TESI DOCTORAL

Títol:

"Adaptive Communications for Next Generation Broadband Wireless Access Systems"

Realitzada per	Ismael Gutiérrez González
en el Centre	Escola Tècnica Superior d'Enginyeria i Informàtica La Salle
i en el Departament	"Departament d'Electrònica i Comunicacions"
Dirigida per	Dr. Joan Lluis Pijoan, Dr. Faouzi Bader

C. Claravall, 1-3
08022 Barcelona
Tel. 936 022 200
Fax 936 022 249
E-mail: urlsc@sec.url.es
www.url.es

Adaptive Communications for Next Generation Broadband Wireless Access Systems

Ph.D. Thesis

Ismael Gutiérrez González

Advisors: Dr. Faouzi Bader and Dr. Joan Lluis Pijoan

Barcelona, June 2009

Author's address:

Ismael Gutiérrez
Advanced Technologies Standardization and Regulation (ATSR)
Samsung Electronics Research Institute (SERI)
Communications House, South Street, Staines, TW18-4QE (UK)

Tel. +44 (0) 1784 428600 Ext. 780
Fax: +44 (0) 1784 428610
E-mail: i.gutierrez@samsung.com and igutierrez@salle.url.edu

This work has been carried out partly in collaboration with the *Centre Tecnològic de Telecomunicacions de Catalunya* (CTTC, Spain).

Centre Tecnològic de Telecomunicacions de Catalunya
Parc Mediterrani de la Tecnologia (PMT)
Avda. Canal Olímpic, s/n
08860 – Castelldefels, Barcelona (Spain)

A Tania y a toda mi familia,
en especial a mis padres Andrés y Carolina.

Abstract

In Broadband Wireless Access systems the efficient use of the resources is crucial from many points of views. From the operator point of view, the bandwidth is a scarce, valuable, and expensive resource which must be exploited in an efficient manner while the Quality of Service (QoS) provided to the users is guaranteed. On the other hand, a tight delay and link quality constraints are imposed on each data flow hence the user experiences the same quality as in fixed networks. During the last few years many techniques have been developed in order to increase the spectral efficiency and the throughput. Among them, the use of multiple antennas at the transmitter and the receiver (exploiting spatial multiplexing) with the joint optimization of the *medium access control* layer and the *physical* layer parameters.

In this Ph.D. thesis, different adaptive techniques for B3G multicarrier wireless systems are developed and proposed focusing on the SS-MC-MA and the OFDM(A) (IEEE 802.16a/e/m standards) communication schemes. The research lines emphasize into the adaptation of the transmission having (Partial) knowledge of the Channel State Information for both; single antenna and multiple antenna links. For single antenna links, the implementation of a joint resource allocation and scheduling strategies by including adaptive modulation and coding is investigated. A low complexity resource allocation and scheduling algorithm is proposed with the objective to cope with *real-* and/or *non-real-* time requirements and constraints. A special attention is also devoted in reducing the required signalling both in the downlink and the uplink. Moreover, for multiple antenna links, the performance of a proposed adaptive transmit antenna selection scheme jointly with space-time block coding selection is investigated and compared with conventional structures. In this research line, mainly two optimizations criteria are proposed for spatial link adaptation, one based on the minimum error rate for fixed throughput, and the second focused on the maximisation of the rate for fixed error rate. Finally, some indications are given on how to include the spatial adaptation into the investigated and proposed resource allocation and scheduling process developed for single antenna transmission.

Acknowledgments

First of all, I would like to express my gratitude to Dr. Faouzi Bader for the many hours that he has dedicated to this Ph.D. Thesis. His perseverance in pushing me to improve and achieve higher levels of demand, as well as his knowledge that has allowed me to go forward in this exciting world of research. Equally, I would like to express all my thanks to Dr. Joan Lluís Pijoan for that first coffee when he encouraged me to go ahead with the doctorate. Now, almost five years later, I realize that after many days (including weekends and holidays) of hard work, I see myself at the place I wanted and that the journey is still at its beginning.

Nevertheless, I will never forget all the people from the *Research Group in Electromagnetism and Communications* (A.M. Sanchez, D. Badia, J. Pajares, J.Mauricio, J.R. Regué, M. Ribó, S. Graells, J.P. Rodríguez, R. Bardají and R.M. Alsina), and especially my old office mates (Ricard Aquilué, Pau Bergadà and Marc Deumal) with whom I shared many happy moments which made these years pass in a second.

I would like to thank all those people and institutions that have helped me during all these years, mainly the Access Technologies Research area of the Centre Tecnològic de Telecomunicacions de Catalunya – (CTTC), the Wireless@KTH department from the Royal Institute of Technology University (KTH), and all the members of the *Advanced Technologies Standardization and Regulation* department of Samsung Electronics Research Institute, UK. Thanks to all of them for their support of the research undertaken in this thesis.

In addition, all my sincere gratitude to all the reviewers of this Ph.D. thesis for their comments and suggestions that help to improve the quality of this dissertation.

Finally, all my acknowledges to the *"Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya"* – (DURSI), for the awarded grantships within the *"Beques doctorals per a la formació de personal investigador"*(FI) and "*Beques per a estades per a la recerca fora de Catalunya"* – (BE) programs.

About the author

Ismael Gutiérrez González received the Bachelor and Master degrees in Telecommunications Engineering in Sept. 2003 and June 2004 respectively from *Enginyeria i Arquitectura La Salle* of *Ramon Llull University* (URL) in Barcelona (Spain). In September, 2004 he was awarded with a Ph.D. scholarship from *Enginyeria i Arquitectura La Salle*, and four months later he was awarded with a Ph. D grantship from the local Catalan Government (*Departament d'Universitats, Recerca i Societat de la Informació*) for the period 2005-2008. During this period he was an active member of the Electromagnetism and Communications Research Group (GRECO) research group within the *Electronics and Communications Department* (DEC), where he carried out his Ph.D. researches in collaboration with the "Access Technologies" research area of the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) of Barcelona.

The author has participated in several European research projects as the *"Wireless Systems Providing High Quality Services"* (WISQUAS) - CP2-035, the "*PHYsical layer for DYnamic AccesS and cognitive radio"* (PHYDYAS)- ICT-211887, and the *COST-289* related with "*Spectrum and Power Efficient Broadband Communications"*, as well as in the industrial project related with the development of a "*broadband OFDM modem for medium and high voltage power lines*" lead by ENDESA Network Factory - Spain.

During his Ph.D. research period, he was granted two international research mobility funds for his researches at Kunglinga Tekniska Högskolan (KTH) University of Stockholm, Sweden in 2006, and at Samsung Electronics Research Institute (SERI), UK in 2008. From September 2003 to January 2009, he exerted as assistant professor in *Wireless and Mobile Communications Networks* at *Enginyeria i Arquitectura La Salle* of URL University.

Since March 2009, he has been a senior research engineer at the *Advanced Technologies Standardization and Regulation* (ATSR) department of Samsung Electronics Research Institute, in Staines, UK.

Author's publications list

The author's contributions are below listed and grouped according to the research topic.

Publications directly related with the Ph.D. thesis:

[1] I. Gutiérrez, F. Bader; R. Aquilué, J. Pijoan, "Continuous Frequency-Time Resource Allocation and Scheduling for Wireless OFDMA Systems with QoS Support", *EURASIP Journal on Wireless Communications and Networking*, Special Issue on OFDMA Architecture, Protocols and Applications, Hindawi. It will be published on May 2009.

[2] I. Gutiérrez, F. Bader, A. Mourad, J. Pijoan, "Transmit antenna and code selection for MIMO systems with linear receivers", *7th International Workshop on Multi-Carrier Systems & Solutions* (MC-SS-2009), Herrsching, Germany, May, 2009.

[3] I. Gutiérrez, F. Bader, J. Pijoan, "Mixed TUSC and Band AMC Permutation Zone in OFDMA Systems with Limited Feedback", *IEEE Vehicular Technology Conference (VTC-Spring'09)*, Barcelona, Spain, April, 2009.

[4] I. Gutiérrez, F. Bader, J.L. Pijoan, "New Prioritization Function for Packet Data Scheduling in OFDMA Systems"**,** *IEEE Radio and Wireless Symposium 2009 (RWS-2009)*, San Diego, USA, Jan. 2009.

[5] T. Lestable, M. Jiang, A. Mourad, D. Mazzarese, S. Han, H. Choi, H. Kang , I. Gutierrez, "Linear Dispersion Codes for Uplink MIMO schemes in IEEE 802.16m", IEEE C802.16m-08/535, July, 2008.

[6] I. Gutiérrez, F. Bader, J.L. Pijoan, "Prioritization Function for Packet Data Scheduling in OFDMA Systems", *Fourth International Wireless Internet Conference (WICON-2008)*, Maui, USA, Nov.2008.

[7] I. Gutiérrez, F. Bader, J.L. Pijoan, "Subchannel Power Allocation and Rate Adaptation for a MC-CDMA System Satisfying Mixed QoS Requirements"**,** *Mosharaka International Conference on Communications, Signals and Coding (MIC-CSC 2008)*, Mosharaka, Jordan, Oct. 2008.

[8] I. Gutiérrez, F. Bader, J.L. Pijoan, "Cross-Layer Design for Resource Allocation under QoS Requirements for Multiuser SS-MC-MA Systems", Editors: Thomas Magedanz, Markus Muek, et al., *Book title: Middleware Technologies for Enabling Next Generation Network Services and Applications*. Chapter: NGN Networking and Security, pp123-136. Multicon Verlag Edition, 2008.

[9] I. Gutiérrez, F. Bader, J.L. Pijoan, "Radio Resource Allocation in MC-CDMA Under QoS Requirements", *Published in the book Multi-Carrier Spread Spetrum,* pp. 207-216*,* Chapter: Adaptive Transmission, edited by S. Plass, A. Dammann, S. Kaiser and K. Fazel, ISBN: 978-1-4020-6128-8 , Springer, 2007. The Netherlands.

[10] I. Gutiérrez, F. Bader, J.L. Pijoan, S. Ben Slimane., "Adaptive Resource Management for a MC-CDMA System with Mixed QoS Classes Using a Cross Layer Strategy", *IEEE Vehicular Technology Conference (VTC-Spring'07)*, Dublin, Ireland, April, 2007.

[11] I. Gutiérrez, F. Bader, J.L. Pijoan, M. Deumal, "Adaptive Bit Loading With Multi-User Diversity in MC-CDMA"**,** *Proceedings European Wireless 2006 (EW2006)*, Athens, Greece, April, 2006.

[12] I. Gutiérrez, F. Bader, J.L. Pijoan, M. Deumal, M., "Performance of a new adaptive grouping and modulation procedure for multiuser group orthogonal MC-CDMA", *in Proc. IASTED: Communications and Systems Networks*, Benidorm, Spain, Sept., 2005.

[13] I. Gutiérrez, J.L. Pijoan, M. Deumal, F. Bader, "User grouping under adaptive modulation mechanisms for GO-MC-CDMA systems", in *Proc. on URSI 2005*, Gandia, Spain, Sept., 2005.

[14] I. Gutiérrez, D. Bartolomé, C. Vilella, "Analysis of different power control methods for CDMA systems based on Game Theory", *in Proc. on URSI 2004*, Barcelona, Spain, Sept., 2004 .

Publications indirectly related with the Ph.D. thesis:

[15] R. Aquilué, I. Gutiérrez, J.L. Pijoan, G. Sánchez, "High Voltage Multicarrier Spread Spectrum Field Test", IEEE Transactions on Power Delivery (to be published).

[16] F. Bader, N. Labed, I. Gutiérrez, "Intercell Interference Investigation in a MC-CDMA System with Iterative Demapping", *in Proc. IEEE Vehicular Technologies Conference (VTC-2006 Fall)*. Montreal, Canada, Sept. 2006.

[17] N. Labed, F. Bader, I. Gutiérrez, "Supported Active Users in a MC-CDMA System under Intercell Interference", *in Proc. 1st IEEE International Conference on Mobile Computing and Wireless Communications (MCWC)*, Sept – 2006, Amman, Jordan, 2006.

[18] I. Gutiérrez, J.L. Pijoan, F. Bader, R. Aquilué, "New Channel Interpolation Method for OFDM Systems by Nearest Pilot Padding", *in Proc. European Wireless 2006 (EW2006)*, Athens, Greece, April, 2006.

[19] P. Bergadà, R. Aquilué, I. Gutiérrez, J.L. Pijoan, "Ionospheric channel estimation between the Antartica and the Ebro Observatory based on OFDM symbols", *in Proc. on URSI 2006*, Oviedo, Spain, 2006.

[20] J. Pijoan, C. Vilella, M. Deumal, P. Bergadà, I. Gutiérrez, R. Aquilué, D. Badia, S. Graells, D. Miralles, J.M. Torta, D. Altadill, "Activities and results from La Salle Engineering at the Antarctica (2003-2006)", *VII Spanish symposium for polar studies*, Granada, Spain, 2006.

[21] C. Vilella, J.C. Socoró, J.L. Pijoan, I. Gutiérrez, D. Altadill, "An Antarctica to Spain HF Link. Oblique Sounding Results", *in Proc. of IEE International Conference on Ionospheric Radio Systems and Techniques (IRST-2006)*, July 2006, London, England, 2006 .

[22] R. Aquilué, P. Bergadà, I. Gutiérrez, J.L. Pijoan, "Channel Estimation for Long Distance HF Communications based on OFDM Pilot Symbols", *in Proc. of IEE International Conference on Ionospheric Radio Systems and Techniques (IRST-2006)*, London, England, July, 2006.

[23] M. Deumal, J.L. Pijoan, I. Gutiérrez, A. Behravan, "Peak reduction of multi-carrier systems by Controlled Spectral Outgrowth", *in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-2006)*, Toulouse, France , May, 2006.

[24] M. Deumal, I. Gutiérrez, J.L. Pijoan, "PAPR reduction in orthogonal MC and MC-SS systems", *Workshop on Special topics on 4G systems, COST 289 Spectrum and Power Efficient Broadband Communications*, Antalya, Turkey, 2005.

[25] I. Gutiérrez, J.L. Pijoan, C. Vilella, "Ionospheric channel characterization at the Antarctica", *in Proc. de la URSI 2004*, Barcelona, Spain, Sept. 2004.

Contents

List of Figures

List of Tables

Acronyms

3GPP-LTE	3rd Generation Partnership Project - long term evolution
AMC	adaptive modulation and coding
AWGN	additive white Gaussian noise
BE	best effort
BLER	block error rate
BER	bit error rate
BPSK	binary phase shift keying
BS	base station
BWA	broadband wireless access
CBR	constant bit rate
CDMA	code division multiple access
CNR	channel to noise ratio
CP	cyclic prefix
CQM	channel quality metric
CSM	collaborative spatial multiplexing
CSI	channel state information
DS-SS	direct sequence spread spectrum
ESINR	effective signal to interference and noise ratio
ESM	effective SNR mapping
EESM	Exponential Effective SNR Mapping
FDD	frequency duplex division
FDMA	frequency division multiple access
FFT	fast Fourier transform
FH-SS	frequency hopping spread spectrum
FTP	file transfer protocol
ICI	inter-carrier interference
IP	internet protocol
IEEE	Institute of Electrical and Electronics Engineers
IMT-2000	International Mobile Telecommunications-2000
ISI	inter-symbol interference
JRAS	joint resource allocation and scheduling
LDC	linear dispersion code
LOS / NLOS	(non-) line of sight
MAC	medium access control
MAI	multiple access interference
MC-CDMA	multicarrier code division multiple access
MC-DS-CDMA	multicarrier direct sequence code division multiple access
MC-SS	multicarrier spread spectrum
MCS	modulation and coding scheme
MI-ESM	mutual information based effective SNR mapping
MIMO	multiple input multiple output
MISO	multiple input single output

ML	maximum likelihood
MMSE	minimum mean square error
MS	mobile station
MSE	mean square error
MT-CDMA	multitone code division multiple access
(n)rt	(non) real time
OFCDM	orthogonal frequency code division multiplexing
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
P/L/M/W-AN	personal/local/metropolitan/wide area network
PAPR	peak to average power ratio
PER	packet error rate
PFS	proportional fair scheduler
PN	pseudo noise
PRB	physical resource block
PUSC	partial usage of the subchannels
O-QAM	offset quadrature amplitude modulation
QoS	quality of service
QPSK	quaternary phase shift keying
RAS	resource allocation and scheduling
RU	(minimum allocable radio) resource unit
SC-FDMA	single carrier frequency division multiple access
SDMA	space division multiple access
S(I)NR	signal to (interference and) noise ratio
SISO	single input single output
SIMO	single input multiple output
SM	spatial multiplexing
STBC	space-time block coding
STTC	space-time Trellis coding
TACS	transmit antenna and space-time code selection
TDD	time division duplex
TDMA	time division multiple access
UGS	unsolicited grant service
UPA	uniform power allocation
VBR	variable bit rate
VRB	virtual resource block
VSF	variable spreading factor
VoIP	voice over IP
WSSUS	wide sense stationary uncorrelated scattering
ZF	zero forcing
Z-QAM	Z-ary rectangular quadrature amplitude modulation

Notation

$(x)^*$	complex conjugate of x
$(x)^+$	is equivalent to $max(0,x)$
$exp(x)$	means e^x
$log(x)$	natural logarithm of x
$log_b(x)$	logarithm in base b of x
$max(x,y)$	gives the maximum value of x and y
$min(x,y)$	gives the minimum value of x and y
$\arg\max_x y(x)$	Value of x that maximizes the function $y(x)$
$\mathrm{E}\{x\}$	statistical expectation of x
$\mathrm{var}\{x\}$	variance of x
$\sim$	distributed according to
$\mathcal{N}(\mu,\sigma^2)$	means a Gaussian (i.e. normal) distribution with mean μ and variance σ^2
$\mathbb{C}\mathcal{N}(\mu,\sigma^2)$	means a complex Gaussian (i.e. normal) distribution with mean μ and variance σ^2
$\mathbb{R}^{n\times m}, \mathbb{C}^{n\times m}$	the set of $n\times m$ matrices with real- and complex- valued entries, respectively
$\Re\{x\}, \Im\{x\}$	the real and imaginary parts of the complex value x.
$\delta(t)$	Dirac delta dunction
$Q\{x\}$	means the Q-function of x
∞	Infinity
$\lvert x\rvert$	Absolute value of x
$\lVert\mathbf{A}\rVert_F^2$	Froebenius norm of the matrix
$\mathbf{A}$	denotes a matrix
$\mathbf{A}(n,m)$ and $\mathbf{A}_{n,m}$	refer to the matrix element in n^{th} row and m^{th} column
$vec(\mathbf{A})$	stacks matrix $\mathbf{A}$ into an $nm\times 1$ vector in column wise order
$\mathbf{A}^{\mathrm{H}}$	is the Hermitian of the matrix.
$\mathbf{A}^t$	is the transpose of the matrix.
$\mathbf{A}^{-1}$	inverse of the matrix.
$\mathbf{I}_x$	means the identity matrix with x rows and columns
$\mathbf{A}\otimes\mathbf{B}$	is the Kronecker tensor product of matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathrm{Tr}\{\mathbf{A}\}$	means the trace of the matrix
$rank(\mathbf{A})$	gives the rank of the matrix

This Ph.D. thesis is divided in two parts. The first part is dedicated to the definition of the Broadband Wireless Access (BWA) communications frameworks, and the introduction of the different transmission schemes and services used within the forthcoming generation of BWA systems. A special emphasis is given to the definition of the quality of service requirements as well as the most common performance metrics, both from the user and the operator perspective. The second part is devoted to the analysis of the different author's research topics and proposed solutions focused on BWA systems mainly related with: resource allocation and scheduling schemes for BWA based on multicarrier transmission schemes with QoS support, and spatial adaptation mechanisms for multiple antennas transmission schemes. Several hints on future researches based on the obtained results are suggested, and finally the most important conclusions are given.

Part I

Overview

Chapter 1. Introduction

After the great success of broadband Internet access and the development of mobile wireless networks, it is expected that the next generation of Broadband Wireless Access (BWA) systems will be able to achieve spectral efficiencies of order of tens of bits/s/Hz, as well as to provide similar quality of service (QoS) than that experienced in fixed networks. However, in order to achieve such requirement two essential research lines need more deep analysis and progress: the dynamic resource allocation and/with more adequate/realistic scheduling techniques to guarantee mixed QoS service requirements, and the implementation of adaptable multiple antenna transmission schemes. Researches undertaken in this thesis focuses on the development of low complexity algorithms (suitable for real time implementation) that could exploit both resource allocation and multiple antenna techniques having as a main objective the optimization of the spectral efficiency, the fairness and QoS provisioning.

1.1. Wireless communications evolution

The history of wireless communications is marked by important technical advances and improvements on the transmission and reception schemes. During the early 80s the first analogue cellular telephony systems appeared into the market, and soon afterwards different mobile communications systems spread over the worldwide such as the Total Access Communication System (TACS) in Europe, the Advanced Mobile Phone Service (AMPS) in the USA or the Nippon Telephone and Telegraph (NTT) system in Japan among others [1][2]. The main characteristics of these systems, later known as the first generation systems, were the analogue frequency modulation employed for transmitting the voice signal (although digital signalling was also used for purposes of registering, billing, call setup, paging, etc.) and the medium access technique based on the Frequency Division Multiple Access (FDMA), where a non-overlapping frequency band was assigned to each connection [3][4].

In 1982 the *European Conference of Postal and Telecommunications Administration* (CEPT) founded a group named "*Groupe Spécial Mobile*" in charge of designing a new generation of cellular systems (second generation - 2G) based on a completely digital technology. The main outcome of this group was the standardization in 1989 of the Global System for Mobile communications (GSM) standard based on a new air interface where digital modulation was used for both voice and data communications. Moreover, the combination of FDMA and Time Division Multiple Access (TDMA) schemes was proposed in order to increase the channel capacity. Afterwards, another technology based on the Spread Spectrum scheme, previously developed for tactic army communications, lead in 1993 to the USA's Interim Standard (IS)-95 [5] for civilian communications. The Spread Spectrum technique consists on using a much larger bandwidth than strictly necessary for data transmission. Since data is spread over a wider channel, the system is then able to combat the interference and the multipath propagation better than with narrowband transmissions [6]. The spectrum spreading can be achieved by different techniques as the Direct Sequence Spread Spectrum (DS-SS), the Frequency Hopping Spread Spectrum (FH-SS) or the combination of both. In both, the DS-SS and/or the FH-SS, a specific code word sequence is assigned to each user, and whether the

spreading sequences have orthogonal propriety or good cross-correlation, they can be used to multiplex several users in the code domain; therefore the multiplexing scheme is referred as *Code Division Multiple Access* (CDMA).

The 90s were the decade of the Internet boom where multimedia communications and services became widely spread. The vertiginous increase on demand of such applications made that the increase on the services offer meet the necessity to achieve more capacity, what lead to the design of new systems offering integrated voice and data services with wider cell coverage and higher spectral efficiency. In 1985, the International Mobile Telecommunications (IMT-) 2000[1] working group was started driven by the ITU (International Telecommunications Union). The suitability of CDMA was investigated since 1988 until 1999 under several projects research programmes (e.g. *Research and development in Advanced Communications technologies* (RACE), the *Code Division Testbed* (CODIT), and the *Future Radio Wideband Multiple Access System* (FRAMES)) which concluded with the suitability of the CDMA (and FDMA access) scheme for the third generation (3G) mobile communication systems and becomes the basis of the further Universal Mobile Telecommunications Systems (UMTS) [7][8][9], launched only a few years ago in EU. The UMTS system offers to mobile telecom operators significant capacity increase and broadband capabilities to support greater numbers of voice and data customers, it also provides higher data rates (up to 2Mbps in indoor environments) at lower incremental cost than the 2G systems. Afterwards, new proposals and modifications have been introduced in order to increase the user data rate and/or the system efficiency into the UMTS system, where the most significant is the actual High Speed Packet Access (HSPA) specified in the 3GPP 25.321 Release 5 as "HSDPA" for the downlink and in the 3GPP 25.828 Release 6 as "HSUPA" for the uplink, with data rates up to 14Mbps and 5.8Mbps in the downlink and in the uplink respectively [10][11].

1.2. Enabling Beyond 3G and 4G technologies

It is observed that approximately every ten to fifteen years a new wireless generation comes to scene. The ITU has promoted the IMT-Advanced working group in charge of defining the different requirements and characteristics of the fourth generation (4G). The IMT-Advanced group has defined the 4G as a future wireless telecommunications technology allowing data transfer rates of 1Gbps at nomadic circumstances and 100Mbps at mobile environment [12]. The 4G mobile communication systems are expected to become commercially available between 2010 and 2015. The bands over which these 4G systems will be applied as well as the bandwidth assigned per channel are still to be decided although advanced progress has been achieved within the European Projects IST-WiNNER[2], IST- WiNNER-II, and WiNNER+ (Celtic) to localise and estimate the needed spectrum in Europe, Asia, and USA for such generation [13]. To achieve the above mentioned throughputs, between 50 and 100 MHz per carrier frequency will be required.

---

[1] The number 2000 has three meanings: it was supposed to represent the year when this technology should be available, it also meant the required data rate (2000 Kbps) as well as the band where this systems should be allocated (2000 MHz).

[2] Wireless World Initiative New Radio, IST-2003- 507581 (WINNER), IST-4-027756 (WINNER II), CP5-026 (WINNER+)

It is also expected that in the 4G systems, users will be able to gain access to high multimedia applications and data rate (voice, video, streaming, internet, etc), as well as being capable to interact in real time with other users over different types of environments (high/slow mobility, cellular/short range coverage, and/or adaptive transmit power) while choosing each time the fitted radio access that guarantees the users' service requirement (in terms of QoS, power saving, etc.) [14]. Figure 1.1 gives a global view of the wireless and mobile communications system evolution and how all of them will converge to the forthcoming 4G framework. Furthermore, parallel to the development of the mobile communications networks, the number of wireless networks for nomadic users (designed initially for data connections offering best effort services) has increased since the mid 90s and different standards have been developed according to the communication range (e.g. IEEE 802.11, IEEE 802.15, IEEE 802.16, etc.). The convergence of wireless data networks and cellular networks is actually another important aspect to consider under an all-IP world (often described under the "seamless" concept), where the packet transmissions basis will replace the circuit-switched based [12][15][16]. This doesn't imply that it will be only one type of network, but the cooperation between heterogeneous networks would create a unique wireless fabric (also known as pervasive network), where fixed, nomadic or mobile wireless users will be served by different operators and different self-configurable (mobile) ad-hoc networks (MANET), where the intelligence and network management will be carried out both in a centralized and distributed manner [18].

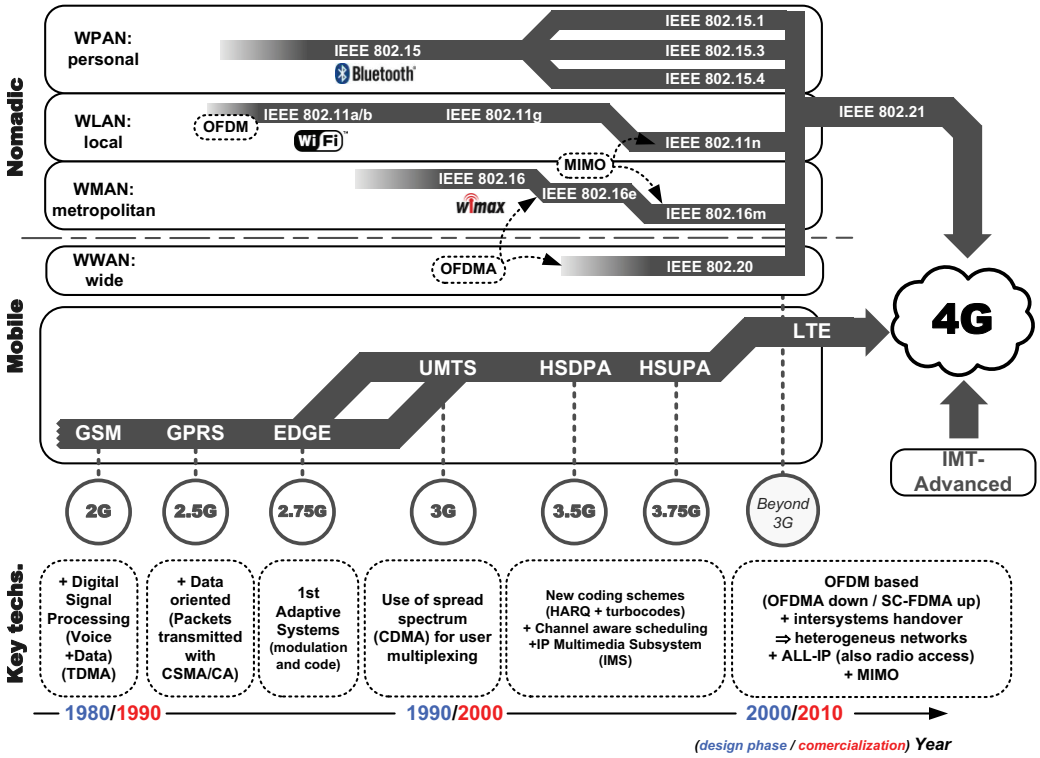


Figure 1.1. Convergence of nomadic and mobile broadband wireless access systems to 4G.

Furthermore, for the different emerged (or under development) standards which might be able to achieve the fourth generation communication requirements, there is a common factor which is the use of the *Orthogonal Frequency Division Multiplexing* (OFDM) scheme. The OFDM is based on the FDMA principle but using orthogonal frequencies. Thanks to this

orthogonality it is possible to increase the spectral efficiency by avoiding the use of guard bands, and also a more simplified design of the transmitter and the receiver is achieved since separate filters are not necessary anymore [19]. The key of the successful of this scheme lies also in its robustness against the multipath channels effects, and its flexibility to adapt according to the channel state, loading different power, modulation and coding scheme as a function of the channel state per subcarrier. The OFDM scheme, first introduced for Digital Subscriber Lines (DSL) connections [20], has demonstrated its high effectiveness in wireless local area networks (e.g. IEEE 802.11a/g [21]) and has been extended to a multitude of other wireless networks operating in different areas (i.e. personal, metropolitan, wide, rural, digital video and audio broadcasting, etc.) [22]-[28]. Nevertheless, the OFDM scheme has been also proposed in the last 3GPP Release 8, as the scheme that will replace the WCDMA scheme used in the UMTS radio access networks [29]. The Evolved UMTS Terrestrial Radio Access Network (E-UTRAN), also named Long Term Evolution (LTE), aims to become the basis of the fourth generation standard for mobile communications combining advanced signal processing and coding techniques, with multiple antenna transmission (multiple input multiple output – MIMO). Actually, the use of spatial multiplexing (achieved by using multiple antennas at the transmitter and at the receiver) techniques increase the channel capacity by a factor proportional to the minimum number of transmit and receive antennas. As a result, spectral efficiencies of more than 20bits/s/Hz can be achieved with sufficient number of transmit and receive antennas and if the environment is sufficiently rich in scatterers [30].

1.3. Broadband Wireless Access (BWA) definition

Broadband Wireless Access (BWA) is the radio access in which the connection(s) capabilities are higher than the primary rate (1.5 - 2Mbits/s) [31]. Then the BWA networks are differentiated in Fixed, Mobile and Nomadic. For the Fixed Wireless Access (FWA) networks, the location of the end-user termination and the network access point to be connected to the end-user are fixed, whereas for the Mobile Wireless Access (MWA) the end user is mobile. In the case of Nomadic Wireless Access, the location of the end-user termination may be in different places but it must be stationary while in use. Examples of BWA are the Wireless LAN, Metropolitan Area Networks, 3G and Wireless WAN which can even be combined in one single device to ensure seamless operation. Therefore, as previously introduced in Section 1.2, almost the systems are evolving into BWA where depending on the system specifications mobility is considered differently.

A typical BWA network provides a pool of bandwidth shared automatically amongst the users. Demand from different users is often statistically low correlated, allowing the BWA network to deliver significant bandwidth-on-demand to many users by accurate resource allocation and scheduling. The range of applications is very wide and evolving quickly. It includes voice, data and entertainment services of many kinds. Each subscriber may require a different mix of services, which is likely to change rapidly as connections are established and terminated.

The schemes studied in this thesis and the different author's proposals reviewed throughout the thesis can be fitted in any kind of BWA system as long as it is based on OFDM.

1.4. Adaptive communications

The high variability of the wireless channel makes that if we define a fixed transmission scheme irrespective of the channel state, the system throughput will be below the channel capacity during most of the time. Then, if we want to come closer to the channel capacity, since the instantaneous channel capacity changes constantly, the transmission parameters must be periodically modified in order to follow the channel changes. In consequence, the adaptation of the transmission in case of having (partial) channel state information at the transmitter has shown to increase the spectral efficiency, as well as the link reliability. Nevertheless, in a multiuser scenario we can combine these power and adaptive modulation and coding techniques with multiuser channel aware scheduling increasing even more the throughput of the system and the spectral efficiency. The general transmitter scheme is indicated in Figure 1.2 where the shaded block indicates those transmitter blocks which are aware of the channel status and accordingly may vary some of their internal parameters or behaviour. In the following subsection the main concepts regarding (spatial) link adaptation and resource allocation and scheduling are resumed. These concepts will be tackled in detail in Chapter 4 to Chapter 6.

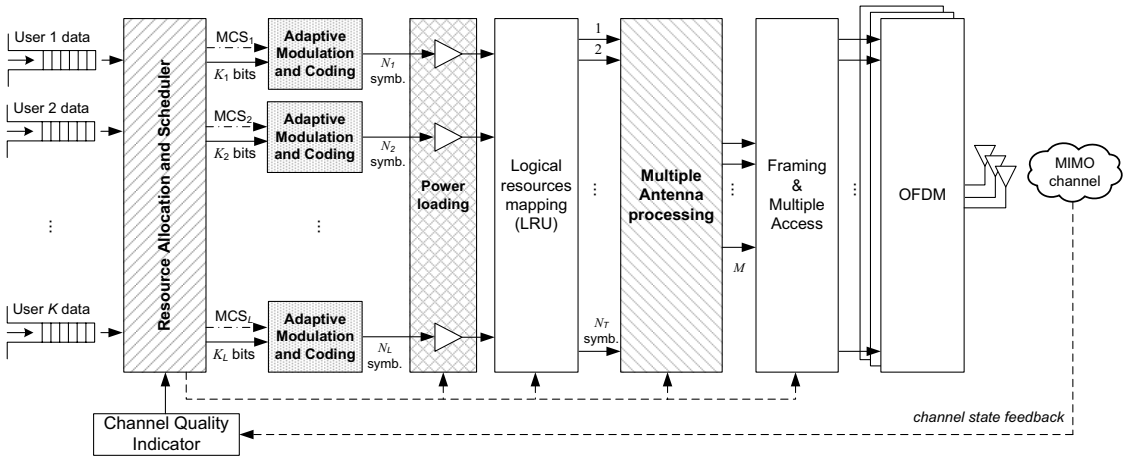


Figure 1.2. System architecture at the transmitter side.

1.4.1. Link adaptation

In BWA systems channel conditions can vary significantly due to the (fast) fadings. It is therefore desirable to adapt the modulation and the coding scheme (referred as Adaptive Modulation and Coding – AMC) as well as the power spectral density (power loading) according to the channel conditions. Traditionally, wireless voice networks have been designed to deliver a fixed bit rate stream that can be decoded correctly by all the users in the network (disregarding their position in the cell). As a result, the worst case (cell edge users) is usually considered to fix the maximum throughput. Thus, the channel capacity of other users close to the base station is much underestimated. However, wireless data networks where the channel access is done periodically (given by the packet rate) use the channel state in order to obtain the maximum attainable throughput given a certain bit or packet error rate (BER or PER). Therefore, the BWA networks can also take advantage of adaptive modulation and coding techniques to improve the overall system throughput.

In a typical adaptive modulation procedure, dynamic variations in the modulation order (constellation size) and forward error correction (FEC) code rate are undertaken. In practice, the receiver feeds back information about the channel state, which is then used to control the adaptation procedure. The AMC can be used in both up- and downlink modes. On the other hand, the power can be also adapted on per a user basis according to the channel gain experienced by each user. Different open-loop and close-loop power control algorithms have been proposed for most of the wireless and mobile communications systems, most of them trying to compensate the channel attenuation by transmitting more power on those faded bands. A number of fast close-loop power control algorithms have been proposed for the UMTS and WiMAX systems where the power is adapted on a per frame basis (e.g. 0.5-5ms). As a result the channel fluctuations can be practically removed using accurate power control procedures [32]. Nevertheless, recent researches on power adaptation have demonstrated that the channel capacity can be increased if the power is assigned to those parts of the spectrum where the channel is more beneficial (studied later in Section 4.3.2). In that case, more bits can be loaded per subchannel achieving higher spectral efficiencies. The techniques that combine power loading and AMC will be studied in Chapter 4.

1.4.2. Spatial Adaptation in multiple antenna links

As it was previously outlined, the BWA systems face two key challenges: the first one is how to provide high-data-rate, and the second is how to guarantee high-quality access over fading channels at almost wireline quality [34]. As it has been shown in several research works [35]-[37], the use of multiple antennas at both transmit and receive sides of the wireless link, in combination with signal processing and coding, increases the channel capacity (the capacity under certain conditions increases linearly with the minimum number of transmit or receiver antennas).

Actually, different data streams can be multiplexed simultaneously by exploiting the spatial domain. Furthermore, the combination of different transmitted and/or received signals can be used to combat the multipath channel efficiently by increasing the (spatial) diversity. Several space-time codes have been designed in the last years that achieve different trade offs between spatial diversity and spatial multiplexing.

However, whether the channel state information is known at the transmitter side, the transmission process can be adapted in order to maximize the channel capacity by for instance; adapting the transmitted power of each antenna (known as transmit antenna selection), to modify the space-time coding scheme according to the multiple antenna channel conditions (referred as spatial adaptation), or to adapt jointly the power, the modulation, and coding scheme of each spatial stream (i.e. link adaptation). These concepts will be studied in depth in Chapter 4 and Chapter 5.

1.4.3. Cross-layer Resource allocation and Scheduling

Usually, when two nodes of a network want to exchange information, following the Open System Interconnection (OSI) reference model, there is a flow of information through seven layers each of which is in charge of a specific part of the communication [38]. At the bottom of the protocol stack is where the interchange of information through a physical medium is

carried out by the physical layer (PHY). The PHY layer then decides according to the current channel state what are the optimum transmission power, as well as the modulation and coding scheme to assign. The medium access control (MAC) layer provides the addressing as well as the multiple channel access mechanisms that allow several terminals to share the radio communication medium.

However, it has been established that when the channel capacity varies constantly, it is more efficient if the MAC layer is informed about the channel state, and in case the channel is in outage, the transmission is postponed [39]. In addition, in BWA where the channel may be highly frequency selective (some parts of the spectrum present much better conditions than others), the MAC layer can decide in which frequencies the transmission should be established. Nevertheless, in a multiuser scenario where several users attempt to access over the channel communication in the uplink mode, or whether the base station must transmit data to the active users, the shared resources (power and time/frequency transmission intervals) must be assigned following a certain strategy to optimise the radio resources. In order to maximize the channel capacity, as well as the QoS required by the users' services, several cross-layer designs have been proposed by Georgiadis et al. in [39], Shakkottai et al. in [40], Song et al. in [41], etc. In these cross-layer designs or approaches, a number of physical layer and access layer parameters are jointly considered. The main benefits (behind the increase of complexity introduced by the implemented cross layer strategies) are higher spectral efficiencies, higher system load, as well as a higher control in reaching the QoS requirements. However, the main drawback of the previously mentioned research works is that they did not consider realistic conditions where the users' requirements may also change within short periods of time, as well as the required signalling in both downlink and uplink derived from the cross-layer based implementations. Hence, despite of its proved theoretical benefits, it is still difficult nowadays to get commercial equipments which include inter/multiple layer interactions.

In a single user OFDM(A)-based wireless network, where the frequency and the time resources are efficiently divided into subcarriers and time symbols respectively, the use of power and modulation and coding scheme adaptation over each subcarrier, as well as the dynamic sub-carrier assignment can significantly improve the performance. On the other hand, in multiuser systems where users' channel characteristics are almost mutually independent, the communication system can also take benefits from the multiuser diversity by assigning dynamically the different subcarriers (i.e. Orthogonal Frequency Division Multiple Access – OFDMA), hence each user is allocated into that subcarriers where higher spectral efficiencies can be achieved.

The dynamic cross-layer resource allocation may be performed according to several criteria such as: the channel capacity maximization, the maximum fairness, the minimum delay, etc. The resource allocation problem usually turns into a Non-Polynomial hard combinatorial problem, thus finding the optimal solution in a real time implementation is almost unfeasible. Several (mathematical) tools have been proposed (e.g. utility theory [42], neural networks [43], genetic algorithms [44], convex programming [45], etc). However, these proposals still imply high computational requirements, and as a result less complex heuristic algorithms are

usually applied in order to solve the multiuser resource allocation problem in real-time applications.

The resource allocation and the scheduling problem in wireless networks will be tackled in detail in Chapter 4, where different optimization criteria and the author's proposals regarding joint resource allocation and scheduling (under rectangular data packing structures and partial feedback) are described and evaluated.

1.5. Thesis layout

The thesis is mainly structured into two main parts referred as; *Overview* and *Contributions*. The "*Overview"* part gives introductions general background and basic concepts for understading how are the different communication structures of several wireless Broadband communication systems, and which are the limitations and the requirements that they have to deal with. This section is principally introduced to help the readers to follow the problematic analysis and the research solutions introduced and developed in the second part. In the "*Contributions"* part, the author describes the research works carried out during the Ph.D. thesis dealing with link adaptation, resource allocation and scheduling in single and multiple antenna transmission schemes. Mapping these topics to the transmitter scheme depicted in Figure 1.2, the "*Contributions"* section deals with those blocks gray shaded whereas the *"Overview"* one deals with the rest of blocks and elements that get involved in the communication process.

Chapter 2 is devoted to the analysis of those radio technologies that are considered as potential candidates for the forthcoming next generation of Broadband Wireless Access systems. Within this chapter, different transmission schemes are described (OFDM, OFDMA, Orthogonal Frequency Code Division Multiplexing, Single Carrier(SC)–FDMA and Filter Bank Multicarrier-FBMC) making a special emphasize on the OFDM(A) scheme which is considered in almost the beyond 3G and fourth wireless networks (IEEE 802.16m, LTE, IEEE 802.11 Very High Throughput - VHT, etc.). Furthemore, the main properties of the wireless (MIMO) broadband radio channel are also tackled with introductions to the different channel models that will be later used within the link and the system level simulations. Chapter 3 principally focus on the definition and the analysis of the framework (i.e. the requirements and the evaluation methodology) that has been used throughout the development of this Ph.D. thesis. The normalization of these metrics as well as using adequate performance metrics is highly required in order to obtain comparable results to other research works.

In Chapter 4, the State of the Art (SoA) of single input single output (SISO) link adaptation and the applied techniques for link level to system level mapping are described. These techniques are later used in one of the two core chapters of this thesis which is Chapter 5. In the latter, the author first reviews the SoA in cross-layered resource allocation and scheduling algorithms. Afterwards, two algorithms developed by the author are studied introducing two later modifications of the algorithms which aim to reduce the required signalling as well as improving the performance in case of mixed low and high mobility users.

The following Chapter 6 constitutes the last core chapter of this thesis where the research carried in case of spatial adaptation for multiple antenna transmission is presented. First, the fundamental trade-off between diversity and spatial multiplexing is reviewed. Afterwards, the principles of space-time coding based on linear dispersion codes are summarized. Then, assuming that the transmitter is awared of (partial) channel status, the two main precoding techniques, i.e. transmit antennas selection and space-time coding scheme selection are analysed and enhanced leading to the Transmit Antenna and Code Selection (TACS) proposed scheme. Two optimization criterias for the TACS scheme are evaluated by means of Monte Carlo simulations. The first optimization criteria follows the previously proposed adaptation schemes that improve the link reliability (i.e. BER or PER minimization), and extends this analysis to any Linear Dispersion Code. The second optimization criteria focuses on the throughput maximization by adapting the space-time code as well as the modulation and the code rate applied to each of the spatial streams. The performance and the system behaviour of these adaptation schemes are evaluated by means of computer simulation.

Finally, Chapter 7 summarises and points out the main research challenges undertaken in this thesis and the major achieved performances related with the proposed adaptive techniques covered in this dissertation.

1.6. Main Ph.D. thesis Contributions

During the development of this thesis the author has contributed mainly to two research topics: resource allocation and scheduling, and spatial adaptation algorithms for multiple antenna transmission. However, most of the human resources were devoted to the first topic and as a result most of the contributions are also to the issue of resource allocation and scheduling.

As it will be latter discussed, the author has used a cross-layer approach in order to solve the multiuser resource allocation problem coping with tight delay constraints. Among the different author's proposals to this issue it is important to highlight that the joint resource allocation and scheduling algorithm developed has shown to achieve much better results that the Proportional Fair Scheduler (which is commonly used as a reference). In addition, the author proposes a new resource allocation scheme in which the packets are prioritized in order to meet the delays in a packet basis. This has shown to be very convenient since, besides leading to low average and maximum delays, the jitter is also reduced. Nevertheless, the author has also highlighted the relevance of the signalling in the resource allocation process, dealing with new schemes that allow meeting a trade of between the required signalling and the throughput. In deed, it is shows that when signalling requirements are included in the computation of the spectral efficiency, the optimum operation point is found considering just a small amount of signalling in the feedback channel (i.e. imited channel state information).

For the multiple antenna transmission case, the author has studied the combination of transmit antenna selection with space-time coding selection. The study has been done for any kind of linear Space Time Block Code (STBC) under the Linear Dispersion Code (LDC) framework. It has been shown that the proposed scheme is able to achieve the maximum diversity order as well as coding gain of up to 2 dB compared to the best coding scheme in

each scenario. Furthermore, it has been shown that the proposed scheme is suited in case of low complex receivers (for instance using linear receivers) since the performance gap between the Maximum Likelihood detector and the linear receiver (e.g. Zero Forcing or Minimum Mean Square Error based detector) is shortened by the use of the proposed technique. In addition, it has been also shown that the signalling requirements are quite low since the number of LDC codes is much reduced. These results may let us conclude that the technique studied and proposed in this thesis for spatial adaptation can be promoted for most of the BWA. Actually, in some standards as the IEEE 802.16e the standard itself leaves room and mechanisms to apply such kind of techniques. Next researches dealing with this topic should then consider more realistic conditions and implementation constraints (channel estimation errors, feedback delay, antenna correlation effects, etc.) in order to asses it benefits and allow its promotion within the previously reviewed BWA standards.

References

[1] J.E. Padgett, C.G. Günther, T. Hattori, “Overview of wireless personal communications”, *IEEE Communications Magazine*, vol.33, pp.28-41, Jan.1995

[2] G.L. Stüber, *Principles of Mobile Communications*. Boston: Kluwer Academic Publishers, 1996.

[3] B. Sklar, *Digital Communications: Fundamentals and Applications.* Prentice Hall, 1988.

[4] J.G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.

[5] R.C. Dixon, *Spread Spectrum Systems with Commercial Applications*. New York: John Wiley & Sons, 1994.

[6] A.J. Viterbi, *CDMA Principles of Spread Spectrum Communication.* Reading: Addison-Wesley, 1995.

[7] P.W. Baier, “CDMA or TDMA? CDMA for GSM?”, in *Proceedings IEEE International Symposim on Personal, Indoor and Mobile Radio Communications (PIMRC’04),* The Hague, The Netherlands, pp.1280-1284, Sept.1994.

[8] A. Baier, U.C. Fiebig, W. Granzow, W. Koch, P. Teder, J. Thielecke, “Design study for a CDMA-based third-generation mobile radio system”, *IEEE Journal on Selected Areas in Communications*, vol.12, pp.733-743, May 1994.

[9] P.W. Baier, P. Jung, A. Klein, “Taking the challenge of multiple access for third generation cellular mobile radio systems – a European view”, *IEEE Communications Magazine*, vol.34, pp.82-89, Feb. 1996.

[10] T.E. Kolding, F. Frederiksen, P.E. Mogensen, "Performance aspects of WCDMA systems with high speed downlink packet access (HSDPA)," *IEEE 56th Vehicular Technology Conference, 2002. Proceedings. (VTC-2002-Fall)*, vol.1, pp. 477- 481 vol.1, 2002

[11] R. Love, A. Ghosh, X. Weimin, R. Ratasuk,"Performance of 3GPP high speed downlink packet access (HSDPA)," *IEEE 60th Vehicular Technology Conference, 2004 (VTC-2004-Fall),* vol.5, pp. 3359- 3363 Vol. 5, 26-29 Sept. 2004.

[12] ITU-R Recommendation M.1645, “Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000,” 2003.

[13] IST-4-027756 WINNER II, “Final Report”, D7.1.5, Feb. 2008.

[14] J.M. Pereira, "Fourth generation: now, it is personal!," *The 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2000. (PIMRC-2000)*, vol.2, pp.1009-1016 vol.2, 2000.

[15] M. Frodigh, S. Parkvall, C. Roobol, P. Johansson, P. Larsson, "Future-generation wireless networks," *Personal Communications, IEEE [see also IEEE Wireless Communications]*, vol.8, no.5pp.10-17, Oct 2001.

[16] L. Ying, Z. Shihua, R. Pinyi, H. Gang, "Path toward next generation wireless Internet - cellular mobile 4G WLAN/WPAN and IPv6 backbone," *TENCON '02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering*, vol.2, no.pp. 1146-1149 vol.2, 28-31 Oct. 2002.

[17] P802.21 Draft D5.2 Issue. June 19 2007

[18] D. K. Kim, "A New Mobile Environment: Mobile Ad Hoc Networks (MANET)," *IEEE Vehic. Tech. Soc. News*, August 2003, pp. 29-35.

[19] R.W. Chang, “Synthesis of band-limited orthogonal signals for multi-channel data transmission”, *Bell System Technical Journal* 46, 1996, pp.1775-1796.

[20] “Asymmetric digital subscriber line (ADSL) transceivers”, ITU-T Recom.. G.992.1, June 1999.

[21] IEEE Std 802.11™-2007, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, New York, USA, 12 June 2007.

[22] IEEE 802.16TM-2004/Cor1-2005 (Amendment and Corrigendum to IEEE 802.16-2004), Part 16: Air Interface for Fixed Broadband Wireless Access Systems; Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1, IEEE, New York, February 2006.

[23] ETSI TS 136 300 V8.4.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRAN); Overall description", April 2008.

[24] M. Wang, "Ultra Mobile Broadband Technology Overview," Communication Networks and Services Research Conference, 2008. CNSR 2008. 6th Annual , vol., no., pp.8-9, 5-8 May 2008

[25] W. Bolton, X. Yang, M. Guizani, "IEEE 802.20: mobile broadband wireless access," *IEEE Wireless Communications*, vol.14, no.1, pp.84-95, Feb. 2007.

[26] R. Laroia, "Fast-hopped OFDM Technology offers a new opportunity for mobile broadband – technology information", *Computer Technology Review*, May, 2001.

[27] ETSI EN 300 401 V1.3.3. European Standard for Radio Broadcasting Systems. Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers. May 2001.

[28] ETSI EN 300 744 V1.4.1. European Standard for Digital Video Broadcasting (DVB). Framing structure, channel coding and modulation for digital terrestrial television. January 2001.

[29] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); 3GPP TS 36.300 V8.7.0 (2008-12).

[30] A. F. Molisch, M. Z. Win, J. H. Winters, A. Paulraj, "Capacity of MIMO systems with antenna selection," *In Proc. IEEE International Conference on Communications (ICC)*, 2:570– 574, 2001.

[31] Vocabulary of terms for wireless access, RECOMMENDATION ITU-R F.1399-1, 2001.

[32] 3GPP TS 25.211 V3.12.0 (2002-09), Physical channels and mapping of transport channels onto physical channels (Release 1999).

[33] P. Whitehead, Broadband Wireless Access (BWA) System Overview, IEEE 802.16 Broadband Wireless Access Working Group, Dec. 1999.

[34] H. Bölcskei, A.J. Paulraj, K.V.S. Hari, R.U. Nabar, W.W. Lu, "Fixed Broadband Wireless Access: State of the Art, Challenges, and Future Directions", IEEE Communications Magazine, Vol.39, pp. 100-108, Jan. 2001.

[35] E. Telatar, "Capacity of multi-antenna Gaussian channels", *European Transactions on Telecommunications*, Nov. 1999.

[36] V. Tarokh, N. Seshadri, A. R. Calderbank, "Space-Time codes for high data rates wireless communication: Performance criterion and code construction", IEEE. Trans. on Information Theory, vol.44, pp.744-765, March, 1998.

[37] S. Sandhu, R. Heath, A. Paulraj, "Space-time block codes versus space-time trellis codes," *IEEE International Conference on Communications*, 2001 (ICC-2001), vol.4, pp.1132-1136 vol.4, 2001.

[38] X.200 Information technology – Open Systems Interconnection – Basic reference model: The basic model (IS 7498), July, 94.

[39] L. Georgiadis , M. J. Neely , L. Tassiulas, "Resource Allocation and Cross-Layer Control in Wireless Networks", Foundations and Trends in Networking, Vol. 1, Iss. 1, 2006.

[40] S. Shakkottai, T. S. Rappaport and P. C. Karlsson, "Cross-Layer Design for Wireless Networks", *IEEE Comun. Mag.* Oct. 2003, pp.74-80.

[41] G. Song and Y. (G). Li, "Cross-layer optimization for OFDM wireless network – part I and part II," *IEEE Trans. WirelessCommun.*, vol. 4, no. 2, pp. 614–634, March 2005.

[42] Haipeng Lei, Lei Zhang, Xin Zhang, Dacheng Yang, "A Packet Scheduling Algorithm Using Utility Function for Mixed Services in the Downlink of OFDMA Systems," IEEE 66th Vehicular Technology Conference, 2007. (VTC-2007-Fall), pp.1664-1668, Sept. 30 2007-Oct. 3 2007.

[43] N. García, J. Pérez-Romero, R. Agustí, "A New OFDMA Scheduler for Delay-Sensitive Traffic Based on Hopfield Neural Networks," *EURASIP Journal on Wireless Communications and Networking*, 2008.

[44] Yu Yu, Wuyang Zhou, "Resource Allocation for OFDMA System Based on Genetic Algorithm," Cross Layer Design, 2007. IWCLD '07. International Workshop on, pp.65-69, 20-21 Sept. 2007.

[45] S. Kibeom, M. Mohseni, J.M. Cioffi, "Optimal Resource Allocation for OFDMA Downlink Systems," *Information Theory, 2006 IEEE International Symposium on* , vol., no., pp.1394-1398, 9-14 July 2006.

Chapter 2. Broadband Wireless Access radio technologies and radio channel characteristics

As introduced in the first chapter of this dissertation, one of the main challenges for the fourth generation (4G) is the design of radio interfaces capable of delivering high throughputs (with high spectral efficiencies), as well as providing the quality of service required by each user. This Chapter 2 reviews some of the best known air interface candidates considered for the forthcoming next generation of BWA systems. In addition, some general knowledge about the mobile radio channel as well as the models that are used throughout the thesis is given. The background provided in this chapter regarding single and multiple antenna BWA systems will allow a smooth flow of the concepts and schemes developed on the forthcoming chapters.

In this chapter, the following issues are addressed:

- We start by reviewing the fundamentals of the radio channel. The basic propagation concepts are explained and a special interest is given to understanding the multipath channels as well as how they can be modelled.
- Afterwards, the more well-known transmission schemes proposed for Broadband Wireless Access systems are explained, showing its main characteristics and differences.
- We conclude the description of the radio interfaces and technologies extracting the commonalities among the reviewed air interfaces. These will be used in the following chapters to test the different author's proposals as well as observing their possible applications in the different BWA standards.

2.1. Radio channel definitions

The signal sent by any transmit antenna suffers different effects and attenuations before arriving to any of the receiver antennas at the mobile station (see the link budget example in Figure 2.1). These effects are mainly the free space loss (referred as *path loss*), the shadowing, the multipath propagation and the multi-antenna correlation. This behaviour is usually classified into long-term and short term fading characteristics. Long term fading depends on the geometrical location of the users in the network, whereas the short-term fading depends on the time variability of the channel. Both types of fadings can be modelled following a deterministic or stochastic approach, the former is applied for specific network planning and deployment when the channel conditions do not suffer large changes, whereas the second one is preferred for mobile users where the channel conditions cannot be predicted precisely or in simulation-based studies.

Actually, all these effects need to be considered in order to obtain the signal to noise ratio at the receiver side (which is actually one of the metrics that influences most the physical layer performance).

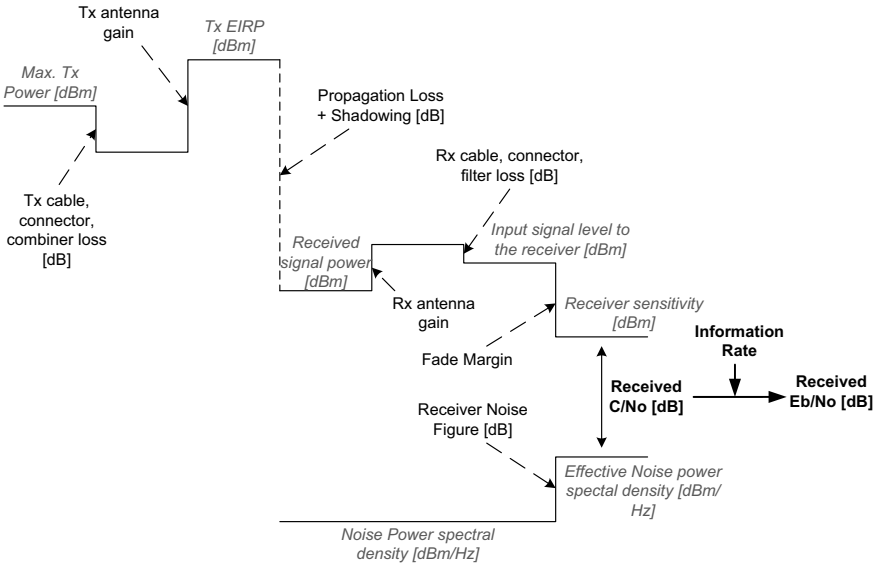


Figure 2.1. Link budget conceptual diagram

2.1.1. The path loss

The path loss is defined as the power loss due to the distance separation between the transmitter and the receiver. It depends on the carrier frequency, the distance between the BS and the MS, and the propagation environment (antenna height, buildings, walls, topography, etc. [1][2]). For a non-changing environment, if the distance and carrier frequency are fixed, the path loss can be determined and is constant. Since the path loss is not linearly dependant with the carrier frequency, different models have been proposed for the different transmission bands. However, all these models assume an attenuation which is proportional to the distance (d) between the transmitter and the receiver powered to n (where $2 \leq n \leq 4$, with n=2 in case of free space propagation and n=4 for dense urban scenarios). The *Path Loss* (L_{PL}) is then given by

$$L_{PL} = \alpha \times d^n \tag{2.1}$$

where α is a constant that depends on the environment, the antennas height, the frequency, etc. The system parameters as antenna gains, cable losses, propagation model, etc. will be described throughout this thesis for each system simulation case.

2.1.2. Shadowing model

Shadowing is the phenomenon caused by an object placed in front of a source of light, sound or electromagnetic waves. In case of electromagnetic waves, the effects of the shadowing over the received signal power are slowly time variant. The shadowing effect is then modelled as a random lognormal variable (with zero mean and variance σ[dB]) that is added (or subtracted) to the transmitted power [3]. Moreover the shadowing effect is correlated in distance, thus for two receivers separated by a distance Δx the expected shadowing will be correlated with a value,

$$R(\Delta x) = e^{-\frac{|\Delta x|}{d_{cor}} \ln 2} \tag{2.2}$$

where d_{cor} is the shadowing de-correlation length defined as $R(d_{\text{cor}})$=0.5. Then the slowly time variant shadowing process for the mobile user can be described by the following equation,

$$L_{SHAD}(t)[dB] = R(\Delta x) \times L_{SHAD}(t-\Delta t)[dB] + \sqrt{1-R^2(\Delta x)} \times X[dB] \quad (2.3)$$

where $L_{SHAD}(t)$ is the loss due to the fading process at time t and X is a log-normally distributed random variable with standard deviation $\sigma_X=\sigma(1-R)$, where σ is the shadowing standard deviation. The term Δt is the difference in time from the previous shadowing estimation and is equal to $\Delta t= d_{cor}/v$, where v is the mobile station velocity. The following Table 2.1 shows the typical decorrelation lengths as well as the shadowing standard deviation defined in [4].

Environment	Indoor	Urban	Rural
Shadowing decorrelation length d_{cor}	5 m	20 m	50 m
Shadowing standard deviation	12 dB	8 dB	6 dB

Table 2.1. Test environment shadowing characteristics.

In many references the path loss and the shadowing are also grouped under the name of large-scale fading whereas multipath effects are referred as small-scale fading. The Figure 2.1 shows all the parameters that should be taken during a link budget calculation where other parameters as the antennas gain, the cable and connectors losses, etc. are also included.

2.1.3. Multipath propagation via reflectors and scatterers

In addition to the path loss and the shadowing, the radio signal transmitted through the mobile radio channel suffers from other effects and distortions that are present during any propagation of electromagnetic waves. Mainly, these effects are caused by the reflection, the diffraction and the scattering of the waves propagating between the transmitter and the receiver [1].

According to [1], two kinds of *reflections* exist: *specular* reflection and *difusse* reflection. When the propagation wave impacts with a smooth surface (a uniform surface compared to the wavelength) the wave is reflected in only one direction as shown in Figure 2.2. However, when the surface presents irregularities (also in terms of the wavelength) the reflected signal takes different directions and in consequence the energy is spread in the space, this case is named *diffuse* reflection. This last reflection is quite difficult to predict and becomes a serious problem in high frequencies i.e. for tens of gigahertz.

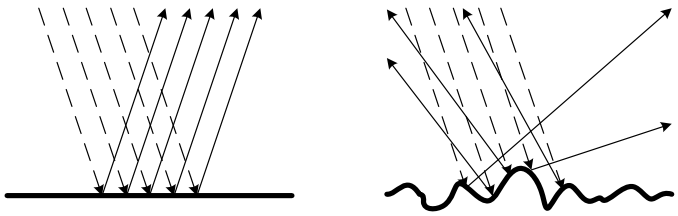


Figure 2.2. Reflection of radio waves over different surfaces.

Diffraction occurs when the propagation wave impacts with an edge of a large obstacle, in this case the wave tends to bend covering areas that where shaded because of the obstacle [1]. This mechanism affects more to the lower frequencies despite it is not negligible at higher frequencies. *Scattering* takes places when the propagation wave hits small objects (smaller than the wavelength) of irregular shapes, i.e. raindrops, snowflakes, leaves, etc. These objects scatters (spreads) the energy in all directions (they act as small antennas) and it is very difficult to predict the amount of energy in each direction.

As a result, according to the different propagation effects described above, the different propagating waves can reach the receiver at different instants since they can follow very different paths. Since each wave will arrive with a specific phase and amplitude, the sum of all the received waves will lead to the final received signal. Furthermore, the waves can arrive with any phase in the range [0, 2π), so the superposition of these waves can be constructive or destructive depending on the carrier frequency, producing large fluctuations in the received signal energy. Therefore, since small changes in the environment can change the way waves are reflected (and in consequence the arrival phase), the fluctuations at the receiver is rapidly varying. These effects are known as the multipath propagation mechanisms, and its effects are named small-scale fading (despite the fadings can be up to 30dB) or fast fading.

In order to model the multipath propagation, it is necessary to introduce the time-variant channel impulse response $h(\tau,t)$ *(throughout this thesis, for mathematical convenience but without loss of generality, all the bandpass signals and channels are replaced by their equivalent lowpass)*. This function describes how the channel reacts at time t to an impulse applied at an instant $\delta(t\text{-}\tau)$. Under a perfect knowledge of the channel impulse response, the output of the channel can be predicted for any input signal. The equivalent function is given in the frequency domain, named time-variant channel transfer function $H(f,t)$, by applying the Fourier transform to $h(\tau,t)$. The mobile channel is assumed to be stationary in the wide-sense (WSS), the time correlation between two instants (t_1,t_2) does not depend on the absolute time but only on the difference $(\Delta t=t_2\text{-}t_1)$. The same property holds also for spatial displacements. As a result, channel parameters such as the mean and variance do not change over time or position.

In multipath conditions, the channel impulse response can be expressed as the composition of a large number of scatter impulses received over N_p different paths,

$$h(\tau,t)=\sum_{p=1}^{Np} a_p e^{j\left(2\pi f_{D,p}t+\varphi_p\right)}\delta\left(\tau-\tau_p\right) \tag{2.4}$$

where a_p, $f_{D,p}$, φ_p and τ_p are the amplitude, the Doppler frequency, the phase and the propagation delay associated to each p^{th} path respectively. The Doppler frequency $f_{D,p}$ is determined as

$$f_{D,p}=\frac{v\times f_c}{c}\times\cos\left(\alpha_p\right), \tag{2.5}$$

where v means the speed of the mobile station, c is the speed of light, f_c is the carrier frequency, and α_p the angle of incidence of the wave associated to the p^{th} path.

2.1.3.1. Statistical analysis of the time-variant channel impulse response

The fast fading process can be characterized by the description of the correlation functions of the channel impulse response [5]. Thus, the autocorrelation of the $\underline{h}(\tau,t)$ is given by

$$R_h(\tau_1,\tau_2,\Delta t)=\frac{1}{2}\mathrm{E}\left\{h(\tau_1,t)h^*(\tau_2,t+\Delta t)\right\}, \tag{2.6}$$

where $(x)^*$ means the complex conjugate of x.

As it was mentioned before, assuming that $h(\tau,t)$ is as WSS process and that $h(\tau_1,t)$ and $h(\tau_2,t)$ are uncorrelated for $\tau_1 \neq \tau_2$ (referred as uncorrelated scattering – US), the autocorrelation function in Eq. (2.6) can be simplified as

$$R_h(\tau_1,\tau_2,\Delta t)=\rho_x(\tau_1,\Delta t)\cdot\delta(\tau_1-\tau_2), \tag{2.7}$$

where $\rho_x(\tau,\Delta t)$ is the delay cross-power spectral density [5], and the channel characterized under these assumptions is referred as WSSUS channel. The Fourier transform of the delay cross-power spectral density in Δt yields the following scattering function [5],

$$S(\tau,f_D)=\int_{-\infty}^{\infty}\rho_x(\tau,\Delta t)e^{-j2\pi f_D\Delta t}\partial\Delta t. \tag{2.8}$$

which provides a measure of the average power output as a function of the delay τ and the Doppler frequency f_D. In case the scattering function is integrated in the Doppler frequency domain, the delay power spectrum is

$$\rho(\tau)=\int_{-\infty}^{\infty}S(\tau,f_D)\,\partial f_D, \tag{2.9}$$

which is equivalent to the delay cross-power spectral density $\rho_x(\tau,\Delta t)$ at Δt=0. The delay power density spectrum, also known as the Average Power Delay Profile (APDP), gives the average power of the channel output as a function of the delay τ. From $\rho(\tau)$, the characteristic parameters of the multipath channel can be obtained (i.e. the mean delay τ_{mean}, the delay spread σ_τ, and the maximum delay τ_{max}). The τ_{mean} is equivalent to the first moment of the delay power density spectrum, it can be obtained as

$$\tau_{mean}=\overline{\tau}=\frac{\int_0^{\infty}\tau\cdot\rho(\tau)\partial\tau}{\int_0^{\infty}\rho(\tau)\partial\tau}, \tag{2.10}$$

where the denominator is used for normalization purposes since $\rho(\tau)$ is not a probability density function. Next the delay spread σ_τ, which is the standard deviation (or the second moment) of the delay power density spectrum, is defined as

$$\sigma_\tau^{\,2}=\frac{\int_0^{\infty}(\tau-\overline{\tau})^2\cdot\rho(\tau)\partial\tau}{\int_0^{\infty}\rho(\tau)\partial\tau}. \tag{2.11}$$

The bandwidth over which the signal propagation characteristics are correlated above a certain value is referred as the coherence bandwidth Δf_c which is proportional to the reciprocal of the delay spread. Mathematically speaking, the coherence bandwidth is defined as the bandwidth over which the frequency correlation function R_h is above 0.5, Δf_c can be approximated by the following formula [6]

$$\Delta f_c \approx \frac{1}{5 \cdot \sigma_\tau}. \tag{2.12}$$

Moreover, when the signal bandwidth W is larger than the coherence bandwidth Δf_c, the channel is said *frequency selective*, in case W is smaller than Δf_c the channel is said *non-frequency selective* or *flat*.

On the other hand, the Doppler power density spectrum $S_{fD}(f_D)$, defined as the integral of the scattering function in the delay domain τ, is given by

$$S_{f_D}(f_D) = \int_{-\infty}^{\infty} S(\tau, f_D)\, \partial\tau, \tag{2.13}$$

hence the average power of the channel output is as a function of the Doppler frequency f_D. Parallel to the time dispersive characteristics defined in Eq. (2.10) and (2.11), the frequency dispersive characteristics of the multipath channel are generally quantified by the maximum occurring frequency Doppler $f_{D,max}$. A common model for the time correlation is the well-known *Clarke's Model* [8]. This model assumes an isotropic scattering and an isotropic antenna at the receiver, thus the angles of arrival are uniformly distributed in the range [0,2π). Under these assumptions the time autocorrelation function follows the expression

$$R_t(\Delta t) = R_h(\tau_1, \tau_2, \Delta t)\big|_{\tau_1=\tau_2} = J_0(2\pi f_D \Delta t), \tag{2.14}$$

where $J_0(x)$ is the zero-order modified Bessel function of first kind evaluated at x. Applying the Fourier Transform ($FT\{x\}$) over (2.14), we obtain the so-called *Jakes Doppler Power Spectrum* [9][10]

$$J(f) = FT\{R_t(\Delta t)\} = \frac{1.5}{\pi f_{D_{\max}} \sqrt{1-\left(\frac{f}{f_{D_{\max}}}\right)^2}}, \quad f \in \{-f_{D_{\max}}, f_{D_{\max}}\}. \tag{2.15}$$

Furthermore, the coherence time Δt_c (time over which the time correlation functions is above 0.5) can be approximated by [6]

$$\Delta t_c \approx \frac{9}{19\pi f_{D_{\max}}}. \tag{2.16}$$

Similar to the previous frequency analysis, if the duration of the transmitted symbols is larger than the coherence time Δt_c , the channel is considered *time selective*, in case the symbols become shorter than Δt_c the channel is considered *non-time selective*.

2.1.3.2. Introduction to multipath channel models

The number of multipath channel models is quite large, each assuming certain probability density function to model the fast fading process. An approximation of the multipath channel effects is obtained considering that there is a large number of scatterers in the channel that contribute to the signal at the receiver. According to the central limit theorem, the summation of these contributions leads to a complex-valued Gaussian channel impulse response. In absence of *Line of Sight* (NLOS) the process is zero-mean. In this case, the magnitude of the corresponding channel transfer function $a(f,t)=|H(f,t)|$ is a random variable, for brevity denoted by a, whose distribution follows the Rayleigh distribution

$$p(a) = \frac{2a}{\Omega} e^{-a^2/\Omega}, \quad a \geq 0, \tag{2.17}$$

where Ω is the average power given by

$$\Omega = \mathrm{E}\left\{ \left| H(f,t) \right|^2 \right\}, \tag{2.18}$$

and where $E\{x\}$ means the expectation of x. The phase is considered uniformly distributed in the interval [0, 2π). This channel is known as *Rayleigh* fading channel and its distribution matches the propagation characteristics in macrocells environments [6].

When a *line of sight* (LOS) or a dominant component exists among the randomly moving scatterers, the channel impulse response is no longer zero-mean valued. In this case, the magnitude of the channel transfer function follows a Ricean distribution defined as

$$p(a) = \frac{2a(K_{Rice}+1)}{\Omega} e^{-K_{Rice} - a^2(K_{Rice}+1)/\Omega} \cdot \mathrm{I}_0\left(2a\sqrt{\frac{K_{Rice}(K_{Rice}+1)}{\Omega}} \right), \quad a \geq 0 \tag{2.19}$$

where K_{Rice} is the power ratio between the dominant path and the scattered paths at the receiver. The average power Ω is obtained as in (2.18) and $\mathrm{I}_0(x)$ is the zero-order modified Bessel function of first kind evaluated at x. The phase is also uniformly distributed in the interval [0, 2π). This channel is said to be *Rician* and fits better to the propagation characteristics of micro- and picocells where line of sight path is present [6].

Moreover, the *Nakagami-m* distribution introduced by Nakagami in [11] describes the magnitude of the channel transfer function $|H(f,t)|$ distribution by a central Chi-square distribution with m ($m \in \mathbb{N}$) degrees of freedom, hence

$$p(a) = \frac{2a^{2m-1}}{\Gamma(m)} \left(\frac{m}{\Omega} \right) e^{-ma^2/\Omega}, \quad a \geq 0 \tag{2.20}$$

with

$$m = \frac{\Omega^2}{\mathrm{E}\left\{ \left(a^2 - \Omega \right)^2 \right\}}, \quad m \geq \frac{1}{2}. \tag{2.21}$$

and where Γ(x) means the Gamma function defined as

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt. \tag{2.22}$$

The average power Ω is obtained as in Eq. (2.18). The parameter m is related to the fading strength. Actually, the Nakagami distribution is a generalization of the Rayleigh and Rician distributions where in case of LOS the Nakagami-*m* distribution approximates the Rice distribution under the following statement,

$$m = \frac{(K_{Rice}+1)^2}{(2K_{Rice}+1)}. \tag{2.23}$$

Moreover, if the term m approaches infinity (given by very high K_{Rice} values), the Nakagami-*m* and the Rice distribution becomes an impulse corresponding to an ideal channel ($h(\tau,t)=\delta(\tau,t)$). In case m is equal to 1, then K_{Rice}=0 and the Nakagami-*m* distribution is equivalent to a Rayleigh distribution. Finally in case $m \in [0.5, 1]$ the fading condition becomes more severe than for the Rayleigh distribution (e.g. if m = 0.5 a one-sided Gaussian distribution is obtained).

As a conclusion, the Nakagami-m is the most flexible distribution to model the channel, and if the parameters m and Ω are obtained, the channel is completely modeled. For instance, the channel propagation models in the CODIT study [12] where the channel for macro-, micro- and picocells are proposed based on the Nakagami-m distribution. In Figure 2.3 the Rice, Rayleigh and the Nakagami distribution functions are plotted for different values of K_{Rice} and m. It is shown that the Nakagami-m distribution is able to characterize any multipath channel conditions.

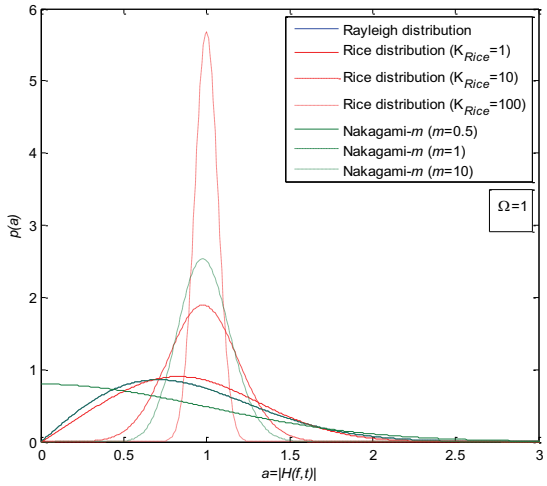


Figure 2.3. Rayleigh, Rice and Nakagami-m probability density function

2.1.4. Broadband Single Input Single Output (SISO) channel modelling

For single input single output channels, different techniques and models have been proposed to model the mobile radio channel. The ray tracing technique is the most complex and is based on computer simulation of the travelling waves where the landscapes are also modelled to estimate the waves' reflections and the scattering. Thus, the ray tracing method usually implies a very high computational cost [14]. On the other hand statistical models based on experiments can be quite realistic and are much more computationally efficient. According to the CODIT and COST207 projects [12][13], the channel can be modelled as the sum of N_p paths each of them with its respective amplitude a_p, Doppler frequency $f_{D,p}$, phase φ_p and propagation delay τ_p. In this case the channel transfer function is expressed as

$$H(f,t) = FT\left\{\sum_{p=1}^{N_p} a_p \cdot e^{j\left(2\pi f_{D,p} t + \varphi_p\right)} \cdot e^{-j2\pi f \tau_p}\right\}, \tag{2.24}$$

where the main advantage from this model is that channel can be directly simulated in frequency domain. For a multicarrier based system[3], assuming that the channel is flat over each sub-carrier and the channel does not change significantly during one symbol, the system can be completely simulated in the frequency domain reducing the computational cost during the simulations. The discrete representation of the channel transfer function becomes

$$H_{n,m} = H\left(n\Delta f, mT_s\right) = \sum_{p=1}^{N_p} a_p \cdot e^{j\left(2\pi f_{D,p} mT_s + \varphi_p\right)} \cdot e^{-j2\pi n\Delta f \tau_p}, \tag{2.25}$$

[3] The system must be properly designed in order to avoid inter-symbol interference and inter-carrier interference.

where Δf is the frequency spacing between subcarriers and T_s is the multicarrier symbol length. However, in most of the last European research projects [4][15] and the IEEE and ETSI working groups [16][17][18] the channel model is based on a discrete version of the scattering function (i..e. the average power delay profile). This model channel model uses a *Tapped Delay Line* (TDL) composed of N_{TAPS} where the number of taps is defined according to each environment. Each tapped delay line i (i={1.. N_{TAPS} }) is characterized by [10]:

- the average gain $\hat{a}_i=|a_i|^2$ of each tap,
- the relative delay τ_i of each tap,
- the Doppler power spectrum $S(\tau,f_D)|_{\tau=\tau i}$ of each tap,
- and the delay spread evaluated on the average power delay profile.

In consequence, the fading channel can be modelled very efficiently as a linear *Finite Impulse Response* (FIR) filter where the weights of the taps a_i are given by [19]

$$g[k]=\sum_{i=1}^{N_{TAPS}} sinc\left(\frac{\tau_i}{T-k}\right)E\left\{\left|a_i[m]\right|^2\right\}=\sum_{i=1}^{N_{TAPS}} sinc\left(\frac{\tau_i}{T-k}\right)\hat{a}_i,\quad k\in\aleph \tag{2.26}$$

where $g[k]$ is the discrete version of the average power delay profile, T is the input sample period and k is the sample index. The fading path gain for each tap $a_i[m]$ is a randomly generated scalar vector drawn from a Rayleigh distribution at intervals equal to T. Next, the obtained random vector is passed through a filter whose transfer function follows the Jakes Doppler spectrum (hence the time correlation follows the Clarke's model).

