UNIVERSITÀ DI PISA

Scuola di Dottorato in Ingegneria "Leonardo da Vinci"

Corso di Dottorato di Ricerca in
Ingegneria dell'Informazione

SSD: ING-INF/03

Tesi di Dottorato di Ricerca

Adaptive Techniques for Packet-Oriented Transmissions in Future Multicarrier Wireless Systems

Autore:

Riccardo Andreotti _____

Relatori:

Prof. Filippo Giannetti _____

Prof. Ruggero Reggiannini _____

Ing. Vincenzo Lottici _____

Prof. Luc Vandendorpe _____

Anno 2013

Sommario

I sistemi wireless dei prossimi anni sono tenuti a fornire trasmissioni caratterizzate da data rate e affidabilità sempre maggiori per sostenere la sempre più crescente richiesta di applicazioni e servizi eterogenei. Inoltre, tali requisiti devono essere soddisfatti nel difficile ambiente di propagazione rappresentato dal canale wireless, e devono far fronte alla scarsità dello spettro radio disponibile. Per tali sistemi, la tecnica di modulazione multiportante nota come orthogonal frequency division multiplexing (OFDM) è emersa come tecnologia chiave a livello fisico grazie alla sua elevata efficienza spettrale, struttura di ricetrasmissione piuttosto semplice e robustezza al fenomeno di multipath fading.
In tale contesto, questa tesi indaga nuove tecniche adattative in cui le risorse e parametri di trasmissione sono adattati in base alle informazioni sullo stato di canale al fine di fornire una consegna affidabile ed efficiente di pacchetti dati su canali selettivi in frequenza. Queste tecniche, note in letteratura come link resource adaptation (LRA) e resource allocation (RA), sono proposte in combinazione con un numero di funzionalità avanzate come l'efficiente tecnica di codifica di canale chiamata bit interleaved coded modulation (BICM) e meccanismi di hybrid automatic repeat request (HARQ). Diversamente dalla maggior parte dei problemi di LRA e RA considerati nella letteratura, questa tesi si basa sul goodput come figura di merito, definito come il numero di bit di informazione consegnati in pacchetti senza errori per unità di tempo. Quest'ultimo rappresenta, infatti, una metrica adeguata per dare un quadro attendibile delle effettive prestazioni del collegamento caratterizzato da modulazione e codifici pratici, trasmissioni a pacchetto e meccanismi di HARQ.
In dettaglio, i contributi principali della tesi sono: la derivazione di una strategia di LRA che assegna modulazione, tasso di codifica e potenza ad un sistema BIC-OFDM cognitivo; un nuovo metodo di predizione delle prestazioni, che sfrutta la metodologia di effective signal-to-noise ratio (SNR), per sistemi BIC-OFDM impieganti protocolli di HARQ con packet combining; un algoritmo di LRA che seleziona la migliore distribuzione dei bit e tasso di codifica per sistemi BIC-OFDM; uno schema equo di RA che assegna potenza, ordine di modulazione, tasso di codifica e sottoportanti agli utenti sul downlink di un sistema BIC orthogonal frequency division multiple access (OFDMA) col fine di ottimizzare le prestazioni dell'utente avente il valore più basso di goodput.

Abstract

Future wireless systems are expected to provide even more high data rates and reliable communications to support the ever increasing demand of heterogeneous applications and services. Furthermore, these strict requirements must be satisfied in the harsh environment of the wireless propagation channel, and must cope with the scarcity of the available radio spectrum. For such systems, the multicarrier modulation technique known as orthogonal frequency division multiplexing (OFDM) emerged as the key technology at the physical layer thanks to its high spectral efficiency, rather simple transceiver structure and robustness to multipath fading.

Within this context, this thesis investigates novel adaptive techniques where the transmission resources and parameters are adapted according to the channel state information in order to provide reliable and efficient data packet delivering over frequency selective channels. These techniques, known in the literature as link resource adaptation (LRA) and resource allocation (RA), are proposed in combination with a number of advanced features such as the efficient channel coding technique represented by the bit interleaved coded modulation (BICM) and hybrid automatic repeat request (HARQ) mechanisms. Differently from the majority of LRA and RA problems considered in the literature, this thesis focuses on the goodput figure of merit, defined as the number of information bits delivered in error-free packets per unit of time. The latter represents in fact an adequate metric to give a reliable picture of the actual link performance, when features such as practical modulation and coding, packet oriented transmissions and HARQ mechanisms are employed.

In detail, the main contributions of the thesis are: the derivation of a LRA strategy that assigns modulation, coding rate and power to a cognitive BIC-OFDM system; a novel performance prediction method, which exploits the effective signal-to-noise ratio (SNR) methodology, for BIC-OFDM systems featuring HARQ protocols with packet combing; a LRA algorithm that selects the best bit loading distribution and coding rate for BIC-OFDM systems; a fair RA scheme that assigns power, modulation order, coding rate and subcarriers to users on the downlink of a BIC orthogonal frequency division multiple access (OFDMA) system to maximize the performance of the user having the lowest goodput value.

Acknowledgments

I would like to express my gratitude to my supervisors for their guidance and constant support. In particular, I would like to thank Prof. Filippo Giannetti and Ing. Vincenzo Lottici. During these three years of Ph.D. they have always been helpful and available to listen to me and, with my great joy, they have taught me how to address problems and shared with me their valuable technical background. I would also like to thank Prof. Luc Vandendorpe from Université catolique de Louvain (UCL) for being my co-supervisor. He has allowed me to have a precious research experience at UCL and, above all, his office door has always been open for me.
A special thanks goes to Ing. Ivan Stupia that supported me throughout all the Ph.D. years, with precious advices and discussions, and that, most important, I am glad to call friend.
Many thanks to my colleagues of the Department of Information Engineering of the University of Pisa. They have contributed to make the lab a stimulating and friendly research environment. Among them, I would like to especially thank Ing. Leonardo Marchetti. I would also like to thank Prof. Aldo Nunzio D′Andrea and Prof. Ruggero Reggiannini that have provided me with all means to carry out my research activities.
I would like to express my sincere thanks to all the members of the research division Electrical engineering (ELEN) of the Institute of Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM) at UCL. My stay at ICTEAM/ELEN has been very pleasant and full of stimulating discussions. Among them, a special thanks goes to Ing. Tao Wang.
Finally, I would like to thank my parents, friends and my girlfriend that have encouraged me to pursue this Ph.D. They have always been close to me during such an important period of my life and, for this, I will be forever grateful to them.

Pisa, April 2013 Riccardo Andreotti

Contents

List of figures **xi**

List of tables **xiii**

List of acronyms **xv**

List of operators **xix**

1 Introduction **1**
- 1.1 Wireless systems evolution 2
- 1.2 Next generation wireless systems 3
- 1.3 Focus of the thesis 5
- 1.4 Organization of the thesis and major contributions 8

2 Overview of the BIC-OFDM system **11**
- 2.1 Packet oriented radio access model 11
  - 2.1.1 Hybrid automatic repeat request 13
- 2.2 BIC-OFDM system description 15
  - 2.2.1 Error probability analysis 19
- 2.3 Link performance prediction techniques 20
  - 2.3.1 The κESM model 22

3 The goodput criterion **25**
- 3.1 Rationale of the goodput criterion 25
- 3.2 Expected goodput derivation 27
  - 3.2.1 Equivalent BIC-OFDM model 27
  - 3.2.2 Metric derivation 28
  - 3.2.3 Inter-round EGP optimization approach 30

4 LRA for cognitive BIC-OFDM systems **33**
4.1 Dynamic spectrum access . . . 33
4.2 BIC-OFDM cognitive radio scenario . . . 35
4.2.1 The underlay and interweave paradigms . . . 35
4.2.2 Power constraints evaluation . . . 36
4.3 LRA problem formulation . . . 37
4.4 Optimal power allocation . . . 39
4.4.1 Formulation of the PA problem . . . 39
4.4.2 Lagrangian dual decomposition approach . . . 40
4.5 Successive Set Reduction approach . . . 43
4.5.1 Rationale of the SSR approach . . . 44
4.5.2 SSR approach for the EGP optimization problem . . . 45
4.5.3 Existence of the SSR solution . . . 47
4.5.4 SSR algorithm . . . 49
4.6 Extreme points criterion . . . 53
4.7 AMC-SSR algorithm . . . 56
4.8 Simulation results . . . 56
4.9 Concluding remarks . . . 61

5 Hybrid ARQ based LRA for BIC-OFDM systems **67**
5.1 Hybrid ARQ with packet combining . . . 67
5.1.1 Background on HARQ . . . 67
5.1.2 HARQ based BIC-OFDM system . . . 68
5.2 The αESM model . . . 69
5.3 Expected goodput in presence of HARQ mechanism . . . 76
5.3.1 Expected goodput formulation . . . 76
5.4 LRA algorithm with instantaneous CSIT . . . 79
5.4.1 Bit loading algorithm . . . 81
5.4.2 AMC with BL algorithm . . . 84
5.5 LRA with limited CSIT . . . 85
5.5.1 AMC algorithm . . . 86
5.6 Simulation results . . . 86
5.7 Concluding remarks . . . 90

6 Fair RA scheme for BIC-OFDMA systems **95**
6.1 Resource allocation in OFDMA systems 95
6.1.1 Margin and rate adaptive approaches 96
6.1.2 The maximum fairness problem 97
6.2 BIC-OFDMA system model 98
6.3 Max-min goodput problem 100
6.3.1 Problem formulation 100
6.3.2 Iterative algorithm for the MMG problem 102
6.4 Power allocation optimization problem 103
6.5 AMC Algorithm 106
6.6 Subcarrier allocation problem 108
6.6.1 Branch and bound based SA-OP 108
6.6.2 ACO based SA-OP 116
6.7 Simulation Results 121
6.8 Concluding remarks 125

A Appendices **131**
A.1 Basics of convex optimization theory 131
A.1.1 Duality 131
A.1.2 A brief summary on the subgradient method 133
A.2 MGF evaluation 134

Bibliography **137**

List of publications **149**

List of Figures

1.1 Wireless systems evolution. 4
1.2 Spectrum activity in North America TV bands. 6

2.1 Simplified radio access model. 12
2.2 Example of multiple-channel ARQ with 8 logical channels. 14
2.3 BIC-OFDM system model. 16
2.4 κESM quality model structure. 23

3.1 Equivalent model for the ARQ BIC-OFDM system. 27

4.1 Spectrum activity 36
4.2 Multi-level water-filling interpretation 44
4.3 Graphical interpretation of the extreme points criterion. 54
4.4 Simulation scenario. Example with $Q_U = 2$, $Q_I = 3$. 59
4.5 LDD vs. SSR algorithm. Performance comparison. 63
4.6 LDD vs. SSR algorithm. Convergence comparison. 63
4.7 Estimated vs. actual goodput comparison. 64
4.8 AMC-SSR vs. AMC-SL algorithm. Performance comparison with convolutional codes. 64
4.9 AMC-SSR vs. AMC-SL algorithm. Performance comparison with turbo codes. 65

5.1 Equivalent BIOS model. 70
5.2 Block diagram of the αESM update. 76
5.3 Equivalent model for the HARQ BIC-OFDM system.. 77
5.4 HybridGP-LRA with BL vs. GP-LRA with BL. AGP Performance comparison. 91

5.5 HybridGP-LRA with BL vs. GP-LRA with BL. Chunk approximation. 91
5.6 HybridGP-AMC with BL vs. HybridGP-AMC and GP-AMC. AGP performance comparison. 92
5.7 HybridGP-LRA with BL vs. GP-LRA with BL. Packet dropped rate comparison. 92
5.8 HybridGP-LRA with BL vs. GP-LRA with BL. Retransmission rate comparison. 93
5.9 HybridGP-LRA with BL vs. GP-LRA with BL. QoS performance comparison. 93

6.1 BIC-OFDMA system model. 98
6.2 Multi-user SW. 100
6.3 Branch and bound toy example. 115
6.4 ACO graph evolution. 121
6.5 PER fitting model. 126
6.6 Performance comparison of B&B and ACO algorithms. 126
6.7 AGP values for different initial couples of (m, r). 127
6.8 Performance comparison for different values of ACO parameters ρ, δ_φ, N_{it}, N_{a}. 127
6.9 Average number of iterations of the CAM-MMG algorithm. 128
6.10 AGP performance of $Q = 2$ users obtained with CAM-MMG and static RA policy. 128
6.11 AGP performance of $Q = 3$ users obtained with CAM-MMG and static RA policy. 129
6.12 PER performance of $Q = 2$ users obtained with CAM-MMG for different QoS thresholds. 129

List of Tables

4.1 Pseudo-code of the subgradient-based EGOPA-OP 42
4.2 Pseudo-code of the SSR algorithm 56
4.3 Pseudo-code of the AMC-SSR algorithm 57
4.4 Parameters and features of the cognitive BIC-OFDM system. 58
4.5 Parameters and features of the radio propagation channel model. . . . 59

5.1 Pseudo-code of the bit loading algorithm 85
5.2 Pseudo-code of the AMC algorithm with BL 86
5.3 Pseudo-code of the AMC algorithm for EGP maximization with limited CSIT 87
5.4 Parameters and features of the HARQ BIC-OFDM system. 88
5.5 Parameters and features of the radio propagation channel model. . . . 89

6.1 Pseudo-code of the coordinate ascent method for the MMG-OP 103
6.2 Pseudo-code of EPA-OP algorithm 106
6.3 Pseudo-code of AMC-OP algorithm 107
6.4 Pseudo-code of RSA-OP algorithm 111
6.5 Pseudo-code of RSA-OP.b algorithm 112
6.6 Pseudo-Code of the ACO based algorithm for SA-OP.2 120
6.7 Parameters and features of the ARQ BIC-OFDMA system. 122
6.8 Parameters and features of the radio propagation channel model. . . . 123
6.9 Parameters values of the PER Model. 123

List of Acronyms

αESM	aggregate ESM
κESM	cumulant moment generating function based ESM
ACK	acknowledgment
ACO	ant colony optimization
AGP	actual goodput
AMC	adaptive modulation and coding
AMPS	advanced mobile phone system
ARQ	automatic repeat request
AWGN	additive white Gaussian noise
B&B	branch and bound
BIC(M)	bit-interleaved coded (modulation)
BIOS	binary input output symmetric
BL	bit loading
BPSK	binary phase-shift keying
BS	base station
CAM	coordinate ascent method
CBS	coded binary symbols
CDMA	code division multiple access
cEGOPA	cost based EGOPA
CMGF	cumulant moment generating function
COP	combinatorial optimization problem
CP	cyclic prefix
CR	cognitive radio

CRC	cyclic redundancy check
CSI	channel state information
CSIT	channel state information at the transmitter
DFT	discrete Fourier transform
DSA	dynamic spectrum access
EESM	exponential based ESM
EGP	expected goodput
EGOPA	EGP oriented PA
ESM	effective SNR mapping
ESNR	effective signal-to-noise ratio
FCC	Federal Communications Commission
FDMA	frequency division multiple access
FEC	forward error correction
FFT	fast Fourier transform
FM	frequency modulation
GP	goodput
GPRS	general packet radio service
GSM	global system for mobile communications
HARQ	hybrid automatic repeat request
HSDPA	high speed downlink packet access
IBI	interblock interference
IDFT	inverse discrete Fourier transform
IP	Internet protocol
IPU	interweave PU
ISM	industrial, scientific and medical
ITU	International Telecommunication Union
KKT	Karush-Kuhn-Tucker
LDD	Lagrangian dual decomposition
LLR	log-likelihood ratio
LOS	line of sight

LRA	link resource adaptation
LTE	long term evolution
LUT	look-up table
MAC	medium access control
MAP	maximum a posteriori
MGF	moment generating function
MIESM	mutual information based ESM
MILP	mixed integer linear programming
MIMO	multiple input multiple output
MMG	max-min goodput
MPEC	mathematical program with equilibrium constraints
NACK	non-acknowledgment
NP	nondeterministic polynomial-time
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OOB	out-of-band
OP	optimization problem
PA	power allocation
PDS	pairwise decoding score
PDU	protocol data unit
PEP	pairwise error probability
PER	packet error rate
PR	protocol round
PU	primary user
QAM	quadrature amplitude modulation
QoS	quality of service
QVI	quasi variational inequality
RA	resource allocation
RHS	right hand side
RLC	radio link control

RR	radio resource
RRM	radio resource management
RSA	relaxed subcarrier allocation
RV	random variable
SA	subcarrier allocation
SBCQ	sequentially bounded constrained qualification
SINR	signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
SL	step ladder
SR	selective repeat
SRx	secondary receiver
SSR	successive set reduction
STx	secondary transmitter
SU	secondary user
SW	stop-and-wait
TACS	total access communication system
TDMA	time division multiple access
TM	transmission mode
TP	transmission parameter
UMTS	universal mobile telecommunications system
UPU	underlay PU
w.g.l.	without loss of generality
w.r.t.	with respect to
Wi-Fi	wireless fidelity
WiMAX	worldwide interoperability for microwave access
WLAN	wireless local area network
WMAN	wireless metropolitan area network
WPAN	wireless personal area network

List of Operators

absolute value	$\lvert\cdot\rvert$
Cartesian product of the sets $\mathcal{X}_i$	$\prod_i \mathcal{X}_i$
composition of functions	$\circ$
conjugate transposition	$(\cdot)^{\mathrm{H}}$
diagonalization of vectors into diagonal matrices or matrices into block-diagonal matrices	$\mathrm{diag}(\cdot)$
elementwise greater	$\succ$
elementwise greater or equal	$\succeq$
floor function (rounding downward)	$\lfloor\cdot\rfloor$
gradient	∇
infimum of function f over set $\mathcal{X}$	$\inf_{x\in\mathcal{X}} f(x)$
maximum between 0 and x	$[x]^+$
projection of x over set $\mathcal{X}$	$[x]_{\mathcal{X}}$
statistical expectation w.r.t. RV x	$\mathrm{E}_x\{\cdot\}$
transposition	$(\cdot)^{\mathrm{T}}$
vector norm	$\lvert\lvert\cdot\rvert\rvert$

Chapter 1

Introduction

In the last years, wireless communication systems have been playing a starring role in everyday life, affecting the social behaviors as well as the working habits of millions of people. Not surprisingly, wireless communication is the fastest growing segment of the communication industry. Many examples of this pervasive phenomenon can be daily experienced. Business users equipped with smartphones can access every time the Internet exploiting either the cellular network or available Wi-Fi hot spots in order to check emails, control bank accounts, public transport time schedule, and so on. Private users have the possibility to share photos, videos, check the latest news and their accounts on social forums. Moreover, such a rapid diffusion is experienced not only by cellular devices but also by computer laptops and tablets, that combine the portability of the cellular phones with the higher performance and multitasking possibility typical of computers. ITU statistics [1] report that in 2011 the number of mobile subscribers reached the 86% of the worldwide population whereas in 2009 the market sales of laptops overcame the ones of desktop PC and a sale of over 230 millions of notebooks is foreseen for the year 2013. With such an ever increasing demand of multimedia services *in every place at every time*, next generation wireless systems, i.e., the advanced fourth generation (4G) and the fifth generation (5G), are required to supply even more high data rates and reliable communications, facing several obstacles such as the efficient utilization of the transmission resources and the harsh propagation conditions of the wireless environment.

1.1 Wireless systems evolution

The first generation of cellular systems (1G), such as Advanced Mobile Phone System (AMPS) in the United States and the Total Access Communication System (TACS) in Europe and Japan, was commercially available in late 1970s and early 1980s [2]. These systems were based on an analog technology and supported voice only services. In particular, frequency modulation (FM) over the 800-900 MHz band was employed for voice transmission and frequency division multiple access (FDMA) was the technique adopted to share the medium. In practice, a portion of the electromagnetic spectrum was entirely and exclusively allocated to a certain user for whole call duration.
The second generation, such as GSM (Global System for Mobile Communications) [3], entered the market in early 1990s and it was based on a digital technology, where the users data were divided in frames and transmitted in specified and exclusive use time slots (time division multiple access technique, or TDMA for short), combined with FDMA. Thanks to the digital technology, the 2G offered lower device costs with respect to the 1G and a greater spectral efficiency. Moreover, its evolution for supporting packet-based communications, i.e., the General Packet Radio Service (GPRS), allowed the possibility to exploit services as e-mails and web browsing. These seemingly little improvements led to the success of the cellular devices, obtaining in 2001 more than 500 millions of subscribers.
The ever increasing demand of multimedia application called for a new generation of cellular networks, the 3G, such as Universal Mobile Telecommunications System (UMTS) [4], [5] in Europe and cdma2000 [6] in U.S.A., able to manage the growing number of users offering higher data rates and improved quality of service. These systems were based on a combined FDMA/CDMA (code division multiple access) technique, where bunch of signals were assigned to different subbands and the ones sharing the same subband were distinguished by quasi-orthogonal spreading codes. For the first time, the radio resource management was no more static, offering the possibility of variable bit rate depending on the application, of supporting asymmetric services in downlink and uplink and of assigning bandwidth on demand, increasing in this way the spectral efficiency of the overall system.

A few years after the 1G entered the market, also the wireless networks make their debut but with a different vision. The exchanged information is in fact based on

data packets, that is, streams of bits coming from a digitized analog signal or directly from a digital source that are eventually grouped into packets. The first commercial devices for the wireless local area networks (WLAN) were introduced in 1985. Though, they had high prices and low performance, due to limitations on the employed transmission power since, operating over the ISM band, they couldn't interfere with the primary services operating over the same band. A new interest and, consequently, a great diffusion of this technology, occurred at the beginning of 2000s towed by the increasing requests of multimedia services and applications. Several standards were developed according to the application type and required coverage. The most popular is probably the IEEE 802.11 WLAN standard family. The IEEE 802.11a [7] and 802.11g [8] standards, operating over the 5 GHz and 2.4 GHz respectively, are based on multicarrier modulation, in particular, on the orthogonal frequency division multiplexing (OFDM). The IEEE802.11g in particular reach up to 54 Mbps. The IEEE 802.11b [9] employs direct-sequence spread spectrum techniques and reaches 11 Mbps. The coverage is up to 1 km in line of sight and the topology can be a star network with a centralized access point or an ad-hoc network. The natural extension of the WLAN is a broadband network able to cover wide areas, such as the wireless metropolitan area networks (WMAN) [10], which relies on the orthogonal frequency division multiple access (OFDMA) technology. Many other standards can be cited such as the family of IEEE 802.15 standards [11], for personal area networks, that includes, for instance, Ultrawideband [12] technology for high-speed and short distance communications.

1.2 Next generation wireless systems

As apparent, up to the early 2000s, the communication industry was very fragmented. On the one hand, the cellular networks, mainly oriented to voice communications, and thus characterized by relatively low data rate and high bit error rate probability, but also by strict delay requirements, a huge total coverage area and mobility. On the other hand, the wireless networks characterized by high data rates and more reliable communications, therefore suitable for application such as data transfer or video streaming, but lower coverage and difficult roaming.
Driven by the ever increasing demand of heterogeneous multimedia applications, which range from voice to data transfer, from real time video to web browsing,

next generation wireless systems, i.e., the advanced 4G and 5G, bring an end to this fragmentation, trying to provide even more high data rates, reliable communications, improved quality of service, enhanced coverage and mobility, integrating and improving the points of strength of these different technologies, as depicted in Fig. 1.1.

The original vision of the so-called 4G [13] was the achievement of at least 100

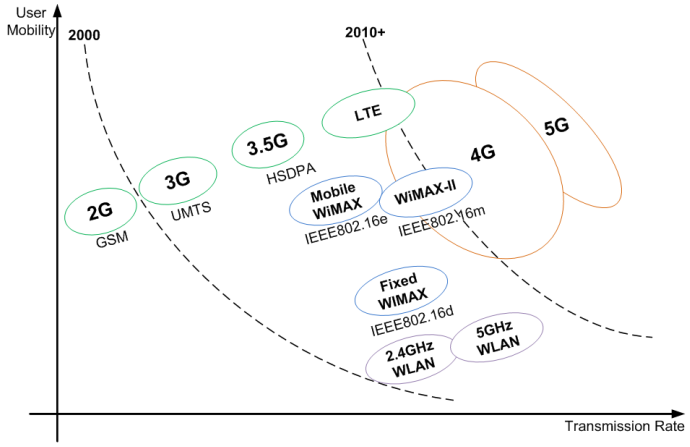

Figure 1.1: *Wireless systems evolution.*

Mbps peak rates in full-mobility wide area coverage and 1 Gbps in low-mobility local area coverage services, the possibility to connect everywhere and every time (ubiquitous systems), a very smooth global roaming among different systems and an entirely IP-based communication applying packet switching method. Finally, multiple wireless air interfaces should be integrated in a single common platform, exploiting the solution offered by the open wireless architecture [14], where the majority of transceiver functions are executed as a software. The 5G, first proposed in [15], pushes even more forward the 4G vision, adding key concepts such as, among the others, cooperative relays, cognitive radio networks and a user-centric more than a service-centric approach [16].

Currently, 4G networks either actually use enhanced 3G technologies or offer only a minimal implementation of 4G technologies [13]. The 3G networks that have been extended to 4G are wideband code division multiple access (W-CDMA), which is likely to be phased out, its evolution into high speed dowlink packet access (HSDPA) and WiMAX [17], based on OFDM technology with multiple-input multiple output

(MIMO) antenna configuration, capable of obtaining downlink peak data rates up to 75 Mbps. In WiMAX, the coverage can reach up to 50 km, allowing users to get broadband connectivity in non line-of-sight conditions. IEEE 802.16e (Mobile WiMAX), designed to fill the gap between WLAN and high mobility cellular wide area networks, comes with enhanced quality of service (QoS) and mobility up to 120 km/h. Finally, other implementations of 4G networks are long term evolution (LTE) and LTE Advanced [13], [17], which are considered to have the most promising future. These standards are still based on MIMO-OFDM technology and offer peak rates up to 100 Mbps in downlink, reduced latency (less than 10 ms) other than flexibility and scalability in deployment.
Even if these existing technologies do not meet all the envisaged requirements, many related advances are likely to affect how 4G and even futuristic 5G networks will be designed. For instance, excluding the W-CDMA, all the other standards employ OFDM as physical layer technology, which is in fact considered as the basis of future generation wireless systems.

1.3 Focus of the thesis

One of the major challenges of the future wireless systems is how to guarantee the stringent services requirements in the harsh environment represented by the wireless communication channel. First of all, the transmitted signals are received after being scattered, reflected and diffracted by the surrounding objects. The receiver therefore observes multiple delayed and attenuated copies of the original signal, that can constructively or destructively combine introducing large variations on the overall signal strength. Moreover, the relative motion between transmitter and receiver makes the channel characteristics randomly vary during time. Another limiting factor is represented by the availability of the radio spectrum. The latter is in fact a very precious resource that is assigned by local government to companies at very high prices. The damage caused by this assignment policy is manifold. First of all, only the biggest companies are able to purchase it, creating an oligopoly that hurts competition and therefore innovation and technology development. Furthermore, the fixed allocation of radio spectrum lead sometimes to its underutilization in a certain location and/or at a certain time, as shown in Fig. 1.2 where the spectrum activity over the North America TV bands is depicted, and other times can be insufficient to

offer the required QoS.

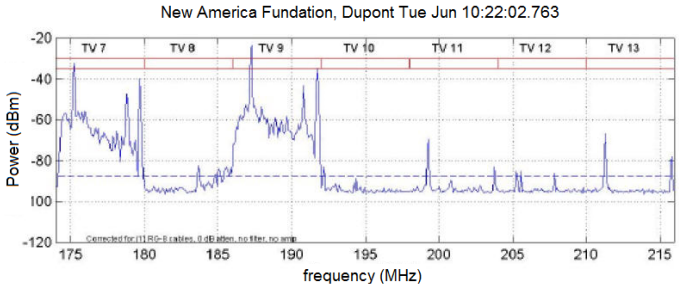

Figure 1.2: *Spectrum activity in North America TV bands.*

Thus, the efficient utilization of the available resources in the limited spectrum is a key point for future wireless systems. For this reason, this thesis studies adaptive transmission techniques, in order to face the above mentioned obstacles and obtain a reliable and fast communication satisfying the required QoS over the constrained set of transmission resources. The OFDM technology in combination with a number of advanced features, that will be the focus of the dissertation, can provide reliable and efficient data packet transmissions over harsh fading channels.

The OFDM modulation is a multicarrier communication technique that allows to efficiently and reliably transmit data over a radio channel, even in multipath environment. The information data is transmitted over a large number of narrowband carriers, or subcarriers. These are closely and equally spaced in a way that they are orthogonal to each other, so that there is not interference between them. The result is a very high spectral efficiency, simple equalization and a fast implementation via fast Fourier transform (FFT). Moreover, OFDM technology is extremely flexible and thus able to adapt to the current channel condition and service requirements. This signaling technique can be efficiently combined with error correcting codes, exploiting for instance the Bit-Interleaved Coded Modulation (BICM) approach [18], that improves code diversity and avoids the need for the complicated and somewhat less flexible design typical of coded modulation.

As outlined above, the radio link quality significantly varies during time due to frequency-selective fading. A static set-up of the transmission parameters (e.g. mod-

ulation order, coding rate, power) can lead to very poor system performance when the channel conditions are bad or, on the contrary, to a waste of the transmission resources when the channel conditions are good. *Link resource adaptation* (LRA) [19], [20] is therefore a fundamental aspect for future wireless systems since it deals with how to efficiently set the transmission parameters of a radio link to handle variations of the radio link quality. In 3G systems, for instance, dynamic transmit-power control has been used, where the allocated power is dynamically adjusted according to the link quality to guarantee a constant data rate (which is a desirable property for voice services). For packet-data traffic, a more appealing property is simply to transmit with the highest possible rate. This can be obtained by means of dynamic rate control, where, in practice, the data rate is controlled by adapting the modulation order and coding rate to the actual channel conditions. Intuitively, the better the channel quality, the higher the modulation orders and code rates employed. For this reason, this link adaptation technique is also referred to as adaptive modulation and coding (AMC) [19].

A further help to the link resource adaptation is given by the advanced retransmission scheme known ase *hybrid automatic repeat request* (HARQ) with soft combining [21], [22]. With this mechanism, the erroneously received copies of a certain packet, detected via cyclic redundancy check (CRC), are stored at the receiver, which in turn requires a new transmission of the same packet. The transmitter sends then a novel coded copy and the receiver combines the bit-level soft metrics of the latter with the ones of the previous stored copies. The combined metrics result therefore more reliable improving probability of successfully decoding. Moreover the transmitter can decide to send the same set of coded bits representing the original message (Chase Combining case) or a different subset (Incremental Redundancy case), trading off in this way reliability and spectral efficiency.

OFDM can also be adopted as a multiplexing technique or multiple-access scheme, with different subsets of the overall set of available subcarriers used to communicate with different mobile terminals. In this case, the term OFDMA is also often used and *Channel-Dependent Scheduling* stems as a powerful paradigm to perform dynamic allocation of the available radio resources [20], [23], such as subcarriers, power, coding rate and bits, to all the users. The basic idea is to take advantage of the rapid and random variations of the frequency-selective fading channel experienced by each user. This principle is usually called multiuser diversity.

In this scenario, information-theoretic performance limits, which rely on ideal assumptions like Gaussian inputs and infinite length codebooks, can be inadequate to give a reliable picture of the actual link performance when practical modulation and coding are applied. For this reason, the thesis will focus on a metric based the number of information bits delivered in error-free packets per unit of time [24], or *goodput* (GP) for short. This figure of merit is in fact capable of exploiting all the above mentioned features such as packet-oriented transmission, ARQ mechanism as well as the effect of practical modulation and coding schemes. Finally, since the above mentioned LRA problems can be represented as constrained optimization problems, where the objective function, i.e., the performance metric, is optimized over the constrained set of the transmission parameters, it is required an effective way to predict the system performance, accounting for both the channel state information (CSI) and the information relevant to the transmitted-signal mode choice (i.e. modulation size, coding rate, etc.). Specifically, for the OFDM system under analysis, where the frequency selective channel makes the signal-to-noise ratio (SNR) level largely vary across the subcarriers, the powerful link performance prediction approach given by the *effective SNR* (ESNR) methodology [25], [26] is exploited.

1.4 Organization of the thesis and major contributions

The dissertation is organized as follows.

Chapter 2 presents a technical background on the BIC-OFDM system under analysis. A brief description of the radio access model is first given and then an overview of the key features of the BIC-OFDM system is reported. Finally, the chapter summarizes the main link performance prediction techniques, that is, how to predict the performance of the system when a certain setting of transmission parameters is employed for a given status of the channel.

Chapter 3 first introduces the rationale of the GP figure of merit. Then, the GP metric is derived resorting to the renewal reward theory. Finally, it is introduced the concept of inter-round GP optimization as an effective method to adapt the transmission parameters of the BIC-OFDM system at each ARQ protocol round.

Chapter 4 focuses on the LRA problem aimed at maximizing the goodput for cog-

nitive BIC-OFDM systems. According to the cognitive paradigm, the BIC-OFDM system can transmit over the same frequencies assigned to the licensed users, as long as the interference caused to the latter is kept below a prescribed threshold. Particular emphasis is thus given to the power allocation (PA) problem. First, it is demonstrated that it is a convex optimization problem and, accordingly, it is solved with conventional numerical methods. Then, in order to circumvent several drawbacks typical of these methods, such as slow convergence and need for parameters tuning, a novel PA framework, called successive set reduction (SSR), is proposed resorting to the Quasi Variational Inequality (QVI) theory and its optimality condition is analytically derived. Based on this PA algorithm, an AMC scheme is eventually derived. The chapter ends with simulation results showing the effectiveness of the proposed SSR-based AMC algorithm. The GP improvements over non-adaptive LRA techniques are indeed highlighted, besides a remarkable complexity reduction compared to conventional numerical methods.

Chapter 5 investigates the impact of the HARQ mechanism with soft combining on the performance of the BIC-OFDM system. First, a novel link performance prediction approach for such a system is proposed based on the effective SNR methodology. Capitalizing on this novel prediction method, tagged as *aggregate* effective SNR mapping technique, or αESM for short, a link adaptation strategy, that at each round of the HARQ protocol sets the transmission parameters to maximize the GP, is derived. Thanks to the novel αESM technique, the link adaptation strategy exploits the information offered by the HARQ feedback channel about the past failed transmissions along with the information on the actual channel conditions. In particular, two different cases of level of knowledge about the actual channel state are considered. When the only channel statistics are available at the transmitter side, the LRA consists of an AMC algorithm. Conversely, when the instantaneous channel conditions are perfectly known at the transmitter side, the proposed LRA algorithm is further refined deriving a bit-loading procedure that assigns different modulation orders across the subcarriers. Finally, experimental results obtained by simulation over a realistic wireless scenario show the improvements of the proposed LRA based on the αSNR technique over the case where only information on the actual channel state are taken into account.

Chapter 6 investigates the downlink of a BIC-OFDMA base station and, in particular, the problem of how to assign the radio resources to each user served by

the base station. The chapter begins with an overview of the main approaches usually adopted for resource allocation. Then, the fairness criterion is introduced as solution to the unbalanced treatment of the users when typical rate-adaptive and margin-adaptive optimization criterions are employed. Hence the resource allocation problem aimed at maximizing the performance of the user with the minimum GP, also called max-min problem, is stated. In particular, the proposed algorithm that solves the max-min problem returns the power, subcarrier, modulation order and coding rate allocation for each user. Each of these allocation problems is investigated, with particular focus on the subcarrier allocation problem. Detailed performance analysis providing guidelines for the algorithm implementation in typical wireless scenario is finally presented. Simulation results show the effectiveness of the proposed approach, highlighting the improvement on the minimum GP value as well as QoS satisfaction.

Chapter 2

Overview of the BIC-OFDM system

In this chapter, the technical background that constitutes the basis of the dissertation is given. In particular, Sections 2.1 and 2.2 describe the radio access model and the BIC-OFDM system characteristics, respectively, while Sect. 2.3 focuses on link performance prediction techniques for multicarrier systems.

2.1 Packet oriented radio access model

In future wireless systems circuit-switching networks will make way to a packet oriented communication. For this reason, the fundamental piece of information to be transmitted over the radio interface is identified in the packet [27]. Since it is expected that transport of Internet protocol (IP) packets will dominate the traffic in wireless systems, a packet will typically correspond to an IP packet. However, it could also correspond to other kinds of information to be communicated over the radio interface, including, e.g., Layer 3 control signaling. Thus, in order to obtain a flexible and simple design avoiding increased cost in development and operation, a key characteristic of the radio-access framework is that all types of information to be transported over the radio interface are processed in a more or less identical way as follows [27]:
1) the *packet processing*, that includes channel coding and the retransmission protocol;
2) the *frame processing*, that includes different multiplexing and modulation steps and maps to the basic physical resource. Assuming OFDM-based transmission, the basic

physical resource can be expressed as a unit in the time/frequency grid.

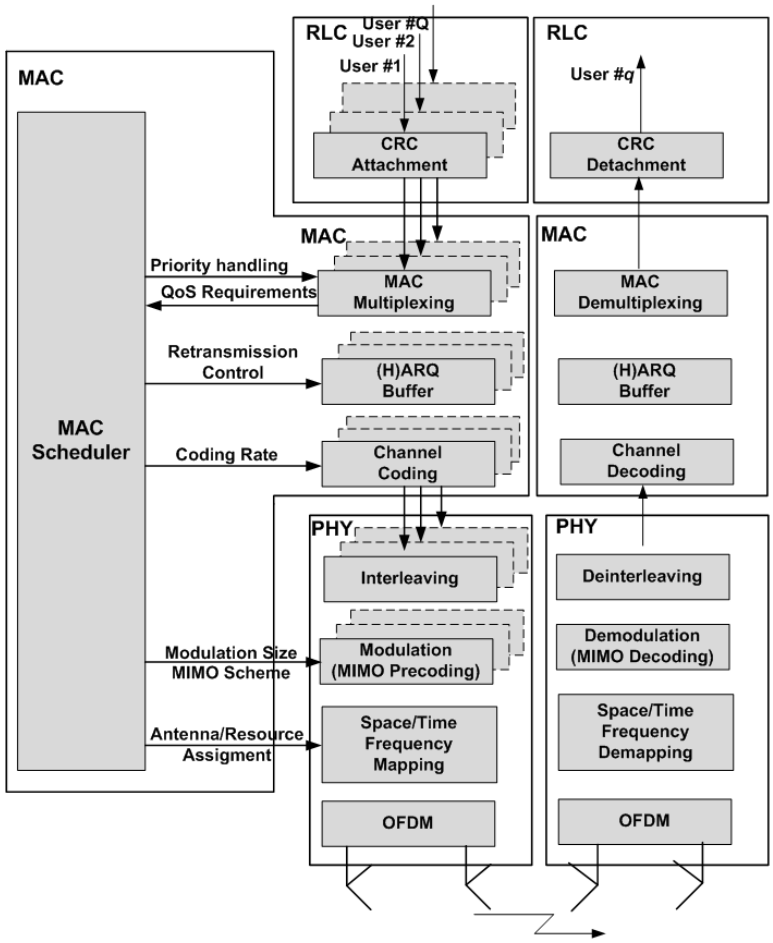

Figure 2.1: *Simplified radio access model.*

This model is depicted in Fig. 2.1, along with the scheduler, a *cross-layer* entity that makes decisions taking into account the QoS requirements, as well as the instantaneous radio-channel conditions (channel-dependent scheduling). The goal of resource allocation/link resource adaptation is in fact to achieve a high "benefit" for the users on the one hand, while making efficient use of the limited resources of the wireless link on the other hand [28].

In detail, the packet processing, that acts on top of the frame processing, provides powerful means to handle varying packet sizes and changing channel conditions, integrating packets handling, retransmissions, and channel coding in one functional block. It relies on the following key design guideline allowing an efficient cross-layer approach: "packets must be kept as integral units as far down in the protocol stack as possible". Accordingly, there is a one-to-one mapping of packets (e.g., IP packets or control messages), previously stored in queues, to retransmission units of the radio link control (RLC) protocol, also denoted as RLC protocol data units (RLC-PDUs). Each of them consists of an header, a payload and a cyclic redundancy check (CRC) field. Then, one RLC-PDU is mapped to exactly one forward error correction (FEC) block, i.e., once again a one-to-one mapping. Thus, the receiver can decode each packet (or RLC-PDU) independently. Finally, the output of the FEC block can be punctured, producing variable-sized fragments, named RLC fragments. This feature can be efficiently combined with the error control mechanism represented by the ARQ mechanism or the HARQ mechanism, to adapt to different channel conditions. One or more RLC fragments of a single user are then grouped to form a MAC PDU, together with MAC PDUs of other users in case of multiple users packets queues. These data are then handed over to the frame processing module. It is evident how link resource adaptation strategies suit very well to be adopted with this design, since the latter optimizes the delivery of packets and not the reception of parts thereof.

In the following Sect. 2.2, a detailed description of the processing at the MAC and physical layers is given for a single user case. The multi-user case will be described later on in Sect. 6.2.

2.1.1 Hybrid automatic repeat request

ARQ is a powerful mechanism that controls the transmission errors in order to deliver error-free data to the users. In its simplest form, the stop-and-wait (SW) [21] scheme, the transmitter sends a packet (i.e., an RLC-PDU) and waits for the reception, on the feedback channel, of an acknowledgment (ACK) or a non-acknowledgment (NACK) before transmitting the following packet. The correct reception of the packet is detected via CRC. In particular, upon the reception of an ACK a novel packet is transmitted, contrariwise, the same packet is transmitted if a NACK is received. This mechanism requires minimum overhead but wastes a lot of resource in that

the system remains idle until the reception of the feedback. The selective repeat (SR) [29] protocol continuously transmits packets and performs retransmissions only for the erroneously received ones. Hence it would be the most efficient protocol, but a receiver should have an infinite buffer so that a transmitter can keep sending data all the time while retransmitting erroneous ones. Thus, in practice, multiple channel SW protocols are adopted [30], [31], where a certain number of parallel SW logical channels over a single link is employed, so that the transmission can be considered continuous. As depicted in Fig. 2.2, when the round trip time (i.e. maximum time allowed to receive an ACK/NACK) is less than 8 transmission time interval, that is the temporal slot assigned to the generic logical channel to transmit the RLC-PDU, then a multiple-channel SW with 8 logical channels allows the transmission to be considered continuous.

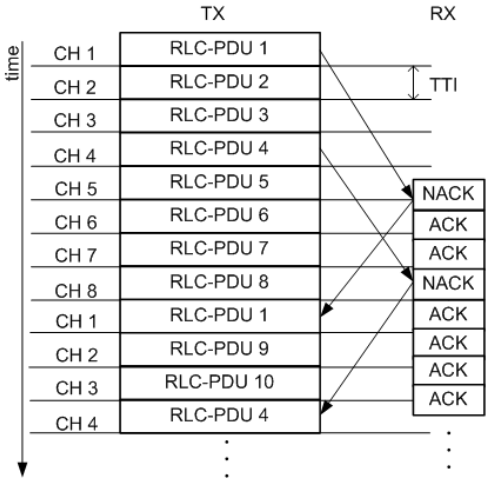

Figure 2.2: *Example of multiple-channel ARQ with* 8 *logical channels.*

In poor channel conditions, the only ARQ mechanism suffers of a great loss of throughput since the error rate is high. For this reason, in nowadays systems the ARQ mechanism is practically always combined with the FEC scheme, which helps in increasing the reliability of the message. Such an approach, which is also called HARQ [21], basically transmits a message after being processed by the CRC block and the FEC encoder. If after the decoding and CRC evaluation the message is received in error, the transmitter sends again the message encoding it with an amount of

redundancy that depends on the particular scheme employed [21], [22]. As detailed in the radio access processing described above, we will rely on this combined error detection and error correction scheme and will refer to it as ARQ for simplicity. The amount of redundancy sent at each transmission is determined by the particular LRA strategy employed, as it will be clear in the following. A further improvement is offered when the receiver stores the received copies relevant to the same packet and combines them to create a packet that is more reliable than its constituent parts. This mechanism is called HARQ with soft combining [20] and will be analyzed in Chapter 5.

2.2 BIC-OFDM system description

Transmitter structure

This section describes the BIC-OFDM system based on the packet and frame processing outlined in Sect. 2.1 and transmitting over a band B composed of N subcarriers belonging to the set $\mathcal{D}_s \triangleq \{1, \cdots, N\}$. The equivalent system model is depicted in Fig. 2.3

Each packet, identified by a RLC-PDU made of $N_{\mathrm{u}} = N_{\mathrm{h}} + N_{\mathrm{p}} + N_{\mathrm{CRC}}$ bits, i.e., including the header, the payload and the CRC sections, of size N_{h}, N_{p} and N_{CRC}, respectively, is sent to the receiver within L attempts, that is, the maximum number of ARQ protocol rounds (PRs). At the generic PR ℓ, each RLC-PDU is processed in two steps.

In the first, called packet processing step, the RLC-PDU is input to the channel encoder whose coding rate $r^{(\ell)}$ is chosen in the set of punctured rates $\mathcal{D}_r \triangleq \{r_0, \cdots, r_{\max}\}$, where r_0 is the mother code rate and $r_{\max}$ the minimum code rate. The resulting block consists of $N_{\mathrm{c}}^{(\ell)} = N_{\mathrm{u}}/r^{(\ell)}$ coded binary symbols (CBS), which are eventually randomly interleaved according to the BICM model.

In the frame processing that follows, the coded information is mapped onto the physical resources available in the time-frequency grid. The bit-level interleaver, along with the labeling map μ, maps the sequence of CBS $\{b_k^{(\ell)}\}_{k=1}^{N_{\mathrm{c}}^{(\ell)}}$ into a sequence of quadrature amplitude modulation (QAM) symbols $\{x_{j,n}^{(\ell)}\}$, where j denotes the OFDM symbol within the OFDM frame, with $1 \leq j \leq N_{\mathrm{ofdm}}$, and n identifies the subcarrier index, with $1 \leq n \leq N$. In detail, dropping for the sake of notation the dependence on the OFDM symbol index j, the interleaved sequence of CBS is

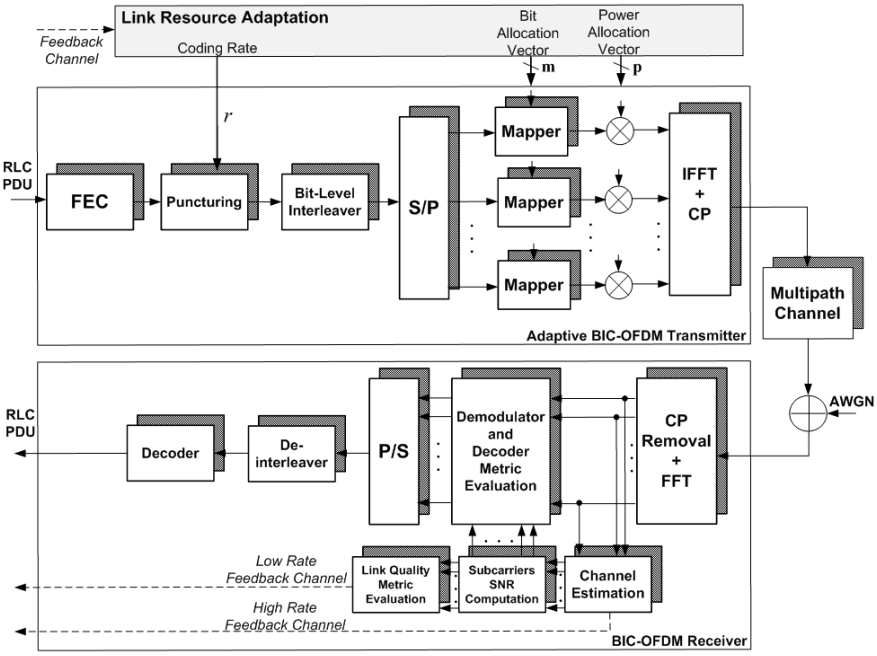

Figure 2.3: *BIC-OFDM system model.*

broken into subsequences of $m_n^{(\ell)}$ bits each, which are Gray mapped onto the unit-energy symbols $x_n^{(\ell)} \in 2^{m_n^{(\ell)}}$-QAM constellation, with $m_n^{(\ell)} \in \mathcal{D}_m = \{2, 4, \cdots, m_{\max}\}$. This means that the index k of the CBS $b_k^{(\ell)}$ is one-to-one mapped into a set of two coordinates (n_k, i_k), that is, $b_k^{(\ell)}$ occupies the i_kth position within the label of the $2^{m_n^{(\ell)}}$-QAM symbol sent on the n_kth subcarrier.

