UNIVERSITAT POLITÈCNICA
DE CATALUNYA

Tesi Doctoral

Feedback-Channel and Adaptive MIMO Coded-Modulations

Autor: Francesc Rey Micolau

Director: Gregori Vázquez Grau

Departament de Teoria del Senyal i Comunicacions
Universitat Politècnica de Catalunya

Barcelona, Novembre 2005

Abstract

When the transmitter of a communication system disposes of some Channel State Information (CSI), it is possible to design linear precoders that optimally allocate the power, inducing high gains either in terms of capacity or in terms of reliable communications. In practical scenarios, this channel knowledge is not perfect and thus the transmitted signal suffers from the mismatch between the CSI at the transmitter and the real channel.

In that context, this thesis deals with two different, but related, topics: the design of a feasible transmitter channel tracker for time varying channels, and the design of optimal linear precoders robust to imperfect channel estimates.

The first part of the thesis proposes the design of a channel tracker that provides an accurate CSI at the transmitter by means of a low capacity feedback link. Historically, those schemes have been criticized because of the large amount of information to be transmitted from the receiver to the transmitter. This thesis focuses, thus, the attention in an accurate design of the return link. The proposed solution is based on the Kalman filter and follows a scheme that reminds the well known DPCM transmitter. The channel variability is processed by two identical linear predictors located at the transmitter and at the receiver, and a feedback link that assists the transmitter with the prediction error. This differential scheme becomes in interest because it allows to track the channel variations with only two or four bits per complex channel coefficient even in fast time-varying channels.

The rest of the thesis covers the second topic, studying different robust power allocation algorithms when the CSI is not perfectly known at the transmitter. For the sake of generality, the problem is formulated for the general MIMO OFDM case, encompassing the single antenna transmission, the beamforming schemes and the frequency-flat fading channels as particular cases.

First, the minimum MSE and the minimum uncoded BER parameters are chosen to be optimized, evaluating the performance of the algorithms in terms of uncoded BER. The basic novelty with respect to previous works that consider the same strategies of design is the proposal of a Bayesian approach for the design of the robust algorithms.

Next, the study is extended by proposing robust power allocation strategies focused on the minimization of the coded BER. For this purpose, information-theoretic criteria are used. Probably, one of the main contributions in the thesis is the proposal of the cut-off rate as a parameter of design whose maximization is directly related to the coded BER. This criterion is introduced as an alternative to the channel capacity and the mutual information for the design of optimal transceivers in the presence of any channel coding stage.

The last part of the thesis proposes a low complexity adaptive interleaver that, making use of the CSI available at the transmitter, reallocates the bits not only to combat the bursty channel errors but also to combat the specific distribution of the faded subcarriers. The design of this interleaver, named as "RCPC interleaver", is based on the Rate-Compatible Punctured Convolutional Codes. As shown by numerical results, the use of this interleaver with convolutional codes, improves the performance of the algorithms when they are compared with the classical block interleavers and pseudo-random interleavers.

Resum

En els sistemes de comunicacions on el transmissor disposa de certa informació sobre l'estat del canal (CSI), és possible dissenyar esquemes lineals de precodificació que assignin la potència de manera òptima induint guanys considerables, sigui en termes de capacitat, sigui en termes de la fiabilitat de l'enllaç de comunicacions. A la pràctica, aquest coneixement del canal mai és perfecte i, per tant, el senyal transmès es veurà degradat degut al desajust entre la informació que el transmissor disposi del canal i el seu estat real.

En aquest context, aquesta tesi estudia dos problemes diferents però alhora estretament relacionats: el disseny d'un esquema pràctic de seguiment del canal en transmissió per canals variants en temps, i el disseny d'esquemes lineals de precodificació que siguin robustos a la incertesa del canal.

La primera part de la tesi proposa el disseny d'un esquema de seguiment de canal que, mitjançant un enllaç de retorn de baixa capacitat, proporcioni al transmissor una informació acurada sobre el seu estat. Històricament, aquest tipus d'esquemes han rebut fortes crítiques degut a la gran quantitat d'informació que és necessari transmetre des del receptor cap el transmissor. Aquesta tesi, doncs, posa especial èmfasi en el disseny d'aquest canal de retorn. La solució que es proposa, basada en el filtre de Kalman, utilitza un esquema que recorda al transmissor DPCM. Les variacions del canal són tractades mitjançant dos predictors lineals idèntics situats en el transmissor i en el receptor, i un canal de retorn que assisteix el transmissor amb l'error de predicció. L'interès d'aquest esquema diferencial és que permet seguir les variacions del canal amb només dos o quatre bits per coeficient complex, fins i tot en canals ràpidament variants.

La resta de la tesi cobreix el segon objectiu, l'estudi de diferents esquemes d'assignació de potències quan el coneixement del canal en transmissió no és perfecte. El problema es planteja per a un sistema MIMO OFDM com a formulació més general, incloent els casos d'una sola antena, de l'esquema beamforming i del canal multiplicatiu com a casos particulars.

Primerament s'ha plantejat l'optimització dels criteris de mínim error quadràtic mig (MMSE) i mínima BER sense codificar. La innovació en el treball presentat a la tesi, respecte a altres treballs que segueixen els mateixos criteris de disseny, ha estat la formulació Bayesiana del problema per al disseny dels algoritmes robustos.

La tesi continua amb el plantejament d'estratègies robustes d'assignació de potència destinades a minimitzar la BER codificada. Per aquesta tasca s'han utilitzat criteris de teoria de la informació. Possiblement una de les principals contribucions d'aquesta tesi ha estat el plantejament del cut-off rate com a paràmetre de disseny. Aquest criteri s'introdueix com alternativa a la capacitat de canal o a la informació mutual per al disseny del transmissor quan s'inclou codificació de canal.

La ultima part de la tesi proposa un interleaver adaptatiu de baixa complexitat que, utilitzant el coneixement del canal disponible en el transmissor, assigna estratègicament els bits no només per combatre les ràfegues d'errors, sinó també per lluitar contra els esvaïments que puguin presentar les diferents portadores del canal per a una realització concreta de canal. El disseny d'aquest interleaver, anomenat "interleaver RCPC" està basat en els codis Rate-Compatible Punctured Convolutional Codes. Com s'il·lustra a partir del resultats numèrics, l'ús d'aquest interleaver conjuntament amb codis convolucionals millora les prestacions dels algoritmes quan es comparen amb les que s'obtindrien si s'utilitzes un interleaver de bloc o un interleaver pseudo-aleatori.

A Imma i als meus pares,

Agraïments

És una joia poder escriure aquestes línies, malgrat ser les primeres del document posen fi a un tot llarg treball de recerca; amb elles vull agrair molt sincerament el recolzament que he rebut al llarg d'aquest anys de totes aquelles persones que m'han envoltat tan en la vida professional com en la vida personal.

Sense més preàmbuls vull començar agraint al Gregori tot el seu suport, confiança i paciència; al llarg d'aquest anys hem mantingut moltes converses no només tècniques i professionals sinó també de la vida, des d'aquí MOLTES GRÀCIES. També es mereix un agraïment especial la Xell, el seu despatx sempre l'he trobat obert per discutir molts dels resultats que presento en aquesta tesi, espero poder-te tornar tot l'ajut incondicional que he rebut per part teva. Una altra porta oberta l'he trobat al despatx del Jaume, estic segur que si hagués fet una tesi en sincronització podria dir de tu el mateix que de la Xell.

També paraules d'agraïment als companys i amics que han fet la tesi al mateix temps que jo. Vull començar especialment pel Xavi Villares, possiblement la persona en la que més m'he recolzat en els moments difícils, els dies en que no ens podíem explicar penes de la tesi ja són anècdota!!! També gràcies al Jose A. López, a l'Enric Muntané i al Josep Bonet per haver revisat aquesta tesi, ànims amb les vostres tesis. Des d'aquí vull recordar també amb gratitud a tot el grup de becaris amb qui vaig començar aquest aventura, especialment al Xavi Mestre i el Gonzalo Seco, dos referents tan a nivell professional com a nivell personal, i al Gustavo Abrego, quantes vegades m'he recordat de tu quan deies: "hacer una tesis es como tener un hijo...".

Més agraïments pels companys del Grup de Processament del Senyal, especialment als de comunicacions, i també als companys de Vilanova, a tots gràcies!!! Espero que puguem treballar junts molts anys més...

Finalment una agraïment molt especial a tota la meva família. Vosaltres no m'heu pogut ajudar tècnicament en l'elaboració d'aquesta tesi, però el suport afectiu i l'estima són en moltes ocasions més necessaris que el recolzament tècnic. MOLTES GRACIES!!!

Aquest treball ha estat parcialemnt finançat pel Comissionat per a Universitats i Recerca del Departament de la Presidència de la Generalitat de Catalunya mitjançant la beca 1998FI 00307.

Contents

Notation

Through the thesis, matrices and vectors are set in uppercase boldface, e.g., $\mathbf{A}$ or $\mathbf{\Phi}$, and lower case boldface, e.g., $\mathbf{a}$ or $\boldsymbol{\phi}$, respectively. Scalar variables are set in normal typeface, e.g, a or ϕ. The subscript k generally denotes that the variable refers to the kth subcarrier, e.g., $\mathbf{F}_k$, $\mathbf{\Phi}_k$ or $\boldsymbol{\phi}_k$. The accent $\hat{(\cdot)}$ denotes an estimate or measured value of the accented variable, i.e., $\hat{\mathbf{h}}$ is the estimated value of vector $\mathbf{h}$. In Chapter 3 superscript $(\cdot)^{(t)}$ over the variables that refer to the channel response will be used to distinguish them from the frequency-domain, i.e., $\mathbf{h}^{(t)}$ is a column vector that contains the channel impulse response, while $\mathbf{h}$ is a vector that stores the complete channel response in the frequency domain. Other notation has been introduced as follows:

$\mathbf{A}^H$, $\mathbf{A}^T$, $\mathbf{A}^*$	Conjugate transpose (Hermitian), transpose and conjugate of the matrix $\mathbf{A}$, respectively.
$\mathbf{A}^{-1}$, $\mathbf{A}^{\#}$	Inverse and Moore-Penrose pseudoinverse of matrix $\mathbf{A}$, respectively.
$\det\{\mathbf{A}\}$	Determinant of matrix $\mathbf{A}$.
$\lvert a\rvert$	Absolute value (modulus) of the scalar a.
$\lvert\mathbf{a}\rvert$	Euclidean norm of vector $\mathbf{a}$: $\lvert\mathbf{a}\rvert = \sqrt{\mathbf{a}^H\mathbf{a}}$.
$\mathrm{Tr}\{\mathbf{A}\}$	Trace of matrix $\mathbf{A}$.
$\mathbb{R}^{M\times N}$, $\mathbb{C}^{M\times N}$	The set of $M \times N$ matrices with real and complex valued entries, respectively. The dimension of the space can be omitted when it is implicit in the context.
$\mathbf{0}_{M\times N}$, $\mathbf{0}_M$, $\mathbf{0}$	A $M \times N$ all-zeros matrix, a $M \times M$ all-zeros square matrix, an all-zeros column vector or matrix of implicit size.
$\mathbf{1}_{M\times N}$, $\mathbf{1}_M$, $\mathbf{1}$	A $M \times N$ all-ones matrix, a $M \times M$ all-ones square matrix, an all-ones column vector or matrix of implicit size.
$\mathbf{I}_M$, $\mathbf{I}$	The $M \times M$ identity matrix and the identity matrix of implicit size.
$\mathbf{i}_m$	Column vector with unit value in its ith entry and zeros elsewhere.
j	Imaginary unit: $j = \sqrt{-1}$.

$\text{vec}(\mathbf{A})$ The vec operator: if $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_d]$ is a $M \times N$ matrix, then $\text{vec}(\mathbf{A}) = [\mathbf{a}_1^T \cdots \mathbf{a}_d^T]^T$ is a MN column vector.

$\text{diag}(\mathbf{a})$ Diagonal matrix whose entries are the elements of vector $\mathbf{a}$.

$\text{diag}(\mathbf{A})$ Column vector whose elements are the diagonal entries of matrix $\mathbf{A}$.

$\text{diag}([\mathbf{A}_1 \ldots \mathbf{A}_K])$ Block diagonal matrix whose entries are matrices $\mathbf{A}_k$:

$$\text{diag}([\mathbf{A}_1 \ldots \mathbf{A}_K]) = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \ddots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{A}_K \end{bmatrix}.$$

$\triangleq$ Defined as.

$\simeq$ Approximately equal.

$\delta(n-l)$ Kronecker delta: $\delta(n-l) = \begin{cases} 1, & n = l \\ 0, & n \neq l \end{cases}$

$\delta(t)$ Dirac delta.

$\log(\cdot)$ Natural logarithm.

$\log_a(\cdot)$ Logarithm in base a.

$\mathbf{A} \otimes \mathbf{B}$ Kronecker product of matrices. If $\mathbf{A}$ is $M \times N$ and $\mathbf{B}$ is $P \times Q$, then the Kronecker product results in a $MP \times NQ$ matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} [\mathbf{A}]_{1,1}\,\mathbf{B} & \ldots & [\mathbf{A}]_{1,l}\,\mathbf{B} \\ \vdots & & \vdots \\ [\mathbf{A}]_{n,1}\,\mathbf{B} & \ldots & [\mathbf{A}]_{n,l}\,\mathbf{B} \end{bmatrix}.$$

$\mathbf{A} \odot \mathbf{B}$ Schur-Hadamard (elementwise) product between matrices $\mathbf{A}$ and $\mathbf{B}$ (they must have the same dimensions).

$\text{Re}\{\cdot\}$, $\text{Im}\{\cdot\}$ Real and imaginary parts. For matrices and vectors are defined elementwise.

$\min_x f(x)$ Value of x (it can be a scalar, vector or matrix) that minimizes the function $f(x)$.

$\max_x f(x)$ Value of x (it can be a scalar, vector or matrix) that maximizes the function $f(x)$.

$\min\{x, y\}$ The smallest of x and y.

$\max\{x, y\}$	The largest of x and y.
$\mathbf{P}_{\mathbf{A}}$, $\mathbf{P}_{\mathbf{A}}^{\perp}$	Projection matrix onto the subspace spanned by the columns of $\mathbf{A}$ and the orthogonal subspace, respectively.
$[a, b], (a, b)$	Closed interval ($a \leq x \leq b$) and open interval ($a < x < b$), respectively.
$[x]^+$	Positive part of x, i.e., $[x]^+ = \max\{0, x\}$. For vectors and matrices the operator is defined elementwise.
$\nabla_x f(x)$	Gradient of the function $f(x)$ with respect to x (it can be a scalar or vector).
$[\mathbf{A}]_{[i,j]}$	The $(i, j)^{th}$ element of matrix $\mathbf{A}$.
$[\mathbf{A}]_{[i,:]}$	The i^{th} row of matrix $\mathbf{A}$.
$[\mathbf{A}]_{[:,i]}$	The i^{th} column of matrix $\mathbf{A}$.
$\mathbf{a}(l)$	The l^{th} element of vector $\mathbf{a}$.
$\mathcal{Q}(\cdot)$	The Gaussian Q-function: $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du$.
$\mathcal{J}_o(x)$	The zero-order Bessel function of the first kind.
$\mathcal{H}(x)$	Entropy of the random variable X.
$\mathcal{I}(X; Y)$	Mutual information of random variables X and Y.
$\mathbf{a} \succeq \mathbf{0}$	Elementwise relation $a(i) \geq 0$.
$P(\mathbf{x} \to \mathbf{y})$	Pairwise error probability of decoding symbol $\mathbf{x}$ when symbol $\mathbf{y}$ was transmitted.
$\mathrm{E}\{\cdot\}$	Statistical expectation. A subscript can be used to indicate the random variable considered for the expectation.
$E_{\mathbf{h}_k \mid \hat{\mathbf{h}}}\{\cdot\}$	Statistical expectation over the random variable $\mathbf{h}_k$ given $\hat{\mathbf{h}}$.
$f_{\mathbf{h}/\hat{\mathbf{h}}}$	Probability density function of the random variable $\mathbf{h}$ given $\hat{\mathbf{h}}$.

Acronyms

ACM	Adaptive Coding and Modulation.
AR	Auto-Regressive.
ARQ	Automatic Repeat Request.
AS	Angular Spread.
ASK	Amplitude Shift Keying.
AWGN	Additive White Gaussian Noise.
2-D, **3-D**	Two-Dimensional, Three-Dimensional.
BER	Bit Error Rate.
BLAST	Bell-labs LAyered Space-Time Architecture.
BPSK	Binary Phase Shift Keying.
CCDF	Complementary Cumulative Distribution Function.
CCM	Constant Coding and Modulation.
CDMA	Code Division Multiple Access.
CSI	Channel State Information.
CSIR	Channel State Information at the Receiver.
CSIT	Channel State Information at the Transmitter.
DA	Data-Aided.
D-BLAST	Diagonal Bell-labs LAyered Space-Time Architecture.
DFT	Discrete Fourier Transform.
DLC-PDU	Data-Link Control Protocol Data Unit.
DoA	Direction of Arrival.
DoM	Direction of Movement.
DPCM	Differential Pulse Code Modulation.

DPS	Doppler Power Spectrum.
FDMA	Frequency Division Multiple Access.
FEC Codes	Forward Error Correcting Codes.
FFT	Fast Fourier Transform.
FIR	Finite Impulse Response.
IEEE	Institute of Electrical and Electronics Engineers.
IFFT	Inverse Fast Fourier Transform.
i.i.d.	Independent and Identically Distributed.
ISI	Inter-Symbol Interference.
KKT	Karush-Kuhn-Tucker.
KLD	Kullback-Leibler pseudo-Distance.
LDPC Codes	Low-Density Parity-Check Codes.
LMS	Least Mean Squares.
LS	Least Squares.
LTV	Linear Time-Varying.
MIMO	Multiple-Input Multiple-Output.
ML	Maximum Likelihood.
MMSE	Minimum Mean Square Error.
MSE	Mean Squared Error.
OFDM	Orthogonal Frequency Division Multiplexing.
OSTBC	Orthogonal Space Time Block Codes.
PAS	Power Angular Spectrum.
PDP	Power Delay Profile.
p.d.f.	Probability Density Function.
PEP	Pairwise Error Probability.
PER	Packet Error Rate.
PSAM	Pilot-Symbol Assisted Modulation.
PSK	Phase Shift Keying.
QoS	Quality of Service.
QAM	Quadrature Amplitude Modulation.

QPSK	Quadrature Phase Shift Keying.
RCPC Codes	Rate-Compatible Punctured Convolutional Codes.
RMS	Root Mean Squared.
SDM	Space Division Multiplexing.
SINR	Signal to Interference plus Noise Ratio.
SISO	Single-Input Single-Output.
SNR	Signal to Noise Ratio.
STBC	Space-Time Block Codes.
STTC	Space-Time Trellis Codes.
ST Turbo Codes	Space-Time Turbo Codes.
SVD	Singular Value Decomposition.
TDD	Time Division Duplex.
TDMA	Time Division Multiple Access.
THP	Tomnlinson-Harashima Precoding.
TV	Time Varying.
US	Uncorrelated Scattering.
V-BLAST	Vertical Bell-labs LAyered Space-Time Architecture.
WLAN	Wireless Local Area Network.
WSS	Wide Sense Stationary.

Chapter 1

Introduction

The modern communication systems, far from the pioneer systems in the 19th century that were composite by a single transmitter and a single receiver connected by a dedicated cable, are given by a combination of multiple transmitters and multiple receivers with multiple users simultaneously transmitting. It is clear that this higher complexity is accompanied by the necessity to solve a higher set of technical challenges. While the main handicap in the middle of 19th century was to get the signal to reach the receiver with a satisfactory quality, nowadays, the demand of high data rates to be provided in very noise environments, which sometimes include conflicts and interferences between users, set the troubles for the engineers.

This context induces a vast number of research areas. We can list, among others, the design of novel channel codes whose performance is close to the Shannon channel capacity, the study of sophisticated receivers based on iterative procedures that jointly decode and demodulate the signal, the proposal of schemes that exploit the sources of diversity, and the research on new transmitter and receiver schemes that mitigate the influence of the propagation channel.

The research carried out in this thesis is concerned with the last topic. Specifically, the thesis contributes to the investigation of algorithms that make use of the channel state information at the transmitter to improve the performance and reliability of the communications link. This interesting area of research has emerged during the last decade, and several choices have been proposed for adapting the transmitter on a given propagation channel (see Section 2.2 for an overview on this topic).

Doubtlessly, the main difficult to implement those adaptive transmitter schemes is that some information on the channel state is required at the transmitter, which opens a new topic to be investigated. The work in this thesis, hence, deals with two independent but related research problems: the study of transmitter designs that mitigate the influence of the propagation channel, and the design of feasible schemes that allow an accurate transmitter channel tracking.

1.1 Motivation and objectives

When the transmitter of a communication system disposes of perfect *Channel State Information* (CSI), it is possible to design linear precoders that optimally allocate the power inducing high gains either in terms of capacity or in terms of reliable communications. Those improvements have been evidenced trough multiple theoretical (e.g., [Fos98, Tel99]) and practical (e.g., [Hon92, Sca99a, Sca99b, Bar00]) studies. In practice, channel state information is always subject to misadjustments and, hence, the assumption of perfect CSI does not hold. The main consequence is that the performances predicted by the algorithms that assume perfect CSI fail because the transmitted signal suffers from the mismatch between the CSI at the transmitter and the real channel.

Considering the existence of those misadjustments between the real and the estimated channel, a reasonable question in this practical context is: How sensitive are the transmitters to the uncertainty in the channel knowledge?

To answer this question we reproduced the algorithm in [Bar00] designing a linear precoder that minimized the *Mean Squared Error* (MSE) for a MIMO OFDM communication system over a frequency-selective fading channel. To evaluate the loss due the CSI error we assumed both, perfect CSI and noisy CSI. The values of the channel uncertainty were based on a realistic assumption of a user moving at 4km/h [1]. Figure 1.1 displays the results in terms of coded BER for different MIMO configurations (i.e., 1x1, 2x2 and 3x3). As it can be seen, the performances of the algorithms are highly degraded when noisy CSI is used to design the transmitter. This influence of the channel uncertainty in the performance of the communication scheme motivates the main objectives of this thesis:

- To study robust algorithms that take into account the possible channel uncertainty when designing the transmitter schemes.

- To design of a feasible scheme that could provide an accurate channel state information at the transmitter.

These two objectives can be tackled separately. The dissertation can be roughly divided, thus, in two independent parts: the design of the algorithm that provides the CSI at the transmitter (Chapter 3) and the design of the robust algorithms (Chapter 4 and Chapter 5).

[1] It is not the aim of this section to present simulation results, but only motivate the necessity of robust designs. Hence, we omit here the information of the simulation parameters. The reader interested in this subject is referred to the first part of Section 4.5 and Section 5.5 where he will found that information.

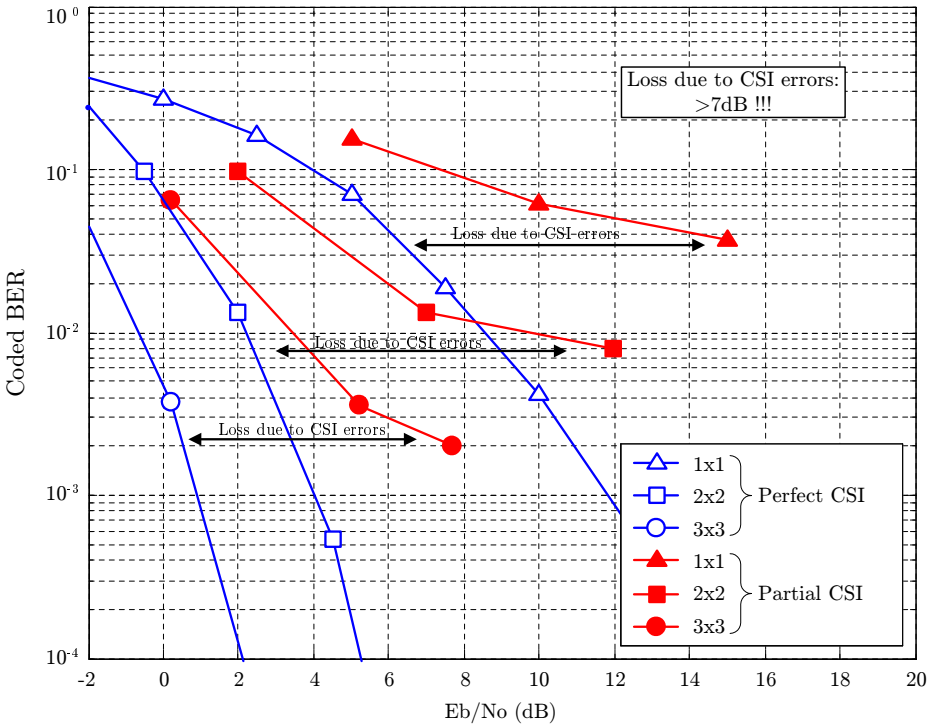


Figure 1.1: Illustration of the loss due to CSI errors. (See Section 5.5 for details on the simulation parameters).

1.2 Thesis outline and research contributions

Below is an outline of the main contributions made in this thesis summarized chapter by chapter, as well as references to previous presentations of the material at conferences and journal papers.

Chapter 2.

This chapter provides an overview of the history and recent development on channel uncertainty and transmit precoding designs. The chapter is structured into two sections. The purpose of the first section is to review the literature on channel uncertainty, focusing the attention on the design of robust algorithms that mitigate the influence of unreliable channel knowledge. After analyzing different viewpoints to design those robust algorithms, the chapter motivates the development of robust algorithms based on a Bayesian approach. The second part of the chapter presents the state-of-the-art on transmit precoding designs classifying those algorithms as a function of the channel state information at the transmitter. The aim of the classification is to provide the reader with an overview of the different transmit precoding algorithms, and to place the algorithms studied in this dissertation (i.e., the robust algorithms based on linear precoding schemes) in context with other methods.

Chapter 3.
This chapter covers the design of a feasible transmitter channel tracker based on some feedback information transmitted from the receiver, using a scheme that reminds the well known *Differential Pulse Code Modulation* (DPCM) transmitter and allows tracking slow and fast varying channels. After introducing the basis of the transmit channel tracker scheme, the chapter reviews some concepts of linear time-varying channels; describes the way to model the time-varying channel as an autoregressive process; and outlines how to predict the channel variations by means of a linear predictor. Next, the most novel and interesting technical contributions, including the design of the transmitter channel tracker using the Kalman filter, are presented. Among the main results we find the set of the minimum amount of information to be transferred trough the feedback link, the study of the quantization of this information, and the feedback channel throughput.

The work of this chapter has been published in a conference paper, in a journal paper that is under preparation, and as part of a deliverable in an European project founded by the European IST program:

- **F. Rey**, M. Lamarca, G. Vazquez, "Transmitter Channel Tracking for Optimal Power Allocation", in *proceedings of the International Conference on Acoustics, Speech and Signal Processing ICASSP'01*, Salt-Lake City (USA), May 2001.

- **F. Rey**, M. Lamarca, G. Vazquez, "Transmitter Channel Tracking for Optimal Power Allocation", in preparation.

- A. Kuzminskiy, M. Lamarca, J. A. López, I. Modonesi, **F. Rey**, "Performance analysis of re-configurable MTMR transceivers for WLAN", *FITNESS IST-2000-30116, Deliverable D3.2.1* available at http://www.telecom.ece.ntua.gr/fitness, November 2002.

Chapter 4.
This chapter deals with the design of robust power allocation algorithms with imperfect channel state information. The chapter focuses on two different design criteria: the minimization of the MSE and the minimization of the uncoded BER. The basic novelty with respect to previous works that consider the same strategies of design, is the proposal of a Bayesian approach for the design of the robust algorithms. Both, for the algorithms that minimize the MSE and for the algorithms that minimize the uncoded BER, a closed-form solution is obtained although the optimum solution for the algorithm that minimizes the uncoded BER requires an iterative algorithms.

The algorithms proposed in this chapter have been previously studied in one journal paper and diverse conference papers:

- **F. Rey**, M. Lamarca, G. Vazquez, "Robust Power Allocation Algorithms for MIMO OFDM systems with imperfect CSI", in *IEEE Transactions on Signal Processing*, Vol. 53, no. 3, pags. 1070-1085, March 2005.
- **F. Rey**, M. Lamarca, G. Vazquez. "A Joint Transmitter-Receiver Design in MIMO systems Robust to Channel Uncertainty for W-LAN Applications", in *proceedings of the IST Mobile&Wireless Telecommunications Summit 2002*, Thessaloniki (Greece). June 2002.
- **F. Rey**, M. Lamarca, G. Vazquez. "Optimal Power Allocation with Partial Channel Knowledge for MIMO Multicarrier Systems", in *proceedings of the 56th Vehicular Technology Conference VTC'02-Fall*. Vancouver (Canada). September 2002.
- **F. Rey**, M. Lamarca, G. Vazquez. "A Generalized Exponential BER Bound for Power Allocation With Imperfect Channel Estimates", in *proceedings of the International Conference on Acoustics, Speech and Signal Processing ICASSP'03*. Hong Kong, (China). April 2003.
- **F. Rey**, M. Lamarca, G. Vazquez. "Transmit Filter Optimization based on Partial CSI Knowledge for Wireless Applications" in *proceedings of the International Conference on Communications ICC'03*. Anchorage (AK-USA). May 2003.
- **F. Rey**, M. Lamarca, G. Vazquez. "A Robust Transmitter Design for MIMO Multicarrier Systems with Imperfect Channel Estimates" in *proceedings of the IEEE Workshop on Signal Processing Advances in Wireless Communications SPAWC'2003*. Rome (Italy). June 2003.

Chapter 5.

This chapter extends the study of Chapter 4 to the design of robust power allocation strategies focused on the minimization of the coded BER. For this purpose, information-theoretic criteria are used (e.g., channel capacity and mutual information). The main contribution of this chapter and, under the author's opinion, one of the main contributions of this thesis, is the proposal of the cut-off rate as a parameter of design whose maximization is directly related to the coded BER. Numerical results compare, in terms of coded BER, all the algorithms proposed in this dissertation and it is shown that, as expected, the algorithm that maximizes the cut-off rate outperforms the others. Analyzing the channel coding stage, the chapter also exhibits the importance of the interleaver and proposes an adaptive interleaver that make use of the channel state information at the transmitter. This result is presented as an alternative to the classical block interleavers and pseudo-random interleavers when CSI is available at the transmitter.

The derivation of this adaptive interleaver is previously unpublished, but other parts of the results in this chapter, specifically those that introduce the cut-off rate as a criterion for power allocation, have been published in some conference papers and will be submitted to a journal paper:

- **F.Rey**, M. Lamarca, G. Vazquez. "A Transmitter Design for Coded Systems in the Presence of CSI Errors". in *proceedings of the Thirty-seven Asilomar Conference on Signals, Systems and Computers*, Pacific Grove (CA-USA). November 2003.
- **F. Rey**, M. Lamarca, G. Vazquez, "Coded BER minimization for MIMO Multicarrier Systems with Imperfect Channel Estimates", in *proceedings of the IEEE Workshop on Signal Processing Advances in Wireless Communications SPAWC'04*, Lisbon (Portugal), July 2004.
- **F. Rey**, M. Lamarca, G. Vazquez, "A Robust Transmitter Design Based on Cutoff Rate for MIMO Multicarrier Systems with Imperfect Channel Estimates", in *proceedings of the 3rd IEEE Sensor Array and Multichannel Signal Processing Workshop SAM'04*, Sitges (Spain), July 2004.
- **F. Rey**, M. Lamarca, G. Vazquez, "A Transmitter Design for Coded Systems in the Presence of CSI Errors", in preparation to be submitted to *IEEE Transactions on Signal Processing*.

Chapter 6.
This chapter concludes the dissertation summarizing the main results of this thesis and comparing the different power allocation strategies that will be studied. Finally, the chapter outlines some ideas for future research lines.

Other contributions.
Into the framework of the author's thesis, but out of the scope of this dissertation, the author had some interesting results in the area of blind equalization and synchronization that were published in several conference papers:

- **F. Rey**, M. Lamarca, G. Vazquez. "Blind Equalization Based on Spatial and Temporal Diversity in Block Coded Modulations", in *proceedings of the 9th IEEE Symposium on Personal, Indoor and Mobile Radio Communications PIMRC'98*. Boston (MA-USA). September 1998.
- G. Vazquez, **F. Rey**, M. Lamarca. "Redundancy in Block Coded Modulations for Channel Equalization Based on Spatial and Temporal Diversity" in *proceedings of the International*

Conference on Acoustics, Speech and Signal Processing ICASSP'99. Phoenix (AZ-USA). March 1999.

- G. Vazquez, **F. Rey**, M. Lamarca, J.R. Fonollosa. "Diversity in Mobile Communications for Blind Detection of Block-Coded Modulations" in *proceedings of the 49th Annual International Vehicular Technology Conference VTC'99*. Houston (TX-USA). May 1999.

- **F. Rey**, G.Vazquez. "Blind Equalization of CDMA Systems using Spatial and Temporal Diversity Receivers", in *proceedings of the Fifth Baiona Workshop on Emerging Technologies in Telecommunications*. Baiona (Spain). September 1999.

- **F. Rey**, G. Vazquez, J. Riba. "Near-Far Resistant CML Propagation Delay Estimation and Multi-user Detection for Asynchronous DS-CDMA Systems", in *proceedings of the 50th Vehicular Technology Conference VTC'99-Fall*. Amsterdam (The Netherlands). September 1999.

- **F. Rey**, G. Vazquez. "Blind Equalization of DS-CDMA and MC-CDMA Modulations in Time-Variant Frequency Selective Channels", in *proceedings of the 50th Vehicular Technology Conference VTC'99-Fall*. Amsterdam (The Netherlands). September 1999.

- **F. Rey**, G. Vazquez, J. Riba. "Joint Synchronization and Symbol Detection in Asynchronous DS-CDMA Systems", in'*proceedings of 10th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*. Pocono Manor, (PA-USA). August 2000.

Reading the dissertation.

Because the thesis need not be read sequentially, Figure 1.2 in the next page summarizes the logical dependencies between the chapters. A solid line indicates that the chapters are closely coupled and should be read in the order indicated. Chapters 5 is, however, largely self-contained and it can be read independently if the reader is interested in the topic that covers: the design of power allocation strategies that try to guarantee good performances in terms of coded BER. The author's recommendation is, however, to read first Section 4.2 to get some insight into the system and channel model used in that chapter.

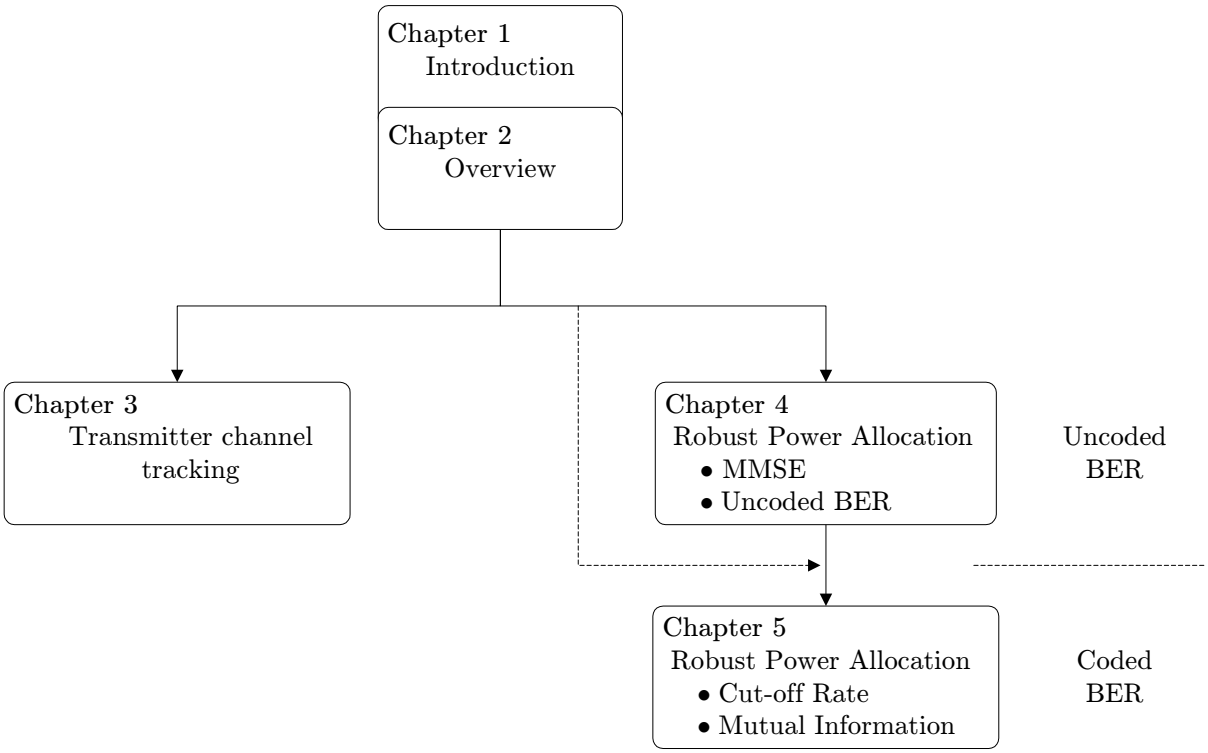


Figure 1.2: Illustration of the dependencies between the chapters.

Chapter 2

Overview

The growing demand for high-capacity communications has required the design of variable transmitter schemes that adapt the signal relative to the the channel state variations. To implement those adaptive transmission schemes and regulate the adaptive policy at the transmitter, accurate information of the channel state at the transmitter or, equivalently, a precise observation of the received signal quality, is required. This identifies two interesting research problems: the study of robust transmitter designs that mitigate the influence of *Channel State Information* (CSI) uncertainty and the design of feasible schemes that allow an accurate transmitter channel tracking in time-varying channels. While the study of closed-loop schemes that assume perfect or partial CSI at the transmitter has been the subject of numerous publications, the works that study and optimize the feedback channel link are scarce. This chapter outlines the state of the art in both topics and refers to Chapter 3 for the contributions on the channel tracking in time-varying channels, and Chapter 4 and 5 for the proposal of novel robust transmitter designs.

2.1 Channel uncertainty in communications

The purpose of this section is to review the literature on channel uncertainty focusing the attention on the way to model this unreliable channel and its influence on the digital communication systems. The section concludes with a survey of the different ways to design robust algorithms that are expected to perform well in the presence of limited channel knowledge.

2.1.1 Model for channel uncertainty

Before modelling the channel uncertainty is appropriate to review the multiple causes of unreliable CSI. These include, as more relevant, the measurement errors, the channel variations, and the inaccuracy in the channel modelling.

- Measurement Errors

  Channel estimation error are the first cause of noisy CSI. Inherent to the estimation process this source of error can be diminished but never suppressed. The quality in the channel estimation depends on the channel characteristics (noise level, interference degree, fading nature, correlation,...) and how the channel estimation is performed. Extensive efforts have been dedicated to the study of the channel estimators, including blind or semiblind estimators [Ton98] and data-aided or pilot assisted estimators [Man01, Mor01, Col02]. Focusing on the pilot assisted schemes, a special attention has been paid in the last years to the channel estimators based on the *Pilot-Symbol Assisted Modulation* (PSAM) technique [Cav91]. This technique inserts the pilot symbols periodically distributed into the information data providing an effective design to estimate the fading variations. This technique has been proposed for single transmit antenna systems [Oie04] and also for *Multiple-Input Multiple-Output* (MIMO) systems [Zho04]. The specific problem of channel estimation for OFDM modulation in a MIMO system, which is the scheme considered in this thesis, is addressed in [Bar02, Lar01, Li99, Li02] (see also [Tun01] for some interesting mathematical results and a MSE bound for channel estimation in OFDM MIMO systems)[1].

- Channel Variation

  As the channel varies in time, the algorithms that provide CSI might be continuously estimating the channel response. Otherwise, outdated channel measures would provide noisy CSI values. This situation is getting worse in modern communication systems. The channel variability, which depends on the Doppler spread and, consequently, is proportional to the carrier frequency and mobile velocity, is enlarged because higher frequency bands and mobile speeds are used. For these fast time-varying channels, CSI can be tracked at expenses of a throughput reduction since more frequent training sequences must be required. Hence, there is tradeoff between the channel uncertainty and the spectral efficiency. Alternatively, channel prediction methods can be used to counteract the time evolution and, therefore, reduce the degradation in channel estimation [Tsa97, DH00] (see also Section 3.1 for a review of the state-of-the-art in channel estimation and prediction of time-varying channels). This topic is extensively covered in Chapter 3 where a transmitter channel tracking scheme is proposed (see Section 3.5).

[1] We remark that we do not seek in this thesis to study algorithms that provide accurate measurements of the channel response. Some references to this area of research can be found in references above. Specifically, we refer the reader to [Ton98] and references therein for a tutorial on this topic.

- Inaccurate Model

  Even when the channel keeps invariant in time and measurement errors are negligible, still remains a channel uncertainty due to the mismatch between the real channel and the channel model. Although this channel uncertainty uses to be inappreciable, when it is present, it is difficult to evaluate and model.

Previous causes of channel uncertainty immediately affect the receiver. When CSI is considered at the transmitter there are several ways for which this information can be obtained at this side of the link. The most disadvantageous in terms of channel quality happens when channel reciprocity between the uplink and the downlink cannot be assumed [Cav00] [2], and consequently, a reliable feedback channel is required to guarantee an accurate CSI at the transmitter side. The degradation in the channel estimate quality is therefore intensified since over the channel uncertainty degree at the receiver we may add the inherent constraints of the feedback link.

- Feedback

  A careful design of the feedback channel link is crucial since it will determine the accuracy in the CSI at the transmitter and, in consequence, the reliability of the digital communication scheme.

  The main impairments in the return channel are the limited capacity of the link [Nar98, Tsa02], which restricts the amount of CSI available at the transmitter, and the feedback delay [DH00] (see also references therein), which forces the transmitter to use an outdated version of the channel state. Another type of feedback nonidealities are the transmission errors trough the return link. However, these errors are not usually considered as they can be diminished by introducing some redundancy with *Forward Error Correcting Codes* (FEC Codes), or by increasing the feedback delay and using an *Automatic Repeat Request* (ARQ) transmission protocol as suggested in [Gol97b]. The complexity of the channel feedback and its importance to provide accurate CSI at the transmitter is well worth a study, which can be found in Chapter 3. For a more reading on the state-of-the-art of the feedback channels, the reader is addresses to the introduction of that chapter.

[2] Channel reciprocity can be assumed in TDD systems when uplink and downlink experience the same propagation conditions, i.e. the TDD frame is shorter than the coherence time of the channel, or equivalently in FDD systems when the transmit and the receive frequencies are within the coherence bandwidth of the channel. Those conditions allow the uplink channel estimates to be used for downlink signal transmission since the channel variations are sufficiently slow.

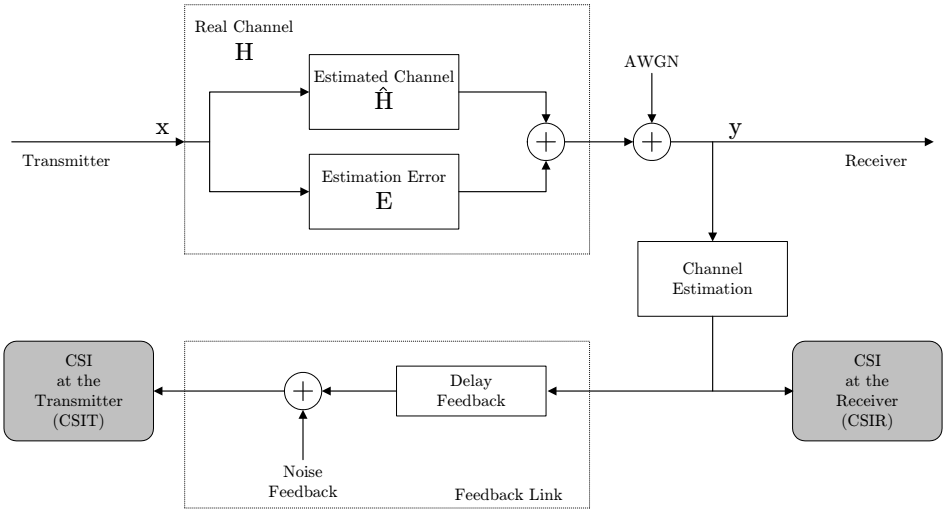


Figure 2.1: Block Diagram for Channel Uncertain Model.

Once we have had a look to the causes of noisy CSI, we will describe the way to model this channel uncertainty. Given partial CSI, the channel response $\mathbf{H}$ is generally modelled as a random process given by the summation of the measured channel $\hat{\mathbf{H}}$ and the channel estimation error $\mathbf{E}$. This channel model is illustrated in Figure 2.1, where $\mathbf{x}$ and $\mathbf{y}$ denote the channel input and output, respectively, and $\mathbf{n}$ is the additive noise term, which is assumed to be statistically independent of the channel and the error. Note that this model allows to consider any of the causes of imperfect CSI previously described.

Several statistical assumptions can be considered for the measured channel and the channel error. A simple model assumes the measured channel $\hat{\mathbf{H}}$ as a deterministic parameter whereas the channel estimation error $\mathbf{E}$ is modelled as a zero mean Gaussian random variable [Vis01]. Most of the references found in the literature assume, however, that the channel estimate is also a Gaussian random process, so that the channel $\mathbf{H}$ and its estimate $\hat{\mathbf{H}}$ are assumed to be jointly Gaussian [Méd00, Jön02, Rey05] . This second model has advantages over the first one since it allows to incorporate into the channel model not only the uncertainty over the channel estimation error, but also over the real channel. Moreover, it allows a Bayesian formulation of the problem that incorporates the estimated channel as a prior knowledge. The most general assumption for that model will, hence, consider the channel $\mathbf{H}$ as a Gaussian random process with nonzero mean and nonwhite covariance.

Focusing on the *Channel State Information at the Transmitter* (CSIT), two extreme cases can be considered [Vis01][3] : *mean feedback* and *covariance feedback*. In the mean feedback case the transmitter assumes knowledge of the channel mean, which is usually based on a noisy

[3]Although the same classification could be extended at the receiver, at this side of the link it is not used since CSIR quality is generally assumed to be almost perfect (only channel estimation errors are considered).

measurement of the channel, and models the covariance as a diagonal matrix proportional to the identity. This assumption only makes sense in time-invariant or slow time-varying channels when the full estimated channel state can be fed back to the transmitter. However, in fast time-varying channels, the limited capacity of the feedback link does not allow the transmitter to accurately track the channel. For those fast-varying scenarios the covariance feedback model assumes a zero mean channel and a non-white covariance that reflects the estimation of the second order channel statistics provided by the receiver. Examples of optimal transmitter schemes for the case of mean feedback can be found in [Nar98, Jön02, Zho02]. For the covariance feedback optimal transmitter algorithms were designed in [Jaf01a, Sam02, Sim03].

Apart from the previous models, which result from a classical channel estimator, other models could be considered for channel uncertainty. A finite state channel with noisy state information at both sides of the link is used in [Sal92], where the channel uncertainty is modelled as a defective binary cell with a certain probability to modify the input variables. In [Kru04] an information theoretic approach is used to model the channel uncertainty, defining four axioms to qualitatively describe the influence of channel uncertainty to the time varying channel information. Finally, authors in [Lap02] (see also [Lap98]) enumerate a list of possible examples of partial CSI.

2.1.2 Influence of channel uncertainty in communications

When perfect CSI is not available either at the transmitter or at the receiver or at both sides of the link, the performance of the algorithms is severely degraded if it is compared with the maximum achievable performance. In practice, CSI is always subject to misadjustments and hence, the study of the effects of this channel uncertainty has motivated the investigation of many researchers. Figure 2.2 displays the classification of the different schemes as a function of the degree of channel knowledge, combining the CSI both at the transmitter and the receiver. Note that, as argued in previous section, the quality of the channel estimation at the receiver is always assumed to be better than at the transmitter.

A vast number of efforts have been dedicated to evaluate the losses in the achievable data rates due to imperfect channel knowledge. Pioneering works studied the extreme cases of perfect CSI at the transmitter and receiver [Sha58, Wol64, Gal68] (see also more recent works in [Gol97a] and [Méd00]) for the single antenna case. More recently, the case of multiple antennas has been considered. Special relevance had the works of [Fos98, Tel99] that studied the capacity for the MIMO systems. On the other hand, the study of capacity when no CSI is available neither at the transmitter nor at the receiver can be found in [Mar99, Etk03, Lap03].

The case of noisy CSI, although more realistic, has not been studied until recently. Uncertainties arising from the receiver, and with no CSI at the transmitter, have been studied in [Méd00] for the case of single antenna, and extended to the MIMO configuration in [Yoo04a].

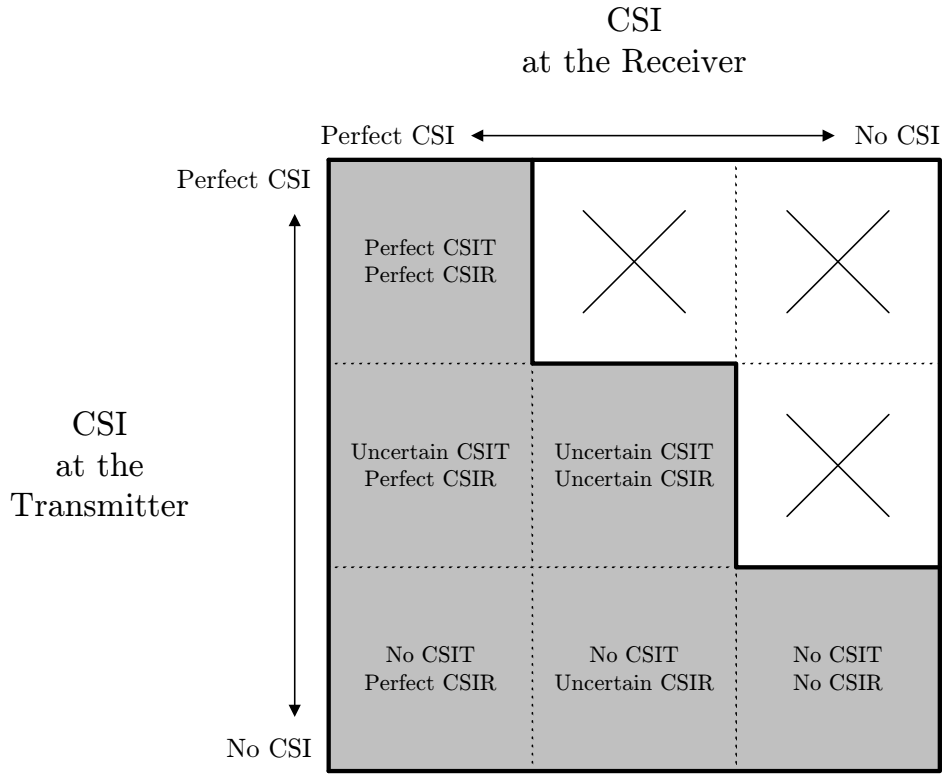


Figure 2.2: Classification of the algorithms as a function of the degree of channel knowledge.

Of particular interest is the study carried out by Médard [Méd00], which presents the effects of the noisy CSI on the mutual information for the simple-user and the multiple-user cases. Under the same considerations for the CSI, in [Lap02] the authors study when side information at the receiver can be considered perfect, concluding that this condition holds when the second moment of the channel uncertainty is negligible compared to the reciprocal of the SNR.

The case of partial CSI at the transmitter due to an imperfect feedback link has been studied from several points of view. The limits of reliable communications that include the effects of the feedback channel were investigated in [Cai99] (for uncertain CSIT) and in [Vis99] (for perfect delayed CSI). When only channel statistics are fed-back to the transmitter (see covariance feedback in the previous section) the channel capacity for multiple antennas was studied in [Jaf01a], whereas the case of nonzero mean (see mean feedback in the previous section) was examined in [Ven03].

Finally, the most general case of partial CSI at the transmitter and receiver was considered in [Yoo04b], which combines the effect of channel estimation error at the receiver with imperfect feedback (mean feedback and covariance feedback), and [Kyr02], which studies how the channel estimation error appears in the capacity formula, and derives mathematically expressions for the first and second order approximations of the error.

Apart from the influence of the channel uncertainty on the capacity, the effects of CSIT imperfections on the performance of the algorithms have been studied in [Agu03, Oie04, Ong01a, Tsa02, PI04b, Zho04]. In particular, [PI04b] study theoretically the degradation, providing an upper bound expression for the worst SNIR reduction and evaluating the degradation of the uncoded *Bit Error Rate* (BER). In [Ong01a, Tsa02] the performance of the closed-loop schemes is compared with the open-loop in terms of averaged uncoded BER. The authors examined the influence of outdated CSI at the transmitter due to the time varying nature of the Rayleigh [Ong01a] and the Ricean [Tsa02] fading channels. The conclusions that can be drawn are not surprising. When perfect CSI is available at the transmitter the optimum closed-loop design always outperforms the open-loop scheme. However, when only a delayed version of the CSI is available, the gain of closed-loop over open-loop schemes may significantly decrease (for Ricean channels closed-loop algorithms are more robust than for Rayleigh channels [Tsa02]) or even become negative at sufficiently fast time-varying channels.

2.1.3 Robustness to imperfect CSI

The influence of channel uncertainty in communications schemes motivates the design of robust algorithms that perform well in the presence of limited channel knowledge [Kas83]. In general, robust signal processing procedures do not perform as well as their corresponding nominal designs (when the nominal model is valid), and hence, an accurate treatment of the channel uncertainty is of paramount importance. For instance, an excessively conservative design could be insensitive to the channel variations and thus, although could be labelled as robust, its performance could be far from the nominal behavior, becoming, thus, a very bad robust design.

Different viewpoints can be followed to design algorithms that perform well when only noisy CSI is available. One design philosophy is based on the optimization of the worst-case effect of the channel uncertainty. This problem was formulated in a maximin approach in [Kas85], and more recently in [Vor03], for the context of robust adaptive beamforming. In [PI04a] this maximin approach was formulated to design a robust transmitter that combined *Orthogonal Space Time Block Codes* (OSTBC) with transmit beamforming. A game-theoretic framework can also be applied to optimize the worst-case performance as it is shown in [Mar83, Pal03c, Pal03a]. Clearly, the algorithms designed in this manner are excessively conservative since they assume the worst possible conditions. In favor of these schemes it can be said that they guarantee a minimum level of performance and thus a minimum Quality of Service (QoS). However, some information of how often this worst-case occurs should be also considered in order to avoid too pessimistic or too restrictive designs that usually spend more power than the strictly necessary.

Alternatively, a less conservative approach is possible by averaging the objective function with respect to the channel uncertainty. When the real channel and its estimate are assumed to be jointly Gaussian, this approach allows a Bayesian formulation that computes the expectation of the objective function with respect to the real channel given the channel estimates $E_{\mathbf{H}|\hat{\mathbf{H}}}\{\cdot\}$ [Nar98, Bel00, Jön02, Zho02, Rey05, Lia05]. Contrary to the maximin formulation, this technique guarantees an averaged QoS reducing the robustness condition and improving the performance when the channel estimate is far from the worst-case specifications.

2.2 Transmit precoder designs

Because perfect CSI is not possible to be available at the transmitter, an active area of study relies on the design of optimum transmission strategies when the transmitter has some, but not perfect, channel knowledge.

The analysis of channel side information at the transmitter was first introduced by Shannon in 1958 [Sha58]. Some years later Turin [Tur65, Tur66] and Viterbi [Vit65] analyzed the potential of a feedback channel link to inform the transmitter of the state of the receiver's uncertainty regarding which signal was sent in a binary transmission and a white Gaussian noise channel. Further, Schalkwijk et al. [Sch66b, Sch66a] proposed a coding scheme for additive noise channels. The first study through a Rayleigh fading multipath channel was done by Hayes at the late 1960's [Hay68]. The proposed scheme was quite similar to the current closed-loop schemes: an adaptive receiver learns the present state of the channel, uses this information to the detection process and sends to the transmitter, via a feedback channel, the information on the channel state in order to modify the transmitted signal in such a manner that the probability of error is minimized subject to a certain power constraint. His results showed that feedback techniques can improve the performance significantly. Moreover, trying to optimize the feedback link for his proposed scheme, he concluded that for a given channel state, only the sum of the squares of the channel gains were required to be fed back, rather than the module of each channel multipath, its phase an its time-delay.

Figure 2.3 illustrates the classification of the algorithms that adapt the transmitter as a function of the degree of channel knowledge when a feedback link is employed. In the two extremes we have the ideal feedback link, which yields algorithms that assume perfect CSIT, and the case of no feedback link. In between these extremes fall the algorithms that only have partial CSI at the transmitter, which can be broadly classified into three categories: mean feedback, covariance feedback and coarse feedback. The two former, defined in Section 2.1.1, are related with robust algorithms that adapt the resources at the transmitter as a function of the channel state. The latter use a low capacity feedback link to adapt a set of parameters at the transmitter like the channel code or the modulation. Next, a brief state-of-the-art on the

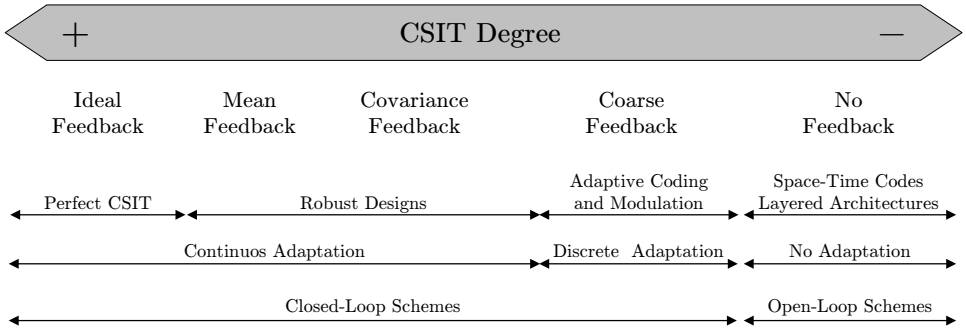


Figure 2.3: Classification of the feedback link as a function of the CSIT degree.

algorithms is described, presenting first the algorithms that do not assume CSI at the transmitter and then the algorithms that assume perfect or partial CSI.

2.2.1 Open-loop designs: No feedback link

When CSI is not available at the transmitter only open-loop algorithms, which do not take into account the channel response during the transmission of the data, can be used. In this case, however, the lack of channel knowledge at the transmitter does not prevent the system to exploit transmit diversity, since multiple transmit antennas combined with channel coding can be used to provide this diversity [Tar98]. Depending on the use made with the diversity provided by the MIMO channel, these algorithms can be classified into two general categories: transmit diversity and spatial multiplexing. The algorithms in the first category use the additional capacity of the MIMO channel to improve the quality and robustness of the link, thereby reducing the BER while keeping constant the data rate. Belong to this first family the space-time codes [Ges03] which include: *Space-Time Trellis Codes* (STTC) [Tar98], *Space-Time Block Codes* (STBC) [Ala98, Tar99, Jaf01b] and Space-Time Turbo Codes [Ste01a, Su01, Lam03]. On the contrary, into the second category fall the algorithms that exploit the multiple antennas to increase the data rate, introducing multiple data substreams on the different antennas. Belong to this second group the *Space Division Multiplexing* (SDM) [Nee00] and the layered architectures (BLAST) like D-BLAST [Fos96] and V-BLAST [Fos99, Loz02].

The decoding of previous algorithms requires the knowledge of the CSI at the receiver. Although this condition commonly holds, when no CSI is available neither at the transmitter nor at the receiver differential detection schemes can be used. For the specific 2×2 MIMO configuration a limited solution was proposed in [Tar00] based on a differential detection scheme. In [Hug00, Gan02a] another differential modulation procedure was proposed without the restrictions in the number of antennas. Finally [Hoc00b, Hoc00a] proposed a nondifferential approach that can be decoded without *Channel State Information at the Receiver* (CSIR) but with the inconvenient that requires exponential encoding and decoding complexities.

2.2.2 Discrete adaptation designs: Coarse feedback link

The algorithms that adapt the transmitter scheme as a function of the channel state can be classified, depending on the discrete or continuous nature of the adaptation, into two categories: the *discrete adaptation* schemes and the *continuous adaptation* schemes.

The *discrete adaptation* schemes include those adaptive techniques based on a finite set of transmission modes, which allow to vary, among others, the constellation size, the symbol transmission rate, the coding rate or the coding scheme, the number of transmit antennas or the transmitted power according to a finite-state power control policy (see [Vuc91, Gol98, Kos00] and references therein for studies of discrete adaptive schemes). The main advantage of these techniques is that they do not require a perfect channel knowledge at the transmitter, allowing to feedback a coarse quantized version of the channel to configure the best transmission mode. This schemes also provide the possibility to determine the adaptive policy at the receiver as a function of the quality of the received signal, and afterwards configure the transmitter using a low-capacity return channel link. The simplicity of these transmit adaptation schemes makes them feasible to be proposed in most of the current communication standards [HL201, IEE99, DVBc, DVBa]. The main disadvantage of these schemes is that if the number of finite transmission modes is small, the transmitted signal is not totally adapted to the channel conditions resulting in insufficient utilization of the full channel capacity.

As an alternative the *continuous adaptation* schemes allow to adapt more efficiently the transmitted signal to the channel state. The main handicap of these methods is that an accurate information of the channel state must be available at the transmitter at expenses of highly complex feedback schemes. The inefficiency of the feedback channel link, if the entire CSI is transmitted, reduced the interest of the researchers in the adaptive transmitter techniques. However, the growing demand for high capacity communications during the last decade has motivated a new interest in adaptive schemes in MIMO systems that utilize perfect or partial CSI at the transmitter to adjust a set of transmitter parameters (see next section).

2.2.3 Linear precoding designs: Perfect and partial CSIT

The historical evolution of transmit precoding designs as a function of the degree of channel knowledge has followed the same steps than the study of the maximum achievable data rates described in Section 2.1.2. Pioneering works analyzed the potential of channel knowledge at the transmitter assuming perfect CSI, whereas the more realistic cases of noisy CSI has not been studied until recently.

One of the earliest works in this area, assuming perfect CSI, was presented in [Sal85], where a *Minimum Mean Square Error* (MMSE) criterion was used to design transmit filters and a linear equalizers for the MIMO channel assuming uncorrelated data and white noise. Similarly, [Hon92] focused the design of transmit and receive filters on the suppression of near and far end crosstalk, generalizing the work in [Sal85] for correlated data symbols and colored noise. In [Yan94] a complete solution to the joint transmit-receive optimization problem with arbitrary dimensions and excess bandwidth was investigated following the approach in [Sal85], obtaining an analytical result for the optimum design in a MMSE sense. Analogous criteria addressing the problem of joint transmitter and receiver design in the context of wireless multiuser communications were proposed in [Jan98] and [Rap95]. While previous works required of an iterative procedure to achieve the optimum design, this was solved in [Sca99b] (see also [Bar00, Sca00]), leading to a closed-form solution for the design of the optimum transceivers following a discrete time-domain matrix formulation. Analytical performance results were presented for the zero-forcing and the MMSE criteria. Although the analysis was derived for the *Single-Input Single-Output* (SISO) case, the formulation was quite general, and the extension to multiple antennas was possible [Sam99]. In most cases, the minimization of the MSE does not reflect good performance in terms of communications reliability. Hence, other optimization criteria as the average uncoded BER [Ong02, Din03] or the mutual information [Dha96, Ral98, Sca99a] has been used as the performance measure, always under the assumption of perfect CSI at the transmitter.

When channel is not perfectly known, robust algorithms, following any of the strategies described in Section 2.1.3, were studied either satisfying a QoS constraint [Pal04] or satisfying any of the previous optimization criteria: MMSE [Rey05], average uncoded BER [Ong02, Rey05] or mutual information [Nar98].

Once the closed-form designs that exploit perfect or partial CSI at the transmitter were understood, some researchers, motivated by the excellent performance of these schemes, and the fact that perfect CSI is not available in practice, investigated the performance of hybrid schemes that combine space-time codes with closed-form precoding schemes. Space-time block codes combined with beamforming at the transmitter were proposed in [Gan02b, Jön02, Zho02, Lei02, Sam02, Tak02, Liu03, PI04a], while the union of space-time trellis codes with beamforming can be found in [Neg99, Li03, Li05]. The main interest in some of these solutions [Neg99, Sam02, Jön02, Zho02, PI04a] is that the design adapts automatically as a function of CSIT quality, taking the maximum benefit of the channel knowledge at the transmitter when it is available, and converging to classical space-time codes when no CSIT is obtained. This feature is common with the algorithms proposed in this thesis [Rey05, Rey04b, Rey04a], that moves from open-loop (the same power is allocated across all subcarriers and antennas) to the closed-loop scheme according to channel uncertainty.

2.2.4 Nonlinear precoding designs

As an alternative to the previous linear precoding designs, which are based on a linear transformation applied over the symbols to be transmitted, non-linear precoding structures can be used. The most common nonlinear precoding technique is the *Tomlinson-Harashima Precoding* (THP) [Tom71, Har72] (see also more recent work [Wes98]). This precoding technique, based on the introduction of a nonlinear predistorion stage at the transmitter, was proposed to combat the ISI as an alternative to a decision feedback equalizer, which could suffer from error propagation. However, this nonlinear structure has also been applied recently in multicarrier modulation techniques to combat inter-carrier interference [Ben02], MIMO channels to combat spatial interferences [Fis02, Win04, Sim04] and multiuser systems to combat multi-user interference [Win04, Wan05]. Schemes based on THP, like other precoding schemes, require perfect CSI at the transmitter. When this information is not totally available, robust techniques have been studied: [Lia05] (assuming delayed CSI and adopting a Bayesian formulation) or [Fis02, Sim04] (exploiting the channel correlation).

Special interest has the work in [Sim04] because uses a formulation that encompasses linear and nonlinear precoding techniques, such as THP, layered architectures (BLAST) and linear preequalization schemes. This generic formulation facilitates a comparison between linear and nonlinear structures (see also [Win04]). Experimental results show that when only channel correlation is known at the transmitter, linear precoding combined with a nonlinear equalizer performs, in general, better than nonlinear precoding schemes.

Chapter 3

Transmitter channel tracking in MIMO systems for optimal power allocation designs

Closed-loop techniques can only be applied when the transmitter has accurate knowledge of the channel response, either by means of a feedback channel from the receiver to the transmitter or by applying reciprocity between the uplink and the downlink channel responses. In this thesis we suppose that channel reciprocity cannot be assumed. In this sense, the chapter describes a feasible scheme to track the channel response based on channel prediction. Moreover, it also deals with the use of a return channel link as a suitable solution in the presence of *Linear Time-Varying* channels (LTV). When channel coefficients do not remain constant in time, it becomes inefficient to update frequently the *Channel State Information* (CSI) from the receiver to the transmitter. Instead, a more interesting design can be achieved if the channel variability is processed with a scheme containing two identical linear predictors at the transmitter and at the receiver, and a feedback link to assist the transmitter with the prediction error. This proposed solution addresses the severe limitations of the feedback channel capacity by means of a scheme which reminds the well known *Differential Pulse Code Modulation* (DPCM) transmitter and allows tracking slow and fast varying channels fitting the predictor to the channel dynamics.

The chapter is organized as follows. Section 3.1 introduces the transmitter channel tracking scheme proposed in this thesis, and reviews the state-of-the-art on the related topics. Section 3.2 and Section 3.3 introduce the signal model and define the most relevant parameters that characterize the linear time-varying channels, respectively. The main contributions of this chapter are covered in Section 3.4 and Section 3.5. While Section 3.4 analyze the way to predict the time evolution and capture the channel dynamics, modelling the LTV channel as a pth-order autoregressive process, Section 3.5 implements the linear predictor by means of a Kalman fil-

ter. That section also determines the minimum amount of information to be transferred trough the feedback link, studies the quantization of the prediction error and determines the feedback channel throughput. According to a rate-distortion criterion an accurate analysis on the return link channel throughput can be done as a function of the normalized Doppler frequency and the channel uncertainty degree. Section 3.6 analyzes the performance of the proposed scheme evaluating the quality of the channel tracker under different conditions. Finally, Section 3.7 summarizes and concludes the chapter.

3.1 Problem statement

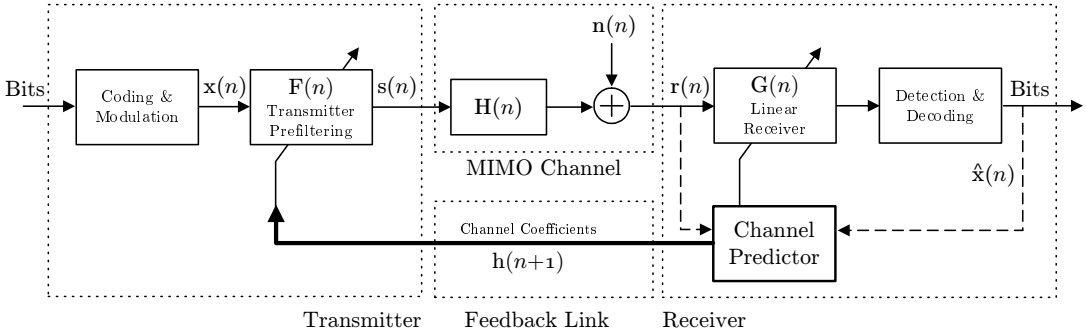
The study of the transmitter channel tracking scheme proposed in this thesis (this scheme was first published in [Rey01]) is decomposed into two stages: a first step that mitigates the feedback delay by means of channel prediction, and a second step that designs the return channel link and sets the minimum amount of information required to track the channel. This section briefly describes the specific channel scheme and summarizes the main advantages when compared with other alternatives. The last part of the section reviews the state-of-the-art on both topics, serving as an introduction to the rest of the chapter.

3.1.1 Transmitter channel tracking scheme

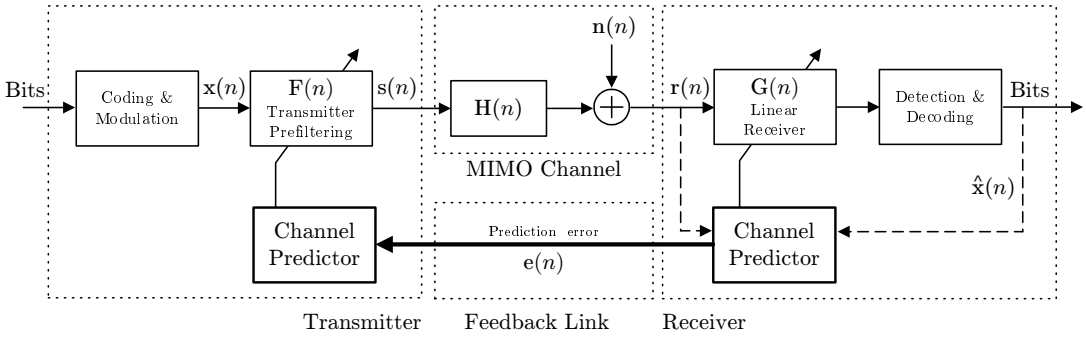
The design of feedback schemes that provide reliable CSI at the transmitter shall carefully address the feedback delay problem and propose efficient schemes that minimize the amount of information to be exchanged through the feedback link. The first issue is addressed by suggesting the use of a linear predictor that, advancing the channel dynamics, mitigates the feedback delay. The second issue is addressed by carefully analyzing the quantization of the feedback link to minimize the throughput at the feedback channel.

The design of the linear predictor requires an accurate model of the time-varying channel. We propose, in this chapter, to model the channel as an autoregressive process similar to that one suggested in [DH00] and references therein. The predictor is implemented by means of a Kalman filter, an efficient recursive method for computing the minimum variance linear filter [And79, p.47]. Furthermore, the structure of this filter allows not only to predict a priori the channel evolution from the transition (or state) equation, but also to estimate a posteriori the true channel values from the measurement equation and the channel output observations.

As important as the prediction of the channel is the quantization of the feedback link. In many cases, a quantized version of the predicted channel response at the receiver is employed to update CSI at the transmitter as illustrates Figure 3.1a (see e.g. [Nar98, Tsa02]). However, in fast time-varying channels, when high bandwidth constraints exist for the feedback link, it



(a)



(b)

Figure 3.1: Return channel link schemes. a) Non-optimum solution: mitigation of the feedback delay. The predicted channel coefficients obtained from a linear predictor at the receiver are fed back to the transmitter. b) Optimum solution: mitigation of the feedback delay and minimization of the amount of information to be fed back. Identical linear predictors at both sides of the link are available. Only the prediction error is fed back, reducing, hence, the throughput in the return link.

becomes inefficient to update frequently the predicted channel fading from the receiver to the transmitter. Instead, a more interesting design is proposed in this thesis with a scheme based on two identical linear predictors at the transmitter and the receiver sides, and a feedback link that assists the transmitter predictor with the prediction error rather than transmitting the channel coefficients themselves (see Figure 3.1b). The main advantage of this novel structure is that the dynamic range of the prediction error is much lower than that one of the channel coefficients. Hence, the amount of information to be transmitted in the feedback link (i.e., the channel throughput) is substantially reduced or, equivalently, for the same throughput, the quality of the CSI at the transmitter is improved. This structure, which reminds the well known DPCM transmitter, appears in a natural way when identical Kalman filters are implemented at both sides of the link. The first step in the Kalman filter (see Section 3.5), which predicts the channel state from the transition equation, can be identically processed at both filters. However,

the computation of the innovation, which corrects the channel prediction from the measurement equation, can only be calculated at the receiver since it requires the symbols at the channel output. Therefore, the Kalman filter at the transmitter has to be aided with this innovation, or equivalently with the channel prediction error, in order to keep both Kalman filters at the same state (i.e., with the same CSI prediction). An analysis on the return link channel throughput can be found in Section 3.5 as a function of the normalized Doppler frequency and the channel uncertainty degree using the rate-distortion theory.

The feasibility of the proposed transmitter channel tracking scheme depends on how fast the channel is changing and how frequent the channel predictor is updated. Next, we analyze the viability of the scheme for two real scenarios: HIPERLAN/2 standard (WLAN applications) and the DVB-S2 standard (satellite applications in the Ka-Band).

Scenario 1: Channel prediction in HIPERLAN/2 standard

Figure 3.2 illustrates the way to introduce the channel tracking scheme within the HIPERLAN/2 framework. The figure shows two consecutive HIPERLAN/2 MAC frames [HL200], and describes how the channel identification and the feedback link can be allocated into the MAC frame structure of HIPERLAN/2.

The MAC in HIPERLAN/2 is based on a TDD/TDMA approach using a MAC frame with duration of 2ms. This MAC frame includes time slots for data transmission in the downlink and the uplink phases, which are allocated dynamically depending on the transmission resources. Focusing on an arbitrary user, the downlink channel can be estimated at the receiver making use of the transmitted data during the corresponding slot, and the feedback channel can be allocated in the next uplink slot. Following the previous scheme, CSI is updated each 2ms (this time is approximate since the uplink and the downlink slots are dynamically allocated and differs form frame to frame).

