Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

Télécom ParisTech

Spécialité : « Électronique et Communications »

présentée et soutenue publiquement par

Sébastien MARCILLE

Soutenance prévue en février 2013

Allocation de ressources pour les réseaux ad hoc mobiles basés sur les protocoles HARQ

Directeur de thèse : **Philippe CIBLAT**
Co-directeur de thèse : **Christophe LE MARTRET**

T
H
È
S
E

Télécom ParisTech

école de l'Institut Télécom – membre de ParisTech

Contents

List of Acronyms

3GPP   3rd Generation Partnership Project

AMC   Adaptive Modulation and Coding

ARQ   Automatic Repeat reQuest

AWGN   Additive White Gaussian Noise

ACK   ACKnowledgment

BEC   Binary Erasure Channel

BER   Bit Error Rate

BI   Binary Input

BICM   Bit Interleaved Coded Modulation

CC   Chase Combining

CDMA   Code Division Multiple Access

CH   Cluster Head

CRC   Cyclic Redundancy Check

CSI   Channel State Information

ED   Early-Drop

EDGE   Enhanced Data Rates for GSM Evolution

FBS   Fragment-Based Strategy

FDD   Frequency Division Duplex

FDMA   Frequency Division Multiple Access

FEC   Forward Error Correction

FH   Frequency Hopping

GBN   Go-Back-N

GNR   Gain to Noise Ratio

GOP   Global OPtimization

GPRS   General Packet Radio Service

HARQ   Hybrid ARQ

HSPA   High Speed Packet Access

IBS   IP-Based Strategy

IP   Internet Protocol

IR   Incremental Redundancy

KKT   Karush-Kuhn-Tucker

LDPC	Low Density Parity Check
LLR	Log Likelihood Ratio
LTE	Long Term Evolution
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
ML	Maximum Likelihood
NACK	Negative ACKnowledgment
NET	Network
OFDM	Orthogonal Frequency Division Multiplex
OFDMA	Orthogonal Frequency Division Multiple Access
OSI	Open Systems Interconnection
PER	Packet Error Rate
PHY	Physical
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RCPC	Rate-Compatible Punctured Convolutional
RCS	Report Credit Strategy
RTT	Round-Trip Time
SNR	Signal to Noise Ratio
SR	Selective Repeat
SW	Stop and Wait
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
UMTS	Universal Mobile Telecommunications System
UWB	Ultra Wide Band
MANET	Mobile Ad Hoc Network

General Introduction

The work presented in this Ph.D. thesis has been produced thanks to the collaboration between the "Communications et Electronique" (COMELEC) department of the Intitut Mines-Télécom / Télécom ParisTech (Paris, France) and the "Systèmes Numériques Embarqués" (SNE/SPM) division of Thales Communications & Security (Gennevilliers, France), within the framework of "Convention Industrielle de Formation par la REcherche" (CIFRE). The thesis started in January 2010.

Problem statement

Ad hoc networks have become an important research field in the wireless digital communications area for the last years. Indeed, unlike cellular communication systems, ad hoc networks do not need any infrastructure and are thus a highly flexible solution for fast and short-lived communications deployment for many situations, from operational military deployments or other critical scenarios to future smart networks. Different strategies can be envisaged to communicate efficiently in a Mobile Ad Hoc Network (MANET). The approach that is retained in this thesis is to divide the network into several clusters, and to organize orthogonal communication schemes to separate the pairs of users inside a cluster. Though suboptimal from an information theoretic point of view, this solution can be easily implemented. Typically, Orthogonal Frequency Division Multiple Access (OFDMA), which combines the so-called Orthogonal Frequency Division Multiplex (OFDM) technique to combat inter-symbol interference due to multipath spread, and Frequency Division Multiple Access (FDMA) to separate the users, has been widely considered since it is a promising solution for future wireless standards.

The problem of resource sharing in orthogonal schemes has been largely addressed in the literature. However, the lack of structure in MANETs makes the resource management difficult compared to cellular systems. This can be mitigated thanks to the clustered organization that provides a centralized coordination of the pairwise communications. Furthermore, it is difficult to provide reliable channel state information at the fusion center of the MANET due to the feedback latency. The recent success of Hybrid ARQ (HARQ) techniques in 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE) makes

these retransmission techniques attractive for enforcing the link performance. Moreover, the main objective of this thesis is to perform **the resource allocation of HARQ-based OFDMA clustered MANETs using only long-term statistics**. Furthermore, in order to cope with industry constraints, we have considered practical coding schemes instead of infinitely long codewords coming from information theory. More precisely, the purpose is to **design and analyze algorithms that optimize the assignment of power, bandwidth, modulation order, and code rate, of the HARQ mechanism at the top of our multiuser communication scheme**.

The beginning of the thesis is dedicated to the study of HARQ performance extending the work initiated in the Ph.D. dissertation of [Le Duc, 2009] taking into account more realistic assumptions. In particular, the latency in MANETs can lead to delayed and corrupted feedback for HARQ. Therefore, **an in-depth study of cross-layer HARQ mechanisms with imperfect feedback** has been conducted in this thesis. In addition, we have also studied the gain in performance brought by the use of the Early-Drop (ED) along with the cross-layer HARQ schemes.

Outline and contributions

This Section depicts the thesis outline and gives some insights on the main contributions. The thesis is organized into six Chapters: the three first are related to the study of HARQ performance, whereas the three last focus on the application of HARQ to resource allocation problems.

In Chapter 1, we give the fundamental notions for the study of HARQ that will be useful until the rest of the thesis. This Chapter is divided into three parts. Firstly, we describe the HARQ mechanisms and briefly review the state of the art. Next, we define the metrics used to measure HARQ performance for each of the three major ways to implement retransmission mechanisms. Finally, the application of HARQ in modern packet-oriented systems and the associated cross-layer approaches, as well as the existing analytic performance, are presented.

A slight improvement of an existing HARQ cross-layer scheme, called ED, is studied in Chapter 2. After a precise description of this technique that is well adapted to fragmented packets, we show its effect on the HARQ performance. New closed-form expressions are developed for the HARQ efficiency, which is the only performance metric affected by ED. Numerical examples reveal that the major gain of the ED is reached when the number of fragments of the IP packet is close to the maximum number of allowed transmissions.

Chapter 3 is devoted to the analysis of cross-layer HARQ schemes with imperfect feedback. We propose a model for two kinds of feedback impairments: errors in the acknowledgment messages, and delayed feedback. New analytic expressions are derived for the main performance metrics of two HARQ schemes, and numerical results show that imperfect feedback has a great impact on the performance of the cross-layer HARQ scheme, whereas it has no influence on the error rate of the noncross-layer one. It is thus of great interest to design cross-layer schemes that still have a gain but are also robust to imperfect feedback, and we propose the definition of a new cross-layer scheme that generalizes the existing one. The analysis of its performance is conducted within a unified framework and we show, using numerical examples, that there is a trade-off between robustness against imperfect feedback and cross-layer gain that can be adjusted by the initial credit distribution of our proposed solution.

Chapter 4 introduces the second part of the thesis, which is dedicated to the resource allocation of HARQ schemes in the paradigm of ad hoc networks. We discuss the design choices and the main assumptions concerning the clustered MANET, which impose to work with channel statistics only. Then, we define the main optimization problem for total cluster power minimization under some Quality of Service (QoS) constraints, which is the mathematical formulation of the Type-I HARQ-based OFDMA clustered MANET resource allocation with statistical Channel State Information (CSI) only. The main originality of this optimization problem is to rely on HARQ measurable performance metrics (such as Packet Error Rate (PER), delay, efficiency, ...) instead of channel capacity. After reviewing the related state of the art, we specify two different implementations of the Physical (PHY) layer in order to take into account several practical constraints: finite-length Gaussian codes and existing modulations and Forward Error Correction (FEC) codes. Four optimization problems are derived from these assumptions and are treated in the two remaining Chapters.

The case of finite-length Gaussian codes is done in Chapter 5, which gives the best performance that one can expect from the proposed clustered OFDMA network using Type-I HARQ. The Chapter is organized into four parts. We firstly compute the distribution of the mutual information of the Rayleigh channel with finite size Gaussian inputs. Then, the error probability of Gaussian codes with finite length over the Rayleigh channel is obtained in closed-form as a byproduct. Based on this new result, we are able to find the optimal resource allocation under minimum rate constraint using an original algorithm from the literature. Finally, the performance of this algorithm are studied through numerical simulations. The framework developed in this Chapter can serve as a basis for Type-I HARQ based OFDMA resource allocation when powerful FEC coding is used.

Finally, in Chapter 6 we develop another framework that is better suited to non-

capacity achieving coding schemes. The Chapter is divided into four parts. We firstly describe the model used for a practical set of Modulation and Coding Scheme (MCS)s. Then, the three remaining optimization problems are solved successively within each part. We begin with the power minimization under rate constraints in the second part, for which optimal solutions are provided when the MCS is fixed, and the optimal MCS selection is addressed next. In the third part, error rate constraints are considered in addition to rate constraints, and optimal as well as practical solutions are proposed. Finally, delay constrained are added to the rate constraints in the fourth part, where optimal solutions are partially characterized and suboptimal but efficient algorithms are discussed.

Publications

Peer-reviewed Journal

J1. C.J. Le Martret, A. Le Duc, S. Marcille and P. Ciblat: "Analytical performance derivation of Hybrid ARQ schemes at IP layer", *IEEE Transactions on Communications*, vol. 60, no. 5, pp. 1305-1314, May 2012.

J2. S. Marcille, P. Ciblat, and C.J. Le Martret: "Resource Allocation for Type-I HARQ based Wireless Ad Hoc Networks", *IEEE Wireless Communications Letters*, vol. 1, no. 6, pp. 597-600, December 2012.

International Conference

C1. S. Marcille, P. Ciblat, and C.J. Le Martret: "Early-Drop based Hybrid ARQ in a Cross-layer context", *in proc. of 22nd IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Toronto (Canada), September 2011.

C2. S. Marcille, P. Ciblat, and C.J. Le Martret: "Performance computation of cross-layer Hybrid ARQ schemes at IP layer in the presence of corrupted acknowledgments", *in proc. of 3rd IEEE International Workshop on Cross-Layer Design (IWCLD)*, Rennes (France), December 2011.

C3. S. Marcille, P. Ciblat, and C.J. Le Martret: "Stop-and-Wait Hybrid ARQ performance at IP level under imperfect feedback", *in proc. of 76th IEEE Vehicular Technology Conference (VTC Fall)*, Québec City (Canada), September 2012.

C4. S. Marcille, P. Ciblat, and C.J. Le Martret: "Optimal resource allocation in HARQ-based OFDMA wireless networks", *in proc. of IEEE Military Communications Conference (MILCOM)*, Orlando (Florida), October 2012.

C5. S. Marcille, P. Ciblat, and C.J. Le Martret: "On OFDMA resource allocation for delay constrained HARQ systems", *in proc. of 46th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove (California), November 2012.

C6. S. Marcille, P. Ciblat, and C.J. Le Martret: "A robust cross-layer HARQ scheme for imperfect feedback context", *in proc. of 46th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove (California), November 2012.

French Conference

C7. S. Marcille, P. Ciblat, et C.J. Le Martret: "Etude au niveau IP d'un protocole ARQ Hybride avec voie de retour imparfaite", *XXIIIème Colloque GRETSI*, Bordeaux (France), September 2011.

Patents

P1. S. Marcille, C.J. Le Martret, P. Ciblat: "Procédé de retransmission de paquets fragmentés", no. 11/03948.

Chapter 1

An overview of Hybrid ARQ techniques

1.1 Introduction

The recent success of the 3GPP LTE standard [Sesia et al., 2009] has exposed HARQ techniques as promising solutions to improve future high data rates mobile systems. It has been adopted from the beginning of 2G evolution General Packet Radio Service (GPRS) and Enhanced Data Rates for GSM Evolution (EDGE), has been part of the corner stones of the High Speed Packet Access (HSPA) modes in 3G Universal Mobile Telecommunications System (UMTS) cellular standards, and it is still included in the most recent towards-4G standards, like IEEE 802.16m (WiMAX) or LTE.

This first Chapter gives the fundamentals for the study of HARQ that will be done in the rest of the thesis. Although a complete overview of all the contributions made for HARQ is out of the scope, the materials introduced in this Chapter are necessary to understand the work that will be presented throughout this thesis.

The Chapter is organized as follows. In Section 1.2, a state of the art of HARQ is given from the retransmission techniques first ideas, to the latest technologies. Section 1.3 details the three major ways to implement retransmission mechanisms, and presents an important discussion on the metrics used to measure HARQ performance in terms of rates. Finally, the application of HARQ in modern packet-oriented systems, and the associated cross-layer approaches, are introduced in Section 1.4.

1.2 From Automatic Repeat reQuest (ARQ) to Hybrid ARQ

1.2.1 ARQ

Automatic Repeat reQuest (ARQ) ideas go back to 1940s with the invention of Van Duuren [Van Duuren, 1943] (as reported in [Schwartz, 1963]), and is based upon a feedback

mechanism that informs the transmitter whether a transmitted packet is correctly received or not. An ACKnowledgment (ACK) or a Negative ACKnowledgment (NACK) is sent back to the transmitter accordingly.

More precisely, ARQ is implemented using an error detection code (usually Cyclic Redundancy Check (CRC)) and retransmits the current data packet upon error detection at the receiver, as depicted in Fig. 1.1. The transmitter of pure ARQ systems encodes the data with the CRC, and then transmits the resulting packet onto the noisy channel. The receiver checks whether the received packet has been corrupted by the channel or not, and feeds back to the transmitter ACK or NACK accordingly. If a NACK occurs, the receiver discards the received packet. In that case, the transmitter receives NACK and retransmits the same packet. Otherwise, the receiver releases the decoded packet and the protocol starts again with the next data packet.
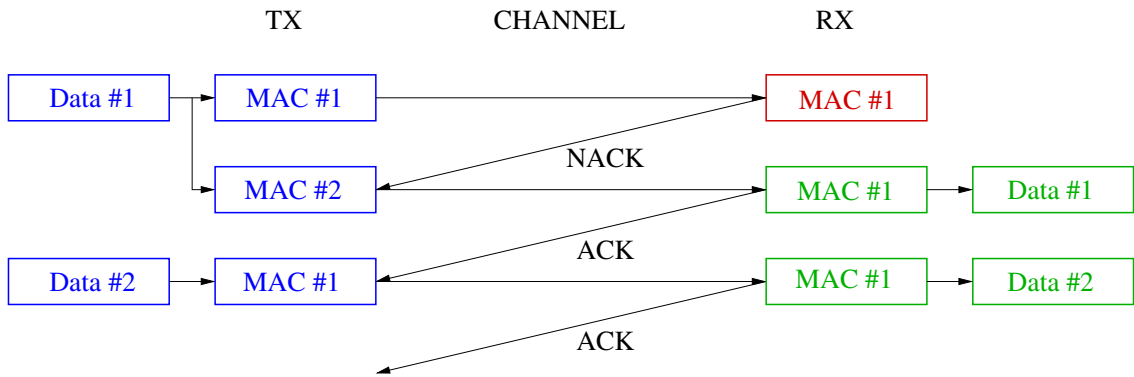
Figure 1.1: A general ARQ scheme.

By construction, ARQ makes the transmission link error-free but at the expense of some extra delay. A trade-off between packet error probability and packet transmission delay can be obtained with the so-called truncated version of ARQ [Adachi et al., 1989], which allows each packet to be transmitted at most L times. The parameter L is called the transmission credit (also known as persistence). ARQ can be implemented at any layer to improve its reliability, and has been used for instance at Medium Access Control (MAC) layer and in the Transmission Control Protocol (TCP) stack of Transport layer, where ARQ is implemented to ensure the reliability of the delivered packets that can be dropped by the routers.

1.2.2   Hybrid ARQ

The main issue with pure ARQ is that the transmitted packets are constituted by the information bits, *i.e.* there is no protection against the channel errors during the transmission. Since the reliability offered by ARQ relies on repetition of the information data, it is very efficient in good channel conditions, *i.e.* when not retransmitting to often. Alternatively,

using FEC coding enables to recover the packet in more and more difficult channel conditions by decreasing the coding rate; however, when the channel is good the overhead due to FEC rate penalizes the transmission efficiency.

Several solutions have been proposed to overcome this issue by using FEC techniques along with the packet repetition mechanism, which gave birth to HARQ. A good historical review of the key papers concerning (H)ARQ is given in [Lin and Costello, 1983]. HARQ is classified into two types depending on the error correcting capability within each transmission of ARQ, according to [Schmitt, 2002]. An equivalent classification is to determine whether there is memory at the receiver side, for packet recombination purposes.

1.2.2.1 Type-I HARQ

Type-I HARQ describes an ARQ mechanism for which the data packet, after CRC encoding, is encoded by a FEC code of given rate R_0. At the receiver side, the received packet is still discarded if NACK occurs. Hence, Type-I defines HARQ schemes with constant error correction capability along the retransmissions. Equivalently, it also describes HARQ schemes for which there is no memory processing at the receiver. The main interest of Type-I HARQ is to use the correction capability of the FEC in order to recover the information bits in more noisy conditions, and to decrease the retransmission probability of the underlying ARQ. However, when the channel is good the coding rate R_0 decreases the amount of received information bits inside each accepted packet.

1.2.2.2 Type-II HARQ

Type-II HARQ gives a satisfying solution to the drawback of Type-I schemes by introducing memory and processing at the receiver. The main difference between Type-II and Type-I schemes is that Type-II performs combining of the multiple packets received within each ARQ transmission, which allows to increase the correction capability of the code (hence more powerful coding gains). Type-II HARQ thus automatically adapts the code rate to the current channel conditions. In practice, the discovery of code combining [Chase, 1985] (better known as Chase Combining) and of rate-compatible codes [Hagenauer, 1988; Kim et al., 2006] enables substantial gains in received information bits per accepted packet.

Chase Combining (CC) Type-II HARQ with CC, or CC-HARQ for short, is a scheme where the same encoded packet is retransmitted if requested. At the transmitter, the same operations are done as for Type-I HARQ, *i.e.* encoding of the data with a code of fixed rate R_0. However, if a NACK occurs, the packet is kept at the receiver. Then, the transmitter retransmits the same encoded packet, which is combined at the receiver to the

previous packets in memory, using the so-called CC scheme[1]. The coding gain brought by CC comes from the increasing correction capability at each retransmission: at the k-th transmission, CC yields a virtual coding rate $R_k = R_0/k$.

Incremental Redundancy (IR)   Type-II HARQ with IR, or IR-HARQ for short, is a scheme where the redundancy is sent piecewise upon error detection. IR-HARQ is the most versatile scheme, and gives the best compromise between ARQ and FEC by finding the coding rate that is adapted to channel conditions.

At the transmitter, after CRC encoding of the data, the packet is encoded by a mother code of rate R_0 and, usually following a puncturing scheme, is split into a sequence of t_0 increments. The transmitter transmits sequentially the first increment up to the t_0-th increment upon error detection. If the data still cannot be decoded after the transmission of the t_0-th increment, the first increment of the sequence is transmitted again and so on.

At the receiver side, the first increment of the sequence is simply decoded and the HARQ process checks if the data can be recovered without error or not. If an ACK occurs, the transmitter restarts the HARQ process with the next data packet. Otherwise, the next increment in the sequence is transmitted over the channel. Its received version is combined to the previously received packets of the sequence, the aggregation of packets is decoded, ACK/NACK is sent back, and so on until the reception of the t_0-th increment, resulting in the decoding of the mother code of rate R_0. If the data is still detected in error and the transmission limit is not reached, the receiver can choose between several strategies as depicted in [Le Duc, 2009, Sec. 1.3.2] and the process starts again. The increasing correction capability given by IR is the result of the decreasing code rate obtained after combination of the increments at the receiver.

1.3   The retransmission protocols

ARQ systems can be derived in several protocols, according to how the data packets are scheduled in relation to the feedback. According to [Lin and Costello, 1983] there exists three basic ARQ protocols: the Stop and Wait (SW), the Go-Back-N (GBN) and the Selective Repeat (SR). Although these three protocols achieve the same level of reliability, their performance in terms of the amount of out-coming data related to incoming data can be very different. In this Section, we will review the internal mechanism of the three protocols as well as their performance. The performance of ARQ schemes are often given in terms of **throughput** [Lin and Costello, 1983; Wicker, 1995]. We first define and make clear what will be called throughput in the sequel.

---

[1]Basically, code combining corresponds to the maximal ratio combining of the Log Likelihood Ratio (LLR) entering into the soft channel decoder.

1.3.1 Throughput, efficiency, and their byproducts

The throughput efficiency has been defined in [Lin and Costello, 1983] as "the ratio of the average number of information bits successfully accepted by the receiver per unit of time to the total number of bits that could be transmitted per unit of time". This definition is reduced to "throughput" in the rest of the book, and the throughput definition has often been misleading in the literature. In what follows, several measure definitions are reviewed and their interrelation is studied. We hope that it may help to construct a comprehensive study of (H)ARQ.

1.3.1.1 Efficiency

The efficiency, denoted by η, has been defined in the thesis work of [Le Duc, 2009][2] as the long-term ratio between the number of received information bits and the number of transmitted bits:

$$\eta := \lim_{b \to \infty} \frac{1}{b} \sum_{k=1}^{N(b)} I_k, \tag{1.1}$$

where $\{N(b), b \in \mathbb{N}\}$ is a discrete-time stochastic process that counts the number of successful packet decoding up to bit transmission b, and I_k is the number of information bits received whenever a packet is decoded with success. Efficiency has no dimension, $\eta \in [0, 1]$ and is interpreted as the long-term average number of bits received per transmitted bit. Using renewal theory [Ross, 2007, Chap. 7], it was found in [Le Martret et al., 2012] that:

$$\eta = \frac{\mathbb{E}[I]}{\mathbb{E}[B]}, \tag{1.2}$$

where I is the random number of received information bits per successful packet, and B is the random number of transmitted bits between two successive packets success.

1.3.1.2 Throughput

The throughput τ is the information bitrate measured in bit/s at the system output. The Round-Trip Time (RTT), defined as the time spent between the transmission of a packet and the reception of its corresponding acknowledgment at the transmitter side, plays a big role in the throughput evaluation. For slotted systems with equal length packets, we denote by $T \geq 0$ the average number of transmissions that could occur during the RTT. Different protocols lead to different RTT values and thus to different throughput measures for the same efficiency. The throughput is proportional to the efficiency multiplied by the raw bitrate $m D_s$ obtained from the number m of bits in a constellation symbol and the symbol rate D_s (in symb/s), hence:

$$\tau = \frac{m \eta}{1 + T} D_s \quad \text{(bit/s)}. \tag{1.3}$$

---

[2]http://pastel.archives-ouvertes.fr/docs/00/55/93/22/PDF/TheseAudeLeDucCouleurs_TitreFrancais.pdf

When the signals are transmitted through a bandwidth W, we can link the throughput to W by introducing an equivalent spectral efficiency ρ:

$$\rho = \frac{\tau}{W} \qquad \text{(bit/s/Hz)}. \tag{1.4}$$

1.3.1.3 Throughput efficiency

The throughput efficiency, which captures the absorption of bits by the system, is the throughput normalized by the maximum bitrate at which the system can transmit $m\,D_s$. Thus, the throughput efficiency τ_{eff} is:

$$\tau_{\text{eff}} := \frac{\tau}{m\,D_s} = \frac{\eta}{1+T}. \tag{1.5}$$

The throughput efficiency has no dimension and $\tau_{\text{eff}} \in [0,1]$.

1.3.1.4 Goodput

Coming from the queuing theory community (see [Doshi and Heffes, 1986] or [Berger, 1991]), the goodput originally measured the quantity of error-free data in the throughput of the queue. Nowadays, goodput captures the idea of useful bits received at the system output within each symbol transmission [Devillers et al., 2008]. The goodput $\tilde{\eta}$ is thus proportional to the efficiency:

$$\tilde{\eta} := m\eta \qquad \text{(bit/symb)}. \tag{1.6}$$

Notice that the so-called "throughput" computed in [Caire and Tuninetti, 2001] announced in bit/s/Hz, was in fact expressed in bit/symb along their paper, was actually the goodput. The goodput is proportional to the spectral efficiency given in Eq. (1.4) by a factor $(1+T)$. When using SW with instantaneous RTT or SR, then $T = 0$ and the goodput is equal to the spectral efficiency.

1.3.2 Selective Repeat protocol

The SR continuously sends packets over the channel, even during the RTT, as shown in Fig. 1.2 where $T \geq 0$ is the fixed RTT duration. In an ideal context, *i.e.* with unlimited buffer size, packets are continuously sent and decoded during the RTT and only erroneous packets are requested at the transmitter. However, SR has the most complex receiver, since if packets must be delivered in their incoming order, the receiver must bufferize all the decoded packets that are waiting for a previous one which was in error. For a Type-I HARQ, the throughput efficiency of SR is computed in [Lin and Costello, 1983]:

$$\tau_{\text{eff}}^{\text{SR}} = R(1 - p_0), \tag{1.7}$$

where p_0 is the packet error probability over the channel, and R is the coding rate.
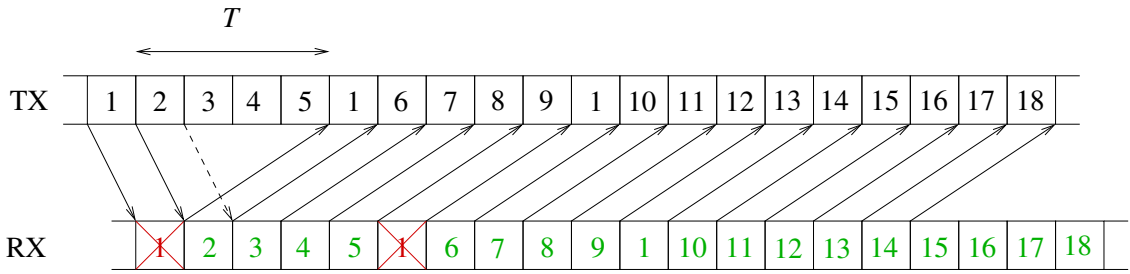
Figure 1.2: The SR protocol.

1.3.3 Go-Back-N protocol

The GBN allows to relax the receiver complexity, at the expense of throughput efficiency. Basically, if a retransmission is requested for a packet, the next N decoded packets are dropped from the receiver memory and the transmitter retransmits them too. Usually $N = 1 + T$, where $T \geq 0$ is the fixed RTT duration, as depicted in Fig. 1.3. Hence, the throughput efficiency is [Lin and Costello, 1983]:

$$\tau_{\text{eff}}^{\text{GBN}} = \frac{\tau_{\text{eff}}^{\text{SR}}}{1 + (N - 1)p_0}. \tag{1.8}$$

Here, we observe a throughput efficiency reduction of $(1 + (N - 1)p_0)$. For $N = 1$, the throughput efficiency equals that of SR.
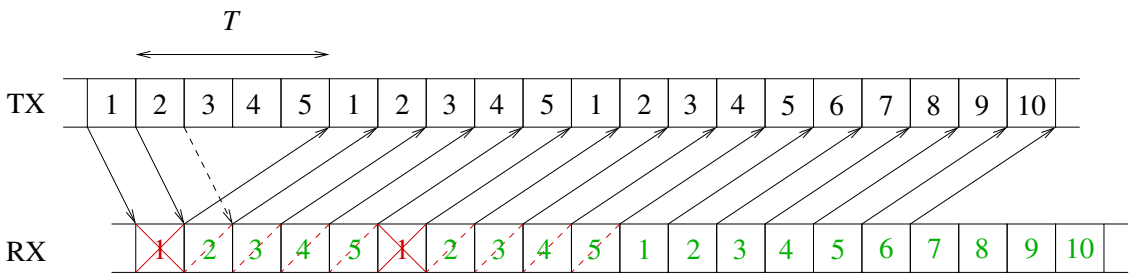
Figure 1.3: The GBN protocol.

1.3.4 Stop and Wait protocol

As the simplest of the three, SW is the worst one too, and is outlined in Fig. 1.4. It consists in keeping the transmitter idle during the RTT of fixed duration $T \geq 0$. Its throughput efficiency is equal to [Lin and Costello, 1983]:

$$\tau_{\text{eff}}^{\text{SW}} = \frac{\tau_{\text{eff}}^{\text{SR}}}{1 + T}. \tag{1.9}$$

When the RTT is assumed to be zero ($T = 0$), then SW, GBN, and SR become equivalent. Otherwise, the SR is the most efficient in terms of throughput efficiency, as it is illustrated by Fig 1.5.
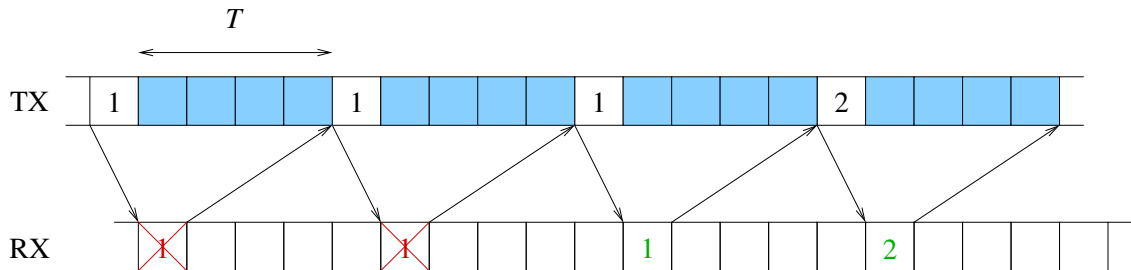
Figure 1.4: The SW protocol.
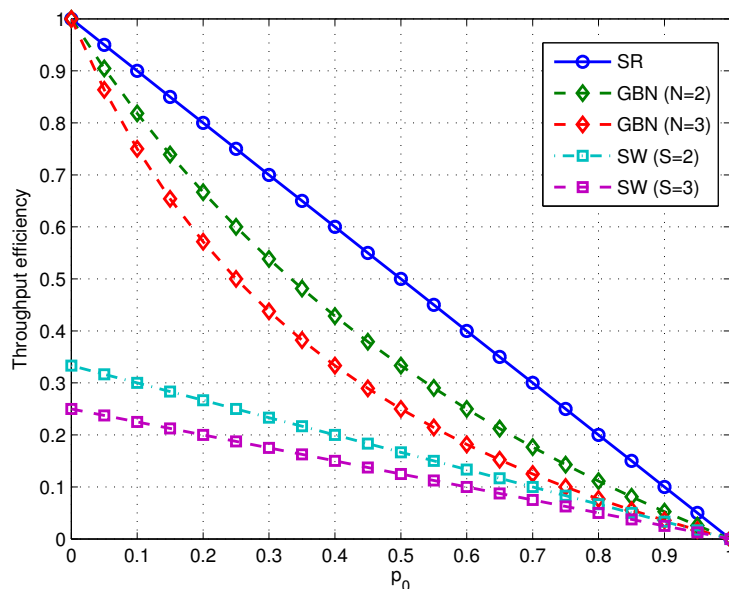
Figure 1.5: The throughput efficiency for SR, GBN and SW (ARQ, $R = 1$) versus p_0.

1.4 Cross-layer HARQ techniques for packet-oriented systems

1.4.1 Layer model

Highly inspired from [Le Duc, 2009], the model considered in this thesis encompasses the three first layers of the seven-layer Open Systems Interconnection (OSI) model, *i.e.* the PHY layer, the MAC layer and the Network (NET) layer. The PHY layer defines the means of transmitting raw bits through a physical propagation medium (the channel),

by describing the shape of the electrical signal, the modulation format, etc. The MAC layer focuses on transferring logical data packets between two nodes in the network, by error checking, and on controlling access to the medium, by carrying out the radio resource management for multiple users. Finally, the NET layer is responsible for packet forwarding in the network, by determining a route from the source node to the destination node, and uses other existing nodes as relays for that. Without loss of generality, the NET protocol will be assumed to be the Internet Protocol (IP) in the rest of the thesis. In the following, let us describe the retransmission process when HARQ is conventionally applied in a layered system, as depicted in Fig. 1.6. At MAC layer, the incoming IP packets of length L_{IP} are assumed to be split into N fragments, of length $L_{\mathrm{MAC}} = L_{\mathrm{IP}}/N$.

Next, each fragment is transmitted following a given HARQ scheme. At the transmitter side, a fragment is first transformed into MAC packet(s) according to the considered HARQ scheme. After adding the MAC overhead, a fragment is possibly encoded by a FEC code and a sequence of t_0 MAC packets is generated. The transmitter transmits sequentially the first MAC packet up to the t_0-th MAC packet upon error detection. If the fragment still cannot be decoded after the transmission of the t_0-th MAC packet, the first MAC packet of the sequence is transmitted again and so on. At PHY layer, the MAC packet ready for transmission is inserted into a frame, modulated according to a given constellation, and sent through the wireless channel.

At the receiver side, the PHY layer demodulates the received signal and pushes forward the resulting MAC packet to the MAC layer. The first MAC packet of the sequence is simply decoded and the HARQ process checks if the fragment can be recovered without errors (using CRC control) or not. The receiver sends back to the transmitter ACK or NACK accordingly. The ACK/NACK is transmitted through the feedback channel, which is assumed ideal in this Chapter. If ACK is received, the transmitter restarts the HARQ process with the next fragment. Otherwise, the next MAC packet in the sequence is transmitted over the channel. Its received version may: either be combined to the previously received packets of the sequence before decoding the aggregation of packets (Type-II); or independently decoded (Type-I). Then ACK/NACK is sent back, and so on until the reception of the t_0-th MAC packet. If the fragment is still detected in error and the transmission limit L is not reached, the receiver memory is flushed and the process starts again. If the last authorized transmission fails, the fragment is dropped and the retransmission process is started again with the next fragment. Finally, once the N fragments have been correctly received, the receiver concatenates them into an IP packet that is released to the NET layer. If at least one fragment is missing, the resulting IP packet is dropped by the reassembly process and is not delivered to the NET layer.
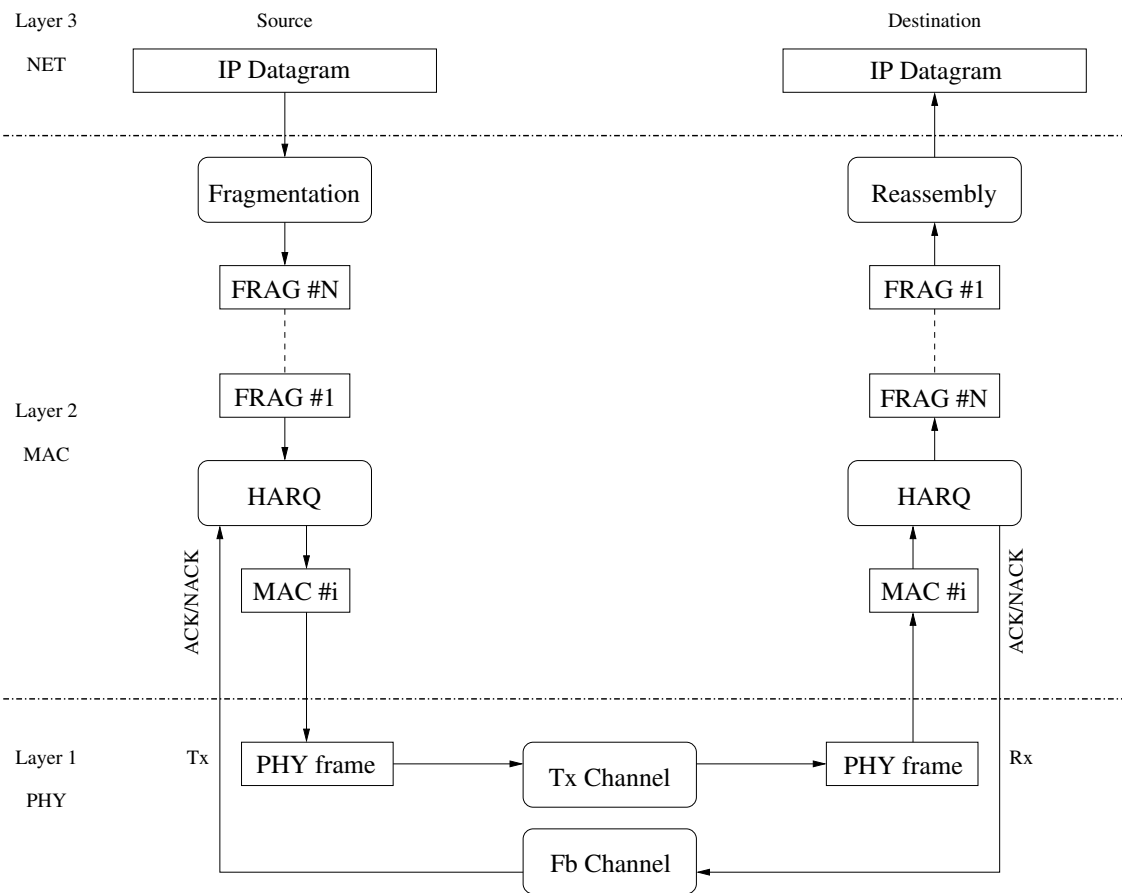
Figure 1.6: Layer model.

1.4.2 Definition of the HARQ performance metrics

Various performance metrics were identified in [Le Duc, 2009] and [Le Martret et al., 2012] in order to give a complete view of the system performance:

- the PER,

- the delay,

- the efficiency.

Although the conventional HARQ design was mainly done at MAC layer in the past (as seen in Section 1.2), the new packet-oriented systems of 4G standards run the IP. It is thus of great interest to study the performance at the NET layer, as proposed in [Rossi and Zorzi, 2003] and [Le Martret et al., 2012], in order to enlarge the vision of practical systems performance. For a notational convenience, a subscript 'IP' (resp. 'MAC') will stand for the metrics defined at NET (resp. MAC) level.

1.4.2.1 Packet Error Rate

The PER, denoted by P, is the probability that an information packet is not transmitted with success within the HARQ process. According to the previous notation, P_{MAC} is the PER of the fragments, whereas if the IP packets are of interest we will use P_{IP} instead. Therefore:

$$P_{\mathrm{MAC}} := \Pr\left\{\text{fragment not successfully received}\right\}, \tag{1.10}$$

$$P_{\mathrm{IP}} := \Pr\left\{\text{IP packet not successfully received}\right\}. \tag{1.11}$$

1.4.2.2 Delay

The delay will be denoted d, and is defined as the average number of MAC packets that have been transmitted, knowing that the information packet is received without errors. Using the previous notations, one can thus obtain:

$$d_{\mathrm{MAC}} := \mathbb{E}\left[\text{\# of MAC packets sent} \mid \text{fragment received without errors}\right], \tag{1.12}$$

$$d_{\mathrm{IP}} := \mathbb{E}\left[\text{\# of MAC packets sent} \mid \text{IP packet received without errors}\right]. \tag{1.13}$$

As remarked in [Le Duc, 2009], this delay definition is not proportional to the inverse of the efficiency.

1.4.2.3 Efficiency

The efficiency has already been discussed in Section 1.3.1. Using the proposed notations, we rewrite:

$$\eta_{\mathrm{MAC}} = \frac{\mathbb{E}\left[\text{\# of information bits received per successful fragment}\right]}{\mathbb{E}\left[\text{\# of bits transmitted between two successive error-free fragments}\right]}, \quad (1.14)$$

$$\eta_{\mathrm{IP}} = \frac{\mathbb{E}\left[\text{\# of information bits received per successful IP packet}\right]}{\mathbb{E}\left[\text{\# of bits transmitted between two successive error-free IP packets}\right]}. \quad (1.15)$$

1.4.3 Cross-layer HARQ techniques

As said previously, since the communication systems tend to be built upon the packet-oriented IP, it is important to design HARQ schemes that improve the performance at NET layer. In particular, it can be achieved by following the cross-layer ideas, resulting in schemes such as the MAC-IP retransmission management developed in [Choi et al., 2005]. The conventional retransmission schemes described before are usually applied at MAC layer where HARQ manages the fragments one after the other independently, as shown in Fig. 1.6. With this approach, if the L-th MAC packet transmission fails, the corresponding fragment is dropped and so the corresponding IP packet too.

It can be interesting to take into account for the fact that the fragments come from a same IP packet, as depicted in Fig. 1.7. The PER can be improved at IP level as proposed in [Choi et al., 2005] by granting the total transmission credit $C = NL$ to the set of fragments belonging to the same IP packet. This credit is decremented by 1 at each MAC packet transmission. So if a fragment does not use the C transmissions, then the remaining credit can be used by the next ones, and the IP packet is only dropped if all the corresponding fragments are not received correctly in at most C transmissions. In the sequel we will refer to the conventional strategy (transmission credit per fragment) as Fragment-Based Strategy (FBS) and to the cross-layer strategy of [Choi et al., 2005] as IP-Based Strategy (IBS).

1.4.4 Brief state of the art on HARQ performance expressions

HARQ schemes have been widely studied in the literature with usually two objectives: *i)* the theoretical performance derivation of existing HARQ schemes, and *ii)* the design of good HARQ schemes. This Section gives a short, non exhaustive review of the main contributions to HARQ. For a good overview of first (H)ARQ principles, one can consult [Lin and Costello, 1983] and [Wicker, 1995].

It is of interest to analyze the performance of existing HARQ schemes through closed-form expressions. Hagenauer pioneered the performance study of IR-HARQ as an application of his Rate-Compatible Punctured Convolutional (RCPC) codes presented in
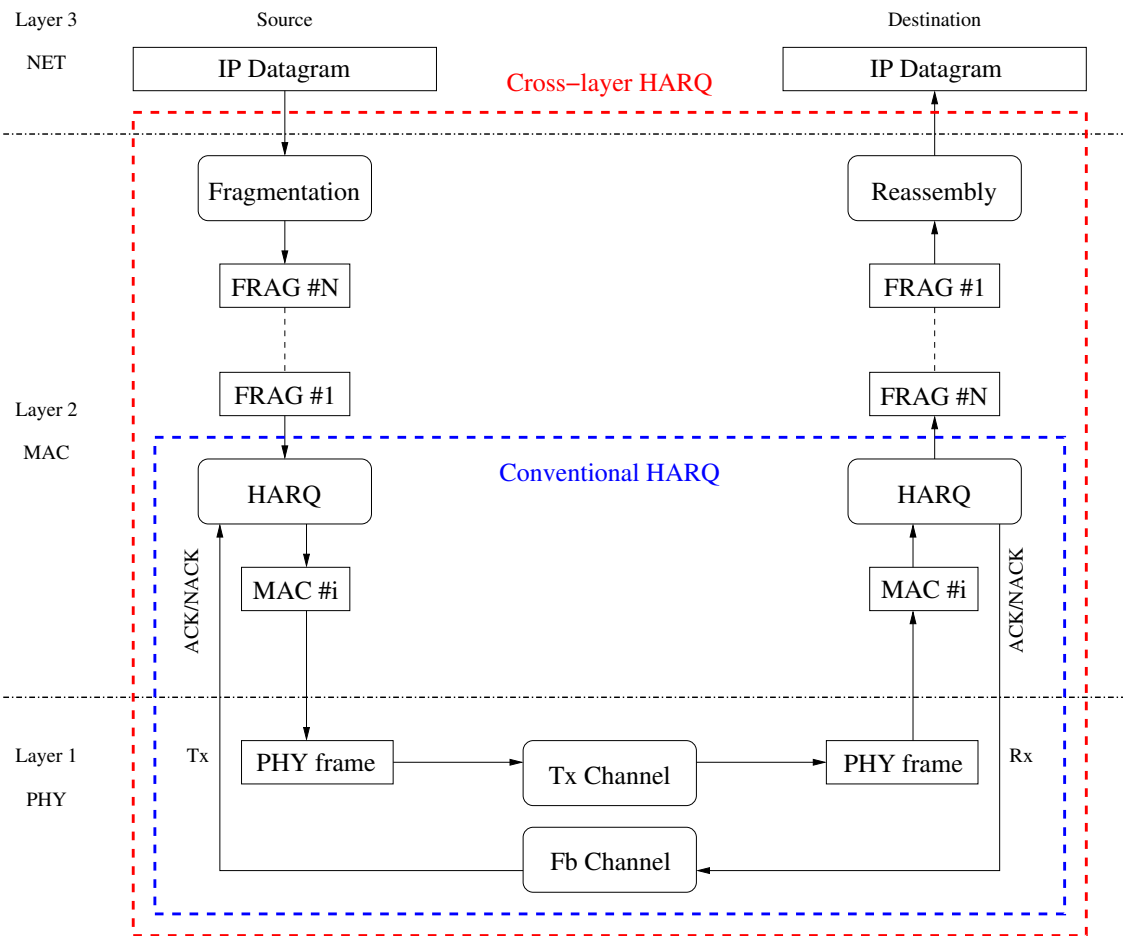
Figure 1.7: Cross-layer model.

[Hagenauer, 1988], whereas [Kallel, 1990] was probably the first to analyze the performance of CC-HARQ, though this scheme was already mentioned by [Chase, 1985] as an application of his code combining. CC-HARQ has been analyzed for interleaved fading channels in [Chen and Fan, 2005], while IR-HARQ was the main purpose of [Levorato and Zorzi, 2008] and [Andriyanova and Soljanin, 2012]. Later, [Cheng, 2006] tried to answer to the interesting question: what scheme, of CC-HARQ or IR-HARQ, is the best one? It results that the answer depends on the link quality as well as the initial code rate R_0.

Interestingly, the systematic analysis of ARQ came after that Type-II HARQ schemes were already under investigation: the framework given by [Zorzi and Rao, 1996] for the analysis of ARQ has revealed to be very fruitful for future investigations. Their work served as the basis for [Caire and Tuninetti, 2001], which presented the ultimate performance of the main HARQ schemes. More recently, these ideas were reused in [Le Martret et al., 2012] to establish the performance of general HARQ schemes at upper layers.

For the rest of the thesis, the following definitions will be useful:

- Let $p_n(k)$, $n \geq 1$, $k \geq n$, be the probability of receiving n fragments without errors in exactly k MAC packets transmissions. We recall that the probabilities $p_n(k)$ are provided in [Le Duc, 2009].

- Let p_j, $j \geq 1$, be the probability that the $(j + 1)$-th MAC packet coming from a given fragment is not decoded, knowing that the j previous MAC packets of the same fragment were not decoded as well. As a particular case, let p_0 be the probability that a MAC packet is not decoded with success.

- Let $q(k)$, $k \geq 1$, be the probability of receiving a fragment with errors after k MAC packet transmissions, hence:

$$q(k) = \prod_{j=0}^{k-1} p_j. \tag{1.16}$$

The expressions that will be presented next were found in [Le Duc, 2009]. For each metric, general expressions that are valid for any HARQ type will be presented first, next simplified expressions that hold true for Type-I will be given. A superscript 'F' (resp. 'I') will stand for the metrics expressed for the FBS (resp. IBS).

1.4.4.1 HARQ with conventional FBS

Packet Error Rate: At MAC level, the error probability of a given fragment is known as:

$$P_{\text{MAC}} = 1 - \sum_{\ell=1}^{L} p_1(\ell), \tag{1.17}$$

and leads to the error probability at IP level:

$$P_{\text{IP}}^{\text{F}} = 1 - (1 - P_{\text{MAC}})^N. \tag{1.18}$$

For Type-I HARQ, $p_1(\ell) = (1 - p_0)p_0^{\ell-1}$ for $\ell \geq 1$, and thus:

$$P_{\text{MAC}} = p_0^L,$$ (1.19)

$$P_{\text{IP}}^{\text{F}} = 1 - (1 - p_0^L)^N.$$ (1.20)

Delay: The delay at MAC level is:

$$d_{\text{MAC}} = \frac{1}{1 - P_{\text{MAC}}} \sum_{\ell=1}^{L} \ell \, p_1(\ell),$$ (1.21)

whereas at IP level it is simply related to d_{MAC}, thanks to the independence of the fragments:

$$d_{\text{IP}}^{\text{F}} = N d_{\text{MAC}}.$$ (1.22)

For Type-I HARQ the delay becomes a simple function of p_0:

$$d_{\text{MAC}} = L + \frac{1}{1 - p_0} - \frac{L}{1 - p_0^L}.$$ (1.23)

Efficiency: A general expression for the efficiency, valid for any retransmission scheme, was given in [Le Duc, 2009]. At any layer l and for any strategy s, the most general efficiency is:

$$\eta_l^s = \frac{L_l \, (1 - P_l^s)}{P_l^s \, \check{d}_l^s \; + \; (1 - P_l^s) \, \hat{d}_l^s},$$ (1.24)

where $\check{d}_l$ (resp. $\hat{d}_l$) is the average number of bits sent knowing that the layer l current packet reception fails (resp. layer l packet has been received without errors). In particular, this expression holds true when the MAC packets are of unequal length. However, due to the high complexity of the $\check{d}$ formula that presents no interest for our dissertation, we report here only the case of equally long MAC packets. At MAC level, neglecting the MAC overhead (CRC length relative to the packet length), the efficiency was expressed as:

$$\eta_{\text{MAC}} = \frac{R(1 - P_{\text{MAC}})}{L \, P_{\text{MAC}} + (1 - P_{\text{MAC}})d_{\text{MAC}}} = \frac{R \sum_{\ell=1}^{L} p_1(\ell)}{L \left(1 - \sum_{\ell=1}^{L} p_1(\ell)\right) + \sum_{\ell=1}^{L} \ell \, p_1(\ell)},$$ (1.25)

where R is the coding rate. As for the delay case, the efficiency can be expressed in simple terms at IP level:

$$\eta_{\text{IP}}^{\text{F}} = \eta_{\text{MAC}}(1 - P_{\text{MAC}})^{N-1}.$$ (1.26)

The efficiency when considering Type-I schemes dramatically simplifies into:

$$\eta_{\text{MAC}} = R \, (1 - p_0).$$ (1.27)

1.4.4.2 HARQ with cross-layer IBS

For the cross-layer IBS, all the metrics are only defined at IP level.

Packet Error Rate: The PER is expressed as a function of the probability $p_N(\ell)$, which is itself a combinatorial function of the elementary probability p_1:

$$P_{\mathrm{IP}}^{\mathrm{I}} = 1 - \sum_{\ell=N}^{C} p_N(\ell) = 1 - \sum_{\ell=N}^{C} \sum_{k \in \mathcal{K}_{\ell,N}} \prod_{n=1}^{N} p_1(k_n), \tag{1.28}$$

where $\mathcal{K}_{m,n} = \left\{ k \in \mathbb{N}_*^n \mid \sum_{j=1}^{n} k_j = m \right\}$ is a combinatorial set of cardinal $\mathrm{Card}(\mathcal{K}_{m,n}) = \binom{m-1}{n-1}$. For Type-I HARQ, we have $p_N(\ell) = \binom{\ell-1}{N-1}(1 - p_0)^N p_0^{\ell-N}$, hence:

$$P_{\mathrm{IP}}^{\mathrm{I}} = 1 - (1 - p_0)^N \sum_{\ell=N}^{C} \binom{\ell-1}{N-1} p_0^{\ell-N}. \tag{1.29}$$

It can be shown that this expression can be simplified to:

$$P_{\mathrm{IP}}^{\mathrm{I}} = I(p_0; C - N + 1, N), \tag{1.30}$$

with $I(x; a, b)$ the normalized incomplete Beta function [Abramowitz and Stegun, 1972, Eq. (8.392)]. The proof is reported in Appendix A.4.

Delay: The delay has the same shape in this case than the delay for FBS. It was found that:

$$d_{\mathrm{IP}}^{\mathrm{I}} = \frac{1}{1 - P_{\mathrm{IP}}^{\mathrm{I}}} \sum_{\ell=N}^{C} \ell \, p_N(\ell), \tag{1.31}$$

which turns, for Type-I HARQ (see Appendix A.5), into:

$$d_{\mathrm{IP}}^{\mathrm{I}} = \frac{N}{1 - p_0} - \frac{(1 - p_0)^{N-1} p_0^\kappa}{\alpha(p_0)}, \tag{1.32}$$

where $\kappa := C - N + 1$ and $\alpha(x) := B(\kappa, N) I(1 - x; N, \kappa)$, with $B(a, b)$ the so-called Beta function [Abramowitz and Stegun, 1972, Eq. (8.390)].

Efficiency: Finally, the efficiency of IBS is given by:

$$\eta_{\mathrm{IP}}^{\mathrm{I}} = \frac{RN(1 - P_{\mathrm{IP}}^{\mathrm{I}})}{C P_{\mathrm{IP}}^{\mathrm{I}} + (1 - P_{\mathrm{IP}}^{\mathrm{I}}) d_{\mathrm{IP}}^{\mathrm{I}}} = \frac{RN \sum_{\ell=N}^{C} p_N(\ell)}{C \left(1 - \sum_{\ell=N}^{C} p_N(\ell) \right) + \sum_{\ell=N}^{C} \ell \, p_N(\ell)} \tag{1.33}$$

and for Type-I schemes by (from Appendices A.4 and A.5 and direct algebraic computations):

$$\eta_{\mathrm{IP}}^{\mathrm{I}} = \frac{RN(1 - p_0)\alpha(p_0)}{(1 - p_0) C B(p_0; \kappa, N) + N\alpha(p_0) - (1 - p_0)^N p_0^\kappa}, \tag{1.34}$$

where $B(x; a, b)$ is the incomplete Beta function [Abramowitz and Stegun, 1972, Eq. (8.391)].

1.5 Conclusion

In this Chapter, we have presented the fundamentals about HARQ that will be useful throughout the thesis. Without being exhaustive, the state of the art presented in this Chapter covers a large amount of the retransmission techniques from the basic concepts of ARQ to most advanced cross-layer HARQ techniques.

This work will be extended in the next two Chapters along the following lines: in Chapter 2 we will study the "early drop" version of IBS which can slightly improve the HARQ efficiency, and in Chapter 3 we will analyze the effects of imperfect feedback by deriving new expressions for the performance metrics.

Chapter 2

An Early-Drop version of cross-layer Hybrid ARQ

2.1 Introduction

New 4G communications standards use an all-IP oriented infrastructure to manage the NET layer. It is thus of interest *i)* to analyze HARQ at NET level, *i.e.*, when the IP packet is the figure of merit, and *ii)* to design HARQ schemes to improve the performance at NET level. A cross-layer ARQ scheme has been designed between the MAC and the NET layers in [Choi et al., 2005]. This scheme, that we reviewed in Chapter 1, improves the PER at NET level and has been extensively studied in a unified framework that extends to HARQ in [Le Duc, 2009].

An improvement of this cross-layer scheme, called ED, has been depicted in [Choi et al., 2005] for ARQ. The basic idea is to stop the IP packet transmission as soon as the number of remaining fragments is larger than the remaining number of transmission attempts. This technique was investigated for ARQ only, and we propose in this Chapter to derive closed-form expressions for the efficiency of the ED based HARQ, for any HARQ type.

The Chapter is organized as follows. The ED is defined in Section 2.2, and next the new expressions of efficiency are computed in Section 2.3. Section 2.4 details some particular cases. The relevance of this technique is finally discussed in Section 2.5 where some numerical results are given.

2.2 Description of the Early-Drop mechanism

The ED technique, applied to an HARQ scheme using IBS, allows the transmission of a given IP packet to be stopped as soon as it is detected that there is not enough credit left in order to transmit the remaining fragments, as depicted in Fig. 2.1. To be more precise, the transmitter discards the IP packet at the j-th fragment if the remaining transmission

credit is less than $(N - j)$. Let $i_k \in \mathbb{N}_*$ be the number of transmissions used by the k-th fragment, and let $\boldsymbol{i} \in \mathbb{N}_*^n$ be a vector with elements i_k that represents a combination of the transmissions done for n fragments. This yields the following definition:

Definition 2.1 (Early-Drop). *A global credit C is granted to an* IP *packet of N fragments. Let us define by $m_{\boldsymbol{i}}(j) := \sum_{k=1}^{j} i_k$ the number of* MAC *packets transmissions that occurred up to fragment #$j \in \{1, \ldots, N\}$ for the combination $\boldsymbol{i}$. Then, the* IP *packet dropped if:*

$$m_{\boldsymbol{i}}(j) \geq C - N + j + 1. \tag{2.1}$$
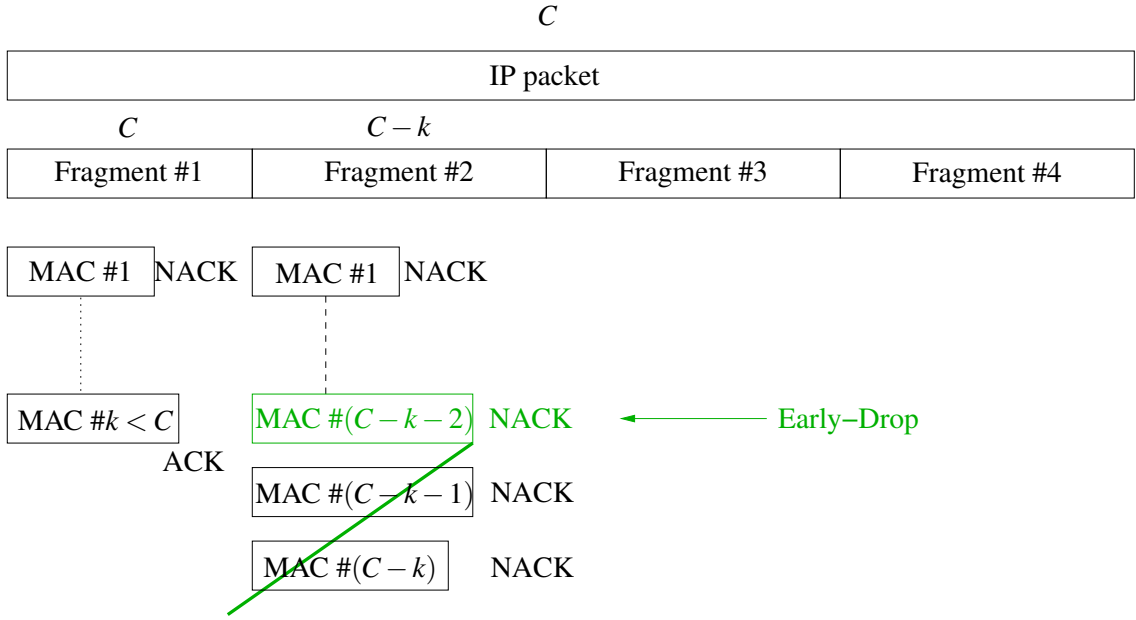
Figure 2.1: The Early-Drop mechanism on an example.