The schematic diagram for this channel simulator is depicted in Figure 2.4, $Z^{-\tau_i \cdot f_m}$ means the signal is delayed $\tau_i \cdot f_m$ samples where f_m means the sampling rate.

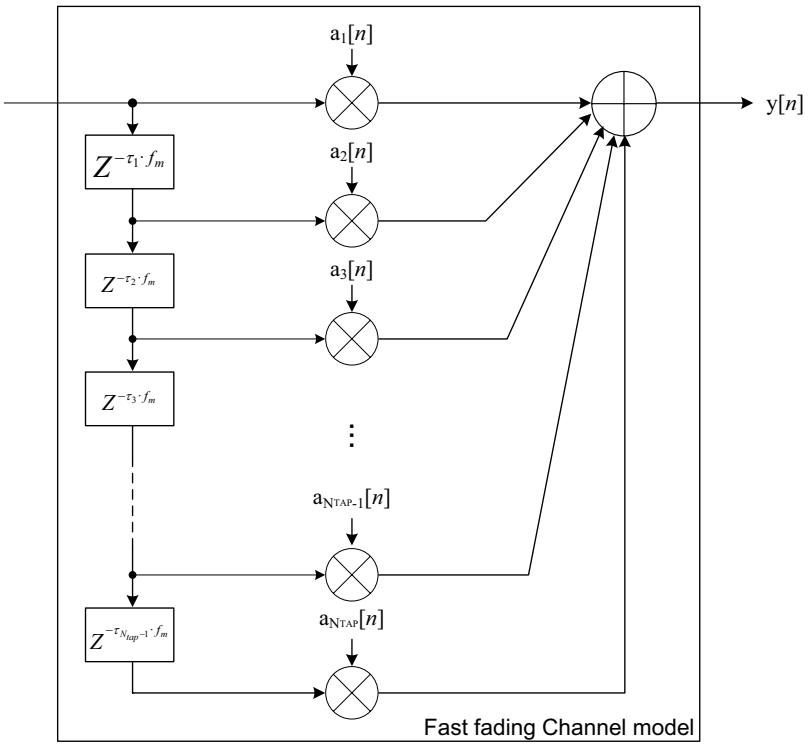


Figure 2.4. Fast fading channel simulator scheme.

In Appendix A, the values given in [4] and [16] for the average power delay profiles used throughout this thesis are given. In case of line-of-sight transmission, a constant gain is added to the line-of-sight tap (usually the first one). For any profile, the sum of all the taps' average gain is equal to one (in other words, the integral of $g[k]$ is unitary). Therefore, the multipath channel gain is normalized independently of the environment and channel model assumed.

2.1.5. Broadband MIMO channel modelling aspects

For the evaluation of multiple antenna techniques (which are envisioned to be a key enabling technology for beyond third generation systems), it is essential to have an accurate model of the Multiple Input Multiple Output (MIMO) channel. According to the literature, the MIMO channel can be modelled as a double-directional channel [20]. The double-directional term comes from two facts:

- the term *directional* indicates that the channel description takes into account the angular (spatial) distribution of the energy at the antennas, whereas a non-directional model only takes into account the temporal spreading of the signal.
- the term *double* means that the there are multiple antennas at both transmitter and receiver sides.

Nevertheless, one important aspect of any MIMOs channel is the channel correlations between pairs of transmit and receive antennas. These correlations limit the MIMO channel capacity as well as the maximum achievable diversity. In the following subsections, all these aspects are reviewed hence we can understand better the MIMO channel matrices used during simulations in the later Chapter 6.

2.1.5.1. Physical MIMO Channel Modelling

As stated in the previous section, the channel impulse response subsumes all the channel propagation effects. In case of ideal omni-directional antennas, all the multipath components (departing and arriving in all angular angles[4]) contribute to the impulse response. Considering the direction of departure (DoD) ϕ and the Angle of Arrival (AoA) θ, the channel impulse response is defined as

$$h(\tau,t,\phi,\theta)=\sum_{l=1}^{L}h_l(\tau,t,\phi,\theta) \tag{2.27}$$

where the contribution of each l multipath component is indicated by h_l which are given by

$$h_l(\tau,t,\phi,\theta)=a_l e^{j\varphi_k}\delta(\tau-\tau_l)\delta(\phi-\phi_l)\delta(\theta-\theta_l) \tag{2.28}$$

with a_l, φ_l, τ_l, ϕ_l and θ_l denoting the amplitude, the phase, the delay, the DoD and the AoA of each multipath component respectively. In order to reduce the model complexity, the multipath component with similar delay, DoD and AoA are grouped (merged) into *clusters* (or *taps*). The polarization effects might be also included by extending the impulse response to a 2×2 polarimetric matrix that describes the coupling between the vertical and the horizontal

[4] The angles that describe the angle of departure and the angle of arrival are spatial angles that correspond to the point on the unit sphere and replace the spherical azimuth and elevation angles.

polarizations. However, the antenna patterns and the system bandwidth are not considered in Eq. (2.27).

Furthermore, the MIMO system channel impulse response can be represented by the following matrix notation

$$\mathbf{H}(\tau,t)=\begin{pmatrix} h_{11}(\tau,t) & h_{12}(\tau,t) & \cdots & h_{1M}(\tau,t) \\ h_{21}(\tau,t) & h_{22}(\tau,t) & \cdots & h_{2M}(\tau,t) \\ \vdots & \vdots & \ddots & \\ h_{NN}(\tau,t) & h_{N2}(\tau,t) & \cdots & h_{NM}(\tau,t) \end{pmatrix}, \tag{2.29}$$

where N and M means the number of receiver and transmitter antennas respectively, and $h_{ij}(\tau,t)$ denotes the time-variant impulse response between the j-th transmitter antenna and the i-th receiver antenna. In this case there is no distinction between (spatially) separate antennas and different polarizations of the same antenna. As a result, when the vertical and the horizontal polarization are considered, the number of antennas is doubled. It is worth mentioning that the channel matrix as expressed in Eq. (2.29) contains the antenna patterns and the frequency filters, hence spatial correlations between antennas are also implicitly expressed.

Both types of channel characterization expressed in Eq. (2.27) and Eq. (2.29) are related by the following expression

$$h_{ij}(\tau,t)=\int_{\tau'}\int_{\phi}\int_{\varphi} h(\tau',t,\phi,\theta)\times G_{Tx}^{j}(\phi)\,G_{Rx}^{i}(\theta)\,f(\tau-\tau')\,\partial\tau'\,\partial\phi\,\partial\theta, \tag{2.30}$$

where the channel impulse response is obtained by the integral of the double-directionally channel impulse response defined in Eq. (2.27), the overall impulse response of the transmit and the receive antennas joint with the frequency filters $f(\tau)$, as well as the transmit and receive antenna patterns $G_{Tx}(\phi)$ and $G_{Rx}(\theta)$.

2.1.5.2. Deriving the MIMO channel matrix

In order to determine parameters such as the delay, the DoD or the AoA from Eq. (2.27), we may succinct to the geometrical distribution of the transmitter and the receiver as well as the surrounding conditions. However, this deterministic approach requires large computational cost and the obtained results are tied to a specific scenario. Hence, non-geometrical stochastic models have been proposed according to several channel sounding and measurement campaigns which characterize (according to specific stochastic distributions) the parameters mentioned before (delay, DoD, AoA, etc.) for different environments. Several models have been proposed in the last years; among them the more relevant are: the COST[5] 259/273 channel model, the 3GPP Spatial Channel Model (SCM), the Stanford University Interim (SUI) channel model and the IST-WINNER[6] project extended spatial channel model (SCME). An extended review of all these channel models can be found in [20] and [21].

For pure stochastic models, and under the WSSUS assumption, it can be assumed that the MIMO channel matrix is frequency flat and non-time variant for bandwidths smaller than the

[5] COST is an abbreviation for *European Cooperation in the field of Scientific and Technical research*.
[6] The European public funded research project IST-WINNER (Wireless-World-Initiative-New-Radio).

coherence bandwidth and during periods of time shorter than the coherence time respectively. Under previous assumptions, Eq. (2.29) is reduced to the following channel matrix expression

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1M} \\ h_{21} & h_{22} & \cdots & h_{2M} \\ \vdots & \vdots & \ddots & \\ h_{N1} & h_{N2} & \cdots & h_{NM} \end{pmatrix}, \tag{2.31}$$

Then spatial correlation matrix is then defined as

$$\mathbf{R}_{\mathbf{H}} = \mathrm{E}\left\{\left(\mathbf{H}^{H}\mathbf{H}\right)^{H}\right\} = \mathrm{E}\left\{vec\left(\mathbf{H}^{H}\right)\ vec\left(\mathbf{H}^{H}\right)^{H}\right\}, \tag{2.32}$$

which is a $MN \times MN$ positive semi-definite Hermitian matrix that describes the correlation between all pairs of transmit-receive channels (i.e. antenna pairs). The receiver and the transmitter side correlation matrices R_{H}^{Rx} and R_{H}^{Tx} are then given by

$$\mathbf{R}_{\mathbf{H}}^{Rx} = \mathrm{E}\left\{\left(\mathbf{H}\mathbf{H}^{H}\right)^{T}\right\}, \tag{2.33}$$

$$\mathbf{R}_{\mathbf{H}}^{Tx} = \mathrm{E}\left\{\left(\mathbf{H}^{H}\mathbf{H}\right)\right\}. \tag{2.34}$$

In case a sufficient spacing between the antenna elements at either the transmitter or receiver sides, the corresponding correlation matrix R_{H}^{Rx} and R_{H}^{Tx} becomes equal to a diagonal matrix.

Following with the expression in Eq. (2.32), we may express the channel matrix as

$$vec\left(\mathbf{H}^{H}\right) = \mathbf{R}_{\mathbf{H}}^{1/2} vec\left(\mathbf{H}_{w}^{H}\right) \tag{2.35}$$

where $\mathbf{H}_w$ is the decorrelated channel matrix which in case of the Rayleigh fading channels is equal to a circularly symmetric complex Gaussian matrix. Furthermore, in case all the individual channels are characterized by the same average power, $\mathbf{R}_{\mathbf{H}}$ is proportional to the identity matrix and in that case all the elements of the matrix are independent and identically distributed (i.i.d). Physically, these conditions occur only in rich scattering environments characterized by independent multipath channel components uniformly distributed in all directions.

However, the expression in Eq. (2.35) is not easily tractable, for that reason the Kronecker model is preferred [20]. Whether it is assumed that the transmitter and the receiver correlation matrices are separable, this is equivalent to restrict the correlation matrix to be defined by

$$\mathbf{R}_{\mathbf{H}} = \mathbf{R}_{\mathbf{H}}^{Rx} \otimes \mathbf{R}_{\mathbf{H}}^{Tx}, \tag{2.36}$$

then it can be shown that the channel matrix can be expressed by [22]

$$\mathbf{H} = \left(\mathbf{R}_{\mathbf{H}}^{Rx}\right)^{1/2}\ \mathbf{H}_{w}\ \left(\mathbf{R}_{\mathbf{H}}^{Tx}\right)^{1/2}. \tag{2.37}$$

Furthermore, under the separable correlations assumption, we can easily generate a correlated MIMO channel matrices since given a spatial uncorrelated matrix ($\mathbf{H}_w$) we can obtain the correlated version by including the information of the transmit and the receive correlation matrices. This approach has been used in this thesis, where the elements of $\mathbf{H}_w$ are obtained directly from a Rayleigh (or Ricean) distribution. In case we need to take into account the frequency and the time correlations, we assume that each element in $\mathbf{H}_w$ comes from a frequency and time sampled correlated SISO channel simulated according to the different channel models exposed in Section 2.1.4. Hereafter, in Eq. (2.38) and (2.39) and examples of

the correlation values between the antenna pairs in case of Uniform Linear Arrays (ULA) are given for four (transmit or receive) antennas having an antenna separation of 4λ at the Base Station, and λ/2 at the Mobile Station (base line test scenario in [16]).

$$R_H^{BS} = \begin{pmatrix} 1 & 0.0113 + j0.0048 & 0.0044 + j0.0053 & -0.005 + j0.002 \\ 0.0113 - j0.0048 & 1 & 0.0076 + j0.0066 & 0.0014 + j0.006 \\ 0.0044 - j0.0053 & -0.005 - j0.002 & 1 & 0.011 + j0.0041 \\ -0.005 - j0.002 & 0.0014 - j0.006 & 0.011 - j0.0041 & 1 \end{pmatrix}, \quad (2.38)$$

$$R_H^{MS} = \begin{pmatrix} 1 & -0.3205 + 0.0182 & 0.2254 - j0.0082 & -0.1876 + j0.0061 \\ -0.3205 - 0.0182 & 1 & -0.3204 + j0.0188 & 0.2262 - j0.0084 \\ 0.2254 + j0.0082 & -0.3204 - j0.0188 & 1 & -0.3198 + 0.01963 \\ -0.1876 - j0.0061 & 0.2262 + j0.0084 & -0.3198 - 0.01963 & 1 \end{pmatrix}. \quad (2.39)$$

The reader is referred to Appendix A in [16] for further details regarding how to obtain these antenna correlation matrices as a function of the azimuth angle spread, the angle of departure and arrival, as well as the distance between the antenna elements.

2.1.6. Uncorrelated (SISO and MIMO) channel models

Usually during the system design, it is necessary to use simple channel models that can be analytically studied in order to obtain insights on the channel effects. For this reason several times an ideal uncorrelated Rayleigh channel is used during the simulations hence the theoretical performance can be easily compared to the results obtained from simulations. Furthermore, the use of analytical channels usually requires less simulation runs in order to obtain stable results. The most used channel model is the i.i.d. Rayleigh also referred as uncorrelated Rayleigh. Other distribution types as the Ricean or the Nakagami-*m* can be used for such purpose, but generally the Rayleigh is preferred since it represents in a certain way the worst-case (non-line of sight).

In the uncorrelated Rayleigh case, the complex channel values that each (signal) sample experiences are drawn from a circularly symmetric complex Gaussian distribution with unitary variance. When this model is applied in SISO channels, it is implicitly assumed that the channel is uncorrelated in time and frequency (which might be accomplished by a sufficiently large time and frequency interleaving). On the other hand, for MIMO channels it is assumed that the channel is uncorrelated also in the spatial domain; in this case all the elements in the MIMO channel matrix are also drawn from an i.i.d. Rayleigh distribution.

2.2. Transceiver schemes and multiple access techniques

It is expected that the forthcoming 4th generation wireless and cellular networks will achieve data transmission rates up to 1Gbps in nomadic circumstances and 100Mbps in mobile (vehicular) ones [23][24]. In this section different air interfaces proposed (some of them currently in use) for wireless broadband access systems are presented. Most of the schemes proposed for such networks are based on Orthogonal Frequency Division Multiplexing (OFDM) which is detailed in Subsection 2.2.2. However, some of the proposed schemes combine OFDM with spread spectrum techniques, and thus, a few details about spread spectrum have been

included accordingly. The following subsections give an overview of the structure of the most important schemes.

For the sake of clarity, the signals transmitted by each scheme are illustrated according a three dimensional time, frequency, and power density representation. The different signals are represented as cubes that are span in time and frequency direction, where the height represents the transmitted energy (Figure 2.5a). However, since it is impossible to confine all the energy in a limited bandwidth during a limited period, the real signal representation should be like in Figure 2.5b and Figure 2.5c depending whether the signal is limited in time or in frequency domain. The duration of each data symbol is T_d, and in consequence the data symbol rate is equal to $1/T_d$. For convention, the energy of each data symbol is normalized to one for all the schemes hence comparisons between them can be done.

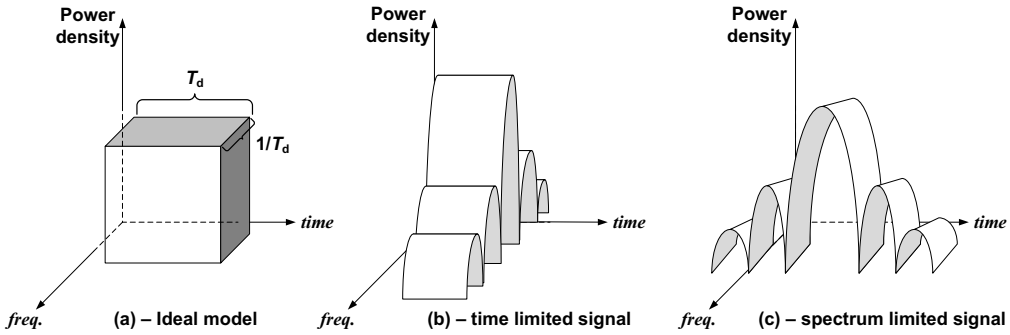


Figure 2.5. Three-dimensional time/frequency/power density representation

2.2.1. Spread Spectrum Techniques

Spread spectrum (SS) techniques were first used in military communications where privacy was the more relevant issue. Since then, they have been used in many applications due to its inherent privacy and capacity to cope well with interferences and multipath channels [25][26]. The first systems applying spread spectrum principles are dated during the mid 1950s. This technology has been developed since then and was already adopted by the 2nd generation Interim Standard (IS)-95 [26] and recently as the medium access technique for the 3G wireless systems (UMTS) [27]-[30]. One of the main benefits of using spread spectrum in cellular networks comes from the fact that neighbouring cells may use the same frequencies thus frequency planning becomes less complex than in previous TDMA-FDD based networks (e.g. GSM). However, the frequency reutilization also implies that power control algorithms as well as call admission control have to be applied to minimize multi-user interference [31].

The principle of spread spectrum techniques is to spread a signal in the frequency domain occupying a much wider bandwidth than minimum required for its transmission [32][33]. Spreading can be accomplished by multiplying the original signal with a wideband signal (code) which is independent from the data, named Direct Sequence Spread Spectrum (DS-SS), or by continuously changing the carrier frequency following a specific hopping pattern, named Frequency Hoping Spread Spectrum (FH-SS). The basic principles of DS-SS are represented in Figure 2.6, where each spreading code consists on L chips of duration T_c ($T_c << T_d$) each. The relationship between the chip rate and the data symbol rate is then the spreading factor, L, or referred as the processing gain ($P_G = T_d/T_c$).

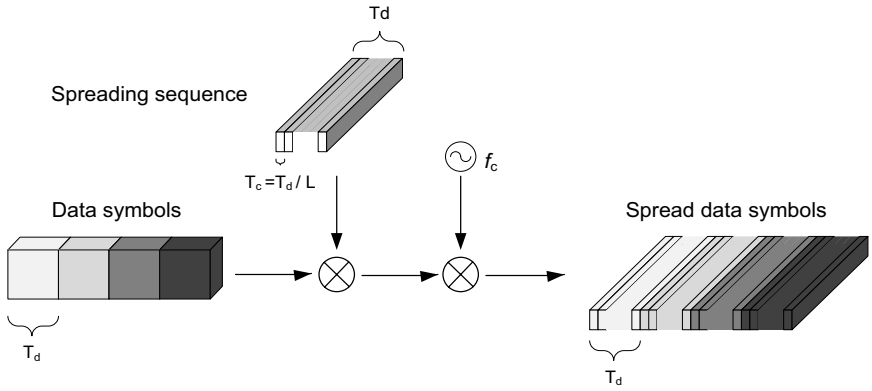


Figure 2.6. Direct Sequence (DS) spreading principles for a single user scenario

For any spread spectrum scheme, the ratio between the information signal bandwidth, W_S, and the bandwidth after spreading, W_{SS}, is known as the processing gain P_G which is defined by,

$$P_G = \frac{W_{SS}}{W_S}. \tag{2.40}$$

Therefore, in case the processing gain is increased and the total transmitted power remains constant, the power spectral density becomes lower. Both techniques can also be combined, i.e. an hybrid DS/FH which applies DS in a relatively wide band (but feasible in the practice), and then applying a FH pattern transmitting the signal it over different bands or channels, achieving even larger processing gain without excessive implementation cost [34].

In case the different active users need to share the channel, they can use different spreading code words or different frequency hopping patterns. In next sub-section, the DS-CDMA scheme is further analyzed since it will be later combined with multicarrier techniques.

2.2.1.1. Direct Sequence Code Division Multiple Access (DS-CDMA)

Using the DS-CDMA scheme, the signal information of each user k is multiplied by its respective codeword $c^{(k)}(t)$. These spreading codes must have good auto- and cross-correlation properties since information signal has to be equally spread in all the assigned bandwidth. Moreover, the spreading sequences or codes should have also a quite flat power spectral density, similar to a white noise signal. In consequence, these sequences are often called pseudo-random noise (PN) sequences [32][33]. Each sequence is then made up from L chips, c_l, each of duration T_c and are defined as,

$$c^{(k)}(t) = \sum_{l=0}^{L-1} c_l^{(k)} p(t - lT_c) \tag{2.41}$$

where $p(t)$ is a rectangular pulse equal to 1 for $0 \le t \le T_c$, and 0 otherwise. After the original signal spreading, the signal belonging to user k is expressed as follows

$$x^{(k)}(t) = d^{(k)} \sum_{l=0}^{L-1} c_l^{(k)} p(t - lT_c), \quad 0 \le t \le T_d \tag{2.42}$$

where $d^{(k)}$ is the transmitted data symbol and T_d the data symbol duration ($T_d = LT_c$). Since information and spreading sequences are synchronized, and assuming users are synchronized

too (i.e. in the downlink), the signal from all users $k=\{0, ..., K-1\}$ (where K is total number of active users) is given by

$$x(t)=\sum_{k=0}^{K-1}x^{(k)}(t) \tag{2.43}$$

The received signal $y(t)$ can be expressed as

$$y(t)=x(t)\otimes h(t)+n(t)=r(t)+n(t)=\sum_{k=0}^{K-1}r^{(k)}(t)+n(t). \tag{2.44}$$

At the receiver side, a matched filter (MF) is applied, thus the signal before the detector becomes

$$\begin{aligned} z^{(k)}(t) &= y(t)\otimes h_{MF}^{(k)}(t)=y(t)\otimes\left(c^{(k)*}(-t)\otimes h^{*}(-t)\right)= \\ &= \underbrace{r^{(k)}(t)\otimes h_{MF}^{(k)}(t)}_{\text{desired symbol}}+\underbrace{\sum_{\substack{g=0\\ g\neq k}}^{K-1}r^{(g)}(t)\otimes h_{MF}^{(k)}(t)}_{\text{multiple access interference}}+\underbrace{n(t)\otimes h_{MF}^{(k)}(t)}_{\text{noise}}. \end{aligned} \tag{2.45}$$

In Eq. (2.45) the signal at the receiver contains, the desired symbol, the multiple access interference (MAI), and the noise component. The MAI term is determined by the PN sequences cross-correlation properties and the power control algorithms (the signals received with higher power will interfere more). Ideally, the MF resolves all multipath propagation of the channel. In practice, a good approximation of the MF is the rake receiver represented in Figure 2.7 which combines the signal from the main D paths or echoes (i.e D = 3–4 arms). However, recovering coherently the energy transmitted by each channel might become very challenging and computational complexity demanding [35][36] when several paths need to be combined.

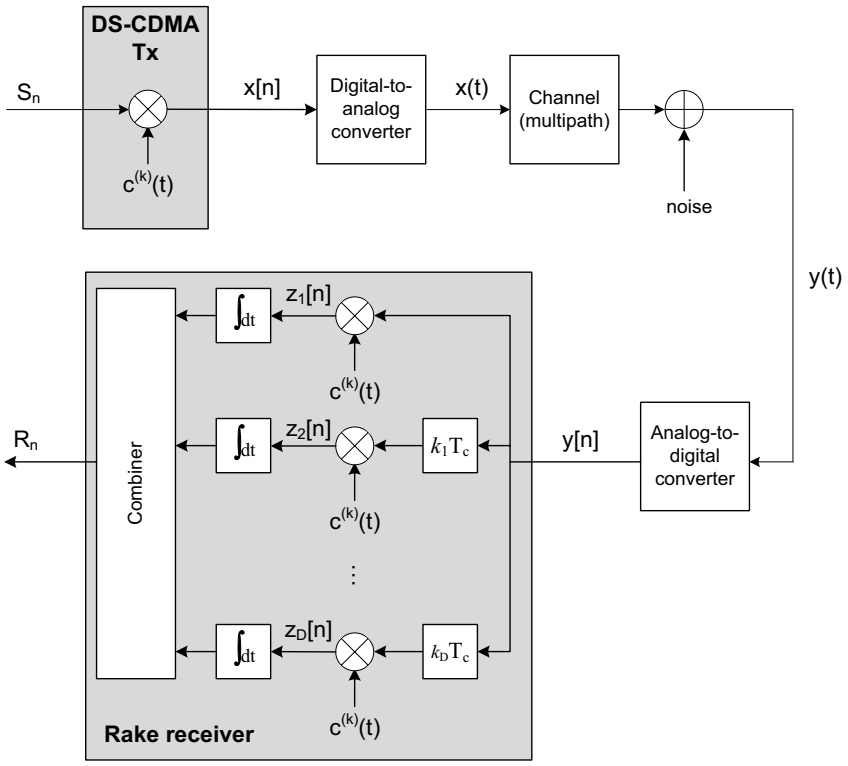


Figure 2.7. DS-CDMA system block diagram with a rake receiver.

As it was already mentioned, one of the major advantages of the spread spectrum scheme in mobile communication networks is the frequency reuse which eases the network deployment and the flexibility in rate adaptation by simply modifying the spreading factor [29]. When spread spectrum is applied in multiuser environments, the multiple access interference becomes the main concerning issue, and the power control algorithms turn out to be essential. Nevertheless, the effects of the multipath channel on the spread spectrum system are diminished since the occupied bandwidth is much wider than the coherence bandwidth. The signal is the provided with intrinsic frequency diversity.

2.2.2. Orthogonal Frequency Division Multiplexing (OFDM)

The multicarrier based transmission scheme (already proposed during the 60s) is becoming the basic framework over which the latest wireless standards are now implemented [39]. The requirement for higher data rates implies that the symbol duration is reduced. When a symbol is transmitted through a multipath channel, different copies of the same symbol may span over intervals larger than one symbol period. These copies can affect several symbols causing high Inter-Symbol Interference (ISI) and in consequence the equalization of the signal becomes extremely challenging. In contrast, whether the data stream is divided into N_c different sub-streams with lower data rates (N_c/T_d), these symbols will be much less affected by the ISI since the percentage of interfered symbol will be reduced in the same factor N_c. This principle is shown in Figure 2.8 for N_c=4. However, splitting the main stream implies that different parallel transmissions must be later performed. The transmission of each sub-stream in a certain f_n frequency is exactly what is done using an FDM access scheme. However in the FDM case due to the guard bands inserted between the streams to minimize the Inter-Carrier Interference (ICI), the spectral efficiency is reduced. On the other hand, applying the Orthogonal Frequency Division Multiplexing (OFDM) scheme, different sub-streams can overlap between themselves reducing the required bandwidth more than 50% compared to the classical FDM scheme [38]. When time-limited symbols are used, orthogonality between sub-streams can be obtained simply by choosing subcarriers with an integer number of cycles within the symbol duration ($f_k=k/T_d$, $k\in\mathbb{N}$). If the minimum spacing between sub-carriers becomes the inverse of the symbol duration a highest spectral efficiency is obtained. The block of N_c data symbols transmitted during T_s seconds is then referred as an OFDM symbol where the data symbol rate is $1/T_d=N_c/T_S$. T_s is then the OFDM symbol duration. By increasing the number of subcarriers while having a fixed T_s, the time-variant power spectral density approaches an ideal rectangular shaping in both in time and frequency [39]. Furthermore, in case the transmitter adapts the power,the modulation, and the coding rate of the transmitted symbols over each subcarrier, the transmission rate may become very close to the Shannon channel capacity [40]. Different adaptation mechanisms for such purpose will be studied and analyzed in Chapter 4 and Chapter 5.

In order to implement the OFDM modulation, the Fast Fourier Transform (FFT) is used due to its very low complexity and is applied in most of the OFDM system transmission schemes [38]. Furthermore, if a guard interval is added at the beginning of each symbol the ISI can be completely avoided even a decrease of the spectral efficiency is experienced. Usually, this guard interval lasts for 5-25% of the OFDM symbol and in order to keep the subcarrier

orthogonality a copy of the symbol tail is appended at the beginning of the symbol, this operation is known as cyclic prefix extension (i.e. CP-OFDM).

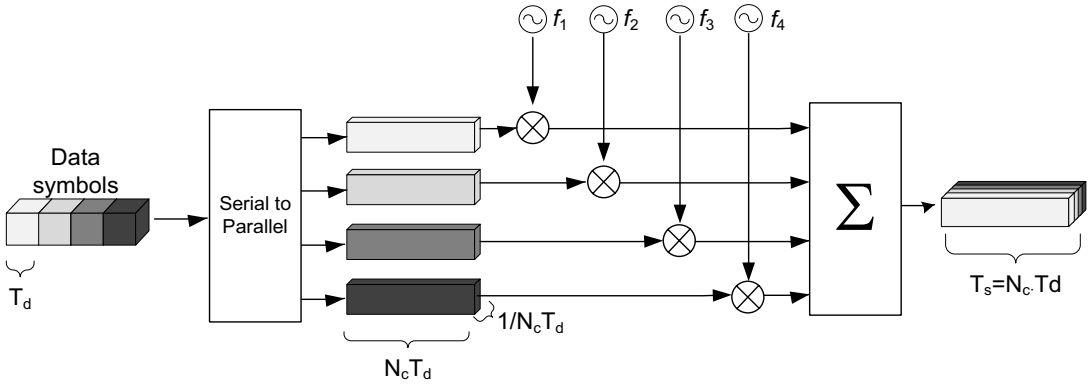


Figure 2.8. Orthogonal Frequency Division Multiplexing (OFDM) transmission schematic

Another advantage of the OFDM scheme is its suitability for broadcast transmissions. In case of having synchronized transmitting stations, the interference from neighbouring cells is experienced as an artificial multipath [41][42]. This has lead to the *Single Frequency Network* (SFN) concept where the transmitted signals by the different stations are exactly the same irrespective of the transmitter location. In such case the active users at the cell border are those more benefited from SFN, since the signals coming from the different stations have experienced same order delay and are received with similar power level. Hence very high frequency and time diversity may be achieved.

The high difference in the Signal to Noise Ratio (SNR) values between the subcarriers forces the transmitted signal to be encoded and interleaved in order to cope with the large fading fluctuations. Thus the system is usually referred as Coded OFDM (COFDM). Furthermore, the maximum channel diversity might be achieved whether soft-decision metrics (i.e. SNR per subcarrier) is taken into account within the channel decoder [43].

Nevertheless, if the channel is considered as non-variant during each symbol (i.e. the OFDM symbol duration T_s is shorter than the coherence time) with a flat fading over each sub-carrier (i.e. the bandwidth of each sub-stream is less than the coherence bandwidth), the received signal can be easy equalized by multiplying each symbol on each sub-carrier by the corresponding inverse channel gain. This complexity reduction at the receiver side as well as its good performance in broadband multipath channels has inclined many of the nowadays systems to use the OFDM scheme. Among others, we may emphasize the use of OFDM in the Wireless Local Area Networks (WLAN) [44], the IEEE 802.16a standard for Wireless Metropolitan Area Networks (WMAN) [45], the Digital Audio and Video Broadcasting (DAB, DVB) [46][47], or the Asynchronous Digital Subscriber Line (ADSL) [48] due to its great success and broad recognition.

2.2.2.1. OFDM system architecture

In the OFDM scheme, the subcarriers over which each sub-stream is transmitted are orthogonal, being the minimum separation between consecutive sub-carriers

$$\Delta f = \frac{1}{T_S}, \tag{2.46}$$

where T_S is the OFDM symbol duration delimited by a rectangular pulse shape. If the number of sub-carriers is N_C, then the transmitted OFDM symbol takes the following forms in time, $x(t)$, and in frequency , $X(f)$, respectively

$$x(t) = \frac{1}{\sqrt{N_C}} \sum_{k=0}^{N_C-1} S_k \cdot e^{j2\pi f_n t}, \quad 0 \le t \le T_S \tag{2.47}$$

$$X(f) = \frac{1}{\sqrt{N_C}} \sum_{k=0}^{N_C-1} S_k \cdot \frac{\sin\left(\pi\left(T_S f - k\right)\cdot\right)}{\pi\left(T_S f - k\right)\cdot} \tag{2.48}$$

where S_k are the complex baseband modulated symbols (i.e. phase and amplitude modulated - QAM), and

$$T_S = N_c \times T_d. \tag{2.49}$$

The following Figure 2.9 shows an example of an OFDM signal where 8 subcarriers have been modulated. The central subcarrier, referred as DC subcarrier, does not bear any information hence the receiver can be implemented using direct conversion. Furthermore, it can be observed that due to the windowing of each symbol, the spectrum of each subcarrier spans over its neighbouring subcarriers following a *sinc* shape, presenting a null on others subcarriers frequency.

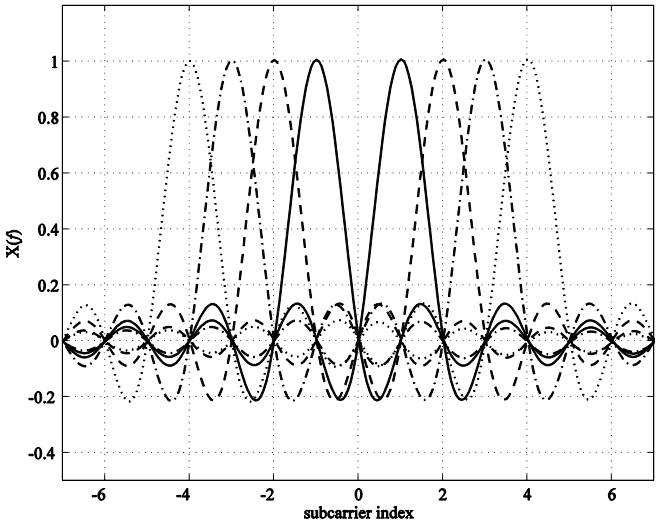


Figure 2.9. Representation of an OFDM signal in the frequency domain.

For large values of N_C the power density spectrum becomes flatter in the region that contains those sub-carriers, whilst for any value of N_C the power density spectrum falls more than 10dB just Δf Hz above or below the last or first sub-carrier respectively. In fact it has been demonstrated that at $N_C \cdot \Delta f/2$ the power density spectrum falls more than 20dB [6].

Despite the OFDM scheme was already proposed during the 60s, it was not until the late 80s when it was possible to apply the IFFT/FFT algorithm with a relative low cost digital signal processing components. A typical OFDM transmitter and receiver block diagram is depicted in Figure 2.10.

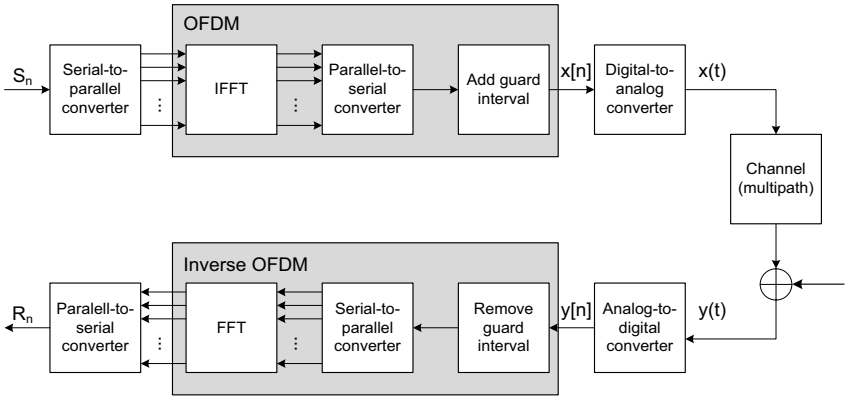


Figure 2.10. Digital OFDM block diagram.

The discrete version of the OFDM signal can be written as follows,

$$x[n] = \frac{1}{\sqrt{N_C}} \sum_{k=0}^{N_C-1} S_k \cdot e^{j2\pi nk/N_C}, \quad 0 \le n \le N_C - 1 \tag{2.50}$$

where the minimum sampling period that fulfils Nyquist is $T_m = T_S / N_C$.

Moreover, whether we increase N_C while the occupied bandwidth remains constant, the OFDM symbol duration T_S becomes larger compared to the delay spread and the amount of ISI gets reduced. Furthermore, to minimize the effects of the ISI, a guard interval can be inserted between the symbols. To maintain orthogonality between subcarriers instead of inserting a silence period, a cyclic extension of each OFDM symbol is copied at the guard interval. The scheme using the cyclic prefix guard interval is also known as Cyclic Prefix (CP-) OFDM. This cyclic prefix is then a replica of the last part of each OFDM symbol. More details about this process as well as benefits and drawbacks of using a cyclic prefix extension can be found in [49].

Afterwards, the digital signal is converted to the analog domain and transmitted over the communications channel. The received signal at the MS can be written as,

$$y(t) = \int_{-\infty}^{\infty} x(\tau) \cdot h(t-\tau) \partial\tau + n(t) = x(t) \otimes h(t) + n(t) \tag{2.51}$$

where $n(t)$ is the Additive White Gaussian Noise (AWGN) at the receiver side, and $\otimes$ denotes the convolution operation. Then this signal is converted to the digital domain and after having done proper synchronization, the samples belonging to the guard interval can be perfectly removed. Applying the FFT, the signal will be again translated to the frequency domain giving,

$$R_k = \sum_{n=0}^{N_C-1} y[n] \cdot e^{-j2\pi kn/N_C}, \quad k = 0, \ldots, N_C - 1. \tag{2.52}$$

Since the Inter-carrier interference (ICI) and the inter-symbol interference (ISI) can be considered negligible in case of perfect synchronization, and with a channel delay spread shorter than the guard interval (GI), each sub-channel can be considered independently. Thus, the received symbols on each sub-channel can be expressed as following the next linear expression

$$R_k = H_k S_k + N_k, \quad k = 0, \ldots, N_C - 1, \tag{2.53}$$

where N_k represents the noise on the k-th sub-carrier and H_k is the flat fading component. The time variable has been omitted for simplicity (i.e. $H_k = h(k\Delta f, t)$).

We can observe that in order to estimate the received symbols, the received signal can be equalized by applying the inverse function of the channel, in that case the amplitude and phase of the symbol is perfectly restored. It is not the focus of this thesis to go into the channel estimation techniques in depth, however some details about channel estimation and interpolation are given in Appendix D as well as the details of an interpolation scheme proposed by the author in [50]. Hence, the estimated symbols are given by

$$\hat{S}_k = \frac{R_k}{H_k} = S_k + N'_k, \quad k = 0, ..., N_C - 1. \tag{2.54}$$

where N'_k is a coloured noise whose standard deviation is the same as for N_k.

Furthermore, whether the ICI and the ISI are considered null, the system may be considered as a discrete time and frequency system with N_C parallel Gaussian channels with different complex value attenuations $H_{k,m}$ as represented in Figure 2.11, where the indexes k,m indicate the sub-carrier and OFDM symbol index respectively.

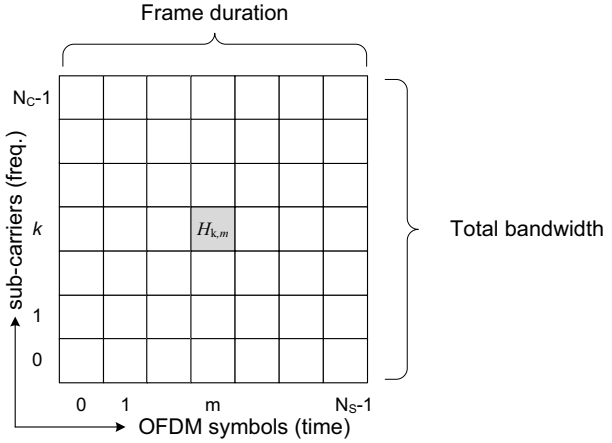


Figure 2.11. Time/Frequency representation of an OFDM frame.

Due to the time/frequency representation exposed in Figure 2.11, the OFDM transmission can also be mathematically described using matrices and vectors. Under this notation, each OFDM symbol can be expressed as a vector of N_C complex values,

$$\mathbf{s} = \left(S_0, S_1, ..., S_{N_C-1}\right)^T, \tag{2.55}$$

whereas the channel is expressed as a diagonal N_C x N_C matrix $\mathbf{H}$ that in absence of ISI and ICI can be expressed as

$$\mathbf{H} = \begin{pmatrix} H_{0,0} & 0 & \cdots & 0 \\ 0 & H_{1,1} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & H_{N_C-1,N_C-1} \end{pmatrix}, \tag{2.56}$$

while the following vector

$$\mathbf{n} = \left(N_0, N_1, ..., N_{N_C-1} \right)^T \tag{2.57}$$

represents the additive white Gaussian noise. With this notation the received OFDM symbol is given by

$$\mathbf{r} = \left(R_0, R_1, ..., R_{N_C-1} \right)^T = \mathbf{Hs} + \mathbf{n}. \tag{2.58}$$

where R_k are the data symbols transmitted within.

Whether the number of data symbols that must be transmitted exceeds N_c, several OFDM data symbols can be concatenated giving an OFDM frame. Actually, one of the main characteristics of the OFDM is that; all the transmitted symbols within the same OFDM symbol have the same destination. In other words, only one stream of data is multiplexed per OFDM symbol. In case of multiple users, some OFDM based schemes accept to multiplex different users by combining OFDM and TDMA access mode [51]. In this case, the minimum allocable Resource Unit (RU) is equal to one OFDM symbol.

To sum up, some of the advantages of using OFDM are; its high spectral efficiency,the minimization of the ICI, and ISI effects (although efficiency is reduced by the guard interval). Moreover, since each symbol is independent from the others, the amount of bits per symbol or the power assigned to each sub-carrier can be estimated in order to maximize the system capacity. Finally, the IFFT/FFT based implementation of the OFDM transmitter and receiver has reduced its computational complexity leading to low cost designs. However, one of the main drawbacks of the OFDM scheme is the high peak-to-average power ratio (PAPR) effect, which for the uplink transmission becomes a limiting factor since the amplifiers must work in the linear zones (degrading the power efficiency). Last but not least, the design of the time and frequency synchronization processes will become crucial thus ISI and ICI do not ruin the system performance.

2.2.3. Orthogonal Frequency Division Multiple Access (OFDMA)

The Orthogonal Frequency Division Multiple Access (OFDMA) scheme can be interpreted as the multiuser access version of the previously reviewed OFDM scheme. The OFDMA scheme was developed to move the Orthogonal Frequency Division Multiplexing (OFDM) scheme from a fixed-access wireless system to a true cellular system with mobility. Nevertheless, it has been also observed that the OFDM-TDMA scheme in combination with the Carrier Sense Multiple Access Collision Avoidance (CSMA-CA) protocol as it was proposed in the Distributed Coordination Function (DCF) of the IEEE 802.11 standard [44], becomes very inefficient as the number of users is increased [51]. Even, if the BS assigns periodically the resources in a TDMA manner as proposed in the Point Coordination Function (PCF), the latency remains quite high to cope with simultaneous real time applications. In order to solve these problems, the OFDMA scheme is able to multiplex several users both in time and frequency domains simultaneously without a lag due to the synchronization processes inserted between users in the OFDM-TDMA scheme.

In the last five years the number of systems and standards for BWA using OFDMA has grown rapidly. We have already introduced the IEEE 802.16 (WiMAX) and the 3GPP-LTE as the main players, however other systems for BWA also use the OFDMA scheme, here we mention some

of them: the Ultra Mobile Broadband (UMB) [53], the IEEE 802.20 also referred as Mobile Broadband Wireless Access (MBWA) [54] which is in fact pretty similar to the Fast Low Latency with Seamless Handover (FLASH-) OFDM system [55][56], the IEEE 802.22 [57], etc. Despite of that, each of the mentioned standards is designed for a specific segment (fixed, mobile, metropolitan areas, backhaul, rural areas, etc.) and business model, the main difference between them is essentially not at the physical layer but at the Medium Access Control (MAC) layer and its relationship with the higher layers. We then emphasize that most of the above standards are based on the same transmitter and receiver architecture structure, i.e. the OFDMA scheme. However, since some of them are more oriented to fixed/high mobility and multiple cell networks, or IP packet transmissions, the OFDMA frame partitioning arise as the main differences in the air interface.

The difference between OFDM and OFDMA are illustrated in the following Figure 2.12. In OFDMA, subcarriers are grouped into larger units, referred as subchannels, and these subchannels are further grouped into bursts which can be assigned to the different active users. Assuming an underlying OFDM scheme with N_C subcarriers, the OFDMA scheme packs the subcarriers into subchannels where each subchannel is a subset of several subcarriers extracted from the total set of available subcarriers. The subcarriers assigned to each subchannel may be distributed along the channel spectrum, or belong to a certain part of the spectrum (thus they are contiguous in the frequency domain in the latter). Figure 2.13 depicts an example of how the different subcarriers are mapped according to the *Interleaved* or the *Localized* subcarrier permutations.

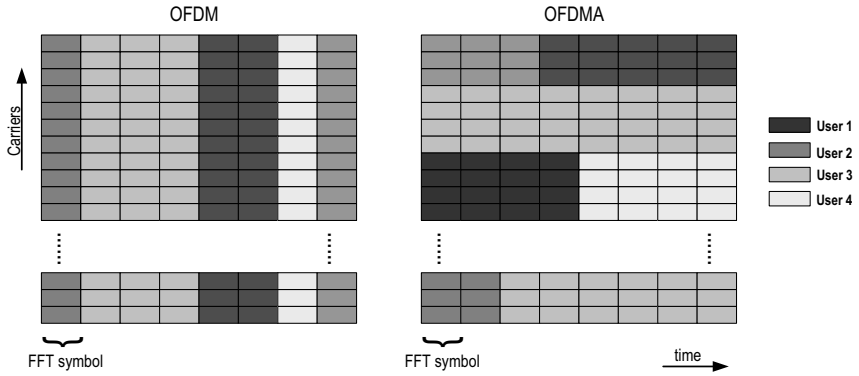


Figure 2.12. OFDM vs OFDMA schemes.

In the *Distributed* or the *Interleaved OFDMA* mapping scheme, the usable subcarrier space is divided into N_G successive groups containing N_E subcarriers each one (hence N_C=N_G×N_E). Then, each subchannel has an assigned subcarrier from each group following a pseudo-random process based on permutations, where N_G is equal to the number of subchannels. The main advantage using these subcarrier permutations is that the frequency diversity is increased, thus the link reliability is increased too. This type of transmission is desirable for high speed users, where difficultly the link adaptation and scheduling algorithms can follow the continuous channel changes [58]. Note that here the main objective is to maximize the frequency diversity. Furthermore, the subcarriers mapped to each subchannel may change from the actual OFDM symbol to the next one, the interferences are then averaged across all the subchannels like in a FH-SS scheme (see 2.2.1). In fact, the transmitted signals within the

same cell are synchronous being orthogonal between them. Thus, only the intercell interference is present at the receivers.

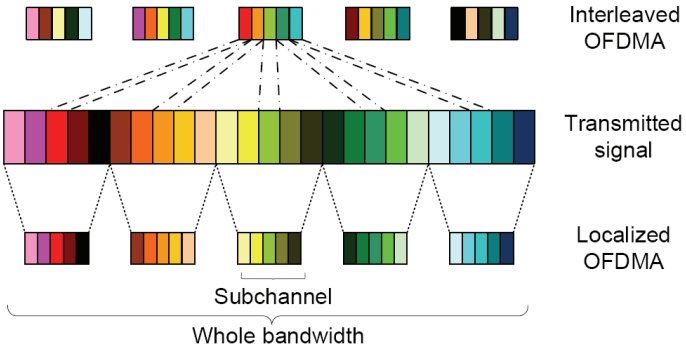


Figure 2.13. Interleaved and localized subcarrier mapping schemes in OFDMA IEEE 802.16e standard.

In some cases, it is preferable to map or to distribute each subchannel into a certain or specific parts of the spectrum. In this case, all the N_E subcarriers within a group are mapped directly over one subchannel. This special map is known as "*Blocked"* or "*Localized" OFDMA* structure. Then, the radio channel over the subchannel is highly correlated and the diversity order is reduced. However, in the localized scheme the optimum modulation and coding scheme can be obtained achieving a high spectral efficiency.

In a multiuser scenario, it is possible to combine the multiuser diversity with frequency scheduling leading to very high spectral efficiencies as it will be demonstrated in the forth-coming Chapter 4 and Chapter 5. However, the *Localized OFDMA* scheme is only appropriate for low speed mobility environment where the channel variations within a frame interval are not significant. Therefore the BS might have reliable channel information to apply a channel aware scheduling.

Either in the Distributed and the Localized OFDMA schemes, several streams are multiplexed in frequency (Frequency Division Multiple Access - FDMA) and time (Time Division Multiple Access - TDMA) domain in a very efficient way. Nevertheless, users far from the base station may concentrate their power into one subchannel, thus they occupied $1/N_G$ of the whole channel bandwidth. As a result, the OFDMA scheme is also beneficial to obtain wide coverage areas, achieving connectivity at distances of tens of kilometres [45].

In Figure 2.14 the OFDMA frame format in the IEEE 802.16 standard [45] is depicted. It is shown that the bandwidth is divided into several subchannels whereas the time domain is divided into several OFDM symbols. The minimum physical Resource Unit (RU) is defined as the basic radio unit for resource allocation hence the data transmitted within a RU has a unique source and one (unicast) or many destinations (multicast). In addition, the data transmitted within a RU (also called slot) follows the same modulation, coding and power. The shape of the RU is defined over both frequency and time domains usually following a rectangular shape. An example of this shape distribution is depicted in Figure 2.14, each RU is composed of by 24×3 subcarriers in frequency and time domains respectively. In order to reduce the signaling load the RUs are grouped into bursts or data regions. All the RUs within

the same burst apply the same modulation and coding, the same source and address, etc. Hence, the required signaling is a function of the number of bursts and not the RUs.

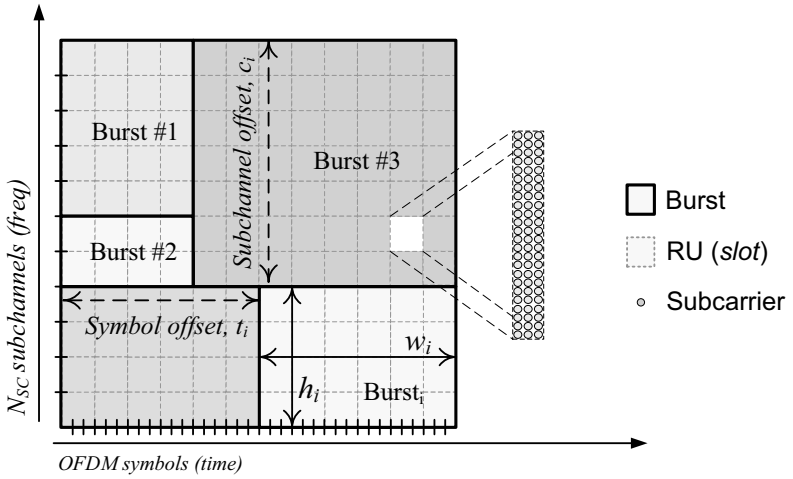


Figure 2.14. OFDMA frame format.

The frame is further divided following a TDM or a FDM scheme into the different permutation zones (mainly *scattered* subcarriers per RU, or *contiguous* subcarriers per RU) [59]. Recently, some hybrid schemes have been proposed which may push even further the system efficiency as well as the flexibility of the resource allocation processes [60]-[64]. In these proposals (largely further discussed in Chapter 5) the scattered and the contiguous subcarriers permutation are combined and in order to achieve it, two new concepts named Virtual Resource Block (VRB) and Physical Resource Block (PRB) have been introduced. A PRB is very similar to a RU with the difference that for a PRB the subcarriers are always consecutive, thus the PRB has a structure based on $N_f \times N_t$ consecutive subcarriers in frequency and in time respectively. Then, a VRB is defined according to the following attributes:

- Size; measured in terms of allocated time-frequency resources,
- Type; *localized* or *distributed*.

The size of the Distributed VRB (DVRB) can be different from the size of the Localized VRB (LVRB). The DVRBs are mapped onto the PRBs in a distributed manner, while frequency hopping in the DVRB mapping can be applied to randomize the intercell interference effect. The LVRBs are mapped onto the PRBs in a localized manner. Two options for multiplexing distributed and localized VRB blocks have been proposed in [61]-[62] and illustrated in Figure 2.15. In *option 1* (left side of the figure), the distributed VRB blocks are first allocated, and then the remaining RUs are allocated using a localized distribution. In *option 2*, the localized VRB blocks are allocated first, and then the remaining RUs are mapped using the distributed manner. Hence under the option 2 the diversity of the Distributed VRB may be severely affected even when the Localized VRB blocks span through large bandwidths. In case of having half the bandwidth for Distributed VRB mapping, a reduction of the performance degradation of 0.3dB to 0.5dB has been reported in [63].

The use of these hybrid schemes is desirable when the number of active users with high-mobility is small compared with those with low mobility. When the number of high-mobility

active users is comparable to those with low-mobility, a TDM multiplexing scheme with pure LVRB and DVRB mapping would be more than desirable [62].

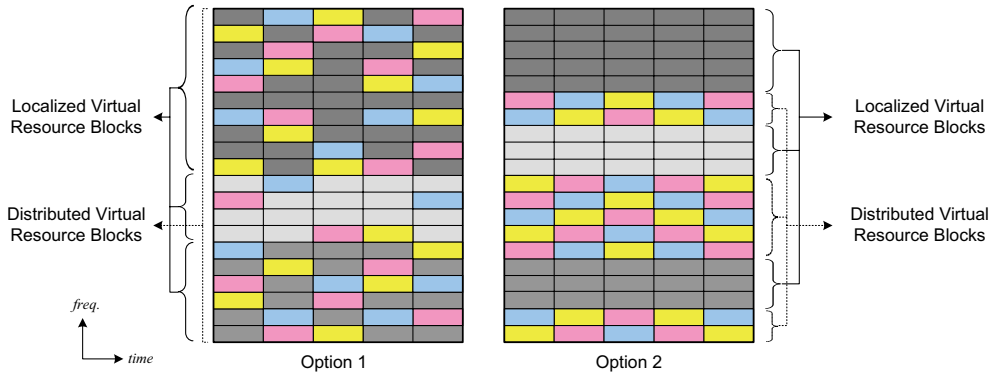


Figure 2.15. Multiplexation options for distributed and localized allocations.

Since the scheduling and the minimum resource allocation block (i.e. PRU or VRB) may change from frame to frame, each user´s data is signalled with its exact frequency and time position at the beginning of each frame within the Media Access Protocol (MAP) message. Thus, some amendments and limitations to the resource allocation process are introduced in order to reduce the signalling without decrease too much the system performances. For instance, in the 3GPP-LTE downlink mode, both the localized and the distributed virtual resource blocks can be used following a hybrid scheme. However if one user is assigned a LVRB or DLRB block, the allocation is fixed for an entire downlink subframe (the downlink subframe duration is 0.5ms, equivalent to 6 or 7 OFDM symbols) [64]. However, the IEEE 802.16 standard proposes to group the VRB blocks of each user into larger structures named "*Data Regions*" or "*bursts*" [45], which are represented in Figure 2.14. The unique restriction to group the VRB blocks is that they must be contiguous according to their logical indexes, and follow a rectangular shape. In that case, instead of signalling each of the VRB block within the frame, only the burst shape (frequency and time dimensions) as well as it position within the frame needs to be signalled. By this signalling reduction strategy larger frames (i.e. 2-20ms) can be used while the resource allocation process is kept very flexible.

2.2.4. Filter Bank Multicarrier (FBMC) scheme

The OFDM modulation with a cyclic prefix as a guard interval (CP-OFDM) is well known for its robustness to multipath time varying channel propagations, and for its wide use in several wireless communications standards as mentioned in previous sections. To remove this guard interval, the prototype function modulating (or filtering) each sub-carrier must be very well localized in the time and in the frequency domain to limits the inter-symbol and the intercarrier interferences. This function must also guarantee the orthogonality between the sub-carriers [67][68]. An alternative to the CP-OFDM is the OFDM with Offset QAM (OFDM-OQAM) modulation which has the advantage to not require the use of a guard interval, which leads to a gain in the spectral efficiency [69]. However, the OFDM-OQAM requires an independent prefiltering over each subcarrier, thus large filter banks are required.

The transmission and reception scheme of a Filter Bank-based Multicarrier system is shown in Figure 2.16. The difference compared to the regular OFDM signal modulation turns out on the *Synthesis Filter Bank* (SFB) and the *Analysis Filter Bank* (AFB) placed after the IFFT and before the FFT blocks respectively. These filters consist on a set of digital filters whose coefficients form the impulse response of the used prototype low-pass filter. These filters, aim to model the spectrum shape of each subcarrier moving from the original *"sinc"* shape function in OFDM to a shape function where each subcarrier spreads the energy in a much narrower bandwidth (see the difference on shape function in Figure 2.17).

As it is seen in Figure 2.17, the OFDM exhibits large ripples due to the rectangular time pulse shape. However as it was explained in Section 2.2.2, the amplitudes of these ripples are null at the subcarrier frequencies. However, in a filter bank multicarrier system the frequency response has negligible amplitudes beyond the center frequency of the adjacent subcarriers. Thus, the FBMC scheme offers high spectrum resolution and provides independent sub-channels, while it maintains or enhances the high data rate capability. Furthermore, this high spectral resolution can be exploited by cognitive systems where the level of interference in adjacent subchannels (i.e. primary users) must be kept into a minimum.

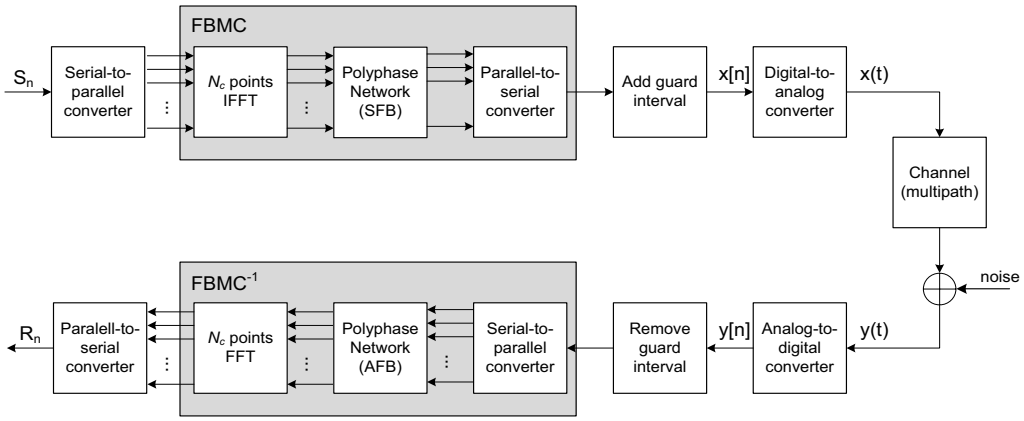


Figure 2.16. System diagram of a Filter Bank-based Multicarrier (FBMC) system.

One of the key issues regarding the FBMC is the design of the proper filter bank. Normally, uniform filter banks are considered where all the subcarriers have the same bandwidth. Efficient uniform filter banks can be implemented using various structures as lapped transforms, lattice structures or polyphase structures [70]. All these structures consist on a prototype filter design with the section which modulates the data going through the filter bank. The prototype filter must control the phase and amplitude distortions in the subcarriers as well as the interference between them. The phase distortion is eliminated whether the prototype filter is linear phase. The amplitude distortion may produce intersymbol interference.

Ideally, the SFB and the AFB should be able to perfectly reconstruct the signal and also without introducing any cross-talk between subcarriers. However, due to imperfect synchronization and also due to the radio channel effects some distortion may be present inevitably at the received signals. In that case, a perfect reconstruction is not essential, and the filter design constraints can be relaxed if the cross-talk between the subcarriers is lower than the

interference introduced by the channel impairments. Hence, for nearly perfect reconstruction, the polyphase networks are the best choice [71][72].

The prototype filter is obtained by the frequency sampling technique, where the values of the filter at certain distances from the subcarrier central frequency are specified. For instance, assuming N_C subcarriers and a oversampling factor of K=4, in [72] and [73] the filter frequency response $G(f)$ is given by

$$\begin{aligned} &G(0)=1, \\ &G\left(\frac{\pm 1}{KN_C}\right)=-0.972, \\ &G\left(\frac{\pm 2}{KN_C}\right)=\frac{\sqrt{2}}{2}, \\ &G\left(\frac{\pm 3}{KN_C}\right)=-\sqrt{1-G^2\left(\frac{1}{KN_C}\right)}=-0.235, \\ &G\left(\frac{\pm k}{KN_C}\right)=0 \quad ; \quad 4\le k\le \frac{KN_C-1}{2}. \end{aligned} \tag{2.59}$$

Hence, with (2.59) we may obtain the digital prototype filter coefficients by applying the IFFT leading to

$$g[0]=0,\ g[n]=1+2\sum_{k=1}^{K-1}(-1)^k\, G(k/KN_C)\cos(2\pi kn/LN_C) \quad \text{for} \quad 1\le n\le KN_C-1 \tag{2.60}$$

The filter transfer function $|G(f)|^2$ is depicted in Figure 2.17. It is shown that the prototype filter attenuates more than 60dB the signal at frequency ranges exceeding 2 subcarriers spacing, an even for the adjacent subcarrier the interference is lower than 40dB. Moreover, the spectrum shape of an OFDM signal is also plotted for comparison purposes. We can notice that the FBMC scheme is able to achieve a much higher frequency resolution that can be used for dynamic access, spectrum sensing, etc.

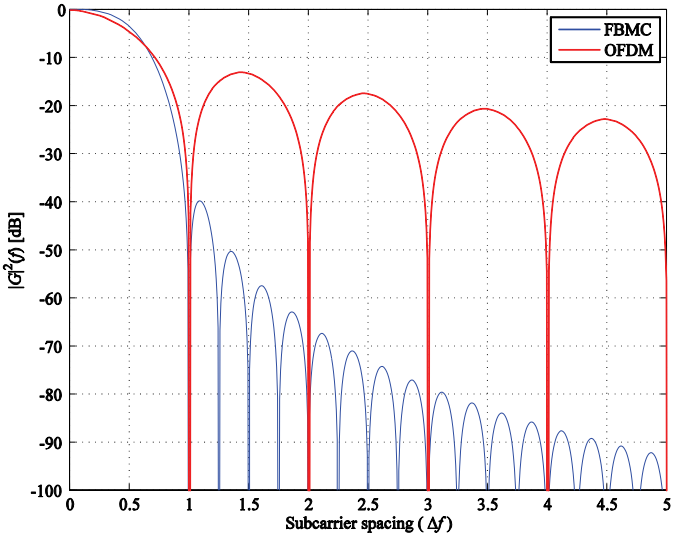


Figure 2.17. FBMC filter frequency response per subcarrier.

Following the design of the filter bank, to better understand the behaviour of a FBMC system, it is necessary to analyze how the filters can be applied efficiently as well as how the data symbols can be transmitted in order to maximize the channel capacity. First, we can write the baseband equivalent of a continuous-time FBMC signal as follows

$$x(t)=\sum_{n\in\mathbf{Z}}\sum_{k=0}^{N_C-1} a_{k,n}\times g\left(t-n\tau_0\right)\times e^{j2\pi kt/2\tau_0}\times e^{j\phi_{k,n}} \tag{2.61}$$

with N_C an even number of subcarriers, the subcarrier spacing $\Delta f=1/2\tau_0$, g is the prototype function from (2.60) and $\phi_{k,n}$ an additional phase term with

$$\phi_{k,n}=\frac{\pi}{2}(k+n)-\pi kn. \tag{2.62}$$

The transmitted symbols $a_{k,n}$ are real valued and obtained from a 2^{2P}–QAM constellation, with $P\in\mathbb{R}$. The real data symbol rate then becomes $N_C/2\tau_0$. The following Table 2.2 shows the value that the phase rotation $\phi_{k,n}$ term introduces to each symbol, where Q and $Q'\in\mathbb{N}$. We observe that for a specific subcarrier the symbols are transmitted in the real and in the imaginary parts alternatively. These phase rotations are introduced in order to relax the requirements imposed on the filter bank where according to the previous design criteria, each real data symbol should introduce a negligible interference over its adjacent (in frequency and time) data symbols.

Moreover, since the input symbols S_n to the FBMC system are complex values and only real values (i.e. $a_{k,n}$) are used to modulate each subcarrier, the input symbols are separated into real and imaginary components, transmitting each one on its pair of subcarriers, following a crossed symbol order (see example in Table 2.2). Nevertheless, despite the data rate loss caused by transmitting real data, the same data rate as in the OFDM system is obtained, since in one OFDM symbol time two real FBMC symbols values are transmitted (i.e. $\Delta f=1/2\tau_0$). This type of modulation is known as Offset (O-) QAM modulation, thus in many scientific literature the filter bank multicarrier based system is also referred as OFDM/OQAM scheme [74].

$e^{j\phi_{k,n}}$	$n=2Q'$	$n=2Q'+1$
$k=KQ$	1	j
$k=(K+1)Q$	j	1
$k=(K+1)Q$	-1	-j
$k=(K+3)Q$	-j	-1

Table 2.2. Example of the phase rotation as a function of the subcarrier and symbol index.

Then, considering a time-limited and time-variant channel impulse response $h(t,\tau)$ with maximum delay spread equal to σ, the received signal can be written as

$$\begin{aligned} y(t)&=x(t)\otimes h(t)+n(t)=\int_{-\infty}^{\infty}x(\tau)\cdot h(t,\tau)\partial\tau+n(t)\\ &=\sum_{n\in\mathbf{Z}}\sum_{k=0}^{N_C-1} a_{k,n}\times e^{j2\pi kt/2\tau_0}e^{j\phi_{k,n}}\times\int_0^{\sigma}h(t,\tau)\,g\left(t-\tau-n\tau_0\right)e^{-j2\pi k\tau/2\tau_0}\partial\tau \end{aligned} \tag{2.63}$$

where ⊗ denotes the convolution operation. On the other hand, whether the coherence bandwidth of the channel is larger than the subcarrier bandwidth, we can assume that the channel will be flat for each subcarrier. Equivalently, the prototype filter $g(t)$ has relatively low variation on time over the interval [0,σ], thus

$$g\left(t-\tau-n\tau_0\right)\approx g\left(t-n\tau_0\right) \quad \text{with } \tau\in\left[0,\sigma\right], \tag{2.64}$$

and we can rewrite Eq. (2.63) as follows

$$y\left(t\right)=\sum_{n\in\mathbf{Z}}\sum_{k=0}^{N_C-1}a_{k,n}\times e^{j\pi kt/\tau_0}e^{j\phi_{k,n}}\times g\left(t-n\tau_0\right)\times H\left(k\right) \tag{2.65}$$

which can be also be expressed as

$$y\left(t\right)=\sum_{n\in\mathbf{Z}}\sum_{k=0}^{N_C-1}a_{k,n}\times g_{k,n}\left(t\right)H\left(k\right), \tag{2.66}$$

where

$$g_{k,n}\left(t\right)=e^{j\pi kt/\tau_0}e^{j\phi_{k,n}}g\left(t-n\tau_0\right). \tag{2.67}$$

The received signal is then a composite of multiple subcarriers signals with a linear combination of time-shifted (by multiples of $\tau_0/2$) and overlapped impulse responses modulated by their respective symbol values $a_{k,n}$, which are also frequency shifted in each k-th subcarrier, and then weighted by the channel value $H(k)$. At the demodulator stage, we get an estimation of the transmitted real symbols using the inner product of the received signal and $g_{k,n}(t)$ (which is equivalent to the matched filter of the prototype function):

$$\begin{aligned}\hat{a}_{k,n}&=\Re\left\langle g_{k,n},y\left(t\right)\right\rangle=\Re\left\{\int_{-\infty}^{+\infty}g_{k,n}^{*}\left(t\right)y\left(t\right)\partial t\right\}=\\&=\Re\left\{\int_{-\infty}^{+\infty}g\left(t-n\tau_0\right)e^{-j\pi kt/\tau_0}e^{-j\phi_{k,n}}y\left(t\right)\partial t\right\}\end{aligned} \tag{2.68}$$

Then, it can be assumed a channel with free distortion $\hat{a}_{k,n}=a_{k,n}$ if and only if the set of functions $g_{k,n}$ has an orthonormal basis of its span [72].

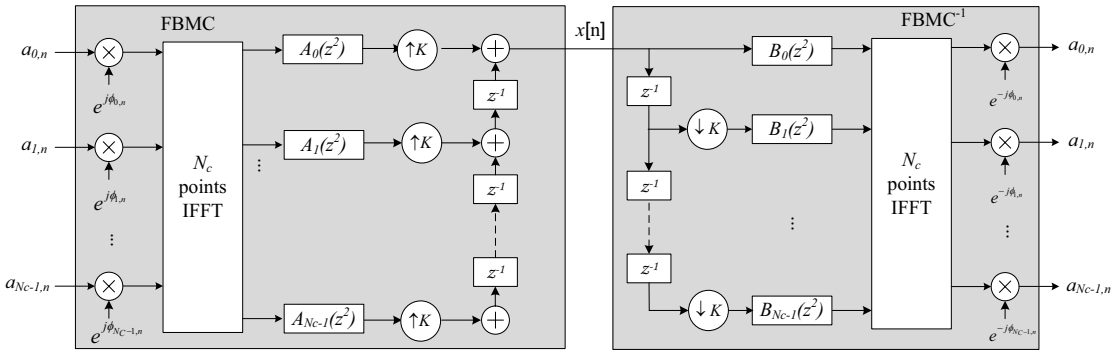


Figure 2.18. OFDM/OQAM modulation and demodulation realized with an IFFT.

Regarding the implementation of the SFB and the AFB, we have already mentioned that one efficient approach is to implement them using a polyphase network where the signal from each subcarrier is filtered by a frequency-shifted prototype function. Actually, an efficient IFFT-based implementation is represented in Figure 2.18 [72].

It is very important to remark that since consecutive symbols not produce intersymbol interference when FBMC is used, a guard interval between the OFDM/OQAM symbols is no longer required. Therefore, if the loss due to the intercarrier interference generated by the FBMC is lower than the gain from the lack of the guard interval, it is expected that the OFDM/OQAM overcomes the performance of the classic CP-OFDM transmission scheme.

2.2.5. Single Carrier – Frequency Division Multiple Access (SC-FDMA)

One of the main drawbacks of the OFDM scheme (and thus from the OFDMA or FBMC too) is its high signal envelope fluctuations that are translated into high PAPR values. Therefore, high linearity is required at the power amplifiers. To achieve this linearity, the amplifiers have to operate with a large back-off from their peak power in order to avoid distortion. The result is a low power efficiency which burdens the battery of the mobile terminals. In consequence, the 3GPP-LTE has adopted the Single Carrier Frequency Division Multiple Access (SC-FDMA) in the uplink [75]. Hence, both the OFDMA and the SC-FDMA will substitute in the downlink and uplink respectively the current Wideband CDMA based air interface (UMTS standard [76]).

The SC-FDMA system uses different orthogonal frequencies to transmit the information symbols. However, instead of transmitting all these symbols in parallel as in OFDM, these are transmitted sequentially leading to lower envelope fluctuations. Nevertheless, since these symbols are transmitted sequentially, the symbol time is shortened compared to the OFDM symbol duration, which could lead to large intersymbol interference (the base station then needs to employ adaptive frequency equalization to overcome the ISI). As a result, using the SC-FDMA it is possible to save the battery of the terminal at the expense of more computational complexity and signal processing at the base station compared to the OFDM scheme.

The schematic of an SC-FDMA based transmitter and receiver is shown at Figure 2.19. We can appreciate that the SC-FDMA scheme is very similar to the OFDM scheme, with the difference of an additional FFT and IFFT blocks at the input and output of the data stream. Following the schematic, we observe that a block of L complex symbols is transformed to its equivalent in the frequency domain. Later, the frequency samples of the input complex symbols are arranged according to the specific subcarrier interleaver and later mapped to the N_c subcarriers of the OFDM scheme (N_c>L). At this point a regular OFDM transmitter and receiver are used where the same concepts developed in Section 2.2.2 can be also applied, for this reason some times the SC-FDMA is also referred as *DFT-spread* OFDMA. Afterwards, the complex data values in frequency recovered at the output of the Inverse OFDM block are passed again to its time domain representation recovering the initial data symbols. The discrete version of the SC-FDMA signal with L occupied subcarriers is given by

$$\begin{aligned} x[n] &= \frac{1}{\sqrt{N_C}} \sum_{k=0}^{N_C-1} DFT\{\mathbf{s}\} \cdot e^{j2\pi nk/N_C} \\ &= \frac{1}{\sqrt{N_C}} \sum_{k=0}^{N_C-1} \left(e^{j2\pi nk/N_C} \sum_{n'=0}^{L-1} e^{-j2\pi n'(k+\Delta_k)/L} S_{n'} \right), \quad 0 \le n \le N_C - 1 \end{aligned} \tag{2.69}$$

where $\mathbf{s} \in \mathbb{C}^{L\times 1}$ is the vector of transmitted symbols. Note that the result from the operation depicted in Eq. (2.69) is the same vector $\mathbf{s}$ upsampled by a N_C/L factor and modulated (i.e. shifted in frequency) to the Δ_k $L/2$ frequency.

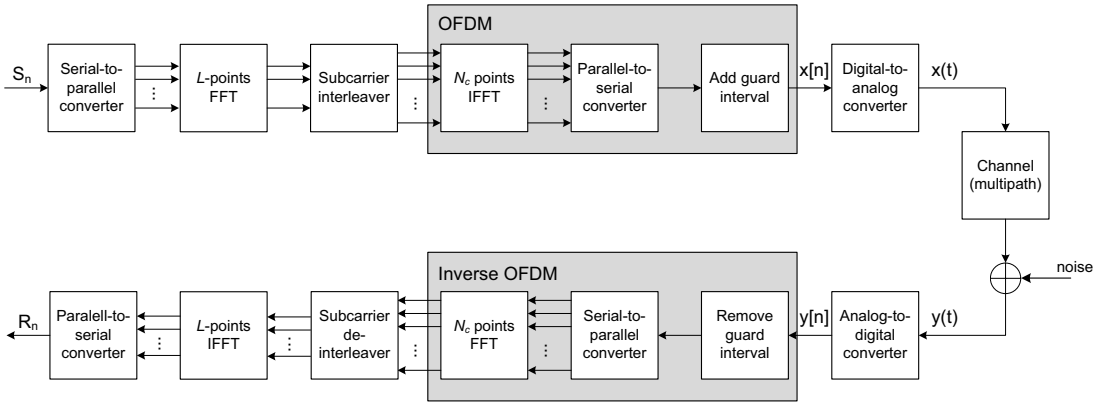


Figure 2.19. Transmitter and Receiver architectures of SC-FDMA systems.

Similarly to the behaviour of the subcarrier permutation scheme in the OFDMA scheme, the subcarrier interleaver may follow an *Interleaved* or a *Localized* subcarrier mapping scheme.

In the *Interleaved* subcarrier mapping, the output of the DFT is allocated across the entire bandwidth by allocating zeros on the unused subcarriers resulting in a non-continuous comb-shaped spectrum. This subcarrier structure leads to the Interleaved Frequency Division Multiple Access (IFDMA) scheme [78][79]. Due to the spanning over all the bandwidth, IFDMA provides the higher frequency diversity as well as the lower PAPR among the mentioned DFT precoded OFDMA schemes [78]. However, IFDMA is the more sensitive against carrier frequency offsets [79]. Further on, there exists a very efficient implementation for signal generation in time domain for IFDMA where each data symbol is phase rotated according to the symbol index and the subcarrier mapping, whose computational complexity is much lower than for a regular OFDMA scheme [80].

On the other hand, the *Localized* subcarrier mapping matches the frequency samples into contiguous subcarriers, thus the transmitted signal results in a contiguous spectrum that occupies a fraction of the total available bandwidth (similar to Localized OFDMA represented in Figure 2.13. This structure is also known as Localized Frequency Division Multiple Access (LFDMA). Due to the blockwise structure, LFDMA provides good robustness to carrier frequency offsets as well as high multi user diversity gains in case of adaptive resource allocation [81]. On the other hand, LFDMA shows low frequency diversity because the subcarriers allocated to a specific user are restricted to a localized fraction of the available bandwidth. However, whether frequency scheduling is applied, the throughput might be increased up to 80% compared to static LFDMA, whereas for the IFDMA the gain obtained by channel dependant scheduling is only 26% [75][82].

Despite the above mentioned advantages, we may point out the discussions that have been carried out recently within the IEEE 802.16m Work Group, where it has been opted to keep on with the OFDMA scheme in both downlink and uplink, since the necessity of the SC-FDMA it is still not justified. Furthermore, some detailed analyses have concluded that when we compare a MIMO-OFDMA with a MIMO-SC-FDMA OFDMA scheme, a gain of up to 1-6dB may be achieved with the use of the first one [83][84]. Moreover, among the mentioned advantages of the OFDMA versus the SC-FDMA in the uplink mode, an emphasis has been addressed on the flexibility of the pilot symbols allocation procedure within the OFDMA frame, the possibility to

exploit reciprocity in TDD links, and the large gains obtained if channel-aware scheduling is applied [85]. Nevertheless, if MIMO techniques are implemented in the uplink, since the channel experienced by the symbols transmitted under the SC-FDMA scheme may not be flat, the complexity of the receiver increases and a large time domain equalization will be required that may limit its implementation for high numbers of transmit and receive antennas.

2.2.6. Multi-Carrier Spread Spectrum (MC-SS)

Spread spectrum techniques have reached a great success with the election of CDMA as the medium access technique for the 3G systems. On the other hand, the application of OFDM in digital broadcasting and wireless LANs since the mid 90s has demonstrated the potential of OFDM systems in to deliver broadband communications.

The performance and the advantages of the CDMA and the OFDM systems in different environments motivated the scientists to consider the combination of both schemes as a possible candidate for 4G mobile communication systems. This combination was first analyzed in 1993 and named as *Multi-carrier Spread Spectrum* (MC-SS) techniques [86]. According to how users and data are multiplexed in the spectrum, a variety of combinations have appeared. The firsts in appearing were the *Multi-Carrier Code Division Multiple Access* (MC-CDMA), the *Multi-Carrier Direct-Sequence Code Division Multiple Access* (MC-DS-CDMA) and the Multitone Code Division Multiple Access (MT-CDMA) [6][87]-[89]. These systems are characterized by the way on how the spreading process is implemented, i.e. in MC-CDMA scheme the symbols are spread in frequency domain, whereas in MC-DS-CDMA scheme they are spread in time domain. Similarly to the MC-DS-CDMA system, in the MT-CDMA scheme, the symbols are also spread in time but here the spread symbols span over more than the sub-carrier assigned bandwidth, thus increasing the spectral efficiency at the expense of increasing the interference between adjacent sub-carriers.