In order to evaluate the feasibility of this proposal, a measure of how fast the channel varies in time (e.g., the normalized Doppler frequency, f_d or the coherence time T_c [Rap96, p.165]) must be computed. Assuming a user moving at $v = 4$km/h, a carrier frequency $f_c = 5$GHz, and a sampling period $T = 2$ms, the normalized Doppler frequency becomes:

$$f_d = \frac{f_c\, v}{c}\, T = 0.037 \tag{3.1}$$

This magnitude is better understood if it is used to compare the channel coherence time T_c with the period at which the CSI is updated $T = 2$ms. Defining the coherence time as a function of the normalized Doppler frequency [Rap96, p.165], the ratio "time coherence to sampling period" gives information of how fast the channel varies (T_c) when compared with the update frequency

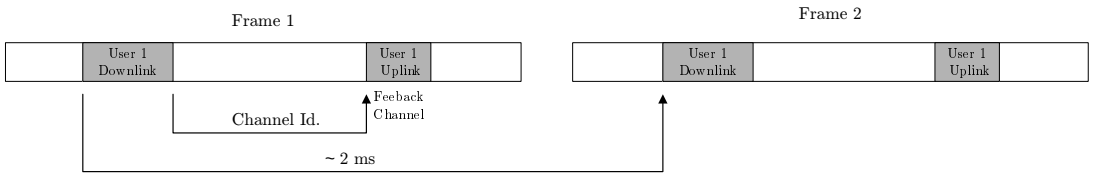


Figure 3.2: HIPERLAN/2 MAC Frame structure.

(T). In our case this ratio is:

$$\frac{T_c}{T} = \frac{9}{16\pi f_d} = 4.839 \tag{3.2}$$

which evidences that the time channel evolution is slow enough to be predicted.

It is worth to remark that HIPERLAN/2 standard does not consider either closed-loop algorithms or MIMO scenarios. Hence the references in this thesis to this standard are limited to the use of the same modulation and channel coding schemes.

Scenario 2: Channel prediction for satellite channels in the Ka-Band

The geostationary satellite channels, excluding specific cases, have been typically considered as ideal propagation channels that only introduce a time invariant attenuation and a constant delay in the communication path. However, for frequencies over 10GHz the propagation channel is vulnerable to variable weather conditions and, thus, the transmitted signal suffers from time varying fadings. In consequence, a special attention has to be paid to the satellite channels when operating over the 10GHz.

Specifically, the DVB-S2 standard operates at the Ka-Band (20-30GHz.), where the rain attenuation and scintillations are the two dominant factors for signal fading [May97, Cho02]. As it is specified in the draft of the DVB-S2 standard [DVBc], under fast fading conditions, the SINR variations in the Ka-band may occur as fast as 0.5 to 1dB/s. Thus, these channel variations may be critical in satellite systems where the time-delay propagation is not negligible (for geostationary systems 250ms). A possible solution to compensate the time lag is to predict the channel, using an autoregressive model as that one proposed in this section, in order to mitigate its variations.

Different arguments can support the use of a channel predictor in the Ka-band. First, the i.i.d. assumption in the channel evolution is not realistic because its time variation is highly correlated. Second, the rain and the scintillations can be modelled as Markov processes and thus, their dynamics are highly predictable. In [Cho02], experimental data have verified that first or second order autoregressive models, as previously proposed, fit to the channel variations. It has also been shown that attenuation in the Ka-band due to the rain and scintillations can be predicted within ±0.5dB one second ahead and ±1dB four seconds ahead. Although

some dependencies between rain attenuation and scintillation can be found in [May97], it is also shown in [Gre99] and references therein that this relationship can be neglected in practical cases. Finally, another source of propagation looses in the satellite link, especially at high frequencies, is the misalignment between antennas. The tracking algorithm that adjusts the antennas induces slow variations in the channel looses, and thus, this misalignment could also be introduced into the prediction model.

3.1.2 State-of-the-art on time-varying channel prediction

In the previous section we have seen the importance of the channel prediction to cope the channel variations. Next, the state-of-the-art on this topic is reviewed. Feedback delay impairment creates a need for predicting the channel variability diminishing the outdated CSI at the transmitter. Different methods have been used to address this problem depending on whether a deterministic model or a stochastic model is use to describe the channel behavior.

The most common approach to predict the channel follows a time-varying deterministic model based on a composite of sinusoids whose amplitude, frequency and phase are slowly variant [Jac74, p.46]. An accurate estimation of these parameters allows to extrapolate the channel evolution according to a deterministic model assuming that the channel structure remains constant. Subspace-based methods have been applied to estimate those channel parameters [Vau00]. Specifically, a modified root-MUSIC algorithm was used in [Hwa98], and an ESPRIT-type algorithm was proposed in [And99] and [Don01]. The deterministic nature of the previous model has been exploited by some researchers to propose a polynomial approximation method [She03], to suggest polynomial predictors for the prediction of the channel values [Tan95], or to introduce a SVD of the channel to separate space and time components, predicting the time evolution components and keeping constant the spatial structure [Gui04]. A generalization of the sinusoidal decomposition of the channel based on a basis expansion of the time varying impulse response was proposed in [Gia98], and used for channel estimation and prediction in [Tsa97].

The main drawbacks of the deterministic channel model are the misadjustments in the estimation of the parameters of the model and the short-range channel predictability that can be performed. Alternatively, a stochastic model introduces a different approach to predict the channel evolution, based on a linear prediction algorithm that tracks the channel dynamics using an autoregressive model. Overviews of the channel predictors based on the *Auto-Regressive* (AR) model can be found in [DH00] and references therein. Specifically, in Duel-Hallen works [Eyc98, DH00] a long-range channel prediction scheme of several tends-to-hundreds of symbols ahead is developed. The key of the large channel predictability is the low sampling rate of the channel estimation, which is much lower than the data rate, in combination with a channel interpolation for an accurate prediction of the channel coefficients. Recently, autoregressive pre-

diction schemes has been combined with *Pilot-Symbol Assisted Modulation* (PSAM) technique [Cav91, Cai05] for single transmit antenna systems [Oie04] and for MIMO systems [Zho04]. The main advantage of this joint study is that some relevant parameters, like the channel prediction error or the correlation between the true and the predicted channel, are related with parameters like the density of the pilot symbols or the amount of power allocated to them. Other forms of linear predictors include, for example, the design of a continuous time predictor [Lym03].

Opposite to the previous linear predictors, nonlinear schemes has also been investigated for the channel prediction. In [Gao96, Gao97] neuronal networks were proposed and it was demonstrated that they perform better than linear predictors when the effects of non-planar waves are considered. Furthermore, the study and extension of the linear predictor to Volterra filters was performed in [Zha97, Ekm99b], and the analysis of a nonlinear predictor using multivariate adaptive regression splines was proposed in [Ekm99a]. It has to be noted that in some cases (typically in most discrete link adaptation schemes) not the complex channel values, but only the channel power is of interest and, consequently, a prediction of the channel power becomes more efficient [Tan95, Gao96, Ste01b, Ekm01, Ekm02, Fal04].

3.1.3 Overview on feedback channel link

Most of the works that effort for accurate CSI at the transmitter focus on the propagation feedback delay, whereas only a few of them consider the limited capacity of the feedback link. It is obvious that the estimated or the predicted channel cannot be directly fed back to the transmitter since it results in an inefficient scheme, specially in fast TV channels when the CSI must be frequently updated.

The characterization of the feedback channel was studied in [Nar98], which considered the case in which the side information consists of a N-bit quantization of the fading channel. In that analysis two measures of performance were used: signal-to-noise ratio and mutual information. In order to diminish the impairments of the limited feedback bandwidth, special attention was addressed in the way how the bits were chosen to maximize the expected SNR or mutual information. The authors proposed the Lloyd algorithm [Llo57] and some heuristic quantization strategies to obtain the quantization regions and codebooks associated to the CSI. In [Tsa02] the influence of the number of quantization bits is studied, showing that for a two-element beamforming scheme, three bits were enough to quantize a frequency-flat fading channel. The authors in [Jöt02] developed a cost efficient procedure to feedback the CSI when the total number of bits available was constrained. The procedure was based on the decomposition of the total channel information into a slow fading term that was slowly tracked at the transmitter, and a fast fading term that was frequently updated. Finally, the work in [Mie05] proposed a vector quantization to reduce the bit rate in the feedback link.

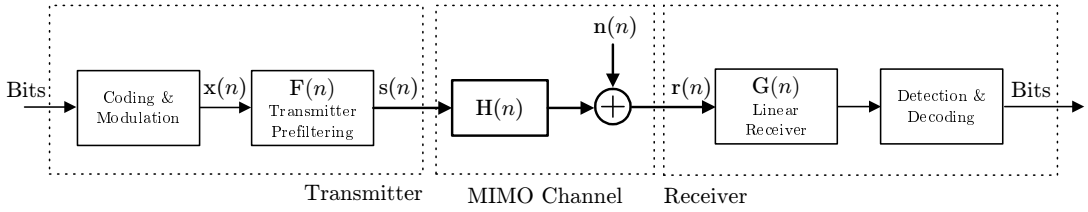


Figure 3.3: Generic discrete-time MIMO communications system.

3.2 Signal model

This section describes the matrix formulation used in the present chapter for the MIMO channel model. This model is introduced in its most general formulation and it is next detailed for specific cases like TDMA, OFDM or frequency flat fading channels.

A generic MIMO communications system with M_T transmitting and M_R receiving antennas is depicted in Figure 3.3. Since the main application of the proposed channel tracker is the design of closed-loop schemes, the formulation introduced in this figure is general and includes in the scheme the pre-filtering matrix $\mathbf{F}(n)$ and the post-filtering matrix $\mathbf{G}(n)$. However, neither the linear transformation matrix $\mathbf{F}(n)$ that optimally allocates the information symbols $\mathbf{x}(n)$ into the different antennas, nor the receiver structure that includes the linear receiver $\mathbf{G}(n)$ will be considered in this chapter (see Chapter 4 and Chapter 5 for the design of optimal linear precoder schemes).

Let $s_i(n)$ be the complex signal containing the transmitted symbols from the ith antenna, and let $h_{ij}^{(t)}(n;l)$ denote the time-varying channel impulse response from the ith transmitting to the jth receiving antenna[1], where n is the time index, and $l \in 1 \dots L_{ij}$ is the tap delay. A linear, discrete-time, baseband channel model can be introduced to describe the received signal at the jth antenna $r_j(n)$ as:

$$r_j(n) = \sum_{i=1}^{M_T} \sum_{l=0}^{L_{ij}-1} h_{ij}^{(t)}(n;l) s_i(n-l) + n_j(n) \tag{3.3}$$

where $n_j(n)$ is a zero mean circularly symmetric complex Gaussian sequence with covariance $E\left\{n_j(n) n_{j'}(m)^H\right\} = \sigma_n^2 \delta(n-m)\delta(j-j')$ that models the *Additive White Gaussian Noise* (AWGN) term.

The input-output relationship is usually written in a matrix notation. Denoting $\mathbf{h}_{ij}^{(t)}(n) = \left[h_{ij}^{(t)}(n;0) \dots h_{ij}^{(t)}(n;L-1)\right]^T$ as the vector that contains the L-length channel impulse response from the ith transmitting to the jth receiving antenna [2], and collecting in the Q column vector

[1] Superscript $(\cdot)^{(t)}$ over the variables that refer to the channel response will be used to distinguish them from the frequency-domain.

[2] In the MIMO case the length of each specific subchannel pair (L_{ij}) can be different. Hence L will denote the

$\mathbf{r}_j(n) = [r_j(nQ) \dots r_j(nQ+Q-1)]^T$ the received signal at the jth antenna, a linear input-output relationship can be written as follows:

$$\mathbf{r}_j(n) = \sum_{i=1}^{M_T} \mathbf{S}_i(n)\mathbf{h}_{ij}^{(t)}(n) + \mathbf{n}_j(n) \tag{3.4}$$

where $\mathbf{S}_i(n)$ is a matrix that contains the complex transmitted symbols from the ith antenna at the time index n, and $\mathbf{n}_j(n) = [n_j(nQ) \dots n_j(nQ+Q-1)]^T$ is the additive noise vector. This matrix formulation is quite general and encompasses TDMA, OFDM or frequency flat fading channels as particular cases. Hence, the specific structure of the matrix $\mathbf{S}_i(n)$ will be latter described for each individual case.

In order to estimate the channel coefficients and track its time evolution, it will be useful to store the complete channel response into a column vector $\mathbf{h}^{(t)}(n)$ that contains the channel impulse response for all transmitting and receiving antenna pairs:

$$\mathbf{h}^{(t)}(n) = \left[\mathbf{h}_{11}^{(t)}(n)^T \dots \mathbf{h}_{1M_R}^{(t)}(n)^T \dots \mathbf{h}_{M_T1}^{(t)}(n)^T \dots \mathbf{h}_{M_TM_R}^{(t)}(n)^T\right]^T \tag{3.5}$$

Hence, collecting all the channel responses into the vector $\mathbf{h}^{(t)}(n)$, defining $\mathbf{r}(n) = [\mathbf{r}_1^T(n) \dots \mathbf{r}_{M_R}^T(n)]^T$ as the column vector that stacks the received symbols over all antennas, the noise as $\mathbf{n}(n) = [\mathbf{n}_1^T(n) \dots \mathbf{n}_{M_R}^T(n)]^T$ and the matrix $\mathbf{S}(n) = [\mathbf{I}_{M_R} \otimes \mathbf{S}_1(n) \dots \mathbf{I}_{M_R} \otimes \mathbf{S}_{M_T}(n)]$, we are able to write the channel input-output relationship for the general MIMO system as follows:

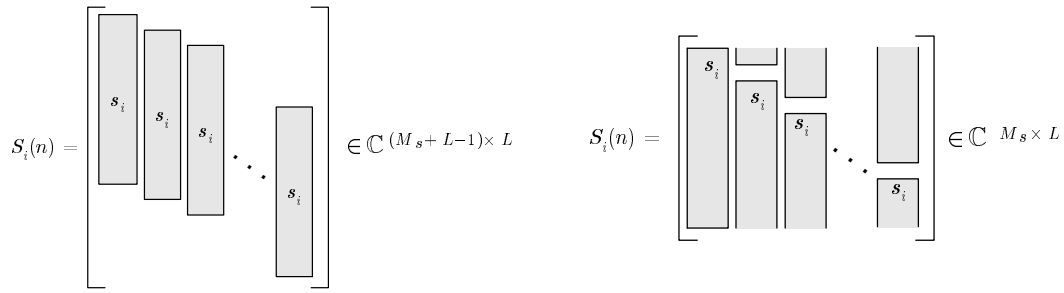
$$\mathbf{r}(n) = \mathbf{S}(n)\mathbf{h}^{(t)}(n) + \mathbf{n}(n) \tag{3.6}$$

There is a large variety of situations that can be formulated by the generic MIMO communications system represented in (3.6). To illustrate the generality of the transmitter channel tracking scheme proposed in the present chapter, next we will explore some of this modulation schemes detailing for each case the structure of matrix $\mathbf{S}_i(n)$.

Case 1: Time-division multiple access (TDMA)

A TDMA-based transmission scheme assumes that a block of M_s consecutive symbols are transmitted followed by guard intervals between successive blocks. Presuming that the length of the guard interval is appropriately chosen to avoid inter-block interference (at least L trailing zeros must be appended at the end of each data symbol), and considering that the channel keeps time-invariant along a block period, equation (3.6) may express the channel input-output relationship when $\mathbf{S}_i(n) \in \mathbb{C}^{(M_s+L)\times L}$, which contains the block of M_s symbols transmitted at the ith antenna, is characterized by the well known Toeplitz structure whose columns are scrolled versions of the vector $\mathbf{s}_i(n) = [s_i(nM_s) \dots s_i(nM_s + M_s - 1)]^T$ (see Figure 3.4a).

maximum length of all channel impulse responses $L = \max\{L_{ij}\}$ $i = 1 \dots M_T$, $j = 1 \dots M_R$. For the sake of simplicity the length of all channel impulse responses will be assumed to be equal.



TDMA (a) OFDM (b)

Figure 3.4: Schematic representation of the structured matrix $\mathbf{S}_i(n)$ in (3.6) for TDMA-based transmission schemes (a) and for OFDM modulation (b). The column vector $\mathbf{s}_i(n) = [s_i(nM_s) \dots s_i(nM_s + M_s - 1)]^T$ contains a block of M_s consecutive symbols transmitted from the ith antenna.

Case 2: Orthogonal frequency division multiplexing (OFDM)

For OFDM, or equivalently for discrete matrix multitone modulation, the previous matrix formulation is also applicable if vectors and matrices are properly defined. To generate the OFDM time waveform at each transmitting antenna the data symbols are first modulated into K frequency bins or subcarriers with the IFFT. Next a cyclic prefix, whose length is adjusted to the channel impulse response length L, is appended at the beginning of each data block and the time waveform is transmitted through the MIMO channel. At the receiver, assuming that the channel keeps invariant within an OFDM symbol, and after discarding the cyclic prefix, the vector that contains the received symbols at the jth antenna $\mathbf{r}_j(n) \in \mathbb{C}^{K\times 1}$ can be written in a similar form to (3.6). If the inverse discrete Fourier transform matrices for each antenna are included into the linear precoder matrix $\mathbf{F}(n)$, the K symbols at the output of the IFFT, previous to add the cyclic prefix, can be stacked to generate the first column in matrix $\mathbf{S}_i(n) \in \mathbb{C}^{(M_s)\times L}$ while the rest of the matrix appear as a circulant structure (see Figure 3.4b).

Case 3: Frequency Flat Fading Multi-Antenna Channel

For the frequency flat fading channel the MIMO model in (3.6) can be straightforwardly applied by defining the matrix that contains the transmitted data symbols $\mathbf{S}(n)$ is defined as:

$$\mathbf{S}(n) = [s_1(n)\ s_2(n)\ \dots s_{M_T}(n)] \otimes \mathbf{I}_{M_R}$$

3.3 Linear time-varying channels

Once the signal model has been described, a brief summary of linear time-varying channels and some general definitions are introduced in this section. If the reader is familiar with the concepts of WSSUS channel, power delay profile and Doppler power spectrum, he can proceed to the Section 3.4 where the model used to describe the TV channel is introduced.

A linear time-varying channel can be fully described by the Bello's model. Denoting the discrete-time channel impulse response function as $h(n;l)$, where index $l \in 1 \dots L$ denotes the multipath delay, and index n refers to the time evolution, the channel input/output relationship can be written as:

$$r(n) = \sum_{l=0}^{L} h(n;l)s(n-l) + n(n) \tag{3.7}$$

where $s(n)$ and $r(n)$ are the transmitted and received symbols, respectively, and $n(n)$ is the additive noise.

Depending on whether the channel evolution can be captured deterministically or not, a deterministic model or a stochastic model can be considered for describing the TV channel [Gia98].

3.3.1 Deterministic model

A deterministic channel may be fully described by the channel impulse response $h(n;l)$. However, a further representation of the TV channel described by the spectral content of the impulse response in both n and l variables provides more information on the channel variability. This representation was worked by Bello [Bel63] who introduced the following set of functions to describe the TV channel.

Delay-Doppler-spread function. The Delay-Doppler spread function describes the channel behavior in the Doppler domain as follows [3]:

$$S(\nu;l) \triangleq \sum_{n=-\infty}^{\infty} h(n;l)e^{-j2\pi\nu n} \tag{3.8}$$

where a finite 2-D support over ν and l domains is used to describe the channel. While the multipath delay l is responsible for the time dispersion (i.e., the frequency distortion), the new variable ν corresponds to the normalized Doppler shifts producing the frequency dispersion (i.e., the temporal variations). The delay-Doppler-spread function is interesting in that it explicitly illustrates both time and frequency dispersion

[3] The original Delay-Doppler-spread function defined by Bello in [Bel63] was given for the continuous-time representation of the channel impulse response $S(\nu;\tau) \triangleq \int_{-\infty}^{\infty} h(t;\tau)\, e^{-j2\pi\nu t} dt$.

of a channel. Moreover parameters in the the Doppler domain (e.g. the maximum Doppler shift) can offer valuable information on the channel variability. Its physical meaning is better understood if the channel impulse response in (3.7) is replaced by the inverse Fourier transform of the Delay-Doppler spread $S(\nu; l)$ with respect to the normalized Doppler shift ν . Hence, the channel input-output relationship is described in terms of $S(\nu; l)$ [Gre92]:

$$r(n) = \int_{-\nu_{max}}^{\nu_{max}} \sum_{l=0}^{L} S(\nu; l) s(n-l) e^{j2\pi\nu n} d\nu \tag{3.9}$$

that is, the symbols at the channel output are represented as the sum of delayed and Doppler shifted replicas of the symbols $s(n)$ at the channel input.

Time variant transfer function. Similar to the Delay-Doppler spread function, a deterministic TV channel can be represented in the transfer domain with respect to the delay index l as:

$$T(n; f) \triangleq \sum_{l=0}^{L-1} h(n; l) e^{-j2\pi f l} \tag{3.10}$$

The time-variant transfer function has the same physical meaning that its time-invariant counterpart, describing the channel frequency response for a certain input signal at time n.

3.3.2 Stochastic model

To fully describe a stochastic model for a time-varying random channel is a difficult task. However, when the aim is to model the time variability, for example by means of an autoregressive process, the second order statistics of the channel impulse response are enough to characterize the channel evolution. Hence, the autocorrelation function of the randomly time-variant complex impulse response is introduced:

$$R_h(n_1, n_2; l_1, l_2) = E\left\{h(n_1; l_1) h^*(n_2; l_2)\right\} \tag{3.11}$$

This channel characterization can be extended by the correlation of any of the Bello's functions. Hence, when the channel is modelled as an stochastic process, an statistical description of the Delay-Doppler-spread function based on the correlation becomes:

$$R_S(\nu_1, \nu_2; l_1, l_2) = E\left\{S(\nu_1; l_1) S^*(\nu_2; l_2)\right\} \tag{3.12}$$

And similarly the correlation function of the time varying transfer function is denoted as:

$$R_T(n_1, n_2; f_1, f_2) = E\left\{T(n_1; f_1) T^*(n_2; f_2)\right\} \tag{3.13}$$

A common simplification of the time-varying random channel is based on the assumption of stationarity and uncorrelation between scatters as next introduced.

Wide sense stationary assumption (WSS). A channel is said to accomplish the wide sense stationary conditions with respect to the time index n if its first two moments (mean and autocorrelation) do not depend on the reference time. That is:

$$\begin{aligned} E\{h(n_1;l)\} &= E\{h(n_2;l)\} \\ R_h(n_1,n_2;l_1,l_2)_{WSS} &= R_h(\Delta n;l_1,l_2) \end{aligned} \tag{3.14}$$

In the dual Doppler domain, the wide sense stationary assumption gives an uncorrelated function in the variable ν:

$$R_S(\nu_1,\nu_2;l_1,l_2)_{WSS} = P_S(\nu_1;l_1,l_2)\delta(\nu_2-\nu_1), \tag{3.15}$$

where the function $P_S(\nu_1;l_1,l_2)$ is the cross spectral density function of $h(n,l_1)$ and $h(n,l_2)$, defined as the Fourier transform of the time autocorrelation function: $P_S(\nu_1;l_1,l_2) = \sum_{\Delta n=-\infty}^{\infty} R_h(\Delta n;l_1,l_2)e^{-j2\pi\nu_1\Delta n}$.

Uncorrelated scattering assumption (US). A channel is said to accomplish the uncorrelated scattering condition when the impulse response is uncorrelated in the delay index l:

$$R_h(n_1,n_2;l_1,l_2)_{US} = R_h(n_1,n_2;l_1)\delta(l_2-l_1) \tag{3.16}$$

US assumption is sometimes called the dual to WSS assumption in the sense that the process is wide sense stationary in the frequency domain f (i.e., frequency correlation of US channels only depends on the frequency separation and not on the absolute frequency values):

$$R_T(n_1,n_2;f_1,f_2)_{US} = R_T(n_1,n_2;\Delta f) \tag{3.17}$$

Wide sense stationary uncorrelated scattering assumption (WSSUS). Combining WSS and US assumptions, a channel is called WSSUS if:

$$\begin{aligned} R_h(n_1,n_2;l_1,l_2)_{WSSUS} &= R_h(\Delta n;l_1)\delta(l_2-l_1) \\ R_S(\nu_1,\nu_2;l_1,l_2)_{WSSUS} &= P_S(\nu_1;l_1)\delta(\nu_2-\nu_1)\delta(l_2-l_1) \end{aligned} \tag{3.18}$$

where $P_S(\nu_1;l_1)$ will be defined as the scattering function of the TV channel.

Although WSSUS assumption is commonly applied for mostly fading channels, it is well known that neither a channel can be rigorously assumed to be stationary, nor US assumption can be accepted for all channels (e.g., some underwater environments [Ilt91] cannot be considered US).

Nevertheless, stationary assumption is suitable for restricted time intervals and limited frequency bands. An accurate model for this channels was suggested by Bello [Bel63] as Quasi-WSSUS. A Quasi-WSSUS channel becomes a WSSUS channel into a bounded region, and outside this region correlation functions are not considered to accomplish the stationary conditions. Further on, we will consider for convenience and simplicity that TV channels satisfy the WSSUS assumptions.

From a practical point of view a set of functions can be introduced to calculate various channel parameters that characterize the multipath effects and the time channel variations. Specifically, the scattering function, the power delay profile and the Doppler power spectrum are next defined:

Scattering function. The scattering function of the TV channel $P_S(\nu;l)$[4] (also named power spectral density function), is defined as the Fourier transform of the time correlation channel under the WSSUS assumption, and provides a measure of the channel behavior as a function of the time delay l and the normalized Doppler frequency ν:

$$P_S(\nu;l) = \sum_{\Delta n=-\infty}^{\infty} R_h(\Delta n;l) e^{-j2\pi\nu\Delta n}. \tag{3.19}$$

This function fully describes the second order statistics of the TV channel under the WSSUS conditions and it is a support function to compute the power delay profile and the Doppler power spectrum.

Power Delay Profile. The *Power Delay Profile* (PDP) (or multipath intensity profile) models the power decay over the delay index l in a WSSUS channel, and it is defined as:

$$P_{PDP}(l) = E\{h(n;l)h^*(n;l)\} = \int_{-\nu_{max}}^{\nu_{max}} P_S(\nu;l)d\nu \tag{3.20}$$

This function is the basis to calculate the delay spread and coherence bandwidth of the channel [Rap96, p.160].

Doppler Power Spectrum. Similarly, in the dual frequency domain, the *Doppler Power Spectrum* function (DPS) describes the time variability of the channel fadings, and is related with the scattering function by the equation:

$$P_{DPS}(\nu) = \sum_{l=0}^{L-1} P_S(\nu;l) \tag{3.21}$$

Some parameters related with this function are the coherence time and the Doppler spread [Rap96, p.165].

[4]Suffixes in the variables ν and l have been dropped as they are not involved in the function.

A typical extension to the WSSUS assumption, to simplify the channel model, is to assume a *separable scattering function*:

$$P_S(\nu; l) = \mathcal{K} \cdot P_{DPS}(\nu) \cdot P_{PDP}(l) \tag{3.22}$$

where $\mathcal{K}$ is an irrelevant constant.

3.3.3 Scattering function in MIMO channels

The time correlation of the channel, required to track the channel evolution, can be computed from the Doppler power spectral density function. Hence, this function is an important characteristic of the channel since it provides both a model for time-varying dynamics and a measure of the channel time-variability. The computation of this channel variability, requires any assumption for the users and/or scatters mobility. Using the Clarke's model for flat fading mobile radio propagation scenarios [Cla68], and assuming a separable scattering function, an expression for the Doppler power spectral density function can be derived for the single antenna case (see [Jac74, p.21] or [Rap96, p.180]) as a function of the mobile or scatters velocity v and the carrier frequency f_c:

$$P_S(\nu; l) = \frac{P_{PDP}(l)}{\pi f_d \sqrt{1 - \left(\frac{\nu}{f_d}\right)^2}} \qquad |\nu| < f_d \tag{3.23}$$

where f_d is the maximum normalized Doppler shift: $f_d = \frac{v}{c} f_c T$ (T denotes the sampling period of the channel evolution). From the Doppler power spectral density function, the time correlation of the process is easily determined by the inverse Fourier transform as:

$$R_h(\Delta n; l) = P_{PDP}(l)\, \mathcal{J}_o\left(2\pi f_d \Delta n\right) \tag{3.24}$$

where $\mathcal{J}_o(\cdot)$ is the zero-order Bessel function of the first kind.

The previous U-shaped Clarke's Doppler spectrum, widely used in literature, is only applied under two assumptions:

- The *Direction of Arrival* (DoA) of the propagation waves is uniformly distributed $[-\pi, \pi)$
- Omidirectional antennas with constant gain are used.

Unfortunately the first assumption can not be, in general, considered for MIMO scenarios where the waves are confined in restricted incident angles. In consequence, a more general Doppler power spectral density function has to be computed as a function of the p.d.f. of the *Power Angular Spectrum* (PAS) (see [Jac74, p.21] or [Rap96, p.180]):

$$P_S(\nu; l) = \frac{P_{PDP}(l)}{f_d \sqrt{1 - \left(\frac{\nu}{f_d}\right)^2}} \left[p(\alpha)G(\alpha) + p(-\alpha)G(-\alpha)\right] \qquad |\nu| < f_d \tag{3.25}$$

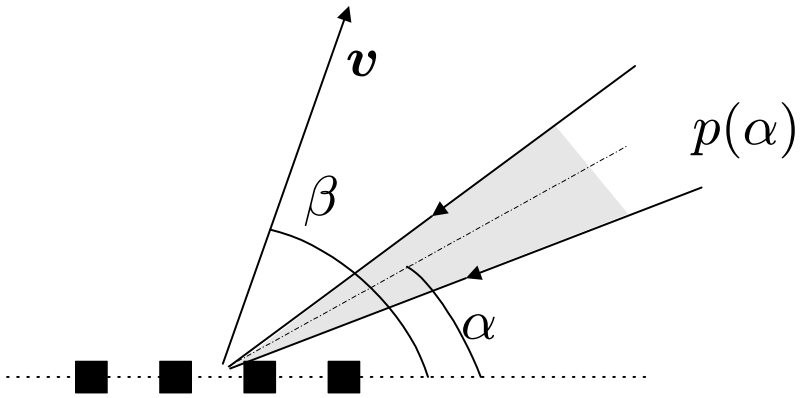


Figure 3.5: Illustration of a scatter impinging the mobile MIMO receiver.

where $p(\alpha)$ is the p.d.f. of the PAS and $G(\alpha)$ is the radiation pattern of the antenna. Note that in the particular case of omnidirectional antennas and rays isotropically distributed, $p(\alpha)G(\alpha) = \frac{1}{2\pi}$ and thus, equation in (3.25) reduces to (3.23).

A study of the Doppler spectrum knowing the PAS was done by [Pet97] (see also [Qin01]) in the case of directional antennas at the base station. The authors concluded that the Doppler spectrum converges to the classical U-Shape when the beamwidth is increased, which is not surprising because under those conditions the PAS tends to be uniform.

A similar analysis is next reproduced for MIMO channels introducing the PAS knowledge (uniform, Gaussian, Laplacian, etc.) into (3.25). Let us assume that the receiver is moving with speed v at an angle β with respect to the endfire, and a cluster of rays is impinging the antenna with a PAS p.d.f. denoted as $p(\alpha)$ as shown in Figure 3.5. Each ray experiences a Doppler shift as a function of the relative angle between β and α, and consequently the incident angle can be isolated as a function of the normalized Doppler shift:

$$\nu = f_d \cos(\beta - \alpha) \Rightarrow \alpha = \beta - \cos^{-1}\left(\frac{\nu}{f_d}\right) \tag{3.26}$$

Substituting (3.26) in (3.25) under the assumption of omnidirectional antennas, an expression for the Doppler power spectral density function, which only depends on the p.d.f. of the power angular spectrum, is derived:

$$P_S(\nu; l) = \frac{P_{PDP}(l)}{f_d\sqrt{1-\left(\frac{\nu}{f_d}\right)^2}}\left[p\left(\beta - \left|\cos^{-1}\left(\frac{\nu}{f_d}\right)\right|\right) + p\left(\beta + \left|\cos^{-1}\left(\frac{\nu}{f_d}\right)\right|\right)\right] \qquad |\nu| < f_d \tag{3.27}$$

Figure 3.6 illustrates the Doppler Spectrum according to previous expression when the impinging scatters are modelled as a Laplacian PAS arriving at the endfire direction (DoA=0°). Several plots are compared for different *Angular Spreads* (AS) and *Directions of Movement* (DoM). Results are compared with the classical U-Shaped Clarke's model. As it can be seen,

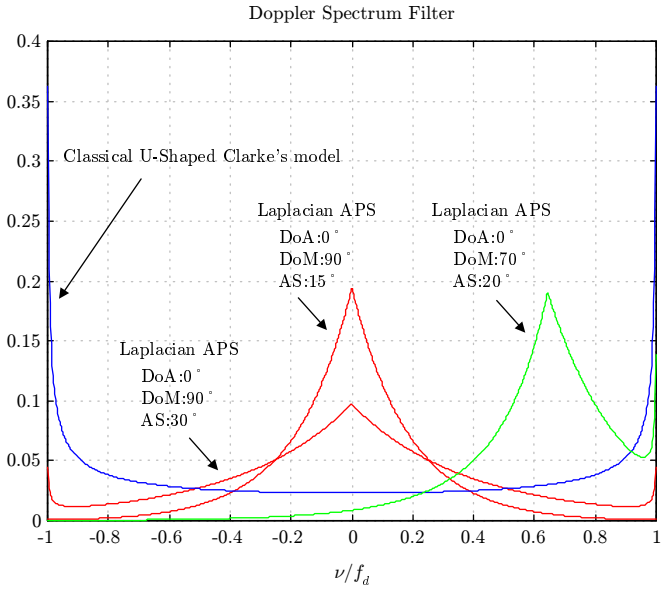


Figure 3.6: Doppler spectra for Laplacian azimuth power spectrum compared with the U-shaped Clarke's model.

when the direction of arrival is constrained in a certain space region (Laplacian PAS), the majority Doppler shifts are confined around a dominant frequency related with the relative angle $(\beta - \alpha)$. Furthermore, the lower the AS is, the narrower the Doppler spectrum becomes.

The above derived expression has assumed a 2-D model, considering that only the azimuth is relevant to model the multiple rays reaching the mobile. Nevertheless, according to [Rep98] for indoor channels not only the azimuth, but also the elevation is relevant and has to be considered to model the whole impinging rays. In accordance, an extended 3-D model should be considered in indoor scenarios. The first author that proposed a three-dimensional multipath scenario was Aulin in [Aul79]. More recently, [Cla97, Vat97] derived a statistical 3-D model, and concluded that assuming uniformly distributed rays in azimuth and elevation, the power spectral density keeps flat between $[-f_d, f_d]$. A generalized Doppler power spectrum for wireless environments considering a propagation model over the whole space (3-D) is derived in [Nar99]. Those general expressions should be managed to determine an accurate Doppler spectrum shape if non-uniformly spatial distributed rays were considered in MIMO indoor scenarios. In such a case, not only the azimuth power spectrum, but also the elevation should be settled, $p(\theta, \phi)$.

Although it has been shown that the classical U-Shaped Clarke's model is not the most appropriate for MIMO scenarios, it will be used as a benchmark channel through this chapter. The performance for more accurate Doppler spectrum shapes induced by multiple antennas will be evaluated later, concluding that U-Shaped Doppler spectrum leads to a worst-case design for the feedback channel.

3.4 Linear channel prediction using an autoregressive model

As outlined in Section 3.1 the design of the transmitter channel tracking scheme is decomposed into two stages: the prediction of the channel evolution and the study of the feedback link. The first stage is further formulated proposing the use a linear predictor based on an autoregressive model of the LTV channel. The second stage, covered in the next section, will use this linear predictor as part of a Kalman. Hence, an accurate selection of the channel predictor will be of paramount importance to ensure a successful design of the return link.

Channel model

The simplicity in the approach of a stochastic model to describe the channel behavior, as well as its capacity of long-range prediction, were the reasons to choose this kind of model over the deterministic one. With this choice, and assuming that the channel time-evolution is described by a lowpass Doppler spectrum, a possible approach in predicting this evolution and capturing the channel dynamics consists on modelling the TV channel as a pth order *Auto-Regressive* (AR) process [Ilt90, Eyc98]. This process satisfies that the present value of the channel state is given by a linear combination of past values plus an error term. Hence, for a specific tap delay l, the impulse response $h^{(t)}(n;l)$ can be modelled as:

$$h^{(t)}(n;l) = \sum_{q=1}^{p} \mathrm{w}_q(l) h^{(t)}(n-q;l) + v(n,l) \tag{3.28}$$

where $\mathrm{w}_q(l)$ are the model coefficients and $v(n,l)$ is a circular complex white Gaussian process with zero mean and variance $\sigma_v^2(l)$. In order to uniquely define the AR model, its parameters must be properly chosen (i.e., the coefficients $\mathrm{w}_q(l)$ and the variance $\sigma_v^2(l)$). These parameters can be fitted to the channel features by solving the Yule-Walker equations [Hay96, p.118]. This set of equations are based on the time correlation sequence of the channel coefficients $R_h(0;l)$, ..., $R_h(q;l)$ and thus, this is the only information required to fit the model.

Note that when classical U-shape Doppler spectrum is assumed, (3.24) shows that the computation of these correlation coefficients is an easy task since only the maximum normalized Doppler shift f_d is necessary. In more complex cases, however, those correlation coefficients should be estimated [Tsa96].

For simplicity in the notation the model in (3.28) has been written for the case of a SISO channel and for a specific tap delay. The extension of the model to the whole MIMO channel is next derived. Using the column vector $\mathbf{h}(n)$ defined in (3.5), which contains the channel impulse response for all transmitting and receiving antenna pairs, the pth order autoregressive process

in (3.28) can be straightforwardly extended as:

$$\mathbf{h}^{(t)}(n) = \sum_{q=1}^{p} \mathbf{W}_q \mathbf{h}^{(t)}(n-q) + \mathbf{v}(n) \tag{3.29}$$

where the AR coefficients are contained in the set of matrices $\{\mathbf{W}_1 \dots \mathbf{W}_p\}$, and $\mathbf{v}(n)$ is a complex white Gaussian process with zero mean and covariance matrix $\mathbf{R_{vv}}$. This model can be simplified assuming WSSUS channels and spatial uncorrelated antennas at both transmitter and receiver sides. Under those assumptions the evolution of each subchannel and tap can be independently modelled, and hence, matrices $\mathbf{W}_q$ and $\mathbf{R_{vv}}$ are diagonal. Moreover, if a separable scattering function with identical power delay profile and Doppler power spectrum functions for all subchannels are also assumed, those matrices are structured as follows:

$$\begin{aligned} \mathbf{W}_q &= \mathbf{I}_{M_T \cdot M_R} \otimes \mathrm{diag}\left\{[\mathrm{w}_q(0) \dots \mathrm{w}_q(L-1)]\right\} \\ \mathbf{R_{vv}} &= \mathbf{I}_{M_T \cdot M_R} \otimes \mathrm{diag}\left\{\left[\sigma_v^2(0) \dots \sigma_v^2(L-1)\right]\right\} \end{aligned} \tag{3.30}$$

Channel predictor

Independently of the assumptions on the channel model it is necessary to design a channel predictor that predicts the future value of the channel given a set of past channel values. For this task we will select, for simplicity, a linear predictor that will compute the predicted channel value $\hat{h}^{(t)}(n;l)$ as:

$$\hat{h}^{(t)}(n;l) = \sum_{q=1}^{p} \mathrm{a}_q(l) \hat{h}^{(t)}(n-q;l) \tag{3.31}$$

where M is the order of the predictor, and $\mathrm{a}_q(l)$ are the coefficients of the predictor.

To optimize the design of those coefficients under a MMSE sense the Wiener-Hopf equations, which require the knowledge of the autocorrelation values of the channel, must be solved [Hay96, p.203]. A simple comparison between those equations, which determine the predictor coefficients, and the Yule-Walker equations, which set the AR coefficients, shows that both sets of equations have the same mathematical form. Hence, when the stochastic process to be predicted obeys a pth order autoregressive model, the best linear predictor, in a MMSE sense, is a pth order predictor whose coefficients take the same values that those of the AR process (i.e., $\mathrm{a}_q(l) = \mathrm{w}_q(l)$). Furthermore, under those conditions, the variance of the predictor is the variance of the white process used in the AR model $\sigma_v^2(l)$.

In real scenarios the channel evolution does not obeys an AR process. In such cases, however, although the AR model is only an approximation, low order AR processes can provide accurate channel tracking performances when the Doppler power spectrum has a narrowband lowpass characteristic. This condition holds when the sampling frequency is significantly higher than the maximum Doppler shift.

	$\mathrm{w}_1(l)$	$\mathrm{w}_2(l)$	$\sigma_v^2(l)$
Random-walk	1	-	$P_{PDP}(l) \cdot 2\left(1 - \mathcal{J}_o\left(2\pi f_d\right)\right)$
AR(1)	$\mathcal{J}_o\left(2\pi f_d\right)$	-	$P_{PDP}(l)\left(1 - \mathcal{J}_o^2\left(2\pi f_d\right)\right)$
AR(2)	$\mathcal{J}_o\left(2\pi f_d\right)\left[\frac{\mathcal{J}_o(2\pi 2 f_d)-1}{\mathcal{J}_o^2(2\pi f_d)-1}\right]$	$\frac{\mathcal{J}_o^2(2\pi f_d)-\mathcal{J}_o(2\pi 2 f_d)}{\mathcal{J}_o^2(2\pi f_d)-1}$	$P_{PDP}(l)\left(1 - \mathcal{J}_o\left(2\pi f_d\right)\mathrm{w}_1(l) - \mathcal{J}_o\left(2\pi 2 f_d\right)\mathrm{w}_2(l)\right)$

Table 3.1: Parameters of the linear channel predictor in (3.31) using random-walk process, first AR process: AR(1) and second order AR process: AR(2).

Hence, the following items can be considered to set the channel model and the linear predictor:

- When no information is available on the channel fluctuations, a simple *random-walk process*, which is modelled assuming that the next channel state equals the previous one, can be considered for slow to moderate LTV channels.
- When more accurate approaches are required, typically, first or second order autoregressive processes are enough to track the channel.

Table 3.1 summarizes the AR parameters/predictor coefficients for the random-walk process and for the AR(1) and AR(2) models under the U-shape Doppler spectrum and separable scattering function assumptions.

3.5 Design of the transmitter channel tracking based on the Kalman filter

In this section we complete the study of the transmitter channel tracking scheme describing the design of the transmitter and the receiver channel predictors using the Kalman filter. The interest in the Kalman filter is motivated by the fact that it is the optimal filter in a minimum variance sense. Moreover, its structure will yield, in a natural way, an interpretation to the minimum amount of information to feedback through the return link.

First, the channel model introduced in the previous section is used to set the Kalman filter equations and variables. These equations will be the used as the basis to predict the channel response using past values of the channel and present observations at the channel output. Later, in the core of this section, it will shown that the prediction error is the minimum amount of information required to track the channel variability at the transmitter. Finally, the last part of the section will discuss the way to transfer this information through the feedback link diminishing the effect of the quantization error.

3.5.1 Statement of the Kalman filter equations

The Kalman filter problem is mathematically formalized from a pair of equations denoted as state equation and measure equation [And79, p.37], [Hay96, p.307]. Those equations are next introduced for the specific case of the linear time-varying channel model. For the sake of clarity in the notation superscript $(\cdot)^K$ will be used to denote the Kalman variables whereas superscript $(\cdot)^{(t)}$ will be used to refer to the channel impulse response.

The *state equation*, which describes the evolution of the state parameter to be tracked, will be obtained from the AR process introduced in the previous section to model the channel dynamics (3.29). Note that when the AR parameters $\{\mathbf{W}_1 \dots \mathbf{W}_p\}$ and the noise covariance $\mathbf{R_{vv}}$ are known, this channel model can be written as:

$$\begin{bmatrix} \mathbf{h}^{(t)}(n) \\ \mathbf{h}^{(t)}(n-1) \\ \vdots \\ \mathbf{h}^{(t)}(n-p+1) \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 & \dots & \mathbf{W}_p \\ \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{h}^{(t)}(n-1) \\ \mathbf{h}^{(t)}(n-2) \\ \vdots \\ \mathbf{h}^{(t)}(n-p) \end{bmatrix} + \begin{bmatrix} \mathbf{v}(n) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.32}$$

Comparing this equation with the structure of the Kalman state equation, it is shown that (3.32) can be understood as the Kalman state equation. Specifically, if the state vector $\mathbf{h}^K(n)$, which typically contains the information to be estimated, the state transition matrix $\mathbf{W}^K$, which relates the state of the system at times n and $n-1$, and the noise process $\mathbf{v}^K(n)$, which contains the uncertainty of the model, are defined as follows:

$$\mathbf{h}^K(n) = \begin{bmatrix} \mathbf{h}^{(t)}(n) \\ \mathbf{h}^{(t)}(n-1) \\ \vdots \\ \mathbf{h}^{(t)}(n-p+1) \end{bmatrix} \;;\; \mathbf{W}^K = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 & \dots & \mathbf{W}_p \\ \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} \end{bmatrix} \;;\; \mathbf{v}^K(n) = \begin{bmatrix} \mathbf{v}(n) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.33}$$

the Kalman state equation can be written as:

$$\textbf{State equation}: \qquad \mathbf{h}^K(n) = \mathbf{W}^K \mathbf{h}^K(n-1) + \mathbf{v}^K(n) \tag{3.34}$$

Of special interest is the Kalman state vector $\mathbf{h}^K(n)$. Note that this vector contains in its first rows the channel impulse response to be predicted/estimated, i.e., $\mathbf{h}^{(t)}(n)$. Hence, in each step special attention will be focused on these rows since contain the information required to design later the linear transmitter $\mathbf{F}(n)$.

On the other hand the *measurement equation*, which provides a noisy observation of a linear transformation applied over the state vector $\mathbf{h}^K(n)$, can be obtained from the channel output $\mathbf{r}(n)$. According to equation (3.6) the vector $\mathbf{r}(n)$ can be expressed as:

$$\mathbf{r}(n) = [\mathbf{S}(n) \;\; \mathbf{0} \ldots \mathbf{0}] \begin{bmatrix} \mathbf{h}^{(t)}(n) \\ \mathbf{h}^{(t)}(n-1) \\ \vdots \\ \mathbf{h}^{(t)}(n-p+1) \end{bmatrix} + \mathbf{n}(n) \tag{3.35}$$

Then, defining the measurement matrix $\mathbf{S}^K(n)$ as:

$$\mathbf{S}^K(n) = [\mathbf{S}(n) \;\; \mathbf{0} \ldots \mathbf{0}] \tag{3.36}$$

we may write the measurement equation as follows:

$$\textbf{Measurement equation}: \qquad \mathbf{r}(n) = \mathbf{S}^K(n)\mathbf{h}^K(n) + \mathbf{n}(n) \tag{3.37}$$

Notice that the transmitted symbols $\mathbf{s}(n)$ are needed to generate the observation matrix $\mathbf{S}^K(n)$ according to the structure defined in Section 3.2. Thus, the information symbols $\mathbf{x}(n)$ (and the linear transformation $\mathbf{F}(n)$) must be known at the receiver in order to generate this matrix. In tracking mode, those symbols can be obtained at the output of the decoder. However, during the acquisition period symbols at the receiver are not reliable, and hence, any auxiliary deterministic blind equalizer or a training sequence might be introduced.

The Kalman model must also consider information about the uncertainties in the state equation (vector $\mathbf{v}^K(n)$) and the measurement equation (vector $\mathbf{n}(n)$). This uncertainties are described by zero mean, white Gaussian processes with known covariances. Specifically, the noise vectors $\mathbf{v}^K(n)$ and $\mathbf{n}(n)$ are statistically independent, so that $E\left\{\mathbf{v}^K(n)\mathbf{n}(m)^H\right\} = \mathbf{0}$ for all n and m, the covariance of the measurement noise is $E\left\{\mathbf{n}(n)\mathbf{n}(m)^H\right\} = \mathbf{R_n}\delta(m-n)$ where $\mathbf{R_n} = \sigma_n^2\mathbf{I}$, and the covariance of the noise process is given by $E\left\{\mathbf{v}^K(n)\mathbf{v}^K(m)^H\right\} = \mathbf{R}^K_{\mathbf{vv}}\delta(m-n)$ where:

$$\mathbf{R}^K_{\mathbf{vv}} = \begin{bmatrix} \mathbf{R_{vv}} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \end{bmatrix} \tag{3.38}$$

and the covariance matrix $\mathbf{R_{vv}}$ was defined in (3.30). Both, the noise process and the measurement noise are assumed to be stationary. Hence, its covariances do not depend on the time index n. This assumption will allow us to make use of the steady-state solution of the Kalman filter and it will also allow to compute some matrices before the Kalman filter starts to operate as it is referenced at the end of Section 3.5.3.

It is worth to note that the covariance matrix $\mathbf{R_{vv}}$, given by the process noise in (3.32), only considers the uncertainty in the AR model of the channel. A more accurate description of the state equation (3.34) might consider the quantization error induced by the quantization of the information transmitted through the feedback link modifying the definition of $\mathbf{v}^K(n)$ in (3.33). For the sake of simplicity, however, we will ignore this error at this point.

3.5.2 Channel prediction using the Kalman state equation

The first step in the Kalman filter aims to predict the CSI at time n using all the measurements up till time $n-1$. This task is described in the state equation (3.32) in which the transition matrix $\mathbf{W}^K$ establishes, in the absence of new channel observations, a relationship between the state vector estimated at time $n-1$ and the future value at time n. Using the same notation that [And79] we shall denote these vectors by $\hat{\mathbf{h}}^K_{n-1|n-1}$ and $\hat{\mathbf{h}}^K_{n|n-1}$, respectively. The prediction error equation that describes this step, is written as:

$$\hat{\mathbf{h}}^K_{n|n-1} = \mathbf{W}^K \, \hat{\mathbf{h}}^K_{n-1|n-1} \tag{3.39}$$

Note that this equation simply describes, in a matrix notation, the channel predictor introduced in (3.31) in which the past values of the channel impulse response have been substituted by their estimations at time $n-1$, and the predictor coefficients are the coefficients of the AR model (see Section 3.4 for a formal justification of this last assignment).

Once the channel has been predicted, it will be necessary to know how good this prediction is. The parameter that measures the quality of this prediction is the covariance matrix $\mathbf{\Sigma}_{n|n-1}$, which is given by:

$$\begin{aligned} \mathbf{\Sigma}_{n|n-1} &= E\left\{ \left(\mathbf{h}^K(n) - \hat{\mathbf{h}}^K_{n|n-1}\right) \left(\mathbf{h}^K(n) - \hat{\mathbf{h}}^K_{n|n-1}\right)^H \right\} \\ &= \mathbf{W}^K \, \mathbf{\Sigma}_{n-1|n-1} \mathbf{W}^{K\,H} + \mathbf{R}^K_{\mathbf{vv}} \end{aligned} \tag{3.40}$$

where $\mathbf{\Sigma}_{n-1|n-1}$, in accordance with the notation used, is the covariance of the estimation error of the vector $\hat{\mathbf{h}}^K_{n-1|n-1}$. Note that the covariance in (3.40) is formed by two terms. The second term contain the uncertainty due to noise process, which is simply the MSE of the used channel predictor (see Section 3.4). Of more interest is, however, the first term. Note that this term adds to the prediction error the propagated error due to the uncertainty in the estimation of the state vector $\hat{\mathbf{h}}^K_{n-1|n-1}$.

Thanks to the structure of the Kalman state vector $\hat{\mathbf{h}}^K(n)$, which stacks in a column vector the last pth values of the channel impulse response (see equation (3.33)), the covariance matrix $\mathbf{\Sigma}_{n|n-1}$ can be structured as a block matrix divided as:

$$\mathbf{\Sigma}_{n|n-1} = \begin{bmatrix} \mathbf{\Sigma}^{1,1}_{n|n-1} & \mathbf{\Sigma}^{2,1}_{n|n-1} & \cdots & \mathbf{\Sigma}^{p,1}_{n|n-1} \\ \mathbf{\Sigma}^{1,2}_{n|n-1} & \mathbf{\Sigma}^{2,2}_{n|n-1} & \ddots & \mathbf{\Sigma}^{p,2}_{n|n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}^{1,p}_{n|n-1} & \mathbf{\Sigma}^{2,p}_{n|n-1} & \cdots & \mathbf{\Sigma}^{p,p}_{n|n-1} \end{bmatrix} \tag{3.41}$$

where each one of the blocks $\mathbf{\Sigma}^{i,j}_{n|n-1}$ is given by:

$$\mathbf{\Sigma}^{i+1,j+1}_{n|n-1} = E\left\{ \left(\mathbf{h}^{(t)}(n-i) - \hat{\mathbf{h}}^{(t)}_{n|n-1}(n-i) \right) \left(\mathbf{h}^{(t)}(n-j) - \hat{\mathbf{h}}^{(t)}_{n|n-1}(n-j) \right)^H \right\} \tag{3.42}$$

This structure provided to the covariance matrix will be used in the next sections to simplify and facilitate the interpretation of some of the derived equations.

3.5.3 Channel estimation using the innovation

When the channel state vector at time n can be observed by means of the measure $\mathbf{r}(n)$, the prediction of the vector $\hat{\mathbf{h}}^K_{n|n-1}$ can be refined making use of that observation. The key to improve the prediction is to make use of the new information contained in the measure at time n which was not available from the observations up till time $n-1$. The fact is that the observation vector $\mathbf{r}(n)$ does not contributes by itself with all the information of vector $\mathbf{h}^K(n)$ since all the predictable information was already contained in $\hat{\mathbf{h}}^K_{n|n-1}$ thanks to the observations up till time $n-1$. In consequence, this new information provided by the present measure, named as the *innovation*, is defined by the prediction error between the real measure $\mathbf{r}(n)$ and the predicted value of this vector at time n using the past observations. Hence:

$$\begin{aligned} \boldsymbol{\alpha}(n) &= \mathbf{r}(n) - \hat{\mathbf{r}}^K_{n|n-1} \\ &= \mathbf{r}(n) - \mathbf{S}^K(n)\, \hat{\mathbf{h}}^K_{n|n-1} \end{aligned} \tag{3.43}$$

This equation is better understood if the structure of matrix $\mathbf{S}^K(n)$ and vector $\hat{\mathbf{h}}^K_{n|n-1}$ are used. According to the definition of these two variables (3.33), the equation for the innovation is simplified as follows:

$$\boldsymbol{\alpha}(n) = \mathbf{r}(n) - \mathbf{S}(n)\, \hat{\mathbf{h}}^{(t)}_{n|n-1} \tag{3.44}$$

Note that this new equation reflects the fact that the innovation only provides information of the channel impulse response at time n since the observation $\mathbf{r}(n)$ only depends on $\mathbf{h}^{(t)}(n)$ and not on the past values of the channel.

Next, we present some properties of the innovation whose derivations are omitted because of its simplicity [Hay96, p.307]:

Property 1 The innovation associated with the observed random variable $\mathbf{r}(n)$ is orthogonal to the past observations $\mathbf{r}(1)\ \ldots\ \mathbf{r}(n-1)$. Thus:

$$E\left\{\boldsymbol{\alpha}(n)\mathbf{r}(m)\right\} = \mathbf{0} \qquad 1 \leq m \leq n-1$$

Property 2 The innovations $\boldsymbol{\alpha}(1)\ \ldots\ \boldsymbol{\alpha}(n)$ are orthogonal to each other. Thus:

$$E\left\{\boldsymbol{\alpha}(n)\boldsymbol{\alpha}(m)\right\} = \mathbf{0} \qquad n \neq m$$

Property 3 There is a one-to-one correspondence between the observation and the innovations in the sense that the set of observation vectors $[\mathbf{r}(1)\ \ldots\ \mathbf{r}(n)]$ and the set of innovations $[\boldsymbol{\alpha}(1)\ \ldots\ \boldsymbol{\alpha}(n)]$ share the same information. In consequence, the Kalman state vector $\mathbf{h}^K(n)$ can be estimated indistinctly for any of the two sets of data. This property is the basis to deduce the Kalman filter equations in [Hay96, p.304].

Property 4 The innovation is a zero mean random process. Hence:

$$E\left\{\boldsymbol{\alpha}(n)\right\} = \mathbf{0}$$

Property 5 The covariance of the innovation is given by:

$$E\left\{\boldsymbol{\alpha}(n)\boldsymbol{\alpha}(n)^H\right\} = \mathbf{S}^K(n)\boldsymbol{\Sigma}_{n|n-1}\mathbf{S}^K(n)^H + \mathbf{R_n}$$

Once the innovation has been introduced, this parameter will be used to update the prediction with the measurement $\mathbf{r}(n)$. The equation that refines the prediction is given by:

$$\hat{\mathbf{h}}^K_{n|n} = \hat{\mathbf{h}}^K_{n|n-1} + \mathbf{K}(n)\ \boldsymbol{\alpha}(n) \tag{3.45}$$

Note that updated value of the state vector is obtained by adding to the predicted Kalman state $\hat{\mathbf{h}}^K_{n|n-1}$ the new information provided by the innovation $\boldsymbol{\alpha}(n)$. This information is given by the term $\mathbf{K}(n)\ \boldsymbol{\alpha}(n)$, where the matrix $\mathbf{K}(n)$, known as the Kalman gain, is defined as follows:

$$\mathbf{K}(n) = \boldsymbol{\Sigma}^H_{n|n-1}\ \mathbf{S}^K(n)^H \left(\mathbf{S}^K(n)\ \boldsymbol{\Sigma}_{n|n-1}\ \mathbf{S}^K(n)^H + \mathbf{R_n}\right)^{-1} \tag{3.46}$$

and using the structure of the covariance matrix $\boldsymbol{\Sigma}_{n|n-1}$ given in (3.41), and the fact that the measurement matrix $\mathbf{S}^K(n)$ contains a large number of zeros (3.36), the Kalman gain can be simplified:

$$\mathbf{K}(n) = \begin{bmatrix} \boldsymbol{\Sigma}^{1,1}_{n|n-1} \\ \boldsymbol{\Sigma}^{1,2}_{n|n-1} \\ \vdots \\ \boldsymbol{\Sigma}^{1,p}_{n|n-1} \end{bmatrix} \mathbf{S}(n)^H \left(\mathbf{S}(n)\ \boldsymbol{\Sigma}^{1,1}_{n|n-1}\ \mathbf{S}(n)^H + \mathbf{R_n}\right)^{-1} \tag{3.47}$$

This simplification in the Kalman gain allows to uncouple the correction equation (3.45) into a set of p independent equations. Hence, the estimated channel vector $\hat{\mathbf{h}}^{(t)}_{n|n}$ can be written as follows:

$$\begin{aligned}
\hat{\mathbf{h}}^{(t)}_{n|n}(n) &= \hat{\mathbf{h}}^{(t)}_{n|n-1}(n) + \boldsymbol{\Sigma}^{1,1}_{n|n-1}\,\mathbf{S}(n)^H\left(\mathbf{S}(n)\,\boldsymbol{\Sigma}^{1,1}_{n|n-1}\,\mathbf{S}(n)^H+\mathbf{R_n}\right)^{-1}\boldsymbol{\alpha}(n)\\
\hat{\mathbf{h}}^{(t)}_{n|n}(n-1) &= \hat{\mathbf{h}}^{(t)}_{n|n-1}(n-1) + \boldsymbol{\Sigma}^{1,2}_{n|n-1}\,\mathbf{S}(n)^H\left(\mathbf{S}(n)\,\boldsymbol{\Sigma}^{1,1}_{n|n-1}\,\mathbf{S}(n)^H+\mathbf{R_n}\right)^{-1}\boldsymbol{\alpha}(n)\\
&\vdots\\
\hat{\mathbf{h}}^{(t)}_{n|n}(n-p+1) &= \hat{\mathbf{h}}^{(t)}_{n|n-1}(n-p+1) + \boldsymbol{\Sigma}^{1,p}_{n|n-1}\,\mathbf{S}(n)^H\left(\mathbf{S}(n)\,\boldsymbol{\Sigma}^{1,1}_{n|n-1}\,\mathbf{S}(n)^H+\mathbf{R_n}\right)^{-1}\boldsymbol{\alpha}(n)
\end{aligned} \tag{3.48}$$

Let us now comment the significance of this set of equations. As it was previously shown, the innovation only provides information of the CSI at time n. This fact is clearly displayed in the first equation of the previous set, in which the predicted value of the CSI $\hat{\mathbf{h}}^{(t)}_{n|n-1}(n)$ is updated thanks to the innovation to get $\hat{\mathbf{h}}^{(t)}_{n|n}(n)$. More surprising can be the result given by the rest of equations. As it is shown, the innovation at time n is also used to update the past channel impulse responses $\hat{\mathbf{h}}^{(t)}_{n|n-1}(n-1)\ \ldots\ \hat{\mathbf{h}}^{(t)}_{n|n-1}(n-p+1)$. This relationship between the innovation and the past channel responses is due to the dependence of the present CSI with the past values of the channel (3.34). Those impulse responses, although they are not required to design the linear transformation $\mathbf{F}(n)$, they are necessary to predict the channel response at time $n+1$ (see equation (3.39)) and thus, its update at both, the transmitter and the receiver, is mandatory.

The last step in the Kalman filter consist in computing the covariance of the estimation error of vector $\hat{\mathbf{h}}^{K}_{n|n}$. This covariance is given by:

$$\begin{aligned}
\boldsymbol{\Sigma}_{n|n} &= E\left\{\left(\mathbf{h}^K(n)-\hat{\mathbf{h}}^K_{n|n}\right)\left(\mathbf{h}^K(n)-\hat{\mathbf{h}}^K_{n|n}\right)^H\right\}\\
&= \boldsymbol{\Sigma}_{n|n-1}-\mathbf{K}(n)\,\mathbf{S}^K(n)\,\boldsymbol{\Sigma}_{n|n-1}\\
&= \boldsymbol{\Sigma}_{n|n-1}-\mathbf{K}(n)\,\mathbf{S}(n)\left[\boldsymbol{\Sigma}^{1,1}_{n|n-1}\ \boldsymbol{\Sigma}^{1,2}_{n|n-1}\ldots\boldsymbol{\Sigma}^{1,p}_{n|n-1}\right]
\end{aligned} \tag{3.49}$$

where the last equality has used the structure of matrices $\mathbf{S}^K(n)$ and $\boldsymbol{\Sigma}_{n|n-1}$.

With the computation of (3.49) finishes the nth iteration of the Kalman filter. At this point, hence, the algorithm is ready to start the next iteration by computing from (3.39) a new prediction of the CSI at time $n+1$ with the observation at time n, i.e.,$\hat{\mathbf{h}}^K_{n+1|n}$. Table 3.2 summarizes the Kalman equations introduced in these sections as well as the initial conditions for the iterative algorithm. Note that in our concerning problem, the computation of the prediction error covariance $\boldsymbol{\Sigma}_{n|n-1}$, the Kalman gain $\mathbf{K}(n)$ and the correction error covariance $\boldsymbol{\Sigma}_{n|n}$ do not depend on the observation $\mathbf{r}(n)$. Hence, these three variables can be computed and stored before the filter starts to operate, simplifying, in consequence, the computational complexity.

Initial Conditions:	$\boldsymbol{\Sigma}_{0\|0}$	$=$	$\alpha\,\mathbf{I} \quad ; \quad \alpha \gg 1$
	$\hat{\mathbf{h}}^K_{0\|0}$	given by:	$h^{(t)}_{i,j}(0;l) = \delta(l) \quad i = 1 \ldots M_T \quad j = 1 \ldots M_R$
			$h^{(t)}_{i,j}(n;l) = 0 \qquad n < 0$
Iteration n:			
Prediction equation:	$\hat{\mathbf{h}}^K_{n\|n-1}$	$=$	$\mathbf{W}^K\,\hat{\mathbf{h}}^K_{n-1\|n-1}$
Prediction error covariance:	$\boldsymbol{\Sigma}_{n\|n-1}$	$=$	$\mathbf{W}^K\,\boldsymbol{\Sigma}_{n-1\|n-1}{\mathbf{W}^K}^H + \mathbf{R}^K_{\mathbf{vv}}$
Innovation:	$\boldsymbol{\alpha}(n)$	$=$	$\mathbf{r}(n) - \mathbf{S}^K(n)\,\hat{\mathbf{h}}^K_{n\|n-1}$
Kalman gain:	$\mathbf{K}(n)$	$=$	$\boldsymbol{\Sigma}^H_{n\|n-1}\,\mathbf{S}^K(n)^H\left(\mathbf{S}^K(n)\,\boldsymbol{\Sigma}_{n\|n-1}\,\mathbf{S}^K(n)^H + \mathbf{R}_{\mathbf{n}}\right)^{-1}$
Correction equation:	$\hat{\mathbf{h}}^K_{n\|n}$	$=$	$\hat{\mathbf{h}}^K_{n\|n-1} + \mathbf{K}(n)\,\boldsymbol{\alpha}(n)$
Correction error covariance:	$\boldsymbol{\Sigma}_{n\|n}$	$=$	$\boldsymbol{\Sigma}_{n\|n-1} - \mathbf{K}(n)\,\mathbf{S}^K(n)\,\boldsymbol{\Sigma}_{n\|n-1}$

Table 3.2: Summary of the Kalman filter equations.

3.5.4 Feedback link

The application of the previous Kalman state representation for the design of the transmitter channel tracking scheme described in Section 3.1.1, needs to identify the minimum amount of information to be fed back to the transmitter. Note that the prediction equation (3.39) can be computed at both, transmitter and receiver. However, the correction equation (3.45), which makes use of the unpredictable information of the channel evolution, can only be computed at the receiver since it is based on the innovation $\boldsymbol{\alpha}(n)$. In accordance, the information contained on the innovation must be fed back to the transmitter to compensate at this side of the link the differences between the real and the predicted channel values.

In order to determine which is the minimum amount of information to feedback we enumerate, next, a set of conditions that must accomplish:

Condition 1 The number of elements to be fed back must be the same that the number of degrees of freedom of the problem. In our particular case this number is given by the length of the CSI vector $\mathbf{h}^{(t)}(n)$, i.e., $L \cdot M_T \cdot M_R$ elements.

Condition 2 The elements to be fed back must be zero mean. Otherwise, it is obvious that unnecessary information would be transmitted.

Condition 3 The elements to be fed back must be uncorrelated. Otherwise, it is obvious that redundant information would be transmitted.

Condition 4 In order to ensure the channel response can be uniquely identified, the matrix that contains the transmitted symbols $\mathbf{S}(n) \in \mathbb{C}^{(M_s+L)M_R \times M_T M_R L}$ must be a *tall* matrix, i.e., $(M_s + L)M_R \geq M_T M_R L$. Hence, the number of elements to be transmitted M_s must be appropriately chosen.

According to equation (3.45), it seems that the innovation $\boldsymbol{\alpha}(n)$ must be transmitted through the feedback channel link to appropriately track the channel at the transmitter. The feedback of all the elements of $\boldsymbol{\alpha}(n)$, however, does not constitutes the minimum information required to update the CSI since the length of this vector, $(M_s + L)M_R$, exceeds the number of elements set in Condition 1.

The term that updates the channel prediction in (3.45), i.e., $\mathbf{K}(n)\boldsymbol{\alpha}(n)$, could be also proposed as a candidate to be fed back. Nevertheless, its length does not satisfies Condition 1 either, and hence its validity is also refused.

A subset of the elements in $\mathbf{K}(n)\boldsymbol{\alpha}(n)$, however, could provide this minimum amount of information. Having in mind the structure in $\mathbf{h}^K(n)$ (see equation (3.33)) it can be seen that the first $M_T M_R L$ elements in $\mathbf{K}(n)\boldsymbol{\alpha}(n)$ are related with the rest of elements. This relationship is clearly displayed in the set of equations (3.48), where the correction equation has been uncoupled into p independent equations. A simple inspection of these equations shows that there is a part of the expression common to all the equations, $\mathbf{S}(n)^H \left(\mathbf{S}(n)\boldsymbol{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H + \mathbf{R_n}\right)^{-1} \boldsymbol{\alpha}(n)$, and another part obtained from the covariance matrix $\boldsymbol{\Sigma}_{n|n}$ different for each equation. Whereas the common part depends on the innovation $\boldsymbol{\alpha}(n)$, and thus it can only be computed at the receiver, the other part does not depend on the observation, and thus it can be computed both at the transmitter and at the receiver. Hence, if only the common term to all the equations, which fulfils Condition 1, were fed back to the transmitter, all the elements in $\mathbf{h}^K(n)$ could be satisfactory updated. This term, however, lacks of a clear interpretation. Alternatively, the whole updating term in the first equation, which also satisfies Condition 1, could be regarded as the vector that defines the prediction error of the channel impulse response at time n.

Defining this vector as:

$$\mathbf{e}(n) = \boldsymbol{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H \left[\mathbf{S}(n)\boldsymbol{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H + \mathbf{R_n}\right]^{-1} \boldsymbol{\alpha}(n) \tag{3.50}$$

the set of equations in (3.48) could be rewritten as:

$$\begin{aligned}
\hat{\mathbf{h}}^{(t)}_{n|n}(n) &= \hat{\mathbf{h}}^{(t)}_{n|n-1}(n) + \mathbf{e}(n) \\
\hat{\mathbf{h}}^{(t)}_{n|n}(n-1) &= \hat{\mathbf{h}}^{(t)}_{n|n-1}(n-1) + \boldsymbol{\Sigma}^{1,2}_{n|n-1}\left(\boldsymbol{\Sigma}^{1,1}_{n|n-1}\right)^{-1}\mathbf{e}(n) \\
&\vdots \\
\hat{\mathbf{h}}^{(t)}_{n|n}(n-p+1) &= \hat{\mathbf{h}}^{(t)}_{n|n-1}(n-p+1) + \boldsymbol{\Sigma}^{1,p}_{n|n-1}\left(\boldsymbol{\Sigma}^{1,1}_{n|n-1}\right)^{-1}\mathbf{e}(n)
\end{aligned} \tag{3.51}$$

Although we have concluded that the elements in $\mathbf{e}(n)$ contain all the information required to update the Kalman state, we have not proven yet that this vector contains the minimum amount of information to be transferred across the feedback channel. To prove this characteristic, it is necessary to show that the vector $\mathbf{e}(n)$ satisfies Condition 2 and Condition 3.

P1: The prediction error $\mathbf{e}(n)$ has zero mean

Proof. First, it can seen that Condition 2 is accomplished by computing the mean of the vector that can be obtained as:

$$E\{\mathbf{e}(n)\} = \mathbf{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H \left[\mathbf{S}(n)\mathbf{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H + \mathbf{R_n}\right]^{-1} E\{\boldsymbol{\alpha}(n)\} = \mathbf{0} \tag{3.52}$$

where the Property 4 of the innovation $E\{\boldsymbol{\alpha}(n)\} = \mathbf{0}$ has been used. ■

P2: The elements of the prediction error $\mathbf{e}(n)$ are uncorrelated

Proof. This condition can be proven showing that the covariance matrix is diagonal (under some reasonable assumptions). This covariance matrix is derived from:

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{e}(n)} &= E\left\{\mathbf{e}(n)\mathbf{e}(n)^H\right\} \\
&= \mathbf{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H \left[\mathbf{S}(n)\mathbf{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H + \mathbf{R_n}\right]^{-1} E\left\{\boldsymbol{\alpha}(n)\boldsymbol{\alpha}(n)^H\right\} \times \\
&\qquad \left[\mathbf{S}(n)\mathbf{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H + \mathbf{R_n}\right]^{-1} \mathbf{S}(n)\mathbf{\Sigma}^{1,1}_{n|n-1} \\
&= \mathbf{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H \left(\mathbf{S}(n)\mathbf{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H + \sigma_n^2\mathbf{I}\right)^{-1} \mathbf{S}(n)\mathbf{\Sigma}^{1,1}_{n|n-1}
\end{aligned} \tag{3.53}$$

where the last equality has been obtained plugging the covariance of the innovation (Property 5) into the previous expression. Next, applying the matrix inversion lema, the covariance matrix can be written as:

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{e}(n)} &= \mathbf{\Sigma}^{1,1}_{n|n-1}\frac{\mathbf{S}(n)^H\mathbf{S}(n)}{\sigma_n^2}\mathbf{\Sigma}^{1,1}_{n|n-1} - \\
&\qquad \mathbf{\Sigma}^{1,1}_{n|n-1}\frac{\mathbf{S}(n)^H\mathbf{S}(n)}{\sigma_n^2}\left[\left(\mathbf{\Sigma}^{1,1}_{n|n-1}\right)^{-1} + \frac{\mathbf{S}(n)^H\mathbf{S}(n)}{\sigma_n^2}\right]^{-1}\frac{\mathbf{S}(n)^H\mathbf{S}(n)}{\sigma_n^2}\mathbf{\Sigma}^{1,1}_{n|n-1}
\end{aligned} \tag{3.54}$$

Note that the contribution of the transmitted symbols into the covariance matrix is given by $\mathbf{S}(n)^H\mathbf{S}(n)$. This matrix can be shown to be diagonal both when a training sequence and when i.i.d. data symbols are transmitted. In the case of training sequences this is a common constraint for the design of the sequences [Tun01], whereas in the case of data symbols, this assumption applies asymptotically for long data sequences thanks to the independence of the symbols. Hence, assuming $\frac{\mathbf{S}(n)^H\mathbf{S}(n)}{\sigma_n^2} = SNR \cdot M_s\,\mathbf{I}$, where M_s is the

number of symbols transmitted in a block, the covariance $\mathbf{\Sigma}_{\mathbf{e}(n)}$ becomes diagonal, and it can be expressed as:

$$\mathbf{\Sigma}_{\mathbf{e}(n)} = SNR \cdot M_s \mathbf{\Sigma}^{1,1}_{n|n-1} \left(\mathbf{\Sigma}^{1,1}_{n|n-1} SNR \cdot M_s + \mathbf{I} \right)^{-1} \mathbf{\Sigma}^{1,1}_{n|n-1} \tag{3.55}$$

■

At this point, once we have shown that Condition 2 and Condition 3 are accomplished, it can be concluded that the elements in vector $\mathbf{e}(n)$ contain the minimum information required to compute the correction equation (3.45), and to track the LTV channel at the transmitter site, with independence of the order of the autoregressive model.

