IEEE Trans. on Information Theory*, vol. IT-18, no. 3, pp. 363–378, May 1972.

[GDF91] J. G. David Forney and M. V. Eyuboglu, "Combined equalization and coding using precoding," *IEEE Communications Magazine*, vol. 29, no. 12, pp. 25–34, Dec. 1991.

[Ges02] D. Gesbert, H. Bölcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. on Communications*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[Gin02] G. Ginis and J. Cioffi, "Vectored transmission for digital subscriber line systems," *IEEE Journal on Selected Areas in Communications (special issue on Twisted Pair Transmission)*, vol. 20, no. 5, pp. 1085–1104, June 2002.

[Goe99] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. on Communications*, vol. 47, no. 6, pp. 844–855, June 1999.

[Gol96] G. H. Golub and C. F. V. Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1996.

[Gol97] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. on Information Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[Gra72] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. on Information Theory*, vol. IT-18, no. 6, pp. 725–730, Nov. 1972.

[Han98] S. V. Hanly and D. N. C. Tse, "Multiaccess fading channels - Part II: Delay-limited capacities," *IEEE Trans. on Information Theory*, vol. 44, no. 7, pp. 2816–2831, Nov. 1998.

[Hir88] W. Hirt and J. L. Massey, "Capacity of the discrete-time gaussian channel with intersymbol interference," *IEEE Trans. on Information Theory*, vol. 34, no. 3, pp. 380–388, May 1988.

[Hir92] A. Hiroike, F. Adachi, and N. Nakajima, "Combined effects of phase sweeping transmitter diversity and channel coding," *IEEE Trans. on Vehicular Technology*, vol. 41, no. 2, pp. 170–176, May 1992.

[Hoc00a] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in rayleigh flat fading," *IEEE Trans. on Information Theory*, vol. 46, no. 2, pp. 543–564, March 2000.

[Hoc00b] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. on Communications*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.

[Hon90] M. L. Honig, K. Steiglitz, and B. Gopinath, "Multichannel signal processing for data communications in the presence of crosstalk," *IEEE Trans. on Communications*, vol. 38, no. 4, pp. 551–558, April 1990.

[Hon92] M. L. Honig, P. Crespo, and K. Steiglitz, "Suppression of near- and far-end crosstalk by linear pre- and post-filtering," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 3, pp. 614–629, April 1992.

[Hon03] Z. Hong, K. Liu, R. H. Jr, and A. Sayeed, "Spatial multiplexing in correlated fading via the virtual channel representation," *IEEE Journal on Selected Areas in Communications: Special Issue on MIMO Systems and Applications*, vol. 21, no. 5, pp. 856–866, June 2003.

[Hor85] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 1985.

[Hug88] B. Hughes and P. Narayan, "The capacity of a vector Gaussian arbitrarily varying channel," *IEEE Trans. on Information Theory*, vol. 34, no. 5, pp. 995–1003, Sept. 1988.

[IEE99] IEEE, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY)," *IEEE Std 802.11a*, Dec. 1999.

[Ise01] A. P. Iserte, A. I. Perez-Neira, and M. A. Lagunas, "Pre- and post-beamforming in MIMO channels applied to HIPERLAN/2 and OFDM," in *Proc. IST Mobile Communications Summit*, pp. 3–8, Barcelona, Spain, Sept. 2001.

[Ise02] A. P. Iserte, A. I. Perez-Neira, and M. A. Lagunas, "Joint beamforming strategies in OFDM-MIMO systems," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2002)*, Orlando, Florida, May 2002.

[Ise03] A. P. Iserte, M. A. Lagunas, and A. I. Perez-Neira, "Robust power allocation for minimum BER in a SISO-OFDM system," in *Proc. IEEE 2003 International Conference on Communications (ICC 2003)*, vol. 2, pp. 1263–1267, Anchorage, AK, May 11-15, 2003.

[Ivr03] M. T. Ivrlac, W. Utschick, and J. A. Nossek, "Fading correlations in wireless MIMO communication systems," *IEEE Journal on Selected Areas in Communications: Special Issue on MIMO Systems and Applications*, vol. 21, no. 5, June 2003.