The application of each scheme was the focus of many researchers during the last decade. It has been shown that MC-CDMA scheme is more suitable for the downlink where the different symbols are synchronized. In the uplink, where the symbols are not synchronized (quasi-synchronous systems), the MC-DS-CDMA scheme and a modified version of the MC-CDMA named Spread Spectrum Multicarrier Multiple Access (SS-MC-MA) [6][88] are preferred, since they make easier the MAI avoidance and thanks of the aid of non-linear receivers (Rake, Serial/Parallel interference cancellation, etc.), channel coding and interleaving in the frequency domain, the performance increased significantly [88]. Hereafter the above mentioned MC-SS combinations are described more in depth.

2.2.6.1. Multicarrier Code Division Multiple Access (MC-CDMA) scheme

The MC-CDMA scheme, also known as OFDM-CDMA scheme, can be considered a classical OFDM system where the information applied to each L subcarriers belongs to the same spread symbol, hence L is the *spreading factor* (L=SF) [88]. The election of the SF determines how the information is spread, and therefore the frequency diversity. Moreover the SF determines the number of users that will share the same bandwidth in case of *fully-loaded* (all the codes are used). When SF becomes smaller than N_C, different groups of sub-carriers can be established (N_C /L) which form the whole bandwidth, thus a FDM component can be recognized between

groups. In this last case, different algorithms can be used to assign users and data to sub-bands according to different parameters as the channel, the required data rate, the QoS, etc [6]. Furthermore, the sub-carriers belonging to the same group are not necessarily consecutive. An schematic example describing the principles of the MC-CDMA scheme is depicted in Figure 2.20.

If the channel in each sub-carrier is flat (or nearly constant) the equalization process is carried as for the OFDM scheme with a simple one-tap equalizer per subcarrier. The main disadvantages of the MC-CDMA scheme are the tight synchronization requirements and the high PAPR (as in OFDM), making MC-CDMA more suitable for the downlink. Nevertheless, in the downlink case, the BS can apply different grouping strategies or adaptive modulation and coding, pre-equalization, power allocation, etc., hence the spectral efficiency can be enhanced. In the uplink case, these strategies can also be applied as long as the mobile terminals are perfectly synchronized, otherwise the SS-MC-MA scheme where each user is allocated a separated su-band known has been proposed [88]. Many proposals and European research projects have encouraged researches on MC-CDMA scheme as a potential access transmission scheme (i.e. DVB-T return channel, IST MATRICE[7] and 4More[8] projects, fixed broadband wireless access, aeronautical communications, etc.) [88][90][91]. Nevertheless, another important benefit of using MC-CDMA is the feasibility of using frequency reuse factor of 1, however, as in any CDMA scheme where several users share the same bandwidth, different power, and admission control techniques will be necessary in order to bound the amount of multiple access interference [92][93].

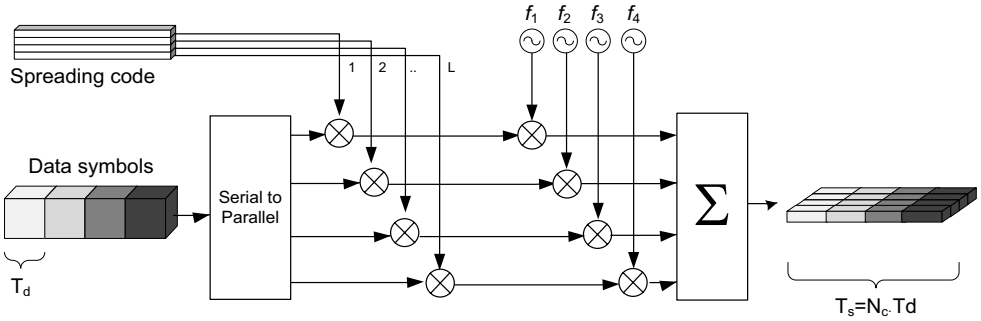


Figure 2.20. MC-CDMA principles for a single user scenario and N_c (N_c=4) subcarriers

The discrete version of a MC-CDMA signal with spreading factor L is given by

$$x[n]=\frac{1}{\sqrt{N_C}}\sum_{k=0}^{N_C-1}e^{j2\pi nk/N_C}\sum_{l=0}^{L-1}c_k^{(l)}S_l,\quad 0\le n\le N_C-1 \tag{2.70}$$

where $c_k^{(l)}$ is the k-th chip from the $c^{(l)}$ spreading sequence. Note that if $L<N_C$ not all the subcarriers are occupied unless different groups of spread symbols are transmitted simultaneously in frequency (up to N_C/L groups). The signal structure of the MC-CDMA scheme is represented in Figure 2.21 where it is shown that the symbols are spread in the frequency domain. Each spread data symbol is represented by a different colour. The number of symbols

[7] Multicarrier CDMA TRansmission Techniques for Integrated Broadband CEllular Systems (***MATRICE***), European research project IST-2001-32620.
[8] 4G MC-CDMA Multiple Antenna System On Chip for Radio Enhancements (**4MORE**), European research project IST-2002-507039.

sharing the same bandwidth (thus interfering between them) is limited by the spreading factor. However it must be carefully chosen since it depends on other factors as the channel conditions, the receiver detector, power control algorithms, spreading codes auto- and cross-correlation properties, etc.

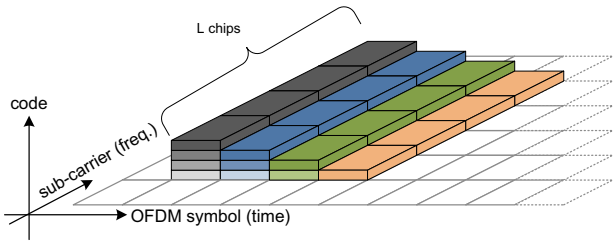


Figure 2.21. MC-CDMA time/frequency signal structure.

2.2.6.2. Multicarrier Direct Sequence Code Division Multiple Access (MC-DS-CDMA) scheme

In the MC-DS-CDMA scheme the high-data rate stream is also converted to low-rate streams, where each stream is applied to one OFDM sub-carrier [94][95]. Each low-rate symbol is then spread over the assigned sub-channel by multiplying the symbol by a specific user spreading codeword. Then different parallel DS-CDMA transmissions (see Figure 2.22), each of them with a bandwidth equal to N_C/T_d, are placed at orthogonal carrier frequencies f_n. The same spreading codes can be utilized in the different sub-carriers.

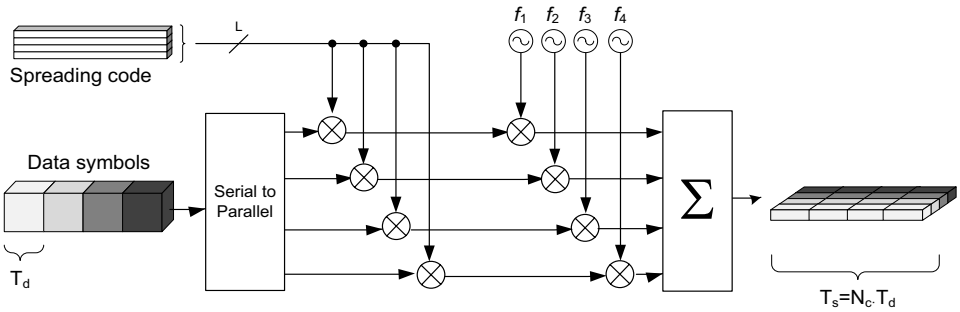


Figure 2.22. MC-DS-CDMA principles for a single user scenario having N_c=4 subcarriers

Furthermore, this scheme can be used to efficiently generate parallel DS-CDMA systems, where the channel on each sub-carrier is not necessarily flat (broadband sub-channel). In this case, regular broadband channel equalizers or rake receivers are required in order to obtain the maximum diversity. However, whether the number of sub-carriers is increased and the total bandwidth remains constant, the symbols are then transmitted over narrow-band sub-channels and simple one tap equalizer are required as in the OFDM scheme. These schemes can only achieve time diversity gain due to the spreading domain.

One of the advantages of the MC-DS-CDMA scheme (compared to the OFDM or the MC-CDMA) is that the experienced PAPR is low, what is makes MC-DS-CDMA more suitable for the uplink communications. Nevertheless, since the spreading process is applied in time direction this scheme is also well-fitted to large time-variant channels (due to high mobility environments for instance) where time diversity gain improves the system performance.

The signal structure obtained with the MC-DS-CDMA scheme is represented in Figure 2.23. Each symbol spread in time is represented with the same colour and is mapped to one subcarrier. In order to increase the spectral efficiency different spreading codes might be used thus several symbols are transmitted in each subcarrier (thus multiple access interference will be present). Therefore, the discrete version of a MC-DS-CDMA signal with spreading factor L is given by

$$x[n] = \frac{1}{\sqrt{N_C}} \sum_{k=0}^{N_C-1} e^{j2\pi nk/N_C} \sum_{l=0}^{L-1} c_n^{(l)} S_l, \quad 0 \le n \le N_C - 1 \tag{2.71}$$

where $c_n^{(l)}$ is the n-th chip from the $c^{(l)}$ spreading sequence. Nevertheless, since the MC-DS-CDMA concentrates all the data within one subcarrier, apart from having low PAPR, this is also useful in order to improve the link budget since more power is transmitted in a shorter bandwidth, increasing then the signal to noise ratio.

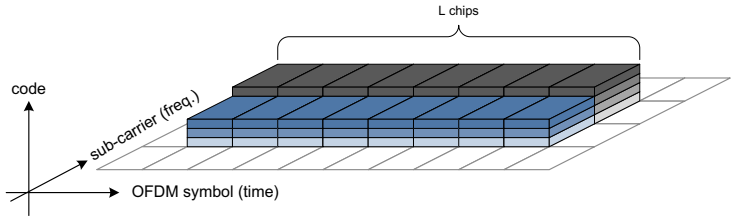


Figure 2.23. MC-DS-CDMA time/frequency signal structure.

A variant of the MC-DS-CDMA scheme has been proposed in [89] by L. Vandendorpe. The scheme is referred as the Multitone Code Division Multiple Access (MT-CDMA) scheme which spread the symbols in time domain as in the MC-DS-CDMA scheme [88]. The MT-CDMA is generated by modulating first a block of N_c symbols and mapping them onto N_c subcarriers. The OFDM generated signal is then multiplied in the time domain by a spreading sequence which is function of the user to which the signal is addressed. The chip duration of the spreading sequence is $T_c \le T_d/N_c$. If we look at the spectrum of this signal, each data symbol is transmitted over a bandwidth close to $1/T_c = N_c/T_d$ while the different subcarriers are distant one from each other 1/ T_d. This means that each symbol benefits from frequency diversity as in the MC-CDMA scheme despite there is a strong overlap of the different sub-carriers. However, due to the orthogonality properties of the spreading sequences, the receiver is able to recover all the data symbols.

2.2.6.3. Multicarrier Spread Spectrum schemes with 2D spreading

Recently, two dimensional (2D) spreading is becoming of great interest since it combines the goodness from MC-CDMA and MC-DS-CDMA, and according to the environment it may adjust the spreading factor en each domain in order to ensure that all chips with spread data are faded independently. These schemes are referred as *Variable Spreading Factor Orthogonal Frequency and Code Division Multiplexing* (VSF-OFCDM) and V*ariable Spreading and Chip Repetition Factors* (VSCRF), both proposed for the downlink and the uplink respectively [98][99][100]. Moreover, the Time Division Code Division Multiplexing OFDM (TD-CDM-OFDM)

scheme departs from the existing system TD-SCDMA (the third generation system deployed in China) and combines 2D spread spectrum with OFDM.

From previous mentioned schemes, it is obvious that spreading in time and/or frequency leads to an increase in diversity which mitigates (even minimizes) the effects of the multipath channel. Furthermore, the spreading length can be adjusted in time and/or frequency to the channel conditions and also to the user requirements (throughput, BER, delay, etc.). In a downlink synchronous transmission case, the best option is to use orthogonal spreading codes (i.e. Walsh-Hadamard codes, Fourier codes, Golay complementary codes, etc.), whereas in the uplink since synchronization between sequences is quite complex, simple PN sequences with good cross-correlation properties (i.e. Golden code, Zadoff-Chu Codes, low-rate convolutional codes, etc.) are used. Focusing on the downlink case, the Walsh-Hadamard codes are generalized since they are very simple and computational efficient by using the Hadamard transform. These codes can be recursively generated by using the following Hadamard matrix generation,

$$C_L = \begin{bmatrix} C_{L/2} & C_{L/2} \\ C_{L/2} & -C_{L/2} \end{bmatrix}, \quad \forall L = 2^m, \quad m \geq 1, \quad C_1 = 1. \tag{2.72}$$

The different codes generated for each spreading factor as well as the relation between codes are depicted in Figure 2.24. We can observe that the codes are generated recursively and the cross-correlation between them also depends on the number of branches in common (the highest number of uncommon branches, the better cross-correlation properties). On the other hand, by using variable spreading codes one is able to achieve a good trade-off between coverage, single-cell/multi-cell environments, and mobility [103]. For high coverage areas with high delay spread, large spreading factors can be applied and for low coverage areas with low delay spread, the smallest spreading factor can be used. These principles, extracted from the UMTS system, are also applicable in VSF-OFCDM systems, where each symbol will be spread in time and in frequency with the respective spreading codes. Obviously, assigning spreading codes becomes more complicate since the orthogonality must be maintained in both domains and at the same time being able to assign different spreading lengths. One important property from the Walsh-Hadamard codes is that different spreading factors can be used while orthogonality is preserved. However, whether a code with spreading factor L_i is used, all its *children branches* (codes with $L_j > L_i$ that are generated from such code) cannot be not used. In Figure 2.25 the signal structure for the VSF-OFCDM scheme is shown and the spreading codes pairs that could be assigned for this example ($\{\mathbf{c}_{4,2};\mathbf{c}_{2,0}\}$, $\{\mathbf{c}_{2,0};\mathbf{c}_{4,2}\}$, $\{\mathbf{c}_{4,3};\mathbf{c}_{4,3}\}$). A more detailed formulation of the transmitted and received signals for the OFCDM transmission scheme is can be found in Section 4.2.4. Note also that the time spreading code assignment is independent from the frequency spreading code assignment.

Nevertheless, since VSF-OFCDM is a more general case of the OFDM, the MC-CDMA or the MC-DS-CDMA, a transceiver designed under the VSF-OFCDM scheme may cope with many different systems, enabling a low cost *generic radio interface* operational in different cell types and in different environments. This was first proposed in 2001 by the Japanese mobile operator NTT DoCoMo, where a bandwidth of approximately 100MHz was used. Throughputs close to 100Mbps were achieved for Signal to Noise Ratio of 13dB and without antenna

diversity [100][102]. In the year 2005 and under the same scheme, the throughput was raised up to 2.5Gbps under certain conditions, and it has been shown to work up to 1Gbps in indoor and 100Mbps in outdoor environments [104]. The maximum spectral efficiency achieved with this system (using a 64-QAM, 6 transmitter and receiver antennas and a MS velocity of 20km/h) is 25bps/Hz which may fit pretty well to what the fourth generation wireless networks will require.

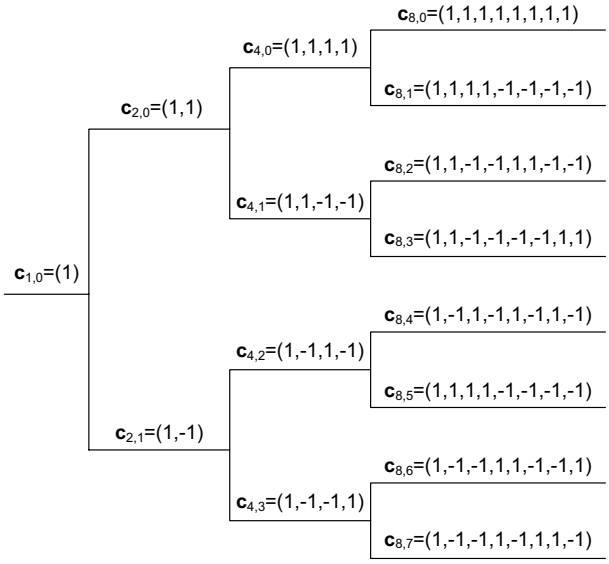


Figure 2.24. Variable length orthogonal spreading code generation.

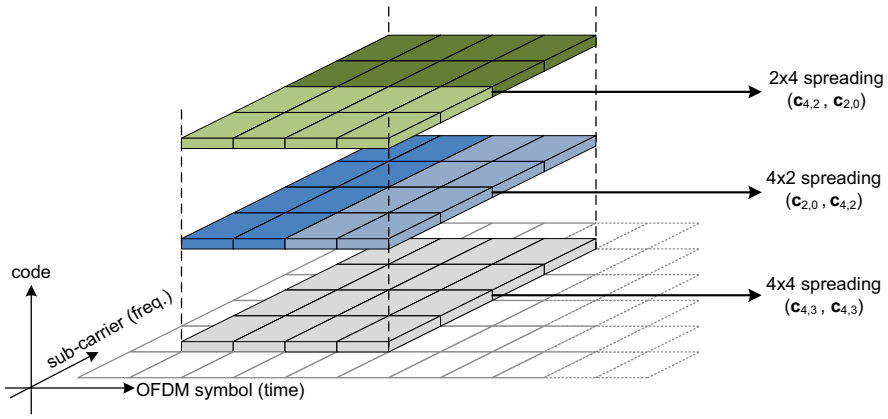


Figure 2.25. VSF-OFCDM time/frequency signal structure.

2.3. Summary of BWA air interface technologies

The forthcoming generation of BWA systems is expected to be a heterogeneous list of air interfaces in which any of them must achieve spectral efficiencies up to 10bits/s/Hz. In this second chapter the different proposed transmission schemes has been reviewed, making a special emphasize on the OFDM scheme which is in a certain way the basis of the other mentioned multicarrier system communications (e.g. OFDMA, SC-FDMA, FBMC, OFCDM, etc.). In deed, most of the research works presented in the remainder chapters are focused on the

use of the OFDMA transmission scheme, but they can be easily extended to any of the other transmission schemes without implying a big impact on the algorithms neither their performance. Nevertheless, it is also expected that multiple antennas are used at both transmitter and receiver in order to increase the channel capacity as it will be further explained in Chapter 6. The different channel models and properties in case of single or multiple antennas have been also explained in this chapter; hence the effects of the channel in the transmitted signal as well as the diversity order can be later perfectly understood.

References

[1] T.S. Rappaport, Wireless Communications: Principles and Practice, Second edition, Prentice Hall, 2002.

[2] M. Hata, “Empirical Formula for propagation loss in land mobile radio services”, *IEEE Trans. Vehicular Technology*, vol.VT-29, Aug. 1980.

[3] M. Gudmundson, “Correlation Model for Shadow Fading in Mobile Radio Systems”, Electron. Letters, Vol.27, No.23, Nov. 1991.

[4] Reference scenario specification: final description, Deliverable 1.3, Multicarrier CDMA TRansmission Techniques for Integrated Broadband CEllular Systems, 2001-4.5.2 Terrestrial wireless systems and networks, IST-2001-4.5.2 Terrestrial wireless systems and networks.

[5] P.A. Bello, “Characterization of randomly time-variant linear channels”, *Transactions on Communications Systems*, vol. 11, pp. 360-393, Dec. 1963.

[6] S. Kaiser, “Multi-Carrier CDMA Mobile Radio Systems - Analysis and Optimization of Detection, Decoding, and Channel Estimation”, *Ph.D. Thesis,* Munich, Germany, 1998.

[7] B. Sklar, “Rayleigh fading channels in mobile digital communication systems Part I: Characterization", *IEEE Communications Magazine*, vol. 35, pp. 90-100, July 1997.

[8] R.H. Clarke, "A Statistical Theory of Mobile-Radio Reception," *Bell Systems Technical Journal,* vol. 47, pp. 957-1000, 1968.

[9] J.W.C. Jakes, *Microwaves Mobile Communications*, New York: John Wiley & Sons, 1974.

[10] J.D. Parsons, A.S. Bajwa, “Wideband characterization of fading mobile radio channels”, IEE Proceedings, Vol. 129, Part F, No.2, April 1982, pp-95-101.

[11] M. Nakagami, “The m-distribution - A general formula of intensity distribution of rapid fading”, *in Statistical Methods of Radio Wave Propagation*, W.C. Hoffman (ed.), pp. 3-36. New York: Pergamon Press, 1960.

[12] *CODIT, Final Propagation Model*. Report R2020/TDE/PS/DS/P/040/b1, 1994.

[13] "Digital Land Mobile Radio Communications - COST 207", Commission of the European Communities, Final Report, 14 March, 1984--13 September, 1988, Office for Official Publications of the European Communities, Luxembourg, 1989.

[14] T. Zwick, “Comparison of ray tracing results with indoor measurements”, COST259, TD(00)031, Valencia, Spain, Jan., 2000.

[15] “Final Report on Link Level and System Level Channel Models”, Deliverable 5.4. v1.4, Wireless World Initiative New Radio, IST-2003-507581 WINNER.

[16] “IEEE 802.16m Evaluation Methodology Document (EMD)”, IEEE 802.16m-08/004r4, IEEE 802.16 Broadband Wireless Access Working Group, Nov. 2008.

[17] I. Siaud, “Indoor Propagation Channel Simulation for the Performance Evaluation of HIPERLAN system”, *ETSI EP BRAN document*, Nº WG3TD46, Oct. 1997.

[18] J. Medbo, H. Andersson, P. Schramm, H. Asplund, H., “Channel Models for HYPERLAN/2 in different Indoor Scenarios”, COST 259 TD98, Bradford, UK, April, 1998.

[19] M.C. Jeruchim, B. Balaban, K. Sam Shanmugan, “*Simulation of Communication Systems”, 2nd* Edition, New York: Kluwer Academic/Plenum, 2000.

[20] C. Oestges, B. Clerck, “MIMO Wireless Communications”, Academic Press, 2007.

[21] P. Almers et al. “Survey of Channel and Radio Propagation Models for Wireless MIMO Systems”, Eurasip Journal on Wireless Communications and Networking, 2007.

[22] C.N. Chuah, J.M. Kahn, D. Tse, “Capacity of multiantenna array systems in indoor wireless environment”, Proc. Globecom 1998, vol.4, p.1894-1899, Sydney, Australia, 1998.

[23] ITU-R Recommendation M.1645, “Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000,” 2003.

[24] W. Mohr, R. Luder, K.H. Mohrmann, "Data rate estimates, range calculations and spectrum demand for new elements of systems beyond IMT-2000," Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on , vol.1, no., pp. 37-46 vol.1, 27-30 Oct. 2002.

[25] R.C. Dixon, *Spread Spectrum Systems with Commercial Applications,* New York: John Wiley & Sons, 1994.

[26] A.J. Viterbi, *CDMA Principles of Spread Spectrum Communication,* Reading, Addison-Wesley, 1995.

[27] TIA/EIA/IS-2000.2-A, Physical Layer Standard for cdma2000 Spread Spectrum Systems, Telecommunications Industry Association, Arlington, VA, March 2000.

[28] A. Baier, U.C. Fiebig, W. Granzow, W. Koch, P. Teder, J. Thielecke, “Design study for a CDMA-based third-generation mobile radio system”, *IEEE Journal on Selected Areas in Communications*, vol.12, pp.733-743, May 1994.

[29] C. Smith, D. Collins, *3G Wireless Networks*, New York, McGraw-Hill Professional, 2001.

[30] P.W. Baier, P. Jung, A. Klein, “Taking the challenge of multiple acces for third generation cellular mobile radio systems – a European view”, *IEEE Communications Magazine*, vol.34, pp.82-89, Feb. 1996.

[31] M.P.J. Baker, T.J. Mouslsley, "Power control in UMTS Release '99", *3G Mobile Communication Technologies, 2000. First International Conference on* (Conf. Publ. No. 471) , pp.36-40, 2000.

[32] B. Sklar, *Digital Communications: Fundamentals and Applications*, Prentice Hall, 1988.

[33] J.G. Proakis, *Digital Communications*, New York, McGraw-Hill, 1995.

[34] L.L. Yang, L. Hanzo, "Slow frequency-hopping multicarrier DS-CDMA for transmission over Nakagami multipath fading channels," *Selected Areas in Communications, IEEE Journal on* , vol.19, no.7, pp.1211-1221, Jul 2001.

[35] D. L. Noneaker, “Optimal combining for Rake reception in mobile cellular CDMA forward links,” in *Proc. IEEE Military Commun. Conf.*, Bedford, MA, 1998.

[36] G.E. Bottomley, T. Ottosson, Y.P.E. Wang, "A generalized RAKE receiver for interference suppression", *Selected Areas in Communications, IEEE Journal on* , vol.18, no.8, pp.1536-1545, Aug 2000

[37] K. Fazel, S. Kaiser, P. Robertson, “OFDM: a key component for terrestrial broadcasting and cellular mobile radio”, *Proceedings International Conference on Telecommunication* (ICT’96), Istanbul, Turkey, pp.576-583. Apr.1996.

[38] S.B. Weinstein, P.M. Ebert, “Data Transmission by Frequency-division multiplexing using the discrete Fourier transform”, *IEEE Transactions on Communication Technology*, vol.19, pp.628-634, Oct.1971.

[39] R.W. Chang, “Synthesis of band-limited orthogonal signals for multi-channel data transmission”, *Bell System Technical Journal* v.46, 1996, pp.1775-1796.

[40] C. E. Shannon, “A mathematical theory of communication”, *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

[41] H. Sari, G. Karam, I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *Communications Magazine, IEEE* , vol.33, no.2, pp.100-109, Feb 1995.

[42] W.Y. Zou, W. Yiyan, "COFDM: an overview", *Broadcasting*, *IEEE Transactions on*, vol.41, no.1, pp.1-8, Mar 1995.

[43] F. Tosato, P. Bisaglia, "Simplified soft-output demapper for binary interleaved COFDM with application to HIPERLAN/2," *Communications, 2002. ICC 2002. IEEE International Conference on*, vol.2, pp. 664-668 vol.2, 2002.

[44] IEEE Std 802.11™-2007, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, New York, USA, 12 June 2007.

[45] IEEE 802.16TM-2004/Cor1-2005 (Amendment and Corrigendum to IEEE 802.16-2004), Part 16: Air Interface for Fixed Broadband Wireless Access Systems; Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1, IEEE, New York, February 2006.

[46] ETSI EN 300 401 V1.3.3. European Standard for Radio Broadcasting Systems. Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers. May 2001.

[47] ETSI EN 300 744 V1.4.1. European Standard for Digital Video Broadcasting (DVB). Framing structure, channel coding and modulation for digital terrestrial television. January 2001.

[48] "Asymmetric digital subscriber line (ADSL) transceivers", ITU-T Recommendation G.992.1, June 1999.

[49] S. Hara, R. Prasad, "Multicarrier Techniques for 4G Mobile Communications", Artech House, USA, 2003.

[50] I. Gutiérrez, J. Pijoan, F. Bader, R. Aquilué, New Channel Interpolation Method for OFDM Systems by Nearest Pilot Padding, in Proc. European Wireless 2006 (EW2006), Athens, Greece, April, 2006.

[51] H. Rohling, R. Grunheid, "Performance of an OFDM-TDMA mobile communication system", *IEEE 46th Vehicular Technology Conference* (VTC-96 Spring), vol.3, pp.1589-1593 vol.3, 28 Apr-1 May 1996.

[52] J. Zyren, W. McCoy, "Overview of the 3GPP Long Term Evolution Physical Layer", White Paper, Freescale Semiconductors, July, 2007.

[53] M. Wang, "Ultra Mobile Broadband Technology Overview", 6th Annual *Communication Networks and Services Research Conference*, 2008 (CNSR-2008), pp.8-9, 5-8 May 2008.

[54] W. Bolton, X. Yang, M. Guizani, "IEEE 802.20: mobile broadband wireless access," *Wireless Communications, IEEE*, vol.14, no.1, pp.84-95, Feb. 2007.

[55] R. Laroia, "Fast-hopped OFDM Technology offers a new opportunity for mobile broadband – technology information", Computer Technology Review, May, 2001.

[56] Flarion technologies, "FLASH-OFDM solutions", www.flarion.com , www.qualcomm.com.

[57] C. Cordeiro, K. Challapali, D. Birru, N. Sai Shankar, "IEEE 802.22: the first worldwide wireless standard based on cognitive radios," *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005* (DySPAN-2005), pp.328-337, 8-11 Nov. 2005.

[58] L. Seung Joon, "Link Adaptation Considering Mobility in OFDMA Systems with Multiple Transmit Antennas," Communications Letters, IEEE , vol.12, no.7, pp.493-495, July 2008.

[59] LG Electronics, Downlink Resource Allocation in EUTRA, R1-060052, 3GPP TSG RAN WG1 AH on LTE. Helsinki, Finland, Jan. 2006.

[60] IEEE 802.16 Broadband Wireless Access Working Group, "IEEE 802.16m System Description Document", IEEE 802.16m-08/003r7, Feb. 2009.

[61] Motorola, "Downlink Localized and Distributed Multiplexing", R1-060396, 3GPP TSG RAN1 Meeting #44, Denver, USA, Feb. 2006.

[62] Samsung, “Text proposal on rules for Mapping VRBs to PRBs”, R1-060345, 3GPP RAN WG1 Meeting #44, Denver, USA, Feb. 2006.

[63] Motorola, “Subcarrier Mapping for Distributed Allocation for EUTRA DL”, R1-060245, RAN 1 LTE Adhoc, Helsinki, Finland, Jan. 2006.

[64] Huawei, “EUTRA Downlink Multiplexing of Localized and Distributed Channels”, R1-051192, 3GPP TSG RAN WG1, San Diego, USA, Oct. 2005.

[65] ETSI TS 136 300 V8.4.0, “Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRAN); Overall description”, April 2008.

[66] M.I. Rahman, R.V. Reynisson, D.V.R. Figueiredo, R. Prasad, "Coordinated Sub-Carrier and Band Hopping in OFDMA based Systems," *4th International Symposium on Wireless Communication Systems, 2007 (ISWCS-2007)*, pp.96-101, 17-19 Oct. 2007.

[67] B. Le Floch, M. Alard, C. Berrou, “Coded Orthogonal Frequency Division Multiplex”, *Proceedings of the IEEE,* vol. 83, no. *6,* June 1995.

[68] P. Siohan, C. Roche, “Cosine modulated filterbanks based on extended Gaussian functions”, *IEEE Trans. On Signal Processing,* vol. 48, no 11, pp. 3052-3061, November 2000.

[69] D. Lacroix, N. Goudard, M. Alard, “OFDM with Guard Interval Versus OFDM/OffsetQAM for High Data Rate UMTS Downlink Transmission,” *Vehicular Technology Conference (VTC) 2001 Fall*, 2001.

[70] P.P. Vaidyanathan, Multirate Systems and filterbanks, Prentice-Hall, 1993.

[71] M. Bellanger, G. Bonnerot, and M. Coudreuse, “Digital filtering by polyphase network: Application to sample-rate alteration and filter banks,” IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pp. 109-114, Apr. 1976.

[72] M.G. Bellanger, "Specification and design of a prototype filter for filter bank based multicarrier transmission," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001, (ICASSP-01), vol.4, pp.2417-2420 vol.4, 2001.

[73] K.W. Martin, "Small side-lobe filter design for multitone data-communication applications," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on* , vol.45, no.8, pp.1155-1161, Aug 1998.

[74] P. Siohan, C. Siclet, N. Lacaille, "Analysis and design of OFDM/OQAM systems based on filterbank theory", *IEEE Transactions on Signal Processing,* vol.50, no.5, pp.1170-1183, May 2002.

[75] H. G. Myung, J. Lim, D. J. Goodman, “Single Carrier FDMA for Uplink Wireless Transmission”, *IEEE Vehicular Technology Magazine*, vol. 1, no. 3, Sep. 2006, pp. 30-38.

[76] 3GPP TS 25.211 V3.12.0 (2002-09), Physical channels and mapping of transport channels onto physical channels (Release 1999).

[77] 3rd Generation Partnership Project (3GPP); Technical Specification Group Radio Access Network; Physical Layer Aspects for Evolved UTRA.

[78] T. Frank, A. Klein, E. Costa, A. Kuehne, “Low Complexity and Power Efficient Space-Time-Frequency Coding for OFDMA”, In *Proc. of 15th Mobile & Wireless Communications Summit*, Mykonos, Greece, June 2006.

[79] T. Frank, A. Klein, E. Costa, E. Schulz, “Robustness of IFDMA as Air Interface Candidate for Future Mobile Radio Systems”, In Advances in Radio Science, Miltenberg, Germany, Oct. 2004.

[80] T. Frank, A. Klein, E. Costa, E. Schulz, “IFDMA – A Promising Multiple Access Scheme for Future Mobile Radio Systems”, In Proc. PIMRC 2005, Berlin, Germany, Sep. 2005.

[81] A. Sohl, A. Klein, “Comparison of Localized, interleaved and Block-interleaved FDMA in terms of pilot multiplexing and channel estimation”, *15th European Signal Processing Conference* (EUSIPCO), 2007.

[82] G. Berardinelli, L.A. Ruiz de Temino, S. Frattasi, M. Rahman, P. Mogensen, "OFDMA vs. SC-FDMA: performance comparison in local area IMT-a scenarios (recent advances and evolution of WLAN and WMAN standards)," *Wireless Communications, IEEE* , vol.15, no.5, pp.64-72, October 2008.

[83] J. Zhang et al., “Comparison of the Link Level Performance between OFDMA and SC- FDMA”, IEEE ChinaCom, 2006.

[84] T. Lestable, A. Mourad, M. Jiang, Y. Cho, J. Lim, H. Choi, Unveiling Myths about SC-FDMA in TGm, IEEE C802.16m-08/066, Jan. 2008.

[85] Y. Choi, R. Yang, J. Wang, T. Harel, Y. Lomnitz, H. Yin, “On the Multiple Access Schemes for IEEE 802.16m: Comparison of SC-FDMA and OFDMA”, IEEE C802.16m-08/045r1, Jan. 2008.

[86] K. Fazel, “Performance of CDMA/OFDM for mobile communication system”, in Proc. IEEE international conference on Universal Personal Communications (ICUPC’93), Ottawa, Canada, pp.975-979, Oct.1993.

[87] S. Hara, R. Prasad, "DS-CDMA, MC-CDMA and MT-CDMA for mobile multi-media communications ", IEEE 46th *Vehicular Technology Conference*, 1996, vol.2, pp.1106-1110, 28 Apr-1 May 1996.

[88] K. Fazel, S. Kaiser, *Multi-Carrier Spread-Spectrum*, Springer, 2004.

[89] L. Vandendorpe, “Multitone spread spectrum multiple access communication systems in a multipath Rician fading channel”, *IEEE Trans. on Vehicular Technology*, IEEE, Vol. 44, n 2, May 1995, pp. 327-337, 1995.

[90] F. Portier et al., “Transmission techniques for downlink multi-antenna MC-CDMA systems in a Beyond-3G context”, *Journal of Communications and Networks*, vol. 7, pp. 157–170, Jun. 2005.

[91] B. Haindl, M. Sajatovic, C. Rihacek, J. Prinz, M. Schnell, E. Haas, I. Cosovic, "B-VHF - a multi-carrier based broadband VHF communications concept for air traffic management", *Aerospace Conference, 2005 IEEE*, pp.1894-1904, 5-12 March 2005.

[92] F. Bauer, E. Hemming, W. Wilhelm, M. Darianian, "Intercell interference investigation of MC-CDMA," *IEEE 61st Vehicular Technology Conference, 2005* (VTC-2005-Spring), vol.5, pp. 3048-3052 Vol. 5, 30 May-1 June 2005.

[93] N. Labed, F. Bader, I. Gutiérrez, “Supported Active Users in a MC-CDMA System Under Intercell Interference”, *in Proc. 1st IEEE International Conference on Mobile Computing and Wireless Communications* (MCWC), Sept – 2006, Amman, Jordan, 2006.

[94] S. Kondo, L.B. Milstein, “On the use of multicarrier direct sequence spread spectrum systems”, *in Proc. IEEE Military Communications Conference* (MILCOM’93), Boston, USA, pp. 52-56, Oct. 1993.

[95] A. Chouly, A. Brajal, S. Jourdan, “Orthogonal multicarrier techniques applied to direct sequence spread spectrum CDMA systems”, *in Proc. IEEE Global Telecommunications Conference* (GLOBECOM’93), Houston, USA, pp.1723-1728, Dec. 1993

[96] S. Tsumura, R. Mino, Y. Hara, S. Hara, "Performance Comparison of OFDM-FH and MCCDM in Single- and Multi-cell Environments," *in Proc. IEEE Vehicular Technology Conference 2005*, Spring, Stockholm, Sweden, (May 2005).

[97] L. Vandendorpe, “Multitone direct sequence CDMA system in an indoor wireless environment”, *in Proc. IEEE First Symposium of Communications and Vehicular Technology*, Delft, The Netherlands, pp.4.1.1-4.1.8, Oct.1993.

[98] H. Atarashi, N. Maeda, S. Abeta, M. Sawahashi, "Broadband packet wireless access based on VSF-OFCDM and MC/DS-CDMA", *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2002., vol.3, pp. 992- 997 vol.3, 15-18 Sept. 2002.

[99] H. Atarashi, N. Maeda, Y. Kishiyama, M. Sawahashi, "Broadband wireless access based on VSF-OFCDM and VSCRF-CDMA and its experiments", *European transactions on telecommunications*, Special Issue on Multi-Carrier Spread-Spectrum 2004, vol. 15, no 3, pp. 159-172

[100] N. Maeda, Y. Kishiyama, K. Higuchi, H. Atarashi, M. Sawahashi, "Experimental evaluation of throughput performance in broadband packet wireless access based on VSF-OFCDM and VSF-CDMA," *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications*, 2003. PIMRC 2003. , vol.1, no.pp. 6- 11 Vol.1, 7-10 Sept. 2003

[101] Z. Kan, H. Lin, W. Wenbo, Y. Guiliang, "TD-CDM-OFDM: Evolution of TD-SCDMA toward 4G," *IEEE Communications Magazine*, vol.43, no.1pp. 45- 52, Jan. 2005

[102] W. Peng, Z. Yu, C. Dazhong, C. "Research on downlink capacity for broadband wireless access - VSF-OFCDM," *The Fifth International Conference on Computer and Information Technology, 2005* (CIT-2005), pp. 454- 458, 21-23 Sept. 2005.

[103] S. Kaiser, "OFDM Code-Division Multiplexing in Fading Channels", *IEEE Transactions on Communications*, Vol.50, No.8, August, 2002.

[104] Z. Yiqing, W. Jiangzhou, M. Sawahashi, "Downlink transmission of broadband OFCDM Systems-part I: hybrid detection," *IEEE Transactions on Communications,* vol.53, no.4, pp. 718-729, April 2005.

Chapter 3. Performance evaluation and system level metrics

Measuring and comparing the performance of different schemes is sometimes a challenging task, thus it is necessary to have a fixed comparison framework where fair comparisons can be provided. This framework could be based on several deterministic or stochastic models depending on each of the involved and analysed processes during the wireless transmission. Usually, the main actors in these scenarios are: the data streams that need to be delivered under certain QoS, the transmission scheme and the channel. Whereas the last two were studied in the previous chapter, this chapter is dedicated to the review of the different streams models that are required for system performance evaluation. Since the BWA systems must cope with the multimedia traffic, different service classes as well as test services are proposed for system design and evaluation. The performance of each combination must be measured according to standardized metrics which show relevant insights of the system behaviour. This chapter aims to overview some of the proposed models and metrics for both system and link performance evaluation. These concepts will be used in the following chapters when developing resource allocation and scheduling algorithms for multiuser systems with multimedia QoS requirements.

The following research issues are addressed in this chapter:

- First, the services that the network must provide are introduced as well as the models proposed for generating synthetic streams of video, voice and data streams. These models will be later used for testing and evaluation of the system performance under quite realistic system conditions.
- Next, the performance metrics measured both at the link level and at the system level are summarized. These metrics attain to characterize the utilization of the radio channel as well as the QoS offered to each user.
- Finally, some notes are given dealing on system fairness and the most common fairness metrics.

3.1. Services and traffic models

Most of the BWA systems are designed in order to cover the different users needs: *voice services* (e.g. VoIP and conference call), *data services* (e.g., email, instant messaging, web browsing, file transfer, internet gaming, white boarding), and *multimedia services* (e.g., audio and/or video streaming, broadcast, interactive conferencing). In fact, the number of services the access network must provide is quite extensive and they are usually grouped into different classes of services according to how delay sensitive the traffic is. On the other hand, each service requires that a certain level of performance is guaranteed. Such performance agreement is known as Quality of Service (QoS). Furthermore, for system evaluation and simulation each stream of data can be modelled according to complex packet arrival rates and packet size distributions. In this section, the classes of service proposed in the main BWA standards reviewed in the previous chapter (i.e. IEEE 802.16e, 3GPP-LTE, etc.) are reviewed as well as the models that define some of the more common traffic flows. These models will be used later throughout this thesis for system and link level simulations of the resource allocation and scheduling algorithms reviewed in the following chapter.

3.1.1. Test service classes.

The BWA systems may offer a larger variety of services in order to compete with their fixed broadband access counter parts. According to [1]-[4], the following classes of services are defined for a BWA network:

- *Unsolicited Grant Service (UGS)* class*:* designed to support real-time services that generate fixed data packets size on a periodic basis (e.g. VoIP),
- *Real-Time Polling Service (rtPS)* class*:* fitted to support real-time services that generate variable data packets size on a periodic basis (e.g. video conference, MPEG, etc.),
- *Non-real Time Polling (nrtPS)* class: similar to the *rtPS* but they are delay tolerant, hence only minimum data rate is specified,
- *Interactive class:* the traffic is always between peers characterized by a request-response pattern, the round trip delay is then the key attribute,
- *and Best Effort (BE)* class: designed to support a data transmission when no minimum service level is required.

Whether a user requires any of these services, the network assigns a connection identifier to that user. Each user may have several active service flows simultaneously, hence each service flow has a service flow identifier assigned and it is mapped to any of the previously mentioned service classes. In order to simplify the designing phase some *test services* are proposed to use during system performance evaluation. In the following subsection the test services properties and requirements are described.

3.1.2. Streaming class (large delay real time services).

This service class groups applications such as video streaming, data streaming, video-on-demand, multicasting, etc. which are one way streams. They are characterized by the time relations (variation) between the information entities (i.e. samples, packets) within a flow which must be preserved (although it does not have any requirements on low transfer delay). The delay variation of the end-to-end flow must be limited to preserve the time relation (variation) between information entities of the proper stream. However, since the stream is normally time aligned at the receiving end (in the user equipment), the highest acceptable delay variation over the transmission media is given by the capability of the time alignment function of the application [1].

The maximum delay variation of the end-to-end flow is typically 300 ms, which is the tolerated delay for the large delay constrained circuit-switched service defined in [2]. However, depending on the buffers implemented at the receiver side, this maximum delay may be increased. The target Bit Error Rate BER_{target} is 10^{-6} or equivalently a Packet Error Rate (PER) lower than 1%. The traffic associated to this service is asymmetric, with large downlink bit rate requirements and low uplink bit rate requirements. Furthermore, a user is defined in outage for streaming video service if the 98th percentile video frame delay is larger than 5 seconds. The system outage requirement is such that no more than 2% of users can be in outage [5].

3.1.2.1. Streaming test service traffic model.

The streaming test traffic proposed in [4] and [7] is a variable bit rate video traffic which simulates the behaviour of the MPEG2 protocol. The period during which the service is active is known as a session. The session arrival is modelled by a Poisson process. Then, each session is composed by different scenes each of them started by a *scene-change frame* and followed by several *intra-scene frames,* with a constant frame rate f_{FR}. The number of scenes N_{sc} within a session is a geometrically random variable with mean μ_{Nsc}, thus the probability density function (*pdf*) becomes,

$$f_x(x) = \frac{1}{\mu_{N_{sc}}}\left(1-\frac{1}{\mu_{N_{sc}}}\right)^{x-1}, \quad for \quad x \in \aleph. \tag{3.1}$$

Next, each scene contains a random number of video frames. The number of video frames follows a truncated Pareto distribution with the parameters α_0, k_0 and m_0, where α_0, k_0 adjust the shape and the location of the probability density function respectively and m_0 fixes the maximum frame size. The probability of the density function for the truncated Pareto distribution (without cut-off) takes the following expression,

$$f_x(x;\alpha,k) = \left\{ \begin{array}{ll} \alpha \cdot \dfrac{k^\alpha}{x^{\alpha+1}}, & for \quad k \le x < m \\ \left(\dfrac{k}{m}\right)^\alpha, & for \quad m \le x \end{array} \right\}, \quad x \in \aleph, \quad \alpha \ge 1 \tag{3.2}$$

where the mean value is given by the following expression,

$$\mu_x = \frac{\alpha k - m\left(\dfrac{k}{m}\right)^\alpha}{\alpha - 1}. \tag{3.3}$$

The size of the frames is also modelled as a random variable with truncated Pareto distribution with the parameters α_1, k_1, m_1 for the *scene-change frames* and with α_2, k_2, m_2 for the *intra-scene frames*. Moreover, in order to obtain correlated video frame sizes during each scene, the sizes of the *intra-scene frames* are correlated according to a Discrete Auto Regressive (DAR) filter with the following transition matrix:

$$P = \rho I + (1-\rho)\left(\begin{bmatrix} f(0;\alpha_2,k_2) \\ f(1;\alpha_2,k_2) \\ \vdots \\ f(m_2;\alpha_2,k_2) \\ \sum_{x=K}^{\infty} f(x;\alpha_2,k_2) \end{bmatrix}_{(K+1)\times 1} \times \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}_{1\times(K+1)} \right)^T \tag{3.4}$$

where I is the identity matrix, ρ is the auto-correlation factor and K corresponds to the peak frame size. The maximum frame size m_1, m_2 is obtained as a function of the Peak Bit Rate (PBR) and the frame rate. The parameter k_2 is determined so that the Average Bit Rate (ABR) of intra-scenes frames equals the GBR. The rest of the parameters have been determined according to experimental tests in [4], and are shown on Table 3.1 for two streaming qualities (e.g. 2Mbps and 384Mbps of average bit rate). The following Figure 3.1 illustrates the model here reviewed.

Nevertheless, some authors have indicated that the synthetic video traces generated by the analytical model are quite different from the reference traces therefore it might be very difficult to capture the core characteristics of the streaming video traffic. Actually, some authors prefer to use real streaming video traces during the simulation environment in order to obtain more realistic results [7].

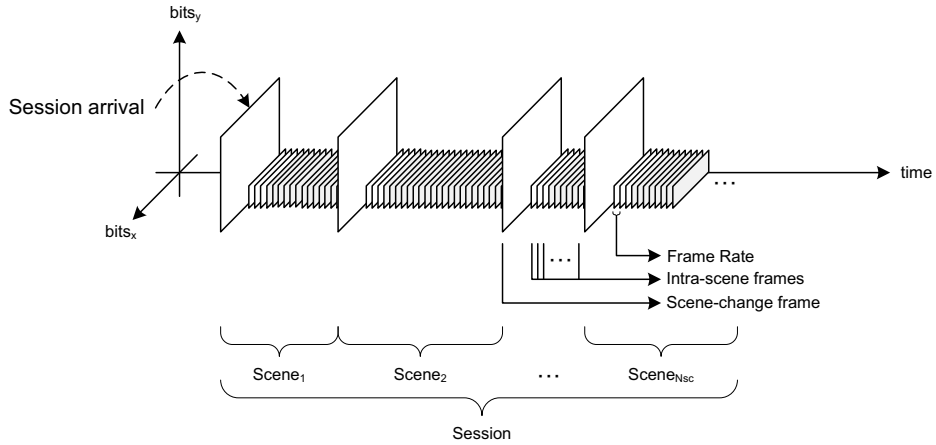


Figure 3.1. Streaming test service traffic model.

Streaming test services		
Peak Bit Rate (PBR)	10 Mbps	2 Mbps
Guaranteed and Average Bit Rates (GBR,ABR)	2 Mbps	384 Mbps
Frame Rate (f_{FR})	10 Hz	
Mean number of scenes per session (μ_{Nsc})	1500	
Video frames per scene (Pareto dist.)	α_0 = 1.1 ; k_0 = 4 m_0 = 100	α_0 = 1.1 ; k_0 = 4 m_0 = 100
Scene-change frames (Pareto dist.)	α_1 = 4 ; k_1 = 50 KB m_1 = 128 KB	α_1 = 4 ; k_1 = 10 KB m_1=25.6 KB
Intra-scene frames (Pareto dist.)	α_2 = 4 ; k_2 = 19 KB m_2 = 128 KB	α_2 = 4 ; k_2 = 3.6 KB m_2=25.6 KB
DAR	K = 59KB, ρ = 0.98	

Table 3.1. Streaming test service traffic model parameters.

3.1.3. Conversational class (low delay real time services).

The conversational class includes applications such as telephony speech, Voice over IP (VoIP), video-conference, etc. Real time conversations are always performed between peers (or groups) of live (human) end-users. In consequence, the maximum transfer delay is given by the human perception of audio and video conversation (i.e. 50 ms for voice) [1][2]. A user is in voice outage if more than 2% of the packets are dropped, erased or not delivered successfully within the delay bound. The reliability of the link usually falls in the range 10^{-4}<BER$_{target}$<10^{-6}. The traffic associated to this service is symmetric, with the same guaranteed bit rate for both downlink and uplink.

3.1.3.1. Video conference test service traffic model.

The video conference traffic is modelled as variable bit rate traffic [4][5]. The period during which the service is active is known as a session (the session has an average duration of 150

seconds). Actually, a video conference is modelled equivalently to a scene from the streaming model. The session arrival is also modelled by a Poisson process. Then, the number of frames within a session is modelled as geometrically random variable with mean μ_{Nfr} where each frame has a size modelled as a random variable with truncated Pareto distribution with the parameters α, k, m (shape, location and maximum frame size respectively). The frame rate is fixed and equal to 10 frames per second. Next Table 3.2 summarizes the relevant values for this test service regarding two video conference qualities (e.g. 2Mbps and 384Kbps of average bit rate).

Video Conference (Low Delay real time) test services		
Peak Bit Rate (PBR)	10 Mbps	2 Mbps
Guaranteed and Average Bit Rates (GBR,ABR)	2 Mbps	384 Kbps
Frame Rate f_{FR}	10 Hz	10 Hz
Mean Number of frames per session μ_{Nfr}	150	150
Intra-scene frames (Pareto dist.)	α = 4 ; k = 19 KB m = 128 KB	α = 4 ; k = 3.6 KB m = 25.6 KB
DAR	K = 59 KB, ρ = 0.98	K = 59 KB, ρ = 0.98

Table 3.2. Video Conference test service traffic model parameters.

3.1.3.2. Voice over IP test service traffic model with activity detection.

Nowadays, one of the hottest topics in broad wireless access networks is the management of Voice over IP (VoIP) traffic in all IP networks [3][5]. Recalling that for voice transmission the delay requirements are very tight, and that the latency and the maximum delay of the network must be guaranteed, an efficient resource management and traffic prioritization needs to be performed to satisfy such QoS requirements. In case the access networks as well as the transport network manage to operate efficiently the voice traffic, VoIP might turn out into the killer application that forces most of the second and third generation mobile networks to migrate to the new fourth generation BWA networks.

The VoIP model with activity detection considers two possible activity states as shown in Figure 3.2. The transitioned probabilities (p_a, p_{ra}, p_i, p_{ri}) are obtained according the steady-state equilibrium where $P_{S0}+P_{S1}=1$, where P_{S0} and P_{S1} are the probabilities of being in the state "0" or the state "1" respectively. In consequence,

$$P_{S_0} = \frac{P_i}{P_i + P_a} \quad ; \quad P_{S_1} = \frac{P_a}{P_i + P_a}. \tag{3.5}$$

With these probabilities, the mean silence period μ_i and the mean talk spurt μ_a (period between entering the active state and leaving it) is given by

$$\mu_i = \frac{1}{P_i} \quad ; \quad \mu_a = \frac{1}{P_a} \tag{3.6}$$

where the mean time between active states entries is then given by

$$\mu_{AE} = \mu_i + \mu_a \tag{3.7}$$

which leads to the following mean arrival rate

$$\bar{R}_{AE} = \frac{1}{\mu_{AE}}. \tag{3.8}$$

The model is then updated at the speech encoder frame rate $R=1/T$, where T is the encoder frame duration (typically T=20ms). As defined in [12], if we stay into the active state, a packet of fixed length (i.e. 49 Bytes for a 12.2Kbps codec rate) is sent during each frame, whereas when staying into an inactive state a *Silence Insertion Descriptor* (SID) is sent periodically.

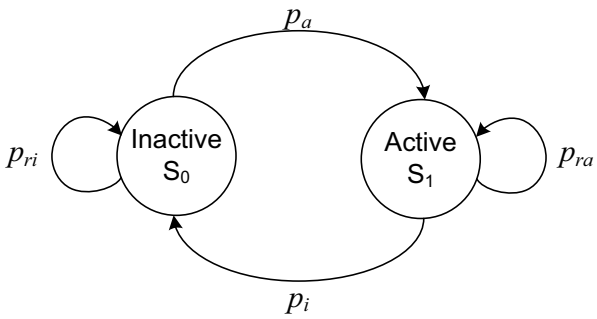


Figure 3.2. Voice over IP activity model

Voice over IP (extended real time) test service	
Codec Rate	12.2Kbps
Encoder Frame length	20ms
Voice Payload	33 bytes for active, 7 bytes for inactive
Voice activity factor (p_{S_1})	50% (p_a=0.01, p_{ra}=0.99)
Encoded Packet Size (when active)	81 Bytes (including headers) or 44 Bytes in case of header compression
Silence Descriptor (including headers) **(sent every 160ms during silence)**	55 Bytes (including headers) or 18 Bytes in case of header compression
Packet encapsulation considered	MAC header + IPv4 + UDP + RTP

Table 3.3. Voice over IP test service traffic model parameters.

3.1.4. Unsolicited Grant Service.

The Unsolicited Grant Service (UGS) is designed to support real-time service flows that generate fixed-size data packets on a periodic basis. Then, a fixed-size grant is provided to the UGS service flows at fixed intervals without additional polling or interactions. In consequence, much of the overhead associated with the polling flow types, as well as the latency, are eliminated. Typical transfer modes and applications that use this kind of services are; the Asynchronous Transfer Mode (ATM), T1/E1 classical PCM (Pulse Coded Modulation) phone signal transmission, Voice over IP communications without silence suppression, etc [1][2].

3.1.5. Interactive class (Non-real time services).

The interactive class could be applied to those connections in which the end-user, that could be either a machine or a human, is on-line requesting data from a remote equipment. This test

services have a model applications such as; the *www* browsing, the data base retrieval, the automatic data base enquiries, emails, automatic measurements, etc.

The interactive traffic is characterized by the request response pattern (round trip delay time is one of the key attributes), and also by the transparent transference of the packets (meaning low BER, e.g. $BER_{target}=10^{-6}$) [1][2].

3.1.5.1. Web browsing test service traffic model.

The traffic associated to a web browsing session is clearly asymmetric, where the peak bit rate for the downlink is much larger than for the uplink. The traffic model proposed in [11] for web browsing model is based on a bursty packet transmission as represented in Figure 3.3. The session arrival is modelled by a Poisson process. Each session is composed by a number N_{pc} of packet calls, where each packet call is formed by a sequence of N_d packets that represent the download of one document (i.e. a web document). The time between packet calls D_{pc} corresponds to the response time (i.e. the reading time for the previous web document), which starts when the last packet of the previous packet call arrives to the BS and finishes when the user makes a new request (the propagation time is considered negligible). The time between packets inside each packet call D_d and the size of each packet S_d are modelled as random variables. Hence N_{pc}, N_d, D_{pc}, D_d are geometrically distributed random variables and S_d is modelled by a Pareto distribution, the values for these distributions are depicted in Table 3.4. The peak bit rate (PBR) for this service is given by

$$PBR = \frac{\mu_{S_d} \cdot \mu_{N_d}}{\left(\mu_{N_d} - 1\right) \cdot \mu_{D_d}} \cdot \tag{3.9}$$

Nevertheless, more accurate traffic models have been recently proposed which may achieve more realistic simulations at the expense of computational cost [9]. Typically, the maximum tolerated packet error rate for this service is 1%.

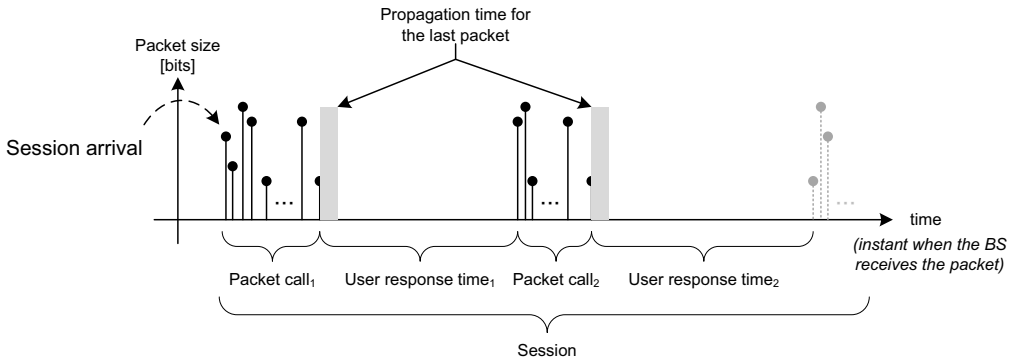


Figure 3.3. Web browsing traffic model description.

Web browsing (www) test services		
Peak Bit Rate	10 Mbps	2Mbps
Avg. number of packet calls in a session ($\mu_{N_{pc}}$)	5	5
Avg. response time between packet calls ($\mu_{D_{pc}}$)	412 s	412 s
Avg. number of packets in a packet call (μ_{N_d})	25	25
Avg. inter-arrival time between packets (μ_{D_d})	381 μs	1.9 ms
Packet size distribution (Pareto dist.)	α = 1.1 ; k = 81.5 B m = 66 KB	α = 1.1 ; k = 81.5 B m = 66 KB

Table 3.4. Web browsing test service traffic model parameters.

3.1.6. Best Effort (or background) QoS class.

The background class applies to those connections where the systems on both sides are quite insensitive to the delivery time. The applications under this test service are those such as background delivery of emails, file transferring, short message service (SMS), download of databases, reception of measurement records, etc. Another characteristic is that the content of these packets must be transparently transferred (i.e. low bit error rate, e.g. $BER_{target}=10^{-6}$) [1][2].

3.1.6.1. File transferring test service traffic model.

This communications are generally unidirectional (downloading or uploading of contents), however the overall traffic generated in the network is considered symmetric. The traffic model proposed in [11] for file transferring is the same as for the interactive class, with the difference that each packet call contains only one packet. The parameters for the background traffic are depicted in Table 3.5. A maximum tolerated packet error rate of 1% is usually required.

FTP test service	
Peak Bit Rate	10 Mbps
Average number of packet calls in a session ($\mu_{N_{pc}}$)	5
Average response time between packet calls ($\mu_{D_{pc}}$)	412 s
Packet size distribution (Pareto dist.)	α = 1.1 ; k = 4.5 KB; m = 2 MB

Table 3.5. FTP test service traffic model parameters.

3.1.6.2. Full buffer test service traffic model.

Finally, the full buffer traffic model is the simplest model where it is assumed that the input buffers have always data bits waiting to be transmitted [12]. For uplink traffic the transmit buffers exist in the users terminal while for the downlink traffic the traffic buffers exist in the network infrastructure (e.g. in the base station). In consequence, there is always a constant amount of data that needs to be transmitted both in the downlink and the uplink. Usually, when considering this traffic model, no delay requirements are imposed despite a maximum packet error rate of 1% is required.

3.2. Criteria for evaluation of the radio interface

When two technologies or algorithms must be compared we usually require some common metrics in order to conclude which one performs better in each scenario. One of the key parameters in which we can support our conclusions is the number of satisfied users for each technology. However, this metric hides much of the exact network performance. Hence the spectrum efficiency, system coverage, bit error rate, system throughput, delay, etc. might be also obtained in order to gain an inside view of the network behaviour and performance. Furthermore, the recommendation ITU-R M.2135 fixes some of the targets for these metrics such that the radio interface can be considered as an IMT-Advanced RF interface candidate [8]. In this section these parameters and measurements are introduced hence no additional formulation will be required in the later chapters in order to compare the different proposals.

3.2.1. Definition of satisfied user

One of the main system metrics is the percentage of satisfied users in the system. This evaluation must be done according to the different services used by each user. For the reviewed packet-switched services in Section 3.1, a satisfied user is defined as the user that has all the following constraints fulfilled [4][9]:

- The user doesn't get blocked when getting to the system, and in case there are inputs buffer (queues), the packets are not lost due to overflow.
- The active session throughput is equal or higher than 95%, 70% and 10% of the guaranteed bit rate for the high delay real time services, low delay real time services, and non-real time services respectively.
- In case of real-time service, the delay does not exceed the maximum tolerated delay for more than 90% and 95% of the packets for streaming and conversational class respectively which means that the traffic is conformant with the guaranteed bit rate.
- The user does not get dropped. For non real time service, a user is dropped if the BER obtained is higher than the BER_{target} for 1% of the frames. On the other hand, for real-time services a call is dropped if the BER is higher than BER_{target} during all the frames transmitted in a period equal to $max(5, 10/(GBR \cdot BER_{target}))$ seconds.

3.2.2. Radio interface performance metrics

In this section some of the more common metrics used to evaluate the performance of the radio interface are described. These metrics should be obtained during link level and system level simulations and should be enough in order to asses which scheme perform best in each scenario.

3.2.2.1. Cell spectral efficiency

The cell spectral efficiency is defined as the sum of all correctly received bits (bits interchanged from the network layer entities) in the network divided by the channel bandwidth divided by

the number of cells (a cell is equivalent to a sector). The cell spectral efficiency is measured in bits/s/Hz/cell.

Hence, the cell spectral efficiency (η) might be obtained by

$$\eta\left[bits/s/Hz/cell\right]=\frac{\sum_{k=1}^{K}b_k}{T\times W\times M} \tag{3.10}$$

where b_k denotes the number of correctly received bits by/from user k during the a period of T seconds, in a system comprising a user population of K users and M cells. W denotes the channel bandwidth.

The following Table 3.6 shows the average spectral efficiencies required for downlink and uplink assuming an antenna configuration of 4×4 and 2×4 respectively [8]. The peak spectral efficiency (computed as the highest theoretical data rate assuming an error-free channel) is also depicted.

Test environment	Downlink (bits/s/Hz)		Uplink (bits/s/Hz)	
	Average	Peak	Average	Peak
Indoor	3	15	2.25	6.75
Urban (<60Km/h)	2.2 - 2.6		1.4 - 1.8	
High Speed (>60Km/h)	1.1		0.7	

Table 3.6. Required cell spectral efficiency for IMT-Advanced.

3.2.2.2. User average service throughput

For each user k, the average user throughput R_k is defined as the ratio between the bits correctly received b_k and the simulation time. The ratio is mathematically expressed as

$$R_k\ [\text{bits/s}]\ =\frac{b_k}{T}. \tag{3.11}$$

3.2.2.3. Cell edge performance metrics

The previous performance metrics can be also computed for the cell edge users which usually fix and limit the maximum system capacity. Hence, the cell edge user spectral efficiency is defined as 5% point of the cumulative distribution function (CDF) of the cell spectral efficiency. Following the recommendation ITU-R M.2134, the values depicted in the must be achieved by any IMT-advanced radio frequency interface [8].

Test environment	Downlink (bits/s/Hz)	Uplink (bits/s/Hz)
Indoor	0.1	0.07
Urban (speed<120Km/h)	0.06-0.075	0.03-0.05
High Speed (speed >120Km/h)	0.04	0.015

Table 3.7. Cell edge user minimum spectral efficiency.

3.2.2.4. System Capacity

When several services are simultaneously provided by the network, the system capacity is defined as the system load when one of the services has exactly 98% of satisfied users and the rest have at least 98% satisfied users. On the other hand, in case only one service is provided the system capacity is defined as the system load where exactly 98% of the users are satisfied.

3.2.2.5. Voice over IP Capacity

The VoIP capacity is the minimum of the calculated capacity for either link direction divided by the effective bandwidth (in TDD system this is the operative bandwidth normalized by the downlink/uplink ratio) in the respective link direction.

The VoIP service assumes is the one described in Section 3.1.3.2 (12.2 kbit/s codec with a 50% activity factor). A user is defined to have experienced a voice outage if less than 98% of the VoIP packets have been delivered successfully to the user within the allowed VoIP packet delay bound of 50 ms. The following Table 3.8 shows the minimum VoIP that should be obtained according to the ITU-R M.2134 recommendation [8] for IMT-Advanced radio interfaces.

Test Environment	Minimum VoIP Capacity (Active Users/sector/MHz)
Indoor	50
Urban (speed<120Km/h)	40
High Speed (speed >120Km/h)	30

Table 3.8. Minimum Voice over IP Capacity.

3.2.2.6. Latency and Packet Delay

The latency is typically measured as the one-way transit time between Layer 3 homologous identities. Hence, we can measure the latency for control/management packets (control plane) and for data packets (user plane). It is required that for the user plane a small IP packet (zero bytes payload) the latency achieved becomes lower that 10ms in unloaded conditions (a single user with a single data stream) [8].

On the other hand, the packet delay is typically measured as the time elapsed since the packet enters the transmitter queue at the BS and the time when the packet arrives successfully at the receiver side. The propagation time is not considered since it is negligible in front of the queuing time. Nevertheless, the cumulative density function (CDF) of the packet delay provides a basis in which maximum delay, x-th percentile, average delay as well as jitter can be derived.

3.2.2.7. Average packet delay per sector

The average packet delay per sector (ω) is given by the ratio of the accumulated delay from all the correctly received packets in the network divided by total number of packets transmitted divided by the number of sectors. The average packet delay per sector is then given by

$$\omega = \frac{1}{P}\sum_{p=1}^{P}\tau_p \tag{3.12}$$

where τ_p is the delay for each p-th packet. P means the total number of received packets.

3.3. Fairness Criteria

On of the objectives that also needs to be considered during system evaluation and optimization is the fairness among users or services flows belonging to the same service class (or under the same QoS requirements) can be also contemplated. The concept of fairness, or inequality, arises in wireless networks when a limited the bandwidth is shared among several individuals/users simultaneously. However, several fairness criteria have been defined in the literature [10]. Among those, one of the best known and accepted fairness metric is the *Jain's Fairness Index* (FI) which is defined by

$$FI_{Jain} = \frac{\left|\sum_{i=1}^{K} x_i\right|^2}{\sum_{i=1}^{K}\left|x_i\right|^2}, \tag{3.13}$$

where x_i is the metric over which we want to estimate the fairness and K is the set of active users that require resources from the network. It is noted that if FI_{Jain}=1 all the users are given the same amount of resource, whereas FI_{Jain}=0 means maximum unfairness. Whether FI_{Jain} is measured during short intervals (lower than 40ms), the minimum value of $FI_{Jain}(t)$ can be used as an indicator of how much fairness is maintained all the time [12].

3.4. Summary of performance evaluation

Since the BWA systems must cope with multimedia traffics, the different service classes as well as the test services proposed for the system design and the evaluation have been resumed in this chapter. Furthermore, for system performance evaluation it is necessary to have models that allow you the simulation of the different behaviours of all the components involved in the communications. For that reason, several models have been outlined here to characterize the different video, voice and data streams that are transferred between the network and the user and vice versa. In addition, since each user requires a certain quality of service, the service requirements as well as the performance metrics have been also considered. These concepts will be used in the following chapters when the proper resource allocation and scheduling algorithms for multiuser systems with multimedia (tight) QoS requirements need to be developed and analysed.

References

[1] ETSI, “Universal Mobile Telecommunication System (UMTS)”; Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03 version 3.2.0)”, TR 101 112 v.3.2.0, April, 1998.

[2] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; QoS Concept and Architecture (3G TS 23.107 version 3.0.0), Oct.1999

[3] Quality of Service (QoS) classes for BWA, IEEE 802.16sc-99/28, IEEE 802.16 Broadband Wireless Access Working Group, July 22, 1999

[4] End-user multimedia QoS categories, ITU-T G.1010, INTERNATIONAL TELECOMMUNICATION UNION, 29 Nov. 2001.

[5] IEEE 802.16m System Requirements, IEEE 802.16m-07/002r7, IEEE 802.16 Broadband Wireless Access Working Group, Dec. 2008.

[6] Trace Based Streaming Video Traffic Model for 802.16m Evaluation Methodology Document, IEEE 802.16m-07/145, IEEE 802.16 Broadband Wireless Access Working Group, July, 2009.

[7] D.P. Heynman, T.V. Lakshman, "Source Models for VBR Broadcast-Video Traffic", *IEEE/ACM Transaction on Networking*, Vol.4, No1, February 1996.

[8] Requirements related to technical performance for IMT-Advanced radio interface(s), Report ITU-R M.2134, 2008.

[9] C. Johansson, L. De Verdier, F. Khan, "Performance of different scheduling strategies in a packet radio system," *IEEE 1998 International Conference on Universal Personal Communications* (ICUPC-98) , vol.1, pp.267-271 vol.1, 5-9 Oct. 1998.

[10] M. Dianati, X. Shen, S. Naik, "A new fairness index for radio resource allocation in wireless networks," *Wireless Communications and Networking Conference, 2005 IEEE*, vol.2, pp. 712-717 Vol. 2, 13-17 March 2005.

[11] Reference scenario specification: final description, Deliverable 1.3, Multicarrier CDMA TRansmission Techniques for Integrated Broadband CEllular Systems, 2001-4.5.2 Terrestrial wireless systems and networks, IST-2001-4.5.2 Terrestrial wireless systems and networks.

[12] “IEEE 802.16m Evaluation Methodology Document (EMD)”, IEEE 802.16m-08/004r4, IEEE 802.16 Broadband Wireless Access Working Group, Nov. 2008.

Part II

Contributions

Chapter 4. Link adaptation in single antenna links

The migration to 4G networks will bring a new level of expectations to wireless communications. Just as the digital wireless revolution in the 1990s made mobile phones available for everyone, the higher speeds and packet delivery of 4G networks will make high-quality multimedia available everywhere. According to the International Telecommunication Union (ITU) definition, Next Generation Networks (NGN) should be able of handling multiple broadband end-to-end QoS-enabled interfaces where packet switching should be used across the entire network [1]. Coming to wireless networks, the ITU has also specified some requirements related with the radio channel, these systems (4G) need higher spectral efficiencies (>10bits/s/Hz) targeting different environments as stationary, pedestrian, or vehicular, and seamless handover plus IP connectivity to other networks [2].

In a BWA system, a full (or Partial) Channel State Information (CSI) from all the users is assumed at the Base Station (BS), link adaptation techniques, e.g. power allocation or adaptive modulation and coding (AMC) can be combined with user scheduling in order to exploit efficiently the multiuser diversity and thus increasing the spectrum efficiency, the data rate, and the cell coverage. However, it is necessary to control the resource allocation and their utilization much more accurately than that in the classical statistics based techniques. An air interface that has gained many supports is the OFDMA based scheme, where the users can be multiplexed efficiently in the time and frequency domains, furthermore, multiple antenna techniques can be also applied without too much extra complexity due to the underlying OFDM scheme (this topic will be covered further in Chapter 6).

In this chapter, the following issues are addressed:

- First the capacity of single antenna links is reviewed as well as the different methods to optimize the spectral efficiency in single user scenarios.
- Afterwards, a brief review of some link level to system level mapping functions is given by developing a closed form expression in case of the general OFCDM transmission scheme.
- Finally, the almost well known techniques for link adaptation as adaptive modulation and coding, power loading, Rate Adaptation (RA), and Margin Adaptation (MA) are explained by bringing some new results. A comparison between the several classical solutions for the RA optimization problem is also provided.

4.1. The channel capacity of single antenna links.

One of the main performance metrics of any wireless data communication system is the spectral efficiency. The spectral efficiency is the amount of error free information (measured in bits/s/Hz) that is transmitted over a given bandwidth following a certain scheme (see Section 3.2.2) and is equivalent to the rate R at which to physical layer entities exchange some information. On the other hand, the maximum amount of information that can be transmitted through the channel is defined by the (Shannon) channel capacity [3] which is a theoretical upper bound based on ideal channel coding with infinite codeword length and ideal modulation schemes (i.e. Gaussian symbol entries).

To understand better these concepts lets first define the system model of a wireless point-to-point (with one antenna at both transmitter and receiver sides). The schematic of the single input single output (SISO) system is depicted in Figure 4.1, the discrete-time channel output is expressed as

$$\mathbf{y} = h \cdot \mathbf{x} + \mathbf{n}. \tag{4.1}$$

where $\mathbf{x}$, $\mathbf{n}$, $\mathbf{y} \in \mathbb{C}^{T\times 1}$ each following a specific distribution, where T is the codeword length.

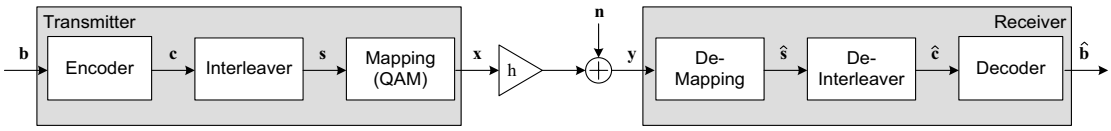


Figure 4.1. Discrete-time system model of a single antenna link.

The capacity of any channel is then defined as the maximum mutual information between $\mathbf{x}$ and $\mathbf{y}$ which is given by

$$C(h) = \max_{p(\mathbf{x})} I(\mathbf{x};\mathbf{y}) = \max_{\mathbf{x}} \left\{ \sum_{\mathbf{x},\mathbf{y}} p(\mathbf{x},\mathbf{y}) \log_2 \left(\frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{x})\,p(\mathbf{y})} \right) \right\}, \tag{4.2}$$

and whether the noise considered is Additive White Gaussian Noise (AWGN), i.e. $\mathbf{n} \sim \mathbb{C}\mathcal{N}(0,\sigma_N^2)$, we could apply the *Law of Large Numbers,* then the channel capacity is expressed as follows

$$C(h) = \log_2 \left(1 + h \frac{E\{\mathbf{x}\mathbf{x}^T\}}{E\{\mathbf{n}\mathbf{n}^T\}} \right) = \log_2 \left(1 + h \frac{\sigma_{\mathbf{x}}^2}{\sigma_N^2} \right) = \log_2 (1 + h\rho), \tag{4.3}$$

where ρ means the Signal to Noise Ratio (SNR). Then, the Shannon's channel coding theorem states that when the rate R is lower than the channel capacity ($R<C$), it exists at least one channel encoder and one decoder for the memory less channel such that the error probability $P_e \to 0$ as $T \to \infty$.

Now, if we consider the continuous-time AWGN channel (i.e. $\mathbf{y}(t)=h\mathbf{x}(t)+\mathbf{n}(t)$), where the random process x(t) spans over a limited bandwidth W[9], the channel capacity is given by

$$C(h) = W \log_2 \left(1 + h \frac{\sigma_{\mathbf{x}}^2}{\eta_0} \right) = W \log_2 \left(1 + h \frac{P_{tx}}{W\eta_0} \right), \tag{4.4}$$

where P_{tx} means the transmission power. Thus, the maximum achievable spectral efficiency Υ can be expressed as

$$\Upsilon_{\max}(h) = \log_2 \left(1 + h \frac{P_{tx}}{W\eta_0} \right) = \log_2 \left(1 + h \frac{R \cdot E_b}{W\eta_0} \right). \tag{4.5}$$

It is then shown, that in limited case of very large bandwidth (i.e. $\Upsilon \to 0$), the minimum $E_b/\eta_0 \geq log(2) = -1.59$. However, as the bandwidth is a scarce resource, in most scenarios we will have to sacrifice the power in favour of the bandwidth efficiency.

[9] the vector $\mathbf{x}(t)$ can represented in the geometric domain as a vector in WT dimension signal space over the complex field, i.e. $\mathbf{x}(t) \in \mathbb{C}^{WT\times 1}$.

4.1.1. The Bit Error Rate

It is clear that the main objective of any communication system is to transfer the information from a transmitter to a receiver. It is also well-known that the minimum information entity is one bit (i.e. a random variable with two possible states and uniform distribution, e.g. a fair coin, has entropy of 1 bit). Then, a established performance measurement of the reliability of a link is the bit error rate (BER).

As it is shown in Figure 4.1, the bits in vector **b** are typically encoded and mapped into symbols **s** which are then sent over the channel. A widely used scheme for symbol mapping is the Z-QAM scheme, where $n=log_2(Z)$ bits are mapped into each symbol by adjusting the phase and the amplitude of two orthogonal waves (usually sinusoids of the same frequency). Closed expressions for the Symbol Error Rate (SER) P_e have been derived for Z-QAM modulation as a function of the SNR in an AWGN channel, by applying the *nearest neighbour union bound* (NNUB) where having an even n, the SER is given by

$$P_e \le \frac{1}{2} N_n(Z) \cdot Q\left(\sqrt{\frac{E_s}{N_0} \frac{d^2_{\min,Z-QAM(Z)}}{2}} \right)^{10} \tag{4.6}$$

where the E_S/N_0 is equivalent to the Signal to Noise Ratio, $N_n(Z)$ means the number of closest neighbours, and d^2_{min} is the minimum squared distance between constellation points. For a rectangular QAM scheme both terms can be obtained as

$$N_n(Z) = 4\left(1 - \frac{1}{\sqrt{Z}}\right) \quad , \quad d^2_{\min} = \frac{6}{Z-1}. \tag{4.7}$$

Moreover, if the bits are mapped following a Gray ordering (neighbour symbols differ in one bit) we can assume that if a symbol is in error, only one bit is incorrect (i.e. the bit error probability $P_b=P_e/n$). Furthermore, for high SNR, the BER is usually spread as

$$P_{b,Z-QAM} \le 0.2 \cdot \exp\left[\frac{-1.6}{(Z-1)} \frac{E_s}{N_0} \right] \tag{4.8}$$

which for $n\ge2$ and BER$\le10^{-3}$ is tight up to 1dB [5].

