Hybrid-ARQ mechanisms in a radio-cognitive context.

Romain Tajan

Université de Cergy-Pontoise
École doctorale Sciences & Ingénierie

THÈSE

présentée en première version en vu d'obtenir le grade de

DOCTEUR

spécialité Sciences et Technologies
de l'Information et de la Communication

par

Romain Tajan

Mécanismes de retransmission Hybrid-ARQ en radio-cognitive.

Directeur de thèse : **Mme Inbar Fijalkow**

Co-Directeur de thèse : **M. Charly Poulliat**

Soutenue le 5 Décembre 2013

Devant la Commission d'Examen

JURY

Rapporteur :	Professeur	M. Philippe Ciblat	Télécom ParisTech
Rapporteur :	Full Professor	M. Michele Zorzi	University of Padova
Examinateur :	Directeur de Recherche Inria	M. Eitan Altman	Sofia Antipolis
Président :	Directeur de Recherche CNRS	M. Pierre Duhamel	LSS Supélec
Examinatrice :	Docteur	Mme Véronique Serfaty	DGA/DS/MRIS

Remerciements

Je tiens à remercier en premier lieu Charly Poulliat, pour son excellent encadrement qui m'a permis de mener à bien cette thèse. Cette thèse n'aurait jamais pu exister sans son implication, ses conseils avisés, sa curiosité scientifique et nos nombreuses discussions. Merci aussi à Inbar Fijalkow, qui a toujours été d'excellent conseil, qui a su être présente malgré la distance et qui m'a souvent aidé à mettre des mots sur mes idées.

Je remercie le CNRS et la DGA pour avoir financé ces travaux, le laboratoire ETIS au sein duquel j'ai effectué ces trois années de thèse et le laboratoire IRIT qui m'a accueilli pour ma dernière année.

Je remercie ensuite Philippe Ciblat et Michele Zorzi pour avoir accepté de rapporter sur ce travail. Leurs remarques et leurs commentaires m'ont aidé à améliorer le contenu de ce document. Je tiens aussi à remercier Eitan Altman et Véronique Serfaty pour avoir accepté de faire parti de mon jury et Pierre Duhamel pour l'avoir présidé.

Tout au long de cette thèse, j'ai eu la chance de pouvoir occuper trois bureaux différents dans deux laboratoires différents. Je souhaite remercier ceux avec qui j'ai partagé ces bureaux : Alexis, Jean-Christophe, Alan, Rémi, Raouia, Sockchenda, Bouchra, Olivier et Hasan. Je remercie aussi tous ceux avec qui j'ai pu partager des moments agréables pendant ces trois années, en particulier Gaël, Ludovic, Leïla, Mathilde, Laurent, Marina, Erbao, David P.Q., Iryna, David P., Romain, Anne, le groupe des doctorants du TéSA ainsi que ceux des équipes SC et RT de l'IRIT.

Au risque de répéter quelques noms, je souhaiterais dire un grand merci pour tous ceux qui m'ont aidé lors de mes venues sur Cergy et pour l'organisation de ma soutenance. Je pense en particulier à JC, Julie, Leïla, Antoine, Alan, Fanny, Liang, Véronica, Annick, Pauline, Suzanne et Damien.

Ce doctorat étant la conclusion de ma scolarité, je souhaite remercier ceux qui m'ont accompagné durant cette période. En particulier je remercie les professeurs M. Joyeux et M. Paul qui ont largement contribué à développer la curiosité et la pugnacité qui ont été nécessaire pendant cette thèse. Je remercie aussi les amis rencontrés pendant cette période : les 87, les amis de la prépa et aussi Fab, Clément, Chris, Brahim et Claire.

Il serait difficile de terminer sans remercier ceux qui ont fait de moi ce que je suis aujourd'hui. Un énorme merci à toute ma famille et ma belle-famille, en particulier mes parents, Florian et Pascal, qui ont toujours été présents pour moi et m'ont toujours soutenu.

Enfin je ne saurais comment te remercier suffisamment, Séverine, tu as été patiente et conciliante, tu m'as accompagné tout au long de cette thèse, tu m'as soutenu dans les moments difficiles et tu m'as encouragé à me surpasser. Pour toutes ces raisons, je te dis merci beaucoup, beaucoup...

Résumé de thèse

Dans les standards actuels tels que HSDPA ou LTE, des protocoles de retransmissions (ARQ : Automatic Repeat reQuest) sont utilisés conjointement au codage de canal afin de palier aux erreurs dues à l'absence ou la mauvaise connaissance sur le canal à la transmission. De tels protocoles sont appelés protocoles de retransmission hybrides (HARQ). On garantit ainsi la fiabilité du lien physique pour les couches OSI supérieures (du moins un taux d'erreur paquet faible). L'objet de cette thèse est de proposer des outils permettant l'analyse et l'optimisation des systèmes de communication en présence de protocoles HARQ avec une emphase particulière sur les systèmes cognitifs.

Dans la première partie, nous étudierons un système point-à-point dans lequel trois différents protocoles HARQ *adaptatifs* seront considérés. Nous considérerons principalement le régime asymptotique (i.e. codes optimaux gaussiens) pour lequel nous étudierons la maximisation du débit moyen sous des contraintes de puissance instantanée et de puissance moyenne. Nous montrerons que les Processus de Décision Markoviens (MDP) sont des outils adaptés aux problèmes d'optimisations considérés.

Dans la seconde partie, nous considérerons un contexte de radio cognitive. La radio cognitive est une approche permettant d'augmenter l'efficacité spectrale des réseaux sans fil. Pour ce faire, des utilisateurs non-licenciés (réseau secondaire) sont autorisés à communiquer dans les mêmes bandes de fréquences que des utilisateurs licenciés (réseau primaire). Les utilisateurs secondaires doivent, en revanche, limiter la quantité d'interférence générée sur les signaux primaires. Nous étudierons, dans cette thèse, le cas où un utilisateur secondaire interfère un utilisateur primaire qui emploie un protocole HARQ. Nous montrerons que si l'état du protocole HARQ est connu du système secondaire, une allocation conjointe de puissance et de rendement est possible, même sans connaissance instantanée du canal à l'émetteur secondaire. Cette allocation permet de maximiser le débit de l'utilisateur secondaire sous différentes contraintes. Nous considérerons en particulier des contraintes de puissance instantanée et de puissance moyenne pour l'utilisateur secondaire et une contrainte de garantie en débit pour l'utilisateur primaire. Nous montrerons, là encore, que les MDP sont des outils adéquats afin de résoudre le problème d'optimisation proposé.

Thesis Abstract

Automatic Repeat Request protocols (ARQ) are widely implemented in current mobile wireless standards such as HSDPA and LTE. In general, ARQ protocols are combined with channel coding to overcome errors caused by the lack of channel knowledge at the transmitter side. These protocols are called Hybrid ARQ protocols (HARQ). HARQ protocols ensure a good reliability (at least a small packet error rate) of the physical layer for the OSI upper layers. The purpose of this thesis is to provide tools for the analysis and the optimization of HARQ communication systems with an emphasis on cognitive systems.

In the first part of this work, we consider a point-to-point context. In this context, we study three different types of *adaptive* HARQ protocols (HARQ with no combining, with code combining, and with Chase combining). Under hypotheses of asymptotically optimal Gaussian codes and Rayleigh block fading channel, we address the maximization of the average throughput under peak and average power constraints. In this part, we show that Markov Decision Processes (MDP) provide a theoretical framework to this optimization problems.

In the second part, we consider the cognitive radio context. Cognitive Radio (CR) is an approach aiming to increase the spectral efficiency of wireless networks. In a CR context, unlicensed users (also called secondary users) are allowed to communicate within the same frequency bands and at the same time as licensed users (also called primary users). Secondary users must however limit the amount of interference generated on the primary users signals. In this thesis, we consider a scenario in which the secondary user interferes a primary user employing a HARQ protocol. When the secondary user knows the state of the primary HARQ protocol, we show that a joint power and rate allocation can be performed. This result holds even without instantaneous channel knowledge at the transmitter side. The power and rate allocation we propose maximizes the throughput of the secondary user under different constraints. In particular, we take into account constraints on the secondary instantaneous and average powers and a constraint on the primary throughput loss. We show again that the MDPs offer a good theoretical framework to solve the proposed optimization problem.

Contents

List of Figures

List of Tables

List of Acronyms

κ-ESM *κ-Effective SNR Mapping*

ACK *ACKnowledgement*

ACMI *ACcumulated Mutual Information*

AMC *Adaptive Modulation and Coding*

ARQ *Automatic Repeat reQuest*

AWGN *Additive White Gaussian Noise*

BIC *Backward Interference Cancellation*

BICM *Bit Interleaved Coded-Modulation*

CC *Chase Combining*

CD *Chain Decoding*

CMDP *Constrained Markov Decision Process*

CQI *Channel Quality Indicator*

CR *Cognitive Radio*

CRC *Cyclic Redundancy Check*

CSI *Channel Side Information*

CSIR *Channel Side Information at the Receiver*

CSIT *Channel Side Information at the Transmitter*

EESM *Exponential Effective SNR Mapping*

FCC *Federal Communications Commission*

FEC *Forward Error Correction*

FER *Frame Error Rate*

GNR *Gain to Noise Ratio*

HARQ *Hybrid Automatic Repeat reQuest*

HSDPA *High Speed Downlink Packet Access*

i.i.d. *independent and identically distributed*

IR *Incremental Redundancy*

LLN *Law of Large Numbers*

LP *Linear Programming*

LTE *Long Term Evolution*

MC *Markov Chain*

MDP *Markov Decision Process*

MIESM *Mutual Information Effective SNR Mapping*

ML *Maximum Likelihood*

MRC *Maximum Ratio Combining*

NACK *Negative ACKnowledgement*

OCSI *Outdated Channel Side Information*

OSA *Opportunistic Spectrum Access*

pdf *probability density function*

PHY *Physical*

POMDP *Partially Observable Markov Decision Process*

PSI-MDP *Partial State Information Markov Decision Process*

PU *Primary User*

QoS *Quality of Service*

SCSI *Statistical Channel Side Information*

SDR *Software Defined Radio*

SIC *Successive Interference Cancellation*

SINR *Signal to Interference plus Noise Ratio*

SNR *Signal to Noise Ratio*

SU *Secondary User*

TV-norm *Total Variation norm*

List of Notation

General Notation

$\mathbb{N}$ Set of positive integers; $\{0, 1, 2, \dots\}$,

$\mathbb{R}$ Set of real numbers,

$\mathbb{R}^+$ Set of positive real numbers,

$\bar{\mathbb{R}}$ Extender real numbers: $\mathbb{R} \cup \{-\infty, +\infty\}$,

$\mathbb{C}$ Set of complex numbers,

$\times$ Cartesian product

$C_b(X)$ Set of bounded continuous functions on X,

$C(X)$ Set of continuous functions on X,

$B(X)$ Set of measurable bounded functions on X,

Measure and Probability Notation

$\mathcal{B}(X)$ Borel σ-algebra of space X,

$\ell(\cdot)$ Lebesgue Measure ($\ell(]a, b[) = b - a$),

$\delta_x(\cdot)$ Dirac measure in x ($\delta_x(A) = \mathbb{1}_A(x)$),

$f_X(\cdot)$ probability density function of the random variable X,

$F_X(\cdot)$ cumulative density function of the random variable X,

$\mathbb{1}_A(\cdot)$ Function that is equal to 1 if x belongs to A and 0 otherwise,

$\mathbb{P}[\cdot]$ Probability,

$\mathbb{E}[\cdot]$ Expectation,

General Introduction

The work presented in this PhD thesis has been done within the *Information, Communications, Imagerie (ICI)* team of the *Equipes Traitement de l'Information et Systèmes (ETIS)* laboratory in Cergy, France. This thesis has been co-founded by the *Centre National de la Recherche Scientifique (CNRS)* and the *Direction Générale de l'Armement (DGA)*.

Problem statement

In the past decade, there has been a dramatic growth in the demand for new wireless services such as video transmissions and high-speed data transmissions. To meet the needs for these new services in terms of Quality of Service (QoS), highly reliable communication systems have to be designed. In modern telecommunication standards such as High Speed Downlink Packet Access (HSDPA) or Long Term Evolution (LTE) these QoS requirements are guaranteed by link layer adaptation techniques. In these telecommunication standards, link layer adaptation techniques are a combination of Adaptive Modulation and Coding (AMC) and HARQ protocols. Although, when considered separately, these techniques are now well understood; their joint optimal design is still a challenging open issue.

AMC techniques exploit Channel Side Information at the Transmitter (CSIT) to adapt the code and modulation to the channel condition. This adaptation is in general performed to guarantee a certain level of Frame Error Rate (FER) for upper-layers. The CSIT required to design an AMC is generally provided through a Channel Quality Indicator (CQI). The CQI represents the data rate supported by the channel using a quantized index representation (see e.g. [Sesia et al., 2011]). Because of the channel versatility, some errors can still occur; these errors are then handled with a HARQ protocol acting as an error control protocol.

The first objective of this thesis is to provide a theoretical framework for analysing and optimizing systems implementing an adaptive HARQ protocol in several contexts of communication.

On the other hand, the command-and-control regulation imposes each service to be allocated in an exclusive band. The main consequence of this command-and-control regulation is that almost all the frequency bands have already been allocated. This makes the wireless spectrum a scarce resource. In parallel, a report of the Federal Communications Commission (FCC)[FCC, 2002], has shown although every bands has been allocated, the spectrum is underutilized. In this context, the Cognitive Radio (CR) paradigm has been proposed in [Mitola III & Maguire, 1999]. The CR paradigm consists in allowing Secondary Users (SUs) to opportunistically access the bandwidth initially dedicated to Primary Users (PUs) (or licensed users) to increase spectrum usage.

In this paradigm, the PUs are considered as the legitimate users of the bandwidth; in consequence, the SUs have to control the degradation done to the PUs performances.

Although PUs should have devices based on modern wireless communication standards, only a few works consider that the PUs implement a HARQ protocol. However, it has been shown in [Eswaran et al., 2007] that the feedback bits used for the HARQ can be used by the SUs to evaluate the throughput-loss of the HARQ protocol of the PUs .

A second objective of this thesis is to provide a theoretical framework for optimizing the throughput of a secondary user who evaluates the throughput-loss of the primary HARQ protocol using only the information fed back by the PU over its feedback channel.

Outline and contributions

This section presents the outline of this thesis as well as the main contributions. This thesis is organized around four chapters that are briefly reviewed in the sequel.

Chapter 1

In Chapter 1, we review the main features of HARQ based telecommunication systems, presenting the main protocols of retransmissions and their performance analysis. In particular, we focus on the ACcumulated Mutual Information (ACMI) introduced in [Cheng et al., 2003] to model the HARQ protocols.

After reviewing the state of the art, this chapter ends with a first contribution for modelling an HARQ protocol:

• It is shown that HARQ protocols can be modelled efficiently using Markov chains on general (not necessarily countable) state spaces. This model allows us to link the dynamical behaviour of HARQ protocols with the Markov Decision

Process (MDP) formalism. We conclude this chapter with the proposal of the power allocation problem to maximize the throughput of HARQ protocols under peak and average power constraints.

Chapter 2

In Chapter 2 we review existing literature about MDPs with a particular attention on CMDPs. The theoretical framework of CMDPs is commonly used to handle problems where decisions are taken sequentially to maximize a long-term reward under long-term cost constraints. To cope with the applications such as resource allocation for the Physical (PHY) layer of HARQ systems, we present the CMDP framework in the general case of possibly uncountable but compact state and action spaces. In this chapter, we provide sufficient conditions for the existence of a solution to the CMDP problem.

Because we are interested in computing solutions for practical problems, we also present how solutions of CMDPs can be approximated using finite linear programs. In particular we provide conditions so that the solutions obtained through finite linear programs converge to a continuous solution.

Generally, CMDPs assume that the controller (who takes actions) can observe the state of the system that is controlled. However, in our applications for HARQ systems, the controller does not have this complete information. In consequence, in this chapter, we also introduce (constrained) Partially Observable Markov Decision Process (POMDP) and (constrained) Partial State Information Markov Decision Process (PSI-MDP) frameworks. We finally present how these partially observable problems can be solved theoretically and numerically.

- The main contribution of this chapter has been to provide a condition so that the (constrained and with continuous state and action spaces) PSI-MDP under long-term average criteria is solvable.

Chapter 3

In Chapter 3 we apply the results presented in Chapter 2 about CMDPs to the power allocation problems considered at the end of Chapter 1. This chapter has mainly two contributions:

- Our first contribution has been to verify that CMDPs constitute a suitable framework for allocating resources in situations in which CQI is continuous and conveys the whole state of the HARQ protocol. We propose simulation results to show the accuracy of the proposed method.

- Our second contribution has been to address the same power allocations problem considering 1-bit feedback. This problem has been addressed within the PSI-MDP framework. Again, we verify by simulations that the PSI-MDP method is accurate.

Chapter 4

In Chapter 4, we address the problem of allocating the power of a SU while mitigating the throughput-loss incurred to the PU. As proposed in [Eswaran et al., 2007], this throughput-loss is evaluated only by listening the feedback bits of the HARQ protocol of the PU. This contribution is twofold:

- We first derive an upper bound of the solution of the initial problem by relaxing the constraint of the throughput-loss evaluation. In particular, we assumed a "genie aided" problem assuming the complete knowledge of the PU state (defined in Chapter 1) by the SU. To solve this problem, we have proposed a modified version of the model presented in Chapter 1 that takes into account interferences generated by the SU. Finally, this optimization problem is solved using the CMDP framework.
- Then we derive a PSI-MDP framework for performing power allocation when the state of the PU is partially known. This case only requires the PU to be compliant with the SU and to broadcast its feedback bits.

In both situations, we have provided simulation results to evaluate the policies obtained numerically from the CMDP framework and the PSI-MDP framework.

Publications

International Journal Papers

IJ1 **Romain Tajan**, Charly Poulliat, Inbar Fijalkow. "Secondary Resource Allocation for Opportunistic Spectrum Sharing with IR-HARQ based Primary Users". (submitted to Transactions on Signal Processing).

International Conference Papers

IC3 **Romain Tajan**, Charly Poulliat, Inbar Fijalkow. "Interference Management for Cognitive Radio Systems Exploiting Primary IR-HARQ: a Constrained Markov Decision Process Approach". *IEEE Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, November 2012.

IC2 **Romain Tajan**, Charly Poulliat, and Inbar Fijalkow. "Opportunistic Secondary Spectrum Sharing Protocols for Primary Implementing an IR Type Hybrid-ARQ protocol".*In the IEEE proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, 15-30 March, 2012.

IC1 **Romain Tajan**, Charly Poulliat, Rodrigue Imad, and Inbar Fijalkow. "A new trellis representation for source-channel rate allocation". *In the IEEE proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 22-27 May, 2011.

National Conference Papers

NC2 **Romain Tajan**, Charly Poulliat, Inbar Fijalkow, "Allocation conjointe de puissance et rendement d'un utilisateur cognitif exploitant les retransmissions d'un utilisateur primaire : le cas du canal en Z", *GRETSI*, 2013.

NC1 **Romain Tajan**, Charly Poulliat, Inbar Fijalkow, "Coexistence de protocoles a retransmissions incrémentales pour un canal cognitif", *GRETSI*, Bordeaux, France, September 2011.

- Chapter 1 -

An overview of HARQ-based telecommunication systems

Contents

1.1 Introduction

In modern communication standards such as High Speed Downlink Packet Access (HSDPA) or Long Term Evolution (LTE), Hybrid Automatic Repeat reQuest (HARQ) protocols are used in conjunction with Adaptive Modulation and Coding (AMC) to guarantee high reliability and high data rates for wireless communications.

In the early 2000s, ACcumulated Mutual Information (ACMI) (see in particular [Caire & Tuninetti, 2001], [Cheng et al., 2003] , [Sesia et al., 2004], and references therein) has been proved to be an appropriate quantity for theoretically analysing HARQ protocols. These models are usually derived from an information theoretical perspective and assume asymptotically long Gaussian codewords. Although this hypothesis might seem restrictive at first glance, such codes has been proved to provide some general insights of what are the key parameters for the communication systems. In particular, the ACMI model has also been extended to coded modulations and Bit Interleaved Coded-Modulations (BICMs) in [Cheng, 2006]. This ACMI model was also found to be a good representation for allocation problems in [Li & Ryan, 2007], [Stiglmayr et al., 2007; Stiglmayry et al., 2008] and [Pfletschinger & Navarro, 2010]. In addition to their optimality in the asymptotic regime, these codes have been shown to provide good insights on modern codes such as Turbo codes (see Buckingham & Valenti [2008]) or Low Density Parity Check codes (see [Marcille, 2013]). The Gaussian codes framework (as in [Caire & Tuninetti, 2001]) has been chosen for a unified treatment of the concepts that are discussed in this thesis.

This chapter is organised as follows. In Section 1.2 we described the single-carrier and single-user framework that will be considered throughout this thesis (with the exception of Chapter 4 that considers multi-user scenarios). In Section 1.2, we define in particular the block-fading channel, the considered channel codes, different Channel Side Information (CSI) assumptions and the Automatic Repeat reQuest (ARQ) protocols. In Section 1.3, we review the concept of a HARQ protocol and we give the definitions of three HARQ protocols. In all cases we present a throughput analysis based on the ACMI. The main contribution of this chapter is given in Section 1.4. In Section 1.4, we model the HARQ protocol by an original Markov chain. This Markov chain will be the cornerstone for the results presented in Chapter 3 and Chapter 4.

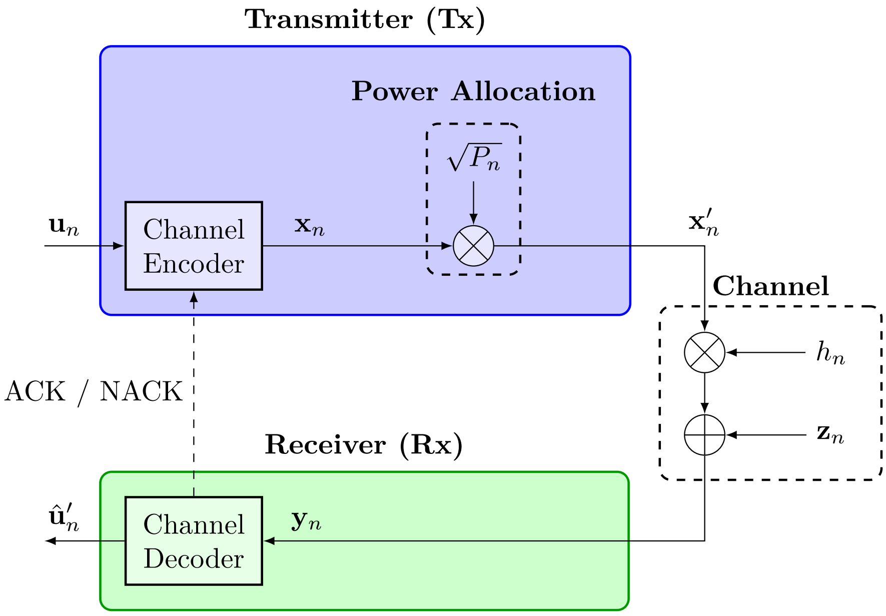

Figure 1.1 – *System model and block fading channel*

1.2 System Model

We consider the single-user, single-carrier communication system depicted in Figure 1.1. In this system, a transmitter (Tx) sends data to a receiver (Rx) over a discrete-time *wireless channel*. We suppose a slotted communication system where slots have duration T seconds. The slots are indexed by $n \in \mathbb{N}$ so that slot n happens between time nT and $(n+1)T$ (see Figure 1.2).

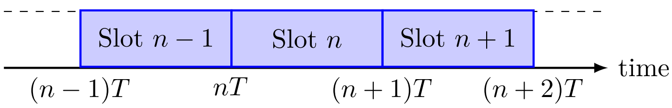

Figure 1.2 – *Time slot model.*

Suppose that the data sent by Tx is composed of *information messages* $\mathbf{u}_n$, each conveying b information bits. Tx encodes $\mathbf{u}_n$ into a coded message $\mathbf{x}_n$ of length $L = \lfloor B_W T \rfloor$ symbols where B_W is the bandwidth and T is the duration of the slot. We suppose here that every possible coded message has unit power. Before sending $\mathbf{x}_n$ over the channel, Tx applies its power allocation by multiplying $\mathbf{x}_n$ by $\sqrt{P_n}$. We suppose that the power allocation is defined as a sequence $\{P_n\}_{n\in\mathbb{N}}$ is known to Tx and Rx. The received signal at Rx is denoted by $\mathbf{y}_n$. This signal is decoded into an information message denoted by $\hat{\mathbf{u}}_n$.