[Jaf01] S. A. Jafar and A. Goldsmith, "On optimality of beamforming for multiple antenna systems with imperfect feedback," in *Proc. IEEE International Symposium on Information Theory (ISIT 2001)*, p. 321, Washington, DC, June 24-29, 2001.

[Jön02] G. Jöngren, M. Skoglund, and B. Ottersen, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. on Information Theory*, vol. 48, no. 3, pp. 611–627, March 2002.

[Jun95] P. Jung and J. Blanz, "Joint detection with coherent receiver antenna diversity in CDMA mobile radio systems," *IEEE Trans. on Vehicular Technology*, vol. 44, no. 1, pp. 76–88, Feb. 1995.

[Kal89] I. Kalet, "The multitone channel," *IEEE Trans. on Communications*, vol. 37, no. 2, pp. 119–124, Feb. 1989.

[Kan02] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 46–55, Jan. 2002.

[Kas85] S. A. Kassam and V. Poor, "Robust techniques for signal processing: A survey," *Proc. of the IEEE*, vol. 73, no. 3, pp. 433–481, March 1985.

[Kas90] S. Kasturia, J. T. Aslanis, and J. M. Cioffi, "Vector coding for partial response channels," *IEEE Trans. on Information Theory*, vol. 36, no. 4, pp. 741–762, July 1990.

[Kav85] M. Kavehrad and J. Salz, "Cross-polarization cancellation and equalization in digital transmission over dually polarized multipath fading channels," *At&T Technical Journal*, vol. 64, no. 10, pp. 2211–2245, Dec. 1985.

[Kay93] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall International, Inc., Englewood Cliffs, NJ, 1993.

[Kle96] A. Klein, G. K. Kaleh, and P. W. Baier, "Zero forcing and minimum mean-square-error equalization for multiuser detection in code-division multiple-access channels," *IEEE Trans. on Vehicular Technology*, vol. 45, no. 2, pp. 276–287, May 1996.

[Kri96] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, July 1996.

[Lag00] M. A. Lagunas, J. Vidal, and A. Pérez-Neira, "Joint array combining and MLSE for single-user receivers in multipath Gaussian multiuser channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 11, pp. 2252–2259, Nov. 2000.

[Lan69] P. Lancaster, *Theory of Matrices*, Academic Press, New York, 1969.

[Lap98] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[Leb97] H. Lebret and S. Boyd, "Antenna array pattern synthesis via convex optimization," *IEEE Trans. on Signal Processing*, vol. 45, no. 3, pp. 526–532, March 1997.

[Lee76] K. H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. on Communications*, vol. COM-24, no. 12, pp. 1283–1290, Dec. 1976.

[Liu96] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspaces," *Signal Processing*, vol. 50, no. 1-2, pp. 83–99, April 1996.

[Liu02] K. Liu and A. M. Sayeed, "Space-time D-block codes via the virtual MIMO channel representation," *IEEE Trans. on Wireless Communications*, submitted, Nov. 2002.

[Lok00] T. M. Lok and T. F. Wong, "Transmitter and receiver optimization in multicarrier CDMA systems," *IEEE Trans. on Communications*, vol. 48, no. 7, pp. 1197–1207, July 2000.

[Lue69] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley, New York, 1969.

[Lup89] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple access channels," *IEEE Trans. on Information Theory*, vol. 35, no. 1, pp. 123–136, Jan. 1989.

[Mag99] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York, 1999.

[Mar79] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, 1979.

[Mes96] D. J. G. Mestdagh and P. M. P. Spruyt, "A method to reduce the probability of clipping in DMT-based transceivers," *IEEE Trans. on Communications*, vol. 44, no. 10, pp. 1234–1238, Oct. 1996.

[Mil00] J. Milanovic, T. N. Davidson, Z.-Q. Luo, and K. M. Wong, "Design of robust redundant precoding filter banks with zero-forcing equalizers for unknown frequency-selective channels," in *Proc. 2000 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2000)*, vol. V, pp. 2761–2764, Istanbul, Turkey, June 2000.

[Mos96] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Communications Magazine: Topics in Personal Communications*, vol. 34, no. 10, pp. 124–136, Oct. 1996.