Since the ED approach only modifies the retransmission processing when the IP packet is not correctly received, the PER and the delay[1] are identical to those given in Chapter 1. On the other hand, the efficiency must be changed.

2.3 Efficiency new closed-form expression

2.3.1 General expression

The most general expression of the efficiency given in Chapter 1 becomes, for IBS:

$$\eta_{\text{IP}}^{\text{I}} = \frac{L_{\text{IP}} \left(1 - P_{\text{IP}}^{\text{I}}\right)}{P_{\text{IP}}^{\text{I}} \, \check{d}_{\text{IP}}^{\text{I}} \; + \; (1 - P_{\text{IP}}^{\text{I}}) \, \hat{d}_{\text{IP}}^{\text{I}}}, \tag{2.2}$$

where:

---

[1]We recall, that the delay is defined as the average number of packet transmissions when an IP packet is successfully received.

- $\hat{d}_{\text{IP}}^{\text{I}}$ is the average number of bits sent knowing that IP packet has been correctly received. Thus, the expression of $\hat{d}_{\text{IP}}^{\text{I}}$ given in [Le Duc, 2009] remains valid for the ED approach.

- $\check{d}_{\text{IP}}^{\text{I}}$ is the average number of bits sent given that the current IP packet reception fails. The expression of $\check{d}_{\text{IP}}^{\text{I}}$ given in [Le Duc, 2009] is modified by the ED approach since the transmission credit is not managed in the same way with or without ED when the IP packet fails.

Therefore the main goal is now to find a new expression of $\check{d}_{\text{IP}}^{\text{I}}$ when ED is used, and let us denote it by $\check{d}_{\text{IP}}^{\text{ED}}$.

2.3.2 Computation of $\check{d}_{\text{IP}}^{\text{ED}}$

For that purpose, all the combinations of fragments leading to a failure of the IP packet must be enumerated. The set of these combinations defines the event $\mathcal{D}$ that can be decomposed as follows:

$$\mathcal{D} = \bigcup_{n=1}^{N} D(n), \tag{2.3}$$

where the events $D(n)$ are defined below:

- $D(1) = \{$Fragment #1 consumes $C - N + 2$ credits$\}$,

- $D(n) = \{$Fragment #1 OK and Fragment #2 OK and ... and less than $(N - n)$ credits left during fragment #n transmission$\}$ for $n \in \{2, \ldots, N - 1\}$,

- $D(N) = \{$Fragment #1 OK and ... and Fragment #$(N - 1)$ OK and Fragment #N KO with the remaining credit$\}$.

In all the Chapter, we assume[2] that all the MAC packets have the same length L_{MAC}. This assumption fits well the CC-HARQ schemes. Considering IR-HARQ, if the mother code has a rate $R_0 = 1/t_0$ and the punctured code rates are equal to $\{1/t\}_{t=1,\ldots,t_0}$, then the equal MAC packet length assumption is satisfied. Now, the probabilities of the events $D(n)$ are detailed:

- $n = 1$: whenever it is received or not, the fragment #1 consumes at least $(C - N + 2)$ trials which leads to:

$$\Pr\{D(1)\} = q(C - N + 1), \tag{2.4}$$

and the number of bits sent during this event is equal to $d(1) = (C - N + 2)L_{\text{MAC}}$.

- $n \in \{2, \ldots, N - 1\}$: assume that the fragment #k (with $k \leq n - 1$) is successfully received and has used i_k transmissions. Then, the consumed transmission credit is

---

[2]The most general case with unequal packet length was done in our paper [C1].

equal to $m_i(k) = \sum_{j=1}^{k} i_j$ for $k \in \{1, \ldots, n-1\}$. The IP packet will not be received if the fragment #n consumes at least $(C - m_i(n-1) - (N-n) + 1)$ credits, whenever it is received or not. Such an event is denoted by $D_i(n)$, and the number of bits sent during this event is denoted by $d_i(n)$. Therefore, we have:

$$\Pr\{D(n)\} = \sum_{i \in \mathcal{T}_n} \Pr\{D_i(n)\}, \tag{2.5}$$

where $\mathcal{T}_n = \{i \in \mathbb{N}_*^{n-1} \,|\, m_i(n-1) = \sum_{k=1}^{n-1} i_k < C - N + n\}$, and

$$\Pr\{D_i(n)\} = q(C - m_i(n-1) - N + n) \prod_{k=1}^{n-1} p_1(i_k). \tag{2.6}$$

One can easily check that:

$$\mathcal{T}_n = \bigcup_{s=n-1}^{C-N+n-1} \mathcal{K}_{s,n}, \tag{2.7}$$

where $\mathcal{K}_{s,n}$ is the subset of $\mathcal{T}_n$ such that $\sum_{k=1}^{n-1} i_k = s$. As a consequence:

$$\sum_{i \in \mathcal{T}_n} d_i(n)\Pr\{D_i(n)\} = \sum_{s=n-1}^{C-N+n-1} \sum_{i \in \mathcal{K}_{s,n}} d_i(n)\Pr\{D_i(n)\}. \tag{2.8}$$

Furthermore, when $i \in \mathcal{K}_{s,n}$, the number of bits sent during the event $D_i(n)$ is equal to:

$$d_i(n) = sL_{\text{MAC}} + (C - s - (N-n) + 1)L_{\text{MAC}} = (C - N + n + 1)L_{\text{MAC}}, \tag{2.9}$$

and thus, putting Eq. (2.6) into Eq. (2.8) and using Eq. (2.9):

$$\sum_{i \in \mathcal{T}_n} d_i(n)\Pr\{D_i(n)\} = L_{\text{MAC}}(C - N + n + 1) \sum_{s=n-1}^{C-N+n-1} \sum_{i \in \mathcal{K}_{s,n}} q(C - s - N + n) \prod_{k=1}^{n-1} p_1(i_k)$$

$$= L_{\text{MAC}}(C - N + n + 1) \sum_{s=n-1}^{C-N+n-1} q(C - s - N + n) p_{n-1}(s). \tag{2.10}$$

- $n = N$: similar derivations lead to

$$\Pr\{D(N)\} = \sum_{i \in \mathcal{T}_N} \Pr\{D_i(N)\}, \tag{2.11}$$

where $D_i(N)$ is defined as in Eq. (2.6) for $n = N$. However, the number of transmitted bits during the event $D_i(N)$ is $d_i(N) = C L_{\text{MAC}}$.

Finally, the term $\check{d}_{\text{IP}}^{\text{ED}}$ is the sum of the number of bits $d_i(n)$ weighted by the probability of the event $\{D_i(n)| \text{ IP packet dropped }\}$, knowing that the IP packet has not been correctly received:

$$\check{d}_{\text{IP}}^{\text{ED}} = d(1)\Pr\{D(1)| \text{ IP packet dropped}\} + \sum_{n=2}^{N} \sum_{i \in \mathcal{T}_n} d_i(n)\Pr\{D_i(n)| \text{ IP packet dropped}\}. \tag{2.12}$$

Using the Bayes' rule leads to:

$$\check{d}_{\text{IP}}^{\text{ED}} = d(1)\frac{\Pr\{D(1)\}}{\Pr\{\text{ IP packet dropped}\}} + \sum_{n=2}^{N}\sum_{i\in\mathcal{T}_n} d_i(n)\frac{\Pr\{D_i(n)\}}{\Pr\{\text{ IP packet dropped}\}}$$

$$= \frac{1}{P_{\text{IP}}^{\text{I}}}\left(d(1)\Pr\{D(1)\} + \sum_{n=2}^{N}\sum_{i\in\mathcal{T}_n} d_i(n)\Pr\{D_i(n)\}\right). \tag{2.13}$$

Finally, we find:

$$\check{d}_{\text{IP}}^{\text{ED}} = \frac{L_{\text{MAC}}}{P_{\text{IP}}^{\text{I}}}\Big((C-N+2)q(C-N+1)$$

$$+ \sum_{n=2}^{N-1}(C-N+n+1)\sum_{s=n-1}^{C-N+n-1} q(C-s-N+n)\,p_{n-1}(s) + C\sum_{s=N-1}^{C-1} q(C-s)\,p_{N-1}(s)\Big). \tag{2.14}$$

2.3.3 Main result

Based on the previous computation, we are able to obtain the following result:

Proposition 2.1.

$$\eta_{IP}^{ED} \geq \eta_{IP}^{I}. \tag{2.15}$$

Proof. In non-ED context, a more precise description than $D_i(n)$ is needed since we must know how the MAC packets #n' (with $n' > n$)) are handled. The event $D_i(n)$ can be decomposed as follows: $D_i(n) = \cup_{i'} D_{i,i'}(n)$ where $D_{i,i'}(n)$ represents a single way of handling the remaining $(N-n-1)$ MAC packets, given that the n first MAC packets are handled as in $D_i(n)$. Fig. 2.2 depicts this decomposition on an example.

Then, we replace in Eq. (2.12):

$$\sum_{i\in\mathcal{T}_n} d_i(n)\Pr\{D_i(n)\}, \tag{2.16}$$

with:

$$\sum_{i\in\mathcal{T}_n}\sum_{i'} d_{i,i'}(n)\Pr\{D_{i,i'}(n)\}, \tag{2.17}$$

where $d_{i,i'}(n)$ is the cost in packets of the event $D_{i,i'}(n)$. Since, when using ED, the transmission stops as soon as $D_i(n)$ occurs, we have:

$$d_{i,i'}(n) \geq d_i^{ed}(n), \tag{2.18}$$

which implies that $\check{d}_{\text{IP}}^{\text{I}} \geq \check{d}_{\text{IP}}^{ed}$ and concludes the proof. $\qquad\square$
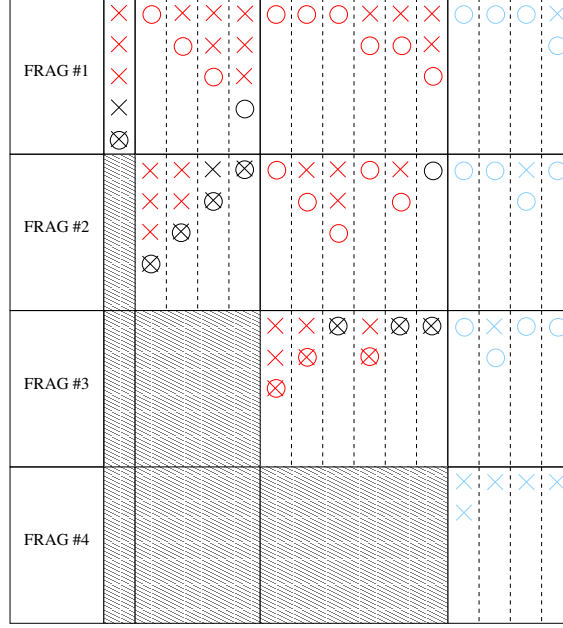
Figure 2.2: ED scenario with $N = 4$ and $C = 5$. Red paths refer to ED, black paths to non-ED, and blue paths appear in both ED and non-ED cases.

2.4   Particular case: Type-I HARQ

For Type-I HARQ, the MAC packets are all identical and are processed independently at the receiver side. Therefore, using Eq. (2.14) and reordering it like Eq. (2.13), one finds:

$$\check{d}_{\text{IP}}^{\text{ED}} = L_{\text{MAC}}(C - N + 1) + \frac{L_{\text{MAC}}}{P_{\text{IP}}^{\text{I}}} \left(\sum_{n=1}^{N} n \Pr\{D(n)\} - \Pr\{D(N)\} \right). \tag{2.19}$$

Due to the simple relation between the MAC packet and the fragment, it is possible to exhibit simple closed-form expression for the terms $\Pr\{D(n)\}$. The MAC packets are identical and handled independently in this context, and we remind that the error probability of any MAC packet is p_0. Thus, we recall from Chapter 1 that:

$$p_1(k) = (1 - p_0) p_0^{k-1}, \tag{2.20}$$
$$q(k) = p_0^k. \tag{2.21}$$

Furthermore, let us recall from Chapter 1 that:

$$P_{\text{IP}}^{\text{I}} = I(p_0; C - N + 1, N), \tag{2.22}$$

where $I(x; a, b) := B(x; a, b)/B(a, b)$ is the regularized Beta function, $B(x; a, b)$ is the incomplete Beta function and $B(a, b) = B(1; a, b)$.

Now, let us compute $\Pr\{D(n)\}$:

$$
\begin{aligned}
\Pr\{D(n)\} &\overset{(a)}{=} \sum_{i \in \mathcal{T}_n} q(C - m_i(n-1) - N + n) \prod_{k=1}^{n-1} p_1(i_k) \\
&\overset{(b)}{=} \sum_{i \in \mathcal{T}_n} p_0^{C - m_i(n-1) - N + n} \prod_{k=1}^{n-1} (1 - p_0) p_0^{i_k - 1} \\
&\overset{(c)}{=} \sum_{i \in \mathcal{T}_n} p_0^{C - m_i(n-1) - N + n} (1 - p_0)^{n-1} p_0^{m_i(n-1) - n + 1} \\
&\overset{(d)}{=} \mathrm{Card}(\mathcal{T}_n)(1 - p_0)^{n-1} p_0^{C - N + 1},
\end{aligned}
\tag{2.23}
$$

where (a) is obtained using Eqs. (2.5)-(2.6), (b) comes from Eqs. (2.20)-(2.21), (c) by using $\sum_{k=1}^{n-1} i_k = m_i(n-1)$, and (d) after factorization of the terms into the sum that are independent of i. By convention, $\mathrm{Card}(\mathcal{T}_1) = 1$. Using Appendix A.1, it can be easily checked that:

$$
\mathrm{Card}(\mathcal{T}_n) = \sum_{s=n-1}^{C-N+n-1} \binom{s-1}{n-2} = \binom{C-N+n-1}{n-1}.
\tag{2.24}
$$

Therefore it remains to calculate:

$$
\sum_{n=1}^{N} n \Pr\{D(n)\} = \sum_{n=1}^{N} n \binom{C-N+n-1}{n-1}(1 - p_0)^{n-1} p_0^{C-N+1}.
\tag{2.25}
$$

In Appendix B, it is shown that:

$$
\sum_{n=1}^{N} n \Pr\{D(n)\} = \frac{p_0 + \kappa(1 - p_0)}{p_0} I(p_0; \kappa, N) - \frac{p_0^{\kappa-1}(1 - p_0)^N}{B(\kappa, N)},
\tag{2.26}
$$

with $\kappa = C - N + 1$. Finally:

$$
\check{d}_{\mathrm{IP}}^{\mathrm{ED}} = \left(\frac{p_0 + K}{p_0} - \frac{p_0^{K-1}(1 - p_0)^N}{B(p_0; K, N)} - \binom{C-1}{N-1}(1 - p_0)^{N-1} p_0^K \right) L_{\mathrm{MAC}}.
\tag{2.27}
$$

Notice that the ED brings a lot of complexity in the derivation compared to the non-ED case for which $\check{d}_{\mathrm{IP}}^{\mathrm{I}} = C L_{\mathrm{MAC}}$ [Le Duc, 2009].

2.5 Numerical results

2.5.1 Simulation settings

For the sake of clarity, we present numerical results for two different HARQ schemes only:

- a pure ARQ, with MAC packets of 124 bits, including 16 bits for CRC,

- and a CC-HARQ with packets of 124 bits of data including CRC-16 that are encoded by a 1/2-rate convolutional code, with generators $(23, 35)_8$.

The bits are modulated over a Quadrature Phase Shift Keying (QPSK) constellation, and then are transmitted through an Additive White Gaussian Noise (AWGN) channel.

The figures are presented versus the Signal to Noise Ratio (SNR). We define the SNR as the ratio E_s/N_0, where E_s is the average energy per coded symbol, and N_0 is the bilateral energy spectral density of the noise, *i.e.* the noise variance is $N_0/2$ per real dimension.

2.5.2 Exact analytic expressions versus simulations

To begin with, let us compare the analytic expression of efficiency with ED with the simulated one. The efficiency expression given in Eq. (2.14) as well as extensive Monte-Carlo simulations are displayed in Fig. 2.3. For the ARQ scheme with $N = 8$ and $C = 16$ as well as for the CC-HARQ scheme with $N = 10$ and $C = 20$, we observe a nice agreement between our expressions and the estimated points. This validates the general expression for the efficiency of any HARQ type using IBS in conjunction with ED.
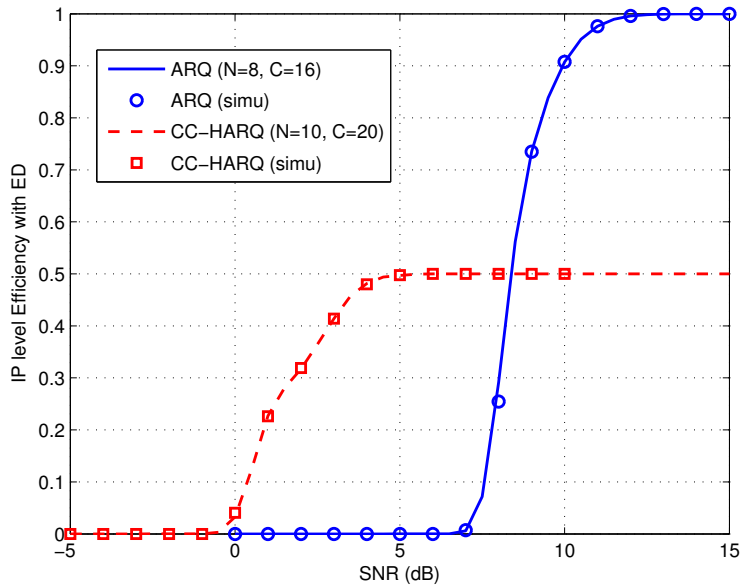
Figure 2.3: ARQ ($N = 8$, $C = 16$) and CC-HARQ ($N = 10$, $C = 20$).

2.5.3 Discussion on the relevance of Early-Drop

In Fig. 2.4 (resp. Fig. 2.5), the IP level efficiency of ARQ (Fig. 2.4a and 2.5a) and of CC-HARQ (Fig. 2.4b and 2.5b) is plotted for $N = 8$ and $C = 16$ (resp. $N = 10$ and $C = 20$). At first glance, we remark the little gain in efficiency, especially CC-HARQ for which

the gain is not visible at this scale. Thus, the ED does not provide a priori a significant gain but just an incremental one. Fig. 2.6 shows that for a fixed N, the gain in efficiency brought by ED depends on the total credit C value: basically, the less is C, the better is the efficiency gain.

(a) ARQ   (b) CC-HARQ
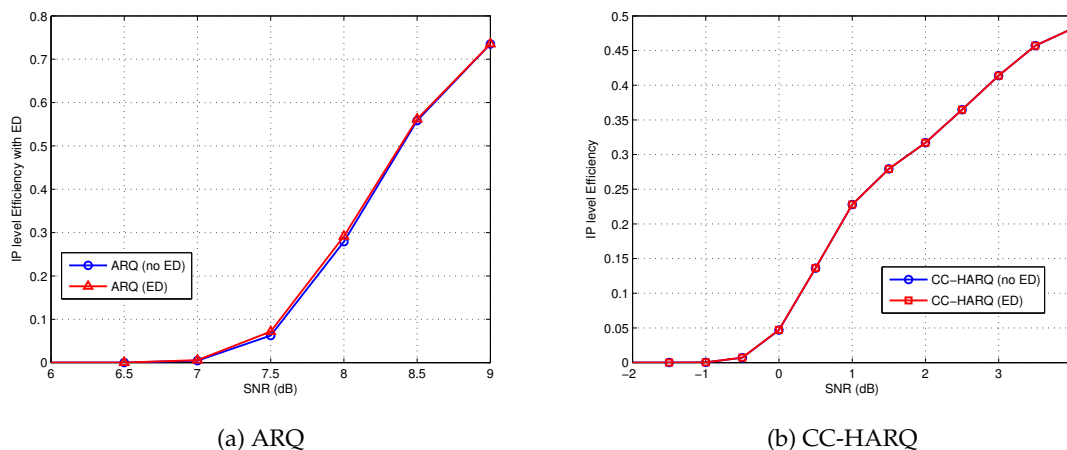
Figure 2.4: IP level Efficiency with/without Early-Drop ($N = 8$ and $C = 16$).

(a) ARQ   (b) CC-HARQ
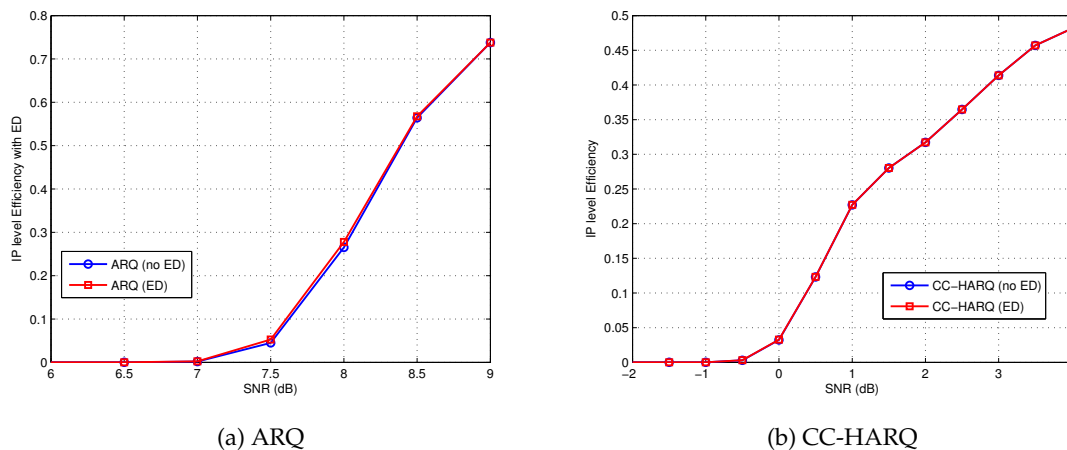
Figure 2.5: IP level efficiency with/without Early-Drop ($N = 10$ and $C = 20$).

More precisely, the terms $P_{\text{IP}}^{\text{I}} \, \check{d}_{\text{IP}}^{\text{ED}}$ and $P_{\text{IP}}^{\text{I}} \, \check{d}_{\text{IP}}^{\text{I}}$, which are the average number of MAC packets that are unsuccessfully transmitted with and without ED, respectively, are plotted in Fig. 2.7a (resp. Fig. 2.8a) for ARQ (resp. CC-HARQ). The IP fragmentation is fixed to $N = 12$. These terms appear in the denominator of the efficiency, thus it must be kept low in order to have a good efficiency. The main difference between $P_{\text{IP}}^{\text{I}} \, \check{d}_{\text{IP}}^{\text{ED}}$ and $P_{\text{IP}}^{\text{I}} \, \check{d}_{\text{IP}}^{\text{I}}$ occurs at low and medium SNR, but as soon as the SNR is large enough, the ED improvement
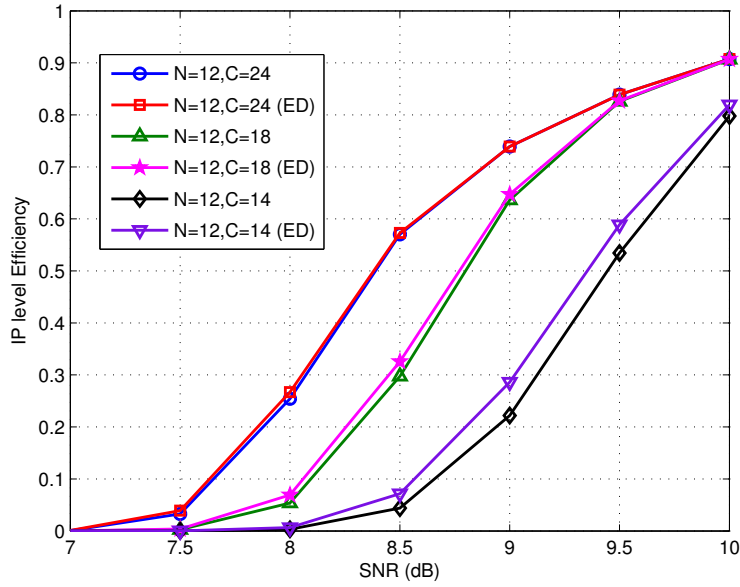
Figure 2.6: IP level Efficiency of ARQ with/without Early-Drop ($N = 12$).

vanishes since $P_{IP}^{I} \check{d}_{IP}^{ED}$ and $P_{IP}^{I} \check{d}_{IP}^{I}$ tend towards the same value.

However, notice that the gain is more important at low SNR (in fact, ED saves ($C-N+2$) transmissions in this area). But, as it can be seen from Fig. 2.7b and 2.8b, the PER in this area is equal to 1, and then the efficiency is almost equal to zero. Hence, the attractive feature of ED which is to suppress useless MAC packets transmissions, occurs in an SNR range where the system is inoperable.
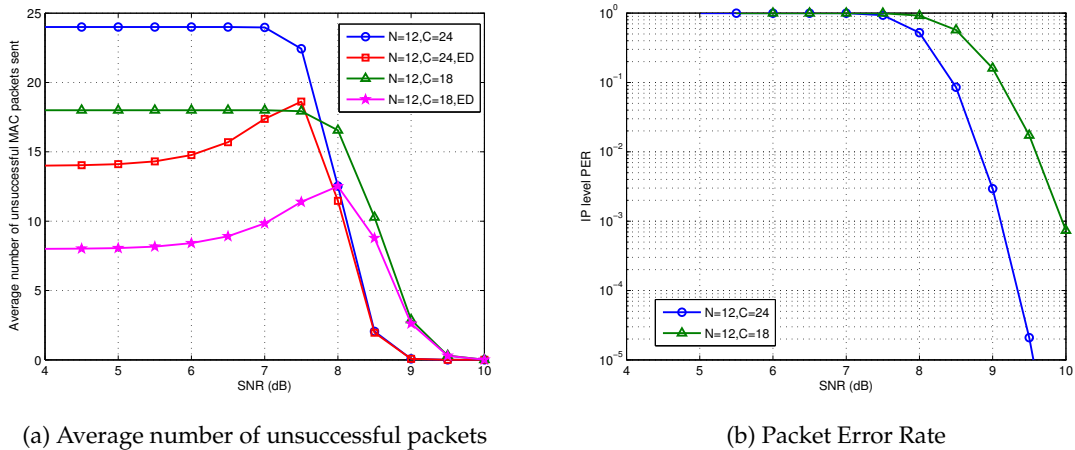
(a) Average number of unsuccessful packets

(b) Packet Error Rate

Figure 2.7: ARQ ($N = 12$).

(a) Average number of unsuccessful packets

(b) Packet Error Rate
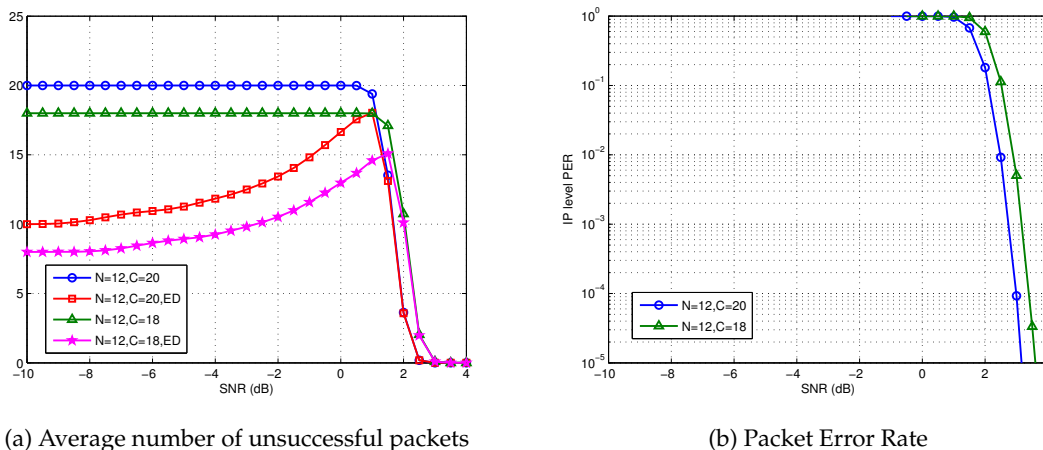
Figure 2.8: CC-HARQ ($N = 12$).

Finally, in Fig. 2.9 we inspect the relative efficiency gain defined as:

$$G := \frac{\eta_{\mathrm{IP}}^{\mathrm{ED}} - \eta_{\mathrm{IP}}^{\mathrm{I}}}{\eta_{\mathrm{IP}}^{\mathrm{I}}}. \tag{2.28}$$

Fig. 2.9a plots the gain G (in %) when $N = 12$ is fixed and C varies. In particular, we see that the ED gains are more important when C decreases towards N. Similarly, Fig. 2.9b shows the gain G (in %) when $C = 20$ is fixed and N varies, and we see again that G increases for increasing values of N towards C. We conclude from these figures that ED has more impact on the efficiency when N and C are getting close together. But in the same time, this leads to lower efficiency values, as shown in Fig. 2.6. Therefore, though the gain may be quite small, it remains of interest given the free-cost implementation of ED.

2.6 Conclusion

In this Chapter, we have presented an ED based version of the cross-layer HARQ with IBS. The mechanism of the ED technique has been clearly exposed, and it comes from its definition ED modifies only the HARQ efficiency metric. Then, we have developed the expression of efficiency that is valid for any HARQ type using ED with equal MAC packet length, and derived some closed-form expressions for the Type-I HARQ case.

The numerical analysis has revealed that the efficiency is only slightly improved when ED is used. The improvement has been measured through the gain G, and we concluded that ED was more helpful when the fragmentation N and the total credit C are close.

Finally, part of this work has been published in [C1].
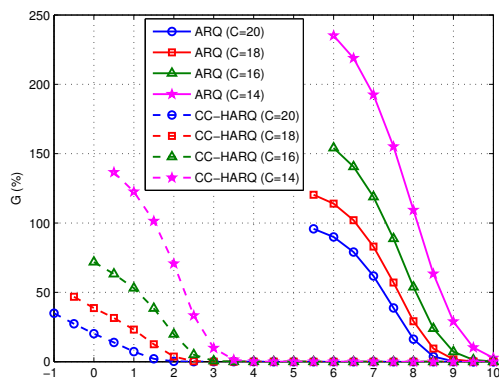
(a) $N = 12$
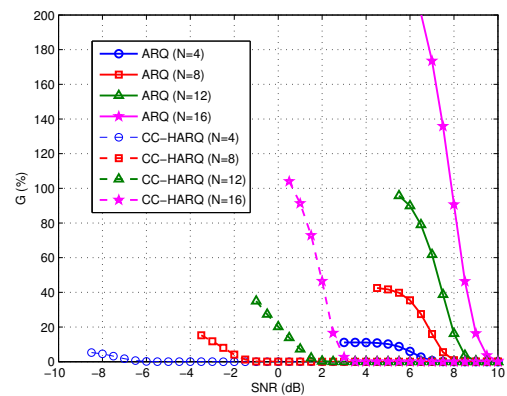
(b) $C = 20$

Figure 2.9: Early-Drop gain in efficiency G (%) for ARQ/CC-HARQ.

Chapter 3

Hybrid ARQ with imperfect feedback

3.1 Introduction

In wireless systems, the feedback message of HARQ systems can be either distorted, lost or delayed due feedback channel conditions or link scheduling congestion. The work presented in this Chapter was developed for the SW HARQ, and thus is appropriate for the context of a wireless ad hoc network using a Time Division Duplex (TDD) mechanism. For instance in systems relying on Ultra Wide Band (UWB) that is a well adapted waveform for military ad hoc networks, the transmitter cannot send packet and listen the acknowledgment channel simultaneously due to its half-duplex nature, thus avoiding the use of the GBN or SR protocols. Furthermore, since we focus on a data-link/network cross-layer paradigm, the effects of errors and delay in the feedback on HARQ performance are measured at IP level.

The Chapter is organized as follows. The literature relative to HARQ with imperfect feedback is reviewed in Section 3.2, and the system model is described in Section 3.3. Section 3.4 depicts a new cross-layer credit retransmission management adapted to the imperfect feedback context, Section 3.5 is devoted to the mathematical developments whereas some particular cases are studied in Section 3.6. Finally, Section 3.7 gives some insightful design guidelines relying on numerical illustrations.

3.2 State of the art

HARQ has already been investigated through numerous works. In Chapter 1, we have presented the principal contributions in the literature relative to the study of HARQ. All these works were done under the assumption of ideal feedback but, in wireless systems, this feedback message can be degraded due to errors and return delays.

Only a small amount of works have analytically studied the impact of non-ideal

feedback on HARQ. Notice that all these existing works focused on the performance at MAC level only, and have studied either noisy feedback or delayed feedback, but never the both jointly.

3.2.1   Unreliable ACK/NACK

When the RTT is assumed to be zero but the feedback is unreliable, then SW, GBN, and SR are equivalent, and therefore only the SW protocol was studied.

If the number of packet retransmissions L is infinite, the delay and efficiency of a Type-I HARQ are given in [El bahri et al., 2005] and [Wicker, 1995]. This is the most known delay expression with imperfect feedback:

$$d_{\mathrm{MAC}} = \frac{1}{1-p_0} + \frac{p_{\mathrm{fb}}}{1-p_{\mathrm{fb}}}, \tag{3.1}$$

where p_{fb} is the feedback error probability.

Secondly, if L is finite, an analytic expression of the efficiency can be found in [Wu and Jindal, 2009] for a Type-I HARQ:

$$\eta_{\mathrm{MAC}} = R(1-p_0)\frac{1-p_{\mathrm{fb}}}{1-p_{\mathrm{fb}}p_0}, \tag{3.2}$$

and of the delay in [Malkamäki and Leib, 2000] for a Type-II HARQ with CC:

$$d_{\mathrm{MAC}} = 1 + \sum_{\ell=1}^{L-1} q(\ell) + \sum_{\ell=1}^{L-1} \left(p_{\mathrm{fb}}^{\ell} - q(\ell)p_{\mathrm{fb}}^{L-\ell} \right), \tag{3.3}$$

where $q(\ell)$ is the packet failure probability after ℓ transmissions (defined in Chapter 1).

Finally, the throughput of the GBN HARQ with unreliable ACK/NACK has been studied in [Ausavapattanakun and Nosratinia, 2007a] using Markov processes.

3.2.2   Non-zero RTT

It was seen in Chapter 1 that a non-zero RTT leads to a reduction of the ARQ efficiency/throughput. To overcome this fixed/deterministic RTT, the GBN or the SR protocols are of interest.

When the RTT is assumed to be non-zero but fixed and known at the transmitter side, the SR protocol was analyzed only by [Badia, 2009] (delay) and by [Ausavapattanakun and Nosratinia, 2007b] (throughput). These works were done through a Markov analysis, but this framework is not very convenient to derive closed-form expressions, even for the most simple SW ARQ.

3.3   Imperfect feedback model

3.3.1   Typical feedback errors

In wireless systems, the feedback communications can be subject to errors and return delays. Three main impairments on the feedback channel can be identified:

i) errors in the acknowledgment message value due to the wireless nature of the feedback channel,

ii) acknowledgment messages that are not received instantaneously due to the RTT inherent to all the ARQ systems,

iii) acknowledgment messages that are lost either because of the feedback channel conditions or dropped in the queues of the feedback link.

In its fundamental description, HARQ is built upon a one-bit feedback. However, in practical systems more bits are sent since the communication slot is not allocated to transmit only a single bit, hence the feedback frames generally contain more bits than the ACK/NACK message, like channel state information, the packet identifier, etc. We assume that the feedback information integrity can be controlled at the transmitter side, by an error detection code for instance, in order to differentiate erroneous from error-free feedback frames.

3.3.1.1   ACK/NACK errors

Assuming that the error detection code is sufficiently powerful to neglect misdetection, the ACK or NACK can be considered as error-free when no error detection occurs. Conversely, if an error is detected, we assume that a NACK is received. This is in accordance with the assumption of neglecting the NACK-to-ACK errors that is widely adopted in the literature, as found in [Malkamäki and Leib, 2000], [Badia, 2009] or [Ausavapattanakun and Nosratinia, 2007b]. Thus, we get:

$$\Pr\{\text{ACK} \to \text{NACK}\} = p_e \tag{3.4a}$$

$$\Pr\{\text{NACK} \to \text{ACK}\} = 0. \tag{3.4b}$$

3.3.1.2   Feedback delay

It is already known that RTT causes delay in the feedback. Moreover if congestion problems (coming from receiver queues, scheduler, routing, etc) occur in the reverse channel, then the transmitted feedback may arrive randomly or may be dropped in the data link queues. We introduce a time-out $\tau_0 > 0$ in order to design the waiting time and to overcome deadlock issues due to lost or dropped feedback messages. If the RTT is larger than the time-out, then the feedback message will be considered to be lost and will be interpreted as a NACK by the transmitter.

3.3.1.3 An example of ARQ transmission with imperfect feedback

Fig. 3.1 depicts an example of ARQ rounds when the feedback is subject to the imperfections described previously. During the first ARQ round ($k = 1$), the data packet #1
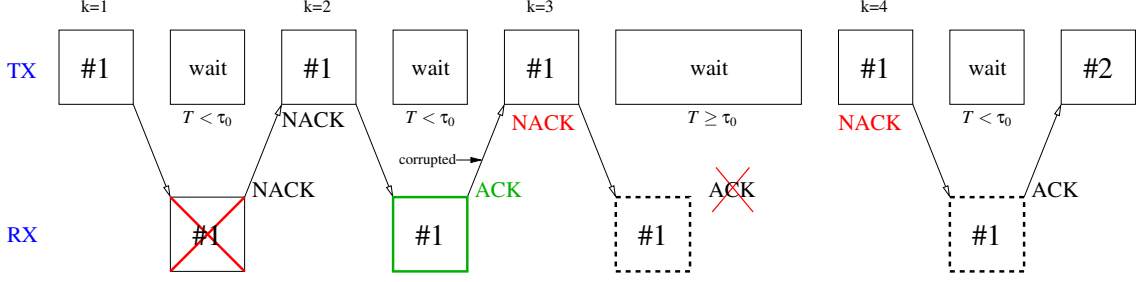
Figure 3.1: ARQ scheme with imperfect feedback.

is received with errors, then the receiver sends a NACK that is received correctly by the transmitter. Therefore, the packet #1 is retransmitted in a second round ($k = 2$) and this time is received without errors. The receiver sends ACK, which is received within the waiting window in a time $T < \tau_0$, but was corrupted by the feedback channel and detected with errors by the transmitter. This leads to the third ARQ round ($k = 3$) where the packet is retransmitted, but discarded by the receiver because it has already been received during the second round. Then, the receiver sends ACK again, but it will never arrive at the transmitter which waited until τ_0. Finally, the packet will be correctly acknowledged in the fourth round and the ARQ goes to packet #2.

3.3.2 Mathematical model

3.3.2.1 RTT model

Due to the random nature of the considered feedback, the RTT is naturally modeled as a random variable $T_0 + T$, where T_0 is the MAC packet duration and T is a continuous random variable defined by its probability density dF_T. When the feedback is not received at the transmitter side, *i.e.*, $T + T_0 \geq \tau_0$, we have:

$$p_c := \Pr\{T \geq \tau_0 - T_0\} = 1 - F_T(\tau_0 - T_0). \tag{3.5}$$

3.3.2.2 Feedback channel

Consequently, the general model adopted in this paper for the unreliable feedback channel is the cascade of a Binary Erasure Channel (BEC) and of a Z channel with ternary input, as shown in Fig. 3.2. The channels have individual transition matrices Σ and Z, respectively,

given by:

$$\Sigma = \begin{pmatrix} 1 - p_c & p_c & 0 \\ 0 & p_c & 1 - p_c \end{pmatrix} \tag{3.6}$$

$$Z = \begin{pmatrix} 1 & 1 & p_e \\ 0 & 0 & 1 - p_e \end{pmatrix}^T \tag{3.7}$$
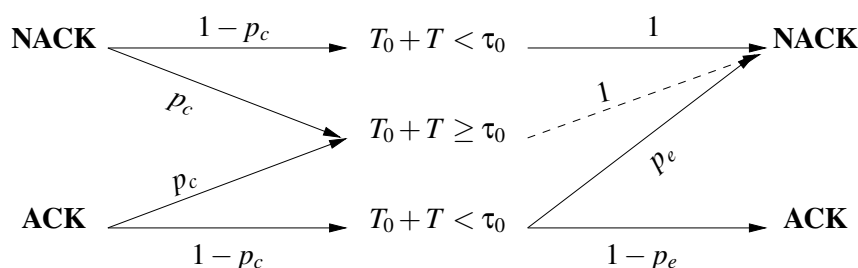
Figure 3.2: Feedback channel modeled as a BEC-Z channel.

3.3.2.3 Equivalent feedback channel

The equivalent transition matrix for the feedback channel is obtained by multiplying Σ and Z [Silverman, 1955], and is given by:

$$F = \begin{pmatrix} 1 & 0 \\ p_{fb} & 1 - p_{fb} \end{pmatrix}, \tag{3.8}$$

where the probability p_{fb} is equal to:

$$p_{fb} = p_c + (1 - p_c)p_e. \tag{3.9}$$

Therefore, the feedback channel is equivalent to a Z channel with crossover probability p_{fb} as depicted in Fig. 3.3.
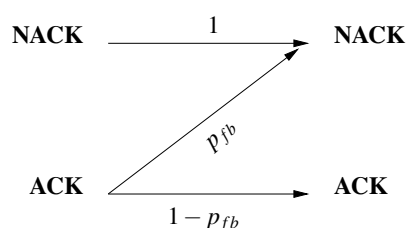
Figure 3.3: Equivalent Z channel for feedback.

The transmitter receives NACK with an average transmission time depending on the initial feedback information (ACK or NACK). We define the average NACK receiving time given that an ACK has been sent:

$$\tau_{r,a} := \mathbb{E}\left[T_0 + T | \text{NACK Rx, ACK Tx}\right], \tag{3.10}$$

the average NACK receiving time given that a NACK has been sent:

$$\tau_{r,n} := \mathbb{E}\left[T_0 + T | \text{NACK Rx, NACK Tx}\right], \tag{3.11}$$

and the average waiting time for ACK:

$$\tau_a := \mathbb{E}\left[T_0 + T | \text{ACK Rx, ACK Tx}\right]. \tag{3.12}$$

These time averages can be computed as follows:

Proposition 3.1.

$$\tau_a = T_0 + \mathbb{E}\left[T | T < (\tau_0 - T_0)\right], \tag{3.13}$$

$$\tau_{r,n} = p_c \tau_0 + (1 - p_c)\tau_a, \tag{3.14}$$

$$\tau_{r,a} = \frac{p_c \tau_0 + p_e(1 - p_c)\tau_a}{p_{\text{fb}}}. \tag{3.15}$$

Proof. Let us define $\text{TO} := \{T_0 + T \geq \tau_0\}$ as the time-out event. Then:

$$\begin{aligned}
\tau_a &= \mathbb{E}\left[T_0 + T | \text{ACK Rx, ACK Tx}, \overline{\text{TO}}\right] \Pr\left\{\overline{\text{TO}} | \text{ACK Rx, ACK Tx}\right\} \\
&= \mathbb{E}\left[T_0 + T | T_0 + T < \tau_0\right] \Pr\left\{\overline{\text{TO}} | \text{ACK Rx, ACK Tx}\right\},
\end{aligned} \tag{3.16}$$

where $\Pr\left\{\overline{\text{TO}} | \text{ACK Rx, ACK Tx}\right\} = 1$ (using Bayesian rule) and $\mathbb{E}\left[T_0 + T | T_0 + T < \tau_0\right] = T_0 + \mathbb{E}\left[T | T < (\tau_0 - T_0)\right]$.

Next, $\tau_{r,a}$ can be decomposed as follows:

$$\begin{aligned}
\tau_{r,a} = &\mathbb{E}\left[T_0 + T | \text{NACK Rx, ACK Tx}, \overline{\text{TO}}\right] \Pr\left\{\overline{\text{TO}} | \text{NACK Rx, ACK Tx}\right\} \\
&+ \mathbb{E}\left[T_0 + T | \text{NACK Rx, ACK Tx, TO}\right] \Pr\left\{\text{TO} | \text{NACK Rx, ACK Tx}\right\}.
\end{aligned} \tag{3.17}$$

Using Bayesian rule again, we find $\Pr\left\{\text{TO} | \text{NACK Rx, ACK Tx}\right\} = p_c/p_{\text{fb}}$. Furthermore:

$$\mathbb{E}\left[T_0 + T | \text{NACK Rx, ACK Tx, TO}\right] = \tau_0, \tag{3.18}$$

$$\mathbb{E}\left[T_0 + T | \text{NACK Rx, ACK Tx}, \overline{\text{TO}}\right] = \mathbb{E}\left[T_0 + T | T_0 + T < \tau_0\right] = \tau_a. \tag{3.19}$$

Putting these three equations into Eq. (3.17) boils down to the $\tau_{r,a}$ expression. $\tau_{r,n}$ is obtained similarly. $\qquad\square$

3.4 A general cross-layer HARQ scheme using report of credit

3.4.1 Description of the proposed scheme

From the IP point of view, it is interesting to share the transmission credit among the fragments belonging to a same IP packet, as proposed by [Choi et al., 2005]. But, for some reasons that will become clear soon, it is of importance to also bound the number of transmissions per fragment. To combine these two dual approaches, we suggest to keep a maximum transmission credit per fragment, while allowing the unused credit of a given fragment to be carried forward to the next fragment.

More precisely, let $L_n^{(0)}$ be the initial maximum transmission credit of fragment n, and let L_n be the maximum transmission credit for the n-th fragment of a given IP packet. The proposed approach, called Report Credit Strategy (RCS), consists in applying the following rule:

$$L_n \leftarrow L_n^{(0)} + (L_{n-1} - k_{n-1}), \ \forall n > 1, \tag{3.20}$$

where $k_n \leq L_n$ denotes the number of transmissions consumed by fragment n. Let us denote by $\boldsymbol{L}^{(0)} = (L_n^{(0)})_{n \in \{1,...,N\}}$ the sequence of initial credits, and $C = \sum_{n=1}^N L_n^{(0)}$ the total initial transmission credit. An instance of this new scheme is displayed in Fig. 3.4. Observe that the HARQ process of a fragment continues until ACK is received, or its own credit is consumed. Then, the HARQ continues with the next fragment.
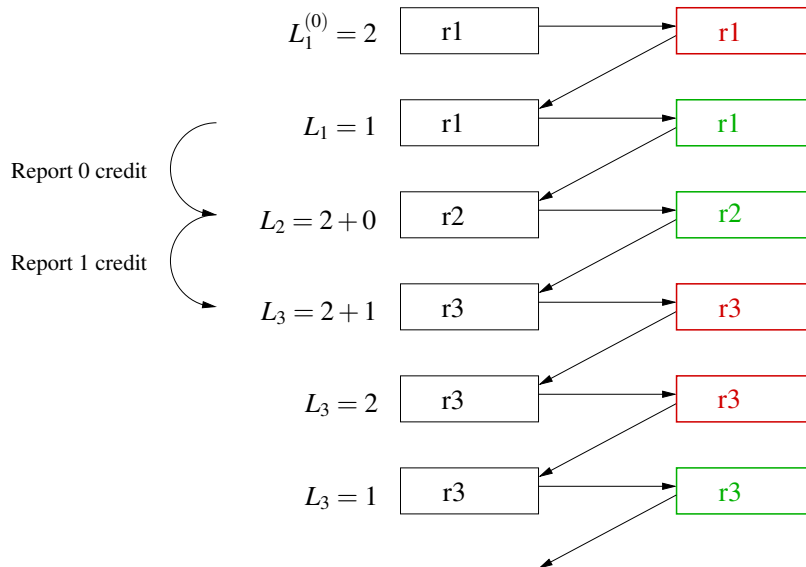
Figure 3.4: RCS scheme example ($N = 3$ and $\boldsymbol{L}^{(0)} = [2, 2, 2]$).

3.4.2 IBS seen as a particular case

The new RCS is a generalization of the previous cross-layer HARQ scheme with IBS introduced by [Choi et al., 2005]. Considering Eq. (3.20), if one sets $L_1^{(0)} = C$ and $L_n^{(0)} = 0$ for $n = 2$ to N, the IBS is obtained as a byproduct. Indeed, if the total credit C is given to the first fragment, the next fragment will not receive the remaining credit until the first fragment transmission is successful, or C is consumed, and so on. This is exactly the scheme depicted in [Choi et al., 2005].

3.4.3 An example of RCS and IBS with imperfect feedback

In this Section we illustrate on an example the drawbacks of IBS when the feedback is imperfect, and in the same time the robustness brought by RCS. In the center of Fig. 3.5 we display channel instances: the left blocks represent the transmission channel and the blocks in the right the feedback channel. A green block means that the transmission succeeds whereas a red block means that transmission fails.
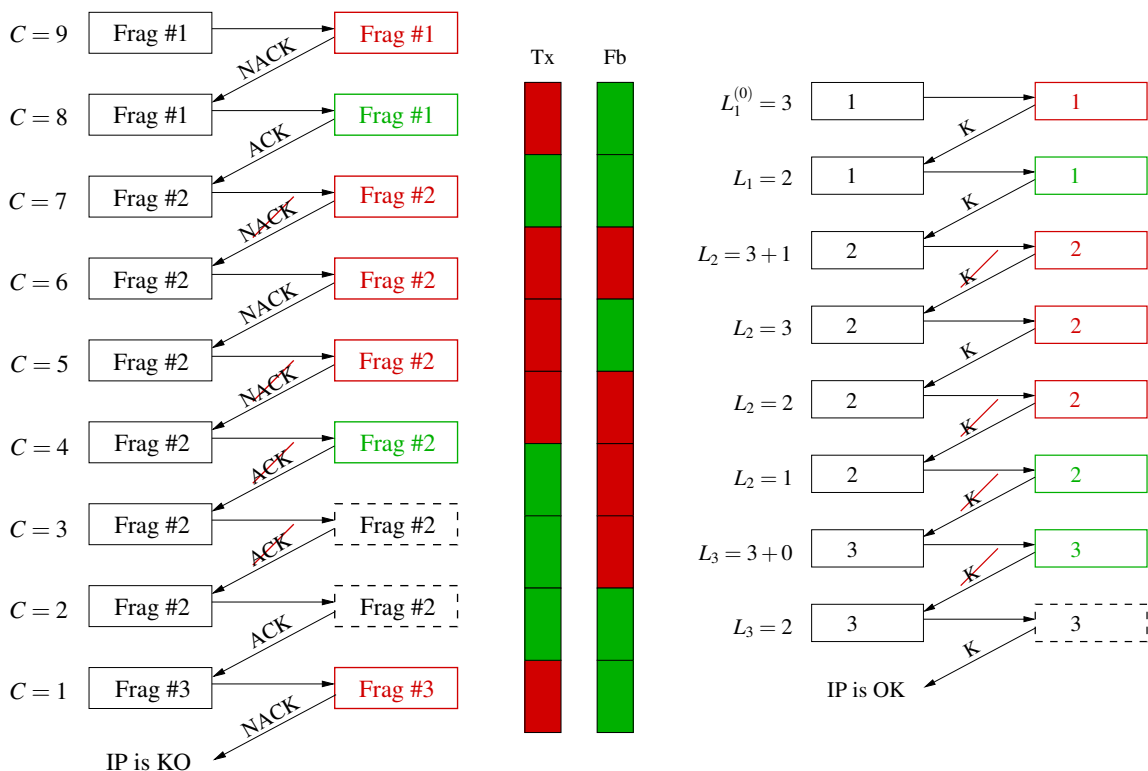
Figure 3.5: ARQ example with IBS (left, $C = 9$) and RCS (right, $L^{(0)} = [3,3,3]$) for $N = 3$ fragments.

On the left side, an ARQ process is run for $N = 3$ fragments and uses IBS with a total of $C = 9$ credits. The first fragment is received in two rounds, whereas the second fragment needs four rounds before being decoded without errors (*i.e.*, ACK is sent).

Observe that the corrupted NACK messages have no influence on the process. But, when the second fragment is finally decoded, the ACK is corrupted by the feedback channel, which leads to the retransmission of this fragment. The bad feedback conditions cause another two retransmissions of fragment #2. Finally, when the transmitter receives ACK for this fragment, there remains only one transmission ($C = 1$). Unfortunately, the third fragment is not decoded at the first time, and thus the IP fails since there is no more credit ($C = 0$).

On the right, the same example is run using RCS with the credit distribution $L^{(0)} =$ $[3, 3, 3]$. There are two key observations:

- Since the fragment #1 is received correctly in two rounds, the remaining credit is carried forward to fragment #2, which owns $L_2 = L_2^{(0)} + 1 = 4$ now. This additional credit helps the second fragment to be received, since three attempts would not be sufficient and would have led to the IP failure.

- Furthermore, since the number of transmissions for fragment #2 is bounded, the ARQ switches to fragment #3 as soon as the credit of fragment #2 is exhausted, regardless of the ACK presence at the transmitter. Therefore, the last fragment can be received without errors, which leads to the success of the IP packet transmission.

3.5   HARQ performance analysis with imperfect feedback

In this Section, we derive analytic expressions for the HARQ figures of merit, *i.e.* the PER, the delay and the efficiency. This is done for the three IP level strategies described previously: the performance of RCS are first derived, then IBS is found as a particular instance of RCS, finally FBS is computed as a particular case of the IBS analysis.

3.5.1   IP level analysis of RCS

Given the credit distribution $L^{(0)}$, we will denote:

- by $\alpha_n(i)$ the average time cost for receiving n fragments, in i MAC packet transmissions,

- and by $\beta_n(i)$ the probability of decoding n fragments in i transmissions.

By definition of RCS, the retransmission scheme continues with the next fragment whenever the ACK is received for the current fragment or the transmission credit of the current fragment is consumed. Therefore, the IP packet can be received without errors even if some ACKs were not received after the good decoding of a fragment. Obviously, it will impact the delay and the efficiency.

3.5.1.1 Packet Error Rate

The packet error rate depends on the quality of the feedback channel. Indeed, the maximum number of transmissions of the n-th fragment is driven by the number of transmissions done by the $(n-1)$ previous fragments. An IP packet is correctly decoded if, and only if, the N fragments are correctly received in i transmissions with $i \in \{N, \dots, C\}$, leading to:

$$P_{\text{IP}}^{\text{R}} = 1 - \sum_{i=N}^{C} \beta_N(i). \tag{3.21}$$

3.5.1.2 Delay

Applying Bayes' rule, the average delay (in time units) is obtained as:

$$d_{\text{IP}}^{\text{R}} = \frac{1}{1 - P_{\text{IP}}^{\text{R}}} \sum_{i=N}^{C} \alpha_N(i). \tag{3.22}$$

3.5.1.3 Efficiency

From the most general expression given in Eq. (1.24) of Chapter 1, the efficiency is:

$$\eta_{\text{IP}}^{\text{R}} = \frac{T_0 R N (1 - P_{\text{IP}}^{\text{R}})}{\check{d}_{\text{IP}}^{\text{R}} P_{\text{IP}}^{\text{R}} + (1 - P_{\text{IP}}^{\text{R}}) d_{\text{IP}}^{\text{R}}}, \tag{3.23}$$

where $\check{d}_{\text{IP}}^{\text{R}}$ is the average delay when IP packet failed, and R is the coding rate.