4.1.1.1. The Bit Error Rate in Rayleigh channels

It can often occur that the experienced wireless channel h is time-variant following a Rayleigh (or Ricean) distribution (details have been presented in Section 2.1.3). It is important to understand how the wireless channel influences the bit error rate behaviour. Whether a radio channel remains constant during the transmission of one symbol (slowly fading channel) the BER can be approximated as in [4] by

$$P_{b,Rayleigh} \approx \frac{1}{\alpha\rho}. \tag{4.9}$$

---

[10] The Q-function is defined as $Q(x) = \frac{1}{2} erfc\left(\frac{x}{\sqrt{2}}\right)$ where $erfc$ is the complementary error function.

where α is a constant that depends on the modulation scheme and the number of bits per symbol. Recently, a closed expression for the BER in case of Z-QAM modulation with Gray coding has been developed in [6] leading to the following closed expression

$$P_{b,Rayleigh} = \frac{1}{\log_2 \sqrt{Z}} \sum_{i=1}^{\log_2 \sqrt{Z}} P_{b,Rayleigh}(i) \tag{4.10}$$

with

$$P_{b,Rayleigh}(i) = \frac{1}{\sqrt{Z}} \sum_{k=0}^{\left(1-2^{-i}\right)\sqrt{Z}-1} \omega(k,i,Z) \left(1 - \frac{\sqrt{\frac{3\rho_b (2k+1)^2 \log_2 Z}{2(Z-1)}}}{\sqrt{\frac{3\rho_b (2k+1)^2 \log_2 Z}{2(Z-1)} + 1}} \right), \tag{4.11}$$

and

$$\omega(k,i,Z) = (-1)^{\left\lfloor \frac{k \cdot 2^{i-1}}{\sqrt{Z}} \right\rfloor} \left(2^{i-1} - \left\lfloor \frac{k \cdot 2^{i-1}}{\sqrt{Z}} + \frac{1}{2} \right\rfloor \right) \tag{4.12}$$

where $\rho_b = \rho / log_2(Z)$ and $\lfloor x \rfloor$ means the smallest integer of x. It is then observed in [6] that for BER<10^{-2}, 3 to 4 dB are required each time the constellation increases one bit onto each quadrature component. This is conclusion can also be drawn if an AWGN channel is used. However, due to the low slope of the BER curve as a function of the SNR, if we want to reduce one order of magnitude the BER, the power must be increased 10dB, whereas in the AWGN channel only 1.57dB would be required.

4.1.1.2. Increasing diversity in fading channels

The negative effects of the fading channels on the BER performances force to use novel techniques that increase the diversity (either in time, frequency or space as it will be shown further in Chapter 6). One of these techniques is the combination of channel coding and interleaving (see Figure 4.1). Typically, the channel encoder adds redundancy into the signal thus in case some bits are corrupted during the transmission, they can be later recovered (i.e. *forward error correction* – FEC). The ratio between the number of bits k at the input of the encoder and the n bits obtained at the output is usually known as the code rate $r = k/n$.

According to the previously described Shannon theorem, it should exist a coding scheme that given a data word $\mathbf{b}$ provides a codeword $\mathbf{s}$ which is able to achieve a spectral efficiency very close to the channel capacity. The number of channel coding schemes that attain this premise is quite extensive (e.g. block codes, convolutional codes, turbo-codes, Low Density Parity Check - LDPC – codes, etc.)[4][7].

Since it is not the focus of this thesis to investigate into coding techniques, the convolutional codes are used for evaluation of the schemes proposed. Indeed, convolutional codes are mandatory use in the IEEE 802.16 standard [8].

The Figure 4.2 and Figure 4.3 show the BER performance using convolutional coding with QPSK and 16QAM modulations in an AWGN and Rayleigh channels as a function of the average SNR (i.e. average E_S/N_0). The (punctured) convolutional code is the one standardized in IEEE 802.16 with a polynomial generators of {133_O, 171_O}, constraint length L=7 and a native code rate r=1/2. The puncturing patterns can be found in [8]. The diversity order is increased which is

appreciated by the increase of the negative slope of the BER(SNR) function. Moreover, *hard decoding* – HD (solid lines) and *soft decoding* – SD (dashed lines) are also compared. When hard decoding is applied, the demapper decides which bits have been received and the decoder searches for errors within the received sequence. On the other hand, when a soft decoding is applied, the decoder is fed with the log-likelihood ratio of each bit, hence the decoder may weight according to the channel experienced by each bit (i.e. following the MRC principle) the input data as a function of their reliability. It is then observed according to the simulations carried out by the author (shown in Figure 4.2 and Figure 4.3) that in case the soft decoding is applied in Rayleigh (or Ricean) channels the diversity order is further increased.

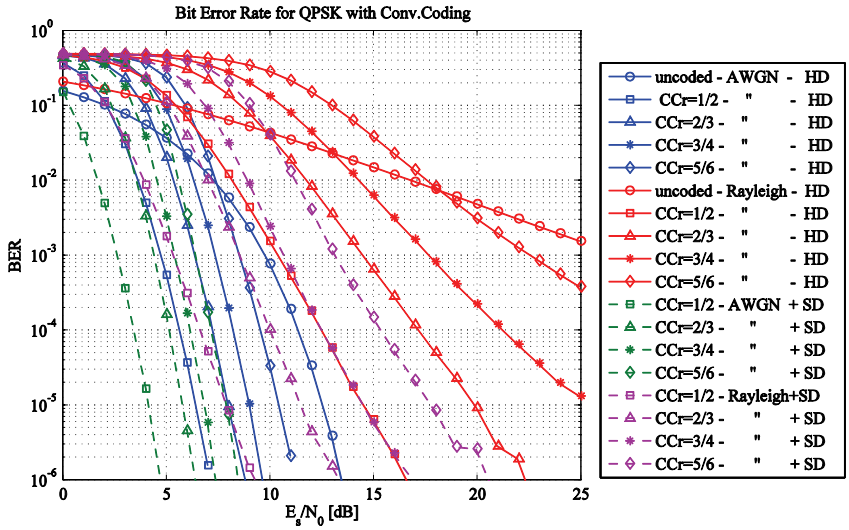


Figure 4.2. BER for QPSK with conv. coding and Hard and Soft Decoding (HD, DS) in case of AWGN and Rayleigh channels (codeword length 4096 bits).

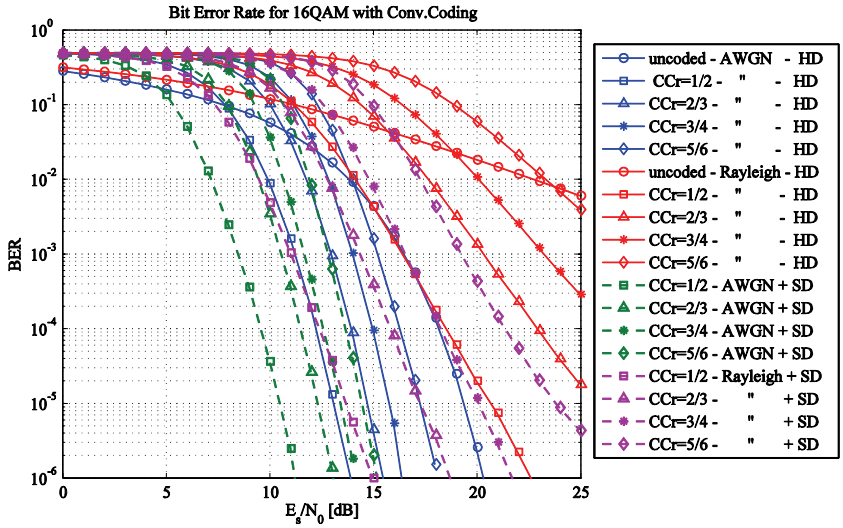


Figure 4.3. BER for 16QAM with conv. coding and Hard and Soft Decoding (HD, DS) in case of AWGN and Rayleigh channels (codeword length 4096 bits).

4.1.2. The Packet Error Rate

Almost all the multimedia traffic types are based on the transmission of synchronous or asynchronous packets of variable sizes. Thus, one of the main performances metric for the

stream performance evaluation is the packet error rate (PER). In general, the PER depends on the post-processing SINR, the packet length, and the Modulation and Coding Scheme (MCS).

If the interference, noise and channel effects are i.i.d. across the packet, the bit errors are also uniformly distributed and hence the PER is expressed by

$$PER = 1 - \left(1 - BER\right)^{N_b}, \tag{4.13}$$

where N_b is the length of the packet in bits and BER is the bit error rate after the channel decoder. Nevertheless, in wireless system where the transmission experiences different BER as a function of the frequency and time interval, it can be considered that several blocks make up the packet where the channel conditions for each block are similar. Therefore, Eq. (4.13) can be rewritten as

$$PER = 1 - \prod_i \left(1 - BLER_i\right), \text{ with } BLER_i = 1 - \left(1 - BER\right)^{N_{b,i}}. \tag{4.14}$$

However, since it is quite difficult to discern which parts of the channel can be considered similar or not, most of the PER estimation techniques are based on channel simulations, and the measurements carried for different packet sizes. Afterwards, the obtained results are summarized in a $PER(SNR,\ MCS,\ N_b)$ function which depends on the channel environment characteristics and the applied transmission scheme.

4.2. Effective SNR Mapping.

Usually, most of the communication systems were optimized considering that the channel conditions do not vary substantially in time and frequency (*in this chapter we consider only single antenna transmission*). In that case, it is possible to select which is the optimum MCS based on the average measured SNR at the receiver. However, actual systems are based on exploiting the instantaneous channel condition to enhance their performances. The adaptive modulation and coding (AMC) techniques as well as the channel resource allocation and scheduling are examples of these introduced adaptable schemes. In consequence, always efficient and more advance techniques to model the dynamic behaviour of the channel are required. Furthermore, the metrics developed for channel modelling need to include all the channel effects, but at the same time being as simple as possible, since as the number of parameters required to compute the metric is increased, the more complex is the adaptation mechanism. In addition, for system level simulations side, the consideration of very efficient adaptive schemes is very important since more accurate system level performances can be obtained without the need of physical layer (i.e. link level) simulation.

Then, since the upper layers are also concerned about the physical layer performance behaviour from the point of view of quality of service (QoS), it is necessary to have models than can predict with high accuracy the block error rate (BLER) or packet error rate (PER) given a specific channel realization. For OFDM-based transmission schemes, a widespread method for channel performance modelling is the *Effective SINR mapping* (ESM). Using the ESM, the channel response matrix (i.e. both frequency and time domains are considered) is compressed into a single value $ESINR$ which is sufficient to predict the system performance (e.g. BLER). The

ESINR would be equivalent to the SNR required in an AWGN channel to obtain the same BLER. According to this premise, the ESINR can be defined as follows

$$ESINR = g(H) \quad \text{with } H \in C^{N_f \times N_t} \tag{4.15}$$

where $g(\cdot)$ is known as the compression function. Then, the BLER becomes a function of the ESINR, the MCS and the packet length,

$$BLER = f(ESINR, MCS, N_b). \tag{4.16}$$

It is important to note that the compression function $g(\cdot)$ should be independent of the multipath channel model and the channel profiles.

4.2.1. Use of Pythagorean means for ESM.

In order to determine the best compression function, one type of compression functions are the Pythagorean means (arithmetic, geometric and harmonic). Despite not having a theoretical justification, the use of this class of means is accepted and widely used in several adaptation mechanisms. The first possible compression function is based on the calculus of the arithmetic mean of the channel values (i.e. the SINR of each transmitted symbol). Then, applying the arithmetic mean over the different SINR values, the $ESINR_{AM}$ is given by

$$ESINR_{AM} = \frac{1}{N_f N_t} \sum_{f=1}^{N_f} \sum_{t=1}^{N_t} SINR_{f,t}. \tag{4.17}$$

Another compression function is the geometric mean, in this case the $ESINR_{GM}$ is given by

$$ESINR_{GM} = \left(\prod_{f=1}^{N_f} \prod_{t=1}^{N_t} SINR_{f,t} \right)^{\frac{1}{N_f N_t}}. \tag{4.18}$$

The $ESINR_{GM}$ can be considered as the arithmetic average of the SINR values in the logarithm domain, which has been used in many applications. Another mean, also proposed in the literature, is the harmonic mean of the SINR samples [12] which is defined as

$$ESINR_{HM} = \frac{N_f N_t}{\sum_{f=1}^{N_f} \sum_{t=1}^{N_t} \frac{1}{SINR_{f,t}}}. \tag{4.19}$$

It is well-known that the $ESINR_{HM}$ gives the lowest mean of the three means (arithmetic, geometric and harmonic). Actually, the $ESINR_{HM}$ is mainly influenced by the worst $SINR_{f,t}$ value.

4.2.2. Generalization of the Exponential ESM.

In order to accurately model the wireless channel, more complex compression functions have been introduced to correctly estimate the performance of the channel. One of such techniques which has gained the interest of many researches (without adding too much complexity) is the Exponential Effective SNR Mapping (EESM) scheme [13][14]. It is shown that over an AWGN channel, the Chernoff bound of the bit error and block error probabilities can be written as follows

$$P_e \leq \exp(-SINR). \tag{4.20}$$

Hence having N parallel AWGN channels, the probability that at least one of the channels produce an error is given by

$$P_e^N = \prod_{n=1}^{N}\left(1-P_e^1\right) \approx SINR\sum_{i=1}^{N}\exp\left(-SINR\right). \tag{4.21}$$

Since, we are looking for the SNR required in a unique AWGN channel, we have to solve the following expression in order to find which is the effective SNR such that

$$N\exp\left(-ESINR\right) = SINR\sum_{i=1}^{N}\exp\left(-SINR\right). \tag{4.22}$$

The generalized solution to the previous equation is then expressed as [13]

$$ESINR_{EESM} = -\beta\log_n\frac{1}{N_f N_t}\sum_{f=1}^{N_f}\sum_{t=1}^{N_t}\exp\left(-\frac{SINR_{f,t}}{\beta}\right). \tag{4.23}$$

where β is a value that is adjusted according the MCS applied and the encoding length. For the previous studied MCS (Z-QAM with Z={4,16} and the convolutional coding scheme described) the values of β proposed in [16] are depicted in Table 4.1.

Modulation	Code Rate	β
QPSK	½	2.9
	¾	2.4
16QAM	½	6.2
	¾	8
64QAM	½	12.6
	¾	27.1

Table 4.1. β values for EESM with QAM modulation and convolutional coding.

4.2.3. Mutual information based ESM (MIESM).

The equivalent channel can be also obtained via the mutual information of the channel as developed in Section 4.1 but considering finite constellation inputs (not Gaussian distributed). The computation of the mutual information per coded bit can be derived from the received symbol-level mutual information leading to the named Received Bit mutual information Rate (RBIR), or the Mean Mutual Information per Bit (MMIB).

In the first case, the RBIR is obtained averaging the symbol mutual information (SI) across the whole transmitted block. The SI function is defined as

$$SI\left(SINR_{f,t},\log_2(Z)\right) = \log_2(Z) - \frac{1}{Z}\sum_{m=1}^{Z}E\left\{\log 2\left(1+\sum_{k=1,k\neq m}^{Z}\exp\left(-\frac{\left|X_k - X_m + U\right|^2 - \left|U\right|^2}{SINR_{f,t}}\right)\right)\right\} \tag{4.24}$$

here X_m, X_k are the values of each constellation point and U is a zero mean Gaussian variable with variance $1/(2SINR_{f,t})$ per component. Then, if we consider that all the symbols apply the same constellation, the RBIR is obtained as

$$RBIR = \frac{\sum_{f,t} SI\left(SINR_{f,t},\log_2(Z)\right)}{\sum_{f,t}\log_2(Z)}. \tag{4.25}$$

The $SI(SINR,Z)$ function can be computed and stored as a Look-Up-Table (LUT) since its values depend only on the constellation function. The reader is referred to [14] for further details

concerning exact value of the function for SINR ranging from -20dB to 27dB, and how to compute the RBIR in case of multiple transmit/receive antennas. The main constraint from the RBIR approach is that it doesn't consider the constellation mapping neither the bit interleaver effects (it just operates on the symbol mutual information).

In the second case, the MMIB scheme calculates directly the mutual information per coded bit and then averages it through all the transmitted bits. If we define the log-likelihood ratio (*LLR*) for each transmitted bit as

$$LLR\left(b_i\right)=\log_n\left(\frac{P\left(y\middle|b_i=1\right)}{P\left(y\middle|b_i=0\right)}\right), \tag{4.26}$$

and in case having a bit interleaver, it is possible to break up the memory of the modulator, thus $log_2(Z)$ independent channels can be assumed. Therefore, the mutual information of the equivalent channel is described as

$$I\left(b,LLR\right)=\frac{1}{\log_2 Z}\sum_{i=1}^{\log_2 Z} I\left(b_i,LLR\left(b_i\right)\right) \tag{4.27}$$

where $I(b_i,LLR_i)$ is the mutual information between each input bit and it output LLR. Extending these results to a whole coded block, the mean mutual information per bit (*MMIB*) measured across the whole codeword is given by

$$MMIB=\frac{1}{\sum_{f,t}\log_2\left(Z\right)}\sum_{f,t}\sum_{i=1}^{\log_2 Z} I\left(b_i,LLR\left(b_i\right)\right). \tag{4.28}$$

The mutual information function described in Eq. (4.27) can be simply described as a function of the modulation scheme (e.g. Z-QAM, Z-PSK, etc.) and the post-processing SINR (i.e. $I_{\log2(Z)}(ESINR)=I(b,LLR)$). This is actually used in several research works as in [4], [5], [14] or [15]. The following Table 4.2 shows the MIB approximations proposed in [14] to obtain the mutual information for QAM modulated symbols where the following $J(x)$ function is defined as

$$J\left(x\right)=\begin{cases}0.0421x^3+0.2092x^2-0.0064x & \text{if } x\le 1.6363\\ 1-\exp\left(0.0018x^3-0.1427x^2-0.0822x+0.055\right) & \text{if } x>1.6363\end{cases}, \tag{4.29}$$

which is a piece-wise linear function of the complementary error function previously described. It is observed in Table 4.2 that the mutual information of high order constellations is described as a mixture of Gaussian distributions that are non-overlapping at high SNR. The *MMIB* is then obtained as the average mutual information of all the transmitted symbols as noted in Eq. (4.28).

Modulation	Numerical approximations
QPSK	$I_{QPSK}=J\left(2\sqrt{ESINR}\right)$
16QAM	$I_{16QAM}=\frac{1}{2}J\left(0.8\sqrt{ESINR}\right)+\frac{1}{4}J\left(2.17\sqrt{ESINR}\right)+\frac{1}{4}J\left(0.965\sqrt{ESINR}\right)$
64QAM	$I_{64QAM}=\frac{1}{3}J\left(1.47\sqrt{ESINR}\right)+\frac{1}{3}J\left(0.529\sqrt{ESINR}\right)+\frac{1}{3}J\left(0.366\sqrt{ESINR}\right)$

Table 4.2. Numerical approximations for MIB mapping with QAM modulation.

4.2.4. ESM for Multicarrier Spread Spectrum systems

As stated in Section 2.2.6, a large number of schemes that combine the OFDM transmission scheme with spread spectrum techniques have been proposed since the 90s. These schemes have been proposed in order to increase frequency and/or time diversity at the expense on requiring complex multiple access interference control and cancelling algorithms. The most general scheme is the OFCDM (see Section 2.2.6.3) where the information is spread in frequency and time.

The effects of the spreading in the computation of the *ESINR* have been investigated first by Tang et al. in [17] in case of MC-CDMA and the Zero-Forcing equalizer. These results have been later extended by the author to the OFCDM scheme in [18]. It is shown that the $ESINR$ in case of Zero-Forcing equalizer (also referred as *orthogonal restoring combiner*) can be obtained exactly by the harmonic mean of the channel values, whereas for the Maximum Ratio Combiner the $ESINR$ is given by the arithmetic mean of the channel values. Furthermore, the concept of the *effective channel coefficient* is introduced. This term condenses the channel effects in time and frequency directions over the bandwidth and the data time occupancy. The effective channel coefficient may be used for resource allocation purposes (as will be described in Section 5.1.3) in order to decrease the computational cost. Furthermore, the effective channel coefficient indicates also the equivalent channel in case of transmitting the information under a single carrier scheme without using a spreading code.

In order to find the $ESINR$ and the effective channel functions/values, a description of the OFCDM system is presented hereafter.

4.2.4.1. Derivation of the ESINR for OFCDM systems with linear receivers

In case of the OFCDM systems (considered as a potential scheme for 4G broadband communications), the whole system bandwidth is divided into N_c subcarriers, each group of L_f adjacent subcarriers are grouped into a sub-band, thus, the total number of sub-bands is $G_f=N_c/L_f$. The downlink frame is formed by N_s OFCDM symbols and is also divided into $G_t=N_s/L_t$ time transmission intervals (TTI). The OFCDM use two dimensional spreading factors which are L_f and L_t for frequency domain and time respectively (see depicted scheme in Figure 2.25).

The OFCDM frame is segmented into several frequency and time resource units (RU), where up to $L_f \times L_t$ different data symbols can be multiplexed if the code division multiplexing is has an unitary energy frequency, and the spreading sequences in frequency and time domains ($\hat{c}^{(l)} \in \mathbb{C}^{L_f \times 1}$ and $\check{c}^{(l)} \in \mathbb{C}^{L_t \times 1}$ respectively) are orthogonal having $c^{\mathrm{H}}c=1$. The complex symbol matrix $\mathbf{S}_{n,m} \in \mathbb{C}^{L_f \times L_t}$ transmitted into the $\{n,m\}$-th RU (n and m indicate the subchannel index and the TTI index respectively) is then modulated by the OFDM block where the inverse fast Fourier transform (IFFT) and the cyclic prefix (CP) are performed. To ensure that the subcarriers are orthogonal, the minimum used subcarrier spacing is $\Delta_f = 1/T_S$, where T_S means the OFCDM symbol duration.

The result from the RU matching the n-th sub-band and the m-th TTI ($t \in m$-th TTI) yields to the following signal (without the CP)

$$x_{n,m}(t)=\sum_{\chi_1}\sum_{\chi_2}\sqrt{\zeta_{n,m}^{\chi_1,\chi_2}}S_{n,m}^{\chi_1,\chi_2}\sum_{l=0}^{L_f-1}\widehat{c}_l^{(\chi_1)}\sum_{r=0}^{L_t-1}\breve{c}_r^{(\chi_2)}\cos\left(\left(\omega_n+l\Delta f\right)t\right)\cdot\left(p\left(t-rT_S\right)\right) \quad (4.30)$$

where $p(t)$ is a rectangular pulse shaping filter with time duration T_S. We assume that the average power of the symbols $S_{n,m}^{\chi_1,\chi_2}$ is unitary, the amount of power assigned to each symbol is represented by $\zeta\pi_{n,m}^{\chi_1,\chi_2}$, where the indexes χ_1 and χ_2 indicate the spreading code index used in frequency and time domains respectively to spread the data symbol. If a perfect time and frequency synchronization is assumed (as well as a flat channel per subcarrier), the channel for a user k within the n-th sub-band and the m-th TTI is expressed as follows

$$H_{n,m}^{(k)}(l,r)=\alpha_{n,m}^{(k)}(l,r)e^{j\varphi_{n,m}^{(k)}(l,r)} \quad (4.31)$$

where $j=\sqrt{-1}$, l and r denote the subcarrier and TTI indexes respectively. The term $\alpha\in\mathbb{R}$ is the Rayleigh fading gain, and φ is the uniformly distributed phase over the interval (0,2π]. At the receiver side an additive white Gaussian noise (AWGN) with a double-power spectral density (PSD) of $\eta_0/2$ is assumed, which leads to the following signal at the k-th mobile terminal,

$$r_{n,m}^{(k)}(t)=\left(\sum_{\chi_1}\sum_{\chi_2}\sqrt{\zeta_{n,m}^{\chi_1,\chi_2}}S_{n,m}^{\chi_1,\chi_2}\sum_{l=0}^{L_f-1}\widehat{c}_l^{(\chi_1)}\sum_{r=0}^{L_t-1}H_{n,m}^{(k)}(l,r)\breve{c}_r^{(\chi_2)}\cos\left(\left(\omega_n+l\Delta f\right)t\right)p\left(t-rT_S\right)\right)+v(t) \quad (4.32)$$

Now, assuming that the k-th user is being transmitted data into the *n,m-th* RU within the spreading code pair $\{\chi_1, \chi_2\}$, and given the channel equalizer $G_{n,m}^{(k)}(l,r)$, the received signal after the FFT process (demodulator) and removing the cyclic prefix can be written as

$$\hat{S}_{n,m}^{(k)}(l,r)=\sum_{\chi_1}\sum_{\chi_2}\sqrt{\zeta_{n,m}^{\chi_2,\chi_2}}S_{n,m}^{\chi_1,\chi_2}G_{n,m}^{(k)}(l,r)H_{n,m}^{(k)}(l,r)\widehat{c}_l^{(\chi_1)}\breve{c}_r^{(\chi_2)}+G_{n,m}^{(k)}(l,r)v_{n,m}^{(k)}(l,r) \quad (4.33)$$

where $v_{n,m}^{(k)}(l,r)$ is the AWGN contribution to the l-th subcarrier in the r-th OFCDM symbol within the n,m-th RU.

Afterwards, the signal is then multiplied by it corresponding pair of spreading codes $\{\chi_1, \chi_2\}$ and doubly integrated (time and frequency domains) over a RU, having the following demodulated signal of user k

$$\Upsilon_{n,m}^{\chi_1,\chi_2(k)}=D_{n,m}^{\chi_1,\chi_2(k)}+I_{n,m}^{\chi_1,\chi_2(k)}+N_{n,m}^{\chi_1,\chi_2(k)}. \quad (4.34)$$

Here, $D_{n,m}^{\chi1,\ \chi2\ (k)}$ is the desired signal, $I_{n,m}^{\chi1,\ \chi2\ (k)}$ is the interference term generated by the other data multiplexed within the same RU and $N_{n,m}^{\chi1,\ \chi2\ (k)}$ means the noise term. The information term can be detailed as

$$D_{n,m}^{\chi_1,\chi_2(k)}=\sqrt{\zeta_{n,m}^{\chi_1,\chi_2}}\frac{S_{n,m}^{\chi_1,\chi_2}}{L_fL_t}\sum_{l=0}^{L_f-1}\sum_{r=0}^{L_t-1}G_{n,m}^{(k)}(l,r)H_{n,m}^{(k)}(l,r). \quad (4.35)$$

The multiple access interference (MAI) term is due the cross-correlation properties between the spreading code pair $\{\chi_1, \chi_2\}$ and any of the $\{\chi_1', \chi_2\}$ or $\{\chi_1, \chi_2'\}$ pairs that have been used in the transmitter with $\chi_1'\neq\chi_1$ and $\chi_2'\neq\chi_2'$. Regarding the interference term, its detailed expression is defined as

$$I_{n,m}^{\chi_1,\chi_2(k)}=\sum_{\chi_1'\neq\chi_1'}\sum_{\chi_2'\neq\chi_2}\sqrt{\zeta_{n,m}^{\chi_1',\chi_2'}}S_{n,m}^{\chi_1',\chi_2'}\cdot\left(\sum_{l=0}^{L_f-1}\widehat{c}_l^{(\chi_1)*}\widehat{c}_l^{(\chi_1')}\cdot\sum_{r=0}^{L_t-1}G_{n,m}^{(k)}(l,r)H_{n,m}^{(k)}(l,r)\breve{c}_r^{(\chi_2)*}\breve{c}_r^{(\chi_2')}\right). \quad (4.36)$$

The resulting noise term in Eq. (4.34) is given as

$$N_{n,m}^{\chi_1,\chi_2}(k) = \sum_{l=0}^{L_f-1} \widehat{c}_l^{(\chi_1)*} \sum_{r=0}^{L_t-1} \breve{c}_r^{(\chi_2)*} G_{n,m}^{(k)}(l,r) \cdot v_{n,m}^{(k)}(l,r). \tag{4.37}$$

Therefore, we can express the post-processing SINR (i.e. the *ESINR*) for the symbol transmitted within the spreading code pair $\{\chi_1, \chi_2\}$ into the n,m RRU to the k-th user as

$$ESINR_{n,m}^{\chi_1,\chi_2(k)} = \frac{\mathrm{var}\left\{D_{n,m}^{\chi_1,\chi_2(k)}\right\}}{\mathrm{var}\left\{I_{n,m}^{\chi_1,\chi_2(k)}\right\} + \mathrm{var}\left\{N_{n,m}^{\chi_1,\chi_2(k)}\right\}}, \tag{4.38}$$

Furthermore, independently of the combiner, the previous equation can be reformulated as follows

$$ESINR_{n,m}^{\chi_1,\chi_2(k)} = \frac{\zeta_{n,m}^{\chi_1,\chi_2}}{\Delta f \cdot \eta_0} \left|\mathrm{H}_{n,m}^{(k)}\right|^2, \tag{4.39}$$

where the $|\mathrm{H}_{n,m}{}^{(k)}|^2$ term is the effective channel in the $\{n,m\}$ RRU for user k. For some linear receivers, the effective channel can be obtained in a closed form as it is described hereafter.

4.2.4.1.1. Using the Maximum Ratio Combiner

The Maximum Ratio Combiner (MRC) is the optimum diversity combining scheme with respect to the bit error rate when the noise power is uniform distributed over the RRU and no other signals are multiplexed or transmitted into the same resource unit (the code division multiplexing is not used). The MRC weights the different samples of the data according to the complex channel coefficients as given in [19] by

$$G_{n,m}^{(k)}(l,r) = \alpha_{n,m}^{(k)}(l,r) e^{-j\varphi_{n,m}^{(k)}(l,r)}. \tag{4.40}$$

Thus, the MRC combiner gives more emphasis to those samples that have experienced good channel conditions. However, the MRC combiner enhances the interference component in Eq. (4.36). As a result the MRC is suitable to be used when only spreading code is applied otherwise interference cancellation techniques are required.

Then replacing Eq. (4.40) into Eq. (4.38), the *ESINR* for the MRC can be obtained as

$$\begin{aligned} ESINR_{n,m}^{\chi_1,\chi_2(k)} &= \frac{\frac{\zeta_{n,m}^{\chi_1,\chi_2}}{\left(L_f L_t\right)^2}\left(\sum_{l=0}^{L_f-1}\sum_{r=0}^{L_t-1}\left(\alpha_{n,m}^{(k)}(l,r)\right)^2\right)^2}{\frac{1}{\left(L_f L_t\right)^2}\sum_{\chi_1'\neq\chi_1'}\sum_{\chi_2'\neq\chi_2}\zeta_{n,m}^{\chi_1',\chi_2'}\cdot\left(\sum_{l=0}^{L_f-1}\sum_{r=0}^{L_t-1}\left(\alpha_{n,m}^{(k)}(l,r)\right)^2\right)^2 + \frac{\Delta f\cdot\eta_0}{\left(L_f L_t\right)}\sum_{l=0}^{L_f-1}\sum_{r=0}^{L_t-1}\left(\alpha_{n,m}^{(k)}(l,r)\right)^2} = \\ &= \frac{\zeta_{n,m}^{\chi_1,\chi_2}\sum_{l=0}^{L_f-1}\sum_{r=0}^{L_t-1}\left(\alpha_{n,m}^{(k)}(l,r)\right)^2}{\left(\sum_{\chi_1'\neq\chi_1'}\sum_{\chi_2'\neq\chi_2}\zeta_{n,m}^{\chi_1',\chi_2'}\cdot\sum_{l=0}^{L_f-1}\sum_{r=0}^{L_t-1}\left(\alpha_{n,m}^{(k)}(l,r)\right)^2 + L_f L_t \Delta f\cdot\eta_o\right)} \end{aligned} \tag{4.41}$$

Furthermore, in case only one symbol is multiplexed into RU with spreading code pair $\{\chi_1'', \chi_2''\}$, the next expression is found

$$ESINR_{n,m}^{\chi_1'',\chi_2''}(k)=\frac{\zeta_{n,m}^{\chi_1'',\chi_2''}}{\Delta f\cdot\eta_o\cdot L_f L_t}\sum_{l=0}^{L_f-1}\sum_{r=0}^{L_t-1}\left(\alpha_{n,m}^{(k)}(l,r)\right)^2=\frac{\zeta_{n,m}^{\chi_1'',\chi_2''}}{\Delta f\cdot\eta_o}\left|\mathrm{H}_{n,m}^{(k)}\right|^2 \quad (4.42)$$

where the $|\mathrm{H}_{n,m}{}^{(k)}|^2$ term represents the effective channel. Equation (4.42) indicates that, as it was stated before, in case of using the MRC and assuming there is no multiple access interference, the effective channel is the arithmetic mean of the channel values across the RU.

4.2.4.1.2. Using the Orthogonal Restoring Combiner

The Orthogonal Restoring Combiner (ORC), also known as Zero-Forcing (ZF) or channel inversion, can eliminate the multiple access interferences by inverting the channel effects. Thus the equalizer is given as in [19] by

$$G_{n,m}^{(k)}(l,r)=\frac{1}{\alpha_{n,m}^{(k)}(l,r)}e^{-j\varphi_{n,m}^{(k)}(l,r)}. \quad (4.43)$$

Now, substituting Eq. (4.43) into Eq. (4.38), the *ESINR* in case of the ORC can be expressed as

$$\gamma_{n,m}^{\chi_1'',\chi_2''}(k)=\frac{\zeta_{n,m}^{\chi_1'',\chi_2''}}{\frac{\Delta f\cdot\eta_o}{L_f L_t}\sum_{l=0}^{L_f-1}\sum_{r=0}^{L_t-1}\frac{1}{\left(\alpha_{n,m}^{(k)}(l,r)\right)^2}}=\frac{\zeta_{n,m}^{\chi_1'',\chi_2''}}{\Delta f\cdot\eta_o}\frac{L_f L_t}{\sum_{l=0}^{L_f-1}\sum_{r=0}^{L_t-1}\frac{1}{\left(\alpha_{n,m}^{(k)}(l,r)\right)^2}}=\frac{\zeta_{n,m}^{\chi_1'',\chi_2''}}{\Delta f\cdot\eta_o}\left|\mathrm{H}_{n,m}^{(k)}\right|^2, \quad (4.44)$$

where the term $|\mathrm{H}_{n,m}{}^{(k)}|^2$ is the effective channel. It is observed that in case of the ORC combiner, the effective channel can be obtained by the harmonic mean of the channel values across the RU. Furthermore, we can observe in Eq. (4.44) that MAI has been completely eliminated although the noise is increased on those sub-carriers with poor channel conditions.

Dealing with a MC-CDMA scheme (which is a special case of the OFCDM having L_t=1), the effects of the spreading length are depicted in Figure 4.4 and Figure 4.5 where the cumulative density function of the effective channel is analyzed. The channel simulated is the BRAN Channel A model, and the subcarrier are spaced with Δf=56.25 KHz. It can be seen in Figure 4.4 that in case having adjacent sub-carriers per group, the obtained results are quite similar from a spreading factor length of L_f =1 up to L_f =32, where L_f =32 matches to half of the coherence bandwidth. If we focus on the obtained results using the ORC, it is observed that the effective channel gets worse values as the spreading length is increased. On the other hand, using a MRC we can see that the effective channel variance is reduced, hence the maximum possible effective channel is also reduced. Afterwards, if we consider the case where sub-carriers are equidistant, it is shown that for the ORC the effective channel is reduced as the spreading factor is increased whereas for the MRC the variance is reduced. Whether we compare the obtained results from Figure 4.4 and Figure 4.5 having a spreading factor $8\leq L_f \leq 32$, it can be observed that higher gains should be expected when adjacent sub-carriers are grouped jointly with adaptive resource allocation since higher channel variability is experienced which benefits the multiuser diversity.

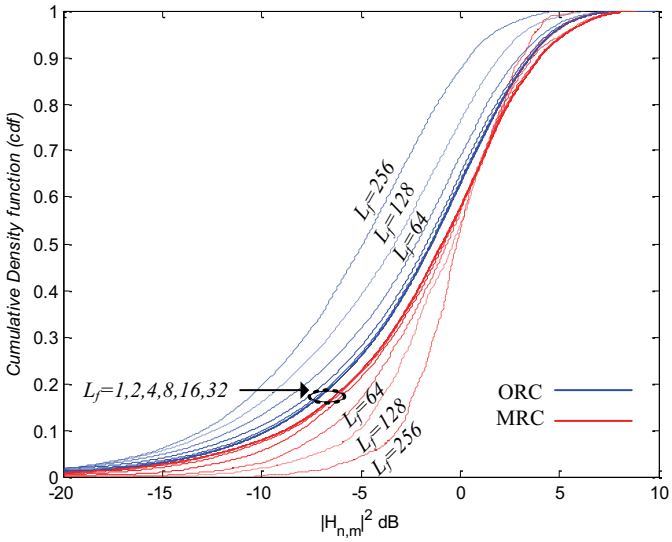


Figure 4.4. Cumulative density function of the effective channel measured over L_f subcarriers and using an MC-CDMA system with spreading factor L_f and adjacent sub-carriers per group with ORC and MRC single user linear detectors.

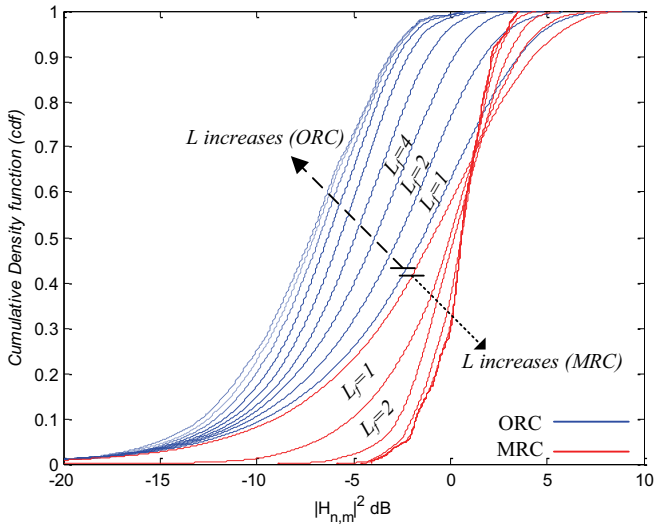


Figure 4.5. Cumulative density function of the effective channel measured over L_f subcarriers and using an MC-CDMA system with spreading factor L_f and equidistant sub-carriers per group with ORC and MRC single user linear detectors.

4.2.5. Empirical parameterized models.

A common way to predict the BLER is to modelling the BLER(SNR) function as a exponential decaying function (this approach is also valid for high SNRs), or as a parameterized complementary error function. Moreover, it has been demonstrated that the BLER is independent of the QAM modulation type, and only the binary coding rate (BCR), and the block length (N_b) need to be considered [14]. Thus, the BLER is defined as a function of the SNR as

$$1)\quad BLER_{BCR,N_b} = \begin{cases} 1 & \text{for } SNR \le b_{BCR,N_b} \\ \exp\left(-a_{BCR,N_b}\left(SNR - b_{BCR,N_b}\right)\right) & \text{for } SNR > b_{BCR,N_b} \end{cases},$$
$$2)\quad BLER_{BCR,N_b} = \frac{1}{2}\left(1 - erf\left(\frac{x - c_{BCR,N_b}}{d_{BCR,N_b}}\right)\right),\ \text{with } d \neq 0. \tag{4.45}$$

The values of a_{BCR}, b_{BCR}, c_{BCR}, d_{BCR} are usually obtained via computer simulation.

4.3. Link adaptation.

In previous sections the author introduced the need of tools that accurately model the behaviour of the channel in terms of bit error rate or block error rate for a specific transmission scheme. Actually, almost all the presented models are based on instantaneous channel knowledge hence the transmission scheme may be adapted to the channel conditions by varying the transmission power, the MCS, or even the assigned subchannels to each user. Then, when the transmitter knows (or can predict) perfectly the channel state, several cost functions can be considered for the system optimization. The main objectives of any wireless system are typically the maximization of the system reliability (i.e. minimization of the BER or the PER), the maximization of the throughput (for a given bandwidth this is equivalent to maximize the spectral efficiency), and/or the minimization of the transmission power (which helps to reduce the interference power and extend/save the batteries life).

As stated above, whether the transmitter needs the full channel state information these knowledge can be acquired via two mechanisms: the channel reciprocity (only in TDD links), or (fast) feedback channels (both for TDD and FDD links). The problematic of the amount of feedbacks and whether it has sense to consider that the transmitter knows perfectly the CSI is tackled at Section 5.3 where a scheme that attains to reduce the feedback channel is also proposed. In the following subsections the different link level adaptation mechanisms are detailed.

4.3.1. Adaptive Modulation and Coding

The adaptive modulation and coding (AMC) technique selects the most appropriate Modulation and Coding Scheme (MCS) depending on the propagation conditions of the communication channel. For example, during good propagation conditions a high order modulation scheme with low coding redundancy is used in order to increase the data rate of the transmission, whereas during a signal fade the system selects a modulation scheme and a coding rate of lower order to maintain both connection quality and link stability without the need of increasing the signal power. The selection criteria for the optimum MCS is defined as follows

$$\begin{aligned} &MCS = \arg\max_{MCS}\left(R\left(MCS\right)\times\left(1 - PER\left(ESINR, MCS\right)\right)\right) \\ &s.t.: PER < \mu \text{ and } P_t \le P_m \end{aligned} \tag{4.46}$$

where P_m is the maximum transmission power permitted. It is then concluded that under the optimization problem defined in Eq. (4.46), the highest MCS (in the sense of spectral efficiency) is transmitted while the QoS and the transmission power are bounded. Most of the

AMC implementations are carried by defining a set of SNR ranges in which the optimum MCS is known a priori. An example of such systems is defined in the IEEE 802.11 standard where depending on the channel state, the MCS is adapted from frame to frame [20]. The following Table 4.3 depicts the minimum SNR values above which each MCS achieves a bit error rate lower than 10^{-4} and 10^{-6} in an AWGN channel (they have been obtained by the author from computer simulation). These BER thresholds are considered reference values for data and multimedia traffic [14].

Min. SINR	BER=10^{-4}	BER=10^{-6}
2PSK,*r*=1/2	0.5	1.75
2PSK, *r*=2/3	2.2	3.3
2PSK, *r*=3/4	3.2	4.5
4PSK,*r*=1/2	3.4	4.8
4PSK, *r*=2/3	5.2	6.4
4PSK, *r*=3/4	6.2	7.6
4PSK, *r*=5/6	7.2	8.6
16QAM,*r*=1/2	9.6	11.2
16QAM,*r*=2/3	11.5	13.2
16QAM,*r*=3/4	12.6	14.1
16QAM,*r*=5/6	13.6	15.3
64QAM,*r*=2/3	17.7	19.2
64QAM,*r*=3/4	18.8	20.2
64QAM,*r*=5/6	20.1	21.8

Table 4.3. Minimum SNR for BER<10^{-4} or 10^{-6} in an AWGN with QAM modulation with convolutional coding using a Soft Decoder.

4.3.2. Power adaptation: The water-filling algorithm

In the AMC strategy, the optimum MCS is selected according to a certain QoS requirement and given a specific SNR per symbol (or subcarrier in OFDM systems). However, we may consider adapting the transmitted power per symbol in order to maximize the system throughput. It is observed in Eq. (4.3) that the channel capacity depends on the signal power spectrum.

We focus on a point to point communication with N independent channels (e.g. N subcarriers), in order to find which optimal power distribution is the most convenient. Then, we define the optimization problem as follows

$$P = \arg\min_{P} \left(-\sum_{i=1}^{N} \log\left(1 + P_i \cdot \frac{|H_i|^2}{W\eta_0} \right) \right), \qquad (4.47)$$

$$s.t.: \quad 1^T \mathbf{P} = \sum_{i=1}^{N} P_i \le P_m, \qquad P_i \ge 0 \quad for \;\; i = 1,...,N,$$

where $\mathbf{P}$ is the power distribution vector, P_m the maximum allowed transmitted power, $|H_i|^2$ is the squared channel transfer function over the i-th channel, W means the channel bandwidth and η_0 is the noise power spectral density.

In order to obtain the optimal power distribution and expressing the channel to noise ratio $C_i=|H_i|^2/(W\eta_0)$, the Lagrange dual function of Eq. (4.47) is given by

$$\mathcal{L}(\mathbf{P},\lambda,\gamma) = -\sum_{i=1}^{N} \log(1 + P_i \cdot C_i) + \lambda \left[\sum_{i=1}^{N} P_i - P_m \right] - \sum_{i=1}^{N} \gamma_i \cdot P_i \qquad (4.48)$$

where λ and γ_i are the Lagrange multipliers. Then, applying the Karush-Kuhn-Tucker (KKT) optimality conditions [21], the optimal power distribution vector $\mathbf{P}^*$ is obtained by

$$\begin{aligned} \frac{\partial \mathcal{L}(P_i,\lambda,\gamma)}{\partial P_i} &= -\frac{C_i}{(1+P_i \cdot C_i)} + \lambda - \gamma_i = 0 \quad for\ i = 1,...,N, \\ P_i^* &\geq 0 \quad for\ i = 1,...,N, \\ \gamma_i^* &\geq 0 \quad for\ i = 1,...,N, \\ \lambda^* \cdot \left[1^T P^* - P_m\right] &= 0, \\ \gamma_i^* \cdot P_i^* &= 0; \end{aligned} \tag{4.49}$$

where λ^* and γ_i^* are the optimal Lagrange multipliers. It can be shown that the solution to the linear equations system defined in Eq. (4.49) is [22]

$$P_i^* = \begin{cases} \frac{1}{\lambda^*} - \frac{1}{C_i} & for \lambda^* \leq C_i, \\ 0 & for \lambda^* > C_i. \end{cases} \tag{4.50}$$

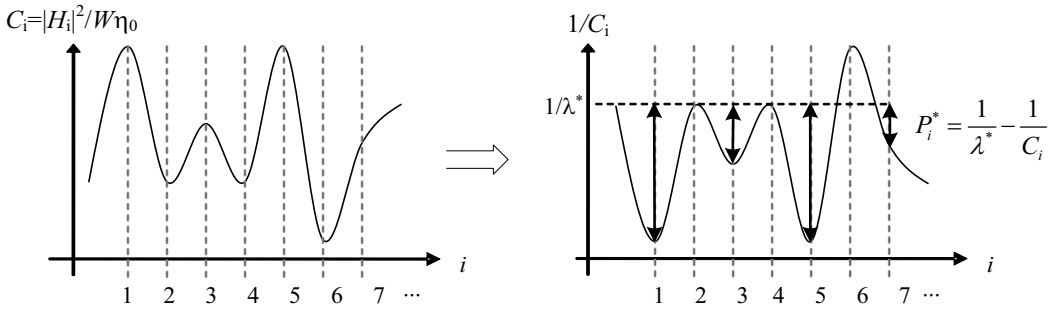


Figure 4.6. Power allocation based on the water-filling principle.

Figure 4.6 describes schematically the behaviour of the power adaptation algorithm which is known as *water-filling* due to its similitude to water-pouring inside an empty bowl. The shape of the bowl is obtained by the inverse of the channel to noise ratio C_i per subcarrier. Those subcarriers where the channel conditions are better will receive more power than those experiencing bad channel conditions. In order to find which is the power allocated to each subcarrier, an initial water-level λ_0[11] is assumed which is iteratively increased until the total power is reached [22]. When all the power is allocated, the optimum water-level λ^* is found which also indicates the power allocated to each sub-carrier. Those subcarriers, where the water level is below the inverse channel to noise ratio, are not assigned any power and in consequence they remain switched off during the transmission.

The solution given in Eq. (4.50) is valid for point to point communications; in case of the broadcast channel (downlink) it is shown that the optimum solution is [22][23]

[11] The initial water-level value proposed in [22] is $\lambda_0 = \frac{P_T}{N} + \min_i\left(\frac{1}{C_i}\right)$.

$$P_i^{(k)*} = \begin{cases} \frac{1}{\lambda^*} - \frac{1}{C_i^{(k)}} & for\ \lambda^* \le C_i^{(k)} \quad and \quad C_i^{(k)} \ge C_i^{(k')}\ \forall k \ne k', \\ 0 & for\ \lambda^* > C_i^{(k)}. \end{cases} \tag{4.51}$$

Therefore, only the best channel to noise ratio per subcarrier (among all the users) is considered. Actually, the equivalent channel that maximizes the channel capacity for the multiple users is given by

$$\left|H_i\right|_{eq}^2 = \max_k \left(\left|H_i^{(k)}\right|^2\right), \quad k = 1,...,K. \tag{4.52}$$

Finally, the power adaptation algorithm might be applied also in multiple access channel (uplink) [22][23][24]. In that case, the power constraint is considered independently for each user, and each user may consider a different maximum transmit power $P_m^{(k)}$. To solve the optimization problem all the powers are normalized, hence the same maximum power $\hat{P}_m$ is considered for all the users, then

$$\hat{P}_m = b^{(k)} \cdot P_m^{(k)}. \tag{4.53}$$

After the power normalization, the optimum power allocation strategy is given by

$$\hat{P}_i^{(k)*} = \begin{cases} \frac{1}{\lambda^*} - b_i^{(k)} \times \frac{1}{C_i^{(k)}} & for\ \lambda^* \le \frac{C_i^{(k)}}{b_i^{(k)}} \quad and \quad \frac{C_i^{(k)}}{b_i^{(k)}} \ge \frac{C_i^{(k')}}{b_i^{(k')}}\ \forall k \ne k', \\ 0 & for\ \lambda^* > C_i^{(k)}. \end{cases} \tag{4.54}$$

Since in this case both the water-level as well as the power normalization factor $b_i(k)$ need to be found, and efficient algorithm is proposed in [22] in order to minimize the number of iterations that get to the optimal solution.

As a conclusion, in either the point-to-point or point-to-multipoint downlink/uplink channels, the optimum power allocation strategy that maximizes the sum capacity is to allocate the resources to the users with the best subcarriers. Such resource allocation strategy leads to the so-called opportunistic scheduling, where the user with the best channel is scheduled [25]. However, this solution may lead to very unfair situations in which users far from the base station will not receive any resources.

4.3.3. Optimal link adaptation: the RA and MA optimization process.

In the previous sections the adaptation of the MCS was introduced as well as the adaptation of the transmitted power per symbol. According to the previous power allocation algorithm, a Gaussian distribution of the symbols is assumed hence when specific MCS are applied (with finite constellations) the capacity is slightly reduced. Algorithms to achieve a throughput close to the channel capacity even with finite constellation have been proposed in [26]-[28] where one of the most extended is the Hughes-Hartogs algorithm [26], in which power is assigned iteratively to the different subcarriers and users in order to increase the constellation order over each iteration. The resource/user that requires less power in each iteration will experiment an increase of its constellation. Other algorithms such as those proposed by J. Campello and Chow et al. in [27] and [28] respectively achieved similar performances with lower computational requirements.

The optimization problem of the power allocation and the AMC has been combined by different authors with more complex QoS requirements such as: the minimum or guaranteed rate (increasing the fairness), the maximum delay, the buffers state, etc. Under such general framework two possible schemes are usually considered. First, in case the bit error rate (or packet error rate) is bounded but the maximum throughput is attained by allocating a certain power to each user with its corresponding MCS. This optimization problem was defined in [25] and is usually referred as *Rate Adaptation* (RA) process. The second optimization process is referred as *Margin Adaptation* (MA) where the optimum power and MCS is defined as that combination that allows the use of the minimum transmission power while at the same time guaranteeing the minimum QoS requirements (i.e. the BER bound).

Then, it is possible to generalize the (multiuser) link adaptation process considering either the RA objective function or the MA objective function. The outcomes of this adaptation process are the resource assigned to each user into the time, the frequency or the space domains as well as the amount of transmitted power per user and the optimum MCS (in this chapter the spatial adaptation is not considered, it will be tackled in the further Chapter 6).

If an OFDM scheme is assumed; the resources can be efficiently assigned to different users without the need of guard band or time gaps (perfect synchronization among mobile stations and the base station is assumed). In fact the whole frequency and time domains are segmented into resource units (RU) that can be assigned arbitrarily to each user. Then, it falls off that the minimum RU is a symbol. The optimization process is illustrated in Figure 4.7. Furthermore, whether spread spectrum or multiple antennas techniques are applied the code and spatial domains can be also exploited. Hence, as depicted in Figure 4.7, the resource allocation algorithm is provided with the channel transfer functions from all the users and the maximum transmission power, and it gives the optimum resource allocation as well as the power and MCS applied to each RU.

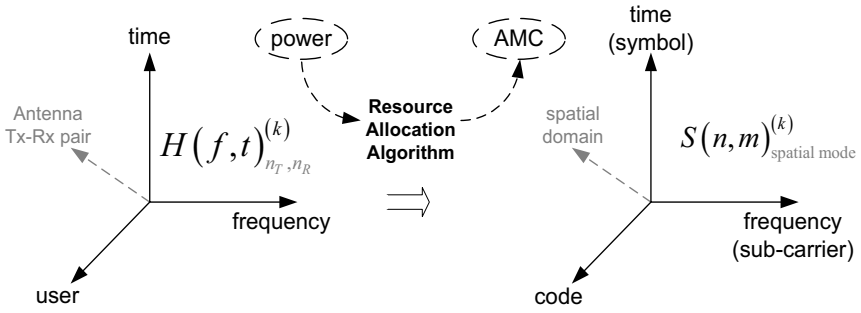


Figure 4.7. Multiuser OFDM-based link adaptation process.

In the downlink case, with K subscribers, the margin adaptation (MA) process is mathematically modelled as follows [29]

$$\min_{P^{(k)}_{n,m}} \sum_{k=1}^{K} \sum_{n=1}^{N_f} \sum_{m=1}^{N_t} P^{(k)}_{n,m}, \quad \text{with } P^{(k)}_{n,m} = \frac{f_1^{(k)}\left(r^{(k)}_{n,m}\right)}{\left|H^{(k)}_{n,m}\right|^2} \tag{4.55}$$

subject to

$$\mathrm{C1}: \quad BER\left(SINR^{(k)}{}_{n,m}, r^{(k)}{}_{n,m}\right) \leq \mu \tag{4.56}$$

where *n,m* index the RU in the frequency and time domain respectively, and *k* means the user index. The function $f_1(\cdot)$ returns the transmission power required to transmit a certain MCS (defined by $r^{(k)}{}_{n,m}$) given the channel conditions ($H^{(k)}{}_{n,m}$). The term $r^{(k)}{}_{n,m}$ then gives the solution to the MA process where the RU assigned to each user are indicated and also how many bits are transmitted from each user in each RU. Since in a pure OFDMA scheme the resources cannot be shared between the users, the following constraint is added to the optimization problem to avoid multiuser interference

$$\mathrm{C2}: \quad \text{if } r^{(k)}{}_{n,m} > 0, \text{ then } r^{(k')}{}_{n,m} = 0 \text{ for } \forall k \neq k'. \tag{4.57}$$

However, some authors have eliminated this constraint by allowing the sharing of the resource either in a time fashion (e.g. TDMA) [29][30] or using spread spectrum techniques (e.g. CDMA) [31].

Furthermore, sometimes the QoS requirements also imply a minimum rate guarantee; this can be included by adding the following constraint to the optimization problem

$$\mathrm{C3}: \quad \sum_{n=1}^{N_f}\sum_{m=1}^{N_t} r^{(k)p,q}_{n,m} \geq R^{(k)} \quad \forall k. \tag{4.58}$$

On the other hand, the link adaptation process can be defined following the RA principle. In that case, the optimization problem is defined as follows [29][30]

$$\max_{r^{(k)}{}_{n,m}} \sum_{n=1}^{N_f}\sum_{m=1}^{N_t} r^{(k)}{}_{n,m} \tag{4.59}$$

subject to

$$\mathrm{C4}: \quad P_m \geq \sum_{k=1}^{K}\sum_{n=1}^{G_f}\sum_{m=1}^{G_t} f_1^{(k)}\left(r^{(k)}{}_{n,m}\right) \tag{4.60}$$

and the $\mathrm{C2}$ and $\mathrm{C3}$ constraints expressed in Eq. (4.57) and Eq. (4.58) respectively. In the following sections, the cited solutions to the RA and MA problems are explained. Furthermore, a scheme proposed by the author to solve the RA problem in case of multiuser OFDMA systems is detailed.

4.3.3.1. Optimum and suboptimum solutions to the RA and MA optimization problems

The number of algorithms and proposals to solve the previous MA and RA problems cited in the scientific literature is quite extensive. However, the resolution of these problems involves usually a Non-Polynomial (NP) hard combinatorial problem, which is not feasible for real-time implementations. Thus, most of the previous works propose sub-optimum solutions, where often are less than 1dB from the optimum solution [29]. These sub-optimum solutions usually split the problem into at least two steps: the RU allocation and the power-bit loading. The main differences are obtained from the RU allocation algorithm since the power and bit loading process is usually performed according to the water-filling algorithm detailed in the previous Section 4.3.2. When the RUs and the power are distributed among users, the optimum power allocation and bit loading algorithm is then performed on user by user mode according to the Greedy algorithm [26]. The Greedy algorithm assigns iteratively the power to

that symbol that requires less power to increment the modulation level. This is done until all the power is assigned.

In fact, the solution for the RA problem can be derived from the MA solution. We can note that the MA problem can be considered as a particular case of the RA, where the power and bit rate constraints are forced to equality. Then the solution to the RA can be obtained performing a last step in the MA problem, where the remaining power is assigned again according to the Greedy algorithm without any constraint since the constraints have been guaranteed and fulfilled by the MA solution.

Then the RU allocation algorithms can be carried by applying one of the following approaches: 1) initial RU assignment and iterative improvement by a RU swapping between users and, 2) resource accounting followed by RU allocation. Both approaches are explained in the following subsection as long with their characteristics.

4.3.3.1.1. Initial RU assignment + iterative improvement

Wong et al. in [29] proposed a two- heuristic steps methods to assign the RUs. One of the main contributions of Wong in [29] is the resolution of Eq. (4.59)-(4.60) where the NP-hard combinatorial problem is modified in order to make it less computational demanding. The requirement that $r^{(k)}_{n,m}$ belongs to a discrete set of MCS is relaxed by fixing $r^{(k)}_{n,m} \in [0, M]$, hence RU sharing is allowed indicated by a sharing factor $\rho^{(k)}_{n,m} \in [0,1]$ over each RU and each *k*-th user. Then, the proposed optimization problem is similar as before, except it is performed over a larger set [29]. Moreover, the RU sharing factor used can be physically interpreted as the time and/or the frequency sharing factor, where in case of having a spread spectrum communications system it can be also accomplished by means of a code assignment into the RU as it was proposed by Trifonov et al. in [31].

Now to solve the RA with the relaxed $r^{(k)}_{n,m}$ values, Wong et al. proposed in [29] an iterative optimization algorithm where the single-user water-filling algorithm is performed iteratively until all the QoS requirements are met. During this iterative procedure, the water-level for each user is incremented with a certain amount of power at each step. Moreover, each time the water-level for one user is modified the sharing factors $\rho^{(k)}_{n,m}$ must be updated. In case of pure OFDMA systems, the solution obtained cannot be directly applied since $r^{(k)}_{n,m}$ cannot be continuous and shared RU is not allowed, hence a final quantization step is carried which may slightly decrease the capacity.

On the other hand, Bakhtiari et al. in [32] proposed an iterative improvement based on RU swapping. In this case after obtaining a starting point solution where all the rate requirements are met, the assigned RUs are swapped between users. Whether the reassignment implies a reduction in the total required power and the user who releases the RU still meets its requirements. All the RUs are then evaluated every round until no more power reduction can be achieved from round to round. This last algorithm step performs better than the Wong's algorithm after 2-3 rounds, and after 4-5 rounds the performance gain in only about 0.1dB lower than the optimum solution.

4.3.3.1.2. Resource Allocation + RU assignment

Yin et al. proposed in [33] to divide the problem into two steps: in the first step a resource allocation process is carried out (i.e. the number of RU and power that should be allocated per user), and in the second step a RU assignment process is performed (which specifies the RU that should be allocated to each user and which MCS is applied to each RU). In the resource allocation step, the required number of RU and power for each user are determined. This step is performed according to the effective SINR for each user. During this process the number of RU per user is iteratively increased until all the required power becomes less than or equal to the maximum transmitted power.

Then according to the number of RU and power assigned to each user, the problem now is to determine the $r^{(k)}_{n,m}$ that maximizes the total data rate (RA) or that minimizes the transmission power (MA). This is equivalent to the classical *Assignment Problem* that can be solved under the Hungarian Algorithm (computational cost $\mathcal{O}(N^4)$ where N is the number of elements) [34] or by other sub-optimum techniques as the RU swapping which reduces the complexity. This approach where the number of resource assigned per user is fixed after the first resource allocation step (although sub-optimum) is much less complex than the iterative in the previous sub-section. Two avoid the complexity burden required by the previous solutions, Assad and Mourad in [16] propose a prioritization mechanism where each user is assigned a priority over each resource unit and the resources are allocated according to such priorities.

The main disadvantage from this approach based on problem partitioning is that since the first step is performed according to the average SNR, the $ESINR$ must be quite independent of the RU position (i.e. high frequency diversity), thus multiuser diversity is minimized.

4.3.3.1.3. Simulation results

Both schemes, the iterative and the two steps of optimization have been compared by the author in [18], where a cross-layer resource allocation scheme has been proposed. The benefits from using cross-layer approach are that; despite of the increase of the computational cost other parameters such as the packets delay or the input buffers status obtained from higher layer (e.g. MAC) are considered into the resource allocation and the scheduling process. This is highly beneficial when the input data streams are sensitive to the delay or follow variable bit rate distributions. In such cases, it is logical to consider that packets that are close to expire require higher priorities than those which are far from the maximum delay. Similarly, when any of the buffers is saturated, it should be prioritized to avoid the packet dropping.This issues will be tacled in detail in the following Chapter 5. Actually, as it is shown by the author in [18], adapting the users requirements within each frame according to the current buffer status improves noticeably the packet delay even more for the *Initial Assignment and Iterative Improvement* schemes reviewed in sub-section 4.3.3.1.1 referred as “IAII” in the Figure 4.8. The system simulation parameters are those depicted in Table 4.4. For variable bit rate streams it is observed that by updating the bit rate requirements according to the instantaneous buffers states (indicated as the cross-layer approach in the plot) the number of simultaneously active users can be increased without affecting the QoS.

MC-CDMA Air Interface and System Level configuration	
Transmission mode	Down-link TDD mode, f_c=5 GHz
Sampling frequency	57.6 MHz
Number of Carriers	N_c = 1024
Cyclic Prefix Time	3.75 μs
Modulation	*M*-QAM, *M*={0,2,4,16,64}
Multipath Channel Model	BRAN channel A
Path Losses model [dB]	46.72 + 39·log10(distance[m])
Downlink Frame Time	0.666 mseg
Spreading Factor (in freq and time.)	L_f=8, L_t=1.
BER target	10^{-3} (*coding is not considered*)
Transmission Power	100mW (20% in pilots and signalling)
Noise power spectral density	-174dBm/Hz
Antenna gains, MS Noise Figure, other losses	2dB, 9dB, 2dB (respectively)
Cell configuration	Square 75m x 75m (*regular office*)
Traffic Type	Low Delay Real Time test service (see 3.1.3)

Table 4.4. MC-CDMA system simulation parameters

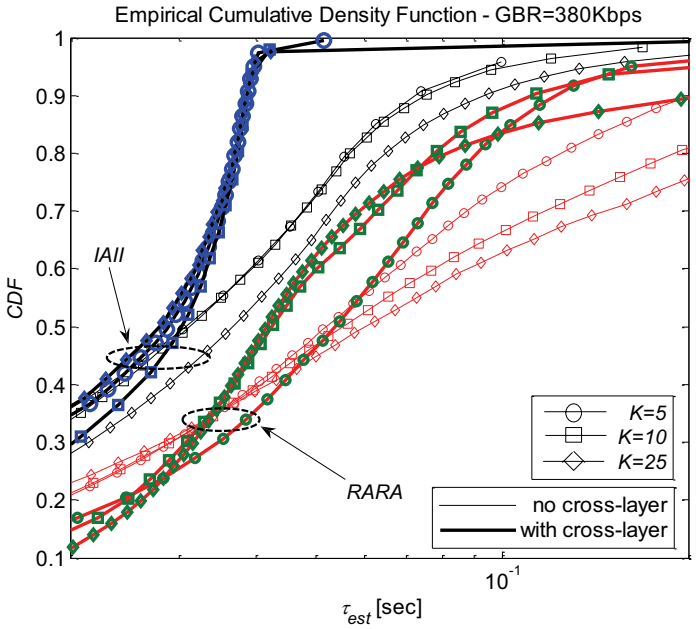


Figure 4.8. Cumulative density function of the delay τ_{est} for the IAII and RARA algorithms with and without a Cross-Layer implementation with K users each requiring a 380Kbps with a maximum packet delay of 50ms.

4.4. Summary of link adaptation techniques for single antenna links

In this chapter the basic concepts to understand the link adaptation techniques have been reviewed. In addition, we have observed that in order to compute the optimum MCS and power per resource unit or the burst, the physical layer abstraction techniques (i.e. the effective SINR concept) could help to reduce the complexity of the resource allocation problem

very useful for the system level evaluation. Finally, the more basic link adaptation techniques (AMC and power loading) have been reviewed and studied in the multiuser context, where the well-known Margin Adaptation and Rate Adaptation optimization problems have been analyzed, giving some new results and comparing the two main solutions for the RA optimization problem. Furthermore, the use of cross-layer techniques has been introduced showing the significant enhancement achieved when both the physical and the medium access layer are designed jointly.

References

[1] ITU-T Recommendation Y.2001 (12/2004) - General overview of NGN, Dec., 2004.

[2] Recommendation ITU-R M.1645, "Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000". June, 2003.

[3] C.E. Shannon, "A mathematical theory of Communication", Bell System Technical Journal, vol.27, pp.379-423, July 1948.

[4] J.G Proakis,"Digital Communications", 4th Ed., Mc.Graw Hill, New York, 2001.

[5] S. T. Chung, A. Goldsmith, "Degrees of Freedom in adaptive Modulation: a Unified View", *The IEEE Transactions on Communications*, vol. 49, pp. 1561-1571, Sept 2001.

[6] W. Lopes, W. Queiroz, F. Madeiro, M. Alencar, "Exact Bit Error Probability of M-QAM modulation Over Flat Rayleigh Fading Channels", Microwave and Optoelectronics Conference, 2007. Oct. 2007.

[7] T. Moon, Error Correcting Coding: Mathematical Methods and Algorithms, John Wiley & Sons, 2005.

[8] IEEE 802.16e-2005, IEEE standard for local and metropolitan area networks, Part 16: Air interface for fixed and mobile broadband wireless access systems, amendment 2, February 2006.

[9] J. Lu, T. Tjhung, C. Chai, "Error probability Performance of *L*-Branch Diversity Reception of MQAM in Rayleigh Fading", IEEE Transactions on Communications, vol. 46, No.2, Feb. 1998.

[10] W. T. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Trans. Commun.*, vol. 43, pp. 2223–2230, Jul. 1995.

[11] S. Nanda, K. M. Rege, "Frame Error Rates for Convolutional Codes on Fading Channels and the Concept of Effective Eb/N0," IEEE Transactions on Vehicular Technology, vol. 47, no. 4, Nov.1998.

[12] M.O. Hasna, M.S. Alouini, "Application of the harmonic mean statistics to the end-to-end performance of transmission systems with relays," *Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE*, vol.2, no., pp. 1310-1314 vol.2, 17-21 Nov. 2002.

[13] R. Sandanalakshmi, T.G. Palanivelu, K. Manivannan, "Effective SNR mapping for link error prediction in OFDM based systems," *IET-UK International Conference on Information and Communication Technology in Electrical* Sciences (ICTES-2007), 2007. pp.684-687, 20-22 Dec. 2007.

[14] IEEE 802.16m-08/004r1, "IEEE 802.16m Evaluation Methodology Document", March, 2008.

[15] K. Brueninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, G.A. Seraji, "Link performance models for system level simulations of broadband radio access systems," *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2005. (PIMRC-2005), vol.4, pp.2306-2311 Vol. 4, 11-14 Sept. 2005.

[16] M. Assaad, A. Mourad, "New Frequency-Time Scheduling Algorithms for 3GPP/LTE-like OFDMA Air Interface in the Downlink," Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE, pp.1964-1969, 11-14 May 2008.

[17] Tang, C.; Stolpman, V.J., "Multiple users adaptive modulation schemes for MC-CDMA," *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE* , vol.6, no., pp. 3823-3827 Vol.6, 29 Nov.-3 Dec. 2004.

[18] I. Gutierrez, F. Bader, J. Pijoan, "Cross-Layer Design for Resource Allocation under QoS Requirements for Multiuser SS-MC-MA Systems", Editors: Thomas Magedanz, Markus Muek, et

al. Book title: Middleware Technologies for Enabling Next Generation Network Services and Applications. Chapter: NGN Networking and Security, pp123-136. Edition Multicon Verlag, 2008.

[19] K. Fazel, S. Kaiser, *Multi-Carrier Spread-Spectrum*, Springer, 2004.

[20] IEEE Std 802.11™-2007, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, New York, USA, 12 June 2007.

[21] S. Boyd, L. Vandenberghe, "Convex Optimization", Cambridge University Press

[22] G. Münz, S.Pfletschinger, J.Speidel, *"An efficient waterfilling algorithm for multiple access OFDM"*, IEEE Global Telecommunications Conference 2002, Taipei (Taiwan), Nov 2002.

[23] C. Zeng, L. Hoo, J. Cioffi, *"Efficient Water-Filling Algorithms for a Gaussian Multiaccess Channel with ISI"*, IEEE Vehicular Technology Conference 2000, p1072-p1077.

[24] J. Jang, K. B. Lee, "Transmit power adaptation for multiuser OFDM systems,," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 171–178, Feb. 2003.

[25] X. Liu, E. K. P. Chong, N. B. Shroff, "A Framework for Opportunistic Scheduling in Wireless Networks," Comp. Net. J., 2002.

[26] D. Hughes–Hartogs. Ensemble Modem Structure for Imperfect Transmission Data. U.S. Patents Nos.4,679,227 (July 1987), 4,731,816 (March 1988) and 4,833,706 (May 1989).

[27] J. Campello. Optimal Discrete Bit Loading for Multicarrier Systems. *Proceedings of IEEE international symposium on information theory*, Cambridge, MA. USA. August, 16-21, 1998.

[28] P.S. Chow, J.M. Cioffi, J.A.C. Bingham. A Practical Discrete Multitone Transceiver Loading Algorithm for Data Transmission over Spectrally Shaped Channels. *IEEE Transactions on Communications.* Vol. 43, No. 2/3/4, Feb.-April, 1995.p. 773-775.

[29] C.Y. Wong, R.S. Cheng, K.B. Letaief, R.D. Murch, Multiuser OFDM with Adaptive Subcarrier, Bit and Power Allocation. *IEEE Journal on Selected Areas in Communications*, vol.17.no.10. Oct.1999. p1747-1757.

[30] W. Rhee, J.M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st , vol.2, no., pp.1085-1089 vol.2, 2000

[31] P. Trifonov, E. Costa, E. Schulz. Adaptive user allocation, bit and power loading in multi-carrier systems. In *Proceedings of the 9th International OFDM-Workshop*, 2004.

[32] E. Bakhtiari, B. Khalaj. A Novel Rate-Adaptation Algorithm for Multiple Access OFDM Systems. 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2004. Vol.4. p.2509-2513.

[33] H. Yin, H. Liu. An efficient multiuser loading algorithm for OFDM-based broadband wireless systems. *IEEE Global Telecommunications Conference, 2000*. Vol.1, Iss., 2000, p:103-107

[34] H. W. Kuhn, The Hungarian method for the assignment problem, Naval Research Logistics Quarterly, pp. 83-97, 1955.

Chapter 5. Cross-Layer Resource Allocation and Scheduling with tight QoS requirements

In the previous chapter the benefits of link adaptation techniques were introduced. However, in order to exploit the channel capacity, the different layers of the protocol stack must be interconnected following a cross-layer or integrated architecture. The benefits of this scheme are unquestionable, however, the analysis and the optimization of the whole system becomes extremely challenging since there are opposite objective functions from the upper to the lower layers. Furthermore, the author observed that many previous cross-layer proposals in the literature do not consider realistic input streams (eg. variable bit rate models, tight delay constraints, packet dropping, etc.), either the complexity associated to the presented solutions in terms of computational cost or required signalling. These aspects are of extremely importance and need to be assessed before stating the benefits of the cross-layer structure. This chapter presents the research carried out by the author in the topic of resource allocation and scheduling based on a cross-layer implementation.

Then, the following issues are addressed throughout this chapter:

- First, the classical scheduling mechanisms in case of multiple users are reviewed, listing some of the best known channel and QoS aware scheduling schemes.
- Two author's proposals for joint resource allocation and scheduling are then introduced, showing by means of system level simulation the benefits of proposed schemes and their performances compared to previous scheduling schemes.
- Finally, two schemes for downlink and uplink for reducing the signalling are presented. Moreover, the combination of adjacent and distributed subcarrier permutations on the same sub-frame is explored, where high benefits in terms of delay and spectral efficiency are observed.

5.1. Cross-Layer Scheduling and Resource Allocation with QoS constraints

A distinctive feature of the cross-layer optimization methodology is the subcarrier, MCS selection, and power allocation at the physical layer jointly with the scheduling at the data link layer. However, it is impossible to achieve the optimality for spectral efficiency, fairness and QoS simultaneously since the increase of one of these characteristics lead to a reduction of the other two. In consequence, a PHY layer optimization (power or bandwidth oriented) resource allocation problem is transformed into a more intricate cross-layer resource allocation problem where the resource allocation implies PHY layer issues such as the power control, the AMC, as well as at the MAC layer issues such as resource allocation and scheduling, and admission control strategies [1][2]. The following Figure 5.1 depicts the schematic system on which the cross-layer resource allocation and scheduling (CL-RAS) are designed, where λ_i means the incoming packet rate of the i-th user.

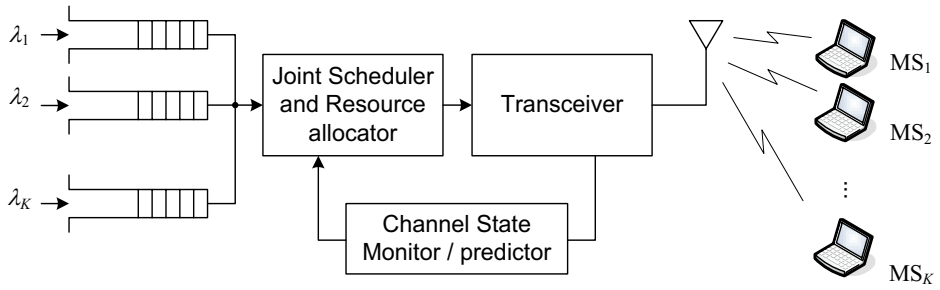


Figure 5.1. Joint Resource Allocation and Scheduling system diagram.

The efficient and fair resource allocation aspects have been well studied in economics analysis, where utility functions have been used to quantify the benefit of the usage of certain resources [3]. The extension of these works to the communication networks field has been proposed to evaluate how a network configuration satisfies specific service requirements of each user's applications, rather than in terms of system-centric quantities like throughput, outage probability, packet drop rate, and power [4]-[6].

Then based on the utility theory, the cross-layer resource allocation and the scheduling function for an OFDMA system can be described as follows

$$\left(MCS_{f,t}^{*(k)},\zeta_{f,t}^{*(k)}\right)=\underset{MCS_{f,t}^{(k)},\zeta_{f,t}^{(k)}}{\arg\max}\left\{\Psi\left(U\left(H_{f,t}^{(k)},\tau^{(k)},Q^{(k)}\right)\right)\right\},k=1,...,K \tag{5.1}$$

where $\tau^{(k)},Q^{(k)}$ indicate the delay and buffer status per user k, whereas $U(\cdot)$ defines the utility-pricing function considered for each user and $\Psi(\cdot)$ collapses the utilities from all the users in the system giving the aggregate utility function of the network. The terms MCS^* and ζ^* indicate the optimum MCS and power assigned to each RU. The utility function per user is usually defined to represent the level of user satisfaction received from the system where different applications/users may have different utility functions according to the current needs of the user [7]. These needs correspond to the previous QoS requirements defined in terms of BER, delay or throughput. However, a more challenging aspect is the definition of the $\Psi(\cdot)$ function in Eq. (5.1). For some applications such as best effort, it is proposed to maximize the sum of the utility functions, hence

$$\Psi\left(U\left(H_{f,t}^{(k)},\tau^{(k)},Q^{(k)}\right)\right)=\sum_{k=1}^{K}U\left(H_{f,t}^{(k)},\tau^{(k)},Q^{(k)}\right). \tag{5.2}$$

On the other hand, when we require fairness among the users, it is shown that the logarithmic sum of the utility function, i.e.

$$\Psi\left(U\left(H_{f,t}^{(k)},\tau^{(k)},Q^{(k)}\right)\right)=\sum_{k=1}^{K}\log\left(U\left(H_{f,t}^{(k)},\tau^{(k)},Q^{(k)}\right)\right), \tag{5.3}$$

provides the required fairness. This is usually referred as the Proportional Fair Scheduling (PFS) function. It is noted that according to Eq. (5.3) if the utility function of a single user is zero, the network utility is minimized ($\Psi\rightarrow$-∞). Then, the PFS maximizes the geometric mean of the users' utility functions. Nevertheless, we may also achieve fairness by maximizing the minimum user utility function, in this case

$$\Psi\left(U\left(H_{f,t}^{(k)},\tau^{(k)},Q^{(k)}\right)\right)=\min_{(k)}U\left(H_{f,t}^{(k)},\tau^{(k)},Q^{(k)}\right). \tag{5.4}$$

Then we can generalize the aggregated utility function as follows

$$\Psi\left(U\left(H_{f,t}^{(k)},\tau^{(k)},Q^{(k)}\right)\right)=\frac{\left(U\left(H_{f,t}^{(k)},\tau^{(k)},Q^{(k)}\right)\right)^{1-\alpha}}{1-\alpha}. \quad (5.5)$$

where in case α=1 it is equivalent to the PFS, for $\alpha\to\infty$ we obtain the max-min criterion, and with α=0 it gives the maximum throughput criterion as described in [9].

However, several aggregate utility functions have been proposed where the combination with specific user utility function lead to some well-known scheduling functions hereafter explained. Furthermore, it has been shown that the solution to the optimization problem in Eq. (5.1) is equivalent to a channel and QoS aware dependant prioritization of the different users [6]-[8].

5.1.1. Classical scheduling functions

When the channel state information, neither the QoS nor the buffers status are unknown at the BS, the usually applied scheduling algorithm is the *Round Robin* (RR) where all the users are scheduled periodically at fixed time intervals. According to the Round Robin algorithm, all the users are warranted with the same number of resources independently of their channel experience and their traffic conditions. Hence, the Round Robin scheduler is a maximum fair scheme the achieved spectral efficiency is very poor since the multiuser diversity is not well exploited.