3.5.5 Application to closed-loop schemes

Once the minimum amount of information has been derived, Figure 3.7 and Tables 3.3 and 3.4 briefly summarize the procedure to implement the proposed transmitter channel tracking algorithm in closed-loop schemes. Note that both, the transmitter and the receiver, work with the same CSI (i.e., $\hat{\mathbf{h}}^{(t)}_{R\,n|n-1} = \hat{\mathbf{h}}^{(t)}_{T\,n|n-1}$ and also $\hat{\mathbf{h}}^{(t)}_{R\,n|n} = \hat{\mathbf{h}}^{(t)}_{T\,n|n}$). Although the receiver has whole information on vector $\mathbf{e}(n)$, a quantized version of this vector is used to update the channel prediction (see the 4th step in Table 3.3) in order to monitor the same CSI that the transmitter. Otherwise, the small differences between the estimated channel at the transmitter and at the receiver due to the cumulative quantization error will induce different CSI's at both sides of the link.

Channel estimation error at the receiver. According to the procedure proposed in Table 3.3 to estimate the CSI at the receiver, the channel estimation error at this side can be obtained from the correction error covariance in (3.49) as:

$$\mathbf{\Sigma}_{Rx}(n) = E\left\{ \left(\mathbf{h}^{(t)}(n) - \hat{\mathbf{h}}^{(t)}_{R\,n|n} \right) \left(\mathbf{h}^{(t)}(n) - \hat{\mathbf{h}}^{(t)}_{R\,n|n} \right)^H \right\} = \mathbf{\Sigma}^{1,1}_{n|n} \tag{3.56}$$

where $\mathbf{\Sigma}^{1,1}_{n|n}$, using the same matrix partition that for the covariance $\mathbf{\Sigma}_{n|n-1}$ in (3.41), is a block of the covariance matrix $\mathbf{\Sigma}_{n|n}$. Hence, using this structure for both, $\mathbf{\Sigma}_{n|n}$ and $\mathbf{\Sigma}_{n|n-1}$, and according to (3.49), it can be shown that $\mathbf{\Sigma}^{1,1}_{n|n}$ is given by:

$$\mathbf{\Sigma}^{1,1}_{n|n} = \mathbf{\Sigma}^{1,1}_{n|n-1} - \mathbf{\Sigma}^{1,1}_{n|n-1} \mathbf{S}(n)^H \left(\mathbf{S}(n) \mathbf{\Sigma}^{1,1}_{n|n-1} \mathbf{S}(n)^H + \sigma^2_n \mathbf{I} \right)^{-1} \mathbf{S}(n) \mathbf{\Sigma}^{1,1}_{n|n-1} \tag{3.57}$$

Next, we summarize some easy, but lengthy manipulations that apply three times the matrix inversion lema, and assume $\frac{\mathbf{S}(n)^H \mathbf{S}(n)}{\sigma^2_n} = SNR \cdot M_s\ \mathbf{I}$ in the second equality (as argued in the previous section, this is a realistic assumption both for long data sequences and for training

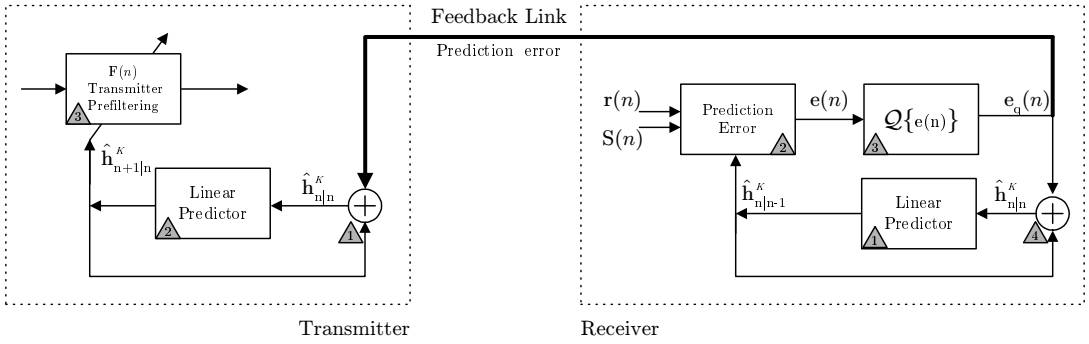


Figure 3.7: Transmitter channel tracking algorithm. Block diagram using the Kalman filter. The numbers into the boxes indicate the order in which the operations are done.

sequences). From (3.57) we have:

$$\begin{aligned}
\boldsymbol{\Sigma}^{1,1}_{n|n} &= \boldsymbol{\Sigma}^{1,1}_{n|n-1} - \frac{\boldsymbol{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H}{\sigma_n^2}\left[\mathbf{I} - \mathbf{S}(n)\left(\left(\boldsymbol{\Sigma}^{1,1}_{n|n-1}\right)^{-1} + \frac{\mathbf{S}(n)^H\mathbf{S}(n)}{\sigma_n^2}\right)^{-1}\frac{\mathbf{S}(n)^H}{\sigma_n^2}\right]\mathbf{S}(n)\boldsymbol{\Sigma}^{1,1}_{n|n-1} \\
&= \boldsymbol{\Sigma}^{1,1}_{n|n-1} + SNR\cdot M_s\boldsymbol{\Sigma}^{1,1}_{n|n-1}\left(\mathbf{I} + \left(SNR\cdot M_s\boldsymbol{\Sigma}^{1,1}_{n|n-1}\right)^{-1}\right)^{-1} SNR\cdot M_s\boldsymbol{\Sigma}^{1,1}_{n|n-1} \\
&\qquad - SNR\cdot M_s\boldsymbol{\Sigma}^{1,1}_{n|n-1}\boldsymbol{\Sigma}^{1,1}_{n|n-1} \\
&= \boldsymbol{\Sigma}^{1,1}_{n|n-1} - \boldsymbol{\Sigma}^{1,1}_{n|n-1}\left(\boldsymbol{\Sigma}^{1,1}_{n|n-1} + \frac{\mathbf{I}}{SNR\cdot M_s}\right)^{-1}\boldsymbol{\Sigma}^{1,1}_{n|n-1} \\
&= \left[SNR\cdot M_s\,\mathbf{I} + \left(\boldsymbol{\Sigma}^{1,1}_{n|n-1}\right)^{-1}\right]^{-1}
\end{aligned} \tag{3.58}$$

From this result, and observing the dependence between the covariances $\boldsymbol{\Sigma}_{n|n-1}$ and $\boldsymbol{\Sigma}_{n-1|n-1}$ (see Table 3.2), it can be concluded that the covariance matrix, which sets the channel estimation error at the receiver, is diagonal. Hence, only the diagonal elements, denoted as $[\cdot]_{[k,k]}$, have to be computed:

$$\boldsymbol{\Sigma}^{1,1}_{n|n\,[k,k]} = \frac{1}{SNR\cdot M_s + \dfrac{1}{\boldsymbol{\Sigma}^{1,1}_{n|n-1\,[k,k]}}} \tag{3.59}$$

The previous equation gives a recursive sequence whose solution converges to the steady-state covariance matrix of the estimated channel. This steady-state covariance matrix is not difficult to be derived in the particular case of a first order autoregressive model, which accomplishes $\boldsymbol{\Sigma}^{1,1}_{n|n-1}$=$\boldsymbol{\Sigma}_{n|n-1}$ and $\boldsymbol{\Sigma}^{1,1}_{n|n}$=$\boldsymbol{\Sigma}_{n|n}$. Substituting $\boldsymbol{\Sigma}_{n|n-1\,[k,k]}$ by its definition (3.40), and assuming that $\boldsymbol{\Sigma}_{n|n\,[k,k]}$ reaches the the steady-state condition: $\boldsymbol{\Sigma}_{n|n\,[k,k]} = \boldsymbol{\Sigma}_{n-1|n-1\,[k,k]} = \boldsymbol{\Sigma}_{\,[k,k]}$ (note that the index n has been omitted since the covariance matrix becomes constant), the diagonal elements in $\boldsymbol{\Sigma}$, are given by:

$$\boldsymbol{\Sigma}_{[k,k]} = \frac{1}{SNR\cdot M_s + \dfrac{1}{\left|\mathrm{W}_{1\,[k,k]}\right|^2\boldsymbol{\Sigma}_{[k,k]} + \mathbf{R}_{\mathbf{vv}[k,k]}}} \tag{3.60}$$

Next, some particular comments can be made on this result:

- A closed form solution exists for the the values of $\mathbf{\Sigma}_{[k,k]}$. This solution can be found by solving a simple second order equation derived from (3.60). It can be shown that only one of the two solutions is valid for our concerning problem, since the other one is always negative.

- In the hypothetic case in which the model for the evolution of the channel is perfectly known, i.e., $\mathbf{R_{vv}} = \mathbf{0}$, the steady-state covariance matrix is $\mathbf{\Sigma}_{[k,k]} = 0$. This results, as expected, says that if the Kalman model is perfectly adjusted to the real channel evolution, the variance on the estimation error can be as small as desired by simply enlarging the number of iterations, which implies to average a large number of observations.

Table 3.3: Transmitter channel tracking algorithm. Procedure at the receiver.

1. *Predict new channel state* $\hat{\mathbf{h}}^{(t)}_{R\,n|n-1}$ *from the from the first elements of the Kalman state vector as defined in* (3.39).

$$\hat{\mathbf{h}}^{(t)}_{R\,n|n-1}(n) = \sum_{q=1}^{p} \mathbf{W}_q \hat{\mathbf{h}}^{(t)}_{R\,n-1|n-1}(n-q)$$

2. *Making use of the observation* $\mathbf{r}(n)$*, the matrix with the transmitted symbols* $\mathbf{S}(n)$*, and the covariance* $\mathbf{\Sigma}_{n|n-1}$*, compute the innovation and the prediction error.*

$$\begin{aligned}\boldsymbol{\alpha}(n) &= \mathbf{r}(n) - \mathbf{S}(n)\hat{\mathbf{h}}^{(t)}_{R\,n|n-1} \\ \mathbf{e}(n) &= \mathbf{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H \left[\mathbf{S}(n)\mathbf{\Sigma}^{1,1}_{n|n-1}\mathbf{S}(n)^H + \mathbf{R_n}\right]^{-1} \boldsymbol{\alpha}(n)\end{aligned}$$

3. *Feedback to the transmitter the quantized prediction error.*

$$\mathbf{e}_q(n) = \mathcal{Q}\left\{\mathbf{e}(n)\right\}$$

4. *Update the CSI with the observation at time* n *and the covariance* $\mathbf{\Sigma}_{n|n-1}$.

$$\begin{aligned}\hat{\mathbf{h}}^{(t)}_{R\,n|n}(n) &= \hat{\mathbf{h}}^{(t)}_{R\,n|n-1}(n) + \mathbf{e}_q(n) \\ \hat{\mathbf{h}}^{(t)}_{R\,n|n}(n-i) &= \hat{\mathbf{h}}^{(t)}_{R\,n|n-1}(n-i) + \mathbf{\Sigma}^{1,i}_{n|n-1}\left(\mathbf{\Sigma}^{1,1}_{n|n-1}\right)^{-1}\mathbf{e}_q(n) \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1 \ldots p-1\end{aligned}$$

(*) Expressions for the covariance matrix $\mathbf{\Sigma}_{n|n-1}$ *are given in* (3.40) *and* (3.41)

- In the case in which there is a large uncertainty in the model used to describe the channel evolution, i.e., $\mathbf{R_{vv}} \longrightarrow \infty$, the steady-state covariance tends to: $\mathbf{\Sigma}_{n|n\,[k,k]} \longrightarrow \frac{1}{SNR \cdot M_s}$. In this case, the Kalman model does not help the receiver to estimate the channel, and thus, the performance cannot be improved when it is compared with algorithms that only use an instantaneous observation to estimate the channel (see e.g., [Tun01]). In other words, this result says that because the Kalman model is very uncertain, the algorithm, in opposite to the previous case, cannot improve its performance when increasing the number of iteration.

Channel estimation error at the transmitter. Following the steps proposed in Table 3.4 to track the channel evolution at the transmitter, the covariance of the channel estimation error $\mathbf{\Sigma}_{Tx}(n)$ can be defined as follows:

$$\mathbf{\Sigma}_{Tx}(n) = E\left\{\left(\mathbf{h}^{(t)}(n) - \hat{\mathbf{h}}^{(t)}_{T\,n|n-1}\right)\left(\mathbf{h}^{(t)}(n) - \hat{\mathbf{h}}^{(t)}_{T\,n|n-1}\right)^H\right\} = \mathbf{\Sigma}^{1,1}_{n|n-1} + \mathbf{\Sigma}_Q \tag{3.61}$$

where the first term contains the estimation error due to the channel evolution, and $\mathbf{\Sigma}_Q$ is the covariance due to the quantization error.

Table 3.4: Transmitter channel tracking algorithm. Procedure at the transmitter.

1. *Update the channel prediction using the feedback channel and the covariance* $\mathbf{\Sigma}_{n|n-1}$.

$$\begin{aligned} \hat{\mathbf{h}}^{(t)}_{T\,n|n}(n) &= \hat{\mathbf{h}}^{(t)}_{T\,n|n-1}(n) + \mathbf{e}_q(n) \\ \hat{\mathbf{h}}^{(t)}_{T\,n|n}(n-i) &= \hat{\mathbf{h}}^{(t)}_{T\,n|n-1}(n-i) + \mathbf{\Sigma}^{1,i}_{n|n-1}\left(\mathbf{\Sigma}^{1,1}_{n|n-1}\right)^{-1}\mathbf{e}_q(n) \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1 \ldots p-1 \end{aligned}$$

2. *Predict next channel state* $\hat{\mathbf{h}}^{(t)}_{T\,n+1|n}$.

$$\hat{\mathbf{h}}^{(t)}_{T\,n+1|n} = \sum_{q=1}^{p} \mathbf{W}_q \hat{\mathbf{h}}^{(t)}_{T\,n+1-q|n}$$

3. *Design the linear precoder matrix* $\mathbf{F}(n)$ *making use of the predicted channel. See next chapters for the design of linear precoder matrices.*

(*) Expressions for the covariance matrix $\mathbf{\Sigma}_{n|n-1}$ *are given in* (3.40) *and* (3.41)

3.5.6 Quantization of the feedback channel

As the prediction error $\mathbf{e}(n)$ is the minimum required information to track the LTV channel modelled as an autoregressive process, we analyze the statistics of the vector and the optimal way to quantize and encode its elements in a finite number of bits according to information theoretical principles. To encode the elements in vector $\mathbf{e}(n)$ without loosing information an infinite number of bits should be required. Alternatively, when a finite number of bits is employed, a quantization distortion is unavoidable. The rate-distortion theorem, which aims at reducing the distortion for a given target bitrate by optimally allocating the bits, is next defined.

> **Rate-distortion.** Defining D as the mean square-error distortion produced when quantizing a random variable x with $\mathcal{Q}(x)$: $D = (x - \mathcal{Q}(x))^2$, and employing R bits to quantize the variable, the minimum number of bits required to quantize a source with distortion less or equal to D is given by the rate-distortion function.

As a consequence of the central limit theorem, real and imaginary elements in vector $\mathbf{e}(n)$ are approximately Gaussian random processes whose mean and variance have been derived in Section 3.5.4. Hence, the vector to be quantized, $\mathbf{e}(n)$, is a circularly symmetric complex Gaussian random vector of length $M_T \cdot M_R \cdot L$, and whose elements are i.i.d. with zero mean. In the particular case of a first order autoregressive model, and under the steady-state condition: $\mathbf{\Sigma} = \mathbf{\Sigma}_{n|n} = \mathbf{\Sigma}_{n-1|n-1}$, an expression for the covariance matrix $\mathbf{\Sigma_e}$ simpler than that one presented in (3.55) is next derived. The expression given in (3.53) for the error covariance $\mathbf{\Sigma_e}$ coincides with the second term in (3.57). Hence, this matrix can be easily computed as:

$$\mathbf{\Sigma}_{\mathbf{e}(n)} = \mathbf{\Sigma}^{1,1}_{n|n-1} - \mathbf{\Sigma}^{1,1}_{n|n}$$

Note that the covariance of the channel prediction error can be understood as the amount of improvement in the knowledge of the state vector (i.e., the improvement in the estimation of the CSI) between the prediction provided by the observations up till time $n-1$ and the estimation provided with the observation up till time n.

Next, substituting the covariance matrix $\mathbf{\Sigma}^{1,1}_{n|n-1}$ by its definition for a first order AR model (3.40), and assuming that the covariance matrices reach the steady-state condition $\mathbf{\Sigma} = \mathbf{\Sigma}_{n|n} = \mathbf{\Sigma}_{n-1|n-1}$ we get:

$$\mathbf{\Sigma}_{\mathbf{e}(n)} = \mathbf{W}_1 \mathbf{\Sigma}_{n-1|n-1} \mathbf{W}_1^H + \mathbf{R_{vv}} - \mathbf{\Sigma}_{n|n} \tag{3.62}$$

Finally, assuming that these matrices are diagonal (this assumption hold true if $\mathbf{S}(n)^H \mathbf{S}(n)$ is diagonal), we obtain a constant value for the covariance matrix $\mathbf{\Sigma}_{\mathbf{e}(n)}$ given by:

$$\mathbf{\Sigma}_{\mathbf{e}(n)} = \left[\left(|\mathrm{W}_1|^2 - \mathbf{I} \right) \mathbf{\Sigma} + \mathbf{R_{vv}} \right] \tag{3.63}$$

For the specific case of a zero mean Gaussian source with variance σ_x^2 the rate-distortion only depends on the variance and is given by [Pro94, p.245]:

$$R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma_x^2}{D} & 0 \leq D \leq \sigma_x^2 \\ 0 & D > \sigma_x^2 \end{cases} \tag{3.64}$$

or equivalently:

$$D(R) = \sigma_x^2 2^{-2R} \tag{3.65}$$

and thus, this value could be easily computed for the elements of vector $\mathbf{e}(n)$ according to results given in (3.55) or (3.63).

The rate-distortion function is a theoretical bound in the sense that it can only be achieved by increasing the encoding-decoding complexity. Vector quantization is the optimal solution to attain (3.65) making use of the Lloyd algorithm [Llo57]. Nevertheless, vector quantization design is extremely complex and only attains the optimal limit when the number of elements to quantize is large. Instead, a suboptimum solution, considering individually real and imaginary parts, and quantizing each vector element separately, is proposed. Hence the total number variables to encode is $2 \cdot M_T \cdot M_R \cdot L$. Because the elements in vector $\mathbf{e}(n)$ are approximately Gaussian random processes, an optimum non-uniform quantizer for Gaussian sources designed by Max [Max60] will be applied. Numerical methods were employed in that paper to minimize the distortion (3.65) when the signal to quantify follows a normal distribution.

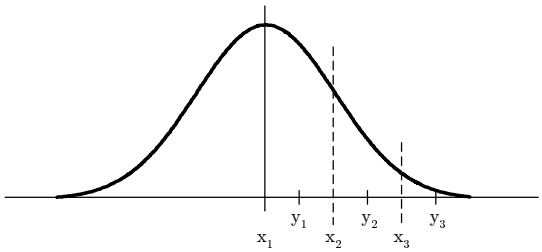
Table 3.5 summarizes the values of the optimum quantizer for different number of bits, displays the expected distortion of the quantization process, and compares the result with the theoretical distortion given by (3.65). The difference between the theoretical distortion and the real distortion using the non-uniform Max's quantizer is justified as the difference between a scalar quantization (based on the quantization single elements) and an optimal quantization (based on the vector quantization), respectively.

It results in interest to evaluate the theoretical gain in terms of bits when the prediction error, instead to the channel impulse response coefficients, is quantized and fed back. In order to evaluate this growth in the number of bits we will first compute the distortion produced by the quantization of the prediction error with R_e bits, and then we will calculate the number of bits R_h that would be required to quantize the channel coefficients producing the same distortion. From (3.65) the minimum distortion due to the quantization of a random variable with variance σ_e^2 and R_e bits is:

$$D = \sigma_e^2 2^{-2R_e},$$

Next, if this distortion value is introduced into (3.64) to guess the number of bits R_h required to quantize the channel coefficients, whose variance is σ_h^2, we get:

$$R_h = \frac{1}{2}\log_2 \frac{\sigma_h^2}{D},$$



x's mark the input interval end points.
y's mark the output levels.

R=1	x	0
	y	±0.7980
	D/σ_e^2 [Max60]	0.3634
	D/σ_e^2 $D(R)$	0.25

R=2	x	0 ±0.9816
	y	±0.4528 ±1.510
	D/σ_e^2 [Max60]	0.1175
	D/σ_e^2 $D(R)$	0.0625

R=3	x	0 ±0.5006 ±1.050 ±1.748
	y	±0.2451 ±0.7560 ±1.344 ±2.152
	D/σ_e^2 [Max60]	0.03454
	D/σ_e^2 $D(R)$	0.0156

R=4	x	0 ±0.2582 ±0.5224 ±0.7996 ±1.099 ±1.437 ±1.844 ±2.401
	y	±0.1284 ±0.3881 ±0.6568 ±0.9424 ±1.256 ±1.618 ±2.069 ±2.733
	D/σ_e^2 [Max60]	0.009497
	D/σ_e^2 $D(R)$	0.0039

Table 3.5: Minimum mean-squared quantization scheme for Gaussian sources with zero mean and unit variance [Max60].

Finally, combining both equations, a relationship between the number of bits required to quantize the channel coefficients, R_h, and the number of bits required to quantize the prediction error, R_e, can be derived:

$$R_h = R_e + \frac{1}{2}\log_2 \frac{\sigma_h^2}{\sigma_e^2} \tag{3.66}$$

Note that because the variance of the prediction error σ_e^2 has a dynamic range lower than the channel coefficients σ_h^2, i.e., $\sigma_h^2 > \sigma_e^2$ (most of the times $\sigma_h^2 \gg \sigma_e^2$), as it was expected, the number of bits required to quantize the channel impulse response coefficients is always greater than the the number of bits required to quantize the channel prediction error. In the next section this difference will be evaluated for each one of the scenarios simulated (see Figures 3.8d, 3.9d and 3.10d).

OFDM symbols and frame structure	According to HiperLAN/2 standard
MIMO configuration	$M_T = 2$ $M_R = 2$
Channel	Exponential PDP with Delay spread 50 ns
SNR	10dB
Normalized Doppler frequency	f_d=0.037
Training sequence	1 symbol OFDM
Doppler Spectrum	Classical U-shape

Table 3.6: Simulated parameters.

3.6 Simulation results

This section illustrates the performance of the proposed scheme to track the TV channel, as a function of the normalized Doppler frequency, the SNR and the length of the training sequence. All the simulations have been computed for different AR models (p=1 and p=2), as well as for the *random-walk* predictor. All plotted simulations in this section refer, without loss of generality, to the first elements in vector $\mathbf{h}$ (i.e., the first tap $l = 1$ of the channel impulse response).

A MIMO wireless scenario has been considered, and the simulated parameters has been selected according to the physical layer of the HIPERLAN/2 standard [HL201]. Table 3.6 summarizes the parameters used in the simulations.

In Figures 3.8-3.10 the subplot labelled as (a) displays the channel estimation error at the transmitter (3.61), the subplot (b) displays the channel estimation error at the receiver (3.56), the subplot (c) displays the variance of the prediction error $\mathbf{e}(n)$ (3.55), and the subplot (d) displays the number of additional bits required to feedback the channel coefficients instead of the prediction error without increasing the distortion (3.66).

3.6.1 Performance evaluation for different normalized Doppler frequencies

Figure 3.8 compares the transmitter channel tracking algorithm as a function of the normalized Doppler frequency while the rest of the simulation parameters are those shown in Table 3.6. Vertical line at $f_d = 0.037$ only denotes the normalized Doppler frequency for HIPERLAN/2 assuming that the user is moving at 4km/h. Figure 3.8d illustrates the gain in terms of bits when the prediction error is fed back instead of the channel coefficients themselves. Note that always exists a gain, and it is inversely proportional to the prediction error variance (figure 3.8c). Hence, the lower the prediction error is, the greater the gain in number of bits we get by transmitting only the prediction error.

Analyzing the performance as a function of the linear predictor order it can be shown that the random-walk predictor and the AR(1) predictor have similar performances for all the simulated range. As it was expected, the AR(2) always outperforms the first order model. See for example the channel estimation error at the transmitter (Figure 3.8a) and at the receiver (Figure 3.8b). There exists, thus, a trade-off between the complexity of the linear predictor and the variance in the channel uncertainty.

Analyzing the performance in terms of the channel estimation error at the receiver, it can be seen in Figure 3.8b that when the normalized Doppler frequency increases (i.e., channel tends to be fast time varying) the channel estimation error tends to be constant. The explanation can be found analyzing the steady-state covariance for the AR(1) model (3.60) (the same arguments are valid for higher AR orders). Note that when channel is fast time varying, the second term in the denominator of the channel estimation error vanishes and thus the channel estimation error at the receiver tends to:

$$\mathbf{\Sigma}_{n|n\,[k,k]} = \frac{1}{SNR \cdot M_s} \tag{3.67}$$

As said in Section 3.5.5, in this case the linear predictor does not help the receiver to estimate the channel, and the estimation can only be done making use of the observed sequence. Nevertheless, when channel is slow time varying, the Kalman linear predictor aids the channel estimator reducing the variance of the estimation error.

The present simulation also allows to set the number of required bits to quantize $\mathbf{e}(n)$. This information can be induced from the channel estimation error at the transmitter (Figure 3.8a). As it can be shown, the gain is inappreciable when more than 3 bits are used, whereas if only 1 bit is used the performance degradation due to quantization error can be appreciated. Consequently, it can be concluded that 2 or 3 bits are enough to get a performance that is close to the unquantized solution.

3.6.2 Performance evaluation for different SNR's

Figure 3.9 compares the transmitter channel tracking algorithm as a function of the noise level, keeping constant the rest of the simulation parameters according to Table 3.6. As in the previous simulation, the random-walk predictor and the AR(1) predictor have close performances, whereas the AR(2) model exhibits a better performance.

The most important conclusion of this simulation is drawn from the channel estimation error at the receiver (Figure 3.9b). The plotted results corroborate that the dominant term to determine the channel estimation error at the receiver is the noise level, obtaining a proportional relationship between the SNR and the channel estimation error as given in (3.67).

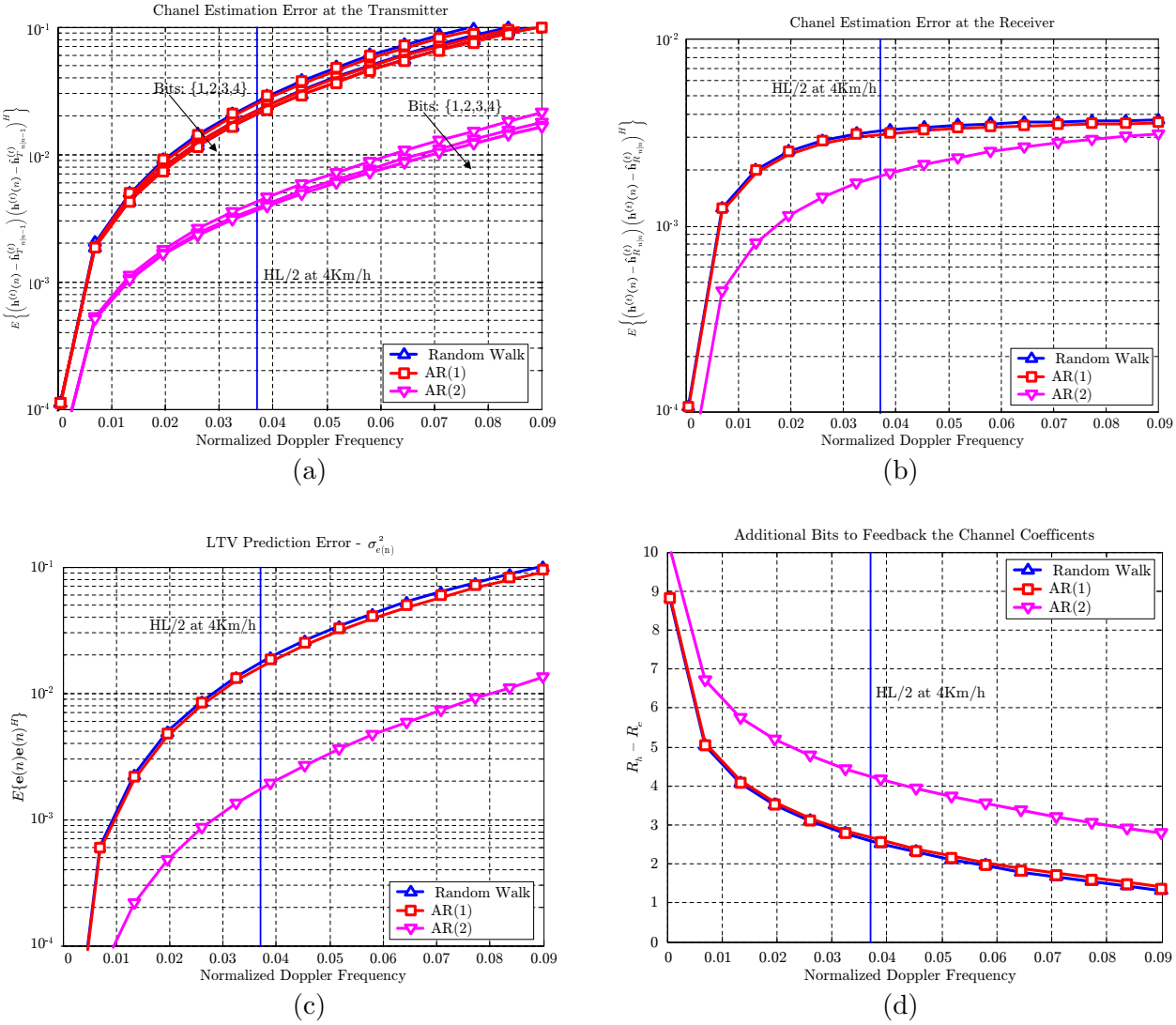


Figure 3.8: Performance evaluation of the first tap in the channel impulse response for different normalized Doppler frequencies. Subplot (a): Channel estimation error at the transmitter (3.61). Subplot (b): Channel estimation error at the receiver (3.56). Subplot (c): Variance of the prediction error $\mathbf{e}(n)$ (3.55). Subplot (d): Number of additional bits required to feedback the channel coefficients instead of the prediction error without increasing the distortion (3.66).

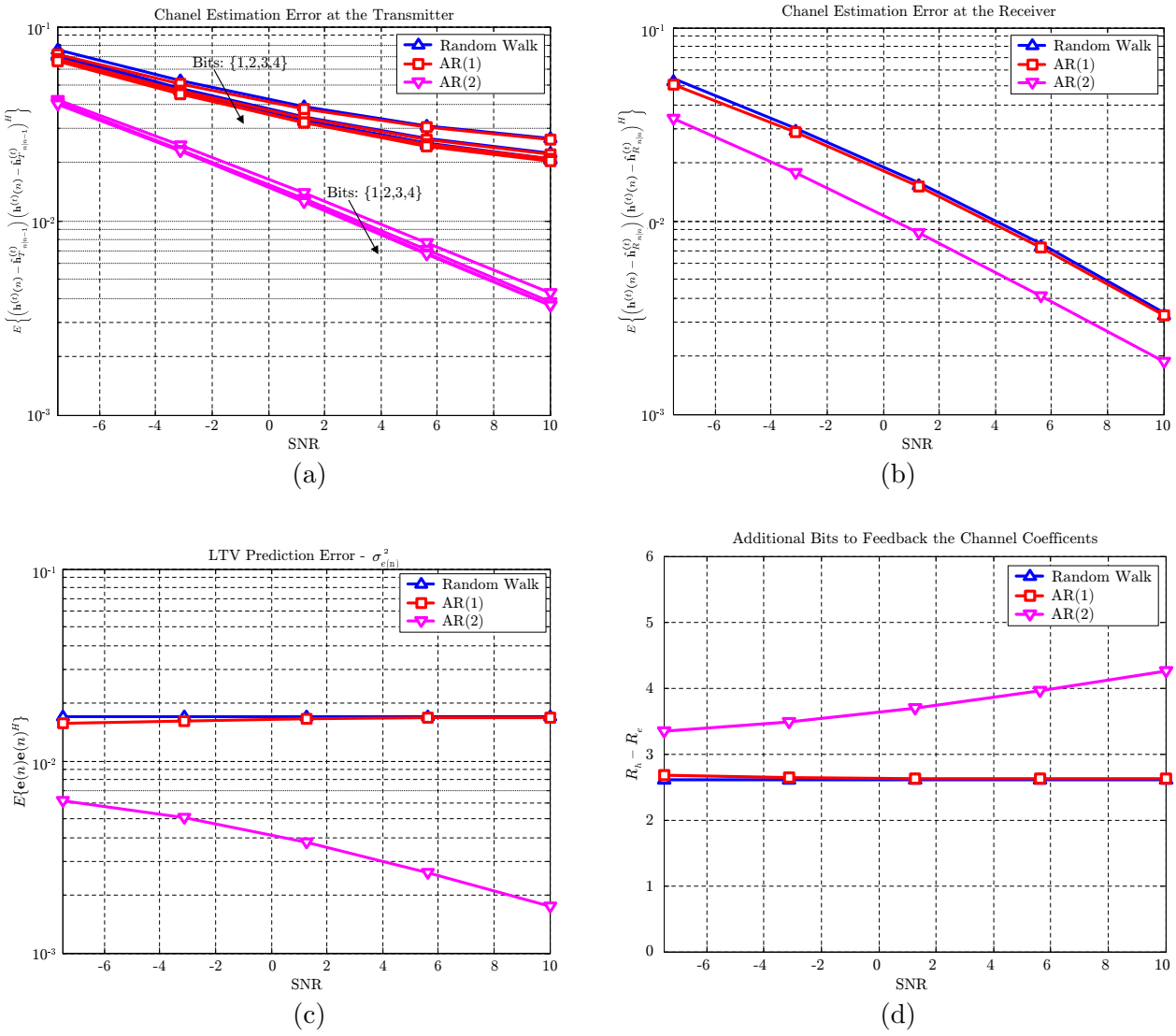


Figure 3.9: Performance evaluation of the first tap in the channel impulse response for different SNR's. Subplot (a): Channel estimation error at the transmitter (3.61). Subplot (b): Channel estimation error at the receiver (3.56). Subplot (c): Variance of the prediction error $\mathbf{e}(n)$ (3.55). Subplot (d): Number of additional bits required to feedback the channel coefficients instead of the prediction error without increasing the distortion (3.66).

It is also relevant to note that when the random-walk predictor is used, the variance of the prediction error $\mathbf{e}(n)$ keeps independent of the noise level (see Figure 3.9c). This is straightforwardly justified form equation (3.63). Note that when $\mathbf{W}_1 = \mathbf{I}$, the covariance matrix $\mathbf{\Sigma}_{\mathbf{e}(n)\,[k,k]}$ only depends on the channel variability and thus, it is insensitive to the rest of parameters.

Finally, the number of bits to quantize the linear prediction error does not depends on the SNR and as previously deduced, 2 or 3 bits are enough to track the channel (see Figure 3.9a).

3.6.3 Performance evaluation for different training sequence lengths

Figure 3.10 compares the performance of the transmitter channel tracking algorithm when the number of OFDM symbols in the training sequence is changed. We can drawn the same conclusions that in Section 3.6.2 referring to the relation between the random-walk, the AR(1) and the AR(2) predictors. Moreover, Figure 3.10c illustrates that the variance of the prediction error $\mathbf{e}(n)$ keeps independent of the training sequence length when the random-walk predictor is used.

The influence of the number of OFDM symbols employed in the training sequence is observed in the channel estimation error at the transmitter (Figure 3.10a) as well as at the receiver (see Figure 3.10b). As expected from equation (3.59), the variance of channel estimation error at the receiver reduces approximately 3dB when the number of OFDM symbols is doubled. This result applies, as argued in Section 3.5.5, at high SNR's or large uncertainties in the model used to describe the channel variability. In both cases, $\mathbf{\Sigma}_{n|n\,[k,k]} \longrightarrow \frac{1}{SNR \cdot M_s}$

3.6.4 Performance evaluation for different normalized Doppler spectrum

For the design of the linear predictor we have assumed that the Doppler spectrum obeys the classical U-shaped Clarke's Doppler spectrum. However, as analyzed in Section 3.3.3 this assumption only holds when the direction of arrival is isotropically distributed, and omnidirectional antennas with constant gain are used. Otherwise, different Doppler spectra are obtained. Under a 2-D assumption of the impinging rays, the p.d.f. of the power angular spectrum (PAS), the *Direction of Arrival* (DoA), the angular spread (AS) and the *Direction of Movement* (DoM) of the mobile determine the final Doppler spectrum.

To evaluate the mismatch effects on the channel tracker when the classical U-shaped, instead of the real Doppler spectrum, is used to design the linear predictor, we present a set of simulations for different Doppler spectra. The uniform Doppler spectrum labeled as *'Doppler spectrum: Uniform'* and the Doppler spectrum derived under a Laplacian PAS with DoA: 0^o, DoM: 70^o and AS: 20^o labeled as *Doppler spectrum: Laplacian* (see Section 3.3.3) are next simulated as a function of the normalized Doppler frequency and assuming an AR(1) predictor. The rest of the simulation parameters are in Table 3.6.

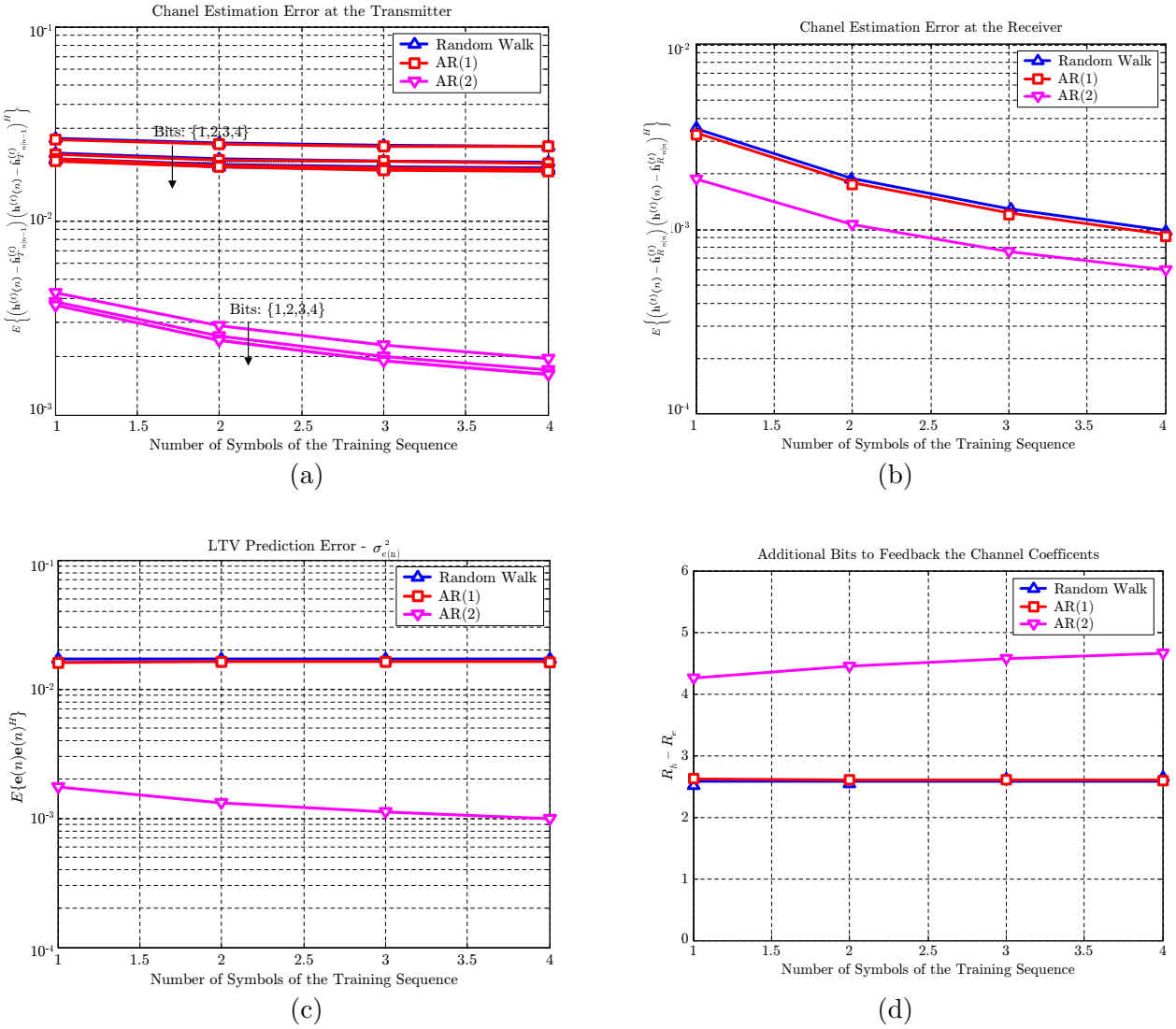


Figure 3.10: Performance evaluation of the first tap in the channel impulse response as a function of the number of OFDM symbols in the training sequence. Subplot (a): Channel estimation error at the transmitter (3.61). Subplot (b): Channel estimation error at the receiver (3.56). Subplot (c): Variance of the prediction error $\mathbf{e}(n)$ (3.55). Subplot (d): Number of additional bits required to feedback the channel coefficients instead of the prediction error without increasing the distortion (3.66).

Figure 3.11 illustrates the channel estimation error at the transmitter (Figure 3.11a) and at the receiver (Figure 3.11b), and the variance of the prediction error (Figure 3.11c), comparing its performance with the classical U-shaped simulations presented in previous subsections.

As it can be seen, the worst performance for the channel estimation error at the transmitter, as well as the higher variance of $\mathbf{e}(n)$, occur for the classical U-shape Doppler spectrum. To understand this behavior, Figure 3.11d displays the time channel correlation for the tree simulated scenarios. This correlation, denoted by $R_h(\Delta n; l)$, is given by the inverse Fourier transform of the scattering function (see Section 3.3.2). The results shown in this Figure 3.11d reveal that the channel is more correlated for the Laplacian and uniform Doppler spectrum when compared with the U-shape and consequently, the time variability of those channels is slower. This result justifies that the variance of $\mathbf{e}(n)$ and the channel estimation error at the transmitter is lower for the new Doppler spectra because the channel evolution is easier to be predicted.

Another conclusion that can be drawn form this figure is that when the channel is fast time varying, the channel estimation error at the receiver is insensitive to the Doppler spectrum (Figure 3.11b). This result, as argued in Section 3.5.5, reveals that when channel state changes rapidly, its estimation at the receiver is independent of the time evolution (i.e., insensitive to the specific Doppler spectrum), and tends to the channel estimation error given in (3.67).

3.7 Conclusions

This chapter has introduced a suitable scheme to assist the transmitter with the CSI in optimal transmitter-receiver equalization designs. The proposed solution is based on two identical linear predictors at both sides of the link, which are based on the Kalman filter, and a feedback channel to aid the transmitter to track the channel variations.

Considering the severe limitations of the feedback channel capacity, this chapter has focused efforts to set the minimum amount of information required to track the channel and to determine the return link throughput. It has been shown that the prediction error contains the minimum information to be transferred through the feedback link. Moreover, it has also been shown that it is possible to track fast fading channels with a low rate feedback link (two or four bits per complex channel coefficient).

The complexity of the channel predictors has also been considered. Simulations have shown 1st order predictors are good in slow LTV channels, while second order predictors are required in fast LTV channels at the expense of higher complexity. Although it has not been considered in the chapter, in the presence of very fast time-varying channels the study could be extended to higher orders.

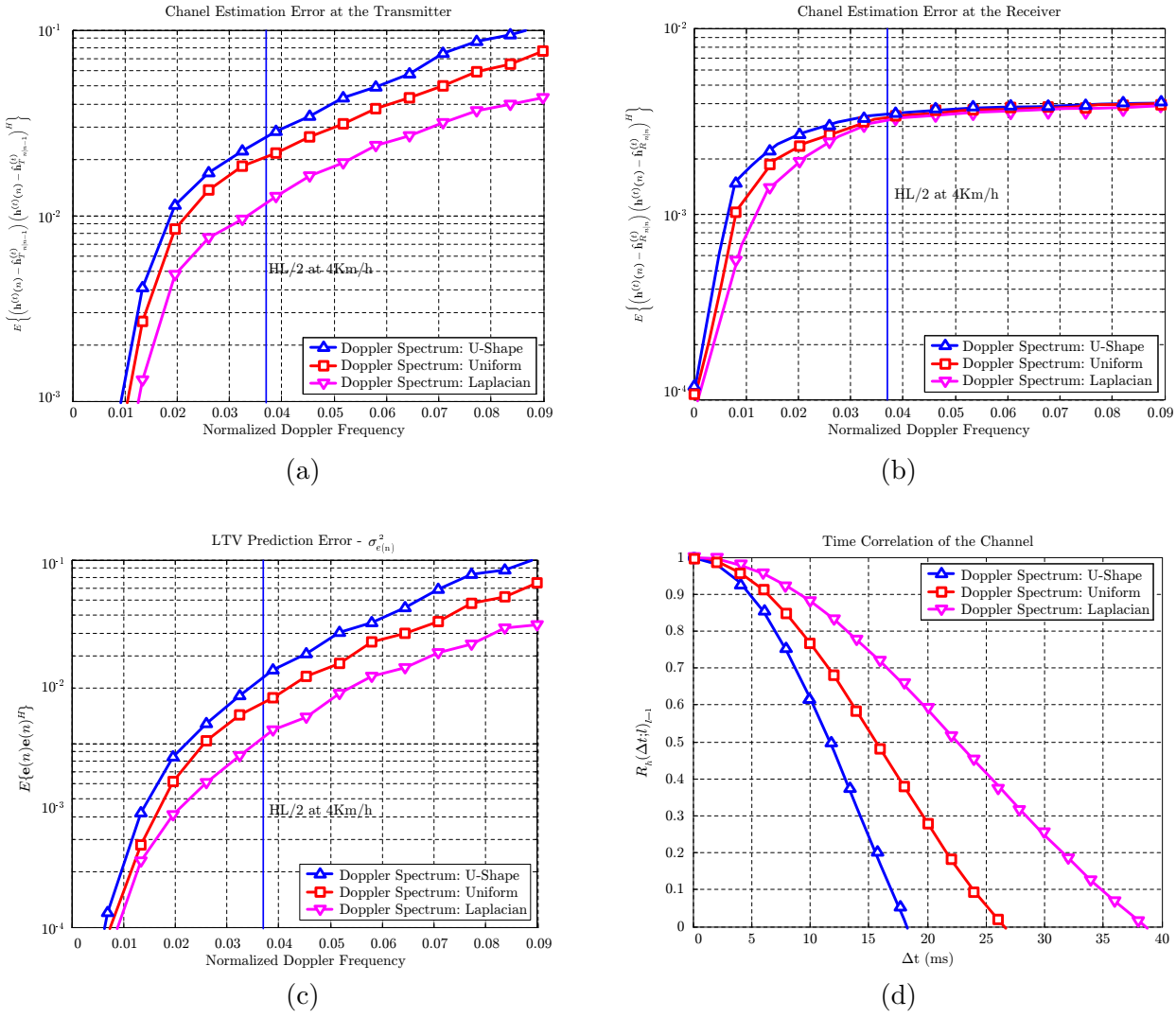


Figure 3.11: Performance evaluation of the first tap in the channel impulse response for different Doppler spectra (U-Shape, Uniform and Laplacian). Subplot (a): Channel estimation error at the transmitter (3.61). Subplot (b): Channel estimation error at the receiver (3.56). Subplot (c): Variance of the prediction error $\mathbf{e}(n)$ (3.55). Subplot (d): Time correlation of the channel for different Doppler spectra.

Chapter 4

Robust power allocation algorithms for MIMO OFDM systems with imperfect CSI

The design of power allocation criteria in fading channels improves the performance of the digital communication systems. A large variety of criteria can be chosen to be optimized, including the zero-forcing, the *Minimum Mean Square Error* (MMSE) or the minimum uncoded and coded BER. These criteria has been widely studied in the literature when perfect CSI is available at the transmitter. However, the study of closed-loop schemes when channel estimates are noisy is still an open area. As suggests this thesis, the performance degradation due to the CSI uncertainty can be diminished by proposing a Bayesian approach to the design of the transmitter prefiltering matrices in closed-loop schemes that becomes robust to the channel estimation errors. The design of these robust linear transmitters, like their nonrobust counterparts, are based on the *Singular Value Decomposition* (SVD) of the channel response (in multicarrier schemes the SVD is independently a applied over each subcarrier). The SVD decomposition is particularly convenient for the OFDM modulation since it allows to allocate power or bits on a subcarrier basis.

This chapter pays slight attention to the channel coding stage, even though, for the sake of completeness, the performance of the proposed algorithms has been compared in terms of coded BER for the particular channel code described in the HIPERLAN/2 standard. Relevant contributions to this topic are left to the next chapter, where robust power allocation strategies related with the minimization of the coded BER will be studied.

The chapter is organized as follows. First, Section 4.2 describes the system architecture under analysis and the channel model. This section introduces the named equivalent channel that exploiting the correlation between the subcarriers and the channel uncertainty structure will try to mitigate the degradation due to the CSI errors. Afterwards, the design of the linear transmitter that minimizes the MSE is presented in Section 4.3 with the following structure: derivation of the cost function, closed-form solution and asymptotic analysis of the solution for very high and very low uncertainty. Section 4.4 is concerned with the design of the linear transmitter that minimizes the uncoded BER. The first part of the section derives, after introducing the cost function, a closed-form solution for the specific case of beamforming, whereas the second part of the section extends the study for the spatial multiplexing case. As other previously proposed solutions in the literature it will be shown that this algorithm converges to the uniform power allocation for very poor CSI estimates. Finally, Section 4.5 and 4.6 report some simulation experiments comparing the performance of the two algorithms in terms of uncoded BER and coded BER, respectively. In order to focus the comparison on the different designs proposed for the transmitter, a *Maximum Likelihood* (ML) receiver has been always used.

4.1 Introduction

The problem of power allocation for single and multiple antenna transmission in fading channels has been widely studied in the literature under a variety of criteria, including zero-forcing and *Minimum Mean Square Error* (MMSE) schemes [Bar00, Teh98], maximum information rate [Sca99a] and also minimum uncoded *Bit Error Rate* (BER) [Ong02, Din03] providing in all cases a design based on the SVD of the channel estimate. However, all these algorithms assume that the channel is perfectly known at the transmitter. This hypothesis does not hold in real systems, since neither the transmitter nor the receiver have access to ideal CSI. As it was shown in the previous chapter, the errors in the channel estimates can be originated from several sources, including the estimation variance due to noise, the time lag between channel estimation/prediction and its use for transmitter design, and the quantization error in the feedback channel. While the first one is common to the transmitter and the receiver, the second and third ones only appear in the transmitter. Thus, depending on the pace of channel variation and depending on the dimensioning of the feedback channel, the channel uncertainty at the transmitter can be larger than that one at the receiver, but in any case none of them will have perfect CSI.

If the errors in the CSI are not taken into account in the transmitter design, the performance of the closed-loop algorithm will degrade and, eventually, may get worse than that one of open-loop transmission. Thus, the potential of linear precoding can only be fulfilled when the reliability of the channel estimates is considered in the cost function [Nar98]. The design of algorithms that take into account partial knowledge of the CSI can be grouped into two cate-

gories: a first group that considers imperfect estimates of the channel impulse response to design the transmitter [Ong02, Jön02], and a second group that designs optimal transmitter schemes based only on the knowledge of the channel statistics [Ong02, Sam02]. The major contribution in [Ong02] was the design of wide-band signaling schemes under a minimum uncoded BER criterion. It was shown that when perfect CSI is available at the transmitter, a beamformer focused to the most dominant channel mode achieves the optimal solution, whereas a set of linearly independent signature codes that transmit on all channel modes are required when CSI is poor. Readers are also referred to [Sam02] and [Jön02] for the design of linear precoders for space-time coded systems.

In the framework on this thesis we propose a Bayesian approach to design the optimal linear transformation when the channel estimates at the transmitter are noisy. Following the structure proposed in the previous chapter, where the receiver predicts and estimates the channel response and feeds back this information to the transmitter, unbalanced CSI quality between the transmitter and the receiver can be assumed. We introduce the idea of the *a priori* design, as the design carried out at the transmitter and based on the predicted channel at this side, and the *a posteriori* design, as the design carried out at the receiver and based on the updated CSI from the observation of the channel output.

As it will be shown in the following sections, the Bayesian formulation leads to a design based on the SVD decomposition of the equivalent channel, that is the result of averaging the channel estimate over the channel uncertainty. The two criteria studied in this chapter (MMSE and minimum uncoded BER) lead to closed-form solutions whose complexity is similar to their non-robust counterparts, although in the case of the uncoded BER some constraints must be introduced. Besides, the algorithm for minimum uncoded BER can be regarded as a transmitter design that adapts automatically to the channel knowledge, moving from the open-loop scheme (the same power is allocated across all subcarriers and antennas) to the closed-loop scheme according to channel uncertainty. This feature is in common with the results in [Ong02, Sam02, Jön02]. Another characteristic of the proposed algorithms is that they can accept different transmission rates, so they can use the space diversity (MIMO channel) to improve channel reliability or to increase the transmission rate.

4.2 System and channel model

This section describes the signal model for a MIMO OFDM communications system over a frequency-selective fading channel expanding the generic discrete-time MIMO communication system introduced in Chapter 3. This formulation is quite general, encompassing the single antenna transmission, the beamforming schemes and the frequency flat fading channels as particular cases.

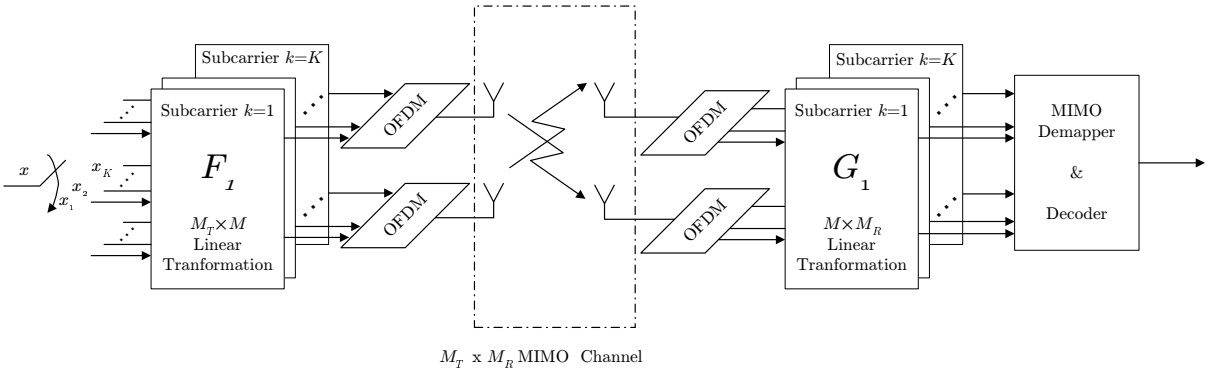


Figure 4.1: Block diagram of the proposed MIMO OFDM system.

The application of the OFDM modulation will allow us to design the optimum precoders in the frequency domain and, hence, the system model will be directly formulated in this domain. According to the scheme in Figure 4.1, the MIMO configuration consists of M_T transmit antennas and M_R receive antennas, M denotes the number of symbols to be transmitted per subcarrier into a transmit block that is bounded by $M \leq min\{M_T, M_R\}$, and K is the total number of subcarriers.

Let $\mathbf{x}$ be the $KM \times 1$ vector containing the KM symbols that are simultaneously transmitted in the M_T OFDM symbols (i.e., one OFDM symbol per antenna). The data are assumed to be i.i.d. symbols in a constellation $\{s_1 \dots s_{\mathcal{N}}\}$ of cardinality $\mathcal{N}$, with zero mean and variance $E\left\{\mathbf{x}\mathbf{x}^H\right\} = \sigma_x^2\mathbf{I}$. If the channel keeps invariant within one OFDM symbol and the cyclic prefix (whose length is appropriately chosen) is removed at the receiver, the system model can be written as follows:

$$\mathbf{r} = \mathbf{GHFx} + \mathbf{Gn} \tag{4.1}$$

where $\mathbf{H}$ is a $KM_R \times KM_T$ block diagonal channel matrix containing the frequency responses of the MIMO channels, $\mathbf{F}$ is a $KM_T \times KM$ matrix that denotes the linear precoder matrix and allocates the power across the K subcarriers and M_T antennas, $\mathbf{G}$ is a $KM \times KM_R$ matrix that combines the signal received at the M_R antennas, and $\mathbf{n}$ is the $KM \times 1$ noise vector in the frequency domain, which has the same Gaussian statistic as its time-domain counterpart (i.e., zero mean and covariance $E\left\{\mathbf{n}\mathbf{n}^H\right\} = \sigma_n^2\mathbf{I}$).

Imposing a certain structure in the transmitter and receiver matrices, the frequency-selective MIMO channel can be decoupled into K independent MIMO frequency-flat fading channels, simplifying the design of the linear transformations $\mathbf{F}$ and $\mathbf{G}$ at expenses of perform a suboptimal solution. As shown in Figure 4.1, these linear transformations are split into K submatrices that process each subcarrier independently. Hence, $\mathbf{G}$,$\mathbf{H}$,$\mathbf{F}$ matrices and $\mathbf{x}$,$\mathbf{r}$ vectors involved in (4.1)

can be structured as:

$$\begin{aligned} \mathbf{H} &= \text{diag}\{[\mathbf{H}_1\mathbf{H}_2\ldots\mathbf{H}_K]\} \\ \mathbf{F} &= \text{diag}\{[\mathbf{F}_1\mathbf{F}_2\ldots\mathbf{F}_K]\} \\ \mathbf{G} &= \text{diag}\{[\mathbf{G}_1\mathbf{G}_2\ldots\mathbf{G}_K]\} \\ \mathbf{x}^T &= [\mathbf{x}_1^T\ldots\mathbf{x}_K^T] \\ \mathbf{r}^T &= [\mathbf{r}_1^T\ldots\mathbf{r}_K^T] \end{aligned} \tag{4.2}$$

where $\mathbf{F}_k$,$\mathbf{H}_k$, and $\mathbf{G}_k$ are, respectively, $M_T \times M$, $M_R \times M_T$, and $M \times M_R$ matrices processing the kth subcarrier, and $\mathbf{x}_k$, $\mathbf{r}_k$ are $M\times 1$ vectors containing the transmitted and received symbols by the kth subcarrier through the different antennas. Following the structure in (4.2), the input-output relationship in (4.1) is rewritten as a set of equations that contain only one subcarrier each:

$$\mathbf{r}_k = \mathbf{G}_k\mathbf{H}_k\mathbf{F}_k\mathbf{x}_k + \mathbf{G}_k\mathbf{n}_k \quad k = 1\ldots K \tag{4.3}$$

The architecture proposed in (4.3) is the basis for the study of the power allocation algorithms followed in this thesis. Note that the block diagonal structure proposed for the linear transformations $\mathbf{F}$ and $\mathbf{G}$ in (4.2) has the advantage of providing a scheme where the spatial prefiltering matrix is applied individually to each subcarrier, reducing to the classic OFDM scheme when only one antenna is used at the transmitter.

A more general scheme could have been possible if the linear transformations $\mathbf{F}$ and $\mathbf{G}$ had been considered as unstructured matrices. Under this condition the transmitted symbols could have been mixed among all subcarriers and antennas reducing the restrictions in the design at expenses of a greater complexity. This complexity, and the fact that this unconstrained design breaks the OFDM structure, justify the structure proposed for $\mathbf{F}$ and $\mathbf{G}$ matrices.

It is worth mentioning that the receiver filtering matrix $\mathbf{G}$ has been included in the model for the sake of completeness and because it is required in the design of the minimum MSE algorithm, although the simulation results do not implement it because a ML detector was used to make a fair comparison of the proposed algorithms. This issue is reviewed in Section 4.3 when the cost function for minimum MSE is addressed. It is also worth to observe that although most of the communication systems incorporate a channel code scheme, this stage has not been taken into account in this chapter and it has not been considered in the design of $\mathbf{F}_k$ and $\mathbf{G}_k$ matrices. The evaluation of the proposed algorithms in terms of coded BER is left to the next chapter.

During subsequent developments, it will be useful to store the complete channel response in a vector. The MIMO channel response for the kth subcarrier and for the multicarrier system

will be denoted $\mathbf{h}_k$ and $\mathbf{h}$, respectively:

$$\begin{aligned} \mathbf{h}_k &= \mathrm{vec}\left(\mathbf{H}_k\right) \\ \mathbf{h} &= \left[\begin{array}{ccc} \mathbf{h}_1^T & \ldots & \mathbf{h}_K^T \end{array} \right]^T \end{aligned} \tag{4.4}$$

According to this new notation, and making use of the identity $\mathrm{vec}\left(\mathbf{ABC}\right) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \mathrm{vec}\left(\mathbf{B}\right)$ [Mag99, p.30], the received vector for the kth subcarrier $\mathbf{r}_k$ can be rewritten as:

$$\mathbf{r}_k = \mathbf{A}_k \mathbf{h}_k + \mathbf{G}_k \mathbf{n}_k \quad k = 1 \ldots K \tag{4.5}$$

where the matrix $\mathbf{A}_k$ is defined as:

$$\mathbf{A}_k = \left(\mathbf{F}_k \mathbf{x}_k\right)^T \otimes \mathbf{G}_k \tag{4.6}$$

4.2.1 Channel model

This section introduces the model for the channel response and channel estimates used to design and simulate the power allocation algorithms. The propagation channel is modelled as a fading channel with uncorrelated coefficients for all taps in the impulse response (i.e., US assumption) and identical power delay profile for all subchannels. This model encompasses the Rayleigh fading and Ricean fading channels as particular cases. For the sake of completeness, spatial correlation is considered in the channel model and also introduced into the formulation of the algorithms (assuming identical correlation for all the channel taps), although the uncorrelation assumption will be required to derive closed-form solutions for the proposed algorithms. The channel vector $\mathbf{h}$ defined in (4.4) is modelled as a multivariate Gaussian process with mean and covariance derived in Appendix 4.A:

$$\begin{aligned} E\left\{\mathbf{h}\right\} &= \mathbf{m}_{\mathbf{h}} \\ E\left\{\left(\mathbf{h} - \mathbf{m}_{\mathbf{h}}\right)\left(\mathbf{h} - \mathbf{m}_{\mathbf{h}}\right)^H\right\} &= \mathbf{C}_{\mathbf{hh}} = \mathbf{P} \otimes \mathbf{R} \end{aligned} \tag{4.7}$$

where matrix $\mathbf{R}$ is the antenna correlation matrix, and matrix $\mathbf{P}$ is an Hermitian Toeplitz matrix whose first row is the *Discrete Fourier Transform* (DFT) of the variance in the channel impulse response coefficients, which is assumed to be known (both matrices are clearly defined in Appendix 4.A).

The design of the linear precoder in the presence of channel estimation errors requires the definition of a model for the estimated channel $\hat{\mathbf{h}}$. It will be assumed that the channel estimation error $\boldsymbol{\varepsilon}$ can be modelled as an additive term as follows:

$$\hat{\mathbf{h}} = \sqrt{\frac{\sigma_h^2}{\sigma_h^2 + \sigma_\epsilon^2}} \left(\mathbf{h} + \boldsymbol{\varepsilon}\right) \tag{4.8}$$

where $\sigma_h^2 = E\left\{\mathbf{h}^H\mathbf{h}\right\}$, $\sigma_\epsilon^2 = E\left\{\boldsymbol{\varepsilon}^H\boldsymbol{\varepsilon}\right\}$ and the term into the squared-root is a terms that keeps the power of the estimated channel independent of the power of $\boldsymbol{\varepsilon}$.

The equation in (4.8) is valid to CSI errors both at the transmitter and at the receiver (see Chapter 3 for details). The transmitter may update CSI based on a feedback link and the prediction of the future channel state from previous CSI when channel is time-varying; thus, the feedback delay and the prediction error are the most important causes of imperfect channel estimates at this side. At the receiver side, the degradation of channel estimate quality is mainly caused by the additive noise. These sources of CSI degradation can be collected in the channel estimation error $\boldsymbol{\varepsilon}$, which is modelled as a zero mean Gaussian process independent of the true channel $\mathbf{h}$ and with covariance matrix (see Appendix 4.A):

$$E\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^H\right\} = \mathbf{E}\otimes\mathbf{I} \tag{4.9}$$

where matrix $\mathbf{E}$, with the same structure as matrix $\mathbf{P}$, contains the DFT of the variance of the channel estimation error for each tap, and it is also assumed to be known.

The Bayesian approach adopted in this thesis relies on the hypothesis that the channel $\mathbf{h}$ and its estimate $\hat{\mathbf{h}}$ are jointly Gaussian. Under this assumption the statistics of $\mathbf{h}$ given $\hat{\mathbf{h}}$ are also Gaussian with conditional mean $\mathbf{m}_{\mathbf{h}|\hat{\mathbf{h}}}$ and conditional covariance $\mathbf{C}_{\mathbf{h}|\hat{\mathbf{h}}}$ defined as [Kay93, p.324]:

$$\begin{aligned}\mathbf{m}_{\mathbf{h}|\hat{\mathbf{h}}} &= \mathbf{m}_{\mathbf{h}} + \mathbf{C}_{\mathbf{h}\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}}\hat{\mathbf{h}}}^{-1}\left(\hat{\mathbf{h}} - \mathbf{m}_{\hat{\mathbf{h}}}\right)\\ \mathbf{C}_{\mathbf{h}|\hat{\mathbf{h}}} &= \mathbf{C}_{\mathbf{h}\mathbf{h}} - \mathbf{C}_{\mathbf{h}\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}}\hat{\mathbf{h}}}^{-1}\mathbf{C}_{\hat{\mathbf{h}}\mathbf{h}}\end{aligned} \tag{4.10}$$

In our particular case, following (4.7)-(4.9), the conditional mean and conditional covariance have been derived in Appendix 4.A:

$$\begin{aligned}\mathbf{m}_{\mathbf{h}|\hat{\mathbf{h}}} &= \sqrt{1+\frac{\sigma_\epsilon^2}{\sigma_h^2}}\left((\mathbf{P}\otimes\mathbf{R})\left(\mathbf{P}\otimes\mathbf{R}+\mathbf{E}\otimes\mathbf{I}\right)^{-1}\right) \quad \hat{\mathbf{h}}\\ &\quad + \quad (\mathbf{E}\otimes\mathbf{I})\left(\mathbf{P}\otimes\mathbf{R}+\mathbf{E}\otimes\mathbf{I}\right)^{-1} \quad \mathbf{m}_{\mathbf{h}}\end{aligned} \tag{4.11}$$

$$\mathbf{C}_{\mathbf{h}|\hat{\mathbf{h}}} = (\mathbf{P}\otimes\mathbf{R})\left(\mathbf{P}\otimes\mathbf{R}+\mathbf{E}\otimes\mathbf{I}\right)^{-1}(\mathbf{E}\otimes\mathbf{I}) \tag{4.12}$$

and for the particular case of uncorrelated antennas (i.e., $\mathbf{R}=\mathbf{I}$), these conditional mean and conditional covariance expressions reduce to:

$$\mathbf{m}_{\mathbf{h}|\hat{\mathbf{h}}} = \sqrt{1+\frac{\sigma_\epsilon^2}{\sigma_h^2}}\left(\mathbf{P}\left(\mathbf{P}+\mathbf{E}\right)^{-1}\otimes\mathbf{I}\right)\hat{\mathbf{h}} + \left(\mathbf{E}\left(\mathbf{P}+\mathbf{E}\right)^{-1}\otimes\mathbf{I}\right)\mathbf{m}_{\mathbf{h}} \tag{4.13}$$

$$\mathbf{C}_{\mathbf{h}|\hat{\mathbf{h}}} = \left(\mathbf{P}\left(\mathbf{P}+\mathbf{E}\right)^{-1}\mathbf{E}\right)\otimes\mathbf{I} \tag{4.14}$$

Next sections will focus on the conditional covariance for a specific subcarrier $\mathbf{h}_k$ under the hypothesis of uncorrelated antennas. In that case matrix, $\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}$ is given by:

$$\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}} = \left[\mathbf{P}\left(\mathbf{P}+\mathbf{E}\right)^{-1}\mathbf{E}\right]_{[k,k]}\mathbf{I} = \omega\mathbf{I} \tag{4.15}$$

where subindex $[\cdot]_{[k,k]}$ refers to the (k,k)th element and, the second equality, which proves that the conditional covariance is given by the scalar ω is the same for all subcarriers, follows from the circulant structure of $\mathbf{P}$ and $\mathbf{E}$ matrices (see Appendix 4.A) that forces all the elements in the diagonal of $\mathbf{P}\left(\mathbf{P}+\mathbf{E}\right)^{-1}\mathbf{E}$ to be equal. The asymptotic values when channel is perfectly known and when it is completely unknown are:

$$\omega = \begin{cases} 0 & \text{Perfect CSI at the Transmitter} \\ \sigma_h^2 & \text{No CSI at the Transmitter} \end{cases} \tag{4.16}$$

4.2.2 Equivalent channel

A detailed analysis of (4.13) shows that the conditional mean can be understood as an equivalent channel that exploits the correlation between subcarriers and the channel uncertainty structure to mitigate the degradation due to CSI errors (see Figure 4.2). We will denote this equivalent channel as $\hat{\mathbf{h}}_k^{eq} = \mathbf{m}_{\mathbf{h}_k|\hat{\mathbf{h}}}$. As shown in Figure 4.2 the equivalent channel for a particular subcarrier, under the assumption of uncorrelated antennas, can be expressed as a linear combination of the channel mean $\mathbf{m}_{\mathbf{h}}$ and channel estimates $\hat{\mathbf{h}}$ for all subcarriers:

$$\hat{\mathbf{h}}_k^{eq} = \sum_{l=1}^{K} \hat{\beta}_k(l)\, \hat{\mathbf{h}}_l + \bar{\beta}_k(l)\, \mathbf{m}_{\mathbf{h}_l} \tag{4.17}$$

where $\hat{\beta}_k(l)$ and $\bar{\beta}_k(l)$ are the coefficients for the linear combination. As shown in Figure 4.2 those coefficients are defined as follows:

$$\begin{aligned} \hat{\beta}_k(l) &= \sqrt{1+\tfrac{\sigma_\epsilon^2}{\sigma_h^2}}\, \mathbf{P}_{[k,:]} \left(\mathbf{P}+\mathbf{E}\right)^{-1}_{[:,l]} \\ \bar{\beta}_k(l) &= \mathbf{E}_{[k,:]} \left(\mathbf{P}+\mathbf{E}\right)^{-1}_{[:,l]} \end{aligned} \tag{4.18}$$

where $\mathbf{A}_{[l,:]}$ denotes the lth row of matrix $\mathbf{A}$, and similarly, $\mathbf{A}_{[:,l]}$ denotes its lth column.

The weighting coefficients $\hat{\beta}_k(l)$ and $\bar{\beta}_k(l)$ exploit the subcarrier correlation structure defined by the power delay profile to mitigate the mismatch between the real and the estimated channels. Note that when perfect channel knowledge is available (i.e., matrix $\mathbf{E}=\mathbf{0}$), the coefficients $\bar{\beta}_k(j)$ are all zero, and $\hat{\beta}_k(j) = \delta\left(k-j\right)$, leading to the equivalent channel: $\hat{\mathbf{h}}_k^{eq} = \hat{\mathbf{h}}_k$. In all other cases, the coefficients $\hat{\beta}_k(j)$ and $\bar{\beta}_k(j)$ weight vectors $\hat{\mathbf{h}}_j$ and $\mathbf{m}_{\mathbf{h}_j}$ for all subcarriers to reduce the uncertainty in the specific kth subcarrier. The asymptotic values when channel is perfectly known and when it is completely unknown are:

$$\hat{\mathbf{h}}_k^{eq} = \begin{cases} \hat{\mathbf{h}}_k = \mathbf{h}_k & \text{Perfect CSI at the Transmitter} \\ 0 & \text{No CSI at the Transmitter} \end{cases} \tag{4.19}$$

$$\mathbf{m}_{\mathbf{h}|\hat{\mathbf{h}}} = \sqrt{1+\frac{\sigma_\epsilon^2}{\sigma_h^2}} \quad \left(\mathbf{P}\left(\mathbf{P}+\mathbf{E}\right)^{-1}\otimes\mathbf{I}\right)\hat{\mathbf{h}} \qquad +\ldots$$

$$\begin{bmatrix}\hat{\mathbf{h}}_1^{eq}\\ \vdots\\ \hat{\mathbf{h}}_k^{eq}\\ \vdots\\ \hat{\mathbf{h}}_K^{eq}\end{bmatrix} = \begin{bmatrix}\beta_1(1)\mathbf{I} & \ldots & \beta_1(k)\mathbf{I} & \ldots & \beta_1(K)\mathbf{I}\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ \beta_k(1)\mathbf{I} & \ldots & \beta_k(k)\mathbf{I} & \ldots & \beta_k(K)\mathbf{I}\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ \beta_K(1)\mathbf{I} & \ldots & \beta_K(k)\mathbf{I} & \ldots & \beta_K(K)\mathbf{I}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{h}}_1\\ \vdots\\ \hat{\mathbf{h}}_k\\ \vdots\\ \hat{\mathbf{h}}_K\end{bmatrix} +\ldots$$

$$where\ \ \beta_k(l) = \sqrt{1+\frac{\sigma_\epsilon^2}{\sigma_h^2}}\,\mathbf{P}_{[k,:]}\left(\mathbf{P}+\mathbf{E}\right)^{-1}_{[:,l]}$$

Figure 4.2: Graphical structure of the first term in the conditional mean. $\mathbf{A}_{[l,:]}$ denotes the lth row of matrix $\mathbf{A}$ whereas $\mathbf{A}_{[:,l]}$ denotes its lth column. A similar structure can be found for the second term

Hereafter we will refer to the equivalent channel in any of its two forms: $\hat{\mathbf{h}}_k^{eq} = \mathrm{vec}\left(\hat{\mathbf{H}}_k^{eq}\right)$, and its SVD will be denoted as:

$$\hat{\mathbf{H}}_k^{eq} = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{V}_k^H \tag{4.20}$$

4.2.3 Unbalanced CSI between the transmitter and the receiver

Some algorithms, as will be the case of the minimum MSE criterion, allow to design, in a single step, the linear transformations $\mathbf{F}_k$ and $\mathbf{G}_k$ when the same imperfect CSI is assumed at both sides of the link. However, when the CSI quality is unbalanced between the transmitter and the receiver, there is no way to introduce, into the transmitter cost function, different levels of CSI. Alternatively, we divide the design in two steps:

- A first step is carried out at the transmitter, which we will call the *a priori* design, and whose aim is to design the transmitter matrix $\mathbf{F}_k$ based on the prediction of the channel impulse response before the signal is actually transmitted (see Chapter 3 for details on how the transmitter is aided from the receiver by means of the feedback channel link).
- A second step is carried out at the receiver and it is called the *a posteriori* design, whose aim is to design the optimal receiver based on the channel knowledge at the receiver and the specific linear transformation $\mathbf{F}_k$, once the transmitted signal has been received and the CSI has been updated.