[Mou95] E. Moulines, P. Dehamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. on Signal Processing*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

[Mui82] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, New York, 1982.

[Mul76] C. T. Mullis and R. A. Roberts, "Synthesis of minimum roundoff noise fixed point digital filters," *IEEE Trans. on Circuits and Systems*, vol. CAS-23, no. 9, pp. 551–562, Sept. 1976.

[Nab02] R. U. Nabar, H. Bölcskei, V. Erceg, D. Gesbert, and A. J. Paulraj, "Performance of multi-antenna signaling techniques in the presence of polarization diversity," *IEEE Trans. on Signal Processing*, vol. 50, no. 10, pp. 2553–2562, Oct. 2002.

[Nar98] 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 Journal on Sel. Areas in communications*, vol. 16, no. 8, pp. 1423–1435, Oct. 1998.

[Nar99] A. Narula, M. D. Trott, and G. W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Trans. on Information Theory*, vol. 45, no. 7, pp. 2418–2432, Nov. 1999.

[Nee93] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. on Information Theory*, vol. 39, no. 4, pp. 1293–1302, July 1993.

[Neg99] R. Negi, A. M. Tehrani, and J. Cioffi, "Adaptive antennas for space-time coding over block-time invariant multi-path fading channels," in *Proc. 49th IEEE Vehicular Technology Conference (VTC'99)*, pp. 70–74, Houston, Texas, May 16-20, 1999.

[Nes94] Y. Nesterov and A. Nemirovsky, "Interior-point polynomial methods for convex programming," *Studies in Applied Mathematics, SIAM*, vol. 13, 1994.

[Ong03] E. N. Onggosanusi, A. M. Sayeed, and B. D. V. Veen, "Efficient signaling schemes for wideband space-time wireless channels using channel state information," *IEEE Trans. on Vehicular Technology*, vol. 52, no. 1, pp. 1–13, Jan. 2003.

[Osb94] M. Osborne and A. Rubinstein, *A Course in Game Theory*, MIT Press, Cambridge, MA, 1994.

[Oza94] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Vehicular Technology*, vol. 43, pp. 359–378, May 1994.

[Pal00] D. P. Palomar, M. Nájar, and M. A. Lagunas, "Self-reference spatial diversity processing for spread spectrum communications," *AEÜ International Journal of Electronics and Communications*, vol. 54, no. 5, pp. 267–276, Nov. 2000.

[Pal01a] D. P. Palomar, J. R. Fonollosa, and M. A. Lagunas, "Capacity results of spatially correlated frequency-selective MIMO channels in UMTS," in *Proc. IEEE Vehicular Technology Conf. Fall (VTC-Fall 2001)*, Atlantic City, NJ, Oct. 7-11, 2001.

[Pal01b] D. P. Palomar and M. A. Lagunas, "Temporal diversity on DS-CDMA communication systems for blind array signal processing," *EURASIP Signal Processing*, vol. 81, no. 8, pp. 1625–1640, Aug. 2001.

[Pal01c] D. P. Palomar and M. A. Lagunas, "Capacity of spatially flattened frequency-selective MIMO channels using linear processing techniques in transmission," in *Proc. 35th IEEE Annual Conference on Information Sciences and Systems (CISS-2001)*, The John Hopkins University, Baltimore, MD, March 21-23, 2001.

[Pal02a] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "On the optimal structure of transmit-receive linear processing for MIMO systems," in *Proc. 40th Annual Allerton Conference on Communication, Control, and Computing*, pp. 683–692, Allerton House, Monticello, Illinois, Oct. 2-4, 2002.

[Pal02b] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum joint transmit-receive linear processing for vectored DSL transmission with QoS requirements," in *Proc. 36th Asilomar Conference on Signals, Systems and Computers*, pp. 388–392, Pacific Grove, CA, Nov. 3-6, 2002.

[Pal03a] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," *IEEE Trans. on Information Theory*, vol. 49, no. 7, pp. 1707–1727, July 2003.