The modulation symbols are re-arranged into the vector $\mathbf{x}^{(\ell)} \triangleq \left[x_1^{(\ell)}, \cdots, x_N^{(\ell)}\right]^{\mathrm{T}}$ and allocated over the N available subcarriers along with a certain amount of power $\mathbf{p}^{(\ell)} = [p_1^{(\ell)}, \cdots, p_N^{(\ell)}]^{\mathrm{T}}$, where $p_n^{(\ell)}$ denotes the power load over the nth subcarrier, and satisfying

$$\sum_{n=1}^{N} p_n^{(\ell)} \leq P, \tag{2.1}$$

with P the available power in transmission.

The resulting vector is padded with $N_v = N_{\mathrm{fft}} - N$ symbols equal to zero in correspondence of the N_v virtual subcarriers, undergoes N_{fft}-sized inverse discrete Fourier

transform (IDFT) processing and the cyclic prefix (CP) insertion in order to mitigate the interblock interference (IBI) and the digital-to-analog conversion. Then, the obtained signal is up-converted at carrier frequency f_c and transmitted over a frequency-selective channel.

Receiver structure

In a mobile-radio scenario, signals experience several degradation factors due to reflection, diffraction, scattering and, in general, to any obstacle that obstructs the line of sight (LOS) between the transmitter and the receiver. This phenomenon is called multipath fading and results in the reception of multiple versions of the transmitted signal, each of them characterized by its attenuation, phase and delay. Moreover, another significant effect is the time variation in the structure of the medium. As a result, the characteristics of the paths experienced by the transmitted signal can vary during time [2]. Statistical models for the channel impulse response of a fading multipath channel have been described in details in literature over the past years. Since its out of the scope of this work to provide a description of such models, we refer to some excellent references on this field such as [32], recalling here simply the adopted model. Since the coherence time T_c of the channel is much greater than the OFDM symbol duration T_s, the channel is affected by slowly-fading and the multipath parameters may be regarded as approximately invariant over many signaling intervals or for all the entire packet transmission. This model is called block-fading frequency-selective channel and, accordingly, the channel impulse response can be expressed as

$$c(\tau) = \sum_{n=1}^{N_{\mathrm{pt}}} c_n \delta(\tau - \tau_n), \tag{2.2}$$

where N_{pt} denotes the channel length and c_n and τ_n the complex-valued channel coefficient and delay relevant to the nth path. Even if all these parameters are random variables (RVs), the statistical models typically adopted for their characterization assume N_{pt} and τ_n fixed, whereas c_n are considered uncorrelated complex-valued Gaussian RVs. In particular, the variance of each coefficient c_n and the values of N_{pt} and τ_n are given by a power-delay profile of the model chosen. In the following, we will refer to the ITU multipath fading models as specified in [33].

At the receiver, the signal is fed to the analog-to-digital converter, the IBI is eliminated removing the CP and the resulting signal samples, after serial-to-parallel (S/P)

conversion, are input to the discrete Fourier transform (DFT) block. Thanks to the effect of the cyclic prefix that makes the channel look like circular convolution instead of linear convolution, the model of the signal samples at the output of the DFT block results

$$z_n^{(\ell)} = \sqrt{p_n^{(\ell)}} x_n^{(\ell)} h_n^{(\ell)} + w_n^{(\ell)}, \quad \forall n \in \mathcal{D}_s \tag{2.3}$$

i.e., the overall OFDM transmission can be viewed as N parallel and non-interfering transmissions, where $h_n^{(\ell)}$ is the complex-valued channel coefficient on subcarrier n, that is obtained as the nth coefficient of the DFT of the channel response (2.2) encompassing also the transmitter and receiver filters, and $w_n^{(\ell)} \in \mathcal{CN}(0, {\sigma_{\ell,n}^{(w)}}^2)$ is the circular-symmetric complex-Gaussian random variable with standard deviation $\sigma_{\ell,n}^{(w)}$, denoting the thermal noise sample on subcarrier n. The instantaneous post processing SNR values are then defined as

$$\gamma_n^{(\ell)} \triangleq p_n^{(\ell)} \frac{|h_n^{(\ell)}|^2}{{\sigma_{\ell,n}^{(w)}}^2}, \quad \forall n \in \mathcal{D}_s \tag{2.4}$$

and, upon collecting them into the post-processing SNR diagonal matrix

$$\boldsymbol{\Upsilon}^{(\ell)} \triangleq \begin{pmatrix} \gamma_1^{(\ell)} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \gamma_N^{(\ell)} \end{pmatrix}, \tag{2.5}$$

the input-output relationship of the BIC-OFDM system described above can be equivalently expressed as

$$\mathbf{z}^{(\ell)} = \left(\boldsymbol{\Upsilon}^{(\ell)}\right)^{\frac{1}{2}} \mathbf{x}^{(\ell)} + \mathbf{w}^{(\ell)}, \tag{2.6}$$

where now $\mathbf{w}^{(\ell)} \triangleq [w_1^{(\ell)}, \cdots, w_N^{(\ell)}]^{\mathrm{T}}$ is the noise vector with $w_n^{(\ell)} \in \mathcal{CN}(0,1)$.

Finally, the receiver performs the soft metric evaluation, followed by de-interleaving and decoding. In particular, the soft metric at the input of the decoder, identified by the log-likelihood ratio (LLR) of the kth coded binary symbol, is expressed by

$$\Lambda_k^{(\ell)} = \log \frac{\sum_{\tilde{x} \in \chi_{b'_k}^{(\ell)(i_k,n_k)}} p\left(z_{n_k}^{(\ell)} | x_{n_k}^{(\ell)} = \tilde{x}, \boldsymbol{\Upsilon}^{(\ell)}\right)}{\sum_{\tilde{x} \in \chi_{b_k}^{(\ell)(i_k,n_k)}} p\left(z_{n_k}^{(\ell)} | x_{n_k}^{(\ell)} = \tilde{x}, \boldsymbol{\Upsilon}^{(\ell)}\right)} \tag{2.7}$$

where

$$p\left(z_{n_k}^{(\ell)}|x_{n_k}^{(\ell)}=\tilde{x},\boldsymbol{\Upsilon}^{(\ell)}\right)\propto\exp\left(-\left|z_{n_k}^{(\ell)}-\sqrt{\gamma_{n_k}^{(\ell)}}\tilde{x}\right|^2\right) \tag{2.8}$$

is the Gaussian-shaped probability density function (p.d.f.) of the received sample value, conditioned on the transmitted symbol $\tilde{x}$ and on $\boldsymbol{\Upsilon}^{(\ell)}$, $\chi_a^{(i,n)}$ represents the subset of all the symbols belonging to the modulation adopted on the nth subcarrier whose ith label bit is equal to a and $b_k^{'}$ denotes the complement of bit b_k.

2.2.1 Error probability analysis

Under ideal interleaving assumption, the BICM channel of the system described in Sect. 2.2 behaves like a binary input output symmetric (BIOS) channel [18]. Accordingly, dropping wl.g. the dependence on the PR index ℓ, the packet error rate (PER) of linear binary codes over BIOS channel can be computed exploiting the union-bound as

$$\mathrm{PER}\leq\sum_{d=d_{\mathrm{free}}}^{N_c}\omega(d)\mathrm{PEP}(d) \tag{2.9}$$

where $\mathrm{PEP}(d)$ is the pairwise error probability of two codewords originating from the same state of the trellis of the BICM decoder and merging after d steps, $\omega(d)$ is the weight of all error events at Hamming distance d and d_{free} is the minimum distance between two codewords.
Thus, the PER performance can be obtained estimating the PEP of the system. Since in a BIOS channel each bit depends only on its corresponding channel output [18], [34], for the system described above employing the bit log-likelihood metrics at the decoder input the PEP in (2.9) can be written as

$$\mathrm{PEP}(d)=\Pr\left\{\sum_{k=1}^{d}\Lambda_k>0\right\}=\Pr\left\{\Theta>0\right\}, \tag{2.10}$$

where we defined $\Theta\triangleq\sum_{k=1}^{d}\Lambda_k$, with Λ_k expressed by (2.7). It is well known that the probability of a random variable can be evaluated through its moment generating function (MGF) [35], so that, introducing

$$\mathcal{M}_\Theta(s)\triangleq\mathrm{E}_\Theta\left\{\exp(s\Theta)\right\}=\left[\mathcal{M}_\Lambda(s)\right]^d, \tag{2.11}$$

as the MGF of the RV Θ, where

$$\mathcal{M}_{\Lambda}(s) \triangleq \mathrm{E}\left\{e^{s\Lambda}\right\} \tag{2.12}$$

is the MGF of the i.i.d. RV Λ_k, the PEP can be numerically evaluated exploiting the following integral[1] [35]- [36]

$$\mathrm{PEP}(d) = \frac{1}{2\pi\jmath}\int_{\sigma-\jmath\infty}^{\sigma+\jmath\infty}\mathcal{M}_{\Theta}(s)\frac{\mathrm{d}s}{s} = \frac{1}{2\pi\jmath}\int_{\sigma-\jmath\infty}^{\sigma+\jmath\infty}\left[\mathcal{M}_{\Lambda}(s)\right]^{d}\frac{\mathrm{d}s}{s}. \tag{2.13}$$

2.3 Link performance prediction techniques

The system described in Sect. 2.2 has several transmission parameters, e.g., modulation order (per subcarrier), coding rate, power per subcarrier, that should be properly set in order to obtain the best possible performance. This is done solving a constrained optimization problem where the objective function, representing the performance metric of the system, is optimized over the constrained set of the transmission parameters. To this end, it is required an effective way to predict the system performance accounting for both the CSI and the information relevant to the transmitted-signal mode choice (i.e. modulation size, coding rate, etc.). Moreover, there should be also take into account information coming from different techniques that further improve the transmission such as the effect of the multiple antennas scheme or the effect of the HARQ mechanism. Finally, the prediction model needs to be not only accurate but also simple, in order to make feasible the constrained optimization problem.

A simple approach to predict the performance, for example in terms of PER, is to consider to ratio between the number of erroneous packets by the total number of received packets during a given observation window. Though, such an estimator assumes clearly slow-varying channel and takes many packets to converge. On the contrary, predicting the future channel state information and deriving from that a PER estimate, leads to a more accurate evaluation and faster convergence. However in multi-carrier systems, such as OFDM, wherein the frequency selective channel introduces large SNR variation across the subcarriers, the above approach would

[1] In (2.13), σ is any real number which ensures that the contour path lies in the region of convergence.

imply to obtain the performance of all the possible transmission modes through full link-level simulations. This approach is practically unfeasible. The basic idea is to have a simple link quality model limited to the additive white Gaussian noise (AWGN) channel performance.
The latter concept is well represented by the link performance evaluation method called effective SNR mapping (ESM) [37]. The objective of this method is to find a compression function that maps a sequence of varying SNRs across the subcarriers to a single value that, in turn, is strongly correlated with the actual PER. The basic principle of ESM is to be able to go from an instantaneous channel state, such as the instantaneous SNR for each subcarrier in case of OFDM, to a corresponding PER, through a compressed value called effective SNR. This scalar value represents the SNR of an equivalent system over AWGN channel, whose PER performance can be simply evaluated (e.g., computed off-line and stored into a look-up table (LUT) or according to analytical models such as in [38]). Depending on the way the compressing function is derived, several ESM techniques have been proposed so far [39].

- The exponential ESM (EESM) [40] exploits an exponential function to obtain a rather simple but effective evaluation of link PER performance but only for the case of binary signalling. For high order modulations it is necessary a fine tuning factor for adjusting the PER estimate and reaching good accuracy for each modulation and coding scheme.

- The capacity ESM [26], [41], yields an optimistic estimate of the real PER function of a link channel realization, since the actual information between transmitter and receiver on every subcarrier is upper bounded by the concerning capacity expression [39].

- The last solution makes up for this coarse approximation using a mutual information approach and so it is referred to as mutual information ESM (MIESM) [42], [43]. Differently from the other ones, it includes two separate models, one for the modulation and the other for the coding, thereby providing good prediction performance for the mixed-modulation as well. Therefore, such a separate modulation and coding model makes radio resources management like power allocation and rate adaptation very convenient, as done in [44]. Unfortunately, a closed expression for calculating the mutual information doe not exist, so a polynomial approximation is essential, as proposed in [43].

Finally, a novel ESM technique is represented by the model whose compression function is based on the cumulant moment generating function (CMGF) of the bit level log-likelihood metrics (2.7) at the input of the soft decoder [45]. This method, called κESM for short, offers several significant features compared with existing ESM techniques, such as:
i) improved accuracy performance compared with the conventional EESM method;
ii) separation of the modulation and coding models, thus making configurations using mixed modulation among subcarriers easy to manage;
iii) the use of a tuning factor (required for turbo codes) to offer even improved accuracy that is independent of the modulation format adopted on each subcarrier;
iv) similar accuracy (or even better for multi-level QAM modulations) as the MIESM method, while offering (unlike the MIESM) a modulation model with a convex and simple closed-form mapping function.
Thus, in the reminder of this thesis, wherever required, we will rely on the κESM model as link performance prediction method and, therefore, before proceeding further, its structure is briefly recalled in the following section.

2.3.1 The κESM model

The κESM method offers an accurate yet manageable PER prediction scheme for BIC-OFDM transmission links over frequency-selective channels relying on an accurate evaluation of the PEP figure of merit in (2.9) through the statistical description of the BIC log-likelihood metrics.
According to the route originally pursued in [34], the numerical evaluation of PEP in (2.13) can be simply approximated as the tail probability

$$\mathrm{PEP}(d) \simeq Q\left(\sqrt{-2d\kappa_{\Lambda}(\hat{s})}\right), \tag{2.14}$$

where $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-t^2/2)dt$,

$$\kappa_{\Lambda}(s) \triangleq \log \mathcal{M}_{\Lambda}(s) \tag{2.15}$$

is the CMGF of the LLR metrics Λ_k, $1 \leq k \leq N_{\mathrm{c}}$, relevant to a sequence of N_{c} received CBS, $\mathcal{M}_{\Lambda}$ is the MGF defined in (2.11) and $\hat{s}$ is the so-called saddlepoint, that, in the case of BIOS channels with MAP metric, is $\hat{s} = 1/2$ as shown in [36].

The MGF (2.12) can be evaluated as show in Appendix A.2, so that (2.15) computed at the saddlepoint turns out

$$\kappa_{\Lambda}\left(\hat{s}\right) \simeq \log\left(\frac{1}{\sum_{n=1}^{N} m_n}\sum_{n=1}^{N}\sum_{\mu=1}^{\sqrt{2^{m_n}}/2}\frac{\psi_{m_n}(\mu)}{2^{m_n-1}}\cdot \mathrm{e}^{-\frac{\gamma_n\left(\mu\cdot d_{m_n}^{(\min)}\right)^2}{4}}\right), \tag{2.16}$$

where $d_{m_n}^{(\min)}$ is the minimum Euclidean distance between the symbols of the 2^{m_n}-QAM constellation adopted on the nth subcarrier and $\psi_{m_n}(\mu)$ defines the number of symbols at distance $\mu d_{m_n}^{(\min)}$ from the nearest neighbor in the complementary subset of the constellation.

It can be noted that equation (2.14) corresponds to the PEP of an equivalent system with binary modulation (BPSK) that experiences a simple AWGN channel with SNR

$$\hat{\gamma} \triangleq -\kappa_{\Lambda}(\hat{s}). \tag{2.17}$$

This observation is exactly the principle described in Sect. 2.3 on which the link performance prediction methods based on the ESM rely. Thus, the performance of the BIC-OFDM system characterized by the post-processing SNRs matrix $\boldsymbol{\Upsilon}$ are estimated evaluating the performance of the equivalent binary system over AWGN channel characterized by the SNR $\hat{\gamma}$, i.e.,

$$\mathrm{PER}(\boldsymbol{\Upsilon}) \simeq \Psi_{\mathrm{AWGN}}(\hat{\gamma}), \tag{2.18}$$

where Ψ_{AWGN} is the PER of the equivalent system over AWGN channel.
In conclusion, the κESM approach is described by the scheme in Fig. 2.4, where, as stated before, the division of the modulation and coding models is well evident. In

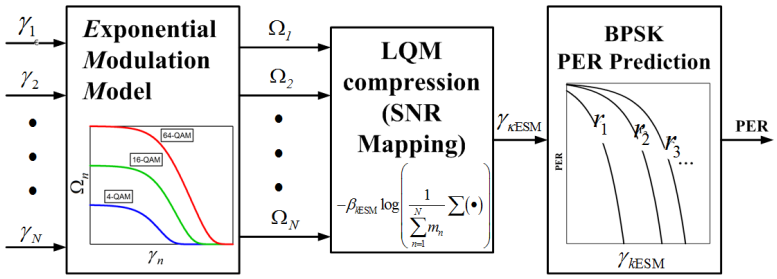

Figure 2.4: *κESM quality model structure.*

fact, it is first evaluated the ESNR as

$$\hat{\gamma} \triangleq -\log\left(\frac{1}{\sum_{n=1}^{N} m_n}\sum_{n=1}^{N}\Omega_n\right), \tag{2.19}$$

where

$$\Omega_n \triangleq \sum_{\mu=1}^{\sqrt{2^{m_n}}/2}\frac{\psi_{m_n}(\mu)}{2^{m_n-1}}\cdot \mathrm{e}^{-\frac{\gamma_n\left(\mu\cdot d_{m_n}^{(\min)}\right)^2}{4}}. \tag{2.20}$$

As apparent for (2.19)-(2.20), $\hat{\gamma}$, given the post-processing SNRs Υ, only depends on the modulation order adopted over each subcarrier and, besides, it is a convex function of the power allocation coefficients p_n, which are contained in γ_n via (2.4). Then $\hat{\gamma}$ is mapped into the PER value either entering into a LUT, where the values of the PER for all the considered coding rates are stored, or evaluating the PER via analytical models. In the latter case, it is worth noting that, in the region of interest, the PER function Ψ_r, with $0 \leq \Psi_r \leq 1$, is monotonically decreasing, analytic and convex (i.e. $\Psi_r' < 0$, $\Psi_r'' > 0$), which is consistent with observations, e.g., in [38], [46] and [47]. Finally, an optimization factor is required for turbo codes, as shown in [48], here neglected w.l.g. for the sake of notation.

Chapter 3

The goodput criterion

3.1 Rationale of the goodput criterion

Link resource adaptation strategies can be essentially described as constrained optimization problems. In fact, the optimization variables, i.e., the transmission parameters, must optimize a certain objective function, that describes the performance of the system, according to the available information and satisfy at the same time constraints on the required QoS. The output of this problem represents the setting of parameters to adopt during the transmission. An effective LRA strategy should be able to take into account, in a cross-layer manner, not only information on the status of the wireless channel but also information coming from the upper layers such as, recalling for instance some features of the systems in Sect. 2.1, the use of HARQ protocols. Moreover, for these systems, denoted by discrete constellations, state-of-the-art codes, ARQ mechanism, packet-oriented transmissions and so on, information-theoretic performance limits, which rely on ideal assumptions like Gaussian inputs and infinite length codebooks, can reveal inadequate to give a reliable picture of the actual link performance [49], [50].

To this end, a significative cross-layer optimization criterion is represented by the *goodput*, i.e., the number of data bits delivered in error-free packets per unit of time, or, offered layer 3 data rate. The goodput approach allows to obtain a trade-off between data rate and link reliability. In fact, in order to deliver a packet, the higher the data rate, the shorter the transmission time in one transmission attempt, but, more likely, the transmission will fail, thus engendering retransmissions. On the other hand, the more robust the transmission strategy, the more likely the packet will be delivered successfully within the retry limit, however, with less efficiency. Thus, a link

adaptation strategy that maximizes the goodput optimally trades-off the probability that the packet will be delivered successfully and the shortest possible transmission time.

Therefore, in the following, we will focus on the *expected goodput* (EGP) figure of merit, defined as the ratio of the expected delivered data payload and the expected transmission time. Many works in literature focused on this metric. Authors in [24], for instance, adapt the pair modulation order and coding rate to the EGP for 802.11a systems, based on a simple table-driven approach. Though, the wireless channel status along the retransmission in the EGP formulation is coarsely approximated by a two-states discrete time Markov chain. The channel variation is in fact evaluated switching between these two states, called good and bad, according to transition probabilities that depend on different wireless channel variation patterns. In [29] and [51], authors focus on the EGP metric considering systems with Viterbi decoding based on hard-decision. In particular, [51] propose using a maximum expected goodput criterion for multiuser diversity scheduling, whereas [29] addresses the problem of how to allocate bits and power among a set of parallel subchannels. The packet error rate to estimate the expected transmission time is then evaluated as a function of the average bit error rate, implicitly assuming a constant wireless channel through the entire packet delivering period. In [45], the power allocation problem for a BIC-OFDM system with soft-Viterbi decoding based on the bit level log-likelihood metrics is addressed. The optimization is aimed at maximizing the EGP, derived under the mild assumption of large retry limit.

All these works show that the performance of a communication system can be improved if the synergy between different layers is exploited, motivating the adoption of the goodput as figure of merit. Thus, in the next section, in order to shed lights on some typical assumptions done for the EGP formulation, we first resort to the renewal reward theory to obtain a general analytical expression of the EGP and then introduce the concept of inter-round optimization approach.

3.2 Expected goodput derivation

3.2.1 Equivalent BIC-OFDM model

In this section, we evaluate the EGP for the system described in Sect. 2.2. To this end, let us refer to the block diagram depicted in Fig. 3.1, representing an equivalent model of the above mentioned ARQ-based system. In this model we

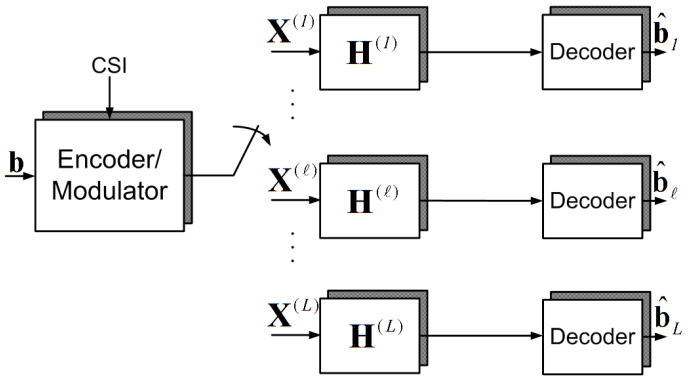

Figure 3.1: *Equivalent model for the ARQ BIC-OFDM system.*

denoted as $\mathbf{H} \triangleq \text{diag}\left\{\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \cdots, \mathbf{H}^{(L)}\right\}$, where $\mathbf{H}^{(\ell)} \triangleq \text{diag}\left\{[h_1^{(\ell)}, \cdots, h_N^{(\ell)}]\right\}$, and $\mathbf{X} \triangleq \text{diag}\left\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \cdots, \mathbf{X}^{(L)}\right\}$, where $\mathbf{X}^{(\ell)} \triangleq \text{diag}\left\{[x_1^{(\ell)}, \cdots, x_N^{(\ell)}]\right\}$, the block diagonal matrices containing the channel realizations that could be experienced by the link during L transmissions and the relevant transmitted symbols throughout the L PRs, respectively[1].

In the ideal and fictitious case of non-causal perfect knowledge of the channel gains evolution, the matrix $\mathbf{H}$ is available at the CSI input of the transmitter in Fig. 3.1, thus enabling the joint adaptation of the transmission parameters to compute $\mathbf{X}$ over the L PRs. Unfortunately, such a joint optimization, though attractive, is clearly impracticable, due to the impossibility to reliably predict the whole matrix $\mathbf{H}$ at the beginning of each RLC-PDU transmission.

Thus, an effective approach usually adopted in literature, inspired by [24] and adopted for instance also in [29]- [48], is to adapt the transmission parameters to the EGP,

[1] If the total number of PRs required for transmitting a given RLC-PDU is $L^* < L$, then $\mathbf{X}_\ell = \mathbf{0}$, for $L^* < \ell \leq L$.

defined as the ratio between the data payload to be delivered and the expected transmission time relying. In particular, we introduce the concept of a *per-round optimization* approach, that can be well understood looking at the equivalent model in Fig. 3.1. This procedure enables the adaptation of the transmission parameters over each branch of the equivalent model in Fig. 3.1, making use of only the instantaneous channel state information currently available and accounting for the time already spent in the previous transmission attempts.

3.2.2 Metric derivation

Based on this premise, let us define the metric at the generic round ℓ, i.e., at the ℓth branch of the equivalent model in Fig. 3.1. To this end, we will resort to the renewal theory [52], first introduced in [53] to analyze throughput performance of a system in presence of ARQ protocols, under the mild assumptions of error and delay free feedback channel and infinite buffer length.

Before proceeding further, it is worth noting that the ℓth branch of the equivalent system corresponds to a system with $L-\ell+1$ PRs still available and where $\ell-1$ packet transmissions previously failed. For such a system, let define a *renewal event* as the event in which the system stops transmitting the current packet either because an ACK is received or because the PR limit $L-\ell$ is reached. The holding time between two consecutive occurrences of the renewal event, or inter-arrival time, is described as a sequence of i.i.d. RVs[2] $\{S_1^{(\ell)}, S_2^{(\ell)}, \cdots\}$.

A packet transmission employing transmission parameters $\boldsymbol{\tau}^{(\ell)} \triangleq \{\mathbf{p}^{(\ell)}, \boldsymbol{\varphi}^{(\ell)}\}$, with $\boldsymbol{\varphi}^{(\ell)} \triangleq \{\mathbf{m}^{(\ell)}, r^{(\ell)}\}$, takes places in

$$T_{\mathrm{u}}(\boldsymbol{\varphi}^{(\ell)}) = \frac{N_{\mathrm{c}}^{(\ell)} T_{\mathrm{B}}}{r^{(\ell)} \sum_{n=1}^{N} m_n{}^{(\ell)}} \tag{3.1}$$

seconds, where T_{B} is the OFDM symbol duration, in seconds. Accordingly, the ith holding time can be written as

$$S_i^{(\ell)} = T_f(\ell-1) + \sum_{j=\ell}^{\ell_i} T_{\mathrm{u}}(\boldsymbol{\varphi}^{(j)}), \tag{3.2}$$

[2]It is worth noting that this in not a restrictive assumption. In fact, recalling the block-fading channel in the system model in Sect. 2.2, the channel remains the same during a packet transmission but independently changes between either different packets or transmitted copies of the same packet.

where the first term on the right hand side (RHS) is the time elapsed over the previous $\ell - 1$ failed transmissions, which is a known quantity at the ℓth branch, whereas $\ell \leq \ell_i \leq L - \ell$ is a RV depending on the number of packet transmissions after which the renewal event happened. A renewal process is then defined as

$$X_t^{(\ell)} \triangleq \sup_{\nu} \{\nu : J_\nu^{(\ell)} < t\}, \tag{3.3}$$

that is, the number of renewal events occurred by time $t \geq 0$, with

$$J_\nu^{(\ell)} \triangleq \sum_{i=1}^{\nu} S_i^{(\ell)} \tag{3.4}$$

the time required for ν renewal events to happen.
Besides, let $\{Z_1^{(\ell)}, Z_2^{(\ell)}, \cdots\}$ be a sequence of random *rewards* associated with the renewal events $\{S_1^{(\ell)}, S_2^{(\ell)}, \cdots\}$. In particular, since we are interested in correctly receiving the N_p information bits out of the $N_\mathrm{c}^{(\ell)}$ transmitted ones, $Z_i^{(\ell)} = N_\mathrm{p}/B$ if the renewal is due to a successful decoding, where B denotes the overall OFDM bandwidth, otherwise $Z_i = 0$.
Before proceeding further, let us introduce $\mathcal{A}_k$ as the event of receiving an ACK at round k, $\bar{\mathcal{A}}_k$ as the event of receiving a NACK at round k and $\mathcal{R}_k$ as the event of having a renewal event after round k. Accordingly, the probability of $\mathcal{R}_k$ is

$$\Pr\{\mathcal{R}_k\} \triangleq \Pr\{\bar{\mathcal{A}}_1, \cdots, \bar{\mathcal{A}}_{k-1}, \mathcal{A}_k\}, \tag{3.5}$$

and, since a renewal event always happens when the retry limit L is reached,

$$\Pr(\mathcal{R}_L) = 1 - \sum_{k=1}^{L-1} \Pr(\mathcal{R}_k). \tag{3.6}$$

On the other hand, defining $\mathcal{N}_k$ as the event of not receiving ACKs in k attempts, with $1 \leq k \leq L$, the more manageable probability $\Pr(\mathcal{N}_k)$ can be introduced,

$$\Pr(\mathcal{N}_k) \triangleq \Pr\{\bar{\mathcal{A}}_1, \cdots, \bar{\mathcal{A}}_k\} = 1 - \sum_{j=1}^{k} \Pr(\mathcal{R}_j). \tag{3.7}$$

It easily follows that

$$\Pr(\mathcal{R}_k) = \Pr(\mathcal{N}_{k-1}) - \Pr(\mathcal{N}_k) \tag{3.8}$$

with $\Pr(\mathcal{N}_0) \triangleq 1$ and the outage probability, i.e., the probability of having unsuccessful decoding within the retry limit L, results

$$P_{\text{out}}(L) \triangleq \Pr\{\bar{\mathcal{A}}_1, \cdots, \bar{\mathcal{A}}_L\} = \Pr(\mathcal{N}_L). \tag{3.9}$$

Denoting $Y_t^{(\ell)} = \sum_{i=1}^{X_t^{(\ell)}} Z_i^{(\ell)}$ the reward earned in $X_t^{(\ell)}$ renewal events, from the renewal theory the following theorem holds.

Theorem 3.1 (Renewal Reward Theorem, [52]) *For a renewal process where $Z_j^{(\ell)}$ is the reward earned at each renewal event $X_j^{(\ell)}$, the reward function $Y_t^{(\ell)} \triangleq \sum_{i=1}^{X_t^{(\ell)}} Z_i^{(\ell)}$ satisfies*

$$\lim_{t\to\infty} \frac{1}{t} \mathrm{E}\{Y_t^{(\ell)}\} = \frac{\mathrm{E}\{Z_1^{(\ell)}\}}{\mathrm{E}\{S_1^{(\ell)}\}}. \tag{3.10}$$

Remark. Theorem 3.1 states that the accumulated reward over time is equivalent to the ratio between the expected reward and the expected time for a renewal event to happen.

3.2.3 Inter-round EGP optimization approach

Capitalizing on Theorem 3.1, the EGP over the generic branch ℓ can be evaluated as follows. Defining with $\mathrm{PER}(\boldsymbol{\tau}^{(\ell)}|\mathbf{H}^{(j)})$ the packet error probability when channel $\mathbf{H}^{(j)}$ is experienced and transmission parameters set $\boldsymbol{\tau}^{(\ell)}$ is employed, we get that

$$\mathrm{E}\{Z_1^{(\ell)}\} = \frac{N_{\text{p}}}{B}\{1 - P_{\text{out}}(L-\ell)\} \tag{3.11}$$

with

$$P_{\text{out}}(L-\ell) = \prod_{j=\ell}^{L} \mathrm{E}_{\mathbf{H}^{(j)}} \left\{ \mathrm{PER}(\boldsymbol{\tau}^{(j)}|\mathbf{H}^{(j)}) \right\}, \tag{3.12}$$

and

$$\mathrm{E}\{S_1^{(\ell)}\} = T_f(\ell-1) + \sum_{j=\ell}^{L} T_u(\boldsymbol{\varphi}^{(j)}) \Pr(\mathcal{N}_{j-1}), \tag{3.13}$$

where

$$\Pr(\mathcal{N}_j) = \prod_{k=0}^{j} \mathrm{E}_{\mathbf{H}^{(k)}} \left\{ \mathrm{PER}(\boldsymbol{\tau}^{(k)}|\mathbf{H}^{(k)}) \right\} \tag{3.14}$$

denotes the probability of having an unsuccessful decoding up to round j, with $\mathrm{PER}(\boldsymbol{\tau}^{(0)}|\mathbf{H}^{(0)}) \triangleq 1$. After substituting equations (3.11)-(3.14) into (3.10), we obtain that the goodput is

$$\eta^{(\ell)}(\boldsymbol{\tau}^{(\ell)}) = \frac{N_{\mathrm{p}}}{B} \frac{1 - \mathrm{PER}(\boldsymbol{\tau}^{(\ell)}|\mathbf{H}^{(\ell)}) \prod_{j=\ell+1}^{L} \mathrm{E}_{\mathbf{H}^{(j)}}\{\mathrm{PER}(\boldsymbol{\tau}^{(j)}|\mathbf{H}^{(j)})\}}{T_{\mathrm{f}}(\{\boldsymbol{\varphi}^{(i)}\}_{i=1}^{\ell-1}) + T_u(\boldsymbol{\varphi}^{(\ell)}) + \mathcal{T}_D(\{\boldsymbol{\tau}^{(j)}, \mathbf{H}^{(j)}\}_{j=\ell}^{L})}, \tag{3.15}$$

where $T_{\mathrm{f}}(\{\boldsymbol{\varphi}^{(i)}\}_{i=1}^{\ell-1})$ represents the time spent in the previous $\ell - 1$ failed attempts and

$$\mathcal{T}_D(\{\boldsymbol{\tau}^{(j)}, \mathbf{H}^{(j)}\}_{j=\ell}^{L}) = \begin{cases} \mathrm{PER}(\boldsymbol{\tau}^{(\ell)}|\mathbf{H}^{(\ell)}) \cdot \Bigg(T_u(\boldsymbol{\varphi}^{(\ell+1)}) + \sum_{k=\ell+2}^{L} T_u(\boldsymbol{\varphi}^{(k)}) \cdot \\ \prod_{j=\ell+1}^{k-1} \mathrm{E}_{\mathbf{H}^{(j)}}\{\mathrm{PER}(\boldsymbol{\tau}^{(j)}|\mathbf{H}^{(j)})\} \Bigg) & \text{if } \ell \leq L-2 \\ T_u(\boldsymbol{\varphi}^{(L)}) \mathrm{PER}(\boldsymbol{\tau}^{(L-1)}|\mathbf{H}^{(L-1)}) & \text{if } \ell = L-1 \\ 0 & \text{if } \ell = L \end{cases} \tag{3.16}$$

is the expected delivering time.

This formulation requires however the knowledge of the channel's p.d.f. for all the possible scenarios, which is a problematic assumption in most cases of interest. Thus, we propose a modified version of the EGP metric, obtained by substituting the expectation on the future channel conditions with the current channel status, i.e., replacing $\mathrm{E}_{\mathbf{H}^{(j)}}\left\{\mathrm{PER}(\boldsymbol{\tau}^{(j)}|\mathbf{H}^{(j)})\right\}$, $\ell < j \leq L$, with $\mathrm{PER}(\boldsymbol{\tau}^{(\ell)}|\mathbf{H}^{(\ell)})$. Accordingly, at each PR, the setting of the transmission parameters $\boldsymbol{\tau}^{(\ell)}$ is the one maximizing the EGP metric

$$\zeta^{(\ell)}(\boldsymbol{\tau}^{(\ell)}) \triangleq \frac{N_{\mathrm{p}}}{B} \frac{1 - \left(\mathrm{PER}(\boldsymbol{\tau}^{(\ell)}|\mathbf{H}^{(\ell)})\right)^{L-\ell+1}}{T_{\mathrm{f}}(\{\boldsymbol{\varphi}^{(i)}\}_{i=1}^{\ell-1}) + T_u(\boldsymbol{\varphi}^{(\ell)}) \sum_{k=\ell}^{L} \left(\mathrm{PER}(\boldsymbol{\tau}^{(\ell)}|\mathbf{H}^{(\ell)})\right)^{k-\ell}}. \tag{3.17}$$

Let us note that the objective function (3.17) is equivalent to the EGP we would have assuming that the packet experiences the current channel conditions $\mathbf{H}^{(\ell)}$ throughout its possible future retransmissions, i.e., long term static channel assumption. As anticipated in Sect. 3.2.1, this assumption, which is usually adopted in these cases, see for instance [29], [45], allows to obtain a manageable metric to optimize over the available set of transmission parameters. Moreover, as it will be shown in the following, simulation results carried over a realistic wireless scenario justify this approach

showing that the actual goodput is very close to the EGP.
Thus, in conclusion, the EGP (3.17) is evaluated at each PR according to the available current CSI and this per-round optimization is performed until the packet is successfully received, or the PR limit L reached.

Chapter 4

LRA for cognitive BIC-OFDM systems

Cognitive radio networks have recently been gaining an ever increasing interest as an effective way to tackle the problem of resource scarcity and inefficiency in the frequency spectrum utilization [54]. Frequency-agile cognitive radios (CRs) [55], [56], employed by unlicensed users, or secondary users (SUs), are the key to this novel radio access paradigm. Such devices adapt their parameters to transmit over segments of spectrum actually owned by licensed users, or primary users (PUs), without causing harmful interference to the latter [57], [58]. In this Chapter, the LRA problem aimed at maximizing the GP of a cognitive BIC-OFDM system is tackled and optimally solved. In particular, the optimal PA strategy is first found through the Lagrangian dual decomposition method. Then, in order to circumvent the drawbacks of conventional numerical methods, a novel iterative yet simple PA algorithm is proposed, whose optimality conditions are analytically demonstrated resorting to the QVI theory. Based on this PA algorithm, an AMC scheme is eventually derived, and the improvements in GP performance are finally highlighted simulating a realistic wireless scenario.

4.1 Dynamic spectrum access

Dynamic spectrum access (DSA) can be defined [59], [60] as a mechanism to adjust the spectrum resource usage in a near-real-time manner in response to the changing environment and objective (e.g. available channel and type of applications), changes

of radio state (e.g. transmission mode, battery status, and location), and changes in environment and external constraints (e.g. radio propagation, operational policy). Furthermore, the unlicensed (i.e., secondary) user spectrum access may encounter incumbent licensed (i.e., primary) users that possess different transmission characteristics. To counter these challenges and operate in a manner that is transparent to the incumbent primary users, the radio must adapt to the varying spectrum operating conditions [61]. Therefore, based on the spectrum sensing information obtained after querying a geo-location database or sensing the radio environment [62], spectrum access schemes have to be designed to efficiently use the radio resource. The main objective of dynamic spectrum allocation is to maximize the QoS performance (e.g., throughput, goodput, packet error rate, etc.) of secondary users while minimizing interference to primary service [63].

In order to address the above issues, a considerable effort has been devoted on optimizing the performance of the secondary link while guaranteeing coexistence with the primary network. In [64], OFDM and multi-antennas are exploited to achieve a trade-off between the benefits of spatial multiplexing for SU transmissions and the interference level caused at the PU receivers. In [65], the optimal solution for the power allocation (PA) problem across the OFDM bandwidth is derived by maximizing the secondary link capacity, under limitations on the out-of-band interference to PUs operating on adjacent bands. It also proposes a low-complexity suboptimal approach, called step-ladder (SL) PA, based on the principle that more power shall be loaded on the subcarriers farther from the PU bands. In [66], the same problem is suitably extended so as to take into account for the constraints about the in-band interference and the available transmit power. Resource allocation schemes in OFDMA cognitive radio networks are considered in [67] and [68], wherein the sum rate maximization problem is addressed for the single and multi-cell case, respectively.

In the sequel of this chapter, we propose a novel framework that allows to adapt the transmission parameters of a secondary BIC-OFDM system in order to optimize its goodput, satisfying at the same time the constraints on the interference caused to the PUs. In particular, it is first analyzed the PA problem which is shown to be a convex optimization problem [69]. However, in order to circumvent the drawbacks of conventional numerical methods [70], such as slow convergence and need for parameter

tuning, used to solve these kind of problems, we turn to the QVI theory [71], [72]. The latter allows to derive a novel PA framework, referred to as SSR, wherein the constrained PA problem is split into elementary subproblems and whose equivalence with the original PA problem is analytically proven. Finally, a simple AMC scheme coupled with the SSR PA is formalized.

4.2 BIC-OFDM cognitive radio scenario

4.2.1 The underlay and interweave paradigms

The scenario under analysis consists of a BIC-OFDM radio link between a secondary transmitter (STx) and its receiver (SRx). Secondary users use the same frequency bands of primary users according to several techniques which depend on the amount of available side information. In practice, due to severe restrictions on the implementation complexity, detailed information such as channel gains or codebooks and messages of PUs will be unlikely available at the STx, which shall instead rely on limited side information only [57]. Thus, coexistence among PUs and SUs is basically ensured by the *underlay* and *interweave* paradigm[1] [57], [58], here briefly recalled.

In the underlay paradigm, SUs are allowed to transmit at the same time over the same bands (referred to as grey spaces) used by PUs, called *underlay* PUs (UPUs), as long as the in-band interference caused to them is kept below a certain threshold which depends on the requested PUs QoS. Clearly, since interference is path-loss dependent, some bands could be forbidden for SUs when they are very close either to PUs or to other SUs that are already using those portions of spectrum [57]. As consequence, it is reasonable to assume, as in [66], that each PU settles an "interference-free" zone of radius R around itself, wherein any SU transmission is interdicted. This allows PUs to protect themselves from too close SUs, which have to operate so that the interference level caused on the edge of each zone is below a given threshold.

The second co-existence scenario refers to the case in which a PU band is present in the close vicinity or adjacent to the SU's one, i.e., PUs and SUs are co-located in the same area with side-by-side bands. Here, SUs have to ensure that their out-of-band (OOB) emissions cause a limited interference to the PUs, called *interweave* PUs

[1] Note that in literature there are different definitions of these paradigms. We refer to the definition given, e.g., in [57] that is different from the one given, for example, in [68].

(IPUs), operating in adjacent bands.

Therefore, adopting the model proposed in [66], the cognitive scenario accounting for both paradigms is described as in Fig. 4.1. The STx transmits over Q_U bands in the set $\mathcal{B}_{\text{UPU}} \triangleq \{B_1, \cdots, B_{Q_U}\}$ used also by Q_U UPUs. Contiguous to these bands, there are Q_I bands belonging to the set $\mathcal{W}_{\text{IPU}} \triangleq \{W_1, \cdots, W_{Q_I}\}$, where Q_I IPUs are exclusively transmitting. Thus, the overall bandwidth is $B_{\text{tot}} \triangleq \sum_{i=1}^{Q_I} W_i + \sum_{u=1}^{Q_U} B_u$, where secondary transmission occurs over $B \triangleq \sum_{u=1}^{Q_U} B_u$.

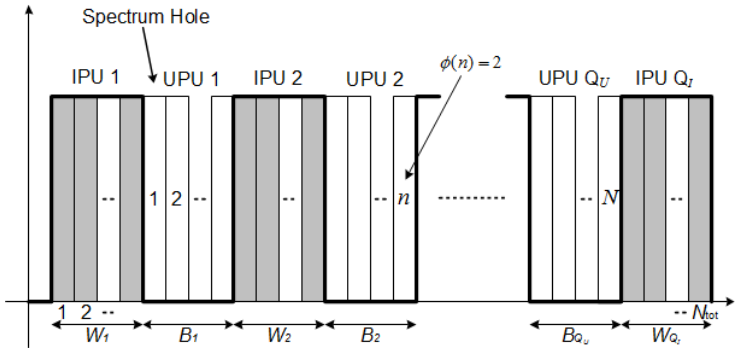

Figure 4.1: *Spectrum activity*

4.2.2 Power constraints evaluation

Recalling the BIC-OFDM system model in Sect. 2.2 and the underlay and interweave paradigms described above, it follows that a few constraints arise, at each PR ℓ, that limit the PA distribution $\mathbf{p}^{(\ell)} \triangleq [p_1^{(\ell)}, \cdots, p_N^{(\ell)}]^{\text{T}}$ that the STx is allowed to allocate over the N available subcarriers.

First, it is recalled constraint (2.1) on the total power that the STx can spend over the subcarriers, that is

$$f_0(\mathbf{p}^{(\ell)}) \triangleq \sum_{n=1}^{N} p_n^{(\ell)} - P \leq 0. \tag{4.1}$$

Then, the in-band interference brought by the STx to the Q_U UPUs described in Sect. 4.2.1 yields

$$f_u(\mathbf{p}^{(\ell)}) \triangleq \sum_{n:\phi(n)=u} p_n^{(\ell)} - T_u^{(\text{U})} \leq 0,\ 1 \leq u \leq Q_U, \tag{4.2}$$

where $\phi(n) = u$ is an indicator function denoting that subcarrier n used by the STx belongs to the subband $B_u \in \mathcal{B}_{\mathrm{UPU}}$ and $T_u^{(\mathrm{U})}$ is the maximum interference (or interference temperature) allowed at the edge of the uth zone scaled by the path loss originated by the STx.

Finally, we focus on the out-of-band interference that can be found within the IPU bands $W_i \in \mathcal{W}_{\mathrm{IPU}}$, $1 \leq i \leq Q_I$, due to the signal transmitted by the STx. Denoting with $P_i(f)$ its power spectral density function, the set of feasible PA coefficients $\mathbf{p}^{(\ell)}$ have to be constrained in order to satisfy [66]

$$f_{Q_U+i}(\mathbf{p}^{(\ell)}) \triangleq \sum_{n=1}^{N} K_{i,n} p_n^{(\ell)} - T_i^{(\mathrm{I})} \leq 0,\ 1 \leq i \leq Q_I, \tag{4.3}$$

where

$$K_{i,n} \triangleq \int_{\Delta f_{n,i}-\frac{W_i}{2}}^{\Delta f_{n,i}+\frac{W_i}{2}} P_i(f) df, \tag{4.4}$$

$\Delta f_{n,i}$ is the frequency distance between the nth subcarrier and the center of the ith IPU band, and $T_i^{(\mathrm{I})}$ is the allowed interference level over that subband, normalized by the path loss of the link STx-ith IPU receiver.

4.3 LRA problem formulation

This section focuses on the formulation of the optimization problem whose solution provides a proper setting of transmission parameters, namely, power distribution across the active subcarriers, modulation order and coding rate for the BIC-OFDM system in the cognitive context described in Sect. 4.2.1.

As outlined in Sect. 3.2.3, we will resort to a per-round optimization, where, at each round, the EGP metric is optimized over the set of possible transmission parameters. In particular, considering perfect CSI at the transmitter and uniform bit loading, i.e., $m_n^{(\ell)} = m^{(\ell)}\ \forall n \in \mathcal{D}_s$, where $\mathcal{D}_s$ is the set of subcarriers, the parameters to be set at each PR ℓ are the pair modulation order and coding rate, identified by transmission mode (TM) $\boldsymbol{\varphi}^{(\ell)} \triangleq \{m^{(\ell)}, r^{(\ell)}\}$, and the power allocation vector $\mathbf{p}^{(\ell)}$.