4.3 MMSE design

This section aims at designing the linear transmitter $\mathbf{F}_k$ that minimizes the MSE at the output of the linear receiver $\mathbf{G}_k$ subject to an average power constraint across all subcarriers and antennas when channel estimates are noisy. According to the classification of the *a priori* and *a posteriori* designs introduced at the end of the previous section, we focus, here, on the *a priori* design, whereas the *a posteriori* design, which designs the optimal receiver based on the ML receiver, is described in Section 4.5. The MMSE cost function requires a linear receiver $\mathbf{G}_k$ which implicitly assumes that the same CSI quality is available at the transmitter and at the receiver. However, in most cases, the CSI quality in unbalanced, and thus, the linear transformations $\mathbf{F}_k$ and $\mathbf{G}_k$ can not be designed in a single step. That is the case of our concerning problem. To solve this situation, in this section we introduce into the optimization problem the linear receiver $\mathbf{G}_k$ but this matrix is never used at the receiver because finally an ML receiver will be used.

Using the system model introduced in (4.3) the objective function that minimizes the MSE between the transmitted symbols $\mathbf{x}_k$ and the received symbols $\mathbf{r}_k$ is given by:

$$\min_{\mathbf{F}_k,\mathbf{G}_k} \quad \sum_{k=1}^{K} E\left\{|\mathbf{r}_k - \mathbf{x}_k|^2\right\} \quad \text{subject to} \quad \sum_{k=1}^{K} \mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\} = P_0 \tag{4.21}$$

where the expectation is computed over the additive Gaussian noise $\mathbf{n}_k$, the discrete-time sequence $\mathbf{x}_k$, and the true channel impulse response $\mathbf{h}_k$ given the channel estimate at the transmitter $\hat{\mathbf{h}}$. For the sake of generality, it is first assumed that the transmitted symbols are known, which is a realistic assumption when training sequences are used, and the purpose is to improve the channel estimation (i.e., in *Data-Aided* (DA) schemes), and the result is later averaged over the vector $\mathbf{x}_k$ to extend the solution to schemes where the transmitted sequences are not known at the receiver. For the sake of clarity we will use the superscript *'Robust-DA'* to refer to the first solution and *'Robust'* to denote the second one.

Under the previous premises, the design of matrices $\mathbf{F}_k$ and $\mathbf{G}_k$ given a transmitted sequence $\mathbf{x}_k$ and a channel estimate $\hat{\mathbf{h}}_k$ is done minimizing:

$$\min_{\mathbf{F}_k,\mathbf{G}_k} \quad \sum_{k=1}^{K} E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{E_{\mathbf{n}_k}\left\{|\mathbf{r}_k - \mathbf{x}_k|^2\right\}\right\} \quad \text{subject to} \quad \sum_{k=1}^{K} \mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\} = P_0 \tag{4.22}$$

Note that this minimization problem differs to those proposed in [Bar00] and [Sam01] on the conditional expectation $E_{\mathbf{h}_k|\hat{\mathbf{h}}}\{\cdot\}$, which mitigates the impact of the channel uncertainty on the algorithm performance, providing a robust design that adapts to the channel estimation quality.

4.3.1 Cost function

This section expands the cost function in (4.22) making use of the conditional mean and covariance given in (4.11) and (4.12). Assuming that the transmitted symbols are known, and

substituting vector $\mathbf{r}_k$ by the expression introduced in (4.5), the expectation over the true channel impulse response $\mathbf{h}_k$ and the additive Gaussian noise $\mathbf{n}_k$ is given by:

$$\begin{aligned}\xi_k^{Robust-DA} &= E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{E_{\mathbf{n}_k}\left\{\left|\mathbf{r}_k-\mathbf{x}_k\right|^2\right\}\right\}\\ &= E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{E_{\mathbf{n}_k}\left\{\left|\mathbf{A}_k\mathbf{h}_k+\mathbf{G}_k\mathbf{n}_k-\mathbf{x}_k\right|^2\right\}\right\}\\ &= E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\left|\mathbf{A}_k\mathbf{h}_k-\mathbf{x}_k\right|^2\right\}+E_{\mathbf{n}_k}\left\{\left|\mathbf{G}_k\mathbf{n}_k\right|^2\right\}\end{aligned} \tag{4.23}$$

where the independence between the transmitted symbols and the noise has been used in the last equality. Next, we will compute the two expectation separately. The first term can be written as:

$$\begin{aligned}E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\left|\mathbf{A}_k\mathbf{h}_k-\mathbf{x}_k\right|^2\right\} &= E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\left|\left(\mathbf{A}_k\left(\mathbf{h}_k-\mathbf{h}_k^{eq}\right)+\mathbf{A}_k\mathbf{h}_k^{eq}\right)-\mathbf{x}_k\right|^2\right\}\\ &= E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\left|\mathbf{A}_k\left(\mathbf{h}_k-\mathbf{h}_k^{eq}\right)\right|^2\right\}+\left|\mathbf{A}_k\mathbf{h}_k^{eq}-\mathbf{x}_k\right|^2\\ &= \mathrm{Tr}\left\{\mathbf{A}_k E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\left(\mathbf{h}_k-\mathbf{h}_k^{eq}\right)\left(\mathbf{h}_k-\mathbf{h}_k^{eq}\right)^H\right\}\mathbf{A}_k^H\right\}+\left|\mathbf{A}_k\mathbf{h}_k^{eq}-\mathbf{x}_k\right|^2\\ &= \mathrm{Tr}\left\{\mathbf{A}_k\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}\mathbf{A}_k^H\right\}+\left|\mathbf{A}_k\mathbf{h}_k^{eq}-\mathbf{x}_k\right|^2\end{aligned} \tag{4.24}$$

where $E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\left(\mathbf{h}_k-\mathbf{h}_k^{eq}\right)\right\}=\mathbf{0}$ has been used in the second equality. The second expectation becomes:

$$\begin{aligned}E_{\mathbf{n}_k}\left\{\left|\mathbf{G}_k\mathbf{n}_k\right|^2\right\} &= \mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k E_{\mathbf{n}_k}\left\{\left|\mathbf{n}_k\right|^2\right\}\right\}\\ &= \sigma_n^2\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}\end{aligned} \tag{4.25}$$

And, substituting (4.24) and (4.25) into the cost function in (4.23) the contribution of kth subcarrier to the robust cost function is given by:

$$\xi_k^{Robust-DA}=\left|\mathbf{A}_k\hat{\mathbf{h}}_k^{eq}-\mathbf{x}_k\right|^2+\sigma_n^2\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}+\mathrm{Tr}\left\{\mathbf{A}_k\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}\mathbf{A}_k^H\right\} \tag{4.26}$$

A simple comparison of equations (4.3) and (4.5) shows that the term $\mathbf{A}_k\hat{\mathbf{h}}_k^{eq}$ can be written as $\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k\mathbf{x}_k$. Hence, a more understandable expression for (4.26) can be written as follows:

$$\xi_k^{Robust-DA}=\left|\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)\mathbf{x}_k\right|^2+\sigma_n^2\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}+\mathrm{Tr}\left\{\mathbf{A}_k\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}\mathbf{A}_k^H\right\} \tag{4.27}$$

and the optimization criterion is finally expressed as:

$$\min_{\mathbf{F}_k,\mathbf{G}_k}\ \sum_{k=1}^{K}\xi_k^{Robust-DA}\quad\text{subject to}\quad\sum_{k=1}^{K}\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\}=P_0 \tag{4.28}$$

This cost function is only useful when sequence $\mathbf{x}_k$ is known at the receiver; however, in most cases, the transmitted symbols are not known. Hence, when only the statistics of the transmitted symbols are known at the receiver, the expectation of (4.27) over the transmitted sequence allows the design of optimal linear transformations $\mathbf{F}_k$ and $\mathbf{G}_k$ independently of the specific information symbols:

$$\xi_k^{Robust}=E_{\mathbf{x}_k}\left\{\left|\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)\mathbf{x}_k\right|^2+\sigma_n^2\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}+\mathrm{Tr}\left\{\mathbf{A}_k\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}\mathbf{A}_k^H\right\}\right\} \tag{4.29}$$

Next, we compute the expectation over the first and third terms separately. The expectation of the first term is a simple task:

$$\begin{aligned} E_{\mathbf{x}_k}\left\{\left|\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)\mathbf{x}_k\right|^2\right\} &= \mathrm{Tr}\left\{\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)^H\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)E_{\mathbf{x}_k}\left\{|\mathbf{x}_k|^2\right\}\right\} \\ &= \sigma_x^2\mathrm{Tr}\left\{\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)^H\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)\right\} \end{aligned} \tag{4.30}$$

whereas the expectation of the third term requires to substitute $\mathbf{A}_k$ by the expression given in (4.6) and use the Kronecker product identity $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=\mathbf{AC}\otimes\mathbf{BD}$. Therefore:

$$\begin{aligned} E_{\mathbf{x}_k}\left\{\mathrm{Tr}\left\{\mathbf{A}_k\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}\mathbf{A}_k^H\right\}\right\} &= E_{\mathbf{x}_k}\left\{\mathrm{Tr}\left\{\left((\mathbf{F}_k\mathbf{x}_k)^T\otimes\mathbf{G}_k\right)\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}\left((\mathbf{F}_k\mathbf{x}_k)^T\otimes\mathbf{G}_k\right)^H\right\}\right\} \\ &= E_{\mathbf{x}_k}\left\{\mathrm{Tr}\left\{\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}\left((\mathbf{F}_k\mathbf{x}_k)^*\otimes\mathbf{G}_k^H\right)\left((\mathbf{F}_k\mathbf{x}_k)^T\otimes\mathbf{G}_k\right)\right\}\right\} \\ &= \mathrm{Tr}\left\{\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}\left(E_{\mathbf{x}_k}\left\{\mathbf{F}_k^*\mathbf{x}_k^*\mathbf{x}_k^T\mathbf{F}_k^T\right\}\otimes\mathbf{G}_k^H\mathbf{G}_k\right)\right\} \\ &= \sigma_x^2\mathrm{Tr}\left\{\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}\left(\mathbf{F}_k^*\mathbf{F}_k^T\otimes\mathbf{G}_k^H\mathbf{G}_k\right)\right\} \end{aligned} \tag{4.31}$$

Substituting (4.30) and (4.31) into the cost function in (4.29) the contribution of kth subcarrier to the robust cost function is given by:

$$\begin{aligned} \xi_k^{Robust} &= \sigma_x^2\mathrm{Tr}\left\{\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)^H\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)\right\} \\ &\quad +\sigma_n^2\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}+\sigma_x^2\mathrm{Tr}\left\{\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}\left(\mathbf{F}_k^*\mathbf{F}_k^T\otimes\mathbf{G}_k^H\mathbf{G}_k\right)\right\} \end{aligned} \tag{4.32}$$

and the whole cost function is finally expressed as:

$$\min_{\mathbf{F}_k,\mathbf{G}_k}\ \sum_{k=1}^{K}\xi_k^{Robust}\quad\text{subject to}\quad\sum_{k=1}^{K}\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\}=P_0 \tag{4.33}$$

It is interesting to analyze the similarities between the derived expression and the cost function appearing in the nonrobust solution in [Bar00] and [Sam01], where the MSE for the kth subcarrier is written as:

$$\begin{aligned} \xi_k^{Nonrobust} &= \sigma_x^2\mathrm{Tr}\left\{\left(\mathbf{G}_k\hat{\mathbf{H}}_k\mathbf{F}_k-\mathbf{I}\right)^H\left(\mathbf{G}_k\hat{\mathbf{H}}_k\mathbf{F}_k-\mathbf{I}\right)\right\} \\ &\quad +\sigma_n^2\ \mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\} \end{aligned} \tag{4.34}$$

The nonrobust cost function, as shown in (4.34), contains two terms: The first one is a measure of the ISI, whereas the second one introduces into the cost function the contribution of the noise. Note that the residual ISI and the noise also appear in the robust cost function derived in (4.32), but there, the channel estimate $\hat{\mathbf{H}}_k$ has been replaced by the equivalent channel $\hat{\mathbf{H}}_k^{eq}$. Thus, the robust cost function exploits the correlation between subcarriers to mitigate the channel uncertainty, improving its performance. A second difference between the robust and the

nonrobust solutions is the third term in (4.32) that does not appear in (4.34). As the covariance matrix $\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}$ is a measure of the channel estimation error, the third term introduces into the score function the mismatch between the real and the estimated channel due to the estimation errors.

4.3.2 Closed-form solution

A solution for the optimization of (4.32) can only be obtained by means of numerical techniques. Unfortunately, a closed-form solution is feasible only for uncorrelated antennas. Hence, in this section, we will assume that the channel is spatially uncorrelated. Substituting covariance matrix $\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}$ given in (4.15) into the third term in (4.32) and using the identity $\mathrm{Tr}\{\mathbf{A}\otimes\mathbf{B}\} = \mathrm{Tr}\{\mathbf{A}\}\otimes\mathrm{Tr}\{\mathbf{B}\}$, we get:

$$\begin{aligned}\sigma_x^2\mathrm{Tr}\left\{\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}}\left(\mathbf{F}_k^*\mathbf{F}_k^T\otimes\mathbf{G}_k^H\mathbf{G}_k\right)\right\} &= \omega\sigma_x^2\mathrm{Tr}\left\{\left(\mathbf{F}_k^*\mathbf{F}_k^T\otimes\mathbf{G}_k^H\mathbf{G}_k\right)\right\}\\ &= \omega\sigma_x^2\mathrm{Tr}\left\{\mathbf{F}_k^*\mathbf{F}_k^T\right\}\otimes\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}\\ &= \omega\sigma_x^2\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\}\otimes\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}\end{aligned} \tag{4.35}$$

and hence, the cost function for the specific case of uncorrelated antennas simplifies to:

$$\begin{aligned}\xi_k^{Robust} &= \sigma_x^2\mathrm{Tr}\left\{\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)^H\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)\right\}\\ &\quad+\sigma_n^2\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}+\omega\sigma_x^2\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\}\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}\end{aligned} \tag{4.36}$$

In Appendix 4.B, it is shown that using the SVD of the equivalent channel matrix $\hat{\mathbf{H}}_k^{eq}$ for each subcarrier $\hat{\mathbf{H}}_k^{eq} = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{V}_k^H$, the cost function in (4.36) is minimized when $\mathbf{F}_k$ and $\mathbf{G}_k$ matrices have the following structure:

$$\mathbf{F}_k = \mathbf{V}_k\mathbf{\Phi}_k\mathbf{T}_k \qquad \mathbf{G}_k = \mathbf{T}_k^H\mathbf{\Gamma}_k\mathbf{\Lambda}_k^{\#}\mathbf{U}_k^H \tag{4.37}$$

where $\mathbf{T}_k$ is an irrelevant unitary matrix (for the sake of simplicity we chose $\mathbf{T}_k = \mathbf{I}$), $\mathbf{\Phi}_k$ is a diagonal matrix that sets the power distribution policy (the square of the elements of the diagonal $\phi_k^2(l)$ define the power allocated to each channel mode), and $\mathbf{\Gamma}_k$ is also a diagonal matrix applied to the symbols at the channel output. According to the SVD decomposition in (4.37), the function to minimize, including all the constraints, becomes:

$$\begin{aligned}\min_{\mathbf{\Gamma}_k,\mathbf{\Phi}_k}\quad & \sum_{k=1}^{K}\sigma_x^2\mathrm{Tr}\left\{\left(\mathbf{\Gamma}_k\mathbf{\Phi}_k-\mathbf{I}\right)^H\left(\mathbf{\Gamma}_k\mathbf{\Phi}_k-\mathbf{I}\right)\right\}+\sigma_n^2\mathrm{Tr}\left\{\mathbf{\Gamma}_k\mathbf{\Gamma}_k^H\mathbf{\Lambda}_k^{\#2}\right\}\\ &\quad+\omega\,\sigma_x^2\mathrm{Tr}\left\{\mathbf{\Phi}_k^H\mathbf{\Phi}_k\right\}\mathrm{Tr}\left\{\mathbf{\Gamma}_k\mathbf{\Gamma}_k^H\mathbf{\Lambda}_k^{\#2}\right\}\\ \text{subject to}\quad & \sum_{k=1}^{K}\mathrm{Tr}\left\{\mathbf{\Phi}_k^H\mathbf{\Phi}_k\right\}=\sum_{k=1}^{K}\sum_{j=1}^{M}\phi_k^2(j)=P_0\\ & \phi_k^2(l)\geq 0\qquad k=1\ldots K\qquad l=1\ldots M\end{aligned} \tag{4.38}$$

It can be easily proven the convexity of the objective function both in $\mathbf{\Gamma}_k$ and $\mathbf{\Phi}_k$ since the Hessian if always semipositive definite. Hence, the optimization problem can be solved by imposing the *Karush-Kuhn-Tucker* (KKT) conditions [Boy04, p.243]. Denoting $\gamma_k(l)$ and $\lambda_k^{eq}(l)$ as the elements of the diagonal matrices $\mathbf{\Gamma}_k$ and $\mathbf{\Lambda}_k$, respectively, the KKT conditions are written as follows:

$$\left(\omega\ \sigma_x^2 \sum_{j=1}^{M} \gamma_k^2(j)|\lambda_k^{eq}(j)|^{-2} - \nu_{kl} + \mu\right)\phi_k(l) + \sigma_x^2\gamma_k^2(l)\phi_k(l) - \sigma_x^2\gamma_k(l) = 0 \tag{4.39a}$$

$$\left(\omega\ \sigma_x^2 \sum_{j=1}^{M} \phi_k^2(j) + \sigma_n^2\right)|\lambda_k^{eq}(l)|^{-2}\gamma_k(l) + \sigma_x^2\phi_k^2(l)\gamma_k(l) - \sigma_x^2\phi_k(l) = 0 \tag{4.39b}$$

$$\sum_{k=1}^{K}\sum_{j=1}^{M} \phi_k^2(j) - P_0 = 0 \tag{4.39c}$$

$$\phi_k^2(l) \geq 0 \qquad k = 1\ldots K \qquad l = 1\ldots M \tag{4.39d}$$

$$\nu_{kl} \geq 0 \qquad k = 1\ldots K \qquad l = 1\ldots M \tag{4.39e}$$

$$\nu_{kl}\phi_k^2(l) = 0 \qquad k = 1\ldots K \qquad l = 1\ldots M \tag{4.39f}$$

where the first and second conditions can be found differentiating the Lagrangian associated with the problem (4.38) with respect to $\phi_k(l)$ and $\gamma_k(l)$, respectively, and equating them to zero; μ is the Lagrange multiplier for the equality constraint (4.39c) and ν_{kl} are the multipliers for the inequality constraint (4.39d); finally, the two last equations (4.39e) and (4.39f) are given by the complementary slackness condition.

In order to solve this equations we will first collapse the first and second conditions into a single equation. Multiplying the first equation by $\phi_k(l)$, and second equation by $\gamma_k(l)$, we get the following identity:

$$\left(\omega\ \sigma_x^2 \sum_{j=1}^{M} \gamma_k^2(j)|\lambda_k^{eq}(j)|^{-2} + \mu\right)\phi_k^2(l) = \left(\omega\ \sigma_x^2 \sum_{j=1}^{M} \phi_k^2(j) + \sigma_n^2\right)|\lambda_k^{eq}(l)|^{-2}\gamma_k^2(l) \tag{4.40}$$

where the constraint (4.39f) has been considered to eliminate the dependency with respect to ν_{kl} variables. From this point μ can be solved as:

$$\mu = \frac{\sigma_n^2}{P_0}\sum_{k=1}^{K}\sum_{j=1}^{L} \gamma_k^2(j)|\lambda_k^{eq}(j)|^{-2} \tag{4.41}$$

Replacing μ into (4.40), the next relation between $\gamma_k(l)$ and $\phi_k(l)$ is derived:

$$\phi_k(l) = C_k|\lambda_k^{eq}(l)|^{-1}\gamma_k(l) \tag{4.42}$$

where C_k is a scalar that must be found. Substituting (4.42) in (4.40), it can be seen that the equality is only accomplished when the scalar C_k is constant for all subcarriers, i.e., $C_k = C$.

Introducing (4.41) and (4.42) into (4.39a), and after some manipulations, it is straightforward to derive an equivalent expression to be solved:

$$\omega\sum_{j=1}^{M}\phi_k^2(j)+\frac{\sigma_n^2}{\sigma_x^2}-\nu_{kl}+\phi_k^2(l)|\lambda_k^{eq}(l)|^2=C|\lambda_k^{eq}(l)| \tag{4.43}$$

Next we will eliminate the slack variables ν_{kl} by forcing the conditions (4.39e)-(4.39f):

- If $\phi_k^2(l)=0$ hence, according to (4.39e), the variables ν_{kl} must satisfy $\nu_{kl}>0$. Substituting this condition into (4.43) we get the next inequality:

$$C|\lambda_k^{eq}(l)|<\omega\sum_{\substack{j=1\\ j\neq l}}^{M}\phi_k^2(j)+\frac{\sigma_n^2}{\sigma_x^2} \tag{4.44}$$

- If $\phi_k^2(l)>0$ we have $\nu_{kl}=0$ in order to satisfy (4.39f). And thus (4.43) becomes:

$$\omega\sum_{j=1}^{M}\phi_k^2(j)+\frac{\sigma_n^2}{\sigma_x^2}+\phi_k^2(l)|\lambda_k^{eq}(l)|^2=C|\lambda_k^{eq}(l)| \tag{4.45}$$

Next we will solve equation (4.45) assuming that $\phi_k^2(l)$ satisfies $\phi_k^2(l)>0$ and we will use the inequality in (4.44) to force $\phi_k^2(l)=0$ when we will get a negative solution. For each subcarrier k, the equality in (4.45) provides a set of M equations that are linear on the $M+1$ unknowns given by the M elements $\phi_k^2(l)$ $(l=1\dots M)$ and the scalar C. Hence, all the unknowns can be found when simultaneously solving the sets of equations given by (4.45) combined with the power constraint. Alternatively, their resolution can be simplified by solving the problem for each subcarrier in two stages, as next shown.

First, (4.45) is solved for each subcarrier:

$$\begin{bmatrix} \omega+|\lambda_k^{eq}(1)|^2 & \omega & \dots & \omega & -|\lambda_k^{eq}(1)| \\ \omega & \omega+|\lambda_k^{eq}(2)|^2 & \ddots & \omega & -|\lambda_k^{eq}(2)| \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \omega & \dots & \dots & \omega+|\lambda_k^{eq}(M)|^2 & -|\lambda_k^{eq}(M)| \end{bmatrix}\begin{bmatrix} \boldsymbol{\phi}_k^2(1) \\ \vdots \\ \boldsymbol{\phi}_k^2(M) \\ C \end{bmatrix}=\begin{bmatrix} -\sigma_n^2/\sigma_x^2 \\ \vdots \\ \vdots \\ -\sigma_n^2/\sigma_x^2 \end{bmatrix} \tag{4.46}$$

Storing the $\phi_k^2(l)$ values that correspond to the same subcarrier in vector $\boldsymbol{\phi}_k^2$, and defining $\mathbf{B}$ matrix, the set of equations can be written as:

$$\mathbf{B}\begin{bmatrix} \boldsymbol{\phi}_k^2 \\ C \end{bmatrix}=-\frac{\sigma_n^2}{\sigma_x^2}\ \mathbf{1}_{M\times 1} \tag{4.47}$$

The solution to the set of equations can be found as the summation of the particular solution given by (4.47) and the kernel of matrix $\mathbf{B}$:

$$\begin{bmatrix} \boldsymbol{\phi}_k^2 \\ C \end{bmatrix} = -\frac{\sigma_n^2}{\sigma_x^2} \mathbf{B}^{\#}\mathbf{1} + D_k\mathbf{v} \tag{4.48}$$

where $\mathbf{v}$ is a vector spanning the null-space of matrix $\mathbf{B}$ and the scalar D_k is an arbitrary constant. Afterwards, the values for the constant D_k are found by solving a second set of equations including the power constraint:

$$\sum_{k=0}^{N-1} \mathbf{1}^T \boldsymbol{\phi}_k^2 = P_0 \tag{4.49}$$

Notice that the power allocated to each subcarrier in vector $\boldsymbol{\phi}_k^2$ should satisfy $\boldsymbol{\phi}_k^2(l) \geq 0$. If any of the components obtained when solving (4.48) were negative, the power allocated to that subcarrier should be set to zero $\boldsymbol{\phi}_k^2(l) = 0$ and the described procedure should be repeated for the rest of subcarriers until all the elements in vector $\boldsymbol{\phi}_k^2$ satisfy $\boldsymbol{\phi}_k^2(l) \geq 0$. A rigorous justification of this step can be found in the inequality (4.44). It is easy to analyze that when $\boldsymbol{\phi}_k^2(l) < 0$ this inequality holds and thus to satisfy (4.39e)-(4.39f) conditions: $\boldsymbol{\phi}_k^2(l) = 0$.

A study of the complexity of this robust solution, as well as a comparison with the nonrobust one, can be found in [IST03]. In that report it is shown that a closed-form solution can be found to solve the problem in (4.47)-(4.49) as a function of $|\lambda_k^{eq}(l)|^{-1}, |\lambda_k^{eq}(l)|^{-2}, |\lambda_k^{eq}(l)|^{-3}, |\lambda_k^{eq}(l)|^{-4}$ $l = 1 \ldots M$ without the need to solve the set of equations in (4.47).

4.3.3 Asymptotic performance

In this section, the robust MMSE algorithm is analyzed for the extreme cases where the uncertainty is very high or very low. When channel knowledge at the transmitter is perfect, i.e., when $\hat{\mathbf{H}}_k^{eq} = \mathbf{H}$ and $\omega = 0$ [see equations (4.16) and (4.19)], the robust cost function in (4.32) converges to the nonrobust solution in equation (4.34) and, therefore, both coincide. When the channel uncertainty is very high, the MMSE design is meaningless since the design is based on the assumption that neither the transmitter nor the receiver know the channel.

4.4 Minimum uncoded BER design

This section proposes the *a priori* design (see Section 4.2.3) for the set of linear precoder matrices $\mathbf{F}_k$ that minimize the uncoded BER in a ML receiver subject to an average power constraint. This algorithm considers the reliability of channel estimation in the cost function, adapting to

the channel uncertainty and providing a solution that goes from the configuration for open-loop to the closed-loop with perfect CSI as estimation errors diminish (in a similar way as [Ong02] and [Jön02]). The filtering matrix at the receiver is set to the identity matrix $\mathbf{G}_k = \mathbf{I}$ since the BER cannot be improved by means of a linear filtering stage at this side. The design is formulated for the general MIMO OFDM case, including single antenna transmission and beamforming [Ong02],[Jön02] as particular cases.

The cost function to be optimized is described for a general constellation $\{s_1 \dots s_{\mathcal{N}}\}$ of size $\mathcal{N}$, even though final equations are shown for the particular case of QPSK modulation. In order to be able to derive a closed-form solution, two main assumptions have been made:

A1) The receiver is operating at high SNR.

A2) The function $\mathcal{Q}\left(\sqrt{x}\right)$ can be approximated with small error as (see Appendix 4.C):

$$\mathcal{Q}\left(\sqrt{x}\right) \simeq \delta e^{-\alpha x} \tag{4.50}$$

For a given channel realization $\mathbf{H}_k$, the uncoded BER for the ML receiver can be written in terms of the *Pairwise Error Probability* (PEP) of detecting symbol $\mathbf{x}_{kj}$ when the symbol $\mathbf{x}_{ki}$ was transmitted:

$$BER(\mathbf{H}_1 \dots \mathbf{H}_K) = \frac{1}{\mathcal{N}^{2M} K} \sum_{k=1}^{K} \sum_{i=1}^{\mathcal{N}^M} \sum_{j=1}^{\mathcal{N}^M} B(\mathbf{x}_{ki}, \mathbf{x}_{kj}) P\left(\mathbf{x}_{ki} \rightarrow \mathbf{x}_{kj} \mid \mathbf{H}_1 \dots \mathbf{H}_K\right) \tag{4.51}$$

where $B(\mathbf{x}_{ki}, \mathbf{x}_{kj})$ denotes the number of bits that are different in vectors $\mathbf{x}_{ki}$ and $\mathbf{x}_{kj}$.

The robustness of the algorithm is obtained averaging the uncoded BER over the channel uncertainty using a Bayesian formulation. Hence, the optimization criterion subject to an average power constraint across all antennas and subcarriers becomes:

$$\begin{aligned} \min_{\mathbf{F}_k} \quad & \frac{1}{\mathcal{N}^{2M} K} \sum_{k=1}^{K} \sum_{i=1}^{\mathcal{N}^M} \sum_{j=1}^{\mathcal{N}^M} B(\mathbf{x}_{ki}, \mathbf{x}_{kj}) E_{\mathbf{h}/\hat{\mathbf{h}}} \left\{ P\left(\mathbf{x}_{ki} \rightarrow \mathbf{x}_{kj} \mid \mathbf{H}_1 \dots \mathbf{H}_K\right) \right\} \\ \text{subject to} \quad & \sum_{k=1}^{K} \mathrm{Tr}\left\{ \mathbf{F}_k^H \mathbf{F}_k \right\} = P_0 \end{aligned} \tag{4.52}$$

4.4.1 Cost function

The pairwise error probability in (4.52) can be written in terms of the Euclidean distance between the transmitted codewords $\mathbf{x}_{ki}$ and $\mathbf{x}_{kj}$ as they appear at the receiver, and is upper bounded by:

$$P\left(\mathbf{x}_{ki} \rightarrow \mathbf{x}_{kj}\right) \leq \mathcal{Q}\left(\sqrt{\frac{1}{2\sigma_n^2} |\mathbf{H}_k \mathbf{F}_k \mathbf{d}_{ijk}|^2} \right) \tag{4.53}$$

where vector $\mathbf{d}_{ijk} = (\mathbf{x}_{ki} - \mathbf{x}_{kj})$ contains the distance between the codewords at the transmitter before the prefiltering. In order to simplify the evaluation of the expectation over the channel response, we introduce the assumption A2) that allows to approximate the pairwise probability by the exponential function[1]:

$$\mathcal{Q}\left(\sqrt{\frac{1}{2\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_{ijk}|^2}\right) \simeq \delta_{ijk}e^{-\alpha_{ijk}\frac{1}{2\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_{ijk}|^2} = \delta_{ijk}e^{-\gamma_{ijk}|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_{ijk}|^2} \tag{4.54}$$

This equation can be rewritten in a more compact form in order to simplify the notation in the subsequent equations:

$$\mathcal{Q}\left(\sqrt{\frac{1}{2\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_{ijk}|^2}\right) \simeq \delta_{ijk}e^{-\mathbf{h}^H\mathbf{M}_{ijk}\mathbf{h}} \tag{4.55}$$

where $\mathbf{M}_{ijk}$ is the $KM_TM_R \times KM_TM_R$ matrix defined as:

$$\mathbf{M}_{ijk} = \gamma_{ijk}\begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{I}_{M_TM_R} \\ \vdots \\ \mathbf{0} \end{bmatrix}\left[\left(\left(\mathbf{F}_k\mathbf{d}_{ijk}\right)\left(\mathbf{F}_k\mathbf{d}_{ijk}\right)^H\right)\otimes\mathbf{I}_{M_R}\right]\begin{bmatrix} \mathbf{0} & \dots & \mathbf{I}_{M_TM_R} & \dots & \mathbf{0} \end{bmatrix} \tag{4.56}$$

and $\mathbf{0}$ is a square all zero matrix, $\mathbf{I}_{M_TM_R}$ is the identity matrix, both of size M_TM_R and $\mathbf{I}_{M_R}$ is the identity matrix of size M_R.

Using the results of the conditional mean (4.13) and the conditional covariance (4.14), and the fact that the channel is modelled as a Gaussian process, the expectation over the channel response, i.e., $E_{\mathbf{h}/\hat{\mathbf{h}}}\{\cdot\}$ can be computed using the probability density function (p.d.f.) $f_{\mathbf{h}/\hat{\mathbf{h}}}(\mathbf{h})$:

$$f_{\mathbf{h}/\hat{\mathbf{h}}}(\mathbf{h}) = \frac{1}{\pi^K \det\left(\mathbf{C}_{\mathbf{h}/\hat{\mathbf{h}}}\right)}\, e^{-\left(\mathbf{h}-\hat{\mathbf{h}}^{eq}\right)^H\mathbf{C}_{\mathbf{h}/\hat{\mathbf{h}}}^{-1}\left(\mathbf{h}-\hat{\mathbf{h}}^{eq}\right)} \tag{4.57}$$

and thus, the averaged PEP becomes:

$$E_{\mathbf{h}/\hat{\mathbf{h}}}\left\{P\left(\mathbf{x}_{ki}\rightarrow\mathbf{x}_{kj}\right)\right\} \leq \frac{\delta_{ijk}}{\pi^K \det\left(\mathbf{C}_{\mathbf{h}/\hat{\mathbf{h}}}\right)}\int_{\mathbf{h}\in\mathbb{C}} e^{-\mathbf{h}^H\mathbf{M}_{ijk}\mathbf{h}}\; e^{-\left(\mathbf{h}-\hat{\mathbf{h}}^{eq}\right)^H\mathbf{C}_{\mathbf{h}/\hat{\mathbf{h}}}^{-1}\left(\mathbf{h}-\hat{\mathbf{h}}^{eq}\right)} d\mathbf{h} \tag{4.58}$$

[1] Although an upper bound of the $\mathcal{Q}(\sqrt{(\cdot)})$ function should strictly be used to minimize (4.53), we propose the use of this lower bound since it greatly reduces the mathematical complexity of the problem. The tightness of this lower bound (approximation) guarantees its correct use in (4.53) (see Appendix 4.C).

where $\hat{\mathbf{h}}^{eq}$, as defined in (4.17), denotes the equivalent channel over all subcarriers. This integral can be easily solved rewriting its integrand as:

$$\int_{\mathbf{h}\in\mathbb{C}} e^{-\left(\mathbf{h}-\mu\right)^H \beta \left(\mathbf{h}-\mu\right) - \eta} d\mathbf{h} \tag{4.59}$$

where:

$$\begin{aligned} \beta &= \mathbf{M}_{ijk} + \mathbf{C}^{-1}_{\mathbf{h}/\hat{\mathbf{h}}} \\ \mu &= \left(\mathbf{M}_{ijk}\mathbf{C}_{\mathbf{h}/\hat{\mathbf{h}}} + \mathbf{I}\right)^{-1} \hat{\mathbf{h}}^{eq} \\ \eta &= \hat{\mathbf{h}}^{eqH} \left(\mathbf{M}_{ijk}\mathbf{C}_{\mathbf{h}/\hat{\mathbf{h}}} + \mathbf{I}\right)^{-1} \mathbf{M}_{ijk}\hat{\mathbf{h}}^{eq} \end{aligned} \tag{4.60}$$

The solution to the integral in (4.59) can be found by comparing its integrand with a complex Gaussian p.d.f., whose integral equals to one. Thus, the averaged PEP in (4.58) becomes:

$$E_{\mathbf{h}/\hat{\mathbf{h}}}\left\{P\left(\mathbf{x}_{ki}\rightarrow\mathbf{x}_{kj}\right)\right\} \leq \frac{\delta_{ijk}}{\det\left(\mathbf{C}_{\mathbf{h}/\hat{\mathbf{h}}}\mathbf{M}_{ijk}+\mathbf{I}\right)}\; e^{-\hat{\mathbf{h}}^{eqH}\left(\mathbf{M}_{ijk}\mathbf{C}_{\mathbf{h}/\hat{\mathbf{h}}}+\mathbf{I}\right)^{-1}\mathbf{M}_{ijk}\hat{\mathbf{h}}^{eq}} \tag{4.61}$$

The derivation of a closed-form solution for this cost function is only feasible when antennas are uncorrelated, requiring numerical techniques to solve the optimization problem for a general spatial correlation between antennas. Hence, the conditional covariance $\mathbf{C}_{\mathbf{h}/\hat{\mathbf{h}}}$ will be hereafter specified for the particular case of uncorrelated antennas. Replacing $\mathbf{M}_{ijk}$ and $\mathbf{C}_{\mathbf{h}/\hat{\mathbf{h}}}$ by its values in (4.56) and (4.15), respectively, the constrained minimization problem becomes:

$$\begin{aligned} \min_{\mathbf{F}_k} \quad & \frac{1}{\mathcal{N}^{2M}K}\sum_{k=1}^{K}\sum_{i=1}^{\mathcal{N}^M}\sum_{j=1}^{\mathcal{N}^M}\frac{B(\mathbf{x}_{ki},\mathbf{x}_{kj})\,\delta_{ijk}}{\left(1+\gamma_{ijk}\omega\left|\mathbf{F}_k\mathbf{d}_{ijk}\right|^2\right)^{M_R}}\; e^{-\frac{\gamma_{ijk}\left|\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k\mathbf{d}_{ijk}\right|^2}{1+\gamma_{ijk}\omega\left|\mathbf{F}_k\mathbf{d}_{ijk}\right|^2}} \\ \text{subject to} \quad & \sum_{k=1}^{K}\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\} = P_0 \end{aligned} \tag{4.62}$$

where the Kronecker product identity $\mathrm{vec}\left(\mathbf{ABC}\right) = \left(\mathbf{C}^T\otimes\mathbf{A}\right)\mathrm{vec}\left(\mathbf{B}\right)$ has been applied.

4.4.2 Closed form solution for beamforming

This section derives a closed-form solution for the minimization of (4.62) by the method of Lagrange multipliers when multiple antennas are used for beamforming (i.e., only a single symbol is spreaded on all antennas for each subcarrier: $M = 1$). For this particular case the high SNR assumption A1) implies that the error probability is dominated by the minimum distance between any pairs of symbols ($\mathbf{x}_{ki}$,$\mathbf{x}_{kj}$). Therefore, the summation in (4.62) can be approximated considering the terms where $\mathbf{x}_{ki} - \mathbf{x}_{kj} = d$, where d is a scalar that refers to the minimum distance between any two constellation symbols. Using this approximation in (4.62), the new

optimization problem can be written as follows:

$$\begin{aligned} \min_{\mathbf{F}_k} \quad & \frac{1}{\mathcal{N}^{2M}K}\sum_{k=1}^{K}\frac{\delta_k}{\left(1+\gamma_k\omega\ d^2\mathbf{F}_k^H\mathbf{F}_k\right)^{M_R}}\ e^{-\frac{\gamma_k\ d^2\left|\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k\right|^2}{1+\gamma_k\omega\ d^2\mathbf{F}_k^H\mathbf{F}_k}} \\ \text{subject to} \quad & \sum_{k=1}^{K}\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\}=P_0 \end{aligned} \tag{4.63}$$

where δ_k and γ_k are the parameters that fit the approximation in (4.54).

Appendix 4.D shows that the cost function is minimized when $\mathbf{F}_k$ focuses the transmitted symbols in the direction of the right singular vector $\mathbf{v}_k$ associated to the largest singular value λ_k^{eq} of the equivalent channel $\hat{\mathbf{H}}_k^{eq}$. Thus:

$$\mathbf{F}_k=\mathbf{v}_k\phi_k \tag{4.64}$$

where ϕ_k is a real positive value related to the power allocated to the kth subcarrier.

Expanding the problem in (4.63) according to the structure of the matrix $\mathbf{F}_k$, the new minimization problem becomes[2]:

$$\begin{aligned} \min_{\phi_k} \quad & \frac{1}{\mathcal{N}^{2M}K}\sum_{k=1}^{K}\frac{\delta_k}{\left(1+\gamma_k\omega\ d^2\phi_k^2\right)^{M_R}}\ e^{-\frac{\gamma_k\ d^2\left|\lambda_k^{eq}\right|^2\phi_k^2}{1+\gamma_k\omega\ d^2\phi_k^2}} \\ \text{subject to} \quad & \sum_{k=1}^{K}\phi_k^2-P_0=0 \\ & \phi_k^2\geq 0 \qquad k=1\ldots K \end{aligned} \tag{4.65}$$

The problem in (4.65) is a convex optimization problem and, consequently, the optimum

[2]Since the function to be minimized depends on $|\phi_k|^2$ we will choose ϕ_k as a real positive value and thus $|\phi_k|^2$ will be hereafter denoted as ϕ_k^2 without loss of generality.

values of ϕ_k can be found deriving the solution that satisfies the KKT conditions:

$$-\frac{1}{\mathcal{N}^{2M}K}\frac{\delta_k\gamma_k\, d^2\phi_k}{\left(1+\gamma_k\omega\, d^2\phi_k^2\right)^{M_R+1}}\, e^{-\dfrac{\gamma_k\, d^2\left|\lambda_k^{eq}\right|^2\phi_k^2}{1+\gamma_k\omega\, d^2\phi_k^2}}\times \left[M_R\omega+\left|\lambda_k^{eq}\right|^2-\frac{\gamma_k\omega\, d^2\left|\lambda_k^{eq}\right|^2\phi_k^2}{1+\gamma_k\omega\, d^2\phi_k^2}\right]-\nu_k+\mu=0 \tag{4.66a}$$

$$\sum_{k=1}^{K}\phi_k^2-P_0=0 \tag{4.66b}$$

$$\phi_k^2\geq 0 \qquad k=1\ldots K \tag{4.66c}$$

$$\nu_k\geq 0 \qquad k=1\ldots K \tag{4.66d}$$

$$\nu_k\phi_k^2=0 \qquad k=1\ldots K \tag{4.66e}$$

where the first condition can be found equating to zero the gradient of the Lagrangian associated with the problem (4.65) with respect to ϕ_k^2, the Lagrange multiplier for the equality constraint (4.66b) is denoted as μ, and ν_k are the multipliers for the inequality constraints (4.66c).

In order to solve these equations we first eliminate the slack variable ν_k. Note that if $\phi_k^2>0$ then (4.66e) can only hold when $\nu_k=0$, which implies that the first condition becomes:

$$-\frac{1}{\mathcal{N}^{2M}K}\frac{\delta_k\gamma_k\, d^2\phi_k}{\left(1+\gamma_k\omega\, d^2\phi_k^2\right)^{M_R+1}}\, e^{-\dfrac{\gamma_k\, d^2\left|\lambda_k^{eq}\right|^2\phi_k^2}{1+\gamma_k\omega\, d^2\phi_k^2}}\left[M_R\omega+\left|\lambda_k^{eq}\right|^2-\frac{\gamma_k\omega\, d^2\left|\lambda_k^{eq}\right|^2\phi_k^2}{1+\gamma_k\omega\, d^2\phi_k^2}\right]+\mu=0 \tag{4.67}$$

After some easy manipulations the following equation must be solved:

$$-\frac{\gamma_k\, d^2\phi_k^2\left|\lambda_k^{eq}\right|^2}{1+\gamma_k\omega\, d^2\phi_k^2}-(M_R+1)\log\left(1+\gamma_k\omega\, d^2\phi_k^2\right)+\log\left(M_R\omega+\frac{\left|\lambda_k^{eq}\right|^2}{1+\gamma_k\omega\, d^2\phi_k^2}\right)=\bar{\mu}-\log\left(\delta_k\gamma_k\, d^2\right) \tag{4.68}$$

A closed-form solution for this identity cannot be derived. However, under the assumption that the channel uncertainty is low, the following approximations can be used: $\log(1+x)\simeq x$ and $1/(1+x)\simeq 1$. Thus (4.68) is simplified as:

$$-\gamma_k\, d^2\left|\lambda_k^{eq}\right|^2\phi_k^2-(M_R+1)\,\gamma_k\omega\, d^2\phi_k^2+\log\left(M_R\omega+\left|\lambda_k^{eq}\right|^2\right)=\bar{\mu}-\log\left(\delta_k\gamma_k\, d^2\right) \tag{4.69}$$

obtaining a closed-form solution for the power allocated to the kth subcarrier ϕ_k^2:

$$\phi_k^2=\left[-\frac{\bar{\mu}-\log\left(\delta_k\gamma_k\, d^2\right)-\log\left(M_R\omega+\left|\lambda_k^{eq}\right|^2\right)}{\gamma_k\, d^2\left((M_R+1)\,\omega+\left|\lambda_k^{eq}\right|^2\right)}\right]^+ \tag{4.70}$$

where the function $[x]^+$, introduced to satisfy (4.66c)-(4.66e), is defined as $\max\{x,0\}$ and $\bar{\mu}$,

that is determined forcing the average power constraint (4.66b), is given by:

$$\bar{\mu} = -\frac{P_0 - \sum_{k=1}^{K} \frac{\log\left(\delta_k \gamma_k\, d^2\right) + \log\left(M_R\omega + \left|\lambda_k^{eq}\right|^2\right)}{\gamma_k\, d^2\left(\left(M_R+1\right)\omega + \left|\lambda_k^{eq}\right|^2\right)}}{\sum_{k=1}^{K} \frac{1}{\gamma_k\, d^2\left(\left(M_R+1\right)\omega + \left|\lambda_k^{eq}\right|^2\right)}} \tag{4.71}$$

4.4.3 Solution for spatial multiplexing

This section derives a suboptimal closed-form solution for the minimization of (4.62) and also proposes an iterative solution based on the Frost algorithm [Fro72]. As the direct optimization of (4.62) leads to very intricate equations, the minimization of the cost function has been obtained under the assumption that the linear transformation at the transmitter has the following structure:

$$\mathbf{F}_k = \mathbf{V}_k \mathbf{\Phi}_k \mathbf{T}_k \tag{4.72}$$

where matrix $\mathbf{V}_k$ contains the right singular vectors of the equivalent channel matrix, $\mathbf{\Phi}_k$ is a diagonal matrix and $\mathbf{T}_k$ is a $M \times M$ unitary matrix whose properties are described next.

Appendix 4.D proves that this configuration is optimal for minimum uncoded BER in a ML receiver when multiple transmit antennas are used for beamforming (i.e., $M = 1$). The same configuration has been shown to be optimal for minimum uncoded BER in a zero forcing receiver with perfect CSI [Din03], and it has also been shown to be optimal in a MMSE receiver when $\mathbf{T}_k$ is set to the DFT matrix [Cha02]. Although it has not been proved to be optimal when multiple symbols are transmitted at the same time, there are several reasons to support its choice. It keeps the same structure as the MMSE solution used in Section 4.3 and other works published previously (see e.g., [Pal03b, Pal03a] and references therein). Besides, a suboptimum solution leads to the same closed-form design given for the beamforming case, as will be shown next. Finally, from the point of view of uncoded BER minimization, it can also be argued that the introduction of matrix $\mathbf{T}_k$ alleviates the main drawback of the use of matrices $\mathbf{V}_k$ and $\mathbf{\Phi}_k$: the loss of space diversity caused by the decomposition of the MIMO channel into a set of parallel multiplicative subchannels. If $\mathbf{T}_k = \mathbf{I}$ and matrix $\mathbf{G}_k$ is selected as the left singular vectors of the equivalent channel matrix, then the MIMO channel reduces to a set of KM parallel independent flat fading subchannels:

$$\mathbf{r}_k(i) = \lambda_k^{eq}(i)\phi_k(i)\mathbf{x}_k(i) + \mathbf{w}_k(i) \tag{4.73}$$

so that the symbols corresponding to small values of $\lambda_k^{eq}(i)\phi_k(i)$ are systematically lost. In the context of frequency flat Rayleigh fading channels, it was shown in [Rai96], [Lam97], [Bou98] that the receiver could benefit from the diversity of the fading channel, provided the transmitter used a linear transformation that spreads the symbols in time is used, obtaining significant

performance gains both in terms of uncoded and coded BER. Similar conclusions would be obtained here if the Rayleigh fading channel statistics were replaced by those of the eigenvalues of $\hat{\mathbf{H}}_k^{eqH}\hat{\mathbf{H}}_k^{eq}$ matrix, but to the best of the author's knowledge the statistics of the eigenvalues of non-central Wishart matrices are not known. We set the unitary matrix $\mathbf{T}_k$ to the DFT matrix, given that it was shown in [Lam97] that the DFT or the Walsh-Hadamard matrices could provide the desired fading diversity with minimum complexity, and besides, the same condition was required for the design of the optimal matrix $\mathbf{T}_k$ in the zero forcing receiver [Din03] and for the minimum MSE design in a linear receiver [Rai96].

Forcing this structure the design of $\mathbf{F}_k$ reduces to the design of the diagonal matrix $\mathbf{\Phi}_k$. Next sections will derive the diagonal elements of this matrix. First, in next section, a closed-form solution which only is valid when $\mathbf{T}_k = \mathbf{I}$ is derived. Then, in Section 4.4.3.2, an iterative solution is proposed when this unitary matrix is different to the identity.

4.4.3.1 Closed-form solution

Repeating the procedure used to derive the closed-form solution for beamforming, we can derive a closed-form solution for the cost function resulting from (4.62) when more that one symbol is transmitted per subcarrier. It will be shown, however, that this closed-form solution is suboptimum since uses an inequality that only holds for high SNR.

Expanding the function to be minimized in (4.62) according to the structure of the matrix $\mathbf{F}_k$, and forcing all the unitary matrices to be equal, i.e., $\mathbf{T}_k = \mathbf{T}$, the new objective function, subject to the average power constraint, becomes:

$$\begin{aligned} \min_{\mathbf{F}_k} \quad & \frac{1}{\mathcal{N}^{2M}K}\sum_{k=1}^{K}\sum_{i=1}^{\mathcal{N}^M}\sum_{j=1}^{\mathcal{N}^M}\frac{B(\mathbf{x}_{ki},\mathbf{x}_{kj})\,\delta_{ijk}}{\left(1+\gamma_{ijk}\omega\left|\mathbf{\Phi}_k\mathbf{T}\mathbf{d}_{ijk}\right|^2\right)^{M_R}}\, e^{-\frac{\gamma_{ijk}\left|\mathbf{\Lambda}_k^{eq}\mathbf{\Phi}_k\mathbf{T}\mathbf{d}_{ijk}\right|^2}{1+\gamma_{ijk}\omega\left|\mathbf{\Phi}_k\mathbf{T}\mathbf{d}_{ijk}\right|^2}} \\ \text{subject to} \quad & \sum_{k=1}^{K}\mathrm{Tr}\left\{\mathbf{\Phi}_k\mathbf{\Phi}_k^H\right\} = P_0 \end{aligned} \tag{4.74}$$

An upper bound of the convex function $\mathcal{Q}\left(\sqrt{x}\right)$ based on the application of Jensen's inequality will allow to simplify the previous cost function and derive the closed-form solution. As it shown in the Appendix 4.E the following inequality holds for high SNR:

$$\frac{1}{\mathcal{N}^{2M}}\sum_{i=1}^{\mathcal{N}^M}\sum_{j=1}^{\mathcal{N}^M}B(\mathbf{x}_{ki},\mathbf{x}_{kj})\mathcal{Q}\left(\sqrt{\frac{\left|\mathbf{\Lambda_k}\mathbf{\Phi_k}\mathbf{T}\mathbf{d}_{ijk}\right|^2}{2\sigma_n^2}}\right) \leq \frac{1}{\mathcal{N}^{2M}}\sum_{i=1}^{\mathcal{N}^M}\sum_{j=1}^{\mathcal{N}^M}B(\mathbf{x}_{ki},\mathbf{x}_{kj})\mathcal{Q}\left(\sqrt{\frac{\left|\mathbf{\Lambda_k}\mathbf{\Phi_k}\mathbf{d}_{ijk}\right|^2}{2\sigma_n^2}}\right) \tag{4.75}$$

where equality applies if and only if the product matrix $\mathbf{\Lambda}_k\mathbf{\Phi}_k$ is proportional to the identity matrix (zero forcing solution) or $\mathbf{T} = \mathbf{I}$. Therefore, if the true uncoded BER depending on

$|\mathbf{\Lambda}_k\mathbf{\Phi}_k\mathbf{T}\mathbf{d}_{ijk}|$ is replaced by an upper bound depending on $|\mathbf{\Lambda}_k\mathbf{\Phi}_k\mathbf{d}_{ijk}|$, an expression is obtained that does not depend on matrix $\mathbf{T}$. Using this replacement in the cost function (4.74) (i.e., substituting $\mathbf{T}=\mathbf{I}$ in the cost function) we get a simple expression that will help us to derive a solution for the diagonal matrix $\mathbf{\Phi}_k$.

Under the high SNR assumption A1) and for $\mathbf{T}=\mathbf{I}$, the error probability is dominated by the minimum distance between any pairs of symbols ($\mathbf{x}_{ki}$,$\mathbf{x}_{kj}$). Therefore, the summation in (4.74) can be approximated, considering the terms where $\mathbf{x}_{ki}-\mathbf{x}_{kj}=d\mathbf{i}(l)$, where $\mathbf{i}(l)$ is one of the columns of the identity matrix, and d is a scalar that refers to the minimum distance between any two constellation symbols. Using this approximation the cost function can be rewritten as:

$$\begin{aligned}
\min_{\phi_k} \quad & \frac{1}{\mathcal{N}^{2M}K}\sum_{k=1}^{K}\sum_{l=1}^{M}\frac{\delta_k(l)}{\left(1+\gamma_k(l)\;\omega\;d^2\phi_k^2(l)\right)^{M_R}}\;e^{-\dfrac{\gamma_k(l)\;d^2\left|\lambda_k^{eq}(l)\right|^2\phi_k^2(l)}{1+\gamma_k(l)\;\omega\;d^2\phi_k^2(l)}} \\
\text{subject to} \quad & \sum_{k=1}^{K}\sum_{l=1}^{M}\phi_k^2(l)-P_0=0 \\
& \phi_k^2(l)\geq 0 \qquad k=1\ldots K \qquad l=1\ldots M
\end{aligned} \tag{4.76}$$

where $\delta_k(l)$ and $\gamma_k(l)$ are the parameters that fit the approximation in (4.54).

Note that previous minimization problem and that one derived for the beamforming case (4.65) are identical except for the summation over the index l. Because a nonnegative weighted sum of convex functions is itself a convex function [Boy04, pp.79], this summation does not modifies the convexity of the objective function. Hence, repeating the equations (4.67)-(4.70) we can derive a closed-form solution for the power allocation parameters:

$$\phi_k^2(l)=\left[-\frac{\mu-\log\left(\delta_k(l)\;\gamma_k(l)\;d^2\right)-\log\left(M_R\omega+\left|\lambda_k^{eq}(l)\right|^2\right)}{\gamma_k(l)\;d^2\left((M_R+1)\,\omega+\left|\lambda_k^{eq}(l)\right|^2\right)}\right]^{+} \tag{4.77}$$

and similarly μ can be determined as in equation (4.71) forcing the average power constraint.

4.4.3.2 Iterative Algorithm

When the unitary matrix $\mathbf{T}$ is set to the identity matrix (i.e., $\mathbf{T}=\mathbf{I}$) it has been shown that the Jensens's inequality (4.75) becomes an equality and, consequently, the previous closed-form solution is a valid solution for high SNR's. However, when matrix $\mathbf{T}$ is a unitary matrix that combines the transmitted symbols per subcarrier over all the antennas this closed-form solution is suboptimum and leads to a very poor performance as it will be shown in Section 4.5.2. For this case an iterative algorithm must be therefore considered to derive the optimum solution.

Optimization problems that include inequality constraints can be solved by means of interior-point methods [Boy04, p.561] (e.g. barrier methods) that transform the inequality constrained

problem into an equality constrained problem to which a gradient method can be applied. The growth in the computational complexity of these methods when the number of inequality constraints is large has motivated, in this thesis, the proposal of a simpler algorithm for solving the optimization problem based on the Frost algorithm [Fro72] (see Appendix 4.F).

When the matrix $\mathbf{T}$ is set to an unitary matrix different from the identity matrix, the objective function remains convex if perfect CSI is available at the transmitter (i.e., $\omega = 0$). However, when only partial CSI is available at the transmitter (i.e., $\omega \neq 0$) we cannot prove the convexity of this function. Under these conditions, although convexity fails, numerical results have proven that the simple Frost algorithm, focused on a minimum local search following a gradient method, results in excellent performance. Consequently, without any claim to get the optimum solution in the objective function (4.74), we will consider as a satisfying algorithm in terms of uncoded BER the solution proposed in this section, which converges to the global minimum when perfect CSI is available at the transmitter and to a local minimum when CSI is noisy.

Let $\boldsymbol{\phi}_k$ be a vector that stores the diagonal terms of the matrix $\boldsymbol{\Phi}_{\mathbf{k}}$ and $L(\boldsymbol{\phi}_k, \mu)$ be the Lagrangian associated to the optimization problem (4.74), which is defined as:

$$L(\boldsymbol{\phi}_k, \mu) = \frac{1}{\mathcal{N}^{2M} K} \sum_{k=1}^{K} \sum_{i=1}^{\mathcal{N}^M} \sum_{j=1}^{\mathcal{N}^M} \frac{B(\mathbf{x}_{ki}, \mathbf{x}_{kj})\, \delta_k(l)}{\left(1 + \gamma_k(l)\, \omega \left|\mathbf{D}_{ijk}\boldsymbol{\phi}_k\right|^2\right)^{M_R}} e^{-\frac{\gamma_k(l) \left|\boldsymbol{\Lambda}_k^{eq}\mathbf{D}_{ijk}\boldsymbol{\phi}_k\right|^2}{1+\gamma_k(l)\,\omega\left|\mathbf{D}_{ijk}\boldsymbol{\phi}_k\right|^2}} - \mu\left[\sum_{k=1}^{K} \boldsymbol{\phi}_k^H \boldsymbol{\phi}_k - P_0\right] \quad (4.78)$$

where $\mathbf{D}_{ijk}$ is a diagonal matrix defined as $\mathbf{D}_{ijk} = \mathrm{diag}\{\mathbf{T}\mathbf{d}_{ijk}\}$ and μ is the Lagrangian multiplier. The gradient of $L(\boldsymbol{\Phi}_{\mathbf{k}}, \mu)$ with respect to the vector $\boldsymbol{\phi}_k^2$ (defined as a vector that stores the diagonal terms of the matrix ${\boldsymbol{\Phi}_{\mathbf{k}}}^H \boldsymbol{\Phi}_{\mathbf{k}}$) is given by:

$$\begin{aligned}\nabla_{\boldsymbol{\phi}_k^2} L(\boldsymbol{\Phi}_{\mathbf{k}}, \mu) =& \frac{1}{\mathcal{N}^{2M} K} \sum_{i=1}^{\mathcal{N}^M} \sum_{j=1}^{\mathcal{N}^M} \frac{-B(\mathbf{x}_{ki}, \mathbf{x}_{kj})\, \delta_k(l)\, \gamma_k(l)}{\left(1 + \gamma_k(l)\, \omega \left|\mathbf{D}_{ijk}\boldsymbol{\phi}_k\right|^2\right)^{M_R+1}} e^{-\frac{\gamma_k(l) \left|\boldsymbol{\Lambda}_k^{eq}\mathbf{D}_{ijk}\boldsymbol{\phi}_k\right|^2}{1+\gamma_k(l)\,\omega\left|\mathbf{D}_{ijk}\boldsymbol{\phi}_k\right|^2}} \\ & \times \mathbf{D}_{ijk}^H \left[M_R\omega\mathbf{I} + {\boldsymbol{\Lambda}_k^{eq}}^H \boldsymbol{\Lambda}_k^{eq} - \frac{\left|\boldsymbol{\Lambda}_k^{eq}\mathbf{D}_{ijk}\boldsymbol{\phi}_k\right|^2 \gamma_k(l)\, \omega\mathbf{I}}{1+\gamma_k(l)\,\omega\left|\mathbf{D}_{ijk}\boldsymbol{\phi}_k\right|^2}\right] \mathbf{D}_{ijk}\mathbf{1}_M - \mu\mathbf{1}_{M\times 1} \quad (4.79) \\ =& \mathbf{a}_k(\boldsymbol{\phi}_k^2) - \mu\mathbf{1}_{M\times 1} \quad (4.80)\end{aligned}$$

Stacking all the power allocation parameters into the vector $\boldsymbol{\phi}^2 = \left[\begin{array}{ccc} {\boldsymbol{\phi}_1^2}^T & \dots & {\boldsymbol{\phi}_K^2}^T \end{array}\right]^T$ the gradient of the cost function with respect to this vector can be written with the same structure that the gradient obtained in Appendix 4.F:

$$\nabla_{\boldsymbol{\phi}^2}\xi = \mathbf{a}\left(\boldsymbol{\phi}^2\right) - \mu\mathbf{C} \quad (4.81)$$

where:

$$\begin{aligned} \mathbf{a}\left(\boldsymbol{\phi}^2\right) &= \left[\begin{array}{ccc} \mathbf{a}_k(\boldsymbol{\phi}_1^2)^T & \ldots & \mathbf{a}_k(\boldsymbol{\phi}_K^{2\,T}) \end{array} \right]^T \\ \mathbf{C} &= \mathbf{1}_{KM\times 1} \end{aligned} \tag{4.82}$$

Hence, according to (4.135)-(4.139) in Appendix 4.F, the iterative equation that minimizes the cost function is given by [3]:

$$\boldsymbol{\phi}^2_{(n+1)} = \mathbf{P}^{\perp}_{\mathbf{C}}\left[\boldsymbol{\phi}^2_{(n)} + \lambda \mathbf{a}\left(\boldsymbol{\phi}^2_{(n)}\right)\right] + \mathbf{C}\left(\mathbf{C}^H\mathbf{C}\right)^{-1} P_0 \tag{4.83}$$

where $\mathbf{P}^{\perp}_{\mathbf{C}}$ is the projection matrix onto the orthogonal subspace spanned by the columns of $\mathbf{C}$. From the inequality constraint $\phi_k(l) \geq 0$ all the elements in the vector $\boldsymbol{\phi}^2$ must be non-negative. Thus, if any of the elements obtained after an iteration were negative, this element should be set to zero before proceed with the next iteration. In accordance, the iterative equation (4.83) should be modified as follows:

$$\begin{aligned} \mathbf{A}_{(n+1)} &= \mathbf{P}^{\perp}_{\mathbf{C}}\left[\boldsymbol{\phi}^2_{(n)} + \lambda \mathbf{a}\left(\boldsymbol{\phi}^2_{(n)}\right)\right] + \mathbf{C}\left(\mathbf{C}^H\mathbf{C}\right)^{-1} P_0 \\ \boldsymbol{\phi}^2_{(n+1)}(l) &= \left[\mathbf{A}_{(n+1)}(l)\right]^+ \end{aligned} \tag{4.84}$$

where $\boldsymbol{\phi}^2_{(n+1)}(l)$ is the lth element of the vector $\boldsymbol{\phi}^2$ at the iteration $n+1$ and the operator $[x]^+ = \max\{0, x\}$ is applied independently at each element of the vector $\mathbf{A}_{n+1}$. Note that by means of this operator we avoid the use of the barrier functions to include the inequality constraints $\phi_k(l) \geq 0$ into the optimization problem. It is worth to remark that this adjustment in the iterative equation perturbates the average power constraint and, consequently, the final solution would violate this equality constraint. As discussed in the Appendix 4.F, the Frost algorithm solves this problem since the algorithm guarantees the constraints in every iteration and admits modifications in the optimum solution from one iteration to the next.

Since we cannot guarantee the convexity of the objective function the initialization of the iterative algorithm can determine the final solution. Next we present the performance achieved by the iterative algorithm for different initialization values. The initialization labelled as *'Uniform'* allocates the same power to all the subcarriers and antennas; the *'Max. Eigv.'* allocates more power in the direction of to the maximum singular value; the *'MMSE'* initializes the algorithm with the MMSE solution derived in Section 4.3; finally *'Random'* initializes the algorithm at a random point using a Gaussian random variable. Specifically, for the 3×3 MIMO configuration $\boldsymbol{\phi}^2_k$ is initialized as shown in Table 4.1. Note that in all cases the initial value satisfies the average power constraint (for the *'Random'* initialization a normalization is applied once the random values are generated to guarantee this constraint).

The performance of the iterative algorithm for the different initialization values has been analyzed for the 3×3 MIMO configuration (see Section 4.5 for a description of the simulation

[3] Do not confuse the vector $\boldsymbol{\phi}^2_{(n+1)}$ that refers to the nth iteration of the vector $\boldsymbol{\phi}^2$ with the vector $\boldsymbol{\phi}^2_k$ defined as a vector that contains the power allocation coefficients for the kth subcarrier

Initialization	$\boldsymbol{\phi}_k^2$
'Uniform'	$\boldsymbol{\phi}_k^2 = [1/3\ 1/3\ 1/3]$
'Max. Eigv.'	$\boldsymbol{\phi}_k^2 = [0.8\ 0.1\ 0.1]$
'MMSE'	See (4.48)
'Random'	$\boldsymbol{\phi}_k^2(l) = \mathcal{N}(1/3, 0.004)$ s.t. $\sum_{k=1}^{K}\sum_{l=1}^{M}\phi_k^2(l) = P_0$

Table 4.1: Different initialization values.

Initialization	BER at Eb/No=6dB	Eb/No required for BER$\leq 10^{-3}$
'Uniform'	$3.476 \cdot 10^{-3}$	6.92 dB
'Max. Eigv.'	$3.638 \cdot 10^{-3}$	7.36 dB
'MMSE'	$3.719 \cdot 10^{-3}$	7.21 dB
'Random'	$3.410 \cdot 10^{-3}$	6.93 dB

Table 4.2: Performance of the iterative algorithm in terms of uncoded BER for different initialization values.

parameters and scenarios) evaluating the uncoded BER for Eb/No=6dB and the minimum Eb/No required to achieve an uncoded BER$\leq 10^{-3}$. Table 4.2 summarizes the results after 50 iterations. Note that the performances of the algorithm for the different initialization values are quite similar. For the sake of simplicity we will initialize the algorithm with the *'Uniform'* value after evidencing that the iterative algorithm is quite insensitive to the initialization values presented in Table 4.1.

4.4.4 Asymptotic performance

In this section the robust uncoded BER algorithm is analyzed for the extreme cases where the uncertainty is very high or very low. When channel knowledge at the transmitter is perfect, i.e., when $\hat{\mathbf{H}}_k^{eq} = \mathbf{H}$ and $\omega = 0$ [see equations (4.16) and (4.19)], and only one symbol is transmitted per subcarrier (i.e., $M = 1$), the solution for the robust cost function in (4.70) coincides with that one of [Ong01b]. Note that the solution does not coincide with the solution proposed in [Din03], since one assumes a ML receiver and the other one a zero-forcing receiver. When the channel uncertainty is very high, in the extreme case of $\hat{\mathbf{H}}_k^{eq} = 0$ and $\omega = 1$, the average PEP function converges to:

$$E_{\mathbf{h}/\hat{\mathbf{h}}}\left\{P\left(\mathbf{x}_{ki} \rightarrow \mathbf{x}_{kj}\right)\right\} = \frac{\delta_{ijk}}{\left(1 + \gamma_{ijk}\left|\mathbf{F}_k\mathbf{d}_{ijk}\right|^2\right)^{M_R}} \tag{4.85}$$

which is the same one found in [Sam02] for uncorrelated antennas. Note that in this case the optimization criterion is simpler, and therefore, there is no need to resort to the high SNR approximations in order to get a closed-form solution. Indeed, it can be seen that the power allocation tends to the open-loop solution (i.e., the same power is allocated to all antennas and subcarriers), as it will be shown in the next section.

4.5 Simulation results

In order to illustrate the performance of the proposed closed-loop algorithms in the presence of imperfect channel estimates, several simulations are presented for different scenarios and channel estimation errors, assuming that CSI quality is unbalanced between the transmitter and the receiver. The aim is to validate the cost function solution given for the MMSE algorithm (4.48) and the solutions proposed for the minimum uncoded BER algorithms when only one symbols is transmitted per subcarrier (4.70) and when more than one symbol is transmitted (4.77),(4.84). In accordance, uncorrelated antennas are always assumed. The two optimization criteria proposed in this chapter, (i.e., MMSE and minimum uncoded BER) are simulated, comparing the differences between the robust algorithms (solid lines) and their nonrobust counterparts (dotted lines) that assume $\omega = 0$. For the specific case of minimum uncoded BER, in order to compare the Chernoff bound with the approximation of the $\mathcal{Q}(\sqrt{(}\cdot))$ function presented in Appendix 4.C, that bound has also been included in the analysis. However, unless the contrary is said, the algorithms that minimize the uncoded BER are based on the exponential approximation given in that Appendix. The labels used to denote the different closed-loop algorithms are: *'MMSE Robust'* and *'MMSE Non-Robust'* for MMSE criteria; *'BER ApproxQ Robust'* and *'BER ApproxQ Non-Robust'* for minimum uncoded BER using the exponential approximation given in Appendix 4.C; *'BER Chernoff Robust'* and *'BER Chernoff Non-Robust'* for minimum uncoded BER using the Chernoff bound. To evaluate the gains of these algorithms tanks to the channel knowledge at the transmitter, the *'Open-loop'* solution, that allocates the same power to all the subcarriers and antennas, has also been simulated.

The simulation parameters are selected according to the HIPERLAN/2 standard [HL201]. As depicted in Figure 4.1 the bit stream to be transmitted, which has been mapped into a QPSK constellation, is multiplexed in the K subcarriers (grouping M symbols per subcarrier), prefiltered by the matrices $\mathbf{F}_k$, and finally modulated in OFDM symbols (including pilot tones and empty subcarriers according to HIPERLAN/2). Although the HIPERLAN/2 standard includes a channel code scheme, specifically a convolutional code is included, this stage has not been taken into account because it has not been considered in the design of $\mathbf{F}_k$ and $\mathbf{G}_k$ matrices. The evaluation of the proposed algorithms in terms of coded BER is left to a further discussion (see Section 4.6) and a detailed study can be found in the next chapter.

In order to have a fair comparison, all algorithms are simulated using a ML detector that takes into account the channel uncertainty at the receiver. As the channel is not perfectly known at the transmitter and/or the receiver, the relationship between the transmitted and the received data (4.3) is not equivalent to a set of independent flat fading subchannels (i.e., $\mathbf{G}_k\mathbf{H}_k\mathbf{F}_k$ is not diagonal). Therefore, the optimum receiver requires the joint estimation of all bits transmitted in the same symbol $\mathbf{x}_k$. This procedure is used to evaluate the uncoded BER, detecting the symbol $\hat{\mathbf{x}}_k$ that maximizes:

$$f_{\mathbf{r_k}|\hat{\mathbf{h}}}\left(\mathbf{r_k} \mid \hat{\mathbf{x}}_k, \hat{\mathbf{h}}, \mathbf{F}_k\right) = \mathcal{K}\, e^{-\frac{\left(\mathbf{r}_k - \hat{\mathbf{H}}_k^{eq}\mathbf{F}_k\hat{\mathbf{x}}_k\right)^H \left(\mathbf{r}_k - \hat{\mathbf{H}}_k^{eq}\mathbf{F}_k\hat{\mathbf{x}}_k\right)}{\omega\hat{\mathbf{x}}_k^H\mathbf{F}_k^H\mathbf{F}_k\hat{\mathbf{x}}_k + \sigma_n^2}} \tag{4.86}$$

where $\mathcal{K}$ refers to a non-relevant constant and the channel uncertainty ω and the equivalent channel $\hat{\mathbf{H}}_k^{eq}$ are computed making use of the channel uncertainty at the receiver.

The simulated Rayleigh MIMO channel obeys an exponential power delay profile with 50ns of delay spread (RMS delay spread of the discrete channel 45.6ns [Rap96, p.160]), modelling a typical office indoor scenario. The simulations model the channel uncertainties due to the errors in the channel estimation process and the errors in the channel prediction when channels are time-varying. In both cases, the variance in the channel estimation error is assumed to be constant for all taps of the channel impulse response. For the sake of clarity, we will denote in this section $\sigma^2_{\epsilon_{Tx}}$ and $\sigma^2_{\epsilon_{Rx}}$ the variances in the channel estimation error at the transmitter and at the receiver, respectively. It is assumed that the main contribution to channel uncertainty at the transmitter is the channel tracking error in fast linear time-varying channels; hence, it is considered to be independent of the SNR, as would be the case of a channel tracker based on a linear predictor (see Chapter 3). Simulations are carried out with $\sigma^2_{\epsilon_{Tx}} = 0.12$.[4] At the receiver the main contribution to channel uncertainty is the estimation variance due to the presence of additive noise. Hence, $\sigma^2_{\epsilon_{Rx}}$ is proportional to the noise variance σ_n^2, as would be the case of a linear channel estimator [Tun01], where the factor of proportionality depends on the training sequence length. Simulations are carried out with $\sigma^2_{\epsilon_{Rx}} = 0.375\sigma_n^2$, which approximately corresponds to the estimation of the simulated channel from a linear channel estimator when the training sequence consists on one OFDM symbol according to HIPERLAN/2 [IST02].

[4] This values has been computed as $\sigma^2_{\epsilon_{Tx}} = L \cdot \sigma^2_{\epsilon_1}$ where $L = 6$ is the number of channel taps and $\sigma^2_{\epsilon_1} = 0.02$, obtained from Figure 3.8 in Section 3.6 (under the assumption that the mobile was moving at 4km/h), is the channel uncertainty at the transmitter for the first tap when an AR(1) model is used. This value of $\sigma^2_{\epsilon_{Tx}}$ is pessimistic since we are assuming that channel uncertainty is uniform for all taps whereas it can be shown from equations in Section 3.5 that the channel uncertainty at the transmitter has a decreasing profile if the power delay profile decreases.