1.2.1 Channel model

For modelling a wireless communication, the channel is often modelled as a *block-fading channel* (see [Tse & Viswanath, 2005] and references therein). In this model, the channel is assumed to be constant for the whole duration of the slot. The channel gains $\{H_n\}_{n\in\mathbb{N}}$ are a discrete time independent and identically distributed (i.i.d.) stochastic process.

Let $\mathbf{X}_n$ be the coded message considered as a complex Gaussian random vector of dimension L with i.i.d.components of zero mean and unit variance. When $\mathbf{X}_n$ is sent over the block-fading channel with power p_n, the received signal $\mathbf{Y}_n$ at Rx is given by

$$\mathbf{Y}_n = \sqrt{P_n} H_n \mathbf{X}_n + \mathbf{Z}_n, \tag{1.1}$$

where the noise vector $\mathbf{Z}_n$ is a length L complex Gaussian random vector with i.i.d. components with zero mean and variance σ_z^2.

In this thesis we will consider a Rayleigh block-fading model. This model represents well situations where there is no line of sight between Rx and Tx. In this case $\{H_n\}_{n\in\mathbb{N}}$ can be considered as a Gaussian process. In this case, the phase of H_n is uniformly distributed in $[0, 2\pi]$, the modulus of H_n is distributed according to a Rayleigh distribution with parameter $\sqrt{\bar{\alpha}}$:

$$f_{|H|}(x) = \frac{2x}{\bar{\alpha}} \exp\left(\frac{-x^2}{\bar{\alpha}}\right), \; x \geq 0$$

and the fading power $\alpha_n = |H_n|^2$ is exponentially distributed with mean $\bar{\alpha}$:

$$f_{\alpha}(x) = \frac{1}{\bar{\alpha}} \exp\left(\frac{-x}{\bar{\alpha}}\right), \; x \geq 0.$$

Other distributions such as the Nakagami-m distribution or Rice distribution can be considered in the literature but are out the scope of this thesis.

1.2.2 Channel State Information

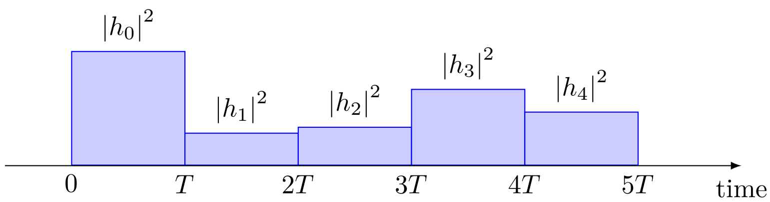

Figure 1.3 – *Block fading channel*

CSI is a cornerstone for communication systems since it defines the knowledge of the communicating devices to operate. In the system presented in Figure 1.1, CSI is defined as the knowledge of H_n at time index n. We call Channel Side Information at the Transmitter (CSIT) or Channel Side Information at the Receiver (CSIR), the fact that Tx or Rx has CSI, respectively. In some references (see for example [Tuninetti, 2011]), this CSI is called Causal CSI because only the current H_n is known but not the future ones $\{H_t\}_{t\geq n+1}$.

In this thesis, Causal CSI and CSI are identical, however we use the term Outdated Channel Side Information (OCSI) for the cases where some information about $\{H_t\}_{t\leq n-1}$ is available but H_n is not. This term is used in particular for resource allocation in the presence of HARQ protocols (see for example [Szczecinski et al., 2011]).

In some situations, H_n is not known but its distribution is. This will be referred to as Statistical Channel Side Information (SCSI). In the literature (see for example [Goldsmith, 2005]), SCSI is sometimes referred to as Channel Distribution Information.

1.2.3 Asymptotically optimal Channel Coding

In the model presented in Figure 1.1, errors introduced by the channel are corrected using *channel coding*. In this section, we suppose that there is CSIR but not CSIT.

Let us consider that every codeword $\mathbf{x}$ at the output of the channel coder span N_T slots and is a complex Gaussian random vector of dimension $N_T L$ with i.i.d. components of zero mean and unit variance. Let $\mathbf{P} = (P_0, P_1, \dots P_{N_T-1})$ be the vector such that P_j is the power allocated for the j^{th} slot. Suppose that $\mathbf{x}$ is split into N_T blocks of size L: $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \dots \mathbf{x}_{N_T-1})$. Each block is then sent over the block fading channel.

For this channel, it has been shown in [Caire & Tuninetti, 2001] that, in the asymptotic regime ($L \to \infty$) the Frame Error Rate (FER) is given by

$$p_f(N_T) = \mathbb{P}\left[\sum_{j=0}^{N_T-1} C\left(P_j \frac{|H_j|^2}{\sigma_z^2}\right) < R\right], \tag{1.2}$$

where $C(x)$ is the capacity of complex valued Additive White Gaussian Noise (AWGN) channel $C(x) = \log_2(1+x)$, R is the equivalent *information rate* of the first transmission given by $R = b/L$, b is the number of information bits conveyed in the information messages $\mathbf{u}$ and σ_z^2 is the noise variance. This probability is referred to as *outage probability* or *information outage probability* (see [Knopp & Humblet, 2000], [Caire & Tuninetti, 2001], [Sesia et al., 2004], [Cheng, 2006] and references therein).

Remarks

• In [Cheng, 2006], the quantity $\sum_{j=0}^{N_T-1} C\left(P_j \frac{|H_j|^2}{\sigma_z^2}\right)$ is referred to as the ACMI. Using the ACMI definition, the author has shown that an equation similar to equation (1.2) holds for modulation constrained AWGN channel (the modulation may change at every slot). In this case, we replace C by the capacity of the corresponding constrained input AWGN channel for each slot. In the same paper, this result is extended to BICM.

• In[Caire & Tuninetti, 2001], equation (1.2) is shown using a typical set decoder. This decoder is suboptimal, (compared to the Maximum Likelihood (ML) decoding rule) in terms of FER for finite length codes, however, it is asymptotically optimal (see [Cover & Thomas, 2006]). The authors have also pointed out two other advantages for HARQ protocols. The first one is that, asymptotically ($L \to \infty$), this decoder performs a perfect error detection. In other words, asymptotically, there is a null probability of not detecting an error. Note that this step is usually performed by adding Cyclic Redundancy Check (CRC) bits to the information message when considering practical channel coding schemes and decoders.

The second advantage is that this decoder can handle puncturing. Suppose that we only send the following blocks: $\mathbf{x}_0, \mathbf{x}_1, \dots \mathbf{x}_{m-1}$. At the decoder, Rx can complete its received message by $N_T - m$ dummy signal blocks $\boldsymbol{\xi}_m, \dots \boldsymbol{\xi}_{N_T-1}$ as long as they are generated independently from the received signal. After this step, Rx uses $\tilde{\mathbf{y}} = \left(\mathbf{y}_0, \mathbf{y}_1, \dots \mathbf{y}_{m-1}, \boldsymbol{\xi}_m, \dots \boldsymbol{\xi}_{N_T-1}\right)$ to decode. This method is similar to classical puncturing techniques; in this case, the FER becomes

$$p_f(m) = \mathbb{P}\left[\sum_{j=0}^{m-1} C\left(P_j \frac{|H_j|^2}{\sigma_z^2}\right) < R\right]. \tag{1.3}$$

From equation (1.2) we observe that, in a block-fading channels, it is impossible to have a zero probability of error even though $L \to \infty$. Outage events will always happens due to deep fades. To ensure reliability of the communication, ARQ based protocols are used.

1.2.4 Automatic Repeat reQuest

These protocols enable the end-to-end delivery of information packets by allowing the same information packet to be sent multiple times. Because the fading coefficients are i.i.d., by retransmitting the same packet multiple times, one can expect to experience a good fade on one of the retransmissions. Consequently, for the ARQ protocol, a feedback channel is necessary to inform the transmitter of decoding errors.

In its simplest implementation, the ARQ protocol only exploits error detection, no channel coding is required. Let $\mathbf{u}_n$ be the information message of b bits that Tx has to send to Rx in slot n. Rx encodes $\mathbf{u}_n$ into a codeword of L symbols $\mathbf{x}_n$ with a channel code that is only used for error detection; Rx receives the corresponding noisy version $\mathbf{y}_n$ and checks the integrity of $\hat{\mathbf{u}}_n$ ($\hat{\mathbf{u}}_n = \mathbf{u_n}$) using the error detection capability of the channel code. If no error is detected, Rx sends an ACKnowledgement (ACK) bit to Tx using a feedback channel. At the reception of this ACK bit, Tx starts the transmission of a new information packet. On the other hand, if an error is detected, Rx sends a Negative ACKnowledgement (NACK) bit over the feedback channel. At the reception of this NACK bit, Tx sends the same message ($\mathbf{x}_{n+1} = \mathbf{x}_n$) over the channel at slot $n+1$. At Rx side, $\mathbf{y}_n$ is discarded and only the new received message $\mathbf{y}_{n+1}$ is decoded. If no error is detected, Rx sends the ACK bit to Tx. If an error is detected, Rx sends the NACK bit to Tx. This protocol continues until no error is detected at the receiver.

Performance evaluation of the ARQ protocol

By definition, the ARQ protocol makes the transmission of information packets reliable since Rx will always end up decoding correctly (if an infinite number of retransmissions is allowed). On the other hand, this reliability is earned at the expense of *delay*. Indeed, in the definition of the ARQ protocol, it may happen that the protocol retransmits a large number of times a packet before decoding it correctly. The average time taken by the ARQ protocol to correctly transmit a packet is called *delay* and is one figure of merit of the performance of ARQ protocols.

A trade-off can be achieved between reliability and delay by imposing a maximum number of retransmissions. This case is often referred to as *truncated ARQ* in the literature. Let $N_T - 1$ be this maximum number of retransmissions (N_T takes also into account the initial transmission). In this case, the overall protocol may fail. This is called an *outage event*. The probability of such event is commonly referred to as outage probability (see Caire & Tuninetti [2001]).

Definition 1.1. The *outage probability* that the ARQ protocol fails after $N_T - 1$ retransmissions is called the *outage probability*

To be completely fair, we have to re-define the delay to take into account the finiteness of N_T. Several definitions of delays may be found in the literature. In this thesis we choose a definition of [Le Duc, 2010]:

Definition 1.2. The *delay* is the mean time between the beginning and the end of an ARQ protocol when a successful decoding occurs.

The third figure of merit usually presented in the literature to evaluate the performance of ARQ protocols is the *throughput*. In this thesis, we will adopt the definition of throughput given in [Lin & Costello, 2004]:

Definition 1.3. The *throughput* is "the ratio of the average number of information digits successfully accepted by the receiver per unit of time to the total number of digits that could be transmitted per unit of time".

Let $b_c(n)$ be defined as the total number of bits correctly decoded up to slot n, the throughput is given by

$$\eta = \frac{\lim_{n\to\infty} \frac{b_c(n)}{nT}}{\frac{L}{T}} = \lim_{n\to\infty} \frac{b_c(n)}{nL} \text{ in b/s/Hz.} \tag{1.4}$$

The computation of this throughput is classically done using the renewal and reward theorem as proposed in [Zorzi & Rao, 1996]. Let $\mathcal{E}$ be the event defined as: "Tx starts the transmission of a new information packet". A random reward $R_{\mathcal{E}}$ is attached to this recurrent event. This random variable is $R_{\mathcal{E}} = b$ (the number of bits contained in $\mathbf{u}_n$) if the packet is successfully decoded by Rx and $R_{\mathcal{E}} = 0$ otherwise. Relying on this event and this reward the renewal-reward theorem states that (see [Ross, 2006])

$$\eta = \frac{\mathbb{E}\left[R_{\mathcal{E}}\right]}{\mathbb{E}\left[D_{\mathcal{E}}\right]}, \tag{1.5}$$

where $\mathbb{E}\left[R_{\mathcal{E}}\right]$ is the expected reward earned in the recurrent event and $D_{\mathcal{E}}$ is the random time between two successive occurrences of $\mathcal{E}$. In the description of the ARQ protocol, we have assumed that once Tx finishes transmitting a packet, it starts the transmission of a new information packet. Under this assumption, $D_{\mathcal{E}}$ is also defined as the time between the start and the end of the transmission of an information packet.

We now further develop the expression of the throughput given in equation (1.5). We introduce $\mathcal{O}_n$ as the following event: "the decoding fails after n transmission attempts". Let $p_s(n)$ and $p_f(n)$ be the probability of decoding success in exactly n transmission attempts and the probability of decoding failure after n transmission attempts. Because of the sequential aspect of the ARQ protocol, these two probabilities $p_s(n)$ and $p_f(n)$ can be written using the event $\mathcal{O}_n$ as follows:

$$p_f(n) = \mathbb{P}\left[\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_{n-1}, \mathcal{O}_n\right] \tag{1.6}$$

$$p_s(n) = \mathbb{P}\left[\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_{n-1}, \overline{\mathcal{O}_n}\right] \tag{1.7}$$

On the other hand, one can note that $R_{\mathcal{E}} = 0$ if and only if the decoding fails after N_T attempts. The probability that $R_{\mathcal{E}} = 0$ is then $p_f(N_T)$, and the probability that $R_{\mathcal{E}} = b$ is then $1 - p_f(N_T)$. Furthermore, because of the slotted nature of the transmission, $D_{\mathcal{E}}$ can only take the values jL where $j \in \{1, \ldots, N_T\}$. By the definition of $D_{\mathcal{E}}$, one can remark that $D_{\mathcal{E}} = jL$ with $j < N_T$ happens only if the decoding is successful after j attempts, hence this event has a probability $p_s(j)$. Finally $D_{\mathcal{E}} = N_T L$ if and only if there has been a failure after $N_T - 1$ attempts; the probability of this event is $p_f(N_T - 1)$. Finally, the throughput is expressed as follows

$$\eta = \frac{b}{L} \frac{1 - p_f(N_T)}{\sum_{l=1}^{N_T-1} l p_s(l) + N_T p_f(N_T - 1)}. \tag{1.8}$$

This equation can be further simplified by introducing the information rate equivalent to the first transmission $R = b/L$ and remarking that

$$\begin{cases} p_s(n) = p_f(n-1) - p_f(n), \ n > 0 \\ p_f(0) = 1 \end{cases} \tag{1.9}$$

Introducing (1.9) in (1.8), we obtain

$$\eta = R \frac{1 - p_f(N_T)}{1 + \sum_{l=1}^{N_T-1} p_f(l)}. \tag{1.10}$$

In this thesis, we will only consider the throughput as a main figure of merit for the performance evaluation of HARQ protocols. However one can refer to [Zorzi & Rao, 1996], [Caire & Tuninetti, 2001], [Sesia et al., 2004], [Le Duc, 2010], [Marcille, 2013], and references therein to have a more complete view and analysis of other performance criteria. We now present some practical issues that have not been considered in the previous analysis.

Practical issues

The description of the ARQ protocol as given above does not taken into account possible imperfections encountered in practice. In practice, several additional delays occur. These delays, depicted in Figure 1.4, are the following:

- τ_p: the one-way delay, which represents the delay incurred by the propagation time between Tx and Rx,
- τ_d: the data processing delay, which represents the delay incurred by the

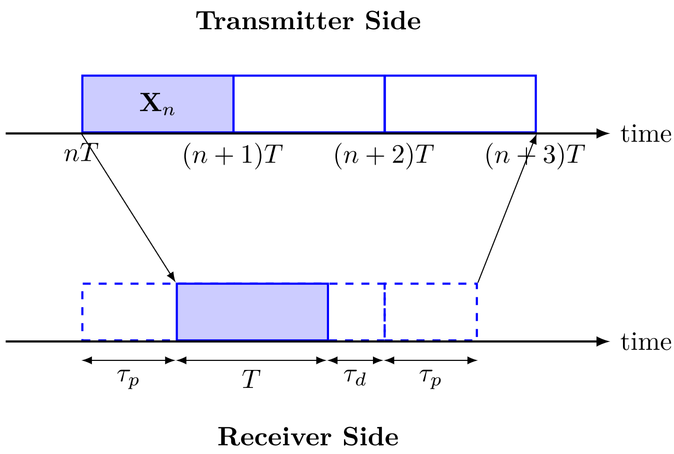

Figure 1.4 – *Delays in the ARQ protocol.*

time that takes Rx for demodulating, decoding, and detecting errors.

The protocol presented at the beginning of this section is called the *Stop-and-Wait* protocol: when Tx has sent one block, it waits to receive the acknowledgement bit to continue its transmission. This protocol is obviously inefficient. This inefficiency can be observed by computing the effective throughput. Let τ be the overhead time for the transmission of one block defined as $\tau = 2\tau_p + \tau_d$. From [Lin & Costello, 2004] we get

$$\eta_{sw} = \frac{T}{T+\tau}\eta, \tag{1.11}$$

where η is the throughput computed without taking into account delays. Note that in the situation depicted in Figure 1.4, we have $\tau = 2T$ and $\frac{T}{T+2\tau_p+\tau_d} = \frac{1}{3}$ so the effective throughput is divided by three compared to the case $\tau = 0$.

To overcome the impact of these delays, several protocols have been proposed in the literature. In particular, in [Lin & Costello, 2004] the Go-Back-N protocol and the Selective and Repeat protocol are described. The main result that has been proved in the literature is the following: under an assumption of infinite buffer at Rx, the throughput of the Selective and Repeat is equal to the throughput of a Stop-and-Wait protocol with $\tau = 0$ (see [Lin & Costello, 2004], [Aoun, 2012], [Marcille, 2013] and references therein). In consequence, in this thesis we will consider a Stop-and-Wait protocol with $\tau = 0$.

Other imperfections can be considered in the ARQ protocol. In particular one may consider that error may occur on the feedback channel. In this context, cross layer optimization techniques have been proposed in [Marcille et al., 2011] and [Marcille et al., 2012b]. One can of course think of mixing the imperfections.

The Go-Back-N protocol under imperfect feedback has been studied in [Zorzi & Rao, 1996].

Through this thesis, we will always suppose that Assumption 1.1 holds. In consequence, every throughput that will be computed in this thesis has to be understood as an upper bound on the possible throughput if imperfections were taken into account.

Assumption 1.1.
(a) *The feedback channel is* instantaneous *and* error-free*.*
(b) *The probability of an undetected error and the probability of detecting an error while the decoding is successful are* 0*.*
(c) *Tx is backlogged: it always has an information packet to transmit in its queue.*

1.3 HARQ Protocols

The main advantage of the ARQ protocol is that it guarantees a high reliability of the communication link. Indeed, when an infinite number of retransmissions is allowed, there is a null outage probability. However, the ARQ protocol does not benefit, neither from error correcting codes, nor from information contained in the earlier transmissions (since it discards the outdated message before decoding the new one) to decrease the probability of successfully decoding a packet.

In this Section, we will study three different ways of combining Forward Error Correction (FEC) and ARQ. This combining approach is referred to as HARQ, where the term "Hybrid" indicates that both FEC and ARQ are used. The first type of HARQ protocol is the so-called Type-I HARQ protocol and is analysed in Subsection 1.3.1. The second method to improve the throughput is to use the information contained in the first retransmissions to decrease the probability of error in the next retransmissions. These protocols are referred to as Type-II HARQ protocols. There exist two main approaches for Type II-HARQ protocols: the Type II-Chase Combining (CC)-HARQ protocol (denoted CC-HARQ in the sequel) and the Type II-Incremental Redundancy (IR)-HARQ protocol (denoted IR-HARQ in the sequel). The CC-HARQ protocol and the IR-HARQ protocol are described in Section 1.3.2 and Section 1.3.3 respectively.

1.3.1 Type-I HARQ protocol

In order to improve the throughput of the ARQ protocol one can think of using FEC to send coded packets instead of non-coded packets. It decreases the outage probability and thus increases the throughput. This newly defined protocol is called the Type-I HARQ protocol. The Type-I HARQ protocol is then equivalent to the truncated ARQ protocol except that the Type-I HARQ protocol uses FEC to correct errors.

In this case, we consider that the codewords of Gaussian codebooks of Section 1.2.3 span one slot. Other schemes could be considered with the same qualitative conclusions. Because the codeword is assumed to span only one slot, and because the j^{th} attempt is independent of every other attempt, the probability that the j^{th} attempt of decoding the information packet fails is given by

$$p_o(j) = \mathbb{P}\left[\log_2\left(1 + \frac{P_j\,|H_j|^2}{\sigma_z^2}\right) < R\right], \tag{1.12}$$

where P_j is the power used on slot j and $R = b/L$. Since all the retransmissions are independent, the probability of failure after n attempts $p_f(n)$ (defined in (1.6)) is given by

$$p_f(n) = \prod_{j=1}^{n} \mathbb{P}\left[\log_2\left(1 + \frac{P_j\,|H_j|^2}{\sigma_z^2}\right) < R\right]. \tag{1.13}$$

One can remark that what has been presented in Section 1.2.4 for the computation of the throughput for the ARQ protocol can also be applied for the Type-I HARQ protocol. The throughput of the Type-I HARQ protocol is finally expressed using equation (1.10) as follows:

$$\eta_{t1} = R\frac{1 - \prod_{j=1}^{N_T} \mathbb{P}\left[\log_2\left(1 + \frac{P_j\,|H_j|^2}{\sigma_z^2}\right) < R\right]}{1 + \sum_{n=1}^{N_T-1}\prod_{j=1}^{n} \mathbb{P}\left[\log_2\left(1 + \frac{P_j\,|H_j|^2}{\sigma_z^2}\right) < R\right]}. \tag{1.14}$$

This expression simplifies if we consider a constant power allocation $P_1 = P_2 = \cdots = P_{N_T} = P$ as follows

$$\eta_{t1} = R\frac{1 - \mathbb{P}\left[\log_2\left(1 + \frac{P\,|H|^2}{\sigma_z^2}\right) < R\right]^{N_T}}{1 + \sum_{n=1}^{N_T-1} \mathbb{P}\left[\log_2\left(1 + \frac{P\,|H|^2}{\sigma_z^2}\right) < R\right]^{n}}, \tag{1.15}$$

where time indexed in H has been dropped since the H_j are i.i.d.. After some calculations (see [Le Duc, 2010]) we obtain

$$\eta_{t1} = R\left(1 - \mathbb{P}\left[\log_2\left(1 + \frac{P\left|H\right|^2}{N_0}\right) < R\right]\right). \tag{1.16}$$

One can remark that the throughput of the Type-I HARQ protocol is equal to the throughput of the same FEC code used without ARQ. However, the Type-I HARQ protocol exploits the ARQ to improve the reliability of its transmission compared to pure FEC. Indeed, the outage probability of the Type-I HARQ protocol is $\mathbb{P}\left[\log_2\left(1 + \frac{P|H|^2}{\sigma_z^2}\right) < R\right]^{N_T}$ whereas it is $\mathbb{P}\left[\log_2\left(1 + \frac{P|H|^2}{\sigma_z^2}\right) < R\right]$ for a transmission using FEC only.

1.3.2 Type-II Chase Combining (CC) HARQ protocol

The major drawback of a Type-I HARQ protocol is that it does not benefit from the different retransmissions. Indeed, all retransmissions are independent and independently processed at Rx. This makes the Type-I HARQ protocol highly inefficient. To improve the efficiency of Type-I HARQ protocol other protocols have been proposed with more complex processing at Rx. We here consider one of these protocols named the Type-II CC HARQ protocol (noted CC-HARQ for the rest of this thesis).

Protocol definition

The main difference between Type-I HARQ and CC-HARQ protocols is the processing done at Rx. From the point of view of Tx, nothing is changed. Tx always sends the same codeword among all the retransmissions. For ease of presentation, suppose that the transmission of the current packet has begun at slot 0 and the current time is nT so that the n attempts to decode has been done without a positive result.