3.5.1.4 Time averages

All the metrics are thus entirely determined by the knowledge of $\alpha_n(i)$, $\beta_n(i)$, and $\check{d}_{\text{IP}}^{\text{R}}$ that are given in the following Propositions (the derivations can be found in Appendices C.1 to C.3):

Proposition 3.2. $\forall n \geq 1, i \geq n,$

$$\alpha_n(i) = \sum_{k=0}^{n} \sum_{(s,t) \in \mathcal{E}_{i-k,n}^B} \left((\sum_\ell s_\ell) \tau_{r,n} + k \tau_a + (\sum_\ell t_\ell) \tau_{r,a} \right) (1 - p_{\text{fb}})^k \prod_{\ell=1}^{n} p_1 (1 + s_\ell) p_{\text{fb}}^{t_\ell}, \tag{3.24}$$

where

$$\mathcal{E}_{p,n}^B = \left\{ (s,t) \in \mathbb{N}^n \times \mathbb{N}^n \mid \sum_{\ell=1}^{n} (s_\ell + t_\ell) = p \text{ and } \forall \ell, \sum_{m=1}^{\ell} (1 + s_m + t_m) \leq \sum_{m=1}^{\ell} L_m^{(0)} \right\}. \tag{3.25}$$

Proposition 3.3. $\forall n \geq 1, i \geq n,$ *we have:*

$$\beta_n(i) = \sum_{x \in \chi_{i,n}} \prod_{j=1}^{n} \sum_{k_j=1}^{x_j} p_1(k_j) p_{\text{fb}}^{x_j - k_j} (1 - p_{\text{fb}})^{\delta\{A_j\}}, \tag{3.26}$$

where

$$\chi_{k,n} = \left\{ \boldsymbol{x} \in \mathbb{N}_*^n \mid \sum_{\ell=1}^{n} x_\ell = k \text{ and } \forall \ell, \sum_{m=1}^{\ell} x_m \le \sum_{m=1}^{\ell} L_m^{(0)} \right\}, \tag{3.27}$$

$A_j = \{\sum_{m=1}^{j} x_m < \sum_{m=1}^{j} L_m^{(0)}\}$, *and* $\delta\{.\}$ *is the Kronecker symbol.*

Proposition 3.4. *The term* $\check{d}_{IP}^R$ *is computed as follows:*

$$\check{d}_{IP}^R = \frac{1}{P_{IP}^R} \sum_{i=L_1+(N-1)}^{C} \sum_{n=0}^{N-1} \theta_n(i), \tag{3.28}$$

where $\theta_n(i)$ *is the average time cost for receiving n correct fragments, and* $(N-n)$ *erroneous fragments, in i transmissions, and where:*

$$\theta_n(i) = \sum_{(\boldsymbol{s},\boldsymbol{t}) \in \mathcal{E}_{i-n,n}^B} \left[\left(\sum_\ell s_\ell \right) \tau_{r,n} + n\tau_a + \sum_\ell t_\ell \right) \tau_{r,a} \right] (1 - p_{fb})^n \delta\{KO\}$$

$$\prod_{\ell=1}^{N} \left(p_1(1 + s_\ell) p_{fb}^{t_\ell} \delta\{B_\ell\} + q(1 + s_\ell + t_\ell) \delta\{\Gamma_\ell\} \right), \quad (3.29)$$

where the events $B_\ell = \left\{ \sum_{m=1}^{\ell}(1 + s_m + t_m) \le \sum_{m=1}^{\ell} L_m^{(0)} \right\}$, $\Gamma_\ell = \left\{ \sum_{m=1}^{\ell}(1 + s_m + t_m) = \sum_{m=1}^{\ell} L_m^{(0)} \right\}$, *and* $KO = \{\exists \ell \in \{1, \dots, N\} \mid \Gamma_\ell\}$, *and* $q(k) = 1 - \sum_{i=1}^{k} p_1(i)$ *is the failure probability of one fragment after k transmissions.*

3.5.2 IP level analysis of IBS

In the following, the delay and PER of IBS are derived from the previous Propositions. However, it is more convenient to write a new proof for the efficiency.

Given the total credit p, we will denote:

- by $\alpha_{n,p}(i)$ the average time cost for receiving n fragments and the corresponding n ACK messages at the transmitter side, in $1 \le i \le p$ MAC packet transmissions,

- and by $\beta_n(i)$ the probability of decoding n fragments in i transmissions and receiving n ACK messages (at the transmitter side).

3.5.2.1 Packet Error Rate

In the IBS case, since the HARQ process continues for a given fragment until ACK is received or the total credit C is consumed, an IP packet is successful if N ACK messages are received at the transmitter side. If the total credit is consumed when an IP is received with success, it means that the last fragment is received correctly before the credit is exhausted, but the ACK is not guaranteed to arrive at the transmitter within the remaining transmissions. Therefore, the PER is obtained as:

$$P_{IP}^I = 1 - \sum_{i=N}^{C} \frac{\beta_N(i)}{(1 - p_{fb})^{\delta\{i=C\}}}, \tag{3.30}$$

since for $i = C$ the transmitter just needs to receive $(N-1)$ ACK messages.

3.5.2.2 Delay

The delay is found as a byproduct:

$$d_{\text{IP}}^{\text{I}} = \frac{1}{1 - P_{\text{IP}}^{\text{I}}} \sum_{i=N}^{C} \alpha_{N,C}(i). \tag{3.31}$$

3.5.2.3 Efficiency

As for the efficiency, we find:

$$\eta_{\text{IP}}^{\text{I}} = \frac{T_0 R N (1 - P_{\text{IP}}^{\text{I}})}{\check{d}_{\text{IP}}^{\text{I}} P_{\text{IP}}^{\text{I}} + (1 - P_{\text{IP}}^{\text{I}}) d_{\text{IP}}^{\text{I}}}, \tag{3.32}$$

where $\check{d}_{\text{IP}}^{\text{I}}$ is the average delay when the IP packet failed.

3.5.2.4 Time averages

The new expressions of $\alpha_{n,p}(i)$, $\beta_n(i)$, and $\check{d}_{\text{IP}}^{\text{I}}$ for IBS are given in the following Propositions (the derivations can be found in Appendices C.4 to C.6):

Proposition 3.5. $\forall n \geq 1, p \geq n, p \geq i \geq n$,

$$\alpha_{n,p}(i) = (1 - p_{\text{fb}})^{n - \delta\{i = p\}} \sum_{k=n}^{i} \left(i\tau_{r,a} - n(\tau_{r,n} - \tau_a) + p_{\text{fb}}\delta\{i = p\}(\tau_{r,a} - \tau_a) - k(\tau_{r,a} - \tau_{r,n}) \right)$$

$$\times \binom{i - k + n - 1}{n - 1} p_n(k) p_{\text{fb}}^{i-k}. \tag{3.33}$$

Proposition 3.6. $\forall n \geq 1, i \geq n$, we have:

$$\beta_n(i) = \sum_{k=n}^{i} \binom{i - k + n - 1}{n - 1} p_n(k) p_{\text{fb}}^{i-k} (1 - p_{\text{fb}})^n. \tag{3.34}$$

Proposition 3.7. The term $\check{d}_{\text{IP}}^{\text{I}}$ is computed as follows:

$$\check{d}_{\text{IP}}^{\text{I}} = \frac{1}{P_{\text{IP}}^{\text{I}}} \left(\sum_{n=1}^{N} \theta_n(C) + \sum_{n=1}^{N-1} \gamma_n(C) \right), \tag{3.35}$$

where $\theta_n(C)$ is the average time cost for receiving $(n-1)$ fragments and the corresponding ACK messages, and n-th fragment is not decoded, in C transmissions. $\gamma_n(C)$ is the average time cost for receiving $(n-1)$ fragments and the corresponding ACK messages, and n-th fragment is decoded, in C transmissions, and where:

$$\theta_n(C) = (1 - p_{\text{fb}})^{n-1} \sum_{\ell=n-1}^{C-1} \sum_{m=1}^{C-\ell} \left((n-1)\tau_a + (C - \ell - m)\tau_{r,a} + (\ell + m - n + 1)\tau_{r,n} \right)$$

$$\times \binom{C - \ell - m + n - 2}{n - 2} p_{n-1}(\ell) p_{\text{fb}}^{C-\ell-m} q(m), \quad \forall n \geq 2, \tag{3.36}$$

and

$$\gamma_n(C) = (1-p_{\text{fb}})^{n-1} \sum_{\ell=n-1}^{C-1} \sum_{m=1}^{C-\ell} \Big((n-1)\tau_a + p_{\text{fb}}\tau_{r,a} + (C-\ell-m)\tau_{r,a} + (\ell+m-n)\tau_{r,n} + (1-p_{\text{fb}})\tau_a \Big)$$

$$\times \binom{C-\ell-m+n-1}{n-1} p_{n-1}(\ell) p_{\text{fb}}^{C-\ell-m} p_1(m), \quad \forall n \geq 1. \quad (3.37)$$

$q(k) = 1 - \sum_{i=1}^{k} p_1(i)$ *is the failure probability of one fragment after k transmissions, and we have* $\theta_1(C) = C \tau_{r,n} q(C)$.

3.5.3 IP level analysis of FBS

3.5.3.1 Packet Error Rate

The correct decoding of a given fragment at the receiver side is not affected by the fact that the ACK is changed into NACK by the feedback link. Since the fragments are processed independently in the FBS, if N fragments are correct at the receiver side, then the corresponding IP packet can be delivered error-free. As a consequence, P_{IP}^{F} **is not modified** by imperfect feedback, as already mentioned in [Malkamäki and Leib, 2000] for the MAC level.

3.5.3.2 Delay

The fragments sent after the reception of a corrupted ACK lead to a delay increase at MAC and IP levels since, for instance, the transmitter will send useless redundant fragments whereas it should have sent new data fragments if the ACK were correctly received. Notice that the average number of MAC packets sent when the fragment is not correctly received is identical to the case of perfect channel feedback. Therefore, under the assumption of i.i.d. fragments, we obtain the average delay (in time units):

$$d_{\text{IP}}^{\text{F}} = N d_{\text{MAC}}, \quad (3.38)$$

where d_{MAC} can be written as follows:

$$d_{\text{MAC}} = \frac{1}{1-P_{\text{MAC}}} \sum_{i=1}^{L} \alpha_{1,L}(i). \quad (3.39)$$

Only $\alpha_{1,L}(i)$ must be evaluated in closed-form. It can be obtained by putting $n = 1$ and $p = L$ in Prop. 3.5 and by replacing P_{IP}^{I} with P_{MAC}. This leads to:

$$d_{\text{MAC}} = \frac{1}{1-P_{\text{MAC}}} \sum_{i=1}^{L} \sum_{k=1}^{i} \Big(i\tau_{r,a} - (\tau_{r,n} - \tau_a) + p_{\text{fb}}\delta\{i=L\}(\tau_{r,a} - \tau_a) - k(\tau_{r,a} - \tau_{r,n}) \Big)$$

$$\times p_1(k) p_{\text{fb}}^{i-k} (1-p_{\text{fb}})^{\delta\{i<L\}}. \quad (3.40)$$

3.5.3.3 Efficiency

Useless retransmissions due to an error in the feedback affect the efficiency as well, which is found as:

$$\eta_{\text{IP}}^{\text{F}} = \frac{T_0 R (1 - P_{\text{MAC}})^N}{\tau_{r,n} L P_{\text{MAC}} + (1 - P_{\text{MAC}}) d_{\text{MAC}}}, \tag{3.41}$$

where $\check{d}_{\text{MAC}}$ is the average delay when a fragment fails, d_{MAC} is the average delay at MAC level (*i.e.* when one fragment is handled), and R is the coding rate.

3.6 Some particular cases

3.6.1 IBS performance at large SNR

It has been seen that the PER was insensitive to errors in the feedback in the FBS case, and then it vanishes at large SNR. However, we show in this Section that unlike the conventional HARQ management, the IBS can be dramatically bad in the presence of feedback errors.

At large SNR, $p_1(k)$ vanishes for all k except for $k = 1$ where it is equal to 1 (indeed, when the channel is extremely good, each packet is correctly decoded in one shot). Thus, we have that $p_n(k)$ vanishes too, for all $k > n$ and $p_n(n) = 1$. When inserting into Eq. (3.34):

$$\lim_{\text{SNR} \gg 1} \beta_n(i) = \binom{i-1}{n-1} p_{\text{fb}}^{\,i-n} (1 - p_{\text{fb}})^n. \tag{3.42}$$

Now, taking the large SNR limit into Eq. (3.30) leads to:

$$\lim_{\text{SNR} \gg 1} P_{\text{IP}}^{\text{I}} = 1 - \sum_{i=N}^{C} \binom{i-2}{N-2} p_{\text{fb}}^{\,i-N} (1 - p_{\text{fb}})^{N-1} = 1 - \left(\frac{1 - p_{\text{fb}}}{p_{\text{fb}}}\right)^{N-1} \sum_{i=N-1}^{C-1} \binom{i-1}{n-1} p_{\text{fb}}^{\,i}. \tag{3.43}$$

Using known results on binomial series (see Appendix A), it drops down to:

$$\lim_{\text{SNR} \gg 1} P_{\text{IP}}^{\text{I}} = I(p_{\text{fb}}; C - N + 1, N - 1), \tag{3.44}$$

where $I(x; a, b)$ is the so-called regularized beta function. Since the function $I(x; a, b)$ vanishes when $x = 0$ or when a or b becomes infinite, for finite transmission credit C and non-zero error probability p_{fb} on the feedback channel, the PER limit is strictly positive. This means that the PER does not vanish at large SNR in the IBS case, contrary to the FBS case.

3.6.2 Instantaneous noisy feedback ($T = 0$)

This is a common assumption made in the related literature ([Malkamäki and Leib, 2000], [El bahri et al., 2005], [Wu and Jindal, 2009]). In that case $p_c = 0$ (so, $p_{\text{fb}} = p_e$) and $\tau_{r,n} = \tau_{r,a} = \tau_a = T_0$ (from Eqs. (3.9)-(3.15) respectively). Without loss of generality, we

assume that $T_0 = 1$ which means that the average delay corresponds to a number of packets, so the metrics become:

$$d_{\text{IP}}^{\text{I}} = \frac{1}{1 - P_{\text{IP}}^{\text{I}}} \sum_{i=N}^{C} i \frac{\beta_N(i)}{(1 - p_{\text{fb}})^{\delta\{i=C\}}} \qquad \text{and} \qquad \eta_{\text{IP}}^{\text{I}} = \frac{RN(1 - P_{\text{IP}}^{\text{I}})}{C P_{\text{IP}}^{\text{I}} + (1 - P_{\text{IP}}^{\text{I}})d_{\text{IP}}^{\text{I}}}, \tag{3.45}$$

$$d_{\text{IP}}^{\text{F}} = \frac{N}{1 - P_{\text{MAC}}} \sum_{i=1}^{L} i \frac{\beta_1(i)}{(1 - p_{\text{fb}})^{\delta\{i=L\}}} \qquad \text{and} \qquad \eta_{\text{IP}}^{\text{F}} = \frac{R(1 - P_{\text{MAC}})^N}{L P_{\text{MAC}} + (1 - P_{\text{MAC}})d_{\text{MAC}}}. \tag{3.46}$$

By setting $p_{\text{fb}} = 0$ in the previous equations (*i.e.*, by considering perfect feedback), all the closed-form expressions given in Chap. 1 can be retrieved.

3.6.3 Type-I HARQ

To be related with the works involving imperfect feedback, we focus on FBS at MAC level ($N = 1$). In the case of Type-I HARQ, we have $p_1(i) = (1 - p_0)p_0^{i-1}$, and then the delay expression of Eq. (3.40) is simplified into:

$$d_{\text{MAC}} = \frac{1 - p_0}{(1 - P_{\text{MAC}})(p_{\text{fb}} - p_0)}\Big(L(p_{\text{fb}}^L - p_0^L) + (1 - p_{\text{fb}})\big(p_{\text{fb}} f_L(p_{\text{fb}}) - p_0 f_L(p_0)\big)\Big), \tag{3.47}$$

with $f_n(x) := \sum_{k=1}^{n-1} k x^{k-1}$.

Furthermore, when $L \to \infty$, we find that:

$$\lim_{L \to \infty} d_{\text{MAC}} = \frac{1}{1 - p_0} + \frac{p_{\text{fb}}}{1 - p_{\text{fb}}}, \tag{3.48}$$

which is in perfect agreement with that given in [Wicker, 1995],[El bahri et al., 2005] and [Wu and Jindal, 2009].

3.7 Numerical results

3.7.1 Simulations setup

We will present the performance of an ARQ and of CC-HARQ, both with FBS, IBS and RCS. Each MAC packet has 128 information bits (including CRC-16), and is encoded with a 1/2-rate convolutional code with generators $(23, 35)_8$ in the CC-HARQ scheme. Next, these bits are sent over a QPSK constellation, and the symbols are transmitted through an AWGN channel.

The random RTT that occurs on the imperfect feedback link is built as follows: T is exponentially distributed with parameter $\lambda > 0$, where $1/\lambda$ is the expected arrival time. In that case, $dF_T(t) = \lambda e^{-\lambda t}dt$ and $\mathbb{E}[T|T < (\tau_0 - T_0)] = \frac{1}{\Pr\{T<(\tau_0-T_0)\}} \int_0^{\tau_0-T_0} t\, dF_T(t)$. After simple algebraic manipulations, we get:

$$p_c = e^{-\lambda(\tau_0 - T_0)} \tag{3.49}$$

$$\mathbb{E}[T|T < (\tau_0 - T_0)] = \frac{1}{\lambda} - \frac{(\tau_0 - T_0)p_c}{1 - p_c}. \tag{3.50}$$

In the sequel, the two causes of imperfect feedback will be studied separately:

(i) On the one hand, we can reasonably assume that the feedback is transmitted over the same channel than the messages, then one can expect that the error probability p_e depends on the SNR. In this case, we set $p_c = 0$ and so $p_{fb} = p_e = f(\text{SNR})$, with f some function.

(ii) On the other hand, the time-out events due to random RTT have a constant error probability p_c, and we set for this case $p_{fb} = p_c$ ($p_e = 0$).

3.7.2 Monte-Carlo simulations

In Fig. 3.6, we check the accuracy of the analytical expressions of the efficiency and of the delay with imperfect feedback. The figures are presented for ARQ and CC-HARQ, for IP packets fragmented in $N = 3$, and for both FBS with $L = 3$ and IBS with $C = 9$. The efficiency expression computed in Fig. 3.6a, for perfect RTT (*i.e.* $T = 0$) and $p_e = 10^{-1}$, is of perfect agreement with the Monte-Carlo simulations. This is also checked in Fig. 3.6b where the delay, computed for a complete imperfect feedback with $\lambda = 1/2$ and $p_e = 10^{-1}$, fits the Monte-Carlo points.

(a) Efficiency ($p_e = 10^{-1}$, $T = 0$)   (b) Delay ($p_e = 10^{-1}$, $\lambda = 1/2$, $\tau_0 = 3$)
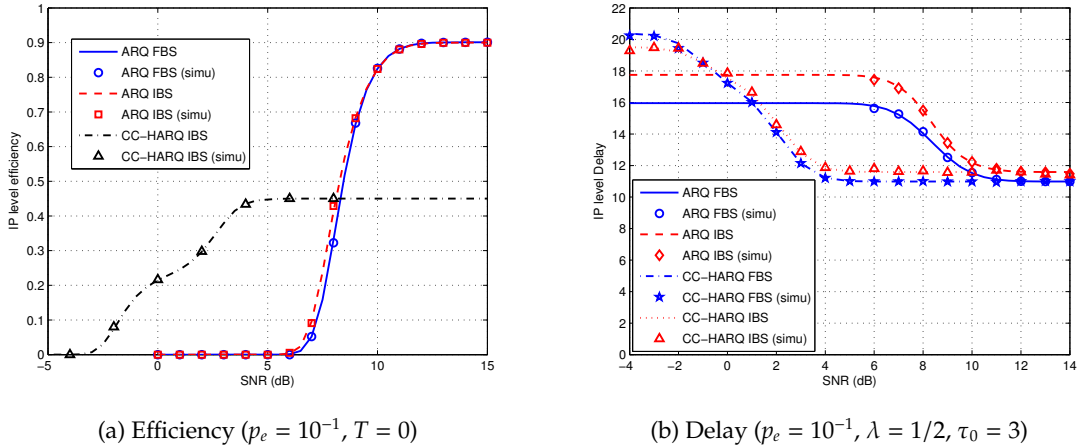
Figure 3.6: Simulations compared with IP level performance of ARQ/CC-HARQ ($N = 3$, $L = 3$, $C = 9$).

Finally, the PER of ARQ for several configurations (N, c) of RCS is plotted versus p_0 in Fig. 3.7.

3.7.3 Discussion on the feedback: effect of p_e

In this Section, we investigate the feedback impairment on a CC-HARQ scheme. Assuming that the feedback is transmitted over the same channel than the data, then we
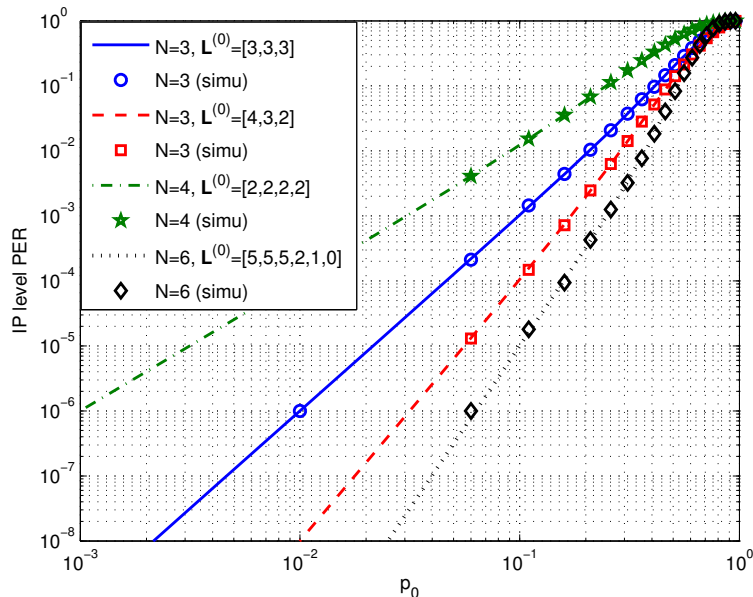
Figure 3.7: Simulations compared with IP level PER versus p_0 for ARQ with RCS.

model the feedback error as Gaussian noise, with probability $p_e = f(\text{SNR})$, where f is some decreasing function of the SNR that depends on the scheme used for the feedback transmission. In what follows, the feedback is constructed in three different ways:

(a) One bit feedback, *i.e.* a single ACK/NACK bit is sent through the noisy channel, and f corresponds to the bit error probability (assuming a detection method is available).

(b) Uncoded feedback frames of 32 bits, including 16 bits for CRC, 1 bit for ACK/ACK, and the remaining 15 bits can be used to transmit the channel state, or the packet number, etc.

(c) Coded feedback frames: 32 bits long frames built as described above, then encoded using the 1/2-rate convolutional code $(23, 35)_8$.

The PER is plotted in Fig. 3.8 for $N = 6$, $L = 3$ and $C = 18$. First of all, the PER of FBS is insensitive to imperfect feedback, as expected. However, the IBS can be highly degraded according to the feedback scheme: even for one bit feedback (a), the PER is significantly increased, but stays lower than the PER of FBS unlike scheme (b). Nevertheless, it is seen that coding (c) recovers the performance, but we notice a slight degradation that remains at low SNR (< 0 dB).

The PER behavior has several consequences on the delay and efficiency performance, as shown in Figs. 3.9 and 3.10. The delay when using the feedback scheme (a) or (c), plotted in Fig. 3.9a for FBS with $L = 3$ and in Fig. 3.9b for IBS with $C = 18$, stays close to the ideal case for the two retransmission strategies. The SNR shift on the IBS delay
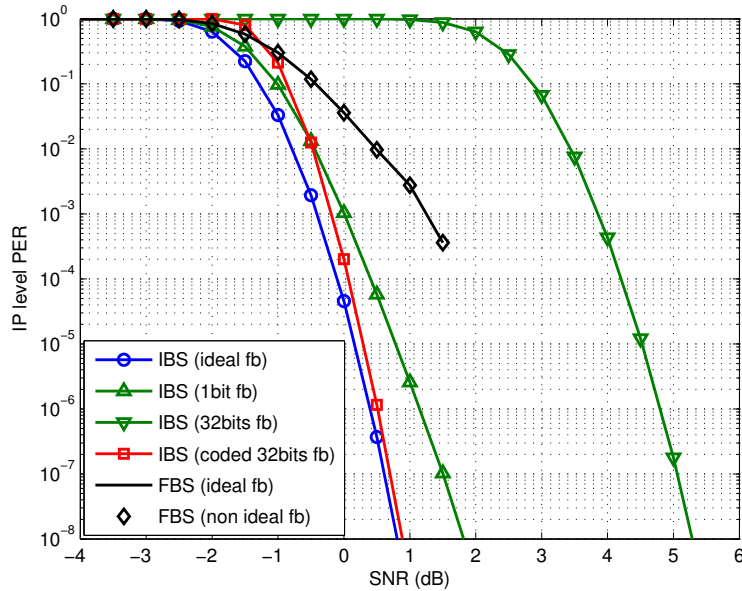
Figure 3.8: Effect of different feedback strategies on IP level PER of CC-HARQ with FBS/IBS ($N = 6$, $L = 3$, $C = 18$, $p_c = 0$).

observed for the feedback scheme (b) is easily explained by the PER figure. However, we notice that the delay of FBS is greatly increased too when uncoded feedback frames are sent.
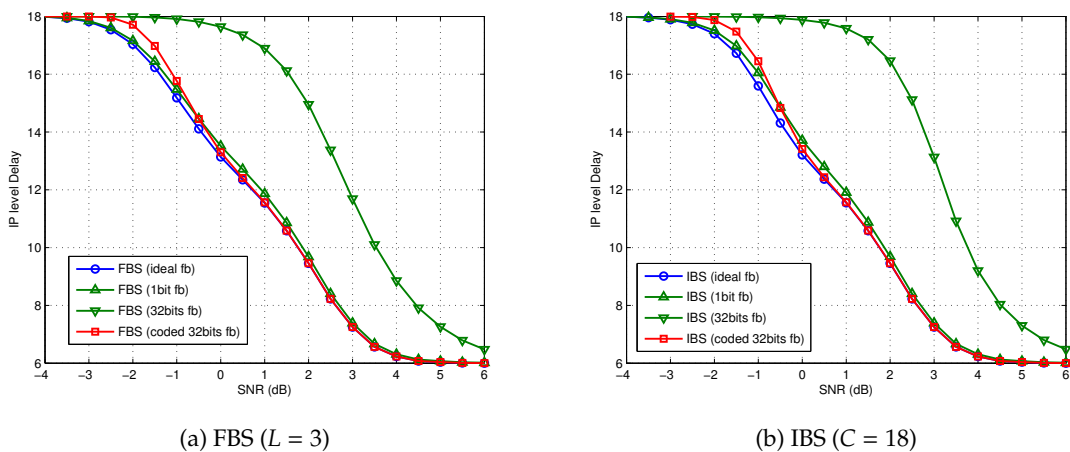
(a) FBS ($L = 3$)

(b) IBS ($C = 18$)

Figure 3.9: Effect of different feedback strategies on IP level delay of CC-HARQ ($N = 6$, $p_c = 0$).

This explains the efficiency of FBS displayed in Fig. 3.10a. Since the delay of FBS is shifted for large SNR values (*i.e.* > 0 dB), this error relative to the ideal case is predominant in the efficiency, which is shifted too. Observe that coding the feedback helps to completely recover the performance in this case. But the PER becomes more predominant in the efficiency at low SNR, hence the ideal and non ideal feedback figures coincide. The delay figure explains also the behaviour of the efficiency of IBS at large SNR. However, Fig. 3.10b shows that the IBS is dramatically degraded also at low SNR due to the poor PER performance of the feedback scheme (b). Notice that even coding cannot retrieve all the efficiency since there is still a shift at low SNR.

(a) FBS (L = 3)   (b) IBS (C = 18)
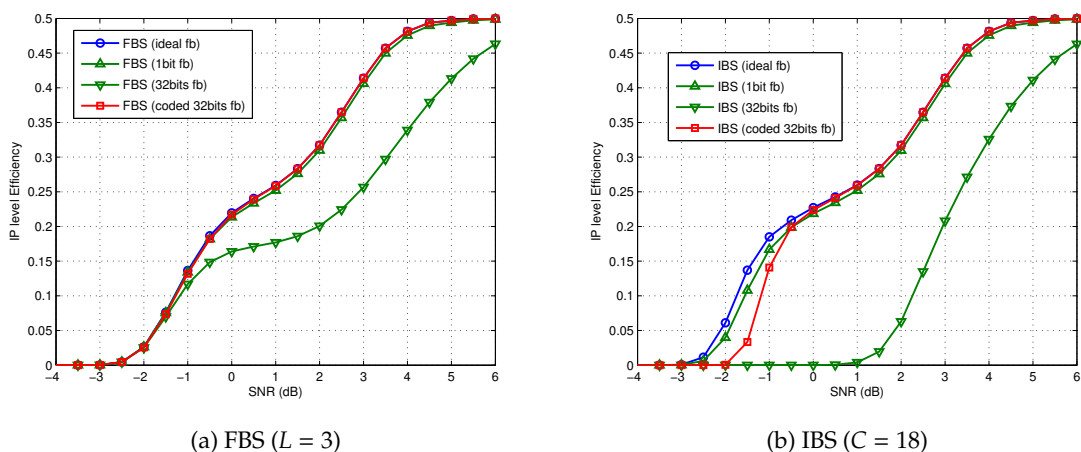
Figure 3.10: Effect of different feedback strategies on IP level efficiency of CC-HARQ ($N = 6$, $p_c = 0$).

3.7.4   Discussion on the time-out value: effect of RTT

In this Section, the effect of feedback impairment *(ii)* is studied on an ARQ and a CC-HARQ schemes. That is, in all this Section $p_e = 0$ and the RTT parameter is set to $\lambda = 1/3$. In Fig. 3.11 we plot the PER of FBS/IBS for $N = 6$, $L = 3$ and $C = 18$. Since the PER of FBS is insensitive to any feedback imperfection, it is displayed for the ideal case as a benchmark. There are three observations. First of all, for the two schemes (ARQ and CC-HARQ) the IBS performance present an error floor, *i.e.* the PER values become constant (to 10^{-3} in our case) beyond some SNR value. This drawback is due to the random RTT that causes a constant feedback error probability p_c (see Section 3.6.1). Secondly, the CC-HARQ with IBS is more degraded than the ARQ scheme. Indeed, Fig. 3.11a shows that before the error floor apparition the PER of IBS is still better than the FBS one, unlike CC-HARQ for which FBS is always better than IBS, as seen in Fig. 3.11b. Finally, these figures confirm the intuition that increasing the time-out value enhance the PER performance (in particular,

the error floor appears later and is lower).

(a) ARQ

(b) CC-HARQ

Figure 3.11: Effect of different time-out values on IP level PER of ARQ/CC-HARQ ($N = 6$, $L = 3$, $C = 18$, $p_e = 0$, $\lambda = 1/3$).

The impact on the delay of the time-out value choice is discussed in Fig. 3.12, where the ideal feedback cases are displayed as benchmarks. In Fig. 3.12a we plot the delay of ARQ and have two remarks. Firstly, for any time-out value the delay of FBS remains lower than the delay of IBS and converges towards the same value at large SNR, as in the ideal case. Secondly, according to the intuition, the delay at low SNR is larger for higher time-out values, but is lower for increasing time-out values at large SNR because of the PER error floors. The delay of CC-HARQ is plotted in Fig. 3.12b. The same remarks hold except for the IBS case at low SNR, where it is shown that it is systematically better than the FBS one, unlike in the ideal case.

The efficiency performance are discussed in Fig. 3.13, where the ideal feedback cases are again displayed as benchmarks. In Fig. 3.13a we plot the efficiency of ARQ, where we observe that the IBS suffers from the delay shift for large SNR values as described above. The efficiency of CC-HARQ is plotted in Fig. 3.13b and we see that although the delay of IBS is better than that of FBS at low SNR, the efficiency of IBS is always lower. Nevertheless, increasing the time-out value enhances the efficiency performance of the IBS, whereas the efficiency is degraded in the FBS case.

3.7.5   PER performance of RCS versus FBS and IBS

The previous analysis of FBS/IBS under several imperfect feedback conditions has revealed many weaknesses of the IBS, compared to the FBS. The performance loss in the PER of the IBS have been identified. Therefore it is of interest to find a retransmission management strategy that still enhance the PER of FBS, while not dramatically affecting

(a) ARQ
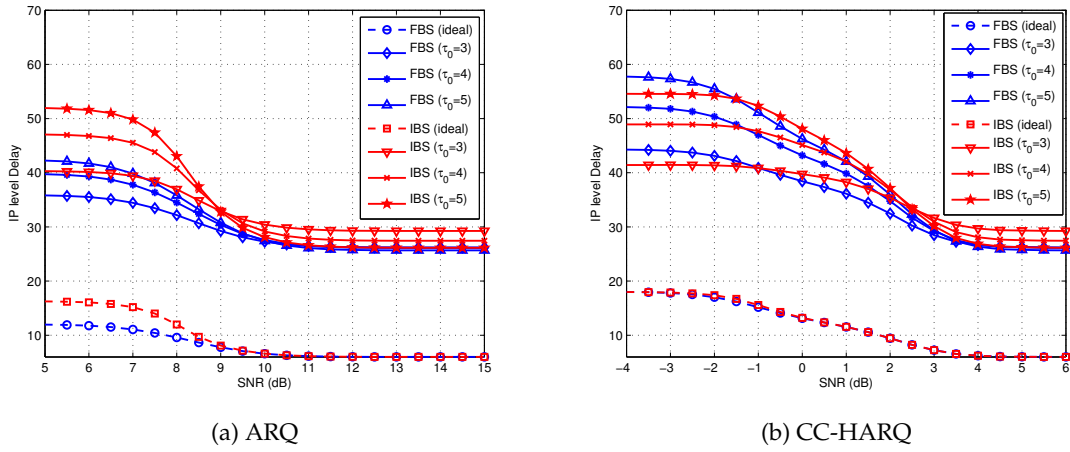
(b) CC-HARQ

Figure 3.12: Effect of different time-out values on IP level delay of ARQ/CC-HARQ ($N = 6$, $L = 3$, $C = 18$, $p_e = 0$, $\lambda = 1/3$).

(a) ARQ
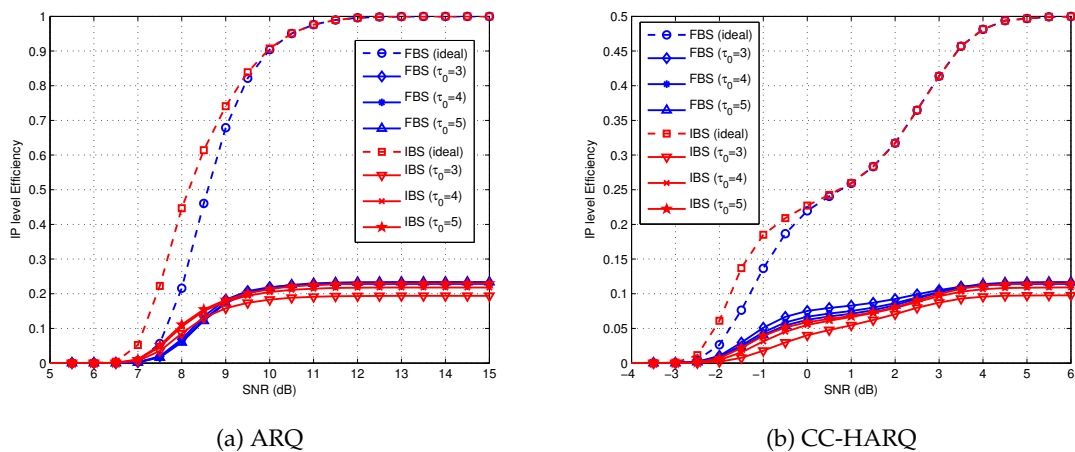
(b) CC-HARQ

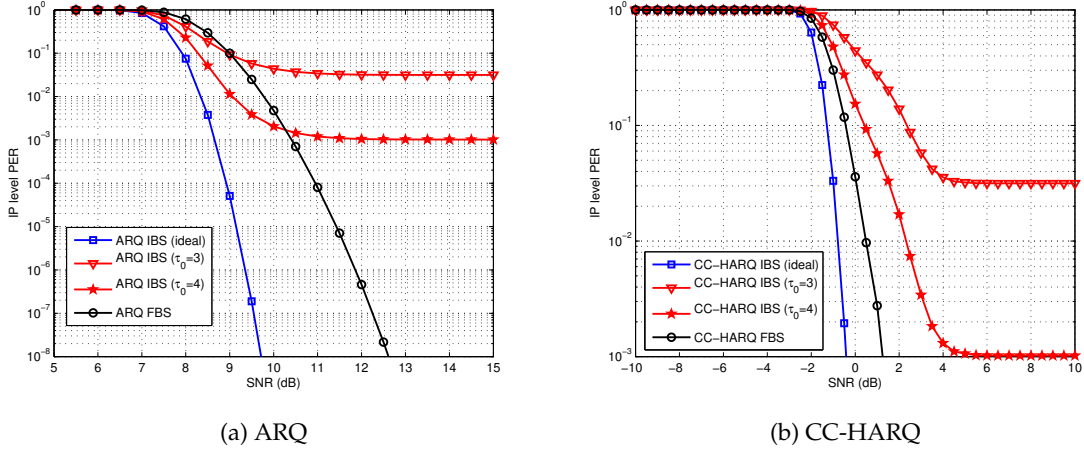Figure 3.13: Effect of different time-out values on IP level efficiency of ARQ/CC-HARQ ($N = 6$, $L = 3$, $C = 18$, $p_e = 0$, $\lambda = 1/3$).

the IBS. This has been achieved with the use of the RCS in Section 3.4, and this Section presents a numerical evaluation of the PER performance of RCS compared to FBS/IBS.

In Fig. 3.14, the PER of RCS is plotted when the feedback has a random RTT with $\lambda = 1/3$ and $p_e = 0$. The fragmentation is $N = 6$ and the credit distribution of RCS is arbitrarily fixed to $L^{(0)} = [3, 3, 3, 3, 3, 3]$. The performance of FBS/IBS are displayed for benchmarking. The PER of an ARQ scheme is plotted in Fig. 3.14a, and of a CC-HARQ in Fig. 3.14b. In both cases, the RCS still brings a gain in PER relative to the FBS, but is far more robust to the imperfect feedback than the IBS.

(a) ARQ	(b) CC-HARQ
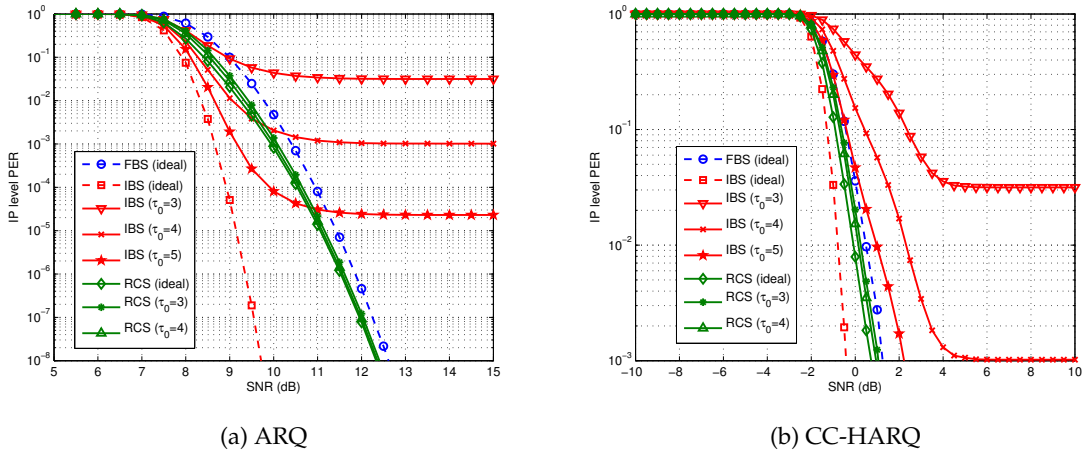
Figure 3.14: Effect of the different cross-layer strategies (FBS/IBS/RCS) on IP level PER of ARQ/CC-HARQ with non-zero RTT ($N = 6$, $L = 3$, $C = 18$, $L^{(0)} = [3, 3, 3, 3, 3, 3]$, $p_e = 0$, $\lambda = 1/3$).

In Fig. 3.15, we discuss on the impact of different credit distributions on the PER of RCS, for $N = 6$. Again, the FBS and IBS are displayed for benchmarking. For the ARQ scheme, Fig. 3.15a, we see that the uniform credit distribution ($L^{(0)} = [3, 3, 3, 3, 3, 3]$) has a little gain over the FBS. Furthermore, the credit distribution $L^{(0)} = [5, 5, 5, 2, 1, 0]$ is more powerful and gets closer to the IBS, that is actually the best credit distribution in the perfect feedback case. In contrast, the distribution $L^{(0)} = [2, 2, 5, 5, 2, 2]$ is even worse than the FBS. These remarks still hold for the CC-HARQ, as seen in Fig. 3.15b, where the distribution $L^{(0)} = [5, 5, 5, 2, 1, 0]$ achieves the IBS performance. We conclude that the PER decreases as the distribution gives more credits to the firsts fragments in the IP packet.

However, such distributions are also less robust to imperfect feedback, as shown in Fig. 3.16, where the PER of an ARQ scheme is plotted versus p_0 for $N = 4$, $L = 2$, $C = 8$. Imperfect feedback is simulated using a fixed probability $p_{\text{fb}} = 10^{-1}$. The figures confirm the previous conclusion, *i.e.* the distribution $L^{(0)} = [3, 2, 2, 1]$ is better than FBS and the uniform distribution $L^{(0)} = [2, 2, 2, 2]$ in the ideal feedback case. But for imperfect feedback, while the uniform distribution is robust and remains very close to the ideal

(a) ARQ
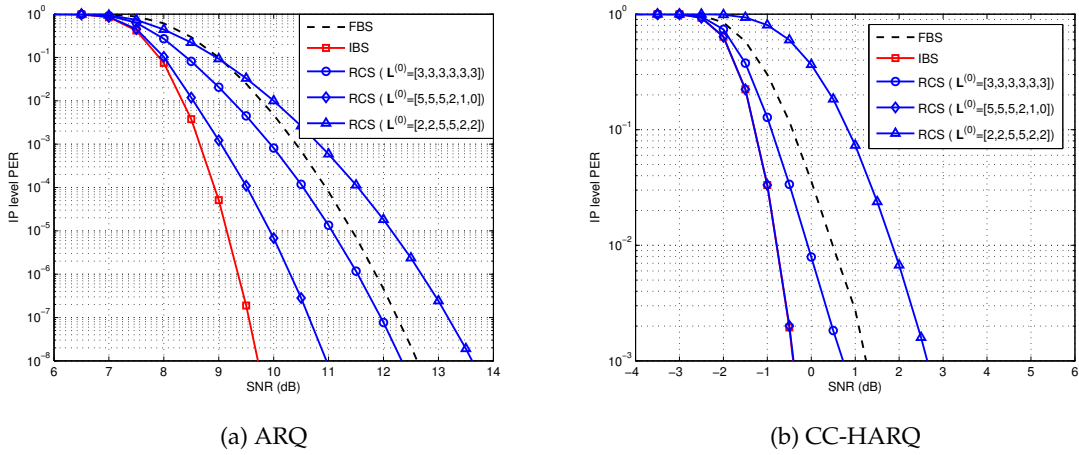
(b) CC-HARQ

Figure 3.15: Effect of different initial distributions $L^{(0)}$ on IP level PER of ARQ/CC-HARQ with RCS ($N = 6$, $L = 3$, $C = 18$, $p_e = 0$, $p_c = 0$).

case, the distribution $L^{(0)} = [3, 2, 2, 1]$ degrades the performance and becomes even worse than FBS at low p_0.
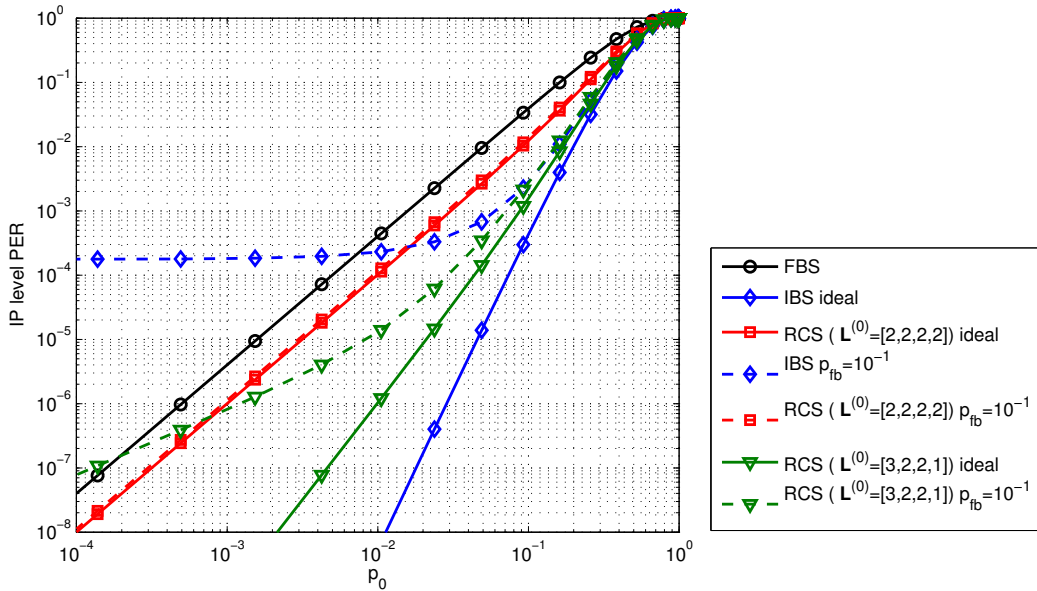
Figure 3.16: Effect of different initial distributions $L^{(0)}$ IP level PER of ARQ versus p_0 with RCS and imperfect feedback ($N = 4$, $L = 2$, $C = 8$, $p_{fb} = 10^{-1}$).

3.8   Conclusion

This Chapter has been devoted to the analysis of cross-layer HARQ schemes with imper-
fect feedback. We have proposed a model for two kinds of feedback impairments: errors
in the acknowledgment messages, and delayed feedback. We have derived new analytic
expressions at IP level for the PER, the delay and the efficiency of any HARQ scheme.

A numerical analysis has shown that such imperfect feedback conditions may have
more or less impact on the HARQ performance according to the chosen cross-layer re-
transmission management (*i.e.*, FBS or IBS). While the PER is not modified by any
feedback imperfection in the FBS case, the PER of IBS is dramatically degraded, and thus
the impact on the two other metrics (delay and efficiency) is significant in this case. If
feedback is transmitted into packets, the best FBS performance are achieved by using
coding inside the feedback and by setting a time-out value that is close to the average
arrival time $1/\lambda$. In contrast, the time-out value must be chosen larger in the IBS case,
and even coding cannot retrieve the ideal performance.

Therefore, it is of great interest to design cross-layer schemes that still have a gain
but are more robust to imperfect feedback than the IBS. This issue has been successfully
addressed with the definition of the RCS scheme, for which the analysis has been con-
ducted within a unified framework. The choice of the initial credit distribution offers a
soft transition from the robustness of FBS against imperfect feedback, to the cross-layer
gain brought by IBS.

Finally, the HARQ performance were studied for the SW protocol only. Since LTE
implements a parallel SW version of the HARQ, it could be interesting to extend the
cross-layer concepts of RCS to parallel SW. This extension is straightforward for FBS
since the ARQ contexts that are placed in parallel are independent. More precisely, the
credit distribution within each context is independent of the other contexts, which is not
true for cross-layer strategies.

We have patented the RCS scheme in [P1]. Part of the materials presented during this
Chapter were published in [C2], [C3], [C6] and [C7].

Chapter 4

Resource allocation problems in mobile ad hoc networks

4.1 Introduction

Unlike cellular systems, which are organized around base stations deployed by the operator, ad hoc networks are relaxed from any fixed infrastructure. Ad hoc networks are thus a highly flexible solution for fast, short-lived communications deployment for many applications, from military ground or other critical scenarios to any future smart network. However, the lack of structure makes the resource management difficult compared to cellular systems, and ad hoc networks suffer from several practical limitations. This Chapter moves towards the second part of the thesis, which is devoted to the resource allocation issue in the paradigm of ad hoc networks. We draw an overview of the ad hoc network context, and the notions tackled here will be used throughout the rest of the thesis.

The Chapter is organized as follows. The context of the study is given in Section 4.2, where we describe the system to which the materials contained within these three last Chapters can be applied, and define the needs and objectives. Section 4.3 presents the main characteristics of the system and formulates the assumptions that are made until the end of the thesis. The associated mathematical model and notations are given in Section 4.4, and will be common to Chapters 5 and 6. Finally, the state of the art is done in Section 4.5 and Section 4.6 formulates the optimization problem.

4.2 Working context

In the rest of the thesis, we investigate a MANET that may be deployed either in a civilian context, or on a military scene. In order to simplify the network management, a cluster-based structure is advocated where a Cluster Head (CH) is elected among the nodes, as illustrated in Fig. 4.1.

The nodes in a clustered MANET are managed by the CH, which collects the trans-
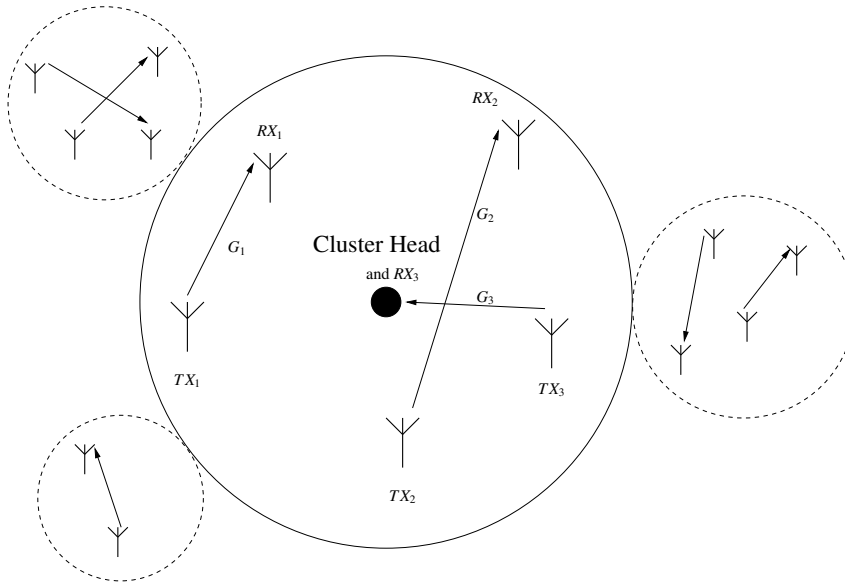
Figure 4.1: Clustered mobile ad hoc network.

mitters' requests and performs a centralized resource allocation accordingly. However, unlike the Base Station which centralizes all the communications in an uplink-downlink fashion in cellular wireless networks, the CH does not relay the information between a pair of users. Instead, in order to avoid the concentration of all the traffic at the CH, peer to peer links are built after each resource allocation stage, and thus pairwise communications are done between the communicating nodes. The transmissions follow a Time Division Multiple Access (TDMA) scheme with specific slots reserved for signaling and data, as done in the specific Thales devices.

The main goal of resource allocation is to organize the available physical resource among the transmitting users. In our case, depicted in Fig. 4.2, the objective is to select the power level and the bandwidth sharing for a communication in order to minimize the total transmit power subject to some long-term QoS constraints.

Total transmit power minimization enables the minimization of the nodes' consumption, as well as the reduction of the frequency spatial footprint of the network, while maintaining the link fidelity. Moreover, power minimization is of great interest to increase the network lifetime, to mitigate the inter-cluster interference and, from a military point of view, to provide low detection capability.

4.3 Clustered mobile ad hoc networks: assumptions

The designer of a communication system may face several issues, that we identified into two classes of concerns:
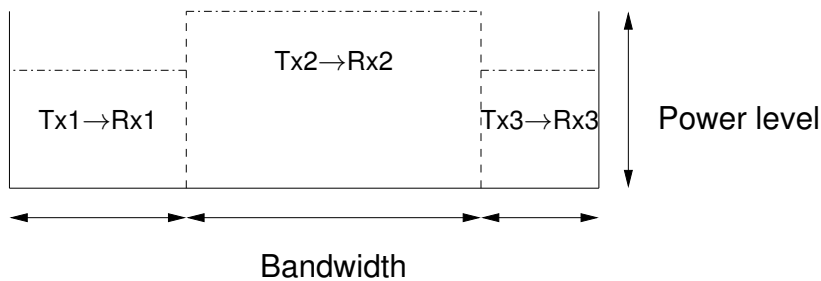
Figure 4.2: How to assign the resource inside the cluster?

(i) How to cope with the interference that is inherent to wireless networks?

(ii) Can we use some CSI to improve the communication performance?

In this Section, we point out what assumptions can be made on the system described in Section 4.2 to address these two questions.

4.3.1 Interference management

Since the communications are wireless, it is obvious that the pairs of users that communicate will interfere with each other. Clustering the network as done in Fig. 4.1 leads to two levels of interference:

(i) interference between the clusters (inter-cluster),

(ii) interference between the users inside a same cluster (intra-cluster).

In what follows we show how we can deal with.

Inter-cluster interference: The clusters composing the overall network are spread over the entire available system bandwidth. Therefore, two users from two distinct clusters that communicate over the same resource will interfere, in spite of any separation or interference treatment inside inside the cluster.

We first neglect the inter-cluster interference in the developments made in Chap. 5 and 6. Taking inter-cluster interference into account, these results can be extended to joint cluster resource allocation following the ideas developed in [Ksairi et al., 2010a] and [Ksairi et al., 2010b].

Intra-cluster interference: The approach that is retained in this thesis is to organize orthogonal communication schemes to separate the pairs inside a cluster. Though suboptimal from an information theoretic point of view [Tse and Viswanath, 2005], this solution can be easily implemented. Typically, OFDMA, which combines the so-called OFDM technique to combat inter-symbol interference due to multipath spread, and FDMA to

separate the users, has been widely considered since it is a promising solution for future wireless standards. We point out that OFDMA is considered in the rest of the thesis. Actually, any other orthogonal scheme like Single Carrier FDMA could be envisaged.

Moreover, perfect synchronization of the underlying OFDM will be assumed, so that the subcarriers are perfectly orthogonal with no frequency shift, as well as perfectly designed cyclic prefix in order to absorb the multipath spread. Therefore, the pairs of communicating users are perfectly separated using OFDMA inside a given cluster.

4.3.2   Channel state information

The channel information is generally difficult to obtain: old 2G networks did not report perfect CSI from the mobile to the base station and so did not performed power control based on perfect CSI. Even the 3G (UMTS) could not have perfect channel knowledge due to the frequency spread of Code Division Multiple Access (CDMA). In spite of the wireless mobile environment, some cellular networks (LTE) where designed so that the channel impulse response may be quickly provided at the transmitter since the Base Station (resource allocator) is always either the transmitter or the receiver. For instance, in downlink TDD mode, the CSI is available at the Base Station without additional cost due to channel reciprocity, and so even if the channel is fast varying (less than a few hundreds Hertz Doppler frequency). The case of Frequency Division Duplex (FDD) is even simpler since the CSI is available on the uplink with less latency.

In peer-to-peer communications based ad hoc networks as described in Fig. 4.1, reporting the CSI at the resource allocator is much more difficult and often irrelevant. Indeed, at least two reasons prevent us to consider instantaneous CSI at the resource allocator:

(i) CSI reporting increases the overhead due to the signaling. Thus the amount of CSI demand for each pairwise link is huge and may flood the network.

(ii) A link between each receiver and the resource allocator must be created for reporting the CSI in an ad hoc network. Non-negligible part of time may be spent to establish this communication, which leads to feedback delays that may be larger than the channel coherence time, leading to outdated CSI.

Exploiting outdated feedback at the resource allocator leads to imperfect CSI knowledge. Some works proved that even imperfect CSI can provide benefit (see [Goldenbaum et al., 2011] or [Szczecinski et al., 2011]). Therefore a lot of works have optimized resource allocation under the assumptions of either full CSI (see [Tse and Viswanath, 2005] and references therein) or imperfect/partial CSI ([Wang and Lau, 2008], [Lau et al., 2008], [Rui and Lau, 2008], [Brah et al., 2008] and [Ho et al., 2009]). Nevertheless, all these methods assume that the CSI imperfection is small enough, *i.e.*, the channel is only slightly different between two CSI reports, and so is highly correlated between the reporting phase and the resource allocation phase. In the case of MANETs, the previous work cannot be

applied since the delay between the CSI measurements and the time it is available at the resource allocator is too large.

Thus, the resource allocation presented in the thesis relies on the long-term average channel conditions instead of instantaneous channel fading. Basically, we assume that the channel statistics vary slow enough, *i.e.* remain constant within few TDMA frames. Nevertheless, dealing with statistical CSI requires some restrictions.

Diversity handling: Since the short-term fluctuations of the channel are not known, the resource allocator cannot determine the allocated power for each subcarrier, nor each subcarrier allocated to which user. Therefore, the so-called multiuser diversity [Tse and Viswanath, 2005] cannot be achieved.

The system is based on OFDM, thus time and frequency diversities can be exploited to counteract the lack of CSI. Since both time/frequency diversity and multiuser diversity are not necessarily cumulative [Tse and Viswanath, 2005], not considering multiuser diversity does not prevent the designed system to work efficiently. To achieve single-user diversity, we resort to:

(i) Frequency Hopping (FH) for handling single-user diversity. Moreover, FH provides an interesting way to counteract eavesdroppers from a military point of view.

(ii) HARQ for enforcing the link performance, since it is a powerful mechanism to cope with the unknown fast channel variations efficiently.

4.4 Mathematical model

This Section introduces the basic notations that will be used throughout the rest of the document. We present the mathematical channel model derived from the assumptions of Section 4.3, and we define the parameters of interest for the optimization. In all the document, the superscript ^{T} stands for the transposition operator, the (multivariate) complex-valued circular Gaussian distribution with mean a and covariance matrix Σ is denoted $\mathcal{CN}(a, \Sigma)$, and f^{-1} is the inverse of any function f with respect to composition.

4.4.1 Channel model

4.4.1.1 OFDM signal

The channel corresponds to the link between the transmitting user k and any receiving node in the network, including the CH. Let $\boldsymbol{h}_k(i) = [h_k(i, 0), \ldots, h_k(i, M-1)]^T$ denote the channel impulse response of link k associated with OFDM symbol i, where M is the number of taps. Let us denote by $\boldsymbol{H}_k(i) = [H_k(i, 0), \ldots, H_k(i, N_c - 1)]^T$ the Fourier Transform of $\boldsymbol{h}_k(i)$, where N_c is the number of OFDM subcarriers. Assuming well-designed OFDM

cyclic prefix and FH pattern, the received signal at OFDM symbol i and subcarrier n for user k is:

$$Y_k(i,n) = H_k(i,n)X_k(i,n) + Z_k(i,n), \tag{4.1}$$

where $X_k(i,n)$ is the transmitted symbol by user k at subcarrier n of OFDM symbol i, and the additive noise $Z_k(i,n) \sim \mathcal{CN}(0, N_0 W/N_c)$ where N_0 is the noise power spectral density and W is the total bandwidth. It is assumed that each channel is an independent Gaussian random process with possibly different variances $\varsigma_{k,m}^2$ for each tap, $i.e.$, $\boldsymbol{h}_k(i) \sim \mathcal{CN}(0, \Sigma_k)$ with $\Sigma_k := \operatorname{diag}_{M \times M}(\varsigma_{k,m}^2)$. Hence, direct calculation shows that the diagonal elements $H_k(i,n)$ of the Fourier Transform matrix $\boldsymbol{H}_k(i)$ are identically distributed[1]:

$$H_k(i,n) \sim \mathcal{CN}(0, \varsigma_k^2), \tag{4.2}$$

with $\varsigma_k^2 := \operatorname{Tr}(\Sigma_k)$. Thus, the subcarriers of a single link are identically distributed.