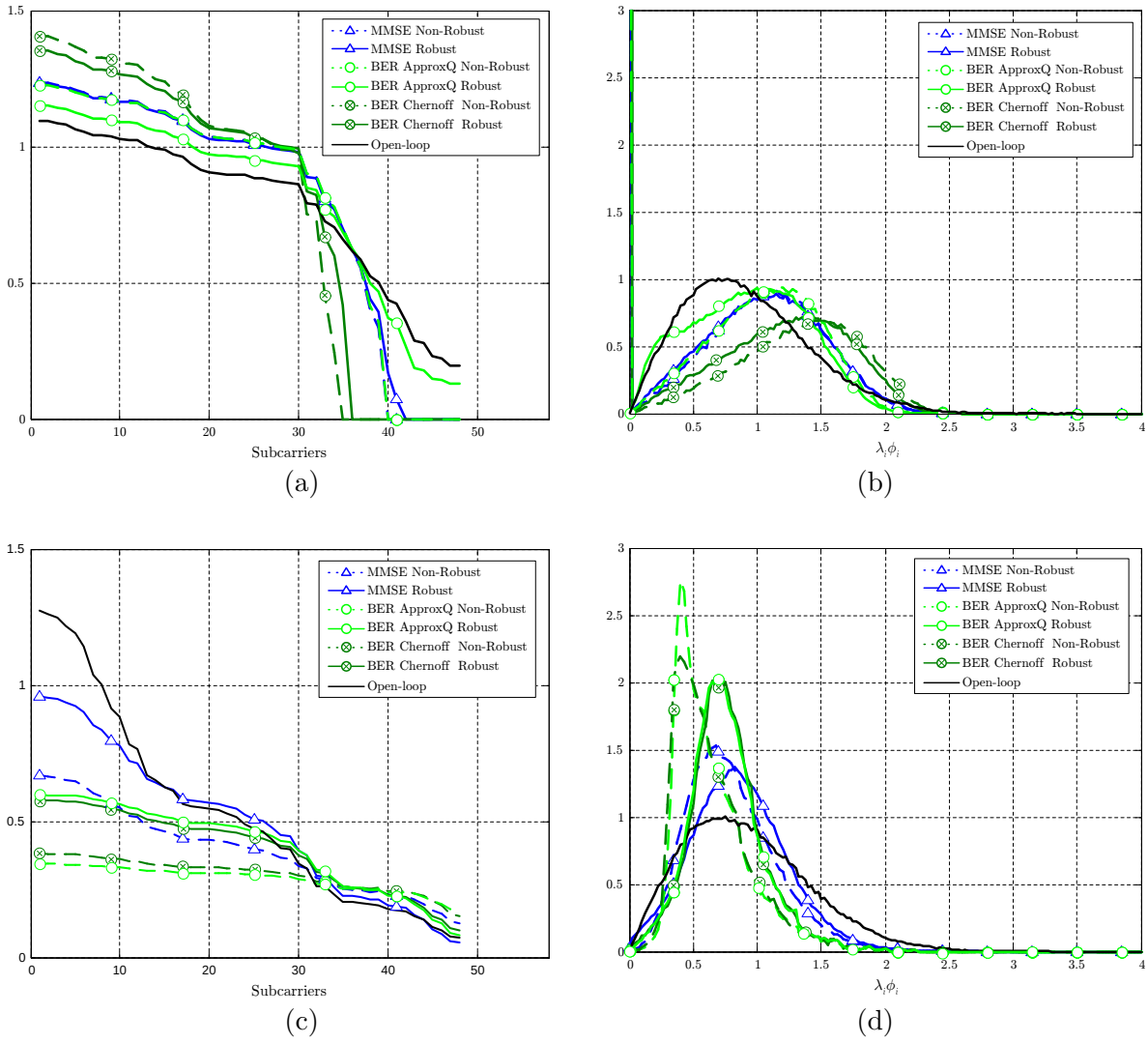


Figure 4.3: An example of $\lambda_i\phi_i$ values (sorted in decreasing order of the equivalent channel subcarriers λ_i $i = 1 \dots K$) for a specific channel realization (a) and (c), and its associated p.d.f. (b) and (d). $M_T = 1$, $M_R = 1$, $M = 1$. Transmitter uncertainty: $\sigma^2_{\epsilon_{Tx}} = 0.12$ ($\rho = 0.1$). SNR=0dB [(a) and (b)] and SNR=20dB [(c) and (d)].

	Eb/No	MMSE Non-Robust	MMSE Robust	BER ApproxQ Non-Robust	BER ApproxQ Robust	BER Chernoff Non-Robust	BER Chernoff Robust
1x1	0dB	7.0	6.0	6.27	<0.05	14.6	11.4
	20dB	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05

Table 4.3: Number of null subcarriers.

4.5.1 Power allocation strategies

Before analyzing the performance of the different algorithms in terms of the uncoded BER we will study the power allocation strategies followed by each one of the algorithms analyzed in this section in order to better understand the results. Next figures display an example of the values of $\lambda_i\phi_i$ (sorted in decreasing order of the equivalent channel subcarriers λ_i $i = 1 \dots K$) for a specific channel realization and its associated probability density function (p.d.f.). Two SNR values have been considered to analyze the performance of the different algorithms in low and high SNR scenarios.

Figure 4.3 illustrates the power allocation strategies for the specific configuration $M_T = M_R = 1$. A comparative study between the MMSE and the uncoded BER algorithms illustrates that the algorithms that minimize the uncoded BER assign less power to the stronger subcarriers and more power to the deepest faded ones, trying to compensate the fading, whereas the MMSE algorithms assign less power to the faded subcarriers. This way to allocate the power, as it will be shown next, will lead the MMSE algorithms to have a very poor performance in terms of uncoded BER for certain subcarriers. This behavior is clearly displayed at high SNR (curves at the bottom), where the algorithms that minimize the uncoded BER tend to force $\lambda_i\phi_i$ to be independent to the channel response values. Figure 4.3(c) exhibits that these curves are *quasi-flat*, whereas the sharpness of the p.d.f. curves in Figure 4.3(d) also reveal that the algorithms tend to force $\lambda_i\phi_i$ to be constant.

A second study that results in interest is to compare the power allocation policies followed by the robust and nonrobust algorithms. As expected, the robust algorithms are more conservative to allocate the power as a function of the channel response. This can be seen both, in the smoothness of the robust curves when they are compared with the nonrobust ones, and also in the number of subcarriers that are nulled (see Table 4.3). It is also worth to note that the *'BER ApproxQ Robust'* algorithm is more conservative that the *'MMSE Robust'* one (compare, for example, the number of null subcarriers for each case), which will have direct consequences in the behavior of each algorithm when the CSI knowledge at the transmitter will tend to be scarce. Finally it can be seen that the differences between the robust and nonrobust algorithms are more evident at high SNR. This behavior can be explained noting that at low SNR the noise influence hides the CSI quality at the transmitter and thus, robust and nonrobust algorithms are quite similar. However, at high SNR, the CSI uncertainty is critical and thus, the robust algorithms, unlike the nonrobust ones, try to diminish its effect.

In order to conclude the study of the results displayed in Figure 4.3 we compare the power allocation strategies followed by the minimum uncoded BER algorithms when the $\mathcal{Q}(\sqrt{(\cdot)})$ function is approximated by the exponential function given in Appendix 4.C, and when this function is upper bounded with the Chernoff bound. As it can be seen, both algorithms follow

similar power allocation policies at high SNR, whereas there are great differences at low SNR. This can be explained because the Chernoff bound is not a tight bound at low SNR as it can be seen in the Appendix. In terms of uncoded BER it will be shown that at low SNR the algorithms that use the Chernoff bound will have a worse performance when they are compared with the algorithms that make use of the exponential approximation. We can expect this performance by noting that at low SNR the algorithm that uses the Chernoff bound forces a higher number of subcarriers to be zero.

Figure 4.4 displays the power allocation strategies when more that one antenna is used at the transmitter and at the receiver. Specifically, the MIMO configuration $M_T = M_R = 3$ has been simulated for two different number of symbols transmitted per subcarrier: $M = 1$ and $M = 3$. Aside from the main conclusions drawn in the previous paragraphs for the SISO configuration, which still hold in this case, we can add some comments that are specific for the MIMO configurations. When $M = 1$ it can be seen that the dissimilarities between the different algorithms have been substantially reduced and, hence, all the algorithms will have similar performances in terms of uncoded BER.

On the other hand, comparing the different algorithms when more than one symbol is transmitted per subcarrier, the importance of the unitary matrix $\mathbf{T}$ is revealed. When matrix $\mathbf{T}$ is omitted the MMSE algorithms perform like the well-known *water-filling* algorithm [Ral98], [Sca99a], which allocates the majority of the power to the best subcarriers. On the contrary, the algorithms that minimize the uncoded BER, trying to compensate the fading, assign less power to the stronger channel modes and more power to the deepest faded ones. This behavior is observed in the sharpness of the uncoded BER algorithms of Figure 4.4(d) when $\mathbf{T}$ is the identity matrix and it can also be seen in Figure 4.4(c) when comparing, for the different algorithms, the values of $\lambda_i\phi_i$ for the best channel modes (i.e., $i \in [1 \dots 48]$). This way to allocate the power will lead the MMSE algorithms to have a very high uncoded BER in some of the subcarriers as it will be shown next.

On the contrary, when $M > 1$ and matrix $\mathbf{T}$ is set to the DFT matrix, as the unitary matrix combines all symbols into all channel modes, the optimal way to distribute the power for all the algorithms, including those that minimize the uncoded BER, is to allocate more power to the stronger channel modes, just as *'BER ApproxQ Robust'* (when $\mathbf{T}$ is the DFT matrix) and *'MMSE Robust'* algorithms do. This change in the power allocation policy, specifically for the *'BER ApproxQ Robust'* algorithm, can be seen when comparing the different number of null subcarriers for the two different choices of the matrix $\mathbf{T}$. Note that when the unitary matrix is used (i.e., $\mathbf{T}$ is the DFT matrix) the algorithms increases the number of null subcarriers, which are always the weakest ones, spending the power to improve the quality of the best channel modes (see in Figure 4.4(c) the values of $\lambda_i\phi_i$ for $i \in [1 \dots 48]$). A remarkable behavior has the *'BER Chernoff Robust'* algorithms that nulls a large number of subcarriers even when the

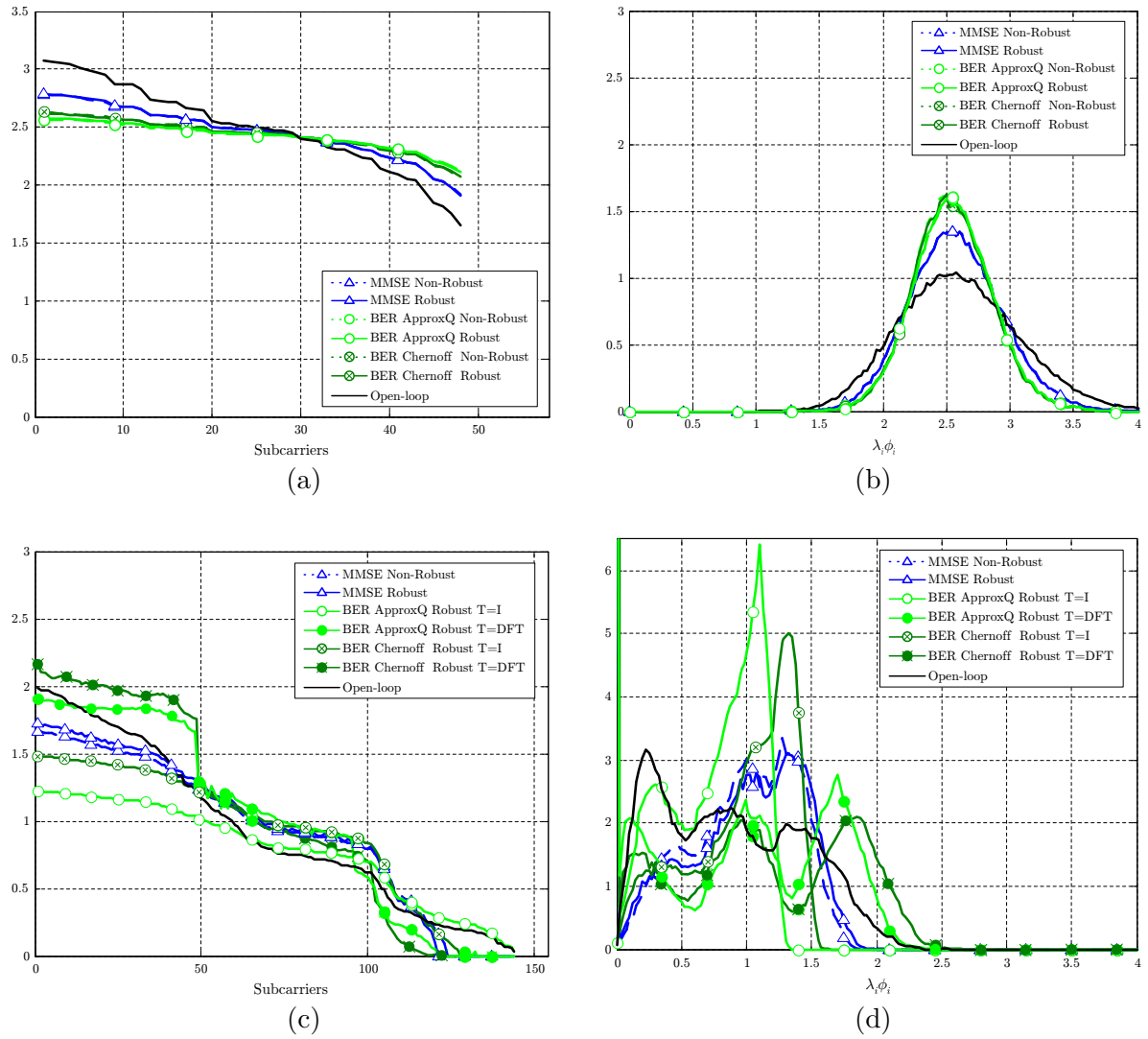


Figure 4.4: An example of $\lambda_i\phi_i$ values (sorted in decreasing order of the equivalent channel modes λ_i $i = 1 \dots M \cdot K$) for a specific channel realization (a) and (c), and its associated p.d.f. (b) and (d). $M_T = 3$, $M_R = 3$. Transmitter uncertainty: $\sigma^2_{\epsilon_{Tx}} = 0.12$ ($\rho = 0.1$). $M = 1$ SNR=0dB [(a) and (b)] and $M = 3$ SNR=5dB [(c) and (d)].

	M	MMSE Non-Robust	MMSE Robust	BER ApproxQ Robust $\mathbf{T = I}$	BER ApproxQ Robust $\mathbf{T} = DFT$	BER Chernoff Robust $\mathbf{T = I}$	BER Chernoff Robust $\mathbf{T} = DFT$
3x3	1	0	0	0	0	0	0
	3	16.8	17.8	<0.1	18.4	15.4	24.2

Table 4.4: Number of null subcarriers.

matrix **T** is the identity matrix. This abnormal behavior will be studied in the next chapter (see Section 5.6) showing that this algorithm, though designed to minimize the uncoded BER, has excellent performances in terms of coded BER.

4.5.2 Uncoded BER performance vs Eb/No

Following the same structure given in Section 4.4 we will first analyze the proposed closed-loop schemes for beamforming (i.e., only one symbol is transmitted per subcarrier $M = 1$) and later we will extend the study for the spatial multiplexing case.

Figure 4.5 shows the performance of the different algorithms for the most simple configuration: $M_T = M_R = 1$. In order to focus on the transmitter design when CSI is noisy at this side, it is assumed that the receiver has perfect CSI knowledge, whereas the channel estimate at the transmitter has a constant variance equal to $\sigma^2_{\epsilon_{Tx}} = 0.12$. As it is expected the robust solutions (solid lines) have best performance than the nonrobust ones (dotted lines). We can profit from the simplicity of this scenario to evaluate the exponential expansion of the $\mathcal{Q}(\cdot)$ function proposed in (4.50). Hence, Figure 4.5 compares the performance of the the generalized uncoded BER bound solution (α_{ijk} and δ_{ijk} obtained according to (4.123)) (*'BER ApproxQ Robust'*), and the Chernoff bound solution ($\alpha_{ijk} = \delta_{ijk} = 1/2$ (*'BER Chernoff Robust'*). As it can be seen, the proposed exponential uncoded BER bound always outperforms others, specifically at low SNR's, where the Chernoff bound exhibits a lower performance. This performance was predicted in the previous section when comparing in Figure 4.3 the power allocation policy of the *'BER ApproxQ Robust'* and *'BER Chernoff Robust'* algorithms. Accordingly, the proposed bound becomes an appropriate alternative to the $\mathcal{Q}(\cdot)$ function extensive to any SNR ratio.

Figure 4.6 displays the performance of the algorithms for the same scenario and transmission rate (only one symbol is transmitted per subcarrier $M = 1$) but increasing the number of transmit and receive antennas. Two conclusions can be drawn from this result. First note, as expected, that all the algorithms improve their performance when the number of antennas is increased. It is also worth to note in this results that when the number of antennas is increased, all the algorithms have similar performances. This behavior can be explained looking at the power allocation strategies in Figure 4.4. As it can be seen, the different power allocation algorithms tend to the same solution when the number of antennas is increased. Roughly speaking we can say that the channel quality of the channel transmission (note that only the best channel mode is used to transmit the symbols) diminishes the influence of how the total power is allocated.

Finally, Figure 4.7 illustrates the performance of the two proposed algorithms in terms of uncoded BER for the 3×3 antenna configuration when more than one symbol is transmitted per subcarrier. The robust (solid lines) and nonrobust (dotted lines) algorithms are plotted for

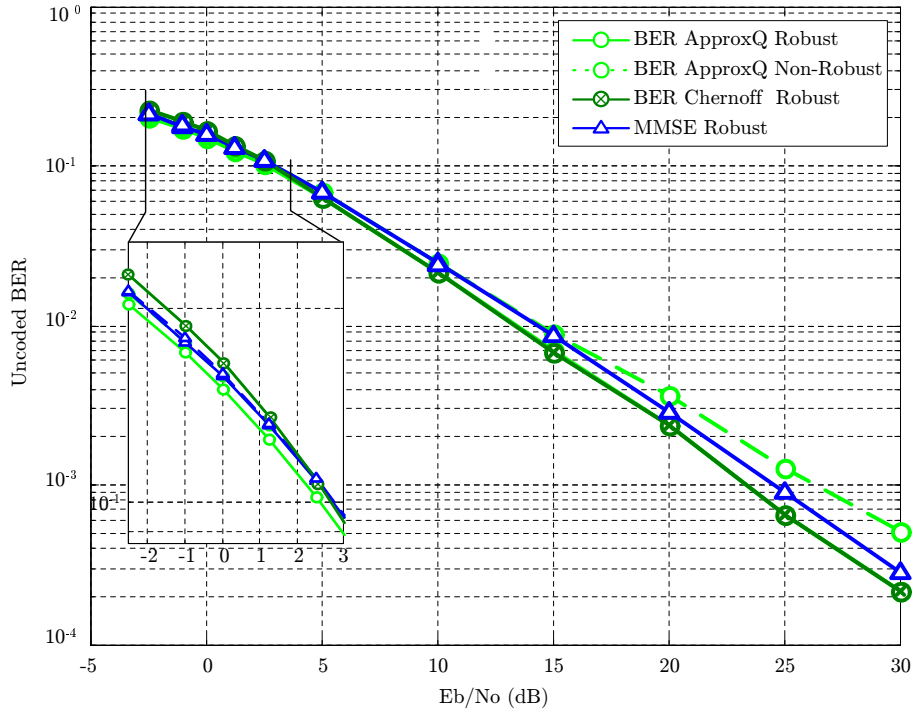


Figure 4.5: Uncoded BER comparison between different power allocation strategies. $M_T = 1$, $M_R = 1$. Transmitter uncertainty: $\sigma^2_{\epsilon_{Tx}} = 0.12$ ($\rho = 0.1$) Receiver uncertainty: Perfect CSI

$M = \{2, 3\}$. For this configuration different results can be obtained as a function of the unitary matrix $\mathbf{T}$. Hence, the performance of the algorithms is compared when this unitary matrix is omitted (i.e.,$\mathbf{T} = \mathbf{I}$) and when the matrix $\mathbf{T}$ is set to the DFT matrix. In the particular case of $\mathbf{T} = \mathbf{I}$ the elements of the diagonal matrix $\mathbf{\Phi}_k$ are obtained according to (4.48) for the MMSE solution and according to (4.77) for the uncoded BER solution. When matrix $\mathbf{T}$ is set to the DFT matrix, the same solution provided by (4.48) is used because, as it has been proved in Appendix 4.B, the MMSE cost function is insensitive to any unitary matrix at the right side of the linear transformation $\mathbf{F}_k$. On the contrary the minimum uncoded BER algorithm has been minimized applying the iterative solution (4.84) since (4.77) is suboptimum when the matrix $\mathbf{T}$ is different from the identity matrix.

The performance comparison in terms of M exhibits the tradeoff between uncoded BER and transmission rate. Observe that the differences between the algorithms is greater as M increases. Unlike the results given in Figure 4.6, when the number of symbols per subcarrier is increased (i.e., $M = 2$ or $M = 3$), the algorithms exhibit different performances, showing that the way to distribute the total power is crucial in these cases.

Some conclusions can also be drawn on the importance of the unitary matrix $\mathbf{T}$ when analyzing the performance losses if this unitary matrix is omitted. As it can be shown, the performance

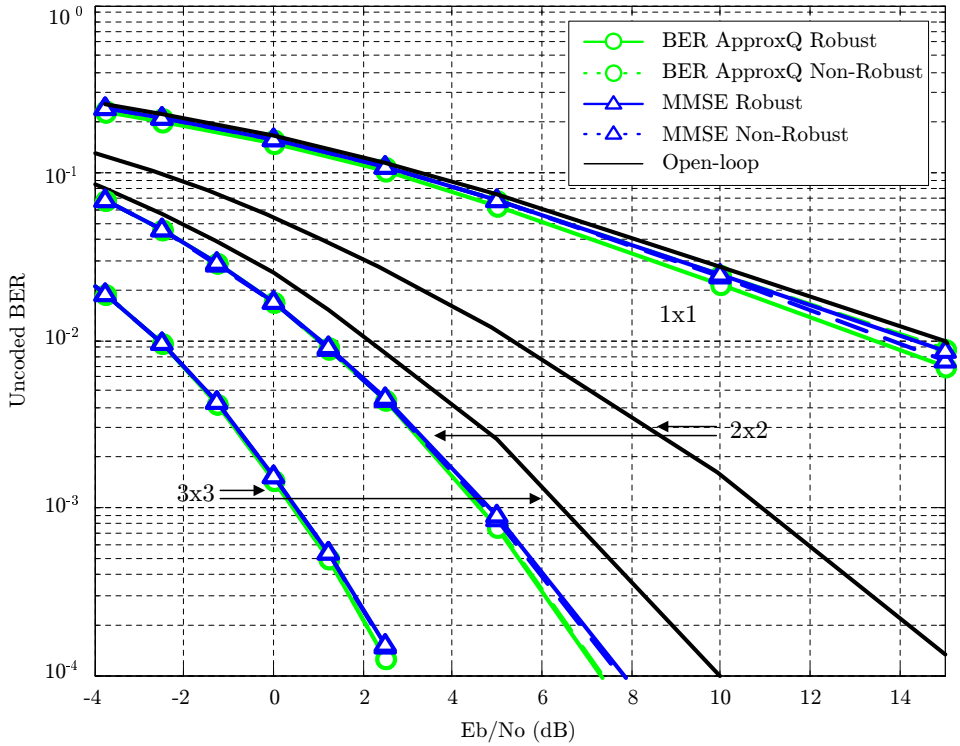


Figure 4.6: Uncoded BER comparison between different power allocation strategies. $M_T \times M_R = \{1 \times 1; 2 \times 2; 3 \times 3\}$, $M = 1$. Transmitter uncertainty: $\sigma^2_{\epsilon_{Tx}} = 0.12$ ($\rho = 0.1$) Receiver uncertainty: Perfect CSI

losses are greater as M or, equivalently, the dispersion of the channel singular values increases. This result confirms that the unitary matrix alleviates loss of space diversity caused by the channel diagonalization.

Finally, Figure 4.7 shows that the *'BER ApproxQ Robust'* algorithm has best performance in terms of uncoded BER, as it was expected. The poor performance of the *'MMSE Robust'* algorithm when the unitary matrix $\mathbf{T}$ is omitted is explained analyzing in Figure 4.4 how this algorithm allocates the power resource.

4.5.3 Performance vs CSI quality at the transmitter

To test the performance of the proposed algorithms when channel uncertainty increases, Figure 4.8 shows the minimum Eb/No required to achieve an uncoded BER$\leq 10^{-3}$. The required Eb/No is plotted as a function of the channel uncertainty degree at the transmitter ρ, defined as:

$$\rho = \frac{\sigma^2_{\epsilon_{Tx}}}{\sigma^2_{\epsilon_{Tx}} + \sigma^2_h} \tag{4.87}$$

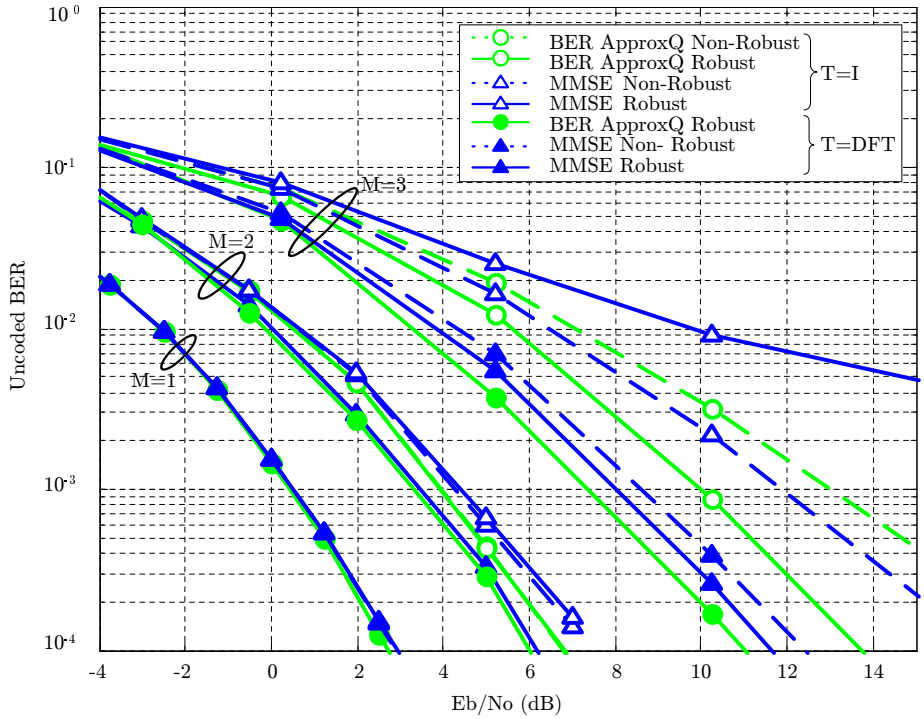


Figure 4.7: Uncoded BER comparison between different power allocation strategies. $M_T = 3$, $M_R = 3$, $M = \{1; 2; 3\}$. Transmitter uncertainty: $\sigma^2_{\epsilon_{Tx}} = 0.12$ ($\rho = 0.1$) Receiver uncertainty: Perfect CSI

where $\rho = 0$ denotes perfect CSI, whereas $\rho = 1$ means no channel knowledge. The channel uncertainty at the receiver is proportional to the noise variance. As a reference, the point $\rho = 0$ was simulated assuming perfect CSI at both transmitter and receiver. Two simulations are presented. A first plot analyzes the minimum uncoded BER design for different MIMO configurations: 1×1; 2×2; 3×3 assuming the maximum transmission rate (i.e., $M = \{1; 2; 3\}$). A second simulation evaluates the MMSE and the minimum BER algorithms for the MIMO configuration 3×3 with $M = \{1; 3\}$.

As shown in Figure 4.8a, the robustness of the proposed algorithm is evidenced since the *'BER ApproxQ Robust'* solution always outperforms *'BER ApproxQ Non-Robust'* one. Note that when CSI quality at the transmitter degrades (i.e., $\rho \to 1$) the *'BER ApproxQ Robust'* design tends to the open-loop solution (i.e., equally power allocation for all subcarriers and antennas) whereas the performance of the *'BER ApproxQ Non-Robust'* algorithm is degraded.

Comparing in Figure 4.8b the MMSE and the minimum uncoded BER algorithms, it is shown that the *'MMSE Robust'* algorithm only exhibits a good performance in terms of uncoded BER for low channel uncertainties. Note that for $M = 3$ the performance of the *'MMSE Robust'* algorithm is very close to the *'BER ApproxQ Robust'* algorithms when the CSI quality is high

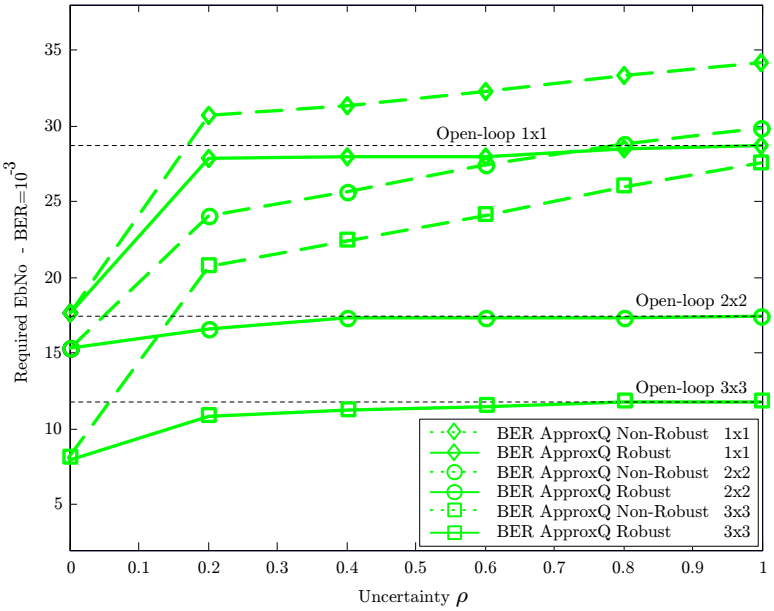
(i.e.,$\rho < 0.2$), but it performs worse than open-loop when the CSI quality degrades. When the transmission rate is decreased (i.e., $M = 1$) the difference in the performances is reduced, but still, when $\rho \to 1$, the *'MMSE Robust'* algorithm degrades rapidly. According to the previous results, and recalling that the *'BER ApproxQ Robust'* algorithm has been computed applying an iterative solution since the closed-form solution in (4.77) is suboptimum when $M > 1$, it can be concluded that when CSI quality at the transmitter is good (i.e.,$\rho < 0.2$), the *'MMSE Robust'* algorithm, which is computed by means of the closed-form solution given by (4.48)-(4.49), is an alternative to the *'BER ApproxQ Robust'* algorithm in terms of uncoded BER. On the contrary, when the CSI quality is decreased, the performance of the *'MMSE Robust'* algorithm degrades rapidly, especially for $M = 3$, and thus, the best solution is the *'BER ApproxQ Robust'* algorithm that adapts to the channel uncertainty and tends to the open-loop solution.

4.6 Discussion on coded BER

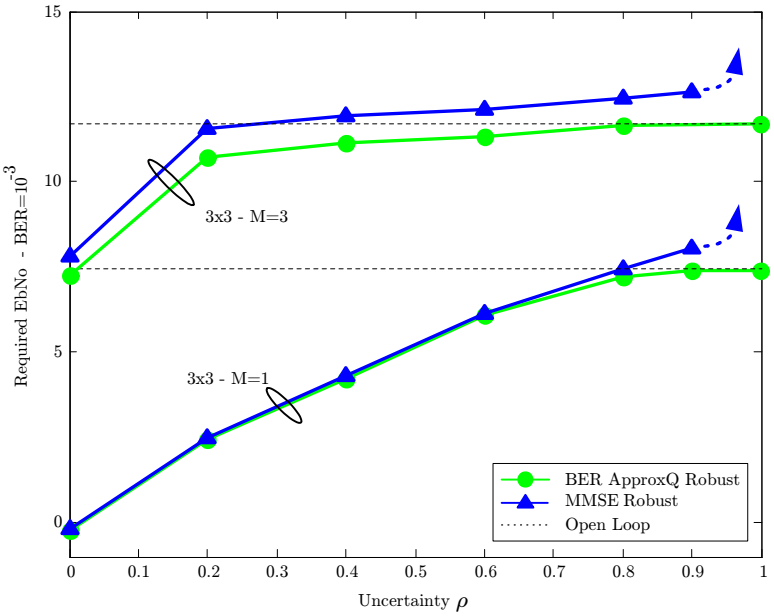
Despite the fact that neither the MMSE algorithm nor the minimum uncoded BER algorithm have been optimized taking into account a channel code, it is interesting to evaluate the proposed algorithms in terms of coded BER since most communication schemes include this stage.

The performance of the algorithms has been compared for the specific channel code described in HIPERLAN/2. For the evaluation of the coded BER, due to the presence of the block interleaver, it is too complex to jointly estimate all the bits transmitted in the symbol $\mathbf{x}_k$ as it was done to evaluate the uncoded BER (4.86). Instead, as the goal is not focused on the receiver architecture but on the design of the transmitter, a simplified decoder has been implemented that provides a bound on the performance achieved by practical decoding schemes. The coded BER results subsequently presented have been obtained using a genie decoder that computes the log-likelihood of a symbol $\mathbf{x}_k(l)$ assuming that all other symbols that were transmitted simultaneously in in vector $\mathbf{x}_k$ are known to the receiver. The coded BER achieved by this scheme bounds and iterative decoder like the one proposed in [Ton00], performing an infinite number of iterations and assuming that no decoding errors where made.

Figure 4.9 depicts the performance of the proposed algorithms in terms of coded BER at the decoder output for the 3×3 ($M = 3$) antenna configuration when the receiver has an estimation variance proportional to σ_n^2 (the study in terms of coded BER of the algorithm that minimizes the uncoded BER using the Chernoff bound is left to the next chapter). It can be seen that when matrix $\mathbf{T}$ is omitted, *'MMSE Robust'* algorithm achieves best performance, followed by *'MMSE Non-Robust'*, *'BER ApproxQ Robust'* and *'BER ApproxQ Non-Robust'* solutions, whereas when matrix $\mathbf{T}$ is set to the DFT matrix, the *'BER ApproxQ Robust'* algorithm outperforms others. This behavior can be intuitively understood comparing, in Figure 4.4, the power allocation strategy followed by the different algorithms. Note that the best algorithms in terms of coded



(a)



(b)

Figure 4.8: Minimum Eb/No that achieves an uncoded BER $\leq 10^{-3}$ vs the transmitter uncertainty ρ. Receiver uncertainty: $\sigma^2_{\epsilon_{Rx}} = 0.375\sigma^2_n$. (a) Uncoded BER algorithm comparison for different MIMO configurations: 1×1 $M = 1$; 2×2 $M = 2$; 3×3 $M = 3$. (b) Uncoded BER and MMSE algorithms comparison for 3×3 $M = \{1, 3\}$.

BER are those that allocate more power to the good subchannels and penalize the weakest ones. This strategy, as opposite to that one followed by the algorithms that minimize uncoded BER, induces some systematic errors in the deepest faded channels. However, if the amount of errors is within the correcting capability of the channel code, these errors can be corrected, explaining the overall increased performance at the decoder output in terms of coded BER.

To get more insight in the performance of the proposed power allocation algorithms in terms of coded BER, Figure 4.10 evaluates the capacity and cut-off rate for each algorithm. Both parameters can be used as an indicator of the coded BER performance that is independent of the specific channel code. While the capacity is the theoretical upper limit on data rates where arbitrarily small coded BER can be achieved with coding, from the practical point of view, it is difficult to attain this upper limit. However, the cut-off rate provides a lower bound on the capacity and, until the proposal of turbo codes, it was considered to be the limit for the coding techniques that had practical interest. Figure 4.10 shows the *Complementary Cumulative Distribution Function* (CCDF) curves of the capacity (in bps/Hz), which is computed as:

$$C = \frac{1}{K}\sum_{k=1}^{K}\log_2\left(\det\left(\mathbf{I}+\frac{\mathbf{F}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{F}_k}{\sigma_n^2/\sigma_x^2}\right)\right) \tag{4.88}$$

and the cut-off rate (in *bits/channel use*), which is computed as (see Chapter 5 for more details):

$$R_0 = -\frac{1}{K}\sum_{k=1}^{K}\log_2\left(\frac{1}{\mathcal{N}^{2M}}\sum_{i=1}^{\mathcal{N}^M}\sum_{j=1}^{\mathcal{N}^M}e^{-\frac{1}{4\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_{ijk}|^2}\right) \tag{4.89}$$

When the unitary matrix $\mathbf{T}$ is omitted, the MMSE algorithms outperform the algorithms that minimize the uncoded BER in terms of capacity and cut-off rate, which explains why the MMSE solutions result in better performance than minimum uncoded BER solutions. To explain the behavior of the solutions when the unitary matrix $\mathbf{T}$ is set to the DFT, note that the capacity in (4.88) is insensitive to any unitary matrix at the right side of the linear transformation $\mathbf{F}_k$, and hence, this parameter cannot be used in our case as a measure of the coded BER performance. On the contrary, the cut-off rate is sensitive to the matrix $\mathbf{T}$ and justifies the relevance of this matrix. As shown in Figure 4.10, when the unitary matrix is set to the DFT, the cut-off rate is increased, which justifies that the coded BER is decreased. Moreover, the performance of the *'BER ApproxQ Robust'* algorithm when $\mathbf{T}$ is the DFT as the best solution in terms of coded BER is explained since this algorithm exhibits the best cut-off rate.

Further the aim of this chapter, the study of the closed-loop algorithms when a channel code is introduced in the communication scheme opens a new line of research based on the design of appropriate transceivers that guarantee good performances in terms of coded BER. The results of the analysis and the conclusions drawn in this section for the MMSE and the minimum uncoded BER algorithms raise interesting questions about how to design an algorithms that minimize the coded BER.

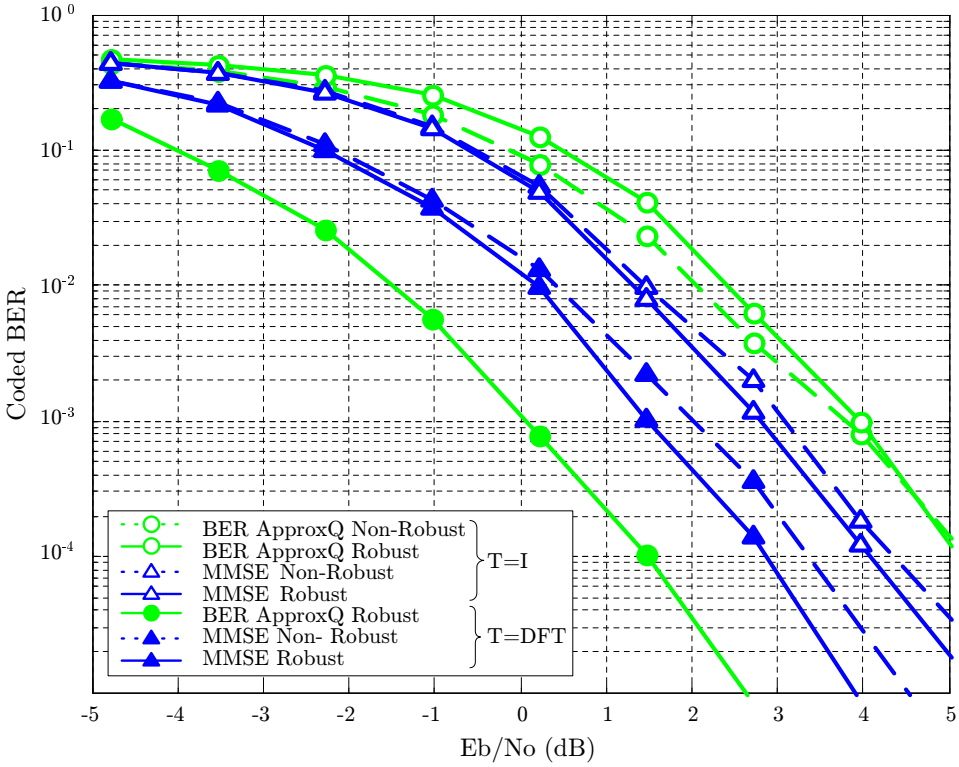
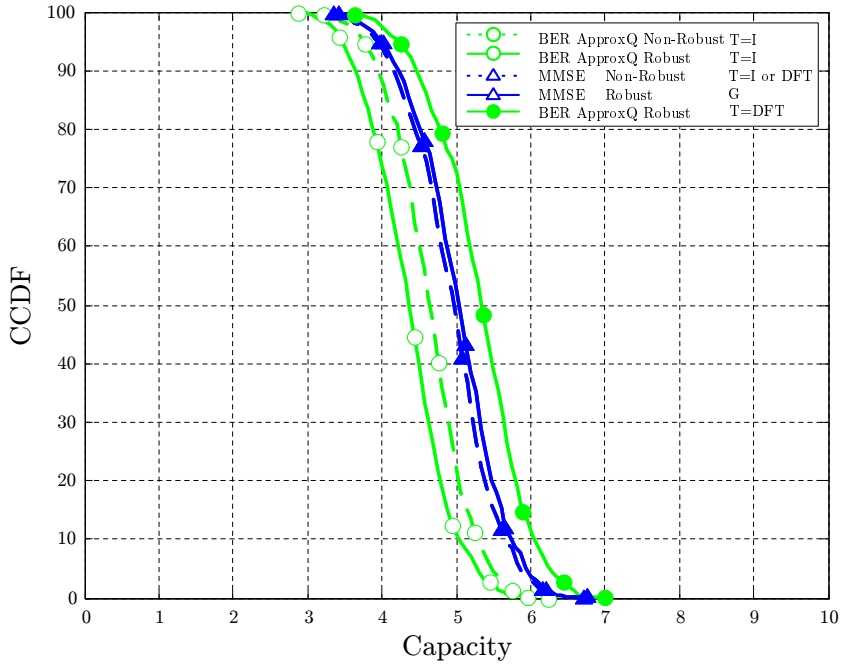


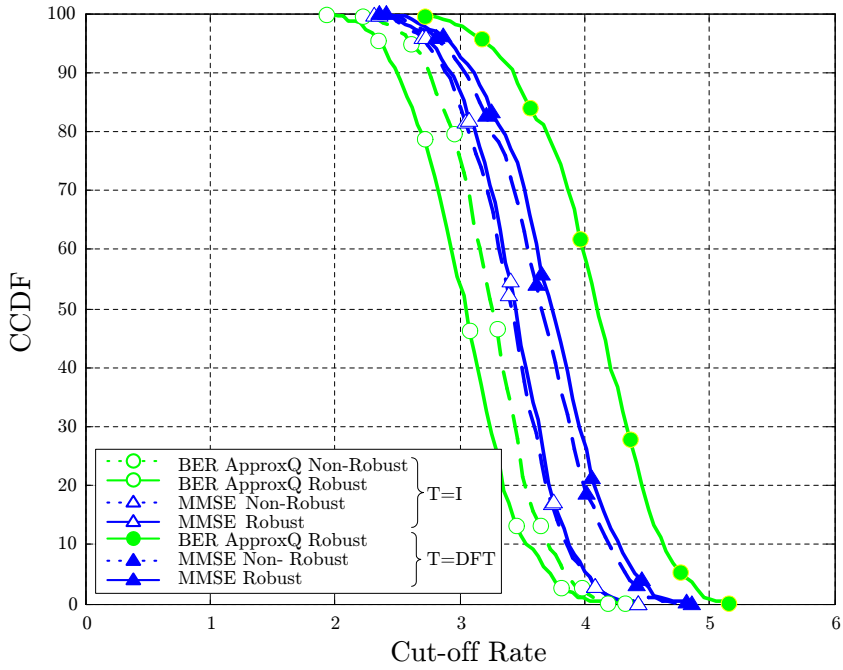
Figure 4.9: Coded BER comparison between different power allocation strategies. $M_T = 3$, $M_R = 3$, $M = 3$. Transmitter uncertainty: $\sigma^2_{\epsilon_{Tx}} = 0.12$ ($\rho = 0.1$). Receiver uncertainty: $\sigma^2_{\epsilon_{Rx}} = 0.375\sigma^2_n$

A first question to consider is how to design the objective function or, equivalently, which might be the parameter to be optimized. Information theoretic criteria could be used for this purpose. A possible choice is to design algorithms based on the channel capacity or the mutual information for a finite constellation, as limits of reliable communications. An alternative to these theoretical limits is the design of an algorithm based on the maximization of the cut-off rate, which has been considered a practical bound when sequential decoding strategies are used. Both, the mutual information and the cut-off rate parameters will be fully studied in the next chapter to answer the question.

A second question that arises, specifically from the power allocation strategies given in Section 4.5.1, is how the null and highly faded subcarriers are arranged in time. As it will be shown in the next chapter an appropriate distribution of the channel errors in the code-word can greatly improve the performance of the error correcting code. In this arrangement the interleaver plays an important role and, hence, the accurate design of new interleavers offers another line of research that will also be treated in the next chapter.



(a)



(b)

Figure 4.10: Complementary cumulative distribution function (CCDF) curves of the capacity (a) and cut-off rate (b) for different power allocation strategies. $M_T = 3$, $M_R = 3$. Transmitter uncertainty: $\sigma^2_{e_{Tx}} = 0.12$ ($\rho = 0.1$). SNR=0dB

4.7 Conclusions and future work

This chapter has presented a Bayesian approach for the design of linear precoding schemes that are robust to channel estimation errors. The proposed linear transformations have been designed according to a minimum MSE and a minimum uncoded BER criterion, both of them subject to an average power constraint across all antennas and subcarriers.

Closed-form solutions that are based on the SVD of the so called equivalent channel have been derived for both algorithms. The prefiltering matrix derived for the minimum MSE criterion has been shown to be optimum, whereas the matrix structure given for the minimum uncoded BER cost function is optimum when one symbol per subcarrier is transmitted (i.e., M=1) and SNR is high. The algorithms converge to the results published previously in the literature in the extreme cases of very high or very low uncertainty in the CSI. Besides, it has been shown that the uncoded BER design can be regarded as a reconfigurable algorithm that adapts the transmitted signal to the available channel knowledge, providing a solution that converges to the open-loop design (i.e.,the same power is allocated across all antennas and subcarriers) for the case of no channel knowledge and to the closed-loop design with perfect CSI for the case of no uncertainty. The minimum uncoded BER design has been formulated based on a generalized exponential bound of the function $Q(\sqrt{x})$ that includes the Chernoff bound as a particular case.

The two algorithms have been compared in terms of uncoded BER and coded BER for the specific channel code described in the HIPERLAN/2 standard using a ML detector at the receiver. Moreover, the capacity and the cut-off rate have also been used as measures of the coded BER performance, which are independent of the specific channel code. The results have shown that the robust algorithms exhibit a lower sensitivity to channel estimation errors when compared with the nonrobust techniques. In terms of uncoded BER it has been proved that, although the algorithm that minimizes the uncoded BER has the best performance, the MMSE algorithm exhibits a good performance for low uncertainty in the CSI. Finally, in terms of coded BER, it has been shown that the robust algorithm that minimizes the uncoded BER outperforms the rest of algorithms.

The performances observed when a channel coding stage is introduced, open a new line of study based on the search of appropriate parameters to be optimized that guarantee good performances in terms of coded BER. This parameters will be widely studied in the next chapter and the performance of the derived designs will be compared with those presented in this chapter.

Appendix 4.A Derivation of some channel statistics

This appendix derives the expression for some channel statistics introduced in Section 4.2.1 and used in Sections 4.3 and 4.4.

Channel mean:

For the sake of completeness in the channel model we will assume that may exist a line-of-sight between the transmitter and the receiver, and hence, the first coefficient in the channel impulse response is non-zero mean. This non-zero mean channel impulse response will force the vector $\mathbf{h}$, which contains the complete channel response in the frequency domain, to be non-zero mean. Hence, the mean of the vector $\mathbf{h}$ will be denoted as:

$$\mathbf{m_h} = E\{\mathbf{h}\} \tag{4.90}$$

Channel covariance:

Due to the structure of vector $\mathbf{h}$, whose elements contain the frequency response of the $M_T M_R$ channels given by the MIMO channel, the covariance of the channel vector $\mathbf{h}$ will be defined by the correlation between the subcarriers and the correlation between the antennas. Denoting matrix $\mathbf{P}$ as a matrix whose entries contain the DFT of the variance in the channel impulse response, and matrix $\mathbf{R}$ as the antenna correlation matrix (a clear definition of both matrices can be found below), the covariance of the vector $\mathbf{h}$ is given by:

$$\mathbf{C_{hh}} = E\left\{(\mathbf{h}-\mathbf{m_h})(\mathbf{h}-\mathbf{m_h})^H\right\} = \mathbf{P}\otimes\mathbf{R} \tag{4.91}$$

Proof. To prove the result in (4.91) we will first compute the covariance between two arbitrary elements of vector $\mathbf{h}$. Denoting $h_k^{i,j}$ as the kth subcarrier frequency response for the channel between the ith transmitting and the jth receiving antenna:

$$h_k^{i,j} = \sum_{l=0}^{L-1} h_{ij}^{(t)}(l) e^{-j\frac{2\pi}{K}kl} \tag{4.92}$$

the covariance $E\left\{\left(h_k^{i,j} - m_{h_k^{i,j}}\right)\left(h_q^{m,n} - m_{h_q^{m,n}}\right)^*\right\}$ is next derived under the following assumptions:

1. The L coefficients of the channel impulse response are supposed to accomplish the uncorrelated assumption (see Section 3.3).
2. Denoting $\rho_{Tx}^{(i,m)}$ as the spatial correlation coefficient between antennas i and m at the transmitter, and $\rho_{Rx}^{(j,n)}$ as the spatial correlation coefficient between

antennas j and n at the receiver, the correlation between two different sets of antenna pairs (i,j) and (m,n) is assumed to be $\rho_{Tx}^{(i,m)} \cdot \rho_{Rx}^{(j,n)}$ [Ped90].

3. The expectation of the lth tap of the channel impulse response $h_{ij}^{(t)}(l)$ is denoted as $\overline{h}_{ij}^{(t)}(l) = E\left\{h_{ij}^{(t)}(l)\right\}$.

and the covariance is computed as follows:

$$
\begin{aligned}
&E\left\{\left(h_k^{i,j} - m_{h_k^{i,j}}\right)\left(h_q^{m,n} - m_{h_q^{m,n}}\right)^*\right\} \\
&\quad = E\left\{\sum_{l=0}^{L-1}\left(h_{ij}^{(t)}(l) - \overline{h}_{ij}^{(t)}(l)\right)e^{-j\frac{2\pi}{K}kl}\sum_{l'=0}^{L-1}\left(h_{mn}^{(t)}(l') - \overline{h}_{mn}^{(t)}(l')\right)^* e^{+j\frac{2\pi}{K}ql'}\right\} \\
&\quad = E\left\{\sum_{l=0}^{L-1}\left(h_{ij}^{(t)}(l) - \overline{h}_{ij}^{(t)}(l)\right)\left(h_{mn}^{(t)}(l) - \overline{h}_{mn}^{(t)}(l)\right)^* e^{-j\frac{2\pi}{K}(k-q)l}\right\} \\
&\quad = \sum_{l=0}^{L-1} E\left\{\left(h_{ij}^{(t)}(l) - \overline{h}_{ij}^{(t)}(l)\right)\left(h_{mn}^{(t)}(l) - \overline{h}_{mn}^{(t)}(l)\right)^*\right\}e^{-j\frac{2\pi}{K}(k-q)l} \\
&\quad = \rho_{Tx}^{(i,m)} \cdot \rho_{Rx}^{(j,n)} \sum_{l=0}^{L-1} E\left\{\left|h^{(t)}(l) - \overline{h}^{(t)}(l)\right|^2\right\}e^{-j\frac{2\pi}{K}(k-q)l} \\
&\quad = \rho_{Tx}^{(i,m)} \cdot \rho_{Rx}^{(j,n)}\, P\left(k-q\right)
\end{aligned}
\tag{4.93}
$$

where assumptions 1, 2 and 3 have been used in the second, fourth and first equalities, respectively, and $P(k)$ has been defined as the kth element of the DFT of the variance in the channel impulse response coefficients:

$$
P(k) = \sum_{l=0}^{L} E\left\{\left|h^{(t)}(l) - \overline{h^{(t)}}(l)\right|^2\right\}e^{-j\frac{2\pi}{N}\,k\,l} \tag{4.94}
$$

where subindex ij that denote the transmitter-receiver antenna pair have been omitted because identical power delay profile is assumed for all the subchannels.

Next, the derivation of the covariance in (4.93) will be expanded to the covariance of vector $\mathbf{h}_k$, which contains the MIMO channel responses for the kth subcarrier. Defining this vector as:

$$
\mathbf{h}_k = \mathrm{vec}\left\{\mathbf{H}_k\right\} = \left[h_k^{1,1} \ldots\ h_k^{1,M_R} \ldots\ h_k^{M_T,1} \ldots\ h_k^{M_T,M_R}\right]^T \tag{4.95}
$$

the solution in (4.93) can be used to compute the covariance:

$$
E\left\{\left(\mathbf{h}_k - \mathbf{m}_{\mathbf{h}_k}\right)\left(\mathbf{h}_q - \mathbf{m}_{\mathbf{h}_q}\right)^H\right\} = \mathbf{R}\, P\left(k-q\right) \tag{4.96}
$$

where $\mathbf{R}$ is the antenna correlation matrix defined as:

$$\mathbf{R} = \begin{bmatrix} \rho_{Tx}^{(1,1)} & \cdots & \rho_{Tx}^{(1,M_T)} \\ \vdots & \ddots & \vdots \\ \rho_{Tx}^{(M_T,1)} & \cdots & \rho_{Tx}^{(M_T,M_T)} \end{bmatrix} \otimes \begin{bmatrix} \rho_{Rx}^{(1,1)} & \cdots & \rho_{Rx}^{(1,M_R)} \\ \vdots & \ddots & \vdots \\ \rho_{Rx}^{(M_R,1)} & \cdots & \rho_{Rx}^{(M_R,M_R)} \end{bmatrix} \tag{4.97}$$

Finally, the covariance of the whole channel vector $\mathbf{h}$ is given by:

$$\begin{aligned} & E\left\{(\mathbf{h}-\mathbf{m_h})(\mathbf{h}-\mathbf{m_h})^H\right\} \\ & = \begin{bmatrix} P(0)\ \mathbf{R} & P(1)\ \mathbf{R} & \cdots & P(K-1)\ \mathbf{R} \\ P(1)^*\ \mathbf{R} & P(0)\ \mathbf{R} & \cdots & P(K-2)\ \mathbf{R} \\ \vdots & \vdots & \ddots & \vdots \\ P(K-1)^*\ \mathbf{R} & P(K-2)^*\ \mathbf{R} & \cdots & P(0)\ \mathbf{R} \end{bmatrix} \\ & = \begin{bmatrix} P(0) & P(1) & \cdots & P(K-1) \\ P(1)^* & P(0) & \cdots & P(K-2) \\ \vdots & \vdots & \ddots & \vdots \\ P(K-1)^* & P(K-2)^* & \cdots & P(0) \end{bmatrix} \otimes \mathbf{R} \\ & = \mathbf{P} \otimes \mathbf{R} \end{aligned} \tag{4.98}$$

where $\mathbf{P}$ has been defined as the Hermitian Toeplitz matrix whose first row is the DFT of the variance in the channel impulse response coefficients. Note, moreover, that since the variance in the channel impulse response is real, the elements of matrix $\mathbf{P}$ satisfy: $P(K-l)^* = P(l)$ and thus matrix $\mathbf{P}$ is a circulant matrix. ■

Error covariance:

Denoting $\boldsymbol{\varepsilon}$ as the channel estimation error in the frequency domain, whose entries are given by:

$$\varepsilon_k^{i,j} = \sum_{l=0}^{L-1} e_{ij}^{(t)}(l) e^{-j\frac{2\pi}{K}kl} \tag{4.99}$$

and assuming that the channel estimation error for each one of the taps in the channel impulse response and for each one of the antenna pairs are uncorrelated, the covariance of the vector $\boldsymbol{\varepsilon}$ is given by:

$$E\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^H\right\} = \mathbf{E} \otimes \mathbf{I} \tag{4.100}$$

Proof. To prove this covariance it suffices to repeat the previous steps followed to derive the channel covariance $\mathbf{C_{hh}}$ substituting the channel response by the channel

estimation error, assuming uncorrelated antennas (i.e., $\mathbf{R} = \mathbf{I}$), and defining matrix $\mathbf{E}$ with the same structure as matrix $\mathbf{P}$ whose entries contain the DFT of the variance of the channel estimation error for each tap:

$$E(k) = \sum_{l=0}^{K-1} E\left\{|e^{(t)}(l)|^2\right\} e^{-j\frac{2\pi}{K}(k)l} \tag{4.101}$$

■

Conditional mean and conditional covariance:

The Bayesian approach adopted in this thesis is based on the definition of the next model for the estimated channel:

$$\hat{\mathbf{h}} = \sqrt{\frac{\sigma_h^2}{\sigma_h^2 + \sigma_\epsilon^2}}\,(\mathbf{h} + \boldsymbol{\varepsilon}) \tag{4.102}$$

and assumes that the channel $\mathbf{h}$ and its estimate are jointly Gaussian.

Under previous assumptions the conditional mean and conditional covariance of the channel $\mathbf{h}$ given $\hat{\mathbf{h}}$ will be computed. Prior to the derivation of those expressions let us to introduce some **intermediate results** that will be latter used:

■
$$\begin{aligned} \mathbf{m}_{\hat{\mathbf{h}}} &= E\left\{\hat{\mathbf{h}}\right\} \\ &= E\left\{\sqrt{\tfrac{\sigma_h^2}{\sigma_h^2+\sigma_\epsilon^2}}\,(\mathbf{h}+\boldsymbol{\varepsilon})\right\} \\ &= \sqrt{\tfrac{\sigma_h^2}{\sigma_h^2+\sigma_\epsilon^2}}\,E\left\{(\mathbf{h}+\boldsymbol{\varepsilon})\right\} \\ &= \sqrt{\tfrac{\sigma_h^2}{\sigma_h^2+\sigma_\epsilon^2}}\,\mathbf{m}_{\mathbf{h}} \end{aligned} \tag{4.103}$$

where the assumption of zero-mean channel estimation error has been used, and $\mathbf{m}_{\mathbf{h}}$, as introduced in (4.90), is the channel mean.

■
$$\begin{aligned} \mathbf{C}_{\mathbf{h}\hat{\mathbf{h}}} &= E\left\{(\mathbf{h}-\mathbf{m}_{\mathbf{h}})\left(\hat{\mathbf{h}}-\mathbf{m}_{\hat{\mathbf{h}}}\right)^H\right\} \\ &= E\left\{(\mathbf{h}-\mathbf{m}_{\mathbf{h}})\sqrt{\tfrac{\sigma_h^2}{\sigma_h^2+\sigma_\epsilon^2}}\,(\mathbf{h}+\boldsymbol{\varepsilon}-\mathbf{m}_{\mathbf{h}})^H\right\} \\ &= \sqrt{\tfrac{\sigma_h^2}{\sigma_h^2+\sigma_\epsilon^2}}\,E\left\{(\mathbf{h}-\mathbf{m}_{\mathbf{h}})(\mathbf{h}-\mathbf{m}_{\mathbf{h}})^H\right\} \\ &= \sqrt{\tfrac{\sigma_h^2}{\sigma_h^2+\sigma_\epsilon^2}}\,\mathbf{C}_{\mathbf{hh}} \end{aligned} \tag{4.104}$$

where the result in (4.103) and the assumption of zero-mean channel estimation error has been used in the second and third equalities, respectively.

$$
\begin{aligned}
\blacksquare\ \mathbf{C}_{\hat{\mathbf{h}}\hat{\mathbf{h}}} &= E\left\{\left(\hat{\mathbf{h}}-\mathbf{m}_{\hat{\mathbf{h}}}\right)\left(\hat{\mathbf{h}}-\mathbf{m}_{\hat{\mathbf{h}}}\right)^{H}\right\} \\
&= E\left\{\sqrt{\tfrac{\sigma_h^2}{\sigma_h^2+\sigma_\epsilon^2}}\left(\mathbf{h}+\boldsymbol{\varepsilon}-\mathbf{m}_{\mathbf{h}}\right)\sqrt{\tfrac{\sigma_h^2}{\sigma_h^2+\sigma_\epsilon^2}}\left(\mathbf{h}+\boldsymbol{\varepsilon}-\mathbf{m}_{\mathbf{h}}\right)^{H}\right\} \\
&= \left(\sqrt{\tfrac{\sigma_h^2}{\sigma_h^2+\sigma_\epsilon^2}}\right)^{2}\left(E\left\{\left(\mathbf{h}-\mathbf{m}_{\mathbf{h}}\right)\left(\mathbf{h}-\mathbf{m}_{\mathbf{h}}\right)^{H}\right\}+E\left\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{H}\right\}\right) \\
&= \tfrac{\sigma_h^2}{\sigma_h^2+\sigma_\epsilon^2}\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)
\end{aligned}
\tag{4.105}
$$

where the result in (4.103) and the assumption of zero-mean channel estimation error has been used in the second and third equalities, respectively.

Under the assumption that the channel $\mathbf{h}$ and its estimate are jointly Gaussian, the **conditional mean** is defined as [Kay93, p.324]:

$$
\mathbf{m}_{\mathbf{h}|\hat{\mathbf{h}}} = \mathbf{m}_{\mathbf{h}} + \mathbf{C}_{\mathbf{h}\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}}\hat{\mathbf{h}}}^{-1}\left(\hat{\mathbf{h}}-\mathbf{m}_{\hat{\mathbf{h}}}\right) \tag{4.106}
$$

and using the intermediate results in (4.103)-(4.105) this conditional mean can be straightforwardly computed as follows:

$$
\begin{aligned}
\mathbf{m}_{\mathbf{h}|\hat{\mathbf{h}}} &= \mathbf{m}_{\mathbf{h}} + \mathbf{C}_{\mathbf{h}\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}}\hat{\mathbf{h}}}^{-1}\left(\hat{\mathbf{h}}-\mathbf{m}_{\hat{\mathbf{h}}}\right) \\
&= \mathbf{C}_{\mathbf{h}\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}}\hat{\mathbf{h}}}^{-1}\hat{\mathbf{h}} + \left(\mathbf{I}-\sqrt{\tfrac{\sigma_h^2}{\sigma_h^2+\sigma_\epsilon^2}}\mathbf{C}_{\mathbf{h}\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}}\hat{\mathbf{h}}}^{-1}\right)\mathbf{m}_{\mathbf{h}} \\
&= \sqrt{\tfrac{\sigma_h^2+\sigma_\epsilon^2}{\sigma_h^2}}\mathbf{C}_{\mathbf{hh}}\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)^{-1}\hat{\mathbf{h}} + \left(\mathbf{I}-\mathbf{C}_{\mathbf{hh}}\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)^{-1}\right)\mathbf{m}_{\mathbf{h}} \\
&= \sqrt{\tfrac{\sigma_h^2+\sigma_\epsilon^2}{\sigma_h^2}}\mathbf{C}_{\mathbf{hh}}\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)^{-1}\hat{\mathbf{h}} + \left(\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)-\mathbf{C}_{\mathbf{hh}}\right)\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)^{-1}\mathbf{m}_{\mathbf{h}} \\
&= \sqrt{\tfrac{\sigma_h^2+\sigma_\epsilon^2}{\sigma_h^2}}\mathbf{C}_{\mathbf{hh}}\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)^{-1}\hat{\mathbf{h}} + \mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)^{-1}\mathbf{m}_{\mathbf{h}} \\
&= \sqrt{1+\frac{\sigma_\epsilon^2}{\sigma_h^2}}\left(\left(\mathbf{P}\otimes\mathbf{R}\right)\left(\mathbf{P}\otimes\mathbf{R}+\mathbf{E}\otimes\mathbf{I}\right)^{-1}\right)\hat{\mathbf{h}} + \left(\mathbf{E}\otimes\mathbf{I}\right)\left(\mathbf{P}\otimes\mathbf{R}+\mathbf{E}\otimes\mathbf{I}\right)^{-1}\mathbf{m}_{\mathbf{h}}
\end{aligned}
\tag{4.107}
$$

Under the jointly Gaussian assumption of $\mathbf{h}$ and $\hat{\mathbf{h}}$, the **conditional covariance** is given by [Kay93, p.324]:

$$
\mathbf{C}_{\mathbf{h}|\hat{\mathbf{h}}} = \mathbf{C}_{\mathbf{hh}} - \mathbf{C}_{\mathbf{h}\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}}\hat{\mathbf{h}}}^{-1}\mathbf{C}_{\hat{\mathbf{h}}\mathbf{h}} \tag{4.108}
$$

and using the results in (4.103)-(4.105) this expression becomes:

$$
\begin{aligned}
\mathbf{C}_{\mathbf{h}|\hat{\mathbf{h}}} &= \mathbf{C}_{\mathbf{hh}} - \mathbf{C}_{\mathbf{h}\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}}\hat{\mathbf{h}}}^{-1}\mathbf{C}_{\hat{\mathbf{h}}\mathbf{h}} \\
&= \mathbf{C}_{\mathbf{hh}} - \mathbf{C}_{\mathbf{hh}}\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)^{-1}\mathbf{C}_{\mathbf{hh}} \\
&= \mathbf{C}_{\mathbf{hh}}\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)^{-1}\left(\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)-\mathbf{C}_{\mathbf{hh}}\right) \\
&= \mathbf{C}_{\mathbf{hh}}\left(\mathbf{C}_{\mathbf{hh}}+\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right)^{-1}\mathbf{C}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} \\
&= \left(\mathbf{P}\otimes\mathbf{R}\right)\left(\mathbf{P}\otimes\mathbf{R}+\mathbf{E}\otimes\mathbf{I}\right)^{-1}\left(\mathbf{E}\otimes\mathbf{I}\right)
\end{aligned}
\tag{4.109}
$$

The **conditional mean and conditional covariance** expressions for the particular case of **uncorrelated antennas** (i.e., $\mathbf{R} = \mathbf{I}$), reduce to:

$$\mathbf{m}_{\mathbf{h}|\hat{\mathbf{h}}} = \sqrt{1+\frac{\sigma_\epsilon^2}{\sigma_h^2}}\left(\mathbf{P}\left(\mathbf{P}+\mathbf{E}\right)^{-1}\otimes\mathbf{I}\right)\hat{\mathbf{h}} + \left(\mathbf{E}\left(\mathbf{P}+\mathbf{E}\right)^{-1}\otimes\mathbf{I}\right)\mathbf{m}_{\mathbf{h}} \quad (4.110)$$

$$\mathbf{C}_{\mathbf{h}|\hat{\mathbf{h}}} = \left(\mathbf{P}\left(\mathbf{P}+\mathbf{E}\right)^{-1}\mathbf{E}\right)\otimes\mathbf{I} \quad (4.111)$$

where the following identities for the Kronecker product has been used [Mag99, p.27]:

$$\begin{aligned}\mathbf{A}\otimes\mathbf{I}+\mathbf{B}\otimes\mathbf{I} &= (\mathbf{A}+\mathbf{B})\otimes\mathbf{I}\\ (\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) &= \mathbf{AC}\otimes\mathbf{BD}\\ (\mathbf{A}\otimes\mathbf{B})^{-1} &= \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}\end{aligned} \quad (4.112)$$

Appendix 4.B Optimum decomposition of $\mathbf{F}_k$ and $\mathbf{G}_k$ matrices to minimize the MSE

This Appendix proves that the optimum decomposition of $\mathbf{F}_k$ and $\mathbf{G}_k$ matrices to minimize the MSE cost function:

$$\begin{aligned}\xi_k^{MMSE} = {} & \sigma_x^2\mathrm{Tr}\left\{\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)^H\left(\mathbf{G}_k\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k-\mathbf{I}\right)\right\}\\ & +\sigma_n^2\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}+\omega\sigma_x^2\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\}\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}\end{aligned} \quad (4.113)$$

is based on the SVD of the equivalent channel $\hat{\mathbf{H}}_k^{eq} = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{V}_k^H$:

$$\mathbf{F}_k = \mathbf{V}_k\mathbf{\Phi}_k\mathbf{T}_k \qquad \mathbf{G}_k = \mathbf{T}_k^H\mathbf{\Gamma}_k\mathbf{\Lambda}_k^{\#}\mathbf{U}_k^H \quad (4.114)$$

where $\mathbf{T}_k$ is any irrelevant unitary matrix and $\mathbf{\Phi}_k$, $\mathbf{\Gamma}_k$ are diagonal matrices.

Proof. It can be easily shown that the cost function (4.113) is convex both in $\mathbf{F}_k$ and $\mathbf{G}_k$ matrices since the Hessian is semipositive definite. Hence, differentiating the objective function with respect to the matrix $\mathbf{G}_k^H$ and equating the result to zero, the function to minimize is solved for $\mathbf{G}_k$ to obtain:

$$\mathbf{G}_k = \mathbf{F}_k^H\hat{\mathbf{H}}_k^{eqH}\left(\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k\mathbf{F}_k^H\hat{\mathbf{H}}_k^{eqH}+\sigma_n^2\mathbf{I}+\omega\sigma_x^2\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\}\mathbf{I}\right)^{-1} \quad (4.115)$$

For the sake of simplicity in the notation let us denote the matrix $\mathcal{A} = \hat{\mathbf{H}}_k^{eq}\mathbf{F}_k$ and the scalar $\mathcal{K} = \sigma_n^2 + \omega\sigma_x^2\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\}$. Using this new notation, matrix $\mathbf{G}_k$ can be written as: $\mathbf{G}_k = \mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}$. Next, substituting matrix $\mathbf{G}_k$ in (4.113), the MSE becomes:

$$\begin{aligned}
\xi_k^{MMSE} &= \mathrm{Tr}\left\{\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\mathcal{A}\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\mathcal{A}\right\} - 2\mathrm{Tr}\left\{\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\mathcal{A}\right\} + \mathrm{Tr}\left\{\mathbf{I}\right\} \\
&\quad +\omega\sigma_x^2\mathrm{Tr}\left\{\mathbf{F}\mathbf{F}^H\right\}\mathrm{Tr}\left\{\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-2}\mathcal{A}\right\} + \sigma_n^2\mathrm{Tr}\left\{\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-2}\mathcal{A}\right\} \\
&= \mathrm{Tr}\left\{\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\mathcal{A}\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\mathcal{A}\right\} - 2\mathrm{Tr}\left\{\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\mathcal{A}\right\} + \mathrm{Tr}\left\{\mathbf{I}\right\} \\
&\quad +\mathcal{K}\mathrm{Tr}\left\{\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-2}\mathcal{A}\right\} \\
&= \mathrm{Tr}\left\{\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\mathcal{A}\right\} - 2\mathrm{Tr}\left\{\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\mathcal{A}\right\} + \mathrm{Tr}\left\{\mathbf{I}\right\} \\
&= -\mathrm{Tr}\left\{\mathcal{A}^H\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\mathcal{A}\right\} + \mathrm{Tr}\left\{\mathbf{I}\right\} \\
&= -\mathrm{Tr}\left\{\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}-\mathcal{K}\mathbf{I}\right)\right\} + \mathrm{Tr}\left\{\mathbf{I}\right\} \\
&= -\mathrm{Tr}\left\{\mathbf{I}-\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\mathcal{K}\right\} + \mathrm{Tr}\left\{\mathbf{I}\right\} = \mathcal{K}\mathrm{Tr}\left\{\left(\mathcal{A}\mathcal{A}^H+\mathcal{K}\mathbf{I}\right)^{-1}\right\}
\end{aligned}$$

And finally, substituting $\mathcal{A}$ and $\mathcal{K}$ by its original values, the new cost function, that only depends on the linear transformation $\mathbf{F}_k$, is:

$$\begin{aligned}
\xi_k^{MMSE} &= \left[\tfrac{\sigma_n^2}{\sigma_x^2}+\omega\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\}\right]\mathrm{Tr}\left\{\left(\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k\mathbf{F}_k^H\hat{\mathbf{H}}_k^{eqH}+\tfrac{\sigma_n^2}{\sigma_x^2}\mathbf{I}+\omega\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{F}_k^H\right\}\mathbf{I}\right)^{-1}\right\} \\
&\qquad -\mu\left[\left(\sum_{k=1}^{K}\mathrm{Tr}\left\{\mathbf{F}_k^H\mathbf{F}_k\right\}\right)-P_0\right]
\end{aligned} \tag{4.116}$$

Next, the optimum linear transformation $\mathbf{F}_k$ will be obtained differentiating (4.116) with respect to the matrix $\mathbf{F}_k^H$ and equating the result to zero. Isolating the matrix $\mathbf{F}_k$ we arrive at:

$$\mathbf{F}_k = \mathcal{K}_2\underbrace{\hat{\mathbf{H}}_k^{eqH}\left(\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k\mathbf{F}_k^H\hat{\mathbf{H}}_k^{eqH}+\mathcal{K}_1\mathbf{I}\right)^{-2}\hat{\mathbf{H}}_k^{eq}}_{\mathbf{P}_F}\mathbf{F}_k \tag{4.117}$$

where $\mathcal{K}_1$ and $\mathcal{K}_2$ are non relevant scalars and the underbraced matrix $\mathbf{P}_F$ is the projection matrix of $\mathbf{F}_k$. Substituting matrices $\mathbf{F}_k$ and $\hat{\mathbf{H}}_k^{eq}$ by its SVD decompositions, denoted as [5] $\mathbf{F}_k = \mathbf{U}_F\mathbf{\Lambda}_F\mathbf{V}_F^H$ and $\hat{\mathbf{H}}_k^{eq} = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{V}_k^H$, the projection matrix $\mathbf{P}_F$ must satisfy $\mathbf{P}_F = \mathbf{U}_F\mathbf{U}_F^H$ and, consequently, the next equality must be accomplished:

$$\frac{1}{\mathcal{K}_2}\mathbf{U}_F\mathbf{U}_F^H = \mathbf{V}_k\mathbf{\Lambda}_k^H\mathbf{U}_k^H\left(\mathbf{U}_k\mathbf{\Lambda}_k\mathbf{V}_k^H\mathbf{U}_F\mathbf{\Lambda}_F\mathbf{\Lambda}_F^H\mathbf{U}_F^H\mathbf{V}_k\mathbf{\Lambda}_k^H\mathbf{U}_k^H+\mathcal{K}_1\mathbf{I}\right)^{-2}\mathbf{U}_k\mathbf{\Lambda}_k\mathbf{V}_k^H \tag{4.118}$$

This equality only holds if $\mathbf{V}_k^H\mathbf{U}_F = \mathbf{P}$, where $\mathbf{P}$ is any permutation matrix, and the elements in the diagonal matrix $\mathbf{\Lambda}_F$ (denoted in (4.114) as $\mathbf{\Phi}_k$) are appropriately defined. In particular, if the permutation matrix $\mathbf{P}$ is set to the identity matrix, therefore, $\mathbf{U}_F = \mathbf{V}_k$ as shown in (4.114). Note that equation (4.118) is independent of the unitary matrix $\mathbf{V}_F$ and, consequently, any unitary matrix (denoted in (4.114) as $\mathbf{T}_k$) can be chosen.

Once the optimum structure for $\mathbf{F}_k$ has been derived, the optimum structure for $\mathbf{G}_k$ can be found substituting into equation (4.115) the matrix $\mathbf{F}_k$ by its SVD decomposition for the particular case of $\mathbf{F}_k = \mathbf{V}_k\mathbf{\Phi}_k\mathbf{T}_k$. Hence, after some manipulations, it is straightforward to verify, as shown in (4.114), that the SVD decomposition of the matrix $\mathbf{G}_k$ satisfy $\mathbf{G}_k = \mathbf{T}_k^H\mathbf{\Gamma}_k\mathbf{\Lambda}_k^{\#}\mathbf{U}_k^H$ where elements in the diagonal matrix $\mathbf{\Gamma}_k$ are appropriately chosen to satisfy (4.115).

■

[5] For the sake of clarity in the notation the subindex k and the label eq have been both omitted.