Afterwards, the simplest channel-aware scheduling mechanism will allocate each resource to the user with better channel conditions (*max-SINR* principle). Then the n-th resource is assigned to the user k that maximizes the priority function $\varphi_n^{(k)}$, hence

$$k=\arg\max \boldsymbol{\varphi}_n^{(k)},$$
$$\text{with}\quad \boldsymbol{\varphi}_n^{(k)}=\frac{P_t}{W\eta_0}\left|H_n^{(k)}\right|^2\propto\varpi_n^{(k)}, \quad (5.6)$$

where $\varpi_n^{(k)}$ is the achievable throughput or rate from the k-th user on the n-th RU and is obtained according the *ESINR* and the available MCSs. Therefore, $\varpi_n^{(k)}$=0 means that the k-th user is experiencing a deep fade in the n-th subchannel. Note that the RU sub-index n may note both the frequency and the time domains (simultaneously) without loss of generality. Then, the *max-SINR* principle scheme is used to maximize the system capacity (higher spectral efficiency) at the expense of sacrificing fairness. Since the resources are allocated to the users experiencing better channel conditions, the scheme is usually referred as *"riding the peaks"* or *opportunistic scheduling* [10].

To improve the fairness, scheduling policies based on Weighted Fair Queuing techniques as those proposed in [9] have been designed for maximizing both the system throughput and fairness among users. One of the best known fair schemes is the *Proportional Fair Scheduler* (PFS) which serves the users under favourable instantaneous radio channel conditions relative to their average ones [11]-[13]. Then, it can provide an appealing trade-off between the cell throughput and the fairness in a time-shared channel for delay tolerant data traffic, thus taking advantage of the temporal variations of the fast fading channel. The PFS could be also considered as the basic scheduler in terms of fairness and throughput and it is applied in the

third generation networks over shared data channels [14]-[16]. For the PFS, Eq. (5.20) is then substituted by

$$\boldsymbol{\varphi}_n^{(k)}(t) = \frac{1}{Th_a^{(k)}(t)} \cdot \frac{\varpi_n^{(k)}}{\varpi_{\max}} \tag{5.7}$$

where the normalization factor ϖ_{max} is the maximum spectral efficiency or the rate achievable by the available MCS. The $Th_a^{(k)}(t)$ is the average throughput obtained by a moving average window with α as the window length and having $Th_i^{(k)}(t)$ as the instantaneous throughput, hence the average throughput is given by

$$Th_a^{(k)}(t) = \frac{1}{\alpha} Th_i^{(k)}(t) + \left(1 - \frac{1}{\alpha}\right) Th_a^{(k)}(t-1), \quad \text{with } Th_i^{(k)}(t) \geq 0. \tag{5.8}$$

Note that the priority assigned to each user depends on the variable t where after each resource allocation process the average throughputs should be updated as well as the users priorities.

Nevertheless, the main drawback of the PFS comes from the fact that it considers full buffers and Constant Bit Rate (CBR) streams (or slow variations in the instantaneous throughput from each user). However, multimedia networks have to deal with different traffic types, often Variable Bit Rate (VBR) streams, and very strict packet delay requirements must be fulfilled for real-time applications.

Then, the simplest scheduling mechanism that takes into account the packets delay is to schedule those packets that have the largest delay. Then, according to the maximum service delay scheduler (also referred as Earliest Deadline First – EDF) the priorities $\varphi_n^{(k)}$ are given as

$$\boldsymbol{\varphi}_n^{(k)} = \max\left(\tau_p^{(k)}\right) \quad \forall p, \tag{5.9}$$

where $\tau_p^{(k)}$ is the delay of the p-th packet. If the queues behave as a First In First Out (FIFO) buffer, the maximum delay can be obtained from the Head-of-Line (HOL) packet (i.e. $max(\tau_p^{(k)}) = \mathcal{W}^{(k)}$).

However, the max-delay scheduler doesn't take into account the channel state information, thus the spectral efficiency that can be achieved is very limited. To solve this aspect, the (Modified-) Largest Weighted Delay First (M-LWDF) scheduler (which combines both the PFS and the max-delay schedulers) has been proposed by Andrews et al. in [17] where the users' priorities are defined by

$$\boldsymbol{\varphi}_n^{(k)}(t) = \frac{\beta^{(k)}}{Th_a^{(k)}(t)} \cdot \frac{\varpi_n^{(k)}}{\varpi_{\max}} \cdot \mathcal{W}^{(k)} \quad , \tag{5.10}$$

the $\beta^{(k)}$ term is designed such that the probability that the HOL delay exceeds the maximum value is lower than a certain value, i.e. $Pr\{\mathcal{W}^{(k)} > \tau_{max}^{(k)}\} \leq \delta^{(k)}$ [17] being $\delta^{(k)}$ the maximum allowable probability of exceeding the maximum tolerable delay $\tau_{max}^{(k)}$, and has the following definition

$$\beta^{(k)} = \frac{-\log\left(\delta^{(k)}\right)}{\tau_{\max}^{(k)}}. \tag{5.11}$$

Shakkotai et al. proposed in [18][19] to use the exponential rule for scheduling, mathematically this is defined as follows

$$\varphi_n^{(k)}(t) = \frac{\beta^{(k)}}{Th_a^{(k)}(t)} \frac{\varpi_n^{(k)}}{\varpi_{\max}} \cdot \exp\left(\frac{\beta^{(k)} W^{(k)}}{1+\bar{W}^{\mu}} \right), \tag{5.12}$$

where $\beta^{(k)}$ is defined as in (5.11). The term $\bar{W}$ means the average head-of-line delays for all the users. Hence, the exponential rule is quite similar to the M-LWDF with the difference that it equalizes the delays when the difference between them is large (overriding the channel state as long as the channel can support a non-zero rate). On the other hand, when the delay difference between the flows is small, the exponential scheduler behaves as the PFS [19].

5.1.2. Cross-layer based scheduling functions for multicarrier based systems

Recent trends in packet scheduling for multicarrier based air interfaces consider channel and QoS aware scheduling designs where the resource allocation and data scheduling are performed together following a cross-layered architecture [20][21]. Srivastava et al. in [20] defined a cross-layer design as any protocol design that violates the reference layered communication architecture with respect to the classical layered architecture. It is also stated that the layered architecture can be violated in the following basic ways: *i)* creation of new interfaces, *ii)* merging of adjacent layers, *iii)* design and coupling without new interfaces and *iv)* vertical calibration across layers. In consequence, a cross-layer resource allocation and scheduling scheme may consider the channel state (PHY level) as well as the input buffers state, e.g. packets delays or the queues lengths (at MAC level).

Liu et al. proposed in [22] a scheduling priority for OFDMA systems similar to that of the M-LWDF with the priority function as

$$\varphi_n^{(k)} = \begin{cases} \beta_{CoS} \cdot \dfrac{\varpi_n^{(k)}}{\varpi_{\max}} \dfrac{1}{F^{(k)}}, & \text{if } F^{(k)} \geq 1 \\ \beta_{CoS}, & \text{otherwise,} \end{cases} \tag{5.13}$$

where $\beta_x \in (0,1]$ is the priority associated to the x service class (i.e. β=1 for *unsolicited grant services*, β=0.8 for *real time services*, β=0.6 for *non-real time services* and β=0.4 for *best effort*), and $F^{(k)}$ means the service satisfaction level associated to each user and is defined as

$$F^{(k)} = \begin{cases} \dfrac{\tau_{\max}^{(k)} - W^{(k)}}{\Delta\tau} & for\ real\ time \\ \dfrac{Th_a^{(k)}(t)}{Th_{\min}^{(k)}} & for\ non\text{-}real\ time \end{cases}, \tag{5.14}$$

and $Th_{min}^{(k)}$ means the minimum rate required by the k-th user, and $\Delta\tau$ a guard time (e.g. $\Delta\tau = 0.25\,\tau_{max}^{(k)}$). These results were extended by Wang et al. in [22] considering the channel state information per subchannel, hence the frequency diversity was exploited by means of the OFDMA scheme. Soo et al. proposed in [24] to prioritize the *real-time* packets according to an *urgency factor* which is the ratio between the packet delay and the maximum delay constraint

$$\upsilon_p^{(k)} = \frac{\tau_p^{(k)}}{\tau_{\max}^{(k)}} \quad \forall p, \tag{5.15}$$

thus any p-th packet with $\upsilon_p^{(k)}>\delta$ (e.g. δ=0.75 in [24]) is marked as a *high priority* packet and scheduled first, the rest of the packets are scheduled following the *max-SINR* principle. Similarly, Ryu et al. in [32] proposed a more complex *urgency factor* for both *real-time* and *non-real-time* applications, hence

$$\varphi_n^{(k)}(t)=\begin{cases}\Pi\left(\dfrac{\tau_p^{(k)}-\tau_{\max}^{(k)}-\Delta\tau}{2\Delta\tau}\right)\cdot\dfrac{\varpi_n^{(k)}}{\varpi_{\max}Th_a^{(k)}(t)}, & \text{if } \tau_p^{(k)}\in \textit{real time},\\ \alpha^{(k)}\exp\left(\alpha^{(k)}\tau_p^{(k)}\right)\cdot\dfrac{\varpi_n^{(k)}}{\varpi_{\max}Th_a^{(k)}(t)}, & \text{if } \tau_p^{(k)}\in \textit{non}-\textit{real time};\end{cases} \tag{5.16}$$

where *real time* applications are scheduled first too, and the remaining resources are used for the *non-real time* applications.

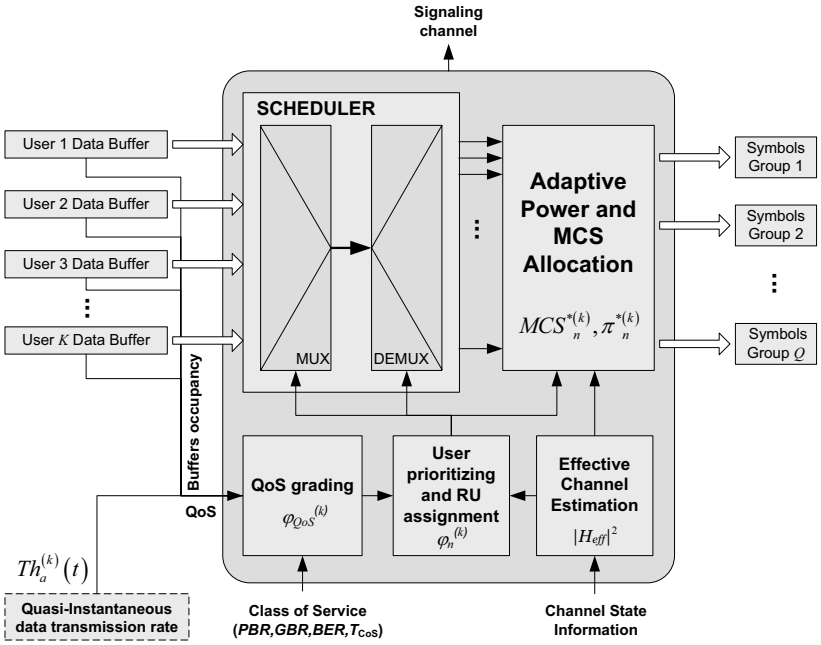


Figure 5.2. Resource Allocation and Scheduling block scheme using a linear combination of CSI and QoS metrics.

5.1.3. Proposed Cross-layer Resource Allocation and Scheduling schemes

The first author's scheme for resource allocation and scheduling is proposed in [25]-[27] which is designed for MC-SS systems despite it could be adapted to any other multicarrier based transmission system. This author proposal is based on a linear combination of the requirements obtained from the QoS metrics and the CSI metrics; hence the priority assigned to each user k in each n-th RU is given by

$$\varphi_n^{(k)}=\varepsilon_{QoS}\cdot\varphi_{QoS}^{(k)}+\varepsilon_{CSI}\cdot\varpi_n^{(k)}, \tag{5.17}$$

where $\varpi_n^{(k)}$ is the already introduced achievable throughput or the rate from the k-th user on the n-th RU which is obtained according the $ESINR$ and the available MCSs (note that in order to obtain the *ESINR* in MC-SS systems the expressions obtained in Eq. (4.42) and Eq. (4.44) should be used for both the MRC and the ZF combiners). The weights ε_{QoS} and ε_{CSI} are used to achieve different tradeoffs between the spectral efficiency and fairness. Hence a scheduler with a total QoS aware has ε_{CSI}=0, and in case of a total channel awareness the scheduler

parameter ε_{QoS} =0. The priority assigned to each user $\varphi_{QoS}^{(k)}$ is obtained based on the previous given priority, and the grade of satisfaction which is measured in terms of delay, hence a satisfied user is that whose packets arrive on time, whereas an unsatisfied user is that whose packets are transmitted after the maximum allowed delay. Mathematically, we can express the priority per user as a function of the QoS as

$$\varphi_{QoS}^{(k)} = \varphi_{QoS}^{(k)} + \mu \cdot sign\left(\frac{L^{(k)}}{Th_a^{(k)}(t)} - \tau_{\max}^{(k)} \right) \tag{5.18}$$

where $L^{(k)}$ is the number of bits queued for the *k*-th user. The term μ is a real constant that fixes the responsiveness of the scheduler to changes within the QoS requirements, hence a high value of μ provides the required QoS faster at the expense of higher jitter (the priority assigned in each iteration may be quite different). As it is observed in Eq. (5.18), the proposed scheme increases the priority if the expected delay is higher than the maximum allowed delay, whereas the priority is decreased. The transmission and the reception block diagram the used system is illustrated in the Figure 5.2.The information from the input buffers and the average data rates (obtained as for the PFS) are introduced into the *QoS grading* block to compute Eq. (5.17). Besides performing the resource allocation and scheduling, the algorithm must also adapted to the MCS, and the power assigned to each resource unit following the Hughes-Hartogs algorithm in [26].

Two downlink transmission schemes, based on the MC-CDMA scheme presented in Section 2.2.6.1, are evaluated by the author in [26] where the impact of the resource allocation is evaluated. The investigated schemes are; the Group Orthogonal (GO-) MC-CDMA, and the Spread Spectrum Multicarrier Multiple Access (SS-MC-MA) scheme. It is shown in Figure 5.3 that the bandwidth is divided into subbands of L_f subcarriers. Following the definition of K. Fazel and S. Kaiser in [19], when the transmitted symbols within each group belong to different users the resulting scheme is named a MC-CDMA with Q-modification, or as a GO-MC-CDMA scheme according to Cai et. al definition in [33]. However, when the symbols within each group belong to the same user, the scheme is usually referred as SS-MC-MA [19]. In both cases the data and the active users are simultaneously multiplexed in the frequency domain (FDMA) and in the code domain (Code Division Multiplexing - CDM). The spread spectrum component is used in order to increase the frequency diversity as well as for minimizing the errors produced by the imperfect channel estimation impact and/or the inter/intra-cell interference [19].

In either the GO-MC-CDMA or the SS-MC-MA scheme the spreading codes are allocated iteratively within the different sub-bands by the *Dynamic Resource Allocation* (DRA) block (see Figure 5.3) according to the previously computed priority value $\varphi_n^{(k)}$defined in Eq. (5.17) . In the GO-MC-CDMA scheme only one code is assigned per user, whereas in the SS-MC-MA scheme each sub-band is exclusively dedicated to only one user. The DRA block determines also the power assigned to each k-th spreading code on the *n*-th RU ($\zeta_n^{(k)}$, see Eq. (4.30)), and the corresponding loaded bits. Therefore, based on the DRA decisions, during each transmitted OFDM symbol each q-th sub-band will have a set of K_g complex assigned symbols (with $K_g \leq L_f$) where the binary data from the selected user(s) is/are mapped according to the assigned modulation scheme(s). Then, the symbol stream $S_q^{(k)}$ of the user k is assigned to the group q

with q={1,...,Q} and k={1,...,K_g} and multiplied by its respective spreading code, the rest of the assigned symbols of each group are summed synchronously. The maximum number of users in the GO-MC-CDMA is equal to K_g, whereas for the SS-MC-MA the number of users in each group is equal to Q=N_c/L_f. Other modifications regarding the MC-CDMA scheme are also possible [19] where the users and the codes are allocated disregarding any predefined structure; hence the codes are allocated without any restriction.

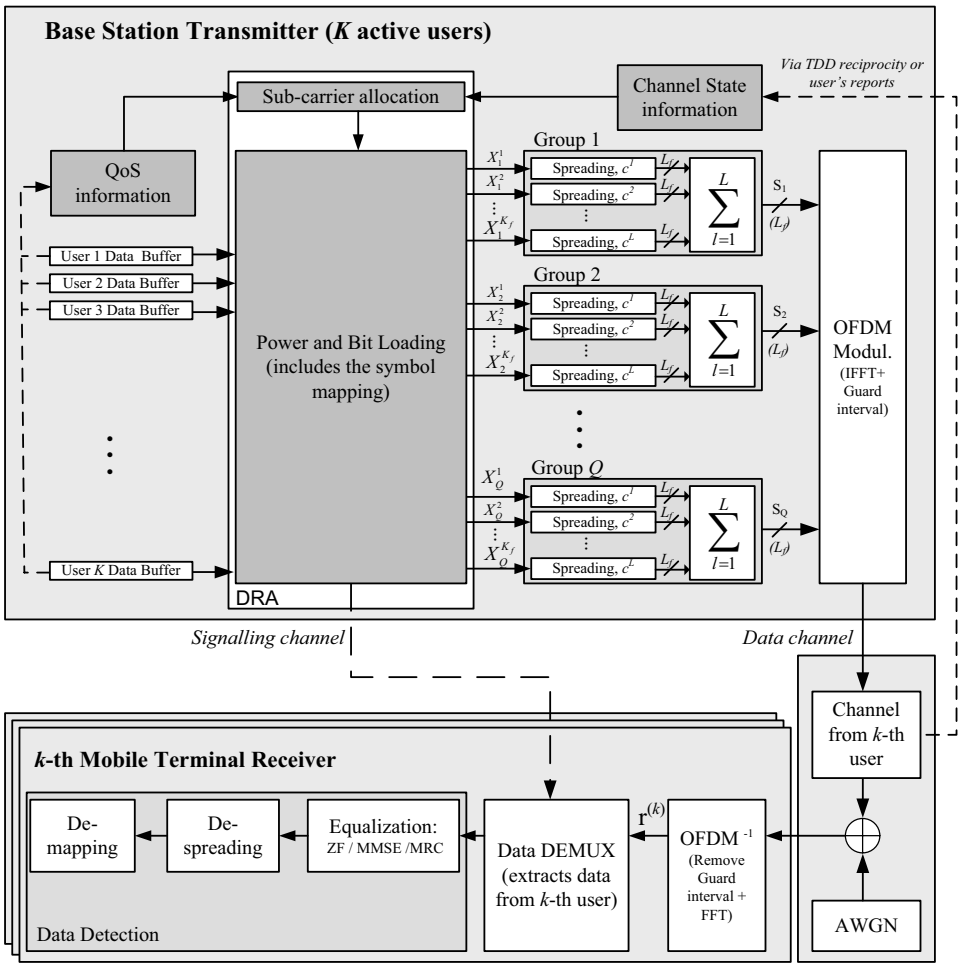


Figure 5.3. Transmission and reception block diagram of the Adaptive SS-MC-MA system.

The performances of both the GO-MC-CDMA and the SS-MC-MA schemes applying the prioritization scheme in Eq. (5.17) and the system parameters depicted in Table 4.4 are shown in Figure 5.4. It is shown that SS-MC-MA is able to achieve much higher spectral efficiencies since each sub-band is allocated for the user with higher channel gain hence users with poor channel conditions do not penalize the system capacity, however it is noted that the difference between GO-MC-CDMA and SS-MC-MA is reduced when the number of users is increased. On the other hand, fairness (measured by the Jain's fairness index in Section 3.3) is greater for the GO-MC-CDMA scheme since although best users are given more resources (i.e. more sub-bands and greater constellations), the bandwidth is necessarily shared among different users, thus avoiding channel monopolization. Moreover, the assigned power as a function of the distance between BS and MS is represented in Figure 5.5. It is shown that for the GO-MC-CDMA scheme more power is assigned for distant users, and thus a greater cell range should be achieved in that case. Moreover, the results obtained from the power distribution as a

function of the distance between the BS and the MSs also reflects the fact that the GO-MC-CDMA distributes more uniformly the resources as was also pointed out in the fairness analysis.

Afterwards, a comparison between the IAII and RARA schemes studied in Section 4.3.3.1 is depicted in Figure 5.6 joint with the performance obtained by the proposed cross-layer prioritization (PRI) algorithm with $\varepsilon_{CSI}=\varepsilon_{QoS}=0.5$. Only the SS-MC-MA scheme is considered in this analysis. The simulations have been carried out under the same parameters used in Section 4.3.3.1 and described in Table 4.4 which follow the values proposed for the IST-MATRICE project [34]. The number of users in the system is K={5,10}. Moreover, the performance of the Proportional Fair Scheduler (PFS) is also shown for comparison purposes where the latency scale (α in Eq. (5.8)) is fixed to 10 frames [11]. As shown in Figure 5.6, the performance of the author's scheme outperforms the other simulated schemes for low number of active users. Furthermore, it has been observed that the PRI algorithm is able to exploit better those instants where the channel is experiencing good propagation conditions leading to higher spectral efficiencies. However in case of higher number of user, the schemes based on prioritization as the PRI and the PFS based on several MAC parameters lead to worse performance than those applied directly at the PHY layer (IAII and RARA) where more accurate controls on the resource allocation process are still possible. This indicates that by adapting the weights ε_{CSI}, ε_{QoS} the system may cope with higher number of active users.

In addition, the trade off between QoS and spectral efficiency achieved by the proposed prioritization algorithm is shown in Figure 5.7 as a function of the ε_{CSI} and ε_{QoS} parameters (having $\varepsilon_{CSI}+\varepsilon_{QoS}$ =1). It is shown that for a low number of active users K, the performance of the system is quite independent from ε_{CSI}, this could be explained by the fact that due there are enough resources (bandwidth and power) to distribute them among all the users, whereas as the number of active users increase best performances are achieved for the following pairs: ε_{CSI}=0.25 with K=10, ε_{CSI}=0.5 with K=25, and ε_{CSI}=0.75 with K=50. Based on the obtained results we can conclude that for a low number of active users ($K\leq10$) it is more beneficial to consider only the *QoS* parameter information at the MAC layer level. However as the number of active users in the system increases, the channel state information becomes more relevant for the resource allocation process since it allows relieving the number of instantaneous service demanding of the active users in a more rapid way.

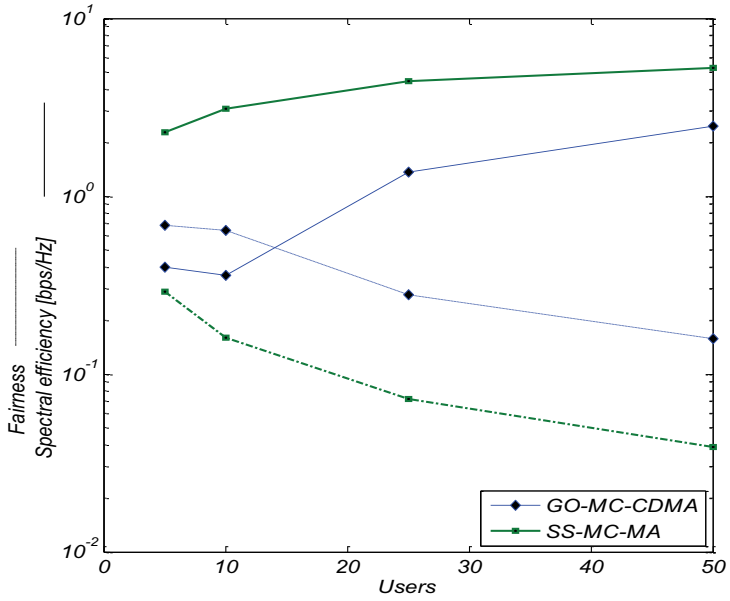


Figure 5.4. Spectral efficiency and fairness of the SS-MC-MA and GO-MC-CDMA schemes using the prioritization scheme in Eq. (5.17).

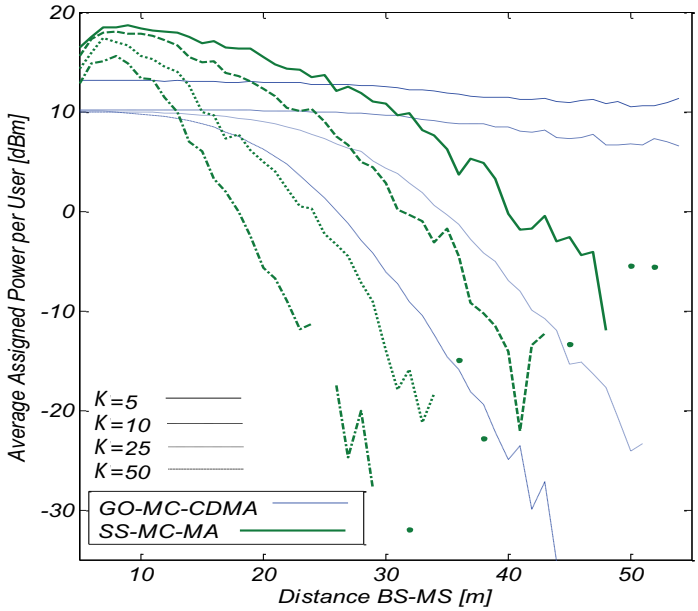


Figure 5.5. Average power assigned per user as a function of the distance BS-MS for both the SS-MC-MA and GO-MC-CDMA schemes using the prioritization scheme in Eq. (5.17).

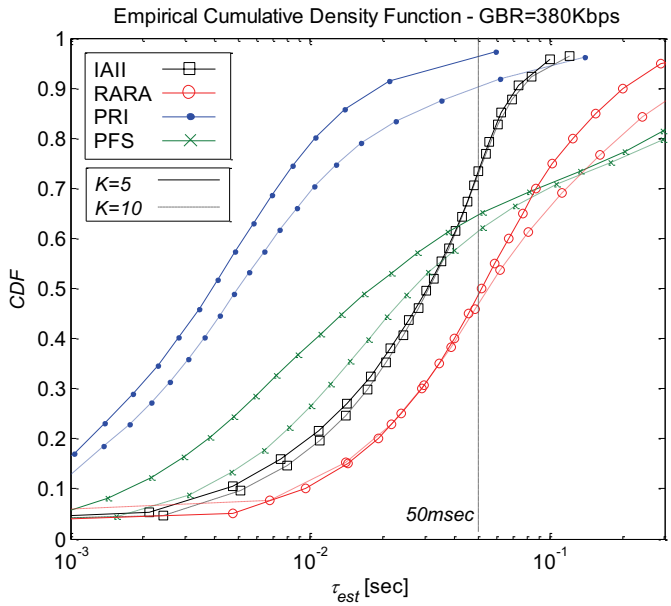


Figure 5.6. Performance comparison using proposed PRI algorithm (with K_{CSI}=0.5) versus the IAII, the RARA and the PFS for low number of active users K={5, 10}.

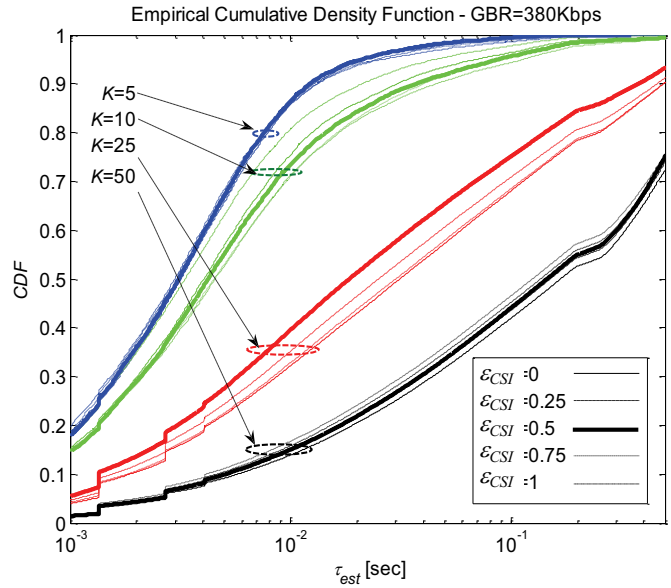


Figure 5.7. Effects of the channel state information parameter (CSI) and the QoS information for different values of α_{CSI} and α_{QoS} over the system performance in terms of the packet delay.

To sum up, in this section it has been shown that, under the proposed prioritization and link adaptation scheme, the SS-MC-MA scheme achieves more than twice the spectral efficiency of the GO-MC-CDMA scheme at the expense of fairness reduction (approx. 1/4 of the Jain's fairness index). On the other hand, the GO-MC-CDMA performs better in terms of cell-range (up to twice the coverage in case of SS-MC-MA at the expense of system capacity, see Figure 5.5). In addition, it has been shown that as the number of users is increased, the channel state information becomes more relevant than the QoS information. Hence, as the system load is increased, the schemes that maximize the spectral efficiency (as the *max-SINR* scheduler) are preferred for other schemes where fairness is maximized. Nevertheless, the small differences observed between the different ε_{CSI} and ε_{QoS} values indicate that the channel as well as the

QoS status effects are very interweaved into the system performance. As a result, a more complex prioritization function is developed by the author in Section 5.1.4 where both channel and QoS metrics are multiplied to obtain the composite priority.

5.1.4. Improved Joint Resource Allocation and Scheduling (JRAS)

As shown in the previous section, two competing aspects exist during the resource allocation and scheduling process, these are *i)* to guarantee the different service flows QoS constraints, and *ii)* to maximize the spectral efficiency. The packet scheduling functions described in Section 5.1 try to maximize both the spectral efficiency and the delay by assigning a priority to each packet or a stream where this priority is a function of the channel state information, and the QoS metrics. Furthermore, according to the research carried by the author in [18] the prevalence of the channel over the QoS or the opposite is difficult to assure and in case of a linear combination of both, the weights should be adapted in accordance with the system load. From system administrator perspective, this approach might be difficult to implement and furthermore it has been shown that when it is the physical layer who decides how many bits are allocated to each user (e.g. following the IAII scheme), the spectral efficiency is further increased. For these reasons, the author in [28] and [29] tackled the problem from a different perspective. Since unavoidably the packets might be fragmented to fit into the physical layer burst. We assume that any packet can be arbitrarily fragmented as many times as necessary (obviously affecting the spectral efficiency due to the fragmentation headers). Then, under this assumption and having each packet delivered within a certain time interval (no matter which class of service it belongs to neither whether it is CBR or VBR) we can obtain the minimum number of bits $b^{(k)}$ that the system should allocate to each user during each frame and is given by

$$b^{(k)} = \begin{cases} T_{frame} \sum_{p=1}^{P} \dfrac{L_p^{(k)}}{\tau_{\max}^{(k)} - \Delta\tau - \tau_p^{(k)}}, \text{ if } \forall p' \rightarrow \ \tau_{p'}^{(k)} < \left(\tau_{\max}^{(k)} - \Delta\tau\right) \\ T_{frame} \sum_{p=1}^{P} \dfrac{L_p^{(k)}}{\tau_{\max}^{(k)} - \Delta\tau - \tau_p^{(k)}} + T_{frame} \sum_{p'} L_{i,p'}, \text{ otherwise} \end{cases} \tag{5.19}$$

where T_{frame} is the frame period in seconds and $L_p^{(k)}$ the number of bits still queued from the p-th packet. Then, if $b^{(k)}$ bits are allocated during each frame to each user, the delay is certainly under its upperbound. Note that if any p' packet has waited more than ($\tau_{max}^{(k)}$-$\Delta\tau$) all the remaining bits from the packet are considered for transmission in the following frame.

Then, the resource allocation and scheduling problem following the *Rate Adaptation* principle presented in Eq.(4.59) can be solved substituting the minimum rate requirements $R^{(k)}$ in (4.58) by the new $b^{(k)}$ requirement in Eq. (5.19). However, since matching these requirements under the power and bit error rate constraints might be unfeasible, the author proposed a further modification in which the number of required bits per frame $b^{(k)}$ is not a tight-constraint. This means that $b^{(k)}$ is only used to determine the priority assigned to each user in each Resource Unit (RU), hence for the k-th user in the n-th subchannel the hereafter allocation priority is defined

$$\boldsymbol{\varphi}_{n\ 1)}^{(k)} = \begin{cases} \min\left(\dfrac{b^{(k)}}{b_{\max}},1\right)\cdot\dfrac{\varpi_n^{(k)}}{\varpi_{\max}}, & \text{if } \forall p' \rightarrow \tau_{p'}^{(k)} < \left(\tau_{\max}^{(k)} - \Delta\tau\right) \\ P_{urgency}\dfrac{\varpi_n^{(k)}}{\varpi_{\max}}, & \text{otherwise} \end{cases} \tag{5.20}$$

where $b_{\max}$ is a normalization factor which is the maximum number of bits that can be transmitted within a frame using the highest MCS. Note that the scheduling priorities are very similar to those proposed for the PFS or the M-LWDF where the normalized attainable channel rate is multiplied by a term that is function of the required QoS. In the PFS this QoS multiplying term was a function of the previous offered QoS whereas in the M-LWDF it was just a function of the head of line delay. We propose to apply here an expression which takes into account the resource required by the user to achieve the expected packet delays. Furthermore, when a packet from the k-th service flow/user is close to exceed its maximum delay, the term $b^{(k)}/b_{\max}$ is substituted by an urgency factor $P_{urgency}$ which is a fixed constant such that $P_{urgency} > (\varpi_{min}/\varpi_{max})^{-1}$. As a result, the packets close to their maximum delay are put ahead in the allocation process in order to avoid the packet drops due to the excessive packet delay. From Eq. (5.19), it is also observed that in case a buffer is empty (i.e. $b^{(k)}$=0), the priority given to that user is zero.

The proposed *Joint Resource Allocation and Scheduling* (JRAS) process is divided into two phase, the first dedicated to allocate the minimum users requirements indicated by $b^{(k)}$, and the second devoted to maximize the spectral efficiency. Then during the first phase the algorithm proceeds by assigning iteratively the RUs to the user with highest priority at each iteration. Once a user has its minimum $b^{(k)}$ bits allocated, the user is extracted from the poll of demanding users (its priority $\varphi_{1)n}^{(k)}$ becomes zero). In case a service flow is not allocated the minimum resources (since all the resources are exhausted or due to the channel conditions) its priority in the next frame will be automatically increased following Eq. (5.20) in order to flush those bits that should have been transmitted in the previous frames. Afterwards when all the users' priorities $\varphi_{1)n}^{(k)}$ are zero (i.e. all the requirements are fulfilled or the channel doesn't allow to transmit) and in case there are still RUs to allocate, the priorities are updated by

$$\boldsymbol{\varphi}_{n\ 2)}^{(k)} = \begin{cases} \dfrac{\varpi_n^{(k)}}{\varpi_{\max}}, & \text{if } b^{(k)} > 0 \\ 0, & \text{otherwise} \end{cases}, \tag{5.21}$$

with

$$b^{(k)} = \sum_p L_p^{(k)}. \tag{5.22}$$

Thus, during this second phase the user with highest carrier to noise ratio is scheduled in each RU hence the spectral efficiency is maximized. As a result, both allocation phases tried to maximize the fairness-efficiency trade-off, by guaranteeing the required resource per frame in the first term, and maximizing the spectral efficiency in the second term.

The end of the JRAS process is achieved when all the RUs have been allocated or when the input buffers have been emptied. As the packets must be received following the correct order, the data from the buffers is extracted from the older packets to the newer packets (as in a

FIFO queue). The delivered packet delay is then measured as the time elapsed between the packet is received for first time at the buffer and when all its bits have been transmitted.

In the following subsection, the author's scheduling function is evaluated by means of computer simulations and compared to the PFS and the priority function proposed by Ling et al. in [22] (referred as PRF in the plots) using the reference bench-mark recommended in [14].

Class of Service	*rtPS*	*nrtPS*	*UGS*	*BE*	
Application	*videocall*	*Streaming*	*voice*	HTTP	FTP
Average bit rate	380Kbps	2Mbps	15Kbps	N/A	N/A
Peak bit rate	2Mbps	10Mbps	15Kbps	2Mbps	10Mbps
Packet rate	10 packets/s	10 packets/s	10 packets/s	Variable	Variable
Max. delay	50ms	300ms	75ms	N/A	N/A
Max. BER	10^{-4}	10^{-4}	10^{-4}	10^{-6}	10^{-6}
Packet dropping ratio	1%	0 – 1%	1%	0	0

Table 5.5. Values for the *rtPS*, the *nrtPS*, the *UGS* ad the *BE* services applied in simulations.

OFDMA Air Interface and System Level configuration	
Carrier Freq, Bandwidth	3.5GHz, 20MHz
Sampling Frequency	22.857Msps
Subcarrier Permutation	Band AMC
FFT length, CP	2048, 12.5%
# of used subcarriers	1728
# of subcarriers per subchannel	18
# of OFDM symbols per RU	3
# of data symbols per RU	48 (efficiency = 8/9)
Modulation	{4,16,64}-QAM
Channel coding	Punctured Convolutional with coding rates1/2, 2/3, 3/4, 5/6
Channel Model	Pedestrian B
MS velocity	10Km/h
Channel estimation and CQI	Ideal
Shadowing standard deviation	5dB
BS Tx power	49dBm
Thermal noise	-174dBm
BS antenna gain and pattern	14dB, Sectorial (70º) (see Appendix A)
MS antenna gain and pattern	0dB, Omnidirectional
Path loss, urban environment	139.57 + 28*log10(R), R in Km.
Other Link budget parameters	BS height = 30m, MS heigh =1.5m, MS Noise Figure = 5dB, Cables = 2dB, Cell Radius = 1000m
Frame duration, T_{frame}	5ms
DL/UL rate	2:1
OFDM symbols in the DL	30

Table 5.6. System simulation parameters

5.1.5. Simulation Results

The simulated system environment and parameters details are depicted in Table 5.6. During each simulation, K users are dropped at different positions following a uniform distribution within the cell area (the maximum cell radius is fixed to 1000m). The positions of the MSs remain fixed during the whole simulation time process, while the speed of the MSs is only used to determine the Doppler effect and the channel coherence time [14]. A simulation time of 50 seconds (10.000 frames) has been chosen to ensure the convergence of the user's service flows and the performance metrics. The full process is repeated with the MSs dropped at new

random locations. The number of simulated drops for each scenario is 25 hence the results become independent of the MS positions. Each user has only one assigned service flow. Five classes of services are considered during the simulations whose main traffics parameters are summarized in Table 5.5 following the traffics models described in Section 3.1. The channel estimation is assumed ideal at the Base Station, and the packet retransmissions not considered. The minimum Effective SINR per MCS with the punctured convolutional coding defined in the IEEE 802.16 standard [8] (constraint length 7 and native code rate 1/2) are depicted in Table 5.7. To obtain the ESINR (which is the SNR required to give the same BER considering an AWGN channel, see Section 4.2 for the details) the channel values inside each subchannel are merged using the geometric mean [12].

MCS	BER=10^{-4}	BER=10^{-6}
2PSK, *r*=1/2	0.5	1.75
2PSK, *r*=2/3	2.2	3.3
2PSK, *r*=3/4	3.2	4.5
2PSK, *r*=5/6	*4.1**	*5.4**
QPSK, *r*=1/2	3.4	4.8
QPSK, *r*=2/3	5.2	6.4
QPSK, *r*=3/4	6.2	7.6
QPSK, *r*=5/6	7.2	8.6
16QAM, *r*=1/2	9.6	11.2
16QAM, *r*=2/3	11.5	13.2
16QAM, *r*=3/4	12.6	14.1
16QAM, *r*=5/6	13.6	15.3
64QAM, *r*=1/2	*15.9**	*17.5**
64QAM, *r*=2/3	17.7	19.2
64QAM, *r*=3/4	18.8	20.2
64QAM, *r*=5/6	20.1	21.8

Table 5.7. Minimum ESINR for Z-QAM modulation and different Convolutional Coding (using a soft-decoder) for a BER=10^{-4} and 10^{-6}.

The obtained performance of the proposed prioritization function is then evaluated and compared to the PFS (see Eq. (5.7) and the Liu's Prioritization Function (PRF) (see Eq. (5.13)). For the PFS and the PRF functions, $b^{(k)}$ is considered as a fixed number equal to the number of buffered bits (see Eq. (5.22)) and a throughput averaged over a window length of 50ms (i.e. T_{frame} = 5ms $\rightarrow$ α=10).

First, the mentioned schedulers (PFS, PRF and the JRAS) are compared in case of *rtPS* traffic. In this case, each user stream requires the real-time test service described in Section 3.1.3.1, with an average bit rate of 380Kbps. For the *rtPS* traffic, when a packet is not transmitted during the maximum delay τ_{max}, the packet is deleted from the queue and discarded. Two parameters are analyzed: the packet delay statistics and the packet loss rate, i.e. the number of discarded packets is divided by the total number of packets (packets discarded plus packets transmitted). Figure 5.8 shows the cumulative density function of the packet delay for 50 and 100 users. We focus first when K=50, as it is shown in Figure 5.8 all the schemes achieve a delay lower than the maximum delay (50ms), in fact the 99th percentile[12] is achieved at 25ms, 25ms and 35ms for the JRAS, the PFS and the PRF respectively. Furthermore, the packet loss rate performance for each scheme is 0 for the JRAS, 1.6·10^{-5} for the PFS and 2.22·10^{-3}·for the PRF. For the K =100

[12] A percentile is the value of a variable below which a certain percent of observations fall. So the n^{th} percentile is the value (or score) below which n percent of the observations may be found.

case, we can observe that the PFS is the algorithm that achieved the lower packet delays whereas the JRAS sends the packets mainly when the urgency factor is applied. During the simulations, the guard time $\Delta\tau$ was fixed equal to $0.2\cdot\tau_{max}^{(k)}$, thus the urgency factor is activated here when $\tau_p^{(k)} > (\tau_{max}^{(k)} - \Delta\tau) = 0.4\text{ms}$. Now, for K =100 the packet loss rate for each scheduling function is 0.0824, 0.1375 and 0.2505 for the JRAS, the PFS and the PRF respectively. Therefore, although most of the packets are sent when they are near to expire with the JRAS scheduler, a lower packet loss rate is achieved.

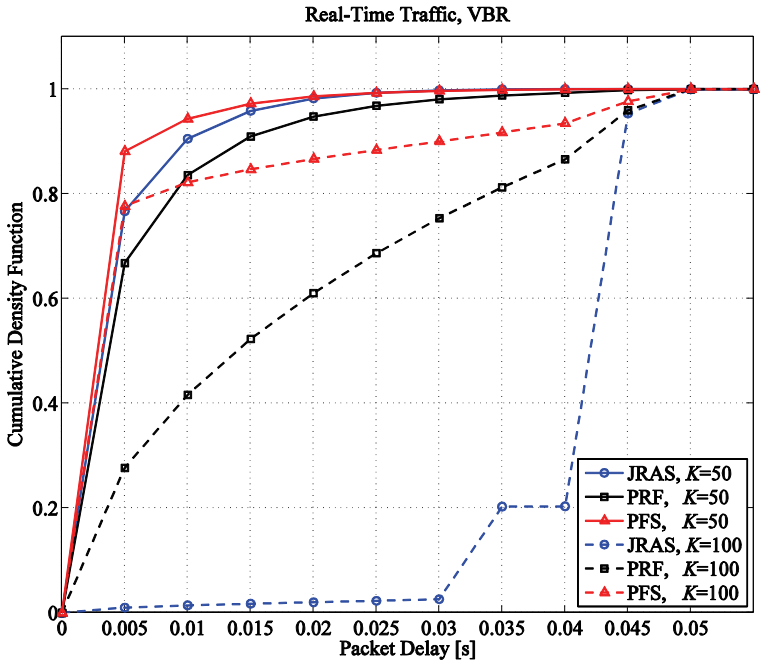


Figure 5.8. Cumulative density functions of the packet delay for the PFS, PRF and JRAS algorithms with *rtPS* traffic with K=50 and 100 users.

In Figure 5.9 the fairness of the proposed scheduling function is evaluated. The simulated traffic is the non-real time test service described in Section 3.1.2. The number of MSs within the cell is K =15. Therefore, taking into account that an average data rate of 2Mbps is demanded per user, an average system throughput of 30Mbps is required. The maximum allowed delay per packet is 300ms, however for the *nrtPS* the packets are not discarded if they exceed this fixed maximum. To evaluate the fairness of the different schedulers, the delays statistics are obtained and analyzed as a function of the distance between the BS and the MS which is sampled at intervals of 100m starting from 50m. It can be observed that the average delay for all the prioritization functions is below the maximum delay. Furthermore, it is also observed that the variation of the average packet delays is lower when the PFS scheduler is used as it is expected. Another important result in Figure 5.9 is the reduced delay obtained using the JRAS scheduler on those users closer to the BS. The reason for that comes from the fact of the two steps design of the resource allocation process, serving first the minimum requirements and later serving those users with higher channel state. On the other hand, when he 99th percentile is obtained at each sampled distance another important conclusion is drawn. Despite a larger variation of the 99th percentile is experienced using the JRAS

prioritization function than using the PFS, it can be observed that at further distances the maximum delay is well upper bounded being able to guarantee a delay lower than 200ms for 99% of the packets. On the other hand, we can observe that the PFS and the PRF cannot guarantee such performance.

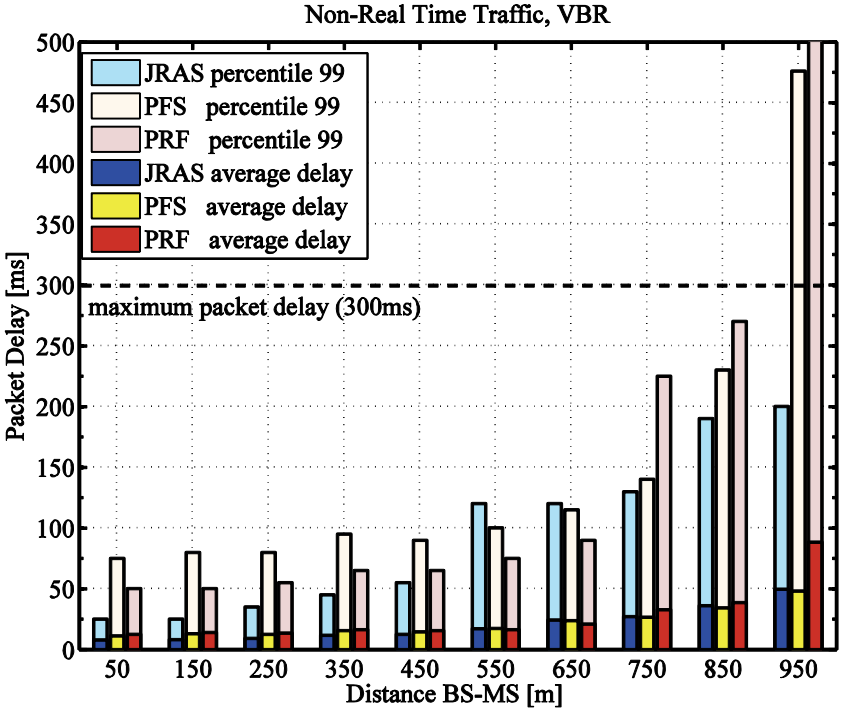


Figure 5.9. Average and 99th percentile of the packet delay for the PFS, PRF and JRAS algorithms with *nrtPS* traffic and K=15 users.

The performance of the proposed JRAS algorithm in case of mixed traffic is shown in Figure 5.10. In this scenario K=50 users are simulated, where 10 users require *non-real time* test service, 13 users require *real-time test service*, 10 users are browsing internet files (World Wide Web service), 5 are downloading files according to the File Transfer Protocol (FTP), and 12 users require UGS connections for applications such as Voice over IP. The average downlink throughput is 26.5Mbps. The maximum delay for each service is indicated in Table 5.5. The delay for the WWW and the FTP services has been assumed as τ_{max}=60s and τ_{max}=90s respectively. It is clearly shown in Figure 5.10 that each traffic type achieves a maximum packet delay lower than its defined maximum value. The 99th percentile for the delay sensitive applications is found at 100ms, 35ms and 20ms for the *nrtPS*, the *rtPS* and the UGS respectively, much lower than its fixed maximum values.

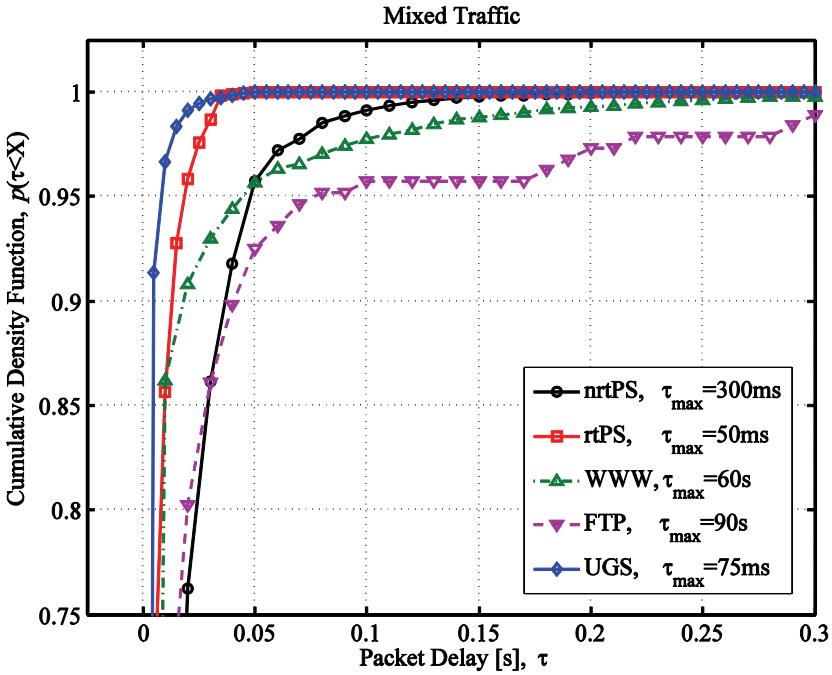


Figure 5.10. Cumulative density functions of the packet delay for the JRAS algorithm with mixed traffic (*rtPS*, *nrtPS*, WWW, FTP and UGS) and K=50 users.

As a result, the proposed Joint Resource Allocation and Scheduling (JRAS) algorithm has shown to guarantee the maximum delay for delay sensitive applications (*rtPS* and *nrtPS*), while at the same time is able to obtain a high spectral efficiency by exploiting multiuser diversity on those unallocated resources. Nevertheless, despite the JRAS initially implies an increase in the computational complexity since it requires more information about the buffers status (i.e. each packet must be time stamped for the JRAS scheduler and we need an independent buffer per service flow), it has been demonstrated its superiority for real and non-real time applications (as well as for mixed traffics). Furthermore, when compared to other prioritization functions, it has shown to outperform them in most scenarios (e.g. half the packet dropping rate or half the 99th percentage delay for the furthest users) at the expense of a slight increase in computational cost. Moreover, for the proposed JRAS scheduling function there is no need to update the priorities each time a RU is allocated, thus the computational complexity is then reduced compared to the PFS and the PRF functions. Nevertheless, the JRAS can easily manage different traffic types by applying different maximum delays to each stream.

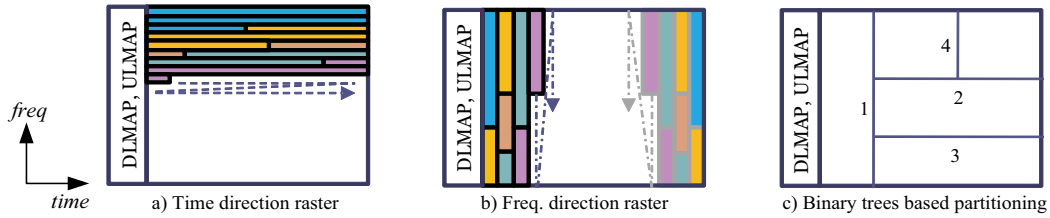
5.2. Contiguous Frequency and Time data grouping based on the JRAS prioritization function

In Section 2.2.3 the benefits and characteristics of the OFDMA scheme have been reviewed. The advantages of OFDMA include the flexibility in subcarrier allocation, the absence of multiuser interference due to subcarrier orthogonality and the simplicity of the receiver among others. The mapping of the subcarriers to subchannels is performed by allocating contiguous or distributed subcarriers, where the first allocation scheme enhances the multiusers diversity, and the second the frequency diversity. Furthermore, following the transmission scheme of the IEEE 802.16a/e standard, the subchannels are grouped into bursts

(also referred as *Data Regions*) following a rectangular structure in frequency and time domains [8]. Then, each burst is mapped to one user (in unicast) or a group of users (in broadcast) and the best MCS is selected for each burst during each frame. This allows the Base Station (BS) to dynamically adjust the bandwidth usage per user according to the users' requirements, i.e. the Quality of Service, and the users' current channel state informations. However, in any of the previously mentioned schemes (including the JRAS) the effects of the resource allocation regarding the required signalling[13] and its payload have been considered neither the need of rectangular shaped bursts for the IEEE 802.16a/e standard.

Several authors have proposed specific resource allocation strategies in order to minimize the number of bursts per frame. In [35], Ben-Shimol et al. proposed to allocate the resources following a "raster approach" such that the resources are allocated first in frequency direction and later in time direction (see Figure 5.11b). Similarly, Erta et al. proposed in [36] to proceed backwards in column-wise order, thus there is no need to book in advance the resources for the transmission within the DL-MAP. Furthermore, the authors in [37] investigated the use of binary trees partitioning in order to decrease the number of burst per frame, hence the time when the receiver is active is minimized which increases the battery life.

However, the research works from [35] to [37] have been conceived considering that the channel within each subchannel is uncorrelated among subcarriers (a subcarrier permutation algorithm is assumed per subchannel), therefore the number of RUs required per user can be determined a priori according to the average SNR (i.e. the MCS is not burst position dependant). Then, although these proposals may achievea good tradeoffs between complexity and spectral efficiency, the gain from frequency scheduling (and multiuser diversity) is minimized since the channel effects have been averaged through all the bandwidth due to subcarrier interleaving.



a) Time direction raster b) Freq. direction raster c) Binary trees based partitioning

Figure 5.11. Examples of burst allocation schemes proposed for OFDMA systems.

A new dynamic radio resource management scheme considering the rectangular burst shape is proposed by the author in [30], which takes into account the position within the frame of the burst. The proposed algorithm can be used indistinctly in case of correlated or uncorrelated channels per subchannel and jointly performs packet scheduling, resource allocation as well as adaptive modulation and coding (AMC) in case a uniform power allocation is assumed. Hereafter, the description and the performance of the proposed scheme is presented.

[13] Each burst is signaled at least by its position in the frame, the number of allocated MRUs in frequency and time, the MCS and the associated service flow or connection identifier (SFID/ CID) [8]. Table 5.8 resumes the fields that are transmitted for each burst.

5.2.1. IEEE 802.16 system model.

We consider the downlink mode in the IEEE 802.16 Point to Multi-Point (PMP) system with one single cell with a total of K MSs within the cell area with free interferences. We focus on the Time Division Duplexing (TDD) scheme, thus channel reciprocity can be assumed between uplink and downlink. The whole TDD frame (depicted in Figure 5.12) is formed by a total of N_s symbols with a T_{frame} duration. The number of downlink and uplink OFDM symbols usually follows the ratio 2:1 or 3:1, however it can be adjusted by the BS according to users' demand [8].

The whole transmission bandwidth W is formed by a total of N_c subcarriers where only N_{used} are active. The active sub-carriers include both the pilot subcarriers and the data subcarriers which will be mapped over different subchannels according to the specific subcarrier permutation scheme described in [8]. For the Full Usage of Subcarriers (FUSC), the pilot subcarriers are allocated first and the remainder are grouped into subchannels where the data subcarriers are mapped. On the other hand, the Partial Usage of Subcarriers (PUSC) and the adjacent subcarrier permutation (usually referred as Band AMC) map all the pilots and data subcarriers within the subchannels, and therefore each subchannel contains its own set of pilot subcarriers. For the FUSC and PUSC, the subcarriers assigned to each subchannel are distant in frequency, whereas for the Band AMC the subcarriers from one subchannel are adjacent. Note that the FUSC and PUSC increase the frequency diversity and average the interference, whereas the Band AMC mapping mode is more convenient for loading and beamforming where the multiuser diversity is increased [22].

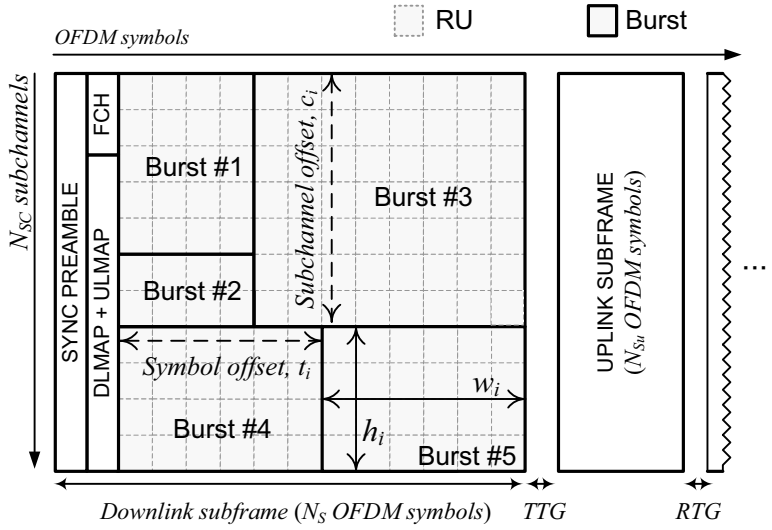


Figure 5.12. IEEE 802.16e OFDMA frame in TDD mode and the different burst structures.

As it is depicted in Figure 5.12, the RUs allocated to any data stream within an OFDMA frame have a two dimensional shape constructed by at least one subchannel and one OFDM symbol. In the IEEE 802.16 standard the specific size of the RU varies according to the permutation scheme, concretely for the Band AMC it may take the shapes 9×6, 18×3 or 27×2 (subcarriers × time symbols respectively), where 1/9 of the subcarriers are dedicated to pilots. We define a RU as a block of N_{sc}×N_{st} symbols in frequency and time domains respectively. Once the size of

the RUs is defined, we can obtain the total number of RUs per frame $Q \times T$, where $Q=N_c/N_{sc}$ is the number of subchannels and $T=N_s/N_{st}$ defines the number of the time slots.

Several RUs may be grouped into a *data region* or *burst* (see Figure 5.12), formed by successive RUs in frequency and in time directions. Both the RU and the data region always follow a rectangular shape structure. Since the MS receiver needs to know how the downlink frame is organized in order to properly decode the data, the downlink control channel includes the number of bursts transmitted as well as the signalling for each burst. In the IEEE 802.16 each burst is signalled by the parameters indicated in Table 5.8. Multicast transmission is addressed by mapping different *Connection Identifiers* (CID) to each burst, where the BS is responsible for issuing the *Service Flow Identifiers* (SFID) and mapping it to single CIDs. These bits signalling are transmitted at the beginning of each frame within the DL-MAP region after the synchronization preamble and the *Frame Control Header* (FCH) [8].

Field	Size [bits]
Number of CIDs	J =8
CIDs (optional)	J·16
MCS	4
OFDMA symbol offset, t_i	8
Subchannel offset, c_i	6
Number of OFDMA symbols, w_i	7
Number of subchannels, h_i	6
Boosting	3

Table 5.8. Signalling required per transmitted burst in the DL-MAP

5.2.2. Rectangular shaped bursts allocation

The main goal of the resource allocation and scheduling mechanisms is to maximize the system throughput (i.e. the spectral efficiency) while guaranteeing the QoS constraints for each user. Nevertheless, one key issue for any resource allocation scheme is to minimize the signalling that is required to inform the receivers how the frame is structured. Following the IEEE 802.16 transmission format, since each burst requires a specific signalling, it is suitable that all the scheduled packets belonging to the same user are transmitted within the minimum number of bursts hence the signalling is minimized.

Thus the optimum shape and position of each burst (with its respective MCS) is explored while the QoS requirements are fulfilled for each user. To reduce the algorithm complexity, the optimization problem formulation considers uniform the power allocation across the subcarriers and that each user is allocated a single burst per frame. According to these premises and considering that there are K active users, the resource allocation and the rate adaptation problem that guarantees the different QoS requirements while maximizing the spectral efficiency can be mathematically expressed as follows

$$\arg\max_{\xi} \left\{ \sum_{k=1}^{K} \sum_{n=1}^{Q} \sum_{l=1}^{T} \varpi^{(k)} \xi^{(k)}(n,l) - K \cdot I_{CC} \right\}, \tag{5.23}$$

subject to

$$C1)\quad R^{(k)}=\sum_{n=1}^{Q}\sum_{l=1}^{T}\varpi^{(k)}\cdot\xi^{(k)}(n,l)\geq b^{(k)},$$
$$C2)\quad \xi^{(k)}(n,l)\cdot\xi^{(k')}(n,l)=0,\ \text{for } k\neq k' \text{ and } n\in[0,Q-1], l\in[0,T-1], \tag{5.24}$$

with

$$\eta^{(k)}\Big|_{BER\leq\mu}=\psi\left(ESINR^{(k)}\right). \tag{5.25}$$

In Eq. (5.23) the term I_{CC} means the number of the required signalling bits transmitted within the control channel for each burst. The minimum required bits per frame $b^{(k)}$ *per user* is obtained by Eq.(5.19) following the scheduling algorithm described in Section 5.1.4. ξ_j is a binary $Q\times T$ matrix which indicates which RUs are allocated to each user (i.e. $\xi^{(k)}(n,l)$=1 means the (n,l) RU has been assigned to the k-th user). In order to force each burst to follows a rectangular shape, the ones in $\xi^{(k)}$ must be placed inside a rectangle. Since each k-th burst must follow a rectangular shape and considering the burst starts at n_k and l_k with h_k and w_k the number of the RUs in frequency and time respectively, the $\xi^{(k)}$ is given by

$$\xi^{(k)}(n,l)=\begin{cases}1, & \text{if } (n_k\leq n\leq n_k+h_k-1) \text{ and } (l_k\leq l\leq l_k+w_k-1)\\ 0, & \text{others}\end{cases}, \tag{5.26}$$

which guarantees that the different bursts do not overlap. Finally, Eq. (5.24) and Eq. (5.25) are used to determine the actual number of bits transmitted for the k-th user $R^{(k)}$. The term ϖ represents the upper layer throughput (in bits) per RU, and it is obtained as a function of the *ESINR* per each burst, the available MCS and the upper bound BER.

5.2.3. Proposed JRAS with rectangular data packing (JRAS-RDP) algorithm

The resolution of Eq.(5.23) to Eq.(5.26) may be performed using non-linear programming techniques. However, such techniques are not feasible for practical systems due to prohibitive computational complexity. Furthermore, the problem as it is defined in (5.23) is very rigid since it forces the number of bursts to be equal to the number of services flows, and in consequence all service flows are scheduled during each frame. However, the number of bursts B should be adapted to the different channel conditions (e.g. a MS transmission experiencing a deep fading should be postponed until the channel is improved). In addition, using a unique burst per user may decrease the spectral efficiency when the burst spans over a large bandwidth due to the effect of frequency selective fading.

To overcome these limitations, the authors propose a low complexity iterative (heuristic) algorithm that adapts the number of bursts for user resource allocation and scheduling purposes. This algorithm uses the priorities per burst defined in Eq. (5.20) and Eq. (5.21) where the resource allocation is divided into two stages: *i)* the minimum requirements fulfilment, and *ii)* the spectral efficiency maximization. Then, given the priority for each user over each subchannel $\varphi_n^{(k)}$ and the minimum bits per frame per user $b^{(k)}$, the RUs are allocated iteratively in order to satisfy the instantaneous rate requirements. Two cases are considered during each iteration; *i)* a new burst might be created, and *ii)* a burst might be increased by allocating another RU (*or a group of*) to the burst. In the second case, when one RU is allocated to an existing burst no extra signalling is required, however, the enlargement of the burst may lead to a reduction on the MCS level.

As it is depicted in Figure 5.12 and Figure 5.13, the bursts equal to one RU are placed in the middle of the frame and may be increased (i.e. new RU are added to the burst) towards four directions, i.e. *top*, *bottom*, *left* and *right* w.r.t its position in the frame (see Figure 5.13) . In order to determine over which direction the increase is more advantageous or suitable, an equivalent priority φ_x ($x \in \{T, B, L, R\}$) is assigned to each direction, where φ_x is obtained by averaging the priority values $\varphi_n^{(k)}$ of the RUs that are covered by the enlarged burst. Whether in the x direction there is any occupied RU or the burst is at the frame boundary then φ_x is forced to "0" meaning that it is not possible to expand the burst on that direction. The burst increasing principle proposed by the authors is also shown in Figure 5.13 where the number inside each rectangle indicates the order in which the resources have been allocated to each burst. In this example, three burst have been created after 15 iterations, where the number indicated inside each MRU indicates the order on which the MRUs have been allocated.

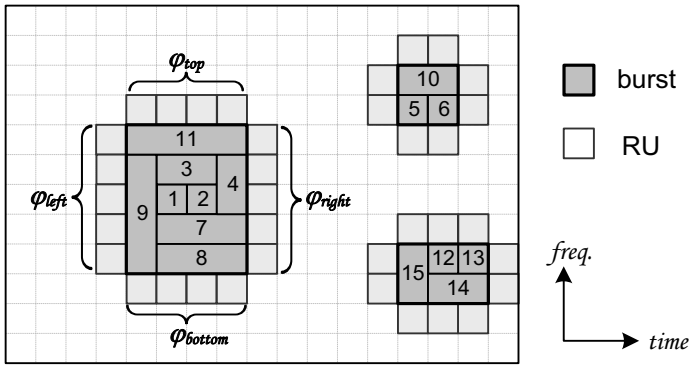


Figure 5.13. Proposed burst increasing options and example of bursts increments after 15 iterations.

The flowchart of the JRAS with rectangular data packing (JRAS-RDP) algorithm is shown in Figure 5.14 and described in Table 5.9. The algorithm is initialized with zero allocated bursts (B=0). For the first burst, the (n,l)-th RU is allocated to the k-th service flow which maximizes the value of $\varphi_n^{(k)}$. The position on the time axis of the RU allocated to the first burst is forced to k=0. Once the first burst is created, the iterative process starts checking the possible increments of the already existing bursts while at the same time it tries to the generate new bursts. Iteratively, the option with highest priority is allocated a new RU (in case of creating a new burst) or a group of RUs (an existing burst is increased). Furthermore, the equivalent priorities associated to each burst increment are multiplied by a P_{burst} factor (e.g. P_{burst}=5) to push forward the enlargement of the existing bursts instead of generating new bursts. In case a new burst is created, since $\varphi_n^{(k)}$ is time independent, its position over the time axis is determined by that time slots whose distance to other already allocated RU is maximized. This fact assures that the new created burst has higher chances to be increased than if it is placed close to the other already existing bursts.

JRAS-RDP algorithm:	
	Start of the algorithm:
1:	Initialize the resource allocation matrices ξ and θ thus all the resource are unallocated and fix the number of allocated burst B=0.
2:	compute the users bit rate requirements $b^{(k)}$ (see Eq. (5.19))
4:	**if** $b^{(k)}>0$ for any k **then**
5:	update the priorities over each n-th RU in order to guarantee the QoS requirements: $\varphi_n^{(k)} = \varphi_n^{(k)}{}_{1)}$
6:	**else**
7:	update the priorities over each n-th RU thus the spectral efficiency is maximized: $\varphi_n^{(k)} = \varphi_n^{(k)}{}_{2)}$
8:	**end if**
9:	**for** b=1:B
10:	obtain the equivalent priority in case the burst is increased as well as the best direction in which it should be increased.
11:	**end for**
11:	obtain which is the k'-th user and n'-th RU pair with the highest priority assigned in case a RU is allocated to a new burst.
12:	assign a RU to the burst with higher equivalent priority obtained in the analyses carried out from lines 9 to 11.
13:	update the resource allocation matrices ξ and θ as well as the number of allocated burst B in case a new burst has been created.
14:	having the modified (new or extended) burst, obtain the ESINR across the burst as well as the highest possible MCS according to the BER constraints of the service flow assigned to the burst.
15:	update the number of transmitted bits per user per frame as well as the bit rate requirements $b^{(k)}$ according to the currently allocated burst.
16:	**if** there are unallocated RU **then**
17:	go to line 2
18:	**end if**
19:	**End of the algorithm.**

Table 5.9. JRAS-RDP algorithm description

The algorithm is then iterated until all the requirements are fulfilled or when all the resources have been allocated. The number of bursts is not fixed and may change from frame to frame depending on the buffers state, the QoS requirements and the channel state conditions. Moreover, since each user may have more than one burst, another auxiliary matrix θ with size ($Q \times T$) is defined. Each value of θ indicates to which burst the RU is allocated. Both matrices ξ and θ are updated each time a new RU is allocated.

The priorities assigned to each RU follows the JRAS scheme presented in Section 5.1.4. Then, when the minimum requirements are satisfied, thus $R^{(k)} \geq b^{(k)}$ for k=1,…,K, and in case there is still some RU pending to be allocated, the JRAS priorities are updated. Thus the users with better channel conditions are assigned a higher priority. In consequence, the end of the joint RRA and packet scheduling process may be achieved due to two factors: *i)* all the RUs have been allocated, or *ii)* the input buffers have been emptied.

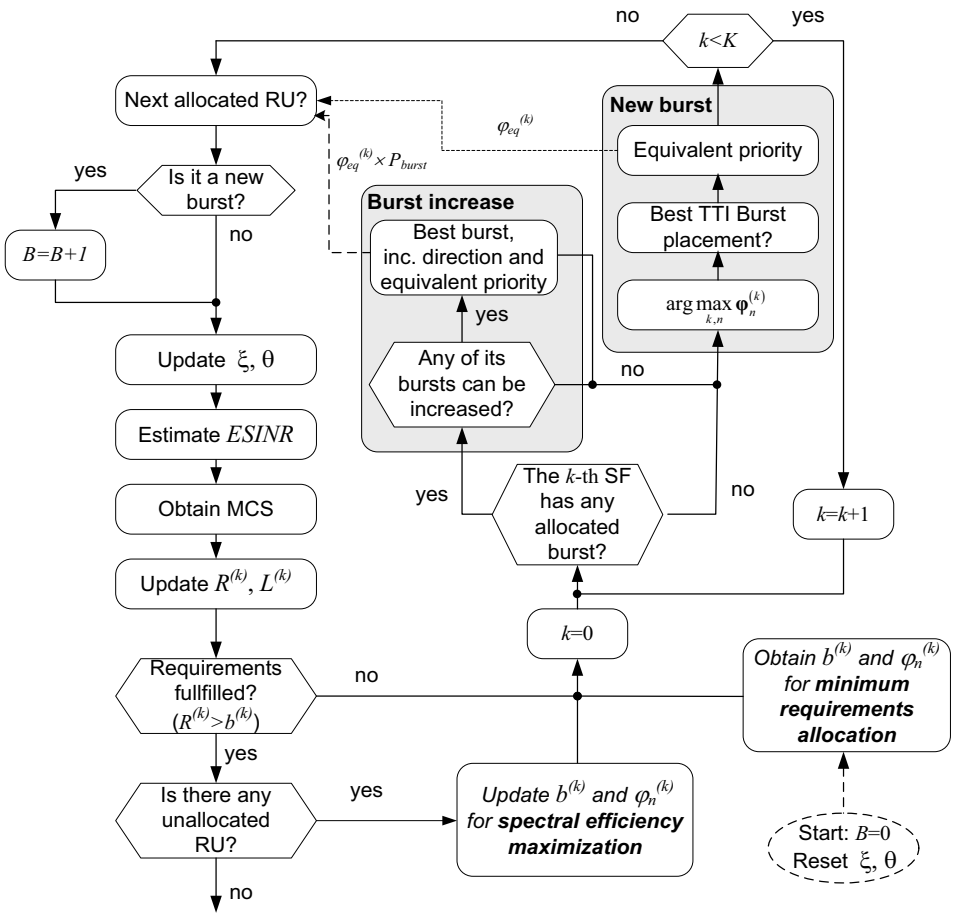


Figure 5.14. Proposed JRAS with rectangular data packing (JRAS-RDP) algorithm.

5.2.4. Simulation results.

The simulated scenario is the same as that explained in Section 5.2.3. Then, the performance of the JRAS-RDP algorithm is evaluated by analyzing the statistics of the number of burst per frame as well as the spectral efficiency and the delivered packets delay as a function of the P_{burst} term which forces the packing of the bursts (higher values of P_{burst} reduces the probability of creating new bursts during the iterative JRAS-RDP algorithm).

The considered scenario is formed by K=15 users, each one requires a *non-real time* test service. The number of bursts per frame is here analyzed as a function of the P_{burst} factor having values of P_{burst}={0, 1, 5, 10, 100}. The prioritization following the JRAS algorithm is here applied. In case P_{burst}=0, each new allocated RU is fitted into a new burst. As a result, with P_{burst}=0 the maximum granularity is achieved but also the signalling is growing up to unaffordable levels (up to $N_{sc} \times N_s$ bursts per frame). It can be observed in Figure 5.15, how for P_{burst}>0 the JRAS-RDP algorithm starts to merge the RUs into bursts. For P_{burst}=1, the same priority is given to the new bursts than for the existing ones. Hence, the number of bursts for P_{burst}=1 is still unaffordable in terms of required signalling. However, for $P_{burst} \geq 5$ it is observed that the number of bursts is lower than 60 for all the simulated frames. In case P_{burst}=5, the achieved number of bursts per frame is lower than 24 in 99% of the transmitted frames, which can be considered as a very good result. Assuming approximately 60 bits are required for signalling for each burst [8] and transmitting the signalling with a QPSK modulation at a code

rate 1/3, the DL-MAP would span less than 2 OFDM symbols which means that the loss due to the downlink signalling is 6.66% for the downlink mode when having a total of 30 OFDM symbols per subframe.

On the other hand, the obtained system throughput using the JRAS-RDP algorithm is plotted in Figure 5.16. Then, it is observed that as P_{burst} increases the throughput decreases. Whether P_{burst}=0, two conclusions are extracted: *i*) most of the frames are sent with a throughput higher than 20Mbps with a significant amount of frames transmitted with instantaneous throughputs higher than 40Mbps, and *ii)* a not negligible amount of frames have been sent nearly empty due to the lack of the buffered bits leading to a low system throughput (peak on the left side of Figure 5.16). Furthermore, when the 99th percentile delay of the delivered packet for each P_{burst} value is computed, the following delay values {160, 185, 250, 275, 725} [ms] have been obtained for P_{burst}={0, 1, 5, 10, 100} respectively.

As a main conclusion, combining the obtained results in figure Figure 5.15 with those depicted in Figure 5.16, it is concluded that P_{burst}=5 achieves the best tradeoff between the granularity (i.e. spectral efficiency), the required signalling, and the QoS requirements. Actually, since the minimum value of $\varpi_n^{(k)}/\varpi_{max}$=0.2, the obtained results indicate that in case of having two users requiring the same date rate, giving a higher priority to the one of them with an already allocated burst achieves the best performance in terms of spectral efficiency and delay. Nevertheless, another observed advantage of the JRAS-RDP algorithm during the simulations is its lower computational time compared when having each RU evaluated in an independent way. Since, in many cases several RUs are allocated in a single iteration, the number of required iterations is reduced as the number of bursts per frame is decreased.

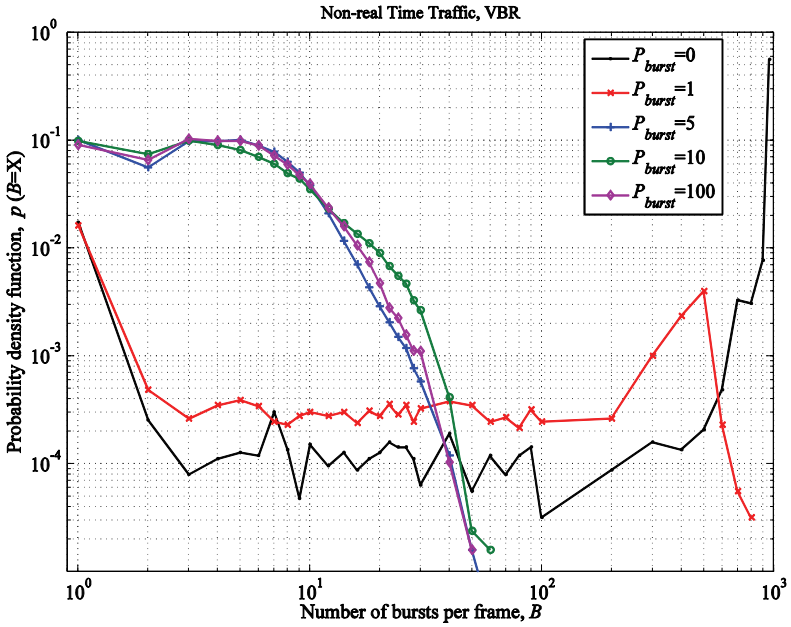


Figure 5.15. Probability density function of the number of bursts per frame using the JRAS-RDP algorithm.

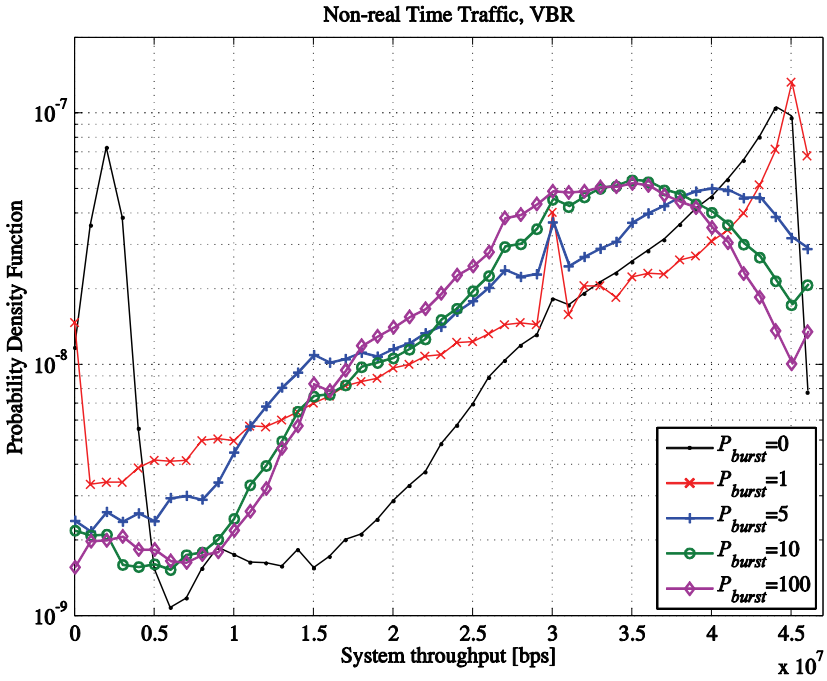


Figure 5.16. Probability density function of the system throughput measured at frame intervals using the JRAS-RDP algorithm.