4.4.1.2   Channel state information

Let $g_k(i,n) := |H_k(i,n)|^2/N_0$ be the instantaneous Gain to Noise Ratio (GNR) of link k at subcarrier n and OFDM symbol i. It is exponentially distributed, with a mean $G_k := \mathbb{E}[g_k(i,n)]$ independent of n, given by:

$$G_k = \frac{\varsigma_k^2}{N_0}. \tag{4.3}$$

We assume that the CH only knows the terms G_k, $i.e.$, the average GNR instead of the instantaneous one, for all the active links. Since G_k depends on k, the users are obviously treated differently, and we assume that the behavior of G_k is driven by the so-called path-loss. Let D_k be the distance between user k and its corresponding receiver. Then:

$$G_k = \frac{\ell(D_k)}{N_0}, \tag{4.4}$$

where $\ell(.)$ is some function dependent on the path-loss model.

Remark on the pairwise links:

The multiuser scheme of the MANET described in Fig. 4.1 is implemented with pairwise communications through the network. At this step, and regardless of the scheduling:

- For one-to-many communications the only difference by still using OFDM is in the received OFDM signal expression:

$$Y_k(i,n) = H_k(i,n)X(i,n) + Z_k(i,n),$$

---

[1]Notice that the elements of $\boldsymbol{H}_k(i)$ are not necessarily independent. Actually, only M elements are independent, the others are obtained by linear combination. However, $(\boldsymbol{H}_k(i))_{i \geq 0}$ is an independent random process (since the channel realization is assumed to be different from one OFDM symbol to an other one), $i.e.$ $\boldsymbol{H}_k(i)$ is independent of $\boldsymbol{H}_k(j)$ for $i \neq j$.

where $X(i, n)$ is the symbol broadcasted to link k at time i over subcarrier n. Thus, the results extension is straightforward. However, one could envisage a more sophisticated use of this downlink (broadcast channel), and in that case the results do not hold anymore.

• For many-to-one communications there is no difference when using OFDM, since the user receives simultaneously (and interference-free) the messages X_k from links k on the assigned subcarriers. Once again, a deeper use of this uplink (multiple access channel) will strongly modify our results.

4.4.2   Power and bandwidth parameters

Let us denote by K the number of links that are active in the considered cluster. Since G_k does not depend on subcarrier n, the resource allocation algorithm will not distinguish between subcarriers for a given users pair k, *i.e.*, the CH cannot attribute **which** subcarriers for user k, but only **how many**. Let n_k be the number of subcarriers assigned to the pair of users k, so the bandwidth proportion occupied by this link is:

$$\gamma_k := \frac{n_k}{N_c}, \tag{4.5}$$

and corresponds to the **bandwidth parameter to be optimized**. By definition, the inequality $\sum_{k=1}^{K} \gamma_k \leq 1$ must hold.

Due to the independence of G_k with respect to the subcarrier n, it is natural for the transmitter k to use the same average power $P_k = \mathbb{E}\left[|X_k(i, n)|^2\right]$ on each subcarrier. Let:

$$E_k := P_k/(W/N_c) \tag{4.6}$$

be the energy consumed to send one symbol on each subcarrier, and

$$\sigma_k^2 := N_0 W/N_c \tag{4.7}$$

be the corresponding noise variance. The energy E_k corresponds to the **power parameter to be optimized**. Then, on each subcarrier user k undergoes an average SNR given by:

$$\text{SNR}_k = \frac{\varsigma_k^2 P_k}{\sigma_k^2} = E_k G_k. \tag{4.8}$$

The definition of these parameters is summarized in Fig. 4.3.

4.4.3   Resource allocation optimization issue

In order to mitigate the power radiated by a single cluster or to minimize battery drain, one can minimize:

$$\mathcal{P}_T = \sum_{k=1}^{K} n_k P_k. \tag{4.9}$$
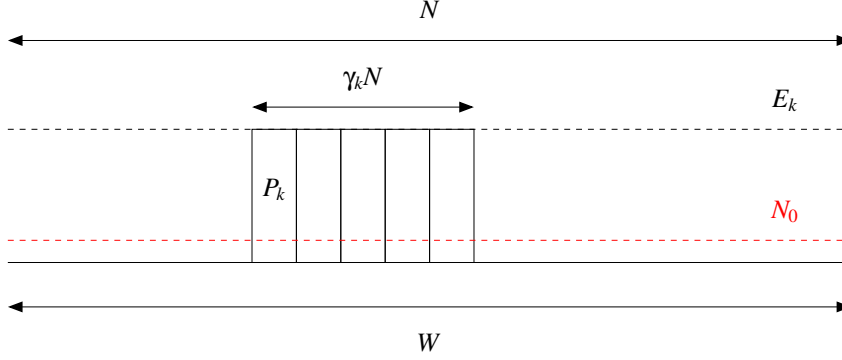
Figure 4.3: Resource assignment to user k.

When the total bandwidth W is fixed the total power of the cluster becomes, using Eqs. (4.5)-(4.6):

$$\mathcal{P}_T = W \sum_{k=1}^{K} \gamma_k E_k. \tag{4.10}$$

Hence, it is equivalent to minimize the total energy $Q_T := \mathcal{P}_T/W$ for sending an OFDM symbol, *i.e.* to minimize the objective function:

$$Q_T = \sum_{k=1}^{K} \gamma_k E_k. \tag{4.11}$$

From Eq. (4.5) γ_k are rational numbers, however for tractability purposes we take $\gamma_k \in [0, 1]$ to make the problem continuous.

Finally, in the sequel the users on link k may request some minimal QoS, which will be denoted by the vector $\mathbf{QoS}_k^{(0)}$. Let us define the bivariate vector function $\mathbf{QoS}_k(.,.)$ that represents a set of QoS constraints: this function will be explicited in the rest of the thesis according to the users' needs, and can be either the rate, or the PER, or the delay, or any combination of them. The optimization problem is formalized in Problem 4.1:

Problem 4.1. *Let us denote* $\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_K]^T$ *and* $\boldsymbol{E} = [E_1, \cdots, E_K]^T$. *The optimization problem boils down to:*

$$\min_{(\gamma, E)} \sum_{k=1}^{K} \gamma_k E_k, \tag{4.12a}$$

$$s.t. \quad \sum_{k=1}^{K} \gamma_k \leq 1, \tag{4.12b}$$

$$\gamma_k \geq 0, \ E_k \geq 0, \ \forall k, \tag{4.12c}$$

$$\mathbf{QoS}_k(\gamma_k, E_k) \geq \mathbf{QoS}_k^{(0)}, \ \forall k. \tag{4.12d}$$

4.5 State of the art

The MCS used in current systems are not often optimal, *i.e.*, capacity achieving. This last practical limitation is not specific to MANETs, and this thesis aims to take up this challenge too. In the following, we review some existing results that fall within, or close to, the context of OFDMA resource allocation with statistical CSI. We have identified three types of results in the literature related to our context:

   (i) information-theoretic tools based allocations when the link is capacity achieving (in Shannon's sense) and when statistical CSI is available. Hence, the metric related to the rate is the so-called ergodic capacity ([Gault et al., 2007]). Other metrics (such as delay and PER) have been taken into account since the delay is theoretically infinite due to the infinite-length of Gaussian random codes and since the PER is arbitrarily small. The works done in this context have been reviewed in Section 4.5.1.

   (ii) information-theoretic tools based allocations when the link is capacity achieving but with finite inputs and when full CSI at the transmitter is available ([Lozano et al., 2006]). We have straightforwardly extended this work to the case of statistical CSI in Section 4.5.2.

   (iii) allocations for links built upon practical modulations and codes, that can be either derived from information theory tools using gaps ([Wong et al., 1999]), or derived from new metrics ([Devillers et al., 2008] and [Wu and Jindal, 2011]). In both cases, the CSI is perfectly known at the transmitter. These works have been summarized in Section 4.5.3.

4.5.1 Information-theoretic tools based allocation with continuous modulation schemes

It is well known that the capacity of ergodic channels with Rayleigh fading is attained for Gaussian input signals ([Tse and Viswanath, 2005]). In the major part of the literature relative to (OFDM) resource allocation the MCS is assumed to be ideal, *i.e.* to consist in infinitely long Gaussian entries, and thus capacity achieving. In that case, the bit loading relies on the Shannon formula, which gives the number of bits that can be reliably transmitted at every channel use. Therefore, the optimization procedures are based on the celebrated $\mathbb{E}\left[\log(1 + \text{SNR})\right]$ function, which exhibits some attractive mathematical features like concavity.

The minimization of an OFDMA cell power, neglecting inter-cell interference, assuming only statistical CSI $(G_k)_{k \in \{1,...,K\}}$ at the Base Station and assuming QoS only related to rate constraint, was investigated in [Gault et al., 2007] through solving the following

optimization problem:

$$\min_{(\gamma, E)} \sum_{k=1}^{K} \gamma_k E_k, \tag{4.13a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \gamma_k \leq 1, \tag{4.13b}$$

$$\gamma_k \geq 0, \ E_k \geq 0, \ \forall k, \tag{4.13c}$$

$$C_k(\gamma_k, E_k) \geq \rho_k^{(0)}/W, \ \forall k, \tag{4.13d}$$

where the users require minimal rates $\rho_k^{(0)}$ driven by the ergodic capacity C_k of their link:

$$C_k = \gamma_k \mathbb{E}\left[\log\left(1 + \frac{|H_k(i,n)|^2 E_k}{N_0}\right)\right]. \tag{4.14}$$

Due to the convexity of the optimization problem, the optimal allocation policy was found and was given as:

$$E_k^* = \frac{1}{G_k} f^{-1}(G_k \lambda^*), \tag{4.15}$$

$$\gamma_k^* = \frac{\rho_k^{(0)}/W}{F(G_k \lambda^*)}, \tag{4.16}$$

where f is the real-valued function given by:

$$f(x) = \frac{x^2 e^{1/x} E_1(1/x)}{x - e^{1/x} E_1(1/x)} - x \tag{4.17}$$

with $E_1(x) := \int_1^{+\infty} (e^{-xt}/t)\,\mathrm{d}t$ the exponential integral[2], and $F(x) := \mathbb{E}\left[\log(1 + f^{-1}(x) X_e)\right]$ with X_e an exponentially distributed random variable with parameter 1. The optimal Lagrange multiplier λ^* is a non-negative scalar that satisfies:

$$\sum_{k=1}^{K} \frac{\rho_k^{(0)}/W}{F(G_k \lambda^*)} = 1. \tag{4.18}$$

4.5.2 Information theoretic tools based allocation with finite-size modulation schemes

The constellation sizes are, in practice, constrained to be integer, and the information theoretic measure must be adapted. Non-Gaussian entries is the first practical limitation that can be studied in resource allocation and, surprisingly, rare works focused on this aspect.

Only [Lozano et al., 2006] gave a complete theory of arbitrary modulation inputs used for power allocation in parallel Gaussian channels, which can be directly applied to

---

[2]Actually, the exponential integral $E_1(.)$ is defined in [Abramowitz and Stegun, 1972] as $E_1(z) := \int_z^{+\infty} (e^{-t}/t)\,\mathrm{d}t$ for $|\arg z| < \pi$, and coincide for $n = 1$ with $E_n(z) = \int_1^{+\infty} (e^{-zt}/t^n)\,\mathrm{d}t$, for $\Re z > 0$.

OFDMA resource allocation with Quadrature Amplitude Modulation (QAM) signals for instance. The authors investigated sum-capacity maximization with full CSI at the Base Station, but their work can easily be applied to total power minimization with statistical CSI through the following optimization problem:

$$\min_{(\gamma, E)} \sum_{k=1}^{K} \gamma_k E_k, \tag{4.19a}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \gamma_k \leq 1, \tag{4.19b}$$

$$\gamma_k \geq 0, \, E_k \geq 0, \, \forall k, \tag{4.19c}$$

$$C_k(\gamma_k, E_k) \geq \rho_k^{(0)}, \, \forall k, \tag{4.19d}$$

where the users require minimal rates $\rho_k^{(0)}$ driven by

$$C_k = \gamma_k \mathcal{I}_k(P_k), \tag{4.20}$$

with $\mathcal{I}_k(x)$ the ergodic mutual information of link k with QAM entries.

In [Weidong et al., 2007], the mutual information of link k with QAM entries of order M_k is given for one channel realization. The ergodic mutual information is just the average of the closed-form expressions provided in [Weidong et al., 2007] over all the channel realizations. Therefore, we have

$$C_k = \gamma_k \mathbb{E}\left[\log\left(1 + \frac{|H_k(i,n)|^2 E_k}{N_0}\right) - \frac{1}{2}\log\left(1 + \left(\frac{|H_k(i,n)|^2 E_k}{N_0 M_k}\right)^2\right)\right]. \tag{4.21}$$

In Appendix D.1, it is proven that optimization problem related to Eqs. (4.19) is convex. Then the optimal policy is given below and is proven in Appendix D.2:

$$E_k^* = \frac{1}{G_k} f_k^{-1}(G_k \lambda^*), \tag{4.22}$$

$$\gamma_k^* = \frac{\rho_k^{(0)}/W}{F_k(G_k \lambda^*)}, \tag{4.23}$$

where f_k^{-1} is the inverse with respect to composition of the real-valued function f_k given by:

$$f_k(x) = x^2 \frac{e^{1/x} E_1(1/x) - C(M_k/x)}{M_k S(M_k/x) - e^{1/x} E_1(1/x)} - x, \tag{4.24}$$

with $C(x) := -\text{ci}(x)\cos(x) - \text{si}(x)\sin(x)$ and $S(x) := \text{ci}(x)\sin(x) - \text{si}(x)\cos(x)$, si (.) and ci (.) are the sine and cosine integrals [Gradshteyn and Ryzhik, 1980], respectively. The function $F_k(x) := \mathbb{E}\left[\log(1 + f_k^{-1}(x) X_e) - (1/2)\log(1 + (f_k^{-1}(x))^2 X_e^2/M_k^2)\right]$ with X_e an exponentially distributed random variable with parameter 1. Again, the optimal λ^* is chosen such that:

$$\sum_{k=1}^{K} \frac{\rho_k^{(0)}/W}{F_k(G_k \lambda^*)} = 1. \tag{4.25}$$

Notice that Eq. (4.21) admits actually a closed-form solution given in Appendix D.3.

4.5.3 Allocation with practical modulation and coding schemes

The previous paragraph has presented some works that have taken into account for the practical limitation of the constellation size. Nevertheless, these works still assume infinitely long codes, which is far more limiting when considering some delay issues. Moreover such coding schemes are capacity achieving so the bits are conveyed reliably at rate $\log(1+\text{SNR})$, *i.e.* the error probability at this rate is made arbitrarily small. Since most practical applications are constrained to non-capacity achieving codes satisfying some QoS like Bit Error Rate (BER) or PER, this Section presents some works that investigated resource allocation regarding the demanded QoS level, but with full CSI knowledge.

The basic idea is to introduce metrics that are able to measure the transmission rate that can be reached at some fixed BER/PER. To the best of the author knowledge, this was initiated in [Wong et al., 1999], where the problem was formulated and solved for full CSI $H_k(i,n)$ available at the Base Station:

$$\mathcal{P}_T^* = \min_{m_{k,n}\in\mathcal{M}} \sum_{k=1}^{K}\sum_{n=1}^{N_c} \frac{N_0}{|H_k(i,n)|^2} f_k(m_{k,n}), \tag{4.26a}$$

$$\text{s.t.} \quad m_{k,n} = \rho_k^{(0)}/W, \ \forall k, \tag{4.26b}$$

$$m_{k,n} = 0 \ \forall k \neq k' \text{ such that } m_{k',n} \neq 0, \ \forall n, \tag{4.26c}$$

where the function $f_k(.)$ gives the power needed to transmit $m_{k,n}$ bits at some required BER, and $m_{k,n}$ are constrained to take values from a discrete ensemble $\mathcal{M}$ that represents the modulation states.

This approach can be generalized by resorting to the so-called SNR gap [Starr et al., 1999]. When using constant energy per symbol E_s at the transmitter, the BER on any subcarrier n of link k can be expressed from the symbol error expression:

$$\text{BER}_k = 4Q\left(\sqrt{\frac{3}{2}\frac{|H_k(i,n)|^2}{2^{m_{k,n}}-1}\frac{E_s}{N_0}}\right), \tag{4.27}$$

so when fixing some target BER to $\text{BER}_k^{\text{target}}$ we can extract:

$$2^{m_{k,n}} - 1 = \frac{|H_k(i,n)|^2 E_s/N_0}{\frac{2}{3}\left(Q^{-1}\left(\frac{\text{BER}_k^{\text{target}}}{4}\right)\right)^2}. \tag{4.28}$$

We define the positive constant:

$$\Gamma_k := \frac{2}{3}\left(Q^{-1}\left(\frac{\text{BER}_k^{\text{target}}}{4}\right)\right)^2 \tag{4.29}$$

as the SNR gap from Shannon capacity for which the number of bits $m_{k,n}$ that can be conveyed over the subcarrier n of link k at $\text{BER}_k^{\text{target}}$ is:

$$m_{k,n} = \left\lfloor \log\left(1 + \frac{|H_k(i,n)|^2 E_s/N_0}{\Gamma_k}\right)\right\rfloor. \tag{4.30}$$

This formalism can be used in conjunction with the waterfilling algorithm to maximize the bit rate under BER constraint (see [Starr et al., 1999] and [Devillers et al., 2008]). However, taking gaps from the capacity is possible only when CSI is available, because Γ_k is computed from expressions that hold true on the AWGN channel, *i.e.* the gap is evaluated knowing the channel coefficient. Therefore in our case where the CSI is not available at the CH, but only statistics are known, adapting such techniques to the ergodic capacity is impossible and the work of [Gault et al., 2007] presented in Section 4.5.1 cannot be extended to practical MCS in this way.

When working with finite-length coding schemes, which do not achieve the ergodic capacity anymore [Tse and Viswanath, 2005], the channel does not support the rates given by the Shannon formula for arbitrarily low error probabilities. Instead, in order to cope with the inevitable packet errors, (H)ARQ can be used on top of the FEC to improve the link reliability. In that case, it has been recently recognized in [Wu and Jindal, 2011] or [Devillers et al., 2008] that the so-called HARQ **goodput** was the meaningful metric to drive the useful bitrate at MAC level since it captures the non-null packet error probability as well. Therefore, our goal is to perform resource allocation with statistical CSI where the goodput dictates the rate requirement of the links.

4.6   Optimization problem

In this Section we detail the QoS constraints function **QoS**$_k$ of Problem 4.1 for two cases:

  (i) finite-length Gaussian codes,

  (ii) practical MCS composed of existing FEC codes and QAM modulations.

Then we derive several optimization problems that will be solved in the next Chapters.

In both cases it will be assumed that each user must be served with a minimum average rate ρ_k, thus there are strictly positive constants $\rho_k^{(0)}$ such that $\rho_k \geq \rho_k^{(0)}$ for all k. In order to cope with the remaining packet errors and fast channel variations, Type-I HARQ is used. We remind from Chapter 1 that the user (error-free) bitrate ρ_k (in bit/s) is proportional to the goodput $\tilde{\eta}_k$:

$$\rho_k = \tilde{\eta}_k W_k, \tag{4.31}$$

where $W_k := \gamma_k W$ is the portion of bandwidth occupied by link k. Thus, in order to remain bandwidth independent, each user requires a minimal goodput $\tilde{\eta}_k \geq \eta_k^{(0)}$ given by:

$$\tilde{\eta}_k(\gamma_k, E_k) = \gamma_k r_k \eta_k(G_k E_k), \tag{4.32}$$

where r_k is the information rate (in bits) conveyed over link k every channel use, and $\eta_k : \text{SNR} \mapsto \eta_k(\text{SNR})$ is the efficiency of link k (as defined in Chapter 1).

4.6.1 Finite-length Gaussian codes

In this case the Type-I HARQ efficiency is:

$$\eta_k(\text{SNR}) = 1 - P_e^{(n)}(\text{SNR}, r_k), \tag{4.33}$$

with $P_e^{(n)}(\text{SNR}, r_k)$ the packet error probability of the Gaussian code of length n and rate r_k. For the sake of clarity, it can be assumed without loss of generality that the users have the same code length n. Thus, taking $\mathbf{QoS}_k = \tilde{\eta}_k$ and using Eqs. (4.32)-(4.33) in Problem 4.1 boils down to the new optimization problem:

Problem 4.2.

$$\min_{(\gamma, E)} \sum_{k=1}^{K} \gamma_k E_k, \tag{4.34a}$$

$$s.t. \quad \gamma_k r_k \left(1 - P_e^{(n)}(G_k E_k, r_k)\right) \geq \eta_k^{(0)}, \ \forall k, \tag{4.34b}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{4.34c}$$

$$\gamma_k \geq 0, \ E_k \geq 0, \ \forall k. \tag{4.34d}$$

Problem 4.2 will be solved in Chapter 5 for a given r_k.

4.6.2 Practical MCS

In this case the Type-I HARQ efficiency is:

$$\eta_k(\text{SNR}) = R_k \left(1 - P_k(\text{SNR})\right), \tag{4.35}$$

with $P_k(\text{SNR})$ the PER of the FEC code of given rate R_k. The QAM order M_k is given, and we have $r_k = m_k$ with $m_k = \log_2(M_k)$ the number of bits conveyed by the constellation. Obviously, M_k and R_k have to be chosen during the resource allocation and this choice will be discussed in details in Chapter 6. Thus, taking $\mathbf{QoS}_k = \tilde{\eta}_k$ and using Eqs. (4.32)-(4.35) in Problem 4.1 boils down to the new optimization problem:

Problem 4.3.

$$\min_{(\gamma, E)} \sum_{k=1}^{K} \gamma_k E_k, \tag{4.36a}$$

$$s.t. \quad \gamma_k m_k R_k \left(1 - P_k(G_k E_k)\right) \geq \eta_k^{(0)}, \ \forall k, \tag{4.36b}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{4.36c}$$

$$\gamma_k \geq 0, \ E_k \geq 0, \ \forall k. \tag{4.36d}$$

Although its mathematical expression depends on the PHY level PER value, the goodput is not enough for completely characterizing a link performance. Indeed, as stated in [Ho et al., 2009], the MAC level PER has also to be kept below a certain threshold $P_k^{\text{MAC},(0)}$. Therefore, taking $\mathbf{QoS}_k = [\tilde{\eta}_k \ P_k^{\text{MAC}}]^T$ in Problem 4.1 leads to:

Problem 4.4.

$$\min_{(\gamma,E)} \sum_{k=1}^{K} \gamma_k E_k, \tag{4.37a}$$

$$s.t. \quad \gamma_k m_k R_k \left(1 - P_k(G_k E_k)\right) \geq \eta_k^{(0)}, \ \forall k, \tag{4.37b}$$

$$P_k^{MAC}(\gamma_k, E_k) \leq P_k^{MAC,(0)}, \ \forall k, \tag{4.37c}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{4.37d}$$

$$\gamma_k \geq 0, \ E_k \geq 0, \ \forall k. \tag{4.37e}$$

Finally, although constraining the packet errors at MAC level enables to roughly control the transmission delay (also constrained by the maximum transmission credit L), it is better to consider the HARQ delay d_k^{MAC} at MAC level as the delay constraint. Taking $\mathbf{QoS}_k = [\tilde{\eta}_k \ d_k^{\text{MAC}}]^T$ in Problem 4.1 leads to the fourth and last problem:

Problem 4.5.

$$\min_{(\gamma,E)} \sum_{k=1}^{K} \gamma_k E_k, \tag{4.38a}$$

$$s.t. \quad \gamma_k m_k R_k \left(1 - P_k(G_k E_k)\right) \geq \eta_k^{(0)}, \ \forall k, \tag{4.38b}$$

$$d_k^{MAC}(\gamma_k, E_k) \leq d_k^{(0)}, \ \forall k, \tag{4.38c}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{4.38d}$$

$$\gamma_k \geq 0, \ E_k \geq 0, \ \forall k. \tag{4.38e}$$

4.7 Conclusion

In this Chapter, we have described the ad hoc context within which the work presented in this thesis can be applied. The rationale to the design choice of the considered MANET has been reviewed, and the different assumptions concerning this system have been carefully discussed, and have led to the channel model that will be considered in the subsequent Chapters.

The main objective is the minimization of total cluster power while fulfilling some QoS constraints. Related techniques from the state of the art have been reviewed, some of them have been selected as a work basis, the others have been rejected after discussion. Finally,

the mathematical formulation of the minimization problem has been done, and will be treated for two different realizations of the PHY layer in the two following Chapters:

- in Chapter 5 we solve the case of finite-length Gaussian codes formalized in Problem 4.2,

- in Chapter 6 we solve Problems 4.3, 4.4 and 4.5 taking into account for realistic MCS.

Chapter 5

Resource allocation for HARQ with finite length codes

5.1 Introduction

The previous Chapter has introduced the background where the designed communication schemes developed in the thesis can take place. In this Chapter, we evaluate the best performance that one could expect from a Type-I HARQ-based clustered OFDMA MANET using statistical CSI since it is assumed that Gaussian codes with a finite block length are used. The Shannon capacity, defined as the largest rate at which one can transmit error-free, is achieved for random coding by letting the block length growing to infinity [Tse and Viswanath, 2005]. When the block length is constrained to a fixed value, for delay purposes for instance, the Shannon capacity is an unreachable limit. Therefore, designing the system based on the ultimate performance available with finite size codes is of great interest.

This Chapter is organized into four parts. In Section 5.2, we review the most contributing techniques for the study of finite length coding, and the framework of information spectrum is chosen as a work basis. Then, in Section 5.3, the error probability of Gaussian codes with finite length is computed in closed-form over the Rayleigh channel. Based on this new result, the optimal power and bandwidth allocation of the users is performed in Section 5.4. Finally, some numerical results are given in Section 5.5.

5.2 Maximum rate codes with finite block length: previous works

This Section reviews the principal attempts of characterizing the error probability of maximal rate coding schemes of finite block length: the Gallager random coding bound, the more recent theory of channel dispersion, and finally the mutual information rate, which is the most promising analysis technique from our objective point of view.

5.2.1 Random coding bound

The error probability of random codes of rate R and length n has an exponential decay, as shown by the Gallager's random coding bound [Gallager, 1968]:

$$P_e^{(n)} \leq 2^{-nE_r(R)}, \tag{5.1}$$

where $E_r(R)$ is the so-called Gallager's exponent that is positive for $R < C$ (where C is the channel capacity defined in [Shannon, 1948]). Unfortunately, the Gallager's exponent does not admit simple closed-form expressions. As a consequence, the resource allocation optimization problem would be intractable.

5.2.2 Channel dispersion

An alternative approach that avoids the use of random exponents to characterize the ultimate coding performance when the code length is finite has been studied in [Polyanskiy et al., 2010]. It is proved that the maximal rate R of such codes, to sustain a given probability of error $P_e^{(n)}$, is related to the channel capacity C as:

$$R = C - \sqrt{\frac{V}{n}} Q^{-1}(P_e^{(n)}) + O\left(\frac{\log n}{n}\right), \tag{5.2}$$

where Q^{-1} is the inverse with respect to composition of the Gaussian tail given by $Q(x) := (1/\sqrt{2\pi}) \int_x^\infty e^{-t^2/2} dt$. The gap $\sqrt{V/n} \, Q^{-1}(P_e^{(n)})$ from channel capacity depends on the channel dispersion V defined in [Polyanskiy et al., 2010] as:

$$V = \lim_{P_e^{(n)} \to 0} \limsup_{n \to \infty} n \frac{(C-R)^2}{2 \log(1/P_e^{(n)})}. \tag{5.3}$$

Thus, a good approximation to the error probability of a rate R code with block length n is given by:

$$P_e^{(n)} \approx Q\left(\frac{\sqrt{n}(C-R)}{\sqrt{V}}\right). \tag{5.4}$$

For the AWGN channel with real-valued Gaussian inputs, characterized by the SNR, the previous approximation holds with [Polyanskiy et al., 2010, Th. 54]:

$$C = \frac{1}{2} \log(1 + \mathrm{SNR}) \tag{5.5}$$

$$V = \frac{\mathrm{SNR}}{2} \frac{2 + \mathrm{SNR}}{(1 + \mathrm{SNR})^2} \log^2 e. \tag{5.6}$$

The authors found the very simple expression Eq. (5.6) after tedious derivations from Eq. (5.3). Since we are interested in the Rayleigh fading channel, averaging the previous derivations from the definition of V seems difficult. Instead we will rely on another approach, more tractable.

5.2.3   Mutual information spectrum

The concept of outage probability is usually dedicated to the study of the block fading channel [Tse and Viswanath, 2005], and more generally of non-ergodic channels. Yet, in [Laneman, 2006], the author argues that even in AWGN channels (and so, in ergodic channels) the mutual information between finite size inputs X and finite size outputs Y is still a random variable, and thus has studied the distribution of the mutual information rate defined as:

$$i(X;Y) := \frac{1}{n} \log \frac{f_{X,Y}(X,Y)}{f_X(X)f_Y(Y)}. \tag{5.7}$$

where X is a codeword of length n symbols and rate R nats per symbol. An outage occurs whenever the coding rate R exceeds the mutual information rate of the transmitted codeword, defining the so-called mutual information spectrum [Han, 2003]:

$$P_o := \Pr\{i(X,Y) \le R\}. \tag{5.8}$$

Since the chance for a codeword to be recoverable is driven by its outage events, these statistics have been proposed in [Buckingham and Valenti, 2008] to represent the error probability of finite length coding schemes over the AWGN channels. Taking codewords from a (n, R) Gaussian codebook, it was found:

$$P_e^{(n)} \approx Q\left(\frac{\sqrt{n}\,(\log(1+\text{SNR}) - R)}{\sqrt{2\,\text{SNR}/(1+\text{SNR})}}\right). \tag{5.9}$$

Noticing that $C = \log(1 + \text{SNR})$ is the capacity of the AWGN channel with complex Gaussian inputs, the similarity between Eq. (5.4) and Eq. (5.9) is interesting and thus the quantity $2\,\text{SNR}/(1 + \text{SNR})$ can be interpreted as a channel dispersion when finite and complex Gaussian codewords are used. Likewise, in the sequel we choose the information spectrum to be the ultimate error probability of finite length codes over the Rayleigh channel.

5.3   The error probability of finite length Gaussian codes over the Rayleigh channel

The distribution of the mutual information rate in a (static) fading channel has been succinctly described in [Laneman, 2006]. However, for our ultimate goal is to use this framework to perform resource allocation over the Rayleigh channel, it is crucial to have an exploitable expression of the mutual information rate of the ergodic Rayleigh fading channel, which is obtained in closed-form in this Section.

5.3.1 Channel model

In this Section we recall the Rayleigh channel model. Denoting by $Y \in \mathbb{C}^n$ the channel output:

$$Y = HX + N, \tag{5.10}$$

where X and N are random vectors of length n with i.i.d. elements X_k and N_k, respectively. X_k is uniformly distributed over $\mathcal{CN}(0, E_s)$, whereas $N_k \sim \mathcal{CN}(0, N_0)$, and H is a $n \times n$ diagonal matrix with i.i.d. elements $H_k \sim \mathcal{CN}(0, \sigma_h^2)$.

Moreover, the channel gains $|H_k|$ are Rayleigh distributed, such that a random SNR can be defined as:

$$\mathrm{SNR}_k = \frac{|H_k|^2 E_s}{N_0}, \tag{5.11}$$

and is exponentially distributed with parameter $1/\overline{\mathrm{SNR}}$, where $\overline{\mathrm{SNR}} = \sigma_h^2 E_s / N_0$ is the average SNR.

5.3.2 The distribution of the mutual information rate

Since the channel model Eq. (5.10) is discrete and memoryless, we can rewrite Eq. (5.7) as the average of mutual information of n scalar inputs:

$$i(X; Y) = \frac{1}{n} \sum_{k=1}^{n} i(X_k; Y_k). \tag{5.12}$$

For the Rayleigh channel defined in Eq. (5.10), one needs only to compute the mutual information $i(X_k; Y_k)$ between two scalar inputs X_k and Y_k thanks to the i.d. property. Conditioning on the channel fading H_k in Eq. (5.10), the random variables $Y_k|(X_k, H_k) \sim \mathcal{CN}(H_k X_k, N_0)$ and $Y_k|H_k \sim \mathcal{CN}(0, |H_k|^2 E_s + N_0)$ are Gaussian. Thus, for a given fading realization $H_k = h_k$, the mutual information between X_k and Y_k is given by:

$$i_{H_k=h_k}(X_k; Y_k) = \log\left(1 + |h_k|^2 \frac{E_s}{N_0}\right) + \frac{|Y_k|^2}{|h_k|^2 E_s + N_0} - \frac{|Y_k - h_k X_k|^2}{N_0}. \tag{5.13}$$

For finite n, and letting H_k to vary, the mutual information rate $i(X, Y)$ is a random variable that we denote by Z_n:

$$Z_n := \frac{1}{n} \sum_{k=1}^{n} \log\left(1 + |H_k|^2 \frac{E_s}{N_0}\right) + \frac{1}{n} \sum_{k=1}^{n} \left(\frac{|Y_k|^2}{|H_k|^2 E_s + N_0} - \frac{|N_k|^2}{N_0}\right). \tag{5.14}$$

It was shown in [Laneman, 2006] that the term $(|Y_k|^2/(|H_k|^2 E_s + N_0) - |N_k|^2/N_0)$ was equivalent to a product of two independent random variables, hence:

$$Z_n = \frac{1}{n} \sum_{k=1}^{n} \log\left(1 + |H_k|^2 \frac{E_s}{N_0}\right) + \frac{1}{n} \sum_{k=1}^{n} \sqrt{\frac{|H_k|^2 E_s / N_0}{1 + |H_k|^2 E_s / N_0}} \, W_k, \tag{5.15}$$

where W_k are i.i.d. Laplace random variables with mean zero and parameter 1, *i.e.* $W_k \sim \mathcal{L}(0,1)$ so that $\mathbb{E}[W_k] = 0$ and $\mathrm{Var}(W_k) = 2$, and W_k is independent of H_k. Hence, we have that $Z_n = (1/n) \sum_{k=1}^n i_k$ with $(i_k)_{k \in \{1...n\}}$ an i.i.d. random process given by:

$$i_k = \log\left(1 + |H_k|^2 \frac{E_s}{N_0}\right) + \sqrt{\frac{|H_k|^2 E_s/N_0}{1 + |H_k|^2 E_s/N_0}}\, W_k. \tag{5.16}$$

For the sake of simplicity, we resort to a Gaussian approximation of Z_n for the Rayleigh channel, which was shown to be accurate in [Buckingham and Valenti, 2008] for the AWGN channel. Thus, as the sum of n i.i.d. random variables, Z_n is approximated with a Gaussian random variable $\mathcal{N}(m_n, \sigma_n^2)$, where:

$$m_n := \mathbb{E}[Z_n], \tag{5.17a}$$

$$\sigma_n^2 := \mathrm{Var}(Z_n). \tag{5.17b}$$

The mean m_n is easily obtained, since Z_n is the sum of i.i.d. random variables:

$$\begin{aligned}
m_n &= \frac{1}{n} \sum_{k=1}^n \mathbb{E}\left[\log\left(1 + |H_k|^2 \frac{E_s}{N_0}\right)\right] + \frac{1}{n} \sum_{k=1}^n \mathbb{E}\left[\sqrt{\frac{|H_k|^2 E_s/N_0}{1 + |H_k|^2 E_s/N_0}}\, W_k\right] \\
&= \frac{1}{n} \sum_{k=1}^n \mathbb{E}[\log(1 + \mathrm{SNR}_k)] + \frac{1}{n} \sum_{k=1}^n \mathbb{E}\left[\sqrt{\frac{|H_k|^2 E_s/N_0}{1 + |H_k|^2 E_s/N_0}}\right] \mathbb{E}[W_k] \\
&= \mathbb{E}[\log(1 + \mathrm{SNR}_k)]
\end{aligned} \tag{5.18}$$

since $\mathbb{E}[W_k] = 0$. Finally, it is well known that this expectation leads to:

$$m_n = e^{1/\overline{\mathrm{SNR}}} E_1(1/\overline{\mathrm{SNR}}), \tag{5.19}$$

where $E_1(x) := \int_1^\infty e^{-xu}/u\, du = \int_x^\infty e^{-t}/t\, dt$ is known as the exponential integral [Abramowitz and Stegun, 1972]. Notice that $C = \mathbb{E}[\log(1 + \mathrm{SNR}_k)]$ is the ergodic capacity of the Rayleigh channel. The variance can be computed from the conditional variance formula [Ross, 2007]:

$$\begin{aligned}
\sigma_n^2 &:= \frac{1}{n^2} \sum_{k=1}^n (\mathbb{E}[\mathrm{Var}(i_k|H_k)] + \mathrm{Var}(\mathbb{E}[i_k|H_k])) \\
&= \frac{1}{n^2} \sum_{k=1}^n \left(\mathbb{E}\left[\frac{\mathrm{SNR}_k}{1 + \mathrm{SNR}_k} \mathrm{Var}(W_k)\right] + \mathrm{Var}(\log(1 + \mathrm{SNR}_k))\right) \\
&= \frac{1}{n}\left(\mathbb{E}\left[\frac{2\,\mathrm{SNR}_k}{1 + \mathrm{SNR}_k}\right] + \mathrm{Var}(\log(1 + \mathrm{SNR}_k))\right).
\end{aligned} \tag{5.20}$$

Thus, $\sigma_n^2 = (1/n)(\varsigma^2 + \kappa^2)$, where $\varsigma^2 := 2\mathbb{E}[\mathrm{SNR}_k/(1 + \mathrm{SNR}_k)]$ and $\kappa^2 := \mathrm{Var}(\log(1 + \mathrm{SNR}_k))$. The following identity holds for the variance:

Proposition 5.1.

$$\sigma_n^2 = \frac{1}{n}\left(2 - \frac{2}{\overline{\mathrm{SNR}}}e^{1/\overline{\mathrm{SNR}}}E_1(1/\overline{\mathrm{SNR}}) - e^{2/\overline{\mathrm{SNR}}}E_1^2(1/\overline{\mathrm{SNR}}) + \int_0^\infty \log^2(1 + \overline{\mathrm{SNR}}\,t)e^{-t}\mathrm{d}t\right).$$
(5.21)

Proof. By direct computation from Eq. (5.20) one finds:

$$
\begin{aligned}
\varsigma^2 &= 2\mathbb{E}\left[\frac{\mathrm{SNR}_k}{1 + \mathrm{SNR}_k}\right] \\
&= 2\int_0^\infty \frac{u}{1+u}\frac{1}{\overline{\mathrm{SNR}}}e^{-u/\overline{\mathrm{SNR}}}\mathrm{d}u \\
&= \frac{2}{\overline{\mathrm{SNR}}}\left(-e^{1/\overline{\mathrm{SNR}}}E_1(1/\overline{\mathrm{SNR}}) + \overline{\mathrm{SNR}}\right),
\end{aligned}
$$
(5.22)

after using [Gradshteyn and Ryzhik, 1980, Eq. (3.353.5)]. Now:

$$
\begin{aligned}
\kappa^2 &= \mathbb{E}\left[\log^2(1 + \mathrm{SNR}_k)\right] - \left(\mathbb{E}\left[\log(1 + \mathrm{SNR}_k)\right]\right)^2 \\
&= \int_0^\infty \log^2(1 + \overline{\mathrm{SNR}}\,t)e^{-t}\mathrm{d}t - m_n^2,
\end{aligned}
$$
(5.23)

where in the first term the integration variable u has been changed into $\overline{\mathrm{SNR}}\,t$, and the second term is easily computed using Eq. (5.19). $\qquad\square$

The variance is more difficult to compute than the mean, since the integrals that are involved are difficult to obtain in closed-form. Nevertheless, the following Proposition gives some insights on the variance, using approximations.

Proposition 5.2. *Tight closed-form approximations of κ^2:*
(Low SNR regime) For $\overline{\mathrm{SNR}} \ll 1$,

$$\kappa^2 \approx \log^2(1 + \overline{\mathrm{SNR}}) - e^{2/\overline{\mathrm{SNR}}}E_1^2(1/\overline{\mathrm{SNR}}).$$
(5.24)

(Large SNR regime) For $\overline{\mathrm{SNR}} \gg 1$,

$$\kappa^2 \approx \frac{\pi^2}{6}\left(Cte_{Euler} - \log(\overline{\mathrm{SNR}})\right)^2 - e^{2/\overline{\mathrm{SNR}}}E_1^2(1/\overline{\mathrm{SNR}}),$$
(5.25)

where $Cte_{Euler} = \lim_{s\to\infty}\left(\sum_{k=1}^{s} 1/k - \log s\right)$ is the Euler-Mascheroni constant.

Proof. From Prop. 5.1 we have:

$$\kappa^2 = \int_0^\infty \log^2(1 + \overline{\mathrm{SNR}}\,t)e^{-t}\mathrm{d}t - e^{2/\overline{\mathrm{SNR}}}E_1^2(1/\overline{\mathrm{SNR}}),$$
(5.26)

so the approximation work is focused on the first term of RHS.

For $\overline{\mathrm{SNR}} \ll 1$ (*i.e.* low average SNR), first order Taylor series expansion gives $\log(1 + \overline{\mathrm{SNR}}\,t) \approx \overline{\mathrm{SNR}}\,t$, so that the function $f(t) = \log^2(1 + \overline{\mathrm{SNR}}\,t) \approx (\overline{\mathrm{SNR}}\,t)^2$ is convex. This

convexity property is used in conjunction with Jensen's inequality, in order to facilitate the integration. Therefore this leads to:

$$
\begin{aligned}
\int_0^\infty \log^2(1 + \overline{\text{SNR}}\,t)e^{-t}\mathrm{d}t &\gtrsim \log^2\left(\int_0^\infty (1 + \overline{\text{SNR}}\,t)e^{-t}\mathrm{d}t\right) \\
&= \log^2\left(1 + \overline{\text{SNR}}\int_0^\infty te^{-t}\mathrm{d}t\right) \\
&= \log^2\left(1 + \overline{\text{SNR}}\right),
\end{aligned}
\tag{5.27}
$$

leading to the first approximation, valid in the low SNR regime. Numerical evaluations will confirm later that this approximation remains close to the true integral value.

Now for $\overline{\text{SNR}} \gg 1$ (*i.e.* large average SNR), $\log(1 + \overline{\text{SNR}}\,t) \approx \log(\overline{\text{SNR}}\,t)$ and from Eq. (5.26) we write:

$$
\begin{aligned}
\int_0^\infty \log^2(1 + \overline{\text{SNR}}\,t)e^{-t}\mathrm{d}t &\approx \int_0^\infty \log^2(\overline{\text{SNR}}\,t)e^{-t}\mathrm{d}t \\
&\overset{(a)}{=} \frac{1}{\overline{\text{SNR}}}\int_0^\infty \log^2(u)e^{-u/\overline{\text{SNR}}}\mathrm{d}u \\
&\overset{(b)}{=} \frac{1}{\overline{\text{SNR}}}\times\overline{\text{SNR}}\left(\frac{\pi^2}{6}\left(\text{Cte}_{\text{Euler}} + \log(1/\overline{\text{SNR}})\right)^2\right),
\end{aligned}
\tag{5.28}
$$

where (a) is obtained after changing $\overline{\text{SNR}}\,t$ into u, and the resulting integral is known from [Gradshteyn and Ryzhik, 1980, Eq. (4.335.1),] giving (b). The final result comes after straightforward simplifications. □

The bounds obtained from Prop. 5.2 are displayed in Fig. 5.1 for $n = 100$. In Fig. 5.2 are compared, for several SNR values, the true distributions of Z_n obtained from Monte-Carlo simulations, with the Gaussian approximation $\mathcal{N}(m_n, \sigma_n^2)$ computed using the Propositions.

5.3.3   Derivations of closed-form expression for the outage probability

Like in [Buckingham and Valenti, 2008], the error probability $P_e^{(n)}$ is computed as the cumulative distribution P_o of the mutual information rate Z_n, for which the distribution has been characterized in Prop. 5.1 and Prop. 5.2:

$$
P_e^{(n)} = \Pr\{Z_n \leq R\} = F_{Z_n}(R),
\tag{5.29}
$$

where F_{Z_n} is the cumulative distribution function of Z_n.

Since the Gaussian random variable $\mathcal{N}(m_n, \sigma_n^2)$ has been introduced as an approximation to Z_n, the distribution of Z_n can be replaced with the Gaussian distribution, leading to:

$$
P_e^{(n)} \approx Q\left(\frac{m_n - R}{\sqrt{\sigma_n^2}}\right),
\tag{5.30}
$$
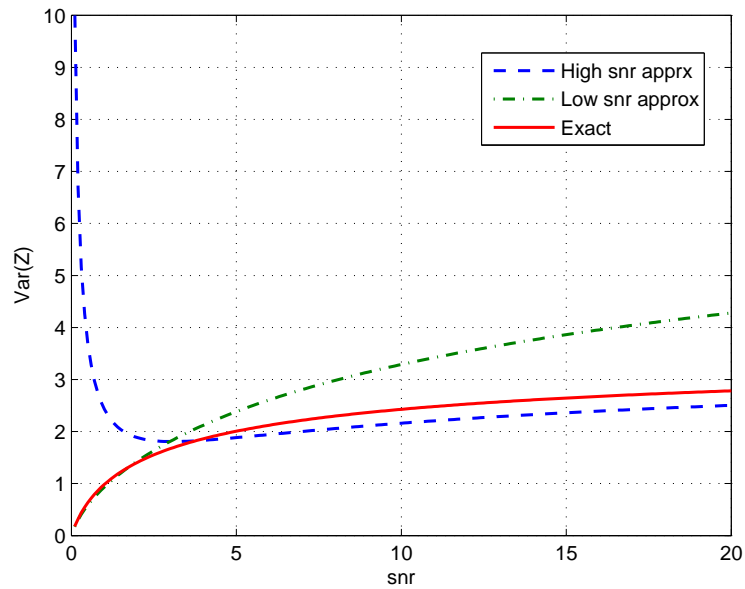
Figure 5.1: Illustration of the bounds for $n = 100$.
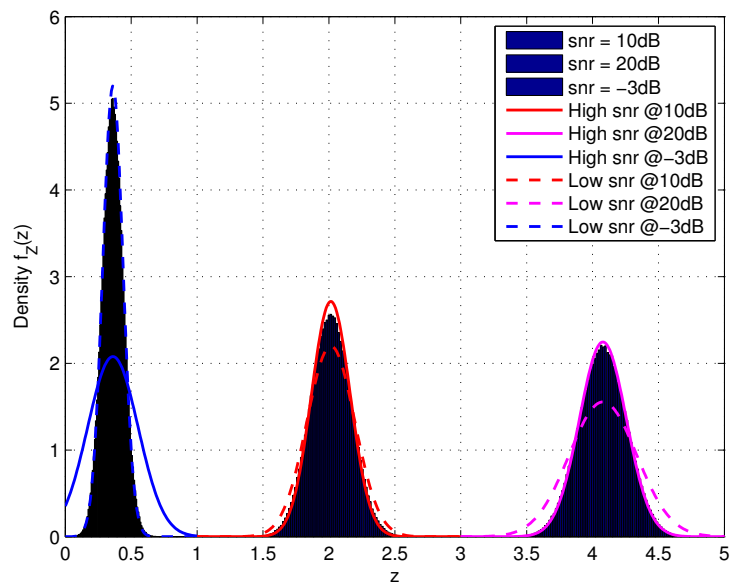
Figure 5.2: Distribution of Z_n for $n = 100$.

where m_n and σ_n^2 are known explicitly from Eq. (5.19), Prop. 5.1 and Prop. 5.2. In Fig. 5.3, the closed-form given in Eq. (5.30) is compared with the true outage probabilities (simulations) for several rates R and blocklength n.
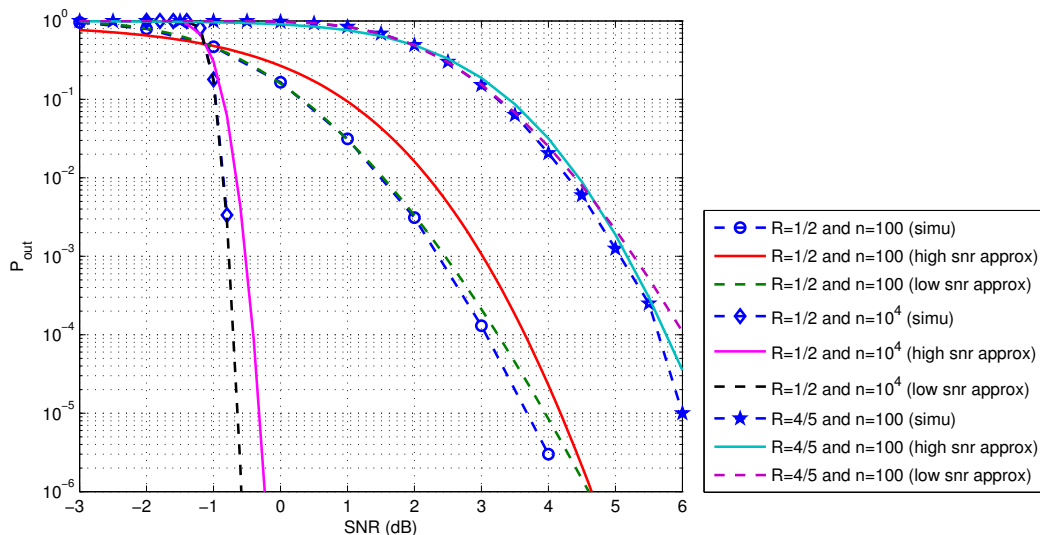
Figure 5.3: Approximate error probability vs Monte-Carlo simulations.

5.4 Resource allocation with finite size codes

In this Section we propose a solution to Problem 4.2, which was defined for Type-I HARQ with finite length Gaussian codes. The optimization problem is first derived using the closed-form expression derived previously, then we present the algorithm from the literature that will be used for the optimal resolution, and finally we show how it is applied to our problem.

5.4.1 Optimization problem

Problem 4.2 has been defined based on the individual goodput:

$$\tilde{\eta}_k(\gamma_k, E_k) = \gamma_k r_k (1 - P_e^{(n)}(G_k E_k, r_k)), \tag{5.31}$$

where $P_e^{(n)}(\overline{\text{SNR}}, R)$ is the error probability, function of $\overline{\text{SNR}}$, of a (n, R) Gaussian code which has been expressed in closed-form in Section 5.3:

$$P_e^{(n)}(\overline{\text{SNR}}, R) = Q\left(\frac{\sqrt{n}(\mu(\overline{\text{SNR}}) - R)}{\sigma(\overline{\text{SNR}})}\right). \tag{5.32}$$

The mean μ and the standard deviation σ are given by (see Eq. (5.19), Prop. 5.1 and Prop. 5.2):

- $\mu(x) := e^{1/x} E_1(1/x),$

- $\sigma(x) := \sqrt{\log^2(1+x) - e^{2/x} E_1^2(1/x) + 2 - (2/x)e^{1/x} E_1(1/x)}.$

It is assumed that the information rate r_k is fixed during the optimization procedure, and the choice of r_k will be discussed in Section 5.5. Let us rewrite Problem 4.2 by composing with the log at both sides of the goodput constraints, as follows:

Problem 5.1.

$$\min_{(\gamma, E)} \sum_{k=1}^{K} \gamma_k E_k, \tag{5.33a}$$

$$s.t. \quad \log \tilde{\eta}_k(\gamma_k, E_k) \geq \log \eta_k^{(0)}, \ \forall k, \tag{5.33b}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{5.33c}$$

$$\gamma_k > 0, \ E_k > 0, \ \forall k. \tag{5.33d}$$

In order to facilitate the optimization procedure, the following conjecture is of great interest (it is more discussed in Appendix E.1):

Lemma 5.1. *For all k, the function $\log(1/\tilde{\eta}_k)$ is biconvex in (γ_k, E_k) [Gorski et al., 2007].*

In the light of Lemma 5.1, Problem 5.1 is the minimization of a biconvex objective function (*i.e.* that is convex in each direction γ and E, but not jointly) over a biconvex set, and thus falls within the class of biconvex optimization problems [Gorski et al., 2007].

The constraints Eq. (5.33d) ensure that the problem will be feasible, *i.e.*, if for instance $\exists k$ such that $\gamma_k = 0$, then $\log \gamma_k$ is not defined, and from $\log \gamma_k \to -\infty$ when $\gamma_k \to 0$, we find that $\log \tilde{\eta}_k < \log \eta_k^{(0)}$. Finally, the next Condition 5.1 provides an equivalence for the problem feasibility. In the rest of the Chapter, it is assumed that Condition 5.1 holds.

Condition 5.1. *Problem 5.1 is feasible if, and only if,*

$$\sum_{k=1}^{K} \frac{\eta_k^{(0)}}{r_k} < 1. \tag{5.34}$$

Sketch of proof. If Problem 5.1 is feasible, then there exists a sequence (γ, E) such that $\forall k, \eta_k^{(0)} \leq \tilde{\eta}_k(\gamma_k, E_k)$ and $\sum_k \gamma_k \leq 1$. This implies that $\eta_k^{(0)} \leq \gamma_k r_k (1 - P_e^{(n)}(G_k E_k)) < \gamma_k r_k$, since $0 < P_e^{(n)}(G_k E_k) < 1$ for $E_k > 0$. So we have

$$\sum_{k=1}^{K} \frac{\eta_k^{(0)}}{r_k} < \sum_{k=1}^{K} \gamma_k \leq 1. \tag{5.35}$$

Conversely, assume that Eq. (5.34) holds. Then, for some sufficiently small $\epsilon > 0$, the problem is feasible by considering $E_k \to \infty$ and $\gamma_k = (\eta_k^{(0)} + \epsilon)/r_k$. $\qquad\square$

5.4.2 Optimal allocation algorithm

The problem of finding globally optimal solutions to the minimization of biconvex functions over a biconvex constraints set has been solved in [Floudas and Visweswaran, 1993]. The optimal solution is reached using a modified primal-dual approach called Global OPtimization (GOP) algorithm. The proposed approach splits the original problem into primal and relaxed dual subproblems which provide valid upper and lower bounds respectively on the global optimum. In a sense that will be detailed, the authors have proven the finite ϵ-convergence of their algorithm towards an ϵ-global optimum.

5.4.2.1 Description of the GOP algorithm

The purpose is to solve the optimization problem:

$$\min_{x,y} f(x,y), \tag{5.36a}$$

$$\text{s.t.} \quad g(x,y) \leq 0, \tag{5.36b}$$

$$x \in \mathcal{X}, \ y \in \mathcal{Y}, \tag{5.36c}$$

where $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^n$ are non-empty, compact, convex sets, and $g(x,y)$ is a p-vector of inequality constraints. Without loss of generality, it can be assumed that $\mathcal{X}$ represents bounds on x and may be incorporated into the constraints g, and so will be dropped from now. The variables y are defined such that the following conditions hold:

Condition 5.2. *The problem defined by Eqs. (5.36) is a biconvex optimization problem, i.e.*

(a) *$f(x,y)$ is convex in x (resp. y) for every fixed value of y (resp. x).*

(b) *$g(x,y)$ is convex in x (resp. y) for every fixed value of y (resp. x).*

Define now the Primal Problem at iteration ℓ:

$$\min_{x} f(x, y^\ell), \tag{5.37a}$$

$$\text{s.t.} \quad g(x, y^\ell) \leq 0, \tag{5.37b}$$

where $y^\ell \in \mathcal{Y}$. Since it is simply the problem described in Eqs. (5.36) solved for fixed values of $y = y^\ell$, the Primal Problem gives an upper bound on the optimal solution to Eqs. (5.36). The Primal Problem is convex in x, therefore its optimal Lagrange multiplier λ^ℓ can be found from convex optimization tools .

Next, applying duality theory and projection concepts leads to the Relaxed Dual Problem at step ℓ:

$$\min_{y \in \mathcal{Y}, \mu_B} \mu_B, \tag{5.38a}$$

$$\text{s.t.} \quad \mu_B \geq \min_{x} \left(f(x,y) + \lambda^{\ell^T} g(x,y) \right), \tag{5.38b}$$

where μ_B is a scalar and λ^ℓ is the optimal multiplier from the ℓ-th Primal Problem. The inner minimization in Eqs. (5.38) involves the Lagrangian $L(x, y, \lambda^\ell)$ formulated at the ℓ-th Primal Problem:

$$L(x, y, \lambda^\ell) = f(x, y) + {\lambda^\ell}^T g(x, y). \tag{5.39}$$

The Relaxed Dual problem defined in Eqs. (5.38) provides a lower bound on the optimal solution to Eqs. (5.36). However it can be very difficult to solve due to the presence of the inner minimization, so the authors define a less complex subproblem.