To this end, let us point out the following observations. Recalling the EGP expression (3.17), it can be noted that the packet error probability $\mathrm{PER}(\boldsymbol{\varphi}^{(\ell)}, \mathbf{p}^{(\ell)} | \mathbf{H}^{(\ell)})$, where $\mathbf{H}^{(\ell)} = \mathrm{diag}\{[h_1^{(\ell)}, \cdots, h_N^{(\ell)}]\}$ collects the channel coefficients at round ℓ, can be

efficiently evaluated by the ESM technique given by the κESM method described in Sect. 2.3.1. Based on (2.18), we thus get

$$\mathrm{PER}(\boldsymbol{\varphi}^{(\ell)}, \mathbf{p}^{(\ell)} | \mathbf{H}^{(\ell)}) = \Psi_{r^{(\ell)}} \left(\hat{\gamma}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}, \mathbf{p}^{(\ell)}) \right), \tag{4.5}$$

where $\Psi_{r^{(\ell)}} \left(\hat{\gamma}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}, \mathbf{p}^{(\ell)}) \right)$ is the packet error rate of the equivalent BPSK system employing coding rate $r^{(\ell)}$ over an AWGN channel denoted by SNR $\hat{\gamma}^{(\ell)} = \hat{\gamma}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}, \mathbf{p}^{(\ell)})$, evaluated as in (2.19)-(2.20)[2].
Therefore, upon plugging (4.5) into (3.17), the EGP metric at the ℓth round results

$$\zeta(\boldsymbol{\varphi}^{(\ell)}, \mathbf{p}^{(\ell)}) = \frac{N_{\mathrm{p}}}{B} \frac{1 - \left[\Psi_{r^{(\ell)}} \left(\hat{\gamma}(\boldsymbol{\varphi}^{(\ell)}, \mathbf{p}^{(\ell)}) \right) \right]^{L-\ell+1}}{T_{\mathrm{f}} + T_u(\boldsymbol{\varphi}^{(\ell)}) \sum_{k=\ell}^{L} \left[\Psi_{r^{(\ell)}} \left(\hat{\gamma}(\boldsymbol{\varphi}^{(\ell)}, \mathbf{p}^{(\ell)}) \right) \right]^{k-\ell}}. \tag{4.6}$$

From now on, w.l.g., the dependence on the PR ℓ will be dropped for the sake of readability. Finally, we are left to the EGP optimization problem (EGP-OP), as stated in the sequel.

EGP-OP. The LRA problem, consisting in finding the TM $\boldsymbol{\varphi}$ and the power vector $\mathbf{p}$ which maximize, at each PR, the EGP (4.6) subject to the available power and interference constraints, can be written as

$$\begin{aligned} (\boldsymbol{\varphi}^*, \mathbf{p}^*) = & \arg\max_{(\boldsymbol{\varphi}, \mathbf{p})} \{\zeta(\boldsymbol{\varphi}, \mathbf{p})\} \\ \text{s.t.} \quad & \mathbf{p} \succeq \mathbf{0} && \text{(4.7.a)} \\ & \boldsymbol{f}(\mathbf{p}) \preceq \mathbf{0} && \text{(4.7.b)} \\ & \boldsymbol{\varphi} \in \mathcal{D}_{\boldsymbol{\varphi}} && \text{(4.7.c)} \end{aligned}, \tag{4.7}$$

where $\boldsymbol{f}(\mathbf{p}) \triangleq [f_0(\mathbf{p}), \cdots, f_{Q_U+Q_I}(\mathbf{p})]^{\mathrm{T}}$ is the $(Q_U + Q_I + 1)$-sized vector including the constraints (4.1)-(4.3) and $\mathcal{D}_{\boldsymbol{\varphi}}$ is the set of allowed transmission modes.

The structure of the EGP-OP (a mixed integer problem) allows to optimally decouple the problem in two parts. First, the optimal PA $\mathbf{p}^*(\boldsymbol{\varphi}_0)$ for a given TM $\boldsymbol{\varphi}_0$ is found, then a simple exhaustive search is done over the finite-size set of the possible TMs to find the pair $\{\mathbf{p}^*(\boldsymbol{\varphi}_0^*), \boldsymbol{\varphi}_0^*\}$ corresponding to the highest EGP value. These issues will be addressed in the following sections.

[2]Let us recall that the ESNR $\hat{\gamma}$, as described in Sect. 2.3.1, depends on $\boldsymbol{\varphi}^{(\ell)}$ only via the modulation oder.

4.4 Optimal power allocation

In this section, the PA problem for EGP optimization, based on (4.7), is formulated and then solved via the well-known dual decomposition method.

4.4.1 Formulation of the PA problem

In order to find the optimal PA distribution $\mathbf{p}^* \triangleq [p_1^*, \cdots, p_N^*]^{\mathrm{T}}$ solving (4.7) for a given TM $\boldsymbol{\varphi} = \boldsymbol{\varphi}_0$, let us adopt the following simplification, as done in [45]. Since in the evaluation of the ESNR (2.19), $\psi_{m_n}(1) > \psi_{m_n}(\mu)$ for $\mu > 1$, let us limit the summation in (2.20) to $\mu = 1$ for simplicity[3], or equivalently, drop the terms relevant to the symbols in the complementary subset at distance greater than $d_{m_n}^{(\min)}$. Moreover, for the sake of notation, let rewrite the coefficients of (2.20) as $\alpha_n \triangleq \psi_{m_n}(1)/2^{m_n-1}$ and $\rho_n \triangleq 4/\left(d_{m_n}^{(\min)}\right)^2$. Then, the following proposition can be stated.

Prop. 4.1 *For a given TM $\boldsymbol{\varphi} = \boldsymbol{\varphi}_0$, the optimal PA $\mathbf{p}^*$ satisfies*

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} \zeta(\boldsymbol{\varphi}_0, \mathbf{p}) = \arg\min_{\mathbf{p}} \psi(\boldsymbol{\varphi}_0, \mathbf{p}) \tag{4.8}$$

where

$$\psi(\boldsymbol{\varphi}_0, \mathbf{p}) \triangleq \sum_{n=1}^{N} \alpha_n e^{-\frac{p_n}{\rho_n}} \tag{4.9}$$

is the argument of the logarithm of the κESM $\hat{\gamma}$ defined in (2.20), limited to $\mu = 1$, that results a convex function of the PA vector $\mathbf{p}$.

Proof In order to prove (4.8), let us note that now the only variable is the power $\mathbf{p}$ that appears in (4.6) only in the packet error rate via $\hat{\gamma}(\mathbf{p})$. Moreover, recalling that the PER $\Psi_r(\hat{\gamma})$ is a monotonically decreasing function of its argument $\hat{\gamma}$, it can be easily seen that the first derivative of (4.6) w.r.t. $\hat{\gamma}$ is always greater that zero, i.e., for a given TM, the EGP monotonically increases with $\hat{\gamma}$. Hence, the PA that maximizes the EGP is the same that maximizes $\hat{\gamma}(\mathbf{p})$. Looking at the expression of $\hat{\gamma}(\mathbf{p})$ given by (2.19)-(2.20) in Sect. 2.3.1, equivalence (4.8) easily follows. □

[3]Let us remark that (2.20) would be a convex function of the PA vector even without limiting μ to 1, and thus without affecting the validity of the following analysis. As demonstrate in [45], this simplification merely allows to obtain a significant reduction of complexity producing a closed-form expression of the optimal PA coefficients in exchange of any appreciable loss of performance.

Accordingly, the EGP oriented PA problem (EGOPA-OP) can be reformulated as follows.

Prop. 4.2 (EGOPA-OP) *Given TM $\boldsymbol{\varphi} = \boldsymbol{\varphi}_0$, the EGP oriented power allocation OP can be formulated as*

$$\begin{aligned} \mathbf{p}^* = & \arg\min_{\mathbf{p}} \{\psi(\boldsymbol{\varphi}_0, \mathbf{p})\} \\ \text{s.t.} \quad & \mathbf{p} \succeq \mathbf{0} \qquad (4.10.\text{a}) \\ & \boldsymbol{f}(\mathbf{p}) \preceq \mathbf{0} \qquad (4.10.\text{b}). \end{aligned} \tag{4.10}$$

Let us notice that, for the sake of notational simplicity, in the following we will denote the objective function as $\psi(\mathbf{p})$, dropping the dependence on the TM $\boldsymbol{\varphi}_0$.

4.4.2 Lagrangian dual decomposition approach

The EGOPA-OP in (4.10) is convex since both the objective function $\psi(\mathbf{p})$ and the set of constraints (4.10.a)-(4.10.b) are convex functions of the PA variable $\mathbf{p}$. Thus, it can be optimally solved applying the Lagrangian dual decomposition (LDD) technique [69], i.e., the OP is solved maximizing the dual function associated to the primal problem (4.10). A summary on the LDD theory is reported in Appendix A.1.1.
Accordingly, upon denoting as ϑ_m the mth Lagrange multiplier relevant to constraints (4.10.b) and stacking them into the (Q_U+Q_I+1)-sized vector $\boldsymbol{\vartheta} \triangleq [\vartheta_0, \cdots, \vartheta_{Q_U+Q_I}]^{\mathrm{T}}$, the Lagrangian associated with the EGOPA-OP (4.10) is

$$L_{\Lambda}(\boldsymbol{\vartheta}, \mathbf{p}) \triangleq \psi(\mathbf{p}) + \boldsymbol{\vartheta}^{\mathrm{T}} \boldsymbol{f}(\mathbf{p}). \tag{4.11}$$

Hence, the LDD approach obtains a lower bound on the optimal value $\psi^* \triangleq \psi(\mathbf{p}^*)$ of the primal problem (4.10) by solving the associated dual problem

$$\boldsymbol{\vartheta}^* = \arg\max_{\boldsymbol{\vartheta} \in \mathcal{D}_{\boldsymbol{\vartheta}}} \{g(\boldsymbol{\vartheta})\}, \tag{4.12}$$

where $\mathcal{D}_{\boldsymbol{\vartheta}} \triangleq \{\boldsymbol{\vartheta} : \vartheta_m \geq 0, 0 \leq m \leq Q_U + Q_I\}$ and

$$g(\boldsymbol{\vartheta}) \triangleq \inf_{\mathbf{p} \in \mathcal{D}_{\mathbf{p}}} \{L_{\Lambda}(\boldsymbol{\vartheta}, \mathbf{p})\}, \tag{4.13}$$

is the dual function, with $\mathcal{D}_{\mathbf{p}} \triangleq \{\mathbf{p} : p_n \geq 0, 1 \leq n \leq N\}$. We note that the difference between the optimal value ψ^* and the optimal dual value $g^* \triangleq g(\boldsymbol{\vartheta}^*)$, i.e., the duality

gap, according to the convex analysis is zero. It follows that strong duality holds, implying that the primal problem (4.10) can be equivalently solved by solving the dual problem (4.12). Then, the optimal PA $\mathbf{p}^*$ can be recovered by $\boldsymbol{\vartheta}^*$ as stated in the following.

Theorem 4.1 *In view of the convexity of the EGOPA-OP, the duality gap is zero and the Karush-Kuhn-Tucker (KKT) conditions are sufficient to derive the optimal PA* $\mathbf{p}^*$ *as*

$$p_n^* = \rho_n \left[\log \frac{1}{\vartheta_0^* + \vartheta_{\phi(n)}^* + \sum_{i=1}^{Q_I} \vartheta_{Q_U+i}^* K_{i,n}} - \log \frac{\rho_n}{\alpha_n} \right]^+ \tag{4.14}$$

$\forall n \in \mathcal{D}_s$, *where* $\boldsymbol{\vartheta}^*$ *is the optimal value of the Lagrange multipliers,* ρ_n *and* α_n *are defined in Sec. 4.4.1, and* $\phi(n)$ *is the indicator function used in (4.2).*

Proof Let associate the Lagrange multipliers μ_n, $1 \leq n \leq N$, and ϑ_m, $0 \leq m \leq Q_U + Q_I$, to the inequality constraints (4.10.a) and (4.10.b), respectively, and cast them into $\boldsymbol{\eta} \triangleq [\mu_1, \cdots, \mu_N, \vartheta_0, \cdots, \vartheta_{Q_U+Q_I}]^{\mathrm{T}}$. Upon applying the KKT conditions, we obtain

$$\nabla_{\mathbf{p}} \psi(\mathbf{p}^*) + \boldsymbol{\vartheta}^{\mathrm{T}} \nabla_{\mathbf{p}} \boldsymbol{f}(\mathbf{p}^*) = 0, \tag{4.15}$$

$$\boldsymbol{\eta} \succeq \mathbf{0}, \tag{4.16}$$

$$\mu_n p_n^* = 0, \quad 1 \leq n \leq N, \tag{4.17}$$

$$\vartheta_m f_m(\mathbf{p}^*) = 0, \quad 0 \leq m \leq Q_U + Q_I. \tag{4.18}$$

Then, we obtain from (4.15), for $1 \leq n \leq N$,

$$-\frac{\alpha_n}{\rho_n} \mathrm{e}^{-p_n^*/\rho_n} - \mu_n + \vartheta_0 + \vartheta_{\phi(n)} + \sum_{i=1}^{Q_I} \vartheta_{Q_U+i} K_{i,n} = 0, \tag{4.19}$$

that, under (4.17), gives

$$\left(\vartheta_0 + \vartheta_{\phi(n)} + \sum_{i=1}^{Q_I} \vartheta_{Q_U+i} K_{i,n} - \frac{\alpha_n}{\rho_n} \mathrm{e}^{-p_n^*/\rho_n} \right) p_n^* = 0, \tag{4.20}$$

whereas, according to (4.16), we get

$$\vartheta_0 + \vartheta_{\phi(n)} + \sum_{i=1}^{Q_I} \vartheta_{Q_U+i} K_{i,n} \geq \alpha_n / \rho_n \mathrm{e}^{-p_n^*/\rho_n}. \tag{4.21}$$

Now, there exist two possible cases. First, when $\vartheta_0 + \vartheta_{\phi(n)} + \sum_{i=1}^{Q_I} \vartheta_{Q_U+i} K_{i,n} < \alpha_n/\rho_n$, then (4.21) is satisfied only if $p_n^* > 0$. This means that the optimal PA results from (4.20) as

$$p_n^* = \rho_n \left(\log \frac{1}{\vartheta_0 + \vartheta_{\phi(n)} + \sum_{i=1}^{Q_I} \vartheta_{Q_U+i} K_{i,n}} - \log \frac{\rho_n}{\alpha_n} \right). \tag{4.22}$$

Alternatively, the condition $\vartheta_0 + \vartheta_{\phi(n)} + \sum_{i=1}^{Q_I} \vartheta_{Q_U+i} K_{i,n} \geq \alpha_n/\rho_n$ is not allowed since if $p_n^* > 0$ it would violate (4.20) and so the only admissible value is $p_n^* = 0$. Hence, combining together these results the optimal PA over the nth subcarrier can be written

$$p_n^* = \rho_n \left[\log \frac{1}{\vartheta_0^* + \vartheta_{\phi(n)}^* + \sum_{i=1}^{Q_I} \vartheta_{Q_U+i}^* K_{i,n}} - \log \frac{\rho_n}{\alpha_n} \right]^+, \tag{4.23}$$

where $\boldsymbol{\vartheta}^*$ is the optimal value of the Lagrange multipliers. Finally, since the EGOPA-OP is convex, from the convex optimization theory it is known that the solution is also unique. □

The following remarks about Theorem 4.1 are now of interest.

Subgradient-based PA Algorithm for EGOPA-OP (4.10)

1. Initialize $i = 0$, $\boldsymbol{\vartheta}^{(0)} = \boldsymbol{\vartheta}_0$, $\boldsymbol{\xi}$, ϵ and $I_{\max}$
2. Repeat
3. Compute $\mathbf{p}_{\boldsymbol{\vartheta}^{(i)}} = \inf_{\mathbf{p} \in \mathcal{D}_{\boldsymbol{p}}} L_\Lambda(\boldsymbol{\vartheta}^{(i)}, \mathbf{p})$
4. Compute $\boldsymbol{\vartheta}^{(i+1)} = \left[\boldsymbol{\vartheta}^{(i)} + \boldsymbol{\xi}^{\mathrm{T}} \cdot \nabla_{\boldsymbol{\vartheta}} g(\boldsymbol{\vartheta}^{(i)}) \right]^+$
5. $i \leftarrow i + 1$
6. Until $\left(||\mathbf{p}^{(i+1)} - \mathbf{p}^{(i)}|| < \epsilon \right)$ or $(i = I_{\max})$
7. Return $\mathbf{p}^* = \mathbf{p}^{(i)}$

Table 4.1: *Pseudo-code of the subgradient-based EGOPA-OP*

1) The optimal variable $\boldsymbol{\vartheta}^*$ is found solving the dual problem (4.12) with the subgradient based updating of the dual variable $\boldsymbol{\vartheta}$, as usually done in these cases (e.g. [70], [73]). The subgradient method [74], detailed in Appendix A.1.2, is an iterative

algorithm that maximizes $g(\boldsymbol{\vartheta})$ updating, at each step $i+1$, all the components of the dual variable $\boldsymbol{\vartheta}^{(i)}$ produced at the previous step i along the search direction defined by the subgradient of $g(\boldsymbol{\vartheta})$ at $\boldsymbol{\vartheta}^{(i)}$. The LDD method based on the subgradient update is summarized in Tab. 4.1. Therein, $\boldsymbol{\vartheta}^{(i)} \triangleq [\vartheta_0^{(i)}, \cdots, \vartheta_{Q_U+Q_I}^{(i)}]^{\mathrm{T}}$ contains the Lagrange multipliers at the ith iteration, with $\boldsymbol{\vartheta}_0$ denoting the chosen initial value, $\boldsymbol{\xi} = [\xi_0, \cdots, \xi_{Q_U+Q_I}]^{\mathrm{T}}$ is the step-size vector, ε is the maximum convergence error and $I_{\max}$ is the maximum number of iterations allowed. The dual variable update is done at line 4, where, at the generic step i, the subgradient, looking at (4.11) and (4.13), results $\nabla_{\boldsymbol{\vartheta}} g(\boldsymbol{\vartheta}^{(i)}) = \boldsymbol{f}(\mathbf{p}_{\boldsymbol{\vartheta}^{(i)}})$, Here, $\mathbf{p}_{\boldsymbol{\vartheta}^{(i)}}$ denotes the optimal PA for the given dual variable $\boldsymbol{\vartheta}^{(i)}$, evaluated using the closed-form expression (4.14) at line 3. The procedure is then iterated until convergence is achieved, in view of the strong duality property of the problem at hand [69].

2) This is obtained at the price of a demanding computational complexity load of $\mathcal{O}(1/\varepsilon^2)$. Further, whenever the problem dimensionality is high (as in our case) it incurs in an extremely slow convergence [70]. The above hard drawbacks will motivate an equivalent yet numerically efficient optimization method, as outlined in Sect. 4.5.

3) The optimal PA offers an interesting multi-level water-filling interpretation [66], [68], depicted in Fig. 4.2. The first term within the square brackets in (4.14), indeed, is the water level per subcarrier, determined by the available power and interference constraints, whereas the second one is the height of the vessel bottom that depends on the inverse of the signal-to-noise-plus-interference ratio (SINR).

4.5 Successive Set Reduction approach

In this section a novel approach to the PA problem for EGP maximization is formalized. Sections 4.5.1 and 4.5.2 provide motivations and the formal description of the SSR problem. In Section 4.5.3, resorting to the QVI theory, we show the existence of the SSR solution. In Sections 4.5.4 and 4.6 an iterative algorithm for PA based on the SSR approach is then introduced and some sufficient conditions for optimality under interweave and underlay interference constraints are also discussed. Finally, the AMC algorithm is described.

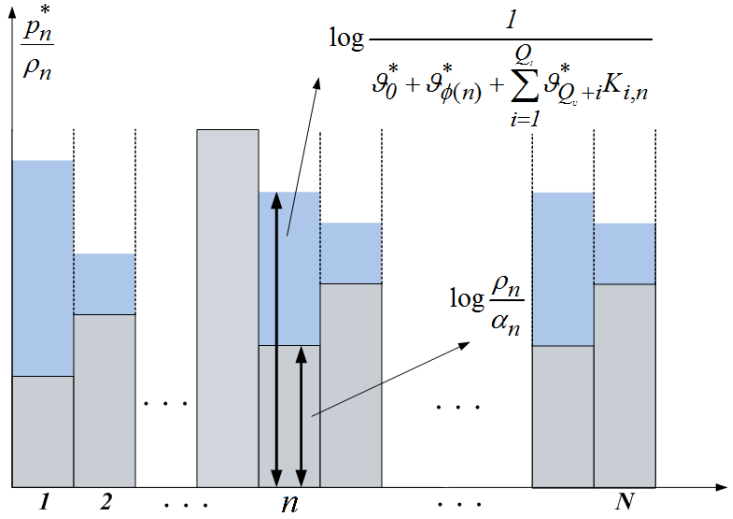

Figure 4.2: *Multi-level water-filling interpretation*

4.5.1 Rationale of the SSR approach

Unfortunately, conventional numerical methods, such as the LDD method described in previous Sect. 4.4.2, can lead to extremely slow convergence. Moreover, a careful off-line tuning of the step size $\boldsymbol{\xi}$ and Lagrange multipliers initialization are often required for guaranteeing a fast convergence. Therefore the following subsections present a novel methodology which overcomes these obstacles relying on a different geometric interpretation of the power allocation problem. Actually, differently from the LDD approach, which builds a lower bound of the objective function and operates over a new set of variables defined by the multipliers, the novel proposed method exploits the *monotonicity* property of the objective function $\psi(\mathbf{p})$ to rearrange the original optimization problem as a mathematical program with equilibrium constraints (MPEC)[4] [75]. This goal can be achieved by deriving a splitting rule for the original set of constraints, together with an optimality condition guaranteeing the equivalence between the original problem and the equivalent MPEC-based re-formulation. This design approach provides then an iterative algorithm that, step by step, moves towards

[4]MPEC constitute a special class of mathematical programs where the decision variables satisfy a finite number of constraints together with an equilibrium condition.

a "promising" region (i.e. a reduced set) where to look for the solution. More specifically, thanks to a proper partition of the set of feasible power vectors, each step of the algorithm produces a *closed-form* power increment that decreases more and more the value of the objective function $\psi(\mathbf{p})$, defined in (4.9), to be minimized. This power increment is obtained by optimizing a properly shifted version of $\psi(\mathbf{p})$ over a reduced set of feasible powers. For this reason, we will refer to this novel methodology as successive set reduction, or SSR for short, and, accordingly, to the relevant iterative algorithm as the SSR algorithm.

4.5.2 SSR approach for the EGP optimization problem

In this subsection, we define the novel SSR approach that offers a solution $\mathbf{p}_J \triangleq \sum_{j=1}^{J} \delta\mathbf{p}_j$, where the power increments $\delta\mathbf{p}_j \triangleq [\delta p_{j,1}, \cdots, \delta p_{j,N}]^{\mathrm{T}}$, $\forall j$, are obtained relying on a proper decomposition of the EGOPA-OP (4.10) into a set of J subproblems, each of which solved through a simple closed-form equation.

First of all, let us define the vectors $\delta\mathbf{p} \triangleq [\delta\mathbf{p}_1^{\mathrm{T}}, \delta\mathbf{p}_2^{\mathrm{T}}, \cdots, \delta\mathbf{p}_J^{\mathrm{T}}]^{\mathrm{T}} \in \Re^{NJ}$ and $\delta\mathbf{p}_{-j} \triangleq [\delta\mathbf{p}_1^{\mathrm{T}}, \cdots, \delta\mathbf{p}_{j-1}^{\mathrm{T}}, \delta\mathbf{p}_{j+1}^{\mathrm{T}}, \cdots, \delta\mathbf{p}_J^{\mathrm{T}}]^{\mathrm{T}} \in \Re^{N(J-1)}$. Thus, introducing the set $\mathcal{J} \triangleq \{1, 2, \cdots, J\}$, the J-dimensional SSR problem can be defined as follows.

Prop. 4.3 (SSR Problem) *The vector $\delta\mathbf{p}^* \in \Re^{NJ}$ is the SSR solution if it jointly solves the following optimization problems, tagged as SSR subproblems, $\forall j \in \mathcal{J}$*

$$\begin{aligned} \min_{\delta\mathbf{p}_j} \quad & \psi_j(\delta\mathbf{p}_j, \delta\mathbf{p}_{-j}) \\ \text{s.t.} \quad & \delta\mathbf{p}_j \succeq \mathbf{0} \qquad (4.24.\text{a}) \\ & u_j(\delta\mathbf{p}_j) \leq 0 \quad (4.24.\text{b}) \\ & w_j(\delta\mathbf{p}_j, \delta\mathbf{p}_{-j}) \leq 0 \quad (4.24.\text{c}) \end{aligned} \tag{4.24}$$

where both u_j and w_j are assumed to be continuously differentiable affine functions (that will be analytically derived in the next section) in the argument $\delta\mathbf{p}_j$ and

$$\psi_j(\delta\mathbf{p}_j, \delta\mathbf{p}_{-j}) \triangleq \sum_{n=1}^{N} \tilde{\alpha}_n(\delta\mathbf{p}_{-j}) \mathrm{e}^{-\frac{\delta p_{j,n}}{\rho_n}} \tag{4.25}$$

represents the objective function associated to the jth increment, with $\delta\mathbf{p}_0 \triangleq \mathbf{0}$ and

$$\tilde{\alpha}_n(\delta\mathbf{p}_{-j}) \triangleq \alpha_n \left(\prod_{\upsilon=0}^{j-1} \mathrm{e}^{-\frac{\delta p_{\upsilon,n}}{\rho_n}} \right). \tag{4.26}$$

Note that the new objective function in (4.25) is a properly shifted version of the original objective function (4.9), while the number of reduced sets J depends on the topology of the original set of the feasible power allocation vectors and on the splitting rule of the original set. The design guidelines for an iterative algorithm producing the reduced sets will be defined in the following sections. Now, let us denote the solution set of the jth SSR subproblem (4.24) as $\mathcal{S}_j(\delta\mathbf{p}_{-j}) \subset \Re^N$. In the sequel, we will refer to (4.25) either as $\psi_j(\delta\mathbf{p}_j, \delta\mathbf{p}_{-j})$ or $\psi_j(\delta\mathbf{p})$ at our best convenience. The SSR solution is then the NJ-tuple $\delta\mathbf{p}^*$ such that $\delta\mathbf{p}_j^* \in \mathcal{S}_j(\delta\mathbf{p}_{-j}^*)\ \forall j \in \mathcal{J}$. Equivalently, a necessary condition for the existence of the solution $\delta\mathbf{p}^*$ to the SSR problem is that there exist some constraint multipliers $\boldsymbol{\lambda} \triangleq [\lambda_1, \lambda_2, \cdots, \lambda_J]^{\mathrm{T}}$ and $\boldsymbol{\mu} \triangleq [\mu_1, \mu_2, \cdots, \mu_J]^{\mathrm{T}}$ such that the KKT systems

$$\nabla_{\delta\mathbf{p}_j}\psi_j(\delta\mathbf{p}_j^*, \delta\mathbf{p}_{-j}^*) + \lambda_j\nabla_{\delta\mathbf{p}_j}w_j(\delta\mathbf{p}_j^*, \delta\mathbf{p}_{-j}^*) + \mu_j\nabla_{\delta\mathbf{p}_j}u_j(\delta\mathbf{p}_j^*) = 0 \tag{4.27}$$

$$\lambda_j \geq 0, \ \ \mu_j \geq 0 \tag{4.27.a}$$

$$\mu_j u_j(\delta\mathbf{p}_j^*) = 0 \tag{4.27.b}$$

$$\lambda_j w_j(\delta\mathbf{p}_j^*, \delta\mathbf{p}_{-j}^*) = 0 \tag{4.27.c}$$

hold $\forall j \in \mathcal{J}$.

Finally, consider the point-to-set map $\mathcal{K}_j : \Re^{N(J-1)} \to \Re^N$, given by

$$\mathcal{K}_j(\delta\mathbf{p}_{-j}) \triangleq \left\{\delta\mathbf{p}_j \in \Re^N : w_j(\delta\mathbf{p}_j, \delta\mathbf{p}_{-j}) \leq 0, \quad u_j(\delta\mathbf{p}_j) \leq 0\right\}. \tag{4.28}$$

Since $\psi_j(\delta\mathbf{p}_j, \delta\mathbf{p}_{-j})$ is convex in $\delta\mathbf{p}_j$, a point $\delta\mathbf{p}_j^* \in \Re^N$, is the solution of the jth SSR subproblem if and only if (*minimum principle*) [71]

$$(\delta\mathbf{p}_j - \delta\mathbf{p}_j^*)^{\mathrm{T}}\nabla_{\delta\mathbf{p}_j}\psi_j(\delta\mathbf{p}_j^*, \delta\mathbf{p}_{-j}) \geq 0 \quad \forall\delta\mathbf{p}_j \in \mathcal{K}_j(\delta\mathbf{p}_{-j}). \tag{4.29}$$

Interestingly, the latter can be interpreted as an instance of

$$(\delta\mathbf{p} - \delta\mathbf{p}^*)^{\mathrm{T}}F(\delta\mathbf{p}^*) \geq 0 \qquad \forall\ \delta\mathbf{p} \in \mathcal{K}(\delta\mathbf{p}^*). \tag{4.30}$$

which is the QVI($\mathcal{K}, F$) problem, where $F(\delta\mathbf{p}) \triangleq [F_1^{\mathrm{T}}(\delta\mathbf{p}), \cdots, F_J^{\mathrm{T}}(\delta\mathbf{p})]^{\mathrm{T}}$: $\Re^{NJ} \to \Re^{NJ}$, with $F_j(\delta\mathbf{p}) \triangleq \nabla_{\delta\mathbf{p}_j}\psi_j(\delta\mathbf{p}_j, \delta\mathbf{p}_{-j})$, and $\mathcal{K}(\delta\mathbf{p}) \triangleq \prod_{j=1}^{J}\mathcal{K}_j(\delta\mathbf{p}_{-j})$. In the next subsection, we will provide sufficient conditions for such a QVI to have a solution.

4.5.3 Existence of the SSR solution

In order to demonstrate the existence of the solution for the SSR problem, we will pursue an approach analogous to the one proposed in [72] for generalized Nash equilibria, which is specifically tailored for sets $\mathcal{K}_j(\delta\mathbf{p}_{-j})$ that are representable by convex inequalities. However, differently from the generalized Nash equilibrium problem, here the equilibrium condition must be satisfied among the reduced subsets instead of among different competitive players. Thus, we will extend that approach to take into account the particular structure of our problem (4.30). In particular, as in [72], we will rely on both the well-known Kakutani's fixed-point theorem [71] and the following Sequentially Bounded Constraints Qualification (SBCQ) assumption [75] for each SSR subproblem $j \in \mathcal{J}$.

Prop. 4.4 SBCQ*: for any bounded sequence of vectors* $\{\delta\mathbf{p}_j^{(k)}\} \in \mathcal{S}_j(\delta\mathbf{p}_{-j}^{(k)})\ \forall k$*, there exists a corresponding bounded sequence* $\{\lambda_j^{(k)}\}$ *of Lagrange multipliers satisfying the* j*th KKT system.*

Remark. Accordingly, for each $\delta\mathbf{p}^{(k)} \triangleq [\delta\mathbf{p}_1^{(k)^{\mathrm{T}}}, \cdots, \delta\mathbf{p}_J^{(k)^{\mathrm{T}}}]^{\mathrm{T}}$ feasible to (4.30), KKT multipliers exist for the solution of the jth SSR problem (4.24), i.e., KKT conditions hold with bounded multipliers on bounded sets.

Let us start by defining the constraint functions u_j and w_j in (4.24.b) and (4.24.c), respectively, as

$$u_j(\delta\mathbf{p}_j) \triangleq -1 + \sum_{n=1}^{N} \frac{\delta p_{j,n}}{P_j}, \tag{4.31}$$

$$w_j(\delta\mathbf{p}_j, \delta\mathbf{p}_{-j}) \triangleq -1 + \sum_{n=1}^{N} \frac{\delta p_{j,n}}{\Theta_{j,n}(\delta\mathbf{p}_{-j})}, \tag{4.32}$$

where $\Theta_{j,n}(\delta\mathbf{p}_{-j})$ are convex differentiable functions and $\{P_j\}_{j=1}^{J}$ are properly defined positive constant values, and let

$$\mathcal{X}_j \triangleq \left\{\delta\mathbf{p}_j \in \Re^N : u_j(\delta\mathbf{p}_j) \leq 0\right\}, \quad \mathcal{X}_{-j} \triangleq \prod_{\upsilon=1,\upsilon\neq j}^{J} \mathcal{X}_\upsilon \tag{4.33}$$

be the sets identified only by the constraints depending on $\delta\mathbf{p}_j$. We now summarize in the following theorem some sufficient conditions which guarantee the existence of

the QVI solution.

Theorem 4.2 (Solution Existence) *Let $F : \Re^{NJ} \to \Re^{NJ}$ be a point-to-point map and let $\mathcal{K} : \Re^{NJ} \to \Re^{NJ}$ a point-to-set map such that*

a) *for each $\delta\mathbf{p}_{-j} \in \mathcal{X}_{-j}$ the set $\mathcal{K}_j(\delta\mathbf{p}_{-j})$ is nonempty, $\forall j \in \mathcal{J}$ (feasibility assumption),*

b) *$\Theta_{j,n}(\delta\mathbf{p}_{-j}) > 0$, $\forall j \in \mathcal{J}$ and $\forall n \in \mathcal{D}_s$,*

c) *the set $\mathcal{X}_j$ is nonempty and bounded, $\forall j \in \mathcal{J}$ (compactness assumption),*

then the QVI($\mathcal{K}, F$) solution is ensured.

Proof At first, note that the SSR subproblem defined as in (4.24), parameterized in $\delta\mathbf{p}_{-j}$, is convex in $\delta\mathbf{p}_j$. Hence, being $\Theta_{j,n}(\delta\mathbf{p}_{-j}) > 0$ $\forall n \in \mathcal{D}_s$ and $\forall j \in \mathcal{J}$ and thus belonging $\delta\mathbf{p}_j$ to a bounded set, there always exists a bounded pair (λ_j, μ_j). This means that the SBCQ assumption holds.
Now, exploiting assumption *c*) we can deduce that

$$\mathcal{X} \triangleq \prod_{j=1}^{J} \mathcal{X}_j \tag{4.34}$$

is a compact, nonempty and convex subset of $\Re^{NJ}$. Hence, the mapping defined as

$$\Phi(\delta\mathbf{p}) \triangleq \prod_{j=1}^{J} \mathcal{S}_j(\delta\mathbf{p}_{-j}) \tag{4.35}$$

is a nonempty, compact and convex subset of $\mathcal{X}$. Even looking at (4.35), since for a given $\delta\mathbf{p}_{-j}$ the original optimization problem is convex, the set $\mathcal{S}_j$ is a singleton. It follows that the QVI($\mathcal{K}$, F) has solution given that the set-valued map Φ has a fixed point, i.e., $\delta\mathbf{p}^* = \Phi(\delta\mathbf{p}^*)$. Then, under the Kakutani's fixed point theorem, demonstrating the existence of the SSR solution amounts to show that Φ is a closed point-to-set map, according to the definition in [76].
As shown in the following, under the SBCQ, there always exist two bounded multipliers vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ such that $\delta\bar{\mathbf{v}}_j \in \mathcal{S}_j(\delta\bar{\mathbf{p}}_{-j})$, $\forall j \in \mathcal{J}$, thus demonstrating the

existence of the SSR solution. First of all, let us introduce the definition of closed point-to-set mapping.

Definition 4.1 (Closed point-to-set mapping, [76]) *The mapping Φ is closed at a point $\delta\bar{\mathbf{p}} \triangleq [\delta\bar{\mathbf{p}}_1^{\mathrm{T}}, \cdots, \delta\bar{\mathbf{p}}_J^{\mathrm{T}}]^{\mathrm{T}} \in \mathcal{X}$ if, given two sequences $\{\delta\mathbf{p}^{(k)}\}$ and $\{\delta\mathbf{v}^{(k)}\}$ belonging to $\mathcal{X}$, with $\delta\mathbf{p}^{(k)} \triangleq [\delta\mathbf{p}_1^{(k)^{\mathrm{T}}}, \cdots, \delta\mathbf{p}_J^{(k)^{\mathrm{T}}}]^{\mathrm{T}}$ and $\delta\mathbf{p}^{(k)} \triangleq [\delta\mathbf{v}_1^{(k)^{\mathrm{T}}}, \cdots, \delta\mathbf{v}_J^{(k)^{\mathrm{T}}}]^{\mathrm{T}}$, such that*

$$\lim_{k\to\infty} \delta\mathbf{p}_j^{(k)} = \delta\bar{\mathbf{p}}_j, \qquad \lim_{k\to\infty} \delta\mathbf{v}_j^{(k)} = \delta\bar{\mathbf{v}}_j \tag{4.36}$$

and $\delta\mathbf{v}_j^{(k)} \in \mathcal{S}_j(\delta\mathbf{p}_{-j}^{(k)})$ $\forall k$ and $\forall j$, then it follows that $\delta\bar{\mathbf{v}}_j \in \mathcal{S}_j(\delta\bar{\mathbf{p}}_{-j})$.

For a given $j \in \mathcal{J}$ define, $\mathbf{z}^{(k)}(j) \triangleq [\mathbf{z}_1^{(k)}(j)^{\mathrm{T}}, \cdots, \mathbf{z}_l^{(k)}(j)^{\mathrm{T}}, \cdots, \mathbf{z}_J^{(k)}(j)^{\mathrm{T}}]^{\mathrm{T}}$ $\forall k$, where

$$\mathbf{z}_l^{(k)}(j) = \begin{cases} \delta\mathbf{v}_j^{(k)} & \text{if } l = j, \\ \delta\mathbf{p}_l^{(k)} & \text{otherwise.} \end{cases} \tag{4.37}$$

From (4.36) we have

$$\bar{\mathbf{z}}_l(j) \triangleq \lim_{k\to\infty} \mathbf{z}_l^{(k)}(j) = \begin{cases} \delta\bar{\mathbf{v}}_j, & \text{if } l = j, \\ \delta\bar{\mathbf{p}}_l, & \text{otherwise.} \end{cases} \tag{4.38}$$

For the sake of notation, from now on the dependence on j will be omitted. Thus, exploiting the assumption a) and b) and the convex nature of optimization problem (4.24), for each k, there exist two KKT multipliers $\lambda_j^{(k)}$ and $\mu_j^{(k)}$ such that the jth system

$$\begin{array}{c} \nabla_{\delta\mathbf{p}_j}\psi_j(\mathbf{z}_j^{(k)}) + \lambda_j^{(k)}\nabla_{\delta\mathbf{p}_j}w_j(\mathbf{z}^{(k)}) + \mu_j^{(k)}\nabla_{\delta\mathbf{p}_j}u_j(\mathbf{z}_j^{(k)}) = 0 \\ \lambda_j^{(k)} \geq 0, \;\; \mu_j^{(k)} \geq 0, \;\; \mu_j^{(k)}u_j(\mathbf{z}_j^{(k)}) = 0, \;\; \lambda_j^{(k)}w_j(\mathbf{z}^{(k)}) = 0 \end{array} \tag{4.39}$$

is solved. Since the SBCQ is satisfied, we can claim that there always exist two bounded KKT multipliers $\bar{\lambda}_j$ and $\bar{\mu}_j$ such that

$$\lim_{k\to\infty} \lambda_j^{(k)} = \bar{\lambda}_j \quad \lim_{k\to\infty} \mu_j^{(k)} = \bar{\mu}_j. \tag{4.40}$$

Thus, for $k \to \infty$, $\delta\bar{\mathbf{v}}_j \in \mathcal{S}_j(\delta\bar{\mathbf{p}}_{-j})$, $\forall j \in \mathcal{J}$, as desired. □

4.5.4 SSR algorithm

We now reformulate the EGP maximization problem in a simpler equivalent form obtained through the SSR approach. To accomplish this task, we first derive the

condition under which the power vector

$$\mathbf{p}_J \triangleq \sum_{j=1}^{J} \delta\mathbf{p}_j^* = [p_{J,1}, \cdots, p_{J,N}]^{\mathrm{T}}, \tag{4.41}$$

referred to as SSR power allocation (SSR-PA) and obtained as the combination of the single components of the SSR solution, solves the EGOPA-OP (4.10).

First of all, let us note that the formulation of the SSR subproblem (4.24) is analogous to the PA problem for EGP optimization originally proposed in [45] for a non-cognitive context. Assuming, w.l.g., that in (5.19) and (4.31) $P_j \geq \Theta_{j,n}(\delta\mathbf{p}_{-j})$, $\forall n \in \mathcal{D}_s$, the following theorem can be stated.

Theorem 4.3 *The solution of the jth SSR subproblem (4.24), $\delta\mathbf{p}_j^* \triangleq [\delta p_{j,1}^*, \cdots, \delta p_{j,N}^*]^{\mathrm{T}}$, $\forall j \in \mathcal{J}$, can be expressed by*

$$\delta p_{j,n}^* = \rho_n \left[\frac{1 + \sum_{i=1}^{N} \tilde{\rho}_{j,i} \log\left(\frac{\tilde{\rho}_{j,i}\, \tilde{\alpha}_{j,n}}{\tilde{\alpha}_{j,i}\, \tilde{\rho}_{j,n}} \right)}{\sum_{k=1}^{N} \tilde{\rho}_{j,k}} \right]^{+} \quad \forall n \in \mathcal{D}_s, \tag{4.42}$$

where $\tilde{\rho}_{j,n} \triangleq \frac{\rho_n}{\Theta_{j,n}}$ and, for the sake of readability, the dependence of $\Theta_{j,n}$ and $\tilde{\alpha}_{j,n}$ to $\delta\mathbf{p}_{-j}$ is omitted.

Proof The reduced set PA problem corresponds to the PA algorithm for EGP optimization proposed [45], where the constraint on the total available power is here substituted by constraint (4.24.c). Accordingly, this problem has a closed-form solution that is obtained as summarized in the following, where the dependence on the step j of the SSR algorithm is dropped w.l.g. Be $\delta\mathbf{p}^*$ the solution of the OP (4.24), its directional derivative must satisfy

$$(\delta\mathbf{p} - \delta\mathbf{p}^*)^{\mathrm{T}} \bigtriangledown \psi_j(\delta\mathbf{p}^*) = \sum_{n=1}^{N} (\delta p_n - \delta p_n^*) \left. \frac{\partial \psi_j(\delta\mathbf{p})}{\partial \delta p_n} \right|_{\delta\mathbf{p}=\delta\mathbf{p}^*} \geq 0, \tag{4.43}$$

which yields

$$\sum_{n=1}^{N} p_n \left. \frac{\partial \psi_j(\delta\mathbf{p})}{\partial \delta p_n} \right|_{\delta\mathbf{p}=\delta\mathbf{p}^*} \geq \sum_{n=1}^{N} \delta p_n^* \left. \frac{\partial \psi_j(\delta\mathbf{p})}{\partial \delta p_n} \right|_{\delta\mathbf{p}=\delta\mathbf{p}^*} \triangleq C. \tag{4.44}$$

Since (4.43) is an affine function of the power $\delta\mathbf{p}$, it can be evaluated on the N extreme points $\{\Theta_n\}_{n=1}^{N}$ of the set $\mathcal{D}_\Theta$ where, if $\delta p_n > 0\ \forall n$, (4.44) holds with strict equality.

Thus, setting $\delta p_k = \Theta_k$, with Θ_k given by (4.59), and $\delta p_n = 0,\ \forall n \neq k$, in (4.44) we get the optimality condition

$$\Theta_k \left.\frac{\partial \psi_j(\delta \mathbf{p})}{\partial \delta p_k}\right|_{\delta \mathbf{p}=\delta \mathbf{p}^*} = C,\ \ \forall k,\ 1 \leq k \leq N. \tag{4.45}$$

This means that all the weighted components of the gradient of the objective function must be equal. Considering N active subcarriers, this condition can be obtained in N steps, where at the generic step i, with $i < N$, the current power $\delta p_k^{(i-1)}$ over all the subcarriers $k \leq i$ is increased of $\delta \tilde{p}_k^{(i)}$, i.e., $\delta p_k^{(i)} = \delta p_k^{(i-1)} + \delta \tilde{p}_k^{(i)}$, so that condition (4.45) is satisfied $\forall k \leq i+1$. In particular, defining

$$\delta \mathbf{p}^{(i)} \triangleq \left[\sum_{k=1}^{i} \delta \tilde{p}_1^{(k)}, \sum_{k=2}^{i} \delta \tilde{p}_2^{(k)}, \cdots, \delta \tilde{p}_i^{(i)}, \overbrace{0, \cdots, 0}^{N-i}\right]^{\mathrm{T}}, \tag{4.46}$$

from (4.44)

$$\Theta_1 \left.\frac{\partial \psi_j(\delta \mathbf{p})}{\partial \delta p_1}\right|_{\delta \mathbf{p}=\delta \mathbf{p}^{(i)}} = \cdots = \Theta_{i+1} \left.\frac{\partial \psi_j(\delta \mathbf{p})}{\partial \delta p_{i+1}}\right|_{\delta \mathbf{p}=\delta \mathbf{p}^{(i)}} \tag{4.47}$$

we get

$$\delta \tilde{p}_i^{(i)} = \rho_i \log\left(\frac{\tilde{\alpha}_i \tilde{\rho}_{i+1}}{\tilde{\rho}_i \tilde{\alpha}_{i+1}}\right)\ \ (4.48.\text{a}),\ \ \frac{\delta \tilde{p}_k^{(i)}}{\rho_k} = \frac{\delta \tilde{p}_i^{(i)}}{\rho_i}\ \ \forall k \leq i\ \ (4.48.\text{b}), \tag{4.48}$$

where $\tilde{\rho}_i \triangleq \frac{\rho_i}{\Theta_i}$. The last increment $\delta \tilde{p}_n^{(N)}$, $1 \leq n \leq N$, is such that constraint (4.24.a) holds with strict equality

$$\sum_{n=1}^{N} \left(\prod_{k=1, k\neq n}^{N} \Theta_k \delta \tilde{p}_n^{(N)}\right) + \sum_{i=1}^{N-1} \sum_{m=i}^{N-1} \left(\prod_{j=1, j\neq i}^{N} \Theta_j \delta \tilde{p}_i^{(m)}\right) = \prod_{i=1}^{N} \Theta_i. \tag{4.49}$$

Substituting $\delta \tilde{p}_n^{(N)} = \rho_n \frac{\delta \tilde{p}_N^{(N)}}{\rho_N}$ according to (4.48.b) into (4.49), $\delta \tilde{p}_N^{(N)}$ is obtained. Plugging its expression into

$$\delta p_n^* \triangleq \sum_{k=1}^{N} \delta \tilde{p}_n^{(k)} = \rho_n \left[\log\left(\frac{\tilde{\alpha}_n}{\tilde{\rho}_n}\right) - \log\left(\frac{\tilde{\alpha}_N}{\tilde{\rho}_N}\right)\right] + \rho_n \frac{\delta \tilde{p}_N^{(N)}}{\rho_N}, \tag{4.50}$$

the closed form (4.42) is eventually found. □

It is then possible to exploit the expression of the jth SSR solution, providing a simple way to compute the SSR-PA. Given a fixed j, let define the subcarrier indices set $\mathcal{N}_j$ as

$$\mathcal{N}_j \triangleq \{n \in \mathcal{D}_s : \delta p_{j,n} \neq 0\},\ \ \forall j \in \mathcal{J}, \tag{4.51}$$

so that, jointly exploiting (4.26), (4.41) and (4.42), we get the SSR-PA

$$\frac{p_{J,n}}{\rho_n} = \log\left(\frac{1}{\lambda_J}\right) + \log\left(\Theta_{J,n}\right) - \log\left(\frac{\rho_n}{\alpha_n}\right) \quad \forall n \in \mathcal{N}_J \tag{4.52}$$

with

$$\log\left(\frac{1}{\lambda_J}\right) = \frac{1 + \sum\limits_{i\in\mathcal{N}_J} \frac{1}{\Theta_{J,n}} \left[\rho_i \log\left(\frac{\rho_i}{\alpha_i\, \Theta_{J,n}}\right) + p_{J-1,i}\right]}{\sum\limits_{k\in\mathcal{N}_J} \frac{\rho_k}{\Theta_{J,k}}}. \tag{4.53}$$

It is now worth to highlight the following result.