Appendix 4.C Exponential expansion of the $\mathcal{Q}(\sqrt{x})$ function

The Gaussian Q-function, or equivalently, the complementary error function, are required for the computation of error probability performances in multiple applications. The solution of the problems formulated by means of this functions are often analytically complex to be solved if the classical definition of this function is used:

$$\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt \tag{4.119}$$

A large study has been done for this function in order to simplify the analysis of such problems. Craig showed a different form of the Gaussian Q-function [Cra91] that allows to average this function over some random parameters [Sim98]. Alternatively, some approximations and bounds can be found for this functions. One of the widely used approximations is the Chernoff bound:

$$\mathcal{Q}(x) = \frac{1}{2}\, e^{-\frac{x^2}{2}} \tag{4.120}$$

However, tighter bound can be found (see e.g., [Erm04] and references therein). In this chapter, we propose an expansion of the $\mathcal{Q}(\sqrt{x})$ function in the neighborhood of the point $x = a$. By a generalized exponential expansion, the $\mathcal{Q}(\sqrt{x})$ function can be expressed as:

$$\mathcal{Q}\left(\sqrt{x}\right) \simeq \delta e^{-\alpha x} \tag{4.121}$$

where constants α and δ can be chosen to set certain constraints. The derived expressions can be applied for any value of δ and α including the Chernoff bound as a particular case for $\alpha = 1/2$ and $\delta = 1/2$. However, it is possible to derive a lower bound for $\mathcal{Q}(\sqrt{x})$ that is very tight in a wide range of values around a. This bound is obtained from a Taylor expansion of $ln\left(\mathcal{Q}(\sqrt{x})\right)$ in the neighborhood of the point a:

$$\begin{aligned} \mathcal{Q}(\sqrt{x}) &= e^{ln\left(\mathcal{Q}(\sqrt{x})\right)} \\ &\simeq e^{\left[ln\left(\mathcal{Q}(\sqrt{a})\right) - \frac{e^{-|a|/2}}{2\sqrt{2\pi a}\mathcal{Q}(\sqrt{a})}(x-a)\right]} \\ &= e^{ln\left(\mathcal{Q}(\sqrt{a})\right)} e^{\frac{e^{-|a|/2}}{2\sqrt{2\pi a}\mathcal{Q}(\sqrt{a})}\, a} e^{-\frac{e^{-|a|/2}}{2\sqrt{2\pi a}\mathcal{Q}(\sqrt{a})}\, x} \\ &= \delta e^{-\alpha x} \end{aligned} \tag{4.122}$$

Where α and δ are given by:

$$\alpha = \frac{e^{-|a|/2}}{2\sqrt{2\pi a}\mathcal{Q}(\sqrt{a})} \;\; ; \;\; \delta = \mathcal{Q}\left(\sqrt{a}\right) e^{\alpha a} \tag{4.123}$$

It is straightforward to show that (4.122) provides a lower bound for the Gaussian Q-function.

Proof. As the exponential function is monotonically increasing, the lower bound is ensured if the Taylor expansion in (4.122) is always lower than $ln\left(\mathcal{Q}(\sqrt{x})\right)$, a condition that always holds since $ln\left(\mathcal{Q}(\sqrt{x})\right)$ is a convex function. ■

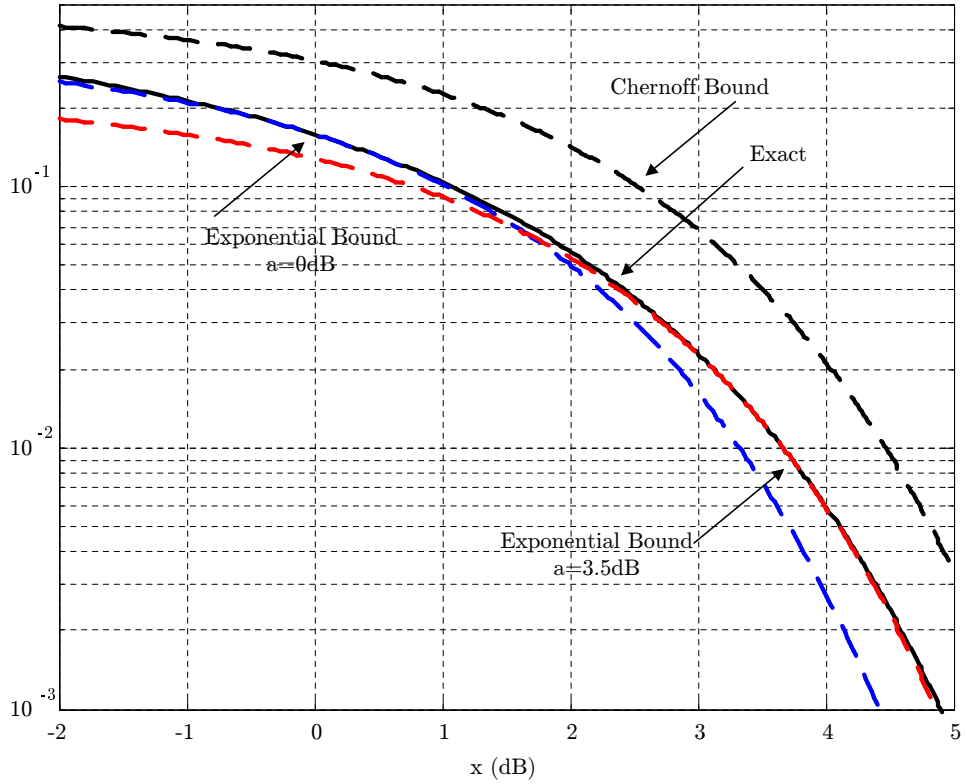


Figure 4.11: Comparison of the obtained exponential bound with the exact Gaussian Q-function and the Chernoff bound.

Figure 4.11 compares the obtained bound with the exact curves and the Chernoff bound. As it can be seen, the proposed bound is tighter than the Chernoff bound in a wide range of values in the neighborhood of the point where the approximation is derived.

Appendix 4.D Optimum decomposition of $\mathbf{F}_k$ matrix to minimize the uncoded BER for beamforming

This Appendix proves that the optimum decomposition of $\mathbf{F}_k$ to minimize the uncoded BER cost function:

$$\xi^{BER} = \mathcal{K} \sum_{k=1}^{K} \frac{\delta_k}{\left(1+\gamma_k \omega\ d^2 \mathbf{F}_k^H \mathbf{F}_k\right)^{M_R}}\ e^{-\frac{\gamma_k\ d^2 \left|\hat{\mathbf{H}}_k^{eq} \mathbf{F}_k\right|^2}{1+\gamma_k \omega\ d^2 \mathbf{F}_k^H \mathbf{F}_k}} - \mu \left[\sum_{k=1}^{K} \mathrm{Tr}\left\{\mathbf{F}_k^H \mathbf{F}_k\right\} - P_0\right] \quad (4.124)$$

for the particular case of $M = 1$ is obtained selecting $\mathbf{F}_k$ to be in the direction of the right singular vector of the equivalent channel $\hat{\mathbf{H}}_k^{eq}$ associated to the largest singular value.

Proof. The proof follows easily substituting the SVD decomposition of $\mathbf{F}_k$ and $\hat{\mathbf{H}}_k^{eq}$, denoted as

$\mathbf{F}_k = \mathbf{u}_{F_k}\phi_k$ and $\hat{\mathbf{H}}_k^{eq} = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{V}_k^H$, into the cost function (4.124) to obtain:

$$\xi^{BER} = \mathcal{K}\sum_{k=1}^{K}\frac{\delta_k}{\left(1+\gamma_k\omega\ d^2\left|\phi_k\right|^2\right)^{M_R}}\ e^{-\frac{\gamma_k\ d^2\left|\phi_k\right|^2\mathbf{u}_{F_k}^H\mathbf{V}_k\mathbf{V}_k^H\mathbf{u}_{F_k}}{1+\gamma_k\omega\ d^2\left|\phi_k\right|^2}-\mu\left[\sum_{k=1}^{K}\left|\phi_k\right|^2-P_0\right]} \quad (4.125)$$

Since all terms in the summation of the previous equation are always positive, and each unit-norm vector $\mathbf{u}_{F_k}$ contributes independently into the cost function, the optimum design of these vectors must be done minimizing each summation term separately. Hence, the minimization of the cost function is achieved selecting the vectors $\mathbf{u}_{F_k}$ that maximize the argument of the exponential term, focusing the transmitted symbols in the direction of the singular vector $\mathbf{v}_k$ associated to the maximum singular value of the channel. In consequence, it must be that $\mathbf{u}_{F_k} = \mathbf{v}_k$ and, therefore:

$$\mathbf{F}_k = \mathbf{v}_k\phi_k \quad (4.126)$$

■

Appendix 4.E Jensen's inequality to the convex function $\mathcal{Q}\left(\sqrt{x}\right)$

In this Appendix we show that for high SNR the following inequality holds

$$\frac{1}{\mathcal{N}^{2M}}\sum_{i=1}^{\mathcal{N}^M}\sum_{j=1}^{\mathcal{N}^M}B(\mathbf{x}_{ki},\mathbf{x}_{kj})\mathcal{Q}\left(\sqrt{\frac{\left|\mathbf{\Lambda_k\Phi_k T d}_{ijk}\right|^2}{2\sigma_n^2}}\right) \leq \frac{1}{\mathcal{N}^{2M}}\sum_{i=1}^{\mathcal{N}^M}\sum_{j=1}^{\mathcal{N}^M}B(\mathbf{x}_{ki},\mathbf{x}_{kj})\mathcal{Q}\left(\sqrt{\frac{\left|\mathbf{\Lambda_k\Phi_k d}_{ijk}\right|^2}{2\sigma_n^2}}\right) \quad (4.127)$$

Proof. Let s_i and s_j bet two elements of a given constellation, let $R(b,d)$ denote the number of symbol pairs (s_i, s_j) that differ in b bits and $d = |s_i - s_j|$ any of the D possible values of the difference between two constellation symbols. Let $\mathbf{i}_m$ denote the mth column of the identity matrix of size M. Under the high SNR condition (assumption A1), the average BER can be approximated taking into account only the pairwise error probability between those pairs of symbols that differ in one coefficient. Thus, for high SNR, the BER for the kth subcarrier can be written as follows:

$$\begin{aligned}\overline{BER}_k &= \frac{1}{\mathcal{N}^{2M}}\sum_{i=1}^{\mathcal{N}^M}\sum_{j=1}^{\mathcal{N}^M}B(\mathbf{x}_{ki},\mathbf{x}_{kj})\mathcal{Q}\left(\sqrt{\frac{\left|\mathbf{\Lambda_k\Phi_k T}\left(\mathbf{x}_{ki}-\mathbf{x}_{kj}\right)\right|^2}{2\sigma_n^2}}\right)\\ &\simeq \frac{1}{\mathcal{N}^{2M}}\sum_{m=1}^{M}\sum_{b=1}^{\log_2\mathcal{N}}\sum_{l=1}^{D}bR(b,d_l)\mathcal{Q}\left(\sqrt{\frac{\left|\mathbf{\Lambda_k\Phi_k T i}_m d_l\right|^2}{2\sigma_n^2}}\right)\end{aligned} \quad (4.128)$$

and reduces to the following expression for $\mathbf{T} = \mathbf{I}$:

$$\begin{aligned}\overline{BER}_k = & \frac{1}{\mathcal{N}^{2M}} \sum_{i=1}^{\mathcal{N}^M} \sum_{j=1}^{\mathcal{N}^M} B(\mathbf{x}_{ki}, \mathbf{x}_{kj}) \mathcal{Q}\left(\sqrt{\frac{\left|\mathbf{\Lambda_k \Phi_k} \left(\mathbf{x}_{ki} - \mathbf{x}_{kj}\right)\right|^2}{2\sigma_n^2}}\right) \\ \simeq & \frac{1}{\mathcal{N}^{2M}} \sum_{m=1}^{M} \sum_{b=1}^{\log_2 \mathcal{N}} \sum_{l=1}^{D} bR(b, d_l) \mathcal{Q}\left(\sqrt{\frac{\left|\mathbf{\Lambda_k \Phi_k} d_l \mathbf{i}_m\right|^2}{2\sigma_n^2}}\right)\end{aligned} \tag{4.129}$$

Comparing equations (4.128) and (4.129) it is clear that proving the inequality in equation (4.127) is equivalent to proving a similar expression for the $\mathcal{Q}(\cdot)$ function that appears in them. It can be shown [Lam97], [Din03] that the DFT matrix satisfies that $|\mathbf{\Lambda}_k \mathbf{\Phi}_k \mathbf{T} \mathbf{i}_m|$ is constant for all m. Therefore,

$$\begin{aligned}& \frac{1}{M} \sum_{m=1}^{M} \mathcal{Q}\left(\sqrt{\frac{d_l^2}{2\sigma_n^2} \left|\mathbf{\Lambda_k \Phi_k T} \mathbf{i}_m\right|^2}\right) \\ & = \mathcal{Q}\left(\sqrt{\frac{d_l^2}{2\sigma_n^2} \frac{1}{M} \sum_{m=1}^{M} \left|\mathbf{\Lambda_k \Phi_k T} \mathbf{i}_m\right|^2}\right) \\ & = \mathcal{Q}\left(\sqrt{\frac{d_l^2}{2\sigma_n^2} \frac{1}{M} \mathrm{Tr}\left\{\mathbf{\Lambda_k \Phi_k T T}^H \mathbf{\Phi}_k^H \mathbf{\Lambda}_k^H\right\}}\right) \\ & = \mathcal{Q}\left(\sqrt{\frac{d_l^2}{2\sigma_n^2} \frac{1}{M} \mathrm{Tr}\left\{\mathbf{\Lambda_k \Phi_k \Phi}_k^H \mathbf{\Lambda}_k^H\right\}}\right) \\ & = \mathcal{Q}\left(\sqrt{\frac{d_l^2}{2\sigma_n^2} \frac{1}{M} \sum_{m=1}^{M} \left|\mathbf{\Lambda_k \Phi_k} \mathbf{i}_m\right|^2}\right) \\ & \leq \frac{1}{M} \sum_{m=1}^{M} \mathcal{Q}\left(\sqrt{\frac{d_l^2}{2\sigma_n^2} \left|\mathbf{\Lambda_k \Phi_k} \mathbf{i}_m\right|^2}\right)\end{aligned} \tag{4.130}$$

where equality in second line applies because $\mathbf{T}$ is a unitary matrix, and last inequality is a result of the application of Jensen's inequality [Cov91, p.25] to the convex function $\mathcal{Q}(\sqrt{x})$. The combination of equations (4.130) and (4.128)-(4.129) leads to equation (4.127). ■

Appendix 4.F The Frost algorithm

The Frost algorithm is a constrained adaptive algorithm that has the ability to keep in each iteration the constraints over the vector of parameters to design. In its essence this algorithm is a simple LMS-type algorithm that includes a set of linear equality constraints in the optimization problem. It was first proposed by O.L. Frost for adaptive array processing [Fro72]. The minimization problem, keeping the notation of the aforementioned paper is formulated as:

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R} \mathbf{w} \qquad \text{subject to} \qquad \mathbf{w}^H \mathbf{C} = \mathcal{F}^H \tag{4.131}$$

where the full rank matrix $\mathbf{C}$ and the vector $\mathcal{F}^H$ define the set of linear constraints, $\mathbf{w}$ is the vector of weights to be designed and $\mathbf{R}$ is a correlation matrix. This optimization problem can be solved by the method of Lagrange multipliers. Hence, the function to be minimized is:

$$\xi = \mathbf{w}^H \mathbf{R} \mathbf{w} - \left(\mathbf{w}^H \mathbf{C} - \mathcal{F}^H\right) \boldsymbol{\mu} \tag{4.132}$$

where $\boldsymbol{\mu}$ is a vector that contains the Lagrange multipliers.

Taking the gradient with respect to $\mathbf{w}^H$ we get:

$$\nabla_{\mathbf{w}^H} \xi = \mathbf{R}\mathbf{w} - \mathbf{C}\boldsymbol{\mu} \tag{4.133}$$

Equaling this gradient to zero and isolating the vector $\mathbf{w}$ a closed-form solution can be found for the minimization of (4.131). However, in some optimization problems this closed-form solution cannot be derived (as is the case of our concerning problem) or requires a computational complexity that makes the solution not suitable for a practical implementation (as was the case of the original work in [Fro72]). Instead, a gradient-descendent algorithm that includes the set of linear constraints can be implemented to find the optimum solution. The main handicap of the constrained iterative algorithms is that although the solution is initialized with a vector that satisfies the constraints, they do not prevent the deviation of the solution from the constraints due to the accumulation of numerical errors (roundoff or truncation errors) in each iteration. The algorithm proposed by Frost solved this problem forcing the Lagrange multipliers to satisfy the constraints in every iteration.

Rewriting the gradient expression as:

$$\nabla_{\mathbf{w}^H} \xi = \mathbf{a}\left(\mathbf{w}\right) - \mathbf{C}\boldsymbol{\mu} \tag{4.134}$$

the iterative algorithm can be written as:

$$\begin{aligned} \mathbf{w}_{n+1} &= \mathbf{w}_n - \lambda \, \nabla_{\mathbf{w}^H} \xi \\ &= \mathbf{w}_n - \lambda \left[\mathbf{a}\left(\mathbf{w}_n\right) - \mathbf{C}\boldsymbol{\mu}\right] \end{aligned} \tag{4.135}$$

where λ is a constant that sets the step-size of the iterative algorithm.

The Lagrange multiplier must be chosen to satisfy the set of linear constraints:

$$\mathbf{w}_{n+1}^H \mathbf{C} = \mathbf{w}_n^H \mathbf{C} - \lambda \mathbf{a}\left(\mathbf{w}_n\right)^H \mathbf{C} + \lambda \boldsymbol{\mu}^H \mathbf{C}^H \mathbf{C} = \mathcal{F}^H \tag{4.136}$$

and isolating $\boldsymbol{\mu}$ the next expression is found:

$$\boldsymbol{\mu} = \left(\mathbf{C}^H \mathbf{C}\right)^{-1} \mathbf{C}^H \mathbf{a}\left(\mathbf{w}_n\right) - \frac{1}{\lambda} \left(\mathbf{C}^H \mathbf{C}\right)^{-1} \left(\mathbf{C}^H \mathbf{w}_n - \mathcal{F}\right) \tag{4.137}$$

Finally, substituting the Lagrange multiplier into the iterative equation (4.135) we get:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \lambda \mathbf{a}(\mathbf{w}_n) + \lambda \mathbf{C}\left(\mathbf{C}^H\mathbf{C}\right)^{-1}\mathbf{C}^H\mathbf{a}(\mathbf{w}_n) - \mathbf{C}\left(\mathbf{C}^H\mathbf{C}\right)^{-1}\mathbf{C}^H\mathbf{w}_n + \mathbf{C}\left(\mathbf{C}^H\mathbf{C}\right)^{-1}\mathcal{F} \quad (4.138)$$

that can be easily simplified defining the orthogonal projection matrix $\mathbf{P}_{\mathbf{C}}^{\perp} = \mathbf{I} - \mathbf{C}\left(\mathbf{C}^H\mathbf{C}\right)^{-1}\mathbf{C}^H$ as follows:

$$\mathbf{w}_{n+1} = \mathbf{P}_{\mathbf{C}}^{\perp}[\mathbf{w}_n - \lambda \mathbf{a}(\mathbf{w}_n)] + \mathbf{C}\left(\mathbf{C}^H\mathbf{C}\right)^{-1}\mathcal{F} \quad (4.139)$$

Note that in the derivation of this iterative equation it has not been assumed that the solution at time n satisfies the linear constraints (i.e., $\mathbf{w}_n^H\mathbf{C} \neq \mathcal{F}^H$) whereas it guarantees the constraints after the nth iteration (i.e.,$\mathbf{w}_{n+1}^H\mathbf{C} = \mathcal{F}^H$). This simple tricky distinguishes the Frost algorithm from the conventional LMS algorithm and allows to modify the optimum solution from one iteration to the next either due to numerical misadjustments or due to external adjustments as is the case of our concerning problem. The differences between the conventional LMS algorithm and the constrained adaptive algorithm can also be regarded analyzing the iterative equation in (4.139). The expression in the brackets corresponds to the conventional LMS solution. Premultiplying this solution by the projection matrix $\mathbf{P}_{\mathbf{C}}^{\perp}$ all the components orthogonal to the constraint plane are removed, and adding the second term, which is the minimum norm vector that satisfies the linear constraints, the solution is returned to the constraint plane. For a detailed geometrical interpretation of this solution see [Fro72].

Chapter 5

Robust power allocation algorithms based on information theoretic criteria. Minimization of the coded BER

Extending the study of Chapter 4, this chapter analyzes robust power allocation strategies related with the minimization of the coded BER. The linear transmitter designs that minimize the MSE or the uncoded BER do not guarantee good bit error performance when any channel correction code is introduced and the QoS is measured in terms of *Packet Error Rate* (PER). Instead, information theoretic criteria could be used for this purpose. A possible choice is to design algorithms based on the channel capacity or mutual information. The drawback of this criterion is that channel capacity represents a theoretical limit of reliable communications that does not take into account the transmitter and the receiver complexity.

The main contribution of this chapter is the proposal of the cut-off rate as a parameter of design whose maximization is directly related to the coded BER. The cut-off rate has been considered for a long time as the practical bound on the performance that could be achieved by feasible codes, providing an exponential upper bound on the error probability for the average over the ensemble of codes with block length N. The goal is, thus, to present a robust power allocation strategy that maximizes the cut-off rate using partial CSI at the transmitter. As in the previous chapter, a Bayesian formulation is used to design an optimal transmitter based on noisy CSI estimates. The result is an algorithm that adapts to the CSI, as other previously proposed solutions, converging to the open-loop transmission scheme for very poor CSI estimates. The maximization of the mutual information can also be considered as a possible design to optimize the coded BER, although a more complex solution is found for it.

The chapter is organized as follows. First, the system and the channel models, which are the same that those employed in the previous chapter, are briefly reviewed in Section 5.1. The main difference in the system model is given by the introduction of a channel coding scheme. Next, two power allocation designs based on the maximization of the cut-off rate and the maximization of the mutual information are presented in Section 5.2 and Section 5.3, respectively. Since this chapter is specifically focused on the first criterion, more emphasis is dedicated to it. Not only analytical results are provided, but also an illustrative interpretation that, making use of the Kullback-Leibler pseudo-distance, relates the maximization of the cut-off rate to the minimization of the pairwise error probability taking into account the CSI reliability. Section 5.4 outlines a comparative study between the different optimization strategies proposed in this chapter and in the previous chapter focusing on their impact on the error probability. Simulation results concerning the algorithms compared in that section can be found in Section 5.5. As it will be shown the cut-off rate provides a good performance in terms of coded BER for the specific channel code used in the simulations. The chapter is ended by studying some surprising results found when comparing the performance of the algorithms in terms of coded BER. Specifically, Section 5.6 justifies the excellent performance of one of the algorithms proposed in Chapter 4 to minimize the uncoded BER: the algorithm that minimizes the uncoded BER using the Chernoff bound. On the other hand, Section 5.7 analyzes the importance of the interleaver and shows that the performance of the algorithms, specifically those that aim to minimized the coded BER, can be improved if an adaptive interleaver that uses the channel knowledge at the transmitter is introduced. Finally, Section 5.8 concludes the chapter.

5.1 System and channel models

As it is shown in Figure 5.1 the same OFDM MIMO communications system considered in the previous chapter (see Section 4.2) is now assumed. Recalling that section, the input-output relationship can be written in terms of matrices that involve only one subcarrier each as:

$$\mathbf{r}_k = \mathbf{H}_k \mathbf{F}_k \mathbf{x}_k + \mathbf{n}_k \quad k = 1 \dots K \tag{5.1}$$

In this case the linear transformation at the receiver, denoted with matrix $\mathbf{G}_k$, has been omitted since this linear receiver was only introduced for the specific MMSE design.

For the design of the linear precoder in the presence of channel estimation error, the same Rayleigh propagation channel is also considered, modelling the channel estimation error $\boldsymbol{\varepsilon}$ as an additive term:

$$\hat{\mathbf{h}} = \sqrt{\frac{\sigma_h^2}{\sigma_h^2 + \sigma_\epsilon^2}} \, (\mathbf{h} + \boldsymbol{\varepsilon}) \tag{5.2}$$

and assuming that the channel $\mathbf{h}$ and its estimate $\hat{\mathbf{h}}$ are jointly Gaussian.

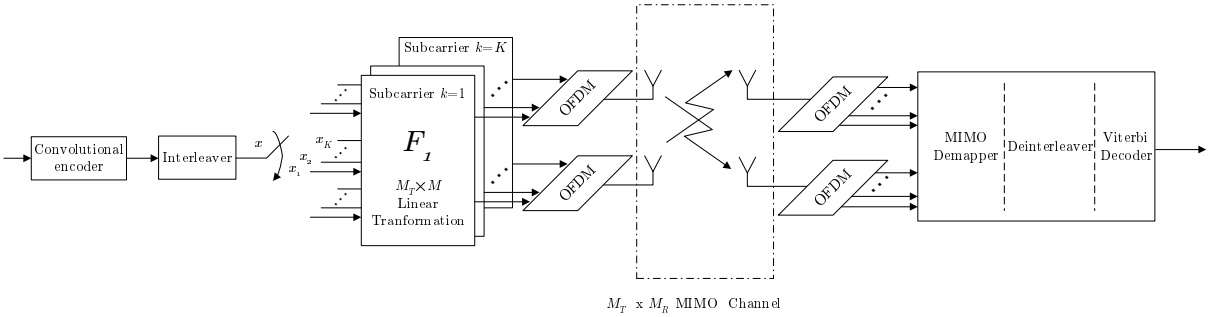


Figure 5.1: Block diagram of the proposed MIMO OFDM system with channel coding scheme

Hence, using the same Bayesian formulation proposed in the previous chapter to design an optimal transmitter based on noisy CSI estimates, the robust design will also depend on the statistics of the channel $\mathbf{h}$ given $\hat{\mathbf{h}}$. More specifically, it will depend on the conditional mean, denoted in Section 4.2.2 as the equivalent channel $\mathbf{m}_{\mathbf{h}|\hat{\mathbf{h}}} = \hat{\mathbf{h}}_k^{eq}$ (Eq.4.17), and the conditional covariance, which under the hypothesis of uncorrelated antennas it was shown to be $\mathbf{C}_{\mathbf{h}_k|\hat{\mathbf{h}}} = \omega\mathbf{I}$ (Eq.4.15).

The novelty of this chapter is the incorporation, in the system model, of a channel coding scheme. As shown in Figure 5.1 this stage comprises a convolutional encoder and an interleaver at the transmitter and a Viterbi decoder at the receiver. In a first study the specific channel code described in HIPERLAN/2 [HL201] and IEEE 802.11a [IEE99] standards has been chosen. According to these standards, the bit stream to be transmitted is first encoded with a convolutional code $k = 7$, $r = 1/2$ and next the encoded data bits are interleaved prior to be multiplexed into the different subcarriers. Although a block interleaver is first used, as it is proposed in the standard, the last part of this chapter (Section 5.7) analyzes the influence of the interleaver in the performance of the different power allocation algorithms in terms of coded BER and proposes an adaptive interleaver based on the *Rate-Compatible Punctured Convolutional Codes* (RCPC) codes, and thus named *'RCPC interleaver'*, which outperforms the classic block interleaver.

5.2 Maximum cut-off rate

This section presents the cut-off rate as an alternative to the channel capacity or mutual information criteria [Sca99b] for the design of optimal transceivers in the presence of channel coding. This parameter was first proposed by [Woz66] and [Mas74] as a sensitive criterion for the design of coded-modulations, arguing that, rather than using the uncoded BER as a modulation criterion, an appropriate parameter could be the cut-off rate. More recently, the cut-off rate has been considered in [Rai96] in the context of open-loop transmission schemes, in [Her01] for the design of space-time codes under the assumption that no CSI is available neither at the

transmitter nor at the receiver, and in [Nab03] for the design of a diagonal matrix that applies a phase shift to each of the symbols to be transmitted exploiting the knowledge of the channel statistics.

The Channel Coding Theorem [Cov91, p.198] states that for any transmission rate below the Shannon channel capacity, the word error probability can be arbitrarily low by increasing the encoder / decoder complexities. The channel capacity defines, hence, the upper bound of reliable data transmission rates. Alternatively, a useful measure of achievable rates can be obtained if the word error rate is upper bounded by an exponential bound. For any discrete memoryless channel, the word error probability $P_w(e)$ averaged over the ensemble of binary block codes of length N and rate R_c, is bounded by [Gal65] (a similar result has been proven for convolutional codes in [Vit67]):

$$E_{\mathcal{C}}\left\{P_w(e)\right\} \leq 2^{-NE(R_c)} \tag{5.3}$$

where $E_{\mathcal{C}}\left\{\cdot\right\}$ denotes the expectation over the ensemble of random codes and the exponent $E\left(R_c\right)$ is called the Reliability Function of the channel, which is a positive function that depends on both the channel and the coding rate, but it does not depend on the specific channel code. Note that an accurate study of the exponent in (5.3) could provide a clear relationship between the bound on the word error probability, the coding rate, the length of the code and the channel characteristics. However, in most scenarios, the reliability function may be difficult to be computed.

A simpler exponential upper bound can be derived if the Reliability Function is lower bounded by a tangent to the function with a negative unitary slope [Woz66]. Thus:

$$E_{\mathcal{C}}\left\{P_w(e)\right\} \leq 2^{-N(R_o-R_c)} \text{ if } R_c \leq R_o \tag{5.4}$$

where R_o is defined as the *cut-off rate* of the transmission scheme. Until the proposal of the turbo codes , this parameter set an upper bound for the rates where it was possible to operate with arbitrarily small probability, and determined also an exponent upper bound of the error probability [Big98, Cos98] (the best exponent upper-bound for rates near to R_o).

Several reasons justify the use of the cut-off rate as an alternative to the channel capacity as a criterion for the linear precoder design:

- The cut-off rate is a lower bound on the channel capacity as will be next analyzed in the simulation results section.
- The cut-off rate provides a useful measure of achievable rates when a feasible coding scheme is used, whereas the channel capacity is a theoretical limit only on the system performance. Contrary to past beliefs, the cut-off rate cannot be considered an upper bound on practical achievable rates since with the advent of the turbo codes, the LDPC codes and the iterative decoding schemes, rates grater than R_o can be achieved [Cos98]. However, the cut-off rate parameter still remains in interest when sequential decoding strategies are used.

- The cut-off rate provides an upper bound on the error probability which does not depend on the specific code. Note that if $R_c < R_o$, the error probability can be arbitrarily reduced by increasing the length of the channel code $N \to \infty$. The use of this parameter allows, hence, to compare different channels in terms of error probability while keeping the comparison independent of the channel code. Moreover, its maximization will offer a good performance in terms of PER for the average over the ensemble of codes.
- Finally, the cut-off rate will allow us to derive a closed-form solution for the robust power allocation strategy using partial CSI at the transmitter.

Next, we provide the expression of R_o for the ensemble of binary block codes. Its derivation makes use of the *Pairwise Error Probability* (PEP) and the Chernoff bound to obtain the final expression. Alternatively, the same expression can be obtained via the union Bhattacharyya bound and also as a specific case of the Gallager bound [Vit79, p.136]. The derivation assumes a general constellation $\{s_1 \dots s_{\mathcal{N}}\}$ of size $\mathcal{N}$, the MIMO OFDM structure described in (5.1), and a ML detector implemented at the receiver. Those assumptions make the derivation to be quite general, encompassing the single antenna transmission, the frequency flat fading channel and the AWGN channel as particular cases.

Let us consider a specific block encoder that maps a set of NR_c information bits to be encoded into one of the 2^{NR_c} possible codewords of length N defined by the block code. Denoting $\mathcal{X}$ as the codeword that was transmitted over the channel, and applying the Theorem on Total Probability, we have that the dependence of the word error probability $P_w(e)$ on the codewords is given by:

$$P_w(e) = \sum_{\mathcal{X}} P_w(e/\mathcal{X}) P(\mathcal{X}) \tag{5.5}$$

where $P_w(e/\mathcal{X})$ is the conditional word error probability assuming that the codeword $\mathcal{X}$ is transmitted.

Because the computation of this word error rate is of prohibitive complexity due to the large number of codewords, we will alternatively compute the average of $P_w(e)$ for all codes of a given rate and length. Hence

$$E_{\mathcal{C}}\{P_w(e)\} = E_{\mathcal{C}}\left\{\frac{1}{2^{NR_c}} \sum_{\mathcal{X}} P_w(e/\mathcal{X})\right\} \tag{5.6}$$

where under the random code hypothesis, it can be assumed that each one of the 2^{NR_c} codewords is equally likely $P(\mathcal{X}) = 2^{-NR_c}$. At the receiver the function of the decoder is to decide which one of the possible codewords was transmitted, Hence, the conditional word error rate can be substituted by the pairwise error probability of deciding the codeword $\hat{\mathcal{X}}$ when $\mathcal{X}$ was transmitted:

$$E_{\mathcal{C}}\{P_w(e)\} = E_{\mathcal{C}}\left\{\frac{1}{2^{NR_c}} \sum_{\mathcal{X}} \sum_{\hat{\mathcal{X}} \neq \mathcal{X}} P(\mathcal{X} \to \hat{\mathcal{X}})\right\} \tag{5.7}$$

Next, assuming the ML detector, an upper bound on the PEP is given by the Chernoff Bound (see Section 4.4.1) and, thus, the error probability is bounded by:

$$E_{\mathcal{C}}\left\{P_w(e)\right\} \leq E_{\mathcal{C}}\left\{\frac{1}{2^{NR_c}}\sum_{\mathcal{X}}\sum_{\hat{\mathcal{X}}\neq\mathcal{X}} e^{-\frac{1}{4\sigma_n^2}\left|\mathbf{HF}\left(\mathcal{X}-\hat{\mathcal{X}}\right)\right|^2}\right\} \tag{5.8}$$

At this point we introduce the MIMO OFDM structure described in (5.1). Assuming that the codeword $\mathcal{X}$ is transmitted through N/K OFDM symbols, assuming that the channel response can only take K different values (i.e. one per subcarrier), and applying the OFDM structure that allows to decouple the MIMO channel into K subcarriers, (5.8) can be written as:

$$\begin{aligned} E_{\mathcal{C}}\left\{P_w(e)\right\} &\leq E_{\mathcal{C}}\left\{\frac{1}{2^{NR_c}}\sum_{\mathcal{X}}\sum_{\hat{\mathcal{X}}\neq\mathcal{X}}\prod_{s=1}^{N/K}\prod_{k=1}^{K} e^{-\frac{1}{4\sigma_n^2}\left|\mathbf{H}_k\mathbf{F}_k\left(\mathcal{X}_{sk}-\hat{\mathcal{X}}_{sk}\right)\right|^2}\right\} \\ &< E_{\mathcal{C}}\left\{\frac{1}{2^{NR_c}}\sum_{\mathcal{X}}\sum_{\hat{\mathcal{X}}}\prod_{s=1}^{N/K}\prod_{k=1}^{K} e^{-\frac{1}{4\sigma_n^2}\left|\mathbf{H}_k\mathbf{F}_k\left(\mathcal{X}_{sk}-\hat{\mathcal{X}}_{sk}\right)\right|^2}\right\} \end{aligned} \tag{5.9}$$

where the first inequality has been upper-bounded by including $\hat{\mathcal{X}}=\mathcal{X}$ in the second summation, and $\mathcal{X}_{sk}$ and $\hat{\mathcal{X}}_{sk}$ are parts of the codewords $\mathcal{X}$ and $\hat{\mathcal{X}}$, respectively.

We now carry out the average over the ensemble of block codes. The random coding hypothesis sets up that:

- two different codewords $\mathcal{X}$ and $\hat{\mathcal{X}}$ are assumed to be i.i.d. random variables;
- the elements into a codeword, say $\mathcal{X}$ or $\hat{\mathcal{X}}$, are also i.i.d. random variables.

Under those hypothesis the vectors $\mathcal{X}_{sk}$ and $\hat{\mathcal{X}}_{sk}$, and also its elements, can be assumed to be i.i.d., and consequently, subindexes s and k are dummy variables in those vectors. Hence (5.9) can be rewritten as:

$$E_{\mathcal{C}}\left\{P_w(e)\right\} < 2^{NR_c}\frac{1}{\mathcal{N}^M}\frac{1}{\mathcal{N}^M}\sum_{\mathbf{x}}\sum_{\hat{\mathbf{x}}}\prod_{k=1}^{K}\left\{e^{-\frac{1}{4\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k\left(\mathbf{x}-\hat{\mathbf{x}}\right)|^2}\right\}^{N/K} \tag{5.10}$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are i.i.d vectors of length M (where M denotes the number of symbols to be transmitted per subcarrier) whose elements are also i.i.d and belong to a constellation of size $\mathcal{N}$. To simplify (5.10) we introduce the error vector $\mathbf{d}=(\mathbf{x}-\hat{\mathbf{x}})$ of length M and the expectation $E_{\mathbf{d}}\left\{\cdot\right\}=\frac{1}{\mathcal{N}^M}\frac{1}{\mathcal{N}^M}\sum_{\mathbf{x}}\sum_{\hat{\mathbf{x}}}(\cdot)$ so that:

$$E_{\mathcal{C}}\left\{P_w(e)\right\} < 2^{NR_c}\prod_{k=1}^{K}E_{\mathbf{d}}\left\{e^{-\frac{1}{4\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k\mathbf{d}|^2}\right\}^{N/K} \tag{5.11}$$

Equaling previous expression with the upper bound given in (5.4), a mathematically expression is derived for the cut-off rate in bits per channel use:

$$R_o = -\frac{1}{K}\sum_{k=1}^{K}\log_2\ E_{\mathbf{d}}\left\{e^{-\frac{1}{4\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k\mathbf{d}|^2}\right\} \tag{5.12}$$

5.2.1 Interpretation of the cost function

This section proposes a robust to CSI design for the linear precoder $\mathbf{F}_k$ that maximizes the cut-off rate (assuming that a ML detector is implemented at the receiver), and analyzes the derived cost function evidencing that the cut-off rate is related with the error probability. When perfect CSI is available at the transmitter, the expression in (5.12), subject to an average power constraint, is the function to be maximized. However, when the channel response is not perfectly known at the transmitter, the performance of the solution obtained with (5.12) will be degraded due to the CSI errors. Alternatively, a solution robust to the CSI errors should be used. The robustness is introduced by averaging the cost function over the channel uncertainty using a Bayesian formulation $E_{\mathbf{h}_k|\hat{\mathbf{h}}}\{\cdot\}$. Hence, the function to be maximized, given a channel estimate, and subject to an average power constraint, becomes:

$$\begin{aligned} \max_{\mathbf{F}_k} \quad & E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{-\frac{1}{K}\sum_{k=1}^{K}\log_2\ E_{\mathbf{d}}\left\{e^{-\frac{|\mathbf{H}_k\mathbf{F}_k\mathbf{d}|^2}{4\sigma_n^2}}\right\}\right\} \\ \text{subject to} \quad & \sum_{k=1}^{K}\mathrm{Tr}\left\{\mathbf{F}_k^H\mathbf{F}_k\right\} = P_0 \end{aligned} \tag{5.13}$$

A detailed analysis of the objective function in (5.13) gives an illustrative interpretation of the maximization of this cost function evidencing that its maximization is related to the minimization of the pairwise error probability taking into account the CSI reliability. Denoting $f_0(\mathbf{F}_k)$ as the function to be maximized in (5.13), the expectation over the cost function can be rewritten as:

$$f_0(\mathbf{F}_k) = -\frac{1}{K}\sum_{k=1}^{K}\int_{\mathbf{h}_k\in\mathbb{C}} f_{\mathbf{h}_k/\hat{\mathbf{h}}}(\mathbf{h}_k)\log_2 Q(\mathbf{h}_k,\mathbf{F}_k)\,d\mathbf{h}_k \tag{5.14}$$

where $f_{\mathbf{h}_k/\hat{\mathbf{h}}}(\mathbf{h}_k)$ is the conditional probability density function of the equivalent channel given the channel estimate, and $Q(\mathbf{h}_k,\mathbf{F}_k)$ is defined as:

$$Q(\mathbf{h}_k,\mathbf{F}_k) = E_{\mathbf{d}}\left\{e^{-\frac{|\mathbf{H}_k\mathbf{F}_k\mathbf{d}|^2}{4\sigma_n^2}}\right\} \tag{5.15}$$

After some straightforward manipulations, the cost function to be maximized, subject to a power constraint, becomes:

$$f_0\left(\mathbf{F}_k\right) = \frac{1}{K}\sum_{k=1}^{K}\int\limits_{\mathbf{h}_k\in\mathbb{C}} f_{\mathbf{h}_k/\hat{\mathbf{h}}}(\mathbf{h}_k)\log_2\frac{f_{\mathbf{h}_k/\hat{\mathbf{h}}}(\mathbf{h}_k)}{Q\left(\mathbf{h}_k,\mathbf{F}_k\right)} - \int\limits_{\mathbf{h}_k\in\mathbb{C}} f_{\mathbf{h}_k/\hat{\mathbf{h}}}(\mathbf{h}_k)\log_2 f_{\mathbf{h}_k/\hat{\mathbf{h}}}(\mathbf{h}_k) \quad (5.16)$$

Notice that the second term can be regarded as the differential entropy of the prior information about the channel $\mathbf{h}_k$ given a channel estimate $\hat{\mathbf{h}}$. This term is irrelevant since it does not contain information of the parameter to design (i.e., matrix $\mathbf{F}_k$). Consequently, the analysis focuses on the first term. The structure of this first term corresponds to the *Kullback-Leibler pseudo-Distance* (KLD) between two probability density functions [Cov91, p.231]. Strictly speaking, we cannot say that this term is a Kullback-Leibler pseudo-distance because a KLD is defined over probability density functions, and $Q\left(\mathbf{h}_k,\mathbf{F}_k\right)$ is not a p.d.f. Regarding, however, this term as the KLD between $f_{\mathbf{h}_k/\hat{\mathbf{h}}}(\mathbf{h}_k)$ and $Q\left(\mathbf{h}_k,\mathbf{F}_k\right)$ functions, its contribution into the cost function in (5.14) could be understood as a measure of the discrepancy between this two functions. Consequently, because the aim is to maximize the cut-off rate, and thus maximize this pseudo-distance, the function $Q\left(\mathbf{h}_k,\mathbf{F}_k\right)$ must be as dissimilar as possible to the channel distribution $f_{\mathbf{h}_k/\hat{\mathbf{h}}}(\mathbf{h}_k)$. As a result, for the most likely channel values (i.e., when $f_{\mathbf{h}_k/\hat{\mathbf{h}}}(\mathbf{h}_k)$ is maximum) the function $Q\left(\mathbf{h}_k,\mathbf{F}_k\right)$, which is related with the pairwise error probability, must be minimized (i.e., the error probability is minimized).

5.2.2 Quadratic approximation of the cost function

It can be proven that the objective function in (5.13) is concave for the specific case of beamforming when CSI is perfectly known at the transmitter. However, it cannot be proven to be concave in a most general case. Therefore, since the concavity of the maximization problem cannot be ensured, an optimization technique for non-concave (non-convex) functions should be used to achieve the optimum solution (e.g., the simulated annealing technique).

We will, however, discard this techniques proposing, as alternative, the maximization of a simple quadratic approximation of the cost function based on the second order Taylor approximation of the cut-off rate (5.20). As it will be shown in the numerical results, this simplified procedure provides a solution that, though it does not necessarily achieves the optimum solution, exhibits good performance in terms of coded BER.

To derive the approximation of the cost function (5.13), we first substitute in (5.13) the expectation over the error vector $\mathbf{d}$ as the summation over all possible values $E_{\mathbf{d}}\left\{\cdot\right\} = \frac{1}{\mathcal{N}^{2M}}\sum_{i=1}^{\mathcal{N}^{2M}}\left(\cdot\right)$ subject to the power constraint:

$$\max_{\mathbf{F}_k} \quad E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{-\frac{1}{K}\sum_{k=1}^{K}\log_2\ \frac{1}{\mathcal{N}^{2M}}\sum_{i=1}^{\mathcal{N}^{2M}} e^{-\frac{\left|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i\right|^2}{4\sigma_n^2}}\right\} \quad (5.17)$$

Next, focusing on the contribution of each subcarrier into the objective function, i.e.:

$$\xi_k = \log_2 \sum_{i=1}^{\mathcal{N}^{2M}} e^{-\frac{|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i|^2}{4\sigma_n^2}}$$

the log-sum-exp term can be approximated by a Taylor expansion in the neighborhood of any arbitrary point a_i as follows:

$$\begin{aligned}\xi_k &= \log_2 \sum_{i=1}^{\mathcal{N}^{2M}} e^{-\theta_i}\Big|_{\theta_i=\frac{|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i|^2}{4\sigma_n^2}} \\ &\simeq C_0 + \sum_i C_{1i}\left(\theta_i - a_i\right) + \frac{1}{2}\sum_i C_{2i}\left(\theta_i - a_i\right)^2 + \frac{1}{2}\sum_{\substack{i,j\\ i\neq j}} C_{3ij}\left(\theta_i - a_i\right)\left(\theta_j - a_j\right)\end{aligned} \tag{5.18}$$

where C_0, C_{1i}, C_{2i}, C_{3ij} are the Taylor expansion coefficients given by:

$$\begin{aligned} C_0 &= f(\mathbf{a}) &&= \log_2 \textstyle\sum_i e^{-a_i} \\ C_{1i} &= \frac{\partial f(\boldsymbol{\theta})}{\partial\theta_i}\Big|_{\theta_i=a_i} &&= -\frac{e^{-a_i}}{\sum_i e^{-a_i}} \\ C_{2i} &= \frac{\partial^2 f(\boldsymbol{\theta})}{\partial\theta_i^2}\Big|_{\theta_i=a_i} &&= \frac{e^{-a_i}}{\sum_i e^{-a_i}} - \frac{e^{-2a_i}}{\left(\sum_i e^{-a_i}\right)^2} \\ C_{3ij} &= \frac{\partial^2 f(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\Big|_{\substack{\theta_i=a_i\\ \theta_j=a_j}} &&= -\frac{e^{-(a_i+a_j)}}{\left(\sum_i e^{-a_i}\right)^2} \end{aligned} \tag{5.19}$$

Finally, substituting the Taylor expansion into the objective function (5.17) a quadratic approximation in the neighborhood of any arbitrary point a_i is obtained as:

$$\begin{aligned}\max_{\mathbf{F}_k} \quad & E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{-\frac{1}{K}\sum_{k=1}^{K}\left(D_0 + \sum_{i=1}^{\mathcal{N}^{2M}}\frac{D_{1i}}{4\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i|^2 + \frac{D_{2i}}{\left(4\sigma_n^2\right)^2}\left(|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i|^2\right)^2 + \right.\right. \\ & \left.\left. \sum_{i=1}^{\mathcal{N}^{2M}}\sum_{j=1}^{\mathcal{N}^{2M}}\frac{D_{3ij}}{\left(4\sigma_n^2\right)^2}\left(|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i|^2\right)\left(|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_j|^2\right)\right)\right\} \\ \text{subject to} \quad & \sum_{k=1}^{K}\mathrm{Tr}\left\{\mathbf{F}_k^H\mathbf{F}_k\right\} = P_0\end{aligned} \tag{5.20}$$

where D_0, D_{1i}, D_{2i}, D_{3ij}, which are related with the Taylor expansion coefficients, are defined as follows:

$$\begin{aligned} D_i &= \frac{e^{-a_i}}{\sum_i e^{-a_i}} \\ D_{1i} &= -D_i\left[1 + a_i - \textstyle\sum_l D_l a_l\right] \\ D_{2i} &= \frac{1}{2}D_i \\ D_{3ij} &= -\frac{1}{2}D_iD_j \\ D_0 &= -2M\log_2\mathcal{N} + \log_2\textstyle\sum_l e^{-a_l} + \sum_l D_l a_l + \sum_l \frac{D_l}{2}a_l^2 - \sum_{l,n}\frac{D_lD_n}{2}a_la_n \end{aligned} \tag{5.21}$$

This approximation allows to compute the expectation over the channel estimation error in order to derive the final cost function (see Appendix 5.A).

5.2.3 Power allocation design

This section analyzes the design of the linear transformation at the transmitter $\mathbf{F}_k$. The maximization of the cost function in (5.13) is obtained assuming the following structure, which is identical to that one proposed in the previous chapter for the MMSE (see Section 4.3.2) and the minimum uncoded BER (see Section 4.4.3) designs:

$$\mathbf{F}_k = \mathbf{V}_k \mathbf{\Phi}_k \mathbf{T}_k \tag{5.22}$$

where $\mathbf{V}_k$ contains the right singular vectors of the equivalent channel $\hat{\mathbf{H}}_k^{eq}$, $\mathbf{\Phi}_k$ is a diagonal matrix and $\mathbf{T}_k$ is a $M \times M$ unitary matrix. The aim of $\mathbf{V}_k$ is to optimally match the power into the best channel modes, leading to the decomposition of the channel into a set of M parallel subchannels. The diagonal matrices $\mathbf{\Phi}_k$ distribute the total power among all subcarriers and antennas. Finally the unitary matrix $\mathbf{T}_k$ recovers the loss of space diversity caused by the decomposition of the MIMO channel into a set of parallel multiplicative subchannels by the use of matrices $\mathbf{V}_k$ and $\mathbf{\Phi}_k$. This structure is optimal for the specific case of beamforming when the channel is perfectly known at the transmitter (see Appendix 4.D in the previous chapter). Otherwise, a suboptimal solution has been developed based on the selection of the same unitary matrix for all subcarriers. As in the previous chapter $\mathbf{T}_k = \mathbf{T}$ matrices has been set to the DFT matrix.

Once the structure of $\mathbf{F}_k$ is forced to be (5.22), the design of the linear transformation at the transmitter reduces to the design of matrix $\mathbf{\Phi}_k$. In accordance, substituting previous structure into the cost function derived from the quadratic approximation (5.20), a closed-form solution based on a set of linear equations is obtained. Appendix 5.A derives the maximization problem that can be written in a compressed expression as follows (where irrelevant constants have been omitted):

$$\begin{aligned} \max_{\boldsymbol{\phi}_k} \quad & -\frac{1}{K}\sum_{k=1}^{K}\left(\mathbf{v}_k^T\boldsymbol{\phi}_k^2 + \boldsymbol{\phi}_k^{2T}\mathbf{A}_k\boldsymbol{\phi}_k^2\right) \\ \text{subject to} \quad & \sum_{k=1}^{K}\mathbf{1}^T\boldsymbol{\phi}_k^2 - P_0 = 0 \\ & \boldsymbol{\phi}_k^2 \succeq \mathbf{0} \qquad k = 1 \ldots K \end{aligned} \tag{5.23}$$

where vector $\mathbf{v}_k$ and matrix $\mathbf{A}_k$ are both derived in Appendix 5.A, and $\boldsymbol{\phi}_k^2$ is a vector that stores the diagonal elements of matrix $\mathbf{\Phi}_k\mathbf{\Phi}_k^H$.

To find the solution in (5.23), we define the Lagrangian associated with the problem as:

$$L(\boldsymbol{\phi}_k, \mu, \boldsymbol{\nu}) = -\frac{1}{K}\sum_{k=1}^{K}\Big(\left(\mathbf{v}_k^T\boldsymbol{\phi}_k^2 + \boldsymbol{\phi}_k^{2T}\mathbf{A}_k\boldsymbol{\phi}_k^2\right) - \mu\left(\mathbf{1}^T\boldsymbol{\phi}_k^2 - P_0\right) - \boldsymbol{\nu}\boldsymbol{\phi}_k^2\Big) \tag{5.24}$$

where μ is a scalar that denotes the Lagrangian multiplier for the equality constraint and $\boldsymbol{\nu}$ is a vector that contains the multipliers for the inequality constraints. This is a convex optimization problem and hence, by imposing the *Karush-Kuhn-Tucker* (KKT) conditions [Boy04, p.243], the optimization problem can be analytically solved. Eliminating the slack variables $\boldsymbol{\nu}$, which force all the elements in vector $\boldsymbol{\phi}_k^2$ to be positive, the gradient of the Lagrangian with respect to $\boldsymbol{\phi}_k^2$ must vanish. In consequence, for each subcarrier a set of M equations must be solved to derive the optimal solution for $\boldsymbol{\phi}_k^2$:

$$\boldsymbol{\phi}_k^2 = -\left(\mathbf{A}_k + \mathbf{A}_k^T\right)^{\#}\mathbf{v}_k + \mu\left(\mathbf{A}_k + \mathbf{A}_k^T\right)^{\#}\mathbf{1} \quad k = 1, \ldots, K \tag{5.25}$$

where μ is determined forcing the power constraint as:

$$\mu = \frac{P_o + \sum_{k=1}^{K}\mathbf{1}^T\left(\mathbf{A}_k + \mathbf{A}_k^T\right)^{\#}\mathbf{v}_k}{\sum_{k=1}^{K}\mathbf{1}^T\left(\mathbf{A}_k + \mathbf{A}_k^T\right)^{\#}\mathbf{1}} \tag{5.26}$$

The power allocated to each subcarrier in vector $\boldsymbol{\phi}_k^2$ should satisfy the inequality constraint $\boldsymbol{\phi}_k^2 \succeq \mathbf{0}$. If any of the components obtained when solving (5.25) were negative, the power allocated to that subcarrier should be set to zero $\boldsymbol{\phi}_k^2(l) = 0$ and the set of equation should be solved again for the rest of subcarriers, until all the elements in vector $\boldsymbol{\phi}_k^2$ satisfy $\boldsymbol{\phi}_k(l)^2 \geq 0$.

Note that, in general, the maximum of the cost function (5.13) differs from the maximum of the quadratic approximation (5.20). Hence, in order to achieve the solution that maximizes (5.13) from the derived closed-form expression, the next procedure must be followed. First the quadratic approximation (5.20) is derived in the neighborhood of an arbitrary point and (5.25) is solved in order to obtain its maximum. The given solution (that does not achieve the maximum cut-off rate) is next used to obtain a new approximation of the cost function (5.13) in the neighborhood of the given solution. Then (5.25) is again solved to obtain a solution that it is close to the maximum of (5.13). This procedure is repeated until the maximum of the quadratic approximation coincides with the maximum of the cost function (5.13). Simulation results have shown that in most cases two or three iterations are enough to achieve the solution that maximizes the cut-off rate.[1]

[1] In the specific case of beamforming, when channel is perfectly known at the transmitter, this procedure can be interpreted as a maximization algorithm based on the Newton's method with equality constraints [Boy04, p.526].

5.3 Mutual information

Although this chapter emphasizes the maximization of the cut-off rate as an algorithm that guarantees good performances in terms of coded BER, other information theoretic criteria could also be used. This section studies, for that purpose, the channel capacity and the mutual information constrained to a specific constellation.

Next we derive a lower bound of the error probability that proves that the maximization of mutual information between the transmitter and the receiver could be considered as a possible design to optimize of the coded BER. Let us consider the following sequence of assumptions:

1. We have a message to be transmitted, $\mathcal{W}$, compound by a sequence of NR_c information bits chosen according to a uniform distribution so that: $P(\mathcal{W}) = 2^{-NRc}$.
2. We have an arbitrary block encoder with rate R_c that maps a specific message $\mathcal{W}$ into a codeword $\mathcal{X}$.
3. We have a memoryless channel through which the codeword $\mathcal{X}$ is received as random sequence $\mathcal{Y}$.
4. We define the word error probability as $P_w(e) = P(\hat{\mathcal{W}} \neq \mathcal{W})$.

From the Fano's inequality [Cov91, p.205], which bounds the entropy $H(\mathcal{W} \mid \mathcal{Y})$:

$$H(\mathcal{W} \mid \mathcal{Y}) \leq 1 + P_w(e)\ N\ R_c, \tag{5.27}$$

and after some manipulations on the conditional entropy, we get that the mutual information is related to a lower bound on the error probability as follows [Cov91, p.206]:

$$P_w(e) \geq \frac{N\ R_c - 1 - \mathcal{I}\left(\mathcal{X};\mathcal{Y}\right)}{N\ R_c} \tag{5.28}$$

Clearly, to improve the lower bound on the error probability, we should maximize the mutual information $\mathcal{I}\left(\mathcal{X};\mathcal{Y}\right)$ between the channel input and the corresponding received data. This statement provides a possible strategy to design linear precoders that is related with the optimization of the coded BER.

Next, the maximization of the mutual information constrained to a specific constellation and the maximization of the channel capacity are studied. The first criterion will be preferred since it takes into account the constellation structure and it is sensitive to the unitary matrix $\mathbf{T}_k$ (see the structure forced to the linear transformation $\mathbf{F}_k$ in Section 5.2.3), which plays an important role in the coded BER as it was shown in the previous chapter (Section 4.6). Unfortunately, to the best of the author's knowledge, a robust power allocation algorithm was not able to be derived neither for the capacity nor for the mutual information constrained to a constellation. Hence,

for the capacity, we present the closed-form solution derived in the literature when prefect CSI is available [Sca99a], [Sca00], whereas for the mutual information we propose a complex numerical solution which is only feasible in simple scenarios (e.g., for beamforming configurations).

5.3.1 Mutual information constrained to a constellation

The mutual information between the transmitted and the received data $\mathcal{I}(\mathbf{F}_k\mathbf{x}_k;\mathbf{r}_k)$ can be understood as the reduction of uncertainty of the transmitted information $\mathbf{F}_k\mathbf{x}_k$ due to the knowledge of the received data $\mathbf{r}_k$, and is given by [Cov91, p.231]:

$$\mathcal{I}(\mathbf{F}_k\mathbf{x}_k;\mathbf{r}_k) = \mathcal{H}(\mathbf{r}_k) - \mathcal{H}(\mathbf{r}_k|\mathbf{F}_k\mathbf{x}_k) \tag{5.29}$$

where $\mathcal{H}(\cdot)$ is the entropy function. While $\mathcal{H}(\mathbf{r}_k|\mathbf{F}_k\mathbf{x}_k)$ corresponds to the entropy of a Gaussian random variable: $\mathcal{H}(\mathbf{r}_k|\mathbf{F}_k\mathbf{x}_k) = \frac{M_R}{\eta}\log_2(\eta\pi\sigma_n^2 e)$, the entropy $\mathcal{H}(\mathbf{r}_k) = -E\{\log_2 p(\mathbf{r}_k)\}$ is more difficult to be computed and, generally, it does not leads to a closed-form solution. In our particular case, because the AWGN channel model, the mutual information in (5.29) can be written as follows [Ung82, Hoc03]:

$$\mathcal{I}(\mathbf{F}_k\mathbf{x}_k;\mathbf{r}_k) = M\log_2\mathcal{N} - \frac{M_R}{\eta}\log_2 e - E\left\{\log_2\sum_{j=1}^{\mathcal{N}^M} e^{-\frac{1}{\eta\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k(\mathbf{x}_{ki}-\mathbf{x}_{kj})+\mathbf{n}_k|^2}\right\} \tag{5.30}$$

where $\eta = 2$ for real constellations (e.g., BPSK or M-ASK) and $\eta = 1$ for complex constellations (e.g., QPSK, 8-PSK, M-QAM). Using a Gaussian random number generator, the mutual information can be evaluated by Monte-Carlo simulations, running the expectation over the additive Gaussian noise $\mathbf{n}_k$ and the discrete symbols $\mathbf{x}_{ki}$.

Cost function

Forcing the structure in (5.22), which has been proven to be optimal for the specific case of beamforming when the channel is perfectly known at the transmitter, the maximization problem, eliminating some irrelevant constants, can be written as follows:

$$\begin{aligned}
\max_{\boldsymbol{\phi}_k} \quad & -\frac{1}{K}\sum_{k=1}^{K} E\left\{\log_2\sum_{j=1}^{\mathcal{N}^M} e^{-\frac{1}{\eta\sigma_n^2}|\mathbf{H}_k\mathbf{V}_k\mathbf{D}_{ijk}\boldsymbol{\phi}_k+\mathbf{n}_k|^2}\right\} \\
\text{subject to} \quad & \sum_{k=1}^{K}\mathbf{1}^T\boldsymbol{\phi}_k^2 - P_0 = 0 \\
& \boldsymbol{\phi}_k^2 \succeq \mathbf{0} \qquad k = 1\ldots K
\end{aligned} \tag{5.31}$$

where $\boldsymbol{\phi}_k$ is a vector that stores the diagonal terms of the matrix $\boldsymbol{\Phi}_\mathbf{k}$, $\boldsymbol{\phi}_k^2$ stores the diagonal terms of $\boldsymbol{\Phi}_\mathbf{k}{\boldsymbol{\Phi}_\mathbf{k}}^H$, and $\mathbf{D}_{ijk}$ is a diagonal matrix defined as $\mathbf{D}_{ijk} = \mathrm{diag}\{\mathbf{T}_k(\mathbf{x}_{ki}-\mathbf{x}_{kj})\}$.

Iterative algorithm

The expectation over the noise is a handicap in the derivation of the closed-form solution for (5.31). Alternatively, an iterative search of the solution is next suggested.
Let $L(\boldsymbol{\phi}_k, \mu)$ be the Lagrangian associated to the optimization problem (5.31), which is defined as:

$$L(\boldsymbol{\phi}_k, \mu) = -\frac{1}{K}\sum_{k=1}^{K} E\left\{ \log_2 \sum_{j=1}^{\mathcal{N}^M} e^{-\frac{1}{\eta\sigma_n^2}|\mathbf{H}_k\mathbf{V}_k\mathbf{D}_{ijk}\boldsymbol{\phi}_k + \mathbf{n}_k|^2} \right\} - \mu\left[\sum_{k=1}^{K}\boldsymbol{\phi}_k^H\boldsymbol{\phi}_k - P_0\right] \quad (5.32)$$

where μ is the Lagrangian multiplier. The gradient of $L(\boldsymbol{\phi}_k, \mu)$ with respect to the vector $\boldsymbol{\phi}_k^H$ is given by:

$$\begin{aligned} \nabla_{\boldsymbol{\phi}_k^H} L(\boldsymbol{\phi}_k, \mu) &= E\left\{ \frac{\sum_{j=1}^{\mathcal{N}^M} e^{-\frac{|\mathbf{H}_k\mathbf{V}_k\mathbf{D}_{ijk}\boldsymbol{\phi}_k + \mathbf{n}_k|^2}{\eta\sigma_n^2}} 2Re\left\{\mathbf{D}_{ijk}^H\boldsymbol{\Lambda}_{\boldsymbol{k}2}\mathbf{D}_{ijk}\boldsymbol{\phi}_k + \mathbf{D}_{ijk}^H\mathbf{V}_k^H\mathbf{H}_k^H\mathbf{n}_k\right\}}{\sigma_n^2 K\log 2\sum_{j=1}^{\mathcal{N}^M} e^{-\frac{|\mathbf{H}_k\mathbf{V}_k\mathbf{D}_{ijk}\boldsymbol{\phi}_k + \mathbf{n}_k|^2}{\eta\sigma_n^2}}} \right\} - \mu\boldsymbol{\phi}_k \\ &= \mathbf{a}(\boldsymbol{\phi}_k) - \mu\boldsymbol{\phi}_k \end{aligned} \quad (5.33)$$

Defining $\boldsymbol{\phi} = \left[\begin{array}{ccc} \boldsymbol{\phi}_1{}^T & \ldots & \boldsymbol{\phi}_K{}^T \end{array}\right]^T$ as the vector that stacks all the power allocation parameters, and vector $\mathbf{A}(\boldsymbol{\phi}) = \left[\begin{array}{ccc} \mathbf{a}(\boldsymbol{\phi}_1)^T & \ldots & \mathbf{a}(\boldsymbol{\phi}_K)^T \end{array}\right]^T$ an iterative algorithm to derive the final solution can be written as [2]:

$$\begin{aligned} \boldsymbol{\phi}_{(n+1)} &= \boldsymbol{\phi}_{(n)} - \lambda\nabla_{\boldsymbol{\phi}^H} L(\boldsymbol{\phi}, \mu) \\ &= \boldsymbol{\phi}_{(n)} - \lambda\Big(\mathbf{A}(\boldsymbol{\phi}_{(n)}) - \mu_{(n)}\boldsymbol{\phi}_{(n)}\Big) \end{aligned} \quad (5.34)$$

where λ is a constant that sets the step-size of the iterative algorithm. The Lagrange multiplier at each iteration must be chosen to satisfy the average power constraint $\boldsymbol{\phi}_{(n+1)}^H\boldsymbol{\phi}_{(n+1)} = P_0$ and therefore, $\mu_{(n)}$ is given by the solution of the next second order equation:

$$P_0 = \left(\boldsymbol{\phi}_{(n)} - \lambda\Big(\mathbf{A}(\boldsymbol{\phi}_{(n)}) - \mu_{(n)}\boldsymbol{\phi}_{(n)}\Big)\right)^H \left(\boldsymbol{\phi}_{(n)} - \lambda\Big(\mathbf{A}(\boldsymbol{\phi}_{(n)}) - \mu_{(n)}\boldsymbol{\phi}_{(n)}\Big)\right) \quad (5.35)$$

Both, the expectation over the noise $\mathbf{n}_k$ and the symbols $\mathbf{x}_k$ are computed numerically using the Monte-Carlo method. Hence, this computation is only feasible when the dimensions of these vectors are not too large.

[2] Do not confuse the vector $\boldsymbol{\phi}_{(n+1)}$ that refers to the nth iteration of the vector $\boldsymbol{\phi}$ with the vector $\boldsymbol{\phi}_k$ whose entries contain the square root of the power allocation coefficients for the kth subcarrier

5.3.2 Capacity

The mutual information in (5.29) is maximized when the transmitted symbols are assumed to be Gaussian rather than constrained to a specific constellation. In that case, the mutual information, denoted as C is given by [Tel99]:

$$C = \frac{1}{K}\sum_{k=1}^{K}\log_2\left(\det\left(\mathbf{I} + \frac{\sigma_x^2}{\sigma_n^2}\mathbf{F}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{F}_k\right)\right) \tag{5.36}$$

When perfect CSI is available at the transmitter, the optimal linear transmitter that maximizes (5.36), subject to an average power constraint, has been widely studied (see, e.g., [Sca99a], [Sca00] for the specific case of SISO configuration, and [Tel99] for the multi-antenna case) but its generalization to the case with imperfect CSI is not straightforward.

Using the structure given in (5.22) for the linear transformation $\mathbf{F}_k$, the optimum solution for 5.36 is given by:

$$\phi_k(l) = \left[\frac{Po/\sigma_n^2 + \sum_{k=1}^{K}\sum_{l=1}^{M}|\lambda_k(l)|^{-2}}{KM} - \frac{1}{|\lambda_k(l)|^2}\right]^+ \tag{5.37}$$

where $[x]^+ = \max\{0, x\}$. Note that the optimum solution is not unique since any unitary matrix $\mathbf{T}_k$ can be used due the insensitivity of the capacity to this matrix (for the sake of simplicity $\mathbf{T}_k = \mathbf{I}$ is generally chosen).

The independence of the capacity with the matrix $\mathbf{T}_k$, which was shown to be relevant to the performance in terms of coded BER, calls into question the validity of the capacity as a power allocation criterion when the assumption of Gaussian symbols does not hold. As argued in [Sca00] the assumption of Gaussian symbols does not provides good performances in terms of BER when a specific constellation is fixed since the power is not used efficiently. Alternatively, in order to take the maximum benefit to this algorithm, different constellation should be used on each subcarrier depending on the goodness of the channel.

5.4 Comparative study

Before presenting some numerical results that evaluate the power allocation designs studied in this thesis, an accurate analysis of the different cost functions will allow to compare the optimization strategies and their impact on the error probability. The comparison includes not only the algorithms based on information theoretic criteria, but also those studied in the previous chapter. Hence, the objective functions to be compared are those that maximize the cut-off rate,

minimize the uncoded BER, minimize the MSE and maximize the mutual information. For the sake of simplicity in the comparison, all the objective functions are rewritten in Table 5.1.

Possibly the most remarkable difference between the cost functions is the different treatment done to the modulation. While the cut-off rate, the mutual information and the uncoded BER are optimized for a specific constellation, the MSE design is insensitive to the modulation and the channel capacity is maximized assuming that the transmitted symbols follow a Gaussian distribution. Next, the different designs are compared taking as a reference the design that maximizes the cut-off rate.

Table 5.1: Summary of the objective functions.

Maximum cut-off rate (c.f. (5.12)*)*

$$\max_{\mathbf{F}_k} \; -\frac{1}{K}\sum_{k=1}^{K}\log_2 \; E_{\mathbf{d}}\left\{ e^{-\frac{1}{4\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k\mathbf{d}|^2} \right\}$$

Minimum uncoded BER (c.f. (4.62)*)*

$$\min_{\mathbf{F}_k} \quad \frac{1}{K}\sum_{k=1}^{K} E_{\mathbf{d}}\left\{ B(\mathbf{x}_k,\hat{\mathbf{x}}_k)\;\delta_k\; e^{-\gamma_k|\mathbf{H}_k\mathbf{F}_k\mathbf{d}|^2} \right\}$$

Minimum MSE (c.f. (4.34)*)*

$$\min_{\mathbf{F}_k} \quad \sum_{k=1}^{K}\sigma_x^2 \mathrm{Tr}\left\{ \left(\mathbf{G}_k\mathbf{H}_k\mathbf{F}_k-\mathbf{I}\right)^H\left(\mathbf{G}_k\mathbf{H}_k\mathbf{F}_k-\mathbf{I}\right)\right\} + \;\sigma_n^2\;\mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\}$$

Maximum Mutual Information (c.f. (5.31)*)*

$$\max_{\mathbf{F}_k} \; -\frac{1}{K}\sum_{k=1}^{K} E\left\{ \log_2 \sum_{j=1}^{\mathcal{N}^M} e^{-\frac{1}{\eta\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k\left(\mathbf{x}_{ki}-\mathbf{x}_{kj}\right)+\mathbf{n}_k|^2} \right\}$$

Maximum Capacity (c.f. (5.36)*)*

$$\max_{\mathbf{F}_k} \quad \frac{1}{K}\sum_{k=1}^{K}\log_2\left(\det\left(\mathbf{I}+\frac{\sigma_x^2}{\sigma_n^2}\mathbf{F}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{F}_k\right)\right)$$

5.4.1 Minimum uncoded BER vs maximum cut-off rate

The main difference between the maximization of the cut-off rate and the minimization of the uncoded BER is how the different subcarriers are combined. Notice that the structure is quite similar in both cost functions except for the treatment given to the subindex k. While the cost function that minimizes the uncoded BER is given by the summation over all the subcarriers of the independent pairwise error probabilities (i.e., the arithmetic mean of the PEP), the cost function that maximizes the cut-off rate can be formulated by means of the product over all the subcarriers of the function that is related with the PEP (i.e., the geometric mean of the PEP). An easy understanding of how the power resources are allocated in both algorithms can be done keeping in mind that in the arithmetic mean the most relevant values are the largest ones, whereas that in a geometric mean the smallest terms are the most relevant. In accordance, the maximization of the cut-off rate will take care of the best subcarriers (i.e., minimize PEP for the subcarriers with low error probabilities), taking into account that the correcting capability of the channel code will allow to correct some systematic errors in the deepest faded subcarriers. On the contrary, the minimization of the uncoded BER must compensate the faded subcarriers allocating more power in them (i.e., minimize PEP for the subcarriers with large error probabilities) in order to guarantee a certain performance in terms of uncoded BER for each subcarrier.

5.4.2 Minimum MSE vs maximum cut-off rate

The minimum MSE cost function includes the summation over all the subcarriers of the MSE for each specific subcarrier. Consequently, the optimum design will preserve the performance of each subcarrier. This behavior influences the performance of the MMSE algorithm in terms of coded BER. As it will be shown in the next section, its performance is much worse than previous power allocation strategies.

5.4.3 Minimum MSE vs maximum mutual information

Recent works published by Verdu et al. showed a simple connection between the mutual information and the MMSE [Guo05]. That paper founds that the mutual information and the MMSE satisfy a very simple relationship. Considering the Gaussian channel $y = \sqrt{snr}\, x + n$, and for any statistical distribution of the information symbols x, it can be proven that:

$$\frac{d}{d\, snr}\, I\left(x; \sqrt{snr}\, x + n\right) = mmse(snr) \tag{5.38}$$

where mutual information is given in nats and the $mmse(snr)$ is the MSE achieved by the conditional mean estimator. This representation of the mutual information has some interesting applications listed in [Guo05].

Next, it is proven that the fundamental relationship in (5.38) also applies to our concerning problem. Let us denote $C(\boldsymbol{\phi}_k^2)$ as the objective function used to maximize the mutual information for the particular case of Gaussian symbols (5.36); $MMSE(\phi_k^2(l))$ as the function that contains the contribution of the power allocation values $\phi_k^2(l)$ into the objective function that minimizes the MSE (i.e.the contribution of the kth subcarrier and lth subchannel mode); and let us decompose the linear transformation at the transmitter as $\mathbf{F}_k = \mathbf{V}_k\boldsymbol{\Phi}_k$, which it has been shown to be optimal both for the maximization of the channel capacity and for the minimization of the MSE. Then, the mutual information and the MMSE satisfy:

$$\frac{\partial}{\partial\phi_k^2(l)} = \frac{1}{K}\frac{\sigma_x^2}{\sigma_n^2}|\lambda_k(l)|^2 MMSE(|\phi_k(l)|^2) \tag{5.39}$$

Proof. Let us first focus the attention on the objective function used to minimize the MSE. In general, the optimal MMSE solution is given by a non-linear receiver. Only in the specific case of Gaussian symbols the optimal receiver is linear. In accordance, because of a linear transformation has been placed at the receiver, we shall assume Gaussian symbols in order to make use of the objective function presented in Chapter 4. Recalling the cost function that minimizes the MSE (4.34):

$$\xi_k^{MMSE} = \sigma_x^2 \mathrm{Tr}\left\{\left(\mathbf{G}_k\hat{\mathbf{H}}_k\mathbf{F}_k - \mathbf{I}\right)^H\left(\mathbf{G}_k\hat{\mathbf{H}}_k\mathbf{F}_k - \mathbf{I}\right)\right\} + \sigma_n^2\ \mathrm{Tr}\left\{\mathbf{G}_k^H\mathbf{G}_k\right\} \tag{5.40}$$

replacing the matrix $\mathbf{G}_k$ by the optimal linear receiver that minimizes the MSE (see equation (4.115) in Appendix 4.B), and decomposing the linear transformation $\mathbf{F}_k = \mathbf{V}_k\boldsymbol{\Phi}_k$, we get:

$$\begin{aligned} \xi_k^{MMSE} &= \sum_{l=1}^{M}\frac{1}{1+\frac{\sigma_x^2|\lambda_k(l)|^2\phi_k^2(l)}{\sigma_n^2}} \\ &= \sum_{l=1}^{M} MMSE(\phi_k^2(l)) \end{aligned} \tag{5.41}$$

On the other hand, let us consider the mutual information for Gaussian symbols (i.e., the channel capacity) introduced in (5.36). Using the decomposition $\mathbf{F}_k = \mathbf{V}_k\boldsymbol{\Phi}_k$ the channel capacity is given by:

$$\begin{aligned} C(\boldsymbol{\phi}_k^2) &= \frac{1}{K}\sum_{k=1}^{K}\log_2\left(\det\left(\mathbf{I}+\frac{\sigma_x^2}{\sigma_n^2}\boldsymbol{\Lambda}_k^2\boldsymbol{\Phi}_k^2\right)\right) \\ &= \frac{1}{K}\sum_{k=1}^{K}\sum_{l=1}^{M}\log_2\left(1+\frac{\sigma_x^2}{\sigma_n^2}|\lambda_k(l)|^2\phi_k^2(l)\right) \end{aligned} \tag{5.42}$$

where the second equality uses the diagonal structure of matrices $\boldsymbol{\Lambda}_k^2$ and $\boldsymbol{\Phi}_k^2$.