Let the first order Taylor series expansion of the Lagrange function at x^ℓ be given by:

$$L(x, y, \lambda^\ell)\big|_{x^\ell}^{lin} = L(x^\ell, y, \lambda^\ell) + \left(\nabla_x L(x, y, \lambda^\ell)\big|_{x^\ell}\right)^T (x - x^\ell)$$

$$= L(x^\ell, y, \lambda^\ell) + \sum_{i=1}^n \nabla_{x_i} L(x, y, \lambda^\ell)\big|_{x^\ell} (x_i - x_i^\ell). \tag{5.40}$$

Furthermore, for $i \in \{1, \dots, n\}$ let x_i^L and x_i^U be the lower and upper bounds on x_i, respectively, and let B_j be a combination of these bounds. Let us denote by x^{B_j} the vector of lower/upper bounds corresponding to the combination B_j, and let CB be the set of all bound combinations. For instance, for $n = 2$, if $x_i^L = 0$ and $x_i^U = 1$ ($i = 1, 2$) then there are 4 combinations $B_j \in CB$ for $j = 1, 2, 3, 4$:

$$\begin{aligned} x^{B_1} &= [0 \quad 0]^T, \quad x^{B_3} = [1 \quad 0]^T, \\ x^{B_2} &= [0 \quad 1]^T, \quad x^{B_4} = [1 \quad 1]^T. \end{aligned} \tag{5.41}$$

For a given combination B_j, the sign of $\nabla_{x_i} L(x, y, \lambda^{\ell'})\big|_{x^{\ell'}}$ when evaluated at $y = y^\ell$, is said to be a qualification constraint of the Lagrange function $L(x, y, \lambda^{\ell'})$ of iteration $\ell' < \ell$ for iteration ℓ:

$$\nabla_{x_i} L(x, y, \lambda^{\ell'})\big|_{x^{\ell'}} \leq 0 \qquad \qquad \text{if } x_i^{B_j} = x_i^U, \text{ and} \tag{5.42a}$$

$$\nabla_{x_i} L(x, y, \lambda^{\ell'})\big|_{x^{\ell'}} \geq 0 \qquad \qquad \text{if } x_i^{B_j} = x_i^L. \tag{5.42b}$$

For all $\ell' < \ell$, let $\mathcal{F}(\ell', \ell)$ be the Lagrange function from iteration ℓ' whose qualification constraint is satisfied at y^ℓ, and let x^{B_j} be the corresponding combination of bounds of the x variables for this Lagrange function.

Finally, the subproblem at iteration ℓ for combination $B_\ell \in CB$ can be defined:

$$\min_{y \in \mathcal{Y}, \mu_B} \mu_B, \tag{5.43a}$$

$$\text{s.t.} \quad \left.\begin{aligned} &\mu_B \geq L(x^{B_j}, y, \lambda^{\ell'})\big|_{x^{\ell'}}^{lin} \\ &\nabla_{x_i} L(x, y, \lambda^{\ell'})\big|_{x^{\ell'}} \leq 0 \quad \text{if } x_i^{B_j} = x_i^U \\ &\nabla_{x_i} L(x, y, \lambda^{\ell'})\big|_{x^{\ell'}} \geq 0 \quad \text{if } x_i^{B_j} = x_i^L \end{aligned}\right\} \forall j \in \mathcal{F}(\ell', \ell) \text{ and } \ell' \in \{1, \dots, \ell\} \tag{5.43b}$$

$$\begin{aligned} &\mu_B \geq L(x^{B_\ell}, y, \lambda^\ell)\big|_{x^\ell}^{lin} \\ &\nabla_{x_i} L(x, y, \lambda^\ell)\big|_{x^\ell} \leq 0 \quad \text{if } x_i^{B_\ell} = x_i^U \\ &\nabla_{x_i} L(x, y, \lambda^\ell)\big|_{x^\ell} \geq 0 \quad \text{if } x_i^{B_\ell} = x_i^L \end{aligned} \tag{5.43c}$$

In [Floudas and Visweswaran, 1993], it is shown that the optimal value of the ℓ-th Relaxed Dual Problem is lower bounded by the minimal value of all the subproblems for all $B_\ell \in CB$ and all $\ell' \in \{1, \ldots, \ell - 1\}$. Solving Eqs. (5.43) in an iterative fashion leads to successive increasing lower bounds on the optimal value of Eqs. (5.38). The GOP algorithm solves alternately the Primal Problem in Eqs. (5.37) and the subproblem in Eqs. (5.43), $\forall B_\ell \in CB$ and $\ell' \in \{1, \ldots, \ell - 1\}$, within an iteration ℓ, until the upper and lower bounds meet or the maximum number of iterations ℓ_M is attained.

Finally, in their paper the authors established the two key theorems:

Theorem 5.2 (Finite ϵ-convergence [Floudas and Visweswaran, 1993]). *If the original problem is biconvex, then, under mild conditions, the GOP algorithm terminates in a finite number of steps for any given gap $\epsilon > 0$ on the bounds.*

Theorem 5.3 (Global Optimality [Floudas and Visweswaran, 1993]). *Under the same conditions, the GOP algorithm will terminate at the global optimum of the original problem.*

5.4.2.2   Application of the GOP to Problem 5.1

Following the GOP approach for Problem 5.1, let us define the Primal Problem at step ℓ:

$$\min_{\gamma} \sum_{k=1}^{K} \gamma_k E_k^\ell, \tag{5.44a}$$

$$\text{s.t.} \quad \log \eta_k^{(0)} - \log \tilde{\eta}_k(\gamma_k, E_k^\ell) \le 0, \ \forall k, \tag{5.44b}$$

$$\sum_{k=1}^{K} \gamma_k \le 1. \tag{5.44c}$$

The Relaxed Dual Problem at step ℓ is:

$$\min_{E>0, \mu_B} \mu_B, \tag{5.45a}$$

$$\text{s.t.} \quad \mu_B \ge \min_{\gamma} L(\gamma, E, \lambda^\ell), \tag{5.45b}$$

where μ_B is a scalar, λ^ℓ are the optimal multipliers of Eqs. (5.44), and the Lagrangian is defined by:

$$L(\gamma, E, \lambda) = \sum_{k=1}^{K} \gamma_k E_k - \sum_{k=1}^{K} \lambda_k \Big(\log r_k + \log \gamma_k + \log(1 - P_e^{(n)}(G_k E_k, r_k)) - \log \eta_k^{(0)} \Big)$$

$$+ \lambda_{K+1} \left(\sum_{k=1}^{K} \gamma_k - 1 \right). \tag{5.46}$$

Eqs. (5.45) are solved using the GOP, thus the following gradient is needed to define the subproblems as in Eqs. (5.43):

$$\forall k \in \{1, \ldots, K\}, \ \nabla_{\gamma_k} L(\gamma, E, \lambda) = E_k - \frac{\lambda_k}{\gamma_k} + \lambda_{K+1}. \tag{5.47}$$

For a fixed precision $\epsilon > 0$, the optimal allocation algorithm is thus summarized in Algorithm 5.1, and the next result holds.

Algorithm 5.1: GOP algorithm for Problem 5.1

---

(1) **Initialization:**

Set the maximum number of iterations to ℓ_M, set $UB = \infty$, $LB = -\infty$.

Set $\mu_B^{stor} = \emptyset$, $\mathcal{F}(\ell', \ell) = \emptyset$, for all $\ell \leq \ell_M$ and $\ell' \leq \ell$.

Initialize CB, $\ell = 1$ and E^1.

(2) **Primal Problem:**

Solve Eqs. (5.44), store the solution into $Q^\ell(E^\ell)$, store the optimal Lagrange multipliers (λ^ℓ) and set $UB = \min(UB, Q^\ell(E^\ell))$.

(3) **Selection of Lagrange functions:**

For $\ell' = 1$ to $(\ell - 1)$, evaluate the sign $\forall k \in \{1, \dots, K\}$ of $\nabla_{\gamma_k} L(\gamma, E^\ell, \lambda^\ell)\big|_{\gamma^\ell}$ for all $B_j \in CB$. Select the one that satisfies the qualification constraints Eq. (5.42) to be in $\mathcal{F}(\ell', \ell)$.

(4) **Relaxed Dual Problem:**

Solve Eqs. (5.43), $\forall B_\ell \in CB$ and $\ell' \in \{1, \dots, \ell - 1\}$, and store the solutions into μ_B^{stor}. Select the minimal value LB from μ_B^{stor} and the associated E^{min}, delete it from μ_B^{stor}, and set $E^{\ell+1} = E^{min}$.

(5) **Check for convergence:**

If $UB - LB > \epsilon$, increment ℓ by 1 and return to Step (2).

Else stop.

Theorem 5.4. *For any $\epsilon > 0$, the allocation algorithm defined in Algorithm 5.1 terminates at an ϵ-global optimum solution to Problem 4.2.*

Proof. By Theorems 5.2 and 5.3 that hold true for the GOP when applied to the equivalent Problem 5.1. □

5.5 Numerical results

In this Section we give some numerical results that illustrate the framework developed in this Chapter. We begin with describing the overall simulation background. Next, GOP results are given for fixed coding rate values when the goodput demand varies. The choice of the coding rate within the allocation procedure is then discussed. Finally, we see that the outage probability computed in the Chapter can tightly approximate the performance of good FEC codes over the ergodic Rayleigh channel.

5.5.1   Simulation settings

Due to the GOP complexity, only $K = 2$ communication links are considered with average SNR configured to 10 dB and 30 dB, respectively. The transmitters use Gaussian codes of given length n and given rates r_k. The coding rates choice will be explained later. For the sake of simplicity the rate request is uniform:

$$\eta_k^{(0)} = \eta_T/K, \tag{5.48}$$

with η_T the total goodput demand of the cluster (in bit/s/Hz). The total spectral efficiency requirement η_T is related to the sum-rate of the cluster ρ (in bit/s) using the bandwidth W (in Hz):

$$\rho = \eta_T W. \tag{5.49}$$

5.5.2   GOP results versus increasing sum-rate demand

In this case, the rate r_k is arbitrarily fixed to a value that satisfies Condition 5.1. If $\eta_T < 1/2$ then $r_k = 1/2$, hence $\sum_{k=1}^{K} \eta_k^{(0)}/r_k = K(\eta_T/K)/(1/2) < 1$. Else if $1/2 \leq \eta_T < 1$ then $r_k = 1$, hence $\sum_{k=1}^{K} \eta_k^{(0)}/r_k = \eta_T/1 < 1$, and so on.

   In Fig. 5.4, we plot the power consumption resulting from Algorithm 5.1 versus the spectral efficiency request η_T. The ergodic capacity based algorithm from [Gault et al., 2007] is displayed as a benchmark. Since $\eta_T \in [0, 1]$ bit/s/Hz (corresponding to a sum-rate $\rho \in [0, 1]$ Mbit/s in $W = 1$ MHz), we have fixed $r_k = 1/2$ for $\eta_T \in [0, 0.5)$ bit/s/Hz and $r_k = 1$ for $\eta_T \in [0.5, 1]$ bit/s/Hz, for all k. Very surprisingly, our optimized Type-I HARQ scheme with finite-length Gaussian code outperforms the capacity-achieving scheme of [Gault et al., 2007] for low spectral efficiency requests (basically, between 0.05 bit/s/Hz and 0.5 bit/s/Hz for $n = 512$ and up to 0.65 bit/s/Hz for $n = 10^4$ and $n = 10^6$).

   This can be explained by the total bandwidth that is allocated to the cluster, as shown in Fig. 5.5 where $\sum_{k=1}^{K} \gamma_k$ is plotted (in %) versus the total spectral efficiency request. We see that the ergodic capacity dictates the allocation of the entire bandwidth for all $\eta_T \in [0, 1]$. In contrast, using the goodput metric in Algorithm 5.1 indicates that the optimum can be reached for $\sum_{k=1}^{K} \gamma_k < 1$ when $\eta_T < 1$. Actually, it decreases the power consumption of the OFDM symbol, for small values of $\eta_k^{(0)}$.

   Elsewhere, the capacity-based scheme of [Gault et al., 2007] has the best performance, as seen from Fig. 5.6. We observe, as expected, that the power consumption of Gaussian coding with finite length decreases for increasing block length n. The performance gain with increasing n is nonetheless limited (see the thin difference between $n = 10^4$ and $n = 10^6$). The remaining gap between our Gaussian scheme and the Shannon limit when $n \gg 1$ is explained by the fact that the information rate r_k is fixed in our case, whereas the ergodic capacity chooses the best information rate. Thus, fixing r_k smartly is of great interest to get closer to the Shannon capacity, and is discussed in the next paragraph.
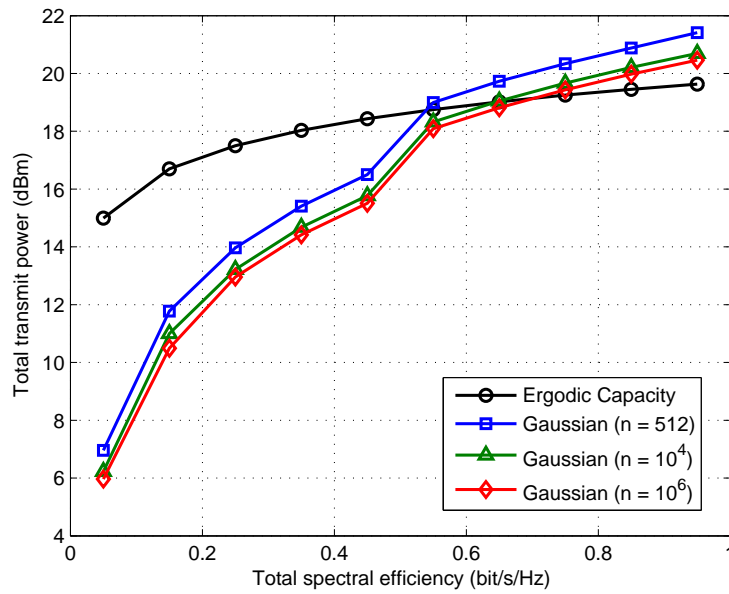
Figure 5.4: Total power from Algorithm 5.1 for several *n* values compared with Ergodic Capacity based algorithm.
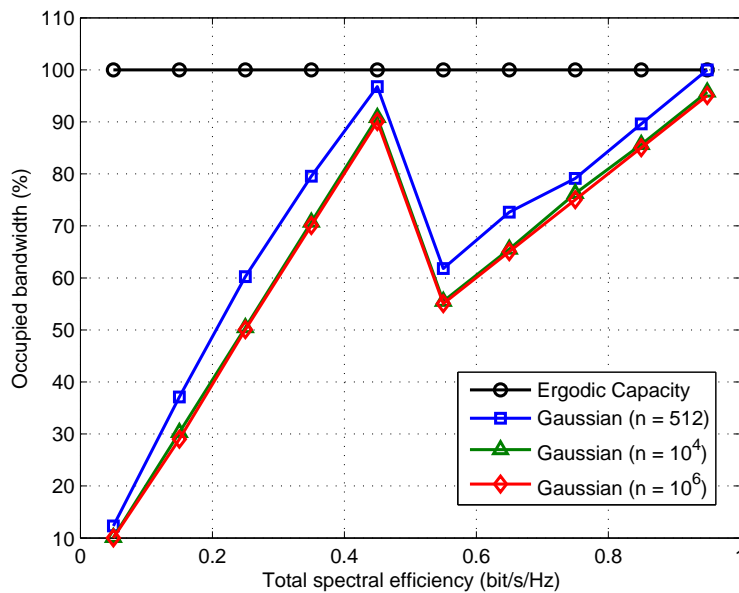

Figure 5.5: Occupied bandwidth from Algorithm 5.1 for several *n* values compared with Ergodic Capacity based algorithm.
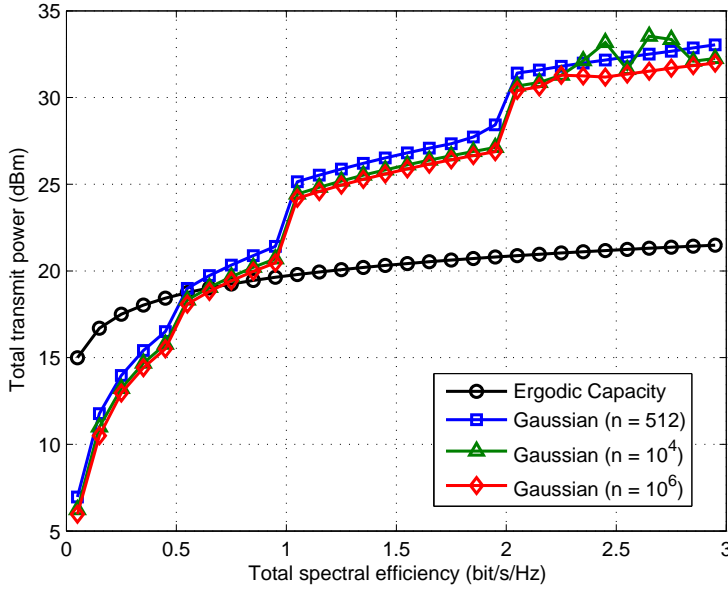
Figure 5.6: Total power from Algorithm 5.1 for several n values compared with Ergodic Capacity based algorithm.

5.5.3 How to choose the information rate r_k?

In a first time, the goodput request is fixed to η_T and we let the information rate r_k vary (inside the feasible interval). In Fig. 5.7 we investigate the power consumption behavior versus r_k in order to find some insights about the way of choosing r_k. Unfortunately, Fig. 5.7 reveals no useful information: the total transmit power Q_T is either nonincreasing or nondecreasing according to the η_T values. In contrast, Fig. 5.8 shows that the occupied bandwidth decreases for increasing values of r_k, for both $\eta_T = 0.5$ bit/s/Hz and $\eta_T = 2$ bit/s/Hz.

From this, we have found no practical way of fixing $r_k \in \mathbb{R}_+$. However, constraining r_k to be chosen from a finite set $\mathcal{R} \subset \mathbb{R}_+$ can improve the performance using a good rate selection algorithm. Fig. 5.9 plots the total transmit power versus η_T for two cases:

(a) The rates r_k are chosen as in Section 5.5.2 among the values 0.5, 1, 2 or 3.

(b) A rate selection algorithm is run over $\mathcal{R} = \{0.5, 1, 1.5, 2, 2.5, 3\}$.

Optimal choice of $r_k \in \mathcal{R}$ thus improve the performance up to 4 dB. Refinement of $\mathcal{R}$ can help to get closer to the Shannon limit. For the sake of simplicity, we used an exhaustive research for the rates selection algorithm. More practical approaches will be discussed in details in Chapter 6.
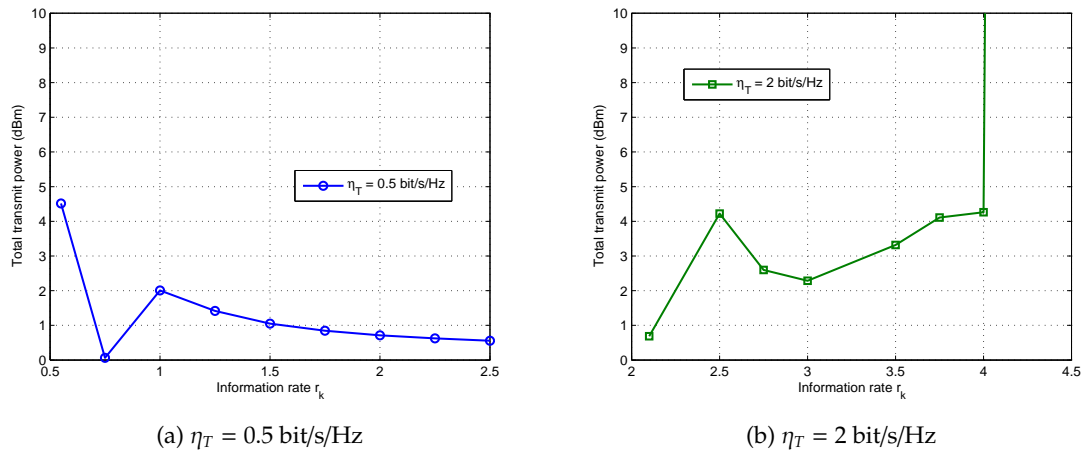
(a) $\eta_T = 0.5$ bit/s/Hz

(b) $\eta_T = 2$ bit/s/Hz

Figure 5.7: Total transmit power of Algorithm 5.1 versus information rate r_k ($n = 512$).
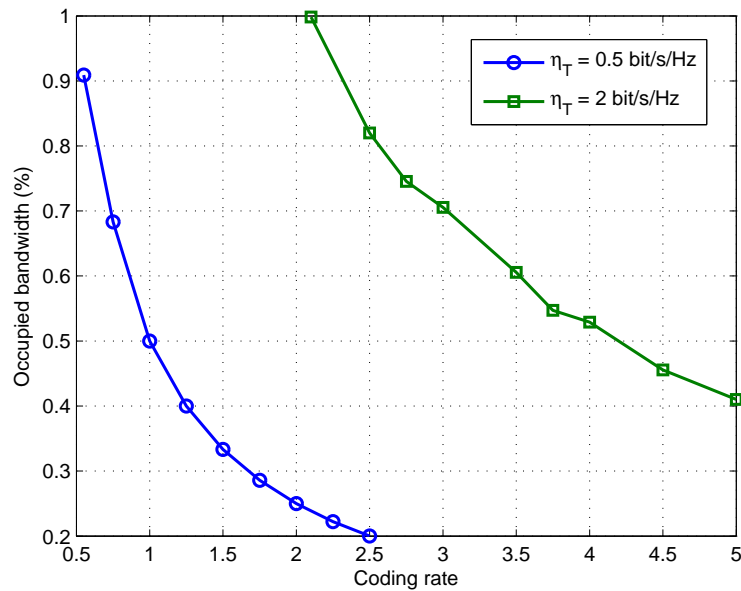

Figure 5.8: Occupied bandwidth of Algorithm 5.1 versus Coding rate ($n = 512$).
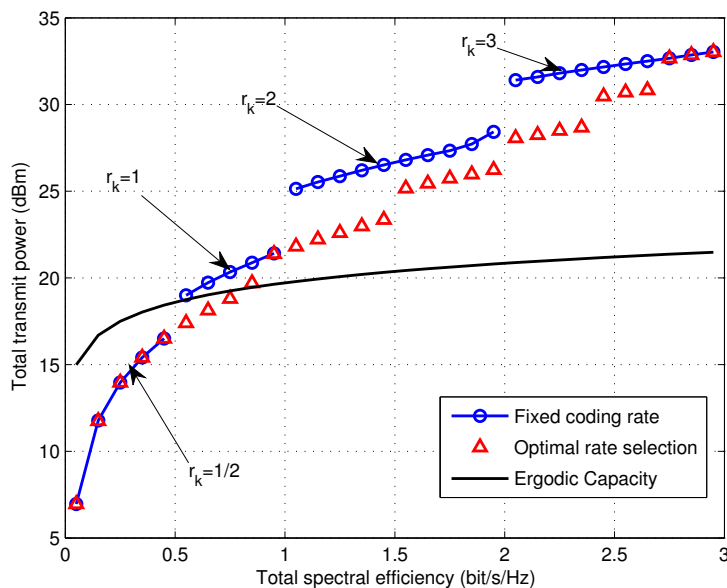
Figure 5.9: Total transmit power versus spectral efficiency request for different rate selections ($n = 512$).

5.5.4   How close are powerful FEC codes?

Finally, we show that the outage probability developed within this framework can be used to predict the performance of powerful FEC codes over the Rayleigh channel[1]. The benefits of the LDPC (and more generally the turbo-like) codes come from the relaxation of the Maximum Likelihood (ML) assumption, and the LLR messages are well modeled by Gaussian random variables under iterative decoding [Ryan and Lin, 2009]. This observation motivates us to describe the waterfall in the LDPC error performance, which is the capacity-achieving region of these codes, using finite length Gaussian codes.

In Fig. 5.10 are plotted the PER figures of two BPSK modulated LDPC codes [2] [Gallager, 1963] versus SNR for $n = 504$ and $R = 1/2$. The error probability of a Gaussian code of length $n = 504$ and $R = 1/2$ is plotted too, as well as its shifted versions using some SNR gaps. It is very interesting to observe a tight approximation of the LDPC performance by shifting the finite-length Gaussian code error probability with a gap, whereas it was impossible to directly resort to gaps on the ergodic capacity function (see Chapter 4).

---

[1]The performance of finite-length Low Density Parity Check (LDPC) and Turbo-like codes over the binary erasure channel were already studied in [Amraoui, 2006], [Andriyanova and Urbanke, 2009], and [Andriyanova, 2009] using the scaling approach. The waterfall of finite-length LDPC codes was already modeled by the $Q(.)$ function in these works. Unfortunately, the formalism is outstandingly hard for any use in our context.

[2]We used a $(3,6)$-regular code and an irregular code generated by Progressive Edge Growth (MacKay's parity-check matrices found on http://www.inference.phy.cam.ac.uk/mackay/codes).

Approximation is very tight for PER between 10^{-3} and 1. The difference is noticeable beyond 10^{-3}, and may be explained by the floor behavior of LDPC, which is not present for Gaussian coding. This is satisfying since it was remarked in [Wu and Jindal, 2011] that the operating point of FEC when optimizing the goodput is generally high (PER of about 10^{-1}).
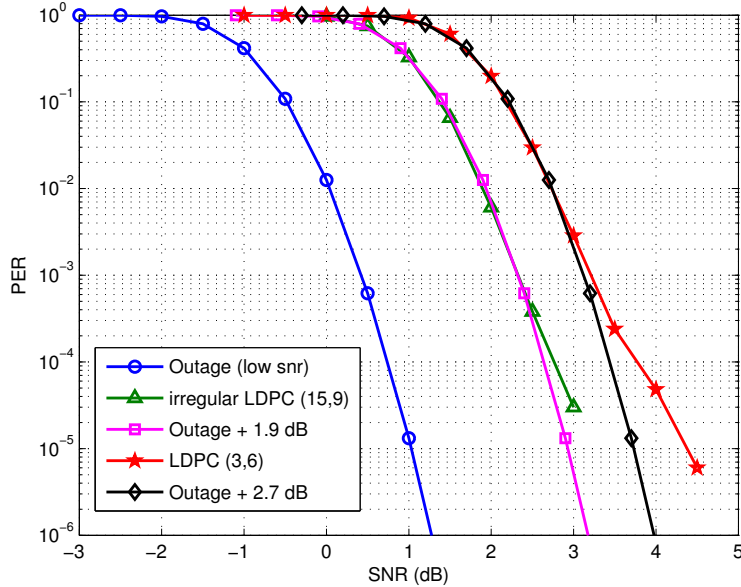
Figure 5.10: PER versus SNR of several FEC codes ($n = 504$ and $R = 1/2$).

However the performance for less powerful FEC code families do not match well with the performance given by the finite length Gaussian codes, as seen in Fig. 5.11 for the 1/2-rate convolutional code of length $n = 512$ with generators $(23, 35)_8$.

5.6   Conclusion

In this Chapter, we have first computed in closed-form the error probability of Gaussian codes with finite length over the Rayleigh channel. This has been done using the information spectrum framework. Based on this new result, we were able to find the optimal power and bandwidth allocation, thus answering to Problem 4.2.

Next, we have illustrated the results of this algorithm, which gives the best performance that one can expect from our clustered OFDMA network using Type-I HARQ. Surprisingly, the numerical results revealed that this scheme can outperform the ergodic capacity limit when the data rate request are low (basically, below 500 kbit/s within a bandwidth $W = 1$ MHz). This is explained by the ability of our algorithm to save the bandwidth: moreover, this ability can be very interesting for frequency reuse purposes.
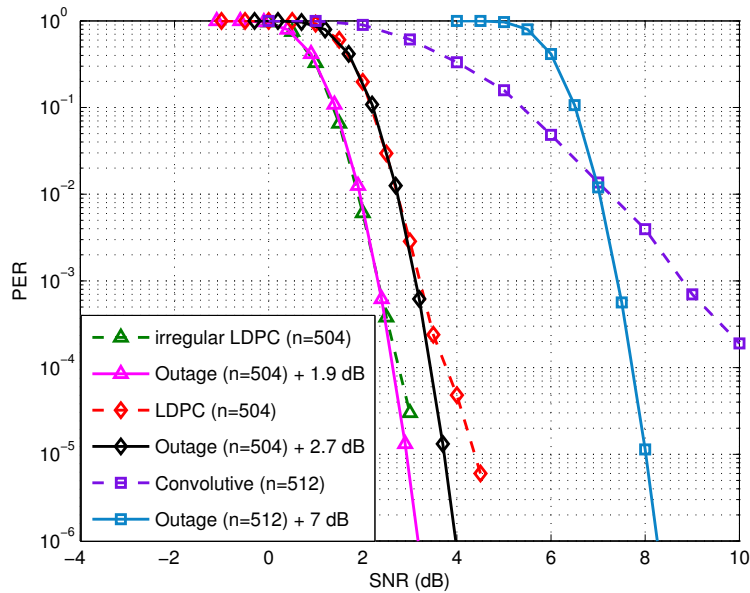
Figure 5.11: PER versus SNR of several FEC codes ($R = 1/2$).

Finally, this framework is well adapted to predict the performance of good FEC codes (LDPC in this Chapter) over the Rayleigh channel. For instance, we know that the power consumption after the resource allocation of an irregular LDPC code of length $n = 504$ and $R = 1/2$ will be only 1.9 dB away from the best achievable performance. Thus, this framework can serve as a basis for Type-I HARQ based OFDMAresource allocation when powerful FEC is used (typically LDPC coding).

However, the GOP steps are cost-computing (the number of constraints increases linearly with the number of iterations) and convergence can be very slow. Though optimal, Algorithm 5.1 is thus not adapted to real applications. Furthermore, it cannot predict the performance of other code families that are not capacity-achieving (in particular the convolutional coding). Therefore we must develop another framework, which is the point of Chapter 6.

Chapter 6

Resource allocation for HARQ with practical MCS

6.1 Introduction

The framework developed in the previous Chapter is useful to perform the resource allocation of schemes that use powerful channel coding like LDPC. However, this framework fails to predict the performance of the other practical schemes (like convolutional coding or other standard algebraic codes that are not capacity achieving). Furthermore the proposed algorithm becomes very complex even for small K and cannot be implemented in real systems.

Therefore, in this Chapter we develop another framework that is best suited to any practical MCS. The work is organized as follows. The model for practical MCS is described in Section 6.2. Then, in Section 6.3 we solve Problem 4.3 by providing an optimal solution to power and bandwidth assignment when the MCS is given. An efficient MCS selection algorithm is proposed in Section 6.3.6. Next, Problem 4.4 is solved in Section 6.4, where optimal as well as practical solutions are proposed. Finally, Section 6.5 is devoted to the resolution of Problem 4.5 for which new efficient suboptimal solutions are developed.

6.2 Practical MCS

This Section explains how to deal with practical MCS, *e.g.* QAM constellations and existing FEC codes like convolutional codes, when they are used for the resource allocation.

The optimization results presented in the subsequent Sections are valid for any PER function satisfying some technical conditions (convex and derivable), and thus are applicable to most of the existing MCSs. It is thus necessary to have an analytic expression of the PER that is derivable as a function of the SNR. For this, we propose to fit the PER versus SNR curves obtained by simulation of the MCSs, with mathematical models. We indicate in the following two different models that can be applied to perform the

curve-fitting in the case of coded packets over Rayleigh channels.

Type-I HARQ for which a single information packet can be sent at most L times is used. It is assumed here, without loss of generality, that all the users are limited by L. The information packet can be either:

(i) uncoded, letting the users k choosing the size 2^{m_k} of their own constellation inside a discrete set $\mathcal{M} \subset \mathbb{N}_*$,

(ii) or encoded using an existing FEC code of rate $R_k \in \mathcal{R}$, with $\mathcal{R} \subset \mathbb{Q}_+$, before modulation using a constellation of size 2^{m_k}.

We recall from Chapter 4 that the Type-I HARQ goodput is:

$$\tilde{\eta}_k(\text{SNR}) = m_k R_k \left(1 - P_k(\text{SNR})\right), \tag{6.1}$$

with $P_k(\text{SNR})$ the PER of the FEC code of given rate R_k sent over the constellation of order 2^{m_k}. A well-designed FH pattern has been assumed in Chapter 4 in order to recover entirely the diversity offered by the channel, *i.e.* at least M, leading to a fast-fading channel model.

In the uncoded case (i), assuming fast-fading channel and that the information packets of n bits are modulated with a 2^{m_k}-QAM, the approximate bit error probability can be found in [Chung and Goldsmith, 2001]. Then, the PER can be written with respect to SNR as (large SNR approximation):

$$P_k(\text{SNR}) \approx \frac{n\, a_{m_k}}{1 + \frac{g_{m_k}}{2^{m_k}-1}} \frac{1}{\text{SNR}}, \tag{6.2}$$

where a_{m_k} and g_{m_k} are two constants depending on the selected constellation m_k and designed to fit the simulated PER curve.

In the coded case (ii), it is known that the diversity in the fast-fading channel is limited by the minimum Hamming distance of the block code if Bit Interleaved Coded Modulation (BICM) is used along with ML decoding [Caire et al., 1998]. Thus, if BICM is assumed in order to retrieve the entire diversity offered by the code, the PER can be written with respect to SNR as:

$$P_k(\text{SNR}) \approx \frac{g_c(m_k, R_k)}{\text{SNR}^{d_f(R_k)}}, \tag{6.3}$$

where $d_f(R_k)$ is the minimal (Hamming or free) distance of the code of rate R_k, and $g_c(m_k, R_k)$ is a coding gain designed to fit the simulated PER curve.

6.3   Rate constrained power minimization

In this Section we solve Problem 4.3 using the convex optimization framework. We begin with evidencing the convexity of Problem 4.3, next optimal solution is obtained in closed-form. The case of imperfect HARQ feedback is also discussed.

6.3.1 Optimization problem formulation

Let Q_k be the average energy consumed to send the part of the OFDM symbol associated with link k. It becomes the new power parameter to be optimized. It can be easily shown that:

$$Q_k = \frac{n_k P_k}{W} = \gamma_k E_k. \tag{6.4}$$

Therefore, the nonconvex objective function $\sum_{k=1}^{K} \gamma_k E_k$ in Problem 4.3 becomes the new objective function $\sum_{k=1}^{K} Q_k$. It is obvious that this function is convex.

Let us define the function $f : x \mapsto 1 - x$ over $[0,1]$, and let us denote the goodput function:

$$\tilde{\eta}_k(\gamma_k, Q_k) = \gamma_k m_k R_k f(P_k(G_k Q_k / \gamma_k)). \tag{6.5}$$

The optimal solution $(\boldsymbol{\gamma}^*, \boldsymbol{Q}^*)$ is such that $\gamma_k^* > 0$ and $Q_k^* > 0$ for all $k \in \{1, \ldots, K\}$. Indeed, if $\exists k$ such that $\gamma_k = 0$, then this user would have no chance to satisfy his rate constraint since $\tilde{\eta}_k(0, Q_k) = 0 < \eta_k^{(0)}$ for any Q_k value. Similarly if $Q_k = 0$, then $P_k(0) = 1$ and thus $\tilde{\eta}_k(\gamma_k, 0) = 0 < \eta_k^{(0)}$.

Therefore, Problem 4.3 boils down to the next Problem 6.1.

Problem 6.1.

$$\min_{(\gamma, Q)} \sum_{k=1}^{K} Q_k, \tag{6.6a}$$

$$s.t. \quad \tilde{\eta}(\gamma_k, Q_k) \geq \eta_k^{(0)}, \ \forall k, \tag{6.6b}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{6.6c}$$

$$\gamma_k > 0, \ Q_k > 0, \ \forall k. \tag{6.6d}$$

6.3.2 Feasibility and convexity properties

Now, let us study the feasibility of Problem 6.1. The next result (proved in Appendix F.1) provides an easy way to check feasibility condition, as an inequality involving m_k, R_k and $\eta_k^{(0)}$ only.

Lemma 6.1. *Problem 6.1 is feasible if, and only if,*

$$\sum_{k=1}^{K} \frac{\eta_k^{(0)}}{m_k R_k} < 1. \tag{6.7}$$

In the rest of the Chapter, we assume that Lemma 6.1 holds. In the next Lemma 6.2, we show that Problem 6.1 is convex as soon as the function $\text{SNR} \mapsto P_k(\text{SNR})$ is convex. Notice that P_k defined in Eqs. (6.2)-(6.3) satisfies the convexity property.

Lemma 6.2. *The constraint function $\tilde{\eta}_k$ defined on $(0,1) \times \mathbb{R}_+^*$ in Eq. (6.5) is concave as long as $P_k : \mathbb{R}_+ \longrightarrow [0,1]$ is a convex function.*

Proof. Assuming the univariate function P_k convex, then $\tilde{\eta}_k$ is concave as the perspective of the concave function $Q_k \mapsto m_k R_k(1 - P_k(G_k Q_k))$ [Boyd and Vandenberghe, 2004]. $\qquad \square$

Thus, Problem 6.1 will be solved in the next Section within the convex optimization framework.

6.3.3 Optimal algorithm with fixed MCS

Assuming that a strictly feasible solution to Problem 6.1 exists[1], the Karush-Kuhn-Tucker (KKT) conditions enable one to exhibit the optimal solution (γ^*, Q^*) to Problem 6.1 since it is convex [Boyd and Vandenberghe, 2004].

After some tedious algebraic manipulations reported in Appendix F.2, we obtain the following result for the optimal power and bandwidth allocation.

Theorem 6.3. *Let us define $F : x \mapsto -(1 - P_k(x))^2 / (f(P_k(x)) P_k'(x)) - x$. The optimal allocation policy (γ^*, Q^*) is:*
If

$$\sum_{k=1}^{K} \frac{\eta_k^{(0)}}{m_k R_k f(P_k(F^{-1}(0)))} \leq 1,$$

then

$$\gamma_k^* = \frac{\eta_k^{(0)}}{m_k R_k f(P_k(F^{-1}(0)))}, \tag{6.8}$$

$$Q_k^* = \frac{\gamma_k^*}{G_k} F^{-1}(0). \tag{6.9}$$

Else:

$$\gamma_k^*(\lambda^*) = \frac{\eta_k^{(0)}}{m_k R_k f(P_k(F^{-1}(\lambda^* G_k)))}, \tag{6.10}$$

$$Q_k^*(\lambda^*) = \frac{\gamma_k^*(\lambda^*)}{G_k} F^{-1}(\lambda^* G_k), \tag{6.11}$$

with $\lambda^* > 0$ chosen such that

$$\sum_{k=1}^{K} \gamma_k^*(\lambda^*) = 1. \tag{6.12}$$

Theorem 6.3 provides a very fast algorithm for the resource allocation: basically, a linear research must be performed in order to find the only scalar value λ^*. It is summarized in Algorithm 6.1. A fast implementation of Algorithm 6.1 is given in Appendix F.3 for the uncoded packets case.

---

[1]This is guaranteed by Lemma 6.1.

Algorithm 6.1: Optimal algorithm for Problem 6.1.

Set $\lambda = 0$ and compute, $\forall k$, $\gamma_k(\lambda)$ from Eq. (6.10) and $Q_k(\lambda)$ from Eq. (6.11).

if $\sum_{k=1}^{K} \gamma_k(\lambda) \leq 1$ **then**

    $\gamma^* = \gamma(\lambda)$ and $Q^* = Q(\lambda)$,

    Exit.

else

    while $\sum_k \gamma_k(\lambda) > 1$ **do**

        Increase λ

        Compute $\gamma_k(\lambda)$ from Eq. (6.10) and $Q_k(\lambda)$ from Eq. (6.11), for all k.

    end

    $\gamma^* = \gamma(\lambda)$ and $Q^* = Q(\lambda)$,

6.3.4 The case of imperfect feedback

We assume in this Section that the HARQ feedback is degraded, as in Chapter 3. As the power dedicated to the direct link does not influence the SNR of the reverse channel devoted to the acknowledgment, we assume erroneous feedback with constant probability p_{fb}. For the sake of simplicity, we have assumed infinite retransmissions ($L = \infty$). In this case, we recall that the Type-I HARQ goodput is given by:

$$\tilde{\eta}_k = \gamma_k m_k R_k f_{\text{fb}}(P_k), \tag{6.13}$$

with $f_{\text{fb}} : x \mapsto 1/(1/(1 - x) + p_{\text{fb}}/(1 - p_{\text{fb}}))$. For $L < \infty$, Eq. (6.13) becomes intractable. The following results hold.

Lemma 6.4. *The constraint function defined by:*

$$\tilde{\eta}_k(\gamma_k, Q_k) = \gamma_k m_k R_k f_{\text{fb}}(P_k(G_k Q_k/\gamma_k)) \tag{6.14}$$

on $[0, 1] \times \mathbb{R}_+$, is concave as long as $P_k : \mathbb{R}_+ \longrightarrow [0, 1]$ is a convex function.

Proof. It is easy to show that the function $f_{\text{fb}}(P_k(x))$ is concave in $x \in \mathbb{R}^+$:

$$f_{\text{fb}}''(x) = \frac{-P_k''(x)}{\left(1 + \frac{p_{\text{fb}}}{1 - p_{\text{fb}}}(1 - P_k(x))\right)^2} - \frac{2(P_k'(x))^2 p_{\text{fb}}/(1 - p_{\text{fb}})}{\left(1 + \frac{p_{\text{fb}}}{1 - p_{\text{fb}}}(1 - P_k(x))\right)^3} \leq 0. \tag{6.15}$$

Hence, the perspective function $\tilde{\eta}_k(\gamma_k, Q_k) = m_k R_k \gamma_k f_{\text{fb}}(P_k(G_k Q_k/\gamma_k))$ is concave [Boyd and Vandenberghe, 2004]. $\qquad\square$

Lemma 6.5. *Problem 6.1 (when f is replaced with f_{fb}) is feasible if, and only if,*

$$\sum_{k=1}^{K} \left(1 + \frac{p_{\text{fb}}}{1 - p_{\text{fb}}}\right) \frac{\eta_k^{(0)}}{m_k R_k} < 1. \tag{6.16}$$

Sketch of proof. If Problem 6.1 is feasible, then there exists a sequence (γ_k, Q_k) such that $\forall k$, $\eta_k^{(0)} \le \eta_k(\gamma_k, Q_k)$ and $\sum_k \gamma_k \le 1$. This implies that $\eta_k^{(0)} \le \gamma_k m_k R_k f_{\text{fb}}(P_k(G_k Q_k/\gamma_k)) < \frac{\gamma_k m_k R_k}{1 + \frac{p_{\text{fb}}}{1 - p_{\text{fb}}}}$. So we have:

$$1 \ge \sum_k \gamma_k > \left(1 + \frac{p_{\text{fb}}}{1 - p_{\text{fb}}}\right) \sum_k \frac{\eta_k^{(0)}}{m_k R_k}. \tag{6.17}$$

Conversely, assume that Eq. (6.16) holds. Then, for some sufficiently small $\epsilon > 0$, the problem is feasible by considering $Q_k \to \infty$ and $\gamma_k = (1 + p_{\text{fb}}/(1 - p_{\text{fb}}))(\eta_k^{(0)} + \epsilon)/(m_k R_k)$. $\square$

Thus, from Lemmas 6.4 and 6.5, Theorem 6.3 still holds for the case $p_{\text{fb}} \ne 0$ by replacing f with f_{fb}. In Corollary 6.6, we derive another interpretation of Lemma 6.5 that characterizes the maximum value of p_{fb} at which the system can work.

Corollary 6.6. *Assuming Eq. (6.16) holds, Problem 6.1 is feasible if, and only if, $p_{\text{fb}} < p_{\text{fb}}^t$ with:*

$$p_{\text{fb}}^t = 1 - \sum_{k=1}^{K} \frac{\eta_k^{(0)}}{m_k R_k}. \tag{6.18}$$

One can remark that, under Eq. (6.7), the threshold is well defined since the inequalities $0 < p_{\text{fb}} < 1$ hold. Furthermore, the proof of Corollary 6.6 is easy since Eq. (6.16) holds if and only if Eq. (6.18) holds.

6.3.5   Numerical results with fixed MCS

In this Section, two practical MCS are considered:

- **MCS1**: uncoded packets of $n = 128$ bits, which are mapped onto a 2^{m_k}-QAM constellation;

- **MCS2**: packets of $n = 512$ bits coming from rate R convolutional encoding.

Each link fixes the order 2^{m_k} of its QAM constellation according to a rule that will be precised for each simulation. For the same reasons as explained in Chapter 5, the data rate request is uniform $\eta_k^{(0)} = \eta_T/K$, with η_T the total goodput demand of the cluster (in bit/s/Hz), and the bandwidth is fixed to $W = 1$ MHz.

6.3.5.1   Performance gap to the Gaussian codes of length n

In Fig. 6.1 (resp. Fig. 6.2), we plot the total transmit power (resp. the total occupied bandwidth) after allocation using:

- Algorithm 6.1 on the MCS2 with $R = 1/2$, and

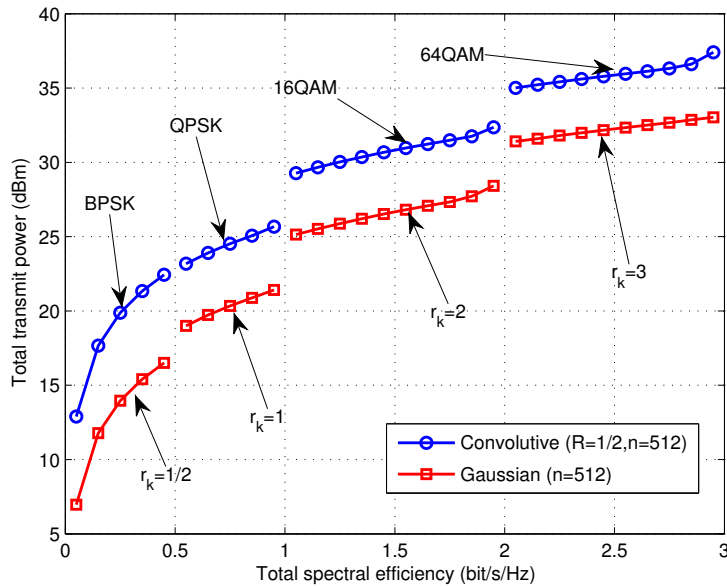- Algorithm 5.1 on a Gaussian code of length $n = 512$.

Figure 6.1: Total transmit power of 1/2-rate convolutional code with QAM versus finite length Gaussian codes ($K = 2$, $n = 512$).
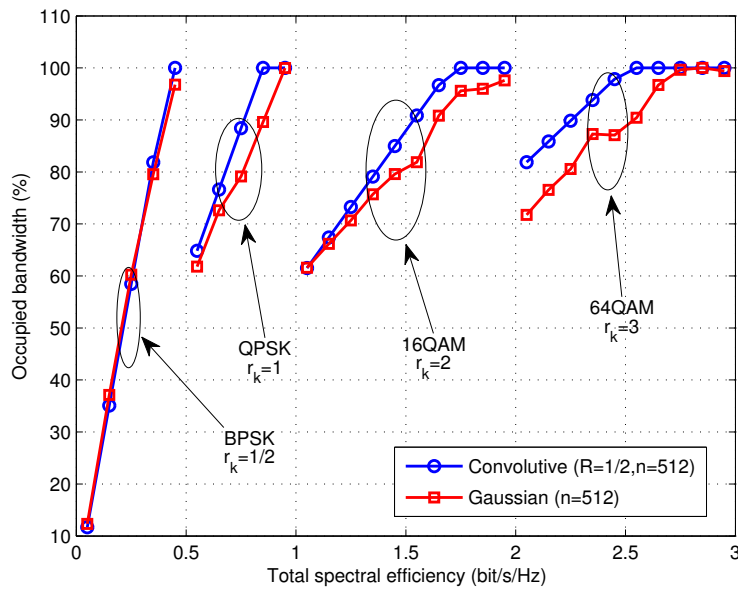

Figure 6.2: Occupied bandwidth of 1/2-rate convolutional code with QAM versus finite length Gaussian codes ($K = 2$, $n = 512$).

In the two cases, (m_k, R_k) are fixed such that Lemma 6.1 holds and the Gaussian rate r_k (see Chapter 5) such that Condition 5.1 holds. As in Chapter 5, $K = 2$ and the average SNRs are fixed to 10 dB and 30 dB.

We have two remarks. Firstly, there is a constant gap of about 5 dB between the performance of scheme (b) and Gaussian coding for the range of [0,3] bit/s/Hz ([0,3] Mbit/s within W). Secondly, the Gaussian code allocation uses less bandwidth than the MCS, and both schemes use an increasing amount of bandwidth at each rate. Thus, frequency reuse is possible when using such coding schemes.

6.3.5.2 Performance gap to efficient suboptimal solutions

Now the optimal Algorithm 6.1 is compared to two suboptimal policies that drastically simplify the optimization procedure. The two suboptimal solutions are obtained by letting the values γ_k^* being proportional to $\eta_k^{(0)}$ as follows:

$$\gamma_k = \frac{\eta_k^{(0)}/(m_k R_k)}{\sum_{k'=1}^{K} \eta_{k'}^{(0)}/(m_{k'} R_{k'})}. \tag{6.19}$$

By definition, the suboptimal solutions place the users in the entire bandwidth:

$$\sum_{k=1}^{K} \gamma_k = \sum_{k=1}^{K} \frac{\eta_k^{(0)}/(m_k R_k)}{\sum_{k'=1}^{K} \eta_{k'}^{(0)}/(m_{k'} R_{k'})} = 1. \tag{6.20}$$

We define two suboptimal policies depending on how Q_k are computed: a) equal Q_k for all k (Suboptimal A), *i.e.*, $Q_k = Q/K$ where Q is chosen such that the rate constraint is satisfied, or b) by computing Q_k such that the rate constraint in Eq. (6.6b) is satisfied:

$$\eta_k^{(0)} = \gamma_k m_k R_k f_{\text{fb}}(P_k\left(\frac{G_k Q_k}{\gamma_k}\right)). \tag{6.21}$$

Definition 6.1 (Suboptimal A). *For each $k \in \{1, \ldots, K\}$, compute γ_k from Eq. (6.19), and $Q_k = Q/K$ with Q such that all the constraints are satisfied.*

Definition 6.2 (Suboptimal B). *For each $k \in \{1, \ldots, K\}$, compute γ_k from Eq. (6.19), and:*

$$Q_k = \frac{\gamma_k}{G_k} P_k^{-1}(f_{\text{fb}}^{-1}\left(\frac{\eta_k^{(0)}}{\gamma_k m_k R_k}\right)). \tag{6.22}$$

When $p_{\text{fb}} = 0$, f_{fb} boils down to f.

The path loss follows the free-space model $\ell(D) = 1/\left((4\pi f_0/c)^2 D^2\right)$ where c is the light celerity and f_0 the carrier frequency [2]. We put $f_0 = 400$ MHz and the noise density power

---

[2]Only free-space is considered for the sake of presentation clarity. Other models have been tested, like the so-called Okomura-Hata path loss for urban propagation models, and it only changes the y-scale, *i.e.* the conclusions remain the same.

is fixed to $N_0 = -170$ dBm/Hz. The distance D_k between both users associated with the k-th link is randomly drawn from a uniform distribution in $[D_m, D_M]$. We have taken $D_m = 50$ m and $D_M = 1$ km. Each point is averaged by Monte-Carlo drawing.

In Fig. 6.3 (resp. Fig. 6.4), we plot the total transmit power versus spectral efficiency for the MCS1 with BPSK (resp. 16-QAM). Suboptimal B outperforms Suboptimal A for both modulations. Interestingly, Suboptimal B performance are close to the optimal for $\eta_T \geq 0.3$ bit/s/Hz. Thus, Suboptimal B provides a fast allocation algorithm that performs quasi-optimally for $\eta_T \geq 0.3$ bit/s/Hz. This can be explained by the bandwidth occupation. In Fig. 6.5 we plot the bandwidth fraction that is occupied after allocation by MCS1. Let us remind from Theorem 6.3 that the bandwidth is not always full at the optimum, especially at low η_T. We observe from Fig. 6.5 that while $\sum_{k=1}^{K} \gamma_k < 90$ % (*i.e.* while $\eta_T < 0.3$ bit/s/Hz) Suboptimal B does not achieve the optimum because it costs up to 90 % more bandwidth.
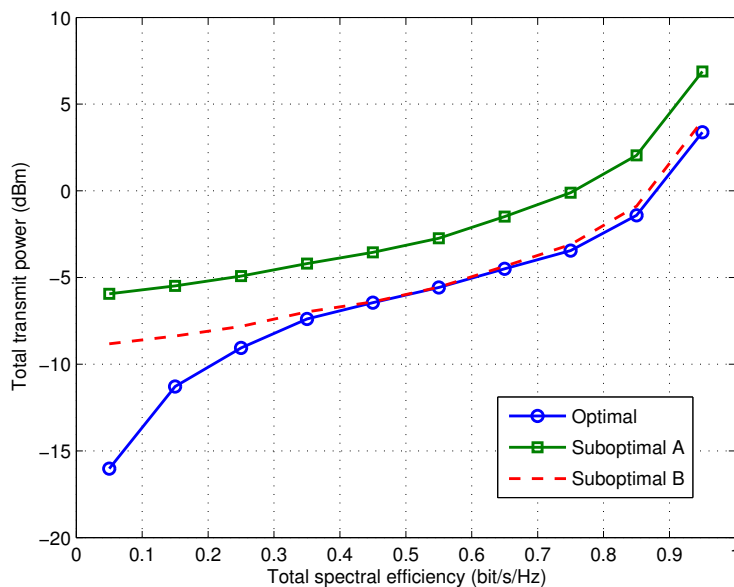
Figure 6.3: Total transmit power versus spectral efficiency (MCS1, BPSK, $K = 4$).

As we will see hereafter, similar remarks can done for the coded case MCS2. Fig. 6.6 displays the total transmit power versus spectral efficiency for the MCS2 with BPSK, and we see again that Suboptimal B is close to optimal for low spectral efficiency. However, when achieving higher spectral efficiency, this policy is not optimal for $\eta_T < 1.5$ bit/s/Hz as seen from Fig. 6.7 which displays the case of 16-QAM. Finally, coding helps to save bandwidth when switching the modulation, as seen from Fig. 6.8), where we plot the bandwidth fraction that is occupied after allocation by MCS2.

Figure 6.4: Total transmit power versus spectral efficiency (MCS1, 16-QAM, $K = 4$).

Figure 6.5: Occupied bandwidth versus spectral efficiency (MCS1, BPSK, $K = 4$).