Prop. 4.5 (cEGOPA-OP) *The SSR-PA* $\mathbf{p}_J$ *given in (4.52) is also the solution to the following convex optimization problem, tagged as cost-based EGOPA-OP or cEGOPA-OP:*

$$\begin{aligned} \min_{\mathbf{p}} \quad & \psi(\mathbf{p}) + \lambda_J\, C_J(\mathbf{p}) \\ \text{s.t.} \quad & \mathbf{p} \succeq \mathbf{0} \end{aligned} \tag{4.54}$$

where

$$C_J(\mathbf{p}) \triangleq \sum_{n=1}^{N} \frac{p_n}{\Theta_{J,n}} - 1 \tag{4.55}$$

plays the role of a linear cost function.

The equivalence can be easily verified by solving the KKT system of the cEGOPA-OP (4.54) and noting that the gradient of the cost function is a constant vector given by $\nabla_{\mathbf{p}} C_J(\mathbf{p}_J) = [1/\Theta_{J,1}, 1/\Theta_{J,2}, \cdots, 1/\Theta_{J,N}]^{\mathrm{T}}$. Interestingly, in the objective function of (4.54) we meet again the objective function $\psi(\mathbf{p})$ of the EGOPA-OP (4.10). Comparing (4.52) with (4.14), we infer that, in general, the SSR approach does not provide solutions to (4.10). An interesting question is whether one can design the reduced set constraints so that the SSR solution coincides with the optimal PA. The answer is given by the following theorem.

Theorem 4.4 (Optimality Condition) *Let the set* $\mathcal{I}(\mathbf{p}_J) \triangleq \{i : f_i(\mathbf{p}_J) = 0\}$ *be the non-empty set of indices associated to the constraints* $f_i(\mathbf{p}_J)$ *that hold with equality in* $\mathbf{p} = \mathbf{p}_J$ *and assume that* $\mathbf{p}_J$ *is a feasible vector to the EGOPA-OP (4.10) and the following condition is satisfied*

$$\frac{\partial C_J(\mathbf{p}_J)}{\partial p_n} = \frac{1}{\Theta_{J,n}} = \frac{A}{\sigma_{\phi(n)} + \mu \sum\limits_{i\in\mathcal{I}(\mathbf{p}_J),\, i>Q_U} K_{i,n}}, \quad \forall n \in \mathcal{N}_J \tag{4.56}$$

where $\sigma_{\phi(n)}$, μ, A are constant coefficients, such that $\sigma_{\phi(n)} \geq 0$, $\mu \geq 0$, $\sigma_{\phi(n)} + \mu > 0$, and $A > 0$.
Then, the power vector $\mathbf{p}_J$ obtained through (4.41) represents the optimal solution of EGOPA-OP.

Proof By writing the KKT system associated to the cEGOPA-OP (4.54) and comparing it with the KKT condition (4.15) of the EGOPA-OP, we obtain that $\mathbf{p}_J$ is the solution of the original problem if and only if it exists a positive vector $\boldsymbol{\vartheta}$ such that

$$\sum_{i \in \mathcal{I}(\mathbf{p}_J)} \vartheta_i \nabla_{\mathbf{p}} f_i(\mathbf{p}_J) = \lambda_J \nabla_{\mathbf{p}} C(\mathbf{p}_J) \tag{4.57}$$

where $\vartheta_i = 0$, $\forall i \notin \mathcal{I}(\mathbf{p}_J)$. In other words, we can claim that $\mathbf{p}_J$ is the solution of our PA problem if $\nabla_{\mathbf{p}} C(\mathbf{p}_J)$ belongs to the convex cone generated by the vectors $\{\nabla_{\mathbf{p}} f_i(\mathbf{p}_J)\}_{i \in \mathcal{I}(\mathbf{p}_J)}$. From the Farka's Lemma, [69], this is true if it does not exist a vector $\mathbf{y} \in \Re^N$ such that

$$\mathrm{Z}_J^{\mathrm{T}} \mathbf{y} \geq 0 \;\; (4.58.\mathrm{a}) \text{ and } \nabla_{\mathbf{p}} C(\mathbf{p}_J)^{\mathrm{T}} \mathbf{y} < 0 \;\; (4.58.\mathrm{b}), \tag{4.58}$$

with $\mathrm{Z}_J \triangleq \left[\nabla_{\mathbf{p}} f_{i_1}(\mathbf{p}_J)^{\mathrm{T}}, \cdots, \nabla_{\mathbf{p}} f_{i_M}(\mathbf{p}_J)^{\mathrm{T}}\right]^{\mathrm{T}}$ and $\{i_1, \cdots, i_M\} \in \mathcal{I}(\mathbf{p}_J)$. Substituting (4.56) into (4.58.a) and (4.58.b) and after some algebra the proof follows. □

Theorem 4.4 provides sufficient conditions for the equivalence between the solutions of the SSR and of the EGOPA-OP, and, since (4.10) is convex, it also states the uniqueness condition for the QVI($\mathcal{K}, F$). Let us note that the sequence in which the J sub-problems are solved does not affect, in general, the optimality of the SSR solution, since it is sufficient for the partition of the set, no matter how obtained, to satisfy the optimality condition at the last increment. Unfortunately, there is no simple procedure to find the point-to-set map $\mathcal{K}(\delta \mathbf{p}^*)$ guaranteeing the optimality conditions. Hence, since the practical implementation of the SSR is viable only if the J-sized vector of power increments can be easily computed, in the next section we focus on a near-optimal criterion, referred to as *extreme points criterion*, to easily design the reduced subsets for all $j \in \mathcal{J}$.

4.6 Extreme points criterion

Let us assume that the values $\{\Theta_{j,n}\}_{n=1}^{N}$ identify the extreme points of the set, i.e., the maximum allowed power increment per each subcarrier that does not violate any

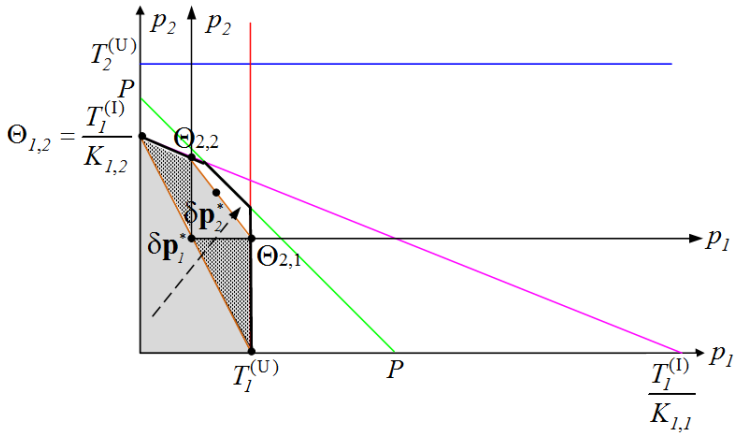

Figure 4.3: *Graphical interpretation of the extreme points criterion.*

of the constraint of the problem, regardless of the other subcarriers. These values are expressed by

$$\Theta_{j,n} \triangleq \min\left\{\bar{P}_{j,n}, \{\bar{I}_{j,i,n}\}_{i=1}^{Q_I}, \bar{T}_{j,n}\right\}, \tag{4.59}$$

with

$$\bar{P}_{j,n} \triangleq P - \sum_{\nu=1}^{N} p_{j-1,\nu}, \tag{4.60}$$

$$\bar{I}_{j,i,n} \triangleq \left(T_i^{(\mathrm{I})} - \sum_{\nu=1}^{N} p_{j-1,\nu} K_{i,\nu}\right) / K_{i,n}, \quad 1 \leq i \leq Q_I, \tag{4.61}$$

and

$$\bar{T}_{j,n} \triangleq T_{\phi(n)}^{(\mathrm{U})} - \sum_{\nu:\phi(\nu)=\phi(n)} p_{j-1,\nu}. \tag{4.62}$$

As apparent, the reduced set $\mathcal{K}_j$ is obtained by the intersection of the halfspace of the positive power increments with the halfspace lying below the hyperplane passing by the extreme points defined as $E_{j,n} \triangleq \{\overbrace{0,\cdots,0}^{n-1}, \Theta_{j,n}, \overbrace{0,\cdots,0}^{N-n}\}$, $\forall n \in \mathcal{D}_s$, properly shifted to take into account the partial solution $\mathbf{p}_j$ obtained so far. Figure 4.3 illustrates the extreme points criterion for a "toy" case with $N = 2$ subcarriers, $Q_U = 2$ UPUs, with $\phi(1) = 1$ and $\phi(2) = 2$, and $Q_I = 1$ IPU. The set delimited by the black thick line represents the set of feasible power values, identified by the intersection of constraints (4.10.a)-(4.10.b). The algorithm starts evaluating the first

reduced set, consisting of the larger grey-shaded triangular-shaped area, and finds the first increment of power that belongs to the edge of this set. Then, the new reduced set is built by re-centering the origin of the axes on the previously found solution $\delta\mathbf{p}_1$ and evaluating the new extreme points. This second step evaluates then a new reduced set, consisting of the smaller grey-shaded triangular-shaped area, and proceeds again by finding the second increment of power $\delta\mathbf{p}_2$ that belongs to the edge of this new set. This procedure is iterated until the last increment lies on the initial set border.

Some observations are now in order.

1) *Features.* The proposed SSR algorithm iteratively reduces the set of feasible power increments until, after J steps, no further power increment that would not violate any of the constraints is feasible, i.e., $\Theta_{J,n} = 0,\ \forall n \in \mathcal{D}_s$. However, in order to avoid possible cases where $J \to \infty$, a safer stop criterion was adopted, such that $||\mathbf{\Theta}_J|| \leq \epsilon_{\text{SSR}}$, where ϵ_{SSR} is a conveniently small value. In particular, at each step, the best local choice is performed solving the SSR subproblem (4.24). This problem is equivalent to the PA problem in [45], with the only difference that the constraint on the total available power is now replaced by the constraint of feasible power increments. In a way, the original constraints (4.10.b) are progressively dumped on $\Theta_{j,n}$, so that, at each step, the problem formulation is analogous to a non-cognitive case.

2) *Optimality.* Let us denote with $\tilde{\mathcal{N}}$ the set of subcarrier indices such that $\Theta_{J,n} > 0$ and assume that $\mathcal{I}(\mathbf{p}_J)$ is a singleton. Comparing equations (4.59)-(4.62) with (4.56), it can be noted that the SSR solution is optimal for those subcarriers that belong to $\tilde{\mathcal{N}}$. As a consequence, if $\mathbf{p}_{J-1}$ is an interior point of the original set, global optimality occurs whenever the set $\mathcal{I}(\mathbf{p}_J)$ is a singleton, i.e. every time that, at the optimal solution, only one constraint in (4.10.b) holds with equality. In this sense, we can claim that the SSR algorithm follows a *near optimality criterion.*

3) *Complexity.* Both the power increment $\delta\mathbf{p}_j$ and the parameters $\{\Theta_{j,n}, \tilde{\alpha}_{j,n}, \tilde{\rho}_{j,n}\},\ \forall n$, at each step j have a closed-form expression, so that the complexity of the algorithm simply reduces to $\mathcal{O}(J)$.

The SSR algorithm is briefly summarized in the pseudo-code of Table 4.2.

SSR Algorithm
1. Initialize $j = 1$, $\Delta\mathbf{p}^{(0)} = \mathbf{0}$
2. Repeat
3. $\quad j \leftarrow j + 1$
4. $\quad$ Evaluate $\mathbf{p}_j$, $\mathbf{\Theta}_j$, $\{\tilde{\alpha}_{j,n}, \tilde{\rho}_{j,n}\}$ $\forall n$
5. $\quad$ Evaluate $\delta\mathbf{p}_j$ according to (4.42)
6. Until $(\|\|\mathbf{\Theta}_j\|\| < \epsilon_{\mathrm{SSR}})$
7. Set $J = j$
8. Return $\mathbf{p}^* = \sum_{j=1}^{J} \delta\mathbf{p}_j$

Table 4.2: *Pseudo-code of the SSR algorithm*

4.7 AMC-SSR algorithm

As outlined in Sect. 4.3, once the PA for a certain TM is evaluated, then the EGP is maximized performing a simple exhaustive search overall the possible TMs. Denoting with $\overline{m}^{(i)}$ the ith element of the set $\mathcal{D}_m$, this procedure, named AMC-SSR algorithm, is made of M steps and is summarized in Table 4.3.

It is worth noting that this exhaustive approach is made possible thanks to the low complexity of the SSR algorithm that in J closed-form steps is able to evaluate the PA $\mathbf{p}^*(\boldsymbol{\varphi})$ for a given $\boldsymbol{\varphi}$. The advantage earned with the AMC approach is great: at each SNR, we are able to always choose the best setting of transmission parameters, so that the GP performance are always greater or equal than the ones obtained keeping a fixed TM.

4.8 Simulation results

In this section, the LRA method is numerically verified for realistic cognitive wireless scenarios. First, the SSR algorithm is checked against the LDD technique to ensure that, practically, both of them yield the same EGP performance. Then, the proposed AMC strategy based on the SSR algorithm is evaluated in terms of the actual GP (AGP) performance.

AMC-SSR Algorithm
1. Initialize $\zeta^{(0)} = 0$
2. For $i = 1 : M$
3. - Set $m^{(i)} = \overline{m}^{(i)}$
4. - Evaluate $\mathbf{p}^{(i)}$ and $\gamma^{(i)}$
5. - Enter $\gamma^{(i)}$ into the LUT and return the coding
6. rate $r^{(i)}$ associated to the best EGP value $\zeta^{(i)}$
7. - If $\left(\zeta^{(i)} > \zeta^{(i-1)}\right)$
8. Set $\mathbf{p}^* = \mathbf{p}^{(i)}$, $m^* = m^{(i)}$, $r^* = r^{(i)}$
9. End For
10. Return $(\mathbf{p}^*, m^*, r^*)$

Table 4.3: *Pseudo-code of the AMC-SSR algorithm*

System setup

Simulations have been carried out on a BIC-OFDM scenario denoted by an overall system radio frequency bandwidth of $B_{\text{tot}} = 20$ MHz, central frequency of $f_0 = 2$ GHz and subcarrier spacing of 15.152 kHz. The simulation parameters setup of the BIC-OFDM system is reported in Tab. 4.4. Accordingly, the total number of TMs is 12 when the convolutional code is employed and 24 under turbo code adoption. The radio propagation channel model is instead reported in Tab 4.5.

The primary network considers Q_U UPUs and Q_I IPUs, transmitting over contiguous bands having the same size, i.e., $B_u = W_i = B_{\text{tot}}/(Q_I+Q_U)$, $1 \leq u \leq Q_U$, $1 \leq i \leq Q_I$. As depicted in Fig. 4.4, the SRx is placed at the origin of the reference system with the STx being 160 m away. The IPUs are set round the secondary receiver within a radius of $R = 200$ m, the UPUs are placed round the secondary receiver at a distance of at least 200 m from it, according to the interference-free zone mentioned in Sect. 4.2.2. The interference threshold, referred at the primary receivers, is set to I_{th} for all the UPUs and IPUs.

Validation of the SSR algorithm

Figure 4.5 compares the optimum EGP, obtained with the LDD technique (empty

Parameter/Feature	Symbol	Value/Description
Payload length	N_p	1024 bits
CRC length	N_{CRC}	32 bits
OFDM (Figs. 4.5, 4.6)		
No. of subcarriers	N	64
FFT size	N_{fft}	64
CP length	N_{cp}	16 samples
OFDM (Figs. 4.7, 4.8, 4.9)		
No. of subcarriers	N	1320
FFT size	N_{fft}	2048
CP length	N_{cp}	160 samples
Transmission Parameters		
Bits per subcarrier	$\mathcal{D}_m$	$\{2, 4, 6\}$
Code type (Figs. 4.5, 4.6, 4.7, 4.8)		64-state Convolutional code
Mother code rate	r_0	1/2
Punctured code rates	$\mathcal{D}_r$	$\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}\right\}$
Code type (Figs. 4.7, 4.9)		Turbo code
Mother code rate	r_0	1/3
Punctured code rates	$\mathcal{D}_r$	$\left\{\frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{4}{7}\frac{2}{3}\frac{3}{4}\frac{4}{5}\frac{6}{7}\right\}$
Available power	P	$[0, 50]$ dBm

Table 4.4: *Parameters and features of the cognitive BIC-OFDM system.*

circles) and the EGP produced by the SSR algorithm (full circles), versus the mean-available-symbol-energy-to-noise-spectral density ratio E_s/N_0 ratio at the SRx, where E_s is evaluated assuming that all the available transmit power is employed. These results are obtained averaging 10^3 independent channel realizations, with the STx employing $N_{\text{tot}} = 64$ subcarriers together with a convolutional encoder and static modulation and coding according to three different TMs, i.e., $\boldsymbol{\varphi}_1 \triangleq \{2, 3/4\}$, $\boldsymbol{\varphi}_2 \triangleq \{4, 1/2\}$ and $\boldsymbol{\varphi}_3 \triangleq \{6, 5/6\}$. The primary network is composed of $Q_U = 1$ UPU and $Q_I = 2$ IPUs, placed at 660, 85 and 52 meters from the STx, respectively. The interference temperature is $I_{\text{th}} = -110$ dBm. The LDD parameters are $\xi_m = \vartheta_m^{(0)}/10$,

Parameter/Feature	Value/Description
Path-loss model	NLOS urban scenario, [77]
Carrier frequency	2 GHz
Base station height	12.5 m
Mobile terminal height	1.5 m
Noise power level	-100 dBm
Short-term fading model	ITU Ped. B, [33]

Table 4.5: *Parameters and features of the radio propagation channel model.*

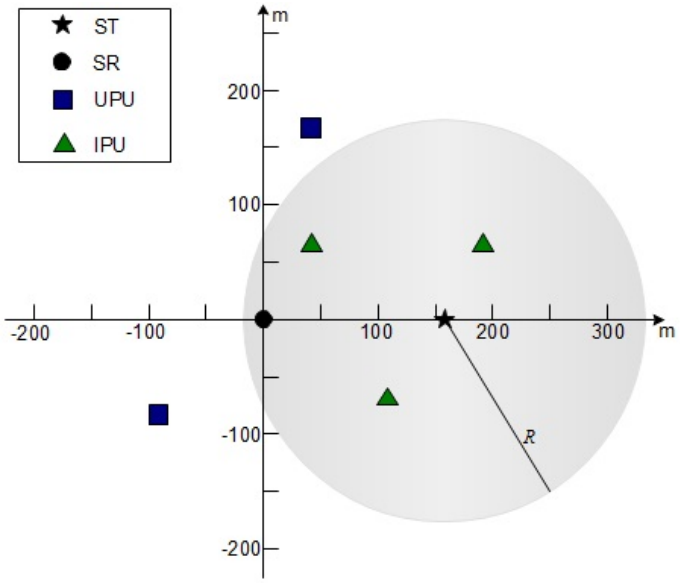

Figure 4.4: *Simulation scenario. Example with $Q_U = 2$, $Q_I = 3$.*

$0 \leq m \leq Q_U + Q_I + 1$, $\varepsilon = 10^{-4}$ and $I_{\max} = 10^6$.

For all the considered TMs, it is apparent that the SSR algorithm suffers a slight performance loss with respect to the LDD method, which can be appreciated for $\boldsymbol{\varphi}_2$ only, while it is negligible for all the other ones. The advantage we earn applying the SSR algorithm instead of conventional numerical methods is enlightened in Fig. 4.6, which depicts the number of iterations required to converge by the SSR (solid line),

LDD (circles) and interior-point (squares), the latter for $\boldsymbol{\varphi}_4 \triangleq \{6, 2/3\}$, algorithms. We observe that the SSR algorithm needs less than 10 iterations to converge for any TM, the LDD method more than some orders of magnitude and the interior-point method about 50 outer iterations and, for each of them, an average of 10 inner iterations. Moreover, each iteration of the SSR is in closed-form, compared to the iterations of the interior-point method that are based on numerical algorithms, obtaining thus great saving in computational complexity.

Further, Fig. 4.7 shows the comparison between the EGP defined in (3.17) (empty marks) and the AGP (full marks), i.e., the average of the ratio between the number of information bits N_p and the time required to successfully deliver them, versus the E_s/N_0 ratio when the AMC-SSR algorithm is adopted. Both convolutional (squares) and turbo (circles) encoders are considered, with $N_{\text{tot}} = 1320$ subcarriers, $Q_U = 2$ UPUs and $Q_I = 3$ IPUs at 400, 597, 85, 52 and 87 meters from the STx, respectively, and $I_{\text{th}} = -110$ dBm. It can be seen how the EGP reliably works in predicting the actual link performance, exhibiting a normalized error that is always lower than 10%. Figure 4.7 also shows the benefit of adopting the AMC algorithm w.r.t. static TMs (dotted grey lines). As matter of fact, if the modulation order and coding rate were fixed, the system performance would incur in one of the following two cases: *i*) if a conservative TM were adopted, i.e., low modulation order and/or high coding rate, the performance would be good at low E_s/N_0, but would prematurely flatten out as the E_s/N_0 increases; ii) on the contrary, if more aggressive TM were employed, we would get low AGP values at low E_s/N_0 and good values only in medium-high E_s/N_0 region. The AMC algorithm instead always selects the best TM making the performance lie over the envelope determined by the static modes. The same behavior was obtained with turbo codes (not reported on the graph for the sake of clearness).

Performance of the AMC-SSR algorithm

In Fig. 4.8 and 4.9, the AGP improvements brought forth by the AMC-SSR algorithm (full marks) are quantified for both convolutional and turbo encoding, respectively, when adopting as performance benchmark the same AMC algorithm based on a conventional SL-PA strategy, referred to as AMC-SL for short (empty marks), recently proposed in [65]. Simulations are carried out averaging 10^4 independent channel realizations, for $N_{\text{tot}} = 1320$, $Q_U = 2$ UPUs and $Q_I = 3$ IPUs, whose positions are randomly determined at each packet transmission according to the scenario in Fig.

4.4. Two different interference thresholds are considered (for UPUs, it is equivalent to change the radius of the interference-free zone), that is to say $I_{\text{th}} = -100$ dBm (squares), i.e. the same as the channel noise level, and $I_{\text{th}} = -110$ dBm (circles). The dashed curve tagged Non Cognitive Scenario stands for the AMC-SSR algorithm without any PU constraint, i.e., with $I_{\text{th}} = +\infty$. As such, this AMC scheme gives the best attainable AGP performance since it allocates all the available power and coincides with the scheme recently proposed in [45] for a non-cognitive scenario.

Both figures show that the proposed AMC-SSR algorithm outperforms the non-adaptive AMC-SL, enabling a maximum relative gain on the AGP of around 25% and 47% for $I_{\text{th}} = -100$ dBm and $I_{\text{th}} = -110$ dBm, respectively. The behavior shown by the AMC-SSR and AMC-SL algorithms can be readily explained as follows. In the low E_s/N_0 regime, the SU link performance is limited by the channel frequency selectivity rather than the interference caused to the PUs, which can be promptly neglected. The AMC-SSR is very close to the non-cognitive case with $I_{\text{th}} = +\infty$, indeed, and upon cleverly distributing the power over the subcarriers with the highest channel gains, it outperforms the non adaptive AMC-SL scheme. In the medium-to-high E_s/N_0 regime, instead, when the available power increases, the main constraint to satisfy turns out to be the interference caused to the PUs. Here, the SL strategy makes the AGP curve flatten out at lower E_s/N_0 values as it inefficiently allocates power among the subcarriers to satisfy the interference constraints without any specific adaptivity. As result, the maximum interference level is prematurely met and no power increase is permitted. On the other side, the AMC-SSR algorithm properly shapes the PA distribution while accounting both for the current channel realization and the topology of the primary network, so that the link performance is significantly boosted. This observation is corroborated by the results obtained both in Fig. 4.8 and 4.9, when the interference threshold I_{th} decreases from -100 to -110 dBm. As a matter of fact, as the interference constraints get tighter, the AGP curves of the AMC-SL flattens out at lower and lower E_s/N_0 values compared with what we have with the proposed AMC-SSR algorithm.

4.9 Concluding remarks

In this chapter, we addressed the LRA problem, i.e., modulation and coding scheme along with power-per-subcarrier, for a cognitive BIC-OFDM system. The resulting

strategy is based upon the optimization of the GP metric under constraints on the total power available at the secondary transmitter and the maximum interference tolerable at the primary receivers. First, the PA optimization problem was analyzed and it was proven that is a convex OP whose solution can be found exploiting conventional numerical methods. However, to overcome some critical drawbacks of these numerical solutions, such as the high computational complexity, the PA problem was re-interpreted through the QVI framework, which yielded the novel iterative SSR algorithm and allowed to prove the existence and uniqueness of its solution. An optimum AMC procedure was also derived to further enhance the link performance. Simulation results corroborated the analytical derivation and showed that the proposed approach gives, practically, identical goodput performance to that offered by the conventional numerical methods, enables fast convergence making it extremely suitable for practical time-varying wireless scenarios and has better AGP performance over conventional non-adaptive PA algorithms.

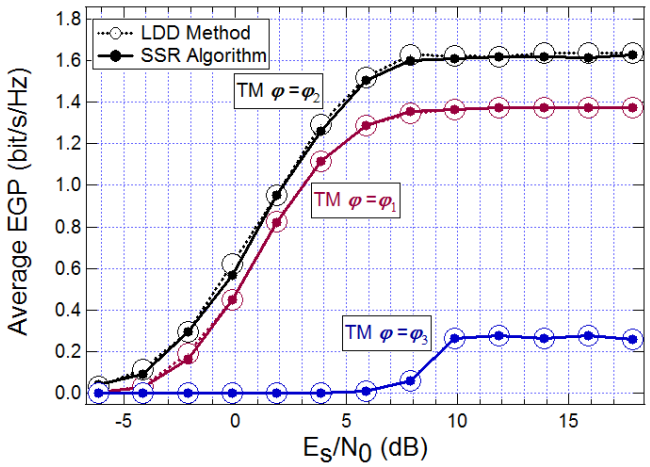

Figure 4.5: *LDD vs. SSR algorithm. Performance comparison.*

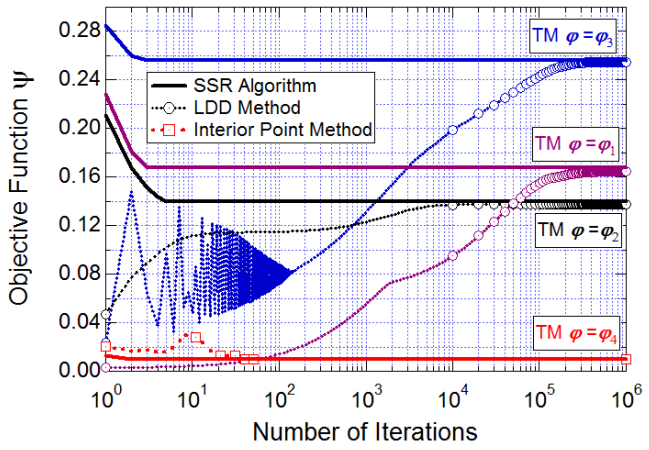

Figure 4.6: *LDD vs. SSR algorithm. Convergence comparison.*

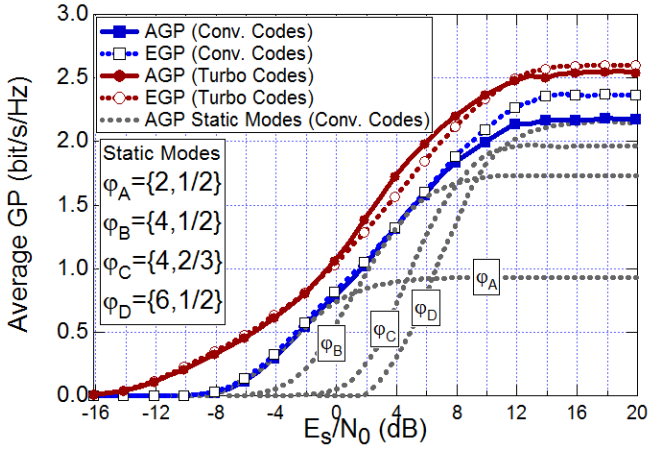

Figure 4.7: *Estimated vs. actual goodput comparison.*

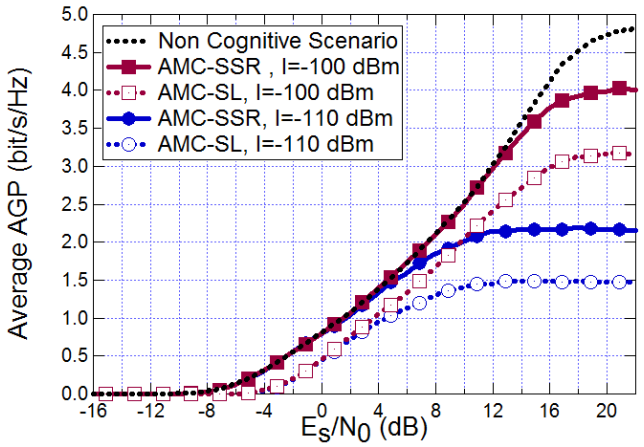

Figure 4.8: *AMC-SSR vs. AMC-SL algorithm. Performance comparison with convolutional codes.*

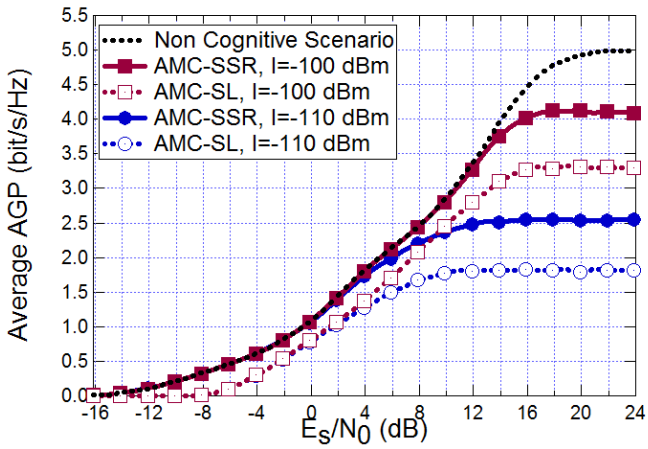

Figure 4.9: *AMC-SSR vs. AMC-SL algorithm. Performance comparison with turbo codes.*

Chapter 5

Hybrid ARQ based LRA for BIC-OFDM systems

In this Chapter, a novel link performance prediction approach, for BIC-OFDM systems employing HARQ mechanisms along with maximal-ratio combining at the receiver, is first proposed. Capitalizing on this novel prediction method, tagged as *aggregate* ESNR mapping technique, or αESM for short, a link adaptation strategy, that at each round of the HARQ protocol sets the transmission parameters to maximize the expected goodput, is then derived. Thanks to the αESM technique, the link adaptation strategy exploits the information offered by the HARQ feedback channel about the past failed transmissions along with information on the actual channel conditions, enabling an improved trade-off between data rate and link reliability compared to the case when only information on the channel status are accounted for.

5.1 Hybrid ARQ with packet combining

5.1.1 Background on HARQ

Next-generation wireless communication systems will implement packet-oriented data transmission [27], with very strict requirements on the data rate and reliability of the communication, according to the desired QoS. This implies that the receiver should be able to exploit all the known features of the communication system, such as interleaving, soft decisions, channel state information, etc., to increase the performance as much as possible. On the other hand, bandwidth is scarce and expensive in a heavily

booming cellular market. This means that the resource "redundancy" must be used with care, applying methods such as adaptive error protection [22].

To this end, a very interesting solution for reliable packet transmission is based on the HARQ concept, which relies on combining the error correction capability of the channel coding and error detection offered by, e.g., the CRC, with the ARQ mechanism [22]. In pure HARQ mechanism, basically, the transmitter sends a message to the receiver which, in turn, sends a 1 bit feedback ACK or NACK to tell the transmitter about a successful or unsuccessful, respectively, reception of the message. In the latter case, the transmitter sends again the packet by encoding it with either the same puncturing pattern or a different subset of redundancy bits. Finally, HARQ mechanism is best exploited when employed with *packet combining*. In fact, in this case, the receiver stores the previously unsuccessfully received copies of the packet and later combines them with the new received copies, creating a single packet that is more reliable than any of its constituent packets [22].

In literature, a considerable effort has been devoted to the analysis of performance limits in the presence of HARQ protocols, mainly focusing on traditional metrics, such as the ergodic capacity and the outage probability, [78]– [79]. For example, an information theoretic throughput analysis of HARQ schemes for the Gaussian collision channel was presented in [80], while in [81] the authors derive the optimal tradeoff among throughput, diversity gain, and delay of the multi-antenna block-fading ARQ channel, with a fixed discrete-input signal constellation. In [79], the long-term transmitted rate of HARQ based systems in a fast fading scenario is analyzed, when only channel statistics are available at the transmitter, and it is also presented a rate adjustment strategy aimed at guaranteeing a certain target outage probability. Other significant examples of works addressing the issue of selecting the transmission parameters on the fly for HARQ protocols are [82]– [83] and the references therein.

5.1.2 HARQ based BIC-OFDM system

Let us consider the BIC-OFDM system described in Sect. 2.2 when Hybrid ARQ protocols with packet combining are employed. In this case the transmitter produces L randomly punctured versions of the message, i.e., the RLC-PDU, as follows. The generic ℓth copy is obtained randomly selecting $N_{\mathrm{c}}^{(\ell)}$ out of the N_{u}/r_0 coded binary symbols at the output of the linear binary code, so that $N_{\mathrm{c}}^{(\ell)} = N_{\mathrm{u}}/r^{(\ell)}$, being $r^{(\ell)}$

the desired coding rate at round ℓ and r_0 the mother code. The ℓth copy is then sent over the block fading channel to the receiver. The latter decodes the received copy combining the soft-metrics, represented by the log-likelihood ratios of the binary coded symbols, of all the ℓ copies received so far in a maximum ratio combining fashion. The equivalence between maximum likelihood decoding and maximum ratio combining was shown for the Type-I HARQ scheme, or Chase Combining [84]. As a matter of fact, the L copies can be seen as L randomly punctured versions of the same message, with soft-metrics associated to the punctured bits equal to 0. If the packet is decoded with errors (detected through the CRC), a NACK is sent on the 1 bit feedback channel, assumed to be error free and delay free, to the transmitter, which in turn is called to transmit the $(\ell+1)$th copy of the packet. This procedure goes on until the transmitter receives an ACK or the maximum number of transmissions L is reached. Whenever one of these events occurs, the message is removed from the buffer and the transmitter moves on sending a new packet. In particular, when all the L copies of the packets are unsuccessfully decoded, that is, at round L a NACK is fed back, the system occurs in an *outage event*.

5.2 The αESM model

In this section, we extend the procedure presented in Sect. 2.2.1 for estimating the pairwise error probability of the maximum likelihood metric decoder to the HARQ context by introducing the concept of aggregate effective SNR mapping αESM. This approach is built around the concept of decoding score, a RV whose positive tail probability yields the pairwise error probability [36], and the equivalent BIOS model of the BIC system. Thus, first of all, let us recall the equivalent BIOS model, depicted in Fig. 5.1, for the BIC-OFDM system described in Sect. 2.2. Here, at each HARQ round ℓ, $1 \leq \ell \leq L$, the BIC-OFDM channel is modeled as a set of

$$B^{(\ell)} \triangleq \sum_{n=1}^{N} m_n^{(\ell)} \tag{5.1}$$

parallel binary input output symmetric channels. The pairwise error probability of the BIC-OFDM system can be expressed as a function of the underlying binary code, so that, if $\mathbf{b}^{(\ell)} \triangleq \{b_1^{(\ell)}, \cdots, b_{N_c^{(\ell)}}^{(\ell)}\}$ is the reference codeword (corresponding to the transmitted RLC-PDU at the ℓth HARQ round) and $\mathbf{b}^{(\ell)\prime} \triangleq \{b_1^{(\ell)\prime}, \cdots, b_{N_c^{(\ell)}}^{(\ell)\prime}\}$ the

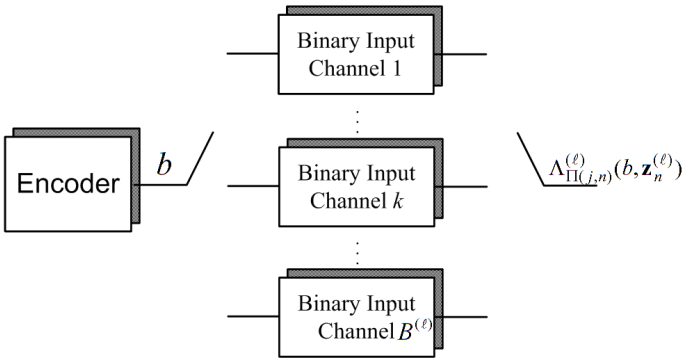

Figure 5.1: *Equivalent BIOS model.*

competing codeword, the PEP results

$$\mathrm{PEP}\left(\mathbf{b}^{(\ell)},\mathbf{b}^{(\ell)\prime}\right) \triangleq \Pr\left\{\phi(\mathbf{b}^{(\ell)\prime},\mathbf{z}^{(\ell)}) > \phi(\mathbf{b}^{(\ell)},\mathbf{z}^{(\ell)})\right\}, \tag{5.2}$$

where $\phi(\cdot)$ is the decoding metric[1]. Relying on the equivalent BIOS model, it can be noted that:

- the input to the kth channel, $1 \leq k \leq B^{(\ell)}$, is the bit $b_j(x_n^{(\ell)})$, that is, the bit in the jth position of the label of the QAM symbol $x_n^{(\ell)}$ sent on subcarrier n, so that we can write the decoder metric as

$$\phi(\mathbf{b}^{(\ell)},\mathbf{z}^{(\ell)}) = \prod_{n=1}^{N}\prod_{j=1}^{m_n^{(\ell)}} \phi_j(b_j(x_n^{(\ell)}), z_n^{(\ell)}), \tag{5.3}$$

  where $\phi_j(b_j(x_n^{(\ell)}), z_n^{(\ell)})$ is the decoding metric associated to bit $b_j(x_n^{(\ell)})$;

- the output is the bit-loglikelihood metric, also named bit score, here rewritten, for the ease of notation, as

$$\Lambda_k^{(\ell)} = \log\frac{\phi_j(b^{(\ell)\prime}_{\Pi(n,j)}, z_n^{(\ell)})}{\phi_j(b^{(\ell)}_{\Pi(n,j)}, z_n^{(\ell)})}, \tag{5.4}$$

[1] In order to avoid a cumbersome notation, we denote with (5.2) the general expression of the PEP at PR ℓ, without any specific assumption on the recombination of the previous copies of message. The analysis developed in the following will clarify how to evaluate the quantities involved according to the chosen decoding strategy.

where $\Pi(n,j)=k$ is the interleaving function that maps the k CBS on the jth position of the symbol sent over the nth subcarrier.

Thus, the pairwise decoding score (PDS) relevant to the transmitted codeword is defined as

$$\Lambda_{\mathrm{pw}}^{(\ell)}=\sum_{n=1}^{N}\sum_{i=1}^{m_n^{(\ell)}}\Lambda_{\Pi(n,i)}^{(\ell)}. \tag{5.5}$$

Clearly, only the bit indices in which the codewords $\mathbf{b}^{(\ell)}$ and $\mathbf{b}^{(\ell)'}$ differ have a non-zero bit score.

Arguably, the optimal receiver, that extends the decoding to account for the combination of all the received copies, should perform a joint decoding of the pairwise decoding scores over all the possible L transmissions but would results in an exponentially increasing complexity with L [85]. Thus, an efficient complexity-performance tradeoff is given by the *bit-level combining receiver* [86]: at the ℓth PR, $1\leq\ell\leq L$, the bit scores $\Lambda_k^{(\ell)}$ are evaluated as in (2.7) and, for each bit k, are added together so that the *aggregate bit score* results

$$\mathcal{L}_k^{(\ell)}=\mathbf{q}_k^{(\ell)\mathrm{T}}\mathbf{\Lambda}_k^{(\ell)}, \tag{5.6}$$

where $\mathbf{\Lambda}_k^{(\ell)}\triangleq\left[\Lambda_k^{(1)},\cdots,\Lambda_k^{(\ell)}\right]^{\mathrm{T}}$ is the vector collecting the bit scores of the CBS k up to round ℓ and $\mathbf{q}_k^{(\ell)}\triangleq\left[q_k^{(1)},\cdots,q_k^{(\ell)}\right]^{\mathrm{T}}\in\{0,1\}^{\ell}$ is the corresponding vector of puncturing variables, that is, $q_k^{(j)}=1$ if bit k has been transmitted at round j, 0 otherwise (i.e., it has been punctured). Hence, the aggregate PDS at round ℓ results

$$\mathcal{L}_{\mathrm{pw}}^{(\ell)}=\sum_{n=1}^{N}\sum_{i=1}^{m_n^{(\ell)}}\mathcal{L}_{\Pi(n,i)}^{(\ell)}, \tag{5.7}$$

whereas the pairwise error probability (5.2), related to the PDS as it can be seen substituting after some algebra (5.6) into (5.3), results

$$\mathrm{PEP}(\mathbf{b}^{(\ell)},\mathbf{b}^{(\ell)'})=\Pr\left(\mathcal{L}_{\mathrm{pw}}^{(\ell)}>0\right), \tag{5.8}$$

i.e., the PEP can be evaluated as the tail probability of the RV $\mathcal{L}_{\mathrm{pw}}^{(\ell)}$. Thus, as it will be clear soon, let us introduce the definition of the CMGF of the bit score

$$\kappa_{\mathcal{L}_k^{(\ell)}}(s)\triangleq\log\left(\mathrm{E}\left\{e^{s\mathcal{L}_k^{(\ell)}}\right\}\right) \tag{5.9}$$

Recalling the features of the HARQ system under analysis, we can assume that, since at each round a random subset of the CBS is selected, the pattern $\mathbf{q}_k^{(\ell)}$ can be modeled as a sequence of ℓ independent and identically distributed (i.i.d.) binary RVs taking values 0 or 1 independently of the bit index k, and as a consequence, it will be designated as $\mathbf{q}^{(\ell)} = [q^{(1)}, \cdots, q^{(\ell)}]^{\mathrm{T}}$.
Accordingly, exploiting the law of total probability and eqn. (5.6), expression (5.9) yields

$$\kappa_{\mathcal{L}_k^{(\ell)}}(s) = \log\left(\sum_{\bar{\mathbf{q}}^{(\ell)} \in \mathcal{Q}^{(\ell)}} \Pr\left(\mathbf{q}^{(\ell)} = \bar{\mathbf{q}}^{(\ell)}\right) \prod_{j=1}^{\ell} \left[\mathrm{E}\left\{e^{s\Lambda_k^{(j)}}\right\}\right]^{\bar{q}^{(j)}}\right), \tag{5.10}$$

where $\mathcal{Q}^{(\ell)}$ is the set of the possible puncturing patterns $\mathbf{q}^{(\ell)}$ over the first ℓ PRs. Finally, recalling that $\kappa_\Lambda(s) \triangleq \log\left(\mathrm{E}_\Lambda\left\{e^{s\Lambda}\right\}\right)$, (5.9) turns into

$$\kappa_{\mathcal{L}_k^{(\ell)}}(s) = \log\left(\sum_{\bar{\mathbf{q}}^{(\ell)} \in \mathcal{Q}^{(\ell)}} \Pr\left(\mathbf{q}^{(\ell)} = \bar{\mathbf{q}}^{(\ell)}\right) \prod_{j=1}^{\ell} \left[e^{\kappa_{\Lambda_k^{(j)}}(s)}\right]^{\bar{q}^{(j)}}\right). \tag{5.11}$$

Now, we can study the pairwise error probability for infinite interleaving. Following the line of reasoning presented in [36], in case of infinite-length interleaving and linear binary code, the bit score $\Lambda_k^{(\ell)}$ are i.i.d RVs over the index k and independent of $q^{(\ell)}$ and the cumulant transform of the aggregate PDS $\kappa_{\mathrm{pw}}^{(\ell)}(s) \triangleq \log\left(\mathrm{E}\left\{e^{s\mathcal{L}_{\mathrm{pw}}^{(\ell)}}\right\}\right)$ can be simply expressed as

$$\kappa_{\mathrm{pw}}^{(\ell)}(s) = d\kappa_{\mathcal{L}_k^{(\ell)}}(s) \tag{5.12}$$

where d represents the Hamming distance between the reference and the competing codewords.
Therefore, the tail probability (5.8) of two codewords differing in d positions can be computed, exploiting relationship (5.12), as

$$\mathrm{PEP}(d) = \frac{1}{2\pi\jmath}\int_{\sigma-\jmath\infty}^{\sigma+\jmath\infty} \frac{1}{s} e^{d\kappa_{\mathcal{L}_k^{(\ell)}}(s)} \mathrm{d}s. \tag{5.13}$$

In general, obtaining exact expressions for the PEP integral (5.13) are extremely difficult and numerical methods are required. However, similarly to what is proposed in [34], we will resort to the so-called Gaussian approximation of the PEP integral, yielding

$$\mathrm{PEP}(d) \simeq Q\left(\sqrt{-2d\kappa_{\mathcal{L}_k^{(\ell)}}(\hat{s})}\right), \tag{5.14}$$

where $\hat{s}$ represents the saddle point, i.e., the value of s that makes the first derivative of the cumulant transform, $\kappa_{\mathrm{pw}}^{(\ell)'}(s)$, equal to zero.
It is worth noting that eqn. (5.14), corresponding to the 0th order Lugannani-Rice expansion of the integral in (5.13), provides the error probability of a BPSK system over AWGN channel with SNR equal to $-\kappa_{\mathcal{L}_k^{(\ell)}}(\hat{s})$. Recalling eqn. (5.11) we can eventually define the *aggregate effective SNR*, or αESNR, as

$$\Gamma_{\alpha}^{(\ell)} = -\log\left(\sum_{\bar{\mathbf{q}}^{(\ell)}\in\mathcal{Q}^{(\ell)}} \Pr\left(\mathbf{q}^{(\ell)}=\bar{\mathbf{q}}^{(\ell)}\right)\prod_{j=1}^{\ell}\left[e^{-\hat{\gamma}^{(j)}}\right]^{\bar{q}^{(j)}}\right), \tag{5.15}$$

where

$$\hat{\gamma}^{(j)} \triangleq -\kappa_{\Lambda_k^{(j)}(\hat{s})}, \quad 1\le j\le \ell \tag{5.16}$$

is the effective SNR relevant to the single jth HARQ round, described in Sect. 2.3.1 .

Exploiting the above ESM concept, expression (5.15) can be further rearranged, leading to the main results of this section as stated in the following.