5.3. Mixed TUSC and Band AMC zone proposal for JRAS with Limited Feedback

It has been previously explained that in the Orthogonal Frequency Division Multiple Access (OFDMA) scheme different permutation schemes have been proposed, mainly the distributed or the contiguous subcarriers per subchannel. Both schemes have been defined in the OFDMA interface of the IEEE 802.16 standard under the names of Partial Usage of the Subchannels (PUSC) and Band Adaptive Modulation and Coding (Band AMC) respectively [8]. Additionally, both schemes have been also considered in the 3GPP-LTE under the names of *distributed* OFDMA (equivalent to the PUSC) and *localized* OFDMA (equivalent to Band AMC) [38].

The Band AMC scheme is preferred in case Channel State Information (CSI) is available at the Transmitter (CSIT) since the optimum power and the MCS can be obtained following the schemes described in section 4.3. Channel reciprocity has been assumed in case of TDD duplexing to obtain the CSI at the BS, however, this implies that every MS must transmit pilots over all the bandwidth during the uplink interval or report the CSI measured by the MS. As a result, the Band AMC scheme implies a large signalling load in the uplink mode to obtain the CSI from all the users, and in the downlink to indicate how the resources have been allocated to each user (i.e. the Downlink-MAP in [8]). In order to reduce the overload signalling two techniques are employed: *i*) the use of limited-rate feedback signals to minimize the uplink feedback requirements, and *ii*) to allocate contiguous sets of Resource Units (RU) leading to a unique *burst* (or *Data Region*) which also preserves the rectangular shape (this aspect has been detailed in the Section 5.2).

In order to reduce the feedback requirements, Han et al. proposed in [39] to send the channel quality metrics (CQM) of the n-best subchannels, as well as sending the CQMs of n-listed

subchannels which are requested by the BS. Moreover, selective feedbacks where only the CQM of those subchannels above a threshold are reported, is also a quite popular technique [40]-[45]. In this case, only the index of the subchannel needs to be sent in the feedback channel. Other researchers have focused on determining which are the minimum CQMs that allow modelling the channel effects and how much they can be compressed [46]. Most of the proposed CQMs are a modified version of the instantaneous ESINR measured at the receiver side, where in almost the cases it can be transmitted over 2-5 bits according to the number of the available MCS.

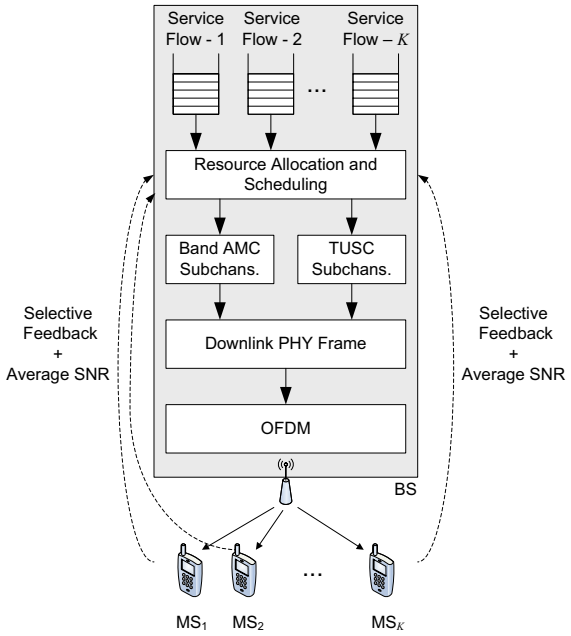


Figure 5.17. Structure of the proposed JRAS-RDP with Low-Rate Feedback algorithm.

The author proposed in [31] to combine these limited feedback techniques with the previous JRAS-RDP algorithm. The scheme applied in this proposal is shown in Figure 5.17. The JRAS-RDP with limited feedback algorithm allocates first those subchannels where selective CQMs are available. In these subchannels, *localized* OFDMA is considered and consequently the best MCS is selected considering the transmission power uniformly distributed in the overall bandwidth. Afterwards, in case there are still subchannels to allocate, the remainder subchannels can be allocated following the already proposed raster approach in [35]. During this second allocation process only the average SNR value is needed, which is feedback periodically from all the Mobile Stations (MS) at a very low rate (e.g. each 15ms). In order to be able to signal the bursts that have been created during the allocation process, an overlapping signalling scheme similar to that proposed in [47] is used, where the signalling order is considered to determine the remaining resources. Moreover, in order to further reduce the feedback requirements, we proposed that only the MS that have packets pending to be scheduled report the CQM of their n-best subchannels, hence the BS announces in the broadcast channel which user have data pending to be delivered.

The assessment of the proposed JRAS-RDP with limited feedback algorithm was carried out through extensive simulations where the author's JRAS-RDP scheme is applied taking

advantage of the proposed mixed *localized* and *distributed* subcarrier permutation zone explained in the following section.

5.3.1. Proposed Mixed TUSC and Band AMC (MTBA) permutation zone

According to the IEEE 802.16e, each downlink or uplink sub-frame is divided into different permutation zones. Each permutation zone is characterized by its specific subcarrier permutation scheme (e.g. PUSC, Band AMC, Full Usage of Subchannels - FUSC, Tile Usage of Subchannels - TUSC, etc.). The transition between zones is indicated in the DL/UL-Map, whereas any burst can span over multiple zones.

As it was previously mentioned, the PUSC scheme (and the Tile Usage of the Subchannels – TUSC - scheme) gives higher frequency diversity, and in consequence they are preferred in case the CSI information is not available at the transmitter. Furthermore, in case of fast moving MSs or unreliable channel estimations, the PUSC and TUSC are able to cope well with those channel uncertainties. On the other hand, when the transmitter knows perfectly the channel from all the users, the Band AMC scheme is preferred since the best MCS can be applied to each subchannel thus maximizing the spectral efficiency. However, having the CSI of all the MS in each subchannel may become unpractical. As a result, the author propose to combine both permutation schemes into a unique zone named Mixed TUSC and Band AMC (MTBA) zone represented in Figure 5.18.

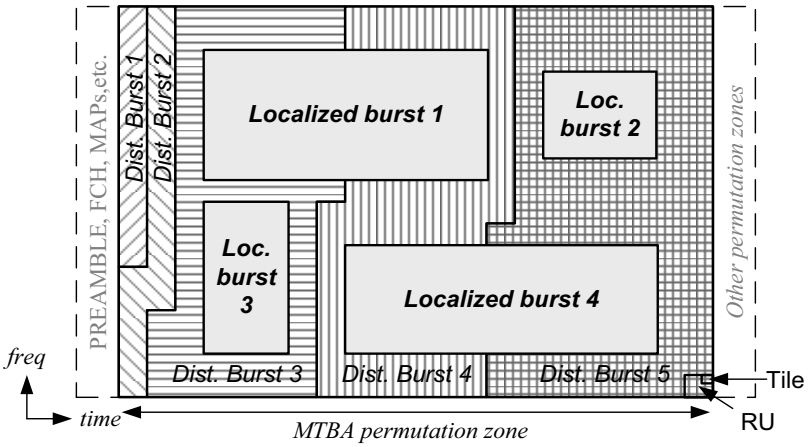


Figure 5.18. Proposed Mixed TUSC and Band AMC (MTBA) zone.

In Figure 5.18, the proposed MTBA zone is constituted by two types of bursts: the *localized bursts* and the *distributed bursts*. For the localized bursts, the RU within the bursts are constructed by N_{sc} adjacent subcarriers during N_{Sd} symbols (e.g. 18×3, 9×6, 27×2, etc.) and in addition, the RUs belonging to the same localized burst are placed in adjacent frequencies and time positions (rectangular data packing). The channel experienced by each MRU is then obtained thanks to the pilot's subcarriers allocated inside each RU.

Having the localized bursts allocated, the remaining subcarriers are divided into tiles. A tile, is a small structure of adjacent subcarriers in frequency and time domains (e.g. 4×3 or 3×3) where the pilot's subcarriers are also allocated. The number of data subcarriers in each tile is fixed and equal to N_{st} (e.g. N_{st}=8). The whole set of tiles are indexed logically thus consecutive tiles are placed at distant positions in frequency and time domain. Afterwards, the RUs for the

distributed bursts are obtained by taking a fixed number N_t of logically consecutive tiles (e.g. N_t =6) which are then spread over the frame in frequency and time directions. Note that, the degree of diversity obtained using the TUSC scheme is slighter lower than that with the PUSC scheme, since in the later the affected channel coefficients over the subcarriers within each RU are all uncorrelated. However, since the PUSC scheme manages independently the pilots and the data subcarriers and the respective mapping to each RU it would be much more difficult to allocate localized and distributed bursts within the same permutation zone if the distributed follow the PUSC permutation scheme. Consequently, we consider that the TUSC scheme offers a good compromise between diversity and computational complexity. Nevertheless, it is worth mentioning that despite the shape of the RUs may change from the localized bursts to the distributed bursts, the amount of data symbols that can be conveyed into any RU is fixed and equal to 48 (i.e. $N_{Sd}\times\mu_p=N_{st}\times N_t\times\mu_p=48$, where μ_p is the pilot efficiency loss). This fact facilitates the resource allocation process since the number of bits that are mapped into each RU can be obtained directly as a function of the channel, the permutation scheme and the available MCS (remember that we assume is uniformly distributed).

Regarding the signalling associated to the proposed MTBA zone, the information that must be transmitted within the DL-MAP can be also reduced for the MTBA zone following a similar scheme to that in [47]. The signalling required for each localized burst remains unchanged compared to the IEEE 802.16e standard. However, for the distributed bursts and assuming that the signalling associated to the localized bursts is sent at the beginning of the DL-MAP and correctly decoded, only the length (i.e. number of RUs) assigned to each distributed burst is required to find its position in the current frame. However, higher redundancy might be necessary to avoid catastrophic error propagation. Thus, if we assume that 10 bits are enough to indicate the burst length, around 30% of the signalling associated to each distributed burst can be reduced using this approach compared to the signalling proposed in the standard without any loss in performance.

5.3.2. JRAS-RDP with limited feedback algorithm

In the previous, it has been stated that the use of localized bursts instead of distributed bursts improves the spectral efficiency of the system since selective fadings can be often avoided. Consequently, localized bursts must be placed first and in case no more localized bursts can be allocated, the remainder resources are assigned to distributed bursts. Furthermore, we are considering that the BS only knows the *n*-best CQM of every user, as well as the average SNR of each user, thus the channel is only known in certain parts of the spectrum.

In order to obtain the SINR thresholds where each MCS gives a BER lower than a certain value, a link level simulation has been carried out where the BER has been computed through several RUs. Two types of channel models have been considered. For the localized bursts the ESINR is obtained hence the SINR thresholds are the same as for the Additive White Gaussian Noise (AWGN) channel (see Table 5.7). On the other hand, for the distributed bursts, only the average SNR is used (as in similar open-loop adaptation mechanisms). We assume that the channel experienced by any RU (hence also by the distributed burst) is an uncorrelated Rayleigh channel. This imply that the channel transfer function $H(n,l)$ can be modelled as a random variable following a zero mean complex Gaussian distribution with unitary standard

deviation. In that case, the optimum MCS is obtained as a function of the arithmetic mean of the SINR values leading to the values in Table 5.10. Those values marked in italic with (x^*) are not used since a MCS with a higher spectral efficiency can be used over the same SNR range. We can observe in Table 5.10 that those MCS where high puncturing rates are applied fail to obtain a low BER for a Rayleigh channel, which indicates that the diversity order is severely affected by the puncturing.

	BER=10^{-4}	BER=10^{-6}
2PSK, *r*=1/2	3.2	5.5
2PSK, *r*=2/3	*6.6**	*9.4**
2PSK, *r*=3/4	*8.8**	*12.2**
2PSK, *r*=5/6	*11.5**	*15.3**
QPSK, *r*=1/2	6.6	9.3
QPSK, *r*=2/3	10.1	13.3
QPSK, *r*=3/4	*12.5**	*16.8**
QPSK, *r*=5/6	*15.4**	*20.3**
16QAM, *r*=1/2	12.4	14.8
16QAM, *r*=2/3	15.8	18.6
16QAM, *r*=3/4	18.4*	21.7*
16QAM, *r*=5/6	21.5*	26.3*
64QAM, *r*=1/2	18.5	21.2
64QAM, *r*=2/3	21.7	24.8
64QAM, *r*=3/4	23.9	27.5
64QAM, *r*=5/6	26.5	30.7

Table 5.10. Average SINR for BER=10^{-4} and 10^{-6} with Z-QAM and Convolutional Coding (using a soft-decoder) in a Rayleigh channel.

JRAS-RDP with limited feedback algorithm:

Start of the algorithm:

1: initialize the resource allocation matrices $\boldsymbol{\xi}$, $\boldsymbol{\theta}$ thus all the resource are unallocated and fix the number of allocated burst B=0.

2: compute the users bit rate requirements $b^{(k)}$ (see Eq. (5.19))

3: divide the OFDMA frame into RU in which contiguous subcarrier permutation is applied.

4: **while** there are unallocated RU where the channel is known **then**

5: apply the JRAS-RDP algorithm

6: **end while**

7: divide the remaining unallocated resource into RU in which distributed subcarrier permutation is applied

8: update the number of transmitted bits per user per frame as well as the bit rate requirements $b^{(k)}$ according to the allocated localized bursts.

9: **if** $b^{(k)}>0$ for any k **then**

10: update the priorities over each n-th RU in order to guarantee the QoS requirements: $\varphi_n^{(k)} = \varphi_n^{(k)}{}_{1)}$

11: **else**

12: update the priorities over each n-th RU thus the spectral efficiency is maximized: $\varphi_n^{(k)} = \varphi_n^{(k)}{}_{2)}$

13: **end if**

14: **if** there are unallocated RUs where the channel is unknown **then**

15: allocate a distributed RU to the user with higher priority $\varphi_n^{(k)}$

16: update the number of transmitted bits per user per frame as well as the bit rate requirements $b^{(k)}$ according to the currently allocated RU.

17: go to line 9

18: **else**

19: go to line 21

20: **end if**

21: **End of the algorithm.**

Table 5.11. JRAS-RDP with low rate feedback algorithm description

In order to allocate the localized bursts, the downlink frame is first segmented into RUs of adjacent subcarriers and symbols, having $Q \times T$ available RU in frequency and time domain respectively. Next, the BS calculates the priorities associated to each user on those known subchannels as following the JRAS prioritization scheme. For the subchannels where the channel is not known an available rate $\varpi_n^{(k)}$=0 is fixed (i.e. $\varphi_n^{(k)}$=0). Then the RUs are iteratively allocated according to users' priority $\varphi_n^{(k)}{}_{1)}$ until every service flow is allocated at least $R^{(k)}$ bits or any other localized burst can be created/increased (whatever happens first). Iteratively, the user whose requirements have not been met yet claim for an unallocated RU such that its priority is the highest. A schematic of the proposed JRAS-RDP with limited feedback algorithm is depicted in Figure 5.19 and described in Table 5.11. Furthermore, the RU are allocated to burst following the rectangular shapes required in the IEEE 802.16 standard (see Section 5.2 for further details regarding the JRAS-RDP).

Then, when all $\varphi_n^{(k)}{}_{1)}$ (or $\varphi_n^{(k)}{}_{2)}$ in case the minimum resources have been satisfied) become zero, it means that the localized bursts can not be increased any more since the BS doesn't know the channel in the remaining RUs. Those remaining resources are then divided into tiles, physically indexed and mapped over N_{dMRU} RUs. Moreover, the N_{dMRU} MRUs are logically indexed such that the tiles of two consecutive RUs are placed at distant positions within the downlink frame. Then, the unallocated RUs are iteratively assigned (irrespective of its index) to the different distributed bursts until all the RUs have been assigned or all the data in the buffers have been allocated. The priorities assigned to each user are updated every iteration considering the ESINR is equal to the average SNR and the channel is Rayleigh (hence the priorities don't depend on the RU position). The maximum number of distributed bursts (according to the signalling scheme proposed in [31][47]) is equal to the number of users, K, in case all the buffers are non-empty. Hence, at the end of this second allocation process, each user might be allocated one distributed burst with up to N_{dMRU} consecutive (logically indexed) MRUs.

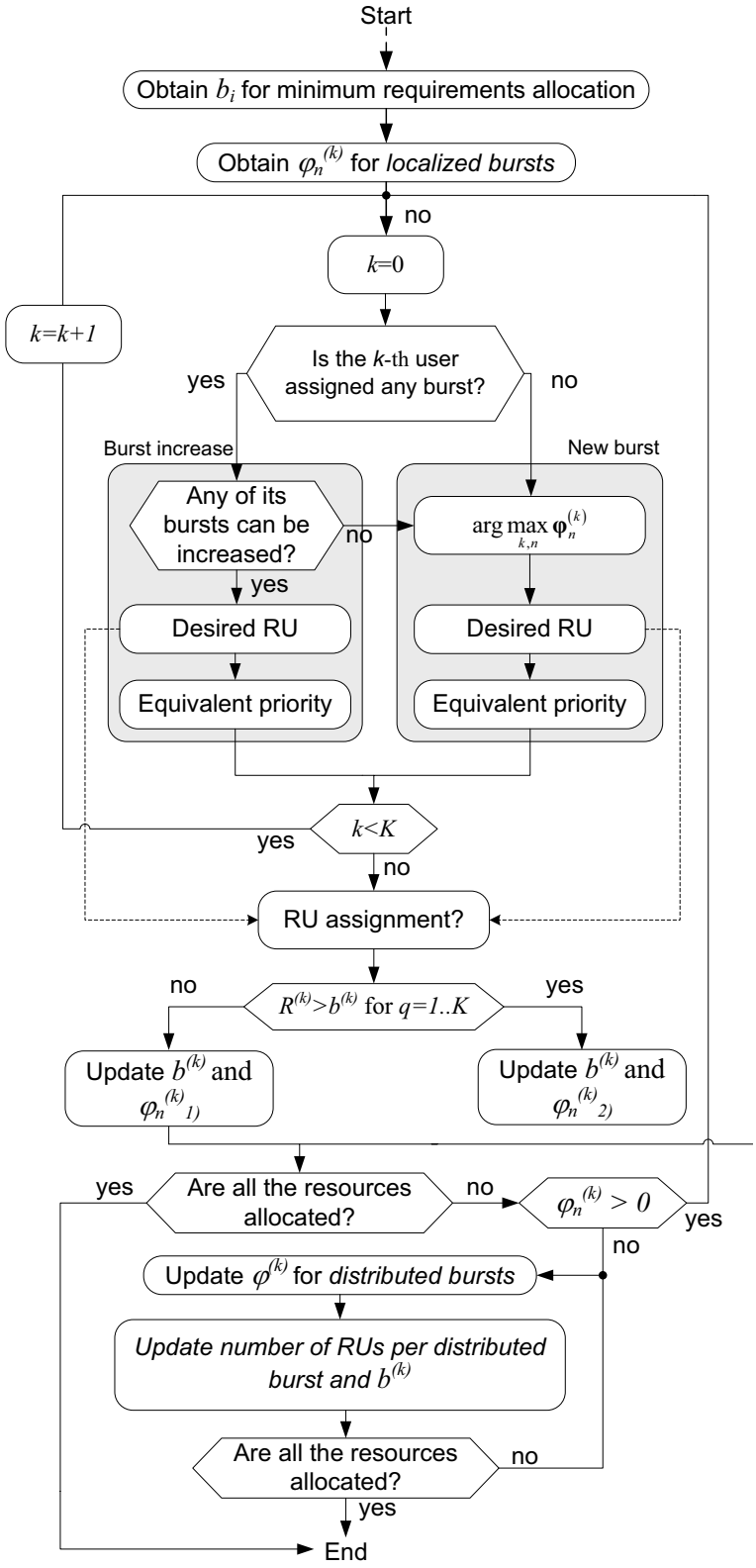


Figure 5.19. Proposed JRAS-RDP with limited feedback algorithm description.

5.3.3. Simulation Results

In order to evaluate the performance of the JRAS-RDP with limited feedback algorithm the system parameters described in Section 5.1.5 have been used. The shape of the RUs for the localized bursts is N_{sc}=18 and N_{Sd} =3, whereas the distributed burst are formed by 6 tiles (i.e. N_t=6) with 3 sub-carriers and 3 symbols each tile (i.e. N_{st}=9). The loss due to pilot subcarriers in both types of RUs is a function of the pilot density μ_p=8/9. We assume that the MTBA zone spans over the whole useful downlink subframe hence N_S=30 OFDM symbols. The performance of the proposal is assessed only for Best Effort traffic varying the number of active users K and the knowledge of the CSI at the BS. The main evaluated performances are the spectral efficiency in order to observe the effects of signalling reduction, and the packet delay to determine if the system is able to guarantee the different QoS requirements.

In order to take into account the loss due to the signalling (channel feedback and DL-MAP transmission), we define the system spectral efficiency v as

$$v\,[bits/s/Hz] = \left(\frac{\sum_{d=1}^{D} B_d}{\sum_{d=1}^{D} \left(\frac{B_d}{r_d} + \frac{B_{sign}}{r_{sign}} \right) + \frac{B_{CSI} + B_{SNR}}{r_{sign}}} \right) \mu_p \qquad (5.27)$$

where B_d is the data bits transmitted in the d-th burst, r_d is the spectral efficiency of the MCS applied to the proper burst, and D means the total number of bursts transmitted during the whole simulation. B_{sign} indicates the signalling bits (transmitted within the DLMAP) required by each allocated burst, whereas B_{CSI} and B_{SNR} are the bits transmitted by the MS in the uplink to report its CSI and the average SNR respectively. It is considered here that the signalling bits are transmitted with a BPSK and a coding rate 1/2, thus the term r_{sign} which means that the spectral efficiency of the MCS used to transmit the signalling is equal to r_{sign}=0.5. The amount of signalling bits required for each burst is 60 and 43 the each localized and distributed burst respectively. Since the signalling bits are transmitted following a PUSC scheme, we can also applying the Rayleigh SNR thresholds values depicted in Table 5.10 being sure that these bits are properly decoded (BER<10^{-6}) for any user within a range of 1.69km. Consequently, the users in the cell border are able to decode properly the DL-MAP. On the other hand, to obtain the term B_{CSI} we have considered that the MSs feed back the SNR value of their n-best subchannels with 6 bits, thus the SNR range of 0-32dB is sampled with steps of 0.5dB. Likewise, the average SNR is fed back from all the MSs each 15ms with a codeword of 6 bits length.

In the first scenario, K different active users (K={25,50,75,100,125,150}) are evaluated where each user's service flow is simulated under the *web browsing test service*. For this service class, the packet delay is unrestricted and the maximum BER for this service class is equal to 10^{-6}. The traffic is modelled such that a web browsing session is simulated (see Section 3.1.5.1 for details on this test service). In our simulation, the mean reading time between two consecutive packet calls has been reduced to 20s to increase the link activity. A simulation time T_{sim} =125s has been used for this scenario in order to include different packets calls.

The system spectral efficiency (see Eq. (5.27)) for different combinations of n (number of fed back subchannels) and k (number of MSs) is depicted in Figure 5.20. Furthermore, the effect of notifying in the downlink which MS have data pending to be transmitted is evaluated. The case where all the MS continuously feedback their CSI information (see solid lines in the plot) are compared with the case where the BS requests the MS to feed back their CSIs only in case there are data for them in the BS queues (dashed lines). For the continuous feedback case, the spectral efficiency is highly reduced as n is increased leading to a (nearly) zero spectral efficiency for the Full-CSI knowledge case. It is then concluded that when the MS have to report their channel independently from the input buffers, the gain due to multiuser diversity is neutralized by the loss caused by the uplink signalling. On the other hand, for the BS requested feedback case, we observed a slight increase of the spectral efficiency as n is increased. This shows that as the knowledge of the channel is improved, the spectral efficiency can be increased if the AMC is applied. Nevertheless, since the traffic simulated is bursty type, we have observed that the number of the MSs that need to be served simultaneously is on average between 1 and 4 for K=25 and K=150 respectively. Clearly, in this case the gain due to multiuser diversity is quite limited as consequence of the bursty nature of the traffic. Larger gains and spectral efficiencies is expected for real-time and non-real time traffic where the number of simultaneously active users is higher.

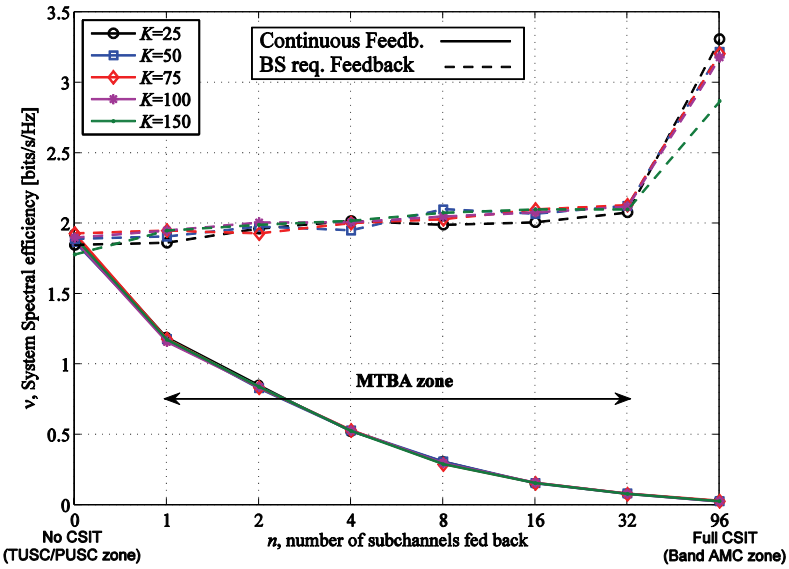


Figure 5.20. Spectral efficiency performance using the JRAS-RDP with limited feedback algorithm and the MTBA zone as a function of the number of subchannels fed back n and the number of users K.

The 99th percentile of the packets delay is depicted in Figure 5.21. From the performances shown we can conclude out that by combining distributed and localized bursts, the delay can be reduced more than one order of magnitude. If we compare the delay in case there is no CSI at the transmitter (i.e. n=0) to the n=1 case, the difference in delay between both schemes is very noticeable (more than one order of magnitude for K>50). This is due to the fact that since the localized bursts have been prioritized over the distributed bursts, it is more than probable that every user is allocated one low rate data channel which guarantees the minimum

throughput. Moreover, it is observed that if n is increased, this minimum throughput is increased reducing the packet delay. However, very slight improvements are achieved for n>4.

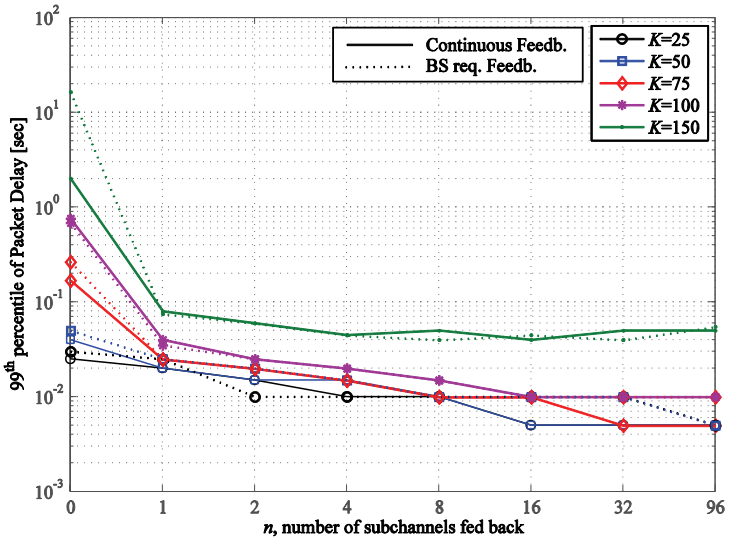


Figure 5.21. 99th percentile evaluation of the packet delay as a function of n and K.

The amount of resources required for downlink and uplink signalling (DL-MAP length and n-best subchannels feedback load) is also analyzed. Figure 5.22 shows the sum of downlink and uplink signalling in terms of number of required OFDM symbols per frame. To obtain the number of OFDM symbols occupied by the signalling we have considered that all the signalling (downlink and uplink) is transmitted with a BPSK modulation and a coding rate equal to 1/2. Therefore, the number of signalling bits that can be fitted within one OFDM symbol (with 1728 active subcarriers) is 768 bits (i.e. $1728\times1/2\times\mu_p$=768). Then, the maximum signalling load obtained through the whole simulation (i.e. the worst case) is represented in Figure 5.22. It is observed that as it was expected as K or n are increased the amount of signalling is increased. If we assume that at most 5 OFDM symbols can be dedicated to signalling within one frame, we can conclude that the best combination pairs of {K, n} are {50,8}, {100,4}, and {150,2}. As a result, by modifying n a tradeoff between the required signalling and the system spectral efficiency (and delay) can be achieved.

Then, two main conclusions are drawn. First, we have observed that the delay can be highly improved in case *localized* and *distributed* bursts are combined. In fact, the use of localized bursts guarantees that unless the system is very saturated every user will have at least one subchannel assigned. This brings up the idea of adapting the number of feedback subchannels according to user requirements or certain fixed priorities (some user may be able to report more channels depending on the requirements). And second, we have shown that by using MTBA zones, a tradeoff between spectral efficiency, delay and load signalling requirements could be obtained by optimizing the number of sub-channels signalled in the uplink. Since Figure 5.22 has been plotted showing the worst case and assuming n fixed during the whole simulation time, we expect that the system performance could be further improved by adapting *n* as a function of the number of users that actually have data to transmit.

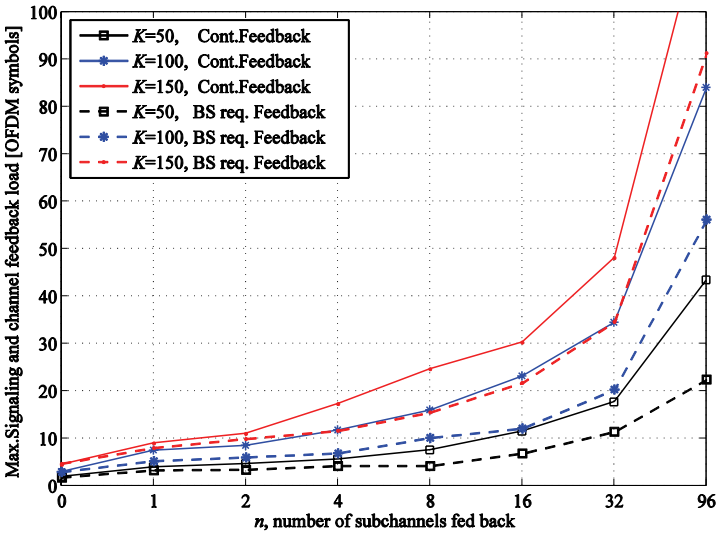


Figure 5.22. Total signalling requirements in the downlink and uplink as a function of n and K.

Finally, it is worth mentioning that the MTBA can be also used to multiplex users with different mobility levels, i.e. the low mobility users can be assigned to the localized bursts, whereas the high mobility users can assigned to the distributed bursts. By doing it so, the low mobility users could achieve high spectral efficiencies whereas the high mobility users would increase the transmit diversity. The advantage then of the MTBA zone comes from the fact that the whole frame could be used efficiently.

5.4. Summary of cross-layer based resource allocation and scheduling schemes

In this chapter, the benefits of joining the resource allocation and the scheduling functions have been investigated. This study has been done to analyse both the MC-CDMA and the OFDMA air interfaces, however it is important that these techniques can be applied to any designed air interface based on the multicarrier transmission. It has been shown that, using the proposed JRAS algorithm, the 99th percentile of the packet delay and the packed dropping ratio can be both reduced up to 50% compared to the Proportional Fair Scheduler (PFS) scheme. Furthermore, the proposed JRAS scheme can cope with the differentiated processing of the service classes as long as each packet delay is supervised at the input buffer level.

However, the complexity and the required signalling seem to be the main drawbacks for such approach. The author's low complexity JRAS with rectangular data packing (RDP) developed algorithm has been extended/adapted in order to fit with the IEEE 802.16 system parameters by allocating contiguous resource units following a rectangular structure. From the obtained simulation results, it is concluded that the JRAS-RDP algorithm is able to pack the data destined to each user with a reduced number of bursts per frame (e.g. lower than 20 bursts per frame for 15 active users) and achieves high spectral efficiency. Furthermore, to the best of the author's knowledge, the proposed scheme is the first that allocates dynamically the resources when contiguous subcarrier permutation is applied. In addition, since the algorithm

operates scaling the priorities of each resource-user mapping, it can be still applied using any other scheduling function (as PFS, MLWDF, etc.).

Since full CSI information at the transmitter is a very idealistic situation, the author also investigated the use of localized and distributed bursts per frame, where each type of burst applies contiguous or interleaved subcarrier permutation respectively. Multiplexing different permutation types in the same subcarrier permutation zone leads to the previous fundamental tradeoff between diversity and efficiency, but in this case the author was able to modify the working point as a function of the amount of the channel feed-backs. Furthermore, a reduction of 90% of the delay has been measured when a combination of the localized and the distributed bursts is used, whereas the uplink signalling has been reduced up to 50% combining selective feedback (only the state of the n best channels is feedback) and the base station notification of incoming packets.

References

[1] S. Shakkottai, T. S. Rappaport and P. C. Karlsson, "Cross-Layer Design for Wireless Networks", *IEEE Comun. Mag.* Oct. 2003, pp.74-80.

[2] G. Song and Y. (G). Li, "Cross-layer optimization for OFDM wireless network – part I and part II," *IEEE Trans. WirelessCommun.*, vol. 4, no. 2, pp. 614–634, March 2005.

[3] R. Wolski, J. S. Plank, J. Brevik, T. Bryan, "Analyzing Market-Based Resource Allocation Strategies for the Computational Grid", International Journal of High Performance Computing Applications, vol.15, pp. 258-281, 2001.

[4] L. Peijuan, R. Berry, M.L. Honig, "A fluid analysis of utility-based wireless scheduling policies," *43rd IEEE Conference on Decision and Control, 2004* (CDC-2004), vol.3, pp.3283-3288 Vol.3, 17-17 Dec. 2004

[5] G. Song; Y. Li, "Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks," *IEEE Communications Magazine,* vol.43, no.12, pp. 127-134, Dec. 2005.

[6] H. Lei; L. Zhang; X. Zhang; D. Yang, "A Packet Scheduling Algorithm Using Utility Function for Mixed Services in the Downlink of OFDMA Systems," *IEEE 66th Vehicular Technology Conference, 2007*, (VTC-2007-Fall), pp.1664-1668, Sept. 30 2007-Oct. 3 2007.

[7] C. Young-June, B. Saewoong, "Delay-Sensitive Packet Scheduling for a Wireless Access Link," *IEEE Transactions on Mobile Computing*, vol.5, no.10, pp.1374-1383, Oct. 2006.

[8] P. Liu, R. Berry, M.L. Honig, "Delay-sensitive packet scheduling in wireless networks," IEEE Wireless Communications and Networking, 2003, (WCNC-2003), vol.3, pp.1627-1632 vol.3, 20-20 March 2003.

[9] J. Mo, J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol.8, no.5, pp.556-567, Oct 2000.

[10] X. Liu, K. Chong, and N. Shroff, "Opportunistic transmission scheduling with resource sharing constraints in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, October 2001.

[11] V. Bharghavan, L. Songwu, T. Nandagopal, "Fair queuing in wireless networks: issues and approaches," *IEEE Personal Communications,* vol.6, no.1, pp.44-53, Feb 1999.

[12] H.J. Kushner, P.A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Transactions on Wireless Communications,* vol.3, n.4, July 2004.

[13] H. Kim, Y. Han, "A proportional fair scheduling for multicarrier transmission systems". *IEEE Communications Letters,* March 2005, Vol. 9, Iss. 3, pp. 210- 212.

[14] A. Jalali, R. Padovani, R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," in *VTC 2000-Spring*, Tokyo, 2000, pp.1854 –1858.

[15] T. Moulsley, "Performance of UMTS HSDPA for data streaming applications",3rd International Conference on 3G Mobile Communications Technologies, 2002.

[16] B. Classon, et al., "Overview of UMTS Air-Interface Evolution", IEEE 64th Vehicular Technology Conference, 2006. Fall. 2006.

[17] M. Andrews, K. Kumaran, K. Ramanan, A. L. Stolyar, R. Vijayakumar, P. Whiting, "Scheduling in a Queueing System with Asynchronously Varying Service Rates," 2000.

[18] S. Shakkottai and A. Stolyar, "Scheduling for Multiple flows Sharing a Time-Varying Channel: The Exponential Rule", American Mathematical Society Translations, Vol. 207, 2002.

[19] S. Shakkottai, A. L. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR", Proc. of 17th International Teletraffic Congress, 2001 pp.793-804.

[20] V. Srivastava, M. Motani, "Cross-layer design: a survey and the road ahead," *IEEE Communications Magazine*, vol.43, no.12, pp. 112-119, Dec. 2005.

[21] K.B. Letaief, Z. Ying Jun, "Dynamic multiuser resource allocation and adaptation for wireless systems," *IEEE Wireless Communications,* vol.13, no.4, pp.38-47, Aug. 2006.

[22] Q. Liu, X. Wang, G. B. Giannakis, "A Cross-Layer Scheduling Algorithm With QoS Support in Wireless Networks", *IEEE Transactions on Vehicular Technology*, Vol. 55, No. 3, May 2006.

[23] L. Wan, W. Ma, Z. Guo, "A Cross-Layer Packet Scheduling and Subchannel Allocation Scheme in 802.16e OFDMA System", *Proc. on Wireless Communication and Networking Conference,* 2007..

[24] S. Soo, D. Geun, W. Sook, "Cross-layer Design of Packet Scheduling and Resource Allocation in OFDMA Wireless Multimedia Networks", *Proc. on IEEE 63rd Vehicular Technology Conference*, May, 2006, Melbourne.

[25] I. Gutiérrez, F. Bader, J.L. Pijoan, S. Ben Slimane, S., "Adaptive Resource Management for a MC-CDMA System with Mixed QoS Classes Using a Cross Layer Strategy ", *IEEE Vehicular Technology Conference (VTC-Spring'07)*, Dublin, Ireland, April, 2007.

[26] I. Gutiérrez, F. Bader, J.L. Pijoan, "Radio Resource Allocation in MC-CDMA Under QoS Requirements", Published in the book Multi-Carrier Spread Spectrum, pp. 207-216., Chapter: Adaptive Transmission, edited by S. Plass, A. Dammann, S. Kaiser and K. Fazel, ISBN: 978-1-4020-6128-8 , Springer, 2007. The Netherlands.

[27] I. Gutiérrez, F. Bader, J.L. Pijoan, "Subchannel Power Allocation and Rate Adaptation for a MC-CDMA System Satisfying Mixed QoS Requirements"**,** *Mosharaka International Conference on Communications, Signals and Coding (MIC-CSC 2008)*, Mosharaka, Jordan, Oct. 2008.

[28] I. Gutiérrez, F. Bader, J.L. Pijoan, "Prioritization Function for Packet Data Scheduling in OFDMA Systems", *Fourth International Wireless Internet Conference (WICON 2008)*, Maui, USA, Nov.2008.

[29] I. Gutiérrez, F. Bader, J.L. Pijoan, "New Prioritization Function for Packet Data Scheduling in OFDMA Systems"**,** *IEEE Radio and Wireless Symposium 2009 (RWS'09)*, San Diego, USA, Jan. 2009.

[30] I. Gutiérrez, F. Bader, R. Aquilue, J.L. Pijoan, "Continuous Frequency-Time Resource Allocation and Scheduling for Wireless OFDMA Systems with QoS Support", *EURASIP Special Issue on OFDMA Architectures, Protocols and Applications* (to be published).

[31] I. Gutiérrez, F. Bader, J.L. Pijoan, "Mixed TUSC and Band AMC Permutation Zone in OFDMA Systems with Limited Feedback", *IEEE Vehicular Technology Conference (VTC'09)*, Barcelona, Spain, 2009.

[32] S. Ryu; B. Ryu; H. Seo; M. Shin, "Urgency and Efficiency based Packet Scheduling Algorithm for OFDMA wireless system," IEEE International Conference on Communications, 2005 (ICC-2005), vol.4, pp. 2779-2785 Vol. 4, 16-20 May 2005.

[33] X. Cai, S. Zhou, G. B. Giannakis. Group-Orthogonal Multicarrier CDMA. *IEEE Transactions on Communications, Vol. 52, No. 1*, Jan. 2004.

[34] Specification of the performance evaluation methodology and the target performance, *Public Deliverable D1.3* , IST-2001-32620 MATRICE.

[35] Y. Ben-Shimol, I. Kitroser, Y. Dinitz, "Two Dimensional Mapping for Wireless OFDMA Systems", IEEE Transactions on Broadcasting, Vol.52, No.3, Sept.2003.

[36] A. Erta, C. Cicconetti, L. Lenzini, "A downlink data region allocation algorithm for IEEE 802.16e OFDMA," *Information, Communications & Signal Processing, 2007 6th International Conference on* , vol., no., pp.1-5, 10-13 Dec. 2007.

[37] C. Desset, E. Batista, G. Lenoir, “WiMAX Downlink OFDMA Burst Placement for Optimized Receiver Duty-Cycling”, Proc. On International Communications Conference (ICC), 2007.

[38] 3GPP TS 36.300 version 8.4.0 Release 8, “E-UTRA and E-UTRAN Overall Description”, April 2008.

[39] Z. Han, Y. Lee, “Opportunistic Scheduling with Partial Channel Information in OFDMA/FDD Systems”, Vehicular Technology Conference, 2004, pp. 511-514 Vol. 1, Sept. 2004.

[40] S. Yoon, O. Somekh, O. Simeone, Y. Bar-Ness, “A Comparison of Opportunistic Transmission Schemes with reduced Channel Information Feedback in OFDMA Downlink”, Proceedings of the 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 2007.

[41] Y. Choi, S. Bahk, “Selective Channel Feedback Mechanisms for Wireless Multichannel Scheduling”, Proceedings of the 2006 Int. symposium on a World of Wireless, Mobile and Multimedia Networks, 2006.

[42] J. H. Kwon, D.Rhee, I.M. Byun, K.S. Kim, K.C. Whang, “Efficient Adaptive Transmission Technique for Multiuser OFDMA Systems with Reduced Feedback Rate”, Wireless Telecommunications Symposium, 2006. April 2006.

[43] K.I. Pedersen, G. Monghal, I.Z. Kovács, T.E. Kolding, A. Pokhariyal, F. Frederiksen, P.Mogensen, “Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback”, Vehicular Technology Conference, 2007, pp.1792-1796, Sept. 2007.

[44] J. Chen, R. A. Berry, M.L. Honig, “Performance of Limited Feedback Schemes for Downlink OFDMA with Finite Coherence Time”, Military Communications Conference, 2007. IEEE MILCOM 2007, Oct. 2007

[45] A. G. Marques, G. B. Giannakis, F.F. Digham, F.J. Ramos, “Power Efficient Wireless OFDMA Using Limited-Rate Feedback”, IEEE Transactions on Wireless Communications, Vol.7, No.2, Feb.2008.

[46] T. Eriksson, T. Ottosson, “Compression of Feedback for Adaptive Transmission and Scheduling”, Proceedings of the IEEE , vol.95, no.12, pp.2314-2321, Dec. 2007.

[47] K. H. Lee et al., “Overlapped two-dimensional resource allocation scheme in OFDMA DL-MAP”, IEEE C802.16e-04/352r2, August, 2004

Chapter 6. Space-Time Adaptation under the LDC framework

The use of multiple antennas at the transmitter and/or receiver sides has been used for years in order to increase the signal to noise ratio at the receiver or for beam steering (also referred as *beamforming*). However, one of the main benefits of using a Multiple Input Multiple Output (MIMO) channel is the increase of the channel capacity. Actually, using different space-time-frequency coding techniques, orthogonal (or quasi orthogonal) virtual paths between transmitter and receiver can be obtained. These virtual paths can be used in order to increase the spectral efficiency or in order to increase the proper diversity (i.e. space diversity). In fact, a fundamental tradeoff between diversity and multiplexing capabilities exists and must be considered when designing a multiple antenna system.

In this chapter, the following issues are addressed:

- We start by describing and reviewing the (ergodic) capacity of the MIMO channel in case of perfect channel state information at the transmitter (CSIT). Afterwards, the outage capacity is studied leading to the diversity-multiplexing tradeoff.
- Then the best known Space-Time coding techniques are reviewed, making emphasis on a class of space-time block codes named Linear Dispersion Codes which are of special interest since they allow meeting different diversity-multiplexing tradeoff under the same scheme.
- Finally, we investigate different spatial adaption and precoding mechanisms which can be used in combination with the adaptation mechanisms explained in the previous chapter. A new spatial adaptation algorithm called *Transmit Antenna and space-time Coding Selection* (TACS) is introduced and evaluated. This new scheme combines the well-known transmit antenna selection techniques with the precoding schemes where the transmitter selects the best space-time code according to the channel state information.

The analysis is done under the general Linear Dispersion Codes framework, hence any linear space-time coding scheme can be included in the study. This kind of spatial adaptation can be very relevant in uplink modes where the MS cannot have more than two radio-frequency chains due to implementation cost and space constrains, hence, although the MS may have up to four antennas, only one or two transmitting antennas can be active simultaneously. These can be extended to the downlink where the BS may have several antennas well spread in the space due to lower space restrictions, but only a short amount of the antennas can be active simultaneously (i.e. antennas with high correlation can be disabled without affecting the performance).

6.1. Capacity of the MIMO Channel

We have already introduced in the Chapter 4 the definition of the channel capacity. Usually, for capacity evaluation of MIMO channels it is assumed that the fading coefficients between antenna pairs are i.i.d. Rayleigh distributed (hence $\mathbf{H}=\mathbf{H}_w$ as stated in Section 2.1.6). Moreover, it is usually assumed that the channel is constant during a codeword of length shorter than the coherence time transmitted within a bandwidth smaller that the coherence bandwidth. Under this assumption the channel is referred as a block fading channel.

For a given channel realization $\mathbf{H}$, the mutual information is given in [1] by

$$I\left(\rho,\mathbf{H},\mathbf{Q}\right)=\log_2\det\left[\mathbf{I}_N+\frac{\rho}{M}\mathbf{HQH}^{\mathrm{H}}\right] \tag{6.1}$$

where ρ is the Signal to Noise Ratio (SNR) and $\mathbf{Q}$ is the input covariance matrix whose trace is normalized to be unitary. Moreover, in case $\mathbf{Q}=\mathbf{I}_{\mathrm{M}}$ this corresponds to a Uniform Power Allocation (UPA) strategy across the transmitter antennas.

Then, given Eq. (6.1), the capacity of the $N\times M$ MIMO channel with perfect Channel State Information (CSI) at the transmitter is written as

$$C\left(\rho,\mathbf{H}\right)=\underset{\mathbf{Q}\geq 0:Tr\{\mathbf{Q}\}=M}{\arg\max}\quad\log_2\det\left[\mathbf{I}_N+\frac{\rho}{M}\mathbf{HQH}^{\mathrm{H}}\right]. \tag{6.2}$$

To obtain the optimum input covariance matrix $\mathbf{Q}^{\Diamond}$, first we apply the *Singular Value Decomposition* (SVD) to the MIMO channel matrix

$$\mathbf{H}=\mathbf{U}_{\mathbf{H}}\mathbf{\Sigma}_{\mathbf{H}}\mathbf{V}_{\mathbf{H}}^{\mathrm{H}}, \tag{6.3}$$

where $\mathbf{\Sigma}_{\mathbf{H}}$=diag{$\sigma_1$,..., σ_{n}} and n=rank($\mathbf{H}$), hence n is equivalent to the number of independent channels between transmitter and receiver. Then we express the optimum input covariance matrix as

$$\mathbf{Q}^{\Diamond}=\mathbf{V}_{\mathbf{H}}diag\left\{p_1^{\Diamond},...,p_n^{\Diamond}\right\}\mathbf{V}_{\mathbf{H}}^{\mathrm{H}}. \tag{6.4}$$

Then, substituting Eq.(6.3) and (6.4) into Eq. (6.2) we rewrite the channel capacity as

$$C\left(\rho,\mathbf{H}\right)=\underset{\{p_k\}_{k=1}^{n}}{\arg\max}\quad\sum_{k=1}^{n}\log_2\left(1+\rho p_k\sigma_k^2\right). \tag{6.5}$$

The expression in Eq. (6.5) is equivalent to the Eq. (4.47) in Section 4.3.2 for subcarrier power adaptation in case single input single output channels. Then, the solution to Eq. (6.5) is the well-known *water-filling* algorithm with the following power allocation strategy

$$p_k^{\Diamond}=\left(\frac{1}{\lambda^{\Diamond}}-\frac{1}{\rho\sigma_k^2}\right)^{+},\quad k=1,...,n \tag{6.6}$$

where $\lambda^{\Diamond}$ defines the water-level and it is chosen such that the following power constraint

$$\sum_{k=1}^{n}p_k^{\Diamond}=1 \tag{6.7}$$

is satisfied.

The transmission scheme that exploits the channel capacity in case of CSI at the transmitter is known as multiple eigenmode transmission. It is shown that if the transmitter multiplies the input vector $x^{M\times 1}$ using $\mathbf{V}^{H}$ as a precoding matrix, and the receiver multiplies the received vector by $\mathbf{U}_{\mathbf{H}}^{H}$, the effective input-output relationship is given by

$$\mathbf{y}=\sqrt{E_s}\mathbf{U}_{\mathbf{H}}^{\mathrm{H}}\mathbf{\Sigma}_{\mathbf{H}}\mathbf{V}_{\mathbf{H}}\mathbf{x}+\mathbf{n}=\sqrt{E_s}\mathbf{\Sigma}_{\mathbf{H}}\mathbf{x}+\mathbf{n}, \tag{6.8}$$

where it is observed that n=*rank*($\mathbf{H}$) parallel (or decoupled) channels can be used to transmit up to n independent complex symbols.

We can conclude that to obtain an optimum solution at low SNR values, the optimum allocation strategy will be to allocate all the available power to the strongest (or dominant) eigenmode, whereas at high SNRs the maximum capacity is obtained by allocating the same power to all the non-zero eigenmodes. Actually, it is proven that UPA is also the optimum

strategy for fast fading channels where the transmitter is not able to capture the instantaneous channel state [1]. Then, the (ergodic) channel capacity is given by

$$C_{UPA}(\rho,\mathbf{H}) = \mathrm{E}\left\{\sum_{k=1}^{n}\log_2\left(1+\frac{\rho}{M}\sigma_k^2\right)\right\} = \mathrm{E}\left\{\log_2\left(1+\frac{\rho}{M}\|\mathbf{H}\|_F^2\right)\right\}. \tag{6.9}$$

Similar research studies have been undertaken regarding the effects of antenna correlation on the channel capacity. It is shown in [2] that for low antenna correlation values, the optimum strategy is to allocate the same power to all the eigenmodes, whereas for high correlation values the optimum strategy is allocating all the power to the strongest eigenmode.

6.2. The Diversity-Multiplexing tradeoff

In the previous chapters, the concept of diversity first appeared in single input single output links where it could be only achieved thanks to coding and interleaving in the frequency and time domains. However, frequency and time diversity (many times obtained through redundancy) incur in a loss in bandwidth or transmission time delay. Alternatively, in case of multiple input multiple output channels, the spatial dimension can be also exploited in order to increase the diversity without losing bandwidth neither increasing the transmission delay.

Some metrics are defined to characterize the diversity. First, the *diversity gain* is defined as the negative asymptotic slope (i.e. for $\rho \to \infty$) of the *log-log* plot of the average error probability P versus the average SNR ρ

$$g_d = -\frac{\log(P)}{\log(\rho)}. \tag{6.10}$$

Furthermore, the diversity gain (or diversity order) is equivalent to the number of independently fading branches. Second, the *array gain* is defined as the increase of average SNR due to coherent combining (*beamforming*) in case of multiple antennas at either transmitter or receiver sides. Hence array gain will be achieved always at the receiver side (it is assumed that the receiver always knows the CSI), whereas it will be achieved at the transmitter side only in case of having the CSI at the transmitter. The array gain is then expressed as

$$g_a = \frac{\rho_{ma}}{\rho_{sa}} \tag{6.11}$$

where ρ_{sa} is the average SNR for the SISO link, and ρ_{ma} is the average SNR for the MIMO link. The *coding gain* is defined as the SNR gain (observed as a left shift of the error curve). Then the coding gain g_c is analytically expressed as

$$P_e = \left(\frac{c}{g_c\rho}\right)^{\alpha g_a}. \tag{6.12}$$

where P_e is error probability, and α and c are scaling constants depending on the modulation level, the coding scheme, and the channel distribution. The following Figure 6.1 depicts the above mentioned concepts over the bit error rate of a QPSK transmission having an AWGN channel and an uncorrelated Rayleigh channel.

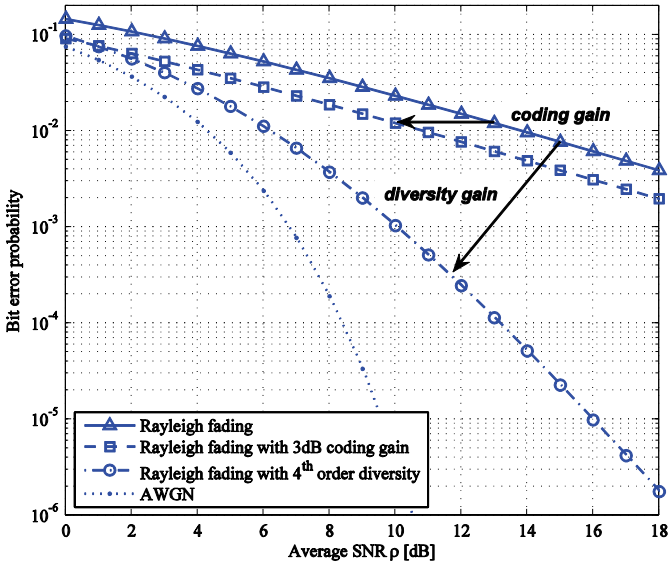


Figure 6.1. Array and diversity gain in Rayleigh fading channels.

Furthermore, in case of multiple antennas at both sides of the link, we have already observed that multiple independent channels exist according to the rank of the channel matrix. Hence, the multiplexing gain g_s is defined as the ratio of the transmission rate $R(\rho)$ to the capacity of an AWGN with array gain g_a

$$g_s = \frac{R(\rho)}{\log_2(1+g_a\rho)} \quad \underset{\rho\to\infty}{\Rightarrow} \quad R(\rho) = g_s \log_2(g_a\rho) \tag{6.13}$$

where $R(\rho)$ is the transmission rate.

For a slow fading channel (i.e. block fading channel), the maximum achievable rate for each codeword is a time variant quantity that depends on the instantaneous channel realizations. In this case, the outage probability P_{out} metric is preferred and is defined as the probability that a given channel realization cannot support a given rate R

$$P_{out}(R) = \inf_{\mathbf{Q}\geq 0: Tr\{\mathbf{Q}\}\leq M} P\left(\log_2 \det\left(\mathbf{I}_N + \frac{\rho}{M}\mathbf{HQH}^{\mathrm{H}}\right) < R\right). \tag{6.14}$$

Then we can gain insights into the channel behaviour by studying the outage probability as a function of the SNR ρ for a given transmission rate. Actually, we can establish a relationship between the diversity gain and the multiplexing gain via the outage probability P_{out} as

$$g_d(g_s,\rho) = -\frac{\partial \log\left(P_{out}(R)\right)}{\partial \log(\rho)} \quad \underset{\rho\to\infty}{\Rightarrow} P_{out}(\rho) = \rho^{-g_d}. \tag{6.15}$$

Both the multiplexing gain and the diversity gain are upper bounded by $g_s \leq min(N,M)$ and $g_d \leq N\times M$. Intuitively, the multiplexing gains indicates the increase of the transmission rate as a function of the SNR, whereas the diversity gain how fast the outage probability decreases with the SNR.

The analysis of Eq. (6.15) at high SNR and uncorrelated Rayleigh channel leads to the diversity-multiplexing trade-off of the channel [3]. It has been shown that $g_d(g_s,\infty)$ is a piece-wise linear function joining the $(g_s,\ g_d(g_s,\infty))$ points with $g_s=\{0,\ldots,\ \min(N,M)\}$ and $g_d=(N-g_s)\times(M-g_s)$. This

trade-off is illustrated in the following 6.2. It is observed that maximum diversity is achieved for a spatial multiplexing gain of zero (i.e. the transmission rate is fixed), whereas the maximum spatial multiplexing gain is achieved when the diversity gain is zero (the outage probability is kept fixed).

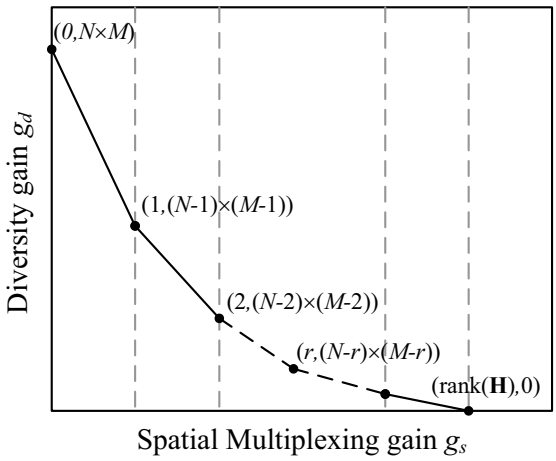


Figure 6.2. Asymptotic diversity-multiplexing trade-off in uncorrelated Rayleigh channels.

6.3. Space-Time Coding over MIMO channels

Analogously to channel coding in SISO links, two types of channel coding have been used for MIMO channels: block coding (referred as Space Time/Frequency Block Coding, STBC/SFBC) and convolutional coding (referred as Space Time Trellis Coding, STTC) [4][5]. For the STBC case, the codeword is only a function of the input bits, whereas the encoder output for the STTC is a function of the input bits and the encoder state (which depend on previously transmitted bits). Both STBC and SFBC are very similar, in STBC the coding is performed along the space and time dimension where the channel must remain static during a certain period larger than the codeword (valid for low mobility environments), whereas for SFBC the coding is performed along the space and frequency domains where the codeword spanning must be less than the coherence bandwidth (hence valid for environments with low delay spread). Since the channel is assumed constant in both for the whole codeword, the analysis of STBC is also valid for SFBC and the same conclusions can be drawn for both.

The inherent memory of the STTC provides an additional coding gain compared to the STBC at the expense of higher computational complexity [8][9]. However, since STBC transforms the MIMO channel into an equivalent scalar AWGN channel [7], the concatenation of traditional convolutional (or Trellis) coding with STBC shows to outperform STTC for low number of receive antennas ($M,N\leq 2$) [8] and the same number of encoder states. Moreover, for a higher number of transmit and receive antennas, the STBC codes designed under the Linear Dispersion Codes (LDC) framework (which are able to preserve channel capacity) combined with convolutional coding are also preferred since they achieve similar performance at lower computational cost compared with the STTC with a high number of Trellis states [10].

Therefore, due to the higher flexibility of the LDC (i.e. linear STBC) codes in order to adjust the diversity-multiplexing tradeoff as well as their low complexity, the author focuses on this class of space-time coding for the subsequent analyses.

6.3.1. Space-Time block coding system model

In the following, a MIMO system with M, N transmitter and receiver antennas respectively communicating over a frequency flat-fading channel is assumed. A codeword $\mathbf{X}$ is transmitted over T channel accesses (symbols) over the M transmitting antennas, hence $\mathbf{X}=[\mathbf{x}_0 \ldots \mathbf{x}_{T-1}]$ with $\mathbf{x}_i \in \mathbb{C}^{M\times 1}$. All the codewords are contained inside a codebook $\mathcal{X}$, and each codeword contains the information from Q complex symbols. The ratio of symbols transmitted per codeword is defined as the spatial multiplexing rate $r_s=Q/T$, where in case of $r_s=M$ the code is referred as *full-rate*. The transmission rate is given by $R=Q\times log_2(Z)/T$ [bits/s/Hz] where $log_2(Z)$ is the number of bits transmitted by each complex symbol. Moreover, the spreading of the symbols in the time and spatial domains lead to the increase of diversity, whereas modifying Q we can modify the spatial multiplexing gain. In consequence, it is possible to obtain different spatial multiplexing-diversity tradeoffs by simply modifying Q [11].

During any i^{th} time instant (equivalent to a channel access), the transmitted and received signals having an SNR ρ are related as

$$\mathbf{y}_i = \sqrt{\frac{\rho}{M}}\mathbf{H}_i\mathbf{x}_i + \mathbf{n}_i, \quad i = \{0,\ldots,T-1\} \tag{6.16}$$

where

$$\begin{aligned} &\mathbf{x} \in \mathbb{C}^{M\times 1} \text{ with } \mathrm{E}\{\mathbf{x}^{\mathrm{H}}\mathbf{x}\} \le M\mathbf{I}_M, \\ &\mathbf{y} \in \mathbb{C}^{N\times 1}, \\ &\mathbf{H} \in \mathbb{C}^{N\times M} \text{ with } \mathrm{E}\{\mathbf{H}^{\mathrm{H}}\mathbf{H}\} = MN, \\ &\text{and } \mathbf{n} \in \mathbb{C}^{N\times 1} \text{ with } \mathbb{C}\mathcal{N}(0,1). \end{aligned} \tag{6.17}$$

and the average transmit power is also normalized such that

$$\mathrm{E}\{\mathrm{Tr}\{\mathbf{X}\mathbf{X}^{\mathrm{H}}\}\} = MT. \tag{6.18}$$

As it was stated before, whether we assume that the channel is flat in frequency and constant in time during the codeword transmission (T_{coh}>>T), the channel dependency on the i^{th} time instant subindex might be dropped and then $\mathbf{H}_i=\mathbf{H}$ for any i={0,…,T-1}. In addition, assuming an i.i.d. (uncorrelated) Rayleigh distributed channel, the following condition is also true

$$\mathbf{H}_{n,m} \in \mathbb{C}\mathcal{N}(0,1), \tag{6.19}$$

where the indexes n,m indicate the receive and transmit antenna respectively. Under these conditions the quasi-static block fading channel model can be assumed, rewriting Eq. (6.16) as follows

$$\mathbf{Y}^T = \sqrt{\frac{\rho}{M}}\mathbf{H}\mathbf{X}^t + \mathcal{V}^t \Rightarrow \mathbf{Y} = \sqrt{\frac{\rho}{M}}\mathbf{X}\mathbf{H}^t + \mathcal{V} \tag{6.20}$$

where $\mathbf{X}\in\mathbb{C}^{T\times M}$ means the space-time transmitted codeword, $\mathbf{Y}\in\mathbb{C}^{T\times N}$ means the received space-time samples and $\mathcal{V}\in\mathbb{C}^{T\times N}$ represents the noise over each receive antenna during each channel access.

6.3.2. Linear Dispersion Codes

The Linear Dispersion Codes (LDC) are a subclass of the STBC codes where the codeword is given by a linear function of the input data symbols [12]. Linear STBC coding schemes as the *Spatial Multiplexing* (i.e. the Bell-Labs Layered Space Time scheme, BLAST) [14], and the *Alamouti* scheme [15] are examples of very popular and used linear STBC codes. Therefore, if the codeword is a linear function of the data symbols, the transmitted codeword can be expressed as

$$\mathbf{X} = \sum_{q=1}^{Q} \left(\alpha_q \mathbf{A}_q + j\beta_q \mathbf{B}_q \right) = \sum_{q=1}^{Q} \left(s_q \mathbf{C}_q + s_q^* \mathbf{D}_q \right), \tag{6.21}$$

where $\mathbf{A}_q$, $\mathbf{B}_q$, $\mathbf{C}_q$, $\mathbf{D}_q \in \mathbb{C}^{T\times M}$ with $\alpha_q = \Re(s_q)$, $\beta_q = \Im(s_q)$. For power normalization purposes, it is considered that the transmitted complex symbols s_q have zero mean and unitary energy ($\mathrm{E}\{s_q^* s_q\}=1$). The matrices $\mathbf{A}_q$,$\mathbf{B}_q$ and $\mathbf{C}_q$,$\mathbf{D}_q$, which are related as follows: $\mathbf{A}_q = \mathbf{C}_q + \mathbf{D}_q$, and $\mathbf{B}_q = \mathbf{C}_q$ - $\mathbf{D}_q$, are referred as the basis matrices and usually are normalized such that Eq. (6.18) holds. Hence the following condition must be fulfilled

$$\sum_{q=0}^{Q-1} \left(Tr\left\{ \mathbf{A}_q \mathbf{A}_q^{\mathrm{H}} \right\} + Tr\left\{ \mathbf{B}_q \mathbf{B}_q^{\mathrm{H}} \right\} \right) = MT. \tag{6.22}$$

If we impose some conditions on the set of basis matrices $\mathbf{A}_q$,$\mathbf{B}_q$ with $q=\{0,\ldots,Q\text{-}1\}$ the mapping between the input symbols and the transmitted codeword $\mathbf{X}$ are unique and the symbols can be perfectly recovered. Note that in addition to the mathematical expression in Eq. (6.21), the following notation is also used in some references regarding the LDC codes

$$\mathbf{X} = \sum_{q=1}^{Q} \left(\Lambda_q s_q \right) \quad \text{with } \Lambda_q \in \mathbb{C}^{T\times M}. \tag{6.23}$$

It can be here observed that this representation leads to a specific type of LDC codes where $\mathbf{A}_q = \mathbf{B}_q = \mathbf{\Lambda}_q$.

Substituting Eq. (6.21) into Eq. (6.20) and applying the *vec* operator on both sides of the expression, the (real valued) system equation can be rewritten as

$$\underbrace{\begin{bmatrix} \Re(\mathbf{y}_0) \\ \Im(\mathbf{y}_0) \\ \vdots \\ \Re(\mathbf{y}_{Q-1}) \\ \Im(\mathbf{y}_{Q-1}) \end{bmatrix}}_{\triangleq \boldsymbol{y}} = \sqrt{\frac{\rho}{M}} \mathcal{H} \underbrace{\begin{bmatrix} \alpha_0 \\ \beta_0 \\ \vdots \\ \alpha_{Q-1} \\ \beta_{Q-1} \end{bmatrix}}_{\triangleq \boldsymbol{s}} + \underbrace{\begin{bmatrix} \mathbf{n}_0 \\ \mathbf{n}_0 \\ \vdots \\ \mathbf{n}_{Q-1} \\ \mathbf{n}_{Q-1} \end{bmatrix}}_{\triangleq \boldsymbol{n}} \quad \Rightarrow \quad \boldsymbol{y} = \sqrt{\frac{\rho}{M}} \mathcal{H}\boldsymbol{s} + \boldsymbol{n}, \tag{6.24}$$

where $\boldsymbol{s}$ is the real input symbols vector and $\boldsymbol{n} \sim \mathcal{N}(0,1/2)$ is the real vector noise i.i.d. components (the same Gaussian distribution it is usually considered for $\boldsymbol{s}$). The equivalent real valued channel matrix $\mathcal{H}$ is then given by

$$\mathcal{H} = \begin{bmatrix} \mathcal{A}_0 h_0 & \mathcal{B}_0 h_0 & \cdots & \mathcal{A}_{Q-1} h_0 & \mathcal{B}_{Q-1} h_0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{A}_0 h_{N-1} & \mathcal{B}_1 h_{N-1} & \cdots & \mathcal{A}_{Q-1} h_{N-1} & \mathcal{B}_{Q-1} h_{N-1} \end{bmatrix}_{2NT \times 2Q} \tag{6.25}$$

$$= \underbrace{\begin{bmatrix} \mathbf{I}_N \otimes \mathcal{A}_0 & \mathbf{I}_N \otimes \mathcal{B}_0 & \cdots & \mathbf{I}_N \otimes \mathcal{A}_{Q-1} & \mathbf{I}_N \otimes \mathcal{B}_{Q-1} \end{bmatrix}}_{2NT \times 4MNQ} \times \underbrace{\begin{bmatrix} \mathbf{I}_{2Q} \otimes \underline{h} \end{bmatrix}}_{4MNQ \times 2Q}.$$

Moreover, the following definitions have been used:

$$\mathcal{A}_q = \begin{bmatrix} \Re\{\mathbf{A}_q\} & -\Im\{\mathbf{A}_q\} \\ \Im\{\mathbf{A}_q\} & \Re\{\mathbf{A}_q\} \end{bmatrix} \in \mathbb{R}^{2T \times 2M},$$

$$\mathcal{B}_q = \begin{bmatrix} -\Im\{\mathbf{B}_q\} & -\Re\{\mathbf{B}_q\} \\ \Re\{\mathbf{B}_q\} & -\Im\{\mathbf{B}_q\} \end{bmatrix} \in \mathbb{R}^{2T \times 2M}, \tag{6.26}$$

$$h_n = \begin{bmatrix} \Re\{\mathbf{h}_n\} \\ \Im\{\mathbf{h}_n\} \end{bmatrix} \in \mathbb{R}^{2M \times 1}, \ \underline{h} = \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{N-1} \end{bmatrix} \in \mathbb{R}^{2MN \times 1},$$

where $\mathbf{h}_n$ is the n-th row of the MIMO channel matrix $\mathbf{H}$. The ergodic channel capacity achieved by a LDC code is then given by

$$C_{LDC}(\rho, \mathcal{H}, T) = \underset{tr\{\mathcal{H}\mathcal{H}\}^{\mathrm{H}} \le MT}{\arg\max} \ \frac{1}{2T} \log_2 \det\left[\mathbf{I}_{2NT} + \frac{\rho}{M} \mathcal{H}\mathcal{H}^{\mathrm{H}} \right], \tag{6.27}$$

hence $C_{\mathrm{LDC}}(\rho,\mathcal{H},T) \le C(\rho,\mathbf{H})$, and analogously to Eq. (6.5), the Eq. (6.27) can be upper bounded by

$$C_{LDC}(\rho, \mathcal{H}, T) \le \frac{1}{2T} \sum_{n=1}^{2Q} \log\left(1 + \frac{\rho}{M} \lambda_n^{\,2}\right) \tag{6.28}$$

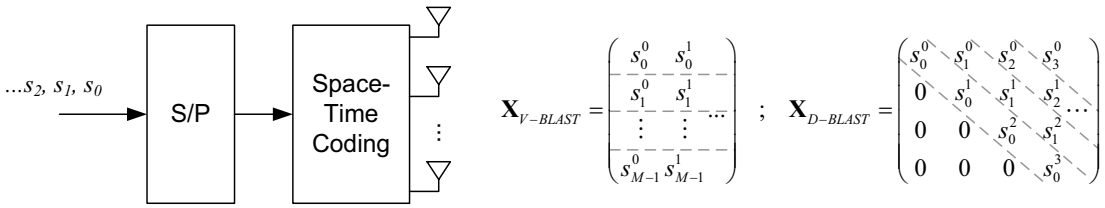
where $\lambda_n^{\,2}$ is the n-th singular value.

Theoretically the maximum number of possible independent streams or channel modes of the effective MIMO channel matrix $\mathcal{H}$ is $2NTx2Q$ (maximum number of singular values different than zero). However, following the LDC design, the number of modes that are excited is equal to the rank of $\mathcal{H}^{\mathrm{T}}\mathcal{H}$ and equal to $2Q$ (i.e. L=$2Q$) in the best case [20]. Furthermore, we can observe in Eq. (6.24) that it is possible to use a linear receiver as long as $Q \le NT$, otherwise the system would be undetermined. If $Q < NT$ the system is over-determined, and in consequence more reliability is given to each estimated symbol (i.e. spatial diversity is increased).

6.3.3. Examples of (well-known) codes expressed as LDC

It has been emphasized that the LDC framework allows expressing all the linear STBCs with the same structure. The first linear STBC proposed follows the Bell-Labs Layered Space Time (BLAST) architecture [14]. The *Spatial Multiplexing* (SM) scheme also referred as V-BLAST, is a full rate code (r_s=M) that consists on transmitting independent data streams on each transmit antenna. A layer is considered as the stream of data transmitted from each transmit antenna, hence each layer experiences a different channel. It is then shown in [11] that the diversity order of V-BLAST is at most $g_d \le N$. An improvement of the V-BLAST scheme is the D-BLAST

where each stream is rotated across the transmitter antennas, thus, the spatial diversity is increased. Furthermore, by introducing zeros at the beginning of the transmission (i.e. the triangular initialization block) the first layer can be obtained with the lowest interference level. Then, the contribution of the first layer is cancelled in order to decode the second layer. The process is repeated for the subsequent layers. Both BLAST structures are depicted in the following Figure 6.3.



$$\mathbf{X}_{V-BLAST}=\begin{pmatrix} s_0^0 & s_0^1 & \cdots \\ s_1^0 & s_1^1 & \\ \vdots & \vdots & \\ s_{M-1}^0 & s_{M-1}^1 & \end{pmatrix} \quad ; \quad \mathbf{X}_{D-BLAST}=\begin{pmatrix} s_0^0 & s_1^0 & s_2^0 & s_3^0 \\ 0 & s_0^1 & s_1^1 & s_2^1 \cdots \\ 0 & 0 & s_0^2 & s_1^2 \\ 0 & 0 & 0 & s_0^3 \end{pmatrix}$$

Figure 6.3. Bell-Labs Layered Space Time (BLAST) architecture.

Then, following the notation

$$\Psi=\begin{bmatrix} vec\left(\mathbf{A}_0\right) & \cdots & vec\left(\mathbf{A}_{Q-1}\right) \\ vec\left(\mathbf{B}_0\right) & & vec\left(\mathbf{B}_{Q-1}\right) \end{bmatrix}^t \tag{6.29}$$

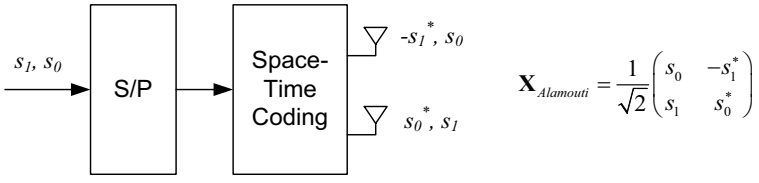
where $\Psi\in\mathbb{R}^{Q\times 2MT}$, the basis matrices of the LDC that models the V-BLAST (SM) for $M/N/T/Q$=2/2/2/4 schemes are given by

$$\mathbf{\Psi}_{SM(2/2/2/4)}=\frac{1}{\sqrt{2}}\begin{bmatrix} 1&0&0&0&1&0&0&0 \\ 0&1&0&0&0&1&0&0 \\ 0&0&1&0&0&0&1&0 \\ 0&0&0&1&0&0&0&1 \end{bmatrix}. \tag{6.30}$$

Moreover, it is shown that the SM scheme is capacity efficient ($\mathbf{\Psi}_{SM}\times\mathbf{\Psi}_{SM}^{\mathrm{T}}=\mathbf{I}_{2\mathrm{M}}/M$)[1], and optimal from point of view of error rate minimization perspective (the spatial streams do not interference between them) [11]. The drawback of the SM scheme is that the diversity gain is minimized. In fact, the maximum diversity gain in case of i.i.d. Rayleigh channel for $N\geq M$ is given by [11]

$$g_d\left(g_s\right)=N\left(1-\frac{g_s}{M}\right), \quad g_s\in\left[0,M\right] \tag{6.31}$$

which for the code expressed in Eq. (6.30) with $M/T/Q$=2/2/4, g_d is equal to zero (i.e. it works on the last segment of the diversity-multiplexing trade-off curve in Figure 6.2).



$$\mathbf{X}_{Alamouti}=\frac{1}{\sqrt{2}}\begin{pmatrix} s_0 & -s_1^* \\ s_1 & s_0^* \end{pmatrix}$$

Figure 6.4. Alamouti Space-Time coding architecture.

Another STBC scheme proposed at the end of the 90s is the well-known Alamouti's scheme [15]. This code was designed for two transmitting antennas and for a spatial multiplexing rate r_s=1. The behaviour of the Alamouti scheme is represented in Figure 6.4. The Alamouti's

scheme falls into an important subclass of LDC codes named Orthogonal-STBC (OSTBC). These codes have the property that the different multiplexed streams can be perfectly decoupled without any interference between them despite information is spread uniformly in space and time domains. As a result, a linear decoder is able to achieve the same performance than the maximum likelihood (ML) decoder. This decoding property is extracted from the following condition

$$\mathbf{X}\mathbf{X}^{\mathrm{H}} = \mathbf{\Psi}^{\mathrm{H}}\mathbf{\Psi} = \frac{T}{QM}\left[\sum_{q=0}^{Q-1}\left|s_q\right|^2\right]\mathbf{I}_M, \tag{6.32}$$

which must be hold for any OSTBC [11]. Then, the basis matrices for the Alamouti scheme (with $M/T/Q$=2/2/2) are expressed as

$$\Psi_{Alamouti} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}. \tag{6.33}$$

Furthermore, the diversity multiplexing trade-off achieved by such OSTBC code in case of i.i.d. Rayleigh fading channels is given by [11]

$$g_d\left(g_s\right) = NM\left(1-\frac{g_s}{r_s}\right), \quad g_s \in \left[0, r_s\right]. \tag{6.34}$$

Thus, the Alamouti scheme in case of $M/T/Q$=2/2/2 achieves full diversity. However, an OSTBC is only capacity optimal in case that the rank of the channel and the spatial multiplexing rate are equal, this only happens when both are equal to one which leads to the initial Alamouti scheme with $M/N/T/Q$=2/1/2/2.

Finally, another well-known linear STBC is the *Golden* code which is able to achieve the optimal diversity-multiplexing trade-off illustrated in Figure 6.2 under the premises that $M/T/Q$=2/2/4, with $g_s \in [0, min(2,N)]$ [16]-[18]. The basis matrices of this code are given as

$$\mathbf{\Psi}_{Golden} = \begin{bmatrix} 0.447 - j0.276 & 0.724 - j0.447 & 0 & 0 \\ 0 & 0 & 0.447 - j0.276 & 0.724 - j0.447 \\ 0 & 0 & -0.724 + j0.447 & 0.447 - j0.276 \\ 0.447 + j0.724 & -0.276 - j0.447 & 0 & 0 \\ 0.447 - j0.276 & 0.724 - 0.447 & 0 & 0 \\ 0 & 0 & 0.447 - j0.276 & 0.724 - j0.447 \\ 0 & 0 & -0.724 + j0.447 & 0.447 - j0.276 \\ 0.447 + j0.724 & -0.276 - j0.447 & 0 & 0 \end{bmatrix}^t. \tag{6.35}$$

All these reviewed codes appear in the IEEE 802.16 standard [19] under the names of **A**, **B** and **C** matrices for the Alamouti, the SM and the Golden code respectively and are widely used in case M=T=2.