Let $\{\mathbf{Y}_0, \mathbf{Y}_1, \ldots, \mathbf{Y}_{n-1}\}$ be the received signals for slots 0 to $n-1$. Whereas the Type-I HARQ protocol only considers $\mathbf{Y}_{n-1}$ at the decoder, the CC-HARQ considers the following combined signal

$$\mathbf{R}_n = \sum_{j=0}^{n-1} g_j \mathbf{Y}_j, \tag{1.17}$$

where g_j are given complex coefficients. Assuming CSIR the optimal choice (see [Brennan, 1959]) for the coefficients g_j is called Maximum Ratio Combining (MRC) and is given by:

$$g_j = \frac{H_j^*}{\sigma_z^2}, \tag{1.18}$$

where $(\cdot)^*$ stands for the complex conjugation. Using equation (1.18) in (1.17) and replacing $\mathbf{y}_j$ by its expression given in equation (1.1), we obtain

$$\mathbf{R}_n = \mathbf{X}\sum_{j=0}^{n-1} \frac{\sqrt{P_j}|H_j|^2}{\sigma_z^2} + \sum_{j=0}^{n-1} \frac{H_j^*}{\sigma_z^2}\mathbf{Z}_j, \tag{1.19}$$

At this point, the Signal to Noise Ratio (SNR) at Rx is denoted γ_n and is defined as

$$\gamma_n = \sum_{j=0}^{n-1} \frac{|H_j|^2 P_j}{\sigma_z^2}. \tag{1.20}$$

One can note here that the CC-HARQ protocol "accumulates" SNR; indeed, the SNR given by equation (1.20) can be rewritten as the following recursion:

$$\begin{cases} \gamma_0 = 0, \\ \gamma_{n+1} = \gamma_n + \dfrac{|H_n|^2 P}{\sigma_z^2}. \end{cases} \tag{1.21}$$

This accumulation result is classical in the literature, see for example [Cheng, 2006].

Performance Analysis

To compute the throughput of the CC-HARQ protocol, we will again use equation (1.10). Note that the throughput expression given in equation (1.10) is dependent on the type of HARQ protocols only through the $p_f(n)$, $n \in \{1, \dots, N_T\}$. We will then compute the throughput by computing these probabilities of failure $p_f(n)$.

Again we present results based on the FEC scheme presented in Section 1.2.3. The only difference between the Type-I HARQ and the CC-HARQ protocol is that the CC-HARQ protocol considers $\mathbf{R}_n$ (the combined signal after n attempts) instead of $\mathbf{Y}_n$ (the last received signal) to decode. Let $R = b/L$ be the information rate corresponding to the initial transmission. The probability of decoding failure after n transmission attempts $p_f(n)$ is computed as follows:

$$\begin{aligned} p_f(n) &= \mathbb{P}\left[\log_2\left(1 + \sum_{j=0}^{n-1} \frac{P_j|H_j|^2}{\sigma_z^2}\right) < R\right] \\ &= \mathbb{P}\left[\sum_{j=0}^{n-1} \frac{P_j|H_j|^2}{\sigma_z^2} < 2^R - 1\right] \end{aligned} \tag{1.22}$$

In the sequel we denote by $\Gamma = 2^R - 1$. Γ is referred to as the decoding threshold for Rx. The throughput of the CC-HARQ protocol is again computed using the renewal and reward theorem. The expression of the throughput is then given by equation (1.5):

$$\eta_{cc} = R\frac{1 - p_f(N_T)}{1 + \sum_{l=1}^{N_T-1} p_f(l)}, \tag{1.23}$$

where $R = b/L$. In this case, the p_f probabilities are given by equation (1.22). In the case of constant power allocation, the expressions of the different $p_f(n)$ are

$$p_f(n) = \mathbb{P}\left[\sum_{j=0}^{n-1} \frac{P|H_j|^2}{\sigma_z^2} < \Gamma\right]. \tag{1.24}$$

To evaluate this expression, one can remark that, because the $|H_j|^2$ are i.i.d. and exponentially distributed, $\sum_{j=0}^{n-1} |H_j|^2$ is gamma-distributed: $\sum_{j=0}^{n-1} |H_j|^2 \sim \text{Gamma}(n, 1/\lambda)$. A closed-form expression of $p_f(n)$ obtained thanks to the this Gamma distribution can be obtained but is out of the scope this thesis.

1.3.3 Type-II Incremental Redundancy (IR) HARQ protocol

The IR-HARQ protocol is different from the Type-I HARQ and the CC-HARQ protocols in the sense that Tx does not send the same packet over the channel. We will now describe the IR-HARQ protocol.

In this section, Tx sends information packets $\mathbf{u}$ of b information bits. The IR-HARQ protocol considers a FEC scheme of Section 1.2.3 where the codewords spans N_T slots. Similarly to the description done in Section 1.2.3, each codeword $\mathbf{x}$ is divided into N_T blocks of equal length:

$$\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_{N_T-1}). \tag{1.25}$$

The main difference between the IR-HARQ protocol and the FEC scheme proposed in 1.2.3 is as follows: the IR-HARQ protocol intends to decode after each slot whereas the FEC scheme of Section 1.2.3 decodes only after the N_T slots.

For ease of presentation, we suppose that the transmission of the current packet begins at slot 0. Tx sends $\mathbf{x}_0$ over the channel. Rx receives the signal $\mathbf{y}_0$ and decodes. Using equation (1.3), the average probability of error is given by

$$p_f(1) = \mathbb{P}\left[\log_2\left(1 + \frac{P_0|H_0|^2}{\sigma_z^2}\right) < R\right]. \tag{1.26}$$

In case of successful decoding Rx sends an ACK bit to Tx who starts the transmission of the next information packet. If the decoding fails, Rx sends a NACK bit to Tx who sends $\mathbf{x}_1$ over the channel. At the reception of $\mathbf{x}_1$, Rx builds $\mathbf{r}_1 = (\mathbf{y}_0, \mathbf{y}_1)$ and tries to decode $\mathbf{r}_1$. Using again equation (1.3), the average probability of error is obtained as

$$p_f(2) = \mathbb{P}\left[\sum_{j=0}^{1} \log_2\left(1 + \frac{P_j |H_j|^2}{\sigma_z^2}\right) < R\right]. \tag{1.27}$$

More generally, the probability of failure after n attempts is given by

$$p_f(n) = \mathbb{P}\left[\sum_{j=0}^{n-1} \log_2\left(1 + \frac{P_j |H_j|^2}{\sigma_z^2}\right) < R\right]. \tag{1.28}$$

From this equation, one can observe that the IR-HARQ protocol accumulates mutual information. In Cheng [2006],

$$i_n = \sum_{j=0}^{n-1} \log_2\left(1 + \frac{|H_j|^2 P}{\sigma_z^2}\right) \tag{1.29}$$

is referred to as ACMI. The throughput of the IR-HARQ protocol is again given by

$$\eta_{ir} = R \frac{1 - p_f(N_T)}{1 + \sum\limits_{l=1}^{N_T - 1} p_f(l)}.$$

where $R = b/L$ and $p_f(n)$ is given by (1.28). Note that in this case, the expression $p_f(n)$ cannot be further simplified for a constant power allocation.

In this section, we defined the Type-I HARQ protocol, the CC-HARQ protocol and the IR-HARQ protocol. In the three cases, we have shown that the throughput can be expressed as

$$\eta = R \frac{1 - p_f(N_T)}{1 + \sum_{l=1}^{N_T - 1} p_f(l)},$$

where $R = b/L$, and the probabilities $p_f(n)$ are given by equation (1.13) for the Type-I HARQ case, by equation (1.22) for the CC-HARQ protocol, and by equation (1.28) for the IR-HARQ protocol.

1.4 A unified analysis based on Markov chains

In the precedent section, we have done the throughput analysis for the Type-I HARQ protocol, the CC-HARQ protocol, and the IR-HARQ protocol. These three analyses are based on the renewal and reward theorem as proposed in

[Zorzi & Rao, 1996]. The main drawback of these analyses is that, except in the Type-I HARQ case, we are not able to derive closed-form expressions for the probabilities $p_f(n)$ for all $n \in 1, \ldots, N_T$ (given by (1.22) and (1.28)). Some closed-form expressions exist (see e.g. [Chaitanya & Larsson, 2011, 2013]), however they are often difficult to compute and to handle. In this thesis we propose a different approach based on a controlled Markov chains with possibly infinite and uncountable state space. Our approach is based on the following observation:

the three HARQ protocols are finally driven by an accumulation of some quantities (SNR, mutual information).

1.4.1 Controlled Markov Chain associated with the Type-I HARQ protocol

Firstly, we focus our study on the Type-I HARQ protocol. Let K_n be a random variable representing the number of attempts done by Tx to transmit its current information packet. For each time nT, K_n can take one of the following $N_T + 1$ values:

- **$'0'$:** Tx starts the transmission of a new information packet after a successful decoding (ACK),
- **$'1'$:** Tx has done 1 attempt and a NACK bit is received,
- **$'2'$:** Tx has done 2 attempts and a NACK bit is received,

⋮

- **$'N_T'$:** Tx has done N_T attempts and NACK is received, this state corresponds to an outage event and, in consequence, to the start of the transmission of a new information packet.

For the sake of completeness, suppose that the power allocation $\pi = \{P_n\}_{n\in\mathbb{N}}$ is a random process so that in each slot n, Tx sends $\mathbf{x}_n$ with a random power P_n. We are now interested in showing that the random process $\{K_n\}$ verifies the following Markov property

$$\begin{aligned}\mathbb{P}\left[K_{n+1} = k_{n+1} | P_n = p_n, K_n = k_n, P_{n-1} = p_{n-1}, K_{n-1} = k_{n-1}, \ldots\right] = \\ \mathbb{P}\left[K_{n+1} = k_{n+1} | P_n = p_n, K_n = k_n\right]\end{aligned} \tag{1.30}$$

Firstly, we introduce the Gain to Noise Ratio (GNR) which is defined similarly in [Marcille et al., 2012a] as follows:

$$\alpha_n = \frac{|H_n|^2}{\sigma_z^2}. \tag{1.31}$$

Let us now interpret the left-hand side of equation (1.30). For the sake of simplicity, suppose first that $N_T > k_n > 0$. By definition, $K_n = k_n$ means that, at the beginning of slot n, Tx had done k_n attempts and that a retransmission is required by Rx (a NACK bit is received). In this case, we have shown that Tx sends the same codeword over the channel. Only two situations can occur: Rx fails to decode its received signal or Rx succeeds in decoding its received signal. We have seen in the preceding section, that the blocks sent over the channel are independent of each other and that (conditionally on $P_n = p_n$) the probability of decoding failure by Rx is given in equation (1.12) and slightly modified to introduce α_n and to stress the dependence on p_n as follows:

$$p_o(p_n) = \mathbb{P}\left[\log_2\left(1 + p_n\alpha_n\right) < R\right]. \tag{1.32}$$

From this probability, we can prove that if the random variable $\Delta(\alpha_n, p_n) = \log_2\left(1 + p_n\alpha_n\right)$ is smaller than R, Rx fails in decoding the information packet. Hence, K_{n+1} will be: 'Tx has done $k_n + 1$ attempts and a NACK bit is received'. The second possible scenario is $\Delta(\alpha_n, p_n) \geq R$. In this case Rx succeeds in decoding the information packet and K_{n+1} will be 'Tx starts the transmission of a new information packet after a successful decoding (ACK)'. More formally it means that

$$\mathbb{P}\left[K_{n+1} = k_{n+1} | P_n = p_n, K_n = k_n, P_{n-1} = p_{n-1}, K_{n-1} = k_{n-1}, \ldots\right] = \begin{cases} \mathbb{P}\left[\Delta(\alpha_n, P_n) \geq R | P_n = p_n\right] \text{if } k_{n+1} = 0 \\ \mathbb{P}\left[\Delta(\alpha_n, P_n) < R | P_n = p_n\right] \text{if } k_{n+1} = k_n + 1 \\ 0 \text{ otherwise.} \end{cases} \tag{1.33}$$

It proves that when $N_T > k_n > 0$, equation (1.30) is verified. Similarly to the proof of the case $N_T > k_n > 0$, we can easily derive the cases $k_n = 0$ and $k_n = N_T$. These proofs are omitted but Table 1.1 summarizes the possible values for K_{n+1} depending on k_n (lines) and on the random variable $\Delta(\alpha_n, P_n)$ (columns).

Finally, one can remark that equation (1.30) is similar to a Markov property except that the probability of transition are dependent on $\{P_n\}_{n\in\mathbb{N}}$. In consequence, we can represent the transition graph of this Markov chain by a graph similar to the one given in Figure 1.5. On this graph, circles represent the possible states for K_n and arrows represents the possible transitions given in Table (1.30).

	$\Delta(\alpha_n, p_n) < D_T$	$\Delta(\alpha_n, p_n) \geq D_T$
$k_n < N_T - 1$	$k_n + 1$	0
$k_n = N_T - 1$	N_T	0
$k_n = N_T$	1	0

Table 1.1 – *Table of rules for the transitions from the state $s_n = k_n$ to s_{n+1} when power Tx communicates at power p_n.*

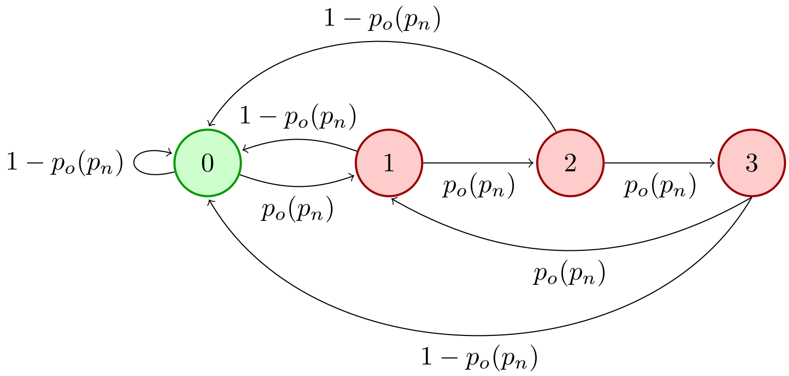

Figure 1.5 – *State Diagram of the Markov chain $\{K_n\}_{n\in\mathbb{N}}$ for a Type-I HARQ protocol with $N_T = 3$.*

Hitherto, we have proposed a controlled Markov chain that represents the state evolution of the random process K_n that model the *state* of the Type-I HARQ protocol. Let $r(K_n)$ be the random variable defined as

$$r(K_n) = \begin{cases} b \text{ if } K_n = 0 \\ 0 \text{ otherwise.} \end{cases} \tag{1.34}$$

From the definition of K_n it is obvious that $r(K_n)$ is the same as defining $R_{\mathcal{E}}$. Indeed, the event $\mathcal{E}$ has been defined in Section 1.2.4 as the start of a transmission of a new information packet and $R_{\mathcal{E}} = b$ only after successful decoding of Rx. Because of this equivalence, the throughput of the controlled Type-I HARQ protocol is rewritten as follows

$$\eta(\pi) = \frac{1}{L} \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} r(K_n),$$

where $\pi = \{P_n\}_{n\in\mathbb{N}}$ is added to show the dependence of the throughput on π (see Table 1.1 or Figure 1.5). We next show that a very similar analysis can be done for the CC-HARQ and the IR-HARQ protocols.

1.4.2 The state of Type-II HARQ protocols

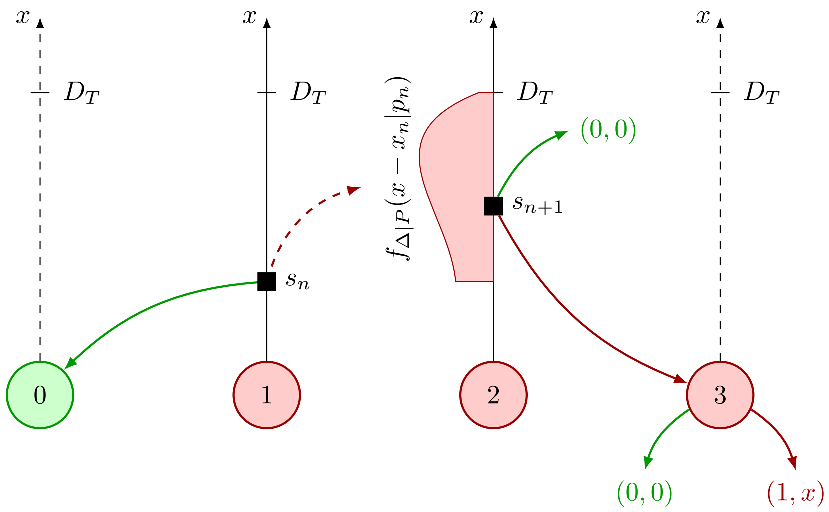

Figure 1.6 – *Transition graph of a Type-II HARQ protocol.*

In Section 1.3.2 we have shown that the main difference between Type-I HARQ protocol and CC-HARQ protocol is that the CC-HARQ protocol accumulates SNR. On the other hand, in Section 1.3.3, we have seen the IR-HARQ protocol differs from the Type-I HARQ protocol by accumulating mutual information. In the sequel we denote by $\Delta(\alpha_n, p_n)$ the increment of the accumulated quantity. For the CC-HARQ protocol we have:

$$\Delta(\alpha_n, p_n) = \alpha_n p_n,$$

while for the IR-HARQ protocol we have:

$$\Delta(\alpha_n, p_n) = \log_2(1 + \alpha_n p_n).$$

The model that is considered for analysing both Type-II HARQ protocols is a simple extension of the model proposed in Section 1.4.1 for Type-I HARQ protocols. This model is based on the random process $\{S_n\}_{n\in\mathbb{N}}$ such that, at every time nT, $S_n = (K_n, X_n)$. K_n has the same definition as in Section 1.4.1, and X_n represents either the accumulated SNR or the ACMI. In the sequel S_n will be referred to as the *state* of the HARQ protocol. Let D_T be defined as a decoding threshold: $X_n < D_T$ indicates a decoding failure and $X_n \geq D_T$ a successful decoding. Because the transmission of a new packet begins after each successful decoding, we must have $S_n = (0,0)$ after a successful decoding. In consequence the state space $\mathcal{S} = \{0,\ldots,N_T\} \times [0, D_T]$ is sufficient to represent the HARQ protocol. One possible representation for this space $\mathcal{S}$ is given in Figure 1.6. In this figure, each possible value for $k \in \{0,\ldots,N_T\}$ is represented by a state (circle). From each k we represent a vertical line that accounts for the possible values for x_n. For example, in Figure 1.6 the state s_n corresponds to a state of the form $(1, x_n)$. In the rest of this section, we will use Figure 1.6 as a transition graph to illustrate our saying.

We now show that for every $B \subset \mathcal{S}$ the following Markov property holds:

$$\begin{aligned}\mathbb{P}\left[S_{n+1} \in B | P_n = p_n, S_n = s_n, P_{n-1} = p_{n-1}, S_{n-1} = s_{n-1}, \ldots\right] = \\ \mathbb{P}\left[S_{n+1} \in B | P_n = p_n, S_n = s_n\right].\end{aligned} \tag{1.35}$$

by proceeding as in Section 1.4.1. Consider the state $s_n = (1, x_n)$ of Figure 1.6. Only two situations can occur: either $\{x_n + \Delta(\alpha_n, p_n) \geq D_T\}$ or $\{x_n + \Delta(\alpha_n, p_n) < D_T\}$. If $\{x_n + \Delta(\alpha_n, p_n) \geq D_T\}$, Rx succeeds in decoding the information packet. In this case Rx sends an ACK bit in the feedback channel and Tx starts a new transmission. Using the state definition, it means that $S_{n+1} = (0,0)$. This transition is represented as the green arrow from s_n to $(0,0)$ in Figure 1.6.

The second case $\{x_n + \Delta(\alpha_n, p_n) < D_T\}$, corresponds to a decoding failure. In this case, a retransmission is requested by Rx. The state of the protocol at time $(n+1)T$ is: $S_{n+1} = (k_n + 1, x_n + \Delta(\alpha_n, p_n))$. This situation is represented in Figure 1.6 by the red dashed arrow leaving s_n. Because $\Delta(\alpha_n, P_n) | P_n = p_n$ is a continuous random variable, it has a probability density function (pdf) denoted by $f_{\Delta|P}(x|p_n)$ that is represented in Figure 1.6. In Figure 1.6, we did not represent a transition $s_n \to (2, x')$ since this transition has 0 probability.

This analysis can be easily generalized for every other couple (k_n, x_n). These case are summarized in Table 1.2. In this table, we enumerate every possible values for S_{n+1} depending on $s_n = (k_n, x_n)$ and on the value of $\Delta(\alpha_n, p_n)$.

To prove equation (1.35), it suffices to remark from Table 1.2 that s_{n+1} can be written as a deterministic function G of s_n, α_n, and p_n:

$$s_{n+1} = G(s_n, \alpha_n, p_n). \tag{1.36}$$

We do not give a more formal expression of G since it will just be an enumeration of the different cases shown in Table 1.2. However, equation (1.36) suffices to prove the Markov property of equation (1.35) (see [Meyn & Tweedie, 2009]).

	$x_n + \Delta(\alpha_n, p_n) < D_T$	$x_n + \Delta(\alpha_n, p_n) \geq D_T$
$k_n < N_T - 1$	$(k_n+1, x_n+\Delta(\alpha_n, p_n))$	$(0,0)$
$k_n = N_T - 1$	$(N_T, 0)$	$(0,0)$
$k_n = N_T$	$(1, \Delta(\alpha_n, p_n))$	$(0,0)$

Table 1.2 – *Possible values for the state of the Type-II HARQ protocol at time $(n+1)T$ (S_{n+1}) depending on $s_n = (k_n, x_n)$ and on the value of $\Delta(\alpha_n, p_n)$.*

The transition probability

$$Q(B|s_n, p_n) = \mathbb{P}[S_{n+1} \in B | S_n = s_n, P_n = p_n] \tag{1.37}$$

is omitted for the sake of clarity but is given in Appendix A.1 since it will be used in Chapter 3.

For the same reason as in the Type-I HARQ case, the throughput of Type-II HARQ protocols can be computed introducing the following function:

$$r(s) = \begin{cases} b \text{ if } s = (0,0) \\ 0 \text{ otherwise.} \end{cases} \tag{1.38}$$

Let $\pi = \{P_n\}$, the throughput initially defined by equation (1.5) is rewritten using the function $r(\cdot)$ as follows:

$$\eta(\pi) = \frac{1}{L} \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} r(S_i), \tag{1.39}$$

where L is the number of symbols sent in one slot, where r is defined in equation (1.38), and where $\pi = \{P_n\}_{n \in \mathbb{N}}$ is added to show the dependence of the throughput on π (see Table 1.2).

1.5 Conclusion

In this section, we have reviewed the different HARQ protocols. In particular, we have given the definitions of the Type-I HARQ protocol, the CC-HARQ protocol, and the IR-HARQ protocol. In the asymptotic context of optimal Gaussian codes, we have presented the classical analysis of the throughput of these protocols in the block-fading channel case.

From this classical analysis, we have established controlled Markov models for the different HARQ protocols. Whereas the proposed model for the Type-I HARQ protocol has been presented in [Levorato et al., 2009], the model for Type-II HARQ model is an original contribution (to the best of the author's knowledge). The models proposed in [Tuninetti, 2011] and [Szczecinski et al., 2011] are similar models to the one presented here for Type-II HARQ protocols except that they do not integrate K_n which is useful for taking into account outage events. On the other hand, these models allow us to highlight easily the impact of a power allocation $\pi = \{P_n\}_{n\in\mathbb{N}}$ on the throughput of the HARQ protocols. A more formal definition of π will be given in Chapter 2. Because of this dependence, one can logically wonder if there exists π that maximizes $\eta(\pi)$ under peak and average power constraints. This problem can be stated as follows:

$$\begin{aligned} &\sup_{\pi=\{P_n\}_{n\in\mathbb{N}}} \quad \eta(\pi) \\ &\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} P_n \leq P_A \\ &P_n \leq P_M \end{aligned} \tag{1.40}$$

In Chapter 2, we present the Constrained Markov Decision Process (CMDP) framework which is a suitable theoretical framework for solving optimization problems similar to the one given in equation (1.40). This optimization problem is then solved in Chapter 3.

- Chapter 2 -

Constrained Markov Decision Processes: an Introduction

Contents

The objective of this chapter is to provide an overview of technical results related to Constrained Markov Decision Processes (CMDPs). These technical results will be used in the next chapters to derive resource allocation for the physical layer in different telecommunication contexts.