Theorem 5.1 *Assuming that the coding rate $r^{(\ell)}$ varies with the PR index ℓ, $1\le\ell\le L$, the αESM $\Gamma_{\alpha}^{(\ell)}$ can be lower-bounded as*

$$\Gamma_{\alpha}^{(\ell)} \ge g(\Gamma_{\alpha}^{(\ell-1)},\xi^{(\ell)}) + f(\hat{\gamma}^{(\ell)},\xi^{(\ell)}), \tag{5.17}$$

with $\Gamma_{\alpha}^{(1)}=\hat{\gamma}^{(1)}$ and $R^{(1)}=r^{(1)}$, where

$$\xi^{(\ell)} \triangleq \frac{r^{(\ell)}}{R^{(\ell-1)}}, \quad R^{(\ell)} \triangleq \min\{R^{(\ell-1)},r^{(\ell)}\} \tag{5.18}$$

$$\begin{aligned} g(x,a) &= \begin{cases} -\log[1+a(e^{-x}-1)], & r^{(\ell)}\le R^{(\ell-1)} \\ x, & r^{(\ell)}> R^{(\ell-1)} \end{cases}, \\ f(x,a) &= \begin{cases} x, & r^{(\ell)}\le R^{(\ell-1)} \\ -\log[1+\frac{1}{a}(e^{-x}-1)], & r^{(\ell)}> R^{(\ell-1)} \end{cases}. \end{aligned} \tag{5.19}$$

Proof In order to prove Theorem 5.1, two different cases are taken into account. Let us start with the case in which the coding rate is monotonically increasing up to the ℓth PR, i.e., $r^{(j)}>r^{(j-1)}$, $1\le j\le\ell$. Denoting as $N_{\mathrm{c}}^{(j)}=N_{\mathrm{u}}/r^{(j)}$, with $N_{\mathrm{c}}^{(j)}\le N_{\mathrm{c}}^{(j-1)}$, the number of coded bits transmitted at the jth PR, $1\le j\le\ell$, the

set $\mathcal{Q}^{(\ell)}$ containing all the possible puncturing patterns among the ℓth PRs can be written as

$$\mathcal{Q}^{(\ell)} \triangleq \{[\underbrace{\mathbf{1}_{\ell}^{T}, \mathbf{0}_{0}^{T}}_{\mathbf{q}_0}]^{T}, [\underbrace{\mathbf{1}_{\ell-1}^{T}, \mathbf{0}_{1}^{T}}_{\mathbf{q}_1}]^{T}, \cdots, \{[\underbrace{\mathbf{1}_{1}^{T}, \mathbf{0}_{\ell-1}^{T}}_{\mathbf{q}_{\ell-1}}]^{T}\}, \tag{5.20}$$

in that a given coded bit can be transmitted at PRs $1, 2, \cdots, \ell$ (pattern $\mathbf{q}_0$), or at PRs $1, 2, \cdots, \ell-1$ (pattern $\mathbf{q}_1$), and so on, or only at PR 1 (pattern $\mathbf{q}_{\ell-1}$). Hence, defining $p^{(j)} \triangleq \Pr\{\mathbf{q}_k^{(\ell)} = \mathbf{q}_j\}$ as the probability that the kth coded bit is punctured at the ℓth PR using the pattern $\mathbf{q}_j$, $0 \leq j \leq \ell-1$, it can be verified that the set of these probabilities can be computed as

$$\mathcal{P}^{(\ell)} \triangleq \left\{ \frac{N_{\mathrm{c}}^{(\ell)}}{N_{\mathrm{c}}^{(1)}}, \frac{N_{\mathrm{c}}^{(\ell-1)} - N_{\mathrm{c}}^{(\ell)}}{N_{\mathrm{c}}^{(1)}}, \cdots, \frac{N_{\mathrm{c}}^{(1)} - N_{\mathrm{c}}^{(2)}}{N_{\mathrm{c}}^{(1)}} \right\}. \tag{5.21}$$

Now, let us prove (5.17) of Theorem 5.1 by induction. It can easily verified that the expression holds for $\ell = 1, 2, \cdots$. Therefore, at the $(\ell+1)$th PR we can write

$$\Gamma_{\alpha}^{(\ell+1)} = -\log\left\{ \frac{N_{\mathrm{c}}^{(\ell+1)}}{N_{\mathrm{c}}^{(1)}} \mathrm{e}^{-\sum_{j=1}^{\ell+1}\hat{\gamma}^{(j)}} + \sum_{k=2}^{\ell+1} \left(\frac{N_{\mathrm{c}}^{(k-1)} - N_{\mathrm{c}}^{(k)}}{N_{\mathrm{c}}^{(1)}} \right) \mathrm{e}^{-\sum_{j=1}^{k-1}\hat{\gamma}^{(j)}} \right\}, \tag{5.22}$$

that after some algebra can be rearranged as

$$\begin{aligned} \Gamma_{\alpha}^{(\ell+1)} = -\log\Bigg\{ & \tfrac{N_{\mathrm{c}}^{(\ell+1)}}{N_{\mathrm{c}}^{(1)}} \mathrm{e}^{-\sum_{j=1}^{\ell}\hat{\gamma}^{(j)}} \left(e^{-\hat{\gamma}^{(\ell+1)}} - 1 \right) + \\ & \tfrac{N_{\mathrm{c}}^{(\ell)}}{N_{\mathrm{c}}^{(1)}} \mathrm{e}^{-\sum_{j=1}^{\ell}\hat{\gamma}^{(j)}} + \sum_{k=2}^{\ell} \left(\tfrac{N_{\mathrm{c}}^{(k-1)} - N_{\mathrm{c}}^{(k)}}{N_{\mathrm{c}}^{(1)}} \right) e^{-\sum_{j=1}^{k-1}\hat{\gamma}^{(j)}} \Bigg\}. \end{aligned} \tag{5.23}$$

Then, considering that the last two terms in the curly brackets of (5.23) correspond to $\mathrm{e}^{-\Gamma_{\alpha}^{(\ell)}}$, and $\dfrac{\mathrm{e}^{-\sum_{j=1}^{\ell}\hat{\gamma}^{(j)}}}{\mathrm{e}^{-\Gamma_{\alpha}^{(\ell)}}} \leq 1$ as the coding rate is increasing, we end up to

$$\Gamma_{\alpha}^{(\ell+1)} \geq \Gamma_{\alpha}^{(\ell)} - \log\left[1 + \frac{R^{(\ell)}}{r^{(\ell+1)}} \left(\mathrm{e}^{-\hat{\gamma}^{(\ell+1)}} - 1 \right) \right], \tag{5.24}$$

where we exploit that $R^{(\ell)} = r^{(1)}$ due to (5.18) and the assumption of increasing coding rate.

In the case the coding rate is not increasing up to the ℓth PR, i.e., $r^{(j)} \leq r^{(j-1)}$, $1 \leq j \leq \ell$, the set of puncturing patterns at the ℓth PRs turns into

$$\mathcal{Q}^{(\ell)} \triangleq \{[\underbrace{\mathbf{0}_{0}^{T}, \mathbf{1}_{\ell}^{T}}_{\mathbf{q}_0}]^{T}, [\underbrace{\mathbf{0}_{1}^{T}, \mathbf{1}_{\ell-1}^{T}}_{\mathbf{q}_1}]^{T}, \cdots, \{[\underbrace{\mathbf{0}_{\ell-1}^{T}, \mathbf{1}_{1}^{T}}_{\mathbf{q}_{\ell-1}}]^{T}\}, \tag{5.25}$$

with probabilities

$$\mathcal{P}^{(\ell)} \triangleq \left\{ \frac{N_{\mathrm{c}}^{(1)}}{N_{\mathrm{c}}^{(\ell)}}, \frac{N_{\mathrm{c}}^{(2)} - N_{\mathrm{c}}^{(1)}}{N_{\mathrm{c}}^{(\ell)}}, \cdots, \frac{N_{\mathrm{c}}^{(\ell)} - N_{\mathrm{c}}^{(\ell-1)}}{N_{\mathrm{c}}^{(\ell)}} \right\}. \tag{5.26}$$

Therefore, following the same line of reasoning adopted above, at the $(\ell+1)$th PR we can write

$$\Gamma_{\alpha}^{(\ell+1)} = -\log \left\{ \frac{N_{\mathrm{c}}^{(1)}}{N_{\mathrm{c}}^{(\ell+1)}} \mathrm{e}^{-\sum_{j=1}^{\ell+1} \hat{\gamma}^{(j)}} + \sum_{k=2}^{\ell+1} \left(\frac{N_{\mathrm{c}}^{(k)} - N_{\mathrm{c}}^{(k-1)}}{N_{\mathrm{c}}^{(\ell+1)}} \right) \mathrm{e}^{-\sum_{j=k}^{\ell+1} \hat{\gamma}^{(j)}} \right\}, \tag{5.27}$$

that after some algebra can be rearranged as

$$\begin{aligned} \Gamma_{\alpha}^{(\ell+1)} = \hat{\gamma}^{(\ell+1)} - \log \Bigg\{ \tfrac{N_{\mathrm{c}}^{(\ell)}}{N_{\mathrm{c}}^{(\ell+1)}} \Big[\tfrac{N_{\mathrm{c}}^{(\ell+1)} - N_{\mathrm{c}}^{(\ell)}}{N_{\mathrm{c}}^{(\ell)}} + \\ \tfrac{N_{\mathrm{c}}^{(1)}}{N_{\mathrm{c}}^{(\ell)}} \mathrm{e}^{-\sum_{j=1}^{\ell} \hat{\gamma}^{(j)}} + \sum_{k=2}^{\ell} \left(\tfrac{N_{\mathrm{c}}^{(k-1)} - N_{\mathrm{c}}^{(k)}}{N_{\mathrm{c}}^{(\ell)}} \right) e^{-\sum_{j=k}^{l} \hat{\gamma}^{(j)}} \Big] \Bigg\}. \end{aligned} \tag{5.28}$$

Then, considering that the last two terms in the curly brackets of (5.28) correspond to $\mathrm{e}^{-\Gamma_{\alpha}^{(\ell)}}$, we end up to

$$\Gamma_{\alpha}^{(\ell+1)} = \hat{\gamma}^{(\ell+1)} - \log \left[1 + \frac{r^{(\ell+1)}}{R^{(\ell)}} \left(\mathrm{e}^{-\Gamma_{\alpha}^{(\ell)}} - 1 \right) \right], \tag{5.29}$$

where $R^{(\ell)} = r^{(\ell)}$ due to the assumption of decreasing coding rate. □

A few remarks about Theorem 5.1 are now discussed.

1. It can be shown that the lower-bound (5.17) is exactly met when $r^{(\ell)} \leq R^{(\ell-1)}$, i.e., in the case the coding rate is decreased along with retransmissions.

2. The updating of $\Gamma_{\alpha}^{(\ell)}$ through (5.17) requires only the knowledge of two aggregate quantities $\Gamma_{\alpha}^{(\ell-1)}$ and $R^{(\ell-1)}$ related to the previous $(\ell-1)$th step, together with the ESNR $\hat{\gamma}^{(\ell)}$, which is evaluated at the current ℓth PR according to (2.19)-(2.20), i.e., based on the SNRs $\boldsymbol{\Upsilon}^{(\ell)}$ and the TM $\boldsymbol{\varphi}^{(\ell)} \triangleq \{\mathbf{m}^{(\ell)}, r^{(\ell)}\}$. In particular, $\boldsymbol{\varphi}^{(\ell)}$ includes the bit distribution across the subcarriers $\mathbf{m}^{(\ell)} \triangleq [m_1^{(\ell)}, \cdots, m_N^{(\ell)}]^{\mathrm{T}} \in \mathcal{D}_{\mathbf{m}}$, $\mathcal{D}_{\mathbf{m}}$ being the set containing all the possible bit loading realizations, with $m_n^{(\ell)} \in \mathcal{D}_m \triangleq \{0, 2, 4, \cdots, m_{\max}\}$, $\forall n$, and the coding rate $r^{(\ell)} \in \mathcal{D}_r \triangleq \{r_0, r_1, \cdots, r_{\max}\}$, so that $\boldsymbol{\varphi}^{(\ell)} \in \mathcal{D}_{\boldsymbol{\varphi}}$, with $\mathcal{D}_{\boldsymbol{\varphi}} = \mathcal{D}_{\mathbf{m}} \times \mathcal{D}_r$ being the set of the allowable TMs. Therefore, let us define the "state" of the HARQ

scheme (or in other words, the memory of the HARQ) $\boldsymbol{\sigma}^{(\ell)} \triangleq \{\Gamma_{\alpha}^{(\ell)}, R^{(\ell)}\}$. This is the reason why the aggregate ESNR at the ℓth PR will depend only on the state $\boldsymbol{\sigma}^{(\ell-1)}$ (related to the past retransmissions up to that of index $\ell-1$) and the TM $\boldsymbol{\varphi}^{(\ell)}$. Accordingly, it will be written as $\Gamma_{\alpha}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})$. The αESM update is depicted in Fig. 5.2.

3. Under the assumption that the coding rate is not adapted, i.e., $r^{(j)} = r^{(j-1)}$, $2 \leq j \leq \ell$, it can be shown that $\Gamma_{\alpha}^{(\ell)} = \sum_{j=0}^{\ell} \hat{\gamma}^{(j)}$. As expected, in view of the HARQ mechanism, the aggregate ESNR is obtained by the accumulation of the ESNRs evaluated at each PR.

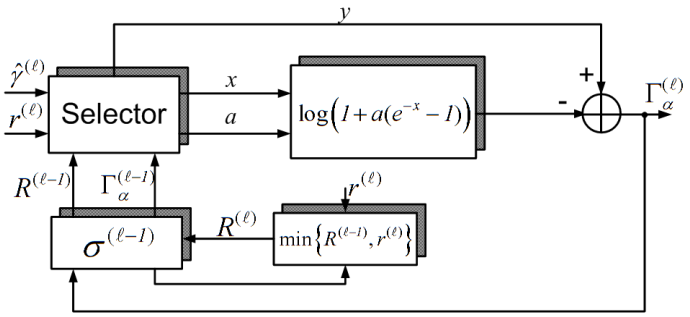

Figure 5.2: *Block diagram of the αESM update.*

5.3 Expected goodput in presence of HARQ mechanism

5.3.1 Expected goodput formulation

As outlined in Chapter 3, we will pursue an inter-round optimization approach, where the EGP metric is optimized over the tunable transmission parameters at each PR. In Sect. 3.2.3, the EGP was derived for the ARQ based system described in Sect. 2.2. Here, the EGP metric is suitably modified to account for the effect of the HARQ mechanism. Let us in fact consider the generic ℓth branch of the equivalent model depicted in Fig. 5.3. Due to the packet-combining mechanism, the PER at the generic round ℓ depends now also on the previously transmitted packets, that is

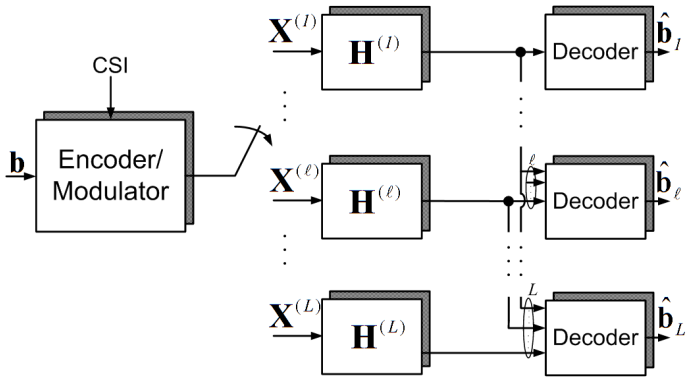

Figure 5.3: *Equivalent model for the HARQ BIC-OFDM system..*

$\mathrm{PER}(\boldsymbol{\varphi}^{(\ell)}|\mathbf{H}^{(\ell)})$ in eqn. (3.15) turns in $\mathrm{PER}(\boldsymbol{\varphi}^{(\ell)}|\{\boldsymbol{\varphi}^{(i)}\}_{i=1}^{\ell-1},\{\mathbf{H}^{(i)}\}_{i=1}^{\ell})$. Thanks to the αESM model derived in the previous section, which is capable of taking into account the effect of the HARQ mechanism, the PER at PR ℓ for the HARQ-based BIC-OFDM system can be evaluated as

$$\mathrm{PER}(\boldsymbol{\varphi}^{(\ell)}|\{\boldsymbol{\varphi}^{(i)}\}_{i=1}^{\ell-1},\{\mathbf{H}^{(i)}\}_{i=1}^{\ell}) = \Psi_{r^{(\ell)}}\left(\Gamma_{\alpha}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})|\mathbf{H}^{(\ell)}\right), \tag{5.30}$$

where the right hand side of (5.30) is the PER of the equivalent BPSK system over AWGN channel that experiences an SNR equal to $\Gamma_{\alpha}^{(\ell)}$.
Capitalizing on equation (5.30), the EGP derivation is then analogous to the one done in Sect. 3.2.3, so that it is here briefly recalled.
Resorting to the long-term static channel assumption, the expectation over the future channel realization $\mathrm{E}_{\mathbf{H}^{(\ell+j)}}\left\{\Psi_{r^{(\ell+j)}}\left(\Gamma_{\alpha}^{(\ell+j)}(\boldsymbol{\varphi}^{(\ell+j)}|\boldsymbol{\sigma}^{(\ell+j-1)})|\mathbf{H}^{(\ell+j)}\right)\right\}$ is substituted with $\Psi_{r^{(\ell)}}\left(\Gamma_{\alpha}^{(\ell+j)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})|\mathbf{H}^{(\ell)}\right)$, that represents the PER we would have if the packet experiences the current channel conditions $\mathbf{H}^{(\ell)}$ throughout its possible future retransmissions.
Thus, based on the above observations, the outage probability (3.12) is now expressed by

$$P_{\mathrm{out}}(\boldsymbol{\varphi}^{(\ell)},\Gamma_{\alpha}^{(\ell)},\cdots,\Gamma_{\alpha}^{(L)}) = \prod_{j=0}^{L-\ell}\Psi_{r^{(\ell)}}\left(\Gamma_{\alpha}{}^{(\ell+j)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})|\mathbf{H}^{(\ell)}\right), \tag{5.31}$$

whereas the expected delivering time (3.16) results

$$T_D(\boldsymbol{\varphi}^{(\ell)},\Gamma_\alpha^{(\ell)},\cdots,\Gamma_\alpha^{(L-1)}) = T_u(\boldsymbol{\varphi}^{(\ell)})\sum_{j=\ell}^{L}\prod_{k=0}^{j-\ell}\Psi_{r^{(\ell)}}\left(\Gamma_\alpha^{(\ell+k-1)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})|\mathbf{H}^{(\ell)}\right), \tag{5.32}$$

with $\Psi_{r^{(\ell)}}\left(\Gamma_\alpha^{(\ell-1)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})|\mathbf{H}^{(\ell)}\right) \triangleq 1$.
Furthermore, approximating the ESNR $\Gamma^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})$ with its lower bound of Theorem 5.1, we have

$$\Gamma_\alpha^{(\ell+j)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)}) = g(\Gamma_\alpha^{(\ell-1)},\xi^{(\ell)}) + (j+1)f[\hat{\gamma}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}),\xi^{(\ell)}] \quad \forall j \geq 0. \tag{5.33}$$

We can now get the following results.

Prop. 5.1 *Replacing (5.33) into (5.32) and (5.31), it turns out that*

1. $T_D(\boldsymbol{\varphi}^{(\ell)},\Gamma_\alpha^{(\ell)},\cdots,\Gamma_\alpha^{(L-1)})$ *in (5.32) can be written as*

$$T_D(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)}) = T_u(\boldsymbol{\varphi}^{(\ell)})\lambda(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)}),$$

where

$$\lambda(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)}) = \sum_{j=\ell}^{L}\prod_{k=0}^{j-\ell}\left\{\Psi_{r^{(\ell)}}(g(\Gamma_\alpha^{(\ell-1)},\xi^{(\ell)}) + kf[\hat{\gamma}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}),\xi^{(\ell)}])\right\}, \tag{5.34}$$

with $\Psi_{r^{(\ell)}}(g(\Gamma_\alpha^{(\ell-1)},\xi^{(\ell)})) \triangleq 1$, *is a monotonically decreasing function of the ESNR* $\hat{\gamma}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)})$*;*

2. $P_{\text{out}}(\boldsymbol{\varphi}^{(\ell)},\Gamma_\alpha^{(\ell)},\cdots,\Gamma_\alpha^{(L)})$ *can be written as*

$$P_{\text{out}}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)}) = \prod_{j=0}^{L-\ell}\Psi_{r^{(\ell)}}\left(g(\Gamma_\alpha^{(\ell-1)},\xi^{(\ell)}) + (j+1)f[\hat{\gamma}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}),\xi^{(\ell)}]\right), \tag{5.35}$$

which is a monotonically decreasing function of the ESNR $\hat{\gamma}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)})$.

Proof In order to show (5.34) and (5.35), it is sufficient to demonstrate that the product of the n PER functions Ψ_r , with $n = j-\ell$ in (5.34), since $\Psi_r \triangleq 1$ for $k=0$, and $n = L-\ell+1$ in (5.35), monotonically decreases with $\hat{\gamma}$. To this end, consider w.l.g. eqn. (5.35) and the case $r^{(\ell)} \leq R^{(\ell)}$. We get that $g(\Gamma_\alpha^{(\ell-1)},\xi^{(\ell)}) = c_\Delta$ is constant and $f(\hat{\gamma}^{(\ell)},\xi^{(\ell)}) = \hat{\gamma}^{(\ell)}$. Thus, recalling the PER $\Psi(\cdot)$ properties in Sect. 2.3.1 and upon

defining $\Psi_{\text{tot}}(\hat{\gamma}^{(\ell)}) \triangleq \prod_{k=0}^{n-1} \Psi_{r^{(\ell)}}\left(c_\Delta + i_k \hat{\gamma}^{(\ell)}\right)$, with $i_k = k$ and $i_k = k+1$ for (5.34) and (5.35), respectively, we get that

$$\frac{\partial}{\partial \hat{\gamma}^{(\ell)}} \Psi_{\text{tot}}(\hat{\gamma}^{(\ell)}) = \Psi_{\text{tot}}(\hat{\gamma}^{(\ell)}) \sum_{j=0}^{n-1} i_j \frac{\Psi'_{r^{(\ell)}}\left(c_\Delta + i_j \hat{\gamma}^{(\ell)}\right)}{\Psi_{r^{(\ell)}}\left(c_\Delta + i_j \hat{\gamma}^{(\ell)}\right)} < 0, \tag{5.36}$$

being $0 \leq \Psi_r \leq 1$ and $\Psi_r^{'} < 0$. Thus, since the derivative of Ψ_{tot} is negative, it follows that Ψ_{tot} is a monotonically decreasing function of $\hat{\gamma}^{(\ell)}$ and, accordingly, equations (5.34) and (5.35) too. □

Collecting these results, the EGP at the ℓth round is finally obtained as

$$\zeta^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)}) = \frac{N_{\text{p}}}{B} \frac{1 - P_{\text{out}}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})}{T_{\text{f}}(\{\boldsymbol{\varphi}^{(i)}\}_{i=1}^{\ell-1}) + T_u(\boldsymbol{\varphi}^{(\ell)}) \lambda(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})}. \tag{5.37}$$

In the following sections, two LRA algorithms will be proposed to maximize the above EGP metric in presence of HARQ mechanism. These two approaches correspond to two different levels of CSI knowledge at the transmitter (CSIT). The first scenario assumes that instantaneous CSI is perfectly known at the transmitter. In this case, the algorithm is able to assign a different modulation over each subcarrier. Since in most practical scenarios the entire instantaneous channel vector can be hardly available, the second scenario, tagged as limited CSIT, only assumes the knowledge of the channel statistics and, accordingly, and AMC algorithm is derived.

5.4 LRA algorithm with instantaneous CSIT

In this section, the LRA problem for the EGP metric (5.37) maximization at the generic PR ℓ is formulated, under the condition that the transmitter has knowledge of the entire SNR matrix $\boldsymbol{\Upsilon}^{(\ell)}$, as defined in (2.5). Accordingly, the EGP optimization is done selecting the best TM $\boldsymbol{\varphi}_{\text{o}}^{(\ell)}$, i.e., the best pair bit loading (BL) distribution $\mathbf{m}_{\text{o}}^{(\ell)}$ and coding rate $r_{\text{o}}^{(\ell)}$, belonging to the finite set $\mathcal{D}_{\boldsymbol{\varphi}}$. The LRA OP, tagged as LRA-OP1, can be formally stated as follows.

$$\mathbf{LRA-OP1}: \quad \begin{array}{l} \boldsymbol{\varphi}_{\text{o}}^{(\ell)} = \arg\max\limits_{\boldsymbol{\varphi}} \left\{\zeta^{(\ell)}(\boldsymbol{\varphi}|\boldsymbol{\sigma}^{(\ell-1)})\right\} \\ \text{s.t.} \quad \boldsymbol{\varphi} \in \mathcal{D}_{\boldsymbol{\varphi}} \end{array}. \tag{5.38}$$

Given the specific structure of the objective function of LRA-OP1, the latter can be split into two decoupled subproblems as formulated in the following.

Prop. 5.2 *The LRA-OP1 at the ℓth PR can be solved in two steps:*

i) *search for the optimal bit loading for each value of the data rate $B^{(\ell)} \triangleq \sum_{n\in\mathcal{D}_s} m_n^{(\ell)}$, where $B^{(\ell)} \in \mathcal{D}_B \triangleq \{0, 2, \cdots, B_{\max}^{(\ell)}\}$ and $\mathcal{D}_s$ defines the set of subcarriers, as*

$$\mathbf{LRA-OP1.a}: \quad \begin{array}{l} \mathbf{m}_{\rm o}(B^{(\ell)}) = \arg\min\limits_{\mathbf{m}} \left\{ \sum_{n\in\mathcal{D}_s} \Omega_n^{(\ell)}(\mathbf{m}) \right\} \\ \text{s.t.} \quad {\mathbf{m}^{(\ell)}}^T \mathbf{1}_N = B^{(\ell)},\ \forall B^{(\ell)} \in \mathcal{D}_B \end{array}, \tag{5.39}$$

where the sequence $\Omega_n^{(\ell)}$, $n \in \mathcal{D}_s$, defines the modulation model in (2.20);

ii) *select the optimal combination of bit loading and coding rate as*

$$\mathbf{LRA-OP1.b}: \quad \begin{array}{l} \boldsymbol{\varphi}_{\rm o}^{(\ell)} = \arg\max\limits_{\boldsymbol{\varphi}} \left\{ \zeta^{(\ell)}(\boldsymbol{\varphi}|\boldsymbol{\sigma}^{(\ell-1)}) \right\} \\ \text{s.t.} \quad \boldsymbol{\varphi} \in \mathcal{D}_{\boldsymbol{\varphi}}^{\star} \end{array}, \tag{5.40}$$

where $\mathcal{D}_{\boldsymbol{\varphi}}^{\star} \triangleq (\mathcal{D}_{\mathbf{m}}^{\star} \times \mathcal{D}_r) \subseteq \mathcal{D}_{\boldsymbol{\varphi}}$, $\mathcal{D}_{\mathbf{m}}^{\star} \subseteq \mathcal{D}_{\mathbf{m}}$ being the set that includes only the results $\mathbf{m}_{\rm o}(B^{(\ell)})$ solving LRA-OP1.a.

Proof In order to show the equivalence of LRA-OP1 and LRA-OP1.a-LRA-OP1.b, first recall the results obtained in Prop. 5.1, i.e., $\lambda(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})$ monotonically decreases with $\hat{\gamma}^{(\ell)} = \hat{\gamma}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)})$, while $1 - P_{\rm out}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})$ monotonically increases with $\hat{\gamma}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)})$. In fact, it can be easily seen that the first derivative of (5.37) w.r.t. $\hat{\gamma}^{(\ell)}$ is always greater than zero, so that (5.37) is monotone increasing with $\hat{\gamma}^{(\ell)}$. Moreover, the ESNR $\hat{\gamma}^{(\ell)}$ does not depend on the coding rate, but only on the bit distribution among the subcarriers, as it can be seen in (2.19)-(2.20). Hence, the equivalence easily follows. In fact, it is first evaluated, for each possible value of data rate $B^{(\ell)}$, the bit distribution $\mathbf{m}_{\rm o}(B^{(\ell)})$ that maximize the ESNR. From (2.19), it follows that $\hat{\gamma}^{(\ell)}(\boldsymbol{\varphi}^{(\ell)})$ is maximized for that $\mathbf{m}_{\rm o}(B^{(\ell)})$ that makes the sum $\sum_{n\in\mathcal{D}_s} \Omega_n^{(\ell)}(\mathbf{m}^{(\ell)})$ minimum, defining accordingly LRA-OP1.a. The solutions to LRA-OP1.a provide the set $\mathcal{D}_{\mathbf{m}}^{\star}$ including the optimal bit distribution $\mathbf{m}_{\rm o}(B^{(\ell)})$, one for each $B^{(\ell)} \in \mathcal{D}_B$. Then, in LRA-OP1.b, the optimal combination of $\mathbf{m}_{\rm o}(B^{(\ell)})$ and coding rate $r^{(\ell)} \in \mathcal{D}_r$ that gives the maximum value of the EGP, among all the possible values $\zeta^{(\ell)}(\mathbf{m}_{\rm o}(B^{(\ell)}), r^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})$, is chosen. This is pursed with an exhaustive search over the set $\mathcal{D}_{\boldsymbol{\varphi}}^{\star}$, such that $|\mathcal{D}_{\boldsymbol{\varphi}}^{\star}| \ll |\mathcal{D}_{\boldsymbol{\varphi}}|$ due to the previous "skimming" action performed by LRA-OP1.a over all the possible combinations of $\mathbf{m}^{(\ell)} \in \mathcal{D}_{\mathbf{m}}$. □

5.4.1 Bit loading algorithm

The idea we pursue in this section is that of reformulating LRA-OP1.a exploiting the embedded structure of matroid, so that it can be optimally and efficiently solved through an iterative greedy approach.
Toward this end, let us recall the definition and some useful propertis of the matroids.

Definition 5.1 (Matroids [87]) *A finite matroid is defined as an ordered pair of sets* $\mathbb{M} \triangleq \{\mathcal{S}, \mathcal{I}\}$*, where* $\mathcal{S}$ *is a finite non-empty set and* $\mathcal{I}$ *is a non-empty family of subsets of* $\mathcal{S}$ *(called the independent subsets) with the properties:*
i) the empty set $\emptyset \in \mathcal{I}$*;*
ii) if $\mathcal{Y} \in \mathcal{I}$ *and* $\mathcal{X} \subseteq \mathcal{Y}$*, then* $\mathcal{X} \in \mathcal{I}$ *(*hereditary *property);*
iii) if $\mathcal{X}, \mathcal{Y} \in \mathcal{I}$*, with* $|\mathcal{X}| < |\mathcal{Y}|$*, there exists an element* $x \in \mathcal{Y} - \mathcal{X}$ *such that* $\mathcal{X} \cup \{x\} \in \mathcal{Y}$ *(*exchange *property).*
Then, a weighted matroid is defined as a matroid $\mathbb{M} \triangleq \{\mathcal{S}, \mathcal{I}\}$ *in which a weight function* $w : \mathcal{S} \to \mathbb{R}^+$ *assigns a strictly positive weight to each element of* $\mathcal{S}$*. Therefore,* $\forall \mathcal{X} \subseteq \mathcal{S}$ *we have* $w(\mathcal{X}) = \sum_{x \in \mathcal{X}} w(x)$*.*

The concept of weighted matroid is usually combined with that of greedy algorithm, a class of (typically simple) iterative methods which can solve a given OP through a sequence of locally optimal steps, i.e., at a given iteration the choice we adopt looks as the best one, irrespective of the effect it may have in the future. More specifically, for a given weighted matroid there always exists a greedy algorithm working on it that enables to find the minimum-weight independent subset at affordable complexity.

Accordingly, dropping for the sake of notation the dependence on the PR index ℓ, in order to find a solution to LRA-OP1.a, let us define $\mathbb{M} \triangleq \{\mathcal{S}, \mathcal{I}\}$ as the weighted matroid such that:

i) $\mathcal{S} \triangleq \{\vartheta_{m,n}, \forall m \in \{2, \cdots, m_{\max} - 2\}, \forall n \in \mathcal{C}\}$, $\vartheta_{m,n}$ representing the action of increasing from $m-2$ to m the number of coded bits loaded on the nth subcarrier, i.e., $0 \to 2$, $2 \to 4$, $(m_{\max} - 2) \to m_{\max}$;

ii) the family $\{\mathcal{X}_a\}$ of independent subsets, with $\mathcal{X}_a \in \mathcal{S}$ and $a \triangleq |\mathcal{X}|$, is composed by all the combinations of the elements of $\mathcal{S}$;

iii) the weight function assigned to each element of $\mathcal{S}$ is defined as

$$w_m(\gamma_n) \triangleq \Omega_n(m|\gamma_n) - \Omega_n(m-2|\gamma_n),\ m \in \{2\,,\cdots,m_{\max}\}\,,\ \Omega_n(0|\gamma_n) \triangleq 0, \tag{5.41}$$

where $\Omega_n(m|\gamma_n)$ is the modulation model (2.20) expressed as a function of the number of bits m loaded over subcarrier n, for a given SNR level equal to γ_n. As it can be shown that $\Omega_n(\overline{m}|\gamma_n) > \Omega_n(m|\gamma_n)$ for $\forall \overline{m} > m$, it can be noted that the weight $w_m(\gamma_n) > 0$.

Due to the above definition the following results follow.

1. The hereditary and exchange property of the matroid $\mathbb{M}$ can be easily verified.
2. Each subset $\mathcal{X} \in \mathcal{I}$ is a *feasible* bit loading distribution if its elements are such that

$$\vartheta_{m,n} \in \mathcal{X} \quad \Rightarrow \quad \vartheta_{m-2,n} \in \mathcal{X},\ \forall m \in \{2\,,\cdots,m_{\max}\}\,,\ \forall n \in \mathcal{D}_s. \tag{5.42}$$

   This means, for instance, that $\mathcal{X}_1 \triangleq \{\vartheta_{4,n}\}$ is not allowed since before the increment $2 \to 4$, the increment $0 \to 2$ has clearly to occur, whereas $\mathcal{X}_2 \triangleq \{\vartheta_{2,n}, \vartheta_{4,n}\}$ and $\mathcal{X}_1 \triangleq \{\vartheta_{2,n}\}$ both obey to (5.42) and thus are feasible bit loading configurations.

3. The minimum-weight independent subset $\mathcal{X}_j^{(\mathrm{o})}$ made of j elements, $1 \le j \le |\mathcal{S}|$, returned by a greedy algorithm working on the corresponding weighted matroid $\mathbb{M}$, is a feasible bit loading distribution (optimality condition), which represents the optimal solution $\mathbf{m}_{\mathrm{o}}(B)$ of LRA-OP1.a, with $B = 2j$, if

$$w_{\overline{m}}(\gamma_n) > w_m(\gamma_n),\ \forall \overline{m} > m,\ \forall n \in \mathcal{D}_s. \tag{5.43}$$

   Thus, $\mathcal{X}_j^{(\mathrm{o})}$ is built starting from $\mathcal{X}_{j-1}^{(\mathrm{o})}$, or equivalently $\mathbf{m}_{\mathrm{o}}(B)$ is derived from $\mathbf{m}_{\mathrm{o}}(B-2)$, and making the local (greedy) optimal choice.

Collecting together the above results, we end up to the following iterative procedure.

1) Sort $\mathcal{S}$ according to the increasing weights of its elements, thus obtaining $\mathcal{G}$;
2) Initialize $\mathbf{m}_{\mathrm{o}}(0) = \mathbf{0}$, $j = 1$;
3) Select the jth element of $\mathcal{G}$ and denote it as $\vartheta_{m,n}$;

4) Return $\mathbf{m}_{\mathrm{o}}(B)$ with $B = 2j$, starting from $\mathbf{m}_{\mathrm{o}}(B-2)$ obtained at the $(j-1)$th iteration and switching from m to $m+2$ bits on the nth subchannel, n and m being given by Step 3);

5) Increment j by 1;

6) If $j \leq |\mathcal{S}|$ then

7) Go to Step 3);

8) Else, the algorithm ends, returning the sequence $\mathbf{m}_{\mathrm{o}}(B)$, $B \in \{0, 2, \cdots, B_{\max}\}$

It is worth noting that the solution to LRA-OP1.a is *optimal*, as at each step j the set composed of j positive elements with the smallest values is selected, and so, no other set of j elements exists whose sum is lower; *efficient*, as having the problem optimal substructure, i.e., $\mathbf{m}_{\mathrm{o}}(B-2)$ is included in $\mathbf{m}_{\mathrm{o}}(B)$, and the greedy choice property, i.e., at each iteration the selection of the element with the smallest weight is locally optimal, we can find the globally optimal solution applying a sequence of $|\mathcal{S}| = \dfrac{m_{\max}}{2}N$ greedy choices, instead of $\left(\dfrac{m_{\max}}{2}+1\right)^N$ required by the exhaustive search.

The above mentioned procedure brings some comments that are worth of being emphasized.
The weights $w_m(\gamma_n)$ for $m = 2, 4, 6$ obey to the optimal condition (5.43) only for $\gamma_n \geq \gamma_{0,m_{\max}}$, where the threshold depends on the highest modulation order $m_{\max}$. In the case of $m_{\max} = 6$, the threshold $\gamma_{0,m_{\max}}$ is around 2 dB. Therefore, whenever for a given subcarrier we have $\gamma_n < \gamma_{0,m_{\max}}$, the greedy procedure does not provide the optimal solution in that it is applied to a problem that does not have optimal substructure, and so, it may give a bit loading configuration that is not feasible. Nevertheless, there always exists a threshold $\gamma_{0,m_{\max}}$ such that the optimal condition (5.43) holds, with

$$\gamma_{0,m_{\max}} \leq G_\gamma(m_{\max}), \tag{5.44}$$

where

$$G_\gamma(m) = \frac{4}{\left[d_m^{(\min)}\right]^2 + \left[d_{\overline{m}-2}^{(\min)}\right]^2 - \left[d_{\overline{m}}^{(\min)}\right]^2} \log\left(\frac{2^{\overline{m}-1}\psi_m(1)}{2^{m-3}\psi_{\overline{m}}(1) + 2^{\overline{m}-1}\psi_{m-2}(1)}\right). \tag{5.45}$$

Proof Let us denote $\alpha_{m_n}(\mu) \triangleq \psi_{m_n}(\mu)/2^{m_n-1}$ and $\beta_{m_n}(\mu) \triangleq \mu^2\left[d_{m_n}^{(\min)}\right]^2/4$ in (2.20), for notation convenience. Substituting (2.20) into (5.41), the weight for a 2^{m_n}-QAM modulation over subcarrier n results

$$w_{m_n} = \sum_{\mu=1}^{\sqrt{2^{m_n}}/2} \left[\alpha_{m_n}(\mu)\mathrm{e}^{-\gamma_n\beta_{m_n}(\mu)} - \alpha_{m_n-2}(\mu)\mathrm{e}^{-\gamma_n\beta_{m_n-2}(\mu)}\right]. \tag{5.46}$$

For sufficiently large SNR values, (2.20) is quite well approximated by the terms relevant to the symbols in the complementary subset at distance $d_{m_n}^{(\min)}$ [45], i.e., the ones relevant to $\mu = 1$. Under this assumption, the optimality condition $\omega_{\overline{m}} > \omega_m$ can be written as

$$\begin{aligned}&\alpha_{m_n}(1)\mathrm{e}^{-\gamma_n\beta_{m_n}(1)} + \alpha_{\overline{m}_n-2}(1)\mathrm{e}^{-\beta_{\overline{m}_n-2}(1)} < \\ &\alpha_{\overline{m}_n}(1)\mathrm{e}^{-\gamma_n\beta_{\overline{m}_n}(1)} + \alpha_{m_n-2}(1)\mathrm{e}^{-\gamma_n\beta_{m_n-2}(1)}.\end{aligned} \tag{5.47}$$

Recalling that $\overline{m}_n > m_n$ and according to the properties of α_{m_n} and β_{m_n} described in Appendix A.2, it follows that $\alpha_{m_n-2}(1) < \alpha_{m_n}(1) \leq \alpha_{\overline{m}_n-2}(1) < \alpha_{\overline{m}_n}(1)$, whereas $\beta_{m_n-2}(1) > \beta_{m_n}(1) \geq \beta_{\overline{m}_n-2}(1) > \beta_{\overline{m}_n}(1)$. Thus, the first term of inequality (5.47) can be lower bounded by $\alpha_{m_n}(1)\mathrm{e}^{-\gamma_n(\beta_{m_n}(1)+\beta_{\overline{m}_n-2}(1))}$. The second term of inequality (5.47) can instead be upper bounded by $(\alpha_{\overline{m}_n}(1) + \alpha_{m_n-2}(1))\,\mathrm{e}^{-\beta_{\overline{m}_n}(1)\gamma_n}$. Substituting these bounds in (5.47), optimality condition (5.44)-(5.45) is eventually found. □

In the low-SNR region where (5.43) is not satisfied, the weights crossing each other are almost overlapped. This fact suggests a slight modification of the procedure that enables a suboptimal yet greedy and feasible bit loading solution: the initial sorting is modified moving $\vartheta_{\overline{m},n}$ just after $\vartheta_{m,n}$ if $w_{\overline{m}}(\gamma_n) < w_m(\gamma_n)$, for $\overline{m} > m$, $\forall n \in \mathcal{D}_s$. The complexity load required by this greedy-based solution is mainly due to the sorting of $\mathcal{S}$, and accordingly, results as $\mathcal{O}(|\mathcal{S}|\log|\mathcal{S}|)$, with $|\mathcal{S}| = \dfrac{m_{\max}}{2}N$.
The the bit loading algorithm solving LRA-OP1.a is summarized in Tab. 5.1, where $\mathcal{G}(j)$ denotes the jth element of the set $\mathcal{G}$.

5.4.2 AMC with BL algorithm

As stated earlier, once the optimal bit loading configuration $\mathbf{m}_o(B)$ for each possible value of data rate B is found, the solution to LRA-OP1 is obtained with a simple exhaustive search over all the pairs $(\mathbf{m}_o(B), r) \in \mathcal{D}_{\boldsymbol{\varphi}}^{\star}$, i.e., solving LRA-OP1.b. This procedure is summarized in Tab. 5.2.

Bit loading algorithm for LRA-OP1.a

Initialization phase

1. Build $\mathcal{G}$ sorting the elements of $\mathcal{S}$ in increasing order
2. For $n = 1 : N$
3. If $w_{\overline{m}}(\gamma_n) < w_m(\gamma_n)$, for $\overline{m} > m$
4. Move $\vartheta_{\overline{m},n}$ just after $\vartheta_{m,n}$ in $\mathcal{G}$
5. End If
6. End For
7. Initialize $\mathbf{m}_o(0) = \mathbf{0}$

Greedy procedure

8. For $j = 1, \cdots, |\mathcal{G}|$
9. Set $\theta_{m,n} = \mathcal{G}(j)$ and $B = 2j$
10. Return $\mathbf{m}_o(B) = \mathbf{m}_o(B-2) + [\overbrace{0, \cdots, 0}^{n}, 2, \overbrace{0, \cdots, 0}^{N-n}]^{\mathrm{T}}$
11. End For

Table 5.1: *Pseudo-code of the bit loading algorithm*

5.5 LRA with limited CSIT

In the LRA problem discussed above, the receiver must communicate on the feedback channel the whole SNR matrix $\boldsymbol{\Upsilon}^{(\ell)}$, since the transmitter requires the full CSI knowledge to perform the bit-loading procedure. An alternative solution that offers a trade-off between the amount of feedback information and quality of solution is given by assigning the same modulation order to all the subcarrier within the same round. In this way, it is sufficient that the receiver feeds back only the value of the index identifying the optimal pair modulation order/coding rate.
Formally, this AMC problem can be defined as

$$\text{OP2}: \quad \begin{array}{l} \boldsymbol{\varphi}_{\mathrm{o}}^{(\ell)} = \arg\max\limits_{\boldsymbol{\varphi}} \left\{ \zeta^{(\ell)}(\boldsymbol{\varphi}|\boldsymbol{\sigma}^{(\ell-1)}) \right\} \\ \text{s.t.} \quad \boldsymbol{\varphi} \in \mathcal{D}'_{\boldsymbol{\varphi}} \end{array}, \tag{5.48}$$

AMC with BL algorithm for LRA-OP1

1. Initialize $\zeta_{\mathrm{o}}^{(\ell)} = 0$
2. Evaluate the optimal BL set $\mathcal{D}_{\mathbf{m}}^{\star}$ with the procedure in Tab. 5.1
3. For $j = 1 : |\mathcal{D}_{\mathbf{m}}^{\star}|$
4. Set $\mathbf{m}_{\mathrm{o}} = \mathcal{D}_{\mathbf{m}}^{\star}(j)$
5. For $k = 1 : |\mathcal{D}_r|$
6. Set $\boldsymbol{\varphi}^{(\ell)} = (\mathbf{m}_{\mathrm{o}}, \mathcal{D}_r(k))$ and evaluate $\zeta^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})$
7. If $\zeta^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)}) > \zeta_{\mathrm{o}}^{(\ell)}$
8. Set $\boldsymbol{\varphi}_{\mathrm{o}}^{(\ell)} = \boldsymbol{\varphi}^{(\ell)}$ and $\zeta_{\mathrm{o}}^{(\ell)} = \zeta^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})$
9. End If
10. End For
11. End For
12. Return $\boldsymbol{\varphi}_{\mathrm{o}}^{(\ell)}$

Table 5.2: *Pseudo-code of the AMC algorithm with BL*

where now the TM $\boldsymbol{\varphi} \triangleq \{m, n\}$, since $m_n = m$, $\forall n \in \mathcal{D}_s$, and $\mathcal{D}'_{\boldsymbol{\varphi}} \triangleq \mathcal{D}_m \times \mathcal{D}_r$.

5.5.1 AMC algorithm

The optimization problem (5.48) can be easily solved through and exhaustive search over all the pairs modulation order and coding rate to find the one that optimizes the EGP. In fact, since all the quantities to be evaluated have a closed-form expression, its complexity simply reduces to $\mathcal{O}(|\mathcal{D}_m| \cdot |\mathcal{D}_r|)$. The AMC algorithm is summarized in the pseudo-code of Tab. 5.3.

5.6 Simulation results

Simulation tests have been carried over a typical wireless link to verify the efficacy of the proposed HARQ-based LRA algorithms. The setting of parameters/features adopted for the simulation is reported in Tabs. 5.4 and 5.5. All the performance

AMC algorithm for LRA-OP2

Initialize $\zeta_{\mathrm{o}}^{(\ell)} = 0$

For $j = 1 : |\mathcal{D}_m|$

  For $k = 1 : |\mathcal{D}_r|$

    Set $\boldsymbol{\varphi}^{(\ell)} = (\mathcal{D}_m(j), \mathcal{D}_r(k))$ and evaluate $\zeta^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})$

    If $\zeta^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)}) > \zeta_{\mathrm{o}}^{(\ell)}$

      Set $\boldsymbol{\varphi}_{\mathrm{o}}^{(\ell)} = \boldsymbol{\varphi}^{(\ell)}$ and $\zeta_{\mathrm{o}}^{(\ell)} = \zeta^{(\ell)}(\boldsymbol{\varphi}^{(\ell)}|\boldsymbol{\sigma}^{(\ell-1)})$

    End If

  End For

End For

Return $\boldsymbol{\varphi}_{\mathrm{o}}^{(\ell)}$

Table 5.3: *Pseudo-code of the AMC algorithm for EGP maximization with limited CSIT*

metrics are considered as a function of the distance between the transmitter (Tx) and the receiver (Rx), averaged over 10^3 independent channel realizations.

Figure 5.4 depicts the average AGP, i.e., the average of the ratio between the number of payload bits N_{p} and the transmission time required to successfully delivering them, offered by the adaptive HARQ-based algorithm with CSIT that exploits the BL procedure, summarized in Tabs. 5.1 and 5.2. This algorithm is tagged as HybridGP-AMC with BL. The AGP performance obtained by the latter are compared with the performance obtained in the case when simple ARQ without packet combining is employed. In the latter case, since there is no recombination, the ESNR adopted to evaluated the goodput is the one relevant to the single PR, i.e., the conventional κESM ESNR, described in Sect. 2.3.1. Interestingly, as it can be seen in Sect. 5.4.1, the BL procedure that maximizes the EGP in presence of HARQ (5.37) is the same that maximizes the EGP with simple ARQ mechanism, that is expressed by (4.6) considering uniform PA. In fact, the BL is obtained by minimizing the sum of the modulation models Ω_n, which is identified by the same expression (2.20) in both cases. Thus, in conclusion, this benchmark is obtained applying the same algorithm in Tabs. 5.1-5.2, where the evaluation of the goodput is now done according to (4.6),

Parameter/Feature	Symbol	Value/Description
RLC-PDU		
Payload length	N_{p}	1024 bits
CRC length	N_{CRC}	32 bits
OFDM		
No. of active subchannels	C	1320
FFT size	N_{FFT}	2048
CP length	N_{CP}	160 samples
Modulation & Coding		
Bits per subcarrier	$\mathcal{D}_m$	$\{0, 2, 4, 6\}$
Code type		PCCC turbo-code
Mother code rate	r	1/3
Punctured code rates	$\mathcal{D}_r$	$\left\{\frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{4}{7}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{6}{7}\right\}$
Transmitted power (Figs. 5.4-5.6)	P	20 dBm
Transmitted power (Figs. 5.7-5.9)	P	40 dBm
Bandwidth	B	20 MHz
ARQ		
ARQ scheme		Multiple-channel Stop & Wait
No. of logical channels		8

Table 5.4: *Parameters and features of the HARQ BIC-OFDM system.*

i.e., without exploiting the αESM methodology, but only the κESM one. We will refer to this case simply with GP-AMC with BL. From Fig. 5.4, it can be argued how the HybridGP-AMC with BL has an edge over the simple GP-AMC with BL, especially when the distance between the Tx and the Rx increases, corresponding to the low SNRs region. Here in fact, since the reliability is lower, the recombination mechanism due to the HARQ protocol is well exploited.