Finally, computing the derivative of the capacity with respect to the power allocation $\phi_k^2(l)$ we proof the relationship in (5.39). ■

Among the multiple consequences of this result, we discuss, next, those that are directly related with the design of power allocation algorithms:

1. *The input distribution that maximizes the mutual information (i.e. Gaussian inputs) maximizes the MSE* [Guo05].

2. *To find closed-form expressions for the gradient of the mutual information with respect to arbitrarily parameters.* [Pal05].
   Specifically those results could be useful to formulate power allocation policies that minimize the mutual information without the use of the explicit mutual information expression (difficult to obtain for some specific constellations). In that sense [Loz05] derives the power allocation policy that maximizes the sum mutual information achieved over parallel independent Gaussian channels with arbitrarily (and not necessarily identical) input constellations.

5.5 Simulation results

This section presents some numerical results that conclude the comparative analysis carried out in Section 5.4 and validates the cut-off rate as a criterion that is related with the coded BER. All the algorithms studied in this thesis are compared in terms of maximum data rates (cut-off rate, capacity and mutual information constrained to a QPSK constellation) and in terms of packet error rate at the decoder output for a specific channel code.

The algorithm that maximizes the cut-off rate based on a Taylor expansion of the cost function (5.25) is labelled as *'Ro Robust'*. The algorithms *'MI-QPSK'* and *'MI Gaussian'* maximize the mutual information constrained to the QPSK constellation (5.34) and the channel capacity (under the assumption of Gaussian symbols) (5.37), respectively. The *'Open-loop'* solution allocates the same power to all the subcarriers and antennas. Finally, the robust algorithms derived in the previous chapter to maximize the MSE *'MMSE Robust'* and minimize the uncoded BER *'BER Chernoff Robust'* and *'BER ApproxQ Robust'* are also evaluated in terms of coded BER. It is worth to note that although the *'BER Chernoff Robust'* algorithm was shown to be irrelevant in terms of uncoded BER, it will be an algorithm of great significance in terms of coded BER.

The simulation parameters have been selected according to the HIPERLAN/2 standard [HL201]. As depicted in Figure 5.1 the bit stream to be transmitted is applied to the classical convolutional code of rate 1/2 and generator polynomials 133_{OCT} and 171_{OCT} [3]. The encoder is initialized to the zero state and returned to it after encoding 864 bits (i.e. 2 DLC-PDUs

[3]This channel code is not specific for HIPERLAN/2 but it is widely used in several standards as for example the IEEE 802.11a standard [IEE99]

of 54 bytes - according to HL/2) by appending 6 tail bits. The coded bits are mapped into a QPSK constellation and interleaved with a symbol-block interleaver. Afterwards, the symbols are multiplexed in the $K = 48$ subcarriers (grouping M symbols per subcarrier), prefiltered by the matrices $\mathbf{F}_k$ and finally modulated in OFDM symbols (including pilot tones and empty subcarriers according to HIPERLAN/2).

All the algorithms are simulated using a ML detector that takes into account the channel uncertainty at the receiver. For the evaluation of the coded BER the presence of the interleaver makes difficult the computation of the Viterbi metrics. Hence, as the goal is only focused on the design of the transmitter, a genie decoder is used that computes the log-likelihood of a symbol $\mathbf{x}_k(l)$ assuming that all other symbols that were transmitted simultaneously in in vector $\mathbf{x}_k$ are known to the receiver.

When more than one symbol is transmitted per subcarrier (i.e., $M > 1$) the unitary matrix $\mathbf{T}_k$, whose importance was revealed in the previous chapter, is set to the DFT matrix except for the *'Open-loop'* algorithm (the unitary matrix is omitted).

Two Rayleigh MIMO channels have been considered obeying an exponential power delay profile. A first channel with 50ns. of delay spread models a typical office indoor scenario, whereas a second channel with 150ns. of delay spread (close to maximum channel length permitted by the cyclic prefix of the OFDM modulation) models an outdoor scenario. As in Chapter 4 the variance of the channel estimation error is assumed to be constant for all the taps and independent of the SNR at the transmitter. For both channel scenarios the simulations are carried out with a channel estimation error at the transmitter $\sigma^2_{\epsilon_{Tx}} = 0.12$ and with a channel uncertainty at the receiver proportional to the noise variance $\sigma^2_{\epsilon_{Rx}} = 0.375\sigma^2_n$ (see results in Section 3.6 for details on these chosen values).

5.5.1 Power allocation strategies

Before analyzing the performance of the different algorithms in terms of coded BER we will proceed to analyze the power allocation strategies followed by the different algorithms as we did in the previous chapter. Next figures display an example of the values of $\lambda_i\phi_i$ (sorted in decreasing order of the equivalent channel subcarriers λ_i $i = 1 \dots K$) for a specific channel realization and its associated p.d.f. We will only focus on the parameters that are relevant for the coded BER and, consequently, we will not repeat in this section some studies that were carried out in the previous chapter (cf. Section 4.5.1). Specifically, the analysis of the different algorithms for several noise levels has been omitted since the main conclusions drawn in that section still hold for the new algorithms introduced in this chapter. On the other hand, the comparison between robust and nonrobust algorithms has also been omitted and only robust algorithms are considered when partial CSI is available at the transmitter.

Figure 5.2 illustrates the power allocation strategies for the specific configuration $M_T = M_R = 1$ (SNR=0dB) comparing the results for two different channel profiles. Focusing the analysis on the algorithm that maximizes the cut-off rate, it can be seen that it allocates the majority of the power to the best subcarriers. This power allocation policy reinforces the highest subcarriers neglecting the weakest ones. This can be appreciated in the curves of Figures 5.2b and 5.2d, where it can be seen that comparing the *'Ro Robust'* algorithm with the rest of algorithms, large values of $\lambda_i\phi_i$ are more likely for the algorithm that maximizes the cut-off rate (see for example the values $\lambda_i\phi_i > 1.5$). The number of nulled subcarriers also reveals this behavior. Note, in Table 5.2, that the *'Ro Robust'* algorithm nulls a number of subcarriers much larger than the *'MMSE Robust'* and *'BER Chernoff Robust'* algorithms. This way to allocate the power is coherent with the comparative study carried out in Section 5.4, where it was argued that the maximization of the cut-off rate, related with the geometric mean of the pairwise error probability, will take care of the best subcarriers assuming that the symbols transmitted in the faded subcarriers will be recovered by the channel code.

Figure 5.3 displays the power allocation strategies for the same channel profiles but increasing the number of antennas $M_T = M_R = 3$ and the number of symbols transmitted per subcarrier $M = 3$ (SNR=5dB). The same conclusions drawn for the SISO configuration still hold for the MIMO case, noting that the algorithms that allocate more power to the best subcarriers, or equivalently null more subcarriers, are the *'Ro Robust'* and *'BER Chernoff Robust'* algorithms. This result will explain the good performances of the *'BER Chernoff Robust'* design in terms of coded BER since its power allocation policy is similar to that one followed by the algorithm that maximized the cut-off rate.

5.5.2 Rate vs SNR

Since the cut-off rate is a lower bound on the channel capacity, it is appropriate to evaluate, for the different algorithms, the channel capacity (serving as a benchmark) (5.36) and the cut-off rate (5.12). However, practical modulation schemes are restricted to discrete distributions (in our particular case QPSK modulation). Hence, as argued in Section 5.3, a fair comparison between the different design criteria must be done evaluating, also, the mutual information for the QPSK constellation (5.30).

Figures 5.4 and 5.5 show, in bits per channel use, the ergodic cut-off rate and the ergodic capacity as a function of the SNR and for the channel with 150ns. of delay spread(for the sake of clarity in the figures the ergodic mutual information constrained to the QPSK modulation has not been plotted and this information is only summarized in the tables). The results for the SISO case are displayed in Figure 5.4, whereas the 3x3 configuration is illustrated in Figure 5.5. In both cases the curves average 500 channels and are plotted when perfect CSI is available at the transmitter and when the channel estimation has a variance equal to $\sigma_\epsilon^2 = 0.12$.

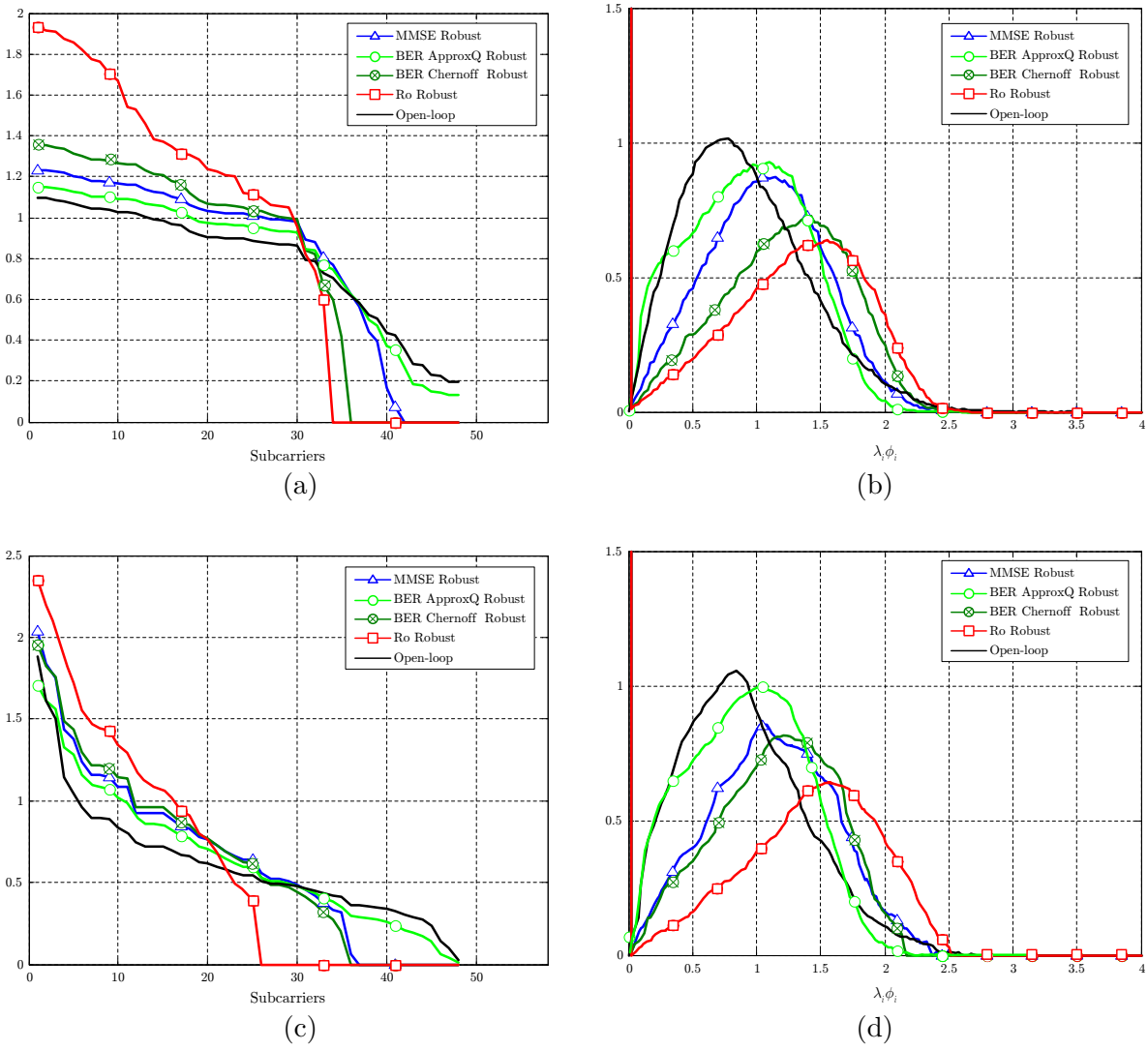


Figure 5.2: An example of $\lambda_i\phi_i$ values (sorted in decreasing order of the equivalent channel subcarriers λ_i $i = 1 \ldots K$) for a specific channel realization (a) and (c), and its associated p.d.f. (b) and (d). $M_T = 1$, $M_R = 1$, $M = 1$. Transmitter uncertainty: $\sigma^2_{\epsilon_{Tx}} = 0.12$ ($\rho = 0.1$). SNR=0dB. Channel Delay Spread 50ns. [(a) and (b)] and 150ns. [(c) and (d)].

	Eb/No	Ro Robust	MMSE Robust	BER ApproxQ Robust	BER Chernoff Robust
1x1	DS:50ns.	14.8	6.0	<0.05	11.4
	DS:150ns.	15.7	6.7	<0.10	8.2

Table 5.2: Number of null subcarriers.

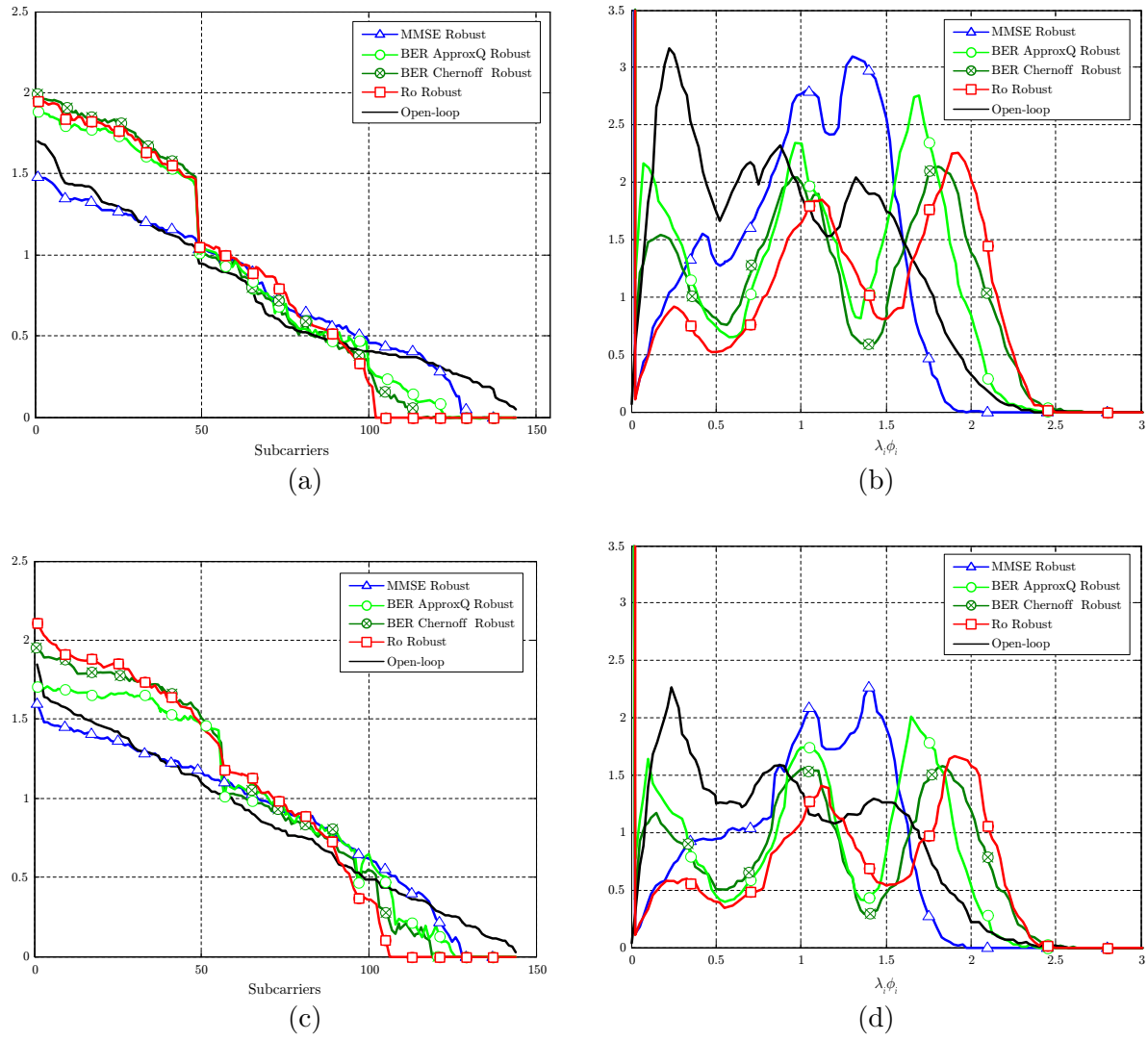


Figure 5.3: An example of $\lambda_i\phi_i$ values (sorted in decreasing order of the equivalent channel subcarriers λ_i $i = 1 \ldots K$) for a specific channel realization (a) and (c), and its associated p.d.f. (b) and (d). $M_T = 3$, $M_R = 3$, $M = 3$. Transmitter uncertainty: $\sigma^2_{\epsilon_{Tx}} = 0.12$ ($\rho = 0.1$). SNR=5dB. Channel Delay Spread 50ns. [(a) and (b)] and 150ns. [(c) and (d)].

	Eb/No	Ro Robust	MMSE Robust	BER ApproxQ Robust	BER Chernoff Robust
3x3	DS:50ns.	37.6	17.8	18.4	24.4
	DS:150ns.	25.2	12.2	12.1	15.6

Table 5.3: Number of null subcarriers.

As it can be shown, the cut-off rate is a lower bound on the channel capacity. Comparing the cut-off rate and the capacity limits, we can see that for a certain data rate, the capacity gives an additional coding gain of 3-4dB over the prediction of the cut-off rate. A comparison of the curves when the channel is perfectly known at the transmitter and when the CSI is noisy, exhibits the robustness of the algorithms. As it can be shown, the performance of the different algorithms only degrades slightly due to the channel uncertainty.

Tables 5.4 and 5.5 illustrate the performance of the different pre-filter design algorithms in terms of cut-off rate (Ro), mutual information for QPSK constellation (MI-QPSK) and channel capacity assuming Gaussian symbols (Capacity), listing the SNR required to achieve a value of Ro, MI-QPSK and Capacity of 66% of the maximum data rate (i.e., 1.33 bits per channel use for the 1x1 configuration, and 4 bits per channel use for the 3x3 configuration). The SNR values are referred to the values required by the *'Ro Robust'* algorithm. Note that, as expected, the *'Ro Robust'* and the *'MI-QPSK'* designs have best performances in terms of Ro and the MI-QPSK, respectively.

Somewhat surprisingly, the *'BER ApproxQ Robust'*, the *'BER Chernoff Robust'* and the *'MMSE Robust'* designs have better performance in terms of mutual information for QPSK constellation when channel is perfectly known at the transmitter, but this gain vanishes when only unreliable CSI estimates are available. It can also be shown that the gain of the closed-loop algorithms compared to the open-loop transmission increases with the number of antennas, thanks to the efficient exploitation of the diversity increase provided by the spatial channel. However, the relative gain of the *'Ro Robust'* design to the *'BER Chernoff Robust'* diminishes with the number of antennas because the presence of the unitary matrix $\mathbf{T}_k$, which breaks the direct link between the apparition of a deep fade and the loss of a constellation symbol. Nevertheless, this gain is preserved for noisy CSI's.

A more accurate analysis of these results will allow to predict poor performances for some algorithms in terms of coded BER. First, note that although the *'MI Gaussian'* design has the best performance in terms of capacity, it is outperformed by the rest of designs in terms of Ro and the MI-QPSK. This result, in accordance with the conclusions given in Section 5.3.2 and [Sca00], explains the very poor performance of this algorithm in terms of coded BER when a specific constellation is forced (see numerical results in the next section). On the other hand, it is also worth to remark that the *'MMSE Robust'* design performs worse than *'Open-loop'* at high SNR's (see Figure 5.5) and, consequently, as it will be justified in the next section, the *'MMSE Robust'* design degrades when noise is reduced.

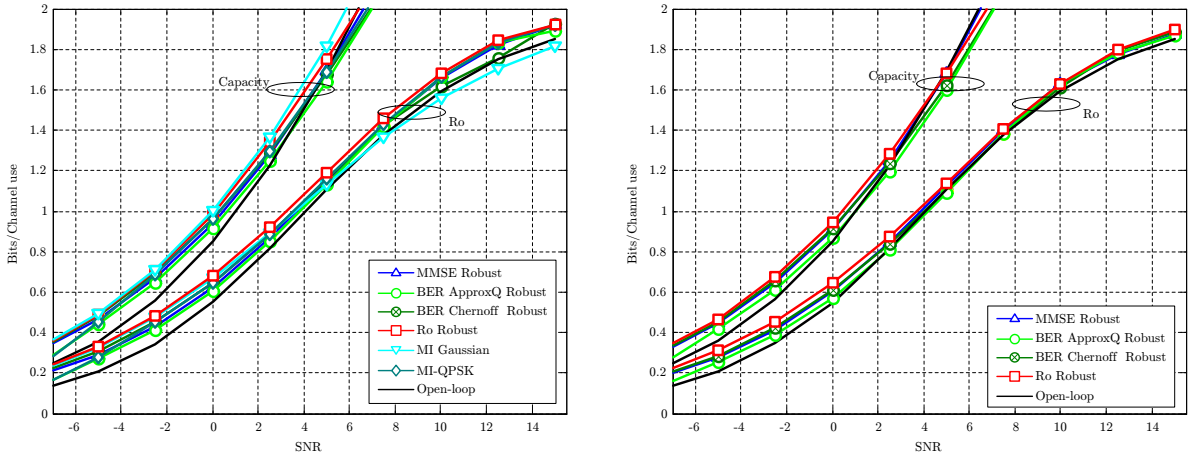


Figure 5.4: Ergodic cut-off rate and channel capacity for different power allocation strategies. $M_T = 1$, $M_R = 1$. Channel: exponential power delay profile with delay spread 150ns. Perfect CSI at the transmitter (a) and noisy CSI at the transmitter $\sigma^2_{\epsilon_{Tx}} = 0.12$ ($\rho = 0.1$) (b).

CSI Uncertainty	Algorithm:	Ro Robust	Unc. BER Chernoff	Unc. BER ApproxQ	MI Gaussian	MI QPSK	MMSE Robust	Open-loop
$\sigma^2_{\epsilon_{Tx}} = 0$	R_o	6.21 (dB)	+0.33	+0.43	+0.83	+0.27	+0.35	+0.76
	MI - QPSK	4.71 (dB)	-0.29	-0.21	+0.77	-0.30	-0.22	+0.24
	Capacity	2.37 (dB)	+0.35	+0.61	-0.19	+0.30	+0.40	+0.64
$\sigma^2_{\epsilon_{Tx}} = 0.12$	R_o	6.78 (dB)	+0.17	+0.19			+0.04	+0.19
	MI - QPSK	4.84 (dB)	-0.04	+0.02			-0.04	+0.11
	Capacity	2.73 (dB)	+0.32	+0.53			+0.17	+0.28

Table 5.4: SNR required to achieve the 66% of the maximum data rate (1.33 bits per channel use) for different information theoretic parameters (Cut-off rate, Mutual Information for QPSK and Channel Capacity). The SNR values are given with respect to the SNR required for the algorithm that maximizes the cut-off rate. MIMO configuration 1x1. Channel: exponential power delay profile with delay spread 150ns.

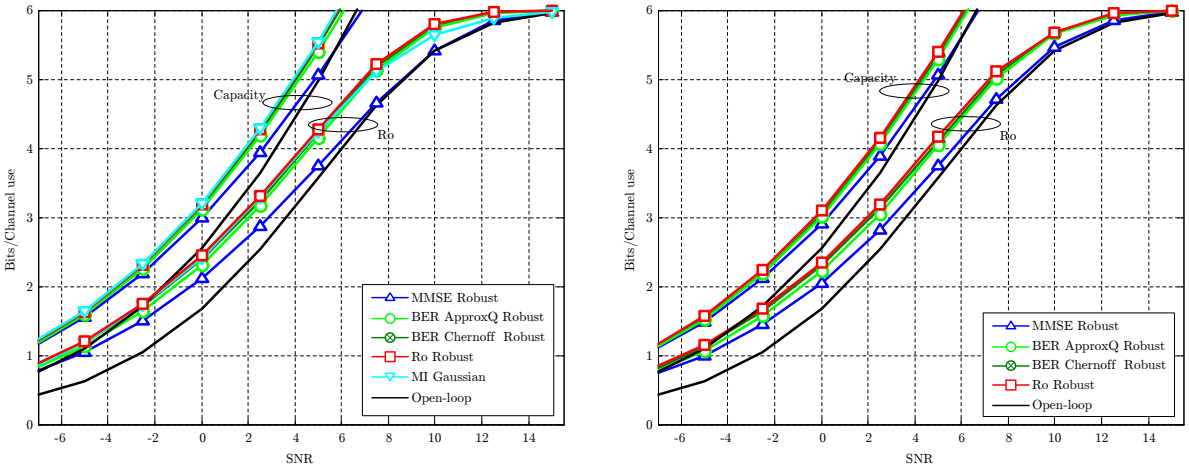


Figure 5.5: Ergodic cut-off rate and channel capacity for different power allocation strategies. $M_T = 3$, $M_R = 3$. Channel: exponential power delay profile with delay spread 150ns. Perfect CSI at the transmitter (a) and noisy CSI at the transmitter $\sigma^2_{\epsilon_{Tx}} = 0.12$ ($\rho = 0.1$) (b).

CSI Uncertainty	Algorithm:	Ro Robust	Unc. BER Chernoff	Unc. BER ApproxQ	MI Gaussian	MMSE Robust	Open-loop
$\sigma^2_{\epsilon_{Tx}} = 0$	R_o	4.19 (dB)	+0.14	+0.34	+0.09	+1.40	+1.74
	MI - QPSK	2.41 (dB)	-0.02	-0.02	-0.05	+0.65	+1.45
	Capacity	1.79 (dB)	+0.12	+0.19	-0.04	+0.74	+1.31
$\sigma^2_{\epsilon_{Tx}} = 0.12$	R_o	2.04 (dB)	+0.12	+2.77		+3.51	+3.89
	MI - QPSK	2.59 (dB)	+0.03	+0.09		+0.62	+1.27
	Capacity	2.04 (dB)	+0.12	+0.23		+0.62	+1.06

Table 5.5: SNR required to achieve the 66% of the maximum data rate (4 bits per channel use) for different information theoretic parameters (Cut-off rate, Mutual Information for QPSK and Channel Capacity). The SNR values are given with respect to the SNR required for the algorithm that maximizes the cut-off rate. MIMO configuration 3x3 channel: exponential power delay profile with delay spread 150ns.

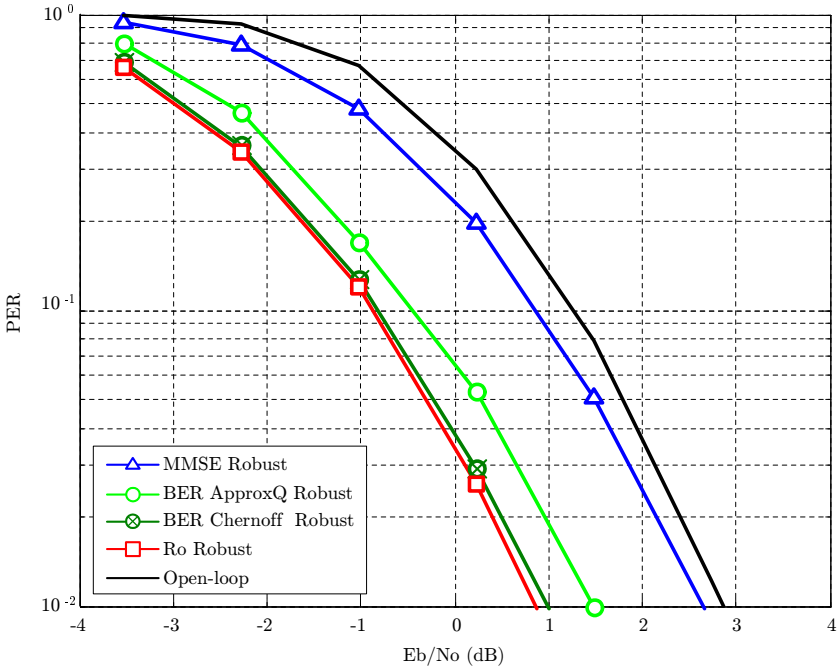
5.5.3 Coded BER vs Eb/No

This section analyzes the performance of the different algorithms in terms of *Packet Error Rate* (PER), where the PER has been evaluate over a packet length of 54 bytes, for different channel profiles and MIMO configurations while keeping constant the channel code.

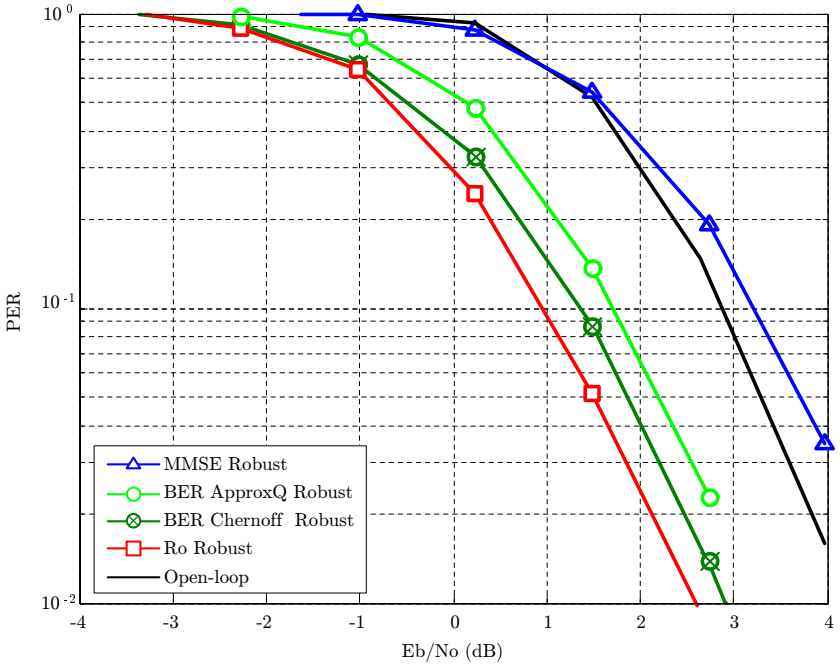
Figure 5.6 shows the PER for the 3x3 antenna configuration and for the different channel delay spreads: 50ns. and 150ns. In both cases the variance of the channel estimation error at the transmitter was set to $\sigma^2_{\epsilon_{Tx}} = 0.12$ and the CSI at the receiver was assumed to be proportional to the noise variance $\sigma^2_{\epsilon_{Rx}} = 0.375\sigma^2_n$. As it is shown in these figures, the *'Ro Robust'* and *'BER Chernoff Robust'* algorithms have similar performances in terms of coded BER and exhibit a remarkable improvement when they are compared with the rest of algorithms. This performance was announced in Section 5.5.1 where the power allocation strategies followed by the different algorithms were analyzed. Several conclusions can be drawn from this result. First, note that the algorithms with the best packet error rate are those that exhibit the best cut-off rate in Table 5.5, evidencing that there exist a direct relationship between the cut-off rate and the performance in terms of the coded BER. In accordance, the maximization of the cut-off rate, although it cannot be used as an optimum algorithm to minimize the coding rate for a specific channel code, guarantees satisfying performances in terms of coded BER. On the other hand, focusing on the *'BER Chernoff Robust'* algorithm, its behavior can result surprising. Note that this algorithm, explicitly designed to minimize the uncoded BER, has excellent performances in terms of coded BER and outperforms the *'BER ApproxQ Robust'* algorithm, whose performance in terms of uncoded BER was superior. The unexpected similarities between the *'Ro Robust'* and *'BER Chernoff Robust'* algorithms are more appreciable in Figure 5.6a (50ns of delay spread) and they will be justified in Section 5.6. In Figure 5.6b it can be seen how at high SNR's the *'BER Chernoff Robust'* algorithm approaches the *'BER ApproxQ Robust'* algorithm, a behavior that is explained from the results given in Section 4.5.1 and Appendix 4.C, where it was shown that at high SNR's both algorithms have similar performances.

It is also worth to analyze the poor performance of the *'MMSE Robust'* algorithm, specifically when the channel delay spread is 150ns. (see Figure 5.6b). This behavior, raised in the previous section, can be understood by noting that for high SNR's the *'MMSE Robust'* solution tends to the zero-forcing solution, resulting in a power allocation policy that assigns less power to the stronger subcarriers and more power to the deepest faded ones. This strategy, opposite to that one followed by the algorithm that maximizes the cut-off rate, is the cause of this unsatisfactory coded BER.

Figure 5.7 shows the PER performance for the SISO case (i.e., $M_T = M_R = M = 1$). In this case perfect CSI was assumed in order to include in the comparison the *'MI Gaussian'* algorithm. Some remarkable results can be found in this figure. First, note that the *'MI Gaussian'* design



(a)



(b)

Figure 5.6: PER comparison between different power allocation strategies. $M_T = 3$, $M_R = 3$, $M = 3$. Transmitter uncertainty:$\sigma^2_{\epsilon_{Tx}} = 0.12$ ($\rho = 0.1$). Receiver uncertainty: $\sigma^2_{\epsilon_{Rx}} = 0.375\sigma^2_n$. Channel Delay Spread 50ns. (a) and 150ns. (b)

has the worst performance in terms of coded BER. This result was predicted from the results given in the previous section where it was shown that this algorithm has the worst performance in terms of cut-off rate and MI-QPSK. As argued in [Sca00] the performance of this design could be improved if different constellations were used on each subcarrier depending on the goodness of the channel.

On the other hand, note that the *'Ro Robust'* is outperformed by the *'MMSE Robust'* algorithm and by the algorithms that minimize the uncoded BER. The explanation for this discouraging result can be found in Table 5.2, where it is shown that the *'Ro Robust'* algorithms nulls much more subcarriers than the rest of the algorithms. Under such conditions the correcting capability of the specific channel code employed in the simulations is not enough to restore the large number of lost symbols that where transmitted in those subcarriers and, consequently, the PER is increased. A simple solution to this problem can be found introducing a more complex channel code with a higher correcting capability. In this thesis we will propose, however, an alternative to improve the performance of the algorithms that null a large number of subcarriers based on the modification of the interleaver while keeping constant the channel code. In accordance, Section 5.7 analyzes the importance of the interleaver and proposes an adaptive interleaver to provide good performances of the *'Ro Robust'* algorithm in terms of coded BER.

Finally, it is worth to remark that this results cannot be conclusive for other channel codes like turbo codes and LDPC codes since those codes allow to achieve data rates greater than R_o. For this kind of channel codes new simulations should be computed to draw any conclusions.

5.5.4 Performance vs CSI quality at the transmitter

Figure 5.8 illustrates the minimum Eb/No required to achieve a $PER \leq 10^{-2}$ as a function of the degree of the channel uncertainty at the transmitter, measured by the coefficient ρ:

$$\rho = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_h^2} \tag{5.43}$$

where $\rho = 0$ denotes perfect CSI, whereas $\rho = 1$ means no channel knowledge. The simulated scenario was a 3x3 MIMO channel with 50ns of delay spread.

As it can be shown, the robustness of the *'Ro Robust'* and *'BER Chernoff Robust'* algorithms is evidenced since both solutions adapt the power allocation to the CSI uncertainty, leading to a solution that tends to *'Open-loop'* when no channel information is available. The similarities between the *'Ro Robust'* and *'BER Chernoff Robust'* algorithms are also displayed. Note that only a slight difference is appreciated for intermediate values of ρ. Figure 5.8 also exhibits the poor performance of the *'MMSE Robust'* algorithm. This algorithm is largely outperformed by the *'Ro Robust'* and *'BER Chernoff Robust'* and it performs worse than *'Open-loop'* when the CSI quality degrades as it was already observed in the previous chapter (cf., Section 4.5.3).

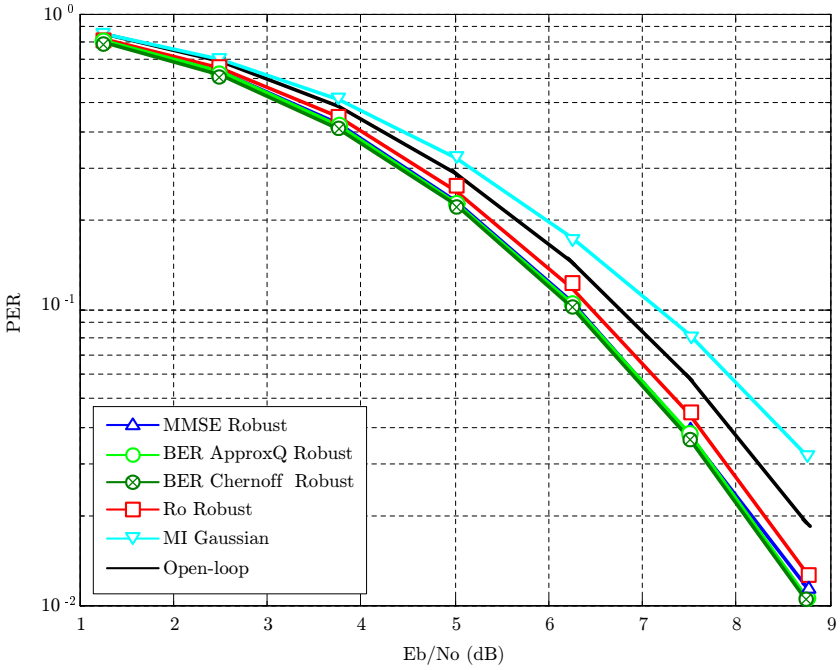


Figure 5.7: PER comparison between different power allocation strategies. $M_T = 1$, $M_R = 1$, $M = 1$. Transmitter uncertainty: Perfect CSI. Receiver uncertainty: Perfect CSI. Channel Delay Spread 150ns.

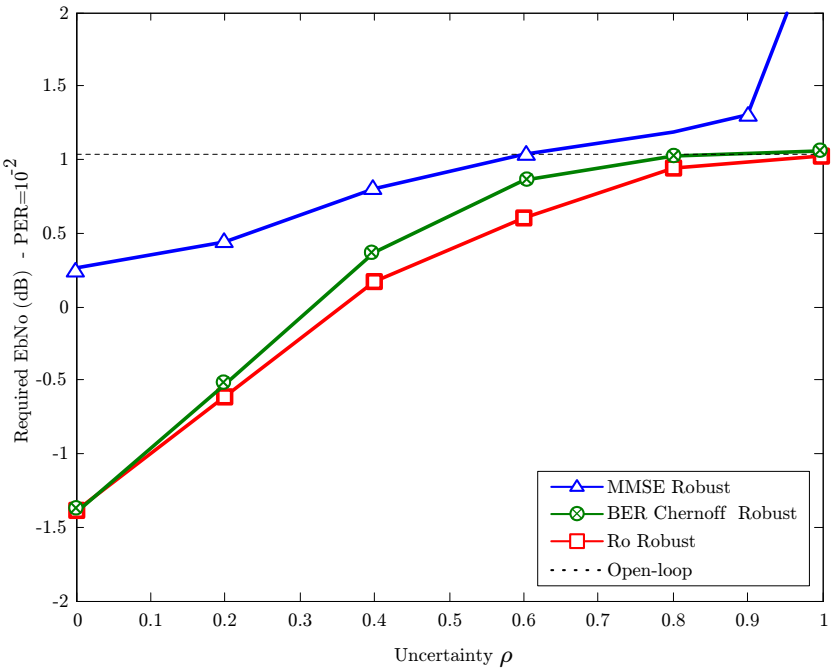


Figure 5.8: Minimum Eb/No that achieves a PER$\leq 10^{-2}$ vs the transmitter uncertainty ρ. Receiver uncertainty: Perfect CSI. $M_T = 3$, $M_R = 3$, $M = 3$

5.6 Minimum uncoded BER with the Chernoff bound

This section continues with the analysis of the numerical results presented in the previous section providing a detailed study of the surprising behavior of the *'BER Chernoff Robust'* algorithm in terms of coded BER. As it was shown in Chapter 4 this algorithm was explicitly designed to minimize the uncoded BER, with the particularity that the Gaussian Q-function was upper bounded by the widely used Chernoff bound. It was also shown in the chapter that, although the Chernoff bound is not a tight bound of the Q-function, this algorithm has good performances in terms of uncoded BER (see, e.g., the comparison between the *'BER Chernoff Robust'* and *'BER ApproxQ Robust'* algorithms in Section 4.5.2). This result, however, is not conclusive of the behavior of the algorithm in terms of coded BER. It could be expected, *a priori*, an acceptable but not optimal behavior of the algorithm when a channel code is introduced (see, e.g., the performance of the *'BER ApproxQ Robust'* algorithm in Figure 5.6). However, this algorithm, compared with algorithms that have been explicitly designed to minimize the coded BER, exhibits similar performances. Details of this behavior can be found in the power allocation policy used by this algorithm (see Figure 5.3), the performance in terms of coded BER (see Figure 5.6) and the maximum achievable rates given by the information theoretic parameters (see Table 5.5).

Next, some arguments that justify this excellent performance in terms of coded BER are presented. First, it is shown that, for the specific case of beamforming and QPSK constellation, the *'BER Chernoff Robust'* algorithm is equivalent to the algorithm the maximizes the cut-off rate. It will also be shown that the power allocation policy followed by this algorithm is similar to that one followed by an algorithm that maximizes the mutual information constrained to a finite constellation. Finally, it will be proven that for the particular case of OFDM, the approximation of the uncoded BER by its Chernoff bound is directly related with the coded BER by means of the union bound on the word-error rate (a criterion that takes into account the distance spectrum of the channel code).

Comparison between the objective functions for 'BER Chernoff Robust' and 'Ro Robust' at high SNR's

The *'BER Chernoff Robust'* design optimizes the cost function presented in Section 4.4.2 for the specific case of beamforming (i.e., $M = 1$). Particularizing this cost function for high SNR's, the error probability is dominated by the minimum distance between any pairs of symbols. Therefore, the objective function in (4.63) can be written for the specific QPSK constellation as

follows:

$$\xi_{UncBER} = \frac{1}{\mathcal{N}^{2M}K} \sum_{k=1}^{K} \frac{1}{2}\, e^{-\frac{2\sigma_x^2 |\mathbf{H}_k\mathbf{F}_k|^2}{4\sigma_n^2}} \tag{5.44}$$

where the Chernoff bound has been considered in the computation of the $\mathcal{Q}(\cdot)$ function (i.e. $\alpha_k = \delta_k = 1/2$) and perfect CSI has been assumed (i.e. $\omega = 0$).

Similarly, the cut-off rate expression (5.12) is next derived under the same conditions and assumptions. For the specific QPSK constellation, and assuming that only one symbol is transmitted per subcarrier (i.e., $M = 1$), the error vector $\mathbf{d}$ becomes an scalar whose values belong to the finite alphabet: $\mathbf{d} \in \{0, \pm 2A, \pm 2Aj, \pm 2A \pm 2Aj\}$, where A is the amplitude of the transmitted symbols. For this specific case the expectation over the error vector $\mathbf{d}$ can be easily computed noting that $\mathbf{d}^2 \in \left\{0(4), 2\sigma_x^2(8), 4\sigma_x^2(4)\right\}$, where the number in parenthesis represents the frequency of the different values. Hence, (5.12) can be simplified as follows:

$$\begin{aligned} R_o = & -\frac{1}{K}\sum_{k=1}^{K} \log_2 \frac{1}{16}\, 4 \left(1 + 2e^{-\frac{2\sigma_x^2}{4\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k|^2} + e^{-\frac{2\sigma_x^2}{4\sigma_n^2}2|\mathbf{H}_k\mathbf{F}_k|^2} \right) \\ = & -\frac{1}{K}\sum_{k=1}^{K} \log_2 \frac{1}{4} \left(1 + e^{-\frac{2\sigma_x^2}{4\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k|^2} \right)^2 \\ = & \; 2 - \frac{2}{K}\sum_{k=1}^{K} \log_2 \left(1 + e^{-\frac{2\sigma_x^2}{4\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k|^2} \right) \end{aligned} \tag{5.45}$$

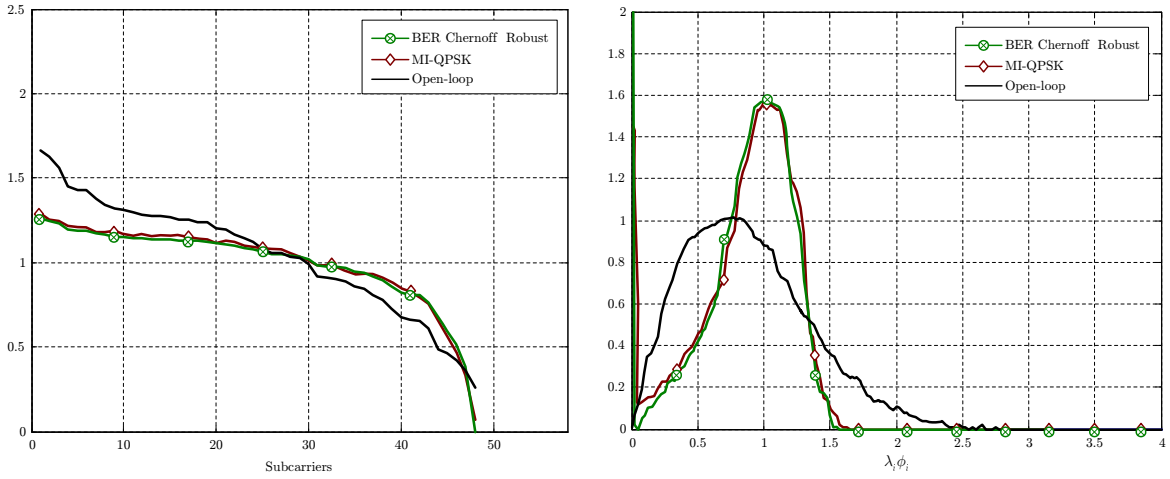
Next the approximation $\log_2(1+x) \simeq x/\log(2)$ can be used to simplify R_o when assuming high SNR:

$$\xi_{Ro} = 2 - \frac{2}{K \cdot \log(2)} \sum_{k=1}^{K} e^{-\frac{2\sigma_x^2}{4\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k|^2} \tag{5.46}$$

Note that under the same assumptions (i.e., high SNR and perfect CSI at the transmitter) previous equation is the same, apart from irrelevant constants, that the objective function used to minimize the uncoded BER under the Chernoff bound assumption. Hence, at high SNR's, the algorithm that maximizes the cut-off rate is equivalent to the algorithm that minimizes the uncoded BER using the Chernoff bound in the design.

Comparison between the power allocation policy for *'BER Chernoff Robust'* and *'MI-QPSK'*

Repeating the same procedure followed in Section 5.5.1 to compare the power allocation strategies followed by the different designs, we compare in this section the *'BER Chernoff Robust'* with the *'MI-QPSK'*. The complexity in the derivation of the design that maximizes the mutual



(a) (b)

Figure 5.9: Comparison of power allocation strategies for the *'BER Chernoff Robust'* and *'MI-QPSK'* algorithms. $M_T = 1$, $M_R = 1$, $M = 1$. Perfect CSI at the transmitter and the receiver. Channel delay spread 150ns. An example of $\lambda_i\phi_i$ values (sorted in decreasing order of the equivalent channel subcarriers λ_i $i = 1 \dots K$) for a specific channel realization (a), and its associated p.d.f. (b)

information restricts the comparison to the SISO configuration and makes use of the iterative algorithm and the Monte-Carlo simulations proposed in Section 5.3.

Figure 5.9 compares the power allocation strategy followed by the two algorithms. As it can be appreciated, the *'BER Chernoff Robust'* design follows the same power allocation policy than the *'MI-QPSK'*, an algorithm that is based on information theoretic criteria. This result, although not conclusive by itself, reinforces the excellent behavior of the *'BER Chernoff Robust'* design in terms of coded BER.

Union Bound on the Word Error Rate

The power allocation algorithms proposed in this thesis do not take into account neither the time/frequency structure of the channel (e.g., how the the fading coefficients are arranged in time or in frequency) nor the structure of the specific channel code that is employed. The consideration, in this thesis, of any design that could be focused on this aspects, was unestimated due to the complexity of its implementation. In [Lam05] a possible criterion that takes into account the distance spectrum of the specific channel code, although not the time/frequency correlation of the channel is outlined for the OFDM modulation. Surprisingly, the algorithm, which evaluates the Union Bound on the Word Error Rate (UB-WER), has a one-to-one mapping with the *'BER Chernoff Robust'* algorithm under certain conditions. In particular, this direct

relationship applies when the maximum distance between two codewords $(\mathbf{c}, \hat{\mathbf{c}})$ used to compute the Union Bound on the Word Error Rate does no exceeds the total number of OFDM symbols, a condition that is not difficult to accomplish in our case [4]. That result is conclusive to justify the excellent performance of the *'BER Chernoff Robust'* in terms of coded BER, since, in the particular case of OFDM, this uncoded BER metric has an interpretation from the perspective of coded BER.

5.7 Adaptive interleaver

This section analyzes the importance of the interleaver and shows that the performance of the algorithms, specifically those that aim to minimize the coded BER, can be improved if an appropriate interleaver is introduced. A simple study of the channel in an OFDM modulation, shows that when the channel is assumed to be constant for several OFDM symbols, it appears a channel structure that is repeated with a periodicity equal to the number of subcarriers. This cyclic structure, inherent to the OFDM modulation, systematically introduces a fixed pattern of damaged symbols and a fixed pattern of good symbols. A specific design of the interleaver should consider this structure to break, not only the bursty channel errors, but also this periodicity. Note, for example, that a simple symbol-block interleaver with deep equal to the channel periodicity, would be totally ineffective to randomize the channel errors. Hence, considering the necessity to design appropriate interleavers, and assuming that CSI is available at the transmitter, we propose the design of an adaptive interleaver well suited for OFDM systems.

The design of this interleaver requires a preliminary analysis of the power allocation strategies followed by the different algorithms. As argued in Section 5.5.1, the algorithms that are focused on information theoretic criteria (i.e., those that maximize the cut-off rate or maximize the mutual information) tend to allocate scarce or null power to the worst channel modes, redistributing this power among the remaining subchannels. This procedure, assuming that the correcting capability of the channel code will restore the information lost in the worst channel modes, results in a power allocation policy that takes care of the bits allocated to the weaker subchannels, neglecting the bits transmitted through the deepest ones (see, e.g., results in Figure 5.2). It is worth to remark, however, that the transmitter design has not been optimized for a specific channel code. In consequence, when a particular code is employed, the power allocation policy previously described could result in an unsatisfactory performance if the code is not able to correct the large number of damaged bits. An example of this can be found in Figure 5.7 where the algorithm that maximizes the cut-off rate (supposedly one of the best algorithms in

[4] It has been proven by simulation, an for the specific scenario considered in this thesis, that although this condition does not holds, the dissimilarities between the two algorithms are inappreciable.

terms of coded BER) is outperformed by the several algorithms. This undesired degradation of the code performance can be diminished if the interleaver rightly decides which are the bits to be allocated in the worst subcarriers and which ones to be allocated to the best ones, a feasible task if the CSI is available at the transmitter. Next, we propose the design of such interleaver while keeping unalterable the channel code.

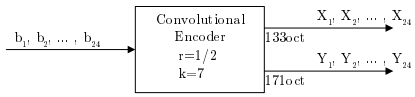
5.7.1 RCPC interleaver

Block interleavers and pseudo-random interleavers are based on the assumption that no CSI is available at the transmitter. However, when the channel impulse response is available, this information can be used to reallocate the bits and improve the performance of the error correcting code. The adaptive interleaver proposed in the last part of this thesis, named *'RCPC interleaver'*, is based on the *Rate-Compatible Punctured Convolutional Codes* (RCPC codes). This family of codes, introduced by Hagenauer in [Hag88], extend the concept of punctured convolutional codes by puncturing a low rate $1/N$ code (in our concerning problem rate $1/2$) to obtain a family of codes with rate $P/(P+l)$ where l can be varied between 1 and $(N-1)P$ (in our case we can get the rates 8/9, 8/10, 8/12, 8/14 and 8/16). A restriction in the puncturing tables ensures that all code bits of a high rate code are used by the lower rate codes. Hence, if a high rate code is not enough powerful to decode the bits with the desired error rate, it suffices to transmit some additional bits, which were previously punctured, to decode the information with a lower error rate.

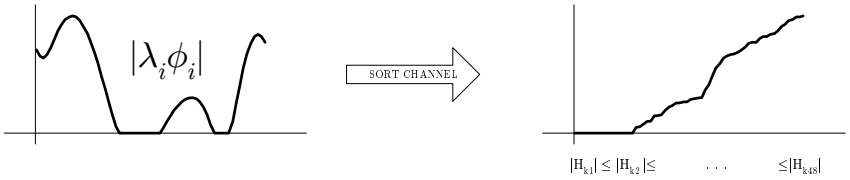
The *'RCPC interleaver'* does not introduce any puncturing in the encoded bits but uses the puncturing tables reported in [Hag88] to reallocate the bits to the different subcarriers. Following the principle that the bits allocated in the deepest subcarriers will be the worst bits at the receiver, the interleaver will assign the coded bits in the same order as they would be punctured in a RCPC code. Thus, the bits that are first punctured in this family of codes are the candidates to be allocated in the worst subcarriers, whereas the bits that are never punctured are allocated in the best subcarriers.

Figure 5.10 displays the procedure employed to deduce the interleaver pattern. For the sake of simplicity we show the pattern for the specific BPSK constellation, but a similar pattern with a sequence of 96 coded bits is easily designed for a QPSK constellation. First, the encoded bits are grouped according to the puncturing sequence given in [Hag88] for the RCPC code. Specifically, for the mother code with rate $1/2$ and memory $M = 6$ (i.e., $k = 7$) the first group of bits is constituted by the encoded bits that should be eliminated to obtain the punctured code of rate $4/7$, which correspond to the bits $Y_4, Y_8, Y_{12}, Y_{16}, Y_{20}, Y_{24}$ in a sequence of 48 encoded bits. Similarly, a second and third groups, with 6 bits each, and a fourth group, with 3 bits, are composed by the bits that should be punctured to obtain the codes of rate $2/3$, $4/5$, and $8/9$,

Encoder:



Channel:



Interleaver:

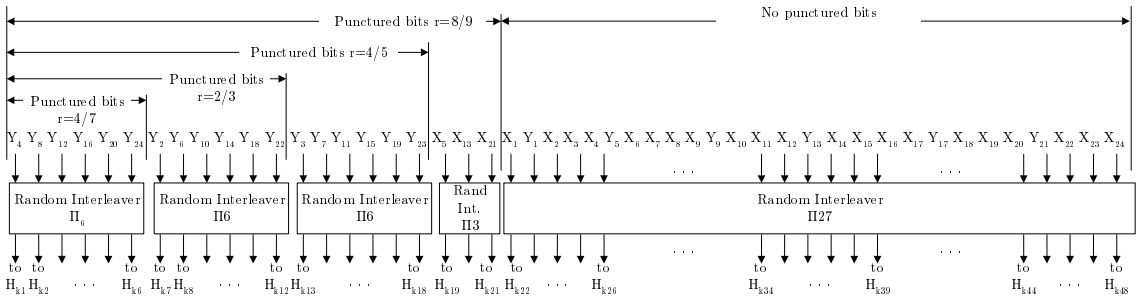
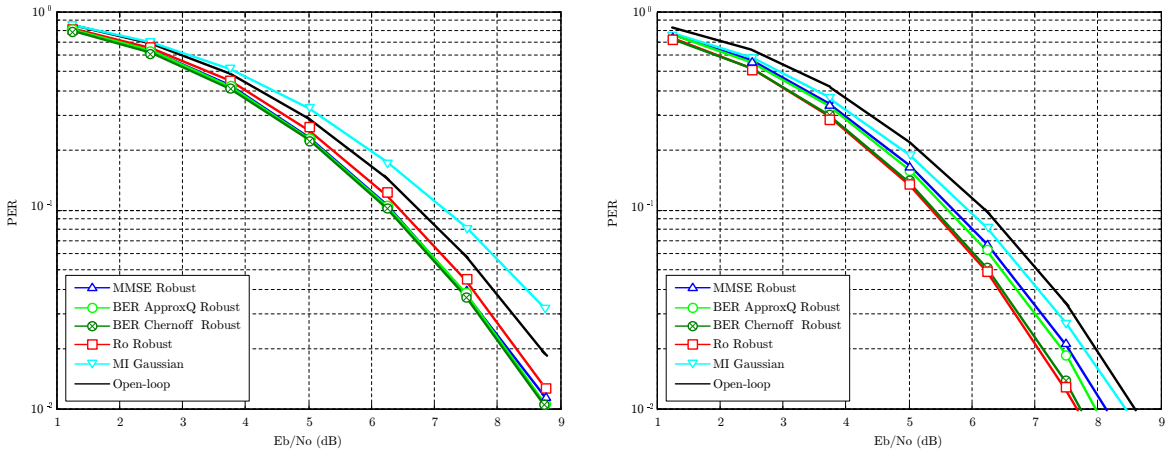


Figure 5.10: Adaptive Interleaver: RCPC Interleaver.

respectively. The remaining bits, which are never punctured, constitute the last group of 27 encoded bits. Next, since there is no priority in the bits that belong to the same group, for each group, a pseudo-random interleaver randomizes the position of the different bits inside its group. Finally, the resulting sequence of 48 interleaved bits are allocated to the different subcarriers in ascending order, so that the first bits to be punctured are allocated to the worst subcarriers, whereas the best subcarriers are assigned to the bits that are never punctured.

Figure 5.11 displays the gain in terms of PER when the symbol-block interleaver (Figure 5.11a) is substituted by the adaptive interleaver (Figure 5.11b). The 1x1 antenna configuration with perfect CSI and channel delay spread of 150ns. has been considered. As it can be seen, all the algorithms improve their performance when the RCPC interleaver is employed. Table 5.6 lists the gain in terms of the Eb/No required to achieve a $PER = 10^{-1}$. Note, as expected, that the algorithms that experiment a larger gain are those that null more subcarriers (i.e., the *'Ro Robust'* and *'MI-QPSK'* algorithms), whereas the *'Open-loop'*, which does not nulls any subcarrier, is the algorithm with the lower gain.



(a) (b)

Figure 5.11: PER comparison between different power allocation strategies for two different interleavers. $M_T = 1$, $M_R = 1$, $M = 1$. Perfect CSI at the transmitter and the receiver. Channel delay spread 150ns. Bloc Interleaver (a) and Adaptive Interleaver (RCPC) (b)

Interleaver	Ro Robust	MMSE Robust	BER ApproxQ Robust	BER Chernoff Robust	MI Gaussian	Open-loop
Bloc	6.47	6.32	6.31	6.27	7.17	6.76
RCPC	5.36	5.71	5.61	5.39	5.94	6.21
Gain (dB)	1.11	0.61	0.70	0.88	1.23	0.55

Table 5.6: Eb/No required to achieve a $PER = 10^{-1}$

5.8 Conclusions

This chapter has proposed a Bayesian approach for the design of linear precoding schemes that are robust to channel estimation errors. Contrary to the previous chapter, the algorithms that have been proposed were based on information theoretic criteria and the goal has been the minimization of the coded BER. Specifically, three different cost functions have been considered: the maximization of the cut-off rate, the maximization of the mutual information constrained to a specific constellation and maximization of the channel capacity, assuming, in all cases, that a ML detector is implemented at the receiver.

An extensive study has been made for the maximization of the cut-off rate, since it achieves the best performance in terms of coded BER, and because a closed form solution has been obtained for it. This solution, based on the SVD of the so called equivalent channel, has been found by means of a quadratic approximation of the cost function using a Taylor expansion.

The approximation has been validated by simulations. The derived design can be regarded as a reconfigurable algorithm that adapts the transmitter to the available channel knowledge, providing a solution that converges to the open-loop design (i.e the same power is allocated across all antennas and subcarriers) for the case of no channel knowledge and to the closed-loop design with perfect CSI for the case of no uncertainty.

The mutual information and the channel capacity has also been studied. However, the cost functions resulted in complex analytical expressions and, consequently, the channel capacity has only been studied when perfect CSI is available, and a complex numerical solution has been proposed for the mutual information that is only feasible in simple scenarios.

All the precoder designs proposed in this thesis has been compared, analyzing the differences in the objective functions, the different power allocation policies, and the performance in terms of data rates and coded BER. The simulation results exhibit that the design that maximizes the cut-off rate, although it is an algorithm that has been optimized based on random coding, provides good performances for the specific convolutional code defined in the HIPERLAN/2 standard. The experiments also reveal that the design that minimizes the uncoded BER using the Chernoff upper bound for the pairwise error probability exhibits excellent performances in terms of coded BER. The behavior of this algorithm has been, thus, accurately studied, and it has been shown that under certain conditions it is related with the maximization of the coded BER.

Finally, this chapter has also analyzed the importance of the interleaver and its influence when the power allocation policy does not allocate power in a large amount of symbols. Assuming that CSI is available at the transmitter, an adaptive interleaver based on the RCPC codes is proposed. Numerical results have shown that all the algorithms improve their performance when the adaptive interleaver is used, specially those that follow a power allocation policy that nulls more subcarriers.

Appendix 5.A Objective function to maximize the cut-off rate

This Appendix derives the maximization problem given in (5.20). First, the expectation over the channel estimation error is computed to derive the quadratic optimization problem. Later, imposing the structure given in (5.22) for the matrix $\mathbf{F}_k$ the final objective function is found.

Recalling the quadratic approximation in (5.20), where the irrelevant constant D_0 has been eliminated, we have:

$$f_0\left(\mathbf{F}_k\right) = E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{-\frac{1}{K}\sum_{k=1}^{K}\left(\sum_{i=1}^{\mathcal{N}^{2M}}\frac{D_{1i}}{4\sigma_n^2}|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i|^2+\frac{D_{2i}}{\left(4\sigma_n^2\right)^2}\left(|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i|^2\right)^2+\sum_{i=1}^{\mathcal{N}^{2M}}\sum_{j=1}^{\mathcal{N}^{2M}}\frac{D_{3ij}}{\left(4\sigma_n^2\right)^2}\left(|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i|^2\right)\left(|\mathbf{H}_k\mathbf{F}_k\mathbf{d}_j|^2\right)\right)\right\} \tag{5.47}$$

The expectation over the channel estimation error reduces to the computation of the following terms:

$$E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\mathbf{d}_i^H\mathbf{F}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i\right\} \tag{5.48a}$$

$$E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\left(\mathbf{d}_i^H\mathbf{F}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i\right)\left(\mathbf{d}_j^H\mathbf{F}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{F}_k\mathbf{d}_j\right)\right\} \tag{5.48b}$$

$$E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\left(\mathbf{d}_i^H\mathbf{F}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i\right)^2\right\} \tag{5.48c}$$

In order to simplify these computations, the channel matrix $\mathbf{H}_k$ can be expressed as: $\mathbf{H}_k = \hat{\mathbf{H}}_k^{eq} + \tilde{\mathbf{H}}_k$, where $\hat{\mathbf{H}}_k^{eq}$ is the conditional mean of the channel derived in Section 4.2.2 and $\tilde{\mathbf{H}}_k$ is a Gaussian random variable that accomplishes: $E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\tilde{\mathbf{H}}_k\right\} = 0$ and $E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\tilde{\mathbf{H}}_k^H\tilde{\mathbf{H}}_k\right\} = M_R\omega\mathbf{I}$. Substituting this decomposition into (5.48a) this term can be evaluated as follows:

$$\begin{aligned} E_{\mathbf{h}_k|\hat{\mathbf{h}}} & \left\{\mathbf{d}_i^H\mathbf{F}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i\right\} \\ &= \mathbf{d}_i^H\mathbf{F}_k^H E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\left(\hat{\mathbf{H}}_k^{eq}+\tilde{\mathbf{H}}_k\right)^H\left(\hat{\mathbf{H}}_k^{eq}+\tilde{\mathbf{H}}_k\right)\right\}\mathbf{F}_k\mathbf{d}_i \\ &= \mathbf{d}_i^H\mathbf{F}_k^H\hat{\mathbf{H}}_k^{eqH}\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k\mathbf{d}_i + M_R\omega\mathbf{d}_i^H\mathbf{F}_k^H\mathbf{F}_k\mathbf{d}_i \end{aligned} \tag{5.49}$$

A more complicate derivation is required to compute the second expectation (5.48b). Making use of the identity $\mathrm{vec}\left(\mathbf{ABC}\right) = \left(\mathbf{C}^T\otimes\mathbf{A}\right)\mathrm{vec}\left(\mathbf{B}\right)$ this term can be written as follows:

$$\begin{aligned} E_{\mathbf{h}_k|\hat{\mathbf{h}}} & \left\{\left(\mathbf{d}_i^H\mathbf{F}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{F}_k\mathbf{d}_i\right)\left(\mathbf{d}_j^H\mathbf{F}_k^H\mathbf{H}_k^H\mathbf{H}_k\mathbf{F}_k\mathbf{d}_j\right)\right\} \\ &= \left[\left(\mathbf{F}_k\mathbf{d}_i\right)^T\otimes\left(\mathbf{F}_k\mathbf{d}_i\right)^H\right]E_{\mathbf{h}_k|\hat{\mathbf{h}}}\left\{\mathrm{vec}\left(\mathbf{H}_k^H\mathbf{H}_k\right)\mathrm{vec}\left(\mathbf{H}_k^H\mathbf{H}_k\right)^H\right\}\left[\left(\mathbf{F}_k\mathbf{d}_j\right)^*\otimes\left(\mathbf{F}_k\mathbf{d}_j\right)\right] \end{aligned} \tag{5.50}$$

Next, the channel matrix $\mathbf{H}_k$ is decomposed as: $\mathbf{H}_k = \hat{\mathbf{H}}_k^{eq} + \tilde{\mathbf{H}}_k$ and, thus, the derivation of the

expectation over the channel estimation error is given by:

$$\begin{aligned}
E_{\mathbf{h}_k|\hat{\mathbf{h}}} & \left\{ \text{vec}\left(\mathbf{H}_k^H \mathbf{H}_k\right) \text{vec}\left(\mathbf{H}_k^H \mathbf{H}_k\right)^H \right\} \\
= & E_{\mathbf{h}_k|\hat{\mathbf{h}}} \left\{ \text{vec}\left(\left(\hat{\mathbf{H}}_k^{eq} + \tilde{\mathbf{H}}_k\right)^H \left(\hat{\mathbf{H}}_k^{eq} + \tilde{\mathbf{H}}_k\right) \right) \text{vec}\left(\left(\hat{\mathbf{H}}_k^{eq} + \tilde{\mathbf{H}}_k\right)^H \left(\hat{\mathbf{H}}_k^{eq} + \tilde{\mathbf{H}}_k\right) \right)^H \right\} \\
= & E_{\mathbf{h}_k|\hat{\mathbf{h}}} \Big\{ \text{vec}\left(\hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq}\right) \text{vec}\left(\hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq}\right)^H + \text{vec}\left(\hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq}\right) \text{vec}\left(\tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k\right)^H \\
& + \text{vec}\left(\hat{\mathbf{H}}_k^{eqH} \tilde{\mathbf{H}}_k\right) \text{vec}\left(\hat{\mathbf{H}}_k^{eqH} \tilde{\mathbf{H}}_k\right)^H + \text{vec}\left(\tilde{\mathbf{H}}_k^H \hat{\mathbf{H}}_k^{eq}\right) \text{vec}\left(\tilde{\mathbf{H}}_k^H \hat{\mathbf{H}}_k^{eq}\right)^H \\
& + \text{vec}\left(\tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k\right) \text{vec}\left(\hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq}\right)^H + \text{vec}\left(\tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k\right) \text{vec}\left(\tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k\right)^H \Big\} \\
= & \text{vec}\left(\hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq} + M_R \omega \mathbf{I}\right) \text{vec}\left(\hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq} + M_R \omega \mathbf{I}\right)^H + M_R \omega^2 \left(\mathbf{I} \otimes \mathbf{I}\right) \\
& + \omega \left(\mathbf{I} \otimes \hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq}\right) + \omega \left(\mathbf{H}_k^{eqT} \mathbf{H}_k^{eq*} \otimes \mathbf{I}\right)
\end{aligned} \tag{5.51}$$

where only the non-null terms have been considered in the second equality. Substituting this expression into (5.50) and after some straightforward derivations, a final expression for (5.48b) is given by:

$$\begin{aligned}
E_{\mathbf{h}_k|\hat{\mathbf{h}}} & \left\{ \left(\mathbf{d}_i^H \mathbf{F}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{F}_k \mathbf{d}_i\right) \left(\mathbf{d}_j^H \mathbf{F}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{F}_k \mathbf{d}_j\right) \right\} \\
= & \mathbf{d}_i^H \mathbf{F}_k^H \left(\hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq} + M_R \omega \mathbf{I}\right) \mathbf{F}_k \mathbf{d}_i \mathbf{d}_j^H \mathbf{F}_k^H \left(\hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq} + M_R \omega \mathbf{I}\right) \mathbf{F}_k \mathbf{d}_j \\
& + M_R \omega^2 \text{Tr}\left\{ \mathbf{F}_k \mathbf{d}_i \mathbf{d}_i^H \mathbf{F}_k^H \mathbf{F}_k \mathbf{d}_j \mathbf{d}_j^H \mathbf{F}_k^H \right\} \\
& + \omega \text{Tr}\left\{ \mathbf{F}_k \mathbf{d}_i \mathbf{d}_i^H \mathbf{F}_k^H \hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq} \mathbf{F}_k \mathbf{d}_j \mathbf{d}_j^H \mathbf{F}_k^H \right\} \\
& + \omega \text{Tr}\left\{ \mathbf{F}_k \mathbf{d}_j \mathbf{d}_j^H \mathbf{F}_k^H \hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq} \mathbf{F}_k \mathbf{d}_i \mathbf{d}_i^H \mathbf{F}_k^H \right\}
\end{aligned} \tag{5.52}$$

Finally, (5.48c) can be computed as the previous derivation for the specific case $i = j$. In accordance, this expectation is given by:

$$\begin{aligned}
E_{\mathbf{h}_k|\hat{\mathbf{h}}} & \left\{ \left(\mathbf{d}_i^H \mathbf{F}_k^H \mathbf{H}_k^H \mathbf{H}_k \mathbf{F}_k \mathbf{d}_i\right)^2 \right\} \\
= & \left(\mathbf{d}_i^H \mathbf{F}_k^H \left(\hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq} + M_R \omega \mathbf{I}\right) \mathbf{F}_k \mathbf{d}_i\right)^2 \\
& + M_R \omega^2 \left(\mathbf{d}_i^H \mathbf{F}_k^H \mathbf{F}_k \mathbf{d}_i\right)^2 \\
& + 2\omega \left(\mathbf{d}_i^H \mathbf{F}_k^H \mathbf{F}_k \mathbf{d}_i \mathbf{d}_i^H \mathbf{F}_k^H \hat{\mathbf{H}}_k^{eqH} \hat{\mathbf{H}}_k^{eq} \mathbf{F}_k \mathbf{d}_i\right)
\end{aligned} \tag{5.53}$$

Once the expectation over the channel estimation error has been computed, the final objective

function is derived substituting (5.49), (5.52) and (5.53) in the quadratic approximation (5.47):

$$\begin{aligned} f_0\left(\mathbf{F}_k\right) = -\frac{1}{K} \quad & \sum_{k=1}^{K}\left(\sum_{i=1}^{\mathcal{N}^{2M}} \frac{D_{1i}}{4\sigma_n^2}\mathbf{d}_i^H\mathbf{F}_k^H\left(\hat{\mathbf{H}}_k^{eqH}\hat{\mathbf{H}}_k^{eq}+M_R\omega\mathbf{I}\right)\mathbf{F}_k\mathbf{d}_i\right. \\ & +\sum_{i=1}^{\mathcal{N}^{2M}}\frac{D_{2i}}{\left(4\sigma_n^2\right)^2}\left[\left(\mathbf{d}_i^H\mathbf{F}_k^H\left(\hat{\mathbf{H}}_k^{eqH}\hat{\mathbf{H}}_k^{eq}+M_R\omega\mathbf{I}\right)\mathbf{F}_k\mathbf{d}_i\right)^2\right. \\ & \qquad \left.+\omega\left(\mathbf{d}_i^H\mathbf{F}_k^H\mathbf{F}_k\mathbf{d}_i\mathbf{d}_i^H\mathbf{F}_k^H\left(2\hat{\mathbf{H}}_k^{eqH}\hat{\mathbf{H}}_k^{eq}+M_R\omega\mathbf{I}\right)\mathbf{F}_k\mathbf{d}_i\right)\right] \\ & +\sum_{i,j=1}^{\mathcal{N}^{2M}}\frac{D_{3ij}}{\left(4\sigma_n^2\right)^2}\left[\mathbf{d}_i^H\mathbf{F}_k^H\left(\hat{\mathbf{H}}_k^{eqH}\hat{\mathbf{H}}_k^{eq}+M_R\omega\mathbf{I}\right)\mathbf{F}_k\mathbf{d}_i\mathbf{d}_j^H\mathbf{F}_k^H\left(\hat{\mathbf{H}}_k^{eqH}\hat{\mathbf{H}}_k^{eq}+M_R\omega\mathbf{I}\right)\mathbf{F}_k\mathbf{d}_j\right. \\ & \qquad +M_R\omega^2\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{d}_i\mathbf{d}_i^H\mathbf{F}_k^H\mathbf{F}_k\mathbf{d}_j\mathbf{d}_j^H\mathbf{F}_k^H\right\} \\ & \qquad +\omega\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{d}_i\mathbf{d}_i^H\mathbf{F}_k^H\hat{\mathbf{H}}_k^{eqH}\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k\mathbf{d}_j\mathbf{d}_j^H\mathbf{F}_k^H\right\} \\ & \qquad \left.\left.+\omega\mathrm{Tr}\left\{\mathbf{F}_k\mathbf{d}_j\mathbf{d}_j^H\mathbf{F}_k^H\hat{\mathbf{H}}_k^{eqH}\hat{\mathbf{H}}_k^{eq}\mathbf{F}_k\mathbf{d}_i\mathbf{d}_i^H\mathbf{F}_k^H\right\}\right]\right) \end{aligned} \tag{5.54}$$

Next, a closed-form solution for this problem is derived forcing the structure given in (5.22): $\mathbf{F}_k = \mathbf{V}_k\mathbf{\Phi}_k\mathbf{T}$, where the same unitary matrix $\mathbf{T}$ has been fixed for all the subcarriers:

$$\begin{aligned} f_0\left(\mathbf{F}_k\right) = -\frac{1}{K} \quad & \sum_{k=1}^{K}\left(\sum_{i=1}^{\mathcal{N}^{2M}} \frac{D_{1i}}{4\sigma_n^2}\mathbf{d}_i^H\mathbf{T}^H\mathbf{\Phi}_k^2\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\right. \\ & +\sum_{i=1}^{\mathcal{N}^{2M}}\frac{D_{2i}}{\left(4\sigma_n^2\right)^2}\left[\left(\mathbf{d}_i^H\mathbf{T}^H\mathbf{\Phi}_k^2\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\right)^2\right. \\ & \qquad \left.+\omega\left(\mathbf{d}_i^H\mathbf{T}^H\mathbf{\Phi}_k^2\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\mathbf{\Phi}_k^2\left(2\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\right)\right] \\ & +\sum_{i,j=1}^{\mathcal{N}^{2M}}\frac{D_{3ij}}{\left(4\sigma_n^2\right)^2}\left[\mathbf{d}_i^H\mathbf{T}^H\mathbf{\Phi}_k^2\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_j^H\mathbf{T}^H\mathbf{\Phi}_k^2\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_j\right. \\ & \qquad +M_R\omega^2\mathrm{Tr}\left\{\mathbf{\Phi}_k^2\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\mathbf{\Phi}_k^2\mathbf{T}\mathbf{d}_j\mathbf{d}_j^H\mathbf{T}^H\right\} \\ & \qquad +\omega\mathrm{Tr}\left\{\mathbf{\Phi}_k^2\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\mathbf{\Phi}_k^2\mathbf{\Lambda}_k^2\mathbf{T}\mathbf{d}_j\mathbf{d}_j^H\mathbf{T}^H\right\} \\ & \qquad \left.\left.+\omega\mathrm{Tr}\left\{\mathbf{\Phi}_k^2\mathbf{T}\mathbf{d}_j\mathbf{d}_j^H\mathbf{T}^H\mathbf{\Phi}_k^2\mathbf{\Lambda}_k^2\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right\}\right]\right) \end{aligned} \tag{5.55}$$

This expression can be simplified making use of the following identities:

Identity 1 *Let* $\mathbf{D}$ *be a diagonal matrix whose diagonal entries are given by vector* $\mathbf{d}$ *and* $\mathbf{A}$ *a generic matrix with the appropriate dimensions, the following equality holds*

$$\mathrm{Tr}\left\{\mathbf{DA}\right\} = \mathrm{diag}\left(\mathbf{A}\right)^T\mathbf{d}$$

where $\mathrm{diag}\left(\mathbf{A}\right)$ *is a column vector whose elements are the diagonal entries of matrix* $\mathbf{A}$

Identity 2 *Let* $\mathbf{D}_1$ *and* $\mathbf{D}_2$ *be two diagonal matrices whose diagonal entries are given by vectors* $\mathbf{d}_1$ *and* $\mathbf{d}_2$*, respectively, and let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two generic matrices with the appropriate dimensions, the following equality holds*

$$\mathrm{Tr}\left\{\mathbf{D}_1\mathbf{A}\mathbf{D}_2\mathbf{B}\right\} = \mathbf{d}_1^T\left(\mathbf{A}\odot\mathbf{B}^T\right)\mathbf{d}_2$$

In accordance with these identities, the cost function in (5.55) can be rewritten as a function of vector ϕ_k^2, which stores the diagonal elements of matrix $\mathbf{\Phi}_k\mathbf{\Phi}_k^H$:

$$\begin{aligned} f_0\left(\mathbf{F}_k\right) = -\frac{1}{K} \quad & \sum_{k=1}^{K}\left(\sum_{i=1}^{\mathcal{N}^{2M}}\frac{D_{1i}}{4\sigma_n^2}\mathrm{diag}\left(\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)^T\phi_k^2\right. \\ & +\sum_{i=1}^{\mathcal{N}^{2M}}\frac{D_{2i}}{\left(4\sigma_n^2\right)^2}\left[\phi_k^{2T}\left(\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\odot\left(\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)^T\right)\phi_k^2\right. \\ & \qquad \left.+\omega\left(\phi_k^{2T}\mathrm{diag}\left(\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)\mathrm{diag}\left(\left(2\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)^T\phi_k^2\right)\right] \\ & +\sum_{i,j=1}^{\mathcal{N}^{2M}}\frac{D_{3ij}}{\left(4\sigma_n^2\right)^2}\left[\phi_k^{2T}\mathrm{diag}\left(\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)\mathrm{diag}\left(\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_j\mathbf{d}_j^H\mathbf{T}^H\right)^T\phi_k^2\right. \\ & \qquad +M_R\omega^2\phi_k^{2T}\left(\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\odot\left(\mathbf{T}\mathbf{d}_j\mathbf{d}_j^H\mathbf{T}^H\right)^T\right)\phi_k^2 \\ & \qquad +\omega\phi_k^{2T}\left(\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\odot\left(\mathbf{\Lambda}_k^2\mathbf{T}\mathbf{d}_j\mathbf{d}_j^H\mathbf{T}^H\right)^T\right)\phi_k^2 \\ & \qquad \left.\left.+\ \omega\phi_k^{2T}\left(\mathbf{T}\mathbf{d}_j\mathbf{d}_j^H\mathbf{T}^H\odot\left(\mathbf{\Lambda}_k^2\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)^T\right)\phi_k^2\right]\right) \end{aligned} \tag{5.56}$$

Although this expression seems to be a little bit confused, it can be compressed as follows:

$$f_0\left(\mathbf{F}_k\right) = -\frac{1}{K}\sum_{k=1}^{K}\left(\mathbf{v}_k^T\boldsymbol{\phi}_k^2+\boldsymbol{\phi}_k^{2T}\mathbf{A}_k\boldsymbol{\phi}_k^2\right) \tag{5.57}$$

where vector $\mathbf{v}_k$ and matrix $\mathbf{A}_k$ are given by:

$$\mathbf{v}_k = \sum_{i=1}^{\mathcal{N}^{2M}}\frac{D_{1i}}{4\sigma_n^2}\mathrm{diag}\left(\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)$$

$$\begin{aligned} \mathbf{A}_k = \quad & \sum_{i=1}^{\mathcal{N}^{2M}}\frac{D_{2i}}{\left(4\sigma_n^2\right)^2}\left[\left(\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\odot\left(\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)^T\right)\right. \\ & \qquad \left.+\omega\left(\mathrm{diag}\left(\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)\mathrm{diag}\left(\left(2\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)^T\right)\right] \\ & -\frac{1}{2\left(4\sigma_n^2\right)^2}\left[\left(\sum_{i=1}^{\mathcal{N}^{2M}}D_i\mathrm{diag}\left(\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)\right)\left(\sum_{i=1}^{\mathcal{N}^{2M}}D_i\mathrm{diag}\left(\left(\mathbf{\Lambda}_k^2+M_R\omega\mathbf{I}\right)\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)\right)^T\right. \\ & \qquad +M_R\omega^2\left(\left(\sum_{i=1}^{\mathcal{N}^{2M}}D_i\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)\odot\left(\sum_{j=1}^{\mathcal{N}^{2M}}D_j\mathbf{T}\mathbf{d}_j\mathbf{d}_j^H\mathbf{T}^H\right)^T\right) \\ & \qquad \left.+\ 2\omega\left(\left(\sum_{i=1}^{\mathcal{N}^{2M}}D_i\mathbf{T}\mathbf{d}_j\mathbf{d}_j^H\mathbf{T}^H\right)\odot\left(\sum_{j=1}^{\mathcal{N}^{2M}}D_j\mathbf{\Lambda}_k^2\mathbf{T}\mathbf{d}_i\mathbf{d}_i^H\mathbf{T}^H\right)^T\right)\right] \end{aligned} \tag{5.58}$$

Chapter 6

Conclusions and future work

This thesis has dealt with two different, but related, topics: the channel tracking at the transmitter and the design of linear precoders robust to imperfect CSI. At the beginning of this dissertation the first topic was studied. It was proposed the design of a transmitter channel tracker based on a feedback channel. Efforts were made to minimize the throughput at the feedback channel. The novelty was the proposal of two identical linear predictors at both sides of the link with a return link that only updates the minimum amount of information required to track the channel impulse response when the channel is time varying. The rest of the thesis covered the second topic studying different robust power allocation algorithms when the CSI is not perfectly known at the transmitter. First, the minimum MSE and the minimum uncoded BER parameters were chosen to be optimized, evaluating the performance of the algorithms in terms of uncoded BER. Next, the study was extended by proposing robust power allocation strategies that minimized the coded BER making use of information theoretic criteria. On this topic we find one of the main contribution of this thesis: the proposal of the cut-off rate as a parameter of design whose maximization is directly related to the coded BER. Next we summarize the results in these subjects and enumerate some open research lines for future works.