2.1 Introduction

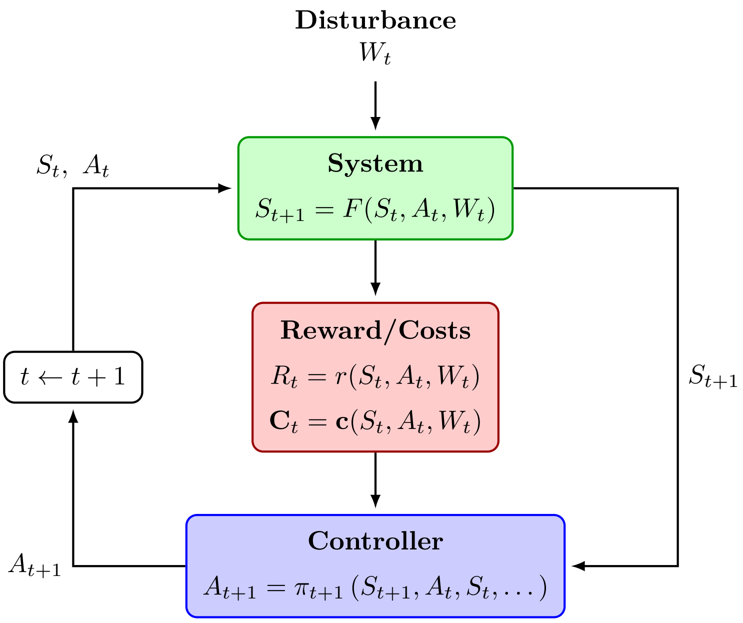

Figure 2.1 – *Markov Control Model*

Markov Decision Processes (MDPs) naturally appear in contexts in which *decisions* (or *actions*) are made sequentially by a *controller* at some discrete times (called *decision epochs*).

The general framework of MDPs is depicted Figure 2.1. The controller observes the *state* (S_t in Figure 2.1) of a *system* and influences this system through its actions (A_t in Figure 2.1). The actions have two effects: they generate some incomes and costs (r and $\mathbf{c}$ in Figure 2.1) and they change the dynamical behaviour of the system (F in Figure 2.1). The first effect is a *short-term* effect while the second one can have *long-term* consequences. The goal of the controller is to find a sequence of actions (a *policy*, π_t in Figure 2.1) for maximizing some long-term reward. The CMDPs framework appears when the controller possesses multiple objectives. For example the controller may want to maximize its long-term reward under a long-term cost constraint.

The MDP framework has been introduced by Richard Bellman in the 1950s (see [Bellman, 1952], [Bellman, 1957]). First applications of MDPs have been proposed early in their development. For example [Manne, 1960] uses MDPs to solve inventory management problems. Nowadays MDPs and CMDPs have applications in numerous areas. In particular, applications exist in robotics where MDPs has been used for autonomous navigation (see [Theocharous & Mahadevan, 2002]). In finance MDPs can be used for portfolio and asset management (see [Puterman, 1994] and [Bertsekas, 2001]). MDPs have also found important applications in the area of communication networks. In this area, it is shown in [Feinberg & Shwartz, 2002, Chapter 16] (and references therein) that MDPs are appropriate for handling problems such as: call admission control, buffer management, packet admission control, flow control, congestion control, routing, scheduling. A few MDPs have also been proposed for optimizing physical layer performance for wireless networks. Among others, we can cite [Negi & Cioffi, 2002] [Karmokar et al., 2006], [Djonin & Krishnamurthy, 2007],[Tuninetti, 2011], [Levorato et al., 2009, 2012], and [Michelusi et al., 2013b,a].

This chapter has one principal objective: introducing the CMDP framework in a suitable context for Physical (PHY) layer resource allocation, in particular for HARQ based systems. This objective comprises three sub-objectives.

- We present a general framework that allows us to manage theoretically continuous values. Indeed, in resource allocation problems for the PHY layer, we may have to cope with ACMI, SNR, power, rate and many other continuous variables. This is the goal of Section 2.2.

- We present a general framework that allows us to manage, in practice, these continuous values. Indeed, in the MDPs literature, general theories for handling continuous spaces are often not suited for practical implementation. This sub-objective is fulfilled by Section 2.4 and 2.5. In Section 2.4 we show that the framework described in Section 2.2 can be viewed as an infinite dimensional Linear Programming (LP). In Section 2.5 we provide finite LP approximations for the infinite dimensional LP.

- We present a general framework that allows us to cope with partially observable problems. This is required because in many situations some parameters of the system cannot be observed. This happens often in problems where the system and the controller (in Figure 2.1) belong to two distinct systems: in the context of Chapter 3 transmitter/receiver, and in the context of Chapter 4 primary user/secondary user. The introduction of partially observable problems is the goal of Section 2.6.1.

2.2 Introduction to the CMDP framework

2.2.1 Model formulation

A CMDP is defined by a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{U}, \mathcal{W}, Q, r, \mathbf{c})$. We will now give the description of each component of this model.

- $\mathcal{S}$ is the *state space*. The elements of $\mathcal{S}$ are denoted by s and are called *states*. Throughout this thesis, we may need to consider finite as well as infinite (possibly not countable) spaces $\mathcal{S}$. So we consider a more general framework including every case. In consequence, we consider that $\mathcal{S}$ is a Borel space. In particular, we will often consider $\mathcal{S}$ as a subset of $\mathbb{R}$ or a finite set or a finite product of these two first cases. When $\mathcal{S}$ is a finite or countable space it is endowed with its *discrete topology*. When $\mathcal{S}$ is a subset of $\mathbb{R}$ it is endowed with $\mathcal{B}(\mathcal{S})$, the Borel σ-algebra. $\mathcal{B}(\mathcal{S})$ is the σ-algebra engendered by the compact subsets of $\mathcal{S}$. When $\mathcal{S}$ is a finite product of these two kinds of spaces, $\mathcal{S}$ is endowed with the product σ-algebra.

- $\mathcal{A}$ is a Borel space called the *action space*. Its elements are called *actions*. For every state $s \in \mathcal{S}$, the set of *admissible actions* is denoted by $A(s)$. For every s in $\mathcal{S}$, $A(s)$ is a *measurable* subset of $\mathcal{A}$.

- $\mathcal{U}$ is the space of all admissible state-action pairs. $\mathcal{U}$ is the measurable subset of $\mathcal{S} \times \mathcal{A}$ defined as $\mathcal{U} = \{(s, a) | s \in \mathcal{S} \text{ and } a \in A(s)\}$. The elements of $\mathcal{U}$ are referred to as admissible *state-action pairs* and are denoted by u.

- $\mathcal{W}$ is a Borel space called the *disturbance space*. For every decision epoch, $t \in \mathbb{N}$, the disturbance W_t is a random element whose distribution is given by the stochastic kernel on $\mathcal{W}$ given $\mathcal{S} \times \mathcal{A}$: $p_W(dw|s, a)$. This means that, on one hand for every $(s, a) \in \mathcal{S} \times \mathcal{A}$, $p_W(\cdot|s, a)$ is a probability measure; on the other hand for every $B \in \mathcal{B}(\mathcal{W})$, $p_W(B|\cdot)$ is a measurable function on $\mathcal{S} \times \mathcal{A}$. The disturbance W_t is acting as a "noise" for the temporal evolution of S_t. This temporal evolution is governed by the following recurrence equation

$$S_{t+1} = g(S_t, A_t, W_t). \tag{2.1}$$

- $Q(B|u)$ is the *transition law*. It is formally defined as a stochastic kernel on $\mathcal{S}$ given $\mathcal{U}$. If the system is in state $S_t = s$ and the controller takes action $A_t = a$, the system moves from s to S_{t+1} with distribution $Q(\cdot|s, a)$. Formally, for every $B \in \mathcal{B}(\mathcal{S})$, $Q(B|s, a)$ is defined as

$$Q(B|s, a) = \mathbb{P}\left[S_{t+1} \in B | S_t = s, A_t = a\right]. \tag{2.2}$$

Using equation (2.1), equation (2.2) is expressed as

$$Q(B|s,a) = \mathbb{P}\left[g(S_t, A_t, W_t) \in B | S_t = s, A_t = a\right]. \tag{2.3}$$

• $r(s,a)$ is called the *one-step reward*. It is a function from $\mathcal{S} \times \mathcal{A}$ to $\mathbb{R}$. For every $s \in \mathcal{S}$ and $a \in \mathcal{A}$, $r(s,a)$ corresponds to the reward earned when the system is in state s and the controller chooses action a. In some situation, it is possible that the natural definition of the reward also depends on the disturbance. In this case, we define the instantaneous reward as

$$r(s,a) = \int_S r'(s,a,w) p_W(dw|s,a). \tag{2.4}$$

The definition of r given by equation (2.4) was proposed in Bertsekas & Shreve [1978] and will not change the results in the sequel.

• $\mathbf{c}$ is vector of n_c *one-step costs.* $\mathbf{c}$ is a function from $\mathcal{S} \times \mathcal{A}$ to $\mathbb{R}^{n_c}$ representing n_c costs incurred if the controller chooses action a while the system is in state s. As for r, if the natural definition of this cost vector $\mathbf{c}'$ also depends on the disturbance w, we introduce

$$\mathbf{c}(s,a) = \int_S \mathbf{c}'(s,a,w) p_W(dw). \tag{2.5}$$

This definition of $\mathbf{c}$ completes the description of the MDP. In the next subsection we present a formal description of policies.

2.2.2 Policies

The formal definition of a policy requires the formal concept of *history*, thus we first defined this notion. In a second time, we will define the concept of policy and describe various kinds of policies. Finally, we will end this subsection by some additional remarks.

History

At every decision epoch $t \in \mathbb{N}$, a history up to time t is the vector defined as

$$h_t = \left(s_0, a_0, s_1, a_1, \ldots, s_{t-1}, a_{t-1}, s_t\right). \tag{2.6}$$

The space of all possible histories up to time t is recursively defined as $\mathcal{H}_0 = \mathcal{S}$ and $\mathcal{H}_t = \mathcal{H}_{t-1} \times \mathcal{U}$.

Policy

A policy is informally defined as a "sequence of rules" for making decisions depending on an observed history. We will now present different kinds of policies that are the general *policies*, the *Markov policies*, the *randomized stationary policies*, and the *deterministic stationary policies*.

- Formally, a general *policy* π is defined as a sequence $\pi = \{\pi_t\}_{t\in\mathbb{N}}$ where, for every $t \in \mathbb{N}$, $\pi_t(da|h_t)$ is a stochastic kernel on $\mathcal{A}$ given $\mathcal{H}_t$. At decision epoch $t \in \mathbb{N}$, the controller chooses an action within the set $A \in \mathcal{B}(\mathcal{U})$ with a probability $\pi_t(A|h_t)$. The set of all policies is denoted by Π.

- A policy $\pi \in \Pi$ is said to be a Markov policy if and only if for all $t \in \mathbb{N}$ it verifies $\pi_t(\cdot|h_t) = \pi_t(\cdot|s_t)$. The set of all Markov policies is denoted by Π_M. For policies of Π_M, there is no need to store the whole history but only the last state s_t.

- A policy $\pi \in \Pi_M$ is a *randomized stationary policy* if for every $t \in \mathbb{N}$, $\pi_t(\cdot|s_t) = \pi(\cdot|s_t)$. The set of all randomized stationary policies is denoted by Π_{RS}.

- A policy $\pi \in \Pi_{RS}$ is a *deterministic stationary policy* if there exists a function from $\mathcal{S}$ to $\mathcal{A}$ denoted by ζ, such that for every $n \in \mathbb{N}$, $\pi(\cdot|s_n) = \delta_{\zeta(s_n)}(\cdot)$. In the last expression, δ stands for the Dirac measure. The set of all deterministic stationary policies is denoted by Π_{DS}. Obviously, we have the inclusion: $\Pi_{DS} \subset \Pi_{RS} \subset \Pi_M \subset \Pi$.

Other categories of policies can be found in the literature, in particular [Altman, 1999] also distinguishes between randomized Markov and deterministic Markov policies.

Remarks

• The first remark concerns the probability space underlying the random processes $\{S_t\}_{t\in\mathbb{N}}$ and $\{A_t\}_{t\in\mathbb{N}}$ when the policy $\pi \in \Pi$ is used by the controller and the initial distribution of S_0 is ν_0. This probability space is $(\Omega, \mathcal{F}, \mathbb{P}_\pi^{\nu_0})$ where $\Omega = (\mathcal{S} \times \mathcal{A})^\infty$ and $\mathcal{F}$ is the product σ-algebra. The Ionescu-Tulcea theorem (see [Hernández-Lerma, 1989, p. 4]) states that there exists a unique probability measure $\mathbb{P}_\pi^{\nu_0}$ on Ω such that

$$\begin{aligned}\mathbb{P}_\pi^{\nu_0}\left[ds_0, da_0, ds_1, da_1, ds_2, da_2 \ldots\right] &= \nu_0(ds_0)\pi_0(da_0|s_0)Q(ds_1|s_0,a_0) \\ &\ldots \pi_1(da_1|s_1,a_0,s_0)Q(ds_2|s_1,a_1)\pi_2(da_2|s_2,a_1,s_1,a_0,s_0)\cdots \quad (2.7)\end{aligned}$$

verifying that for all $B \in \mathcal{B}(\mathcal{S})$ and all $C \in \mathcal{B}(\mathcal{A})$

$$\mathbb{P}_\pi^{\nu_0}\left[\mathcal{H}_\infty\right] = 1 \tag{2.8}$$

$$\mathbb{P}_\pi^{\nu_0}\left[s_0 \in B\right] = \nu_0(B) \tag{2.9}$$

$$\mathbb{P}_\pi^{\nu_0}\left[a_t \in C|h_t\right] = \pi_t(C|h_t) \tag{2.10}$$

$$\mathbb{P}_\pi^{\nu_0}\left[s_{t+1} \in B|a_t, h_t\right] = Q(B|s_t, a_t), \tag{2.11}$$

where $\mathcal{H}_\infty = \mathcal{U}^\infty$ is the set of all admissible histories. In the sequel, $\mathbb{E}_\pi^\mu[\cdot]$ denotes the expectation associated with $\mathbb{P}_\pi^{\nu_0}[\cdot]$.

• The second remark is an interesting property of the Markov policies. This property is that every policy $\pi \in \Pi_M$ induces a Markov Chain (MC) on $\mathcal{S}$. Indeed, let $\pi \in \Pi_M$ be a Markov policy and ν_0 an initial distribution, we have

$$\mathbb{P}_\pi^{\nu_0}\left[S_{t+1} \in B|h_t\right] = \int_{\mathcal{A}} \mathbb{P}_\pi^{\nu_0}\left[S_{t+1} \in B|h_t, a_t\right] \mathbb{P}_\pi^{\nu_0}\left[da_t|h_t\right] \tag{2.12}$$

$$= \int_{\mathcal{A}} Q(B|s_t, a_t)\pi_t(da_t|s_t). \tag{2.13}$$

The first equality comes from the law of total probability and the second one comes from (2.10) and (2.11). The transition kernel of this MC on $\mathcal{S}$ is the following

$$Q_{\pi,t}(\cdot|s) = \int_{\mathcal{A}} Q(\cdot|s, a)\pi_t(da|s). \tag{2.14}$$

For every randomized stationary policy φ, this transition kernel is time-homogeneous and will be denoted by

$$Q_\varphi(\cdot|s) = \int_{\mathcal{A}} Q(\cdot|s, a)\varphi(da|s). \tag{2.15}$$

2.2.3 Performance criteria

Hitherto, we have defined the CMDP model and different sets of policies. To complete the definition of a CMDP we need to define the long term reward and costs. In the sequel, ν_0 is the initial distribution, π a policy and T a positive real number called *horizon*. Various kinds of performance criteria are encountered in the literature (see [Altman, 1999; Bertsekas & Shreve, 1978; Hernández-Lerma, 1989; Hernández-Lerma & Lasserre, 1996, 1999]). We now give the most common performance criteria.

We first present a *finite-horizon performance criterion*. Let r_T be some reward depending on the terminal state, the finite-horizon performance criterion is defined as

$$R(\nu_0, \pi) = \mathbb{E}_{\pi}^{\nu_0}\left[\sum_{t=0}^{T-1} r(S_t, A_t) + r_T\left(S_T\right)\right], \tag{2.16}$$

where $\mathbb{E}_{\pi}^{\nu_0}[\cdot]$ is the expectation associated with the probability measure $\mathbb{P}_{\pi}^{\nu_0}[\cdot]$ defined in equation (2.7).

We now present two different infinite horizon performance criteria. The first one is called the *infinite-horizon discounted-cost performance criterion*. It is defined as follows

$$R(\nu_0, \pi) = \lim_{T\to\infty} \mathbb{E}_{\pi}^{\nu_0}\left[\sum_{t=0}^{T-1} \alpha^t r(S_t, A_t)\right], \tag{2.17}$$

where $\alpha \in [0, 1[$ is called the discount factor.

We finally present a last performance criterion that will be the one considered in the sequel. This last criterion is called the *infinite-horizon average-cost performance criterion*. This performance criterion is defined as:

$$R(\nu_0, \pi) = \liminf_{T\to\infty} \frac{1}{T}\mathbb{E}_{\pi}^{\nu_0}\left[\sum_{t=0}^{T-1} r(S_t, A_t)\right]. \tag{2.18}$$

As pointed out in [Hernández-Lerma, 1989], the choice of considering $'\liminf'$ and not $'\lim'$ in equation (2.18) is motivated by the fact that the limit may not exist. The choice of $'\liminf'$ over $'\limsup'$ is on the other hand motivated by the fact that we will consider the maximisation of $R(\nu_0, \pi)$ and that $'\liminf'$ is somewhat similar to considering a worst-case scenario. Similarly, for long-term costs, it will be preferable to use $'\limsup'$ to consider a worst-case scenario. Consequently, the long-term cost is defined as:

$$\mathbf{C}(\nu_0, \pi) = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_{\pi}^{\nu_0}\left[\sum_{t=0}^{T-1} \mathbf{c}(S_t, A_t)\right]. \tag{2.19}$$

2.2.4 Constrained optimisation problem

Henceforth, the initial distribution ν_0 is supposed to be known. The constrained optimization problem associated with the CMDP is the following:

$$R^\star(\nu_0) = \sup_{\pi\in\Pi} \quad R(\nu_0, \pi) \tag{2.20}$$

$$s.t. \quad \mathbf{C}(\nu_0, \pi) \leq \mathbf{V} \tag{2.21}$$

In view of future applications, we are interested in finding $R^\star(\nu_0)$ as well as finding an optimal policy $\pi^\star$ (if it exists). An optimal policy $\pi^\star \in \Pi$ is a policy verifying equations:

$$R(\nu_0, \pi^\star) = R^\star(\nu_0) \tag{2.22}$$

$$\mathbf{C}(\nu_0, \pi^\star) \leq \mathbf{V}. \tag{2.23}$$

In the sequel, Θ represents the set of all admissible policies:

$$\Theta = \{\pi \in \Pi : \mathbf{C}(\nu_0, \pi^\star) \leq \mathbf{V}\}. \tag{2.24}$$

The next sections of this chapter will be devoted to giving sufficient conditions for the existence of an optimal policy $\pi^\star$ in Θ. In the sequel we will also show that the optimization problem (2.20) is equivalent to an infinite dimensional LP. We then show that finite approximations of this LP can be performed to obtain "near-optimal" policies. Finally we will show that these finite approximations are asymptotically optimal.

2.3 Existence of an optimal policy

The goal of this section is to provide sufficient conditions, that can be checked in practical applications, for the existence of an optimal policy for equation (2.20). Every result shown in this section has been shown in [Kurano et al., 2000]; in consequence, this chapter is highly inspired by this article. In this section, we first provide assumptions that will we be our theoretical framework, not only for this section but also for the following sections. Based on these assumptions and on [Hernández-Lerma, 1989], we first remind some general properties about policies in Π_{RS} when the CMDP is *ergodic*. We then couple these properties together with properties on the components of the CMDP to prove that Π_{RS} is a dominating class of policies. We finally prove that an optimal policy exists within Π_{RS}. This methodology, to prove the existence of an optimal policy, has the same structure as in [Hernández-Lerma & Lasserre, 1996] and [Hernández-Lerma & Lasserre, 1999] for MDPs, [Altman, 1999] for countable CMDPs, and [Kurano et al., 2000] and [Hernández-Lerma et al., 2003] for Borel spaces CMDPs.

2.3.1 Assumptions on the CMDP model

As far as we know, it is impossible to prove the existence of a solution for the problem (2.20) in the general case depicted in Section 2.2. On the other hand, under mild assumptions we will show that the optimization problem (2.20) has a solution. This subsection is devoted to the exposition of these assumptions.

Assumption 2.1.
(a) *$\mathcal{S}$ and $\mathcal{U}$ are compact.*
(b) *r is continuous on $\mathcal{S} \times \mathcal{A}$ and there exists $r_M \in \mathbb{R}^+$ such that for every $(s,a) \in \mathcal{S} \times \mathcal{A}$, $|r(s,a)| \leq r_M$.*
(c) *For every $i \in [1, n_c]$, c_i is a continuous function on $\mathcal{S} \times \mathcal{A}$ and there exists $c_{i,M}$ such that for every $(s,a) \in \mathcal{S} \times \mathcal{A}$, $|c_i(s,a)| \leq c_{i,M}$.*
(d) *For every bounded continuous function v on $\mathcal{S}$, $\bar{v}(s,a) = \int_{\mathcal{S}} v(s')Q(ds'|s,a)$ is a bounded continuous function on $\mathcal{S} \times \mathcal{A}$. Following [Meyn & Tweedie, 2009], we will say that Q verifies the* weak Feller *property, or simply Q is weak feller.*
(e) *For every policy $\varphi \in \Pi_{RS}$, the random process $\{S_t\}_{n \in \mathbb{N}}$ is a uniformly ergodic Markov chain (see [Meyn & Tweedie, 2009]).*

Assumptions 2.1(a) -2.1(d) are the same as Assumptions (ii)-(iii) of [Kurano et al., 2000]. Assumption 2.1(e) is similar to the unichain assumption in the literature of countable CMDPs (see e.g. [Altman, 1999, assumption ($B1$) p. 143]). Sufficient conditions for Assumption 2.1(e) are presented in [Hernández-Lerma, 1989]. In particular, the two conditions presented hereafter are special cases of the Doeblin's condition (see [Meyn & Tweedie, 2009, p. 402]).

Condition 2.1 (Hernández-Lerma [1989])**.**
(a) *There exist $s_0 \in \mathcal{S}$ and $\epsilon \in \mathbb{R}^+$ such that $Q(\{s_0\}|s,a) \geq \epsilon$, with $\epsilon > 0$.*
(b) *There exists a measure ν on $\mathcal{S}$ such that $\nu(\mathcal{S}) > 0$ and,*

$$Q(B|s,a) \geq \nu(B), \ \forall (s,a) \in \mathcal{U} \text{ and } \forall B \in \mathcal{B}(\mathcal{S}). \tag{2.25}$$

It can be shown (see Hernández-Lerma [1989]) that both Condition 2.1(a) and Condition 2.1(b) are sufficient for Assumption 2.1(e).