In order to further reduce complexity, the overall number of subcarriers can be grouped in N/U *chunks*, each composed of U adjacent subcarriers spanning a frequency interval smaller than the channel coherence bandwidth. As a result, the bit loading procedure is simplified since the subcarriers of each chunk are loaded with the

Parameter/Feature	Value/Description
Path-loss model	NLOS urban scenario, [77]
Carrier frequency	2 GHz
Base station height	12.5 m
Mobile terminal height	1.5 m
Noise power level	-100 dBm
Long-term fading model	Log-normal distribution
Variance of the shadowing	6 dB
Short-term fading model	ITU Ped. B, [33]

Table 5.5: *Parameters and features of the radio propagation channel model.*

same number of bits. Figure 5.5 proves that the AGP for both the HybridGP-AMC with BL and GP-AMC with BL slightly degrades when using $U = 4$ when compared to $U = 1$, i.e., the bit loading at subcarrier level.
In Fig. 5.6, the performance of the HybridGP-AMC algorithm with limited CSIT described in Tab. 5.3 are compared with the ones obtained with the HybridGP-AMC with BL algorithm and the ones obtained with simple GP-AMC algorithm. It is evident that, at low distances, where the SNR level is higher, the recombination mechanism does not influence much the results. In fact, the algorithms without BL have almost the same performance, whereas the one with BL produce a better AGP due to the finest mechanism with whom it assigns different modulation orders per subcarrier. When the distance increases, so that the SNR level gets lower, the recombination mechanism starts playing an important role, in fact the performance obtained by two algorithms based on HARQ with packet combining get closer outperforming the GP-AMC algorithm.
Figures 5.7-5.9 are about the HybridGP-AMC with BL and the GP-AMC with BL when a QoS constraint is imposed, which specifies the maximum delay in the delivery of each packet to the end-user as $T_{\mathrm{QoS}} = 0.5$ms. From Fig. 5.7, it can be noticed that the packet drop rate, namely the ratio between the number of discarded packets that are not delivered within the maximum allowed delay and the total number of transmitted packets, is lower for the HybridGP-AMC with BL, still providing the same value of AGP. This competitive feature can be explained as follows, also looking

at Fig. 5.9 as well, which depicts the retransmission rate. Actually, while the GP-AMC with BL is more conservative and transmits with lower data rate to guarantee adequate PER performance, the HybridGP-AMC with BL adopts a stronger resource allocation strategy in virtue of the retransmission model included in the EGP objective function. Consequently, whenever a NACK is received, different copies of the same packet are transmitted within the allowed interval T_{QoS}, thus efficiently exploiting the HARQ recombination mechanism. This is why the HybridGP-AMC with BL has both higher retransmission rate than the simple GP-LRA with BL and enables better data rate performance.

5.7 Concluding remarks

An efficient LRA policy suitable for improving the link performance of BIC-OFDM systems in presence of HARQ schemes with packet combining was proposed and properly outlined. First of all, a novel link performance prediction method was derived relying on the ESM approach, called aggregate ESM, or αESM, capable of taking into account the entire channel history up to the actual PR through a simple scalar value. By exploiting a lower bound of the αESM, characterized by a simple first-order recursive equation, the adaptive modulation and coding problem aimed at maximizing the goodput was properly formalized and solved, either in the case of full CSIT and limited CSIT. In the former case, the AMC strategy derived is able to assign a different modulation order per subcarrier and the optimality of this bit loading procedure was proven resorting to the matroids theory. Simulation results confirmed that the proposed schemes remarkably boost link performance compared to AMC strategies that do not account for the recombination mechanism of the HARQ scheme with packet combining and, besides, when QoS constrained applications are considered, the proposed strategy helps in increasing the number of packets meeting the QoS constraint.

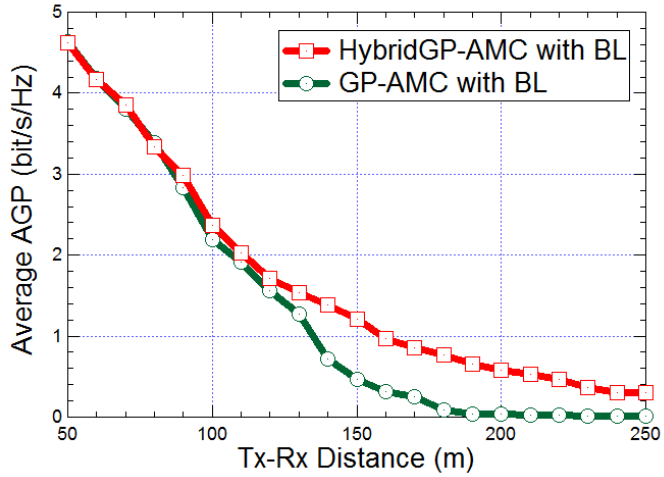

Figure 5.4: *HybridGP-LRA with BL vs. GP-LRA with BL. AGP Performance comparison.*

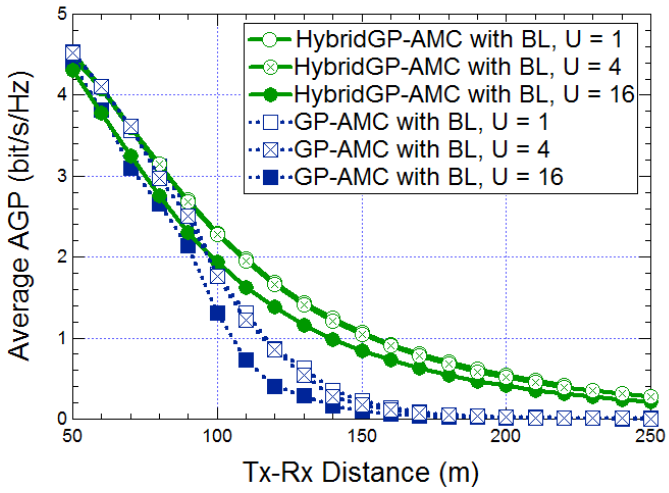

Figure 5.5: *HybridGP-LRA with BL vs. GP-LRA with BL. Chunk approximation.*

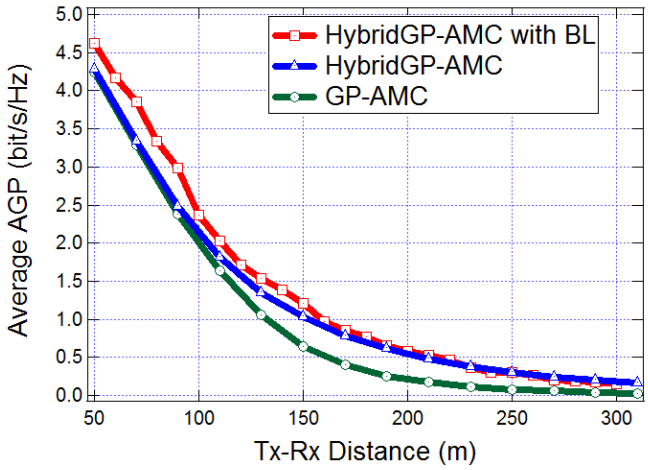

Figure 5.6: *HybridGP-AMC with BL vs. HybridGP-AMC and GP-AMC. AGP performance comparison.*

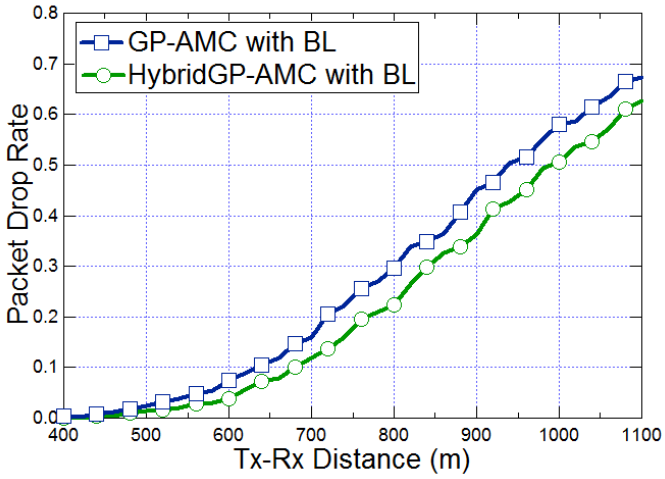

Figure 5.7: *HybridGP-LRA with BL vs. GP-LRA with BL. Packet dropped rate comparison.*

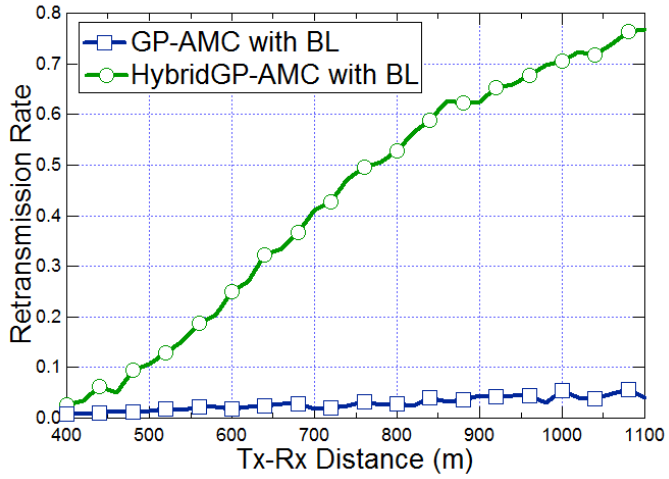

Figure 5.8: *HybridGP-LRA with BL vs. GP-LRA with BL. Retransmission rate comparison.*

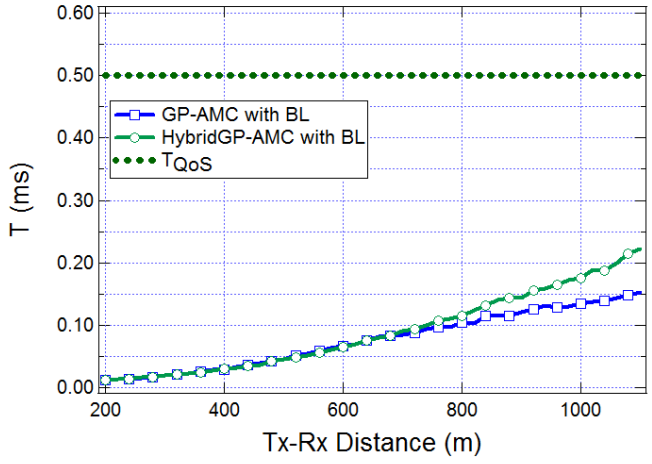

Figure 5.9: *HybridGP-LRA with BL vs. GP-LRA with BL. QoS performance comparison.*

Chapter 6

Fair RA scheme for BIC-OFDMA systems

In this Chapter, a novel RA algorithm, aimed at optimizing the performance of the user with the lowest GP value on the downlink of a BIC-OFDMA base station, is proposed. In detail, after an overview on the main RA approaches usually adopted in literature, the BIC-OFDMA system under analysis and the max-min RA problem are introduced. The RA problem consists in particular of three subproblems, i.e., the power allocation problem, the AMC problem and the subcarrier allocation problem, which are analyzed and optimally solved in Sections 6.4-6.6. For the latter problem, an efficient solution based on the Ant Colony Optimization framework is also provided. Finally, Sect. 6.7 shows simulation results obtained over realistic wireless propagation scenarios, which corroborate the effectiveness of the proposed RA algorithm.

6.1 Resource allocation in OFDMA systems

Radio resource management (RRM) is the set of functionalities whose aim is to provide services according to the QoS negotiated for each application over the area covered by the system and to optimize the system performance through the choice of the best radio resource sharing among users. Scheduling, together with some other well known functionalities such as power control, handover, admission control, congestion and load control and link adaptation (LA), belongs to RRM.

A general definition of the radio resource that should be assigned is quite hard to give, because it depends on the particular wireless system. However, a radio resource

(RR) could be proposed as the signal format necessary to define how a certain amount of data can be transmitted over the wireless medium. For instance, in a TDMA system, an RR is identified by the time slot over which transmission is allowed, the carrier frequency and the relevant bandwidth, the modulation and coding format, the power level and the transmitting spatial dimension. In a FDMA system, instead of the time slot, the RR depends on the portions of spectrum assigned to each user for transmission.

In the last decade, the principles of multiuser downlink and MAC designs have been changed. Instead of assigning a fixed amount of resources to each data flow, channel-aware scheduling strategies [20] have been proposed to adaptively transmit data and dynamically assign wireless resources based on CSI [88]. The key idea is the multiuser diversity, that consists in taking advantage of the independent channel variation across users choosing the ones with good channel conditions to transmit packets [88], improving in this way network performance. For instance, multiuser diversity is exploited by the concept of opportunistic scheduling [89], where some form of priority is given to users with (temporarily) better channels. A major problem for multiuser adaptation schemes is how to design practical algorithms that achieve the multiuser diversity gain while supporting diverse QoS requirements under realistic channel scenarios.

6.1.1 Margin and rate adaptive approaches

As anticipated in Sect. 1.2, OFDMA is not only preferred in many current standards, but is also a strong candidate for next generation cellular systems thanks to its flexibility to perform dynamic allocation of the available radio resources. Here, the multiuser diversity is particularly exploited in the frequency domain where channels for different users fade independently, so that, diversity gain arises from the fact that in a system with many users, there is likely to be a user whose channel is near its peak at any one time. Taking advantage of knowledge of the CSIT, the OFDMA systems can employ the following adaptive resource allocation (RA) technique [88]: AMC, adaptive power allocation, which have already been discussed in the previous chapters, and dynamic subcarrier assignment, where the base station dynamically assigns subcarriers according to CSI or/and QoS requirements.

In literature, there can be distinguished two main approaches to the RA problem.

One is *margin-adaptive* which minimizes the total power consumption subject to prescribed rate requirements for users. This is the approach developed for instance in [90], [91], [92]. The other one is a *rate-adaptive* approach, where the aim is to maximize the sum rate of all users, given a total transmit power constraint [93], [94]. In addition, proportional rate constraints can be imposed to guarantee proportional fairness among users [95], [96].

6.1.2 The maximum fairness problem

The above mentioned approaches can however result unfair in the way resources are assigned to users. In fact, different mobile terminals will experience different average channel conditions, due, for example, to shadow fading and distance dependent path loss that significantly affect the average received signal strength. In this case, different terminals can experience different channel conditions for a long time, so that users with worse channel conditions will receive less resources and obtain poor performance. This situation is usually called starvation. A practical example can be a rate adaptive approach in which resources have to be shared among two users, one very close to the base station (BS) and the other, on the contrary, close to the cell border. This is a typical case in which the former user, that is characterized by an higher average SNR level thanks to its proximity to the BS, obtains almost all the transmission resources.

Thus, to achieve the maximum fairness among users, a *max-min approach* can be adopted to provide similar performance to users [97], which means solving a max-min optimization problem. In literature, the max-min problem has so far focused on providing similar rates (or capacity) to all the users served by an OFDMA BS. In particular, this problem has been tackled in [98], where the subcarriers are dynamically allocated to the users, and in [99], by optimizing the subcarriers and bits assigned to each user. In [97] a multi-cell OFDMA scenario is considered, where this goal is obtained by jointly optimizing coordinated BSs subcarrier and power allocation.

Here, differently to what has been done so far, the focus is on the maximum fairness problem aimed at providing similar GP performance to each user on the downlink of a BIC-OFDMA BS, by dynamically allocating subcarrier, power, modulation order and coding rate.

6.2 BIC-OFDMA system model

In this section, the BIC-OFDMA system that serves Q users in the set $\mathcal{Q} \triangleq \{1, \cdots, Q\}$ over N available subcarriers in the set $\mathcal{D}_s \triangleq \{1, \cdots, N\}$ is described. The qth link is is depicted in Fig. 6.1. Since the processing is similar to the one of the single user BIC-OFDM system described in Sect. 2.2, only the main differences will be highlighted, referring to Sect. 2.2 for the analogous parts. The Q packets, one per

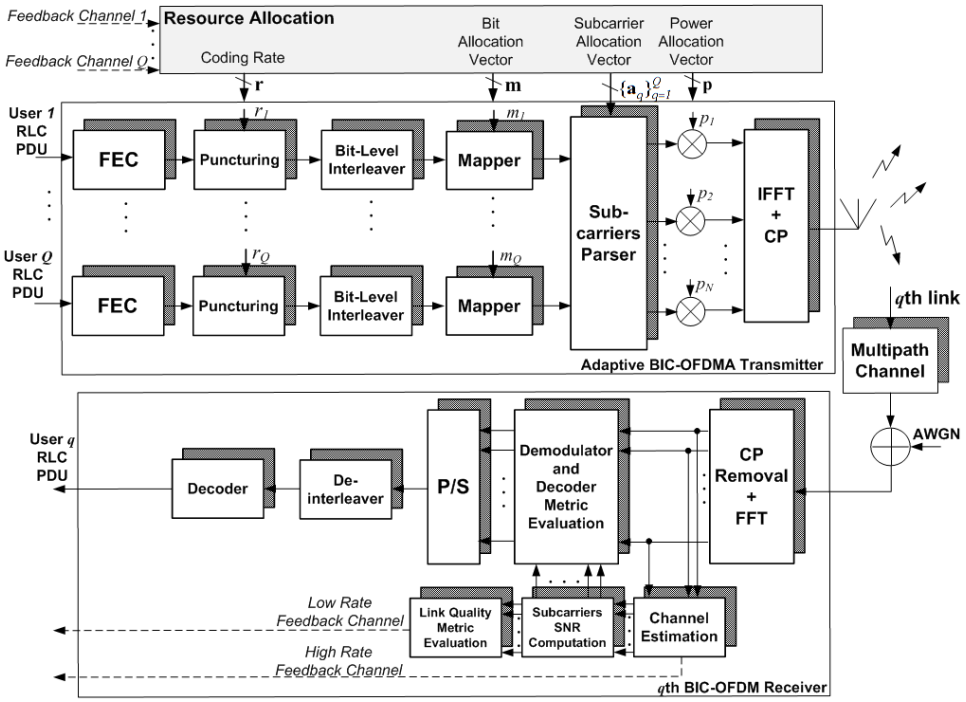

Figure 6.1: *BIC-OFDMA system model.*

each user, coming from the upper layers are processed in two steps. In the former, i.e., the packet processing, the generic packet belonging to user q, which is $N_{\mathrm{u}}^{(q)} = N_{\mathrm{h}}^{(q)} + N_{\mathrm{p}}^{(q)} + N_{\mathrm{CRC}}^{(q)}$ bits long, i.e., the sum of the lengths of the header, payload and cyclic redundancy check, respectively, is the input of a FEC encoder with mother code r_0 and it is transmitted in a maximum of L ARQ PRs. At each PR[1] ℓ, $1 \leq \ell \leq L$,

[1]Packets of different users may be at different stages of the retransmission mechanism. It means that ℓ is actually ℓ_q, but, to avoid a cumbersome notation and keeping in mind this observation, we

the output of this block is punctured according to the coding rate $r_q^{(\ell)} \in \mathcal{D}_r \triangleq \{r_0, \cdots, r_{\max}\}$ assigned to the qth user, obtaining $N_{\mathrm{c}}^{(\ell,q)} \triangleq N_{\mathrm{u}}^{(\ell,q)}/r_q^{(\ell)}$ CBS that, in turn, are randomly interleaved with a bit-level interleaver.

Let us denote the subcarrier allocation vector as $\mathbf{a}_q^{(\ell)} \triangleq [a_{q,1}^{(\ell)}, \cdots, a_{q,N}^{(\ell)}]^{\mathrm{T}} \in \{0,1\}^N$, where $a_{q,n}^{(\ell)} = 1$ if subcarrier n is assigned to user q, 0 otherwise, and with $m_q^{(\ell)} \in \mathcal{D}_m \triangleq \{2, \cdots, m_{\max}\}$ the number of bits that user q loads on its subcarriers. According to the so-called orthogonality principle of the OFDMA paradigm, each subcarrier is assigned to at most one user at time, so that it must hold

$$\sum_{q=1}^{Q} a_{q,n}^{(\ell)} \leq 1, \quad \forall n \in \mathcal{D}_s. \tag{6.1}$$

Thus, in the frame processing, the CBS are mapped into the $(N_{\mathrm{c}}^{(\ell,q)}/m_q^{(\ell)})$-sized sequence $\{s_{q,i}^{(\ell)}\}$ made of unit-energy symbols, with $s_{q,i}^{(\ell)} \in 2^{m_q^{(\ell)}}$-QAM. This sequence is then allocated over the $N_q^{(\ell)} \triangleq \sum_{n=1}^{N} a_{q,n}^{(\ell)}$ available subcarriers and transmitted within $N_{\mathrm{ofdma},q}^{(\ell)} \triangleq N_{\mathrm{c}}^{(\ell,q)}/(m_q^{(\ell)} N_q^{(\ell)})$ OFDMA symbols as follows. For convenience, let us first introduce the N-sized vector $\mathbf{x}_q^{(\ell)} \triangleq [x_{q,1}^{(\ell)}, \cdots, x_{q,N}^{(\ell)}]^{\mathrm{T}}$ built in this way. At the generic jth OFDMA symbol, the $N_q^{(\ell)}$ positions of $\mathbf{x}_q^{(\ell)}$, corresponding to the subcarrier indexes n for which $a_{q,n}^{(\ell)} = 1$, contain the sequence of QAM symbols $\{s_{q,k}^{(\ell)}\}$, with $\left((j-1)N_q^{(\ell)} + 1\right) \leq k \leq \left(jN_q^{(\ell)}\right)$, and the other positions are set to 0. The overall symbol vector is then built as $\mathbf{x}^{(\ell)} \triangleq \sum_{q=1}^{Q} \mathbf{x}_q^{(\ell)}$ and allocated on the N subcarriers along with a certain amount of power $\mathbf{p}^{(\ell)} \triangleq [p_1^{(\ell)}, \cdots, p_N^{(\ell)}]^{\mathrm{T}}$ such that $\sum_{n=1}^{N} p_n^{(\ell)} = P$, where P is the total available power at each PR. The resulting vector is processed by the IFFT and the CP insertion blocks. Finally, the obtained signal is transmitted over a block fading channel.

The sample received on subcarrier n by the qth user results

$$z_{q,n}^{(\ell)} = a_{q,n}^{(\ell)} \sqrt{p_n^{(\ell)}} h_{q,n}^{(\ell)} x_{q,n}^{(\ell)} + w_{q,n}^{(\ell)}, \tag{6.2}$$

where $h_{q,n}^{(\ell)}$ is the complex-valued channel coefficient at the qth receiver on subcarrier n and $w_{q,n}^{(\ell)} \in \mathcal{N}(0, {\sigma_{\ell,q,n}^{(w)}}^2)$ represents the ambient noise for user q on subcarrier n. Therefore, the diagonal post-processing SNRs matrix results $\boldsymbol{\Upsilon}_q^{(\ell)} \triangleq \mathrm{diag}\left\{[\gamma_{q,1}^{(\ell)}, \cdots, \gamma_{q,N}^{(\ell)}]\right\}$, where $\gamma_{q,n}^{(\ell)} = p_n^{(\ell)} |h_{q,n}^{(\ell)}|^2 / {\sigma_{\ell,q,n}^{(w)}}^2$ is the SNR experienced by user q over subcarrier n at round ℓ.

---

will simply write ℓ in the following.

6.3 Max-min goodput problem

6.3.1 Problem formulation

In this section the RA problem, based on the fairness criterion described in Sect. 6.1.2, is introduced. The problem aims at deriving a proper setting of transmission parameters (TPs) $\boldsymbol{\tau}_q^{(\ell)} \triangleq \{r_q^{(\ell)}, m_q^{(\ell)}, \mathbf{a}_q^{(\ell)}, \mathbf{p}^{(\ell)}\}$, $\forall q \in \mathcal{Q}$ at each scheduling period, which corresponds to the generic transmission attempt ℓ of user q, as shown by the SW mechanism in Fig. 6.2. In particular, with reference to the inter-round optimization

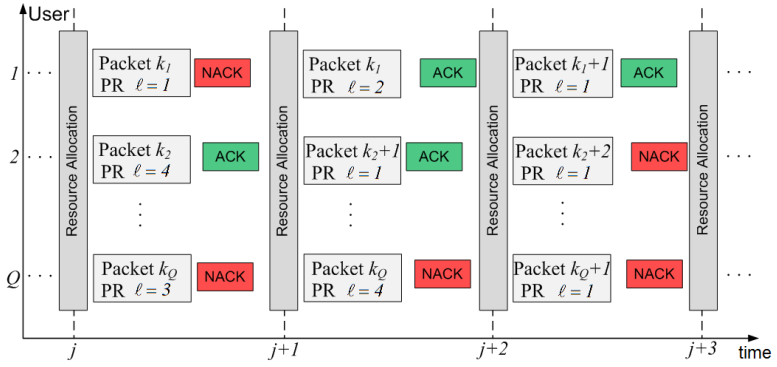

Figure 6.2: *Multi-user SW.*

criterion in Sect. 3.2.3, we consider as figure of merit the limiting behavior of the EGP (3.17) for large retry limit, i.e., $L \to \infty$. In this case, it can be easily shown that, for each user, the optimization of (3.17) is equivalent to optimize the second term at the denominator of (3.17). Thus, solving the series for $L \to \infty$ and exploiting the κESM methodology to evaluate the PER for the generic user q as described in Sect. 2.3.1, i.e. $\mathrm{PER}_{r_q^{(\ell)}}(\boldsymbol{\tau}_q^{(\ell)}|\mathbf{H}_q^{(\ell)}) = \Psi_{r_q^{(\ell)}}(\hat{\gamma}_q(\boldsymbol{\tau}_q^{(\ell)}))$, with $\mathbf{H}_q^{(\ell)} \triangleq \mathrm{diag}\left\{[h_{q,1}^{(\ell)}, \cdots, h_{q,N}^{(\ell)}]\right\}$, the objective function, in (bits/OFDMA symbol), results

$$\zeta_q^{(\ell)}(\boldsymbol{\tau}_q^{(\ell)}) = \zeta_q^{(0)} r_q^{(\ell)}\, m_q^{(\ell)} \left(\sum_{n=1}^{N} a_{q,n}^{(\ell)}\right) \left[1 - \Psi_{r_q^{(\ell)}}\left(\hat{\gamma}_q(\boldsymbol{\tau}_q^{(\ell)})\right)\right], \tag{6.3}$$

where $\zeta_q^{(0)} \triangleq N_{\mathrm{p}}^{(q)}/N_{\mathrm{u}}^{(q)}$.

Actually, the ESNR described in Sect. 2.3.1 is evaluated for a single user OFDM system. Though, since the OFDMA paradigm allows only one user at a time to

transmit over each subcarrier, the evaluation of the ESNR for each user is straightforward. In fact, given the TP setting $\{\boldsymbol{\tau}_q^{(\ell)}\}_{q\in\mathcal{Q}}$, it is as if there were Q independent and non-overlapped OFDM systems, where the generic qth system is characterized by $\sum_{n=1}^{N} a_{q,n}$ subcarriers, and hence the ESNR $\hat{\gamma}_q^{(\ell)}(\boldsymbol{\tau}_q^{(\ell)})$ of each user can be simply evaluated from (2.19) as

$$\hat{\gamma}_q^{(\ell)}(\boldsymbol{\tau}_q^{(\ell)}) \triangleq -\log\left(\frac{1}{m_q^{(\ell)}\sum\limits_{i=1}^{N} a_{q,i}^{(\ell)}}\sum_{n=1}^{N} a_{q,n}^{(\ell)}\,\Omega_{q,n}(m_q^{(\ell)}, p_n^{(\ell)})\right), \qquad \forall q\in\mathcal{Q}, \tag{6.4}$$

where

$$\Omega_n(m_q^{(\ell)}, p_n^{(\ell)}) \triangleq \sum_{\mu=1}^{\sqrt{2^{m_{q,n}^{(\ell)}}}/2} \frac{\psi_{m_n}(\mu)}{2^{m_q^{(\ell)}}-1}\cdot \mathrm{e}^{-p_n^{(\ell)}\gamma_{q,n}^{(\ell)}\beta_{q,n}^{(\ell)}}, \tag{6.5}$$

is as in (2.20), with $\beta_{q,n}^{(\ell)} \triangleq \left(\mu\cdot d_{m_n}^{(\min)}/2\right)^2/p_n^{(\ell)}$, here expressed as a function of $m_q^{(\ell)}$ and $p_n^{(\ell)}$.

Hence, dropping w.l.g. the dependence on the PR index ℓ, we can now introduce the RA problem that achieves the maximum fairness among the users by maximizing the minimum user EGP, tagged as max-min goodput optimization problem (MMG-OP), as follows.

$$\begin{aligned}
\textbf{MMG-OP}: \quad \max_{\boldsymbol{\tau}\in\mathcal{D}_\tau} \quad & f(\boldsymbol{\tau}) = \min_{q\in\mathcal{Q}} \zeta_q(\boldsymbol{\tau}_q) & & \\
\text{s.t.} \quad & \textstyle\sum_{n=1}^{N} p_n = P & & \text{(6.6.a)} \\
& \textstyle\sum_{q\in\mathcal{Q}} a_{q,n} \le 1 & \forall n, & \text{(6.6.b)} \\
& \Psi_{r_q}\left(\hat{\gamma}_q(\boldsymbol{\tau}_q)\right) \le \overline{\Psi}_q & \forall q, & \text{(6.6.c)} \\
& \boldsymbol{\tau}_q \in \mathcal{D}_\tau & \forall q, & \text{(6.6.d)}
\end{aligned} \tag{6.6}$$

where $\boldsymbol{\tau} \triangleq \{\boldsymbol{\tau}_1, \cdots, \boldsymbol{\tau}_Q\}$, $\mathcal{D}_\tau \triangleq \mathcal{D}_p \times \mathcal{D}_a \times \mathcal{D}_m \times \mathcal{D}_r$ is the set of feasible transmission modes, with $\mathcal{D}_p \triangleq \{p_n | 0 \le p_n \le P,\ \forall n\}$, $\mathcal{D}_a \triangleq \{a_{q,n} | a_{q,n} \in \{0,1\},\ \forall q, n\}$, constraint (6.6.b) is the orthogonality constraint (6.1) and (6.6.c) is a constraint on the QoS, implying that the PER has to be kept below a desired limit $\overline{\Psi}_q$, $\forall q$. It is worth noting that this is a constrained nonlinear problem with both integer and continuous variables. In addition, $f(\boldsymbol{\tau})$ is not concave w.r.t. $\boldsymbol{\tau}$, meaning that multiple local

optima can exist. Finding the global optimum of such a function results therefore prohibitive. In the next section, the proposed RA algorithm to solve (6.6) will be highlighted.

6.3.2 Iterative algorithm for the MMG problem

The MMG-OP (6.6) is solved resorting to a coordinate ascent method (CAM) [74]. This method allows to reduce the complexity of problem (6.6), iteratively optimizing the objective function f w.r.t. one variable (i.e., the power allocation or the subcarrier allocation or the pair modulation and coding rate) keeping the others fixed in the meanwhile. Denoting with $\mathbf{a} \triangleq [\mathbf{a}_1^{\mathrm{T}}, \cdots, \mathbf{a}_Q^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{m} \triangleq [m_1, \cdots, m_Q]^{\mathrm{T}}$ and $\mathbf{r} \triangleq [r_1, \cdots, r_Q]^{\mathrm{T}}$ the overall subcarrier, bit and coding rate vectors, respectively, and with $\boldsymbol{\tau}_{-\mathbf{y}}$ the TP setting without transmission parameter $\mathbf{y}$, at the generic ith step of the algorithm, three subproblems are solved, returning respectively

i) the power allocation $\mathbf{p}^{(i+1)}$, given $\boldsymbol{\tau}_{-\mathbf{p}}^{(i)} \triangleq \{\mathbf{a}^{(i)}, \mathbf{m}^{(i)}, \mathbf{r}^{(i)}\}$, tagged as PA-OP;

ii) the subcarrier allocation $\mathbf{a}^{(i+1)}$ given $\boldsymbol{\tau}_{-\mathbf{a}}^{(i)} \triangleq \{\mathbf{p}^{(i+1)}, \mathbf{m}^{(i)}, \mathbf{r}^{(i)}\}$, named SA-OP;

iii) the pair $(\mathbf{m}^{(i+1)}, \mathbf{r}^{(i+1)})$ given $\boldsymbol{\tau}_{-(\mathbf{m},\mathbf{r})}^{(i)} \triangleq \{\mathbf{p}^{(i+1)}, \mathbf{a}^{(i+1)}, \mathbf{r}^{(i)}\}$, tagged as AMC-OP.

The algorithm is described in Tab. 6.1, where I_{CAM} is the maximum number of iterations, ϵ is a small constant, identifying the accuracy interval of the solution, and $\boldsymbol{\tau}^*$ is the solution to the MMG-OP (6.6).

In the following sections, each of these three subproblems is optimally solved so that, at each step, a nondecreasing value of the objective function is produced. For this reason, the algorithm ends when it is reaches either the maximum number of iterations I_{CAM} or a local optimum, within the accuracy interval ϵ. In detail, PA-OP is solved in Sect. 6.4, AMC-OP is solved in Sect. 6.5, whereas SA-OP is solved in Sect. 6.6. The algorithm also requires a good initialization point $\boldsymbol{\tau}^{(0)}$, in order to enhance the possibility to reach the global optimum or at least a better local optimum. This issue motivates the search of an effective $\boldsymbol{\tau}^{(0)}$ that has been carried out and described in Sec. 6.7.

Remark. The solution to the OP (6.6) can be an empty set since there may not be a TP setting that simultaneously satisfies all the constraints. In this case, it means

Coordinate ascent method for the MMG-OP

1. Initialize: $i = 0$, $\boldsymbol{\tau} = \boldsymbol{\tau}^{(0)}$, ϵ, I_{CAM}
2. Repeat
3. (PA-OP): Evaluate $\mathbf{p}^{(i+1)} = \underset{\mathbf{p}\in\mathcal{D}_p}{\arg\max}\, f(\mathbf{p}|\boldsymbol{\tau}^{(i)}_{-\mathbf{p}})$
4. (SA-OP): Evaluate $\mathbf{a}^{(i+1)} = \underset{\mathbf{a}\in\mathcal{D}_a}{\arg\max}\, f(\mathbf{a}|\boldsymbol{\tau}^{(i)}_{-\mathbf{a}})$
5. (AMC-OP): Evaluate $(\mathbf{m}^{(i+1)}, \mathbf{r}^{(i+1)}) = \underset{(\mathbf{m},\mathbf{r})\in\mathcal{D}_m\times\mathcal{D}_r}{\arg\max}\, f(\mathbf{m},\mathbf{r}|\boldsymbol{\tau}^{(i)}_{-(\mathbf{m},\mathbf{r})})$
6. $i = i + 1$
7. Until $\|f(\boldsymbol{\tau}^{(i+1)}) - f(\boldsymbol{\tau}^{(i)})\| < \epsilon$ or $i = I_{\text{CAM}}$
8. Return: $\boldsymbol{\tau}^* = \boldsymbol{\tau}^{(i)}$

Table 6.1: *Pseudo-code of the coordinate ascent method for the MMG-OP*

that the system is overloaded and the problem is relaxed dropping the packets not satisfying all the constraints.

6.4 Power allocation optimization problem

In this section, the algorithm that provides the optimal solution $\mathbf{p}^* = [p_1^*, \cdots, p_N^*]^{\text{T}}$ to the PA-OP is derived resorting to the Lagrangian dual decomposition method. Let us start by explicitly describing the PA-OP previously introduced.

PA-OP. Given $\boldsymbol{\tau}_{-\mathbf{p}}$, the MMG-OP (6.6) reduces to the following power allocation problem, or PA-OP for short:

$$\begin{aligned} \max_{\mathbf{p}\in\mathcal{D}_p} \quad & \min_{q\in\mathcal{Q}} \zeta_q(\mathbf{p}) \\ \text{s.t.} \quad & \sum_{n=1}^{N} p_n \leq P, \\ & \hat{\gamma}_q(\mathbf{p}) \geq \overline{\gamma}_q \qquad \forall q, \end{aligned} \tag{6.7}$$

where $\overline{\gamma}_q \triangleq \Psi_{r_q}^{-1}\left(\overline{\Psi}_q\right)$. This optimization problem is solved as stated in the following

proposition.

Theorem 6.1 *The PA-OP (6.7) is a convex optimization problem and the solution can be found solving this equivalent convex optimization problem, tagged as EPA-OP.*

$$\begin{aligned} \textbf{EPA-OP}: \quad \min_{\mathbf{p}\in\mathcal{D}_p,t} \quad & -t \\ \text{s.t.} \quad & \sum_{n=1}^{N} p_n - P \leq 0, && \text{(6.8.a)} \\ & t - \zeta_q(\mathbf{p}) \leq 0 \quad \forall q, && \text{(6.8.b)} \\ & \overline{\gamma}_q - \hat{\gamma}_q(\mathbf{p}) \leq 0 \quad \forall q. && \text{(6.8.c)} \end{aligned} \tag{6.8}$$

Proof Given $\boldsymbol{\tau}_{-\mathbf{p}}$, for the generic user u the EGP can be rewritten as follows:

$$\zeta_q(\mathbf{p}) = c_{1,q}\left[1 - \Psi_{r_q}\left(-\log\left(\frac{1}{c_{2,q}}\sum_{n=1}^{N} a_{q,n}\Omega_{q,n}(p_n)\right)\right)\right], \tag{6.9}$$

where $c_{1,u} \triangleq \zeta_q^{(0)} r_q m_q \sum_{n=1}^{N} a_{q,n}$ and $c_{2,q} \triangleq m_q \sum_{n=1}^{N} a_{q,n}$ are constant terms. Be, for the ease if notation, $g(\mathbf{p}) \triangleq \frac{1}{c_{2,q}} \sum_{n=1}^{N} a_{q,n}\Omega_{q,n}(p_n)$, and $h(\mathbf{p}) \triangleq -\log \circ g(\mathbf{p})$. For the composition rule of convexity [69], since $g(\mathbf{p})$ is convex according to (2.20) and $-\log$ is convex and nonincreasing, then h is convex. For the same rule, since Ψ_{r_q} is convex and nonincreasing, we get that $\Phi = \Psi \circ h$ is convex and thus $\zeta(\mathbf{p})$ is a concave function of the power. Since the constraints in (6.7) define a convex set and the minimum of a concave function is still concave, the maximization of this function over the set of feasible powers is a convex optimization problem. Finally, as shown e.g. in [97], the max-min problem (6.7) can be equivalently rearranged in its epigraph form (6.8). □

The optimal solution $\mathbf{p}^*$ can be found resorting to the Lagrangian duality theory for convex optimization problems [69], briefly recalled in Appendix A.1.1, by solving the dual optimization problem associated to the so-called primal problem (6.8). In fact, since the objective function and all the constraints in (6.8) are differentiable, the KKT conditions hold and thus strong duality holds too, i.e., the difference between the optimal primal and dual values is zero.

Thus, in order to introduce the dual OP, let us first associate the Lagrange multiplier θ to constraint (6.8.a) and the multipliers ω_q and ϕ_q, $\forall q \in \mathcal{Q}$, to constraint (6.8.b) and

(6.8.c), respectively, whose feasible sets are $\mathcal{D}_\theta \triangleq \{\theta|\theta \geq 0,\}$, $\mathcal{D}_\omega \triangleq \{\omega_q|\forall q\ \omega_q \geq 0\}$, $\mathcal{D}_\phi \triangleq \{\phi_q|\forall q\ \phi_q \geq 0\}$. Collecting these values in the vector $\mathbf{\Theta}$, the dual function is defined as

$$\inf_{\mathbf{p}\in\mathcal{D}_p,t} \left\{ \left(\sum_{q\in\mathcal{Q}} \omega_q - 1 \right) t + \mathcal{L}_p(\mathbf{\Theta}, \mathbf{p}) \right\}, \tag{6.10}$$

where $\mathcal{L}_p(\mathbf{\Theta}, \mathbf{p}) \triangleq \theta f_1(\mathbf{p}) - \sum_{q\in\mathcal{Q}} \omega_q \zeta_q(\mathbf{p}) + \sum_{q\in\mathcal{Q}} \phi_u f_{2,q}(\mathbf{p})$ and $f_1(\mathbf{p})$ and $f_{2,q}(\mathbf{p})$ denote, for the sake of readability, the left-hand side of constraints (6.8.a) and (6.8.c), respectively. Since the infimum of a linear function is $-\infty$, unless it is identically zero, it must hold $\sum_{q\in\mathcal{Q}} \omega_q - 1 = 0$, so that the dual function results

$$g_p(\mathbf{\Theta}) = \inf_{\mathbf{p}\in\mathcal{D}_p} \mathcal{L}_p(\mathbf{\Theta}, \mathbf{p}), \tag{6.11}$$

with the set of feasible ω_q that modifies as follows: $\mathcal{D}_\omega = \{\omega_q|\omega_q \geq 0,\ \forall q,\ \sum_{q\in\mathcal{Q}} \omega_q = 1\}$.

Accordingly, the optimal dual value $\mathbf{\Theta}^*$ is found solving the dual optimization problem stated as follows

$$\begin{array}{ll} \max & g_p(\mathbf{\Theta}) \\ \text{s.t.} & \mathbf{\Theta} \in \mathcal{D}_\Theta \end{array}. \tag{6.12}$$

Upon referring to

$$\mathbf{p}(\mathbf{\Theta}) \triangleq \arg\min_{\mathbf{p}\in\mathcal{D}_p} \mathcal{L}_p(\mathbf{\Theta}, \mathbf{p}) \tag{6.13}$$

as the optimal solution to (6.11) for a given multiplier $\mathbf{\Theta}$, the primal optimal solution to the EPA-OP (6.8) is thus found as $\mathbf{p}^* = \mathbf{p}(\mathbf{\Theta}^*)$, thanks to the strong duality of the problem at hand.

Subgradient Method for the Dual OP (6.12).

The dual OP (6.12) can be solved, as usually done in cases [73], [70], resorting to the subgradient-based update of the dual variable $\mathbf{\Theta}$, briefly recalled in Appendix A.1.2 and referring to [74] for further details. This iterative method consists in maximizing $g_p(\mathbf{\Theta})$ by updating, at each step $i+1$, all the components of the dual variable $\mathbf{\Theta}^{(i)}$ produced at the previous step i along the search direction defined by the subgradient of $g_p(\mathbf{\Theta})$ at $\mathbf{\Theta}^{(i)}$. Looking at (6.11) and at the definition of $\mathcal{L}_p$, the latter results $\nabla_{\mathbf{\Theta}} g_p(\mathbf{\Theta}^{(i)}) = [f_1(\mathbf{p}), -\zeta_1(\mathbf{p}), \cdots, -\zeta_Q(\mathbf{p}), f_{2,1}(\mathbf{p}), \cdots, f_{2,Q}(\mathbf{p})]|^{\mathrm{T}}_{\mathbf{p}=\mathbf{p}(\mathbf{\Theta}^{(i)})}$, with $\mathbf{p}(\mathbf{\Theta}^{(i)})$ evaluated via (6.13). Then, the dual variable

update is $\boldsymbol{\Theta}^{(i+1)} = \left[\boldsymbol{\Theta}^{(i)} + \delta_{\boldsymbol{\Theta}} \nabla_{\boldsymbol{\Theta}} g_p(\boldsymbol{\Theta}^{(i)})\right]_{\mathcal{D}_\Theta}$, where $[\cdot]_{\mathcal{D}_\Theta}$ denotes the projection operation over the set of feasible multipliers values $\mathcal{D}_\Theta \triangleq \mathcal{D}_\theta \times \mathcal{D}_\phi \times \mathcal{D}_\omega$ and $\delta_{\boldsymbol{\Theta}}$ is the step size of the subgradient method, chosen sufficiently small to allow the algorithm to converge. The dual method based on subgradient update is summarized in Table 6.2. The projection operation at Line 4 $[\mathbf{x}]_{\mathcal{D}_\Theta}$ simply reduces to $x_i = [x_i]^+$, $\forall\ x_i$

EPA-OP algorithm
1. Initialize: $i = 0$, $\boldsymbol{\Theta} = \boldsymbol{\Theta}^{(0)}$, δ_Θ, ϵ, $i_{\max}$
2. Repeat
3. $\quad \mathbf{p}(\boldsymbol{\Theta}^{(i)}) = \arg\min_{\mathbf{p}\in\mathcal{D}_p} \mathcal{L}_p(\mathbf{p}, \boldsymbol{\Theta}^{(i)})$
4. $\quad \boldsymbol{\Theta}^{(i+1)} = \left[\boldsymbol{\Theta}^{(i)} + \delta_\Theta \nabla_{\boldsymbol{\Theta}} g_p(\boldsymbol{\Theta}^{(i)})\right]_{\mathcal{D}_\Theta}$
5. $\quad i = i + 1$
6. Until $\left\|\mathbf{p}(\boldsymbol{\Theta}^{(i+1)}) - \mathbf{p}(\boldsymbol{\Theta}^{(i)})\right\| \leq \epsilon$ or $i = i_{\max}$
7. Return $\mathbf{p}^* = \mathbf{p}(\boldsymbol{\Theta}^{(i)})$

Table 6.2: *Pseudo-code of EPA-OP algorithm*

projected over the sets $\mathcal{D}_\theta$ or $\mathcal{D}_\phi$, whereas it is equivalent to $x_i = [x_i + \epsilon_\omega]^+$, where ϵ_ω is a scalar such that $\sum_i [x_i + \epsilon_\omega]^+ = 1$, $\forall\ x_i$ projected over the set $\mathcal{D}_\omega$. The scalar ϵ_ω is found solving the last equality with the bisection method. The solution to the convex optimization problem at Line 3 can be carried out via conventional optimization algorithms like the steepest descent method or the ellipsoid method [69].

6.5 AMC Algorithm

The AMC-OP that returns the optimal pair of bits and coding rates vectors $(\mathbf{m}^*, \mathbf{r}^*)$, with $\mathbf{m}^* = [m_1^*, \cdots, m_Q^*]^{\mathrm{T}}$ and $\mathbf{r}^* = [r_1^*, \cdots, r_Q^*]^{\mathrm{T}}$, can be stated as follows.

AMC-OP. Given $\boldsymbol{\tau}_{-(\mathbf{m},\mathbf{r})}$, the MMG-OP (6.6) reduces to the following adaptive

modulation and coding problem, or AMC-OP for short:

$$\begin{array}{ll} \max\limits_{(\mathbf{m}_q,\mathbf{r}_q)\in\mathcal{D}_m\times\mathcal{D}_r,\ \forall q} & \min\limits_{q\in\mathcal{Q}} \zeta_q(m_q,r_q) \\ \text{s.t.} & \Psi_{r_q}\left(\hat{\gamma}_q(m_q)\right) \leq \overline{\Psi}_q \qquad \forall q. \end{array} \tag{6.14}$$

This problem can be solved through a simple iterative method due to the discrete nature of the sets $\mathcal{D}_m$ and $\mathcal{D}_r$. In fact, it is sufficient to select the user u with the lowest goodput value and perform exhaustive search over the set $\mathcal{D}_m \times \mathcal{D}_r$ to check if there is a pair (m, r) that increases the goodput of u satisfying the QoS constraint. Then, a new control on the goodput values is done to see if there is a new user with the lowest goodput value. In this case, the same search is performed and the algorithm stops when the user with the lowest goodput cannot improve its performance anymore. This algorithm is summarized in Tab. 6.3, where $\mathcal{D}(l)$ denotes the lth element of the set $\mathcal{D}$.