[Pal03b] D. P. Palomar, J. M. Cioffi, M. A. Lagunas, and A. P. Iserte, "Convex optimization theory applied to joint beamforming design in multicarrier MIMO channels," in *Proc. IEEE 2003 International Conference on Communications (ICC 2003)*, vol. 4, pp. 2974–2978, Anchorage, AK, May 11-15, 2003.

[Pal03c] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. on Signal Processing*, (submitted Feb. 2002, revised Dec. 2002) to appear 2003.

[Pal03d] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game-theoretic approach," in *Proc. IEEE 2003 International Symposium on Information Theory (ISIT'03)*, p. 271, Pacifico, Yokohama, Japan, June 29-July 4, 2003.

[Pal03e] D. P. Palomar and M. A. Lagunas, "Simplified joint transmit-receive space-time equalization on spatially correlated MIMO channels: A beamforming approach," *IEEE Journal on Selected Areas in Communications: Special Issue on MIMO Systems and Applications*, vol. 21, no. 5, pp. 730–743, June 2003.

[Pal03f] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. on Signal Processing*, (submitted May 2002, revised March 2003) to appear 2003.

[Pau97] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, vol. 14, no. 6, pp. 49–83, Nov. 1997.

[Ped00] K. I. Pedersen, J. B. Andersen, J. P. Kermoal, and P. E. Mogensen, "A stochastic multiple-input-multiple-output radio channel model for evaluation of space-time coding algorithms," in *Proc. IEEE Vehicular Technology Conference VTC-2000 Fall*, Vol. 2, pp. 893–897, Boston, MA, Sept. 2000.

[Pet78] W. W. Peterson and J. E. J. Weldon, *Error-correcting codes*, The MIT Press, Cambridge, Massachusetts, and London, England, second edition, 1978.

[Pro95] J. G. Proakis, *Digital Communications*, McGraw-Hill, 3rd edition, 1995.

[Qur85] S. U. H. Qureshi, "Adaptive equalization," *Proceedings of the IEEE*, vol. 73, no. 9, pp. 1349–1387, Sept. 1985.

[Ral98] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. on Communications*, vol. 46, no. 3, pp. 357–366, March 1998.

[Rey02] F. Rey, M. Lamarca, and G. Vázquez, "Optimal power allocation with partial channel knowledge for MIMO multicarrier systems," in *Proc. 56th Vehicular Technology Conference (VTC'02-Fall)*, Vancouver, Canada, Sept. 2002.

[Rey03] F. Rey, M. Lamarca, and G. Vázquez, "Transmit filter optimization based on partial csi knowledge for wireless applications," in *Proc. IEEE 2003 International Conference on Communications (ICC 2003)*, vol. 4, pp. 2567 –2571, Anchorage, AK, May 11-15, 2003.

[RF98a] F. Rashid-Farrokhi, K. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1437–1450, Oct. 1998.

[RF98b] F. Rashid-Farrokhi, L. Tassiulas, and K. R. Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays," *IEEE Trans. on Communications*, vol. 46, no. 10, pp. 1313–1324, Oct. 1998.

[Rhe01a] W. Rhee and J. M. Cioffi, "Ergodic capacity of multi-antenna Gaussian multiple access channels," in *Proc. 35th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 4-7, 2001.

[Rhe01b] W. Rhee, W. Yu, and J. M. Cioffi, "The optimality of beam-forming in uplink multiuser wireless systems," *IEEE Trans. on Wireless Communications*, to appear 2003. (See also "On the Asymptotic Optimality of Beam-forming in Multi-antenna Gaussian Multiple Access Channels" in Proc. IEEE Globecomm, San Antonio, TX, Nov. 25-29, 2001).

[Roc70] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, NJ, 2nd edition, 1970.

[Roo68] W. L. Root and P. P. Varaiya, "Capacity of classes of Gaussian channels," *SIAM J. Appl. Math.*, vol. 16, no. 6, pp. 1350–1393, Nov. 1968.

[Rui92] A. Ruiz, J. Cioffi, and S. Kasturia, "Discrete multiple tone modulation with coset coding for the spectrally shaped channel," *IEEE Trans. on Communications*, vol. 40, no. 6, pp. 1012–1029, June 1992.