2.3.2 Properties of randomized stationary policies

Assumption 2.1(e) (*the uniform ergodicity assumption*) implies that for every $\varphi \in \Pi_{RS}$ there exists a unique probability measure on $\mathcal{S}$ denoted by Q^∞_φ, such that

$$Q^\infty_\varphi(B) = \int_S Q_\varphi(B|s)Q^\infty_\varphi(ds). \tag{2.26}$$

For every $\varphi \in \Pi_{RS}$ the n step transition kernel recursively defined as

$$\begin{cases} Q^1_\varphi(B|s) = Q_\varphi(B|s) \\ Q^n_\varphi(B|s) = \int_{\mathcal{S}} Q^{n-1}_\varphi(B|s')Q(ds'|s) \end{cases} \tag{2.27}$$

converges to $Q_\varphi^\infty(B)$ in the Total Variation norm (TV-norm) at a geometrical rate: there exist $\rho \in [0, 1[$ such that

$$||Q_\varphi^\infty - Q_\varphi^n||_{TV} \leq 2\rho^n. \tag{2.28}$$

In the literature, ρ is called the ergodicity rate. For every signed measure μ, the TV-norm is the norm defined as

$$||\mu||_{TV} = \sup_{B \in \mathcal{B}(\mathcal{S})} \mu(B) - \inf_{B \in \mathcal{B}(\mathcal{S})} \mu(B). \tag{2.29}$$

For two probability measures μ_1 and μ_2, $||\mu_1 - \mu_1||_{TV}$ is given by

$$||\mu_1 - \mu_2||_{TV} = 2 \sup_{B \in \mathcal{B}(\mathcal{S})} |\mu_1(B) - \mu_2(B)|. \tag{2.30}$$

The last consequence of the uniform ergodicity condition is that the long-term functions ($R(\nu_0, \varphi)$ and $\mathbf{C}(\nu_0, \varphi)$), are independent of ν_0 and are expressed using the Law of Large Numbers (LLN) by

$$R(\varphi) = \int_{\mathcal{S} \times \mathcal{A}} r(s, a)\varphi(da|s)Q_\varphi^\infty(ds), \tag{2.31}$$

$$\mathbf{C}(\varphi) = \int_{\mathcal{S} \times \mathcal{A}} \mathbf{c}(s, a)\varphi(da|s)Q_\varphi^\infty(ds). \tag{2.32}$$

2.3.3 Domination of randomized stationary policies

This section is devoted to showing that Π_{RS} is a dominating class of policies for the optimization problem (2.20). This is the statement of the following lemma:

Lemma 2.1. *Under Assumptions 2.1, Π_{RS} is a* dominating class *of policies.*

Proof. This domination property is proved by showing that for every policy $\pi \in \Pi$, there exists $\varphi \in \Pi_{RS}$ such that

$$\begin{cases} R(\varphi) \geq R(\nu_0, \pi) \\ \mathbf{C}(\varphi) \leq \mathbf{C}(\nu_0, \pi). \end{cases} \tag{2.33}$$

For every initial distribution ν_0 and every policy $\pi \in \Pi$, an occupation measure is defined as

$$\mu_t(U) = \frac{1}{t} \sum_{j=0}^{t-1} \mathbb{P}_\pi^{\nu_0}\left[(s_j, a_j) \in U\right], \ \forall U \in \mathcal{B}\left(\mathcal{S} \times \mathcal{A}\right). \tag{2.34}$$

For every $t \in \mathbb{N}$, μ_t is concentrated on $\mathcal{U}$; i.e. $\mu_t(\bar{\mathcal{U}}) = 0$. Using this occupation measure, the long-term reward and cost functions are written as follows

$$R(\nu_0, \pi) = \liminf_{t\to\infty} \int r d\mu_t \tag{2.35}$$

$$\mathbf{C}(\nu_0, \pi) = \limsup_{t\to\infty} \int \mathbf{c} d\mu_t \tag{2.36}$$

We now use the occupation measures of equation (2.34) to build a randomized policy $\varphi \in \Pi_{RS}$ that dominates π. The rest of the proof is the same as in [Hernández-Lerma & Lasserre, 1996] and is omitted here. □

A direct consequence of Lemma 2.1 is that the optimization problem (2.20) can be constrained to the set of randomized stationary policies without loss of optimality. In other words we have

$$R^\star = \sup_{\varphi\in\Omega_{RS}} R(\varphi), \tag{2.37}$$

where $\Omega_{RS} = \Omega \cap \Pi_{RS}$.

A second consequence of Lemma 2.1 is that $R^\star$ is independent of ν_0. This is due to the fact that for every $\varphi \in \Pi_{RS}$, $R(\nu_0, \varphi) = R(\varphi)$ and $\mathbf{C}(\nu_0, \varphi) = \mathbf{C}(\varphi)$.

2.3.4 Existence of a solution

We now conclude this section by showing that there exists a solution to the optimization problem (2.20).

Theorem 2.1. *Under Assumption 2.1 and if Ω is non-empty, there exists a solution of the optimization problem* (2.20) *in* Π_{RS}.

Proof. By Assumption, we suppose the Ω is non-empty. From Lemma 2.1, it follows that Ω_{RS} is also non-empty. Let $\{\epsilon_t\}$ be a sequence of positive real numbers such that $\lim_{t\to\infty} \epsilon_t = 0$. By equation (2.37), for every $t \in \mathbb{N}$, there is φ_t in Ω_{RS} such that

$$R(\varphi_t) = \int r dm_t \geq R^\star + \epsilon_t, \tag{2.38}$$

where for every $t \in \mathbb{N}$, $B \in \mathcal{B}(\mathcal{S})$, and $C \in B(\mathcal{S})$,

$$m_t(B \times C) = \int_B \varphi_t(C|s) Q^\infty_{\varphi_t}(ds). \tag{2.39}$$

By the same arguments of the proof of Lemma 2.1, we have that there exists a converging subsequence $\{m_{t_1}\}$ such that

$$R^\star \leq \limsup_{t_1} \int r dm_{t_1} \leq \int r dm^\star. \tag{2.40}$$

It remains to show that there exists $\varphi^\star \in \Omega_{RS}$ such that

$$m^\star(d(s,a)) = \varphi^\star(da|s)Q^\infty_{\varphi^\star}(ds). \tag{2.41}$$

The rest of this proof is the same as proposed in [Hernández-Lerma & Lasserre, 1996] and is omitted here. □

There are two important conclusions for this section. The first one is stated in Lemma 2.1: under Assumption 2.1, the optimization problem given in equation 2.20 can be solved in Π_{RS} without loss of optimality. The second conclusion of this section is given by Theorem 2.1: under Assumption 2.1 and if the set of admissible policies is non-empty, there exists an optimal policy for the optimization problem given in equation 2.20.

2.4 Infinite Linear Programming

In Section 2.3 we have shown that there exists an optimal policy $\varphi^\star \in \Pi_{RS}$ for the optimization problem (2.20). In this section, we will show that under Assumption 2.1, the optimization problem (2.37) is equivalent to an infinite dimensional LP. As it is shown in [Hernández-Lerma & Lasserre, 1996] for unconstrained MDPs, the main idea is to embed Ω into suitable vector spaces.

2.4.1 Dual pairs of vector spaces

In this subsection, we provide a complete description of the different vector spaces that will be used through this section. However, we will not present a general description of LP on general vector spaces. For more information about LP on general spaces, one can refer to [Anderson & Nash, 1987] or [Hernández-Lerma & Lasserre, 1996, Chapter 6]. This section uses the results of [Hernández-Lerma & Lasserre, 1996, Chapter 6] for unconstrained MDPs and of [Hernández-Lerma et al., 2003] for CMDPs.

To correctly define the concept of LP on general vector spaces, we will need some definitions and notation. First, let $\mathcal{F}(\mathcal{U})$ be the vector space of bounded measurable functions on $\mathcal{U}$, i.e. the set of all measurable functions v on $\mathcal{U}$ such that

$$\sup_{(s,a)\in\mathcal{U}} v(s,a) < \infty. \tag{2.42}$$

Let $\mathcal{M}(\mathcal{U})$ be the vector space of all bounded signed measures concentrated on $\mathcal{U}$, i.e. the set of measures m verifying

$$||m||_{TV} < \infty, \tag{2.43}$$

$$m(\mathcal{U}^c) = 0, \tag{2.44}$$

where $||\cdot||_{TV}$ is the TV-norm defined by equation (2.29) and

$$\mathcal{U}^c = \{(s,a) \in \mathcal{S} \times \mathcal{A} | (s,a) \notin \mathcal{U}\}.$$

For every $m \in \mathcal{M}(\mathcal{U})$ and every $v \in \mathcal{F}(\mathcal{U})$, let

$$\langle m, v \rangle = \int v dm. \tag{2.45}$$

It is shown in [Hernández-Lerma & Lasserre, 1996] that equation (2.45) defines a bilinear form on $\mathcal{M}(\mathcal{U}) \times \mathcal{F}(\mathcal{U})$ as long as we take the convention that every function $v \in \mathcal{F}(\mathcal{U})$ is measurably extended to $\mathcal{S} \times \mathcal{A}$ in such a way that for every $m \in \mathcal{M}(\mathcal{U})$

$$\int_{\mathcal{U}^c} v dm = 0.$$

Under these conditions $(\mathcal{M}(\mathcal{U}), \mathcal{F}(\mathcal{U}))$ is a dual pair. Under similar conditions, the pair $(\mathcal{M}(\mathcal{S}), \mathcal{F}(\mathcal{S}))$ is also a dual pair. We now use these vector spaces to define an equivalent LP for the optimization problem (2.37).

2.4.2 LP associated with CMDP

In this subsection, we prove that the optimization problem (2.37) is equivalent to a LP on $\mathcal{M}(\mathcal{U})$. This approach to solve CMDP on Borel spaces is mainly inspired from the following works: Altman [1999], Kurano et al. [2000], and Hernández-Lerma et al. [2003]. The proof that the optimization problem (2.37) is equivalent to a LP is divided in two parts. In the first part, we give a characterization of the probability measures of $\mathcal{M}(\mathcal{U})$ that define policies in Π_{RS}. In the second part we use this characterization to build a LP. We start by proving the following result.

Lemma 2.2. *Under Assumption 2.1, if m is a probability measure in $\mathcal{M}(\mathcal{U})$ such that for all $B \in \mathcal{B}(\mathcal{S})$:*

$$m(B \times \mathcal{A}) - \int_{\mathcal{S}\times\mathcal{A}} Q(B|s,a) m(d(s,a)) = 0. \tag{2.46}$$

then there is a policy $\varphi \in \Pi_{RS}$ such that for all $B \in \mathcal{B}(\mathcal{S})$, $Q_\varphi^\infty(B) = m(B \times \mathcal{A})$.

Proof. For every probability measure m in $\mathcal{M}(\mathcal{U})$, there exists a stochastic kernel φ on $\mathcal{A}$ given $\mathcal{S}$ so that m is disintegrated as

$$m(B \times C) = \int_B \varphi(C|s) \hat{m}(ds), \ \forall B \in \mathcal{B}(\mathcal{S}), C \in \mathcal{B}(\mathcal{A}), \tag{2.47}$$

with $\hat{m} = m(B \times \mathcal{A})$(see [Hernández-Lerma & Lasserre, 1996, Appendix D8]). Since φ is a stochastic kernel on $\mathcal{A}$ given $\mathcal{S}$, it can be viewed as a policy in Π_{RS}. Using equation (2.47) in equation (2.46) gives that $\hat{m}$ verifies

$$\hat{m}(B) - \int_{\mathcal{S}} \int_{\mathcal{A}} Q(B|s,a)\varphi(da|s)\hat{m}(ds) = 0. \tag{2.48}$$

Because of the uniqueness of the invariant probability measure of Q_φ, we have $\hat{m} = Q_\varphi^\infty$. □

We now show that the optimization problem (2.37) is equivalent to an infinite dimensional programming. First, let L_0 be the linear map from $\mathcal{M}(\mathcal{U})$ to $\mathcal{M}(\mathcal{S})$ defined as follows:

$$L_0 m(B) = m(B \times \mathcal{A}) - \int_{\mathcal{S}\times\mathcal{A}} Q(B|s,a)m(d(s,a)). \tag{2.49}$$

In Section 2.3.3, we have shown that under Assumption 2.1, $R(\varphi) = \langle m, r\rangle$ and $\mathbf{C}(\varphi) = \langle m, \mathbf{c}\rangle$ where $m(d(s,a)) = \varphi(da|s)Q_\varphi^\infty(ds)$. The equivalence between the optimization problem (2.37) and an infinite dimensional LP is stated in the following theorem.

Theorem 2.2. *The optimization problem (2.37) is then equivalent to:*

$$\begin{aligned} R^\star \quad &= \quad \sup_{(m,\boldsymbol{\alpha})} \langle (m,\boldsymbol{\alpha}), (r,\mathbf{0})\rangle \\ s.t. \quad & L_0 m = 0 \\ & \langle m, 1\rangle = 1 \\ & \langle m, c_j\rangle + \alpha_j = V_j, \ \forall j \in \{1 \dots n_c\} \\ & (m,\boldsymbol{\alpha}) \in \mathcal{M}_+(\mathcal{U}) \times \mathbb{R}^{n_c}. \end{aligned} \tag{2.50}$$

In equation (2.50), for every $(m,\mathbf{x}) \in \mathcal{M}(\mathcal{U}) \times \mathbb{R}^{n_c}$ and $(v,\mathbf{y}) \in \mathcal{F}(\mathcal{U}) \times \mathbb{R}^{n_c}$

$$\langle (m,\mathbf{x}), (v,\mathbf{y})\rangle = \langle m, v\rangle + \sum_{j=1}^{n_c} x_j y_j, \tag{2.51}$$

where n_c is the number of long-term cost constraints.

Proof. The reward function $\langle (m,\boldsymbol{\alpha}), (r,\mathbf{0})\rangle$ is already linear. In consequence, to show that equation (2.50) defines a LP, we have to verify that the linear map L from $\mathcal{M}(\mathcal{U}) \times \mathbb{R}^{n_c}$ to $\mathcal{M}(\mathcal{S}) \times \mathbb{R} \times \mathbb{R}^{n_c}$ defined as

$$L(m,\boldsymbol{\alpha}) = (L_0 m, \langle m, 1\rangle, \langle m, c_1\rangle + \alpha_1, \dots, \langle m, c_{n_c}\rangle + \alpha_{n_c}) \tag{2.52}$$

is continuous. Let $L^\star$ be the adjoint map of L. For every triplet $(u,\beta,\boldsymbol{\lambda}) \in$

$\mathcal{F}(\mathcal{S}) \times \mathbb{R} \times \mathbb{R}^{n_c}$, $L^\star$ is defined as follows:

$$\langle L(m, \boldsymbol{\alpha}), (u, \beta, \boldsymbol{\lambda}) \rangle = \langle (m, \boldsymbol{\alpha}), L^\star(u, \beta, \boldsymbol{\lambda}) \rangle . \tag{2.53}$$

To show the continuity of the map L, it is sufficient to prove that $L^\star$ defines a map from $\mathcal{F}(\mathcal{S}) \times \mathbb{R} \times \mathbb{R}^{n_c}$ to $\mathcal{F}(\mathcal{U}) \times \mathbb{R}^{n_c}$. In our case, under Assumption (2.1), this trivially holds. Indeed, one can easily verify that $L^\star$ is the map defined as:

$$L^\star(u, \beta, \boldsymbol{\lambda}) = \left(u(s) - \int_S u(s')Q(ds'|s,a) + \beta + \sum_{j=1}^{n_c} \lambda_j c_j(s,a), \boldsymbol{\lambda} \right) . \tag{2.54}$$

□

We now use the map $L^\star$ to define the dual LP of (2.50). This dual LP is expressed as:

$$\begin{aligned} \beta^\star &= \inf_{(\boldsymbol{\lambda},u,\beta)} \beta \qquad (2.55) \\ s.t.\ \beta + u(s) &\geq r(s,a) + \sum_{j=1}^{n_C} \lambda_j \left(V_j - c_j\left(s,a\right)\right) + \int_{\mathcal{S}} u(y)Q(dy|s,a) \\ \boldsymbol{\lambda} &\geq 0 \\ (u, \beta, \boldsymbol{\lambda}) &\in \mathcal{F}(\mathcal{S}) \times \mathbb{R} \times \mathbb{R}^{n_c}. \end{aligned}$$

In this subsection, we have shown that the optimization problem (2.37) can be viewed as a infinite dimensional LP. We have given this LP and its dual. In the next subsection we will give some properties between the primal LP (2.50) and its dual (2.56).

2.4.3 Properties of the dual LP

In this subsection, we derive some results given in [Kurano et al., 2000]. In particular, we show that the dual LP is consistent. This means that there exists $(u, \beta, \boldsymbol{\lambda}) \in \mathcal{F}(\mathcal{S}) \times \mathbb{R} \times \mathbb{R}^{n_c}$ admissible for equation (2.56). We then prove that $\beta^\star$ is equal to $R^\star$. To finish, we prove that there exists a triplet $(u^\star, \beta^\star, \boldsymbol{\lambda}^\star)$ optimal for the programming (2.56).

We first prove that under Assumption 2.1, the dual LP (2.56) is consistent. To prove this consistency we just note that, because of the boundedness of the function r, $(0, r_M, \mathbf{0})$ is admissible. Using classical results on LP (see [Anderson & Nash, 1987]) we have that:

$$\beta^\star \geq R^\star. \tag{2.56}$$

This property is called *weak duality*. We will now show that this weak duality property can be strengthened by showing that there is no duality gap.

Theorem 2.3. *Under Assumptions 2.1, there is no duality gap:* $R^\star = \beta^\star$.

Proof. The proof of this theorem is based on a result given in [Anderson & Nash, 1987]. This result is the following: proving the closeness of the set H defined hereafter is equivalent to proving Theorem 2.3.

$$H = \left\{ \left(L(m, \boldsymbol{\alpha}), \langle (m, \boldsymbol{\alpha}), (r, 0) \rangle + \iota \right), m \in \mathcal{M}_+(\mathcal{U}), \boldsymbol{\alpha} \in \mathbb{R}_+^{n_c}, \iota \in \mathbb{R}^+ \right\} \tag{2.57}$$

The closeness of H is proven considering three sequences $\{m_n\}_{n\in\mathbb{N}}$,$\{\boldsymbol{\alpha}_n\}_{n\in\mathbb{N}}$, $\{\iota_n\}_{n\in\mathbb{N}}$ such that as $n \to \infty$:

$$m_n(\mathcal{U}) \to a_1 \tag{2.58}$$
$$L_0(m_n) \to m_1 \tag{2.59}$$
$$\langle m_n, r \rangle + \iota_n \to r_1 \tag{2.60}$$
$$\langle m_n, c_j \rangle + \alpha_{j,n} \to v_j, \ \forall j \in \{1 \dots n_c\} \tag{2.61}$$

and showing that there exists $(m, \boldsymbol{\alpha}, \iota)$ such that

$$m(\mathcal{U}) = a_1 \tag{2.62}$$
$$L_0(m) = m_1 \tag{2.63}$$
$$\langle m, r \rangle + \iota = r_1 \tag{2.64}$$
$$\langle m, c_j \rangle + \alpha_j = v_j, \ \forall j \in \{1 \dots n_c\}. \tag{2.65}$$

This proof is identical to the one proposed in [Hernández-Lerma & Lasserre, 1996] and is omitted here. □

We now prove that there exists a triplet $(u^\star, \beta^\star, \boldsymbol{\lambda}^\star)$ optimal for the programming (2.56). This property is called *strong duality.*

Assumption 2.2 (Slater Condition)**.**
There exists a policy $\varphi_0 \in \Pi_{RS}$ such that $\mathbf{C}(\varphi_0) < \mathbf{V}$.

Theorem 2.4. *Under Assumptions 2.1 and 2.2, the dual programming is solvable: there exists a triplet* $(\beta^\star, \boldsymbol{\lambda}^\star, u^\star)$ *optimal for the LP* (2.56).

Proof. This proof is given in Appendix B.3. □

As it is remarked in [Hernández-Lerma et al., 2003], the strong duality property implies

Before concluding this section we given two propositions that will be useful in Section 2.5

Proposition 2.1. *Under Assumption 2.1 and Assumption 2.2,*

$$R^\star = \inf_{\boldsymbol{\lambda} \geq \mathbf{0}} \sup_{\varphi \in \Pi_{RS}} \left(R(\varphi) + \sum_{j=1}^{n_c} \lambda_j \left(V_j - C_j(\varphi) \right) \right) \tag{2.66}$$

The proof of this proposition is given in the proof of Theorem 2.4 in Appendix B.3. The second proposition is a bound on the value of $\boldsymbol{\lambda}^\star$.

Proposition 2.2. *Under Assumption 2.1 and Assumption 2.2,*

$$\boldsymbol{\lambda}^\star \leq \frac{r_M - R(\varphi_0)}{\sup_{j \in [1, n_c]} V_j - C_j(\varphi_0)}, \tag{2.67}$$

where φ_0 is the policy of Assumption 2.2.

Proof. This inequality is inspired by a similar inequality proposed in [Altman, 1999, equation (13.4)]. However, since our proof is slightly different, it is given as follows. Let $(\beta^\star, \boldsymbol{\lambda}^\star, u^\star)$ be an optimal triplet for the dual LP. From Proposition 2.1 we obtain

$$\sup_{\varphi \in \Pi_{RS}} \left(R(\varphi) + \sum_{j=1}^{n_c} \lambda_j^\star \left(V_j - C_j(\varphi) \right) \right) \leq R^\star = r_M \tag{2.68}$$

where the second inequality comes from the boundedness of $r(s, a)$. In particular, for the policy φ_0 of Assumption 2.2 we have

$$R(\varphi_0) + \sum_{j=1}^{n_c} \lambda_j^\star \left(V_j - C_j(\varphi_0) \right) \leq r_M. \tag{2.69}$$

Because, we have $\mathbf{C}(\varphi_0) < \mathbf{V}$ and $\boldsymbol{\lambda}^\star \geq \mathbf{0}$, we obtain that for every $i \in [1, n_c]$

$$\lambda_i^\star \leq \langle \boldsymbol{\lambda}^\star, \mathbf{1} \rangle \leq \frac{r_M - R(\varphi_0)}{\min_{j \in [1, n_c]} V_j - C_j(\varphi_0)}, \tag{2.70}$$

where $\mathbf{1} = (1, 1, \ldots, 1)$ is the vector composed by n_c ones. □

In this section we have proved that the nature of the optimization problem (2.37) is a LP on the infinite dimensional space of bounded measures. We have also shown that the dual of this LP is consistent, solvable, and that there is no duality gap. On the other hand, this formulation does not provide an optimal policy for the optimization problem (2.37). However, this formulation can be used to propose approximations for the optimal policy. This is the object of the next section.

2.5 Discrete Approximations

It has been shown in [Chow, 1989] that a continuous (unconstrained) MDP can be approximated by solving discrete (and finite) MDPs. Their method first defines a finite partition of the state and action spaces. Afterwards, they propose a finite MDP on the finite partition. This new MDP is discrete and finite; consequently it can be solved numerically. Finally they prove that the procedure converges to a solution of the continuous MDP when the grid becomes finer. To prove this result, they propose a new MDP which has the same state space as the original MDP and which is equivalent to the discrete one. This method is schematically represented in Figure 2.2. We now show that this procedure is still valid for CMDPs. This chapter has been inspired by the work [Chow, 1989].

2.5.1 Discrete approximation of a CMDP

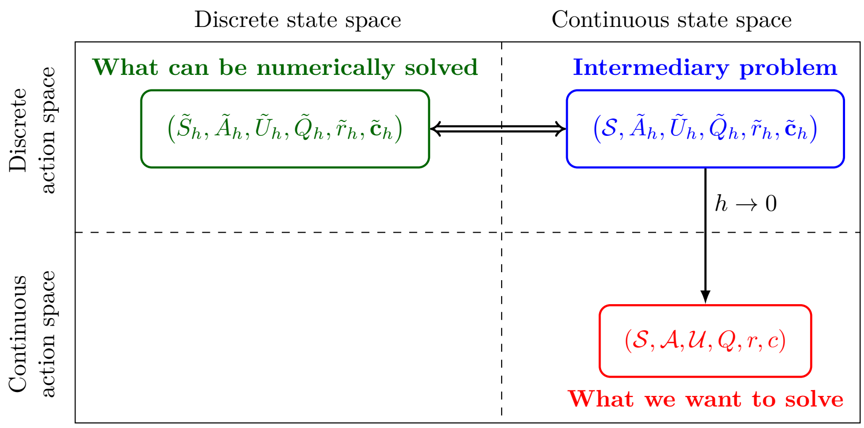

Figure 2.2 – *Schematic representation of the discretization procedure proposed in [Chow, 1989]. The initial CMDP is represented in red. The discrete version of the CMDP is represented in green. In blue is a intermediary problem that is equivalent to the discrete CMDP (green) but has the same state space as the continuous problem (red).*

Let $(\mathcal{S}, \mathcal{A}, \mathcal{U}, Q, r, c)$ be a CMDP defined similarly as in Section 2.2. In this model, $\mathcal{W}$ is omitted since it only plays an indirect role for discretization results. Indeed, the disturbances have been considered only for the definitions of Q, r and $\mathbf{c}$. In the rest of this section, we suppose that Assumption 2.1 holds as well as Assumption 2.2, so that every result proved in the previous sections remain valid. In particular, we assume that $\mathcal{S}$ and $\mathcal{A}$ are compact. To simplify the exposition, we will assume that $\mathcal{S} = [0, 1]^{n_s}$ and $\mathcal{A} = [0, 1]^{n_a}$. This method remains valid for any CMDP with $\mathcal{S}$ and $\mathcal{A}$ compact subsets of $\mathbb{R}^{n_s}$ and $\mathbb{R}^{n_a}$, respectively. This method can be also extended if $\mathcal{S}$ is a subset of a compact set of $\mathbb{R}^{n_s}$ (see [Chow, 1989]).