AMC-OP algorithm
1. Initialize: $k = 1$, $u^{(0)} = 0$ and $(\mathbf{m}, \mathbf{r}) = (\mathbf{m}_0, \mathbf{r}_0)$
2. Evaluate $\zeta_q(m_q, r_q)$, $\forall q$
3. Set $u^{(k)} = \arg\min_q \{\zeta_q\}_{q\in\mathcal{Q}}$
4. If $u^{(k)} \neq u^{(k-1)}$
5. For $i = 1 : \lvert\mathcal{D}_m\rvert$
6. For $j = 1 : \lvert\mathcal{D}_r\rvert$
7. If $\zeta_{u^{(k)}}(\mathcal{D}_m(i), \mathcal{D}_r(j)) > \zeta_{u^{(k)}}(m_{u^{(k)}}, r_{u^{(k)}})$ and $\Psi_{\mathcal{D}_r(j)}(\hat{\gamma}_{u^{(k)}}(\mathcal{D}_m(i))) \leq \overline{\Psi}_{u^{(k)}}$
8. Set $m_{u^{(k)}} = \mathcal{D}_m(i)$, $r_{u^{(k)}} = \mathcal{D}_r(j)$
9. Set $k = k + 1$ and Go to Step 2
10. End If
11. End For
12. End For
13. End If
14. Return $(\mathbf{m}^*, \mathbf{r}^*) = (\mathbf{m}, \mathbf{r})$

Table 6.3: *Pseudo-code of AMC-OP algorithm*

6.6 Subcarrier allocation problem

The subcarrier allocation problem SA-OP introduced in Sect. 6.3.2 can be formalized as follows.

SA-OP. Given $\boldsymbol{\tau}_{-\mathbf{a}}$, the SA problem the maximizes the performance of the user with the minimum EGP can be formalized as:

$$\begin{aligned} \max_{\mathbf{a}_q \in \mathcal{D}_a,\, \forall q} \quad & \min_{q \in \mathcal{Q}} \zeta_q(\mathbf{a}_q) & & \\ \text{s.t.} \quad & \sum_{q \in \mathcal{Q}} a_{q,n} \leq 1 & \forall n, \quad & \text{(6.15.a)} \\ & \hat{\gamma}_q(\mathbf{a}_q) \geq \overline{\gamma}_q & \forall q. \quad & \text{(6.15.b)} \end{aligned} \tag{6.15}$$

This optimization problem is a NP-hard problem, in particular it is a nonlinear integer programming. In order to solve this problem, first a relaxation of the subcarrier allocation variable is carried out and it is demonstrated that such a relaxed version is a convex optimization problem. Then, an algorithm to evaluate its optimal solution will be derived. In fact, the relaxed optimization problem does not converge to a solution in the discrete domain $\mathcal{D}_a$, nevertheless it represents an upper bound of the original problem, since we are optimizing over a lager continuous set, i.e., $\mathbf{a} \in \mathcal{D}_a^{(c)} = \{a_{q,n} | a_{q,n} \in [0,1],\ \forall q, n\}$, that contains the original set $\mathcal{D}_a$. Capitalizing on this observation, the discrete solution can be found resorting to the branch and bound approach [100], where a sequence of upper and lower bounds of the original problem are generated until their difference is within the required accuracy interval. Thus, in the remainder of this section it is first described the optimal solution solving the relaxed SA problem and refining the obtained solution with a branch and bound method. Then, since the complexity of this approach can be high, a practical algorithm, based on the Ant Colony Optimization framework, is also proposed.

6.6.1 Branch and bound based SA-OP

In this section, the relaxed version of the SA-OP (6.15) is first analyzed. Then, the optimal solution $\mathbf{a}_q^* = [a_{q,1}^*, \cdots, a_{q,N}^*]^{\mathrm{T}}$, $\forall q \in \mathcal{Q}$, is found via branch and bound, mapping the continuous solution back to the original discrete domain.

Relaxed SA-OP

Let us start by rewriting $\zeta_q(\mathbf{a}_q)$, with $\mathbf{a}_q \in \mathcal{D}_a^{(c)}$, $\forall q \in \mathcal{Q}$, pointing out the dependence on $\mathbf{a}_q$

$$\zeta_q(\mathbf{a}_q) = \varsigma_q \left(\sum_{n=1}^{N} a_{q,n} \right) \left[1 - \widehat{\psi}_q \left(\frac{\sum_{n=1}^{N} a_{q,n} \overline{\Omega}_{q,n}}{\sum_{i=1}^{N} a_{q,i}} \right) \right], \tag{6.16}$$

where $\varsigma_q \triangleq \zeta_q^{(0)} r_q m_q$, $\overline{\Omega}_{q,n} \triangleq \Omega_{q,n}/m_q$, $\forall q, n$, are constant terms w.r.t. to $\mathbf{a}_q$, and $\widehat{\psi}_q \triangleq \Psi_{r_q} \circ -\log$.

It is worth noting that the presence of vector $\mathbf{a}_q$ in (6.16) is twofold, i.e., both as a simple unweighed sum and as a weighted sum, by $\overline{\Omega}_{q,n}$, of its components. This observation paves the way to the main result of this section, outlined in the following proposition.

Theorem 6.2 *Defining the slack variables $s_q > 0$, $\forall q$, it holds*

$$\tilde{\zeta}_q(s_q, \mathbf{a}_q) \triangleq \varsigma_q s_q \left[1 - \widehat{\psi}_q \left(\frac{\sum_{n=1}^{N} a_{q,n} \overline{\Omega}_{q,n}}{s_q} \right) \right] \leq \zeta_q(\mathbf{a}_q), \tag{6.17}$$

whenever

$$s_q \leq \sum_{n=1}^{N} a_{q,n}, \ \forall q \tag{6.18}$$

and these properties follows:

P1) relationship (6.17) under constraint (6.18) denotes a goodput lower bound;

P2) strict equality holds in (6.17) when $s_q = \sum_{n=1}^{N} a_{q,n}$;

P3) $\tilde{\zeta}_q(s_q, \mathbf{a}_q)$ is jointly concave w.r.t. $(s_q, \mathbf{a}_q)$.

Proof For a given $\mathbf{a}_q$, $\tilde{\zeta}_q(s_q)$ is a monotone nondecreasing function of s_q. In fact, with reference to (6.17), be, for the ease of notation and neglecting w.l.g. the constant terms, $g(s_q) \triangleq -\log(1/s_q)$, and note that, since $s_q > 0$, it holds $g'(s_q) > 0$. Being $\widehat{\psi}_q \triangleq \Psi_{r_q} \circ g$, with $0 \leq \widehat{\psi}_q \leq 1$, and recalling that $\Psi'_{r_q} < 0$, we have that $\widehat{\psi}'_q = \Psi'_{r_q}(g(s_q))\, g'(s_q) < 0$. Hence $\tilde{\zeta}_q(s_q) = \varsigma_q s_q [1 - \widehat{\psi}_q(s_q)]$ is a monotone nondecreasing function of s_q since $\tilde{\zeta}'_q(s_q) = (1 - \widehat{\psi}_q(s_q)) - s_q \widehat{\psi}'_q \geq 0$. Thus, relying on this property and on constraint (6.18), properties P1 and P2 follow.

In order to prove property P3, let us note that $\overline{\psi}_q(\mathbf{a}_q) \triangleq 1 - \widehat{\psi}_q\left(\sum_{n=1}^{N} \overline{\Omega}_{q,n} a_{q,n}\right)$ is concave in $\mathbf{a}_q$. In fact, defining the convex function $h(\mathbf{a}_q) \triangleq -\log(\sum_{n=1}^{N} a_{q,n}\overline{\Omega}_{q,n})$, according to the properties of Ψ_{r_q} and to the rule of composition of convex function [69], we have that $\widehat{\psi}_q(\mathbf{a}_q) = \Psi_{r_q} \circ h$ is a convex function. Therefore it easily follows that both $\overline{\psi}_q(\mathbf{a}_q)$ and $\tilde{\zeta}_q(\mathbf{a}_q)$ are concave functions of $\mathbf{a}_q$. Let us also note that

$$\tilde{\zeta}_q(s_q, \mathbf{a}_q) = \varsigma_q s_q \overline{\psi}_q\left(\frac{\mathbf{a}_q}{s_q}\right). \tag{6.19}$$

Since $s_q > 0$, then $\tilde{\zeta}_q(s_q, \mathbf{a}_q)$ corresponds to the perspective of the function $\overline{\psi}_q$. Since the perspective of a concave function is still concave [69], this proves P3. □

Thus, due to property P3, the relaxed subcarrier allocation problem can be optimally solved maximizing the minimum of (6.19), and, thanks to properties P1-P2, this solution coincides with the one that maximizes the minimum of (6.16). This problem can be summarized as follows, tagged as relaxed SA-OP (RSA-OP).

$$\begin{aligned}
\textbf{RSA-OP}: \quad \max_{\mathbf{a}\in\mathcal{D}_a^{(c)},\, \mathbf{s}} \quad & \min_{q\in\mathcal{Q}} \tilde{\zeta}_q(\mathbf{a}_q, s_q) & & \\
\text{s.t.} \quad & \sum_{q\in\mathcal{Q}} a_{q,n} \leq 1 & \forall n, \quad & \text{(6.20.a)} \\
& \sum_{n=1}^{N} \bar{\Omega}_{q,n} a_{q,n} \leq e^{-\overline{\gamma}_q} s_q & \forall q, \quad & \text{(6.20.b)} \\
& \sum_{n=1}^{N} a_{q,n} \geq s_q & \forall q, \quad & \text{(6.20.c)}
\end{aligned} \tag{6.20}$$

where $\mathbf{s} \triangleq [s_1, \cdots, s_Q]^{\mathrm{T}}$ and constraint (6.20.b) is the QoS constraint (6.6.c) expressed as a function of $\mathbf{a}_q$ and s_q. The optimal solution to problem (6.20) can be found via the coordinate ascent method. In fact, since the objective function is jointly concave w.r.t. to $(\mathbf{a}_q, s_q)$, the iterations between the two concave subproblems lead to the global optimum of the function. The algorithm is summarized in Table 6.4.

The solution to RSA-OP.a is trivial, in fact, as previously demonstrated, since $\tilde{\zeta}_q(s_q)$ is monotonically increasing with s_q, the optimal solution is $s_q^{(j+1)} = \sum_{n=1}^{N} a_{q,n}^{(j)}$. Thus, the complexity of the RSA-OP (6.20) only lies in the optimization problem RSA-OP.b. Here, it is possible to exploit the relation, as in the PA case, between the primal and the dual function to find its solution. The algorithm is summarized in Tab. 6.5, and, since it is analogous to the one of the PA case, it is briefly described in the following.

Coordinate ascent method for the RSA-OP

1. Initialize: set $j = 0$, $\mathbf{a}_q = \mathbf{a}_q^{(0)}$, $\forall u$, ϵ, J
2. Repeat
3. (RSA-OP.a): Compute $s_q^{(j+1)} = \max_{s_q} \tilde{\zeta}_q(s_q, \mathbf{a}_q^{(j)})$ s.t. (6.20.b)-(6.20.c)
4. (RSA-OP.b): Compute $\mathbf{a}_q^{(j+1)} = \max_{\mathbf{a}_q} \tilde{\zeta}_u(\mathbf{a}_q, s_q^{(j+1)})$ s.t. (6.20.a)-(6.20.c)
5. Until $||\mathbf{s}^{(j+1)} - \mathbf{s}^{(j)}|| < \epsilon$ or $j = J$
6. Output: $\mathbf{a}_q^* = \mathbf{a}_q^{(j)}$, $\forall q$.

Table 6.4: *Pseudo-code of RSA-OP algorithm*

LDD Method for RSA-OP.b

Problem RSA-OP.b can be written as

$$\begin{aligned} \min_{\{\mathbf{a}_q\}_{q\in\mathcal{Q}},v} \quad & -v \\ \text{s.t.} \quad & \sum_{q\in\mathcal{Q}} a_{q,n} - 1 \leq 0 \qquad \forall n, \quad (6.21.a) \\ & s_q - \sum_{n=1}^{N} a_{q,n} \leq 0 \qquad \forall q, \quad (6.21.b) \\ & \sum_{n=1}^{N} \overline{\mathcal{M}}_{q,n} a_{q,n} - e^{-\overline{\gamma}_q} s_q \leq 0 \quad \forall q, \quad (6.21.c) \\ & v - \tilde{\zeta}_q(\mathbf{a}_q) \leq 0 \qquad \forall q, \quad (6.21.d) \end{aligned} \tag{6.21}$$

Defining $h_{1,q}(a_{q,n}) \triangleq (a_{q,n} - 1/Q)$, and denoting with $h_{2,q}(\mathbf{a}_q)$, and $h_{3,q}(\mathbf{a}_q)$ the left-hand side of constraints (6.21.b)-(6.21.c), respectively, the dual function results

$$\inf_{\mathbf{a}\in\mathcal{D}_a^{(c)},v} \left(\sum_{q\in\mathcal{Q}} \mu_q - 1 \right) v + \sum_{q\in\mathcal{Q}} \mathcal{L}_q^{(a)}(\mathbf{a}, \boldsymbol{\Xi}) \tag{6.22}$$

with $\mathcal{L}_q^{(a)}(\mathbf{a}_q, \boldsymbol{\Xi}) \triangleq \sum_{n=1}^{N} \psi_n h_{1,q}(a_{q,n}) + \nu_q h_{2,q}(\mathbf{a}_q) + \vartheta_q h_{3,q}(\mathbf{a}_q) - \mu_q \tilde{\zeta}_q(\mathbf{a}_q)$ and $\boldsymbol{\Xi}$ the vector collection all the Lagrangians values. According to the considerations done for the power allocation, $\sum_{q\in\mathcal{Q}} \mu_q - 1 = 0$ otherwise (6.22) is unbounded below. Thus, the dual function can be written as

$$g_s(\boldsymbol{\Xi}) = \inf_{\mathbf{a}\in\mathcal{D}_a^{(c)}} \sum\nolimits_{q\in\mathcal{Q}} \mathcal{L}_q^{(a)}(\mathbf{a}_q, \boldsymbol{\Xi}) \tag{6.23}$$

RSA-OP.b algorithm

1. Initialize: $i = 0$, $\boldsymbol{\Xi} = \boldsymbol{\Xi}^{(0)}$, ϵ_{sa}, $i_{\max}$
2. Repeat
3. $\quad \mathbf{a}_q(\boldsymbol{\Xi}^{(i)}) = \underset{\mathbf{a}_q \in \mathcal{D}_a}{\arg\min}\, \mathcal{L}_q^{(a)}(\mathbf{a}_q, \boldsymbol{\Xi}^{(i)}),\ \forall q$
4. $\quad \boldsymbol{\Xi}^{(i+1)} = \left[\boldsymbol{\Xi}^{(i)} + \delta_{\Xi} \nabla_{\boldsymbol{\Xi}} g_s(\boldsymbol{\Xi}^{(i)})\right]_{\mathcal{D}_{\Xi}}$
5. $\quad i = i + 1$
6. Until $i = i_{\max}$ or $\left\|\mathbf{a}(\boldsymbol{\Xi}^{(i+1)}) - \mathbf{a}(\boldsymbol{\Xi}^{(i)})\right\| \leq \epsilon_{\text{sa}}$
7. Return $\mathbf{a}^* = \mathbf{a}(\boldsymbol{\Xi}^{(i)})$

Table 6.5: *Pseudo-code of RSA-OP.b algorithm*

where $\boldsymbol{\Xi} \in \mathcal{D}_{\Xi} = \mathcal{D}_{\psi} \times \mathcal{D}_{\vartheta} \times \mathcal{D}_{\nu} \times \mathcal{D}_{\mu}$, being $\mathcal{D}_{\psi} \triangleq \{\psi_n | \psi_n \geq 0\ \forall n\}$, $\mathcal{D}_{\nu} \triangleq \{\nu_q | \nu_q \geq 0\ \forall q\}$, $\mathcal{D}_{\vartheta} \triangleq \{\vartheta_q | \vartheta_q \geq 0\ \forall q\}$ and $\mathcal{D}_{\mu} \triangleq \left\{ \mu_q | \mu_q \geq 0\ \forall q \text{ and } \sum_{q \in \mathcal{Q}} \mu_q = 1 \right\}$. The solution $\mathbf{a}^*$ to (6.21) is found first in the dual domain solving the dual OP, i.e., $\boldsymbol{\Xi}^* = \underset{\boldsymbol{\Xi} \in \mathcal{D}_{\Xi}}{\arg\max}\, g_s(\boldsymbol{\Xi})$, and then evaluating $\mathbf{a}^* = \mathbf{a}(\boldsymbol{\Xi}^*)$, where $\mathbf{a}(\boldsymbol{\Xi}) = \arg\min_{\mathbf{a} \in \mathcal{D}_a^{(c)}} \sum_q \mathcal{L}_q^{(a)}(\mathbf{a}, \boldsymbol{\Xi})$. The maximization of the dual function is obtained via the subgradient update algorithm and it is reported in Tab. 6.5, where, in the projected subgradient update of the Lagrange multipliers at Line 4, the projection operator $[\cdot]_{\mathcal{D}_{\Xi}}$ is such that $x_i = [x_i]^+$, $\forall x_i$ projected over $\mathcal{D}_{\psi}$, $\mathcal{D}_{\vartheta}$ and $\mathcal{D}_{\nu}$, whereas it is equivalent to $\sum_i [x_i + \epsilon_{\mu}]^+ = 1$, $\forall x_i$ projected over $\mathcal{D}_{\mu}$, with ϵ_{μ} evaluated via bisection method. The parameter δ_{Ξ} is a sufficiently small value to ensure the convergence of the algorithm. Still at line 4, $\nabla_{\boldsymbol{\Xi}} g_s(\boldsymbol{\Xi}^{(i)})$ denotes at subgradient of g_s at $\boldsymbol{\Xi} = \boldsymbol{\Xi}^{(i)}$. Finally, at line 3, the optimization problem can be solved resorting to conventional optimization algorithms like, for instance, the steepest descent method or the ellipsoid methods [69].

Branch ad bound refinement step

In order to find a feasible solution to the SA-OP (6.15) from the relaxed solution of the RSA-OP (6.20), a branch and bound (B&B) algorithm is now introduced. The B&B is an iterative algorithm that builds a tree whose levels are composed by a certain

number of nodes. At each node, an upper bound, ζ_{ub}, and a lower bound, ζ_{lb}, of the optimal value $\zeta^*(\mathbf{a}^*)$ of the original SA-OP (6.15) are evaluated and the algorithm ends when, at a certain node, the difference between these two values is within the required accuracy interval. In detail, each node corresponds to a subset $\mathcal{A} \subset \mathcal{D}_a^{(c)}$, where some variable values are fixed to 0 or 1 and the others assume continuous values in $[0, 1]$. To show how the B&B algorithm works, some useful sets must now be defined. Let denote with $\mathcal{B}_k$ the set collecting the sets associated to each node of the graph at step k. Be $\mathcal{I}_{k,\mathcal{A}}$ the set of the pairs of indexes (q, n) identifying the subcarrier indicators $a_{q,n}$ that have already been fixed to 0 or 1 in the set $\mathcal{A}$ of step k. The algorithm is initialized with $\mathcal{B}_0 = \{\mathcal{D}_a^{(c)}\}$ and $\mathcal{I}_{0,\mathcal{B}_0} = \{\emptyset\}$, i.e., at the beginning we have only one node where all the subcarriers indexes are variable and can assume continuous values in $\mathcal{D}_a^{(c)}$. Then, the algorithm starts and, at each step k, consists of three phases.

i) In the first, the branching phase, the algorithm chooses a set $\mathcal{A} \in \mathcal{B}_k$ associated to a node and generates two child nodes, whose relevant sets $\mathcal{A}_1$ and $\mathcal{A}_2$ are built as follows. A variable $a_{q,n} \in \mathcal{A}$ such that $(q, n) \notin \mathcal{I}_{k,\mathcal{A}}$ is picked and the two child nodes sets are built as $\mathcal{A}_1 = \mathcal{A} \cap \{a_{q,n} = 1, a_{\nu,n} = 1 \; \forall \nu \neq q\}$ and $\mathcal{A}_2 = \mathcal{A} \cap \{a_{q,n} = 0\}$, with $\mathcal{I}_{k,\mathcal{A}_1} = \mathcal{I}_{k,\mathcal{A}_1} \cup \{(\nu, n)\}_{\nu=1}^{Q}$ and $\mathcal{I}_{k,\mathcal{A}_2} = \mathcal{I}_{k,\mathcal{A}_2} \cup \{(q, n)\}$. The set $\mathcal{B}_{k+1}$ is obtained removing from $\mathcal{B}_k$ the set $\mathcal{A}$ and adding $\{\mathcal{A}_1, \mathcal{A}_2\}$.

ii) The bounding phase follows, where the upper and lower bounds at both the child nodes are evaluated and the best bounds found so far are computed as $U_{k+1} = \min\{U_k, \zeta_{\text{ub}}(\mathcal{A}_1), \zeta_{\text{ub}}(\mathcal{A}_2)\}$, $L_{k+1} = \min\{L_k, \zeta_{\text{lb}}(\mathcal{A}_1), \zeta_{\text{lb}}(\mathcal{A}_2)\}$, where $\zeta_{\text{ub}}(\mathcal{A})$ and $\zeta_{\text{lb}}(\mathcal{A})$ denotes the upper bound and a lower bound, respectively, on the solution of the original optimization problem (6.15) when $\mathbf{a}$ is constrained to assume values in $\mathcal{A}$. Soon, there will be given details on how to evaluate these bounds.

iii) At last, in the pruning phase, all the nodes in $\mathcal{B}_{k+1}$, whose upper bound is lower than the best known lower bound L_{k+1}, are pruned, that is, are removed from $\mathcal{B}$. Finally, if $U_{k+1} - L_{k+1} < \epsilon$ the algorithm stops, otherwise it starts again from the branching phase.

The B&B algorithm is summarized as follows.

1) Initialize $k = 0$, $\mathcal{B}^{(0)} = \{\mathcal{D}_a^{(c)}\}$ and evaluate $U_0 = \zeta_{\text{ub}}(\mathcal{D}_a^{(c)})$ and $L_0 = \zeta_{\text{lb}}(\mathcal{D}_a^{(c)})$;

2) If $U_k - L_k < \epsilon$ stop, otherwise go to Step 3;

3) Branching Phase:

a) be $\mathcal{A} \in \mathcal{B}_k$ the set with the lowest associated upper bound,
b) pick $a_{q,n} \in \mathcal{A}$ such that $(q,n) \notin \mathcal{I}_{k,\mathcal{A}}$,
c) generate the child nodes $\mathcal{A}_1$ and $\mathcal{A}_2$ and update $\mathcal{I}_{k,\mathcal{A}_1}$ and $\mathcal{I}_{k,\mathcal{A}_2}$,
d) update $\mathcal{B}_{k+1} = \mathcal{B}_k \backslash \{\mathcal{A}\} \cup \{\mathcal{A}_1, \mathcal{A}_2\}$;

4) Bounding Phase:
a) evaluate $\zeta_{\text{ub}}(\mathcal{A}_1)$, $\zeta_{\text{ub}}(\mathcal{A}_2)$, $\zeta_{\text{lb}}(\mathcal{A}_1)$ and $\zeta_{\text{lb}}(\mathcal{A}_2)$
b) set $U_{k+1} = \min\{U_k, \zeta_{\text{ub}}(\mathcal{A}_1), \zeta_{\text{ub}}(\mathcal{A}_2)\}$ and $L_{k+1} = \min\{L_k, \zeta_{\text{lb}}(\mathcal{A}_1), \zeta_{\text{lb}}(\mathcal{A}_2)\}$;

5) Pruning Phase:
$\forall \mathcal{A} \in \mathcal{B}_{k+1}$ such that $\zeta_{\text{ub}}(\mathcal{A}) \leq L_{k+1}$, update $\mathcal{B}_{k+1} = \mathcal{B}_{k+1} \backslash \{\mathcal{A}\}$;

6) Set $k = k + 1$ and go to Step 2.

Some remarks are now in order.

- Thanks to the orthogonality constraint, whenever we generate the child node with $a_{q,n} = 1$, we can already set $a_{\nu,n} = 0$, $\forall \nu \neq q$, since it is the only way to satisfy condition (6.15.a).

- The upper bound $\zeta_{\text{ub}}(\mathcal{A})$ is obtained solving the RSA-OP (6.20), with $\mathbf{a} \in \mathcal{A}$. That is, we solve the relaxed SA problem with, possibly, some variable fixed to 0 or 1. In practice, when the two child nodes are generated, we obtain two subproblems, in the sense that they are the same optimization problem tackled at the parent node but with some variables fixed.

- The lower bound $\zeta_{\text{lb}}(\mathcal{A})$ can be obtained rounding each element of the solution of the upper bound problem to the nearest integer [100].

- At Step 3.b of the B&B algorithm, the variable $a_{q,n} \in \mathcal{A}$ chosen is the most ambivalent [100]. In fact, the solution $\zeta_{\text{ub}}(\mathcal{A})$ produce, neglecting the variables whose value is already fixed to 0 or 1, variables with values in $[0, 1]$. Among these, since we are interested to find a discrete solution, intuitively the most critic are those whose value is close to 1/2. Thus, in order to dispel this doubt, the most ambivalent is chosen, i.e., the one for which its value is the closest one to 1/2.

- The solution returned corresponds to the one giving the lower bound L_{k+1} such that $L_{k+1} \leq \zeta^* \leq U_{k+1}$, with $U_{k+1} - L_{k+1} \leq \epsilon$.

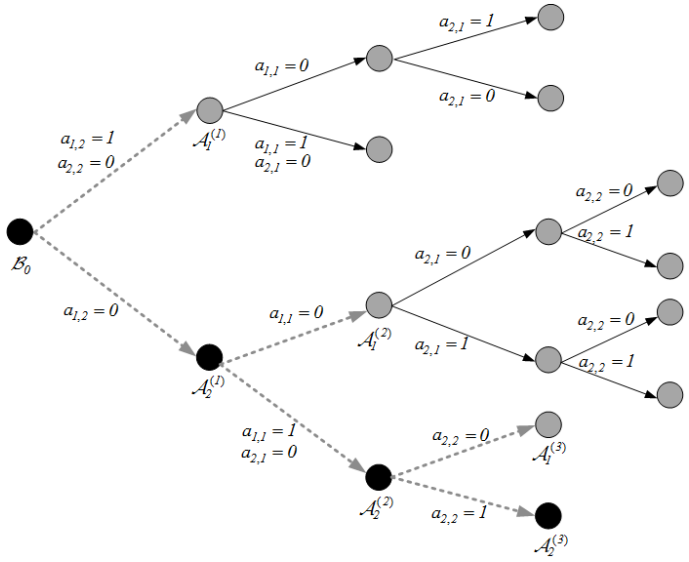

Figure 6.3: *Branch and bound toy example.*

Figure 6.3 depicts a toy example representing both the whole tree produced by an exhaustive search and the path built by the B&B algorithm, for $Q = 2$ users and $N = 2$ subcarriers. Suppose that the optimal solution is identified by the SA indexes connecting the black nodes. The path produced by the B&B algorithm is marked with the tick dashed line, and it is generated as follows. At step $k = 0$, we have $\mathcal{B}_0 = \mathcal{D}_a^{(c)}$ and $\mathcal{I}_{0,\mathcal{B}_0} = \{\emptyset\}$. At the first step, the variable $a_{1,2}$ is picked selecting $(1,2) \notin \mathcal{I}_{0,\mathcal{B}_0}$, so that two child nodes $\mathcal{A}_1^{(1)}$ and $\mathcal{A}_2^{(1)}$ are produced, with index sets $\mathcal{I}_{1,\mathcal{A}_1^{(1)}} = \{(1,2),(2,2)\}$ and $\mathcal{I}_{1,\mathcal{A}_2^{(1)}} = \{(1,2)\}$, respectively, and the set $\mathcal{B}_1 = \{\mathcal{A}_1^{(1)}, \mathcal{A}_2^{(1)}\}$ is built. After evaluating the upper and lower bounds on the EGP at each node, at step $k = 2$ the set $\mathcal{A}_2^{(1)}$ is picked since characterized by the lowest upper bound. Accordingly, the two child nodes $\mathcal{A}_1^{(2)}$ and $\mathcal{A}_2^{(2)}$, with index sets $\mathcal{I}_{1,\mathcal{A}_1^{(2)}} = \{(1,2),(1,1)\}$ and $\mathcal{I}_{1,\mathcal{A}_2^{(2)}} = \{(1,2),(1,1),(2,1)\}$, are produced selecting $(1,1) \notin \mathcal{I}_{1,\mathcal{A}_2^{(1)}}$ and the set $\mathcal{B}_2 = \{\mathcal{A}_1^{(1)}, \mathcal{A}_1^{(2)}, \mathcal{A}_2^{(2)}\}$ is built. Following the same line of reasoning, at step $k = 3$ the set $\mathcal{A}_2^{(2)}$ is picked from $\mathcal{B}_2$, the child nodes $\mathcal{A}_1^{(3)}$ and $\mathcal{A}_2^{(3)}$ are generated selecting the variable $a_{2,2}$ as well as the index sets $\mathcal{I}_{1,\mathcal{A}_1^{(3)}} = \mathcal{I}_{1,\mathcal{A}_2^{(3)}} = \{(1,2),(1,1),(2,1),(2,2)\}$. Here, the algorithm stops returning the optimal SA.

6.6.2 ACO based SA-OP

Rationale of the ACO framework

The ant colony optimization (ACO) is meta-heuristic framework that offers a valid set of algorithms to solve NP-hard combinatorial optimization problems [101]. It belongs to the swarm intelligence algorithms branch [102], which constitutes a relative new approach to problem solving taking inspiration from the social behaviors of insects and of other animals. The ACO framework, in particular, is inspired by the ants foraging behavior and relies on a reinforcement learning mechanism. The ants, in fact, deposit a substance called pheromone on the ground in order to mark some favorable path. Other ants perceive the presence of pheromone and tend to follow paths where its concentration is higher. Through this mechanism, ants are able to transport food to their nest in a remarkably effective way.

Deneubourg et al. [103] thoroughly investigated this behavior in an experiment known as the double bridge experiment. In this experiment, the nest of a colony of Argentine ants is connected to a food source by two bridges of equal lengths. At the beginning, ants start to explore the surroundings of the nest and eventually reach the food source. Along their path between food source and nest, ants deposit pheromone. Initially, each ant randomly chooses one of the two bridges. However, due to random fluctuations, after some time one of the two bridges presents a higher concentration of pheromone than the other and, therefore, attracts more ants. This brings a further amount of pheromone on that bridge making it more attractive with the result that after some time the whole colony converges toward the use of the same bridge. Goss et al. [104] considered a variant of the double bridge experiment, in which one bridge is significantly longer than the other. In this case, the stochastic fluctuations in the initial choice of a bridge are much reduced and a second mechanism plays a key role: ants randomly choosing the short bridge are the first to reach the nest. The short bridge receives, therefore, pheromone earlier than the long one and this fact increases the probability that further ants select it rather than the long one.

The model proposed by Deneubourg and co-workers for explaining the foraging behavior of ants was the main source of inspiration for the development of the ACO framework [101]. In ACO, the optimization problem is coded over a graph where a number of artificial ants build solutions to the problem at hand and exchange information on the quality of these solutions via a communication scheme that is

similar to the above mentioned foraging behavior of real ants. The ACO framework is summarized in the following.

Ant Colony Optimization Algorithm Survey

Ant Colony Optimization (ACO) framework [101] efficiently solves a combinatorial optimization problem (COP), described by the model COP$\triangleq \{\mathcal{S}, \Omega, f\}$, where:

1. $\mathcal{S}$ denotes the discrete search space over V variables x_i that assume values in $\mathcal{V}_i \triangleq \{v_i^{(1)}, \cdots, v_i^{(|\mathcal{V}_i|)}\}$, with $i = 1, \cdots, V$;
2. Ω refers to the set of constraints among the V decision variables;
3. $f : \mathcal{S} \to \mathcal{R}^+$ is the objective function.

A feasible solution $t \in \mathcal{S}$ to COP is a complete assignment $x_i = v_i^{(j)}$ of values to each variable satisfying all the constraints in Ω.

ACO find a solution describing the COP on a graph $G(V, E)$, made of V vertices and E edges. A feasible solution to COP is a complete path on the graph: *i*) all the vertices are connected; *ii*) each vertex is visited only once.

The solution is found through N_{it} iterations, each composed of N_a agents independently exploring the graph. Each agent a: *i*) builds a feasible solution starting from $t^{(a)} = \{\emptyset\}$ and randomly selects the initial vertex; *ii*) selects with probability $\pi_{i,j}$ the edge $e_{i,j} \in \mathcal{N}_e$, (being $\mathcal{N}_e$ the set of edges connecting to the vertices not visited yet), moving from vertex i to vertex j; *iii*) updates $t^{(a)} = t^{(a)} \leftarrow \{e_{i,j}\}$ and $\mathcal{N}_e \leftarrow \mathcal{N}_e \backslash e_{i,j}$, and continues until a complete path is obtained. The probability $\pi_{i,j}$ depends on two quantities: the local desirability $\eta_{i,j}$ and the pheromone $\varphi_{i,j}$. The former is a quantity depending on the particular edge, whereas the latter depends on the quality of the solution to which that edge contributed. At the end of each iteration, the pheromone over all the edges evaporates with a rate $1 - \rho$ and the agent that produced the best solution t_{best} increases the pheromone of a value $\Delta\varphi$ over all the edges $e_{i,j} \in t_{\text{best}}$, so that $\varphi_{i,j} = (1-\rho)\varphi_{i,j} + \Delta\varphi$ if $e_{i,j} \in t_{\text{best}}$, $\varphi_{i,j} = (1-\rho)\varphi_{i,j}$ otherwise. At the end of the N_{it} iterations, a stable path emerges on the graph, corresponding to the solution of the COP.

Solution of SA-OP

In order to solve the SA-OP (6.15) via an efficient algorithm based on the ACO

framework, let us first model the PER Ψ_r as

$$\Psi_r(\gamma) = e^{-\sigma_r(\gamma-\gamma_{0,r})} \tag{6.24}$$

being $\gamma_{0,r}$ and σ_r proper constant values obtained minimizing the relative quadratic error of the fitting between the simulated values of the PER and the negative exponential function in the region of interest (whose validity will be shown in the simulation results section). To this end, it can be noted that, substituting (6.24) into (6.3) and rounding σ_{r_q} with its nearest lower integer, $\bar{\sigma}_{r_q} \triangleq \lfloor \sigma_{r_q} \rfloor$, after some algebra the EGP can be written as[2]

$$\zeta_q(\mathbf{a}_q) \simeq \bar{\zeta}_q(\mathbf{a}_q) = \underbrace{r_q \left[\sum\nolimits_{n=1}^{N} a_{q,n} \Delta m_{q,n} \right]}_{\lambda_q(\mathbf{a}_q)} \cdot \underbrace{\left[1 + \sum_{k=1}^{\bar{\sigma}_{r_q}-1} \left(\frac{\sum_{n=1}^{N} a_{q,n} \overline{\Omega}_{q,n}}{\sum_{i=1}^{N} a_{q,i}} \right)^k \right]}_{\Lambda_q(\mathbf{a}_q)}, \tag{6.25}$$

where $\Delta m_{q,n} \triangleq (m_q - \Omega_{q,n} e^{\gamma_{0,r_q}})$ are constant terms. The following properties are of interest:

P4) since $\Lambda_q(\mathbf{a}_q) \geq 1$, $\lambda_q(\mathbf{a}_q)$ results a lower bound of the EGP;

P5) if $\Lambda_q(\mathbf{a}_q)$ could be neglected, the EGP optimization would simply reduce to a integer programming, since $\lambda_q(\mathbf{a}_q)$ is a weighted sum of the discrete variables $a_{q,n}$;

P6) $\Lambda_q(\mathbf{a}_q)$, on the other hand, depends on the entire allocation vector $\mathbf{a}_q$.

With reference to properties P5 and P6, it can be pointed out the analogy between the two quantities at the basis of the ACO framework, i.e., the local desirability and the pheromone, with λ_q and Λ_q, respectively. This suggests that the SA-OP can be tackled resorting to the ACO framework. Even if the latter is intrinsically heuristic, its potentiality can be efficiently exploited when it is coupled with a deterministic algorithm. The analysis conducted, recalling property P4 too, meets the above observation, so that we can solve the SA-OP (6.15) in two steps as follows.

[2]It is exploited the geometric series $(1-x)\sum_{k=0}^{i} x^k = 1 - x^{i+1}$.

SA-OP.1 The first step consists in maximizing the minimum goodput lower bound $\lambda_q(\mathbf{a}_q)$, denoted SA-OP.1,

$$\begin{array}{lcll} \max\limits_{\mathbf{a}\in\mathcal{D}_a,l} & l & & \\ \text{s.t.} & \sum_{q\in Q} a_{q,n} \leq 1 & \forall n, & (6.26.\text{a}) \\ & \sum_{n=1}^{N} a_{q,n}\left(\overline{\Omega}_{q,n} - m_q e^{\overline{\gamma}_q}\right) \leq 0 & \forall q, & (6.26.\text{b}) \\ & l \leq \lambda_q(\mathbf{a}_q) & \forall q, & (6.26.\text{c}) \end{array} \tag{6.26}$$

where (6.26.b) is constraint (6.15.b) explicit as a function of $\mathbf{a}_q$, that results a mixed integer linear programming (MILP), solvable with largely employed MILP tools. Let denote with $\mathbf{a}^*_{\text{lb}}$ the solution obtained solving the MILP (6.26) and λ^*_{lb} its associated value, that is, the best lower bound found so far.

SA-OP.2 The second step, denoted SA-OP.2, consists in solving the SA-OP (6.15) through the ACO framework, where the solution to SA-OP.1 is used as starting point, so that the ACO algorithm enhances the performance improving the minimum goodput lower bound found.
Thus, SA-OP.2 is coded, according to the ACO framework, on a graph $G(\mathcal{V},\mathcal{E})$ made of the vertices set $\mathcal{V}$ and the edges set $\mathcal{E}$. Thanks to the structure of (6.25), these mappings are in order, and the resulting graph is depicted in Fig. 6.4.

- The vertices of the graph correspond to the N subcarriers, so that $\mathcal{V} = \{1, \cdots, N\}$, and each agent sequentially visits all the vertices.
- As a result, the set of edge connecting vertex $n-1$ to vertex n corresponds to the set of user that can be assigned to vertex n, i.e., $\mathcal{E}_n = \{e_{q,n}\}_{q=0}^{Q}$, with $\mathcal{E} = \mathcal{E}_1 \times \cdots \times \mathcal{E}_N$. Thus, selecting edge $e_{q,n}$ means $a_{q,n} = 1$ and $a_{v,n} = 0$, $\forall v \neq q$ ($e_{0,n}$ means that no user is assigned to subcarrier n, or, $a_{q,n} = 0$, $\forall q$).

The ACO based SA algorithm is summarized in Table 6.6 and described in the following. As previously outlined, for N_{it} iterations N_{a} agents explore the graph, each of them building a complete path $t_j^{(i)}$, with $i = 1, \cdots, N_{\text{it}}$ and $j = 1, \cdots, N_{\text{a}}$. Due to the correspondence between the pairs vertices-edges and subcarriers-users, each path univocally identifies a SA vector: $t_j^{(i)} \leftrightarrow \mathbf{a}^{(i,j)}$. The probability $\pi_{q,n}$ to select edge $e_{q,n}$ is

$$\pi_{q,n} = \frac{\eta_{q,n}\varphi_{q,n}}{\sum\limits_{(k,n)|e_{k,n}\in\mathcal{E}_n} \eta_{k,n}\varphi_{k,n}}, \tag{6.27}$$

ACO based algorithm for SA-OP.2

1. For $i = 0, \cdots, N_{\text{it}}$
2. For $j = 0, \cdots, N_{\text{a}}$
3. Set $t_j^{(i)} = \emptyset$
4. For $n = 1, \cdots, N$
5. Select edge $e_{u,n}$ with probability $\pi_{u,n}$
6. Update $t_j^{(i)} = t_j^{(i)} \cup e_{u,n}$
7. End For
8. End For
9. Evaluate $\zeta_{\text{best}}^{(i)}$, $t_{\text{best}}^{(i)}$ and $\mathbf{a}_{\text{best}}^{(i)}$
10. Pheromone evaporation $\varphi_{q,n} = [(1-\rho)\varphi_{q,n}]_{\varphi_{\min}}^{\varphi_{\max}}$, $\forall q, n$
11. If $\zeta_{\text{best}}^{(i)} > \lambda_{\text{lb}}^{*}$
12. Evaluate $\Delta\zeta(\mathbf{a}_{\text{best}}^{(i)})$
13. Update $\varphi_{q,n} = [\varphi_{q,n} + \delta_{\varphi}\Delta\zeta(\mathbf{a}_{\text{best}}^{(i)})]_{\varphi_{\min}}^{\varphi_{\max}}$, $\forall e_{q,n} \in t_{\text{best}}^{(i)}$
14. End If
15. End For
16. Output: $t_{\text{best}}^{(N_{\text{it}})}$

Table 6.6: *Pseudo-Code of the ACO based algorithm for SA-OP.2*

where $\eta_{q,v}$ is the local desirability and $\varphi_{q,v}$ the pheromone. Recalling properties P4-P5 and the graph structure where SA-OP.2 is mapped, the following correspondences are straightforward:

- The local desirability $\eta_{q,n}$ is a function of the weights $\Delta m_{q,n}$, since they are local quantities depending on the specific user q assigned to the specific subcarrier n. In particular, $\eta_{q,n} \triangleq [\Delta m_{q,n}]_{\bar{\eta}}$, so that $\eta_{q,n}$ is always greater than a minimum value $\bar{\eta}$ allowing at least a minimum probability to each path to be explored.
- The pheromone $\varphi_{q,v}$ depends instead on Λ_q, that is in fact, a global quantity, depending on the quality of the solution found.

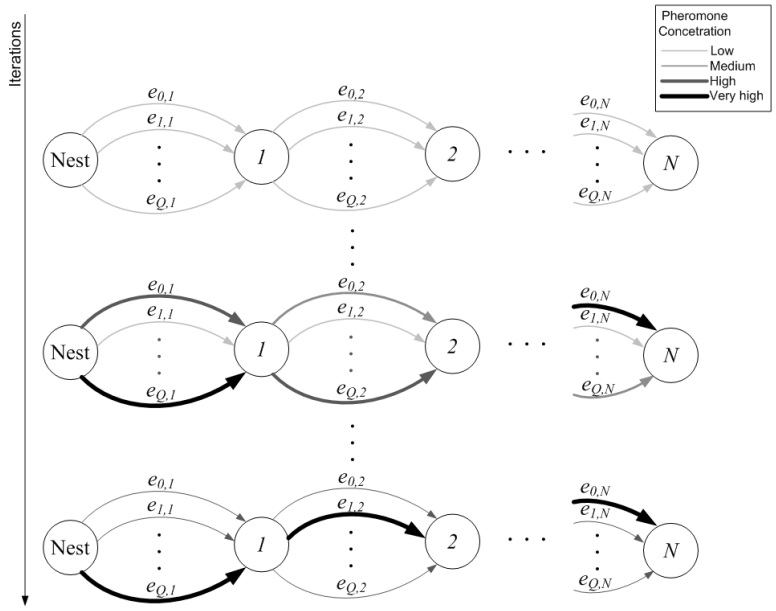

Figure 6.4: *ACO graph evolution.*

In particular, denoting with $\mathbf{a}^{(i,j)}$ the solution found by agent j at iteration i, let us introduce at the end of each iteration $\zeta_{\text{best}}^{(i)} = \max_{\mathbf{a}^{(i,j)}, \forall j} \min_{q} \zeta_q(\mathbf{a}^{(i,j)})$ as the highest value of the objective function found and $t_{\text{best}}^{(i)}$ and $\mathbf{a}_{\text{best}}^{(i)}$ as the related path and SA, respectively. If $\zeta_{\text{best}}^{(i)} > \lambda_{\text{lb}}^*$ the pheromone, $\forall e_{q,n} \in t_{\text{best}}^{(i)}$, is updated as $\varphi_{q,v} = [\varphi_{q,v} + \delta_\varphi \Delta\zeta(\mathbf{a}_{\text{best}}^{(i)})]_{\tau_{\min}}^{\tau_{\max}}$, with $\Delta\zeta(\mathbf{a}_{\text{best}}^{(i)}) \triangleq \zeta_{\text{best}}^{(i)}/\lambda_{\text{lb}}^*$.
The higher the value of the solution w.r.t. the one obtained solving SA.OP.1, the higher the pheromone increment released on the edges that contributed to that solution, driving in this way the search over the graph. At the end of the iterations, thanks to the reinforcement learning mechanism, a path $t_{\text{best}}^{(N_{\text{it}})}$ emerges on the graph, corresponding to the solution of the SA-OP (6.15).

6.7 Simulation Results

In this section, simulation results carried over a realistic wireless scenario are shown in order to support the analytical results obtained in the previous sections. The parame-

ters settings of the OFDMA system and radio channel adopted for the simulations are reported in Tabs. 6.7 and 6.8, respectively. In the following, two reference scenarios are taken into account. In scenario S1 we consider $Q = 2$ users and the performance are evaluated, averaging over 10^3 channel realizations, as a function of the user 2 average symbol energy to noise power spectral density ratio E_s/N_0, whereas user 1 has a constant $E_s/N_0 = 27$ dB. Scenario S2 has the same characteristics of scenario S1 extended to $Q = 3$ users, with user 3 $E_s/N_0 = 7.63$ dB. The goodput curves depicted in the following figures have been normalized w.r.t. B/Q, so that each curve physically represents a spectral efficiency in (bit/s/Hz).

Parameter/Feature	**Symbol**	**Value/Description**
Payload length	N_p	1024 bits
CRC length	N_{CRC}	32 bits
No. of subcarriers	N	64
Bits per subcarrier	$\mathcal{D}_m$	$\{2, 4, 6\}$
Code type		64-state Convolutional code
Mother code rate	r_0	1/2
Punctured code rates	$\mathcal{D}_r$	$\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}\right\}$
Available power	P	34 dBm
Bandwidth	B	20 MHz
ARQ scheme		Multiple-channel Stop & Wait
No. of logical channels		8

Table 6.7: *Parameters and features of the ARQ BIC-OFDMA system.*

PER Model.

Figure 6.5 shows the actual PER (dotted curves) and the analytical PER curves described by eqn. (6.24) (lines with marks), vs. the E_s/N_0 ratio experienced at the receiver. The actual PER curves, averaged over 10^4 channel realizations, have been obtained simulating a BPSK system transmitting over an AWGN channel and employing the 64-state convolutional code whose rates are detailed in Tab. 6.7. The values of σ_r and $\gamma_{0,r}$ are reported in Table 6.9.