Figure 6.6: Total transmit power versus spectral efficiency (MCS2, $R = 1/2$, BPSK, $K = 4$).
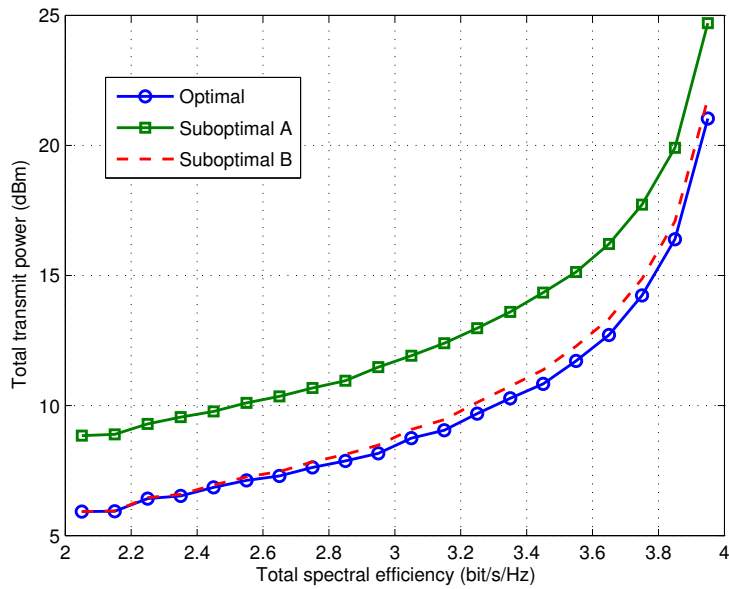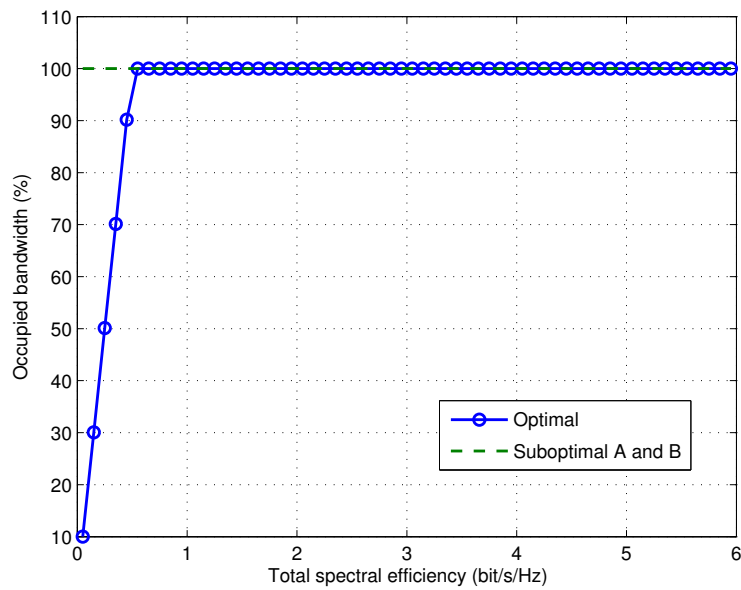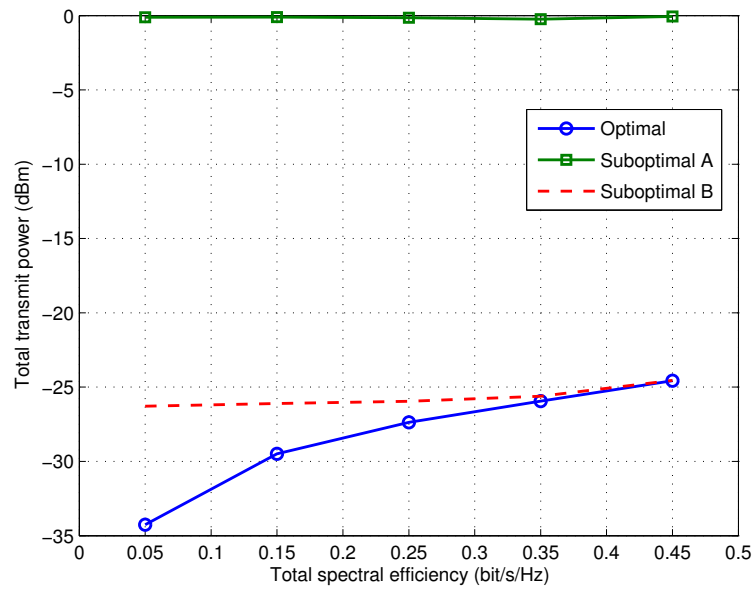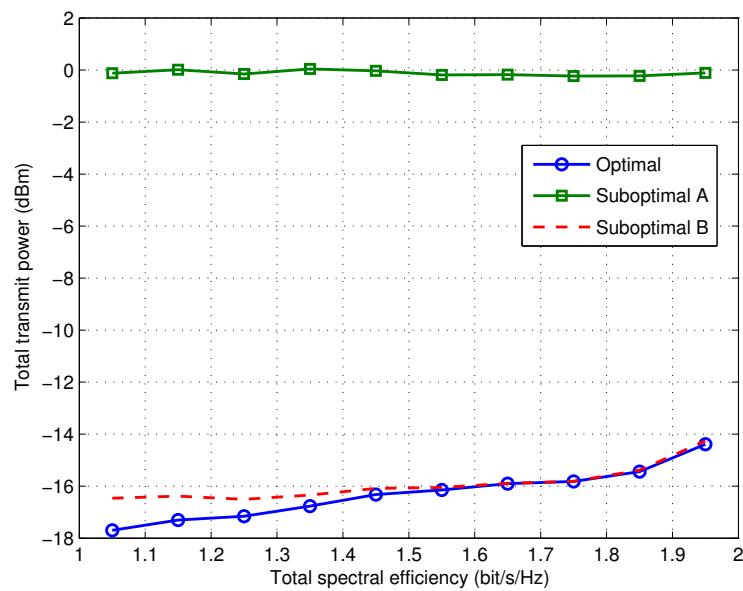

Figure 6.7: Total transmit power versus spectral efficiency (MCS2, $R = 1/2$, 16-QAM, $K = 4$).
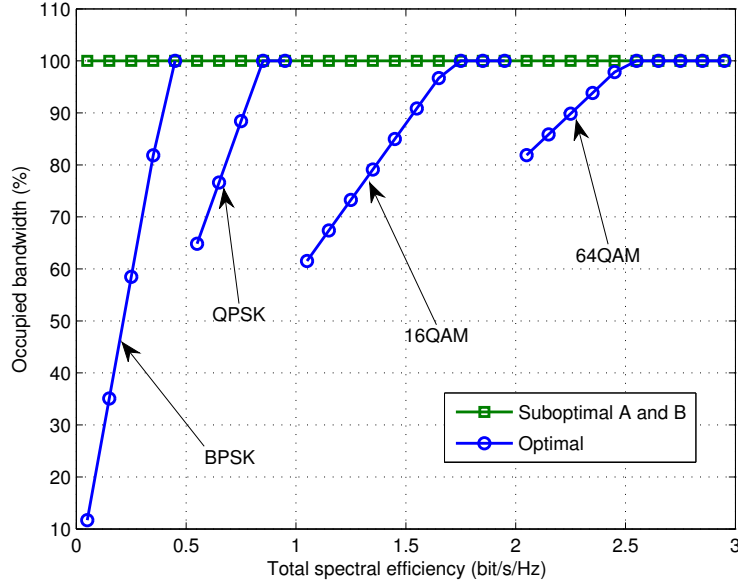
Figure 6.8: Occupied bandwidth versus spectral efficiency (MCS2, $R = 1/2$, BPSK to 64QAM, $K = 4$).

6.3.5.3   Imperfect feedback case

In order to analyze the influence of p_{fb}, we define the power loss (in dB) as:

$$10 \log_{10}\left(\frac{Q_T^*(p_{\text{fb}})}{Q_T^*(0)}\right), \tag{6.23}$$

where $Q_T^*(p_{\text{fb}})$ is the optimal total transmit power when the feedback error probability value is p_{fb}. In Fig. 6.9, we plot the power loss versus p_{fb} when MCS1 (with BSPK and $K = 4$) is used. For this simulation, we have configured $\mathbf{D} = [50, 100, 500, 700]$ m and $\eta^{(0)} = [0.2, 0.2, 0.4, 0.1]$ bit/s/Hz, and BPSK is used. According to Corollary 6.6 we know that $p_{\text{fb}}^t = 0.1$ for this distribution $\eta^{(0)}$. We observe that the power loss grows exponentially and becomes too huge when p_{fb} is close to p_{fb}^t.

Figs. 6.10 and 6.11 show that the suboptimal performance are pretty much robust when the probability of erroneous feedback is fixed to $p_{\text{fb}} = 0.95\, p_{\text{fb}}^t$.

6.3.6   Modulation and coding scheme selection

Now, we address the problem of selecting the best MCS for each user in Problem 4.3. The extension is straightforward for Problems 4.4 and 4.5, therefore we will only present the results with Problem 4.3.

Let us remind that the available modulations and code rates are described by the sets $\mathcal{M}$ and $\mathcal{R}$, respectively. The MCS of link k is denoted by $\text{mcs}_{i_k} = (m_{i_k}, R_{i_k})$ with $m_{i_k} \in \mathcal{M}$
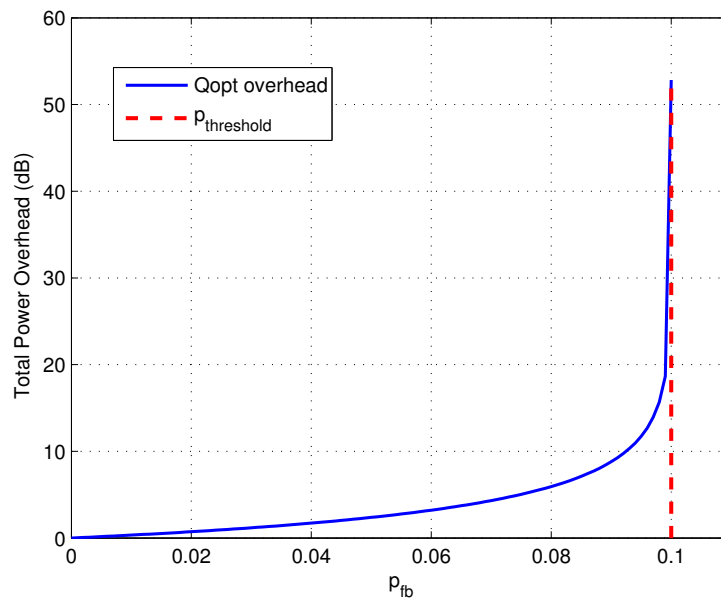
Figure 6.9: Power loss (in dB) versus the feedback error probability.
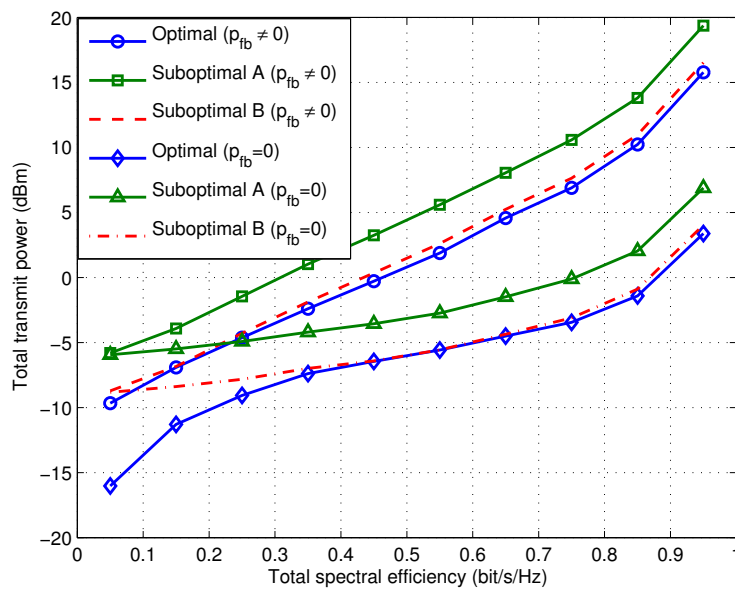
Figure 6.10: Total transmit power versus spectral efficiency (MCS1, BPSK, $p_{\text{fb}} = 0.95\, p_{\text{fb}}^{t}$, $K = 4$).

and $R_{i_k} \in \mathcal{R}$. It is assumed that the MCS combinations $\text{mcs}_i \in \mathcal{G} \subset (\mathcal{M} \otimes \mathcal{R})$ are ordered
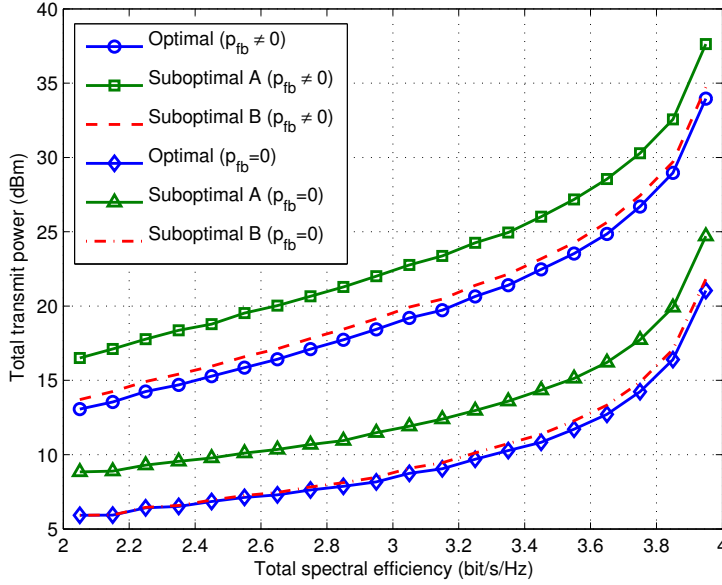
Figure 6.11: Total transmit power versus spectral efficiency (MCS1, 16-QAM, $p_{\text{fb}} = 0.95\, p_{\text{fb}}^t$, $K = 4$).

such that, in order to achieve a target PER:

$$\text{SNR}(\text{mcs}_i) < \text{SNR}(\text{mcs}_j) \qquad \text{for } i > j. \tag{6.24}$$

Finally, let us denote a MCS vector by $\mathbf{mcs} = [(m_{i_1}, R_{i_1}), \ldots, (m_{i_K}, R_{i_K})]^T$.

6.3.6.1   Optimization problem

The solution to Problem 6.1 optimizes the power and bandwidth allocation when the MCS is given. However, when doing the resource allocation the choice of m_k and R_k for each link k must be optimized as well. The global problem of interest can be written as follows:

Problem 6.2.

$$(\gamma^*, Q^*, \mathbf{mcs}^*) = \arg \min_{(\gamma, Q) \in C, \mathbf{mcs} \in \mathcal{G}^K} Q_T(\mathbf{mcs}) \tag{6.25}$$

where $C = \{(\gamma, Q) \in (0, 1)^K \times \mathbb{R}_{+*}^K \mid$ (6.6b)-(6.6c) satisfied $\}$.

Actually, Theorem 6.3 is the key for providing the minimum total energy given $\mathbf{mcs} \in \mathcal{G}^K$:

$$Q_T^*(\mathbf{mcs}) = \min_{(\gamma, Q) \in C} Q_T(\mathbf{mcs}). \tag{6.26}$$

Thus each optimal solution to the next Problem 6.2a leads to a globally optimal solution to Problem 6.2.

Problem 6.2a.

$$\mathbf{mcs}^* = \arg \inf_{\mathbf{mcs} \in \mathcal{G}^K} Q_T^*(\mathbf{mcs}). \tag{6.27}$$

This is a combinatorial optimization problem since the optimization space is discrete, and it is well known that in general this class of problems is difficult (NP-hard). An exhaustive search among all the settings in $\mathcal{G}^K$ is a trivial and optimal way to find it, but its complexity is out of reason and this approach is intractable for large enough users. Combining Adaptive Modulation and Coding (AMC) and HARQ for link adaptation has been already studied in the literature: the case of Type-I HARQ is done in [Liu et al., 2004], and Type-II HARQ is treated in [Lagrange, 2010]. These techniques could be adapted to our situation. In what follows, we propose a suboptimal solution inspired by [Devillers et al., 2008] that reduces the exhaustive search complexity.

6.3.6.2   A Greedy approach towards an efficient MCS selection

The idea is to try the next $\mathrm{mcs}_{i_k+1} \in \mathcal{G}$ in the MCS list for each user k, and to select the users with the new MCS that has the lowest power. The approach is greedy in the sense that the operation continues while the power decreases, and stops otherwise. Let $\mathbf{mcs}^{(0)} = [\mathrm{mcs}_{i_1}^{(0)}, \cdots, \mathrm{mcs}_{i_K}^{(0)}]^T \in \mathcal{G}^K$ be such that $\forall k$, $\mathrm{mcs}_{i_k}^{(0)}$ is the first feasible MCS in the list. Basically, this can be quickly checked using Lemma 6.1. Finally the algorithm, called "Greedy MCS selection" is summarized in Algorithm 6.2.

Algorithm 6.2: Greedy MCS selection.

Set $Q_T^* = \infty$ and $\mathbf{mcs} = \mathbf{mcs}^{(0)}$

1. MCS testing:

for $k = 1$ **to** K **do**
  Let $\mathbf{mcs}^{(k)} \in \mathcal{G}^K$ with $\mathrm{mcs}_{i_k} \leftarrow \mathrm{mcs}_{i_k+1}$
  Compute $Q_T^*(\mathbf{mcs}^{(k)})$ according to Eq. (6.26) using Theorem 6.3
end

2. Select $k^* = \arg \inf_k Q_T^*(\mathbf{mcs}^{(k)})$

3. Greedy heuristic:

if $Q_T^* > Q_T^*(\mathbf{mcs}^{(k^*)})$ **or Lemma 6.1 does not hold then**
  $Q_T^* \leftarrow Q_T^*(\mathbf{mcs}^{(k^*)})$, $\mathbf{mcs} \leftarrow \mathbf{mcs}^{(k^*)}$, and go back to step **1.**
else
  Exit

6.3.7 Numerical results for MCS selection

6.3.7.1 Simulation settings

We combine the optimal bandwidth and power allocation of Theorem 6.3 with three MCS selections:

(i) A fixed selection of the MCS that trivially satisfies Lemma 6.1. The first MCS mcs_{i_k} such that $m_{i_k} R_{i_k} > \eta_T$ is taken, hence $\sum_{k=1}^{K} \eta_k^{(0)}/(m_k R_k) < K(\eta_T/K)/\eta_T = 1$ and Lemma 6.1 holds. For instance, those users requiring $\eta_T = 1$ bit/s/Hz should use the MCS with $m_k = 4$ and $R_k = 1/2$.

(ii) An exhaustive search of the best MCS associated with Problem 6.2a (tractable for small K).

(iii) The "Greedy MCS selection" summarized in Algorithm 6.2.

QAM constellations with $\mathcal{M} = \{1, 2, 4, 6\}$, and a punctured convolutional code with rates $\mathcal{M} = \{3/4, 2/3, 1/2\}$ have been considered in the simulation. The four modulations shown in Tab. 6.1 are actually associated with the (uncoded) MCS1, and the six MCS reported in Tab. 6.2 are associated with the (coded) MCS2. In Fig. 6.12, we plot the empirical PER for the MCS given in Tab. 6.2 and the associated theoretical PER given by Eq. (6.3). The parameters g_c and d_f defined in Eq. (6.3) are determined by applying a curve-fitting method. We remark that the theoretical PER expression predicts well the empirical performance which justifies its use in our derivations.

MCS name	BPSK	QPSK	16-QAM	64-QAM
m	1	2	4	6
max bit/s/Hz	1	2	4	6

Table 6.1: Practical MCS used in the uncoded MCS1 framework.

MCS name	MCSc1	MCSc2	MCSc3	MCSc4	MCSc5	MCSc6
m	1	2	2	4	6	6
R	1/2	1/2	2/3	1/2	1/2	3/4
max bit/s/Hz	0.5	1	1.33	2	3	4.5

Table 6.2: Practical MCS used in the coded MCS2 framework.

6.3.7.2 Simulation results

In Fig. 6.13, (resp. Fig. 6.14) we display the total transmit power versus the spectral efficiency for the MCS defined in Tab. 6.1 when $p_{\text{fb}} = 0$ (resp. $p_{\text{fb}} = 0.95 p_{\text{fb}}^t$).
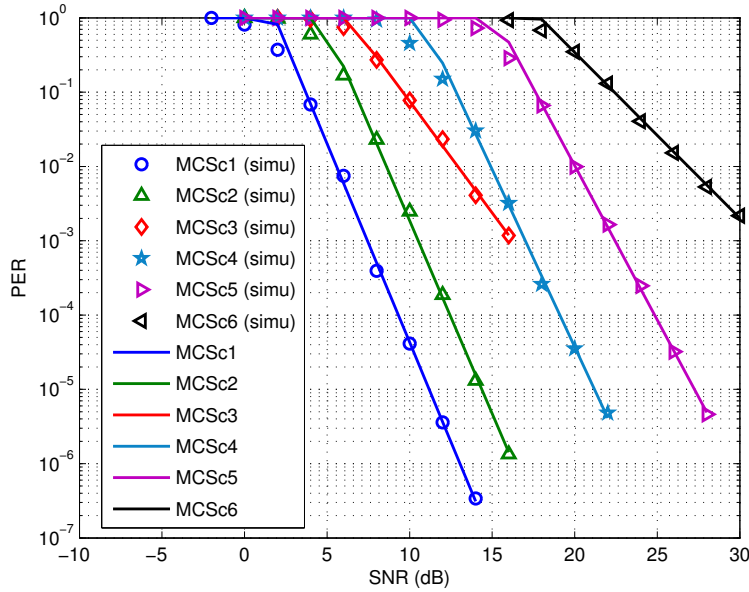
Figure 6.12: MCS used with the coded MCS2.

First of all, Algorithm 6.2 performs close to the exhaustive algorithm (which is optimal) and is far less complex. On the other hand, the fixed MCS selection performs quite poorly, especially when the feedback is imperfect. For fixed MCS, notice the huge increase in power when the total goodput requirement η_T is close to the value of an element $m \in \mathcal{M}$. These high peaks around 1, 2, 4 and 6 bit/s/Hz are explained by Lemma 6.1. Indeed, the power must be strongly increased whenever $\sum_{k=1}^{K} \eta_k^{(0)}/m$ goes close to 1 in the fixed case, i.e. when $\sum_{k=1}^{K} \eta_k^{(0)}$ goes close to m. Actually, switching the MCS as done in Algorithm 6.2 enables large power savings.

In Fig. 6.15, we plot the total transmit power versus the spectral efficiency for the MCS defined in Tab. 6.2. Due to the complexity of the exhaustive selection algorithm, we consider only $K = 2$ links with arbitrary average SNR configured to 10 dB and 30 dB, respectively. The same conclusions hold in this case, i.e., Algorithm 6.2 performance are close to the optimum though far less complex than exhaustive selection, and that a good MCS selection algorithm can greatly improve the performance compared to a fixed MCS solution.

6.4 Packet error and rate constrained power minimization

In this Section we focus on the resolution of Problem 4.4, for which an additional PER constraint has been defined:

$$P_k^{\text{MAC}}(E_k) \leq P_k^{\text{MAC},(0)}. \tag{6.28}$$
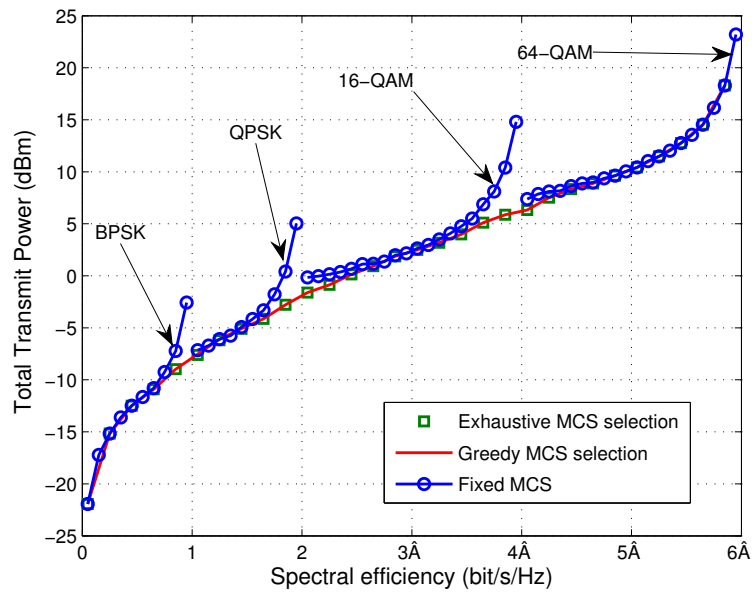
Figure 6.13: Total transmit power versus the spectral efficiency (MCS from Tab. 6.1, $p_{\text{fb}} = 0$, $K = 4$).
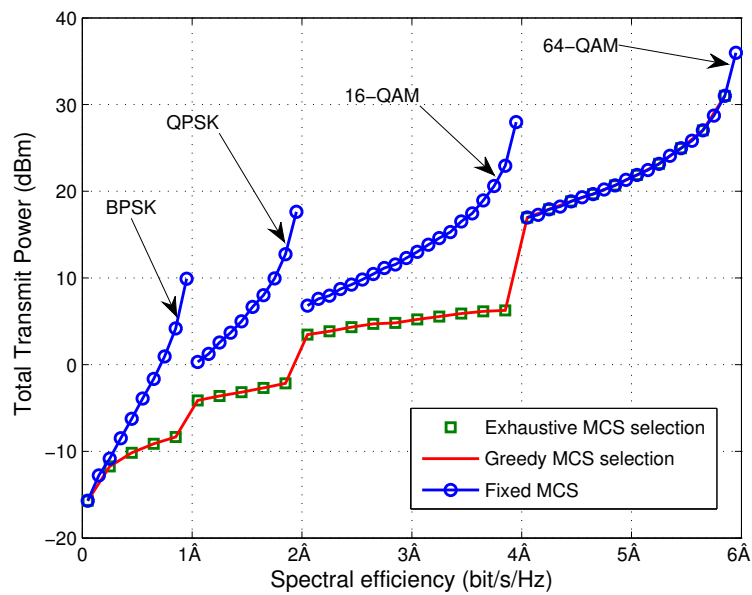
Figure 6.14: Total transmit power versus the spectral efficiency (MCS from Tab. 6.1, $p_{\text{fb}} = 0.95p_{\text{fb}}^t$, $K = 4$).
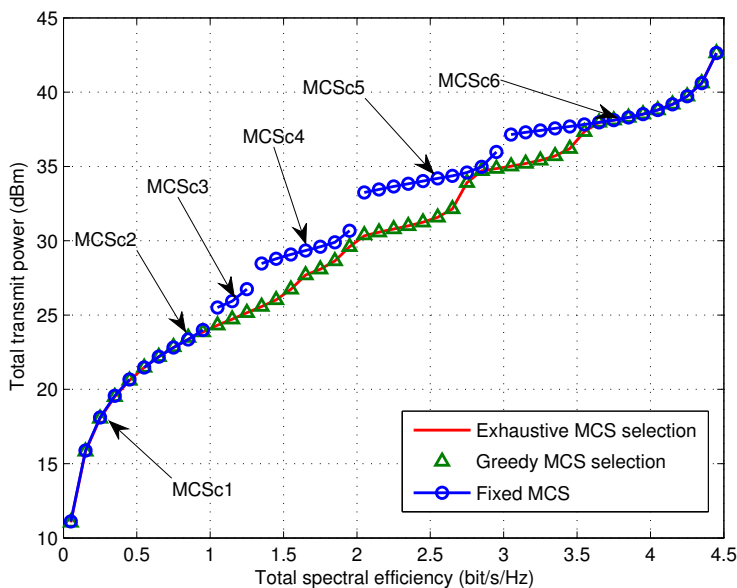
Figure 6.15: Total transmit power versus the spectral efficiency (MCS from Tab. 6.2, $p_{\text{fb}} = 0$, $K = 2$, $G_1 = 10$ dB and $G_2 = 30$ dB).

Recalling that the MAC level PER of Type-I HARQ with L transmission credit is:

$$P_k^{\text{MAC}}(E_k) = P_k(G_k E_k)^L, \tag{6.29}$$

the MAC level constraint boils down to the PHY level PER constraint:

$$P_k(G_k E_k) \leq P_k^{(0)}, \tag{6.30}$$

where $P_k^{(0)} = \left(P_k^{\text{MAC},(0)}\right)^{1/L}$.

The problem is first rewritten as a function of (γ_k, Q_k) in order to exhibit once again a convex objective function. Then the optimal algorithm is derived, and suboptimal solutions are given too.

6.4.1 Optimization problem formulation

Rewritting Problem 4.4 using Eq. (6.4) and Eq. (6.30) leads to Problem 6.3. Notice that the unfeasibility of the all-zero solutions ($\gamma = 0, Q = 0$) applies as well.

Problem 6.3.

$$\min_{(\gamma, Q)} \sum_{k=1}^{K} Q_k, \tag{6.31a}$$

$$s.t. \quad \tilde{\eta}(\gamma_k, Q_k) \geq \eta_k^{(0)}, \; \forall k, \tag{6.31b}$$

$$P_k(G_k Q_k / \gamma_k) \leq P_k^{(0)}, \; \forall k, \tag{6.31c}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{6.31d}$$

$$\gamma_k > 0, \; Q_k > 0, \; \forall k. \tag{6.31e}$$

6.4.2 Feasibility and structure properties

It is easy to show that Lemma 6.1 holds for Problem 6.3 too. Based on Lemma 6.7 proved in Appendix F.4, Problem 6.3 is the minimization of a convex function **over a convex set**.

Lemma 6.7. *The constraint functions defined on $[0,1] \times \mathbb{R}_+$ by:*

$$(\gamma_k, Q_k) \mapsto \tilde{\eta}_k(\gamma_k, Q_k), \tag{6.32}$$

$$(\gamma_k, Q_k) \mapsto P_k(G_k Q_k / \gamma_k), \tag{6.33}$$

are respectively concave and quasi-convex [Greenberg and Pierskalla, 1971], as long as $P_k : \mathbb{R}_+ \longrightarrow [0,1]$ is a convex function.

In Appendix F.5, we prove the following Lemma.

Lemma 6.8. *The set:*

$$\mathcal{F} = \{(\gamma, Q) \in [0,1]^K \times \mathbb{R}_+^K \, | \, Eqs. \, (6.31b)\text{-}(6.31d) \text{ are satisfied}\} \tag{6.34}$$

is convex and Slater's assumption holds[3]. Moreover, the nondegeneracy condition:

$$\nabla \tilde{\eta}_k(\gamma_k, Q_k) \neq 0 \quad \forall(\gamma, Q) \in \mathcal{F} \text{ such that } \eta_k^{(0)} - \tilde{\eta}_k(\gamma_k, Q_k) = 0, \tag{6.35a}$$

$$\nabla P_k(G_k Q_k / \gamma_k) \neq 0 \quad \forall(\gamma, Q) \in \mathcal{F} \text{ such that } P_k(G_k Q_k / \gamma_k) - P_k^{(0)} = 0, \tag{6.35b}$$

holds for every $k = 1, \ldots, K$.

Since Lemma 6.8 holds, it is proven in [Lasserre, 2010, Th. 2.3] that Problem 6.3 can be solved optimally by using the so-called KKT conditions. Therefore in next Subsection, we will develop the optimal algorithm by deriving the KKT conditions in closed-form.

---

[3]Basically, Slater's condition requires that a strictly feasible solution must exist. This is guaranteed for Problem 6.3 by Lemma 6.1

6.4.3 Optimal algorithm

Tedious derivations available in Appendix F.6 lead to the following characterization of optimal solutions to Problem 6.3:

Theorem 6.9. *The optimal allocation policy* (γ^*, Q^*) *satisfies:*

$$\gamma_k^* m_k R_k (1 - P_k(G_k Q_k^*/\gamma_k^*)) - \eta_k^{(0)} = 0, \tag{6.36}$$

$$\left(\Theta(G_k Q_k^*/\gamma_k^*) - \lambda^* G_k \right) \left(P_k(G_k Q_k^*/\gamma_k^*) - P_k^{(0)} \right) = 0, \tag{6.37}$$

$$\lambda^* \left(\sum_{k=1}^{K} \gamma_k^* - 1 \right) = 0, \tag{6.38}$$

where, $\forall x \in \mathbb{R}_{+*}$:

$$\Theta(x) := \frac{x P_k'(x) + 1 - P_k(x)}{P_k'(x)}. \tag{6.39}$$

To deduce an algorithm from Theorem 6.9, it is essential to know if the PER constraint Eq. (6.31c), which is related to Eq. (6.37), is active or not. If the PER constraint is inactive then $\Theta(G_k Q_k^*/\gamma_k^*) = \lambda^* G_k$ from Eq. (6.37). Notice that $\sum_{k=1}^{K} \gamma_k$ is decreasing when λ increases, thus if the PER constraints are inactive at the optimum for $\lambda^* > 0$, then $\sum_{k=1}^{K} \gamma_k^* = 1$ considering Eq. (6.38). However, there may be some users who were computed through Θ that would lead to unfeasible PER values $P_k(\Theta^{-1}(\lambda^* G_k)) > P_k^{(0)}$. In such cases, some users among these unfeasible users should have made their PER constraints active, *i.e.* $Q_k^*/\gamma_k^* = (1/G_k)P_k^{-1}(P_k^{(0)})$. Therefore, all the configurations on Eq. (6.37) are tested, *i.e.*, either the first factor is zero (then PER constraint is inactive) or not (then PER constraint is active), and we select eventually the best one with respect to the total transmit power. More precisely, in a first step, we consider that only $n \in \{0, \cdots, K\}$ user(s) have the PER constraint active. Then in a second step, we compute the total transmit power for all the tested configurations and select the best one. Let $\mathcal{U}_n$ be the set of all the sets of n users out of K. If u is a subset of K users, u^c is the associated complementary subset. For $n = 0$, the loop on u is implemented only once by considering $u = \emptyset$ and $u^c = \{1, \cdots, K\}$. Algorithm 6.3 achieves optimality but requires $O(2^{K-1})$ operations, and thus becomes cost-computing for large enough K. Therefore, we next propose two suboptimal but computationally tractable algorithms.

6.4.4 Suboptimal algorithms

6.4.4.1 Suboptimal KKT based Algorithm (SKA)

In order to reduce the complexity of the optimal Algorithm 6.3, we force the algorithm to operate as if the constraint on the PER were not active, *i.e.* the left-hand term associated with the Lagrange multiplier vanishes in Eq. (6.37). We remind that the constraint related to the goodput is always active (see Eq. (6.36)). Finally, the suboptimal KKT based algorithm is resumed in Algorithm 6.4.

Algorithm 6.3: Optimal algorithm for Problem 6.3.

---

for $n = 0$ **to** K **do**

    for each $u \in \mathcal{U}_n$ **do**

        $\forall k \in u$,

        $\gamma_k = \eta_k^{(0)}/(m_k R_k (1 - P_k^{(0)}))$, and $Q_k = \gamma_k P_k^{-1}(P_k^{(0)})/G_k$,

        $\forall k' \in u^c$,

        $\gamma_{k'} = \eta_{k'}^{(0)}/(m_{k'} R_{k'} (1 - P_{k'}(\Theta^{-1}(\lambda G_{k'}))))$, and $Q_{k'} = \gamma_{k'}\Theta^{-1}(\lambda G_{k'})/G_{k'}$,

        for $\lambda \in \mathbb{R}_*^+$ such that $\sum_{k \in u} \gamma_k + \sum_{k' \in u^c} \gamma_{k'} = 1$ (if no λ leads to equality, put

        first $\lambda = 0$ and test the condition $\sum_k \gamma_k < 1$. If the condition is not satisfied,

        then put $\lambda = \infty$)

        if $\exists k' \in u^c$, **s.t.** $P_{k'}(G_{k'} Q_{k'}/\gamma_{k'}) > P_{k'}^{(0)}$ **or** $\lambda = \infty$ **then**

           | $Q_T(u) = \infty$

        end

    end

end

Choose u minimizing $Q_T(u)$.

Algorithm 6.4: Suboptimal KKT based Algorithm (SKA) for Problem 6.3.

Set $\lambda = 0$, $Q_k = \epsilon > 0$ and $\gamma_k = 1$, $\forall k$.

while $\sum_k \gamma_k > 1$ **or** $\exists k$, $P_k(G_k Q_k/\gamma_k) > P_k^{(0)}$ **do**

    1. $\forall k$, $\gamma_k = \eta_k^{(0)}/(m_k R_k (1 - P_k(\Theta^{-1}(\lambda G_k))))$

    2. $Q_k = \gamma_k \Theta^{-1}(\lambda G_k)/G_k$

    3. Increase λ

end

6.4.4.2 Separate Linear Algorithm (SLA)

By remarking that Problem 6.3 comes from the equivalent form of Problem 4.4 that was written with respect to (γ, E), we hereafter propose another way to exhibit a sub-optimal algorithm. As P_k is a decreasing bijective function, Eq. (6.31c) boils down to $E_k \geq P_k^{-1}(P_k^{(0)})/G_k$ and so leads to the following equivalent optimization problem:

Problem 6.4. *Problem 6.3 is equivalent to:*

$$\min_{(\gamma, E)} \sum_{k=1}^{K} \gamma_k E_k, \tag{6.40a}$$

$$s.t. \quad \gamma_k \geq \eta_k^{(0)} / \left(m_k R_k \left(1 - P_k(G_k E_k) \right) \right), \forall k, \tag{6.40b}$$

$$E_k \geq P_k^{-1}(P_k^{(0)})/G_k, \forall k, \tag{6.40c}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{6.40d}$$

$$\gamma_k > 0, \ E_k > 0, \ \forall k. \tag{6.40e}$$

Problem 6.4 is actually more difficult than Problem 6.3 since the objective function is no longer convex. It is easy to see that it is a biconvex optimization problem, but the GOP solution remains too expensive. However, we can use a suboptimal approach which consists in optimizing the objective function separately on each parameter. Therefore we propose to split Problem 6.4 into two subproblems.

Problem 6.4a (on E). *For fixed γ, the subproblem is:*

$$E^* = \arg \min_{E} \sum_{k=1}^{K} \gamma_k E_k \tag{6.41}$$

subject to Eq. (6.40c) and Eq. (6.40e).

Problem 6.4b (on γ). *For fixed E, the subproblem is:*

$$\gamma^* = \arg \min_{\gamma} \sum_{k=1}^{K} \gamma_k E_k \tag{6.42}$$

subject to Eq. (6.40b) and Eqs. (6.40d)-(6.40e).

The solution to Problem 6.4a is $E_k^* = P_k^{-1}(P_k^{(0)})/G_k, \forall k$. Constraint (6.40b) has been removed from Problem 6.4a to avoid a deadlock issue. Indeed, if Eq. (6.40b) is added to Problem 6.4a and is active for user k, then the solution to Problem 6.4b is actually equal to the value of γ_k initializing Problem 6.4a, and so the optimal γ_k is given by the initialization step.

The solution to Problem 6.4b can be efficiently obtained by using linear programming tool, for instance, the Simplex method [Boyd and Vandenberghe, 2004]. However, for

some E, Problem 6.4b may not have a feasible solution since all the constraints may not be satisfied simultaneously. To overcome this issue, we suggest to increase E until a feasible solution to Problem 6.4b is found as follows: if Problem 6.4b is not feasible, add a small increment δ to each E_k as many times as necessary. Algorithm 6.5 summarizes these steps.

Algorithm 6.5: Suboptimal Separate Linear Algorithm (SLA) for Problem 6.3.

Initialize δ, and $E_k^* = P_k^{-1}(P_k^{(0)})/G_k$, $\forall k$.

1. $\forall k$, $\gamma_k \longleftarrow$ solution of Problem 6.4b (linear programming).

2. If a feasible solution has been found in step **1.**, then Exit.

3. Else, increase E_k by δ and go to step **1.**

6.4.5   MCS selection

Since Algorithm 6.2 can be used straightforwardly for the selection of MCS, we will not discuss about it and consider only BPSK modulation. Thus, the target goodput will not exceed 1 bit/s/Hz in the next Subsection devoted to numerical results.

6.4.6   Numerical results

In this Section, numerical results have been obtained for the scheme (a) presented in Section 6.3.5, and the path loss follows the free-space model. In Fig. 6.16 we plot the total transmit power versus spectral efficiency. It has been computed using the optimal Algorithm 6.3, the SKA (Algorithm 6.4) and the SLA, in order to evaluate the performance of the suboptimal algorithms (SKA and SLA). It can be seen that for the two PER constraints ($P_k^{\mathrm{MAC},(0)} = 10^{-2}$ or $P_k^{\mathrm{MAC},(0)} = 10^{-4}$), the SLA performance are extremely close to the optimal performance. The SKA performance are slightly lower than SLA for required goodput larger than 0.5 bit/s/Hz. We conclude that SLA is a very good heuristic to approach the optimal performance.

In Fig. 6.17 (resp. Fig. 6.18) we plot the PER (resp. the total transmit power) after allocation versus the goodput requirement. Fig. 6.17 displays the PHY level PER P_k as well as the MAC level PER P_k^{MAC} for $L = 3$, obtained using Algorithms 6.1 and 6.5. It can be observed that Algorithm 6.1 was not able to guarantee the target PER of $P_k^{\mathrm{MAC},(0)} = 10^{-2}$ for the links (in particular, it exceeds this constraint for $\eta_T < 0.8$ bit/s/Hz). Beyond 0.8 bit/s/Hz, the two problems achieve the same PER since the solution is driven by the goodput constraint in this area. However, the solution to Problem 4.4 (computed through the SLA algorithm) returns $P_k^{\mathrm{MAC}} \leq P_k^{\mathrm{MAC},(0)}$ for all η_T, as expected. Fig. 6.18 shows that the desired QoS is achieved at a little increase in consumed power (around 2 dB for a factor 10 gain in PER).
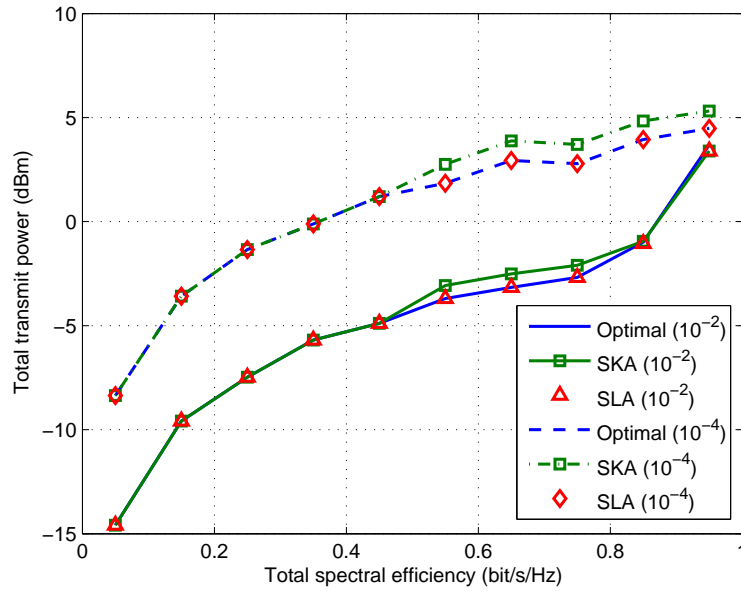
Figure 6.16: Total transmit power versus spectral efficiency computed with different algorithms (MCS1, BPSK, $K = 4$).
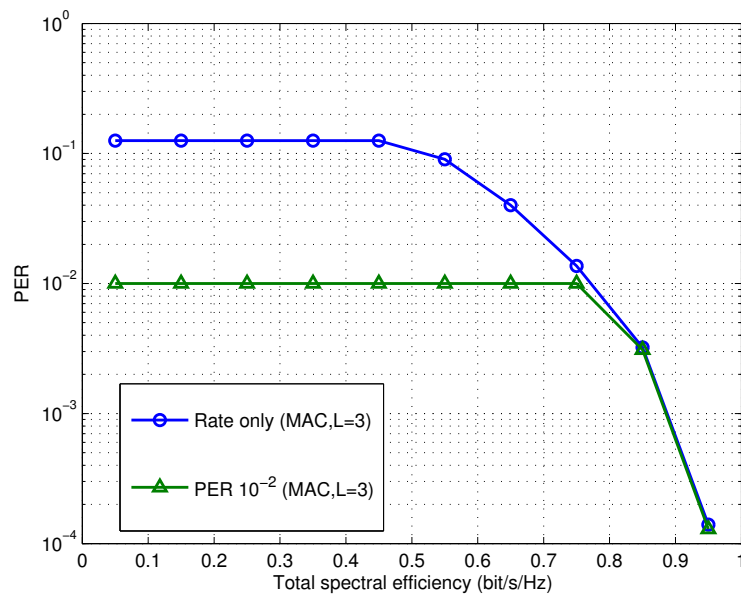

Figure 6.17: PER versus spectral efficiency (MCS1, BPSK, $K = 4$).
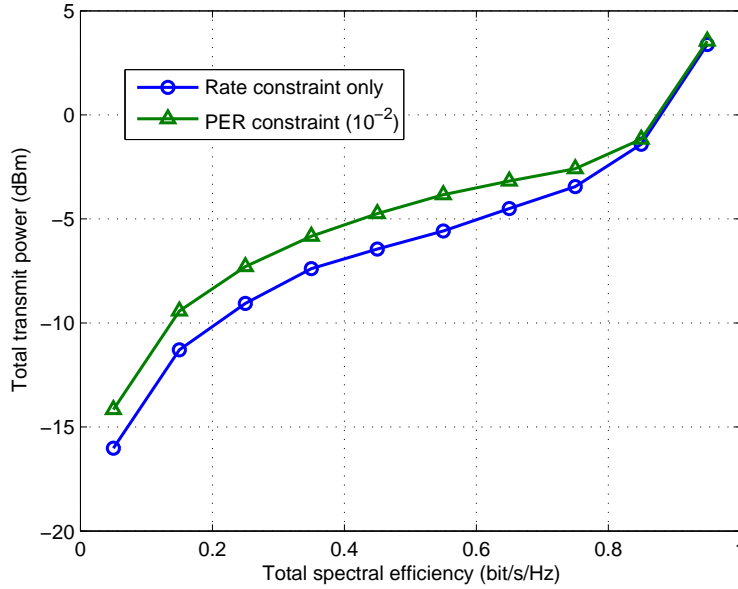
Figure 6.18: Total transmit power versus spectral efficiency (MCS1, BPSK, $K = 4$).

6.5   Delay and rate constrained power minimization

This Section is devoted to the study of Problem 4.5, for which an additional delay constraint has been defined as:

$$d_k^{\text{MAC}}(\gamma_k, E_k) \le d_k^{(0)}. \tag{6.43}$$

Due to the ARQ mechanism, the successful information packets associated with link k are received after $\delta(P_k)$ packet transmissions, where $x \mapsto \delta(x) = 1/(1-x) - Lx^L/\left(1 - x^L\right)$ [Le Martret et al., 2012]. The term $\delta(P_k)$ corresponds to the so-called "delay" in HARQ literature. Actually, as the bandwidth is never entirely assigned to a single link, the true delay is $\delta(P_k)$ divided by the bandwidth occupation rate γ_k. Therefore the delay for each successful information packet of link k, denoted by d_k, is given by:

$$d_k^{\text{MAC}}(\gamma_k, E_k) = \frac{1}{\gamma_k}\delta(P_k(G_k E_k)). \tag{6.44}$$

6.5.1   Optimization problem formulation

Next we rewrite Problem 4.5 using Eq. (6.4) and Eq. (6.44). Notice that the optimal solution (γ^*, Q^*) is such that $\gamma_k > 0$ and $Q_k > 0$ for all k. Indeed, if $\exists k$ such that $\gamma_k = 0$, then this user would have no way to satisfy its goodput nor its delay requirements ($\tilde{\eta}_k = 0$ whereas $d_k^{\text{MAC}} \to \infty$), and such a point would be unfeasible. Thus $\forall k \in \{1, \dots, K\}$, $\gamma_k > 0$. Now, since $Q_k = \gamma_k E_k$, we would have $Q_k = 0 \Rightarrow E_k = 0$ and hence $P_k = 1$. Once again, this would lead to an unfeasible solution. This boils down to Problem 6.5.

Problem 6.5.

$$\min_{(\gamma, Q)} \sum_{k=1}^{K} Q_k, \tag{6.45a}$$

$$s.t. \quad \tilde{\eta}(\gamma_k, Q_k) \geq \eta_k^{(0)}, \; \forall k, \tag{6.45b}$$

$$d_k^{MAC}(\gamma_k, Q_k) \leq d_k^{(0)}, \; \forall k, \tag{6.45c}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{6.45d}$$

$$\gamma_k > 0, \; Q_k > 0, \; \forall k. \tag{6.45e}$$

6.5.2 Feasibility property

Now, let us study the feasibility of Problem 6.5. The next result provides an easy way to check feasibility condition, as an inequality involving m_k, R_k, $\eta_k^{(0)}$ and $d_k^{(0)}$ only. Its proof follows the same lines than Lemma 6.1, and only a sketch of proof is given below. In the rest of the Chapter, we assume that Lemma 6.10 holds.

Lemma 6.10. *Problem 6.5 is feasible if, and only if,*

$$\sum_{k=1}^{K} \max\left(\frac{\eta_k^{(0)}}{m_k R_k}, \frac{1}{d_k^{(0)}}\right) < 1. \tag{6.46}$$

Sketch of proof. Only if part. If Problem 6.5 is feasible, then there is $(\gamma, Q) \in (0, 1)^K \times \mathbb{R}_{+*}^K$ such that for all $k \in \{1, \dots, K\}$:

$$\begin{cases} \eta_k^{(0)} \leq \gamma_k m_k R_k f(P_k(G_k Q_k / \gamma_k)) \\ d_k^{(0)} \geq \frac{1}{\gamma_k} \delta(P_k(G_k Q_k / \gamma_k)) \\ \sum_k \gamma_k \leq 1 \end{cases} \Rightarrow \begin{cases} \eta_k^{(0)} < \gamma_k m_k R_k \\ d_k^{(0)} > \gamma_k \end{cases} \tag{6.47}$$

since $P_k(G_k Q_k / \gamma_k) > 0$, so $1 - P_k(G_k Q_k / \gamma_k) < 1$ and $\delta(P_k) > 1$, hence we have:

$$\sum_{k=1}^{K} \max\left(\frac{\eta_k^{(0)}}{m_k R_k}, \frac{1}{d_k^{(0)}}\right) < \sum_{k=1}^{K} \gamma_k \leq 1. \tag{6.48}$$

If part. Conversely, let us take, $\forall k \in \{1, \dots, K\}$:

$$\gamma_k = \max\left(\frac{\eta_k^{(0)} + \epsilon}{m_k R_k}, \frac{1}{d_k^{(0)} - \epsilon}\right), \tag{6.49}$$

for a sufficiently small $\epsilon > 0$ such that by assumption, $\sum_{k=1}^{K} \gamma_k < 1$. When $Q_k \to \infty$, we have $P_k(G_k Q_k / \gamma_k) \longrightarrow 0$ and $\delta(P_k) \longrightarrow 1$, and one obtains:

$$\begin{cases} \tilde{\eta}_k \longrightarrow m_k R_k \max\left(\frac{\eta_k^{(0)} + \epsilon}{m_k R_k}, \frac{1}{d_k^{(0)} - \epsilon}\right) \geq \eta_k^{(0)} + \epsilon > \eta_k^{(0)} \\ d_k^{MAC} \longrightarrow \min\left(\frac{m_k R_k}{\eta_k^{(0)} + \epsilon}, d_k^{(0)} - \epsilon\right) \leq d_k^{(0)} - \epsilon < d_k^{(0)} \end{cases} \tag{6.50}$$

which concludes the proof. □

Finally, although constraints Eq. (6.45b), Eq. (6.45d) and Eq. (6.45e) are convex, the delay constraint Eq. (6.45c) is not convex. Problem 6.5 is thus nonconvex and can be difficult to solve efficiently. In the next Sections, we develop two suboptimal algorithms to solve Problem 6.5.

6.5.3   KKT based algorithm (KBA)

By looking numerically, the bivariate function $(x, y) \mapsto d_k^{\mathrm{MAC}}(x, y)$ seems very close to be a quasi-convex function. It can be observed in Fig. 6.19 which plots the sublevel sets $S_\alpha := \{(\gamma_k, Q_k) \in (0, 1) \times \mathbb{R}_{+*} \mid d_k^{\mathrm{MAC}}(\gamma_k, Q_k) \leq \alpha\}$ versus (γ_k, Q_k). Therefore, using the KKT conditions seems to be a relevant way even if we are not able to guarantee their optimality [Lasserre, 2010].
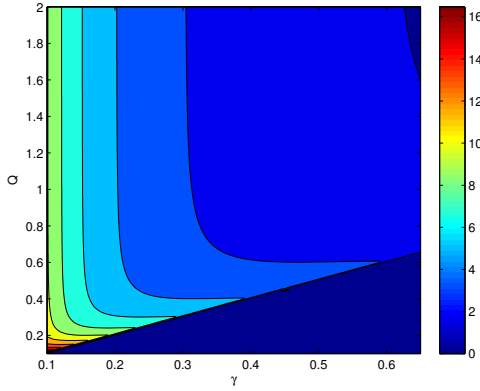
(a) The sublevel sets seem convex.   (b) Zooming shows the nonconvex areas in the sublevel sets.
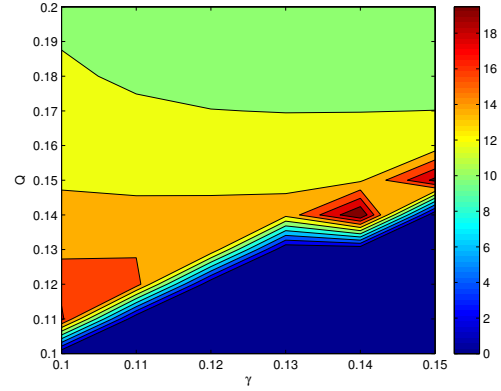
Figure 6.19: Sublevel sets of the delay function $d_k^{\mathrm{MAC}}(\gamma_k, Q_k)$.

Tedious algebraic manipulations given in Appendix F.7 leads to the following characterization of the KKT solution (γ, Q, λ):

Theorem 6.11. *The KKT point (γ, Q, λ) of Problem 6.5 satisfies:*

$$\left(M(G_k Q_k / \gamma_k) - \lambda G_k\right) \left(\eta_k^{(0)} - \gamma_k m_k R_k \left(1 - P_k(G_k Q_k / \gamma_k)\right)\right) = 0 \tag{6.51}$$

$$\left(\Theta(G_k Q_k / \gamma_k) - \lambda G_k\right) \left(\frac{\delta\left(P_k(G_k Q_k / \gamma_k)\right)}{\gamma_k} - d_k^{(0)}\right) = 0, \tag{6.52}$$

$$\lambda\left(\sum_{k=1}^{K} \gamma_k - 1\right) = 0, \tag{6.53}$$

where, $\forall x \in \mathbb{R}_{+*}$*:*

$$M(x) = -\frac{x\delta'(P_k(x))P_k'(x) + \delta(P_k(x))}{\delta'(P_k(x))P_k'(x)} \tag{6.54}$$

$$\Theta(x) = -\frac{xP_k'(x) + 1 - P_k(x)}{P_k'(x)}. \tag{6.55}$$

Now, it is shown in Appendix F.8 how to deduce a simple and efficient algorithm from the KKT characterization Eqs. (6.51)-(6.53). The steps are summarized in Algorithm 6.6.

Algorithm 6.6: KKT based algorithm (KBA) for Problem 6.5.

Set $\lambda = 0$, and $\forall k$,

$\gamma_k(0) = \max\left\{\dfrac{\eta_k^{(0)}}{m_k R_k(1 - P_k(\Theta^{-1}(0)))}, \dfrac{\delta(P_k(\Theta^{-1}(0)))}{d_k^{(0)}}\right\}$,

$Q_k(0) = \dfrac{\gamma_k}{G_k}\Theta^{-1}(0)$

if $\sum_{k=1}^{K} \gamma_k(0) < 1$ **then**
  | Exit

else

  | $\mathbb{K}_M = \{k \in \{1, \dots, K\} \mid d_k^{(0)} \geq 1/\eta_k^{(0)}\}$ and $\mathbb{K}_\Theta = \{1, \dots, K\} \backslash \mathbb{K}_M$
  | **while** $\sum_{k=1}^{K} \gamma_k(\lambda) > 1$ **do**

    | **1.** $\forall k \in \mathbb{K}_M$, compute

    | $\gamma_k(\lambda) = \dfrac{\eta_k^{(0)}}{m_k R_k(1 - P_k(\Theta^{-1}(\lambda G_k)))}$, and
    | $Q_k(\lambda) = \dfrac{\gamma_k(\lambda)}{G_k}\Theta^{-1}(\lambda G_k)$.

    | **2.** $\forall k \in \mathbb{K}_\Theta$, compute
    | $m_\lambda = M^{-1}(\lambda G_k)$, $\theta_\lambda = \Theta^{-1}(\lambda G_k)$

    | **if** $\delta(P_k(m_\lambda))(1 - P_k(m_\lambda)) > \dfrac{\eta_k^{(0)} d_k^{(0)}}{m_k R_k}$ **then**
      | $\gamma_k(\lambda) = \delta(P_k(m_\lambda))/d_k^{(0)}$, and $Q_k(\lambda) = \dfrac{\gamma_k(\lambda)}{G_k}m_\lambda$.

    | **else if** $\delta(P_k(\theta_\lambda))(1 - P_k(\theta_\lambda)) < \dfrac{\eta_k^{(0)} d_k^{(0)}}{m_k R_k}$ **then**
      | $\gamma_k(\lambda) = \eta_k^{(0)}/(m_k R_k(1 - P_k(\theta_\lambda)))$, and $Q_k(\lambda) = \dfrac{\gamma_k(\lambda)}{G_k}\theta_\lambda$.

    | **else**

      | Compute the root $x^* \in (0, 1)$ of Eq. (F.62).
      | $\gamma_k(\lambda) = \delta(x^*)/d_k^{(0)}$, and $Q_k(\lambda) = \dfrac{\gamma_k(\lambda)}{G_k}P_k^{(-1)}(x^*)$.
    | **end**
    | **3.** Increase λ
  | **end**
end

6.5.4   Ping-Pong algorithm (PPA)

The second algorithm consists in rewriting Problem 6.5 as a function of (γ, E). Then, the function to minimize becomes biconvex and solutions can be easily obtained on each direction as done in Section 6.4.

Problem 6.6. *Problem 6.5 is equivalent to:*

$$\min_{(\gamma, E)} \sum_{k=1}^{K} \gamma_k E_k, \tag{6.56a}$$

$$s.t. \quad \tilde{\eta}(\gamma_k, E_k) \geq \eta_k^{(0)}, \ \forall k, \tag{6.56b}$$

$$(1/\gamma_k)\delta(P_k(G_k E_k)) \leq d_k^{(0)}, \ \forall k, \tag{6.56c}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{6.56d}$$

$$\gamma_k > 0, \ E_k > 0, \ \forall k. \tag{6.56e}$$

Indeed, assuming fixed γ (resp. E) the problem is linear in E (resp. γ). More precisely, we split Problem 6.6 into two subproblems which are solved alternately.

Problem 6.6a (on E). *For fixed $\gamma^{(i-1)}$, the subproblem at iteration i is:*

$$E^{(i)} = \arg\min_{E} \sum_{k=1}^{K} \gamma_k^{(i-1)} E_k, \tag{6.57a}$$

$$s.t. \quad E_k \geq (1/G_k)P_k^{-1}\Big(1 - \eta_k^{(0)}/(m_k R_k \gamma_k^{(i-1)})\Big), \forall k, \tag{6.57b}$$

$$E_k \geq (1/G_k)P_k^{-1}(\delta^{-1}(\gamma_k^{(i-1)} d_k^{(0)})), \forall k. \tag{6.57c}$$

Problem 6.6b (on γ). *For fixed $E^{(i-1)}$, the subproblem at iteration i is:*

$$\gamma^{(i)} = \arg\min_{\gamma} \sum_{k=1}^{K} \gamma_k E_k^{(i-1)}, \tag{6.58a}$$

$$s.t. \quad \gamma_k \geq \max\Big(\eta_k^{(0)}/\big(m_k R_k(1 - P_k(G_k E_k^{(i-1)}))\big), \delta\big(P_k(G_k E_k^{(i-1)})\big)/d_k^{(0)}\Big), \forall k, \tag{6.58b}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1. \tag{6.58c}$$

The solution $E^{(i)}$ of the i-th iteration of Problem 6.6a is given by:

$$E_k^{(i)} = (1/G_k)\max\Big(P_k^{-1}\big(1 - \eta_k^{(0)}/(m_k R_k \gamma_k^{(i-1)})\big), P_k^{-1}\big(\delta^{-1}(\gamma_k^{(i-1)} d_k^{(0)})\big)\Big). \tag{6.59}$$

The solution $\gamma^{(i)}$ to Problem 6.4b can be efficiently obtained by using linear programming tool, for instance the Simplex method [Boyd and Vandenberghe, 2004]. The iterative algorithm is summarized in Algorithm 6.7. We initialize the bandwidth occupation terms by assuming infinite energy consumption. We remind, we are not able to guarantee the optimality of this algorithm.