Inspired by the work of [Chow, 1989], we define two CMDPs. The first one is $(S, \tilde{A}_h, \tilde{U}_h, \tilde{Q}_h, \tilde{r}_h, \tilde{c}_h)$. This CMDP is built on the same state space as the initial CMDP. The second CMDP is $(\tilde{S}_h, \tilde{A}_h, \tilde{U}_h, \tilde{Q}_h, \tilde{r}_h, \tilde{c}_h)$, it is a discrete and finite CMDP. This second CMDP is equivalent to $(S, \tilde{A}_h, \tilde{U}_h, \tilde{Q}_h, \tilde{r}_h, \tilde{c}_h)$ except that it can be numerically solved using finite LP. These two CMDPs are described as follows.

• Let $n_h > 1$ be the number of sets constituting the partition of $[0, 1]$. Since we consider a uniform grid, the *grid size* is defined as $h = n_h^{-1}$. For this h, let $\mathcal{I}_h$ be the partition of $[0, 1]$ constituted by the sets of the form $[jh, (j + 1)h[$ and by the set $[(n_h - 1)h, 1]$. By extension, the space $\mathcal{S}$ is then partitioned as follows:

$$S_h = \{I_1 \times I_2 \times \dots I_{n_s} | I_j \in \mathcal{I}_h\}.$$

S_h is composed by a finite number of sets which are denoted as S_h^i for $i \in [0, n_h n_s - 1]$. For every set of S_h, we choose a representative. Let $\tilde{s}_i$ be the representative of the set S_h^i for every $i \in [0, n_h n_s - 1]$. For every $s \in \mathcal{S}$, $S_h(s)$ is the element of S_h to which s belongs. Similarly, for every $s' \in \mathcal{S}$ we define $\tilde{s}_h(s')$ the representative of the set $S_h(s')$. $\tilde{S}_h$ denotes the set of all representatives.

• The action set $\mathcal{A}$ is discretized so that every $\tilde{a} \in \tilde{A}_h$ can be written as a vector of size n_a with all its coordinates multiple of h. For every $s' \in \mathcal{S}$, the set $\tilde{A}_h(s')$ is defined as follows:

$$\tilde{A}_h(s') = \left\{ \tilde{a} \in \tilde{A}_h | \, ||u - \tilde{u}||_\infty \leq \frac{h}{2} \text{ for some } u \in \mathcal{U}\left(\tilde{s}_h\left(s'\right)\right) \right\}. \tag{2.71}$$

The set of all admissible state-action pairs is defined as

$$\tilde{U}_h = \left\{ (s, \tilde{a}) | s \in \mathcal{S} \text{ and } \tilde{a} \in \tilde{A}_h(s) \right\}.$$

• In this section, we will suppose that Q has a density $Q(s'|s, a)$ with respect to the Lebesgue measure. The density $\tilde{Q}_h(s'|s, \tilde{a})$ is defined as follows

$$\tilde{Q}_h(s'|s, \tilde{a}) = \frac{1}{h} \int_{S_h(s')} Q(s'|\tilde{s}_h(s), \tilde{a}) ds' \tag{2.72}$$

• Finally, the reward and cost functions are defined as

$$\tilde{r}_h(s, \tilde{a}) = r(\tilde{s}_h(s), \tilde{a}) \tag{2.73}$$

$$\tilde{\mathbf{c}}_h(s, \tilde{a}) = \mathbf{c}(\tilde{s}_h(s), \tilde{a}) \tag{2.74}$$

2.5.2 Assumptions

Obtaining results on the convergence of discrete states and action approximations requires assumptions that are stronger than Assumption 2.1. In this section, Assumption 2.1 is then replaced by Assumption 2.3 defined as follows.

Assumption 2.3.
(a) *The spaces* $\mathcal{S} = [0,1]^{n_s}$ *and* $\mathcal{A} = [0,1]^{n_a}$*, respectively.*
(b) *For all* $s, s' \in \mathcal{S}$ *and* $a, a' \in \mathcal{A}$*, there exist* r_M *and* K_r *such that* r *is a bounded by* r_M *and* r *is a* K_r*-Lipschitz function:*

$$\begin{cases} |r(s,a)| \leq r_M \\ |r(s,a) - r(s',a')| \leq K_r \, ||(s,a) - (s',a')||_\infty . \end{cases} \tag{2.75}$$

(c) *For all* $j \in \{1, ..., n_c\}$*, for all* $s, s' \in \mathcal{S}$ *and* $a, a' \in \mathcal{A}$*, there exist* M_{c_j} *and* K_{c_j} *such that* c_j *is a bounded by* $c_{M,j}$ *and* c_j *is* K_{c_j}*-Lipschitz function:*

$$\begin{cases} |c_j(s,a)| \leq c_{j,M} \\ |c_j(s,a) - c_j(s',a')| \leq K_{c_j} \, ||(s,a) - (s',a')||_\infty . \end{cases} \tag{2.76}$$

(d) *For all* $s, s' \in \mathcal{S}$ *and* $a, a' \in \mathcal{A}$*, there exists* K_Q *such that* Q *is* K_Q*-Lipschitz function in the total variation norm:*

$$||Q(\cdot|s,a) - Q(\cdot|s',a')||_{TV} \leq K_Q \, ||(s,a) - (s',a')||_\infty . \tag{2.77}$$

(e) *For every policy* $\varphi \in \Pi_{RS}$*, the MC induced on* $\mathcal{S}$ *by* φ *is a uniformly ergodic.*

We note by K the constant defined as

$$K = \max\left(M_r, K_r, \left\{\left(M_{c_j}, K_{c_j}\right)\right\}_{j \in n_c}, K_Q\right). \tag{2.78}$$

Henceforth, the functions r, c_j, and Q are all K-Lipschitz and the functions r and c_j are bounded by K.

Assumption 2.4. For every $\tilde{s} \in \tilde{S}_h$ and every $\tilde{a} \in \tilde{A}_h$ $\tilde{Q}_h$ verifies:

$$\left|\left|Q(\cdot|\tilde{s},\tilde{a}) - \tilde{Q}_h(\cdot|\tilde{s},\tilde{a})\right|\right|_{TV} \leq K'_q h \tag{2.79}$$

In [Chow, 1989], it has been shown that if the density $s' \mapsto Q(s'|s,a)$ is bounded by K and piecewise K-Lipshitz, then the quantized kernel defined by (2.72) satisfies Assumption 2.4.

2.5.3 Solving the discrete CMDP

Suppose that the (unconstrained) MDP $(\mathcal{S}, \mathcal{A}, \mathcal{U}, Q, r)$ verifies Assumption 2.1(e). It this case, it is shown in [Chow, 1989] that the discrete MDP, $(\tilde{S}_h, \tilde{A}_h, \tilde{U}_h, \tilde{r}_h)$ inherits this ergodicity condition (Assumption 2.1(e)). Based on this result the $(\tilde{S}_h, \tilde{A}_h, \tilde{U}_h, \tilde{r}_h)$ has been shown to be solvable and its optimal value $R^\star_h$ verifies:

$$|R^\star_h - R^\star| \leq K' h, \tag{2.80}$$

where h is the grid size and K' is a constant depending on K and K'_Q.

We now extend the result established by [Chow, 1989] for unconstrained MDPs to the case of CMDPs. Consider first the discretized problem

$$(\tilde{S}_h, \tilde{A}_h, \tilde{U}_h, \tilde{Q}_h, \tilde{r}_h, \tilde{\mathbf{c}}_h). \tag{2.81}$$

For every $l \in [0, N_U - 1]$, we denote by $\tilde{s}(\tilde{u}_l)$ and $\tilde{a}(\tilde{u}_l)$ the components of $\tilde{u}_l$ on $\tilde{S}_h$ and $\tilde{A}_h$ respectively; i.e. $\tilde{u}_l = (\tilde{s}(\tilde{u}_l), \tilde{a}(\tilde{u}_l))$. Using this notation, we define M as the matrix of size $N_S \times N_U$ which is composed by the following elements:

$$M_{i,l} = \begin{cases} 1, & \text{if } \tilde{s}(\tilde{u}_l) = \tilde{s}_i \\ 0, & \text{otherwise.} \end{cases}$$

Let also $\tilde{L}_{0,h}$ be the matrix of size $N_S \times N_U$ given by

$$\tilde{L}_{0,h} = M - \tilde{Q}_h. \tag{2.82}$$

Following [Altman, 1999], we have that solving the discrete CMDP is equivalent to solving the following finite LP:

$$\begin{aligned} \tilde{R}^\star_h = \sup_{\mathbf{m} \in \mathbb{R}^{N_U}} & \mathbf{R}\mathbf{m}^T \\ \text{s.t. } \tilde{L}_{0,h}\mathbf{m}^T &= \mathbf{0} \\ \mathbf{1}\mathbf{m}^T &= 1 \\ \mathbf{C}_j\mathbf{m}^T &\leq V_j \\ \mathbf{m} &\geq \mathbf{0}, \end{aligned} \tag{2.83}$$

where the notation $\mathbf{R}$, $\mathbf{C}_j$ and $\mathbf{1}$ stand for the vectors of size N_U containing respectively $r(\tilde{u}_l)$, $\mathbf{c}(\tilde{u}_l)$, and 1 for every $l \in [0, N_U - 1]$. Suppose now that there exists a solution to the finite LP proposed in equation (4.27) and let $\mathbf{m}^\star$ be a solution. From $\mathbf{m}^\star$, we build a policy for the continuous state and action spaces CMDP by considering the following policy:

$$\tilde{\varphi}(da|s) = \sum_{\tilde{a}} \frac{\mathbf{m}^\star(\tilde{s}_j, \tilde{a})}{\sum_{\tilde{s}} \mathbf{m}^\star(\tilde{s}, \tilde{a})} \delta_{\tilde{a}}(da), \ \forall s \in S_j. \tag{2.84}$$

To prove $\tilde{R}_h^\star$ converges to $R^\star$ when h tends to 0.

Theorem 2.5. *If the CMDP* $(\mathcal{S}, \mathcal{A}, \mathcal{U}, Q, r, \mathbf{c})$ *verifies Assumptions 2.2 and 2.3, if the discrete CMDP* $(S_h, \tilde{A}_h, \tilde{U}_h, \tilde{Q}_h, \tilde{r}_h, \tilde{\mathbf{c}}_h)$ *verifies Assumptions 2.2 and 2.4, then there exists* $K'' > 0$ *such that*

$$\left|\tilde{R}_h^\star - R^\star\right| \leq K'' h. \tag{2.85}$$

Proof. This proof is given in Appendix B.4. □

2.6 Partially Observable Markov Decision Processes

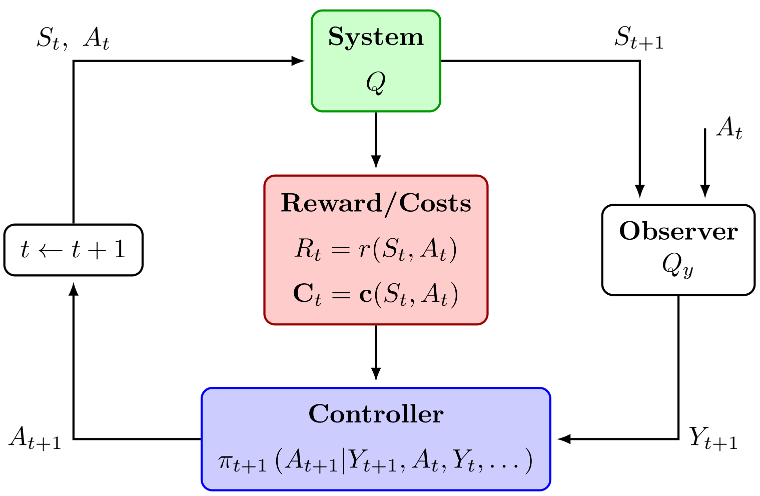

Figure 2.3 – *Partially Observable Markov Control Model*

In many real-life applications, the state of the system S_t may not be completely available to the controller. The controller may have access to S_t only through an observation Y_t. This situation is depicted in Figure 2.3. The Figure 2.3 is similar to 2.1 except that the state evolution is directly denoted by the stochastic kernel Q. Similarly, in Figure 2.3, the observation is also given by a stochastic kernel: Q_y. Since the controller has only access to the process $\{Y_t\}_{t\in\mathbb{N}}$ its policy takes only into account Y_t. The framework that encompasses this "partial observability" is called Partially Observable Markov Decision Process (POMDP). In this thesis, we will focus on a particular type of POMDPs where the state space $\mathcal{S}$ can be written as $\mathcal{X} \times \mathcal{Y}$ where $\mathcal{X}$ and $\mathcal{Y}$ respectively represent the non-observable part and the completely observable part of $\mathcal{S}$. In [Arapostathis et al., 1993], this model is referred to as the partial state information model. Consequently it will be denoted by Partial State Information Markov Decision Process (PSI-MDP) in the sequel. Since PSI-MDPs are a special case for POMDPs, every results shown (in particular in [Hernández-Lerma, 1989, Chapter 6]) remains valid.

The most common way for handling POMDP is to build an equivalent (completely observable) MDP(see [Bertsekas & Shreve, 1978, Chapter 10], [Hernández-Lerma, 1989, Chapter 6], [Arapostathis et al., 1993, Section 7], and references therein). The state space of this equivalent MDP is the set $\mathcal{P}(\mathcal{S})$ and is referred to as the *belief space*(the set of all probability measures on $\mathcal{S}$). In consequence, solving POMDPs is *theoretically* the same as solving MDPs.

In this section, we will face three major difficulties that are the following.

• The practical applications considered in Chapter 3 and Chapter 4 consider the long-term average criterion (to compute the throughput). As it is shown in [Yu & Bertsekas, 2004], the long-term average case is more difficult to handle than others. Indeed, even if in the general case we can provide equations to solve to obtain an optimal policy, it is generally difficult to guarantee the existence of a solution for these equations. In this chapter, we propose a sufficient condition, *based on our applications*, for guaranteeing the existence of an optimal solution.

• The second difficulty is that the space $\mathcal{P}(\mathcal{S})$ is generally not suited for numerical implementations. When the state space is finite with cardinality $|\mathcal{S}|$, the set $\mathcal{P}(\mathcal{S})$ is a continuous space of dimension $|\mathcal{S}|$; in this case, many methods has been proposed (see the survey given in [Aberdeen, 2003] and references there in). In our case, $\mathcal{S}$ is continuous; this makes $\mathcal{P}(\mathcal{S})$ to be infinite dimensional and most methods for finite $\mathcal{S}$ not appropriate. However, it has been observed in [Roy et al., 2005] that we need not consider the whole space $\mathcal{P}(\mathcal{S})$. This is due to the fact that the beliefs (the elements of $\mathcal{P}(\mathcal{S})$) of interest for the application lie in a space of much smaller dimension (this is referred to as "belief compression" in [Roy et al., 2005]). Based on that idea, a heuristic method for solving continuous POMDP has been proposed in [Zhou et al., 2010]. Their method is the following: i) project the beliefs on the exponential family of densities, this leads to a MDP of much smaller dimension, ii) solve this low dimensional MDP with classical tools.

• The last difficulty is that most applications proposed for POMDPs/PSI-MDPs are unconstrained. Although there exist methods for handling Constrained-POMDP(see [Isom et al., 2008], [Kim et al., 2011], and references therein), they are all given in the case of finite state space rarely adaptable to cases with continuous state spaces. In this section we extend the method proposed in [Zhou et al., 2010] to constrained PSI-MDPs.

2.6.1 Partially Observable Model

When adapting the general definition of POMDPs for problems with constraints, we get the following definition. A (constrained) POMDP is defined by a tuple $(\mathcal{S}, \mathcal{Y}, \mathcal{A}, Q, Q_y, r, \mathbf{c})$. For ease of presentation, we suppose that the initial state s_0 is known. Each component of the POMDP is defined as follows.

- $\mathcal{S}$ is a Borel space called the state space.
- $\mathcal{Y}$ is a Borel space called the observation space
- $\mathcal{A}$ is a Borel space called the action space. In this section we will suppose that for every state $s \in \mathcal{S}$, $A(s) = \mathcal{A}$.
- Q is the transition kernel of equation (2.11):

$$Q(B|s, a) = \mathbb{P}\left[S_{t+1} \in B | S_t = s, A_t = a\right].$$

- Q_y is the observation kernel. The definition of Q_y is the following:

$$Q_y(B|s, a) = \mathbb{P}\left[Y_t \in B | S_t = s, A_{t-1} = a\right], \forall B \in \mathcal{B}(\mathcal{Y}).$$

- r is the instantaneous reward
- $\mathbf{c}$ is the instantaneous cost vector

The classical definition of POMDP only differs from this definition by not having $\mathbf{c}$.In this POMDP context, the controller bases the choices of its actions with the *observable history* defined as:

$$h_t = (s_0, a_0, y_1, a_1, y_2, a_2 \ldots, y_{t-1}, a_{t-1}, y_t). \tag{2.86}$$

The different kinds of policies described in Section 2.2 remain the same except that they are defined on the observable history instead of the complete history. The long-term reward and cost functions ($R(s_0, \pi)$ and $\mathbf{C}(\pi, s_0)$) remain also unchanged.

2.6.2 The Partial State Information model

The PSI-MDP model is a particular type of POMDP in which the state space can be written as $\mathcal{S} = \mathcal{X} \times \mathcal{Y}$ where $\mathcal{X}$ and $\mathcal{Y}$ represent the non-observable and fully observable parts of the state space. For clarity of exposition, we will suppose that the spaces $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{A}$ are $[0, 1]^{n_x}$, $[0, 1]^{n_y}$, $[0, 1]^{n_s}$, and $[0, 1]^{n_a}$, respectively. In addition to these definitions, we will suppose that the transition kernel Q possesses a density q with respect to the Lebesgue measure. Consequently, Q has the following expression

$$Q(B|s,a) = \int_B q(x',y'|s,a)dx'dy'. \tag{2.87}$$

In the sequel we decompose $q(x',y'|x,y,a)$ as follows

$$q(x',y'|x,y,a) = q_X(x'|x,y,a,y')q_Y(y'|x,y,a) \tag{2.88}$$

Finally, a PSI-MDP model is defined by the tuple $(\mathcal{X},\mathcal{Y},\mathcal{A},Q,r,\mathbf{c})$, where each component (except $\mathcal{X}$ and $\mathcal{Y}$ defined above) is defined as in the definition of POMDP. Because y_t still represents the observation, policies for PSI-MDPs models are still established on the observable history given in equation (2.86). The set of all policies built on $\mathcal{H}_o$ is denoted by Π_o. A general policy $\pi_o \in \Pi_o$ is defined as a sequence $\pi_o = \{\pi_{o,t}(da_t|h_t)\}_{t\in\mathbb{N}}$, where $h_t \in \mathcal{H}_o$ (see Section 2.2).

Equivalent CMDP model

We will now show that the PSI-MDP model $(\mathcal{X},\mathcal{Y},\mathcal{A},Q,r,\mathbf{c})$ presented above is equivalent to a CMDP model $(\mathcal{Z},\mathcal{A},Q_Z,r',\mathbf{c}')$ such that:

- the state space $\mathcal{Z}$ is $\mathcal{P}(\mathcal{X}) \times \mathcal{Y}$, where $\mathcal{P}(\mathcal{X})$ is the space of all probability measures on $\mathcal{X}$ and $\mathcal{Y}$ is the observable part of $\mathcal{S}$,
- the action space is $\mathcal{A}$,
- Q_Z is a stochastic kernel on $\mathcal{Z}$ given $\mathcal{Z} \times \mathcal{A}$ that will be defined later,
- $r'(z,a)$ and $\mathbf{c}'(z,a)$ are the instantaneous reward and cost functions that will be defined later.

At some decision epoch $t \in \mathbb{N}$, let $z_t \in \mathcal{Z}$ be defined as the couple $(b_{x,t}, y_t)$, where $b_{x,t} = f(x_t|h_t)$ is the *a posteriori distribution* of X_t given an observable history $\mathcal{H}_t = h_t$ and Y_t is the fully observable part of S_t. We now give the expression of the stochastic kernel Q_Z defined, for every $B \in \mathcal{B}(\mathcal{Z})$, as

$$Q_Z(B|z_{t-1},a_{t-1}) = \mathbb{P}\left[Z_t \in B|Z_{t-1} = z_{t-1}, A_{t-1} = a_{t-1}\right]. \tag{2.89}$$

To build Q_Z, we will first describe the evolution of $b_{x,t}$, and in a second time we will study the evolution of y_t. We will now show that the dynamical evolution of $b_{x,t}$ can be described by means of the following deterministic relationship:

$$b_{x,t} = H_b(z_{t-1},a_{t-1},y_t). \tag{2.90}$$

This relationship (as proved in Appendix B.5) has the following expression:

$$b_{x,t}(x_t) = \frac{\int_{\mathcal{X}} q\left(x_t, y_t | x_{t-1}, y_{t-1}, a_{t-1}\right) b_{x,t-1}(x_{t-1}) dx_{t-1}}{\int_{\mathcal{X}} \int_{\mathcal{X}} q\left(x'_t, y_t | x_{t-1}, y_{t-1}, a_{t-1}\right) b_{x,t-1}\left(x_{t-1}\right) dx_{t-1} dx'_t}. \tag{2.91}$$

To complete the definition of Q_Z we now build Q'_Y, a stochastic kernel on $\mathcal{Y}$ given $\mathcal{Z} \times \mathcal{A}$. Q'_Y is the probability that Y_t belongs to C given that $Y_{t-1} = y_{t-1}$, X_{t-1} is distributed according to $b_{x,t-1}$ and the controller chooses action A_{t-1}. For all $C \in \mathcal{B}(\mathcal{Y})$, Q'_Y is defined as follows

$$\begin{aligned} Q'_Y(C|z_{t-1}, a_{t-1}) &= \mathbb{P}\left[Y_t \in C | Z_{t-1} = z_{t-1}, A_{t-1} = a_{t-1}\right], & (2.92) \\ &= \int_X Q_Y(C|x_{t-1}, y_{t-1}, a_{t-1}) b_{x,t-1}(x_{t-1}) dx_{t-1} & (2.93) \end{aligned}$$

where Q_Y is the marginal distribution of Q defined as

$$Q_Y(C|x_{t-1}, y_{t-1}, a_{t-1}) = \int_C q_Y\left(y_t | x_{t-1}, y_{t-1}, a_{t-1}\right) dy_t. \tag{2.94}$$

Consequently, the stochastic evolution of Z_t can be described by a stochastic kernel Q_Z on $\mathcal{Z}$ given $\mathcal{Z} \times \mathcal{A}$. For every $B \in \mathcal{B}(\mathcal{P}(\mathcal{X}))$ and $C \in \mathcal{B}(\mathcal{Y})$, Q_Z is defined as follows:

$$\begin{aligned} &Q_Z(B \times C|z_{t-1}, a_{t-1}) = \mathbb{P}\left[Z_t \in B \times C | Z_{t-1} = z_{t-1}, A_{t-1} = a_{t-1}\right] \\ &\quad = \int_C \mathbb{P}\left[B_{x,t} \in B | Z_{t-1} = z_{t-1}, A_{t-1} = a_{t-1}, Y_t = y_t\right] Q'_Y\left(dy_t | z_{t-1}, a_{t-1}\right) \\ &\quad = \int_C \mathbb{1}\left(H_b(z_{t-1}, a_{t-1}, y_t) \in B\right) Q'_Y\left(dy_t | z_{t-1}, a_{t-1}\right) \qquad (2.95) \end{aligned}$$

To complete the definition of the equivalent CMDP, $(\mathcal{Z}, \mathcal{A}, Q_z, r', \mathbf{c}')$, it remains to define the one-step reward and cost functions r' and $\mathbf{c}'$. By analogy to the literature (see [Bertsekas & Shreve, 1978] and [Hernández-Lerma, 1989]) we define these functions as

$$\begin{cases} r'(z, a) = \int_{\mathcal{X}} r(x, y, a) b_x(dx) \\ \mathbf{c}'(z, a) = \int_{\mathcal{X}} \mathbf{c}(x, y, a) b_x(dx), \end{cases} \tag{2.96}$$

where $z = (b_x, y)$.