Parameter/Feature	Value/Description
Path-loss model	NLOS urban scenario [IEEE 802.16]
Carrier frequency	2 GHz
BS/Mobile heights	12.5/1.5 m
Noise power level	-100 dBm
Short-term fading model	ITU Ped. B

Table 6.8: *Parameters and features of the radio propagation channel model.*

	$r = 1/2$	$r = 2/3$	$r = 3/4$	$r = 5/6$
$\gamma_{0,r}$	0.7198	1.064	1.309	1.633
σ_r	10.23	7.205	5.63	4.581

Table 6.9: *Parameters values of the PER Model.*

ACO Algorithm for the SA Problem.
Figure 6.6 shows the minimum EGP curves obtained solving the SA problem (6.16) both with the optimal algorithm based on B&B and with the ACO algorithm described in sections 6.6.1 and 6.6.2, respectively, for reference scenario S1.
Depending on the number of iterations N_{it} and on the number of agents N_{a} that explore the graph at each iteration, the ACO algorithm stems as a valid candidate to solve the SA-OP (6.16), offering at the same time very good performance and reduced complexity w.r.t. the optimal algorithm.

MMG Algorithm Initialization.
In this section, some guidelines are offered in order to find a good initial transmission mode $\boldsymbol{\tau}^{(0)}$ for the CAM based MMG algorithm described in Tab. 6.1, tagged as CAM-MMG for short. The initial subcarrier allocation is done as follows: each user selects, in round robin fashion, the subcarrier where it has the best channel gain among the ones not yet chosen. This initialization is justified by the observation that, for each user, the PER is dominated by term relevant to the worst channel gain value.

The power is instead uniformly distributed over all the subcarriers, since, to the best of authors knowledge, there are not other PA algorithms that maximizes the minimum GP. This uniform PA is thus also exploited as benchmark to show the improvements obtained with the proposed algorithm, as usually done in these cases. The impact of the initial value of the modulation order and coding rate has been investigated by simulation, as shown in Fig. 6.7, referring to scenario S1.

These curves are obtained running the CAM-MMG for 3 different pairs of initial modulation order and coding rate. Two main observation can be done. First, for the pair $(m = 4, r = 3/4)$, Fig. 6.7 depicts both the EGP ζ_q in (6.3) and the actual goodput (AGP) value, i.e. the average of the ratio between the number of payload bits $N_{\mathrm{p}}^{(q)}$ and the transmission time required to successfully delivering them. As anticipated in Sect. 3.2.3, their value is very close justifying the adoption of the long term static assumption. Secondly, the best initial choice is given by the pair $(m = 4, r = 3/4)$, which offers a good trade-off between data rate and link reliability. In fact, a too aggressive choice like $(m = 6, r = 5/6)$ only performs well when both users are in the high SNR region. Conversely, an extremely conservative choice, i.e., $(m = 2, r = 1/2)$, works well in the low SNR region but makes the algorithm stuck in a lower local optimum when users experiences high SNR values.

The impact of the value of the parameters of the ACO algorithm, adopted to solve the SA problem at Line 4 of the CAM-MMG code in Tab. 6.1, on the final performance has been evaluated through an intensive off-line simulation. Part of these results are shown in Fig. 6.8, from which the following setting of values is obtained: $\rho = 0.1$, $\delta_{\varphi} = 0.1$, $N_{\mathrm{it}} = 50$, $N_{\mathrm{a}} = 50$.

Finally, Fig. 6.9 shows the average number of iterations required by the CAM-MMG algorithm to converge, for two values of user 2 E_s/N_0. Interestingly, after a few iterations, we get the most significant improvements in the minimum EGP value, then only slight gains are obtained. Thus, the value of I_{CAM} can be reduced to trade off complexity and quality of solution.

MMG Algorithm Performance.

Figures 6.10 compares the AGP performance of user 1 and 2 for scenario S1 obtained with both the CAM-MMG algorithm and a static RA policy. In particular, the latter refers to the case in which the setting of transmission parameters $\boldsymbol{\tau}$ is not adapted to the actual channel conditions but is instead kept constant and, in particular, equal

to the setting $\boldsymbol{\tau}^{(0)}$ chosen to initialize the CAM-MMG algorithm.
From Fig. 6.10 it is apparent how unfair RA approaches, such as the static RA policy, penalize all the user that are not denoted by high SNR levels, like user 1 for the case at hand. Conversely, it can be noted that the CAM-MMG algorithm successfully satisfies the fairness criterion, with both users characterized by nearly the same AGP. The performance of the user with the minimum goodput (dotted line with empty triangles) in the static RA case are in fact remarkably enhanced. The same behavior can be found in Fig. 6.11 that refers to scenario S2 where $Q = 3$ users are involved.
Finally, Fig. 6.12 shows that the CAM-MMG effectively keeps the PER under the prescribed threshold, depicting the PER curves of user 2 for scenario S1 for two different QoS constraints, namely, $\overline{\Psi}_q^{(1)} = 10^{-1}$ and $\overline{\Psi}_q^{(2)} = 10^{-2}$, $\forall q \in \mathcal{Q}$.

6.8 Concluding remarks

In this chapter, a novel RA algorithm aimed at maximizing the performance of the user with the lowest value of goodput on the downlink of a BIC-OFDMA system was derived and properly investigated. Thanks to the CAM, the RA problem was tackled through an iterative algorithm where, at each step, three subproblems are solved: the PA problem, the SA problem and the AMC problem. All of these subproblems were analyzed and optimally solved. In detail, the PA problem was shown to be convex and thus solved resorting to the LDD approach. The AMC problem was tackled adopting an exhaustive search. The SA problem was first optimally solved thanks to a novel analysis on the objective function that allowed to bring back the problem to the convex optimization framework. Then, a novel algorithm that offers a trade-off between complexity and quality of solution was also proposed resorting to the ACO framework. Simulation results over a typical wireless scenario certified the effectiveness of the proposed approach.

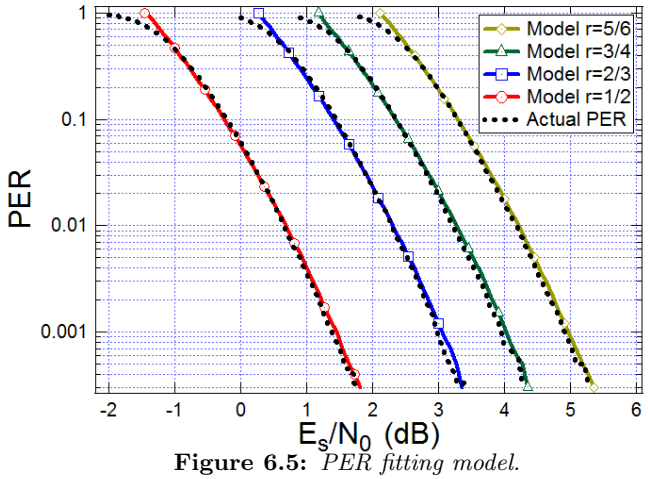

Figure 6.5: *PER fitting model.*

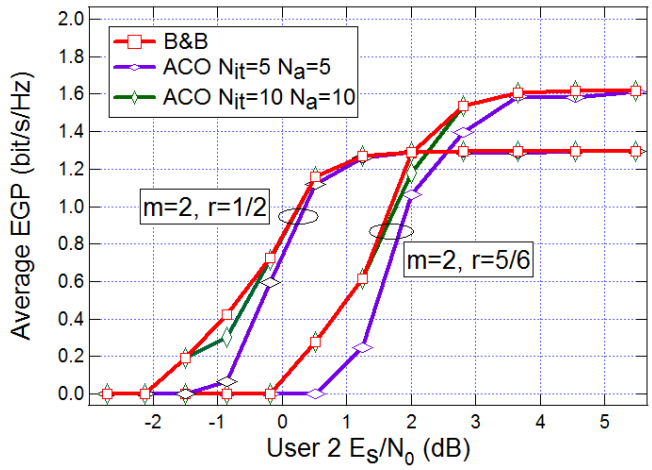

Figure 6.6: *Performance comparison of B&B and ACO algorithms.*

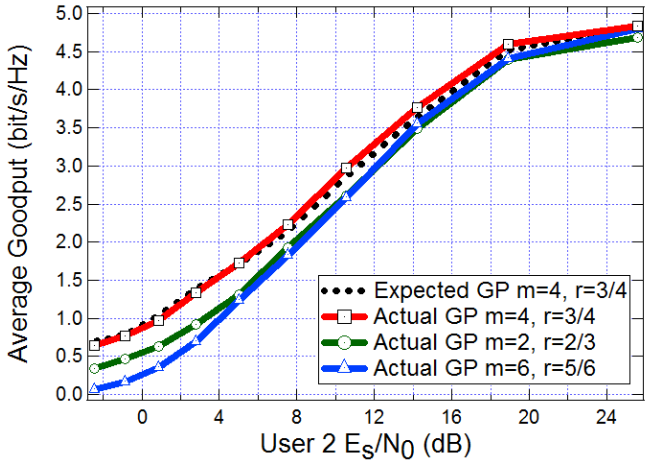

Figure 6.7: *AGP values for different initial couples of* (m, r).

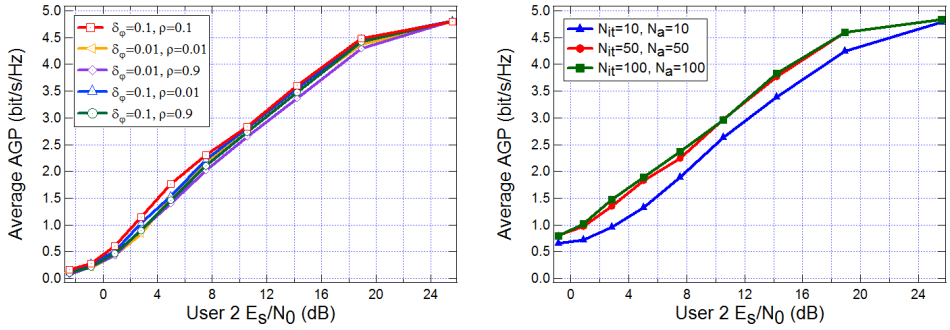

Figure 6.8: *Performance comparison for different values of ACO parameters* ρ, δ_{φ}, N_{it}, N_{a}.

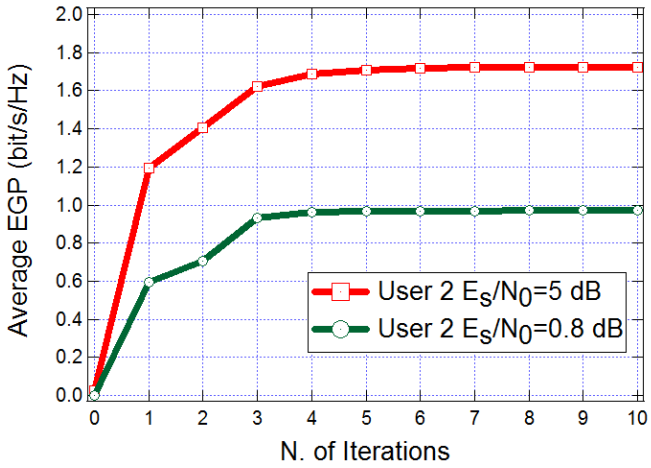

Figure 6.9: *Average number of iterations of the CAM-MMG algorithm.*

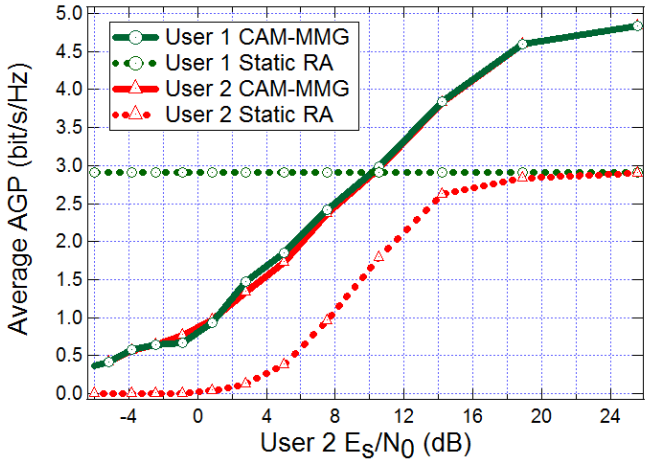

Figure 6.10: *AGP performance of $Q = 2$ users obtained with CAM-MMG and static RA policy.*

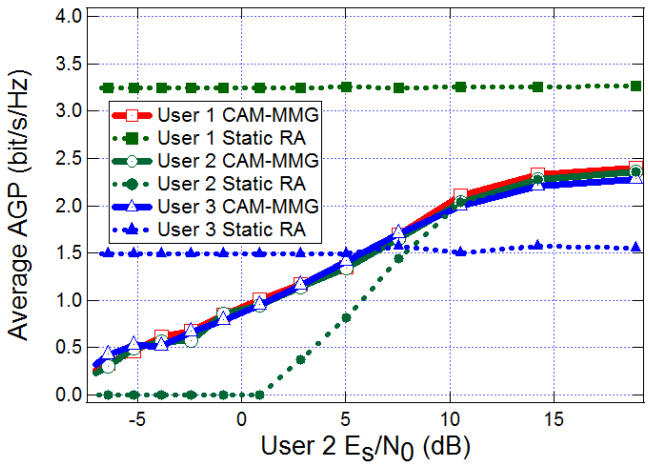

Figure 6.11: *AGP performance of $Q = 3$ users obtained with CAM-MMG and static RA policy.*

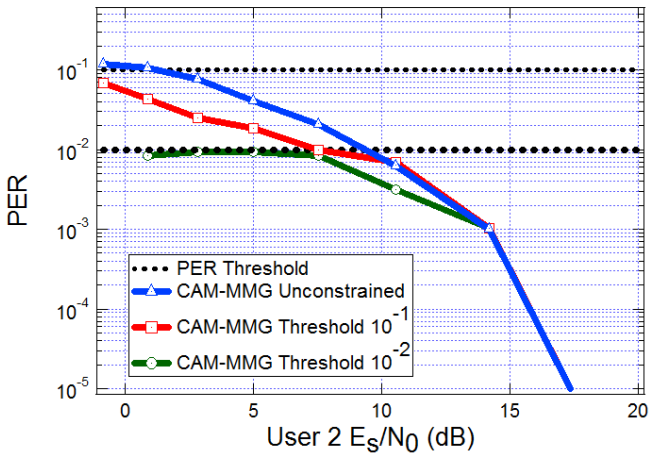

Figure 6.12: *PER performance of $Q = 2$ users obtained with CAM-MMG for different QoS thresholds.*

Appendix A

Appendices

A.1 Basics of convex optimization theory

A.1.1 Duality

Consider the optimization problem

$$\begin{array}{lll} \min\limits_{\mathbf{x}} & f_0(\mathbf{x}) & \\ \text{s.t.} & f_i(\mathbf{x}) \leq 0 \quad i = 1, \cdots, m & (A.1.\text{a}) \\ & h_i(\mathbf{x}) = 0 \quad i = 1, \cdots, p & (A.1.\text{b}) \end{array} \tag{A.1}$$

where $f_0(\mathbf{x})$ is the objective function, $f_i(\mathbf{x})$ and $h_i(\mathbf{x})$ are the inequality and equality constraints, respectively, and $\mathbf{x} = [x_1, \cdots, x_n]^{\mathrm{T}} \in \mathbb{R}^n$ is the optimization variable.
This problem is convex if $\{f_i(\mathbf{x})\}_{i=0}^m$ are convex functions and $\{h_i(\mathbf{x})\}_{i=1}^p$ are affine. A point $\mathbf{x}$ is feasible if it satisfies constraints (A.1.a) and (A.1.b). Let p^* be the optimal value of (A.1).
The *Lagrangian* associated to (A.1) is

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\upsilon}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \upsilon_i h_i(\mathbf{x}) \tag{A.2}$$

where $\boldsymbol{\lambda} \triangleq [\lambda_1, \cdots, \lambda_m]^{\mathrm{T}}$ and $\boldsymbol{\upsilon} \triangleq [\upsilon_1, \cdots, \upsilon_p]^{\mathrm{T}}$ are the *Lagrange multipliers* associated with the inequality and equality constraints, respectively.
Then, the *dual function* of the (primal) problem (A.1) is defined as

$$g(\boldsymbol{\lambda}, \boldsymbol{\upsilon}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\upsilon}). \tag{A.3}$$

For any $\boldsymbol{\lambda} \succeq \mathbf{0}$ and any $\boldsymbol{\upsilon}$, the dual function yields lower bounds on the optimal value p^* of the primal problem. This property is easily verified by noting that, according

to constraints (A.1.a) and (A.1.b), in (A.2) we have

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \upsilon_i h_i(\tilde{\mathbf{x}}) \leq 0, \tag{A.4}$$

where $\tilde{\mathbf{x}}$ is any (primal) feasible point. Upon defining the domain of the dual function as $\mathbf{dom}\ g \triangleq \{(\boldsymbol{\lambda}, \boldsymbol{\upsilon}) : g(\boldsymbol{\lambda}, \boldsymbol{\upsilon}) > -\infty,\ \boldsymbol{\lambda} \succeq \mathbf{0}\}$ (i.e. neglecting the trivial case of g unbounded below), then the pair $(\boldsymbol{\lambda}, \boldsymbol{\upsilon}) \in \mathbf{dom}\ g$ is said dual feasible.
Thus, the best lower bound on p^* is obtained solving the *Lagrange dual problem*

$$\begin{array}{ll} \max & g(\boldsymbol{\lambda}, \boldsymbol{\upsilon}) \\ \text{s.t.} & \boldsymbol{\lambda} \succeq \mathbf{0} \end{array}. \tag{A.5}$$

The pair $(\boldsymbol{\lambda}^*, \boldsymbol{\upsilon}^*)$ is referred to as dual optimal or optimal Lagrange multipliers if they are optimal for problem (A.5) and the associated optimal value of the dual function is referred to as d^*.
In general, $d^* \leq p^*$ and the difference $p^* - d^*$ is called *duality gap*. When $d^* = p^*$ the duality gap is zero, or, equivalently, *strong duality* holds. Thus, supposing that strong duality holds, we have

$$\begin{aligned} & f_0(\mathbf{x}^*) = g(\boldsymbol{\lambda}^*, \boldsymbol{\upsilon}^*) \\ & = \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{i=s1}^{p} \upsilon_i^* h_i(\mathbf{x}) \right) \\ & \leq f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \upsilon_i^* h_i(\mathbf{x}^*) \\ & \leq f_0(\mathbf{x}^*). \end{aligned}$$

From this chain we conclude that

1. since the inequality at the third line is an equality, then $\mathbf{x}^*$ is a minimizer of $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\upsilon}^*)$;

2. $\sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) = 0$ and, since each term of the sum is lower or equal than zero, it follows that $\lambda_i^* f_i(\mathbf{x}^*) = 0$, $i = 1, \cdots, m$.

Assuming that the function $f_0, \cdots, f_m, h_1, \cdots, h_p$ are differentiable then, according to point 1 of the previous observations, since $\mathbf{x}^*$ is a minimizer of $L(\cdot)$, then its gradient must vanish at $\mathbf{x}^*$, i.e.

$$f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \upsilon_i^* \nabla h_i(\mathbf{x}^*) = 0.$$

Thus, collecting all the conditions, we have

$$\begin{aligned} f_i(\mathbf{x}^*) &\leq 0 \quad i = 1, \cdots, m \\ h_i(\mathbf{x}^*) &= 0 \quad i = 1, \cdots, p \\ \lambda_i^* &\geq 0 \quad i = 1, \cdots, m \\ \lambda_i^* f_i(\mathbf{x}^*) &= 0 \quad i = 1, \cdots, m \\ \nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \upsilon_i^* \nabla h_i(\mathbf{x}^*) &= 0, \end{aligned}$$

which are called *Karush-Khun-Tucker* (KKT) *conditions*.
To summarize, the KKT conditions represent a necessary condition on the pair of primal and dual optimal points for any optimization problem with differentiable objective and constraints functions for which strong duality holds. If the optimization problem is *convex*, then the KKT conditions are also *sufficient* for the points to be primal and dual optimal.

A.1.2 A brief summary on the subgradient method

Often, optimization problems of the form in (A.1) are solved maximizing the dual function in the dual domain. This is due to the fact that the dual problem is usually easier to solve, since the dual function is always a concave function of its optimization variable, even when problem (A.1) is not. Moreover, when the primal problem (A.1) is convex and KKT holds, the duality gap is zero and then the solution found is both primal and dual optimal.
The dual OP (A.5) can be solved, as usually done in cases, resorting to the *subgradient-based update* of the dual variables $(\boldsymbol{\lambda}, \boldsymbol{\upsilon})$, here briefly recalled and referring to [74] for further details. This iterative method consists in maximizing $g(\boldsymbol{\lambda}, \boldsymbol{\upsilon})$ by updating, at each step $i+1$, all the components of the dual variable $(\boldsymbol{\lambda}^{(i)}, \boldsymbol{\upsilon}^{(i)})$ produced at the previous step i along the search direction defined by the subgradient of $g(\boldsymbol{\lambda}, \boldsymbol{\upsilon})$ at $(\boldsymbol{\lambda}^{(i)}, \boldsymbol{\upsilon}^{(i)})$.
For the ease of notation, let us stuck the Lagrange multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\upsilon}$ into the vector $\mathbf{y} = [\boldsymbol{\lambda}^{\mathrm{T}}, \boldsymbol{\upsilon}^{\mathrm{T}}]^{\mathrm{T}}$, whose domain is identified by $\mathbf{Y}$, and the relevant functions into $\mathbf{h}(\mathbf{x}) \triangleq [f_1(\mathbf{x}), \cdots, f_m(\mathbf{x}), h_1(\mathbf{x}), \cdots, h_p(\mathbf{x})]^{\mathrm{T}}$. Accordingly, the dual function (A.3) can be rewritten as

$$g(\mathbf{y}) = \inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{x}} f_0(\mathbf{x}) + \mathbf{y}^{\mathrm{T}} \mathbf{h}(\mathbf{x}). \tag{A.6}$$

As previously shown, $g(\mathbf{y})$ is a concave function of $\mathbf{y}$ since it is a pointwise minimum over $\mathbf{x}$ of an affine function of $\mathbf{y}$, namely $L(\mathbf{x}, \mathbf{y})$. For a given multipliers vector $\mathbf{y}$, suppose that $\mathbf{x}_\mathbf{y}$ minimizes $L(\mathbf{x}, \mathbf{y})$ over the set of feasible values of $\mathbf{x}$. Then $\mathbf{h}(\mathbf{x}_\mathbf{y})$ is a subgradient of $g(\mathbf{y})$ at $\mathbf{y}$, i.e., the following inequality holds according to the definition of a subgradient of a concave function in [74]

$$g(\mathbf{y}') \leq g(\mathbf{y}) + (\mathbf{y}' - \mathbf{y})^{\mathrm{T}} h(\mathbf{x}_\mathbf{y}), \tag{A.7}$$

with $\mathbf{y}' \in \mathbf{Y}$. The subgradient method generates a sequence of feasible dual variables according to the iteration

$$\mathbf{y}^{(i+1)} = \left[\mathbf{y}^{(i)} + s h(\mathbf{x}_\mathbf{y})\right]_\mathbf{Y} \tag{A.8}$$

where i and s denotes the iteration number and the positive scalar step size used to produce $\mathbf{y}^{(i+1)}$, and $[\cdot]_\mathbf{Y}$ stands for the operator of projection into $\mathbf{Y}$. What makes the subgradient method work is that for a sufficiently small s, the distance of $\mathbf{y}^{(i)}$ to the optimal dual variable is reduced as the number of iterations increases, as justified by [74].

A.2 MGF evaluation

In order to demonstrate (2.16), let us start by substituting (2.7) into (2.12), neglecting, w.l.g. the PR index ℓ, so that

$$\mathcal{M}_\Lambda(s) = \mathrm{E}_k \left\{ \left[\frac{\sum_{\tilde{x} \in \chi_{b'_k}^{(i_k, n_k)}} \exp\left(- \left| \sqrt{\gamma_{n_k}} \left(x_{n_k} - \tilde{x} \right) + w_{n_k} \right|^2 \right)}{\sum_{\tilde{x} \in \chi_{b_k}^{(i_k, n_k)}} \exp\left(- \left| \sqrt{\gamma_{n_k}} \left(x_{n_k} - \tilde{x} \right) + w_{n_k} \right|^2 \right)} \right]^s \right\}, \tag{A.9}$$

where the expectation is taken over all the parameters depending on the kth bit, i.e., the noise, the modulation symbol and the position of the coded symbol within the label of the QAM symbols.

Similarly to the approach suggested in [34], the m.g.f. (A.9) can be upper-bounded and, it is also demonstrated that, at high SNRs, the bound is dominated by the term relevant to the nearest neighbor (in the sense of Euclidean distance) of x_{n_k} in the complementary subset $\chi_{b'_k}^{(i_k, n_k)}$. As a consequence, we can apply the dominated convergence theorem [34] and, after taking the expectation with respect to the noise,

(A.9) turns into

$$\mathcal{M}_{\Lambda}(s) \simeq \mathrm{E}_k\left\{\exp\left(-\gamma_{n_k} d^2\left(x_{n_k}, \overline{x}\right)\left(s - s^2\right)\right)\right\}, \tag{A.10}$$

where $\overline{x}$ is the nearest neighbor of x_{n_k} in the complementary subset $\chi_{b'_k}^{(i_k,n_k)}$, and $d(y,w)$ is the Euclidean distance between the complex-valued symbols y and w.
To compute the expectation in (A.10), it is worth noting that

1. due to the Gray mapping rule, we have $d\left(x_{n_k}, \overline{x}\right) = \mu\, d_{m_{n_k}}^{(\min)}$, where μ is a positive integer and $d_{m_{n_k}}^{(\min)}$ is the minimum Euclidean distance between the symbols of the $2^{m_{n_k}}$-QAM constellation adopted on the n_kth subcarrier;

2. each of the m_{n_k} label bits conveyed by subcarrier n_k has $2^{m_{n_k}-1}$ symbols on its complementary subset, so that the total number of terms to be averaged results $m_{n_k} \cdot 2^{m_{n_k}-1}$;

3. the distance $d\left(x_{n_k}, \overline{x}\right)$ takes $\sqrt{2^{m_{n_k}}}/2$ distinct values so that $1 \leq \mu \leq \sqrt{2^{m_{n_k}}}/2$;

4. by defining the number of symbols at distance $\mu\, d_{m_{n_k}}^{(\min)}$ from the nearest neighbor in the complementary subset as $\psi_{m_{n_k}}(\mu)$, we get $\sum_{\mu=1}^{\sqrt{2^{m_{n_k}}}/2} \psi_{n_k}(\mu) = m_{n_k} \cdot 2^{m_{n_k}-1}\ \forall\, n \in \mathcal{D}_s$, where $\psi_{m_{n_k}}(\mu)\Big|_{m=2} = 4\delta_{\mu-1}$, $\psi_{m_{n_k}}(\mu)\Big|_{m=4} = 24\delta_{\mu-1} + 8\delta_{\mu-2}$, $\psi_{m_{n_k}}(\mu)\Big|_{m=6} = 112\delta_{\mu-1} + 48\delta_{\mu-2} + 16\delta_{\mu-3} + 16\delta_{\mu-4}$ for the 4-, 16- and 64-QAM formats, respectively;

5. the probability that the generic coded symbol is sent through the n_kth subchannel is $m_{n_k} / \sum_{j=1}^{N} m_j$.

Therefore, collecting the above results together, the MGF in (A.10) turns out to be

$$\mathcal{M}_{\Lambda}(s) \simeq \frac{1}{\sum_{j=1}^{N} m_j} \cdot \sum_{n=1}^{N} \sum_{\mu=1}^{\sqrt{2^{m_n}}/2} \frac{\psi_{m_n}(\mu)}{2^{m_n-1}} e^{-\gamma_n \left[\mu\, d_{m_n}^{(\min)}\right]^2 (s-s^2)}. \tag{A.11}$$

Bibliography

[1] International Telecommunication Union, "Measuring the information society," 2012.

[2] M.-O. Pun, M. Morelli, and C.-C. J. Kuo, *Multi-Carrier Techniques for Broadband Wireless Communications. A Signal Processing Perspective.* Imperial College Press, London, UK, 2007.

[3] M. Rahnema, "Overview of the GSM system and protocol architecture," *IEEE Communications Magazine*, vol. 31, no. 4, pp. 92 –100, April 1993.

[4] K. W. Richard, "UMTS overview," *IEE Electronics and Communication Engineering Journal*, vol. 12, no. 3, pp. 93 – 100, June 2000.

[5] E. Berruto, G. Colombo, P. Monogioudis, A. Napolitano, and K. Sabatakakis, "Architectural aspects for the evolution of mobile communications toward UMTS," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 8, pp. 1477 –1487, Oct. 1997.

[6] D. Knisely, S. Kumar, S. Laha, and S. Nanda, "Evolution of wireless data services: IS-95 to cdma2000," *IEEE Communications Magazine*, vol. 36, no. 10, pp. 140 –149, Oct. 1998.

[7] IEEE 802.11a, "IEEE Std 802.11a-1999, Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer in the 5 GHz band," pp. i–82, Sept. 1999.

[8] IEEE 802.11g, "IEEE Std 802.11g-2003, Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," pp. i–67, 2003.

[9] B. Crow, I. Widjaja, J. G. Kim, and P. Sakai, "IEEE 802.11 Wireless Local Area Networks," *IEEE Communications Magazine*, vol. 35, no. 9, pp. 116 – 126, Sept. 1997.

[10] H. Yaghoobi, "Scalable OFDMA physical layer in IEEE 802.16 WirelessMAN," 2004.

[11] IEEE 802.15, "IEEE Std 802.15.1-2005, Part 15.1a: Wireless medium access control (MAC) and physical layer (PHY) specifications for wireless personal area networks (WPAN)," p. i, 2005.

[12] S. Roy, J. Foerster, V. Somayazulu, and D. Leeper, "Ultrawideband radio design: the promise of high-speed, short-range wireless connectivity," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 295 – 311, Feb. 2004.

[13] U. Varshney, "4G wireless networks," *IT Professional*, vol. 14, no. 5, pp. 34 –39, Sept.-Oct. 2012.

[14] W. Lu, "An open baseband processing architecture for future mobile terminal design," *IEEE Wireless Commun.*, vol. 15, no. 2, pp. 110 –119, April 2008.

[15] T. Janevski, "5G mobile phone concept," in *6th IEEE Consumer Commun. and Networking Conf., 2009. CCNC 2009*, Jan. 2009, pp. 1 –2.

[16] L.-C. Wang and S. Rangapillai, "A survey on green 5G cellular networks," in *2012 Intern. Conf. on Signal Process. and Commun. (SPCOM)*, July 2012, pp. 1 –5.

[17] V. Muntean and M. Otesteanu, "WiMAX versus LTE - An overview of technical aspects for next generation networks technologies," in *9th Intern. Symp. on Electronics and Telecommun. (ISETC), 2010*, Nov. 2010, pp. 225 –228.

[18] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 927 –946, May 1998.

[19] A. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. on Commun.*, vol. 45, no. 10, pp. 1218 –1230, Oct. 1997.

[20] E. Dahlman, S. Parkvall, J. Skld, and P. Beming, *3G Evolution. HSPA and LTE for Mobile Broadband.* Academic Press, Elsevier, Oxford, 2008.

[21] S. Lin, D. Costello, and M. Miller, "Automatic-repeat-request error-control schemes," *IEEE Communications Magazine*, vol. 22, no. 12, pp. 5 –17, Dec. 1984.

[22] J. Costello, D.J., J. Hagenauer, H. Imai, and S. Wicker, "Applications of error-control coding," *IEEE Transactions on Information Theory,*, vol. 44, no. 6, pp. 2531 –2560, Oct. 1998.

[23] J. Lai-Wei, W. Wei, and F. Wei, "An overview of resource allocation algorithms for OFDM systems," in *2012 Second Intern. Conferen. on Instrumentation, Measurement, Computer, Communication and Control (IMCCC)*, Dec. 2012, pp. 429 –432.

[24] D. Qiao, S. Choi, and K. Shin, "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs," *IEEE Transactions on Mobile Computing*, vol. 1, no. 4, pp. 278 – 292, Oct.-Dec. 2002.

[25] S. Nanda and K. Rege, "Frame error rates for convolutional codes on fading channels and the concept of effective E_b/N_0," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 4, pp. 1245 –1250, Nov. 1998.

[26] E. Tuomaala and H. Wang, "Effective SINR approach of link to system mapping in OFDM/multi-carrier mobile network," in *2nd International Conference on Mobile Technology, Applications and Systems, 2005*, Nov. 2005, pp. 1–5.

[27] D. Astely, E. Dahlman, P. Frenger, R. Ludwig, M. Meyer, S. Parkvall, P. Skillermark, and N. Wiberg, "A future radio-access framework," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 693 – 706, March 2006.

[28] R. Grunheid, B. Chen, and H. Rohling, "Joint layer design for an adaptive OFDM transmission system," in *IEEE 58th Vehicular Technology Conference, 2003. VTC 2003-Fall.*, vol. 1, Oct. 2003, pp. 542 – 546 Vol.1.

[29] B. Devillers, J. Louveaux, and L. Vandendorpe, "Bit and power allocation for goodput optimization in coded parallel subchannels with ARQ," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3652 –3661, Aug. 2008.

[30] L. Kleinrock and H. Opderbeck, "Throughput in the ARPANET–Protocols and measurement," *IEEE Transactions on Communications*, vol. 25, no. 1, pp. 95 – 104, Jan. 1977.

[31] WiMAX Forum, "System evaluation methodology," pp. 1–209, July 2008.

[32] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY: McGraw-Hill, 2001.

[33] 3GPP, "Technical specification group radio access network," TR 25.890 V1.0.0.

[34] A. Martinez, A. Guillen i Fabregas, and G. Caire, "Error probability analysis of bit-interleaved coded modulation," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 262 – 271, Jan. 2006.

[35] E. Biglieri, *Coding for Wireless Channels*, 9th ed. New York, NY: Springer, 2005.

[36] A. Guillen i Fabregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," Foundations and Trends in Communications and Information Theory, 2008.

[37] S. Simoens, S. Rouquette-Léveil, P. Sartori, Y. Blankenship, and B. Classon, "Error prediction for adaptive modulation and coding in multiple-antenna OFDM systems," *Signal Process.*, vol. 86, no. 8.

[38] L. Song and N. Mandayam, "Hierarchical SIR and rate control on the forward link for CDMA data users under delay and error constraints," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 1871 –1882, Oct. 2001.

[39] K. Brueninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link performance models for system level simulations of broadband radio access systems," in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005.*, vol. 4, Sept. 2005, pp. 2306 –2311 Vol. 4.

[40] Y. Blankenship, P. Sartori, B. Classon, V. Desai, and K. Baum, "Link error prediction methods for multicarrier systems," in *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall.*, vol. 6, Sept. 2004, pp. 4175 – 4179.

[41] WINNER, "Assessment of advanced beamforming and MIMO technologies," Information society technologies, D2.7, ver1.0. Standardization Document, EU project, 2003.

[42] R. Srinivasan et al, "Project 802.16m evaluation methodology document (EMD)," IEEE 802.16m-08/004r5, Tech. Rep., Jan. 2009.

[43] S. Tsai, "Effective-SNR mapping for modeling frame error rates in multiple-state channels," 3GPP2-C30-20030429-010, April 2003.

[44] W. Hai, W. Lei, and L. Min, "HSDPA link adaptation based on novel quality model," in *2005 IEEE 61st Vehicular Technology Conference, 2005. VTC 2005-Spring.*, vol. 1, May-June 2005, pp. 334 – 338 Vol. 1.

[45] I. Stupia, V. Lottici, F. Giannetti, and L. Vandendorpe, "Link resource adaptation for multiantenna bit-interleaved coded multicarrier systems," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3644 –3656, July 2012.

[46] S. Kumar, S. Nanda, and A. Falconi, "Reverse link power control performance for high speed data over CDMA," TIA Contribution, TR45.5.4/97.11, Tech. Rep.

[47] H. Seo and B. G. Lee, "Optimal transmission power for single- and multi-hop links in wireless packet networks with ARQ capability," *IEEE Transactions on Communications*, vol. 55, no. 5, pp. 996 –1006, May 2007.

[48] I. Stupia, F. Giannetti, V. Lottici, R. Andreotti, L. Vandendorpe, and A. D'Andrea, "A greedy algorithm for goodput-oriented AMC in turbo-coded OFDM," in *In Proc. of Future Network and Mobile Summit, 2010*, June 2010, pp. 1 –8.

[49] A. Lozano, A. Tulino, and S. Verdu, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3033 – 3051, July 2006.

[50] C. Xiao, Y. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector gaussian channels," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3301 –3314, July 2011.

[51] M. Realp, A. Perez-Neira, and C. Mecklenbrauker, "A cross-layer approach to multi-user diversity in heterogeneous wireless systems," in *2005 IEEE International Conference on Communications, 2005. ICC 2005.*, vol. 4, May 2005, pp. 2791 – 2796 Vol. 4.

[52] K. Sigman, "Introduction to renewal theory," Lecture Notes on Stochastic Modeling, 2009.

[53] M. Zorzi and R. Rao, "On the use of renewal theory in the analysis of ARQ protocols," *IEEE Transactions on Communications*, vol. 44, no. 9, pp. 1077 –1081, Sept. 1996.

[54] FCC, "02-48: First report and order," U. S. Federal Communications Commission, Tech. Rep. FCC 02-48, 2002.

[55] I. Mitola, J. and J. Maguire, G.Q., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13 –18, Aug. 1999.

[56] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201 – 220, Feb. 2005.

[57] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894 –914, May 2009.

[58] Z. Tian, G. Leus, and V. Lottici, "Joint dynamic resource allocation and waveform adaptation for cognitive networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 443 –454, Feb. 2011.

[59] R. Prasad, P. Pawelczak, J. Hoffmeyer, and H. Berger, "Cognitive functionality in next generation wireless networks: standardization efforts," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 72 –78, April 2008.

[60] E. Hossain, D. Niyato, and Z. Han, *Dynamic Spectrum Access and Management in Cognitive Radio Networks*, 1st ed. New York, NY: Cambridge University Press, 2009.

[61] D. Datla, A. Wyglinski, and G. Minden, "A spectrum surveying framework for dynamic spectrum access networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 4158 –4168, Oct. 2009.

[62] FCC, "Second report and order and memorandum opinion and order," U. S. Federal Communications Commission, Tech. Rep. FCC 08-260, 2008.

[63] D. Niyato, E. Hossain, and Z. Han, "Dynamic spectrum access in IEEE 802.22-based cognitive wireless networks: a game theoretic model for competitive spectrum bidding and pricing," *IEEE Wireless Communications*, vol. 16, no. 2, pp. 16 –23, April 2009.

[64] R. Zhang and Y.-C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 88 –102, Feb. 2008.

[65] G. Bansal, M. Hossain, and V. Bhargava, "Optimal and suboptimal power allocation schemes for OFDM-based cognitive radio systems," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4710 –4718, Nov. 2008.

[66] Z. Hasan, G. Bansal, E. Hossain, and V. Bhargava, "Energy-efficient power allocation in OFDM-based cognitive radio systems: A risk-return model," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 6078 –6088, Dec. 2009.

[67] P. Cheng, Z. Zhang, H.-H. Chen, and P. Qiu, "Optimal distributed joint frequency, rate, and power allocation in cognitive OFDMA systems," *IET Comm.*, vol. 2, no. 6, pp. 815–826, July 2008.

[68] Y. Ma, D. I. Kim, and Z. Wu, "Optimization of OFDMA-based cellular cognitive radio networks," *IEEE Transactions on Communications*, vol. 58, no. 8, pp. 2265 –2276, Aug. 2010.

[69] S. Boyd and L. Vandenberghe, *Convex Optimization*, 8th ed. New York, NY: Cambridge University Press, 2004.

[70] P. Tsiaflakis, I. Necoara, J. Suykens, and M. Moonen, "Improved dual decomposition based optimization for DSL dynamic spectrum management," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2230 –2245, April 2010.

[71] F. Facchinei and J. Pang, *Finite-Dimensional Variational Inequalities and Complementary Problem*, 9th ed. New York, NY: Springer-Verlag, 2003.

[72] J.-S. Pang and M. Fukushima, "Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games," *Computational Management Science*, 2005.

[73] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1310 –1322, July 2006.

[74] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Massachusetts: Athena Scientifics, 2003.

[75] Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical Programs with Equilibrium Constraints*.

[76] W. W. Hogan, "Point-to-set maps in mathematical programming," *SIAM Review*, vol. 15, July 1973.

[77] IEEE 802.16m, "Evaluation methodology document," Tech. Rep.

[78] J.-F. Cheng, "Coding performance of hybrid ARQ schemes," *IEEE Transactions on Communications*, vol. 54, no. 3, p. 574, March 2006.

[79] P. Wu and N. Jindal, "Performance of hybrid-ARQ in block-fading channels: A fixed outage probability analysis," *IEEE Transactions on Communications*, vol. 58, no. 4, pp. 1129 –1141, April 2010.

[80] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *Information Theory, IEEE Transactions on*, vol. 47, no. 5, pp. 1971 –1988, July 2001.

[81] A. Chuang, A. Guillen I Fabregas, L. Rasmussen, and I. Collings, "Optimal throughput-diversity-delay tradeoff in MIMO ARQ block-fading channels," *Information Theory, IEEE Transactions on*, vol. 54, no. 9, pp. 3968 –3986, Sept. 2008.

[82] H. Zheng and H. Viswanathan, "Optimizing the ARQ performance in downlink packet data systems with scheduling," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 2, pp. 495 – 506, March 2005.

[83] T. Chaitanya and E. Larsson, "Outage-optimal power allocation for hybrid ARQ with incremental redundancy," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 7, pp. 2069 –2074, July 2011.

[84] J.-F. Cheng, "Coding performance of HARQ with BICM – Part I: Unified performance analysis," in *2010 IEEE 21st Intern. Symposium on Personal Indoor and Mobile Radio Commun. (PIMRC)*, Sept. 2010, pp. 976 –981.

[85] E. Jang, J. Lee, H.-L. Lou, and J. Cioffi, "On the combining schemes for MIMO systems with hybrid ARQ," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 836 –842, Feb. 2009.

[86] J. Lee, H.-L. Lou, D. Toumpakaris, E. Jang, and J. Cioffi, "Transceiver design for MIMO wireless systems incorporating hybrid ARQ," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 32 –40, Jan. 2009.

[87] T. H. Cormen, C. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms, Third Edition*, 2nd ed. Cambridge, Massachusetts: The MIT Press, Sept. 2002.

[88] G. Song and 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.

[89] K. Letaief and Y. J. Zhang, "Dynamic multiuser resource allocation and adaptation for wireless systems," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 38 –47, Aug. 2006.

[90] C. Y. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747 –1758, Oct. 1999.

[91] M. Ergen, S. Coleri, and P. Varaiya, "QoS aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless access

systems," *IEEE Transactions on Broadcasting*, vol. 49, no. 4, pp. 362 – 370, Dec. 2003.

[92] Y.-B. Lin, T.-H. Chiu, and Y. Su, "Optimal and near-optimal resource allocation algorithms for OFDMA networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, pp. 4066 –4077, Aug. 2009.

[93] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 171 – 178, Feb. 2003.

[94] H.-W. Lee and S. Chong, "Downlink resource allocation in multi-carrier systems: frequency-selective vs. equal power allocation," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3738 –3747, Oct. 2008.

[95] Z. Shen, J. Andrews, and B. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2726 – 2737, Nov. 2005.

[96] Y. Ma, "Rate maximization for downlink OFDMA with proportional fairness," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 5, pp. 3267 –3274, Sept. 2008.

[97] T. Wang and L. Vandendorpe, "Iterative resource allocation for maximizing weighted sum min-rate in downlink cellular OFDMA systems," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 223 –234, Jan. 2011.

[98] W. Rhee and J. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *2000 IEEE 51st Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo.*, vol. 2, 2000, pp. 1085 –1089 vol.2.

[99] I. Kim, I.-S. Park, and Y. Lee, "Use of linear programming for dynamic subcarrier and bit allocation in multiuser OFDM," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 4, pp. 1195 –1207, July 2006.

[100] S. Boyd and J. Mattingley, "Branch and bound methods," Lecture Notes, 2007.

[101] M. Dorigo, M. Birattari, and T. Stützle, "Ant colony optimization – Artificial ants as a computational intelligence technique," *IEEE Compt. Intell. Mag.*, vol. 1, pp. 28–39, 2006.

[102] E. Bonabeau, M. Dorigo, and G. Theraula, *Swarm Intelligence: From Natural to Artificial Systems*, 1st ed. Oxford University Press, 1999.

[103] J. Deneubourg, S. Aron, S. Goss, and J. Pasteels, "The selforganizing exploratory pattern of the argentine ant," *Journal of Insect Behavior*, vol. 3, p. 159, 1990.

[104] S. Goss, S. Aron, J. Deneubourg, and J. M. Pasteels, "Self-organized shortcuts in the argentine ant," *Naturwissenschaften*, vol. 76, pp. 579–581, 1989.

List of Publications

1. R. Andreotti, I. Stupia, V. Lottici, F. Giannetti, L. Vandendorpe, "Goodput-based Link Resource Adaptation for Reliable Packet Transmissions in BIC-OFDM Cognitive Radio Networks", *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2267-2281, May 2013. DOI 10.1109/TSP.2013.2250970.

2. R. Andreotti, V. Lottici, F. Giannetti, I. Stupia, L. Vandendorpe, "A Game Theoretical Approach for Reliable Packet Transmission in Noncooperative BIC-OFDM Systems", *IEEE International Conference on Communication* (ICC 2013), Budapest, Hungary, 2013.

3. I. Stupia, R. Andreotti, V. Lottici, L. Vandendorpe, "A Consensus Approach for Cooperative Communications in Cognitive Radio Networks, *European Signal Processing Conference* (EUSIPCO 2012), Bucharest, Romania, 2012.

4. I. Stupia, L. Vandendorpe, R. Andreotti, V. Lottici, "A Game Theoretical Approach for Coded Cooperation in Cognitive Radio Networks", *5th International Symposium on Communications, Control and Signal Processing* (ISCCSP 2012), Rome, Italy, 2012.

5. R. Andreotti, V. Lottici, F. Giannetti, I. Stupia, L. Vandendorpe, "Optimal and sub-optimal power allocation algorithms for goodput optimization in cognitive multi-channel wireless systems", *32nd WIC Symposium on Information Theory in the Benelux* (WIC-SP 2011), Brussels, 2011.

6. F. Giannetti, I. Stupia, V. Lottici, R. Andreotti, N. A. DAndrea, L. Vandendorpe, "A performance prediction model for BIC-OFDM transmissions with AMC over nonlinear fading channels", *8th International Workshop on Multi-carrier Systems and Solutions* (MC-SS 2011), Herrsching, Germany, 2011.

7. I. Stupia, R. Andreotti, V. Lottici, F. Giannetti, "Adaptive Hybrid ARQ for Goodput Optimization in BIC-OFDM Systems", *8th International Workshop on Multicarrier Systems and Solutions* (MC-SS 2011), Herrsching, Germany, 2011.

8. F. Giannetti, I. Stupia, V. Lottici, R. Andreotti, L. Vandendorpe, "Trading-off Troughput vs. Energy in Adaptive BIC-OFDM over Nonlinear Channels: a Novel Approach", *N++ /COST 2100 Joint Workshop on Wireless Communications*, Paris, France, 2011.

9. R. Andreotti, I. Stupia, F. Giannetti, V. Lottici, L. Vandendorpe, "Resource Allocation in OFDMA Underlay Cognitive Radio Systems Based on Ant Colony Optimization", *IEEE International Workshop on Signal Processing Advances for Wireless Communications* (SPAWC 2010), pp 1-5, Marrakesh.

10. I. Stupia, F. Giannetti, V. Lottici, R. Andreotti, L. Vandendorpe, N. A. D′Andrea, "A Greedy Algorithm for Goodput-Oriented AMC in Turbo-Coded OFDM", *Future Network & Mobile Summit 2010* (FUNEMS 2010), pp 1-8, Firenze, 2010.