Algorithm 6.7: Ping-Pong algorithm (PPA) for Problem 6.5.

Initialize $\epsilon > 0$, $E_k^{(0)} = \infty$,

$$\gamma_k^{(0)} = \frac{\max\left(\eta_k^{(0)}/(m_k R_k), 1/d_k^{(0)}\right)}{\sum_{k'=1}^{K} \max\left(\eta_{k'}^{(0)}/(m_{k'} R_{k'}), 1/d_{k'}^{(0)}\right)}, \tag{6.60}$$

and

$$E_k^{(1)} = (1/G_k) \max\left(P_k^{-1}\left(1 - \eta_k^{(0)}/(m_k R_k \gamma_k^{(0)})\right), P_k^{-1}\left(\delta^{-1}(\gamma_k^{(0)} d_k^{(0)})\right)\right). \tag{6.61}$$

Initialize $i = 1$

while $\|E^{(i)} - E^{(i-1)}\|_\infty > \epsilon$ **do**

$\quad E^{(i)} \longleftarrow$ Eq. (6.59)

$\quad \gamma^{(i)} \longleftarrow$ solution to iteration i of Problem 6.4b

$\quad i \longleftarrow i + 1$

end

6.5.5 Numerical results

For the same reasons exposed in Section 6.4.6, we only present numerical results for the MCS1 with BPSK described in Section 6.3.5, and the path loss follows the free-space model. The delay will be measured in seconds by multiplying d_k^{MAC} and $d_k^{(0)}$ by the time for transmitting a single MAC packet of $L_{\mathrm{MAC}}/\log_2(M)$ symbols over N_c OFDM subcarriers [4], given by:

$$\tau := \frac{L_{\mathrm{MAC}}}{\log_2(M)W}. \tag{6.62}$$

For $W = 1$ MHz and since BPSK is considered, we have $\tau = 128\ \mu$s. Therefore, $d_k^{\mathrm{MAC}}\tau$ corresponds to the average time for receiving the data packet without errors when transmitting over $\gamma_k N_c$ subcarriers.

In Fig. 6.20 (resp. Fig. 6.21), we plot the total transmit power (resp. the occupied bandwidth) versus the total goodput demand, for two delay requirements: $d_k^{(0)}\tau = 1024\ \mu$s $\approx$ 1 ms or $d_k^{(0)}\tau = 2560\ \mu$s $\approx$ 2.5 ms, for all k. As expected, at larger goodput requirement (basically beyond 0.5 bit/s/Hz) the performance are driven by the goodput constraints, which define the convex problem of Section 6.3, and the two algorithms 6.6 and 6.7 have the same performance (bandwidth saturation and comparable power consumption). However, for lower goodput requirement the algorithms behavior are very different since driven by the delay constraints. As expected, a larger delay constraint ($d_k^{(0)}\tau = 2.5$ ms) is less restrictive and so consumes less bandwidth and less power.

Let us recall from Appendix F.8 that the delay constraint is never active when $\eta_k^{(0)}/(m_k R_k) \geq 1/d_k^{(0)}$, $\forall k$, and so the nonconvex delay constraint can be deleted from Problem 6.5, which is therefore convex and KKT becomes optimal. When $d_k^{(0)}\tau = 1$ ms,

---

[4]$T_s = N_c/W$ is the symbol duration (in seconds) over one subcarrier.

the KBA is optimal as soon as η_T is larger than $\sum_{k=1}^{K} 1/d_k^{(0)} = K/8 = 0.5$ bit/s/Hz as observed in Fig. 6.20. When $d_k^{(0)}\tau = 2.5$ ms, optimality of the KBA is guaranteed as soon as η_T is larger than 0.2 bit/s/Hz. On the other hand, the proposed PPA outperforms the KBA for $d_k^{(0)}\tau = 1$ ms when the users are not in the convex area.
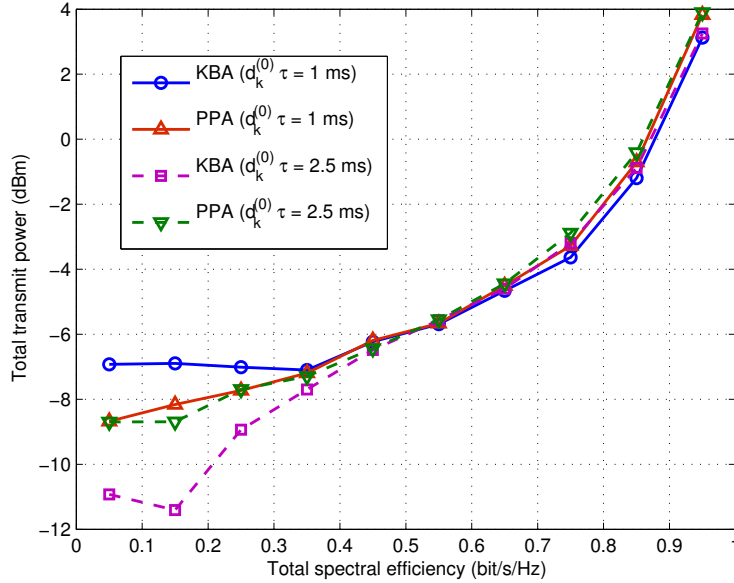
Figure 6.20: Total transmit power versus spectral efficiency (MCS1, BPSK, $K = 4$, $\tau = 128\ \mu$s).

6.6   Conclusion

The study of Type-I HARQ-based OFDMA network resource allocation with statistical CSI at the cluster head ends with this last Chapter. Upon building the Type-I HARQ using practical MCS, we have proposed a general framework for:

- power and bandwidth allocation for minimizing the total power emitted by the cluster, when the links are subject to minimum rate constraints (Problem 4.3, solved by Theorem 6.3 and Algorithm 6.1),

- smart MCS selection among predefined MCS for total power minimization under rate constraints (Problem 6.2a solved by Algorithm 6.2),

- power and bandwidth allocation for minimizing the total power emitted by the cluster, when the links are subject to minimum rate as well as maximum PER constraints (Problem 4.4 solved by Theorem 6.9 and Algorithm 6.3),
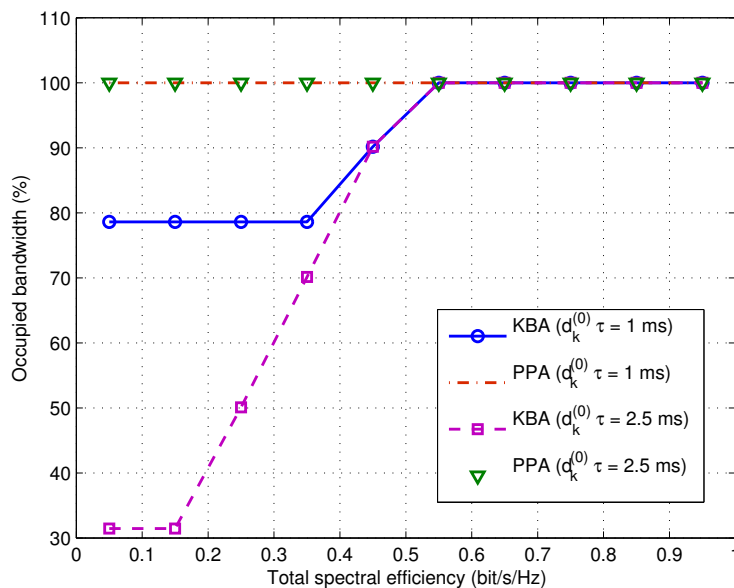
Figure 6.21: Occupied bandwidth versus spectral efficiency (MCS1, BPSK, $K = 4$, $\tau = 128\ \mu$s).

- power and bandwidth allocation for minimizing the total power emitted by the cluster, when the links are subject to minimum rate as well as maximum delay constraints (Problem 4.5 suboptimally solved by Algorithms 6.6 and 6.7).

The numerical results were computed using QAM modulated signals associated with convolutional coded packets, with a free-space path loss propagation model (squared distance $1/D^2$ decay), a carrier frequency $f_c = 400$ MHz, and within a bandwidth $W = 1$ MHz, illustrate a military context.

The main novelty of this framework, greatly influenced by some new trends in the related literature (especially [Devillers et al., 2008] and [Wu and Jindal, 2011], or [Rodriguez, 2003]), is the abandonment of the celebrated Shannon capacity (as used in [Gault et al., 2007]) as the measure of information rate. Instead, we have considered the so-called **goodput** as the figure of merit for information rate, as well as measurable performance metrics for the constraints (goodput, PER, delay).

Moreover, one important point of the proposed method is that the results are general enough to handle any practical MCS for which its performance (PER versus SNR) are available through simulation points only.

Finally, these results have been published in [C4], [C5] and [J2].

Conclusions and Perspectives

The work carried out in this thesis dealt with the resource allocation of HARQ-based OFDMA clustered MANETs, which are a flexible solution for fast and short-lived communications deployment for military applications or future smart networks. Industrial concerns about the realism of the proposed solutions led us to consider practical and existing communications schemes instead of the commonly used capacity tools. In particular, the main objective was to design and analyze algorithms that optimize the assignment of power, bandwidth, modulation order, and code rate, of the HARQ mechanism at the top of the proposed multiuser communication scheme. Due to the presence of HARQ in the proposed network, parts of the thesis were devoted to the extension of the analysis of HARQ performance done in [Le Duc, 2009] in some particular new contexts.

The fundamentals about HARQ that are useful throughout the thesis were presented in Chapter 1. Without being exhaustive, the state of the art exposed in this Chapter covers a large amount of the retransmission techniques from the basic concepts of ARQ to more advanced cross-layer HARQ techniques. We reviewed the main works related to the study of Type-I and Type-II HARQ, and the different retransmission protocols of ARQ (SW, GBN, and SR). We defined the metrics of interest for the study of HARQ performance: the PER, the delay, and the efficiency. We discussed on the relation between the throughput, the goodput, and the efficiency. Finally, we reviewed the HARQ performance study within the cross-layer concept.

In Chapter 2 we studied an ED based version of the cross-layer HARQ with IBS. The mechanism of the ED technique was exposed, and we saw that ED modifies only the HARQ efficiency metric. Then, we developed a new expression of efficiency that is valid for any HARQ type with equal MAC packet length, and derived a closed-form expression for the Type-I HARQ case. The numerical analysis revealed that the efficiency is only slightly improved when ED is used. We concluded that the ED improvement was more helpful when the fragmentation N and the total credit C are close together.

Chapter 3 was devoted to the analysis of cross-layer HARQ schemes with imperfect feedback. We proposed a model for two kinds of feedback impairments, *i.e.* errors

in the acknowledgment messages and delayed feedback, and we derived new analytic expressions at IP level for the PER, the delay and the efficiency of the HARQ schemes. The impact of such imperfect feedback on the HARQ performance was studied through numerical examples. While the PER is not modified by any feedback imperfection in the FBS case, the PER of IBS is dramatically degraded, and thus the impact on the two other metrics (delay and efficiency) is significant in this case. The best performance of the FBS scheme are achieved by using coding inside the feedback packets and by setting a time-out value that is close to the average arrival time of the feedback. In contrast, a larger time-out value must be chosen in the IBS case, and even coding cannot retrieve the ideal performance. Therefore we defined the RCS scheme, for which the analysis was conducted within a unified framework. Numerical results revealed that the choice of the initial credit distribution offered a soft transition from the robustness of FBS against imperfect feedback, to the cross-layer gain brought by IBS.

Chapter 4 moved towards the second part of the thesis and was dedicated to the resource allocation issue in the paradigm of ad hoc networks. We reviewed the different causes to the design choice of the considered MANET and discussed the main assumptions concerning this system, which led to a statistical channel knowledge only. Furthermore, several practical limitations were envisaged to make the design more realistic. Precisely, in this Chapter we established the framework for resource allocation with statistical CSI in a Type-I HARQ-based OFDMA clustered MANET, with either finite-length Gaussian codes or practical MCS. The main objective was the minimization of total cluster power based on the HARQ performance metrics as figures of merit. Finally, the mathematical formulations of the minimization problem have been led to Problems 4.2, 4.3, 4.4, and 4.5.

In Chapter 5 we solved the case of finite-length Gaussian codes formalized in Problem 4.2. We firstly computed in closed-form the error probability of Gaussian codes with finite length over the Rayleigh channel. Based on this new result, we were able to find the optimal power and bandwidth allocation given in Algorithm 5.1, which gives the best performance that one can expect from our clustered OFDMA network using Type-I HARQ. Numerical results revealed that the ergodic capacity limit can be outperformed when the data rate request are low (basically, below 500 kbit/s within a bandwidth $W = 1$ MHz), which was explained by the ability of our algorithm to save the bandwidth. The framework developed in this Chapter can serve as a basis for Type-I HARQ based OFDMA resource allocation when powerful FEC coding is used: the performance of LDPC codes can be assessed by applying a power penalty to the algorithm output (for instance, 1.9 dB for the 1/2-rate irregular LDPC code of length $n = 504$ on the Rayleigh channel).

In Chapter 6 we solved the case of practical MCS formalized in Problems 4.3, 4.4, and 4.5. We proposed a general framework for:

- power and bandwidth allocation for minimizing the total power emitted by the cluster, when the links are subject to minimum rate constraints (Problem 4.3, solved by Algorithm 6.1),

- smart MCS selection among predefined MCS for total power minimization under rate constraints (Problem 6.2a solved by Algorithm 6.2),

- power and bandwidth allocation for minimizing the total power emitted by the cluster, when the links are subject to minimum rate as well as maximum PER constraints (Problem 4.4 solved by Algorithms 6.3, 6.4 and 6.5),

- power and bandwidth allocation for minimizing the total power emitted by the cluster, when the links are subject to minimum rate as well as maximum delay constraints (Problem 4.5 suboptimally solved by Algorithms 6.6 and 6.7).

The main novelty of this framework, greatly influenced by some new trends in the related literature, is the consideration of the goodput as the figure of merit. The numerical results have been computed, without loss of generality, for QAM modulated signals associated with convolutional coded packets. It has been shown that the performance of the 1/2-rate convolutional code of length $n = 512$ were within 5 dB of the optimal finite-length Gaussian performance.

Perspectives

The part of this thesis devoted the resource allocation has raised up several issues that would deserve to be treated in future research.

To begin with, the frameworks developed in this thesis for the resource allocation could be extended to more general systems:

- In Chapters 4, 5, and 6, we assumed Type-I HARQ for facilitate the resolution of the optimization problems. Using instead Type-II HARQ would lead to better performance and would be closer to practical systems. However the closed-form expressions of the metrics will change and our results do not hold anymore. These expressions actually lead to a more complicate optimization problem.

- The proposed framework solves optimization problems whose the QoS were defined at the MAC level. Extension from MAC level defined constraints to IP level defined constraints could be of interest. In the case of HARQ with FBS, this extension is straightforward since the IP level metrics are trivial functions of the MAC level metrics. However, the case of HARQ with IBS or RCS would be more difficult given the metrics expressions, and it could be interesting to know if this strategy can outperform the power minimization based on the FBS metrics.

- OFDM signals were used assuming perfect timing and frequency synchronization. Relaxing this assumption would lead to a more realistic model, and it could be interesting to evaluate the performance losses incurred by some desynchronization. This study is left to future research since desynchronization would lead to a loss of orthogonality amongst the OFDM subcarriers which thus reintroduces multiuser interference inside the cluster. The optimization problems will thus be strongly modified since the SNR has to be replaced with the SINR depending on the resource allocation.

Adding some instantaneous CSI at the Transmitter side would provide other interesting extensions. For instance one could extend the work of [Szczecinski et al., 2011] for outdated CSI to multiuser schemes. Secondly, merging the expected goodput concept used in [Stupia et al., 2012] with predictions based on statistical CSI can be of great interest in order to enhance the performance of each HARQ round.

Finally, several more long-term perspectives could be envisaged: i) since the objectives of system designer associated with a specific application could be multiple (for instance, increasing the data rate and minimizing the delay simultaneously), multi-criteria optimization can be useful and should be analyzed in the future. ii) The multiuser scheme depicted in Chapter 4 (Fig. 4.1) leads to centralized optimization problems via the common resource allocator (actually, the cluster head). In order to be robust against the resource allocator (CH) failure (especially in military context), it would be of interest to develop decentralized allocation schemes using distributed optimization based on consensus-like approach.

Appendix A

Appendix related to Chapter 1

A.1 Proposition A.1

Proposition A.1. $\forall n, p \in \mathbb{N}$,

$$\sum_{k=0}^{p} \binom{n+k}{n} = \binom{p+n+1}{n+1}.$$ (A.1)

Proof. Let $n, p \in \mathbb{N}$ be any integers. $\forall k, 1 \leq k \leq p$, the Pascal triangle formula gives:

$$\binom{n+k+1}{n+1} = \binom{n+k}{n+1} + \binom{n+k}{n}.$$ (A.2)

From this, one finds:

$$\sum_{k=1}^{p} \binom{n+k}{n} = \sum_{k=1}^{p} \binom{n+k+1}{n+1} - \sum_{k=1}^{p} \binom{n+k}{n+1}$$

$$= \sum_{k=1}^{p} \binom{n+k+1}{n+1} - \sum_{k=0}^{p-1} \binom{n+k+1}{n+1},$$ (A.3)

and it remains after straightforward simplifications:

$$\sum_{k=1}^{p} \binom{n+k}{n} = \binom{p+n+1}{n+1} - \binom{n+1}{n+1}.$$ (A.4)

Yet,

$$\sum_{k=0}^{p} \binom{n+k}{n} = \binom{n}{n} + \sum_{k=1}^{p} \binom{n+k}{n}$$

$$= \binom{n}{n} + \binom{p+n+1}{n+1} - \binom{n+1}{n+1},$$ (A.5)

where Eq. (A.4) has been used to obtain the second line. Finally, the result comes after recalling that $\forall n > 0$, $\binom{n}{n} = 1$. $\qquad\square$

A.2 Proposition A.2

Proposition A.2. *Let $n, p \in \mathbb{N}$ and $f : [0, 1] \longrightarrow \mathbb{R}_+$ be the function defined by:*

$$f(x) = \sum_{k=0}^{p} \binom{n+k}{n} x^{n+k+1}. \tag{A.6}$$

Then, for $\mu = \binom{n+p+1}{n+1}$, f satisfies the differential equation:

$$(E_\mu) \quad y - \frac{x(1-x)}{n+1} y' = \mu \, x^{n+p+2}. \tag{A.7}$$

Proof. Starting from the Pascal triangle relation, we have $\forall k, 0 < k \le p$ and $\forall x \in [0, 1]$:

$$\sum_{k=1}^{p} \binom{n+k+1}{n+1} x^{n+k+1} = \sum_{k=1}^{p} \binom{n+k}{n+1} x^{n+k+1} + \sum_{k=1}^{p} \binom{n+k}{n} x^{n+k+1}$$

$$\Leftrightarrow x^{n+1} + \sum_{k=1}^{p} \binom{n+k+1}{n+1} x^{n+k+1} = \sum_{k=0}^{p-1} \binom{n+k+1}{n+1} x^{n+k+2} + x^{n+1} \sum_{k=1}^{p} \binom{n+k}{n} x^{n+k+1}$$

$$\Leftrightarrow x \sum_{k=0}^{p} \binom{n+k+1}{n+1} x^{n+k} = x^2 \sum_{k=0}^{p} \binom{n+k+1}{n+1} x^{n+k} - \binom{n+p+1}{n+1} x^{n+p+2} + \sum_{k=0}^{p} \binom{n+k}{n} x^{n+k+1}. \tag{A.8}$$

Denoting by $f(x) = \sum_{k=0}^{p} \binom{n+k}{n} x^{n+k+1}$, then $\forall x \in [0, 1]$:

$$f'(x) = \sum_{k=0}^{p} \binom{n+k}{n} (n+k+1) x^{n+k}$$

$$= (n+1) \sum_{k=0}^{p} \binom{n+k+1}{n+1} x^{n+k}. \tag{A.9}$$

Thus, by expressing the binomial sums using exclusively f and f' we find:

$$\frac{x}{n+1} f'(x) = \frac{x^2}{n+1} f'(x) - \binom{n+p+1}{n+1} x^{n+p+2} + f(x). \tag{A.10}$$

Lastly, $\mu := \binom{n+p+1}{n+1}$ and then Eq. (A.10) $\Leftrightarrow$ (E_μ). $\square$

A.3 Proposition A.3

Proposition A.3. $\forall \mu \in \mathbb{R}$, *the solutions of (E_μ) are the functions $f_\mu : (0, 1) \longrightarrow \mathbb{R}$ defined as:*

$$f_\mu(x) = \left(\frac{x}{1-x} \right)^{n+1} \left(\lambda - (n+1)\mu B(x; p+1, n+1) \right), \quad \forall \lambda \in \mathbb{R}, \tag{A.11}$$

where $B(x; a, b)$ is the incomplete Beta function.

Proof. Direct resolution using classic tools from linear differential equations theory. $\square$

A.4 Proof of Eq. (1.30)

We recall from Chap. 1 that $P_{\text{IP}}^{\text{I}} = 1 - (1 - p_0)^N \sum_{\ell=N}^{C} \binom{\ell-1}{N-1} p_0^{\ell-N}$, thus for $n = N - 1$ and $p = C - N$, one finds:

$$P_{\text{IP}}^{\text{I}} = 1 - \left(\frac{1-p_0}{p_0}\right)^N f(p_0). \tag{A.12}$$

From Prop. A.2, $f(p_0)$ is a solution of (E_μ) over [0,1] for $\mu = \binom{C}{N}$, and using Prop. A.3, $\forall p_0 \in (0,1)$:

$$P_{\text{IP}}^{\text{I}} = 1 - \lambda + N\mu B(p_0; C - N + 1, N), \quad \mu, \lambda \in \mathbb{R}. \tag{A.13}$$

Yet, the function P_{IP}^{I} is uniquely determined using the two limiting conditions:

$$\lim_{p_0 \to 0} P_{\text{IP}}^{\text{I}} = 0, \tag{A.14}$$

$$\lim_{p_0 \to 1} P_{\text{IP}}^{\text{I}} = 1. \tag{A.15}$$

Since from Eq. (A.13) we have that $P_{\text{IP}}^{\text{I}} \to (1-\lambda)$ when $p_0 \to 0$, then from Eq. (A.14) we find $\lambda = 1$, hence $P_{\text{IP}}^{\text{I}} = N\mu B(p_0; C-N+1, N)$ for $\mu \in \mathbb{R}$. Similarly, since $P_{\text{IP}}^{\text{I}} \to N\mu B(C-N+1, N)$ when $p_0 \to 1$, from Eq. (A.15) we find $\mu = 1/(NB(C-N+1, N))$. Finally, remark that:

$$\mu = \frac{1}{NB(C-N+1,N)} = \frac{\Gamma(C+1)}{N\Gamma(C-N+1)\Gamma(N)} = \frac{C!}{N!(C-N)!} = \binom{C}{N} \tag{A.16}$$

to conclude that $P_{\text{IP}}^{\text{I}} = B(p_0; C-N+1, N)/B(C-N+1, N) = I(p_0; C-N+1, N)$.

A.5 Proof of Eq. (1.32)

From Chap. 1, we rewrite $d_{\text{IP}}^{\text{I}} = ((1-p_0)/p_0)^N/(1-P_{\text{IP}}^{\text{I}}) \sum_{k=0}^{C-N} \binom{k+N-1}{N-1}(k+N) p_0^{k+N}$. Using the definition of f for $n = N - 1$ and $p = C - N$, it gives:

$$d_{\text{IP}}^{\text{I}} = \frac{1}{1-P_{\text{IP}}^{\text{I}}}\left(\frac{1-p_0}{p_0}\right)^N p_0 f'(p_0), \tag{A.17}$$

where f' is obtained using (E_μ):

$$f'(p_0) = \frac{N}{p_0(1-p_0)} f(p_0) - \frac{p_0^C}{(1-p_0)B(C-N+1,N)}, \tag{A.18}$$

where κ is defined as $\kappa := C-N+1$. Using Eq. (1.30), and that $1-I(p_0; \kappa, N) = I(1-p_0; N, \kappa)$, it comes:

$$d_{\text{IP}}^{\text{I}} = \frac{1}{I(1-p_0; N, \kappa)}\left(\frac{1-p_0}{p_0}\right)^N\left(\frac{N}{1-p_0}I(1-p_0; N, \kappa)\left(\frac{p_0}{1-p_0}\right)^N - \frac{p_0^{C+1}}{(1-p_0)B(\kappa, N)}\right). \tag{A.19}$$

The final result drops down after some algebraic computations.

Appendix B

Appendix related to Chapter 2

The purpose is to find in closed-form the term $p_0^{C-N+1} f(1-p_0)$, where f is the function defined for $x \in [0,1]$ by:

$$f(x) = \sum_{n=1}^{N} n \binom{C-N+n-1}{n-1} x^{n-1}. \tag{B.1}$$

Our approach will be to find a closed-form expression for a primitive F of f:

$$F(x) = \sum_{n=1}^{N} \binom{C-N+n-1}{n-1} x^n, \tag{B.2}$$

and then to compute its derivative. First, by using the symmetry property $\binom{n}{k} = \binom{n}{n-k}$, $\forall n \geq k$, and an index reorganization, we obtain that $F(x) = G(x)/x$ where G is defined by:

$$G(x) = \sum_{n=0}^{N-1} \binom{C-N+n}{C-N} x^{C-N+n+1}. \tag{B.3}$$

Then, by using Prop. A.2 for $n = C-N$ and $p = N-1$, the function G satisfies the following differential equation:

$$y - \frac{x(1-x)}{K} y' = \mu x^{C+1} \tag{B.4}$$

with $\mu = \binom{C}{K}$ and $K = C - N + 1$. Prop. A.3 can be used to solve Eq. (B.4), and we have:

$$G(x) = \left(\frac{x}{1-x}\right)^K (\lambda - K\mu B(x;N,K)). \tag{B.5}$$

Due to some binomial properties [Abramowitz and Stegun, 1972, Eq. (3.1.4)], we have $\mu K = 1/B(N,K)$. Thus

$$F(x) = \frac{x}{(1-x)^K} (\lambda - I(x;N,K)). \tag{B.6}$$

In order to characterize λ, let us consider P_{IP}^{I}. One can check that $P_{\text{IP}}^{\text{I}} = p_0^K/(1-p_0)F(1-p_0) = \lambda - I(1-p_0;N,K)$. When $p_0 \to 1$, we know that $P_{\text{IP}}^{\text{I}} \to 1$. Thus, $I(1-p_0;N,K) \to 0$ and we find $\lambda = 1$. Then:

$$F(x) = \frac{x}{(1-x)^K} (1 - I(x;N,K)). \tag{B.7}$$

Moreover, $1 - I(x; b, a) = I(1 - x; a, b)$ and we finally have:

$$F(x) = \frac{x}{(1 - x)^K} I(1 - x; K, N). \tag{B.8}$$

The final result drops down by taking the derivative of $F(.)$ given in Eq. (B.8).

Appendix C

Appendix related to Chapter 3

C.1   Proof of Proposition 3.2

In order that n fragments are received with success, each fragment #ℓ must be acknowledged according to the following ACK/NACK sequence: first, the reception of $s_\ell \geq 0$ NACKs coming from NACKs (and corresponding to the number of decoding failures), then the reception of $t_\ell \geq 0$ NACKs coming from ACKs (and corresponding to the number of feedback errors) and possibly one ACK. Thus, for n fragments for which $0 \leq k \leq n$ ACKs are received, this enumeration leads to the total of $\sum_{\ell=1}^{n}(s_\ell + t_\ell) = i - k$ transmissions. Hence, the expected transmission time $\alpha_n(i)$ is distributed over the combinatorial set:

$$\mathcal{E}_{p,n}^{B} = \left\{ (\boldsymbol{s}, \boldsymbol{t}) \in \mathbb{N}^n \times \mathbb{N}^n \,|\, \sum_{\ell=1}^{n}(s_\ell + t_\ell) = p \text{ and } \forall \ell, \sum_{m=1}^{\ell}(1 + s_m + t_m) \leq \sum_{m=1}^{\ell} L_m^{(0)} \right\}. \quad \text{(C.1)}$$

It is written (simply by enumeration):

$$\alpha_n(i) = \sum_{k=0}^{n} \sum_{(\boldsymbol{s},\boldsymbol{t}) \in \mathcal{E}_{i-k,n}^{B}} \left((\sum_{\ell} s_\ell) \tau_{r,n} + k\tau_a + (\sum_{\ell} t_\ell)\tau_{r,a} \right) \Pr\{(\boldsymbol{s},\boldsymbol{t})\}. \quad \text{(C.2)}$$

Since fragment #ℓ is received in $1 + s_\ell + t_\ell$ transmissions, but only k ACKs have been received during the transmission of the n fragments, one finds $\Pr\{(\boldsymbol{s},\boldsymbol{t})\} = (1 - p_{\text{fb}})^k \prod_{\ell=1}^{n} p_1(1 + s_\ell) p_{\text{fb}}^{t_\ell}$. The final expression is straightforward.

C.2   Proof of Proposition 3.3

Direct enumeration gives, for $n \geq 1$ and $i \geq n$:

$$\beta_n(i) = \sum_{\boldsymbol{x} \in \chi_{i,n}} \prod_{j=1}^{n} \Pr\{\text{fragment } \#j \text{ received in } x_j\}, \quad \text{(C.3)}$$

where $\chi_{k,n} = \left\{ \boldsymbol{x} \in \mathbb{N}_*^n \,|\, \sum_{\ell=1}^{n} x_\ell = k \text{ and } \forall \ell, \sum_{m=1}^{\ell} x_m \leq \sum_{m=1}^{\ell} L_m^{(0)} \right\}$, with $A_j = \{\sum_{m=1}^{j} x_m < \sum_{m=1}^{j} L_m^{(0)}\}$. The event {fragment #$j$ received in x_j} is split as follows: k_j transmissions until

correct decoding of the fragment (at the receiver side), next $(x_j - k_j)$ transmissions for correct ACK reception at the transmitter side. It does not matter if NACK has still been received at the $(x_j - k_j)$-th transmission since the fragment has been correctly decoded. This leads to:

$$\Pr\left\{\text{fragment } \#j \text{ received in } x_j\right\} = \sum_{k_j=1}^{x_j} p_1(k_j) p_{\text{fb}}^{\,x_j - k_j} (1 - p_{\text{fb}})^{\delta\{A_j\}}. \tag{C.4}$$

C.3   Proof of Proposition 3.4

The IP packet fails when at least one fragment is received with errors. At least $L_1 + (N-1)$ transmissions are needed (the first fragments fails with its total L_1, and the last $(N-1)$ fragments are received in one shot), up to C. Considering N fragments for which $0 \le n \le (N-1)$ ACKs are received, we enumerate over the set $\mathcal{E}_{i-n,n}^B$:

$$\theta_n(i) = \sum_{(s,t)\in\mathcal{E}_{i-n,n}^B} \left(\left(\sum_\ell s_\ell\right)\tau_{r,n} + n\tau_a + \sum_\ell t_\ell\right)\tau_{r,a}\right) \Pr\{(s,t)\}. \tag{C.5}$$

There may be a failure for fragment $\#\ell$ only if its credit is consumed, *i.e.*, if $\sum_{m=1}^\ell (1 + s_m + t_m) = \sum_{m=1}^\ell L_m^{(0)}$. Therefore, let us define:

$$B_\ell = \left\{ \sum_{m=1}^\ell (1 + s_m + t_m) \le \sum_{m=1}^\ell L_m^{(0)} \right\}, \tag{C.6}$$

$$\Gamma_\ell = \left\{ \sum_{m=1}^\ell (1 + s_m + t_m) = \sum_{m=1}^\ell L_m^{(0)} \right\}, \tag{C.7}$$

and $\text{KO} = \{\exists \ell \in \{1,\dots,N\} \mid \Gamma_\ell\}$, and thus:

$$\Pr\{(s,t)\} = (1 - p_{\text{fb}})^n \delta\{\text{KO}\} \prod_{\ell=1}^N \left(p_1(1 + s_\ell) p_{\text{fb}}^{t_\ell} \delta\{B_\ell\} + q(1 + s_\ell + t_\ell)\delta\{\Gamma_\ell\} \right). \tag{C.8}$$

C.4   Proof of Proposition 3.5

Success in the IBS setting consists in the acknowledgment of n **consecutive** fragments with a total of i transmissions, $i \in \{n,\dots,p\}$. First, let us focus on $i < p$ (the case $i = p$ will be completed afterward). Prop. 3.2 when applied to IBS gives:

$$\alpha_{n,p}(i) = \sum_{(s,t)\in\mathcal{E}_{i-n,n}^B} \left(\left(\sum_\ell s_\ell\right)\tau_{r,n} + n\tau_a + \left(\sum_\ell t_\ell\right)\tau_{r,a}\right)(1 - p_{\text{fb}})^n \prod_{\ell=1}^n p_1(1 + s_\ell) p_{\text{fb}}^{t_\ell}, \tag{C.9}$$

since the transmitter must have been received n ACKs in order the n fragments be all received. Moreover, the sets $\mathcal{E}_{i-n,n}^B$ and $\mathcal{E}_{i-n,n} = \left\{ (s,t) \in \mathbb{N}^n \times \mathbb{N}^n \mid \sum_{m=1}^n (s_m + t_m) = i - n \right\}$

are equal since the condition $\sum_{m=1}^{\ell}(1 + s_m + t_m) \le \sum_{m=1}^{\ell} L_m^{(0)}$ is satisfied for all $\ell \ge 1$ in the IBS case.[1]

As a set of vectors couples, $\mathcal{E}_{i-n,n}$ can be expressed using the Cartesian product:

$$\mathcal{E}_{i-n,n} = \bigcup_{\ell=0}^{i-n} \left(\mathcal{S}_{\ell,n} \times \mathcal{S}_{i-n-\ell,n} \right), \tag{C.10}$$

where $\mathcal{S}_{\ell,n}$ are the combinatorial sets $\mathcal{S}_{\ell,n} = \left\{ s \in \mathbb{N}^n \mid \sum_{m=1}^{n} s_m = \ell \right\}$. Thus we obtain:

$$\alpha_{n,p}(i) = \sum_{\ell=0}^{i-n} \left(\ell\tau_{r,n} + n\tau_a + (i-n-\ell)\tau_{r,a} \right) \sum_{s\in\mathcal{S}_{\ell,n}} \sum_{t\in\mathcal{S}_{i-n-\ell,n}} (1-p_{\text{fb}})^n \prod_{m=1}^{n} p_1(s_m+1)p_{\text{fb}}{}^{t_m}. \tag{C.11}$$

Let us focus on the set $\mathcal{S}_{\ell,n}$, which gathers all the ways of writing any natural integer ℓ as the sum of n non-negative integers (also known as a **weak n-composition**). By use of the bijection defined by $k_m = s_m + 1$, this set is equinumerous [Stanley, 1997], hence equivalent, to the set $\mathcal{K}_{\ell+n,n} = \left\{ k \in \mathbb{N}_*^n \mid \sum_{m=1}^{n} k_m = \ell + n \right\}$ that represents all the ways of writing any integer $\ell + n > 0$ as the sum of n positive integers (n-composition). Hence,

$$\begin{aligned}
\alpha_{n,p}(i) &= \sum_{\ell=0}^{i-n} \left(\ell\tau_{r,n} + n\tau_a + (i-n-\ell)\tau_{r,a} \right) \sum_{k\in\mathcal{K}_{\ell+n,n}} \sum_{r\in\mathcal{K}_{i-\ell,n}} (1-p_{\text{fb}})^n \prod_{m=1}^{n} p_1(k_m)p_{\text{fb}}{}^{r_m-1} \\
&= \sum_{\ell=0}^{i-n} \left(\ell\tau_{r,n} + n\tau_a + (i-n-\ell)\tau_{r,a} \right)(1-p_{\text{fb}})^n \left(\sum_{k\in\mathcal{K}_{\ell+n,n}} \prod_{m=1}^{n} p_1(k_m) \right) \left(\sum_{r\in\mathcal{K}_{i-\ell,n}} p_{\text{fb}}{}^{\sum_m(r_m-1)} \right) \\
&= \sum_{\ell=0}^{i-n} \left(\ell\tau_{r,n} + n\tau_a + (i-n-\ell)\tau_{r,a} \right)(1-p_{\text{fb}})^n \left(\sum_{k\in\mathcal{K}_{\ell+n,n}} \prod_{m=1}^{n} p_1(k_m) \right) \left(\sum_{r\in\mathcal{K}_{i-\ell,n}} p_{\text{fb}}{}^{i-n-\ell} \right).
\end{aligned} \tag{C.12}$$

We recall that the set of all n-compositions of $k \in \mathbb{N}$ has cardinality $\binom{k-1}{n-1}$ [Stanley, 1997]. Furthermore, the term $\sum_{k\in\mathcal{K}_{\ell+n,n}} \prod_{m=1}^{n} p_1(k_m)$ is identified as the probability $p_n(\ell+n)$. After changing ℓ into $\ell + n$, it gives:

$$\alpha_{n,p}(i) = \sum_{\ell=n}^{i} \left((\ell-n)\tau_{r,n} + n\tau_a + (i-\ell)\tau_{r,a} \right)p_n(\ell)\binom{i-\ell+n-1}{n-1}p_{\text{fb}}{}^{i-\ell}(1-p_{\text{fb}})^n. \tag{C.13}$$

Lastly, for $i = p$ the transmitter receives $(n-1)$ ACKs and the last feedback message is ACK with probability $(1 - p_{\text{fb}})$. Hence, at the p-th transmission the transmitter waits $(1 - p_{\text{fb}})\tau_a + p_{\text{fb}}\tau_{r,a}$. Some simple algebra leads to the result.

---

[1]Indeed, if $(s, t) \in \mathcal{E}_{i-n,n}$ then $\sum_{m=1}^{n}(1 + s_m + t_m) = i \le p$, and because IBS is used one has $L_1^{(0)} = p$ and $L_\ell^{(0)} = 0$ for all $\ell > 1$. Thus, for all $\ell \le n$, $\sum_{m=1}^{\ell}(1 + s_m + t_m) \le p = \sum_{m=1}^{\ell} L_m^{(0)}$, so that $\mathcal{E}_{i-n,n} \subset \mathcal{E}_{i-n,n}^B$. The case $\mathcal{E}_{i-n,n}^B \subset \mathcal{E}_{i-n,n}$ is straightforward.

C.5 Proof of Proposition 3.6

Prop. 3.5, by setting $\tau_{r,n} = \tau_a = \tau_{r,a} = 1$ in Eq. (3.33), leads to:

$$\alpha_{n,p}(i) = (1 - p_{\text{fb}})^{n - \delta\{i=p\}} \sum_{k=n}^{i} i\binom{i - k + n - 1}{n - 1} p_n(k) p_{\text{fb}}^{i-k}. \tag{C.14}$$

Since it is the average number of transmissions when feedback is unreliable, write it $\alpha_{n,p}(i) = \sum_{k=n}^{i} i\beta_n(k)/(1 - p_{\text{fb}})^{\delta\{i=p\}}$ to obtain the result.
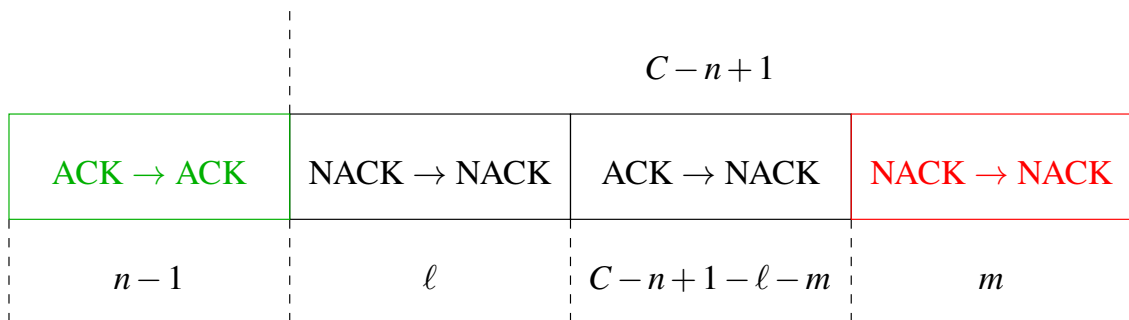
C.6 Proof of Proposition 3.7

Fig. C.1 depicts a typical event for acknowledging $(n - 1)$ fragments (green box) and not decoding the n-th fragment (red box) with C transmissions:

- $n - 1$ ACKs are received with a probability $(1 - p_{\text{fb}})^{n-1}$ leading to a waiting time $(n - 1)\tau_a$ and $C - n + 1$ transmissions left.

- $m \geq 1$ transmissions are spent for the last fragment, which is not decoded with a probability $q(m)$.

- thus, it remains at most $(C - n)$ transmissions: $\ell \geq 0$ of them are used to successfully decode the $(n-1)$ first fragments, with a probability $p_{n-1}(\ell+n-1)$, and a time $(\ell+m)\tau_{r,n}$ is spent to transmit the NACKs that are received as NACKs; the $(C - n + 1 - \ell - m)$ other transmissions are used to transmit the ACKs that are received as NACKs (with a probability $p_{\text{fb}}^{C-n+1-\ell-m}$, in a time $(C - n + 1 - \ell - m)\tau_{r,a}$.

- the $(C - n + 1 - \ell - m) \geq 0$ ACKs that are transformed into NACKs are divided into the $(n - 1)$ fragments in $\binom{C-\ell-m-1}{n-2}$ ways.

Finally, we find:

$$\theta_n(C) = \sum_{\ell=0}^{C-n} \sum_{m=1}^{C-n+1-\ell} \Big((n - 1)\tau_a + (C - n + 1 - \ell - m)\tau_{r,a} + (\ell + m)\tau_{r,n} \Big)$$
$$\times \binom{C - \ell - m - 1}{n - 2} p_{n-1}(\ell + n - 1)q(m)p_{\text{fb}}^{C-n+1-\ell-m}(1 - p_{\text{fb}})^{n-1}, \tag{C.15}$$

and changing index ℓ into $(\ell + n - 1)$ gives the desired expression. $\gamma_n(C)$ is obtained along the same lines.

$\text{ACK} \rightarrow \text{ACK}$	$\text{NACK} \rightarrow \text{NACK}$	$C-n+1$ $\text{ACK} \rightarrow \text{NACK}$	$\text{NACK} \rightarrow \text{NACK}$
$n-1$	ℓ	$C-n+1-\ell-m$	m

Figure C.1: Simple counting in $\theta_n(C)$.

Appendix D

Appendix related to Chapter 4

D.1 Proof of problem convexity

Basically, we aim to solve the problem defined in Eq. (4.19) with the individual user rate constraints driven by the links capacity:

$$C_k(\gamma_k, E_k) = \gamma_k \mathbb{E}\left[\log\left(1 + \frac{|H_k(i,n)|^2 E_k}{N_0}\right) - \frac{1}{2}\log\left(1 + \left(\frac{|H_k(i,n)|^2 E_k}{N_0 M_k}\right)^2\right)\right]. \tag{D.1}$$

Hence, the mathematical optimization problem is:

$$\min_{(\gamma, E)} \sum_{k=1}^{K} \gamma_k E_k, \tag{D.2a}$$

$$\text{s.t.} \quad C_k(\gamma_k, E_k) \geq \frac{\rho_k^{(0)}}{W}, \ \forall k, \tag{D.2b}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{D.2c}$$

$$\gamma_k \geq 0, \ E_k \geq 0, \ \forall k. \tag{D.2d}$$

Introducing the variable $Q_k := \gamma_k E_k$, the problem turns out to:

$$\min_{(\gamma, Q)} \sum_{k=1}^{K} Q_k, \tag{D.3a}$$

$$\text{s.t.} \quad C_k(\gamma_k, Q_k) \geq \frac{\rho_k^{(0)}}{W}, \ \forall k, \tag{D.3b}$$

$$\sum_{k=1}^{K} \gamma_k \leq 1, \tag{D.3c}$$

$$\gamma_k \geq 0, \ Q_k \geq 0, \ \forall k. \tag{D.3d}$$

In the light of the following Property, the previous problem falls within the class of convex minimization problems, which can be easily solved using convex optimization tools [Boyd and Vandenberghe, 2004]:

Property D.1. C_k *is concave in* (γ_k, Q_k).

Proof. Let us study the concavity of the function f defined as:

$$f(x) = \log(1 + x) - \frac{1}{2}\log\left(1 + \frac{x^2}{M^2}\right).$$ (D.4)

Deriving f twice, we find after little algebra:

$$f''(x) = \frac{-1/M^2}{1 + x^2/M^2}\left(\frac{1 - x^2/M^2}{1 + x^2/M^2}\right) - \frac{1}{(1 + x)^2}.$$ (D.5)

It is easy to see that $f''(x) \leq 0$ when $x \leq M$. However it is more difficult for $x > M$, but it can be conjectured from Fig. D.1. Moreover, the function $E_k \mapsto f\mathbb{E}\left[(E_k|H_k(i,n)|^2/N_0)\right]$ is concave since it corresponds to the average of positive concave functions. Finally, since $(\gamma_k, Q_k) \mapsto C_k(\gamma_k, Q_k)$ is the perspective of the concave function $E_k \mapsto \mathbb{E}\left[f(E_k|H_k(i,n)|^2/N_0)\right]$, then C_k is concave in (γ_k, Q_k) [Boyd and Vandenberghe, 2004].
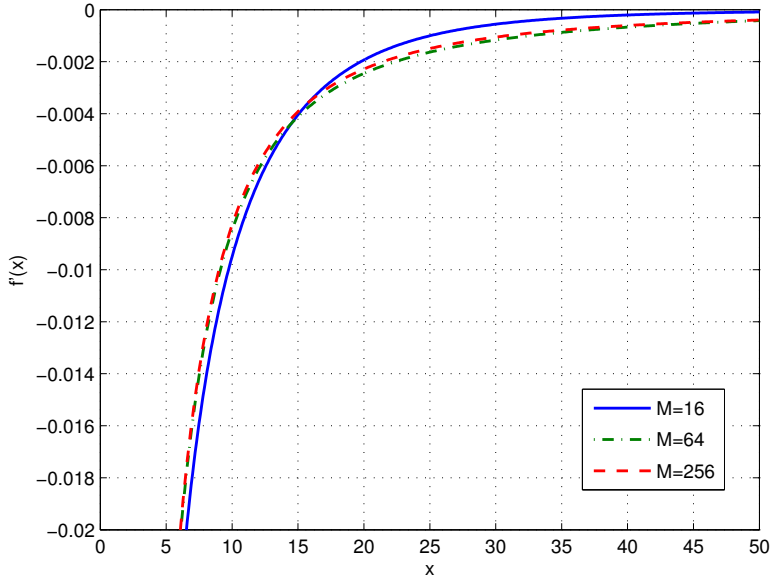

Figure D.1: Sign of f'' for several values of M.

$\square$

D.2   Solution of the convex optimization problem

The minimization problem under investigation exhibits a convex structure from Property D.1. Thus, an optimal solution is obtained by solving the KKT optimal conditions:

$$\nabla\left(\sum_{k=1}^{K} Q_k\right) - \sum_{k=1}^{K} \mu_k \nabla C_k(\gamma_k, Q_k) + \lambda\nabla\left(\sum_{k=1}^{K} \gamma_k\right) = 0.$$ (D.6)

This multivariate can be expanded as a set of $2K$ scalar equations $\mu_k \partial C_k / \partial \gamma_k = \lambda$ and $\mu_k \partial C_k / \partial Q_k = 1$. After developing the two partial derivatives of C_k:

$$C_k(\gamma_k, Q_k) = \gamma_k \left(\mathbb{E}\left[\log\left(1 + \frac{|H_k(i,n)|^2 Q_k}{N_0 \gamma_k}\right) - \frac{1}{2}\log\left(1 + \left(\frac{|H_k(i,n)|^2 Q_k}{N_0 M_k \gamma_k}\right)^2\right)\right]\right), \tag{D.7}$$

which gives, denoting by $g_k := |H_k(i,n)|^2/N_0$:

$$\frac{\partial C_k}{\partial \gamma_k} = \mathbb{E}\left[\log\left(1 + g_k \frac{Q_k}{\gamma_k}\right)\right] - \frac{1}{2}\mathbb{E}\left[\log\left(1 + \frac{g_k^2 Q_k^2}{M_k^2 \gamma_k^2}\right)\right] \tag{D.8}$$

$$- \mathbb{E}\left[\frac{g_k Q_k/\gamma_k}{1 + g_k Q_k/\gamma_k}\right] + \mathbb{E}\left[\frac{g_k^2 Q_k^2/(M_k^2 \gamma_k^2)}{1 + g_k^2 Q_k^2/(M_k^2 \gamma_k^2)}\right] \tag{D.9}$$

$$\frac{\partial C_k}{\partial Q_k} = \mathbb{E}\left[\frac{g_k}{1 + g_k Q_k/\gamma_k}\right] - \mathbb{E}\left[\frac{g_k^2 Q_k/(M_k^2 \gamma_k)}{1 + g_k^2 Q_k^2/(M_k^2 \gamma_k^2)}\right], \tag{D.10}$$

reordering and observing that $Q_k/\gamma_k = E_k$ leads to the following set of equations:

$$\mu_k \mathbb{E}\left[\log(1 + g_k E_k) - \frac{g_k E_k}{1 + g_k E_k}\right] - \frac{1}{2}\mathbb{E}\left[\log\left(1 + \frac{g_k^2 E_k^2}{M_k^2}\right) - \frac{g_k^2 E_k/M_k^2}{1 + g_k^2 E_k^2/M_k^2}\right] = \lambda \tag{D.11}$$

$$\mu_k \mathbb{E}\left[\frac{g_k}{1 + g_k E_k} - \frac{g_k^2 E_k/M_k^2}{1 + g_k^2 E_k^2/M_k^2}\right] = 1. \tag{D.12}$$

Notice the two correction terms in M_k introduced by the finite-modulation approach, compared to the set of equations that were derived in [Gault et al., 2007].

Finally, solving for the multiplier μ_k gives, for all k:

$$f_k(E_k) = \lambda, \tag{D.13}$$

where the functions f_k are defined by:

$$f_k(x) = \frac{\mathbb{E}\left[\log(1 + xg_k) - \frac{1}{2}\log(1 + x^2 g_k^2/M_k^2)\right]}{\mathbb{E}\left[\frac{g_k}{1 + xg_k} - x\frac{g_k^2/M_k^2}{1 + x^2 g_k^2/M_k^2}\right]} - x. \tag{D.14}$$

Since g_k is an exponential random variable with mean G_k, in order to compute f_k in closed-form one can compute the function f defined as:

$$f(x, M) = \frac{\mathbb{E}\left[\log(1 + xX_e) - \frac{1}{2}\log(1 + x^2 X_e^2/M^2)\right]}{\mathbb{E}\left[\frac{X_e}{1 + xX_e} - x\frac{X_e^2/M^2}{1 + x^2 X_e^2/M^2}\right]} - x, \tag{D.15}$$

where X_e is an exponential random variable with mean 1. Therefore, f is related to f_k through $f_k(x) = (1/G_k)f(G_k x, M_k)$. The numerator is known from Appendix D.3 as the capacity with M-QAM input signals:

$$\mathbb{E}\left[\log(1 + xX_e) - \frac{1}{2}\log\left(1 + \frac{x^2 X_e^2}{M^2}\right)\right] = e^{1/x}E_1(1/x) + \text{ci}\left(\frac{M}{x}\right)\cos\left(\frac{M}{x}\right) + \text{si}\left(\frac{M}{x}\right)\sin\left(\frac{M}{x}\right). \tag{D.16}$$

Now, the two remaining terms at the denominator are computed:

$$\mathbb{E}\left[\frac{X_e}{1+xX_e}\right] = \int_0^\infty \frac{t}{1+xt}e^{-t}\mathrm{d}t = \frac{1}{x^2}\int_0^\infty \frac{u}{1+u}e^{-u/x}\mathrm{d}u = \frac{1}{x^2}(-e^{1/x}E_1(1/x)+x), \quad \text{(D.17)}$$

where the last equality comes from [Gradshteyn and Ryzhik, 1980, Eq. (3.353.5)], and:

$$\begin{aligned}\mathbb{E}\left[\frac{X_e^2/M^2}{1+x^2X_e^2/M^2}\right] &= \int_0^\infty \frac{(t/M)^2}{1+(xt/M)^2}e^{-t}\mathrm{d}t \\ &= \frac{1}{x^3}\int_0^\infty \frac{u^2}{M^2+u^2}e^{-u/x}\mathrm{d}u \\ &= \frac{1}{x^3}(-M(\mathrm{ci}\,(M/x)\sin(M/x)-\mathrm{si}\,(M/x)\cos(M/x))+x), \quad \text{(D.18)}\end{aligned}$$

for which the last line is found from [Gradshteyn and Ryzhik, 1980, Eq. (3.356.2)]. Hence, combining these results into Eq. (D.15), and defining $C(x) := -\mathrm{ci}\,(x)\cos(x)-\mathrm{si}\,(x)\sin(x)$ and $S(x) := \mathrm{ci}\,(x)\sin(x)-\mathrm{si}\,(x)\cos(x)$, leads to:

$$f(x) = x^2\frac{e^{1/x}E_1(1/x)-C(M/x)}{M\,S(M/x)-e^{1/x}E_1(1/x)} - x. \quad \text{(D.19)}$$

D.3   Approximate closed-form expressions for ergodic mutual information with QAM entries

.

Let $Y \in \mathbb{C}^n$ be the channel output:

$$Y = HX + N, \quad \text{(D.20)}$$

where X and N are random vectors of length n with i.i.d. elements X_k and N_k, respectively. X_k is uniformly distributed over a discrete subset $\mathcal{X} \subset \mathbb{C}$ of M elements, and such that $\mathbb{E}\left[|X_k|^2\right] = E_s$, whereas $N_k \sim \mathcal{CN}(0,N_0)$, and H is a $n \times n$ diagonal matrix with i.i.d. elements $H_k \sim \mathcal{CN}(0,\sigma_h^2)$.

Moreover, the channel gains $|H_k|$ are Rayleigh distributed, such that a random SNR can be defined as:

$$\mathrm{SNR}_k = \frac{|H_k|^2E_s}{N_0}, \quad \text{(D.21)}$$

and is exponentially distributed with parameter $1/\overline{\mathrm{SNR}}$, where $\overline{\mathrm{SNR}} = \sigma_h^2E_s/N_0$ is the average SNR.

A nice approximation can be intuited from the observation that the capacity with discrete entries will be reduced compared to the capacity with continuous inputs. The Shannon capacity, found as the maximal mutual information shared between the source and the destination, is achieved for Gaussian inputs. Therefore a quantification of the

input symbols to $M > 4$ states, induced by the finite nature of the alphabet, leads to a capacity reduction for the M-QAM AWGN channel evaluated in [Weidong et al., 2007]:

$$C_{G,QAM} \approx \log(1 + \text{SNR}) - \frac{1}{2}\log\left(1 + \frac{\text{SNR}^2}{M^2}\right). \tag{D.22}$$

The capacity $C_{\mathcal{R},QAM}$ of the Rayleigh channel is obtained by averaging:

$$C_{\mathcal{R},QAM} = \mathbb{E}\left[\log(1 + \text{SNR})\right] - \frac{1}{2}\mathbb{E}\left[\log\left(1 + \frac{\text{SNR}^2}{M^2}\right)\right]. \tag{D.23}$$

Next, we give a closed-form expression for Eq. (D.23). Since the first term (ergodic Rayleigh capacity) is already known, it remains only to compute the second term in closed-form:

$$A := \frac{1}{2}\mathbb{E}\left[\log\left(1 + \frac{\text{SNR}^2}{M^2}\right)\right] = \frac{1}{2}\int_0^\infty \log\left(1 + \frac{t^2}{M^2}\right)\frac{e^{-t/\overline{\text{SNR}}}}{\overline{\text{SNR}}}\mathrm{d}t = \int_0^\infty \frac{t}{M^2 + t^2}e^{-t/\overline{\text{SNR}}}\mathrm{d}t, \tag{D.24}$$

where the last equality is obtained after partial integration and simplification. The value of this integral is provided in [Gradshteyn and Ryzhik, 1980, Eq. (3.354)]:

$$A = -\text{ci}\left(\frac{M}{\overline{\text{SNR}}}\right)\cos\left(\frac{M}{\overline{\text{SNR}}}\right) - \text{si}\left(\frac{M}{\overline{\text{SNR}}}\right)\sin\left(\frac{M}{\overline{\text{SNR}}}\right), \tag{D.25}$$

where $\text{ci}(x) = -\int_x^\infty \frac{\cos t}{t}\mathrm{d}t$ and $\text{si}(x) = -\int_x^\infty \frac{\sin t}{t}\mathrm{d}t$ are the cosine integral and sine integral, respectively [Gradshteyn and Ryzhik, 1980]. These functions are related to the exponential integral through $\text{ci}(x) = -(E_1(ix) + E_1(-ix))/2$ and $\text{si}(x) = (E_1(ix) - E_1(-ix))/(2i)$. Hence, inserting this into Eq. (D.24) boils down to:

$$C_{\mathcal{R},QAM} \approx e^{1/\overline{\text{SNR}}}E_1\left(1/\overline{\text{SNR}}\right) - \frac{E_1\left(iM/\overline{\text{SNR}}\right) + E_1\left(-iM/\overline{\text{SNR}}\right)}{2}\cos\left(\frac{M}{\overline{\text{SNR}}}\right)$$
$$+ \frac{E_1\left(iM/\overline{\text{SNR}}\right) - E_1\left(-iM/\overline{\text{SNR}}\right)}{2i}\sin\left(\frac{M}{\overline{\text{SNR}}}\right). \tag{D.26}$$

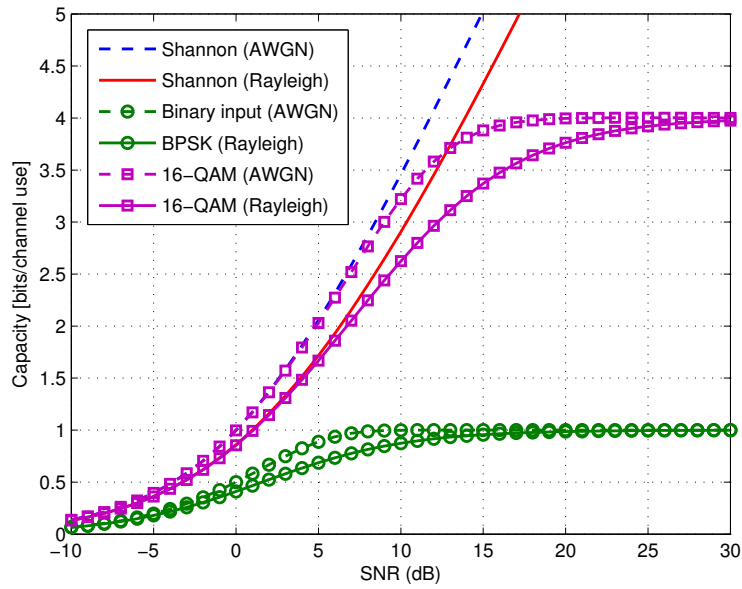The previous closed-form expressions are illustrated in Fig. D.2.