In the end, we obtain the following CMDP: $(\mathcal{Z}, \mathcal{A}, Q_z, r', \mathbf{c}')$. A history for this CMDP model is referred to as an *information vector* and is defined as:

$$\iota_t = \left(x_0, y_0, a_0, b_{x,1}, y_1, a_1, \ldots b_{x,t}, y_t\right), \tag{2.97}$$

where we recall that x_0 is part of the known initial state s_0. The set of these information vectors is denoted by $\mathcal{H}_\iota$. A policy for this CMDP is referred to as an *information policy*. The set of all information policies is denoted by Π_ι. A general policy $\pi_\iota \in \Pi_\iota$ is defined as a sequence $\pi_\iota = \{\pi_{\iota,t}(da_t|\iota_t)\}_{t\in\mathbb{N}}$ (see Section 2.2).

The proof of the equivalence of this CMDP and the initial PSI-MDP is the same as the one presented in [Arapostathis et al., 1993] hence the complete proof is omitted here. Their proof is done by observing that the beliefs $b_{x,1} \dots b_{x,t}$ of every ι_t can be obtained from a particular h_t by using equation (2.91). This is then used to prove that for every π_ι, there exists π_o such that $\mathbb{E}^{s_0}_{\pi_\iota}[\cdot] = \mathbb{E}^{s_0}_{\pi_o}[\cdot]$. This proves that for every π_ι, there exists π_o such that $R(\pi_o, s_0) = R(\pi_\iota, s_0)$ and $\mathbf{C}(\pi_o, s_0) = \mathbf{C}(\pi_\iota, s_0)$. Since showing that for every π_o, there exists π_ι such that $R(\pi_o, s_0) = R(\pi_\iota, s_0)$ and $\mathbf{C}(\pi_o, s_0) = \mathbf{C}(\pi_\iota, s_0)$, they have proved the equivalence between the initial PSI-MDP and the CMDP $(\mathcal{Z}, \mathcal{A}, Q_z, r', \mathbf{c}')$.

2.6.3 Existence of a solution of the CMDP

We now give sufficient conditions on the initial PSI-MDP so that the CMDP $(\mathcal{Z}, \mathcal{A}, Q_z, r', \mathbf{c}')$ is solvable. These conditions are principally designed so that $(\mathcal{Z}, \mathcal{A}, Q_z, r', \mathbf{c}')$ verifies Assumption 2.1.

Condition 2.2.

(a) *$\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{A}$ are compact sets.*

(b) *r and $\mathbf{c}$ are bounded and continuous functions.*

(c) *Q is weak Feller (see Assumption 2.1(d)).*

(d) *H_b is a continuous function on $\mathcal{Z} \times \mathcal{A} \times \mathcal{Y}$.*

(e) *i) There exists a measure ν_Y such that $\nu_Y(\mathcal{Y}) > 0$ and for all $s \in \mathcal{S}$ and all $a \in \mathcal{A}$, $Q_Y(C|s, a) \geq \nu_Y(C)$.*

ii) There exists a set $C_0 \in \mathcal{B}(\mathcal{Y})$ with $\nu_Y(C_0) > 0$ and a measure $m_0 \in \mathcal{P}(\mathcal{X})$ such that for every $y \in C_0$, $z \in \mathcal{Z}$ and every $a \in \mathcal{A}$, $H_b(z, a, y) = m_0$.

Conditions 2.2(a)-2.2(d) are almost identical to Assumptions of [Hernández-Lerma, 1989]; the only differences are that we impose the compactness of $\mathcal{X}$ and $\mathcal{Y}$ which is not required in [Hernández-Lerma, 1989] and that we do not impose Q_Y to be weak Feller since in our case this is obtained automatically by showing that Q is weak Feller. Conditions 2.2(a)-2.2(d) are meant to guarantee that the CMDP given by $(\mathcal{Z}, \mathcal{A}, Q_z, r', \mathbf{c}')$ verifies Assumptions 2.1(a)-2.1(d) (see [Hernández-Lerma, 1989]). In consequence, to prove that the CMDP $(\mathcal{Z}, \mathcal{A}, Q_z, r', \mathbf{c}')$ is solvable, it only remains to prove that it verifies Assumption 2.1(e). As proposed in the following therorem, Assumption 2.1(e) is a consequence of Condition 2.2(e) (to the best of the author knowledge this condition is specific to this thesis).

Theorem 2.6. *If Condition 2.2(e) is satisfied, then the CMDP* $(\mathcal{Z}, \mathcal{A}, Q_z, r', \mathbf{c}')$ *verifies Assumption 2.1(e).*

Proof. The proof of this theorem is given in Appendix B.6. □

We have finally shown that Condition 2.2 implies that the CMDP given by $(\mathcal{Z}, \mathcal{A}, Q_z, r', \mathbf{c}')$ verifies Assumption 2.1. Using Theorem 2.1, we have that if there exists $\pi \in \Pi_o$ such that $\mathbf{C}(\pi, s_0) \leq \mathbf{V}$, the CMDP $(\mathcal{Z}, \mathcal{A}, Q_z, r', \mathbf{c}')$ is solvable.

To conclude this subsection, we will do the following remarks:

Remark 2.1. *[Hernández-Lerma & Lasserre, 1996, Chapter 6]: If* $\mathcal{Y}$ *is countable, then Condition 2.2(c) implies Condition 2.2(d).*

This remark will be of practical interest in Chapter 3 and 4. Indeed this remark means that we will only need to prove that Q is weak Feller to prove Condition 2.2(c) and Condition 2.2(d).

Remark 2.2. *If there exists a state* $s_0 = (x_0, y_0)$ *and* $\epsilon > 0$ *such that: for every* $s \in \mathcal{S}$ *and every* $a \in \mathcal{A}$*,* $Q(\{s_0\} | s, a) > \epsilon$ *and* $q_X(dx'|x, y, a, y_0) = \delta_{x_0}(dx')$*, then Condition 2.2(e) is checked.*

Proof. This proof is given in Appendix B.7. However, we can do two remarks: i) $Q(\{s_0\} | s, a) > \epsilon$ is equivalent to saying that s_0 is accessible from every other states for any actions, and ii) $Q_X(dx'|x, y, a, y_0) = \delta_{x_0}(dx')$ is equivalent to saying that the state $s_0 \in \mathcal{S}$ is completely observable. This provides a physical interpretation of 2.2: if there is a completely observable recurrent state, then Condition 2.2(e) is checked. □

So far, we have given theoretical results concerning PSI-MDPs. In the next subsection we give a heuristic way of solving PSI-MDPs.

2.6.4 Numerical solutions for Partial State Information models

In this section, we do a brief presentation of the heuristic method presented in [Zhou et al., 2010] to approximate continuous (unconstrained) POMDP. Since PSI-MDP are a special case of POMDP, we will present this method from the PSI-MDP point of view. We then prove that without any change, this method can be applied to constrained PSI-MDPs.

The main idea behind the method presented in [Zhou et al., 2010] is that the beliefs $b_{x,t}$ live in a finite dimensional subset of $\mathcal{P}(\mathcal{X})$. Hence they decide to approximate the continuous MDP, $(\mathcal{Z}, \mathcal{A}, Q_z, r')$ by projecting the beliefs $b_{x,t}$

on a space of small dimension. This projected-MDP can then be solved by any usual methods.

Their method is the following. Let Ψ be a set of parametrized pdfs, $\Psi = \{f(\cdot,\theta), \theta \in \Theta_0\}$, where $\Theta_0 \subset \mathbb{R}^{n_\theta}$. They propose to project the beliefs using the *Kullback-Liebler divergence* defined as follows

$$D_{KL}(b_x||f) = \int_{\mathcal{X}} b_x(x') \log \frac{b_x(x')}{f(x')} dx'. \tag{2.98}$$

For every $b_x \in \mathcal{P}(\mathcal{X})$, the projection of b_x on Ψ is given by

$$\text{Proj}_\Psi(b_x) = \arg\min_{g \in \Psi} D_{KL}(b_x||g), \tag{2.99}$$

or equivalently, considering the θ parameter

$$\hat{\theta}(b_x) = \arg\min_{\theta \in \Theta_0} D_{KL}(b_x||f(\cdot,\theta)). \tag{2.100}$$

From this projection, [Zhou et al., 2010] has proposed to solve the following projected MDP $(\mathcal{Z}', \mathcal{A}, Q'_z, r'')$ where each component is defined as follows:

- $\mathcal{Z}' = \Theta_0 \times \mathcal{Y}$,
- $\mathcal{A}$ is unchanged,
- Q'_z is defined, for every $z' = (\theta', y')$ and every $y \in \mathcal{Y}$ as

$$Q'_Z(B \times C|z', a) = \int_C \mathbb{1}\left(H_\theta(z', a, y) \in B\right) Q'_Y\left(dy|z', a\right), \tag{2.101}$$

where $H_\theta(z', a, y)$ is the function from $\Theta_0 \times \mathcal{Y} \times \mathcal{A} \times \mathcal{Y}$ to Θ_0 defined as

$$H_\theta(\theta', y', a, y) = \hat{\theta}\left(H_b\left(f\left(\cdot, \theta'\right), y', a, y\right)\right), \tag{2.102}$$

- r'' is defined as follows: $r''(\theta, y, a) = r'(f(\cdot,\theta), y, a)$

This MDP has a continuous state space of dimension $n_\theta + n_y$ which is finite. Furthermore, in [Zhou et al., 2010], they propose to take Ψ as a parametrized set of policy from the exponential family of pdfs. In this case n_θ is often small, making the whole state of small dimension if $\mathcal{Y}$ has a small dimension. Finally, this MDP is solved using any method.

In prevision of telecommunication applications, we need to be able to take into account constraints. The method in [Zhou et al., 2010] can be extended to constrained PSI-MDP by considering $(\mathcal{Z}', \mathcal{A}, Q'_z, r'', \mathbf{c}'')$, where $\mathbf{c}''$ is obtain by the same transform as r''. This time $(\mathcal{Z}', \mathcal{A}, Q'_z, r'', \mathbf{c}'')$ is a finite dimensional

CMDP. In this thesis, we have applied the method of Section 2.5 to solve this CMDP.

2.7 Conclusions

To conclude this chapter, we have done a review on the CMDPs framework. We have in particular introduced the following notions as parts of the definition of a CMDP:

- $\mathcal{S}$ is the state space,
- $\mathcal{A}$ is the action space,
- $\mathcal{U}$ is the space of all admissible state-action pairs,
- $\mathcal{W}$ is the disturbance space,
- $Q(B|s,a)$ is the *transition law*,
- $r(s,a)$ is the short-term reward function,
- $\mathbf{c}(s,a)$ is the vector of short-term cost function.

We have introduced the notion of policy and specified the following types of policies:

- Π the general set of policies,
- Π_M the set of Markov policies,
- Π_{RS} the set of Randomized Stationary policies,
- Π_{DS} the set of Deterministic Stationary policies.

We have introduced two infinite-horizon average-cost performance criteria, R and $\mathbf{C}$ defined as

$$\begin{cases} R(\nu_0,\pi) = \liminf_{T\to\infty} \frac{1}{T}\mathbb{E}_\pi^{\nu_0}\left[\sum_{t=0}^{T-1} r(S_t,A_t)\right] \\ \mathbf{C}(\nu_0,\pi) = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_\pi^{\nu_0}\left[\sum_{t=0}^{T-1} \mathbf{c}(S_t,A_t)\right]. \end{cases}$$

From these performance criteria we have introduced a constrained optimization problem:

$$\begin{aligned} R^\star(\nu_0) = \sup_{\pi\in\Pi} \quad & R(\nu_0,\pi) \\ s.t. \quad & \mathbf{C}(\nu_0,\pi) \leq \mathbf{V}. \end{aligned}$$

We have shown that under Assumption 2.1, this optimization problem is solvable and that a solution can be found in Π_{RS}. We have then shown that, under Assumption 2.1, solving the optimization problem in Π_{RS} is equivalent to solving an infinite dimensional LP. Approximations of this LP are obtained by discretizing the state and action spaces and solving a finite LP. Based on the infinite dimensional LP and on the relationship with its dual, we have provided Assumption 2.3 that guarantees the convergence of the finite approximations to the infinite solution. Finally, we have introduced the constrained POMDP and constrained PSI-MDP models. For the PSI-MDP model, we have shown that these problems can be reduced to CMDP with an infinite dimensional state space. Based on this new CMDP, we finally have presented a heuristic method to solve constrained PSI-MDPs. This method is inspired from the literature of the unconstrained PSI-MDP.

The main contribution of this chapter has been to provide a new condition, Condition 2.2, for ensuring that the CMDP corresponding to the PSI-MDP model is solvable. This condition is verifiable in practice, for the applications with a completely observable state accessible from every other states.

Perspectives and future works

Because many resource allocation for the PHY layer of telecommunications systems can be seen as a problem where decisions are made sequentially, MDPs (in the wide sense) are an appropriate tool. On the other hand, when using the MDP framework to resource allocation for the PHY layer of telecommunications systems, we generally have to address difficult problems: in fact, in the end, we have considered a setting with: a constrained + partially observable + Borel state and action space + under average-cost MDP. In [Yu & Bertsekas, 2004], the partially observable problem with finite state space and long-term average-cost is told to be a difficult problem. In consequence, there are still challenging research to do on MDPs, even after 60 years of abundant research.

In particular, we did not provide conditions so that the projected-CMDP of Section 2.6.4 is solvable. To prove that the proposed approximation can (actually) be solved will be part of future work. Our first idea for this work is that the following condition should suffice:

Condition 2.3.
(a) *The Condition 2.2 is verified for the PSI-MDP.*
(b) *For every function on $\mathcal{X}$ v which is bounded and continuous, the function $\bar{v}$ defined as*

$$\bar{v}(\theta) = \int_{\mathcal{X}} v(x) f(x,\theta) dx, \theta \in \Theta_0 \tag{2.103}$$

is continuous on Θ_0.

The second perspective of this work is the following: it is shown in [Chow, 1989] that the bounds on the approximations presented in Section 2.5 are loose for $h > 0$ (even if they are good when $h \to 0$). This means that we cannot use these bound to determine h that guarantees to be ϵ-close to the optimum. For a lot of applications, this can be a real problem. Methods that provide better approximations and/or better bounds exist in the literature, however they are rarely suited for constrained problems. Finding approximations that are implementable and that provide a good insight on how close we are for the optimal solution is a challenging open problem.

- Chapter 3 -

Power Allocation for HARQ protocols

Contents

In this chapter, we show that the CMDP framework presented in Chapter 2 is an appropriate framework to solve power allocation problems for various HARQ protocols.

3.1 Introduction

AMC has been widely adopted for link adaptation to improve the spectral efficiency of modern communication systems. This technique dynamically adapts the modulation (type and size) and coding (rate) parameters to the variations of the channel. In modern wireless standards, AMC is assisted by the use of a Channel Quality Indicator (CQI) feedback. The CQI feedback is a quantized information on the quality of the channel and the parameters of the receiver (including HARQ protocols for example). In CQI-based systems, the transmitter (Tx) selects its modulation and coding schemes according to the CQI. In our context, Figure 3.1 is a generalization of the model proposed in Chapter 1 taking into account explicitly the CQI.

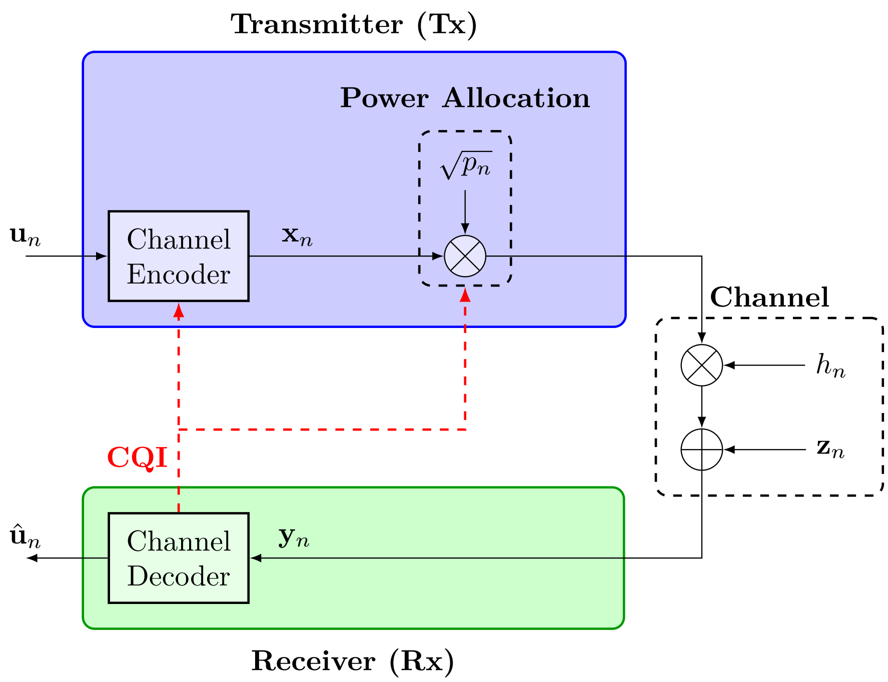

Figure 3.1 – *CQI-based system model*

Although the role of CQI is clearly defined, the choice of the quantity mapped by the CQI is left open in standards such as LTE. It has been shown in the literature that in case of varying channels, taking the mean SNR may not be sufficient. Indeed, in systems where the codewords span multiple fading blocks (such as HARQ systems), different sequences with the same average SNR may lead to very different FER. To overcome this difficulty, other types of CQI have been proposed. The most common indexes are Exponential Effective SNR Mapping (EESM) and Mutual Information Effective SNR Mapping (MIESM) (see [Wan et al., 2006] and references therein). The EESM has the advantage of being simple to compute but seems to be less accurate than the MIESM. A trade-off between complexity and accuracy has been proposed more recently in [Stupia et al., 2009] and is referred to as κ-Effective SNR Mapping (κ-ESM). In every cases (EESM, MIESMor κ-ESM) indexes are all "effective SNRs" and can be interpreted as: the SNR of the equivalent AWGN channel which gives the same FER. The main advantage of such types of CQI is that we only need to know the performance of the coded modulations for the AWGN channel. In the spirit of MIESM, other methods have been proposed based on ACMI that do not require a mapping to an effective SNR. In particular, a mutual information based bit-loading algorithm has been proposed in [Li & Ryan, 2007]. In [Cheng et al., 2003], [Stiglmayr et al., 2007], and [Pfletschinger & Navarro, 2010], heuristic methods are proposed for choosing the modulation type and dimensioning the length of incremental blocks for adaptive IR-HARQ protocols. In the same context, a more formal optimization has been proposed for power allocation in adaptive IR-HARQ protocols in [Tuninetti, 2011].

In this chapter, we consider the system depicted in Figure 3.1 except that the CQI only conveys information about the state of the HARQ protocol at time n, not about the current channel state. This means that throughout the communication Tx has only OCSI and SCSI. In our case the CQI is a quantized version of the state s_n of the HARQ protocol in the feedback channel described in Section 1.4.2. This context has already been proposed in [Szczecinski et al., 2011] to propose a rate adaptive IR-HARQ protocol. Our work differs from [Szczecinski et al., 2011] in the fact that we propose a power allocation.

The rest of this chapter is organized as follows. In Section 3.2 we will show that the problem of allocating power to optimize the throughput of a HARQ protocol under a constraint on the average power is a CMDP, in the case of Type-I HARQ. We will show that this CMDP verifies the assumptions of Chapter 2. These assumptions imply that we can approximate the solution of the CMDP with finite linear programming. These assumptions also imply that approximations provided by the finite linear programming converge to the optimal power allocation of the initial optimization problem. In Section 3.3, the same analysis is performed for the IR-HARQ protocol, and the CC-HARQ protocol.

3.2 Power allocation for the Type-I HARQ protocols

In this section, we propose to analyse, under the scope of CMDPs presented in Chapter 2, the power allocation for the maximization of the throughput under average power and peak power constraints.

3.2.1 CMDP model associated with the Power allocation for the Type-I HARQ protocols

In this subsection, we first introduce the CMDP model associated with the power allocation for the maximization of the throughput under average power and peak power constraints. Secondly, we will show that the proposed CMDP verifies Assumption 2.3. This assumption provides sufficient conditions for proving that the discrete approximations converge to the solution obtained with continuous action space (we will see that in this case $\mathcal{S}$ is already discrete).

Definition of the CMDP

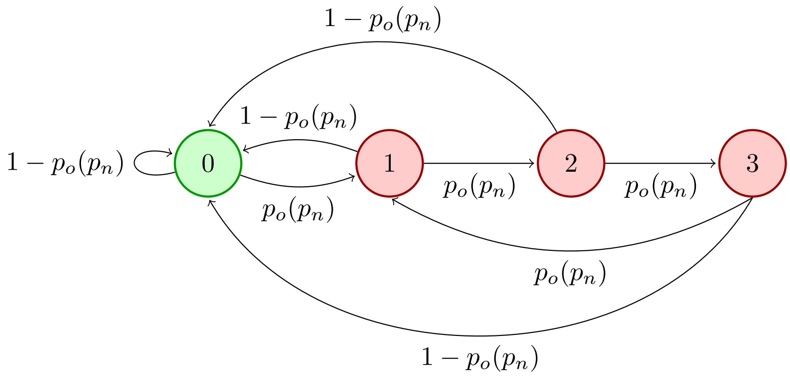

Figure 3.2 – *State Diagram of the Markov chain $\{K_n\}$ for a Type-I HARQ protocol with $N_T = 3$.*

We now define the components $(\mathcal{S}, \mathcal{A}, \mathcal{U}, \mathcal{W}, Q, r, \bar{p})$ of the CMDP associated with the Type-I HARQ protocol. The state space of the Type-I HARQ protocol has been suggested in Section 1.4.1 and is the following $\mathcal{S} = \{0, 1, \ldots, N_T\}$. We have in particular shown that these $N_T + 1$ states have the following meanings:

- **$'0'$:** Tx starts the transmission of a new information packet after a successful decoding (ACK),
- **$'1'$:** Tx has done 1 attempt and a NACK bit is received,
- **$'2'$:** Tx has done 2 attempts and a NACK bit is received,

⋮

- **$'N_T'$:** Tx has done N_T attempts and NACK is received, this state corresponds to an outage event and, in consequence, to the start of the transmission of a new information packet.

The action space $\mathcal{A}$ is defined as the set of available powers to Tx. In our case we will consider $\mathcal{A} = [P_{min}, P_{max}]$. The powers P_{min} and P_{max} constitute two peak power constraints. Although P_{max} is a classical peak power constraint, P_{min} is a much less classical constraint. In fact, we will see that this constraint has only a theoretical range. We will assume $P_{min} > \epsilon$ with $\epsilon \geq 0$. We will further suppose that Tx can always use the whole set of powers. This implies that $A(s) = \mathcal{A}$ for every $s \in \mathcal{S}$ and $\mathcal{U} = \mathcal{S} \times \mathcal{A}$.

In Section 1.4.1, we have shown that the state evolution is based on the random GNR: α. Consequently we take $\mathcal{W} = \mathbb{R}^+$, the set of every possible value for the GNR. For the Rayleigh block fading channel, the channel gain

α has an exponential distribution with average $\bar{\alpha} = \mathbb{E}[\alpha]$. Also, based on the value of α and a, we have given in Table 1.1 the rules of the evolution of the state. These rules have lead to the transition law given in equation (1.33). This transition law has the following expression:

$$Q(s'|s,a) = \begin{cases} 1 - p_o(a) \text{ if } s' = 0 \\ p_o(a) \text{ if } s' = s+1,\ s < N_T \\ p_o(a) \text{ if } s' = 1,\ s = N_T, \\ 0 \text{ otherwise} \end{cases} \tag{3.1}$$

where $p_o(a) = \mathbb{P}[\log_2(1 + a\alpha) < R]$. For the Rayleigh block fading channel $p_o(a)$ can be computed analytically as

$$p_o(a) = 1 - \exp\left(-\frac{2^R - 1}{\bar{\alpha}a}\right). \tag{3.2}$$

For every $s \in \mathcal{S}$ and every $a \in \mathcal{A}$, we have given the short-term reward as the reward associated with the computation of the (long-term) throughput. This reward has the following expression:

$$r(s,a) = \begin{cases} b \text{ if } s = 0 \\ 0 \text{ otherwise.} \end{cases} \tag{3.3}$$

In equation (3.3) we wrote $r(s,a)$ to be faithful to the definition of short-term reward of Chapter 2. However, in our case, r only depends on s.