6.1 Transmitter channel tracking

The first topic studied in this thesis was the design of a feasible scheme to track the CSI at the transmitter by means of a low capacity feedback link. This problem was formulated and solved in Chapter 3. Historically, those schemes have been criticized because of the large amount of information to be transmitted from the receiver to the transmitter, specially in fast TV channels when the whole channel impulse response is supposed to be transmitted. The work in Chapter 3 focused, thus, the attention in an accurate design of the return channel. The proposed solution was based on a scheme that reminding the well known DPCM transmitter only fed back the prediction error and allowed to track the channel variations with only two or four bits per complex channel coefficient.

The design was divided into two stages. First, a model to predict the channel variations was proposed. After reviewing the different alternatives proposed in the literature, a pth order autoregressive model was suggested. This model was next used to design two identical linear predictors located at the transmitter and at the receiver to predict the future values of the channel given the past channel knowledge. Finally, those linear predictors were placed as part of a Kalman filter to design the transmitter channel tracker. Special attention was paid on the minimum amount of information required to aid the linear predictor at the transmitter. It was shown that the prediction error provides this information. Once the information to be fed back was derived, closed-form expressions for the channel estimation error at the transmitter and at the receiver were provided. The last part of the chapter studied the quantization of the feedback link. Theoretical results of the minimum distortion produced when quantizing the information with a certain number of bits, as well as practical quantization procedure were detailed.

6.2 Power allocation algorithms

Minimum MSE. The design of a robust power allocation strategy that minimizes the MSE was covered in Chapter 4. This was one of the first criterion used in the literature to design linear transmitters. It has been widely studied when perfect CSI is available at the transmitter but no when CSI was degraded. Chapter 4, hence, extended the derivation of this criterion when only imperfect CSI is available, proposing a Bayesian approach to design the linear transformation at the transmitter assuming a ML detector at the receiver.

A detailed study of the robust objective function manifested that the optimization procedure takes into account, apart to the noise and the ISI, the mismatch between the real and the estimated channel due to the imperfect CSI. A closed form solution was found based on a set of linear equations whose complexity is comparable with the nonrobust solution.

The analysis of this power allocation strategy showed that this optimization criterion assigns more power to the strongest channel modes and penalizes the deepest ones. This way to allocate the power explains the behavior of the algorithm both in terms of uncoded BER and coded BER. As it was shown, in terms of uncoded BER the algorithm exhibits a poor performance mainly due to the decomposition of the channel into a set of parallel multiplicative subchannels that results in a loss of space diversity. This drawback was alleviated introducing a unitary matrix that spreads the transmitted symbols over all the subcarriers. In the presence of this unitary matrix the performance of the algorithm in terms of uncoded BER was shown to be similar to that one of the algorithms designed to minimize the uncoded BER with the benefit of have a solution substantially simpler.

Despite this good performance in terms of uncoded BER, we cannot conclude that the minimization of the MSE is, in general, a good optimization criterion, specially in MIMO channels, because of several reasons:

- In terms of coded BER the results in Chapter 5 showed that the algorithm is significantly outperformed by other optimization criteria like the minimization of the uncoded BER or the maximization of the cut-off rate.

- Unlike other algorithms studied in this thesis, the algorithm that minimizes the MSE does not tends to the open-loop solution (i.e., the same power is allocated to all the subcarriers and antennas) when the channel uncertainty grows. Hence, when the transmitter only disposes of a poor CSI, the algorithm is highly degraded and it can perform worse that the open-loop solution as it was shown in some simulation results of Chapter 5.

- The poor performance of the algorithm at high SNR's and MIMO channels was also evidenced when the performance of the algorithm was evaluated in terms of theoretic information parameters (cut-off rate or capacity). As it was shown in Figure 5.5 and Table 5.5 the behavior of the algorithm that minimizes the MSE is worse that the rest of the power allocation strategies studied in the thesis.

Minimum uncoded BER. The design of linear transmitters based on the minimization of the uncoded BER was proposed in Chapter 4. The first drawback that we found was the presence of the not easily tractable Gaussian Q-function $\mathcal{Q}(\sqrt{x})$ in the objective function. As a solution, two different algorithms were proposed based on the approximation of that error function by an exponential function. A first solution used the well known Chernoff bound while a second solution was based on a tighter approximation of the $\mathcal{Q}(\sqrt{x})$ function. In both cases a closed form solution was derived for the beamforming case while an iterative solution was required for the spatial multiplexing. Similarities and differences between both algorithms were displayed by numerical results. In terms of uncoded BER it was shown that both solutions have similar performances at high SNR's while they perform in a different way in noisy environments. This behavior was explained after analyzing the approximation of the $\mathcal{Q}(\sqrt{x})$ function and the power allocation policy of both algorithms.

In terms of CSI quality at the transmitter it was shown that, contrary to the algorithm that minimizes the MSE, the algorithms that minimize the uncoded BER provide a solution that converges to the open-loop design (i.e.,the same power is allocated across all antennas and subcarriers) for the case of no channel knowledge and to the closed-loop design with perfect CSI for the case of no uncertainty.

Somewhat surprising was the performance of the algorithm based on the Chernoff bound in terms of coded BER. Numerical results showed that the algorithm had similar performances when compared with algorithms that were explicitly designed to minimize the coded BER. In Chapter 5 this algorithm was studied in detail and its excellent behavior in terms of coded BER was justified from different points of view.

The main conclusions that can be drawn for these algorithms are:

- The algorithm based on the Chernoff bound is preferred against the algorithm that uses a tight approximation of the $\mathcal{Q}\left(\sqrt{x}\right)$ function. Although the second one slightly outperforms the first one in terms of uncoded BER the small differences are only appreciable at low SNR's. On the contrary, in terms of coded BER the second one highly degrades its performance when compared with the first one.
- The algorithm based on the Chernoff bound has an excellent behavior either in terms of uncoded BER and in terms of coded BER.
- Despite its good performance, it does not results an interesting algorithm (when compared with other algorithms like the algorithm that maximizes the cut-off rate) since an iterative algorithm is, in general, required to found the optimal power allocation.

Maximum cut-off rate. The cut-off rate was proposed in Chapter 5 as a power allocation algorithm whose maximization was directly related with the coded BER. This criterion was introduced as an alternative to the channel capacity and the mutual information for the design of optimal transceivers in the presence of any channel coding stage. An interpretation of the objective function based on the Kullback-Leibler pseudo-Distance, as well as an accurate analysis of that function, evidenced that its maximization is related with the minimization of the pairwise error probability taking into account the CSI reliability. Because the original cost function did not allowed an analytical treatment, a quadratic approximation based on a Taylor expansion was proposed to derive the cost function. Numerical results confirmed the correspondence between the maximization of the cut-off rate and the satisfactory performances in terms of coded BER. It was shown that the power allocation strategy followed by the algorithm that maximizes the cut-off rate allocates the majority of the power to the best subcarriers, neglecting the weakest ones. This power allocation policy can be understood in terms of coded BER noting that the transmitter does not takes care of the symbols transmitted in the more fades subcarriers, assuming that this information will be recovered at the receiver thanks to the channel code. This strategy involves a certain amount of risk since the transmitter design has not been optimized for a specific channel code. In consequence, when a particular code is employed, the power allocation policy previously described could result in a unsatisfactory performance if the channel code is

not able to correct the large number of damaged symbols. To confirm this situation Section 5.7 simulated a specific scenario in which the algorithm that maximizes the cut-off rate was outperformed by other algorithms. This undesired behavior was, however, solved by introducing an adaptive interleaver.

Maximum mutual information. Apart from the maximization of the cut-off rate, Chapter 5 studied other information theoretic criteria for the design of the power allocation algorithms. Specifically, the mutual information constrained to QPSK and capacity (mutual information assuming Gaussian symbols) were considered. However, to the best of our knowledge, none of these criteria become feasible for the design of robust power allocation algorithms. For the mutual information constrained to QPSK constellation only an iterative algorithm, whose complexity grows with the number of symbols transmitted per subcarrier, was derived. This solution, hence, was only used to justify the behavior of the rest of the algorithms. On the other hand, for the capacity a closed form solution was derived only when perfect CSI is available. Its performance, however, was shown to be unsatisfactory in terms of coded BER. This poor performance could be explained because a specific constellation (i.e., QPSK) was used, whereas the algorithm was designed assuming Gaussian symbols.

Adaptive interleaver. The last part of Chapter 5 proposed a low complexity adaptive interleaver that making use of the CSI available at the transmitter, reallocated the bits not only to combat the bursty channel errors but also to combat the specific distribution of the faded subcarriers. The design of this interleaver, named *'RCPC interleaver'*, was based on the *Rate-Compatible Punctured Convolutional Codes* (RCPC codes). Simulation results showed, as expected, that the performance of all the algorithms was improved in the presence of the new interleaver, which results an interesting alternative to the classical block interleavers and pseudo-random interleavers when CSI is available at the transmitter. It was also shown that the larger gain was experiment in those algorithms that null more subcarriers, in which we find the algorithm that maximizes the cut-off rate.

	Robust?	Closed-form solution?	High uncertainty	Uncoded BER	Coded BER
MSE	Yes	Yes	High degradation	3rd	4th
Unc. BER (Chernoff)	Yes	Only for beamforming	Open-loop	2nd	2nd
Coded BER (Approx Q)	Yes	Only for beamforming	Open-loop	1st	3rd
Cut-off rate	Yes	Yes (Taylor expansion)	Open-loop	-	1st
MI Gaussian	No	No	-	4th	5th

Summary. The table summarizes and compares the main characteristics of the algorithms studied in this thesis. The column labelled as *Robust* indicates if robust to CSI solution has been found. The column labelled as *Closed-form solution* refers to the disposal of a closed-form solution to design the power allocation strategy. The third column, labelled as *High uncertainty* summarizes the performance of the algorithm when the CSI quality degrades: Open-loop reveals that the power allocation policy tends to the open-loop solution. Finally, the algorithms are sorted as a function of its performance both in terms of uncoded BER and coded BER.

6.3 Other future research topics

Besides the specific research topics that have already been outlined in the previous section, other more general research lines, beyond the scope of this thesis, can be devised and listed below. The two first items are related with the first part of the thesis, which proposed the transmitter channel tracking scheme, whereas the rest of the topics make reference to the study of the power allocation strategies.

1. *Introduce constraints or external conditions in the feedback channel.*

   Chapter 3 studied the minimum amount of information to be fed back; the way to quantize this information; the distortion produced when quantizing the data with a certain number of bits; and its impact on the channel estimation error at the transmitter. This information was used to set the return link channel throughput assuming that all the channel coefficients were updated with the same number of bits at each time interval. Alternatively, the number of bits could be modified either using more bits to quantize certain channel coefficients, or freezing the update of some channel coefficients during a certain lag of time. This procedure would allow, for example, a jointly design of the channel estimation error at the transmitter and the throughput at the return link, given some external constraints or cost/benefit conditions.

2. *Feedback the minimum amount of information required by the transmitter to design the linear transformation* $\mathbf{F}$.

   Chapter 3 proposed an algorithm to track the channel impulse response at the transmitter. A detailed analysis of the use of this channel response at the transmitter manifests, however, that not all the information contained in the channel response is used to design the linear transformation matrix. A fast reviewing of the algorithms studied in this thesis shows that the dependence of the cost function with respect to the channel is given by the hermitian matrix $\mathbf{H}^H\mathbf{H}$. This result, in accordance with the first works in [Hay68] and the results in [Fan04], evidences that only the channel correlation and not the whole channel impulse response is required. Moreover, the hermitian structure of this matrix would allow to reduce the amount of information to be fed back.

3. *Analyze the behavior of the algorithms proposed in this thesis in the presence of other channel codes different that the convolutional code.*

   As it was said in Chapter 5, in the presence of iterative decoding schemes, the cut-off rate cannot be considered anymore an upper bound on the practical achievable rates. Hence, the results of the different algorithms, specially that one that allocates the power under the maximization of the cut-off rate, cannot be conclusive when other codes are used. An immediate research line should be, thus, to review the conclusions drawn for the different algorithm by new simulations with other codes like turbo codes or LDPC codes.

4. *Determine the best design criterion for the minimization of the coded BER or PER.*

   Several criteria were analyzed in Chapter 5 to define the power allocation policy under a minimum coded BER or PER criterion when the transmitter has perfect or partial knowledge of the channel response. However, as none of the optimization criteria studied in the Chapter has a direct relationship with the ultimate goal of minimizing the coded BER or PER, the discussion on what is the best design criterion remains still unclear until the quality of each parameter as a PER indicator will be analyzed.

   In [Lam05] the name of *PER indicator* was used to refer to a metric or parameter that summarizes in a single value the dependence of all the parameters that have influence on the coded BER or PER (e.g., the Eb/No, the channel statistics like Doppler or delay spread, the specific channel impulse response, ...). Under this definition, an indicator is said to be reliable when there is a one-to-one relationship between the parameter and the coded BER or PER. The works that in the literature study those indicators [Lam02, Ale02, Lam05] evaluate its reliability by simulations, plotting the correspondence between the simulated coded BER or PER and the indicator values, and measuring somehow the fuzziness of the curve (e.g, by means of visual criteria, variance measures or mutual information measures).

Note that the study of different power allocation strategies as quality indicators of Coded BER or PER could determine a systematic procedure to conclude what is the best criterion to optimize the coded BER or PER for a given scenario: if there is a one-to-one relationship between the coded BER or PER and the parameter, the minimization of the first could be obtained by minimizing or maximizing the corresponding metric. The election of the best indicator, however, remains still open because all the parameters that have been studied, mainly those proposed in this thesis for the design of power allocation strategies, exhibit some deficiencies in their ability to predict the coded BER or PER for certain scenarios and there is no clear criteria for selecting the most appropriate.

The relevance of this line of research goes beyond the scope of this thesis since the selection of a reliable PER indicator could be used not only to design power allocation policies, but also to solve link adaptation problems [Lam02] or to set an interface between physical layer and system level simulations that will allow to significantly reduce the complexity of the simulations [IST03, Ale02].

5. *Use the Gallager function as alternative to the cut-off rate.*

   The Gallager's random coding bound for block codes and for a discrete memoryless channel states as follows [Gal65]:

   > Consider a discrete memoryless channel with an input alphabet of K symbols x_i for $i = 1 \dots K$; an output alphabet of J symbols: x_j for $j = 1 \dots J$; and a transition probabilities $P(y_j \mid x_i)$. For any block length N, any number of codewords $M = 2^{NR}$, and any probability distribution on the use of the code words, the average block error probability of these codes for a ML decoding satisfies:
   >
   > $$E_{\mathcal{C}}\left\{P_w(e)\right\} \leq 2^{-N(E_0(\rho,Q)-\rho R_c)}$$
   >
   > where $E_0(\rho, Q)$, denoted as the Gallager function, is given by:
   >
   > $$E_0(\rho, Q) = -\log_2 \sum_j \left(\sum_i P(y_j \mid x_i)^{\frac{1}{1+\rho}} Q_i \right)^{1+\rho},$$
   >
   > ρ is an arbitrary number $0 \leq \rho \leq 1$, and Q_i is an arbitrary probability value $0 \leq Q_i \leq 1$.

   In particular, the tightest bound for the error probability can be found by minimizing the exponent over ρ and Q_i. Hence:

   $$E_{\mathcal{C}}\left\{P_w(e)\right\} \leq 2^{-NE_G(R_c)}$$

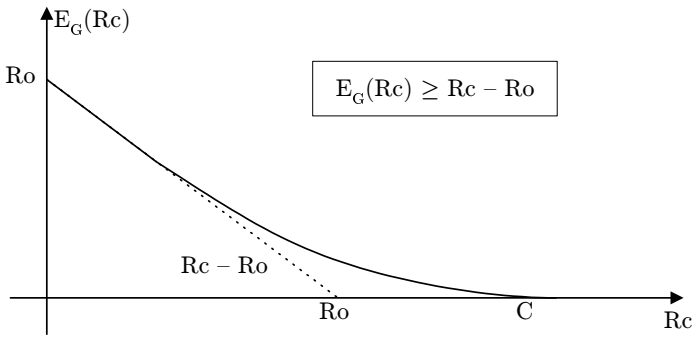
   where the exponent function, denoted as the Gallager exponent, is given by:

   $$E_G(R_c) = \max_{\rho} \max_{Q_1 \dots Q_K} \left\{E_0(\rho, Q) - \rho R_c\right\}$$

The cut-off R_o rate can be devised as a particular case of the Gallager function for $\rho = 1$:

$$R_o = E_G(0) = \max_{Q_1 \cdots Q_K} \{E_0(1, Q)\}$$

that satisfies:



Notice that while the cut-off rate yields a simpler exponential upper bound of the word-error rate averaged over the ensemble of block codes, the Gallager function yields a tighter bound. Hence, if a solution could be found to manipulate the Gallager function, a tighter bound for the the averaged block error probability could be used. Moreover, other bounds like the expurgated exponent could also be interesting for further research.

6. *Apply the connections between estimation theory and information theory to the power allocation strategies.*

   Recent works published by Verdú et al. showed a simple connection between the mutual information and the MMSE [Guo05]. This novel result may open a line of research and new interpretations for the design of power allocation strategies. A first example can be found in [Loz05] where the authors derive a power allocation policy, named mercury/waterfilling, that maximizes the sum mutual information over parallel channels with arbitrary input constellations.

7. *Combine the power allocation strategies with the adaptive coding and modulation schemes.*

   In this thesis the system has been constrained to keep the same constellation and channel code for all the subcarriers, modifying only the amount of power allocated to each channel mode. Alternatively, an unconstrained system could be constrained at expenses of a higher complexity. This more general scheme, which would allow to adapt not only the power but also the constellation and channel code, would allow to combine the discrete and the continuous adaptation schemes (see Section 2.2.2) to fully adapt the transmission rate to the channel capacity as a function of the CSI. Moreover, the extension of the single-user problem to a multi-user scenario, and the extension of the point-to-point connection to ad hoc networks, would open up a whole new range of areas of potential future research that

could include, among others, scheduling and routing schemes aimed to reach the maximum achievable rate by the network.

Observing the natural evolution of some standards, for example the Digital Video Broadcasting (DVB) standards that has evolved from the *Constant Coding and Modulation* (CCM) [DVBb] to the *Adaptive Coding and Modulation* (ACM) [DVBc], we can expect that this topic of research will be of interest in a next future since it could set the basis for future communication standards.

Bibliography

[Agu03] A. Aguilar, H. Karl, A. Wolisz, H. Miesmer, "A Framework for Evaluating Effects of Channel Prediction Inaccuracy on the Performance of Channels Adaptive Techniques", *Proceedings of International Conference on Wireless Networks 2003*, pags. 3–9, Las Vegas (Nevada - USA), Jun. 2003.

[Ala98] S.M. Alamouti, "A Transmit Diversity Technique for Wireless Communications", *IEEE Journal on Selected Areas in Communications*, Vol. 16, nº 8, pags. 1451–1458, Oct. 1998.

[Ale02] A. Alexiou, R. Karimi, K. Peppas, F. Lazarakis, D. Bourse, "System level simulation methodology for 4G wireless communication systems enhanced through space-time techniques", *7th WWRF Meeting*, Eindhoven (The Netherlands), Dec. 2002, available at http://www.telecom.ece.ntua.gr/fitness.

[And79] B.D.O. Anderson, J.B. Moore, *Optimal Filtering*, Prentice Hall, 1979.

[And99] J.B. Andersen, J. Jensen, S.H. Jensen, F. Frederiksen, "Prediction of Future Fading Based on Past Measurements", *Proceedings of VTC-fall'99*, pags. 151–155, Amsterdam, (The Netherlands), Sep. 1999.

[Aul79] T. Aulin, "A modified model for the fading signal at a mobile radio channel", *IEEE Trans. on Vehicular Technology*, Vol. 28, nº 3, pags. 182–203, Aug. 1979.

[Bar00] S. Barbarossa, A. Scaglione, "Time-Varying Channels", G. B. Giannakis, Y. Hua, P. Stoica, L. Tong (eds.), *Signal Processing Advances in Wireless Communications*, Vol. II: Trends in Single- and Multi-User Systems, chap. 1, Prentice-Hall, 2000.

[Bar02] I. Barhumi, G. Leus, M. Moonen, "Optimal Training Sequences for Channel Estimation in MIMO-OFDM Systems in Mobile Wireless Channels", *Proceedings of Int. Zurich Seminar on Broadband Communications*, Zurich (Switzerland), Feb. 2002.

[Bel63] P.A. Bello, "Characterization of Randomly Time-Variant Linear Channels", *IEEE Trans. on Circuits and Systems*, Vol. 11, nº 4, pags. 360–393, Dec. 1963.

[Bel00] K.L. Bell, Y. Ephraim, H.L.V. Trees, "A Bayesian Approach to Robust Adaptive Beamforming", *IEEE Trans. on Signal Processing*, Vol. 48, nº 2, pags. 386–398, Feb. 2000.

[Ben02] N. Benvenuto, S. Tomasin, "Efficient Pre-coding Schemes for FMT Broadband Wireless Systems", *Proceedings of PIMRC'02*, Lisboa, (Portugal), Sep. 2002.

[Big98] E. Biglieri, J. Proakis, S. Shamai, "Fading Channels: Information-Theoretic and Communications Aspects", *IEEE Trans. on Information Theory*, Vol. 44, nº 6, pags. 2619–2692, Oct. 1998.

[Bou98] J. Boutros, E. Viterbo, "Signal Space Diversity: A Power- and Bandwidth- Efficient Diversity Technique for the Rayleigh Fading Channel", *IEEE Trans. on Information Theory*, Vol. 44, nº 4, pags. 1453–1467, Jul. 1998.

[Boy04] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[Cai99] C. Caire, S. Shamai, "On te Capacity of Some Channels with Channel State Information", *IEEE Trans. on Information Theory*, Vol. 45, nº 9, pags. 2007–2019, Sep. 1999.

[Cai05] X. Cai, G.B. Giannakis, "Adaptive PSAM Accounting for Channel Estimation and Prediction Errors", *IEEE Trans. on Wireless Communications*, Vol. 4, nº 1, pags. 246–256, Jan. 2005.

[Cav91] J.K. Cavers, "An Analisys of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels", *IEEE Trans. on Vehicular Technology*, Vol. 40, nº 4, pags. 686–693, Nov. 1991.

[Cav00] J.K. Cavers, "Single-User and Multiuser Adaptive Maximal Ratio transmission for Rayleigh Channels", *IEEE Trans. on Vehicular Technology*, Vol. 49, nº 6, pags. 2043–2050, Nov. 2000.

[Cha02] S. Chan, T.N. Davidson, K.M.Wong, "Assymptotically Minimum Bit Error Rate Block Precoders for Minimum Mean Square Error Equalization", *Proceedings of SAM'02*, pags. 140–144, Rosslyn (VA-USA), Aug. 2002.

[Cho02] J.P. Choi, V.W.S. Chan, "Predicting and adapting satellite channels with weather-induced impairments", *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 38, nº 3, pags. 779–790, Jul. 2002.

[Cla68] R.H. Clarke, "A statistical theory of mobile radio reception", *Bell. Syst. Tech. Journal,*, Vol. 47, nº 6, pags. 957–1000, July/Aug. 1968.

[Cla97] R.H. Clarke, W.L. Khoo, "3-D Mobile Radio Channel Statistics", *IEEE Trans. on Vehicular Technology*, Vol. 46, nº 3, pags. 798–799, Aug. 1997.

[Col02] S. Coleri, M. Ergen, A. Puri, A. Bahai, "Channel Estimation Techniques Based on Pilot Arrangement in OFDM Systems", *IEEE Trans. on Broadcasting*, Vol. 48, nº 3, pags. 223–229, Sep. 2002.

[Cos98] D.J. Costello, J. Hagenauer, H. Imai, S.B. Wicker, "Applications of Error-Control Coding", *IEEE Trans. on Information Theory*, Vol. 44, nº 6, pags. 2531–2560, Oct. 1998.

[Cov91] T. Cover, J.A.Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[Cra91] J.W. Craig, "A new, imple and exact results for calculating the probability of error for two-dimensional signal constellations", *Proceedings of MILCOM 1991*, pags. 25.5.1–25.5.5, Boston-MA (USA), Nov. 1991.

[DH00] A. Duel-Hallen, S. Hu, H. Hallen, "Long-Range Prediction of Fading Signals", *IEEE Signal Processing Magazine*, Vol. 17, nº 3, pags. 62–75, May. 2000.

[Dha96] Naofal Al Dhahir, 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, pags. 137–160, Jan. 1996.

[Din03] Y. Ding, T.N. Davidson, Z Luo, K.M.Wong, "Minimum BER Block Precoders for Zero-Forcing Equalization", *IEEE Trans. on Signal Processing*, Vol. 51, nº 9, pags. 2410–2423, Sep. 2003.

[Don01] L. Dong, G. Xu, H. Ling, "Prediction of Fast Fading Mobile Radio Channels in Wideband Communication Systems", *Proceedings of GLOBECOM'01*, pags. 3287–3291, San Antonio (TX-USA), Nov. 2001.

[DVBa] DVBRCS, "ETSI EN 301 790 V1.2.2 (2000-12), Digital Video Broadcasting (DVB). Interaction channel for satellite distribution systems.", *European Telecomm. Standards Institute (ETSI)*.

[DVBb] DVBS, "ETSI EN 300 421 (V.1.1.2): "Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for 11/12 GHz satellite services.", *European Telecomm. Standards Institute (ETSI)*.

[DVBc] DVBS2, "Draft ETSI EN 302 307 v1.1.1 (2004-06), Digital Video Broadcasting (DVB). Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications.", *European Telecomm. Standards Institute (ETSI)*.

[Ekm99a] T. Ekman, G. Kubin, "Nonlinear Prediction of Mobile Radio Channels: Measurements and MARS model designs", *Proceedings of ICASSP'99*, pags. 2667–2670, Phoenix, (AZ-USA), Mar. 1999.

[Ekm99b] T. Ekman, G. Kubin, M. Sternad, A. Ahlén, "Quadratic and Linear Filters for Radio Channel Prediction", *Proceedings of VTC-fall'99*, pags. 146–150, Amsterdam, (The Netherlands), Sep. 1999.

[Ekm01] T. Ekman, *Prediction of Mobile Radio Channels*, PhD Thesis, Uppsala University, Upsala, Sweeden, Oct. 2001.

[Ekm02] T. Ekman, M. Sternad, A. Ahlén, "Unbiased Power Prediction of Rayleigh Fading Channels", *Proceedings of VTC-fall'02*, pags. 280–284, Vancouver, (Canada), Sep. 2002.

[Erm04] N. Ermolova, S.G. Haggman, "Simplified Bounds for the Complamenatry Error Function; Application to the Performance Evaluation of Signal-Processing Systems", *Proceedings of EUSIPCO 2004*, pags. 1087–1090, Vienna (Austria), Sep. 2004.

[Etk03] R. Etkin, D. Tse, "Degrees of Freedom in Underspread MIMO Fading Channels", *Submitted to the IEEE Transactions on Information Theory*, (submitted May 2003), (See also "Degrees of Freedom in Underspread MIMO Fading Channels", in Proceedings of ISIT'03, pag. 323, Yokohama, (Japan), Jun. 2003).

[Eyc98] T. Eyceoz, A. Duel-Hallen, H. Hallen, "Deterministic Channel Modeling and Long Range Prediction of Fast Tading Mobile Radio Channels", *IEEE Commuications Letters*, Vol. 2, nº 9, pags. 254–256, Sep. 1998.

[Fal04] S. Falahati, A. Svensson, T. Ekman, M. Sternad, "Adaptive Modulation Systems for Predicted Wireless Channels", *IEEE Trans. on Communications*, Vol. 52, nº 2, pags. 307–316, Feb. 2004.

[Fan04] Z. Fan, L. Scharf, J.A. Gubner, "Analog Precoder and Equalizer Design for Multichannel Communication", *Proceedings of SPAWC'04*, pags. 130–134, Lisbon (Portugal), Jul. 2004.

[Fis02] R.F.H. Fischer, C. Windpassinger, A. Lampe, J. Huber, "Tomlinson-Harashima Precoding in Space-Time Transmission for Low-Rate Backward Channel", *Proceedings of International Zurich Seminar on Broadband Communications (IZS '02)*, Zurich, (Switzerland), Feb. 2002.

[Fos96] G. J. Foschini, "Layered Space-Time Architecture for Wireless Communications in a Fading Environment when using Multi-element Antennas", *Bell Labs Tech. J*, pags. 41–59, Autumn 1996.

[Fos98] G.J. Foschini, M.J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas", *Wireless Personal Communications*, , nº 6, pags. 315–335, 1998.

[Fos99] G. J. Foschini, G. D. Golden, R. A. Valenzuela, P. W. Wolnianski, "Simplified Processing for High Spectral Efficiency Wireless Communication Employing Multi-element Arrays", *IEEE Journal on Selected Areas in Communications*, Vol. 17, nº 11, pags. 1841–1852, Nov. 1999.

[Fro72] O.L. Frost, "An Algorithm for Linearly Constrained Adaptive Array Processing", *Proc. of IEEE*, Vol. 60, nº 8, pags. 926–935, Aug. 1972.

[Gal65] R.G. Gallager, "A Simple Derivation of teh Coding Theorem and Some Applications", *IEEE Trans. on Information Theory*, Vol. 11, nº 1, pags. 3–18, Jan. 1965.

[Gal68] R.G. Gallager, *Information Theory and Reliable Communications*, John Wiley, 1968.

[Gan02a] G. Ganesan, P. Stoica, "Differential Modulation Using Space-Time Block Codes", *IEEE Signal Processing Letters*, Vol. 9, nº 2, pags. 57–60, Feb. 2002.

[Gan02b] G. Ganesan, P. Stoica, E.G. Larsson, "Diagonally Weighted Orthogonal Space-time Block Codes", *Proceedings of 36th Asilomar Conf. Signals, Syst. Comput.*, pags. 1147–1151, Pacific Grove (CA-USA), Nov. 2002.

[Gao96] X.M. Gao, J.M.A. Tanskanen, S.J. Ovaska, "Comparison of Linear and Neuronal Network-based Power Prediction Schemes for Mobile DS/CDMA Systems", *Proceedings of VTC'96*, pags. 61–65, Atlanta, (GA-USA), May 1996.

[Gao97] X.M. Gao, X.Z. Gao, J.M.A. Tanskanen, S.J. Ovaska, "Power Control for Mobile DS/CDMA Systems using a Modified Elman Neuronal Network", *Proceedings of VTC'97*, pags. 750–754, Phoenix, (AZ-USA), May. 1997.

[Ges03] D. Gesbert, M. Shafi, D. Shiu, P.J. Smith, A. Naguib, "From Theory to Practice: An Overview of MIMO Space-Time Coded Wireless Systems", *IEEE Journal on Selected Areas in Communications*, Vol. 21, nº 3, pags. 281–302, Apr. 2003.

[Gia98] G.B. Giannakis, C. Tepedelenlioğlu, "Basis Expansion Models and Diversity Techniques for Blind Identification and Equalization of Time-Varying Channels", *Proceedigs of the IEEE*, Vol. 86, nº 10, pags. 1969–1986, Oct. 1998.

[Gol97a] A. Goldsmith, P.P. Varaiya, "Capacity of Fading Channels with Channel Side Information", *IEEE Trans. on Information Theory*, Vol. 43, nº 6, pags. 1986–1992, Nov. 1997.

[Gol97b] A.J. Goldsmith, S. Chua, "Variable-Rate Variable-Power MQAM for Fading Channels", *IEEE Trans. on Communications*, Vol. 45, nº 10, pags. 1218–1230, Oct. 1997.

[Gol98] A.J. Goldsmith, S. Chua, "Adaptive Coded Modulation for Fading Channels", *IEEE Trans. on Communications*, Vol. 46, nº 5, pags. 595–602, May 1998.

[Gre92] D.J. Grenwood, L. Hanzo, *Characterization of Mobile Radio Channels*, Chapter 2 in Mobile Radio Communications. Edited by R. Steele. John Wiley & Sons, 1992.

[Gre99] B. Gremont, M. Filip, P. Gallois, S. Bate, "Comparative analysis and performance of two predictive fade detection schemes for Ka-band fade countermeasures", *IEEE Journal on Selected Areas in Communications*, Vol. 17, nº 2, pags. 180–192, Feb. 1999.

[Gui04] M. Guillaud, D.T.M. Slock, "A Specular Approach to MIMO Frequency-Selective Channel Tracking and Prediction", *Proceedings of SPAWC'04*, pags. –, Lisbon (Portugal), Jul. 2004.

[Guo05] D. Guo, S. Shamai, S. Verdú, "Mutual Information and Minimum Mean-square Error in Gaussian Channels", *IEEE Trans. on Information Theory*, Vol. 51, nº 4, pags. 1261–1283, Apr. 2005.

[Hag88] J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and their Applications", *IEEE Trans. on Communications*, Vol. 36, nº 4, pags. 389–400, Apr. 1988.

[Har72] H. Harashima, H. Miyakawa, "Matched-Transmission Technique for Channels with Intersymbol Interference", *IEEE Trans. on Communications*, Vol. 20, nº 4, pags. 774–780, Aug. 1972.

[Hay68] J.F. Hayes, "Adaptive Feedback Communications", *IEEE Trans. on Communications*, Vol. 16, nº 1, pags. 29–34, Feb. 1968.

[Hay96] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 3rd ed., 1996.

[Her01] A.O. Hero, T.L. Marzetta, "Cutoff Rate and Signal design for the Quasi-Static Rayleigh-Fading Space-Time Channel", *IEEE Trans. on Information Theory*, Vol. 47, nº 6, pags. 2400–2416, Sep. 2001.

[HL200] HL2, "ETSI TS 101 761-1 v.1.1.1 - Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer. Part 1: Basic Data Transport Functions", *European Telecomm. Standards Institute (ETSI)*, 2000.

[HL201] HL2, "ETSI TS 101 475 v1.2.2 - Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) layer", *European Telecomm. Standards Institute (ETSI)*, 2001.

[Hoc00a] B. M. Hochwald, T. L. Marzetta, "Unitary Space-Time Modulation for Multiple Antenna Communications in Rayleigh Flat Fading", *IEEE Trans. on Information Theory*, Vol. 46, nº 2, pags. 543–564, Mar. 2000.

[Hoc00b] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldons, R. Urbanke, "Systematic Design of Unitary Space-Time Constellation", *IEEE Trans. on Information Theory*, Vol. 46, nº 6, pags. 1961–1973, Sep. 2000.

[Hoc03] B.H. Hochwald, S. Brink, "Achieving Near-Capcity on a Multiple-Antenna Channel", *IEEE Trans. on Communications*, Vol. 51, nº 3, pags. 389–399, Mar. 2003.

[Hon92] M.L. Honig, P. Crespo, 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, pags. 614–629, Apr. 1992.

[Hug00] B. L. Hughes, "Differential Space-Time Modulation", *IEEE Trans. on Information Theory*, Vol. 46, nº 7, pags. 145–149, Nov. 2000.

[Hwa98] J.K. Hwang, J.H. Winters, "Sinusoidal Modeling and Prediction of Fast Fading Processes", *Proceedings of GLOBECOM'98*, pags. 882–897, Sydney, (Australia), Nov. 1998.

[IEE99] IEEE802.11, "IEEE Std 802.11a-1999. Supplement to IEEE Standard for Information technology. Part 11: Wireless LAN Medium Access Control(MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5 GHZ Band", *Institute of Electrical and Electronics Engineers,(IEEE)*, 1999.

[Ilt90] R.A. Iltis, "Joint Estimation on PN Code Delay and Multipath Using the Extended Kalman Filter", *IEEE Trans. on Communications*, Vol. 38, nº 10, pags. 1677–1685, Oct. 1990.

[Ilt91] R.A. Iltis, A.W. Fuxjaeger, "A Digital DS Spread-Spectrum Receiver with Joint Channel and Doppler Shift Estimation", *IEEE Trans. on Communications*, Vol. 39, nº 8, pags. 1255–1267, Aug. 1991.

[IST02] FITNESS IST-2000-30116, "Performance analysis of re-configurable MTMR transceivers for WLAN", *Deliverable D3.2.1 available at http://www.telecom.ece.ntua.gr/fitness*, Nov. 2002.

[IST03] FITNESS IST-2000-30116, "MTMR Baseband Transceivers Needs for Intra-system and Inter-system (UMTS/WLAN) Reconfigurability", *Deliverable D3.3.1 available at http://www.telecom.ece.ntua.gr/fitness*, Nov. 2003.

[Jac74] W.C. Jackes, *Microwave Mobile Communications*, New York: IEEE Press, 1974.

[Jaf01a] S.A. Jafarand, S. Vishwanath, A. Goldsmith, "Channel Capacity and Beamforming for Multiple Tranmsit and Receive Antennas with Covariance Feedback", *Proceedings of ICC'01*, pags. 2266–2270, Helsinki (Finland), Jun. 2001.

[Jaf01b] H. Jafarkhani, "A Quasi-Orthogonal Space-Time Block Code", *IEEE Trans. on Communications*, Vol. 49, nº 1, pags. 1–4, Jan. 2001.

[Jan98] W.M. Jang, B.R. Vojcic, R.L. Pickholtz, "Joint Transmitter-Receiver Optimzation in Synchronous Multiuser Communications over Multipath Channels", *IEEE Trans. on Communications*, Vol. 46, nº 2, pags. 269–278, Feb. 1998.

[Jön02] G. Jöngren, M. Skoglund, B. Ottersten, "Combining Beamforming and Orthogonal Space-Time Block Coding", *IEEE Trans. on Information Theory*, Vol. 48, nº 3, pags. 611–627, Mar. 2002.

[Jöt02] C.A. Jötten, P.W. Baier, M. Meurer nad T. Weber, M. Haardt, "Efficient Representation and Feedback Signaling of Channel State Information in Frequency Division Duplexing MIMO Systems", *Proceedings of 5th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pags. 444–448, Honolulu (HI-USA), Oct. 2002.

[Kas83] S.A. Kassam, H.V. Poor, "Robust Signal Processing for Communication Systems", *IEEE Communications Magazine*, Vol. 21, nº 1, pags. 20–28, Jan. 1983.

[Kas85] S.A. Kassam, H.V. Poor, "Robust Techniques for Signal Processing: A Survey", *Proceedings of the IEEE*, Vol. 73, nº 3, pags. 433–481, Mar. 1985.

[Kay93] Steven M. Kay, *Fundamentals of Statistical Signal Processing:Estimation Theory*, Prentice Hall International Ed., 1993.

[Kos00] C. Kose, L. Goeckel, "On Power Adaptation in Adaptive Signaling Systems", *IEEE Trans. on Communications*, Vol. 48, nº 11, pags. 1769–1773, Nov. 2000.

[Kru04] Z.B. Krusevac, P.B. Rapajic, R.A. Kennedy, "On Channel Uncertainty Modeling - An Information Theoretic Approach", *Proceedings of Globecom'04*, pags. 410–414, Dallas (TX-USA), Dec. 2004.

[Kyr02] P. Kyritsi, R. A. Valenzuela, D.C. Cox, "Channel and Capacity Estimation Errors", *IEEE Communications Letters*, Vol. 6, nº 12, pags. 517–519, Dec. 2002.

[Lam97] M. Lamarca, G. Vazquez, "Transform Modulations for Mobile Communications", *Proceedings of PIMRC'97*, pags. 462–466, Helsinki (Finland), Sep. 1997.

[Lam02] M. Lampe, H. Rohling, J. Eichinger, "PER prediction for link adaptation in OFDM systems", *OFDM Workshop*, Hamburg (Germany), Sep. 2002.

[Lam03] M. Lamarca, J.A. Lopez-Salcedo, "Decoding Algorithms for Reconfigurable Space-Time Turbo Codes", *Proceedings of ICASSP'03*, pags. 129–132, Hong Kong (China), Apr. 2003.

[Lam05] M. Lamarca, F. Rey, "Indicators for PER prediction in wireless systems: A comparative study", *Proceedings of VTC-Spring'05*, Stockholm (Sweeden), Jun. 2005.

[Lap98] A. Lapidoth, P. Narayan, "Reliable Communication Under Channel Uncertainty", *IEEE Trans. on Information Theory*, Vol. 44, nº 6, pags. 2148–2177, Oct. 1998.

[Lap02] A. Lapidoth, S. Shamai, "Fading Channels: How Perfect Need "Perfect Side Information" Be?", *IEEE Trans. on Information Theory*, Vol. 48, nº 5, pags. 1118–1134, May. 2002.

[Lap03] A. Lapidoth, S. Moser, "Capacity Bounds via Duality with Applications to Multiple-Antenna Systems on Flat-Fading Channels", *IEEE Trans. on Information Theory*, Vol. 49, nº 10, pags. 2426–2467, Oct. 2003.

[Lar01] E. Larsson, J. Li, "Preamble Design for Multiple-Antenna OFDM-Based WLANs With Null Sub-carriers", *IEEE Signal Processing Letters*, Vol. 8, nº 11, pags. 599–602, Nov. 2001.

[Lei02] Z. Lei, F.P.S. Chin, Y.C. Liang, "Combined Beamforming with Space-time Block Coding for Wireless Downlink Transmission", *Proceedings of VTC-Fall'02*, pags. 2145–2148, Vancouver (Canada), Sep. 2002.

[Li99] Y. Li, J. Chuang, N. Sollenberger, "Transmit Diversity for OFDM Systems and its Impact on High-rate Data Wireless Networks", *IEEE Journal on Selected Areas in Communications*, Vol. 17, nº 7, pags. 1233–1243, Jul. 1999.

[Li02] Y. Li, "Simplified Channel Estimation for OFDM Systems With Multiple Transmit Antennas", *IEEE Trans. on Wireless Communications*, Vol. 1, nº 1, pags. 67–75, Jan. 2002.

[Li03] Y. Li, B. Vucetic, A. Santoso, Z. Chen, "Space-time Trellis Codes with Adaptive Weighting", *Electronic Letters*, Vol. 39, nº 25, pags. 1833–1834, Dec. 2003.

[Li05] Y. Li, B. Vucetic, "Combined Space-time Trellis Codes and Beamforming on Fast Fading Channels", *Proceedings of VTC-Spring'05*, Stockholm (Sweeden), Jun. 2005.

[Lia05] A.P. Liavas, "Tomlinson-Harashima Precoding with Partial Channel Knowledge", *IEEE Trans. on Communications*, Vol. 53, nº 1, pags. 5–9, Jan. 2005.

[Liu03] J. Liu, E. Gunawan, "Combining Ideal Beamforming and Alamouti Space-time Block Codes", *Electronic Letters*, Vol. 39, nº 17, pags. 1258–1259, Aug. 2003.

[Llo57] S.P. Lloyd, "Least Squares Quantization in PCM", *Bell Labs. Technical Note*, 1957, unpublished. (See *IEEE Trans. on Information Theory*, Vol. 28, nº 2, pags. 129-137, Mar. 1982).

[Loz02] A. Lozano, C. B. Papadias, "Layered Space-Time Receivers for Frequency-Selective Wireless Channels", *IEEE Trans. on Communications*, Vol. 50, nº 1, pags. 65–73, Jan. 2002.

[Loz05] A. Lozano, A.M. Tulino, S. Verdú, "Mercury/Waterfilling: Optimum Power Allocation with Arbitrary Input Constellations", *Proceedings of ISIT'05*, pags. 1773–1777, Adelaide (Australia), Set. 2005.

[Lym03] R.J. Lyman, "Optimal Mean-Square Prediction of the Mobile-Radio Fading Envelope", *IEEE Trans. on Signal Processing*, Vol. 51, nº 3, pags. 819–824, Mar. 2003.

[Mag99] J.R. Magnus, H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics,*, John Wiley, 1999.

[Man01] J.H. Manton, "On Optimal Channel Identification by use of Training Sequences and Pilot Tones", *International Symposium on Signal Processing and its Applications*, pags. 599–602, Aug. 2001.

[Mar83] J.C. Martin, M.Mintz, "Robust Filtering and Prediction for Linear Systems with Uncertain Dynamics: A Game-Theoretic Approach", *IEEE Trans. on Automatic Control*, Vol. 28, nº 9, pags. 888–896, Sep. 1983.

[Mar99] T.L Marzetta, B. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading", *IEEE Trans. on Information Theory*, Vol. 45, nº 1, pags. 139–157, Jan. 1999.

[Mas74] J.L. Massey, "Coding and Modulation in Digital Communiations", *Proc. of Int. Zurich Seminar*, pags. E2.1–E2.4, Zurich, Mar. 1974.

[Max60] J. Max, "Quantizing for Minimum Distortion", *IRE Trans. on Information Theory*, Vol. 6, pags. 7–12, Mar. 1960.

[May97] C.E. Mayer, B.E. Jaeger, R.K. Crane, Xuhe Wang, "Ka-band scintillations: measurements and model predictions", *Proceedings of the IEEE*, Vol. 85, nº 6, pags. 936–945, Jun. 1997.

[Méd00] M. Médard, "The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel", *IEEE Trans. on Information Theory*, Vol. 46, nº 3, pags. 933–946, May 2000.

[Mie05] B. Mielczarek, W. Krzymien, "Flexible Channel Feedback Quantization in Multiple Antenna Systems", *Proceedings of VTC-Spring'05*, Stockholm (Sweeden), Jun. 2005.

[Mor01] M. Morelli, U. Mengali, "A Comparison of Pilot-Aided Channel Estimation Methods for OFDM Systems", *IEEE Trans. on Signal Processing*, Vol. 49, nº 12, pags. 3065–3073, Dec. 2001.

[Nab03] R.U. Nabar, H. Bolcskei, A.J. Paulraj, "Cut-off Rate Based Transmit OPtimization for Spatial Multiplexing on General MIMO Channels", *Proc. of ICASSP'03*, pags. 61–64 (Vol.5), Hong Kong, Apr. 2003.

[Nar98] A. Narula, M.J. Lopez, M.D. Trott, G.W. Wornell, "Efficient Use of Side Information in Multiple-Antenna Data Transmission over Fading Channels", *IEEE Journal on Selected Areas in Communications*, Vol. 16, nº 8, pags. 1423–1436, Oct. 1998.

[Nar99] R. Narasimhan, D.C. Cox, "A Generalized Doppler Power Spectrum for Wireless Environments", *IEEE Communications Letters*, Vol. 3, nº 6, pags. 164–165, Jun. 1999.

[Nee00] R. van Nee, A. van Zelst, G.A. Awater, "Maximum Likelihood Decoding in a Space Division Multiplex System", *Proceedings of VTC-Spring'00*, Tokyo (Japan), May 2000.

[Neg99] R. Negi, A.M. Tehrani, J. Cioffi, "Adaptive Antennas for Space-Time Coding over Block-Time Invariant Multipath Fading Channels", *Proceedings of VTC-Spring'99*, pags. 70–74, Houston (TX-USA), May 1999.

[Oie04] G.E. Oien, H. Holm, K.J. Hole, "Impact of Channel Prediction on Adaptive Coded Modulation Performance in Rayleigh Fading", *IEEE Trans. on Vehicular Technology*, Vol. 53, nº 3, pags. 758–769, May. 2004.

[Ong01a] E. N. Onggosanusi, A. Gatherer, A.G. Dabak, S. Hosur, "Performance Analysis of Closed-Loop Transmit Diversity in the Presence Of Feedback Delay", *IEEE Trans. on Communications*, Vol. 49, nº 9, pags. 1618–1630, Sep. 2001.

[Ong01b] E. N. Onggosanusi, A. M. Sayeed, B. D. Van Veen, "High Throughput Wideband Space-Time Signaling Using Channel State Information", *Proceedings of ICASSP'01*, pags. 2421–2424, Salt-Lake City (USA), May. 2001.

[Ong02] E.N. Onggosanusi, A.M. Sayeed, D. Van Veen, "Optimal Antenna Diversity Signaling for Wide-Band Systems Utilizing Channel Side Information", *IEEE Trans. on Communications*, Vol. 50, nº 2, pags. 341–353, Feb. 2002.

[Pal03a] D.Perez Palomar, *A Unified Framework for Communications through MIMO Channels*, PhD Thesis, UPC, May. 2003, Advisor : M.A. Lagunas.

[Pal03b] D.Perez Palomar, J.M. Cioffi, M.A. Lagunas, "Joint Tx-Rx Beamforming Design for Multicarrier MIMO Channels: A Unified Framework for Convex Optimization", *IEEE Trans. on Signal Processing*, Vol. 51, nº 9, pags. 2381–2401, Sep. 2003.

[Pal03c] D.Perez Palomar, J.M. Cioffi, M.A. Lagunas, "Uniform Power Allocation in MIMO Channels: A Game-Theoretic Approach", *IEEE Trans. on Information Theory*, Vol. 49, nº 7, pags. 1707–1727, Jul. 2003.

[Pal04] D.Perez Palomar, M.A. Lagunas, J.M. Cioffi, "Joint Tx-Rx Beamforming Design for Multicarrier MIMO Channels: A Unified Framework for Convex Optimization", *IEEE Trans. on Signal Processing*, Vol. 52, nº 5, pags. 1179–1197, May 2004.

[Pal05] D.P. Palomar, S. Verdú, "Gradient of Mutual Information in Linear Vector Gaussian Channels", *accepted in IEEE Trans. on Information Theory*, submitted April 2005.

[Ped90] K.I. Pedersen, J.B. Andersen, J.P. Kermoal, P.E. Mogensen, "A Stochastic Multiple-Input-Multiple-Output Radio Channel Model for Evaluation of Space-Time Coding Algorithms", *Proceedings of VTC-fall'99*, pags. 893–897, Boston-MA (USA), Sep. 1990.

[Pet97] P. Petrus, J.H. Reed, T.S. Rappaport, "Effects of Directional Antennas at the Base Station on the Doppler Spectrum", *IEEE Communications Letters*, Vol. 1, nº 2, pags. 40–42, Mar. 1997.

[PI04a] A. Pascual-Iserte, A.I. Perez-Neira, M.A. Lagunas, "A Maximum Approach for Robust MIMO Design: Combining OSTBC and Beamforming with Minimum Transmission Power Requirements", *Proceedings of SAM'04*, Sitges (Spain), Jul. 2004.

[PI04b] A. Pascual-Iserte, A.I. Perez-Neira, M.A. Lagunas, "On Power Allocation Strategies for Maximum Signal to Noise and Interference Ratio in an OFDM-MIMO System", *IEEE Trans. on Wireless Communications*, Vol. 3, nº 3, pags. 808–820, May. 2004.

[Pro94] John G. Proakis, Masoud Salehi, *Communication Systems Engineering*, Prentice Hall Int. Ed., 1994.

[Qin01] C. Qing, C.L. Law, S. Yoshida, "On the Doppler Power Spectrum at the Mobile Unit Employing Directional Antenna", *IEEE Communications Letters*, Vol. 5, nº 1, pags. 13–15, Jan. 2001.

[Rai96] D. Rainish, "Diversity Transform for Fading Channels", *IEEE Trans. on Communications*, Vol. 44, nº 12, pags. 1653–1661, Dec. 1996.

[Ral98] G.G. Raleigh, J.M. Cioffi, "Spatio-Temporal Coding for Wireless Communication", *IEEE Trans. on Communications*, Vol. 46, nº 3, pags. 357–366, Mar. 1998.

[Rap95] P.B. Rapajic, B.S. Vucetic, "Linear Adaptive Transmitter-Receiver Structures for Asynchronous CDMA Systems", *European Trans. on Telecommunications*, Vol. 6, nº 1, pags. 21–28, Jan./Feb. 1995.

[Rap96] T.S. Rappaport, *Wireless Communications: Principles and Practice*, Englewood Cliffs, NJ: Prentice Hall, 1996.

[Rep98] ETSI Technical Report, "Selection procedures for the choice of radio transmission technologies of the UMTS ", *(UMTS 30.03 version 3.2.0), TR 101 112 V3.2.0*, Nov. 1998.

[Rey01] F. Rey, M. Lamarca, G. Vázquez, "Transmitter Channel Tracking for Optimal Power Allocation", *Proceedings of ICASSP'01*, Salt-Lake City (USA), May. 2001.

[Rey04a] F. Rey, M. Lamarca, G. Vazquez, "A Robust Transmitter Design Based on Cutoff Rate for MIMO Multicarrier Systems with Imperfect Channel Estimates", *Proceedings of SAM'04*, Sitges (Spain), Jul. 2004.

[Rey04b] F. Rey, M. Lamarca, G. Vazquez, "Coded BER minimization for MIMO Multicarrier Systems with Imperfect Channel Estimates", *Proceedings of SPAWC'04*, Lisbon (Portugal), Jul. 2004.

[Rey05] F. Rey, M. Lamarca, G. Vazquez, "Robust Power Allocation Algorithms for MIMO OFDM systems with imperfect CSI", *IEEE Trans. on Signal Processing*, Vol. 53, nº 3, pags. 1070–1085, Mar. 2005.

[Sal85] J. Salz, "Digital Transmission Over Cross-coupled Linear Channels", *AT&T Technical Journal*, Vol. 64, nº 6, pags. 1147–1159, Jul./Aug. 1985.

[Sal92] M. Salehi, "Capacity and Coding for Memories with Real-Time Noisy Defect Information at Encoder and Decoder", *IEE Proceedings-I*, Vol. 139, nº 2, pags. 113–117, Apr. 1992.

[Sam99] H. Sampath, A. J. Paulraj, "Joint Transmit and Receive Optimization for High Data Rate Wireless Communication using Multiple Antennas", *Procedings of 33st Asilomar Conf. Signals, Syst. Comput.*, pags. 215–219, Pacific Grove, (CA-USA), Oct. 1999.

[Sam01] H. Sampath, P. Stoica, A. Paulraj, "Generalized Linear Precoder and Decoder design for MIMO Channels Using the Weighted MMSE Criterion", *IEEE Trans. on Communications*, Vol. 49, nº 12, pags. 2198–2206, Dec. 2001.

[Sam02] H. Sampath, A. Paulraj, "Linear Precoding for Space-Time Coded Systems With Known Fading Correlations", *IEEE Communications Letters*, Vol. 6, nº 6, pags. 239–241, Jun. 2002.

[Sca99a] A. Scaglionne, S. Barbarossa, G.B. Giannakis, "Filterbank Transceivers Optimizing Information Rate in Block Transmissions over Dispersive Channels", *IEEE Trans. on Information Theory*, Vol. 45, nº 3, pags. 1019–1032, Apr. 1999.

[Sca99b] A. Scaglionne, G.B. Giannakis, S. Barbarossa, "Redundant Filterbank Precoders and Equalizers. Part I: Unification and Optimal Designs", *IEEE Trans. on Signal Processing*, Vol. 47, nº 7, pags. 1988–2006, Jul. 1999.

[Sca00] A. Scaglione, G.B. Giannakis, S. Barbarossa, "Linear Precoding for Estimation and Equalization of Frequency-Selective Channels", G. B. Giannakis, Y. Hua, P. Stoica, L. Tong (eds.), *Signal Processing Advances in Wireless Communications*, Vol. I: Trends in Channel Estimation and Equalization, chap. 9, Prentice-Hall, 2000.

[Sch66a] J.P.M. Schalkwijk, "A Coding Scheme for Additive Noise Channels with Feedback - Part II: Band-Limited Signals", *IEEE Trans. on Information Theory*, Vol. 12, nº 2, pags. 183–189, Apr. 1966.

[Sch66b] J.P.M. Schalkwijk, T. Kailath, "A Coding Scheme for Additive Noise Channels with Feedback - Part I: No Bandwith Constraint", *IEEE Trans. on Information Theory*, Vol. 12, nº 2, pags. 172–182, Apr. 1966.

[Sha58] C. Shannon, "Channels with Side Information at the Transmitter", *IBM Journal of Research and Development*, Vol. 2, pags. 289–293, Oct. 1958.

[She03] Z. Shen, J.G. Andrews, B.L. Evans, "Short Range Wireless Channel Prediction Using Local Information", *Procedings of 37th Asilomar Conf. Signals, Syst. Comput.*, pags. 1147–1151, Pacific Grove, (CA-USA), Nov. 2003.

[Sim98] M.K. Simon, M. Alouini, "A Unified Approach to the Performance Analysis of Digital Communications over Generalized Fading Channels", *Proceedigs of the IEEE*, Vol. 86, nº 9, pags. 1860–1877, Sep. 1998.

[Sim03] S.H. Simon, A.L. Moustakas, "Optimizing MIMO Antenna Systems With Channel Covariance Feedback", *IEEE Journal on Selected Areas in Communications*, Vol. 21, nº 3, pags. 406–417, Apr. 2003.

[Sim04] O. Simeone, Y. Bar-Ness, U. Spagnolini, "Linear and Nonlinear Preequalization/Equalization for MIMO Systems with Long-Term Channel State Information at the Transmitter", *IEEE Trans. on Wireless Communications*, Vol. 3, nº 2, pags. 373–378, Mar. 2004.

[Ste01a] A. Stefanov, T.M.Duman, "Turbo-Coded Modulation for Systems with Transmit and Receive Antenna Diversity over Block Fading Channels: System Model, Decoding Approaches, and Practical Considerations", *IEEE Journal on Selected Areas in Communications*, Vol. 19, nº 5, pags. 958–968, May 2001.

[Ste01b] M. Sternad, T. Ekman, A. Ahlén, "Power Prediction on Broadband Channels", *Proceedings of VTC-spring'01*, pags. 2328–2332, Rhodes Island, (Greece), May 2001.

[Su01] H. Su, E.Geraniotis, "Space-Time Turbo Codes with Full Antenna Diversity", *IEEE Trans. on Communications*, Vol. 49, nº 1, pags. 47–57, Jan. 2001.

[Tak02] Y. Takatori, K. Cho, T. Hori, "Downlink Beamforming method using STBC for Mobile Propagation Environments", *Proceedings of VTC-Spring'02*, pags. 1916–1920, Birmingham (AL-USA), May. 2002.

[Tan95] J.M.A. Tanskanen, A. Huang, T.I. Laakso, S.J. Ovaska, "Prediction of Received Signal Power in CDMA Cellular Systems", *Proceedings of VTC'95*, pags. 922–926, Chicago (IL-USA), Jul. 1995.

[Tar98] V. Tarokh, N. Seshadri, A.R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction", *IEEE Trans. on Information Theory*, Vol. 44, nº 2, pags. 744–765, Mar. 1998.

[Tar99] V. Tarokh, H. Jafarkhani, A.R. Calderbank, "Space-Time Block Coding for Wireless Communications: Performance Results", *IEEE Journal on Selected Areas in Communications*, Vol. 17, nº 3, pags. 451–460, Mar. 1999.

[Tar00] V. Tarokh, H. Jafarkhani, "A Differential Detection Scheme for Transmit Diversity", *IEEE Journal on Selected Areas in Communications*, Vol. 18, nº 7, pags. 1169–1174, Jul. 2000.

[Teh98] A.M. Teherani, A. Hassibi, J. Cioffi, S. Boyd, "An Implementation of Discrete Multi-Tone Over Slowly Time-Varying MIMO Channels", *Globecom*, pags. 2806–2811, Nov. 1998.

[Tel99] I.E. Telatar, "Capacity of Multi-antenna Gaussian Channels", *European Trans. on Telecommunications*, Vol. 10, nº 6, pags. 585–595, Nov. 1999.

[Tom71] M. Tomlinson, "New Automatic Equaliser Employing Modulo Arithmetic", *Electronic Letters*, Vol. 7, nº 5/6, pags. 138–139, Mar. 1971.

[Ton98] L. Tong, S. Perreau, "Multichannel Blind Identification: From Subspace to Maximum Likelihood Methods", *Proceedigs of the IEEE*, Vol. 86, nº 10, pags. 1951–1968, Oct. 1998.

[Ton00] A.M. Tonello, "Space-time bit-interleaved coded modulation with an interative decoding strategy", *Proceedings of Vehicular Technology Conference Fall 2000*, pags. 473–478, Boston (USA), Sep. 2000.

[Tsa96] M.K. Tsatsanis, G.B. Giannakis, G. Zhou, "Estimation and Equalization of Fading Channels with Random Coefficients", *Signal Processing*, Vol. 53, pags. 211–229, 1996.

[Tsa97] M.K. Tsatsanis, G.B. Giannakis, "Subsapace Methods for Blind Estimation of Time-Varying FIR Channels", *IEEE Trans. on Signal Processing*, Vol. 45, nº 10, pags. 3084–3093, 1997.

[Tsa02] J. Tsai, R.M. Buehrer, B.D. Woerner, "Performance of Transmit Beamforming in Ricean fading Channels: Robustness on Quantization Error and Feedback Error", *Proceedings of VTC-fall'02*, pags. 1792–1795, Vancouver (Canada), Sep. 2002.

[Tun01] T. Tung, K.Yao, R.E.Hudson, "Channel Estimation and Adaptive Power Allocation for Performance and Capacity Improvement of Multiplie-Antenna OFDM Systems", *Proceedings of SPAWC'01*, pags. 82–85, Taoyuan (Taiwan), Mar. 2001.

[Tur65] G.L. Turin, "Signal Design for Sequential Detection Systems with Feedback", *IEEE Trans. on Information Theory*, Vol. 11, pags. 401–408, Jul. 1965.

[Tur66] G.L. Turin, "Comparison of Sequential and Nonsequential Detection Systems with Uncertainty Feedback", *IEEE Trans. on Information Theory*, Vol. 12, nº 1, pags. 5–8, Jan. 1966.

[Ung82] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals", *IEEE Trans. on Information Theory*, Vol. 28, nº 1, pags. 55–67, Jan. 1982.

[Vat97] F. Vatalaro, A. Forcella, "Doppler Spectrum in Mobile-to-Mobile Communications in the Presence of Three-Dimensional Multipath Scattering", *IEEE Trans. on Vehicular Technology*, Vol. 46, nº 1, pags. 213–219, Feb. 1997.

[Vau00] R. Vaughan, P. Teal, R. Raich, "Short-term Mobile Channel Prediction using Discrete Scatterer Propagation Model and Subspace Signal Processing Algorithms", *Proceedings of VTC-fall'00*, pags. 751–758, Boston (MA-USA), Sep. 2000.

[Ven03] S. Venkatesan, S. Simon, R. Valenzuela, "Capacity of a Gaussian MIMO Channel with Nonzero Mean", *Proceedings of VTC-fall'03*, pags. 1767–1771, Orlando (FL-USA), Oct. 2003.

[Vis99] H. Viswanathan, "Capacity of Markov Channels with Receiver CSI and Delayed Feedback", *IEEE Trans. on Information Theory*, Vol. 45, nº 2, pags. 761–771, Mar. 1999.

[Vis01] E. Visotsky, U. Madhow, "Space-Time Transmit Precoding with Imperfect Feedback", *IEEE Trans. on Information Theory*, Vol. 47, nº 6, pags. 2632–2639, Sep. 2001.

[Vit65] A.J. Viterbi, "The Effect of Sequential Decision Feedback on Communication Over the Gaussian Channel", *Information and Control*, Vol. 8, nº 1, pags. 80–92, Feb. 1965.

[Vit67] A.J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm", *IEEE Trans. on Information Theory*, Vol. 13, nº 2, pags. 260–269, Apr. 1967.

[Vit79] A.J. Viterbi, J.K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, 1979.

[Vor03] S.A. Vorobyov, A.B. Gershman, 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, pags. 313–323, Feb. 2003.

[Vuc91] B. Vucetic, "An Adaptative Coding Scheme for Time-Varying Channels", *IEEE Trans. on Communications*, Vol. 39, nº 5, pags. 653–663, May 1991.

[Wan05] D. Wang, E.A. Jorswieck, A. Sezgin, E. Costa, "Joint Tomlinson-Harashima Precoding with Diversity Techniques for Multiuser MIMO Systems", *Proceedings of VTC-Spring'05*, Stockholm, (Sweeden), May 2005.

[Wes98] R.D. Wesel, J.M. Cioffi, "Achievable Rates for Tomlison-Harashima Precoding", *IEEE Trans. on Information Theory*, Vol. 44, nº 2, pags. 824–831, Mar. 1998.

[Win04] C. Windpassinger, R.F.H. Fischer, T. Vencel, J.B. Huber, "Precoding in Multiantenna and Multiuser Communications", *IEEE Trans. on Wireless Communications*, Vol. 3, nº 4, pags. 1305–1316, Jul. 2004.

[Wol64] J. Wolfowitz, *Coding Theorems of Information Theory*, Springer-Verlag, 2nd ed., 1964.

[Woz66] J.M. Wozencraft, "Modulation and Demodulation for Probabilistic Coding", *IEEE Trans. on Information Theory*, Vol. 12, nº 3, pags. 291–297, Jul. 1966.

[Yan94] J. Yang, S. Roy, "On Joint Transmitter and receiver Optimization for Multiple-Input-Multiple-Output (MIMO) Trasmission Systems", *IEEE Trans. on Communications*, Vol. 42, nº 12, pags. 3221–3231, Dec. 1994.

[Yoo04a] T. Yoo, A. Goldsmith, "Capacity of Fading MIMO Channels with Channel Estimation Error", *Proceedings of ICC'04*, pags. 808–813, Paris (France), Jun. 2004.

[Yoo04b] T. Yoo, E. Yoon, A. Goldsmith, "MIMO Capacity with Channel Uncertainty: Does Feedback Help?", *Proceedings of Globecom'04*, pags. 96–100, Dallas (TX-USA), Dec. 2004.

[Zha97] Y.S. Zhan, D.B. Li, "Volterra Adaptive Prediction of Multipath Fading Channel", *Electronic Letters*, Vol. 33, nº 9, pags. 754–755, Apr. 1997.

[Zho02] S. Zhou, G. Giannakis, "Optimal Transmitter Eigen-Beamforming and Space-Time Block Coding Based on Channel Mean Feedback ", *IEEE Trans. on Signal Processing*, Vol. 50, nº 10, pags. 2599–2613, Oct. 2002.

[Zho04] S. Zhou, G.B. Giannakis, "How Accurate Channel Prediction Needs to be for Transmit-Beamforming with Adaptive Modulation over Rayleigh MIMO Channels?", *IEEE Trans. on Wireless Communications*, Vol. 3, nº 4, pags. 1285–1294, Jul. 2004.