6.3.4. Detection techniques: Linear vs Non-linear schemes

Recovering the transmitted symbols within each codeword might become a challenging task depending on the set of basis matrices used. Moreover, we have already observed that the ergodic channel is also a function of the basis matrices set. Therefore, a compromise between complexity and performance (in terms of spectral efficiency or decoding errors) exists which motivates the implementation and use of different LDC codes and decoding schemes. In the

following subsections different decoding techniques for linear STBC systems given as in Eq. (6.24) are reviewed.

6.3.4.1. Maximum Likelihood (ML) decoding:

Optimum signal detection requires the maximization of the likelihood function over the discrete set of the code alphabet [14]. Mathematically, this can be expressed as

$$\hat{s} = \arg\min_{\hat{s}\in\Sigma} \left\| y - \sqrt{\frac{\rho}{M}}\mathcal{H}s \right\|_F^2 \tag{6.36}$$

where Σ is the space of all the transmitted symbol vectors where all the input data combinations are equally probable. Regarding the computational complexity of the ML detector, since each vector has a set of $2Q$ symbols each one mapped over $log_2(Z)$ bits, the computational complexity is exponential in $Q\times log_2(Z)$.

The performance of the ML decoder can be measured in terms of the pair-wise codeword error probability (PEP) which is given by [20]

$$P\left(s \to s'\right) \le \frac{1}{\left| I_{4MN} + \frac{\rho}{4} R_S \otimes I_{2N} \right|}, \text{ with } R_S = \left(\mathcal{H}s - \mathcal{H}s'\right)\left(\mathcal{H}s - \mathcal{H}s'\right)^t, \tag{6.37}$$

Then, it is derived from [20] that for high SNR the largest PEP using the ML decoder between any two different symbol vectors $\boldsymbol{s}$ and $\boldsymbol{s}'$ can be upper-bounded by

$$P\left(s \to s'\right) \le \frac{1}{\underbrace{\left(\frac{\rho}{4}\right)^{N\times rank\left(R_S\right)}}_{slope} \underbrace{\prod_{k=1}^{rank\left(R_{\mathbf{S}'}\right)} \lambda_k^N}_{shift}}, \tag{6.38}$$

whose Chernoff bound falls at $1/\rho^v$, where $v=N\times min(rank(\mathrm{R}_S))$. Thus, it is worth mentioning that the improvement in the slope (i.e. diversity) is determined by the smallest product $N\times rank(R_S)$, whereas the shift improvement (or coding advantage) is determined by the smallest product of the singular values of $\mathcal{H}$. Thus it is concluded that, for the ML detector, the eigenvalues of the equivalent channel matrix should be bounded away from zero in order to avoid ambiguity errors produced by system degeneracy [23].

Furthermore, the average pair wise error probability, when the s and s' are chosen with entries from independent zero-mean real Gaussian densities with a variance 1/2, can be upper-bounded according to [12] by

$$P_e \le \mathrm{E}\left\{ \det\left(I + \frac{\rho}{2M}\mathcal{H}^t\mathcal{H} \right)^{-1/2} \right\}. \tag{6.39}$$

Since the pair wise error probability is the probability that at least one of the $Qlog_2(Z)$ bits transmitted per codeword has an error, for $R=Q\cdot log_2(Z)/T$ and applying a union bound, the average bit error probability is then defined as,

$$P_b \le 2^{Q\log_2(Z)-1}\mathrm{E}\left\{ \det\left(I + \frac{\rho}{2M}\mathcal{H}^t\mathcal{H} \right)^{-1/2} \right\} = \frac{1}{2}\mathrm{E}\left\{ e^{\left(-\frac{1}{2}\log\det\left[I + \frac{\rho}{2M}\mathcal{H}^t\mathcal{H} \right] + Q\log_2(Z) \right)} \right\}. \tag{6.40}$$

Both Eq. (6.38) and Eq. (6.40) lead (as expected) to the same conclusion which is that the diversity order $g_d \le N\cdot min(M,T)$ [14]. Moreover, it is observed that T should equal M to achieve

full diversity. However, increasing T requires increasing also the value of Q to maintain the same rate R, which leads to an increase in memory requirements, computational complexity, and decoding time delay. Thus, for LDC codes, there is also a tradeoff between achievable diversity and the decoding complexity.

However, none of the expressions from Eq. (6.38) to (6.40) allow obtaining what is the channel experienced by each input symbol. Hence the author proposes the following expression

$$ESINR_q^{(ML)} \approx \frac{P_{rx} \cdot \left\| \mathbf{H}_q \right\|_F^2 \cdot T / MQ'}{\alpha_n N_0 + \alpha_i \cdot P_{rx} \cdot \left\| \mathbf{H}_q \right\|_F^2 \cdot T / MQ'}, \tag{6.41}$$

to model the Effective Signal to Interference and Noise Ratio (*ESINR*) per transmitted symbol q when the ML detector is used. P_{rx} is the received power, $\|\mathrm{H}_q\|_\mathrm{F}^2$ is the Froebenius norm of the channel matrix obtained considering only the transmitting antennas where the stream is mapped. Q' means the number of streams transmitted per antenna, α_n and α_i are the noise and interference weighting terms respectively which depend on the modulation order as well as the LDC used code. The values shown on Table 6.1 have been obtained by computer simulation for the Alamouti code, the SM scheme and the Golden code's optimized LDC code for the BPSK, QPSK and 16QAM.

Modulation	LDC_{SIMO} (R=1)		$LDC_{Alamouti}$ (R=1)		LDC_{SM} (R=2)		LDC_{Golden} (R=4)	
	α_n	α_i	α_n	α_i	α_n	α_i	α_n	α_i
BPSK	1	0	1	0	0.9	0.111	1	0.066
QPSK	1	0	1	0	1	0.071	1.6	0.035
16QAM	1	0	1	0	1.9	0.013	2.5	0.010

Table 6.1. Values of K_N and K_I estimated for linear STBC with $M/N/T$ = 2/2/2

The evaluation of this model has been carried out by (computer based) link level evaluation comparing the bit error rate obtained by simulation with that obtained assuming the *ESINR* in Eq. (6.41) and the theoretical BER expression for an AWGN channel. Only the performance of the Spatial Multiplexing and Golden codes are illustrated (the correctness for the SIMO and Alamouti can be verified mathematically). It can be observed in Figure 6.5 that within the SNR range where the modulations are used in order to achieve a Packet Error Rate lower than 1%, the bit error rate curves match the theoretical values very tightly. The main advantage of this *ESINR* approximation is that we can use Eq. (6.41) to estimate the optimum modulation and coding scheme (MCS) according to instantaneous channel conditions on a per stream (i.e. symbol) basis.

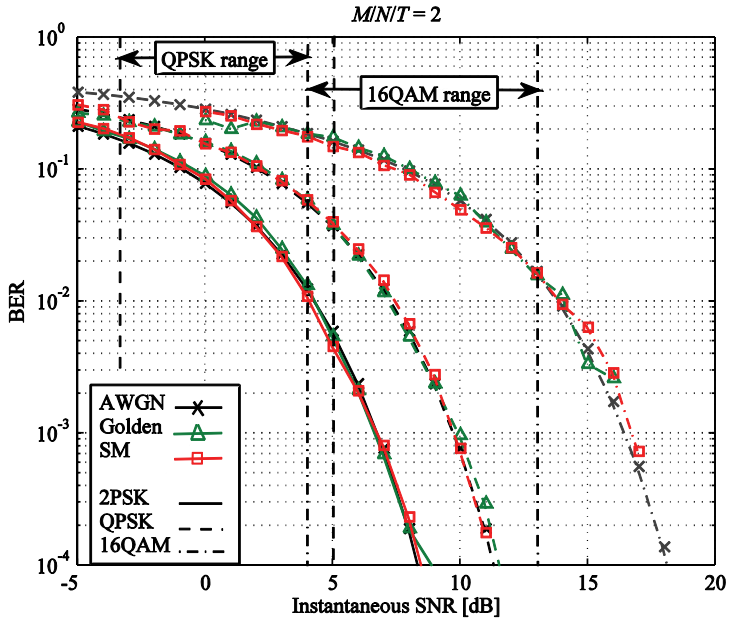


Figure 6.5. Instantaneous SINR versus BER mapping using.

6.3.4.2. Linear detectors: Zero Forcing (ZF) and Minimum Mean Square Error (MMSE)

The high computational cost of the ML receiver ($\mathcal{O}(2^{Q\log_2(Z)})$) leads to the use of less computational demanding receiving techniques such as those based on linear processing [13]. Following the expression in Eq. (6.24), a linear relationship between input and output symbols is noted, hence as long as $Q \leq NT$, solving the system of linear equations defined we can estimate the received symbols [23]. Hence, we can recover the transmitted stream of symbols by applying

$$\hat{\boldsymbol{s}} = \mathcal{G}\boldsymbol{y} = \sqrt{\frac{\rho}{M}}\mathcal{G}\mathcal{H}\boldsymbol{s} + \mathcal{G}\boldsymbol{n}, \tag{6.42}$$

where $\mathcal{G} \in \mathbb{R}^{2Q\times 2N}$ is the equalizer matrix which compensates the MIMO channel effects. Similar to frequency equalization, the equalizer matrix might be designed to suppress the inter-symbol interference (despite the noise vector might be increased due to the equalization) or to minimize the mean square error (i.e. the MMSE). The first design criterion is known as the Zero-Forcing equalization where

$$\mathcal{G}_{ZF} = \sqrt{\frac{M}{\rho}}\mathcal{H}^{\dagger}.^{14} \tag{6.43}$$

On the other hand, according to the MMSE criterion the following equalizer is obtained

$$\mathcal{G}_{MMSE} = \sqrt{\frac{M}{\rho}}\left(\mathcal{H}^{t}\mathcal{H} + \left(\frac{\rho}{M}\right)^{-1}\mathbf{I}_{Q}\right)^{-1}\mathcal{H}^{\mathrm{H}}. \tag{6.44}$$

One of the main benefits of using these linear receivers (apart from their low computational cost, $\mathcal{O}(M^3)$ [13]) is that the channel effects can be completely decoupled on a per stream basis, hence a closed expression for the $ESINR$ per stream can be obtained as

---

[14] $\mathbf{X}^{\dagger}$ denotes the Moore-Penrose pseudo-inverse of matrix $\mathbf{X}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}$, then $\mathbf{X}^{\dagger}=\mathbf{V}\boldsymbol{\Sigma}^{+}\mathbf{U}^{\mathrm{H}}$, where $\boldsymbol{\Sigma}^{+}$ is the element by element inverse of $\boldsymbol{\Sigma}$ (the inverse of zero elements is also zero).

$$ESINR_q = \frac{diag\left[\mathcal{D}\mathcal{D}^t\right]_q}{diag\left[\frac{1}{\rho}\mathcal{G}^t\mathcal{G} + \mathbf{I}_{self}\mathbf{I}_{self}^H\right]_q}. \tag{6.45}$$

where $\mathcal{D}=diag[\mathcal{G}\mathcal{H}]$, and $\mathrm{I}_{self}=\mathcal{G}\mathcal{H}-\mathcal{D}$ is the self-interference term.

Then, substituting Eq. (6.43) or Eq. (6.44) into Eq. (6.45), the following $ESINR$ expressions for both *Zero Forcing* and *MMSE* linear receivers are obtained:

$$\begin{aligned} ESINR_{q,\,ZF} &= \frac{\rho}{M\left[\mathcal{H}^t\mathcal{H}\right]_{q,q}^{-1}}, \\ ESINR_{q,\,MMSE} &= \frac{\rho}{M\left[\mathcal{H}^H\mathcal{H} + 2\rho^{-1}\mathbf{I}_{2Q}\right]_{q,q}^{-1}} - 1. \end{aligned} \tag{6.46}$$

Furthermore, for some specific codes as the Orthogonal STBC (with maximum diversity) or SM schemes, the previous equations can be simplified as follows

$$\begin{aligned} ESINR_q^{(ZF)}\Big|_{OSTBC} &= \frac{\rho}{M}\cdot norm(\mathbf{H}), \\ ESINR_q^{(ZF)}\Big|_{SM} &\geq \frac{\rho}{M}\cdot \lambda_{\min}{}^2(\mathbf{H}). \end{aligned} \tag{6.47}$$

Another important conclusion drawn in [12] is that, in case the equivalent channel matrix $\mathcal{H}$ is orthogonal, the maximum-likelihood and the linear receiving techniques perform equally well because the estimation errors are decoupled. In terms of diversity gain this means that $g_d=MN$ for the OSTBC with linear receivers. For the non-orthogonal schemes such as SM the diversity order is $g_d=N-M+1$, hence when having the same number of transmit and receive antennas the diversity order is just $g_{d,M=N}=1$ which is the diversity of a SISO channel [12]. The authors in [13] propose to use Lattice Reduction techniques in order to obtain at the receiver the equivalent channel matrices that are close to $\mathcal{H}^{\mathrm{T}}\mathcal{H}=\mathbf{I}_M$, in that case the diversity order is increased up to $g_d=N$.

Under the previously described $ESINR$ conditions per stream (see Eq. (6.46)), the per stream error probability P_e can be tightly approximated following the *Nearest Neighbour Union Bound* (NNUB) criterion, which it is defined in [24] as

$$P_e \leq 1 - \mathrm{E}\left(1 - N_n(r)\cdot \boldsymbol{Q}\left(\sqrt{ESINR\frac{d_{\min}^2(Z)}{2}}\right)\right)^Q \tag{6.48}$$

where d^2_{min} is the squared minimum Euclidian distance between any two points of the constellation assuming an average transmission power of 1Watt, whereas N_n is the average number of nearest neighbour constellation points. For instance, for the Z-QAM modulation (where Z = 4, 16, 64, …), $d_{min}{}^2(Z)$ and N_n are given by

$$\begin{aligned} d_{\min}^2(Z) &= \frac{6}{(Z-1)}, \\ N_n(Z) &= 4\left(1 - \frac{1}{\sqrt{Z}}\right). \end{aligned} \tag{6.49}$$

Then, in case $P_e<10^{-2}$, Eq. (6.48) be approximated by

$$P_e \le Q \cdot N_n(Z) \cdot \mathrm{E}\left\{ Q\left(\sqrt{ESINR \frac{d^2_{\min}(Z)}{2}} \right) \right\} \tag{6.50}$$

and whether Gray bit mapping is applied then the bit error probability is approximated by $P_b \cong P_e/Q$.

6.3.5. On the design of the LDC basis matrices.

Several researches have previously focused on designing linear STBC code and their extension to LDC codes that maximize some channel and performance parameters as the channel capacity, the diversity or the link reliability. This section aims to introduce what are the different cost functions over which the LDCs have been optimized hence the further obtained results with these codes can be justified.

Previous analyses have stand out the flexibility of the LDC codes in achieving an arbitrary tradeoff between spatial multiplexing and spatial diversity. Moreover, the LDC optimization must be done for each combination of M, N, T, Q (in the following each code will be referred by its optimization set $M/N/T/Q$). In [12], Hassibi and Hochwald proposed to optimize the LDC basis matrices in order to maximize the mutual information between the transmitted symbol vector $\boldsymbol{s}$ and the received vector $\boldsymbol{y}$. Then, from Eq. (6.27) the convex optimization problem is formulated as,

$$\begin{aligned} C_{LDC}(\rho, M, N, T, \mathcal{H}) &= \underset{\mathbf{A}_q, \mathbf{B}_q, q=1,\ldots,Q}{\arg\max} \quad \frac{1}{2T} \log_2 \det\left[\mathbf{I}_{2NT} + \frac{\rho}{M} \mathcal{H}\mathcal{H}^{\mathrm{H}} \right], \\ st: \quad (i) \quad & \sum_{q=1}^{Q} \left(tr\mathbf{A}_q^* \mathbf{A}_q + tr\mathbf{B}_q^* \mathbf{B}_q \right) \le 2MT, \\ (ii) \quad & tr\mathbf{A}_q^* \mathbf{A}_q \le \frac{MT}{Q}, \quad tr\mathbf{B}_q^* \mathbf{B}_q \le \frac{MT}{Q}, \; q = 1,\ldots,Q, \\ (iii) \quad & \mathbf{A}_q^* \mathbf{A}_q \le \frac{T}{Q} \mathbf{I}_M, \quad \mathbf{B}_q^* \mathbf{B}_q \le \frac{T}{Q} \mathbf{I}_M, \; q = 1,\ldots,Q. \end{aligned} \tag{6.51}$$

The first constraint (i) in Eq. (6.51) bounds the transmission power thus $\mathrm{E}\{\mathrm{tr}(\mathbf{X}\mathbf{X}^*)\} = MT$. The second constraint (ii) ensures that each symbol (the real and imaginary components) is transmitted with the same power over each channel access, the last constraint (iii) is the most stringent and forces the energy from each symbol to be spread uniformly in both time and space domains. In [12] it is also noted that each optimized LDC (i.e. a specific $M/N/T/Q$ set) code performs well in quite wide range of SNR values.

Heath and Paulraj pointed out in [20] that the designed codes for maximizing the capacity may not always guarantee a good performance in terms of error probability. In order to obtain codes that minimize the error probability, it is observed in Eq. (6.38) that the rank and the determinant of R_S must be also maximized.

However, one of the main drawbacks of the resulted codes through the optimization of Eq. (6.51) is that the maximum-likelihood (ML) detection is implicitly assumed. As it was explained in Section 0, the complexity of the ML decoder is $\mathcal{O}(2^{Q\log_2 Z})$. For that reason, Gohary and

Davidson proposed in [23] to look for codes that can be used with linear receivers, hence they defined the optimization problem in order to minimize the Mean Square Error (MSE) as

$$\begin{aligned} &\arg\min_{\mathbf{A}_q,\mathbf{B}_q,q=1,\ldots Q,\mathcal{G}_{MMSE}} \quad MSE, \\ &\text{with} \quad MSE = E\left\{Tr\left((s-\hat{s})(s-\hat{s})^t\right)\right\} = Q - NT + \frac{1}{2}Tr\left(\left(I_{2NT} + \frac{\rho}{M}\mathcal{H}\mathcal{H}^t\right)^{-1}\right). \end{aligned} \tag{6.52}$$

After some simplifications the minimization problem is written as follows [23]

$$\begin{aligned} &\arg\min_{\mathbf{A}_q,\mathbf{B}_q,q=1,\ldots,Q} E_{\mathbf{H}}\left\{\det\left(\mathcal{H}\mathcal{H}^t\right)\right\}, \\ &s.t.: \quad \mathbf{A}_q\mathbf{A}_q^t = \mathbf{B}_q\mathbf{B}_q^t = \frac{M}{Q}\mathbf{I}_T. \end{aligned} \tag{6.53}$$

Furthermore, it is noted in [23] that the basis matrices obtained by Eq. (6.53) also maximize the mutual information as well as minimize the average pairwise error probability defined in Eq. (6.39). Actually, we can observe that the Gohary's optimization problem defined by Eq.(6.53) is equivalent at high SNRs to the Hassibi's optimization problem in Eq.(6.51) where a strongest constraint is imposed. At low SNR ranges, the Gohary's formulation penalizes more the bad equivalent channel events, thus more robust codes are obtained.

6.4. Exploiting (Partial) Transmit Channel Knowledge

It has been already stated in Section 6.1 that in case the transmitter has perfect channel information knowledge, the SNR at the receiver is maximized if the transmitter applies all the power over the dominant eigenmode of the channel. Furthermore, in order to increase the spatial multiplexing rate it may be preferable to transmit over all the non-zero eigenmodes of the channel allocating to each mode a power obtained following the water-filling algorithm described in Section 6.1 [11].

However, both beamforming techniques (dominant/multiple eigenmode transmission) require that the transmitter knows perfectly the channel state, and that the channel doesn't change during a sufficiently large period to estimate the channel and to apply the beamforming. In consequence, the beamforming might be only applied for low mobility scenarios and where the channel can be estimated (and fed back if necessary) with high reliability.

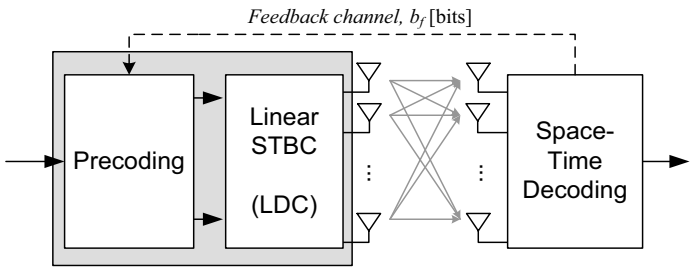


Figure 6.6. Linear space-time precoding.

In order to obtain the channel information at the transmitter, the receiver may sent the measured channel information through a feedback channel (close loop), or take advantage of the channel reciprocity in case of TDD transmission (open loop). However, even we use a

channel reciprocity it will consume bandwidth since some resources need to be allocated for the channel estimation. Furthermore, in open loop techniques a complex calibration of the transmitter and the receiver chains are required in order to equalize the RF hardware differences [20]. In consequence, closed loop schemes have gained more attention especially in the design of techniques that allow reducing the feedback load (quantizing channel information, channel condition number, Demmel condition number, channel rank, etc.). Furthermore, as it is depicted in Figure 6.6, the schemes where the input symbols are arranged/modified according to the channel status are known as precoders. The precoder then increases the number of available LDC codes by introducing certain changes according to the CSI status. Actually, both precoding and space-time block coding can be considered into the same block where given a set of codes (defined by the codebook) one of them is selected each *Time Transmission Interval* (defined by T or multiples of) according to the b_f feedback bits.

6.4.1. Transmit (and Receive) Antenna Selection

Besides the increase of the capacity or the reliability by any of the before mentioned precoding techniques (beamforming, codeword selection based on finite codebooks, etc.), the simplest precoding technique is to select which antenna (or subset of antennas) should be used according to any optimization criterion (capacity, reliability, etc.). Since antenna selection aids to reduce the hardware cost as well as the signal processing requirements, it is very well fitted to be applied in the mobile stations where space and cost must be seriously taken into account. Obviously, the reduction of the number of antennas reduces the array gain, however when the channel in any of these antennas is experiencing a deep fade, the capacity loss by not using such antenna is minimal. In consequence, antenna selection at both transmitter and receiver helps in reducing the implementation costs while retaining most of the benefits of MIMO technology [21].

The MIMO system model considering antenna selection is depicted in Figure 6.7, where M and N are the number of transmitter and receiver RF chains respectively, whereas the available antennas are referred by M_a and N_a for transmitter and receiver respectively (with $M \leq M_a$, $N \leq N_a$).

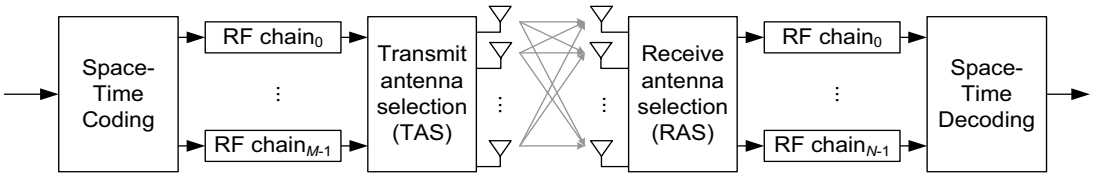


Figure 6.7. Antenna selection in MIMO systems with M_a available transmit and N_a receive antennas.

For the SIMO case, it is shown in [11] and [21] that the array gain using a Maximum Ratio Combiner (MRC) when no Receive Antenna Selection (RAS) is applied is equal to $g_a = N_a$, whereas when RAS is applied (e.g. the antenna with better channel is selected) the array gain is given by

$$g_a = N\left(1 + \sum_{j=N+1}^{N_a} \frac{1}{j}\right). \tag{6.54}$$

As a result, we can note that RAS implies a loss in the SNR which becomes larger as the difference between N and N_a is increased. However, the diversity order for both schemes is exactly the same and there is only a coding gain difference [21]. For the MISO case, the same conclusions are obtained when a Transmit Antenna Selection (TAS) is applied.

However, for the MIMO case, transmit and/or receive antenna selection has further implications than just a reduction of the array gain. Actually, the previous spatial multiplexing-diversity tradeoff leads to different optimization criteria: diversity optimization (i.e. select the set of antennas that gives a higher Frobenius norm of the channel), improve the link reliability (discard antennas that produces large fadings in any eigenmode), maximize the Shannon capacity, etc. Furthermore, in [21] it is stated that the diversity gain obtained by transmit antenna and receive antenna selection is the same as without selection, hence $g_d=(N-g_s)\times(M-g_s)$ with $g_s=\{0,\ldots,\min(N,M)\}$.

6.4.1.1. Transmit Antenna Selection in MIMO systems

Transmit antenna selection techniques were first proposed during the very late 90s in the context of MIMO links in order to improve the array gain. During that period the antenna selection was derived according to the class of ST coding scheme that was involved. Heath et al. in [24] focused on the antenna selection in case of Spatial Multiplexing (i.e. V-BLAST) for linear receivers. The optimization criterion in [24] was to maximize the post-processing SNR in order to minimize the bit error probability. Later it has been shown that the difference between optimizing the $ESINR$ is only 0.5dB better than optimizing the lowest eigenvalue (the ESINR in case of Zero-Forcing is lower bounded by the minimum eigenvalue of $\mathcal{H}$ see Eq. (6.47)). Similar works have been carried out for OSTBCs which in this case concluded that maximizing the Frobenius norm of the active channel was the optimum strategy [25]. More recently, Deng et al. extended these transmit antenna selection schemes under the LDC framework concluding that the best selection criteria for minimize the bit error probability is based on maximizing the post-processing (or Effective) SNR.

Finally, an interesting application of transmit antenna selection has been proposed by Freitas et al. in [27] where different spatial layers are assumed combining spatial diversity and spatial multiplexing. In [27] the different branches are disposed in parallel hence both spatial diversity and multiplexing gains can be simultaneously achieved. The antennas subsets are then assigned to the spatial layers in order to minimize the bit error probability, where the (more susceptible) SM based layers are first assigned their subset of antennas and the remaining are assigned to the OSTBC layers.

6.4.2. MIMO precoding based on LDC codes

In the previous section, TAS precoding scheme has been introduced for some of the existing STBC and LDC. Therefore, given a specific code, the number of bits that must be fed-back from the receiver f_b to indicate the optimum transmit antennas set is

$$f_b = \binom{M_a}{M} = \frac{M_a!}{M!(M_a - M)!}. \tag{6.55}$$

However, if we can afford for sending few more bits over the feedback channel, the transmitter/receiver may also be able to select which code is more suitable according to the current channel state, or to choose how many spatial streams can be transmitted according to the channel rank. One of the first research works deriving the optimal space-time coding strategy (diversity or multiplexing) has been carried out by Heath and Paulraj in [28] where only one feedback bit (f_b=1) is required per narrow-band MIMO channel. It has been demonstrated that based on the minimum Euclidean distance (minimum distance between constellation points at the receiver) it is possible to define which code is better in terms of the bit error rate. In [28] a new channel metric named the *Demmel condition number* $\mathcal{K}_D$ is defined in [28] as follows

$$\mathcal{K}_D = \frac{\|\mathbf{H}\|_F^2}{\lambda_{\min}(\mathbf{H})}. \tag{6.56}$$

Then, defining the minimum distance between constellation point as $d_{min,sm}$ and $d_{min,sd}$ for the spatial multiplexing (SM, e.g. VBLAST) and spatial diversity (OSTBC, e.g. Alamouti) codes respectively, it is shown that the SM scheme is preferred if

$$\mathcal{K}_D \le \frac{d_{\min,sm}}{d_{\min,sd}}. \tag{6.57}$$

where the minimum distance of each constellation is a function of the modulation scheme and the rate. Furthermore, whether the Frobenius norm of the channel $\|\mathbf{H}\|_F^2=1$ the SM is preferred if

$$d_{\min,sd} \le \lambda_{\min}^2(\mathbf{H}) d_{\min,sm}, \tag{6.58}$$

where $\lambda_{min}^2(\mathbf{H})$ is the minimum eigenvalue of $\mathbf{H}$. Hence the multiplexing is preferred as long its minimum scaled distance remains greater than the minimum distance of the diversity constellation.

Recent researches have extended the space-time coding selection (i.e. codebook based precoding) into the LDC framework [29]-[34]. In [29], Che et al. proposed to select the LDC that maximizes the differential entropy according to the instantaneous channel realization, thus the optimum LDC_i is given by

$$\max_i \left(u\left(\sqrt{\frac{\rho}{M} Tr\left\{ \mathcal{H}_i \mathcal{H}_i^{\mathrm{H}} \right\}} \right) \right) \tag{6.59}$$

where $u(\cdot)$ denotes the differential entropy. Nevertheless, the optimality of this approach relies on the design of the codebook. In order to do so, most authors rely on the Grassmanian subspace packing approach, where the set of the precoding matrices are obtained in order to maximize the minimum distance between any pairs of linearly precoded dispersion matrices. Furthermore, it is demonstrated in [31] that $f_b=log_2(M)$ feedback bits are enough to achieve full diversity. Other selection criteria have been proposed such as the maximization of the minimum singular value decomposition

$$\max_i \lambda_{\min}(\mathcal{H}_i), \tag{6.60}$$

where $\lambda_{\min}(\cdot)$ denotes the minimum singular value of a matrix [32], the minimization of the pairwise error (defined in Eq.(6.39) and Eq. (6.48) for the ML and MMSE receivers respectively) [33], etc. Yang et al. in [34] shown that 2dB improvement on the average SNR can be achieved compared to the open loop scheme with only 3 feedback bits (i.e. with 8 basis matrices).

6.5. Proposed Transmit Antenna and Code Selection (TACS) scheme

As it has been explained in the previous Section 6.4, when partial CSIs information is available at the transmitter two common selection techniques could be applied which are: the space-time code selection, and the transmit antenna selection. One of the first works joining both concepts is that presented by Heath et al. in [35] where the number of the spatial streams (in the SM case) are adapted by selecting the best set of transmitter antennas (i.e. f_b=M). Furthermore, it was stated that if the optimum number of streams are transmitted from the optimum selected antenna set, the diversity gain is also maximized ($g_d \leq MN$). The selection criteria in Eq. (6.57) are then extended to include the transmit antenna selection leading to

$$\min_p P_e = \min\left\{1-\mathrm{E}\left(1-N_e\left(Z\right)\cdot \mathrm{Q}\left(\sqrt{ESINR_p\,\frac{d^2_{\min}\left(Z_p\right)}{2}}\right)\right)^{Q_p}\right\}, \tag{6.61}$$

$$\min_p ESINR_p d^2_{\min}\left(Z_p\right), \tag{6.62}$$

where Eq. (6.62) is a simplified version of Eq. (6.61), and p denotes the antenna subset. Then, given an antenna subset and a fixed rate, the required constellation could be determined as well as the number of spatial streams. A simplification of this optimization problem is given in [36] where each stream is switched on/off when the post-processing of the SNR value of the stream is above/below a fixed threshold which is related with the rate. Further extensions of space-time code selection with TAS are given by Machado et al. in [37] where the available codes in the codebook are; the Alamouti code, the SM with M=2, the Quasi-OSTBC with M=3 and single antenna transmission.

In addition, the space-time code selection with transmit antenna selection has been generalized by the author in [38][39][40] under the LDC framework considering both the linear and the ML receivers and developed within the IEEE 802.16m framework [41]. This generalization allows us to use any type of linear STBC (independently of the optimization criteria) and determine which codes are used most times and under which channel conditions. Two optimizations criteria are developed by the author, one following the classical bit error rate optimization (minimizing of the scaled minimum Euclidian distance), and a second one is based on the throughput maximization given a fixed link quality (i.e. fixed packet error rate or bit error rate). This second optimization criterion can be used for resource allocation and scheduling purposes as it will be further discussed. Nevertheless, it is also shown that for low multiplexing rates the classic STBC codes (i.e. Alamouti, SM and Golden) with transmit antenna selection are sufficient to explore the Grassmanian subspace.

In the following subsections, the spatial adaptation schemes based on transmit antenna selection and LDC selection (referred as Transmit Antenna and Code Selection – TACS) is explained, where the performance results obtained through computer simulations are analyzed.

6.5.1. The TACS Selection Criteria

Given the $ESINR$ per stream in Eq. (6.41) and Eq. (6.46) for the ML and linear receivers respectively, and the average pairwise error probabilities in Eq. (6.40) and Eq. (6.48), two

different code and antenna subset optimization scenarios, namely *Minimizing the bit error rate* and *Maximizing the throughput*, are evaluated showing that in both the diversity order and the system throughput are maximized under each respective constraints.

In the first scenario, we consider that the same modulation is applied to all the symbols with a fixed rate R. In that case, and since transmission power is also fixed, we are interested in selecting the transmit antenna subset and the LDC code that minimizes the error rate probability (i.e. the bit error rate – BER) while the modulation that is required by each LDC is adapted in order to achieve the required rate R. In that case, since the *Q*-function is monotonically decreasing as a function of the input, the optimization problem is defined as follows

$$\max_{LDC_i, p_i} \min_{q} \left\{ ESINR_q \left(H, LDC_i, p_i \right) d^2_{\min} \left(Z_i \right) \right\}, \tag{6.63}$$

where i means the LDC index, p_i denotes the transmitting antenna subset (set of antennas that can be used according to the number of transmitter antennas M_a and the number of antennas required by the LDC) and q refers to the spatial stream (i.e. the symbol) index. It is also noted that the constellation is a function of the LDC.

In the second scenario, the optimization is performed in order to maximize the system throughput considering a certain quality of service requirement (i.e. a maximum Block Error Rate - BLER). In that case, the problem is formulated as follows

$$\max_{LDC_i,\, p_i,\, MCS_j} \min_{q} R\left(1 - BLER\left(ESINR_q\right)\right) \quad \text{s.t.: } BLER \le \mu \tag{6.64}$$

where j means the Modulation and Coding Scheme (MCS) index that maximizes the spectral efficiency for the specific channel state subject to a maximum Block Error Rate (BLER). Actually, the selection of the optimum MCS is carried out assuming that the $ESINR$ is the SNR that would be obtained at the receiver in case having an Additive White Gaussian Noise (AWGN) channel. Under that assumption, there is a direct mapping between each MCS and the obtained BLER for each $ESINR$ (this is usually implemented by a Look-Up-Table search).

The following Figure 6.8 illustrates the scheme of the MIMO system applying the TACS selection algorithm.

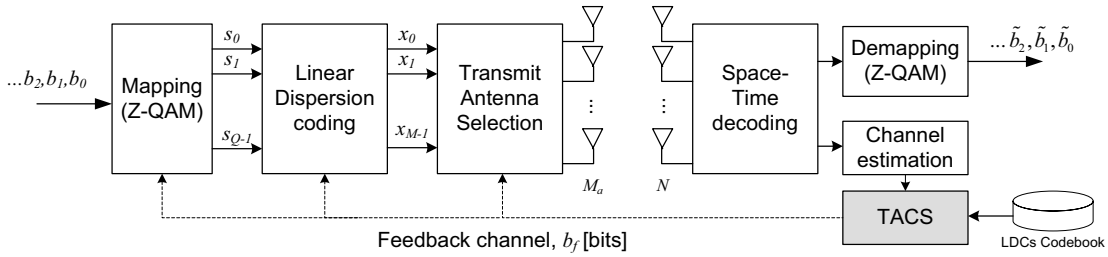


Figure 6.8. Proposed TACS spatial adaptation scheme and integration into the transmission scheme.

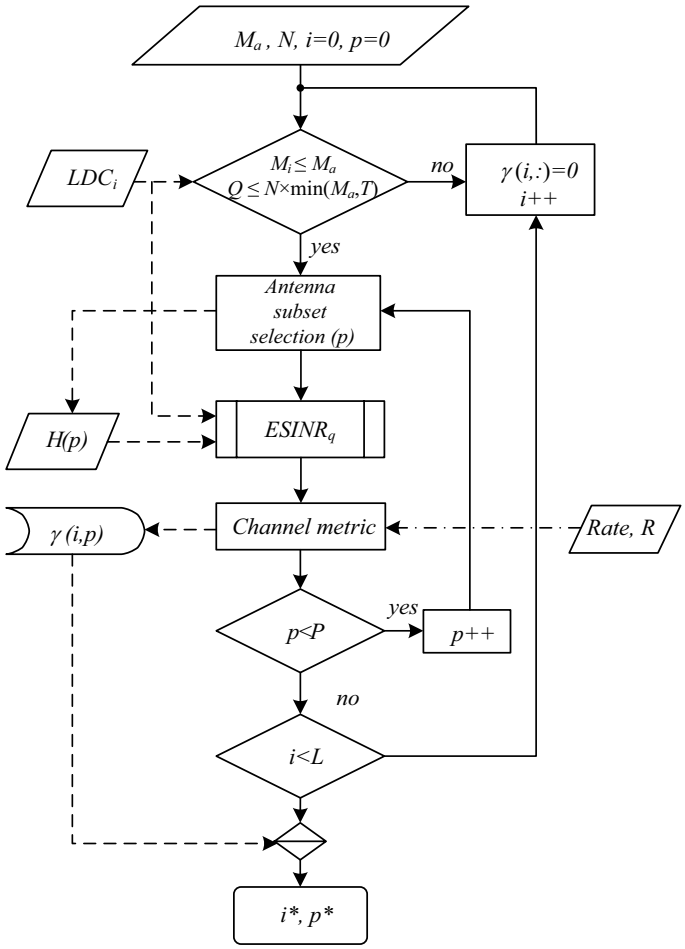


Figure 6.9. TACS flow chart algorithm.

6.5.2. TACS Exhaustive Search algorithm description

This section describes the algorithm proposed for transmit antenna and LDC code selection. An Exhaustive Search (ES) is performed over the whole combination of transmit antennas and for each available LDC. The flowchart of the Transmit antenna and space-time Code Selection (TACS) is shown in Figure 6.9 and detailed here below in Table 6.2.

The *ESINR* depends on the type of used receiver. On the other hand, the channel metric block gives a value $\gamma(i,p)$ which is a function of the code, the antenna subset and the optimization criteria (Eq. (6.63),(6.64)). The exhaustive search method gives the optimum combination, however computational cost is relatively quite high. Other (non-optimal) searches as the incremental/decremental search algorithms have been proposed by Gershman et al. in [21]. The feasibility of these lower complexity schemes in the TACS context is left for further studies.

In order to be able to apply the previously described MIMO space-time block codes into the BWA systems described in Section 2.2, the author assumes that the channel behaves as narrow-band and stationary over the transmitted signal (i.e. $\mathbf{H}$ is constant in frequency and time at least during one codeword). Recovering the frequency-time OFDM transmission grid introduced in Section 2.2.2.1, in case of multiple transmit antennas we have one OFDM grid per transmitter antenna. The main question is then: how to map the symbols in the grid/frame in order to achieve the maximum diversity while at the same time guarantee that the MIMO

channel is constant for the whole codeword? The straight away scheme is to consider each subcarrier as an independent subchannel where the codewords are mapped independently as depicted in Figure 6.10a. This scheme is suitable as long as the time coherence is (approximately) one order of magnitude higher than the codeword length. On the other hand, the mapping can be also carried in frequency direction in case the codeword does not spans more than a small fraction of the coherence bandwidth as represented in Figure 6.10b. In most BWA with multiple antenna support a combination of both is applied, where the symbols belonging to each codeword are arranged in both frequency and time directions as in Figure 6.10c. This last scheme is the more flexible since larger codewords (i.e. high values of T) can be allocated while the necessary condition of flat and stationarity MIMO channel can still be achieved despite frequency or time selectivity.

TACS exhaustive search algorithm:

Start of the algorithm:
for i=1:L (L means the number of LDC codes) **then**
  if the code can be used given the number of available transmitter antennas M_a and the receiver antennas **then**
    for p=1:P (P means the number of possible antenna combinations given a certain LDC and the number of available transmit antennas) **then**
      compute the $ESINR_q$ per each q-th stream according to the receiver detector and channel **H**
      if optimization criteria is *BER minimizing* **then**
        compute the constellation size Z according to the required rate R and the LDC code i
        $\gamma(i,p) = \min_q \left\{ ESINR_q \cdot d^2_{\min}\left(Z_i\right)\right\}$
      Else
        $\max_{MCS_j} \min_q R\left(1 - BLER\left(ESINR_q\right)\right)$ s.t.: $BLER \le \mu$
      end if
    end for
  Else
    $\gamma(i,p)=0$ for $\forall p$
  end if
end for
obtain the code i^* and antenna subset p^* such that: $\max_{i^*,p^*} \gamma\left(i,p\right)$
End of the algorithm.

Table 6.2. TACS algorithm description.

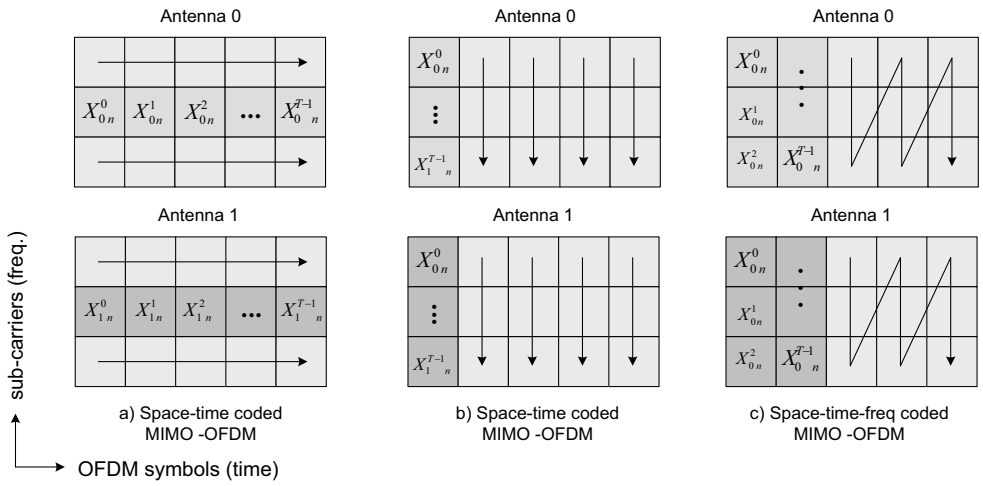


Figure 6.10. Symbol mapping in space-time coded MIMO-OFDM and M=2.

6.5.3. Simulation Results

In order to evaluate the performance of the TACS scheme, the uplink of the IEEE 802.16 standard is used following the simulation parameters depicted in Table 6.3, where *MS*s are assumed perfectly synchronized and there is only one cell (no inter-cell interference). The employed modulation is a Z-QAM (Z={2,4,16,64}) with Gray mapping where all the symbols within a codeword map the same number of bits (or otherwise stated).

According to the CSI measured the BS determines *i*) the antenna subset, *ii*) the LDC subset and (optionally) *iii*) the MCS that maximizes the rate for a maximum Block Error Rate - BLER (second optimization criterion). The codebook is composed mainly of the Single Input Multiple Output (with Maximum Ratio Combining at the receiver), the Alamouti 's Spatial Diversity (SD) coding scheme (referred as G2 in below plotted figures), the SM (pure spatial multiplexing) and the Golden code. It is expected that the diversity order of the TACS scheme is equal to the maximum available ($g_d = N \cdot \min(M,T)$).

The performance of the system is evaluated over 100.000 channel realizations, where for each realization a tile or a subchannel (specified in each analysis) is transmitted. In case of Partial Usage Subcarrier permutation (PUSC), the tile is formed by 4 subcarriers and 3 symbols, where 4 tones are dedicated to pilots as defined in IEEE 802.16e [19]. For the Band Adaptive Modulation and Coding (AMC) permutation scheme, each bin (equivalent to the tile concept) is comprised by 9 subcarriers where 1 tone is used as pilot. Perfect channel estimation is assumed at the receiver. Every $log_2(Z)$ bits are mapped to one symbol. The channel models used are uncorrelated Rayleigh ($\mathbf{H} \sim \mathbb{C}\mathcal{N}(0,1)$) and the ITU Pedestrian A [41]. In both cases the channel is considered constant within a tile (*block fading channel model*). In case of uncorrelated Rayleigh the channel between tiles is uncorrelated, whereas in the ITU PedA case the channel is correlated in frequency and time between consecutive tiles.

OFDMA Air Interface and System Level configuration	
Subcarrier Permutation	Distributed (PUSC) and Contiguous (Band AMC)
FFT length, CP	2048, 12.5%
# of used subcarriers	1728
Modulation	{4,16,64}-QAM
Channel coding	Turbo coding with rates: 1/3, 1/2, 2/3, ¾
Channel model	Rayleigh and ITU Pedestrian A
Channel estimation and CQI	Ideal
Frame duration, T_{frame}	5ms
DL/UL rate	2:1
OFDM symbols in the DL	30

Table 6.3. TACS evaluation framework system parameters

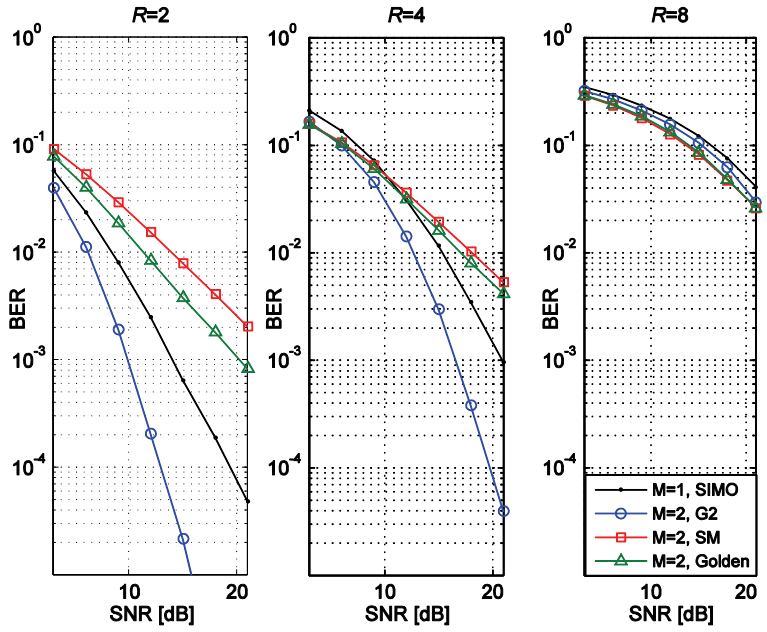


Figure 6.11. BER for uncorrelated Rayleigh channel with MMSE detector and N=2.

In Figure 6.11 the reference performance for a fixed rated is depicted for N=2 when no transmit antenna selection neither code selection are used. For uncorrelated Rayleigh channel, we can observe that for low data rate, i.e. R={2,4}, the Alamouti code outperforms the rest of the schemes. This is strictly related to the diversity order that G2 achieves equal to g_d=N×M=4, whereas the SM and the Golden code with a linear receiver get a diversity order of g_d=($N-M+1$)=1 . At higher data rates (R>8), all the codes perform similarly in the analyzed SNR range despite of the different diversity order between them.

Hereafter the instantaneous post-SNR (or ESINR) is analyzed. In Figure 6.12, the cumulative density function of the instantaneous ESINR using the MMSE receiver is depicted. Clearly, it can be observed that the instantaneous ESINR per stream using the SM and the Golden codes might be up to 10dB lower than for the SIMO and the Alamouti (i.e. G2) codes. This difference in the instantaneous post-SNR can be explained by the factors of: *i*) the higher number of streams or symbols per codeword (Q=4 for SM and Golden codes), and *ii*) the effects of non-orthogonality. However for the proposed evaluation metric $min\{ESINR_q d_{min}^2\}$ (see Eq.(6.62))

which includes the effect of the symbol mapping, the difference between the codes is shortened and for high data rate (e.g. R={4,8}) the SM and the Golden code codes may achieve higher metrics than the SIMO and G2 codes. For Z-QAM modulation, where $Z=2^{RT/Q}$, the minimum Euclidian distance between any two constellation points is depicted in Table 6.4. Combining both the statistics of the $ESINR_{min}$ and the minimum distance values in Table 6.4, we can intuit which scheme will be selected more and in which conditions.

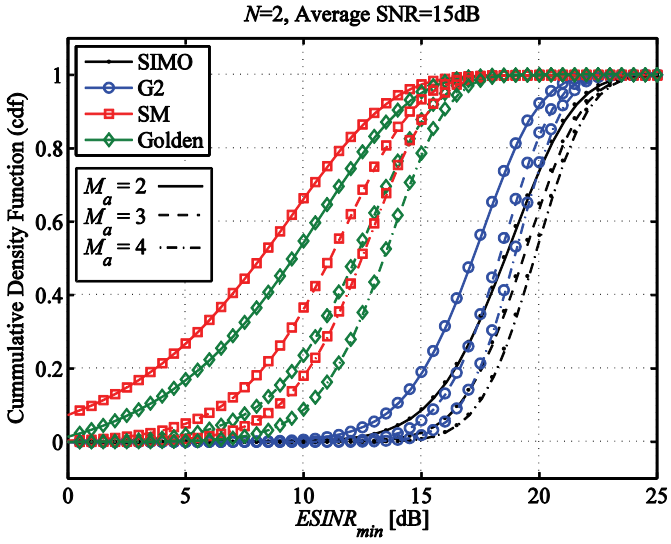


Figure 6.12. Cumulative density function of the minimum instantaneous ESINR for an average SNR = 15dB, N=2, uncorrelated MIMO Rayleigh channel and the MMSE detector.

	R=2	R=4	R=8
SIMO (Q=T=2)	4, +3	16, -4	256, -16.3
G2 (Q=T=2	4, +3	16, -4	256, -16.3
SM (Q=4,T=2)	2, +6	4, +3	16, -4
Golden (Q=4,T=2)	2, +6	4, +3	16, -4

Table 6.4. Constellation size and minimum Euclidian squared distance pairs (Z, d_{min}^2[dB]) per LDC and rate R.

6.5.3.1. TACS performance under bit error rate minization criterion.

In Figure 6.13 and Figure 6.14, the bit error rate performance with the TACS algorithm described in Figure 6.9 is studied having a fixed rate R=4. Figure 6.13 shows the improvement due to the increase in M_a and also the performance achieved when combined with code selection. It can be observed how the TAS increases the diversity order, leading to a large performance increase for the SM and Golden subsets. It is very important to notice that despite the diversity increase for all the LDC subsets, SD and SIMO schemes still perform better when each code is evaluated independently. However, in Figure 6.14, we can observe that when the code selection is switched on, SIMO and Golden subsets are selected most times, while the usage of SIMO increases with the SNR and the usage of SM and the Golden code increases with M_a. Furthermore, the achieved improvement by the TACS is clearly appreciated in Figure 6.13, where an SNR improvement of approximately 1dB is obtained for M_a={3,4}. It is also surprising that the SM code is rarely selected knowing that the Golden code should always

outperform SM since it obtains a higher diversity. However, as it is observed in Figure 6.14, for less than 5% of the channel realizations the SM may outperform slightly the Golden code. Whether the singular value decomposition of the effective channel $\mathcal{H}$ is analyzed when SM is selected, it has been observed that when all singular values are very close, both the SM and the Golden code lead to very similar performances, therefore no matter which one is selected.

In Figure 6.15 and Figure 6.16, the performance using the TACS is again analyzed for R=8. In Figure 6.15 the different diversity orders of SD, SM, and the Golden Code are illustrated. We can appreciate here that the SM and the Golden code show the best performance when M_a={3,4}, and also for M_a=2 when SNR≤18dB. Furthermore the increase in the diversity order due to TACS can be observed in both Figure 6.13 and Figure 6.15. The maximum diversity order ($g_d = M_aN$) is achieved since at least one LDC (SIMO and G2) from those in the codebook are able to achieve the maximum diversity order. Moreover, the BER using the TACS is equivalent to that obtained from the SISO scheme (referred as $SISO_{eq}$ in the plots) over a Rayleigh fading channel with the same rate R, a diversity order g_d=M_aN and a coding gain equal to Δ. The performance of this *equivalent* SISO scheme, in terms of the bit error rate probability P_b, can be obtained directly by close expressions that are found in the scientific literature [42][43] and applying the Craig's formula in [44],

$$P_b = \frac{1}{\log_2\sqrt{Z}} \sum_{i=1}^{\log_2\sqrt{Z}} P_b(i) \tag{6.65}$$

$$P_b(i) = \frac{2}{\sqrt{Z}} \sum_{k=0}^{\left(1-2^{-i}\right)\sqrt{Z}-1} \omega(k,i,Z) \frac{1}{\pi} \int_0^{\pi/2} \left(1+(2k+1)^2 \frac{3\rho_b}{2(Z-1)\sin^2\theta} \right)^{-g_d} d\theta \tag{6.66}$$

$$\omega(k,i,Z) = (-1)^{\left\lfloor \frac{k\cdot 2^{i-1}}{\sqrt{Z}} \right\rfloor} \left(2^{i-1} - \left\lfloor \frac{k\cdot 2^{i-1}}{\sqrt{Z}} + \frac{1}{2} \right\rfloor \right) \tag{6.67}$$

where $\rho_b = \Delta \cdot \rho / log_2(Z)$, $\lfloor x \rfloor$ means the smallest integer of x, and Z is the modulation order of the Z-QAM modulation.

Δ		N=2	N=3	N=4
R = 4	M_a=2	2.66	3.9	6.31
	M_a=3	3.20	5.2	8.41
	M_a=4	3.75	6.2	9.44
R = 8	M_a=2	4.20	9	14
	M_a=3	6.75	14.5	23
	M_a=4	9.00	19	28.5

Table 6.5. Coding gain Δ for the TACS proposal with M_a={2,3,4}, N={2,3,4}, and R={4,8}.

The value of Δ for different combinations of M_a={2,3,4}, N={2,3,4} and R={4,8} are depicted in Table 6.5. These values have been obtained adjusting the BER approximation in Eq. (6.65) to the empirical BER. As shown in Figure 6.13 and Figure 6.15 the performance of the TACS schemes is perfectly parameterized under the equivalent SISO model. Notice also that the power gain is constant across the whole SNR range.

In Figure 6.17, the linear MMSE and ML detectors performances are compared. Since the ML achieves higher diversity order than the MMSE detector, the performance is clearly superior. However, the computational complexity of the ML is exponential with R (and T). As a result,

for data rates R>4 it is unfeasible to implement ML in a practical system. It is also observed that the Golden code and the SM are the most benefited from the ML detector since the diversity order is increased, leading to lower BER. In case of ML decoding, the superiority in performance of the Golden code w.r.t. SM is evident. When comparing the Golden code and SD, it is observed that for data rates $R\geq4$ the Golden code outperforms SD. This can be attributed to the fact that in case R=4 the modulation used with the Golden code is a 16QAM whereas for SD a 256QAM is required. So, although the SNR for each stream is higher with the SD, the BER is increased due to the higher order modulation. Furthermore, in Figure 6.18 the benefits of antenna selection can be observed with the ML decoder. The LDCs that benefit more from TAS are then the Golden code and the SM. Also the improvement achieved by using the code selection criterion before proposed is established. The performance obtained by the ML+TACS combination and the MMSE+TACS combination are less than 0.5dB away, to observe this small difference the equivalent SISO performance with the MMSE detector is plotted for reference purposes. It can be concluded from the obtained results that the MMSE is from a computational cost point of view an affordable solution that, when combined with TACS scheme, gives a quasi-optimal solution.

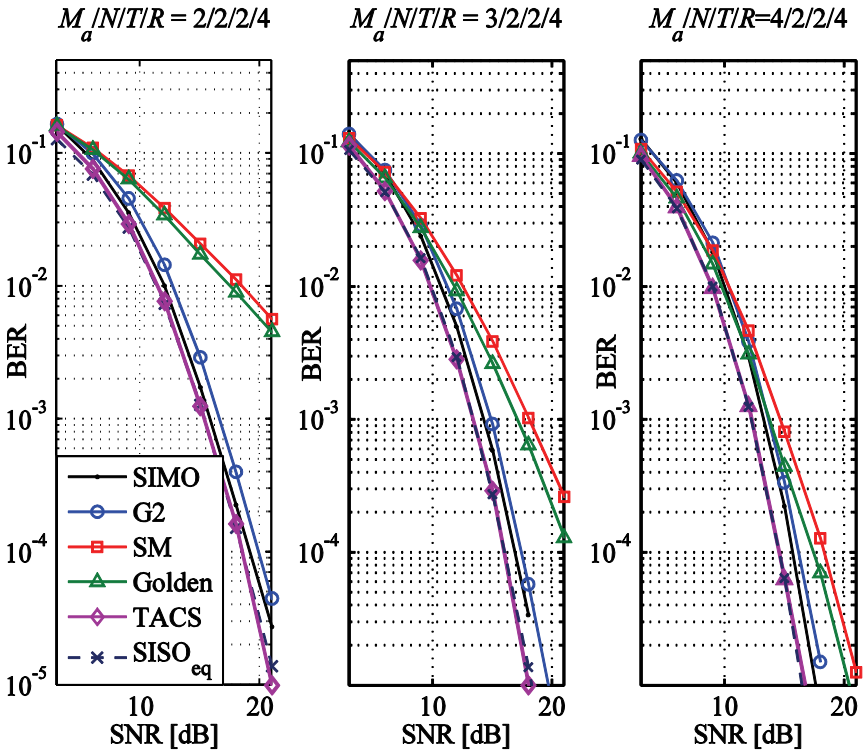


Figure 6.13. BER performance when N=2, R=4, M_a={2,3,4} for uncorrelated MIMO Rayleigh channel and MMSE linear receiver.

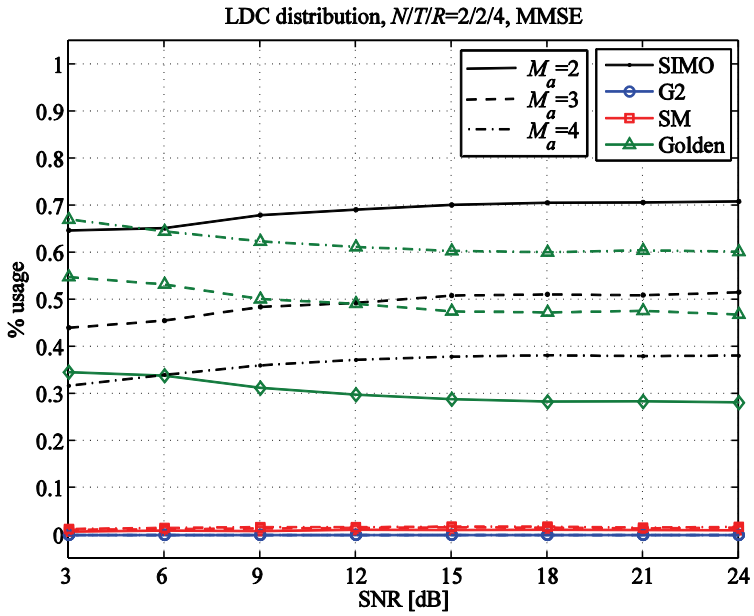


Figure 6.14. LDC selection statistics when N=2, R=4, M_a={2,3,4} for uncorrelated MIMO Rayleigh channel and MMSE linear receiver.

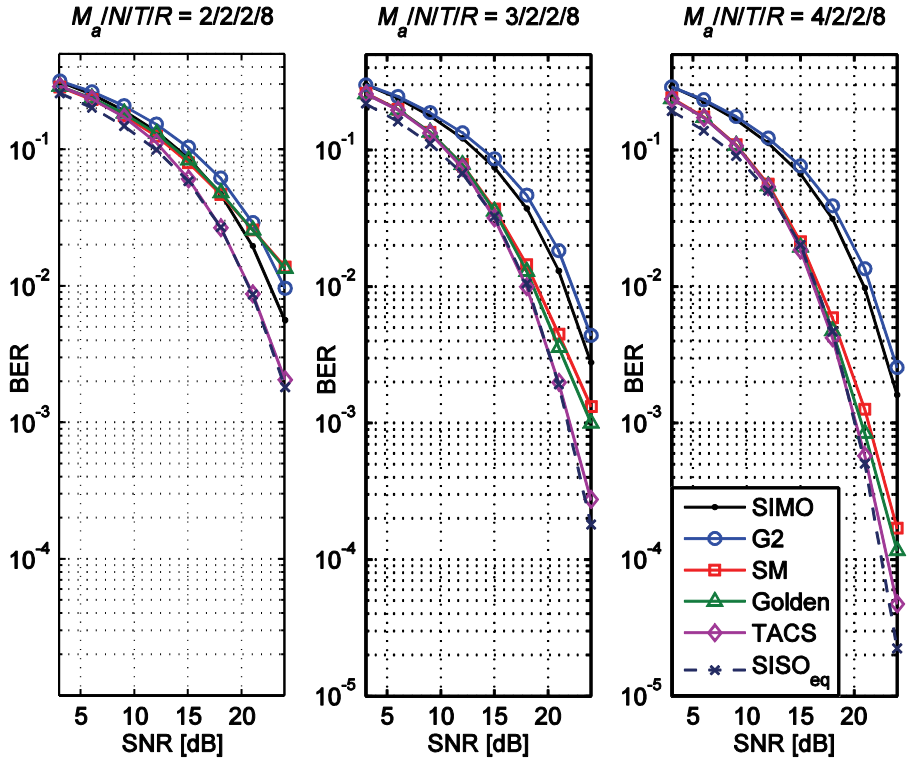


Figure 6.15. BER performance when N=2, R=8, M_a={2,3,4} for uncorrelated MIMO Rayleigh channel and MMSE linear receiver.

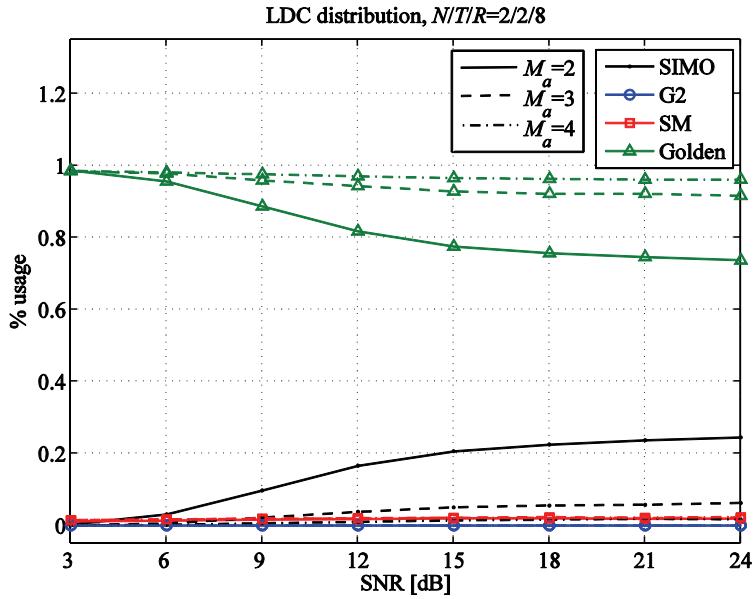


Figure 6.16. LDC selection statistics when N=2, R=8, M_a={2,3,4} for uncorrelated MIMO Rayleigh channel and MMSE linear receiver.

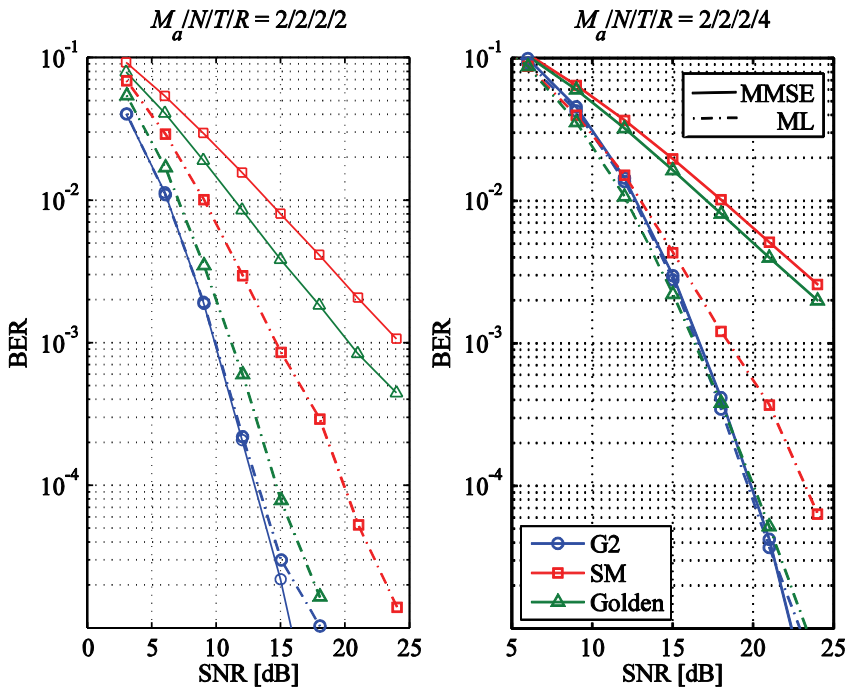


Figure 6.17. BER performance with M_a=2, N=2, R={2,4} and uncorrelated MIMO Rayleigh channel, for MMSE and ML detectors.

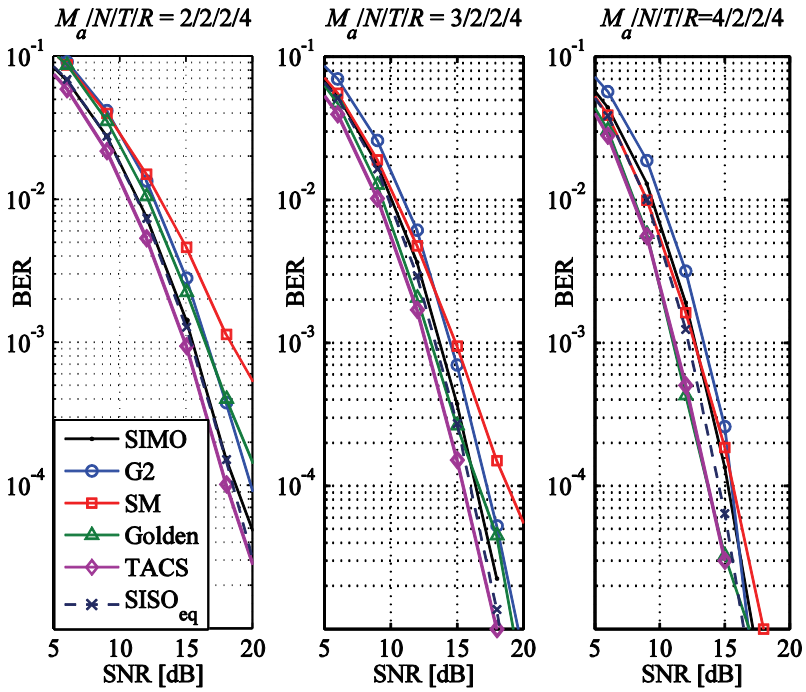


Figure 6.18. BER performance with $M_a=\{2,3,4\}$, $N=2$, $R=4$ and uncorrelated MIMO Rayleigh channel for the ML detector.

Previous results have been obtained by optimizing the antenna and the code subset for one tile. Usually it is not possible to optimize over such small radio resource unit. For instance, in the IEEE 802.16 PUSC scheme the minimum resource unit is a subchannel, which is formed by 6×1 tiles in frequency and time respectively, whereas for the Band AMC scheme one subchannel may be comprised by {1×6, 2×3 or 3×2} bins (equivalent to a tile) in frequency and time respectively [19]. According to the permutation scheme applied, i.e. PUSC or Band AMC, the channel between the different tiles within a subchannel may be uncorrelated (PUSC case) or correlated (Band AMC).

In Figure 6.19 and Figure 6.20 the effects of optimizing over a larger number of tiles is investigated. Using an MMSE linear receiver, it can be appreciated in Figure 6.19 that the TACS scheme performs better in case of correlated channels, something obvious since it is based on a "*maxmin*" cost function and as the degrees of freedom in the optimization set are increased (i.e. the PUSC scheme with higher number of tiles) the minimum channel value across the set also decreases. The performance difference between the TACS for correlated and uncorrelated channels is between 1dB for M_a=2, and 3dB for M_a=4. Nevertheless, looking to Figure 6.20 may clarify the difference in performance for both channels. When we observe the LDC usage distribution in the correlated channel (Figure 6.20 - right plot), we can conclude that the TACS is not so much affected by the optimization over a larger set of tiles. However, in case the channel is uncorrelated we can observe how the spatial diversity scheme (G2) becomes the most frequently used LDC code as the number of tiles per subchannel is increased. This result could be also verified, since for uncorrelated channels with a large number of tiles it is not clear that one antenna subset may perform well for all the tiles, so the best choice is to use the scheme with higher native diversity (i.e. higher diversity without transmit antenna selection), and this is the G2. So we can conclude from the above analysis that the TACS proposed scheme

is aware of channel correlation inside each tile, and also between tiles, selecting the optimum code accordingly.

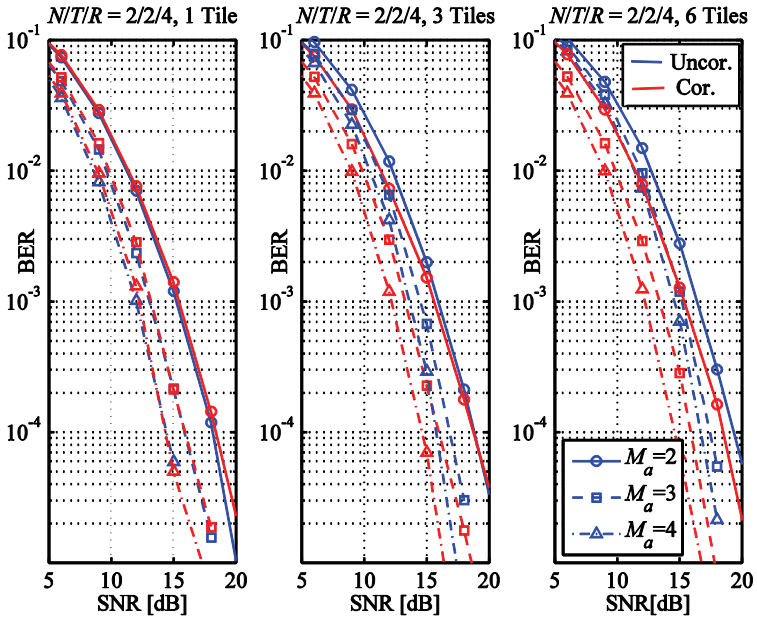


Figure 6.19. Effects of optimization over broadband MIMO channels (MMSE detector).

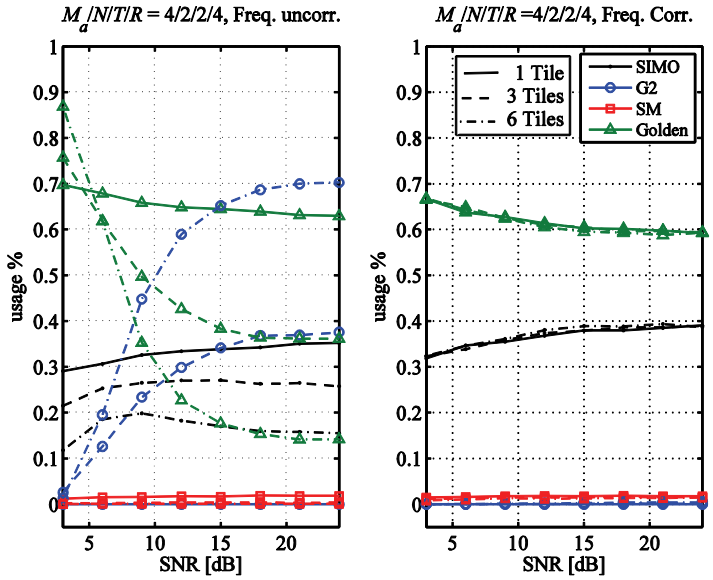


Figure 6.20. Usage of the LDC codes when optimizing over broadband MIMO channels.

Moreover, during the previous analysis, one question arises through the evaluation the TACS adaptation scheme regarding the optimality of the codebook data set. We introduced in section 6.4.2 that schemes based on Grassmanian subspace packing are optimal in the sense that for a fixed number of feedback bits, the space-time code spans uniformly over the whole space of possible codes maximizing the distance between the different codes.

Hence, the optimality of the previous *basic codebook* (SIMO, G2, SM and Golden) must be checked to guarantee that the TACS scheme maximizes the capacity of the channel given such codebook. In order to do so, in this section the TACS algorithm under a fixed rate constraint is

tested with an *extended codebook* to evaluate the convenience of the proposed *basic codebook*. The extended codebook contains all the different block space-time codes standardized in [19][41] for M={2,3,4}.

In Figure 6.21, the performance of all the **A**, **B**, **C** space-time coding matrices is shown for the Rayleigh MIMO channel and a rate R=8. The **A** matrices are designed to maximize diversity whereas **B** matrices maximize multiplexing. Any of $\{\mathrm{A,B,C}\}\text{-}\mathrm{M}_y$ matrix is designed for M transmit antennas where the y sub-index means the permutation order of the matrix. Notice that the *A*2, *B*2, C2 are the same as those used in G2, SM, and the Golden code respectively. By introducing all these matrices into the selection algorithm, the performance of the TACS scheme is again depicted and compared to the performance obtained in previous analysis where only the *SIMO*, *A*2, *B*2 and the Golden code where used. The analysis has been repeated for $R<8$ leading to the same results. The main conclusion from this study is clear: the TACS scheme is optimal with the proposed basic codebook as long as $R\leq 8$.

Furthermore, the same analysis has been also carried considering other LDC optimized according to different parameters (channel capacity, pairwise error probability, etc. [11][12][22][23]) . The results are shown in Figure 6.22 where the performances of several LDC codes which are optimized according to different cost functions are compared (a MMSE detector is also applied). Again the benefit of increasing the codebook is lower than 0.2dB (in terms of SNR) which leads to the same previous conclusion. However, it is expected that as the rate is increased the relevance of space-time coding matrices with high spatial multiplexing rate r_s will be crucial, hence extended sets with codes with Q>4 will be required.

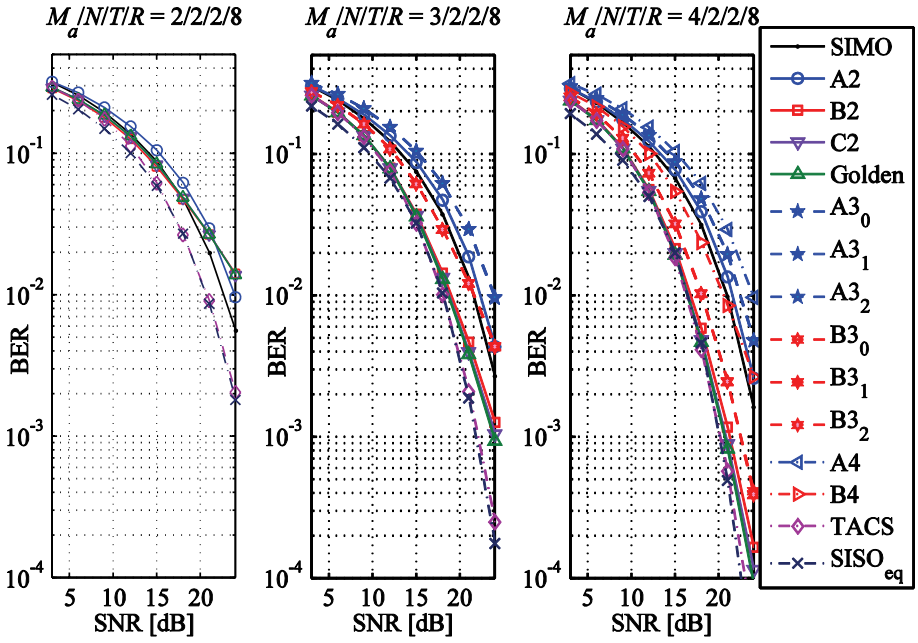


Figure 6.21. BER performance when N=2, R=8, M_a={2,3,4} for uncorrelated MIMO Rayleigh channel, MMSE detector and the codes in the IEEE 802.16 standard.

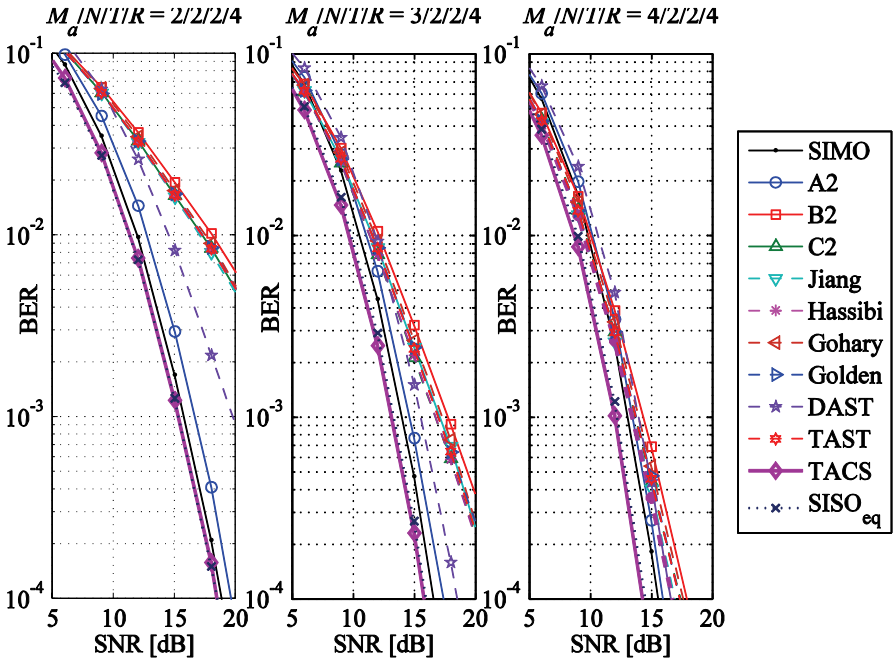


Figure 6.22. BER performance when N=2 and R=4 for uncorrelated MIMO Rayleigh channel, MMSE detector and differently optimized LDC codes.

6.5.3.2. TACS performance under Throughput Maximization criterion.

In this section the performance of the TACS adaptation scheme in case the throughput is maximized (see Eq. (6.64)) is analyzed. Then, for such adaptation scheme, the antenna set and the LDC code that maximizes the throughput is selected. In addition, the highest MCS (in the sense of spectral efficiency) that achieves a BLER<0.01 (1%) is also selected. The look-up-table used for mapping the ESINR to the BLER is included in Appendix C for reference purposes. In these simulations the minimum allocable block length according the IEEE 802.16e standard was selected [19] (i.e. the number of subchannels N_{sch} occupied per block varies between 1 and 4). The number of available antennas is M_a=2 whereas N=2. Linear detection with the MMSE and ML detection are also compared.