[Sal85] J. Salz, "Digital transmission over cross-coupled linear channels," *At&T Technical Journal*, vol. 64, no. 6, pp. 1147–1159, July-Aug. 1985.

[Sal94] J. Salz and J. H. Winters, "Effect of fading correlation on adaptive arrays in digital mobile radio," *IEEE Trans. on Vehicular Technology*, vol. 43, no. 4, pp. 1049–1057, Nov. 1994.

[Sam01] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. on Communications*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[Say02] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[Sca99a] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. on Information Theory*, vol. 45, no. 3, pp. 1019–1032, April 1999.

[Sca99b] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers Part I: Unification and optimal designs," *IEEE Trans. on Signal Processing*, vol. 47, no. 7, pp. 1988–2006, July 1999.

[Sca02] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. on Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[Sch91] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley Publishing Company, Inc., 1991.

[Sch01] L. Schumacher, J. P. Kermoal, F. Frederiksen, K. I. Pedersen, A. Algans, and P. E. Mogensen, "MIMO channel characterisation," *Deliverable D2 V1.1 of IST-1999-11729 METRA project (available at http://www.ist-metra.org)*, pp. 1–57, Feb. 2001.

[Söd89] T. Söderström and P. Stoica, *System Identification*, Prentice Hall International, London, UK, 1989.

[Ses94] 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 Inform. Networks*, vol. 1, no. 1, pp. 49–60, Jan. 1994.

[Sha48] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July-Oct. 1948.

[Shi98] D. Shiu and J. M. Kahn, "Power allocation strategies for wireless systems with multiple transmit antennas," *Internal report of the University of California, Berkley and Lucent Technologies*, July 1998.

[Shi99] D. Shiu and J. M. Kahn, "Design of high-throughput codes for multiple-antenna wireless systems," *submitted to IEEE Trans. on Information Theory*, Jan. 1999.

[Shi00] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems," *IEEE Trans. on Communications*, vol. 48, no. 3, pp. 502–513, March 2000.

[Sim03] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE Journal on Selected Areas in Communications: Special Issue on MIMO Systems and Applications*, vol. 21, no. 3, pp. 406–417, April 2003.

[Sta88] W. E. Stark and R. J. McEliece, "On the capacity of channels with block memory," *IEEE Trans. on Information Theory*, vol. 34, no. 2, pp. 1273–1289, March 1988.

[Sta99] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*, Prentice Hall, Upper Saddle River, NJ, 1999.

[Tar98] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Trans. on Information Theory*, vol. 44, no. 2, pp. 744–765, March 1998.

[Tar99a] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. on Information Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[Tar99b] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Combined array processing and space-time coding," *IEEE Trans. on Information Theory*, vol. 45, no. 4, pp. 1121–1128, May 1999.

[Tel95] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommunications (ETT)*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999. (See also a previous version of the paper in *AT&T Bell Labs Internal Tech. Memo*, June 1995.).

[Tel98] J. Tellado and J. M. Cioffi, "Efficient algorithms for reducing PAR in multicarrier systems," in *Proc. IEEE International Symposium on Information Theory (ISIT'98)*, Cambridge, MA, Aug. 1998.

[Tel00] J. Tellado, *Multicarrier Modulation With Low PAR: Applications to DSL and Wireless*, Kluwer Academic Publishers, 2000.

[Tel01] I. E. Telatar, "Multi-user multi-antenna Gaussian noise fading channels," *Private Communication*, 2001.

[Tsa96] M. K. Tsatsanis and G. B. Giannakis, "Optimal decorrelating receivers for DS-CDMA systems: a signal processing framework," *IEEE Trans. on Signal Processing*, vol. 44, no. 12, pp. 3044–3055, Dec. 1996.

[Tsa97] M. K. Tsatsanis and G. B. Giannakis, "Transmitter induced cyclostationarity for blind channel equalization," *IEEE Trans. on Signal Processing*, vol. 45, no. 7, pp. 1785–1794, July 1997.

[Tse98] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels - Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. on Information Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.

[Ulu01] S. Ulukus and R. D. Yates, "Iterative construction of optimum signature sequence sets in synchronous CDMA systems," *IEEE Trans. on Information Theory*, vol. 47, no. 5, pp. 1989–1998, July 2001.