Figure D.2: Capacity with discrete entries.

Appendix E

Appendix related to Chapter 5

E.1 Proof of Lemma 5.1

From Eq. (5.31), we find that:

$$\log \tilde{\eta}_k(\gamma_k, E_k) = \log(\gamma_k) + \log(r_k) + \log(1 - P_e^{(n)}(G_k E_k)), \tag{E.1}$$

thus the goodput is clearly log-concave in $\gamma_k > 0$. The purpose is the study of the concavity of the function $f(x) := \log\left(1 - Q(u(x))\right)$, where Q is the queue of the Normal distribution, and u is the centering/scaling function given by $u(x) := \sqrt{n}(\mu(x) - R)/\sigma(x)$. In order that the logarithm to be defined, $(1 - Q)$ must be non-zero, which is guaranteed for $x > 0$.
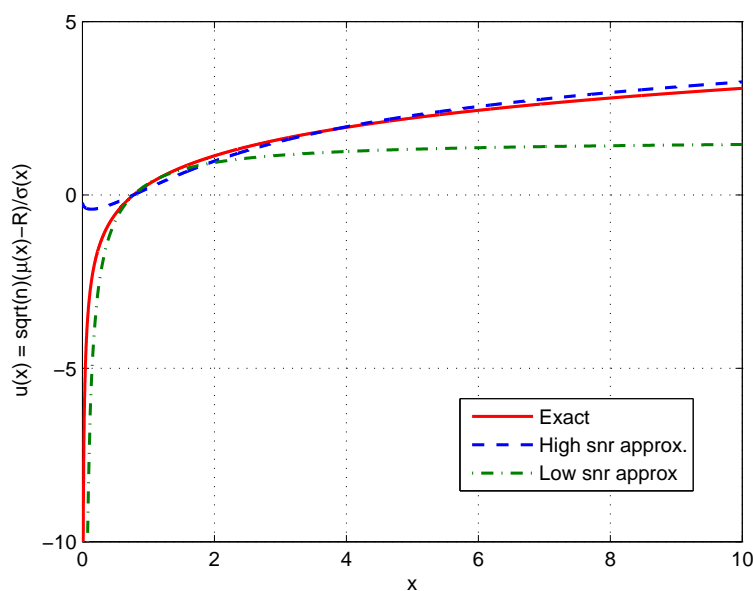

Figure E.1: Concavity of the function $u(x)$ ($n = 10$).

First, we compute the two first derivatives of f:

$$f'(x) = \frac{u'(x)e^{-u^2(x)/2}}{1 - Q(u(x))}, \tag{E.2}$$

$$f''(x) = \left(u''(x) - (u'(x))^2\right)\frac{e^{-u^2(x)/2}}{1 - Q(u(x))} - (u'(x))^2\frac{e^{-u^2(x)}}{(1 - Q(u(x)))^2}. \tag{E.3}$$

Hence, if u is concave ($u''(x) \leq 0$), then f is concave:

$$u'(x) = \sqrt{n}\,\frac{\mu'(x)\sigma(x) - \sigma'(x)(\mu(x) - R)}{\sigma^2(x)}, \tag{E.4}$$

$$u''(x) = \sqrt{n}\,\frac{\Delta(x)}{\sigma^4(x)}, \tag{E.5}$$

where the sign of u'' is given by the sign of Δ as defined below:

$$\Delta(x) := \sigma^2(x)\left(\mu''(x)\sigma(x) - \sigma''(x)(\mu(x) - R) - \mu'(x)\sigma'(x) - \mu(x)\sigma'(x)\right)$$
$$+ 2(\sigma'(x))^2\sigma(x)(\mu(x) - R), \tag{E.6}$$

and

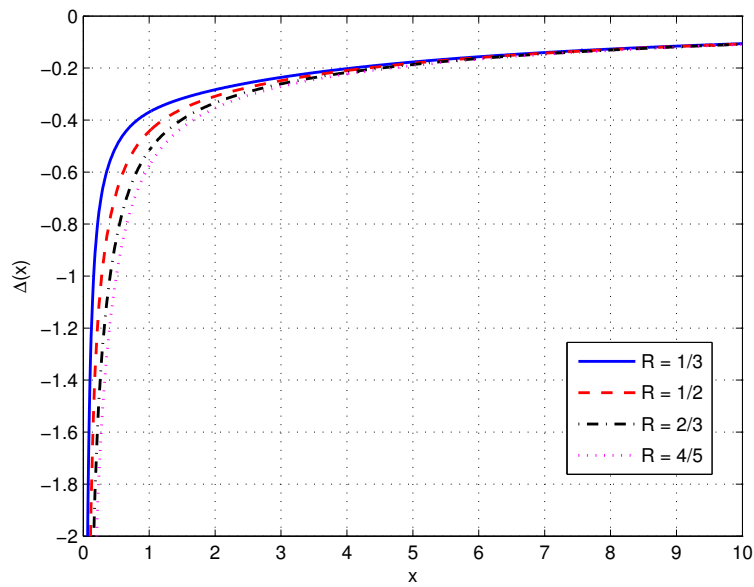$$\sigma'(x) = \frac{1}{\sqrt{(n)}}\frac{v'(x)}{2\sqrt{v(x)}}, \tag{E.7}$$

$$\sigma''(x) = \frac{1}{\sqrt{n}}\frac{2v''(x)\sqrt{v(x)} - (v'(x))^2/\sqrt{v(x)}}{4v(x)}, \tag{E.8}$$

$$v(x) = \int_0^\infty \left(\log^2(1 + xt) - \mu^2(x) + \frac{xt}{1 + xt}\right)e^{-t}dt, \tag{E.9}$$

$$v'(x) = \int_0^\infty \frac{2t}{1 + xt}\left(\log(1 + xt) - \mu(x) + \frac{1/2}{1 + xt}\right)e^{-t}dt, \tag{E.10}$$

$$v''(x) = \int_0^\infty \left(\frac{-2t^2}{(1 + xt)^2}\left(\log(1 + xt) - \mu(x) + \frac{1/2}{1 + xt}\right)\right.$$
$$\left. + \frac{2t}{1 + xt}\left(\frac{t}{1 + xt} - \mu'(x) - \frac{t/2}{(1 + xt)^2}\right)\right)e^{-t}dt. \tag{E.11}$$

Proving that $\Delta(x) \leq 0$ for all $x > 0$ is difficult, nevertheless it can be conjectured from Fig. E.2.

Figure E.2: Sign of Δ.

Appendix F

Appendix related to Chapter 6

F.1 Proof of Lemma 6.1

Only if part. If Problem 6.1 is feasible, then there is $(\gamma, Q) \in (0,1)^K \times \mathbb{R}_{+*}^K$ such that for all $k \in \{1, \dots, K\}$:

$$\begin{cases} \eta_k^{(0)} \leq \gamma_k m_k R_k f(P_k(G_k Q_k/\gamma_k)) \\ \sum_k \gamma_k \leq 1 \end{cases} \Rightarrow \eta_k^{(0)} < \gamma_k m_k R_k \qquad \text{(F.1)}$$

since $1 - P_k(G_k Q_k/\gamma_k) < 1$, hence we have:

$$\sum_{k=1}^{K} \frac{\eta_k^{(0)}}{m_k R_k} < \sum_{k=1}^{K} \gamma_k \leq 1. \qquad \text{(F.2)}$$

If part. Conversely, let us define the open set:

$$O = \left\{ \eta \in \mathbb{R}_{+*}^K \mid \sum_{k=1}^{K} \frac{\eta_k}{m_k R_k} < 1 \right\}, \qquad \text{(F.3)}$$

and thus $\mathbf{t}_0 = \eta^{(0)} \in O$. Therefore, there is an $\epsilon > 0$ such that the ball:

$$B(\mathbf{t}_0, \epsilon) = \{ \mathbf{t} \in \mathbb{R}_{+*}^K \mid \|\mathbf{t} - \mathbf{t}_0\| < \epsilon \} \subset O, \qquad \text{(F.4)}$$

where $\|.\|$ is the L_∞-norm. In particular, we have $\{ \eta \in \mathbb{R}_{+*}^K \mid \forall k \in \{1, \dots, K\}, |\eta_k - \eta_k^{(0)}| < \epsilon \} \subset B(\mathbf{t}_0, \epsilon)$. Now let us consider, $\forall k \in \{1, \dots, K\}$,

$$\gamma_k = \frac{\eta_k^{(0)} + \epsilon/2}{m_k R_k}. \qquad \text{(F.5)}$$

Since $\{\eta_k^{(0)} + \epsilon/2\} \in B(\mathbf{t}_0, \epsilon)$, we have $\sum_{k=1}^{K} \gamma_k < 1$.

Furthermore let $\tilde{\eta}_k = \gamma_k m_k R_k (1 - P_k(G_k Q_k/\gamma_k))$. When $Q_k \to \infty$, one obtains:

$$\tilde{\eta}_k \longrightarrow m_k R_k \frac{\eta_k^{(0)} + \epsilon/2}{m_k R_k} = \eta_k^{(0)} + \frac{\epsilon}{2} > \eta_k^{(0)} \qquad \text{(F.6)}$$

which concludes the proof.

F.2   Proof of Theorem 6.3

Let (μ, λ) be the non-negative Lagrangian multipliers associated with Eqs. (6.6b)-(6.6c), respectively. The KKT conditions lead to the following equality:

$$\nabla\left(\sum_{k=1}^{K} Q_k\right) - \sum_{k=1}^{K} \mu_k \nabla \tilde{\eta}_k(\gamma_k, Q_k) + \lambda \nabla\left(\sum_{k=1}^{K} \gamma_k\right) = 0, \tag{F.7a}$$

$$\mu_k(\tilde{\eta}_k(\gamma_k, Q_k) - \eta_k^{(0)}) = 0, \; \lambda\left(\sum_k \gamma_k - 1\right) = 0. \tag{F.7b}$$

where ∇ stands for the gradient operator.

Before working on the KKT equations, we compute the gradients:

$$\frac{\partial P_k}{\partial Q_k} = G_k \frac{1}{\gamma_k} P'_k(G_k Q_k/\gamma_k), \tag{F.8}$$

$$\frac{\partial P_k}{\partial \gamma_k} = -G_k \frac{Q_k}{\gamma_k^2} P'_k(G_k Q_k/\gamma_k). \tag{F.9}$$

We express the gradient of $\tilde{\eta}_k$ as:

$$\nabla \tilde{\eta}_k = m_k R_k \gamma_k f'(P_k(G_k Q_k/\gamma_k))\nabla P_k + \begin{bmatrix} 0 & m_k R_k f(P_k(G_k Q_k/\gamma_k)) \end{bmatrix}^T. \tag{F.10}$$

Stability Eq. (F.7a) is thus equivalent to the $2K$ scalar equations:

$$1 - \mu_k m_k R_k \gamma_k f'(P_k(G_k Q_k/\gamma_k))\frac{\partial P_k}{\partial Q_k} = 0, \tag{F.11}$$

$$-\mu_k m_k R_k \gamma_k f'(P_k(G_k Q_k/\gamma_k))\frac{\partial P_k}{\partial \gamma_k} - \mu_k m_k R_k f(P_k(G_k Q_k/\gamma_k)) + \lambda = 0. \tag{F.12}$$

Putting the first line into the second enables the elimination of the K multipliers $\mu_k \neq 0$, hence the optimal Q_k^* satisfies:

$$F(G_k Q_k^*/\gamma_k^*) = \lambda G_k, \tag{F.13}$$

with $F : x \mapsto -(1 - P_k(x))^2/(f(P_k(x))P'_k(x)) - x$. Moreover since $\mu_k \neq 0$, Eq. (F.7b) indicates that the goodput constraint Eq. (6.6b) must be active, which gives for the optimal γ_k^*:

$$\gamma_k^* m_k R_k f(P_k(G_k Q_k^*/\gamma_k^*)) = \eta_k^{(0)}. \tag{F.14}$$

The optimal (Q_k^*, γ_k^*) thus satisfy the set of equations (F.13)-(F.14) involving only one multiplier λ.

Finally, the optimal Lagrangian multiplier $\lambda^* \geq 0$ is found. First of all, it is straightforward to show that F is strictly increasing over $\mathbb{R}^+$, hence the inverse function F^{-1} of F with respect to the composition exists. From Eqs. (F.13)-(F.14) one finds:

$$\gamma_k^*(\lambda) = \frac{\eta_k^{(0)}}{m_k R_k f(P_k(F^{-1}(\lambda G_k)))}, \tag{F.15}$$

and from the monotonic behavior of F, f and P_k, we have that $\gamma_k^*(\lambda)$ decreases when λ increases. In contrast, when λ increases then $Q_k^*(\lambda)$ increases too. Then two cases occur for $\lambda = 0$:

(i) if $\sum_{k=1}^{K} \gamma_k^*(0) \leq 1$, then a feasible solution is found, and $\lambda^* = 0$. Otherwise, stopping at a larger λ^* would give larger Q_k^* such that $\sum_{k=1}^{K} Q_k^*(\lambda^*) > \sum_{k=1}^{K} Q_k^*(0)$, which contradicts the optimality of λ^*.

(ii) else, by increasing λ until a feasible solution is found leads to the minimum for $\lambda^* > 0$ such that $\sum_{k=1}^{K} \gamma_k^*(\lambda^*) = 1$ (for Eq. (F.7b) must be satisfied).

F.3 Calculations leading to fast implementation of Algorithm 6.1 in uncoded packet case

Based on Eq. (6.2), the PER derivative becomes:

$$P_k'(G_k E_k) = \frac{-n\, G_k}{(1 + G_k E_k)^2} = \frac{-G_k}{n} P_k^2(E_k). \tag{F.16}$$

Using Eq. (F.16), and since the PER is bijective, we can solve Eq. (F.13) with respect to the variable P_k. Thus we obtain after some algebra that P_k^* is a root of the polynomial ϕ:

$$\phi(X) = \left(\lambda - \frac{1}{G_k} - \frac{n}{G_k}\frac{p_{\text{fb}}}{1 - p_{\text{fb}}}\right) X^2 + \frac{2n}{G_k}\frac{1}{1 - p_{\text{fb}}} X - \frac{n}{G_k}\frac{1}{1 - p_{\text{fb}}}. \tag{F.17}$$

The next result guarantees that the solution P_k^* to $\phi(P_k^*) = 0$ exists and is unique:

Proposition F.1. ϕ *has exactly one root* $P_k^* \in (0,1)$. *Let us denote the constants* $a := \lambda - \frac{1}{G_k} - \frac{n}{G_k}\frac{p_{\text{fb}}}{1 - p_{\text{fb}}}$, $b := \frac{2n}{G_k}\frac{1}{1 - p_{\text{fb}}}$, $c := -b/2$, *and* $\Delta := b^2 - 4ac$. *Then,* $P_k^* = (-b + \sqrt{\Delta})/(2a)$.

Proof. Since ϕ is a polynomial of degree two, it has at most two distinct roots:

$$\Delta := b^2 - 4ac = b(b + 2a) = 2b\left(\frac{n_S - 1}{G_k} + \lambda\right) > 0.$$

Hence, ϕ has exactly two distinct roots $x_1 = (-b + \sqrt{\Delta})/(2a)$ and $x_2 = (-b - \sqrt{\Delta})/(2a)$. Let us investigate the two cases:

- If $\lambda < \frac{1}{G_k} + \frac{n}{G_k}\frac{p_{\text{fb}}}{1 - p_{\text{fb}}}$, *i.e.* $a < 0$. Then,

$$\begin{aligned}
-b + \sqrt{\Delta} &= \frac{-2n}{G_k}\frac{1}{1 - p_{\text{fb}}} + \sqrt{\frac{4n}{G_k}\frac{1}{1 - p_{\text{fb}}}\left(\frac{n-1}{G_k} + \lambda\right)} \\
&< \frac{-2n}{G_k}\frac{1}{1 - p_{\text{fb}}} + \sqrt{\frac{4n^2}{G_k^2}\frac{1}{1 - p_{\text{fb}}}\left(1 + \frac{p_{\text{fb}}}{1 - p_{\text{fb}}}\right)} \\
&= \frac{-2n}{G_k}\frac{1}{1 - p_{\text{fb}}} + \sqrt{\frac{4n^2}{G_k^2}\frac{1}{(1 - p_{\text{fb}})^2}} \\
&\leq 0.
\end{aligned}$$

Thus, $x_1 > 0$. Finally, since $b > 0$, we have $\Delta > 0 \Leftrightarrow -b < 2a$. Hence,

$$x_1 = \frac{-b + \sqrt{\Delta}}{2a} = \frac{2a - (b + 2a) + \sqrt{b(b + 2a)}}{2a} = 1 + \frac{\sqrt{b + 2a}\left(\sqrt{b} - \sqrt{b + 2a}\right)}{2a} < 1$$

since $a < 0$ and then $\sqrt{b} - \sqrt{b + 2a} > 0$. For the other solution, suppose that $x_2 < 1$, then (since $-2a > 0$):

$$b + \sqrt{\Delta} < -2a \Leftrightarrow \sqrt{b(b + 2a)} < -(b + 2a)$$
$$\Rightarrow b(b + 2a) < (b + 2a)^2$$
$$\Leftrightarrow b < b + 2a$$
$$\Leftrightarrow 0 < 2a,$$

which is in contradiction with $a < 0$. Thus, $x_2 > 1$ and this solution is impossible.

- Or suppose that $\lambda > \frac{1}{G_k} + \frac{n}{G_k}\frac{p_{\text{fb}}}{1 - p_{\text{fb}}}$, *i.e.* $a > 0$. Then,

$$-b + \sqrt{\Delta} > \frac{-2n}{G_k}\frac{1}{1 - p_{\text{fb}}} + \sqrt{\frac{4n^2}{G_k^2}\frac{1}{(1 - p_{\text{fb}})^2}} \geq 0,$$

thus $x_1 > 0$. Moreover, suppose that $x_1 > 1$. This is equivalent to:

$$-b + \sqrt{\Delta} > 2a \Leftrightarrow \sqrt{b(b + 2a)} > b + 2a$$
$$\Rightarrow b(b + 2a) > (b + 2a)^2$$
$$\Leftrightarrow b > b + 2a$$
$$\Leftrightarrow 0 > 2a,$$

which is in contradiction with $a > 0$. Therefore, $x_1 < 1$. In this case, the other solution $x_2 = -(b + \sqrt{\Delta})/(2a) < 0$ is not consistent.

$\square$

F.4   Proof of Lemma 6.7

To begin with, let us give a little recall about quasi-convexity (see [Greenberg and Pierskalla, 1971] and [Boyd and Vandenberghe, 2004, § 3.4]). A function f defined on dom f is **quasi-convex** if, and only if, its sublevel sets:

$$S_t = \{x \in \text{dom } f \mid f(x) \leq t\} \tag{F.18}$$

are convex for all $t \in \mathbb{R}$.

The following characterization of quasi-convex functions will be used in the proof. f is quasi-convex if, and only if, for all $x, y \in \text{dom } f$ and $t \in [0, 1]$, f satisfies Jensen's inequality [Boyd and Vandenberghe, 2004]:

$$f(tx + (1 - t)y) \leq \max\{f(x), f(y)\}. \tag{F.19}$$

Now, let us prove Lemma 6.7. The concavity of $\tilde{\eta}$ has already been established in Lemma 6.2. In what follows we focus on the PER function. For all $(\gamma, Q), (\gamma', Q') \in [0, 1] \times \mathbb{R}_+$ and $t \in [0, 1]$, we have:

$$\frac{tQ + (1-t)Q'}{t\gamma + (1-t)\gamma'} = s\frac{Q}{\gamma} + (1-s)\frac{Q'}{\gamma'}, \tag{F.20}$$

where $s := \frac{t\gamma}{t\gamma + (1-t)\gamma'} \in [0, 1]$, hence:

$$
\begin{aligned}
P_k\left(\frac{tQ + (1-t)Q'}{t\gamma + (1-t)\gamma'}\right) &= P_k\left(s\frac{Q}{\gamma} + (1-s)\frac{Q'}{\gamma'}\right) \\
&\overset{(a)}{\leq} sP_k(Q/\gamma) + (1-s)P_k(Q'/\gamma') \\
&\leq \max\{P_k(Q/\gamma), P_k(Q'/\gamma')\},
\end{aligned} \tag{F.21}
$$

where (a) follows from the convexity assumption on $P_k : \mathbb{R}_+ \rightarrow [0, 1]$. Thus, P_k is quasi-convex in the two variables (γ_k, Q_k).

F.5   Proof of Lemma 6.8

The set $\mathcal{F}$ is convex. Indeed, let us recall from Lemma 6.7 that P_k are quasi-convex functions. By definition, the sublevel sets of P_k denoted by:

$$S_t = \{(\gamma, Q) \in [0, 1]^K \times \mathbb{R}_+^K \mid P_k(G_k Q_k/\gamma_k) - P_k^{(0)} \leq t\}, \tag{F.22}$$

are convex for all $t \in \mathbb{R}$. In particular, S_0 is convex. Thus $\mathcal{F}$ is convex by intersection of S_0 with trivially convex sets defined by Eq. (6.31b) and Eq. (6.31d) [Boyd and Vandenberghe, 2004].

Now, we check that the nondegeneracy condition holds for Eqs. (6.31b)-(6.31d) defined over the convex set dom $c = (0, 1)^K \times \mathbb{R}_{+*}^K$. The nondegeneracy of $(\eta_k^{(0)} - \tilde{\eta}_k)$ and $(\sum_{k=1}^K \gamma_k - 1)$ immediately follows from the convexity of these functions [Lasserre, 2010]. Next, we show that if $P_k(G_k Q_k/\gamma_k) - P_k^{(0)} = 0$ then $\nabla P_k \neq 0$. Assume that:

$$P_k(Q_k/\gamma_k) = P_k^{(0)} < 1, \tag{F.23}$$

and that:

$$\nabla P_k = 0 \Leftrightarrow \begin{cases} \dfrac{\partial P_k}{\partial Q_k} = 0 \\[2mm] \dfrac{\partial P_k}{\partial \gamma_k} = 0 \end{cases} \Leftrightarrow \begin{cases} \dfrac{1}{\gamma_k} P_k'(G_k Q_k/\gamma_k) = 0 \\[2mm] \dfrac{-Q_k}{\gamma_k^2} P_k'(G_k Q_k/\gamma_k) = 0. \end{cases} \tag{F.24}$$

But since $(\gamma, Q) \in$ dom c we have $Q_k > 0$ and $\gamma_k > 0$, and from $P_k(G_k Q_k/\gamma_k) < 1$ we have $P_k'(G_k Q_k/\gamma_k) \neq 0$ which contradicts Eq. (F.24). Therefore, $\nabla P_k \neq 0$.

F.6 Proof of Theorem 6.9

From [Lasserre, 2010, Th. 2.3], any KKT point of Problem 6.3 is optimal. It remains to show that any KKT point is a solution of Eqs. (6.36)-(6.38). Let $(\boldsymbol{\mu}, \boldsymbol{\theta}, \lambda)$ be the nonnegative Lagrange multipliers associated with the $2K + 1$ constraints Eqs. (6.31b)-(6.31d), respectively. Then, a KKT point $(\boldsymbol{\gamma}, \boldsymbol{Q}, \boldsymbol{\mu}, \boldsymbol{\theta}, \lambda)$ is given by the following equations:

$$\nabla(\sum_{k=1}^{K} Q_k) - \sum_{k=1}^{K} \mu_k \nabla \tilde{\eta}_k(\gamma_k, Q_k) + \sum_{k=1}^{K} \theta_k \nabla P_k(G_k Q_k / \gamma_k) + \lambda \nabla(\sum_{k=1}^{K} \gamma_k) = 0, \tag{F.25a}$$

$$-\mu_k(\tilde{\eta}_k(\gamma_k, Q_k) - \eta_k^{(0)}) = 0, \ \theta_k(P_k(G_k Q_k / \gamma_k) - P_k^{(0)}) = 0, \ \lambda(\sum_{k=1}^{K} \gamma_k - 1) = 0. \tag{F.25b}$$

Before working on the KKT equations, we compute the gradients:

$$\frac{\partial P_k}{\partial Q_k} = G_k \frac{1}{\gamma_k} P_k'(G_k Q_k / \gamma_k), \tag{F.26}$$

$$\frac{\partial P_k}{\partial \gamma_k} = -G_k \frac{Q_k}{\gamma_k^2} P_k'(G_k Q_k / \gamma_k). \tag{F.27}$$

We remark that the gradient of $\tilde{\eta}_k$ can be expressed as:

$$\nabla \tilde{\eta}_k = -m_k R_k \gamma_k \nabla P_k + \begin{bmatrix} 0 & m_k R_k(1 - P_k) \end{bmatrix}^T. \tag{F.28}$$

Stability Eq. (F.25a) is thus equivalent to the $2K$ scalar equations:

$$1 + (m_k R_k \gamma_k \mu_k + \theta_k) \frac{\partial P_k}{\partial Q_k} = 0, \tag{F.29}$$

$$(m_k R_k \gamma_k \mu_k + \theta_k) \frac{\partial P_k}{\partial \gamma_k} - m_k R_k(1 - P_k)\mu_k + \lambda = 0, \tag{F.30}$$

that are equivalent to the matrix identity:

$$S \begin{bmatrix} \mu_k \\ \theta_k \end{bmatrix} + \begin{bmatrix} 1 \\ \lambda \end{bmatrix} = 0, \tag{F.31}$$

where:

$$S = \begin{bmatrix} m_k R_k \gamma_k \frac{\partial P_k}{\partial Q_k} & \frac{\partial P_k}{\partial Q_k} \\ m_k R_k \left(\gamma_k \frac{\partial P_k}{\partial \gamma_k} - (1 - P_k) \right) & \frac{\partial P_k}{\partial \gamma_k} \end{bmatrix}. \tag{F.32}$$

Therefore, we are provided with K independent linear equations in (μ_k, θ_k), which are easily solved (S is 2×2) if and only if $\det S \neq 0$ (this is guaranteed by the next result):

Lemma F.1. $\det S < 0$.

Proof. By direct computation:

$$\det S = m_k R_k(1 - P_k) \frac{\partial P_k}{\partial Q_k}$$

$$= (1 - P_k) G_k \frac{P_k'(G_k Q_k / \gamma_k)}{\gamma_k}$$

$$< 0 \tag{F.33}$$

since the PER is a decreasing function of SNR, *i.e.* $P_k'(G_kQ_k/\gamma_k) < 0$.   □

The solution of Eq. (F.31) is easy to write from $\begin{bmatrix} \mu_k \\ \theta_k \end{bmatrix} = -S^{-1}\begin{bmatrix} 1 \\ \lambda \end{bmatrix}$, hence after some basic algebra:

$$\begin{bmatrix} \mu_k \\ \theta_k \end{bmatrix} = \frac{-1}{\det S}\begin{bmatrix} \frac{-1}{\gamma_k}G_kP_k'(G_kQ_k/\gamma_k)\left(\frac{Q_k}{\gamma_k} + \lambda\right) \\ m_kR_kG_kP_k'(G_kQ_k/\gamma_k)\left(\frac{Q_k}{\gamma_k} + \lambda\right) + m_kR_k(1 - P_k) \end{bmatrix}. \tag{F.34}$$

The complementary slackness equations:

$$\begin{cases} -\mu_k(\tilde{\eta}_k(\gamma_k, Q_k) - \eta_k^{(0)}) = 0, \\[2mm] \theta_k(P_k(G_kQ_k/\gamma_k) - P_k^{(0)}) = 0, \\[2mm] \lambda\left(\sum_{k=1}^K \gamma_k - 1\right), \end{cases} \tag{F.35}$$

are thus equivalent to:

$$\left(-G_k\frac{Q_k}{\gamma_k} - \lambda G_k\right)\left(\eta_k^{(0)} - \gamma_k m_k R_k(1 - P_k(G_kQ_k/\gamma_k))\right) = 0, \tag{F.36}$$

$$(\Theta(G_kQ_k/\gamma_k) - \lambda G_k)\left(P_k(G_kQ_k/\gamma_k) - P_k^{(0)}\right) = 0, \tag{F.37}$$

$$\lambda(\sum_{k=1}^K \gamma_k - 1) = 0, \tag{F.38}$$

where, $\forall x \in \mathbb{R}_{+*}$:

$$\Theta(x) := \frac{xP_k'(x) + 1 - P_k(x)}{P_k'(x)}. \tag{F.39}$$

Finally, if the goodput constraint is inactive at the optimal point (γ^*, Q^*), then by Eq. (F.36) we have:

$$-G_k\frac{Q_k^*}{\gamma_k^*} - \lambda^*G_k = 0 \Leftrightarrow -\frac{Q_k^*}{\gamma_k} = \lambda^*. \tag{F.40}$$

But $\lambda^* \geq 0$ (by definition), hence $Q_k^* = 0$ which contradicts the fact that $Q^* \neq 0$. Thus the goodput constraint is always active at the optimal, *i.e.* $\tilde{\eta}(\gamma_k^*, Q_k^*) - \eta_k^{(0)} = 0$.

F.7   Proof of Theorem 6.11

Let (μ, θ, λ) be the Lagrange multipliers associated with the $2K+1$ constraints Eqs. (6.45b)-(6.45d), respectively. Then, the KKT conditions are written:

$$\nabla\sum_{k=1}^K Q_k - \sum_{k=1}^K \mu_k\nabla\tilde{\eta}_k(\gamma_k, Q_k) + \sum_{k=1}^K \theta_k\nabla d_k^{\text{MAC}}(\gamma_k, Q_k) + \lambda\nabla(\sum_{k=1}^K \gamma_k) = 0, \tag{F.41a}$$

$$-\mu_k(\tilde{\eta}_k(\gamma_k, Q_k) - \eta_k^{(0)}) = 0, \quad \theta_k(d_k^{\text{MAC}}(\gamma_k, Q_k) - d_k^{(0)}) = 0, \quad \lambda(\sum_{k=1}^K \gamma_k - 1) = 0. \tag{F.41b}$$

Before working on the KKT equations, we compute the gradients:

$$\frac{\partial P_k}{\partial Q_k} = G_k \frac{1}{\gamma_k} P'_k(G_k Q_k / \gamma_k), \tag{F.42}$$

$$\frac{\partial P_k}{\partial \gamma_k} = -G_k \frac{Q_k}{\gamma_k^2} P'_k(G_k Q_k / \gamma_k). \tag{F.43}$$

We remark that the gradients of $\tilde{\eta}_k$ and d_k^{MAC} can be expressed as:

$$\nabla \tilde{\eta}_k = -m_k R_k \gamma_k \nabla P_k + \begin{bmatrix} 0 & m_k R_k (1 - P_k) \end{bmatrix}^T, \tag{F.44}$$

$$\nabla d_k^{\mathrm{MAC}} = \frac{\delta'(P_k)}{\gamma_k} \nabla P_k + \begin{bmatrix} 0 & -\frac{d_k^{\mathrm{MAC}}}{\gamma_k} \end{bmatrix}^T, \tag{F.45}$$

with the function $\delta' : x \mapsto 1/(1-x)^2 - L^2 x^{L-1}/(1-x^L)^2$ for $x \in [0, 1]$.

As a consequence, Eq. (F.41a) leads to the following $2K$ scalar equalities:

$$1 + \left(m_k R_k \gamma_k \mu_k + \frac{\delta'(P_k)}{\gamma_k} \theta_k \right) \frac{\partial P_k}{\partial Q_k} = 0, \tag{F.46}$$

$$\left(m_k R_k \gamma_k \mu_k + \frac{\delta'(P_k)}{\gamma_k} \theta_k \right) \frac{\partial P_k}{\partial \gamma_k} - m_k R_k (1 - P_k) \mu_k - \frac{d_k^{\mathrm{MAC}}}{\gamma_k} \theta_k + \lambda = 0. \tag{F.47}$$

This set of $2K$ equations in Eqs. (F.46)-(F.47) are equivalent to the following K independent 2-by-2 matrix linear identities on (μ_k, θ_k):

$$S \begin{bmatrix} \mu_k \\ \theta_k \end{bmatrix} + \begin{bmatrix} 1 \\ \lambda \end{bmatrix} = 0, \tag{F.48}$$

where:

$$S = \begin{bmatrix} m_k R_k \gamma_k \frac{\partial P_k}{\partial Q_k} & \frac{\delta'(P_k)}{\gamma_k} \frac{\partial P_k}{\partial Q_k} \\ m_k R_k \left(\gamma_k \frac{\partial P_k}{\partial \gamma_k} - (1 - P_k) \right) & \frac{\delta'(P_k)}{\gamma_k} \frac{\partial P_k}{\partial \gamma_k} - \frac{d_k^{\mathrm{MAC}}}{\gamma_k} \end{bmatrix}. \tag{F.49}$$

This matrix identity can be easily solved if and only if $\det S \neq 0$. This property is guaranteed by the next result.

Lemma F.2. $\det S > 0$.

Proof. By direct computation:

$$\det S = -m_k R_k d_k^{\mathrm{MAC}} \frac{\partial P_k}{\partial Q_k} + m_k R_k (1 - P_k) \frac{\delta'(P_k)}{\gamma_k} \frac{\partial P_k}{\partial Q_k}$$

$$= \frac{-m_k R_k G_k P'_k(G_k Q_k / \gamma_k)}{\gamma_k^2} \left(\delta(P_k) - (1 - P_k) \delta'(P_k) \right). \tag{F.50}$$

Since the packet error rate is a decreasing function of SNR, $P'_k(G_k Q_k / \gamma_k) \leq 0$. Thus $\det S$ has the same sign than $\delta(P_k) - (1 - P_k)\delta'(P_k)$:

$$\delta(x) - (1 - x)\delta'(x) = \frac{-Lx^L}{1 - x^L} + (1 - x)L^2 \frac{x^{L-1}}{(1 - x^L)^2}$$

$$= \frac{L^2(1 - x)x^{L-1} - Lx^L(1 - x^L)}{(1 - x^L)^2}$$

$$= \frac{Lx^{L-1}}{(1 - x^L)^2} \left(x^2 - (L + 1)x + L \right). \tag{F.51}$$

Finally, since the polynomial $x^2 - (L+1)x + L = (x-1)(x-L) > 0$ for $0 < x < 1$ (remind that $\pi_k < 1$ since $\gamma_k > 0$ and $Q_k > 0$), then $\det S > 0$. $\qquad\square$

After some simple algebra, we obtain the following solutions for Eq. (F.48):

$$\begin{bmatrix} \mu_k \\ \theta_k \end{bmatrix} = -\frac{S'}{\det S} \tag{F.52}$$

where:

$$S' = \begin{bmatrix} \frac{-1}{\gamma_k^2}\left(\delta'(P_k)P_k'(G_kQ_k/\gamma_k)G_k\left(\frac{Q_k}{\gamma_k} + \lambda\right) + \delta(P_k)\right) \\ \\ m_kR_kP_k'(G_kQ_k/\gamma_k)G_k\left(\frac{Q_k}{\gamma_k} + \lambda\right) + m_kR_k(1 - P_k) \end{bmatrix}. \tag{F.53}$$

Hence, the complementary slackness equations Eq. (F.41b) are equivalent to:

$$(M(G_kQ_k/\gamma_k) - \lambda G_k)\left(\eta_k^{(0)} - \gamma_k m_k R_k\left(1 - P_k(G_kQ_k/\gamma_k)\right)\right) = 0 \tag{F.54}$$

$$(\Theta(G_kQ_k/\gamma_k) - \lambda G_k)\left(\frac{\delta\left(P_k(G_kQ_k/\gamma_k)\right)}{\gamma_k} - d_k^{(0)}\right) = 0, \tag{F.55}$$

$$\lambda\left(\sum_{k=1}^{K} \gamma_k - 1\right) = 0, \tag{F.56}$$

where, $\forall x \in \mathbb{R}_{+*}$:

$$M(x) = -\frac{x\delta'(P_k(x))P_k'(x) + \delta(P_k(x))}{\delta'(P_k(x))P_k'(x)} \tag{F.57}$$

$$\Theta(x) = -\frac{xP_k'(x) + 1 - P_k(x)}{P_k'(x)}. \tag{F.58}$$

F.8 Calculations leading to Algorithm 6.6

We deduce a simple and efficient algorithm from the KKT characterization Eqs. (6.51)-(6.53). First of all, it is worth to emphasize that if link k satisfies $d_k^{(0)} \geq (m_kR_k)/\eta_k^{(0)}$, then its delay constraint is inactive. Indeed, the delay expression can be split as follows:

$$d_k^{\mathrm{MAC}}(\gamma_k, Q_k) = \frac{m_kR_k}{\tilde{\eta}_k(\gamma_k, Q_k)} + \frac{1}{\gamma_k}\left(L - \frac{L}{1 - (P_k(G_kQ_k/\gamma_k))^L}\right). \tag{F.59}$$

Since $P_k(G_kQ_k/\gamma_k) \leq 1$, the term $L - \frac{L}{1 - (P_k(G_kQ_k/\gamma_k))^L} \leq 0$, and one obtains:

$$d_k^{\mathrm{MAC}}(\gamma_k, Q_k) \overset{(a)}{\leq} \frac{m_kR_k}{\eta_k^{(0)}} \overset{(b)}{\leq} d_k^{(0)}, \tag{F.60}$$

where (a) boils down from the feasibility $\tilde{\eta}_k(\gamma_k, Q_k) \geq \eta_k^{(0)}$, and (b) follows by assumption. As a consequence, the right term in the LHS of Eq. (6.51) is equal to zero while the left term in the LHS of Eq. (6.52) is equal to zero. These both identities characterize the associated γ_k and Q_k (see Item 1. in Algorithm 6.6).

Otherwise (*i.e.*, $d_k^{(0)} < (m_kR_k)/\eta_k^{(0)}$), we have two cases:

(i) Let us assume that the goodput constraint is inactive, *i.e.*, $\eta_k^{(0)} < \gamma_k m_k R_k (1 - P_k(G_k Q_k/\gamma_k))$. Then, by Eq. (6.51) we have $M(G_k Q_k/\gamma_k) = \lambda G_k$ which leads to $Q_k = \frac{\gamma_k}{G_k} M^{-1}(\lambda G_k)$. Then two cases are possible: the delay constraint is active or not.

– if the delay constraint is active, then $\gamma_k = \delta(P_k(M^{-1}(\lambda G_k))/d_k^{(0)}$.

– if the delay constraint is inactive, then $\Theta(G_k Q_k/\gamma_k) = \lambda G_k$ which implies that $M^{-1}(\lambda G_k) = \Theta^{-1}(\lambda G_k)$, which is not possible.

(ii) Let us assume that the goodput constraint is active, *i.e.*, $\eta_k^{(0)} = \gamma_k m_k R_k (1 - P_k(G_k Q_k/\gamma_k))$. Once again, two cases are possible: the delay constraint is active or not.

– if the delay constraint is active, then $\delta(P_k(G_k Q_k/\gamma_k)) = \gamma_k d_k^{(0)}$ which implies that there is some P_k such that (due to the active goodput constraint):

$$m_k R_k (1 - P_k)\delta(P_k) = \eta_k^{(0)} d_k^{(0)}. \tag{F.61}$$

According to the closed-form expression of δ, the corresponding P_k (in $(0, 1)$) is a root of the polynomial equation:

$$L x^{L+1} - (L + 1 - d_k^{(0)}\eta_k^{(0)}/(m_k R_k))x^L + 1 - d_k^{(0)}\eta_k^{(0)}/(m_k R_k) = 0. \tag{F.62}$$

– if the delay constraint is inactive, then $\Theta(G_k Q_k/\gamma_k) = \lambda G_k$. Therefore, $\gamma_k(\lambda) = \eta_k^{(0)}/(m_k R_k(1 - P_k(\Theta^{-1}(\lambda G_k))))$ (thanks to the active goodput constraint) and $Q_k(\lambda) = \frac{\gamma_k(\lambda)}{G_k}\Theta^{(-1)}(\lambda G_k)$ (thanks to the inactive delay constraint).

Notice that when the goodput constraint is inactive and the delay constraint is active, we have:

$$\delta(P_k(M^{-1}(\lambda G_k)))(1 - P_k(M^{-1}(\lambda G_k))) > \frac{\eta_k^{(0)} d_k^{(0)}}{m_k R_k}. \tag{F.63}$$

Similarly, when the goodput constraint is active and the delay constraint is inactive, we have:

$$\delta(P_k(\Theta^{-1}(\lambda G_k)))(1 - P_k(\Theta^{-1}(\lambda G_k))) < \frac{\eta_k^{(0)} d_k^{(0)}}{m_k R_k}. \tag{F.64}$$

The algorithm is initialized with $\lambda = 0$ and assuming that only the goodput constraint is active. If the delay constraint is not satisfied, we choose a higher γ_k for satisfying this constraint even if the constraint Eq. (6.45d) does not hold anymore. Then we will increase λ until a solution that satisfies Eq. (6.45d) is found.

Bibliography

M. Abramowitz and I. A. Stegun, **Handbook of mathematical functions, with formulas, graphs, and mathematical tables**, 20th ed., ser. Applied Mathematics. National Bureau of Standards, Jun. 1972, vol. 55. Cited pages 22, 70, 81, and 141

F. Adachi, S. Ito, and K. Ohno, "Performance analysis of a time diversity ARQ in land mobile radio," vol. 37, no. 2, pp. 177–187, Feb. 1989. Cited page 8

A. Amraoui, "Asymptotic and finite-length optimization of LDPC codes," Ph.D. dissertation, EPFL, 2006. Cited page 95

I. Andriyanova, "Finite-length scaling of turbo-like code ensembles on the binary erasure channel," **IEEE J. Sel. Areas Commun.**, vol. 27, no. 6, pp. 918–927, Aug. 2009. Cited page 95

I. Andriyanova and E. Soljanin, "Optimized IR-HARQ schemes based on punctured LDPC codes over the BEC," **IEEE Trans. Inf. Theory**, vol. 58, no. 10, pp. 6433–6445, Oct. 2012. Cited page 20

I. Andriyanova and R. Urbanke, "Waterfall region performance of punctured LDPC codes over the BEC," in **Information Theory (ISIT), 2009 International Symposium on**. IEEE, Jun. 2009. Cited page 95

K. Ausavapattanakun and A. Nosratinia, "Analysis of selective-repeat ARQ via matrix signal-flow graphs," **IEEE Trans. Commun.**, vol. 55, no. 1, pp. 198–204, Jan. 2007. Cited pages 38 and 39

——, "Analysis of go-back-N ARQ in block fading channels," **IEEE Trans. Wireless Commun.**, vol. 6, no. 8, pp. 2793–2797, Aug. 2007. Cited page 38

L. Badia, "On the effect of feedback errors in Markov models for SR ARQ packet delays," in **Global Communications Conference (GLOBECOM)**. IEEE, 2009. Cited pages 38 and 39

A. W. Berger, "Comparison of call gapping and percent blocking for overload control in distributed switching systems and telecommunications networks," **IEEE Trans. Commun.**, vol. 39, no. 4, pp. 574–580, Apr. 1991. Cited page 12

S. Boyd and L. Vandenberghe, **Convex Optimization**, 7th ed. Cambridge University Press, 2004. Cited pages 102, 103, 121, 128, 149, 150, 162, and 163

F. Brah, L. Vandendorpe, and J. Louveaux, "Constrained resource allocation in OFDMA downlink systems with partial CSIT," in **International Conference on Communications (ICC)**. IEEE, 2008, pp. 4144–4148. Cited page 64

D. S. Buckingham and M. C. Valenti, "The information-outage probability of finite-length codes over AWGN channels," in **Information Sciences and Systems (CISS), 2008 42nd Annual Conference on**, Nov. 2008. Cited pages 79, 81, and 83

G. Caire and D. Tuninetti, "The throughput of Hybrid-ARQ protocols for the Gaussian collision channel," **IEEE Trans. Inf. Theory**, vol. 47, no. 5, pp. 1971–1988, Jul. 2001. Cited pages 12 and 20

G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," **IEEE Trans. Inf. Theory**, vol. 44, no. 3, pp. 927–946, May 1998. Cited page 100

D. Chase, "Code combining – A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," **IEEE Trans. Commun.**, vol. 33, no. 5, pp. 385–393, May 1985. Cited pages 9 and 20

Q. Chen and P. Fan, "Performance analysis of Hybrid ARQ with code combining over interleaved Rayleigh fading channel," **IEEE Trans. Veh. Technol.**, vol. 54, no. 3, pp. 1207–1214, May 2005. Cited page 20

J.-F. T. Cheng, "Coding performance of hybrid ARQ schemes," **IEEE Trans. Commun.**, vol. 54, no. 6, pp. 1017–1029, Jun. 2006. Cited page 20

Y. Choi, S. Choi, and S. Yoon, "MSDU-based ARQ scheme for IP-level performance maximization," in **Global Communications Conference (GLOBECOM)**, vol. 5. IEEE, Oct. 2005, pp. 2495–2499. Cited pages 18, 25, 43, and 44

S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view," **IEEE Trans. Commun.**, vol. 49, no. 9, pp. 1561–1571, Sep. 2001. Cited page 100

B. Devillers, J. Louveaux, and L. Vandendorpe, "Bit and power allocation for goodput optimization in coded parallel subchannels with ARQ," **IEEE Trans. Signal Process.**, vol. 56, no. 8, pp. 3652–3661, Aug. 2008. Cited pages 12, 69, 73, 113, and 131

B. T. Doshi and H. Heffes, "Overload performance of several processor queuing disciplines for the M/M/1 queue," **IEEE Trans. Commun.**, vol. 34, no. 6, pp. 538–546, Jun. 1986. Cited page 12

W. El bahri, H. Boujemâa, and M. Siala, "Effects of noisy feedback on the performance of HARQ schemes over multipath block fading channels for DS-SSS," in **Personal, Indoor and Mobile Radio Communications (PIMRC), 2005 16th International Symposium on**. IEEE, 2005, pp. 2552–2556. Cited pages 38, 50, and 51

C. A. Floudas and V. Visweswaran, "A primal-relaxed dual global optimization approach," **Journal of Optimization Theory and Applications**, vol. 78, no. 2, p. 187, 1993. Cited pages 87 and 89

R. G. Gallager, "Low-Density Parity-Check codes," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1963. Cited page 95

——, **Information Theory and Reliable Communication**. Wiley, 1968. Cited page 78

S. Gault, W. Hachem, and P. Ciblat, "Performance of OFDMA on Rayleigh fading channels in a multi-cell environment," **IEEE Trans. Commun.**, vol. 55, no. 4, pp. 740–751, Apr. 2007. Cited pages 69, 73, 91, 131, and 151

M. Goldenbaum, R. Akl, S. Valentin, and S. Stanczak, "On the effect of feedback delay in the downlink of multiuser OFDM systems," in **Information Sciences and Systems (CISS), 2011 45th Annual Conference on**. IEEE, 2011, pp. 1–6. Cited page 64

J. Gorski, F. Pfeuferret, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: a survey and extensions," **Mathematical Methods of Operations Research**, vol. 66, no. 3, pp. 373–407, 2007. Cited page 86

I. Gradshteyn and I. Ryzhik, **Table of integrals, series, and products**, 7th ed. Academic press New York, 1980. Cited pages 71, 82, 83, 152, and 153

H. J. Greenberg and W. P. Pierskalla, "A review of quasi-convex functions," **Operations research**, vol. 19, no. 7, pp. 1553–1570, Nov. 1971. Cited pages 118 and 162

J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," **IEEE Trans. Commun.**, vol. 36, no. 4, pp. 389–400, Apr. 1988. Cited pages 9 and 20

T. S. Han, **Information spectrum methods in information theory**. Springer, 2003. Cited page 79

Z. K. Ho, V. K. Lau, and R. S. Cheng, "Cross-layer design of FDD-OFDM systems based on ACK/NACK feedbacks," **IEEE Trans. Inf. Theory**, vol. 55, no. 10, pp. 4568–4584, Oct. 2009. Cited pages 64 and 75

S. Kallel, "Analysis of a type II Hybrid ARQ scheme with code combining," **IEEE Trans. Commun.**, vol. 38, no. 8, pp. 1133–1137, Aug. 1990. Cited page 20

J. Kim, W. Hur, A. Ramamoorthy, and S. W. McLaughlin, "Design of rate compatible irregular LDPC codes for incremental redundancy Hybrid ARQ systems," in **Information Theory (ISIT), 2006 International Symposium on**. Seattle, USA: IEEE, Jul. 2006. Cited page 9

N. Ksairi, P. Bianchi, P. Ciblat, and W. Hachem, "Resource allocation for downlink cellular OFDMA systems - Part I: Optimal allocation," **IEEE Trans. Signal Process.**, vol. 58, no. 2, pp. 720–734, Feb. 2010. Cited page 63

——, "Resource allocation for downlink cellular OFDMA systems - Part II: Practical algorithms and optimal reuse factor," **IEEE Trans. Signal Process.**, vol. 58, no. 2, pp. 735–749, Feb. 2010. Cited page 63

X. Lagrange, "Throughput of HARQ protocols on a block fading channel," **IEEE Commun. Lett.**, vol. 14, no. 3, pp. 257–259, Mar. 2010. Cited page 113

J. N. Laneman, "On the distribution of mutual information," in **Information Theory and its Applications (ITA), Workshop on**, Feb. 2006. Cited pages 79 and 80

J. B. Lasserre, "On representations of the feasible set in convex optimization," **Optimization Letters**, no. 4, 2010. Cited pages 118, 126, 163, and 164

V. K. Lau, W. K. Ng, and D. S. W. Hui, "Asymptotic tradeoff between cross-layer goodput gain and outage diversity in OFDMA systems with slow fading and delayed CSIT," **IEEE Trans. Wireless Commun.**, vol. 7, no. 7, pp. 2732–2739, Jul. 2008. Cited page 64

A. Le Duc, "Performance closed-form derivations and analysis of Hybrid-ARQ retransmission schemes in a cross-layer context," Ph.D. dissertation, Télécom ParisTech, Paris, France, 2009. Cited pages 2, 10, 11, 14, 17, 20, 21, 25, 27, 31, and 133

C. J. Le Martret, A. Le Duc, S. Marcille, and P. Ciblat, "Analytical performance derivation of Hybrid ARQ schemes at IP layer," **IEEE Trans. Commun.**, vol. 60, no. 5, pp. 1305–1314, May 2012. Cited pages 11, 17, 20, and 124

M. Levorato and M. Zorzi, "Performance analysis of type II Hybrid ARQ with Low-Density Parity-Check codes," in **Communications, Control and Signal Processing (ISCCSP), 2008 3rd International Symposium on**, 2008, pp. 804–809. Cited page 20

S. Lin and D. J. Costello, **Error Control Coding: Fundamentals and Applications**. Englewood Cliffs, NJ: Prentice-Hall, 1983. Cited pages 9, 10, 11, 12, 13, and 18

Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," **IEEE Trans. Wireless Commun.**, vol. 3, no. 5, pp. 1746–1755, Sep. 2004. Cited page 113

A. Lozano, A. M. Tulino, and S. Verdù, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," **IEEE Trans. Inf. Theory**, vol. 52, no. 7, pp. 3033–3051, Jul. 2006. Cited pages 69 and 70

E. Malkamäki and H. Leib, "Performance of truncated type-II Hybrid ARQ schemes with noisy feedback over block fading channels," **IEEE Trans. Commun.**, vol. 48, no. 9, pp. 1477–1487, Sep. 2000. Cited pages 38, 39, 49, and 50

Y. Polyanskiy, H. V. Poor, and S. Verdù, "Channel coding rate in the finite blocklength regime," **IEEE Trans. Inf. Theory**, vol. 56, no. 5, pp. 2307–2359, May 2010. Cited page 78

V. Rodriguez, "An analytical foundation for resource management in wireless communications," in **Global Communications Conference (GLOBECOM)**. IEEE, 2003. Cited page 131

S. M. Ross, **Introduction to probability models**, 9th ed. Academic Press, 2007. Cited pages 11 and 81

M. Rossi and M. Zorzi, "Analysis and heuristics for the characterization of Selective Repeat ARQ delay statistics over wireless channels," **IEEE Trans. Veh. Technol.**, vol. 52, no. 5, pp. 1365–1377, Sep. 2003. Cited page 17

W. Rui and V. K. N. Lau, "Combined cross-layer design and HARQ for multiuser systems with outdated channel state information at transmitter (CSIT) in slow fading channels," **IEEE Trans. Wireless Commun.**, vol. 7, no. 7, pp. 2771–2777, Jul. 2008. Cited page 64

W. E. Ryan and S. Lin, **Channel Codes: Classical and Modern**. Cambridge University Press, 2009. Cited page 95

M. P. Schmitt, "ARQ systems for wireless communications," Ph.D. dissertation, Technischen Universität Darmstadt, Darmstadt, Germany, Sep. 2002. Cited page 9

L. S. Schwartz, "Feedback for error control and two-way communication," **IEEE Transactions on Communication Systems**, pp. 49–56, Mar. 1963. Cited page 7

S. Sesia, I. Toufik, and M. Baker, **LTE: the Long Term Evolution - From theory to practice**. Wiley, 2009. Cited page 7

C. E. Shannon, "A mathematical theory of communication," **Bell Syst. Tech. J.**, vol. 27, no. 379, p. 623, 1948. Cited page 78

R. A. Silverman, "On binary channels and their cascades," **IRE Transactions — Information Theory**, pp. 19–27, Dec. 1955. Cited page 41

R. P. Stanley, **Enumerative combinatorics**. Cambridge University Press, Apr. 1997, vol. 1. Cited page 145

T. Starr, J. M. Cioffi, and P. J. Silverman, **Understanding digital subscriber line technology**. Englewood Cliffs, NJ: Prentice-Hall, 1999. Cited pages 72 and 73

I. Stupia, V. Lottici, F. Giannetti, and L. Vandendorpe, "Link resource adaptation for multiantenna bit-interleaved coded multicarrier systems," **IEEE Trans. Signal Process.**, vol. 60, no. 7, pp. 3644–3656, Jul. 2012. Cited page 136

L. Szczecinski, P. Duhamel, and M. Rahman, "Adaptive incremental redundancy for HARQ transmission with outdated CSI," in **Global Communications Conference (GLOBECOM)**. IEEE, 2011. Cited pages 64 and 136

D. N. Tse and P. Viswanath, **Fundamentals of Wireless Communications**. Cambridge University Press, 2005. Cited pages 63, 64, 65, 69, 73, 77, and 79

H. C. A. Van Duuren, "Printing telegraph systems," U.S. Patent 2313980, Mar. 1943. Cited page 7

R. Wang and V. K. N. Lau, "Robust optimal cross-layer designs for TDD-OFDMA systems with imperfect CSIT and unknown interference: state-space approach based on 1-bit ACK/NAK feedbacks," **IEEE Trans. Commun.**, vol. 56, no. 5, pp. 754–761, May 2008. Cited page 64

L. Weidong, Y. Hongwen, and Y. Dacheng, "Approximation formulas for the symmetric capacity of M-ary modulations," Beijing University of Posts and Telecommunications, Apr. 2007, http://www.paper.edu.cn/en. Cited pages 71 and 153

S. B. Wicker, **Error Control Systems for Digital Communications and Data Storage**. Englewood Cliffs, NJ: Prentice-Hall, 1995. Cited pages 10, 18, 38, and 51

C. W. Wong, R. S. Cheng, K. Ben Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," **IEEE J. Sel. Areas Commun.**, vol. 17, no. 10, pp. 1747–1758, Oct. 1999. Cited pages 69 and 72

P. Wu and N. Jindal, "Coding versus ARQ in fading channels: how reliable should the PHY be ?" in **Global Communications Conference (GLOBECOM)**. IEEE, 2009. Cited pages 38, 50, and 51

——, "Coding versus ARQ in fading channels: how reliable should the PHY be ?" **IEEE Trans. Commun.**, vol. 59, no. 12, pp. 3363–3374, Dec. 2011. Cited pages 69, 73, 96, and 131

M. Zorzi and R. R. Rao, "On the use of renewal theory in the analysis of ARQ protocols," **IEEE Trans. Commun.**, vol. 44, no. 9, pp. 1077–1081, Sep. 1996. Cited page 20