Then, from Figure 6.23 to Figure 6.27 the spectral efficiency achieved by TACS with adaptive Modulation and Coding (AMC) as well as the LDC statistics are shown. Two encoding schemes named *Vertical Encoding* (VE) and *Horizontal Encoding* (HE) are considered. For the first scheme, VE, the symbols within the codeword apply the same MCS format, whereas for the second, HE, each symbol may apply a different MCS. Clearly the first is more restrictive since is limited by the worst stream ($min(ESNR_q)$) whereas the second is able to exploit inter-stream diversity at the expense of higher signalling requirements (at least twice as that required with VE in case of M=2).

In case of MMSE receiver (Figure 6.23 to Figure 6.25), it is shown that at low SNRs (SNR<13dB) the SIMO and Alamouti achieved the highest spectral efficiencies (something that has been already announced in several works [11]). However, as the SNR is increased, the codes with higher multiplexing capacity (e.g. the SM and the Golden code) are preferred. We also observed that the SM with VE implies a loss of around 2dB compared to the Golden code, but when HE is used, the Golden code is around 0.5dB worse than the SM-HE. Actually, the

evolution of the measured ESINR for the different streams is plotted in Figure 6.25 where the x-axis is scaled in channel accesses (i.e. time slots). We can observe that SIMO is the scheme that achieves the highest ESINR during any time instant. Then a few dBs below is the Alamouti where, since is using both transmitter antennas, the Frobenius channel norm dictates the ESINR. Then, the ESINR achieved by the different streams of the SM is plotted separately, where we can observe that the best stream (SM_q, q=1) gets 3dB less than the ESINR of the SIMO, this difference is due to the higher number of transmitter antennas that receive half the transmitting power. The second stream (SM_q, q=2) may fall several dB's (up to 10dB) compared to the best stream. This is the main reason why the SM-VE is so limited compared to the SM-HE. Finally, it is observed that the ESINR of the Golden code falls between the ESINR of both SM streams, however the Golden code's ESINR is lower than the arithmetic mean of the ESINR of both SM streams.

To gain further insights of the TACS behaviour the statistics of LDC selection as a function of the average SNR are plotted in Figure 6.24. We can clearly appreciate that at low SNR the preferred scheme is SIMO where all the power is concentrated in the best antenna, while as the SNR is increased full rate codes ($Q=M$) are more selected since they permit to use lower size constellations. Moreover, comparing SM-VE with SM-HE, we can observe that SM-HE is able to exploit the stream's diversity and hence achieves a higher spectral efficiency than with the Golden code. Actually, at average SNR=12, the SM with HE is the scheme selected for most frames, even more than SIMO.

Next, the performance of the TACS with throughput optimization when using the ML decoder is shown in Figure 6.26 and Figure 6.27. Clearly, there is a large gain when using the ML compared to the previous MMSE detector. Something that is quite relevant is that when using the ML decoder, the Golden code achieves a capacity very close to the one achieved with the TACS selection scheme (less than 1dB improvement due to the TACS). This is a very logical result since the Golden code has been optimized assuming a ML decoder in order to maximize the capacity when $M=T=2$ and Z-QAM modulation which is the same case as analyzed during the simulations. If we have a look at the LDC selection statistics in Figure 6.27, similar results as in the MMSE case are obtained. It is observed that at low SNR values the SIMO is the preferred scheme, whereas at high SNR both the Golden and the SM are chosen similarly (60% and 40% respectively). Furthermore, it is observed that in the SNR range of 10dB to 18dB and following the TACS scheme, the SM-HE is the one selected most times (even more than the Golden code). Again, the high spectral efficiencies achieved by SM-HE in this SNR range come from the fact that this code is able to exploit the diversity between streams when AMC is applied.

As a conclusion, when using TACS with AMC and ML detection the benefits are not so clear even more if we take into account the signalling associated to the TACS scheme. Hence the focus in this case should be to use LDCs optimized for each M,T *pair*. However, we have shown that in case of linear receivers (e.g. MMSE) the TACS scheme with AMC gives a noticeable SNR gain (up to 3dB) in a large SNR margin (SNR from 6 to 18dB) and also is a good technique to achieve a smooth transition between diversity and multiplexing. Then it seems also logical to consider the TACS scheme with linear receivers for the downlink where computational complexity at the mobile station must be kept as low as possible in order to save the batteries. Furthermore, the number of antennas from which the BS can transmit is usually quite large.

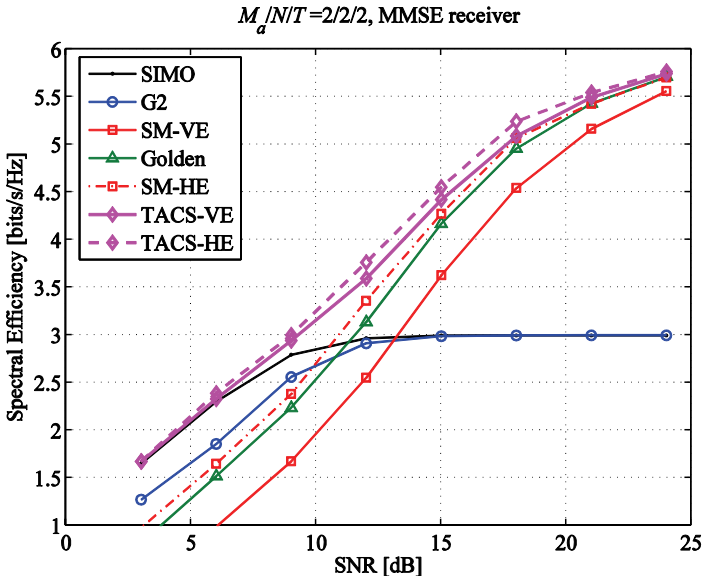


Figure 6.23. Spectral efficiency under TACS with throughput maximization criterion with M_a=2, N=2, adaptive MCS and MMSE receiver for an uncorrelated MIMO Rayleigh channel.

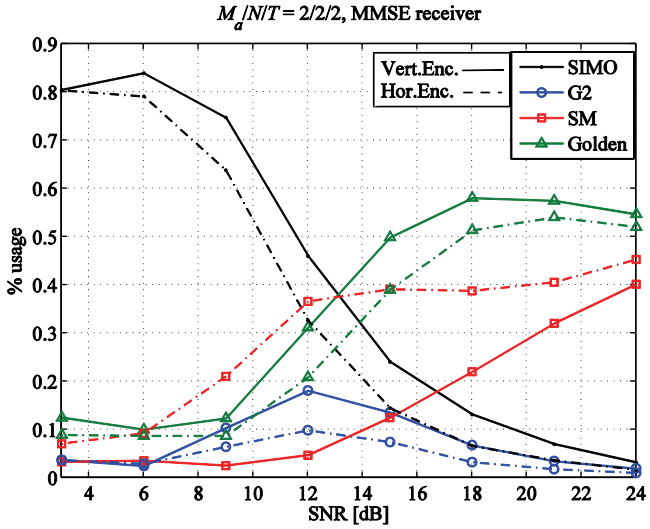


Figure 6.24. LDC selection statistics under TACS with throughput maximization criterion with M_a=2, N=2, adaptive MCS and MMSE receiver for an uncorrelated MIMO Rayleigh channel.

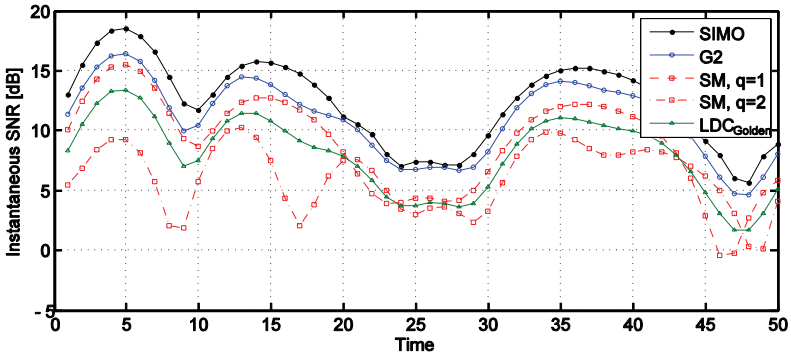


Figure 6.25. Evolution of the ESINR per stream for different space-time codes with M_a=2, N=2, and time-correlated MIMO Rayleigh channel.

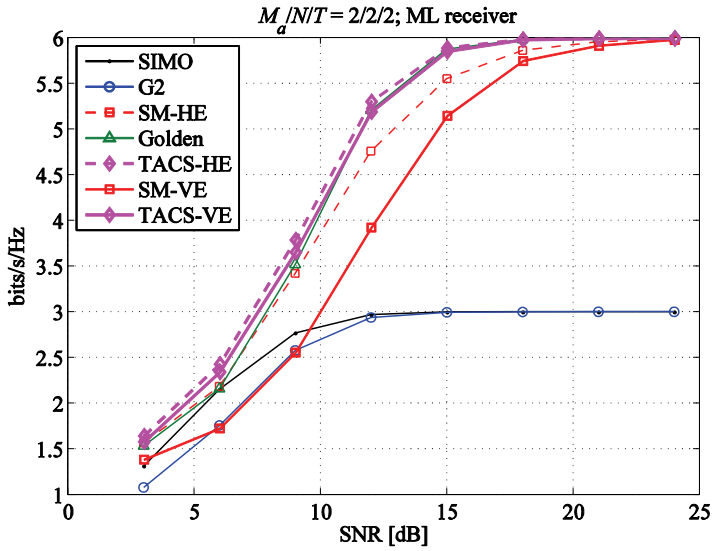


Figure 6.26. Spectral efficiency under TACS with throughput maximization criterion with M_a=2, N=2, adaptive MCS and ML receiver for an uncorrelated MIMO Rayleigh channel.

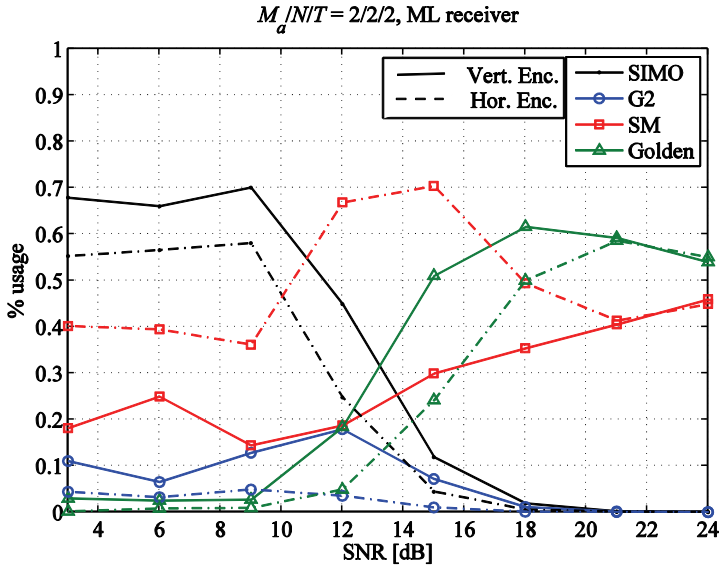


Figure 6.27. LDC selection statistics under TACS with throughput maximization criterion with M_a=2, N=2, adaptive MCS and ML receiver for an uncorrelated MIMO Rayleigh channel.

6.6. Summary of spatial adaptation techniques

The use of multiple antenna techniques at transmitter and receiver sides is still considered as a hot topic for research where the channel capacity can be also increased if multiple streams are multiplexed in the spatial domain. Nevertheless, looking for a trade-off between diversity and multiplexing has motivated the emergence of many different space-time coding architectures where the Linear Dispersion Codes arise as a potential and well-positioned space-time coding scheme for next generation Broadband Wireless Access technologies due to its simplicity and flexibility for obtaining different tradeoffs between diversity and spatial multiplexing. Furthermore, as it was shown by the author in the previous sections, when the transmitter disposes of partial channel state information, the link reliability (i.e. lower bit error rates) as

well as the system throughput can be highly increased. One of the simplest adaptation techniques is the use of antenna selection. It has been also observed that antenna selection increases the diversity of the system up to the maximum available ($g_d=M_a\times N_a$). On the other hand, when transmit antenna selection is combined with code selection a rising coding gain is achieved (more than 2dB and 5dB compared to the Alamouti scheme in case of rates 4bits/s/Hz and 8bits/s/Hz respectively).

The author has proposed a joint Transmit Antenna and space-time Coding Selection (TACS) scheme, where two optimization criteria have been studied: one for bit error rate minimization, and the second for throughput maximization. Moreover, in case of optimization over a subband (i.e. a subchannel in WiMAX terminology), it has been shown that the effective SINR can be improved with more than 2dB in case of uncorrelated channels, and up to 4dB for correlated channels whether the number of transmitting antennas is increased from two to four. Furthermore, another important result obtained from the TACS scheme is that the number of the required space-time coding schemes are very limited. It has been observed that only with the SIMO scheme the Alamouti, the SM, and the Golden codes achieve the best performances in case having a transmission rate lower than 8bits/s/Hz (for higher rates, codes with higher spatial rate would be more appreciated). Nevertheless the low performances achieved in case of using linear receivers (e.g. ZF, MMSE) can be compensated by the TACS scheme, which allows to achieve performances close to those obtained with the non linear receivers (e.g. the Maximum Likelihood) with much lower computational requirements. As a final conclusion, it can be considered that transmit antenna selection with linear dispersion code selection can be an efficient spatial adaptation technique whose low feedback requirements make it feasible for most of the Broadband Wireless Access systems in case of low mobility scenarios.

References

[1] E. Telatar, “Capacity of multi-antenna Gaussian channels”, *European Transactions on Telecommunications*, Nov. 1999.

[2] A. Saad, M. Ismail, N. Misran, “Correlated MIMO Rayleigh Channels: Eigenmodes and Capacity Analyses”, *International Journal of Computer Science and Network Security*, Vol. 8 No. 12 pp. 75-81, Dec. 2008.

[3] L. Zheng, D. Tse, “Diversity and multiplexing: a fundamental tradeoff in multiple antenna channels”, *IEEE Trans. On Information Theory*, May, 2003.

[4] V. Tarokh, N. Seshadri, A. R. Calderbank, “Space-Time codes for high data rates wireless communication: Performance criterion and code construction”, IEEE. Trans. on Information Theory, vol.44, pp.744-765, March, 1998.

[5] V. Tarokh, H. Jafarkhani, A.R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory,* vol.45, no.5, pp.1456-1467, Jul 1999.

[6] G. Bauch, "Space-time block codes versus space-frequency block codes," *The 57th IEEE Semiannual Vehicular Technology Conference, 2003* (VTC-2003-Spring), vol.1, pp. 567-571 vol.1, 22-25 April 2003.

[7] G. Ganesan, P. Stoica, “Space-time diversity scheme for wireless communications”, in Proc. ICASSP, 2000.

[8] S. Sandhu, R. Heath, A. Paulraj, "Space-time block codes versus space-time trellis codes," *IEEE International Conference on Communications*, 2001 (ICC-2001), vol.4, pp.1132-1136 vol.4, 2001.

[9] J. Cheng; H. Wang; M. Chen; S. Cheng, "Performance comparison and analysis between STTC and STBC," the 54th *IEEE Vehicular Technology Conference*, (VTC 2001 Fall), vol.4, no., pp.2487-2491 vol.4, 2001.

[10] L. Yu, P.H.W. Fung, W. Yan, S. Sumei, "Performance analysis of MIMO system with serial concatenated bit-interleaved coded modulation and linear dispersion code," *IEEE International Conference on Communications*, 2004, vol.2, no., pp. 692-696 Vol.2, 20-24 June 2004.

[11] C. Oestges, B. Clerckx, “MIMO Wireless Communications”, Academic Press, Elsevier, USA, 2007.

[12] B. Hassibi, B.M. Hochwald, “High Rates Codes that are Linear in Space and Time”, April 2001.

[13] W. Zhang, X. Ma, B. Gestner, D. V. Andreson, “Designing Low Complexity Equalizers for Wireless Systems”, *IEEE Communications Magazine*, January, 2009, pp. 56-62.

[14] G. J. Foschini, “Layered space-time architecture for wireless communications in a fading environment when using multiple antennas”, Bell Lab Tech. J. v.1., n.2, 1996.

[15] S. M. Alamouti, “A simple transmit diversity technique for wireless communications”, IEEE J. Selected Areas in Communications, vol. 17, pp. 1451-1458, Oct. 1998.

[16] P. Dayal, M.K. Varanasi: "An Optimal Two Transmit Antenna Space-Time Code and its Stacked Extensions", Proceedings of Asilomar Conf. on Signals, Systems and Computers, Monterey, CA , November 2003.

[17] H. Yao, G.W. Wornell: "Achieving the Full MIMO Diversity-Multiplexing Frontier with Rotation-Based Space-Time Codes", Proceedings of Allerton Conf. on Communication, Control and Computing , October 2003.

[18] J.C. Belfiore, G. Rekaya, E. Viterbo: "The Golden Code: A 2 x 2 Full-Rate Space-Time Code with Non-Vanishing Determinants," IEEE Transactions on Information Theory, vol. 51, n. 4, pp. 1432-1436, Apr. 2005.

[19] IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1, IEEE Std 802.16e™-2005, Feb.2006.

[20] M. Vu, A. Paulraj, "MIMO Wireless Linear Precoding", IEEE Signal Processing Magazine,Vol. 4, no.5, pp.87-105, Sept. 2007.

[21] A.B. Gershman, N.D. Sidiropoulos, "Space Time Processing for MIMO Communications", John Wiley & Sons, UK, 2005.

[22] R. W. Heath, A.J. Paulraj, "Linear Dispersion Codes for MIMO Systems Based in Frame Theory", IEEE Transactions on Signal Processing, Vol.50, n.10, Oct.2002

[23] R. Gohary, T. Davidson, "Design of Linear Dispersion Codes: Asymptotic Guidelines and Their Implementation", IEEE Transactions on Wireless Communications, Vol.4, No.6, Nov.2005.

[24] R. Heath, S. Sandhu, A. Paulraj, "Antenna Selection for Spatial Multiplexing Systems with Linear Receivers", IEEE Communications Letters, Vol.5, no.4, April 2001.

[25] D.A. Gore, A.J. Paulraj, "MIMO antenna subset selection with space-time coding," *IEEE Transactions on Signal Processing,* vol.50, no.10, pp. 2580-2588, Oct 2002.

[26] D. Deng, M. Zhao, J. Zhu, "Transmit Antenna Selection for Linear Dispersion Codes Based on Linear Receiver", Proc. Vehicular Technology Conference, 2006. VTC 2006-Spring.

[27] W.C. Freitas, F.R.P. Cavalcanti, R.R. Lopes, "Hybrid MIMO Transceiver Scheme with Antenna Allocation and Partial CSI at Transmitter Side", *IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications, 2006*, pp.1-5, 11-14 Sept. 2006.

[28] R.W. Heath, A.J. Paulraj, "Switching between diversity and multiplexing in MIMO systems," *IEEE Transactions on Communications*, vol.53, no.6, pp. 962-968, June 2005.

[29] L. Che, V.V. Veeravalli, "A Limited Feedback Scheme for Linear Dispersion Codes Over Correlated MIMO Channels," *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007*, (ICASSP-2007), vol.3, pp. 41-44, 15-20 April 2007.

[30] A. Osseiran, V. Stankovic, E. Jorswieck, T. Wild, M. Fuchs, M. Olsson, "A MIMO framework for 4G systems: WINNER concept and results", *IEEE 8th Workshop on Signal Processing Advances in Wireless Communications, 2007*, (SPAWC-2007), pp.1-5, 17-20 June 2007.

[31] D.J. Love, R.W. Heath, T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol.49, no.10, pp. 2735-2747, Oct. 2003.

[32] D. Deng, J. Zhu, "Linear Dispersion Codes Selection Based on Grassmannian Subspace Packing",submitted to IEEE Journal of Selected Topics in Signal Processing, 2007.

[33] R. Machado, B.F. Uchoa-Filho, T.M. Duman, "Linear Dispersion Codes for MIMO Channels with Limited Feedback," *IEEE Wireless Communications and Networking Conference*, 2008 (WCNC-2008), pp.199-204, March 31 2008-April 3 2008.

[34] D. Yang, N. Wu, L.L. Yang, L. Hanzo, "Closed-loop linear dispersion coded eigen-beam transmission and its capacity," *Electronics Letters*, vol.44, no.19, pp.1144-1146, September 11 2008.

[35] R.W. Heath, D. J. Love, "Multimode Antenna Selection for Spatial Multiplexing Systems with Linear Receivers", *IEEE Transactions on Signal Processing*, Vol.53, no.8, Aug.2005.

[36] K. Youngwook, C.Tepedelenlioglu, "Threshold-Based Substream Selection for Closed-Loop Spatial Multiplexing", *IEEE Transactions on Vehicular Technology*, vol.57, no.1, pp.215-226, Jan. 2008.

[37] R. Machado, B.F. Uchôa-Filho, "Extended Techniques for Transmit Antenna Selection with STBCs", Journal of Communications and Information Systems, vol.21, n.3, pp.118-195, 2006.

[38] T. Lestable, M. Jiang, A. Mourad, D. Mazzarese, Uplink MIMO Schemes for IEEE 802.16m, IEEE S80216m-08_534r2, July, 2008.

[39] T. Lestable, M. Jiang, A. Mourad, D. Mazzarese, S. Han, H. Choi, H. Kang , I. Gutierrez, "Linear Dispersion Codes for Uplink MIMO schemes in IEEE 802.16m", IEEE C802.16m-08/535, July, 2008.

[40] I. Gutierrez, F. Bader, J. Pijoan, A. Mourad, "Transmit Antenna and Code Selection for MIMO Systems with Linear Receivers", 7th International Workshop on Multi-Carrier Systems & Solutions (MC-SS 2009), May, 2009 (to be published).

[41] R. Srinivasan et al., "Evaluation Methodology for P802.16m-Advanced Air Interface", IEEE 802.16m-07/037r2.

[42] W. Lopes, W. Queiroz, F. Madeiro, M. Alencar, "Exact Bit Error Probability of M-QAM modulation Over Flat Rayleigh Fading Channels", Microwave and Optoelectronics Conference, 2007. Oct. 2007.

[43] J. Lu, T. Tjhung, C. Chai, "Error probability Performance of *L*-Branch Diversity Reception of MQAM in Rayleigh Fading", IEEE Transactions on Communications, vol. 46, No.2, Feb. 1998.

[44] M. K. Simon, "Evaluation of Average Bit Error Probability for Space-Time Coding based on a Simpler Exact Evaluation of Pairwise Error Probability", Journal of Communications and Networking, 3 (3): 257-267, Sept. 2001.

Chapter 7. Conclusion

For the next generation of Broadband Wireless Access networks, spectral efficiencies in the range of 20-30 bits/s/Hz are required in order to achieve the promised fourth generation throughputs (up to 1Gbps and 100Mbps in nomadic and mobile environments respectively [1]). It is clear that the forthcoming communication air interfaces will be based on multicarrier transmission schemes and will use multiple antenna techniques in order to achieve such tremendous spectral efficiency. The number of research works concerning both topics is very wide, covering many of the physical layer aspects (space-time coding, channel estimation, synchronization, etc.). The research works regarding the medium access methods such as user scheduling, resource allocation (i.e. power and bandwidth), adaptive modulation and coding, etc. are still growing and much of the work proposed in the past assume too optimistic scenarios while the computational cost that imply is usually unaffordable in real time implementations.

Then, from the link layer perspective, since the channel conditions as well as the users' requirements may change rapidly in time, it has been observed that fast link adaptation (including multiple antenna adaptation) as well as adaptive resource allocation and scheduling schemes will be required in order to achieve the most from the theoretical channel capacity. This is even more noticeable in cases of multimedia communications where the required throughput and delay could be very different between the services and even inside each service during proper the connection (variable bit rate). Nevertheless, many of the proposed schemes are not tackled in the standards but are left open, hence each operator may improve its own network management (i.e with proprietary scheduling functions or certain resource management functionalities) leading to very different performances between networks having the same communication standard. This Ph.D. thesis fills some of the research gaps observed and analysed in the field of dynamic resource allocation and scheduling for both single antenna and multiple antenna links.

7.1. Main research results

The author stands out the main research results achieved in this thesis which are:

In **Chapter 5** the design of new resource allocation and scheduling schemes has been studied leading to the introduction and the development of new resource allocation and scheduling algorithms based on a cross-layer strategy design. The following conclusions are drawn:

- For Multicarrier Spread Spectrum communications, it has been shown that in case of dynamic resource allocation and scheduling, the SS-MC-MA scheme doubles the spectral efficiency of the GO-MC-CDMA scheme despite the fairness is reduced.
- For low number of active users ($K \leq 10$), the prioritization schemes based on the provided QoS information achieve better performances in terms of QoS than those that focus only on channel state information. However, as the number of users is increased, it becomes more relevant to take into account the channel status by prioritizing those users with better channel state information.

- When the resource allocation process is performed at the bit level (not at the packet level) as in the schemes proposed by the author, higher flexibility is achieved during the resource allocation process since very small and efficient structures can be allocated to each user. As a result, higher spectral efficiencies and better QoS performances are achieved (e.g. less than half the maximum packet delay obtained with packet prioritization schemes). Furthermore, in case of having mixed traffic services, it has been shown that, if the resources that have not been allocated to users with very low delay constraints are allocated to the users with high (or slight delay constraints) constraints, the packet delay statistics are improved significantly. Then, as it has been shown in the proposed JRAS analysis, the scheme achieves half the packet dropping rate or half the 99^{th} percentage delay compared with other well-known schemes as the Proportional Fair Scheduler in case of *real time* and *non real time* services respectively.
- Moreover, a significant difference from previous research works reviewed in Chapter 3 and Chapter 4 is that we have focused on packet switched networks where the communications is based on packets which are marked with a delay constraint. Thus, every packet is marked by a maximum delay (despite this can be as high as infinite for best effort traffic) and supervised during all the time it is queued. This significant change in the paradigm on how to differentiate the different traffic classes (typically classified into real time, non real time, best effort, etc. – see Section 2.3) makes possible to consider in equally all kind of traffic with the unique difference of the maximum packet delay metric. Actually, this is closer to the user perspective (i.e. the application layer) where the point-to-point delay is the main performance metric.
- Two enhanced resource allocation and scheduling algorithms have been proposed in Figure 5.14 and Figure 5.19, the first dealing with the signalling required in the downlink, and the second in the uplink. For the downlink signalling reductions. It has been shown that, using the data region structures defined in the WiMAX standard, the number of data regions required in the downlink can be adjusted to less than 20 bursts per frame without severely affecting to the spectral efficiency. On the other hand, the use of selective feedback joint with BS notification of incoming packets provide an impressive reduction of the uplink signalling requirements, which could be interpreted as an increase in the spectral efficiency. Furthermore, the proposed schemes can be an off-the-shelf solution which can be rapidly applied in any commercial WiMAX system (or other systems based on the OFDMA scheme).
- Nevertheless, a new permutation zone combining the advantages of distributed subcarrier permutation and contiguous subcarrier permutation has been proposed. It has been shown that (apart from increasing the spectral efficiency) the packet delays can also take advantage from the new permutation zone since two pipes of data are considered. Then, since the localized burst are prioritized over the distributed, every user is allocated one of such localized bursts during each frame, enabling low rate data channels between the transmitter and the receiver. This effect can be clearly appreciated in the simulations for high number of users, where a reduction of more than 90% of the packet delay in case of BE services has been measured having only K'/96 of the resources reserved for the localized bursts (K' means the number of users with data pending to be delivered).

In **Chapter 6**, link adaptation techniques in case of multiple antennas at the transmitter and the receiver have been investigated and developed. Then assuming that the transmitter has (partial) channel state information, new transmit antenna selection techniques combined with space-time coding selection have been proposed. These new adaptation schemes lay on the group of precoding techniques; however the schemes considered are much simpler than the classical beamforming techniques and the amount of feedback is just a small fraction of that required for beamforming. Then, regarding the spatial adaptation issues, the following results have been obtained:

- *Transmit antenna selection* is considered as a key technology which has shown large improvements in terms of spectral efficiency or link reliability, at the expense of very few signalling and complexity. It has been stated that the channel capacity as well as the diversity is a function of the number of transmit and receive antennas. However, adding more antennas in the mobile terminal (apart from being technically difficult due to the required space[15]) requires additional RF and digital processing circuitry at the transmitter and the receiver sides (hence also increasing the power consumption). In consequence, multiple antenna techniques are limited by physical implementation constraints. This is an identified restriction in the mobile stations, where the cost and the size are the essential handicaps and only 2-4 transmit antennas are expected in the forthcoming fourth generation terminals. However, under transmit/receive antenna selection techniques the diversity order is limited by the number of available antennas and not by the number of RF chains. Moreover, this increase of diversity order implies a saving of 10dB in transmitted power when using the SM scheme at BER=10^{-3} and increasing the number of transmitting antennas from two to four. For the Alamouti scheme and the same increase in transmitting antennas, the average SNR is improved by up to 3dB.
- Afterwards, the combination of transmit antenna selection with space-time coding selection has been analysed by the author and unified under the Linear Dispersion Code (LDC) framework leading to the proposed Transmit Antenna and space-time Code Selection scheme (TACS). It has been shown by the simulations carried out that the TACS scheme is able to perfectly double the capacity of a SIMO scheme when two transmitter antennas are used (baseline mobile station), and compared to other MIMO schemes without spatial adaption up to a 3dB improvement in the signal to noise ratio is observed even when a linear receiver is used.
- It is observed that switching between different space-time codes (on top of transmit antenna selection) brings an additional coding gain that can reach 2dB at (uncoded) BER=10^{-3} compared with the best (in terms of average SNR) space-time coding scheme. Furthermore, in case of rate adaptation and linear receivers, an important coding advantage is observed in the range of 8-15dB of SNR where a smooth transition between the spatial diversity schemes (i.e. SIMO and Alamouti) and the spatial multiplexing schemes (SM and Golden code) is obtained.
- Finally, and most importantly is that a small set of space-time codes is required for achieving the maximum capacity or the minimum error rate, concluding that with the

[15] The antennas should be separated at least $\lambda/4$ in order to obtain uncorrelated channels between antennas.

SIMO, the Alamouti, the BLAST and the Golden codes the Grassmanian sampling requirements are nearly satisfied. This is an important result since the same codes used in the open loop mode can be applied also in the closed loop mode, simplifying the transmitter/receiver architectures.

Nevertheless, during the development of this thesis, some necessary simplifications have been introduced in order to boost the results without inquiring in any loss or results misinterpretation. These are here listed, emphasizing its benefits and drawbacks:

- *Need for simple ESINR mapping functions*. During the development of the resource allocation and scheduling algorithms, the author didn't focus on the kind of physical abstraction mechanisms applied. Actually, the geometric mean and the harmonic mean have been used. However, it has been observed by the author (in simulations not included in this thesis) that these simple functions predict very well the behaviour of the channel under certain transmission conditions. Then, it would be necessary to determine when these functions could be used, or when more complex approximations as the EESM or the MIESM are required. These will save a lot of efforts dedicated to the tuning of the EESM or MIESM for any a posteriori applied modulation and coding scheme.

- Further fine-tuning of the *ESINR mapping function proposed for the ML receivers*. This function, despite predicting the results obtained by the simulations very tightly, still needs further developments (mainly theoretical work) in order to get a good understanding/definition on which elements need to be taken into account to estimate the ESINR in case of Linear Dispersion Codes and Maximum Likelihood receivers.

- *Multiuser capacity region under QoS constraints*. In Chapter 5, several studies and proposals have been brought by the author in order to improve the system performance. However, there is not a clear upper bound or reference measurement hence it is difficult to know how far or close to the optimum point we are. This complicates very much the evaluation of these schemes which can be only carried out through comparisons with other proposals. However, many of other authors´ proposals are done under different assumptions and conditions. Therefore, it appears necessary to develop a general framework in which the optimum spectral efficiency, capacity or QoS requirement are clear and the different works can be fairly compared.

- *Characterization of the transmit antenna and code selection gain*. It has been observed in Chapter 6 that large gains can be obtained due to antenna selection and space-time coding selection. It has been noted that this gain is function of the number of transmit and receive antennas as well as the transmission rate and is constant across all the investigated SNR range. However, the derivation of a closed form expression to obtain this gain for any general case (i.e. any number of transmit and receive antennas) has turned unsolvable to us. Furthemore, the gain also depends on the transmission rate

which complicates more the generalization of the expression. As a result, further researches are required in order to find these channel capacity gains.

7.2. Future work

The topics covered within this thesis can be extended towards other research directions, here below are listed some important open questions derived from the research carried during this thesis.

In Chapter 4 and Chapter 5, the problem of resource allocation and scheduling in SISO channels has been deeply studied from different points of views, where heuristic (but efficient) resource allocation and scheduling algorithms have been proposed. It has been shown that in case of multiple users and broadband channels, the throughput (either from the downlink or the uplink) can be increased whether frequency and time dependent scheduling is applied in order to exploit the channel state information. Furthermore, when the number of transmit and receive antennas is greater than one, another degree of freedom must be considered which accounts for the space domain. In Chapter 6, it has been shown that the number of possible multiplexed streams in the space domain is up to $g_s \leq min(M,T)$. Two possible approaches are forecasted to tackle *the* ***multidimensional (space-time-frequency) resource allocation and scheduling*** issue:

- Allocating a frequency-time resource unit to a unique user according to a cost function that determines the revenues of such allocation resource to that user [2][3]. In this case, the classical FDMA+TDMA structure must be followed and, in order to increase the spectral efficiency, the capacity of the MIMO channel is exploited by each user independently on each resource unit applying the diversity and the multiplexing techniques already described in Chapter 6.
- Multiplexing different users within the same resource unit via spatial multiplexing which is referred as Spatial Division Multiple Access (SDMA). Moreover, an uplink system with K terminals (with one transmit antenna each) and N receive antennas at the base station can be considered as a virtual $K \times N$ MIMO channel [4]-[7]. Thus, up to $min(K,N)$ streams could be multiplexed if collaborative spatial multiplexing (CSM) is applied. The same principles can be also applied in the downlink where several (precoding) techniques have been developed to avoid inter-layer interference (e.g. interference cancellation, dirty-paper coding, etc. [8][9]) since the number of receiving antennas might be smaller than the number of spatial layers. Recently, the LDC framework has been also proposed for multiplexing users by assigning a pair of basis matrices ($\mathbf{A}_q$, $\mathbf{B}_q$) to each user [10], which is referred as *Linear-Dispersion Division Multiple-Access* (LDDMA).

Moreover, the minimum allocable resource unit (RU) used throughout this thesis follows a rectangular shape which is beneficial from a signalling point of view since several RU can be packed without increasing the signalling required. However, previous research works on channel estimation have shown that triangular and hexagonal structures are able to follow better the channel variations [11][12]. Furthermore, the triangle is the simplest polygon whose combination gives a large list of flexible structures (squares, rectangles, rhomboids, hexagons,

etc.). Then, more studies on the ***optimum RU structures*** are still required. As a result, the following items will require more deeply researches:

- Design of new RU structures based on simple geometric structures (e.g. triangles), that combined will achieve new flexible bursts structures (see example in Figure 7.1).
- Despite this increase in resource allocation flexibility, it would be necessary to evaluate in detail and balance the benefits and the drawbacks from such scheme, since the signalling requirements may be increased compared to the traditional rectangular RU and burst structures.

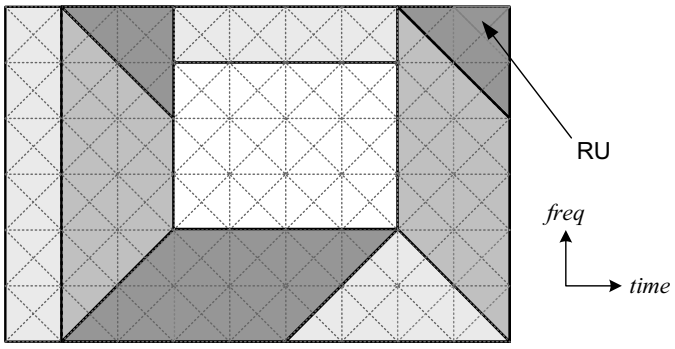


Figure 7.1. Possible OFDM frame decomposition into triangular resource units.

In addition, since the resource allocation and the schemes studied in this thesis consider a unique cell without any intercell interference, it would be also of great interest to extend our research to a ***multiple cell network*** where the handover between the cells as well as macro-cell diversity could be investigated within the same context taking into account the signal to interference and noise ratios measured from each access point as well as the load in each cell. Actually, two types of intercell coordination have already been proposed named static and semi-static coordination, where part of the resources are used for cell-edge user with full-power transmission and orthogonal between the neighbour cells, whereas users in the inner part of the cell are assigned the rest of the resources but with restricted power to avoid interference in the neighbour cells [13]-[16]. The number of resources booked for cell edge users might be fixed or may change according to the traffic load conditions for static and semi-static coordination respectively. Then, the main questions that arise regarding multiple cell coordination and resource allocation are:

- How the previous resource allocation algorithms will be affected by the multiple cell scenarios?
- In addition, since the BS will require information about the interference, this must be measured at the MS and reported in the uplink, forecasted from neighbour base stations, or measured somehow by the base station?
- Furthermore, there is a complexity issue regarding the multiple cell coordination due to the amount of information that a central node needs to manipulate in case of a centralized approach. Otherwise, in a distributed scenario, which information should the base stations share/know?

Finally, the large channel fluctuations (i.e. due to multipath and shadowing), as well as the changes on the instantaneous resource requirements in case of multimedia transmission make it very difficult to predict with accuracy the maximum number of users that can be simultaneously admitted into the system. Then, efficient ***connection admission control*** (CAC) algorithms are indispensable in order to guarantee the QoS without sacrificing spectral efficiency (e.g. considering simply the worst or average channel conditions and traffic requirements). Mainly two types of CAC schemes are mentioned in the scientific literature: *pro-active* and *reactive*. The *pro-active* schemes predict the users demands and as function of the system load, a new user is allowed or not to get into the network. The *reactive* schemes work on real-time performance measurements monitoring the QoS delivered to each user. In that case when the system is close to reach the overload, some of the resources are released (users are disconnected). Then, although the CAC has a great impact on the resource allocation and scheduling, the CAC algorithms have not been considering in any of the author's proposals since they were out of the scope of this thesis. However, the relevance of these schemes into the whole resource allocation process is clear and the following further studies should be considered:

- Evaluate the effects of the CAC schemes into the resource allocation and scheduling algorithm, and how the CAC is affected depending on the traffic type (e.g. non-real time services may accept a very high packet delay until disconnection).
- Define, in case of releasing resources, which is the most fair strategy as well as the most beneficial one.
- Conclude which one the previous approaches (pro-active or reactive) should be used in each case.

References

[1] ITU-R Recommendation M.1645, “Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000,” 2003.

[2] W. Huang, K.B. Letaief, “A Cross-layer Resource Allocation and Scheduling for Multiuser Space-Time Block Coded MIMO/OFDM Systems”, *Communications, 2005. ICC 2005. 2005 IEEE International Conference on* , vol.4, pp. 2655-2659, May 2005.

[3] X. Jian, K. Jongkyung, P. Wonkyu, S. Jong-Soo, "Adaptive Resource Allocation Algorithm with Fairness for MIMO-OFDMA System," *IEEE 63rd Vehicular Technology Conference, 2006* (VTC-2006-Spring), vol.4, no., pp.1585-1589, 7-10 May 2006.

[4] H. Boche, E.A. Jorswieck, "Sum capacity optimization of the MIMO Gaussian MAC," *The 5th International Symposium on Wireless Personal Multimedia Communications, 2002*, vol.1, pp. 130-134 vol.1, 27-30 Oct. 2002.

[5] M. Dohler, J. Dominguez, H. Aghvami, "Link capacity analysis for virtual antenna arrays," *IEEE 56th Vehicular Technology Conference, 2002. Proceedings. VTC 2002-Fall. 2002*, vol.1, pp. 440-443 vol.1, 2002.

[6] A. Nosratinia, T.E. Hunter, A. Hedayat, "Cooperative communication in wireless networks," *IEEE Communications Magazine*, vol.42, no.10, pp. 74-80, Oct. 2004.

[7] M. Baek, H. Song, "Cooperative diversity technique for MIMO-OFDM uplink in wireless interactive broadcasting," *IEEE Transactions on Consumer Electronics*, vol.54, no.4, pp.1627-1634, November 2008.

[8] S. Vishwanath, N. Jindal, and A. Goldsmith, “On the capacity of multiple input multiple output broadcast channels,” in IEEE Int. Conf. on Communications (ICC) New York, NY, vol. 3, Apr. 2002, pp. 1444-1450.

[9] Q.H. Spencer, C.B. Peel, A.L. Swindlehurst, M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Communications Magazine*, vol.42, no.10, pp. 60-67, Oct. 2004.

[10] D. Deng, X. Lv, J. Zhu, “Linear-Dispersion Division Multiple-Access for MIMO systems”, *Journal of Electronics (China)*, Science Press, pp. 433-438, vol.25, no.4, July, 2008.

[11] M.J. Fernandez-Getino Garcia, J.M. Paez-Borallo, S. Zazo, “Pilot Patterns for Channel Estimation in OFDM”, *Electronic Letters*, vol. 36, no 12, June 2000, pp. 1049-1059.

[12] F. Bader, R. Gonzalez, “Pilot Time-Frequency Location Adjustment in OFDM Systems Based on the Channel Variability Parameters”, Multi-Carrier Spread-Spectrum, Springer Netherlands, pp.257-264, 2006.

[13] R1-050764,”Inter-cell Interference Handling for EUTRA”, Ericsson, August 28-September 2, 2005.

[14] R1-061374,“Downlink inter-cell interference coordination /avoidance-evaluation of frequency reuse”, Ericsson, May 8- 12,2006

[15] R1-05-0272 “OFDM air interface with QOS at cell edge”, Alcatel, meeting #40bis, Beijing, China, 4th – 8th April 2005.

[16] R1-05-0694, “Multi-cell Simulation Results for Interference Co-ordination in new OFDM DL”, Alcatel, 3GPP TSG RAN WG1 #42, London, UK, 29th August – 2nd September 2005.

[17] M. Rahman, H. Yanikomeroglu, "Interference Avoidance through Dynamic Downlink OFDMA Subchannel Allocation using Intercell Coordination," *IEEE Vehicular Technology Conference, 2008. VTC Spring 2008.*, pp.1630-1635, 11-14 May 2008.

[18] D. Niyato, E. Hossain, "Call admission control for QoS provisioning in 4G wireless networks: issues and approaches," *IEEE Network*, vol.19, no.5, pp. 5-11, Sept.-Oct. 2005

[19] B. Tarek, N. Nidal, “Efficient call admission control scheme for 4G wireless networks”, Wireless Communications and Mobile Computing, 2008, John Wiley & Sons, Ltd.

[20] M. Castrucci et al., “Connection Admission Control in WiMAX networks”, ICT-MobileSummit 2008 conference Proceedings, June, 2008, Stockholm, Sweeden.

Appendixes

Appendix A. Multipath channel power delay profiles

In this section the two multipath channel models used throughout this thesis are shown: the BRAN-A channel model based on the HIPERLAN/2 project (Broadband Radio Access Network – BRAN - channels) [1][2], and those proposed by the ITU (ITU Pedestrian and Vehicular channel models) [3].

A.1. Broadband Radio Access Network (BRAN) channel A

This model is issued from the ETSI BRAN tapped delay line model designated as channel A [1][2]. From the ETSI BRAN models this is the one with the smaller delay spread. It corresponds to a typical office environment and transmitter to receiver distances larger than 100m (indoor close area) with mobile station moving not faster than 3 km/h. The Doppler power spectrum follows the Clarke's model for each tap.

Indoor close area				
Name	**APDP Type**	**LOS/NLOS**	**Delay Spread**	**Coherence Bandwidth**
Channel A	Typical	OLOS/NLOS	50 ns	4 MHz
Tap Number		**Delay [ns]**		**Power [dB]**
1		0		0
2		10		-0.9
3		20		-1.7
4		30		-2.6
5		40		-3.5
6		50		-4.3
7		60		-5.2
8		70		-6.1
9		80		-6.9
10		90		-7.8
11		110		-4.7
12		140		-7.3
13		170		-9.9
14		200		-12.5
15		240		-13.7
16		290		-18
17		340		-22.4
18		390		-26.7

Table A.1. Indoor close area tapped delay line BRAN channel mode.

A.2. ITU channel models.

The ITU developed in [3] a short list of multipath channel models for outdoor-to-indoor and outdoor communications. These two cases are differentiated by the users' mobility (pedestrian and vehicular). Furthermore, measurements campaigns showed that in outdoor environments

the r.m.s. delay spread can vary over an order of magnitude within the same environment. Hence, to accurately model the variability of delay spread two multipath channels are defined for each test environment. Within one test environment channel A is the low delay spread case that occurs frequently, and channel B is the median delay spread case that also occurs frequently. Each of these two channels are expected to be encountered for some percentage of time in a given test environment. The percentage of time the particular channel may be encountered with the associated *root mean square* average delay spread for channel A and channel B for each terrestrial test environment is 40% and 55% respectively. The following Table A.2. depicts the APDP for these channel models.

Power Delay Profile		**Pedestrian-A**		**Pedestrian-B**		**Vehicular-A**		**Vehicular-B**	
Number of Paths		4		6		6		6	
Terminal velocity		3-10Km/h				Up to 120Km/h			
Relative Path power (dB)	Delay (ns)	0	0	0	0	0	0	-2.5	0
		-9.7	110	-0.9	200	-1.0	310	0	300
		-19.2	190	-4.9	800	-9.0	710	-12.8	8900
		-22.8	410	-8.0	1200	-10.0	1090	-10.0	12900
				-7.8	2300	-15.0	1730	-25.2	17100
				-23.9	3700	-20.0	2510	-16.0	20000

Table A.2. Power Delay Profiles of ITU channel models

References

[1] J. Medbo and P. Schramm, “Channel models for HIPERLAN/2,” ETSI/BRAN document no. 3ERI085B.

[2] J. Medbo, H. Andersson, P. Schramm, H. Asplund, “Channel Models for HYPERLAN/2 in different Indoor Scenarios”, COST 259 TD98, Bradford, UK, April, 1998.

[3] ETSI, “Universal Mobile Telecommunication System (UMTS)”; Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03 version 3.2.0)”, TR 101 112 v.3.2.0, April, 1998.

Appendix B. Antenna Pattern for sectorization

In order to increase the spectral efficiency sectorization can be applied on each cell. Among other advantages sectorization makes possible to increase the number of users per cell and what is more, decreases the transmitted power since it improves the link budget. As the main drawback, the economic cost becomes higher and it requires more radio resources. The defined antenna pattern in [1] and [2] only considers the azimuth (horizontal) pattern diagram. According to this model, the power attenuation $A(\theta)$ (in dB) as a function of the angle θ between the antenna pointing direction and MS position can be obtained by

$$A(\theta)[\mathrm{dB}] = -\min\left[12\left(\frac{\theta}{\theta_{3\mathrm{dB}}}\right)^2, A_m\right] \tag{B.1}$$

where $\theta_{3\mathrm{dB}}$ is the angle at which power is attenuated 3dB and A_m is the maximum considered attenuation. Typically $\theta_{3\mathrm{dB}}$=70° and A_m =20dB for the urban and microcellular environment, and $\theta_{3\mathrm{dB}}$=70° and A_m =12dB for the indoor environment [2].

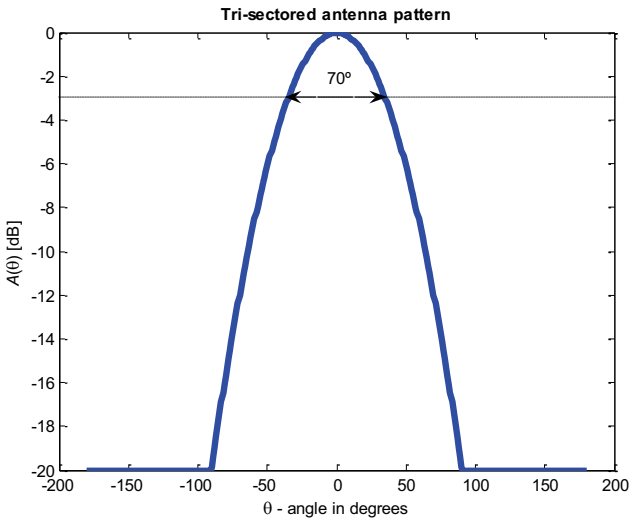


Figure A.1. Antenna pattern for sectorization.

References

[1] *Specification of the performance evaluation methodology ant the target performance, Deliverable 1.3*, Multicarrier CDMA TRansmission Techniques for Integrated Broadband CEllular Systems, 2001-4.5.2 Terrestrial wireless systems and networks, IST-2001-4.5.2 Terrestrial wireless systems and networks.

[2] *Deliverable 6.13.1. v1.0*, Wireless World Initiative New Radio, IST-2002-2.3.1.4 Mobile and wireless systems beyond 3G.

Appendix C. Packet Error Rate for QAM modulation with turbo-coding in AWGN channels

The following Table C.1. depicts the mapping between the SNR and the Block Error Rate (BLER) for packets using N_{sch} subchannels in the uplink, with a QPSK or 16QAM modulation and turbo-coding with coding rate cr.

SNR	QPSK $cr = 1/3$	QPSK $cr = 1/2$	QPSK $cr = 2/3$	16QAM $cr = 1/2$	16QAM $cr = 2/3$	16QAM $cr = 3/4$
4	1	1	1	1	1	1
-3.5	0.99	1	1	1	1	1
-3	0.95	1	1	1	1	1
-2.5	0.91	1	1	1	1	1
-2	0.7098	1	1	1	1	1
-1.5	0.4654	1	1	1	1	1
-1	0.22375	1	1	1	1	1
-0.5	0.0755	0.98	1	1	1	1
0	0.0179	0.82815	1	1	1	1
0.5	0.00275	0.61275	1	1	1	1
1	0	0.3481	1	1	1	1
1.5	0	0.1442	0.99	1	1	1
2	0	0.0422	0.92	1	1	1
2.5	0	0.00855	0.69395	1	1	1
3	0	0.0014	0.3547	1	1	1
3.5	0	0	0.11395	1	1	1
4	0	0	0.0239	1	1	1
4.5	0	0	0.00295	0.97	1	1
5	0	0	0.00065	0.93	1	1
5.5	0	0	0	0.7282	1	1
6	0	0	0	0.50605	1	1
6.5	0	0	0	0.29205	1	1
7	0	0	0	0.1304	1	1
7.5	0	0	0	0.0488	0.99	1
8	0	0	0	0.0133	0.9166	1
8.5	0	0	0	0.00305	0.6312	0.99
9	0	0	0	0	0.2562	0.88315
9.5	0	0	0	0	0.06	0.69535
10	0	0	0	0	0.0084	0.4478
10.5	0	0	0	0	0.0018	0.22725
11	0	0	0	0	0	0.08975
11.5	0	0	0	0	0	0.02975
12	0	0	0	0	0	0.0084
12.5	0	0	0	0	0	0.0026
13	0	0	0	0	0	0.00105
13.5	0	0	0	0	0	0

Table C.1. Packet Error Rate for Z-QAM modulation with turbo coding in AWGN channels.

Appendix D. Channel estimation and interpolation techniques in multicarrier communications

In Chapter 2 the relevance of the OFDM scheme as the transmission technique in which most the forthcoming air interfaces are designed was discussed. Among the more important advantages of the OFDM scheme is the simplicity in signal equalization, however, perfect channel knowledge is required at the receiver in order to properly decode the signal. In the following subsections the principles of channel estimation at the receiver as well as some channel interpolation techniques used in multicarrier communications are reviewed. A new interpolation techniques proposed by the author is also explained.

D.1. Channel Estimation in OFDM

The main idea behind the use of the OFDM scheme is to transform a high data rate transmission (with high bandwidth occupancy) into multiple low bit rate streams where each stream is transmitted by a specific sub-carrier. When the coherence bandwidth of the channel is much greater than the bandwidth occupied per subcarrier, the channel can be considered flat since the subcarriers do not suffer from amplitude distortion and only one-tap equalizer is required. The use of differential phase-shift keying (DPSK) in OFDM systems avoids need to track a time varying channel; however, it limits the number of bits per symbol and results in a 3 dB loss in signal-to-noise ratio (SNR). Coherent modulation allows arbitrary signal constellations, but efficient channel estimation strategies are required for coherent detection and decoding. Many different channel estimation techniques can be found in the literature [1][2][3]. Most of these channel estimation techniques can be divided mainly in two groups: blind channel estimation techniques and pilot assisted channel estimation. The first group of channel estimation techniques usually exploits the cyclostationarity properties of the OFDM signal [3]. Actually, when cyclic prefix extension is applied two copies of the same transmitted signal can be found in the received signal passed through the channel. The cross-correlation of these signals along with some signal processing can give us the channel impulse response [3]. However, blind channel estimation usually requires more complexity at the receiver side. In consequence, many standards insert pilot symbols into the data stream (*known as Pilot Symbols Assisted Modulation* -PSAM-) in order to easily estimate the channel.

For the PSAM schemes and assuming a perfect synchronization at the receiver, recalling equation (2.53), each $S_{n,l}$ symbol received within the OFDM frame (n,l means the subcarrier index and the OFDM symbol index respectively) can be written as

$$R_{n,l} = H_{n,l} S_{n,l} + N_{n,l}, \quad n = 0,\ldots,N_C - 1; \quad l \in \mathrm{Z} \tag{D.1}$$

where $H_{n,l}$ is the frequency response of the channel at the n-th sub-carrier during the l-th and $N_{n,l}$ is the AWGN term with a variance σ_N^2. Since the channel term is time and frequency dependent function, in order to avoid aliasing the channel estimation process should be carried out using a sampling rate at or above the Nyquist frequency in both time and frequency domains. An over-sampling may be carried out to improve the channel estimation accuracy. The, the two dimensional spacing in frequency (N_f) and time (N_t) directions between two consecutive pilots are given by [4]

$$N_f \le \frac{1}{2\cdot\Delta f\cdot\tau_{\max}}, \tag{D.2}$$

$$N_t \le \frac{1}{2\cdot f_{D_{\max}}\cdot T_{symbol}}. \tag{D.3}$$

Then, we may recover the channel coefficients by applying the *Least Squares* (LS) estimator hence the channel matrix is obtained by

$$\arg\min_{\bar{\bar{H}}_\Omega}\left\{\left(\bar{\bar{R}}_\Omega-\bar{\bar{H}}_\Omega\bar{\bar{S}}_\Omega\right)^H\left(\bar{\bar{R}}_\Omega-\bar{\bar{H}}_\Omega\bar{\bar{S}}_\Omega\right)\right\},\quad \forall\{k',l'\}\in\Omega \tag{D.4}$$

which leads to the following:

$$\hat{H}_{k',l'}\Big|_{LS}=\frac{R_{k',l'}}{S_{k',l'}}=H_{k',l'}+\frac{N_{k',l'}}{S_{k',l'}},\quad \forall\{k',l'\}\in\Omega, \tag{D.5}$$

where Ω means the subset of the pilot positions within the OFDM frame. The performance of the LS can be further improved if Equation (D.4) is substituted by one minimizing the mean-square error (MMSE) as follows

$$\arg\min_{\bar{H}_\Omega}\left\{E\left\{\left|\left(\bar{R}_\Omega-\bar{H}_\Omega\bar{S}_\Omega\right)\right|^2\right\}\right\} \tag{D.6}$$

where $\mathrm{E}\{x\}$ means the expectation of x. The solution to Eq. (D.6) is then given by [ref]

$$\hat{H}\Big|_{MMSE}=\mathrm{P}_{HH}\left[\mathrm{P}_{HH}+\sigma_N^2\left(\bar{S}_l\bar{S}_l^H\right)^{-1}\right]\hat{H}\Big|_{LS}, \tag{D.7}$$

and $\mathrm{P}_{HH}=E\left\{\bar{H}\bar{H}^H\right\}$ is the autocovariance matrix of the channel.

It is noted in Eq. (D.7) that the performance of the MMSE channel estimator is that of the *LS* at high SNR, whereas at low SNR the channel is obtained taking into account the *LS* channel samples and including the correlation between them in order to minimize the mean square error.

Since the insertion of the pilots implies a reduction of the system efficiency, a fundamental tradeoff between channel estimation quality and spectral efficiency is given. The chosen pilot density depends on the channel conditions and the required QoS (i.e. the bit error rate) at the receiver. Three main pilot insertion schemes are found in the literature named *block-type*, *comb-type* and a combination of both named *scattered* which are depicted in Figure B. In the block-type pilot insertion one OFDM symbol is dedicated in its whole to channel estimation. On the other hand, the comb-type pilot insertion dedicates some of the subcarrier for channel estimation during all the transmission. Finally under the scattered scheme, the pilots are spread within the OFDM following a regular structure. Several investigations have been undertaken with the objective to distribute and to develop appropriate pilot patterns (*rectangular, diagonal, hexagonal, etc.*) within the OFDM data frames [5][6]. Results obtained by Colieri and al. in [7] have demonstrated the direct influence of pilot location within the frames of the OFDM on the communication system performances, in particular during the channel estimation procedure.

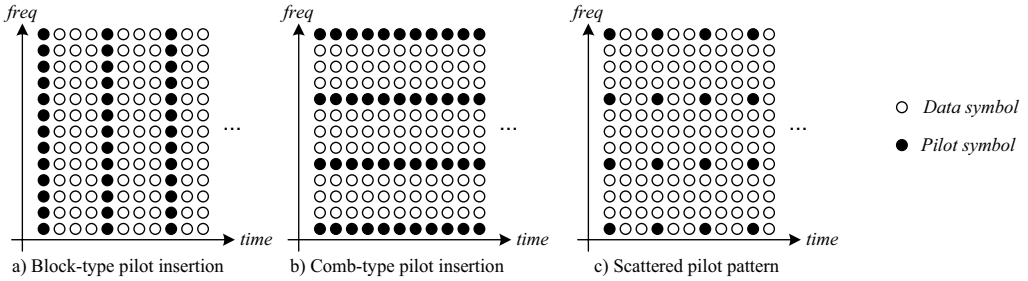


Figure C.1. Pilot symbols insertion in OFDM schemes.

Furthermore, the block-type pilot insertion can be used for synchronization purposes as well as fine channel estimation. This kind of pilot insertion might be found at the beginning of a bursty transmission where at the beginning the receiver has no knowledge of the channel at the transmission is completely asynchronous. The following comb-type pilot insertion scheme is used for channel tracking where the channel at certain frequencies is estimated. This second pilot scheme is also used for frequency and time offset corrections during the demodulation of the transmission. Finally the scattered pilot pattern is a downsampled version of both block-type and comb-type schemes and can be used also for time and frequency synchronization purposes.

D.2. Interpolation methods

The channel samples obtained according to Eq. (D.5) or Eq. (D.7) just give the value of the channel at the pilot positions whereas for data demodulation a complete grid of the channel matrix transfer function ($H(n,l)$) is required. Then, channel coefficients have to be estimated at the data positions interpolating those values known at the pilot positions.

In the following subsections some well-known channel interpolation procedures are studied and compared. Actually, the schemes studied are the optimal filtering using the minimum mean square error (MMSE) function also known as the Wiener filter, the zero padding interpolation (ZPI) scheme, biharmonic splines based interpolation (BSI), the IFFT based interpolation process (IFFT-I), and finally the nearest pilot padding interpolation (NPI) technique proposed by the author in [9].

D.2.1. Optimum Filtering (MMSE)

Recalling the signal in Eq. (D.1), the two dimensional received sequence can be written as

$$R_{n,l} = H_{n,l}X_{n,l} + N_{n,l}, \quad n = 0,\ldots,N_C - 1; \quad l \in \mathrm{Z}. \tag{D.8}$$

It can be seen that the data symbol $X_{n,l}$ multiplied by the channel coefficient $H_{n,l}$ is disturbed by an additive white Gaussian noise $N_{n,l}$. The received value $R_{n,l}$ is observed only at the pilot locations k',l' where $X(k',l')$ are known. Then, the whole estimated channel matrix $\tilde{H}$ can be obtained by filtering the channel sampled at k',l' and the optimum filter W [4]

$$\tilde{H}(k,l) = \sum_{\{k',l'\}} W(k,l;k',l') \cdot R(k',l') \tag{D.9}$$

such that

$$\arg\min_{W} E\left\{\left|\tilde{H}(k,l)-\tilde{H}(k,l)\right|^2\right\} \tag{D.10}$$

The optimum two dimensional (2D) filter usually referred as the Wiener filter is then given by

$$W_{opt}(k,l)^T = \theta^T(k,l)\cdot\phi^{-1} \tag{D.11}$$

where θ and ϕ^{-1} denote the cross-variance vector ($\theta \in \mathbb{C}^{Ntap \times 1}$), and the auto-covariance matrix ($\phi \in \mathbb{C}^{Ntap \times Ntap}$) of the channel matrix respectively. N_{tap} means the numbers of paths of the multipath channels and is used instead of N_C to reduce the computational cost.

Further simplifications can be obtained in order to reduce the filter order, i.e. by taking only the nearest pilots or the highest correlated pilots. When working with a constant pilot pattern, θ and ϕ can be pre-computed since their values only depend on the time/frequency pilot distances, i.e. the auto-covariance and cross-covariance are depicted for a specific symbol, but for other symbols placed at the same relative position inside the grid the previously computed θ and ϕ are also valid. The following Figure C.2 illustrates the channel interpolation process carried out in case of scattered pilot pattern by the Wiener filter if only the closest four pilot symbols are used for interpolation purposes.

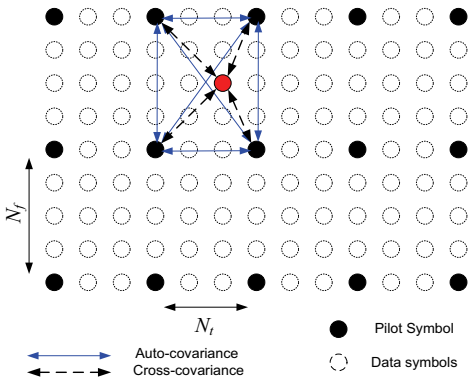


Figure C.2. Channel Estimation using the 2D Wiener Filter with reduced complexity.

D.2.2. Zero Padding Interpolation (ZPI).

This technique is also based on two dimensional filtering. The zero padding process is carried out by filling the entire estimated channel matrix $\tilde{H}$ with zeros except on those positions where the pilots are located. Once the transfer function matrix is filled with the channel estimation from the received pilot symbols using Eq. (D.5) or Eq. (D.7), interpolation of the other samples is performed using a two dimensional filter with digital cut-off frequencies subject to the following restrictions:

$$f_{c\_TIME} = \frac{1}{2(N_t-1)}, \qquad f_{c\_FREQ} = \frac{1}{2(N_f-1)}. \tag{D.12}$$

The total channel estimation $\tilde{H}(k,l)$ should be obtained by a two dimensional convolution process between the filter impulse response $h_r(k,l)$, and the channel transfer function matrix padded with zeros $\hat{H}(k,l)$. Thus the final expression for the estimated channel is given by

$$\tilde{H}(k,l) = \sum_{k,l} \sum_{k',l'} h_r(k,l) \cdot \hat{H}(k',l'). \tag{D.13}$$

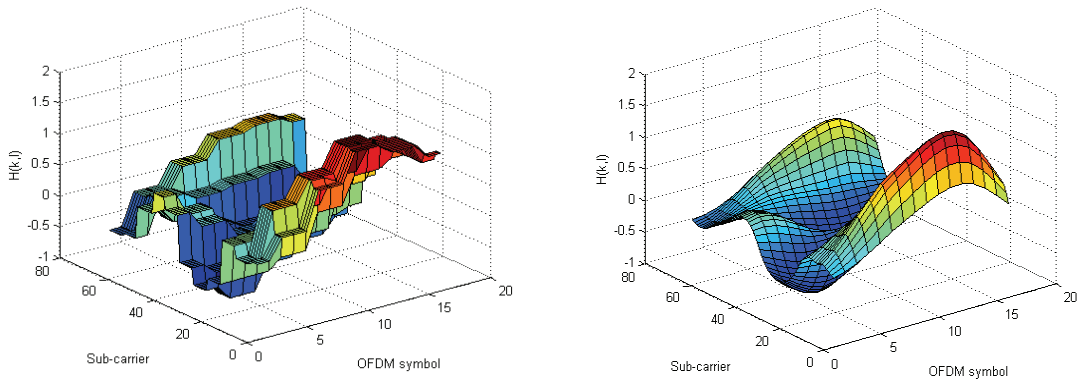
D.2.3. Nearest pilot padding and interpolation (NPI)

In [9], the author proposed another interpolation technique similar to the Zero Padding and Interpolation (ZPI) (reviewed in Section D.2.2) in terms of complexity and performance. Under the Nearest pilot padding and interpolation (NPI) technique, the non-pilot positions of the channel matrix are filled with the information from the nearest (or highest correlated) pilot. The result of this process is represented in Figure C.3a. As it can be noticed, the obtained estimated channel presents strong discontinuities and should be smoothed. Then, a two dimensional filter is applied whose cut-off frequencies are the coherence bandwidth (for the frequency domain) and the coherence time (for the time domain). Translated to digital domain where the frequency domain is sampled at Δf and T_s in frequency and time respectively, the equations for the digital cut-off frequencies are given by

$$f_{c\_FREQ} = \tau_{\max} \cdot \Delta f \cdot N_f, \tag{D.14}$$

$$f_{c\_TIME} = f_{D_{\max}} T_s \cdot N_t. \tag{D.15}$$

In case we have sampled the channel just at Nyquist frequency, both cut-off frequencies becomes *0.5*. The results of the filtering process can be observed in Figure C.3b (with N_c=64, N_f=6, N_t=3 and 20 OFDM symbols), where the channel frequency response has been smoothed and the noise is filtered too.



a) Before filtering *b) After filtering*

Figure C.3. Nearest Pilot Padding effects on a Typical Urban channel.

D.2.4. Biharmonic Splines Interpolation (BSI)

A spline is a special function defined piecewise by polynomials. In interpolating problems, spline interpolation is often preferred to polynomial interpolation because it yields similar results with lower complexity as well as avoiding Runge's phenomenon for higher degrees. On the other hand Green functions of the biharmonic operator, in one and two dimensions, can

be used equivalently for minimum curvature interpolation of irregularly spaced data points. The interpolating curve (or surface) is a linear combination of Green functions centered at each data point. The amplitudes of the Green functions are found by solving a linear system of equations. In one (or two) dimensions this technique is equivalent to cubic spline interpolation. The basis of the biharmonic splines interpolation method, detailed by Sandwell in [8], is to find a two dimensional bi-harmonic function which passes through N data points, where N is the number of pilot symbols within a frame. The solution to this problem is a weighted elastic beam or spline where the forces applied to the beam keep it bent. For small displacements, the fourth derivative of the spline is zero except where the weights are placed. The point force the Green function $\phi(x)$ for the spline satisfies the following bi-harmonic equation:

$$\frac{\partial^4 \phi}{\partial x^4} = 6\delta(x) \tag{D.16}$$

where $\delta(.)$ means the delta function. When interpolating in two dimensions, the Green function is defined by

$$\phi(x) = |x|^2 \left(\ln|x| - 1\right), \tag{D.17}$$

while its gradient is given by

$$\nabla\phi(x) = x\left(2\ln|x| - 1\right). \tag{D.18}$$

Then, for N sample points located at $x_{n',l'}$ the problem is defined by Eq. (D.19) where ∇^4 is the bi-harmonic operator.

$$\nabla^4 H\left(x_{n',l'}\right) = \sum_{n',l' \in \Omega} \alpha_{n,l}\delta\left(x - x_{n',l'}\right) \quad \text{with} \quad H\left(x_{n',l'}\right) = H_{n',l'} \tag{D.19}$$

The general solution to the previous system of linear equations leads to

$$H\left(x_{n',l'}\right) = \sum_{n',l' \in \Omega} \alpha_{n',l'} \cdot \phi\left(x - x_{n',l'}\right), \tag{D.20}$$

where $\alpha_{n',l'}$ is the strength of each point force. Afterwards, the channel matrix H can be estimated by

$$H_{n,l} = \sum_{n',l' \in \Omega} \alpha_{n',l'} \cdot \phi\left(x_{n,l} - x_{n',l'}\right). \tag{D.21}$$

D.2.5. Time Domain Interpolation

The Time Domain Interpolation (TDI) is a high-resolution based on zero-padding and IFFT/FFT. Following this channel interpolation method, the channel transfer function is obtained first for all those OFDM symbols where pilots are available. Then, the subsampled channel frequency responses obtained following Eq. (D.5) or Eq. (D.7) is transformed to time domain by the IFFT. The resulting impulse response is then interpolated up to N_c points. Since for most channels the average power delay can be known a priori, those samples for a delay larger than the maximum delay spread are forced to zero. Afterwards the resulting filtered and upsampled impulse response is transformed to the frequency domain again by the FFT.

Finally, the different frequency responses obtained for each OFDM symbol where the pilots were present, are interpolated in the time domain using traditional one dimensional data

interpolation techniques. In consequence, the TDI is a double one dimensional (2x1D) interpolation technique.

D.2.6. Comparison of the different interpolation algorithms

In this section the performance of the interpolation techniques before mentioned are compared under the simulation parameters described in D) The performance criterion is the Mean Square Error (MSE) averaged over all symbols (the edge effects are hence included) and the Mean Phase Error (MPE) also averaged over all the symbols. The representation of both measurement criteria is shown in Figure C.4. Despite most of the methods are designed to minimize the MSE, it might be important to evaluate both measurements since many modulations transmit the information in both the amplitude and the phase (though the amplitude component has a deep impact on the signal to noise ratio).

As shown in Figure C.5 and Figure C.6, the two dimensional Wiener filtering obtains the best performance for low signal to noise ratios (SNR) values, however it cannot improve the performance of other evaluated methods at higher SNR ranges due to the non-perfect knowledge of the covariance and auto-correlations matrices which leads to channel estimation errors. It has been also observed that the IFFT based interpolation process (i.e. TDI) gives the best performance in environments with long power delay profiles where the paths are well separated between them. On the other hand, the Biharmonic Spline Interpolation (BSI) gives the best performance in terms of MSE in case of urban channel profiles where the coherence bandwidth is relatively high and achieves quite good performance for the rural channel profile, hence it might be a good amplitude channel estimator (in terms of MSE) independently of the channel profile. However, the performance of the BSI in terms of MPE is very bad as can be observed in Figure C.5 and Figure C.6. In consequence, the BSI might be preferred for amplitude modulations. Next, we can also observe that the performance of the ZPI and the NPI are very similar in terms of MSE and MPE. However, we observe that for the MSE criterion the ZPI outperforms the NPI, whereas for the MPE criterion is the opposite. This means than when using phase modulation the NPI should be selected, whereas for amplitude and phase modulations ZPI would perform slightly better.

These results are in agreement with those in [7] and [10] where, in case of comb-type or scattered pilot patterns, the ZPI is preferred due to its lower computational complexity and its good trade-off between MSE and MPE. Nevertheless, we have shown that the proposed NPI may outperform the ZPI when the phase of each data symbol is critical.

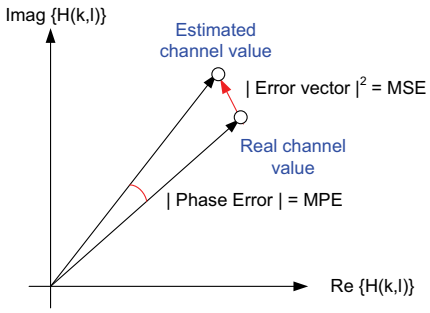


Figure C.4. MSE and MPE representation

Parameter	Value
Carrier Frequency	1.8GHz
Maximum Doppler	320Hz
One-sided Maximum Delay	10 μs
Sub-carrier Spacing (Δf)	4.1667KHz
Guard interval	20 μs
Frequency pilot spacing (N_f)	6
Time pilot spacing (N_t)	3
Number of sub-carriers (N_c)	115
Number of OFDM symbols per frame (N_s)	58
N_{taps} *(Wiener)*	10

Table D.1. System parameters used during channel estimation simulations.

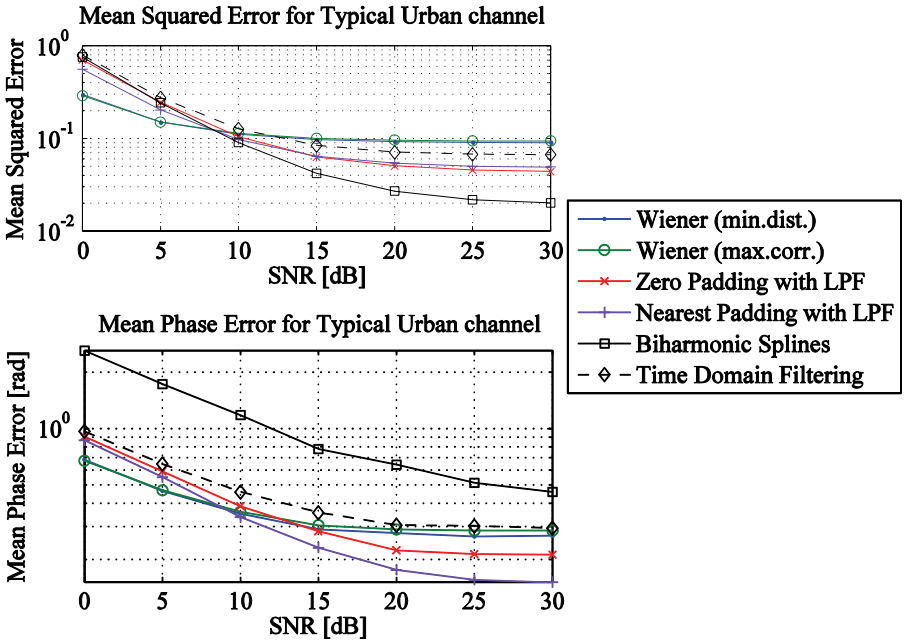


Figure C.5. MSE and MPE evaluation for a Typical Urban channel profile.

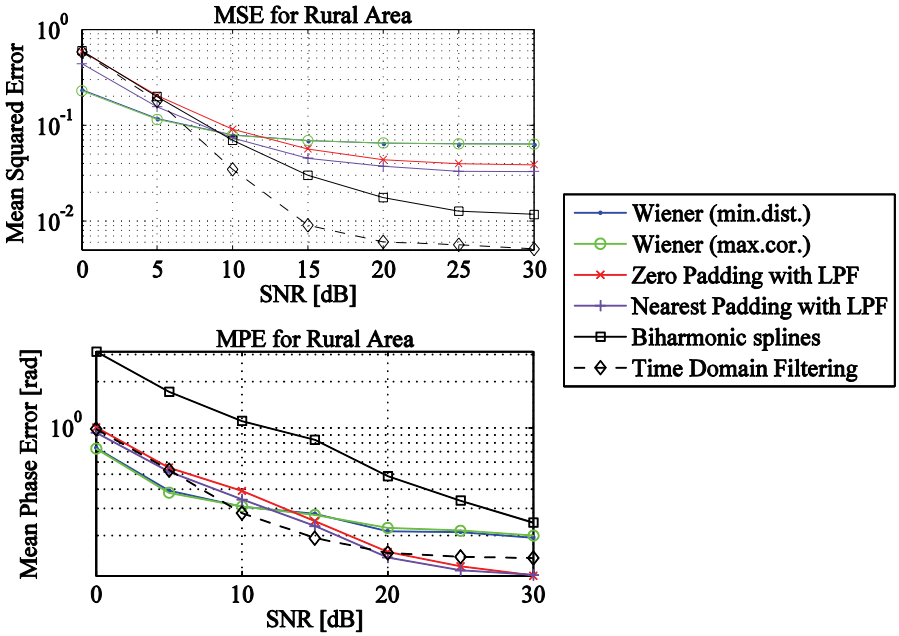


Figure C.6. MSE and MPE evaluation for a Rural Area channel profile.

References

[1] J.J. Van de Beek, O. Edfors, M. Sandell, S.K. Wilson, P.O. Borjesson, "On channel estimation in OFDM systems," IEEE 45th Vehicular Technology Conference, 1995, vol.2, pp.815-819 vol.2, 25-28 Jul 1995.

[2] Ye Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems ," *Vehicular Technology, IEEE Transactions on* , vol.49, no.4, pp.1207-1215, Jul 2000.

[3] R.W. Heath, G.B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *Signal Processing, IEEE Transactions on* , vol.47, no.3, pp.848-856, Mar 1999

[4] P. Höeher, S. Kaiser, P. Robertson, "*Two-Dimensional Pilot-Symbol-Aided Channel Estimation by Wiener Filtering*", *The IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP'97), Vol. 3, pp.1845 - 1848, 21-24 April 1997.

[5] R. González, F. Bader, M. A. Lagunas, "*Adaptive Distribution of Pilot Symbols Aided Channel Estimation in an OFDM System*"**.** The E2R Workshop on Reconfigurable Mobile Systems and Networks Beyond 3G. Barcelona, Spain. Sept 2004.

[6] A. G. Marques, E. Morgado, A. Cano, A. Caamaño, F. J. Ramos, *"Efficient Pilot Patterns in OFDM systems for wireless time and frequency selective channels", URSI '*2005, Sept 2005, Gandía-Valence, Spain.

[7] S. Coleri, M. Ergen, A. Puri, A. Bahai, "*A Study of Channel Estimation in OFDM Systems*". The IEEE Vehicular Technology Conference (VTC 2002-Fall).Vol. 2, pp. 24-28. Sept 2002. Vancouver, BC, Canada.

[8] D.T. Sandwell, "*Biharmonic Spline Interpolation of GEOS-3 and SEASAT Altimeter Data*", Geophysical Research Letters, 2, pp. 139-142, 1987.

[9] I. Gutiérrez, J. Pijoan, F. Bader, R. Aquilué, New Channel Interpolation Method for OFDM Systems by Nearest Pilot Padding, in Proc. European Wireless 2006 (EW2006), Athens, Greece, April, 2006.

[10] Y. Shen, E. Martinez, "Channel Estimation in OFDM Systems", AN-3059, Freescale Semiconductor, Jan. 2006.

Universitat Ramon Llull

Aquesta Tesi Doctoral ha estat defensada el dia ____ d _____ de ____

al Centre _____

de la Universitat Ramon Llull

davant el Tribunal format pels Doctors sotasignants, havent obtingut la qualificació:

President/a

_____

Vocal

_____

Vocal

_____

Vocal

_____

Secretari/ària

_____

Doctorand/a

C. Claravall, 1-3
08022 Barcelona
Tel. 936 022 200
Fax 936 022 249
E-mail: urlsc@sec.url.es
www.url.es