[Ung74] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. on Communications*, vol. COMM-22, no. 5, pp. 624–636, May 1974.

[Vee88] B. D. V. Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, April 1988.

[Ver84] S. Verdú and V. Poor, "On minimax robustness: A general approach and applications," *IEEE Trans. on Information Theory*, vol. IT-30, no. 2, pp. 328–340, March 1984.

[Ver89] S. Verdú, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Trans. on Information Theory*, vol. 35, no. 4, pp. 733–751, July 1989.

[Ver98] S. Verdú, *Multiuser Detection*, Cambridge University Press, New York, 1998.

[Vis99a] E. Visotsky and U. Madhow, "Optimum beamforming using transmit antenna arrays," in *Proc. IEEE 49th Vehicular Technology Conference (VTC'99)*, vol. 1, pp. 851–856, Houston, TX, May 1999.

[Vis99b] P. Viswanath and V. Anantharam, "Optimal sequences and sum capacity of synchronous CDMA systems," *IEEE Trans. on Information Theory*, vol. 45, no. 6, pp. 1984–1991, Sept. 1999.

[Vis99c] P. Viswanath, V. Anantharam, and D. N. C. Tse, "Optimal sequences, power control, and user capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Trans. on Information Theory*, vol. 45, no. 6, pp. 1968–1983, Sept. 1999.

[Vis01] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. on Information Theory*, vol. 47, no. 6, pp. 2632–2639, Sept. 2001.

[Vis02] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[Vor03] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem," *IEEE Trans. on Signal Processing*, vol. 51, no. 2, pp. 313–324, Feb. 2003.

[Wan00] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: Where Fourier meets Shannon," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 29–48, May 2000.

[Win94] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. on Communications*, vol. 42, no. 2/3/4, pp. 1740–1751, Feb./March/April 1994.

[Win98] J. H. Winters, "The diversity gain of transmit diversity in wireless systems with rayleigh fading," *IEEE Trans. on Vehicular Technology*, vol. 47, no. 1, pp. 119–123, Feb. 1998.

[Wit91] A. Wittneben, "Base station modulation diversity for digital SIMULCAST," in *Proc. I41st IEEE Vehicular Technology Conf.*, pp. 848–853, 1991.

[Wit95] A. Wittneben, "Optimal predictive Tx combining diversity in correlated fading for microcellular mobile radio applications," in *Proc. IEEE Global Telecommunications Conf.*, pp. 48–54, 1995.

[Wol78] J. Wolfowitz, *Coding Theorems of Information Theory*, Springer-Verlag, Berlin, Germany, 3rd edition, 1978.

[Won01] K. K. Wong, R. S. K. Cheng, K. B. Letaief, and R. D. Murch, "Adaptive antennas at the mobile and base stations in an OFDM/TDMA system," *IEEE Trans. on Communications*, vol. 49, no. 1, pp. 195–206, Jan. 2001.

[Wu96] S.-P. Wu, S. Boyd, and L. Vandenberghe, "FIR filter design via semidefinite programming and spectral factorization," in *Proc. IEEE Conf. on Decision and Control*, pp. 271–276, Dec. 1996.

[Yan93] K. Yanagi, "Optimal mutual information for coders and jammers in mismatched communication channels," *SIAM J. Control and Optimization*, vol. 31, no. 1, pp. 41–51, Jan. 1993.

[Yan94a] J. Yang and S. Roy, "Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback," *IEEE Trans. on Information Theory*, vol. 40, no. 5, pp. 1334–1347, Sept. 1994.

[Yan94b] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. on Communications*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

[Yu01a] W. Yu and J. M. Cioffi, "On constant-power water-filling," in *Proc. IEEE International Conference on Communications (ICC 2001)*, vol. 6, pp. 1665–1669, Helsinki, Finland, June 11-14, 2001.

[Yu01b] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple access channels," *IEEE Trans. on Information Theory*, to appear 2003. (See also "Iterative Water-filling for Gaussian Vector Multiple Access Channels" in Proc. IEEE International Symposium on Information Theory, Washington, DC, June 24-29, 2001.).

[Zhe03] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.