Similarly to what we have proposed for the short-term reward, we introduce a short-term cost associated with the average power. This cost is simply defined as follows:

$$\bar{p}(s,a) = a. \tag{3.4}$$

This short-term cost definition concludes the description of the component of the CMDP: $(\mathcal{S}, \mathcal{A}, \mathcal{U}, \mathcal{W}, Q, r, c)$. This CMDP definition also settles the history and policy definitions (see Section 2.2).

Finally, the power allocation problem for maximizing the throughput under peak power and average power constraints can be written under the CMDP formalism as follows. Without loss of generality, we assume that the initial state is $s_0 = 0$. The throughput of the policy π when the initial state is s_0 is defined as follows

$$\eta(s_0, \pi) = \liminf_{T\to\infty} \frac{1}{T} \mathbb{E}_\pi^{s_0}\left[\sum_{t=0}^{T-1} r(S_t, A_t)\right]. \tag{3.5}$$

Similarly to the definition of the throughput, we introduce the average power as

$$\bar{P}(s_0, \pi) = \liminf_{n\to\infty} \frac{1}{n}\mathbb{E}_{\pi}^{s_0}\left[\sum_{i=0}^{n-1} \bar{p}(S_t, A_t)\right].$$

Finally, the power allocation problem is written as

$$\begin{aligned} \eta^\star(s_0) = \sup_{\pi\in\Pi} \quad & \eta(s_0, \pi) \\ s.t. \quad & \bar{P}(s_0, \pi) \leq P_A, \end{aligned} \tag{3.6}$$

where P_A is the average power constraint. Note that in the optimization problem given by equation (3.6), the peak power constraints does not appear explicitly. These constraint are in fact contained in Π because of the definition of $\mathcal{A}$.

Properties of the proposed CMDP model

We now analyse this CMDP model to prove that Assumption 2.3 is verified. This assumption will, in turn, allow us to build a finite LP that solves the power allocation problem.

Because of the finiteness of $\mathcal{S}$, Assumption (2.3(a)) trivially holds. Since the short-term reward and cost functions given by equations (3.3) and (3.4) are respectively constant and linear in a, the Lipschitz continuity of these functions trivially holds and Assumptions 2.3(b) and 2.3(c) are verified. In case of Rayleigh fading channel, we show in Appendix C.1 that Assumption 2.3(d) holds.

It finally remains to prove Assumption 2.3(e). We prove this Assumption by showing that Condition 2.1(a) holds. The accessible state needed by Condition 2.1(a) is $s_0 = 0$, indeed, for every $s \in \mathcal{S}$ we have

$$\begin{aligned} Q(s_0|s, a) &= 1 - p_o(a) \\ &\geq 1 - p_o(P_{min}) \\ &= \exp\left(-\frac{2^R - 1}{\bar{\alpha}P_{min}}\right) \\ &> 0, \end{aligned}$$

where the last inequality holds because $P_{min} > 0$.

3.2.2 The Finite Linear Programming

In the preceding subsection, we have shown that the CMDP corresponding to the power allocation problem for Type-I HARQ protocol verifies Assumption 2.3. This has mainly two consequences: i) since Assumption 2.3 implies Assumption

2.1, Theorem (2.1) guarantees that the optimization problem given by (3.6) is solvable at the condition that $P_A \geq P_{min}$.The condition $P_A \geq P_{min}$ ensures that the power allocation with a constant power of P_{min} is admissible and then that the set of admissible policies (Ω in Theorem (2.1)) is non-empty.

ii) Under Assumption 2.3, Theorems 2.1 and 2.2 imply that the optimization problem given by equation (3.6) is equivalent to an infinite dimensional LP.

We now build the finite LP approximating the infinite LP of Theorem 2.2. To do so, we introduce the following finite CMDP: $(\tilde{S}_h, \tilde{A}_h, \tilde{U}_h, \tilde{Q}_h, \tilde{r}_h, \tilde{\tilde{p}}_h)$. As we have already observed in the preceding section, for Type-I HARQ, $\mathcal{S}$ is already discrete. We then only need to discretize the set $\mathcal{A}$. The discretized action space $\tilde{A}_h$ is defined, as proposed in Section 2.5, as

$$\tilde{A}_h = \{P_{min}, P_{min} + h, P_{min} + 2h \cdots, P_{min} + (N_a - 1)h = P_{max}\},$$

where P_{min} and P_{max} are the peak powers, N_a is the number of actions considered in the discrete set, and h is defined as

$$h = \frac{P_{max} - P_{min}}{N_a - 1}. \tag{3.7}$$

Note that in our case, $\mathcal{A}$ is not $[0, 1]$ but the discretization procedure is identical. Moreover, since $\mathcal{U} = \mathcal{S} \times \mathcal{A}$, the set $\tilde{U}_h$ is in our case, the set $\mathcal{S} \times \tilde{A}_h$. In this case the number of elements contained in $\tilde{U}_h$ is $N_U = N_T N_a$.

Since $\mathcal{S}$ is discrete, the discretization of the transition law is trivial: Q is unchanged. To have a more compact notation, we introduce for every $\tilde{a} \in \tilde{A}_h$, the notation $\tilde{Q}_{\tilde{a}}$; $\tilde{Q}_{\tilde{a}}$ is the transition matrix, parametrized by $\tilde{a}$, that is built as follows:

$$\tilde{Q}_{\tilde{a}}(s, s') = Q(s'|s, \tilde{a}). \tag{3.8}$$

Because $\mathcal{S}$ is already finite, the reward and cost functions are also unchanged: $\tilde{r}_h = r$ and $\tilde{c}_h = c$.

With these notations, the finite LP is finally given by:

$$\begin{aligned} \tilde{\eta}^\star = & \sup_{\mathbf{m} \in \mathbb{R}^{N_U}} \mathbf{R}\mathbf{m}^T \\ \text{s.t.} & \sum_{s \in \mathcal{S}} m(s, \tilde{a}) - Q_{\tilde{a}} \mathbf{m}^T = \mathbf{0}, \ \forall \tilde{a} \in \tilde{A}_h \\ & \mathbf{1}\mathbf{m}^T = 1 \\ & \bar{\mathbf{P}}\mathbf{m}^T \leq P_A \\ & \mathbf{m} \geq \mathbf{0}, \end{aligned}$$

where $\mathbf{R}$, $\bar{\mathbf{P}}$ and $\mathbf{1}$ are the vectors of size N_U containing respectively $r(\tilde{u})$ for all $\tilde{u} \in \tilde{U}_h$, $\bar{p}(\tilde{u})$ for all $\tilde{u} \in \tilde{U}_h$ and only ones. Once the optimal $\mathbf{m}^\star$ is found, the corresponding policy consists in choosing randomly a power in $\tilde{A}_h$ with a distribution:

$$\varphi(\tilde{a}|s) = \frac{\mathbf{m}^\star(s, \tilde{a})}{\sum_{s'} \mathbf{m}^\star(s', \tilde{a})}. \tag{3.9}$$

3.2.3 Simulation results

The finite approximation method detailed in the preceding section has been evaluated by simulation. We consider a HARQ protocol with a maximum number of transmission attempts $N_T = 5$ and an initial rate $R = 3$ bpcu. In these results, we consider without loss of generality, that the GNR α is exponentially distributed with average $\bar{\alpha} = 1$. When a power allocation is performed, we consider a set $\tilde{A}_h$ from $P_{min} = -10dBW$ to $P_{max} = 10dBW$ with $N_a = 128$. In Figure 3.3 we compare three results. These three results are all compared at a given average SNR. Note that since in each slot a_n is independent of α_n, the average SNR can be computed as $P_A\bar{\alpha} = P_A$.

• The first result is depicted in Figure 3.3 as the blue curve with triangle marker. This curve corresponds to the result of the finite LP: $\tilde{\eta}^\star$ as a function of P_A.

• The second result is depicted in Figure 3.3 as the red curve with triangle marker. This curve corresponds to the result of the simulation of the Type-I HARQ protocol using the power allocation $\varphi(\tilde{a}|s)$ computed thanks to the finite LP. These simulations has been realized by the Monte-Carlo method over 10^4 slots.

• The third result is depicted in Figure 3.3 as the violet curve with square marker. This curve corresponds to the constant power allocation case. This curve represents what is achieved in the classical Type-I HARQ systems. As expected, we observe in Figure 3.3 that the power allocation obtained by means of finite LP, outperforms the constant power allocation. Before drawing general conclusions, we give the same three simulation results for $R = 1$ bpcu, every other parameters remaining unchanged. With the result given in Figure 3.4, we can remark that the power allocation obtained by the finite LP of the preceding subsection is really efficient for low average SNR. To corroborate this conclusion, we have performed the following simulation: within the same framework as the two preceding simulations, we have computed the throughput for a range of R going from 0.75 bpcu to 5 bpcu. For every average SNR value we only kept the best throughput (among every initial rate R). These results are given in Figure 3.5 and confirm the intuition of the preceding simulation.

This new result leads us to the following conclusion: when dealing with HARQ protocols, choosing the appropriate initial rate R gives in general better

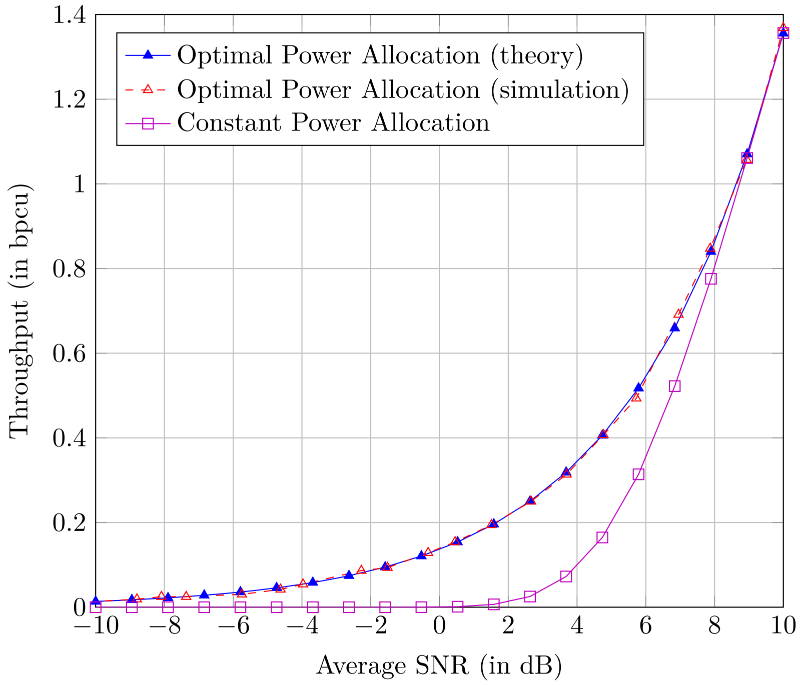

Figure 3.3 – *Comparison of the throughput of constant power allocation and variable power allocation as a function of Average SNR for a Type-I HARQ. Simulations parameters:* $N_T = 5$*,* $R = 3$ *bpcu,* $\bar{\alpha} = 1$*,* $P_{min} = -10dBW$*,* $P_{max} = 10dBW$*, and* $N_a = 128$*.*

results than allocating power. However, in most modern communication systems, the appropriate rate R may not be available among the different coded modulations. In this case, a power allocation is useful to compensate the absence of such coded modulation.

3.3 Power allocation for the throughput maximization of Type-II HARQ protocols

In this section, we propose to analyse the power allocation for the maximization of the throughput under average power and peak power constraints for Type-II HARQ based systems.

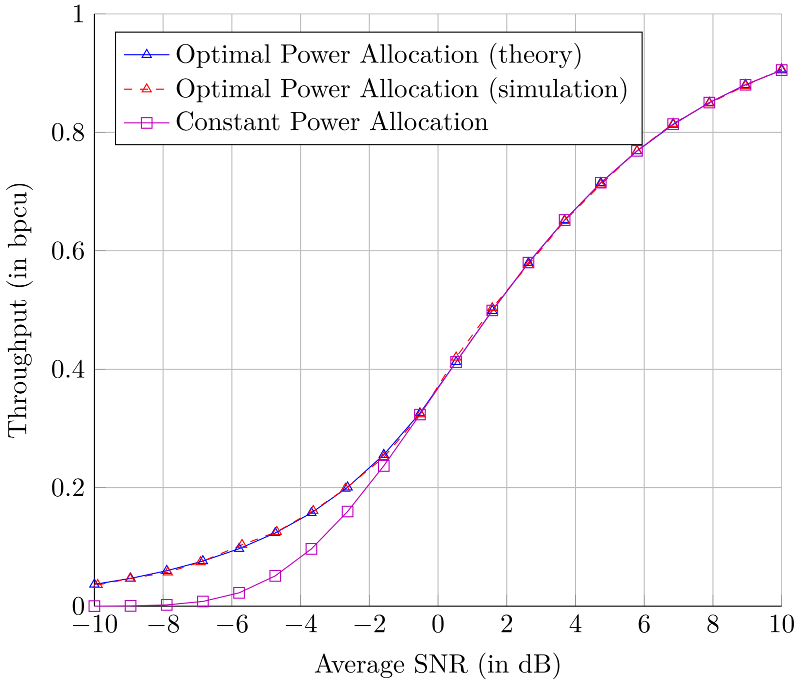

Figure 3.4 – *Comparison of the throughput of constant power allocation and variable power allocation as a function of Average SNR for a Type-I HARQ. Simulations parameters:* $N_T = 5$, $R = 1$ *bpcu,* $\bar{\alpha} = 1$, $P_{min} = -10dBW$, $P_{max} = 10dBW$, *and* $N_a = 128$.

3.3.1 CMDP model associated with the Power allocation for the Type-II HARQ protocols

Following the same reasoning as the one proposed for Type-I HARQ protocols in the preceding section, we first define every component of the CMDP associated with the power allocation problem for maximizing the throughput of the Type-II HARQ protocols. This CMDP has the following form: $(\mathcal{S}, \mathcal{A}, \mathcal{U}, \mathcal{W}, Q, r, \bar{p})$ where each component is described in the sequel. In Section 1.4.2, we have analysed the CC-HARQ protocol and the IR-HARQ under the same formalism. In consequence, we will present every theoretical result of this section under the same framework as in Section 1.4.2; we will specify our analysis only for the simulation results.

Definition of the CMDP

In Section 1.4.2, we have shown that a HARQ protocol can be efficiently modelled using a Markov chain on the space $\mathcal{S} = \{0, \dots, N_T\} \times [0, D_T]$ where N_T is the maximal number of decoding attempts and D_T is the decoding threshold.

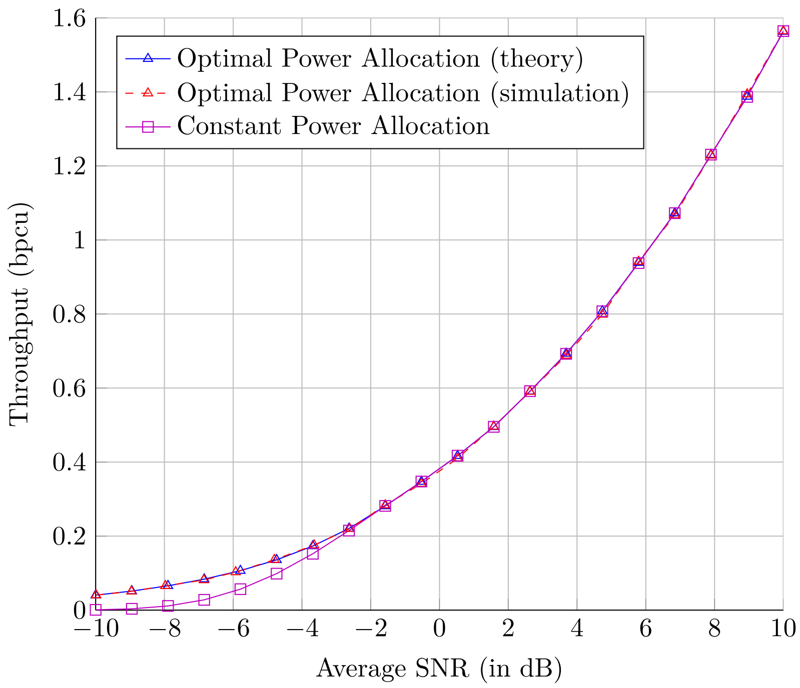

Figure 3.5 – *Comparison of the throughput of constant power allocation and variable power allocation as a function of Average SNR for a Type-I HARQ. Simulations parameters:* $N_T = 5$, $\bar{\alpha} = 1$, $P_{min} = -10dBW$, $P_{max} = 10dBW$, *and* $N_a = 128$. *The curves presented, are obtained by taking the maximum over values of* R *going from* 0.75 *bpcu to* 5 *bpcu*

The set $\mathcal{S} = \{0, \ldots, N_T\} \times [0, D_T]$ is the state space of the CMDP. We can already remark that $\mathcal{S}$ is compact, which verifies a part of Assumption 2.3(a).

Similarly to what we have presented in Section 3.2, the action space of the CMDP is taken as the set of available power levels for Tx, this set is again $\mathcal{A} = [P_{min}, P_{max}]$. Again, we will suppose that Tx can use the whole set of powers in each state so that $\mathcal{U} = \mathcal{S} \times \mathcal{A}$. In the sequel we again suppose that $P_{min} \geq \epsilon$ with $\epsilon > 0$ and that $P_{min} \leq P_{max}$. Directly from the definitions of $\mathcal{A}$ and $\mathcal{U}$, we can remark that the spaces $\mathcal{A}$ and $\mathcal{U}$ are compact, which proves that Assumption 2.3(a) is verified.

When the Type-II HARQ protocol is in state s_n and when action a_n is taken by Tx, we have seen in Section 1.4.2 that the state of the HARQ evolves according to rules defined as in Table 1.2. This table is given again in Table 3.1 for ease of presentation. The only difference between Table 1.2 and Table 3.1 is that in Table 3.1 we use the CMDP notation.

One can note that writing Table 3.1 is equivalent to writing that the random variables S_n evolve according to the following deterministic equation: $S_{n+1} =$

	$x_n + \Delta(\alpha_n, a_n) < D_T$	$x_n + \Delta(\alpha_n, a_n) \geq D_T$
$k_n < N_T - 1$	$(k_n+1, x_n+\Delta(\alpha_n, a_n))$	$(0, 0)$
$k_n = N_T - 1$	$(N_T, 0)$	$(0, 0)$
$k_n = N_T$	$(1, \Delta(\alpha_n, a_n))$	$(0, 0)$

Table 3.1 – *Table of rules for the transitions from the state $s_n = (k_n, x_n)$ to s_{n+1} when power Tx takes action a_n.*

$F(S_n, A_n, \alpha_n)$. In consequence, the GNR α_n plays again the role of disturbance. The disturbance space $\mathcal{W}$ is then defined as the range of the GNR: $[0, \infty)$. The transition kernel Q is again defined in (A.3). The proof that Q verifies 2.3(d) for the Rayleigh channel is given in Appendix C.2.

The short-term reward associated with the computation of the throughput as well as the short-term cost associated with the computation of the average power are still unchanged compared to the case of Type-I HARQ. The short-term reward is defined as:

$$r(s) = \begin{cases} b \text{ if } s = (0, 0) \\ 0 \text{ otherwise.} \end{cases}$$

The short-term cost is defined as:

$$\bar{p}(s, a) = a. \tag{3.10}$$

The reward and the cost functions are again trivially bounded and Lipschitz continuous functions on $\mathcal{U}$.

We now give the long term reward and cost and the associated optimization problem. Without loss of generality, we assume that the initial state is $s_0 = (0, 0)$. The throughput of the policy π when the initial state is s_0 is defined as

$$\eta(s_0, \pi) = \liminf_{T \to \infty} \frac{1}{T} \mathbb{E}_{\pi}^{s_0} \left[\sum_{t=0}^{T-1} r(S_t, A_t) \right]. \tag{3.11}$$

Similarly to the definition of the throughput, we introduce the average power as

$$\bar{P}(s_0, \pi) = \liminf_{n \to \infty} \frac{1}{n} \mathbb{E}_{\pi}^{s_0} \left[\sum_{i=0}^{n-1} \bar{p}(S_t, A_t) \right].$$

Finally, the power allocation problem is written as

$$\begin{aligned} \eta^\star(s_0) = \sup_{\pi\in\Pi} & \quad \eta(s_0,\pi) \\ s.t. & \quad \bar{P}(s_0,\pi) \le P_A, \end{aligned} \tag{3.12}$$

where P_A is the average power constraint. Note that, here also, the peak power constraint does not appear explicitly.

Properties of the proposed CMDP model

We now analyse this CMDP model to prove that Assumption 2.3 is verified. This assumption will, in turn, allow us to build a finite LP that solves the power allocation problem.

Along with the description of the CMDP problem associated with the Type-II HARQ power allocation, we have shown that Assumption (2.3(a)), Assumption 2.3(b), Assumption 2.3(c) holds. The verification of Assumption2.3(d) is done in Appendix C.2.

It finally remains to verify Assumption 2.3(e). We again prove this Assumption by showing that Condition 2.1(a) holds. The accessible state needed by Condition 2.1(a) is $s_0 = (0,0)$, indeed, for every $s \in \mathcal{S}$ we have

$$\begin{aligned} Q(s_0|s,a) &= \mathbb{P}\left[x + \Delta\left(a\alpha\right) \ge D_T\right] \\ &\ge \mathbb{P}\left[x + \Delta\left(P_{min}\alpha\right) \ge D_T\right] \\ &\ge \mathbb{P}\left[\Delta\left(P_{min}\alpha\right) \ge D_T\right] \\ &> 0, \end{aligned} \tag{3.13}$$

where the second inequality holds because Δ is an increasing function of a, the third inequality holds because x is positive and the last inequality holds because $P_{min} > 0$.

3.3.2 The finite linear programming

In the preceding subsection, we have shown that the CMDP corresponding to the power allocation problem for Type-II HARQ protocol verifies Assumption 2.3. This again proves that the optimization problem given by (3.12) is solvable at the condition that $P_A \ge P_{min}$. Under Assumption 2.3, Theorems 2.1 and 2.2 imply that the optimization problem given by equation (3.12) is equivalent to an infinite dimensional LP. We now build finite LP approximating the infinite LP (see Theorem 2.2). To build the finite LP, we introduce the following finite CMDP: $(\tilde{S}, \tilde{A}, \tilde{U}, \tilde{Q}, \tilde{r}, \tilde{p})$. For this finite CMDP, we drop the index h since it is more convenient in practice to consider two different grids, one of size h for $\mathcal{S}$, the other of size ℓ for $\mathcal{A}$. Theorem 2.5 remains unchanged except that we consider that $h' = \min(h, \ell) \to 0$ instead of $h \to 0$. We now describe each component of $(\tilde{S}, \tilde{A}, \tilde{U}, \tilde{Q}, \tilde{r}, \tilde{p})$.

• The partition of $\mathcal{S}$, denoted as S_h, is built from the following partition of the set $[0, D_T]$:

$$[0, D_T] = \bigcup_{j=0}^{N_I-2} [jh,\ (j+1)h[\ \cup\ [(N_I-1)h, D_T]\,, \tag{3.14}$$

where N_I represents the number of sets in the partition of $[0, D_T]$ and $h = D_T/N_I$ defines the length of each interval. S_h is composed by the sets $S_{k,j}$ of the form

$$S_{k,j} = \{k\} \times [jh, (j+1)h[\,. \tag{3.15}$$

The set $\tilde{S}$ is the set of every representatives of the sets $S_{k,j}$. We suppose in this section that the elements of $\tilde{S}$ have the following form:

$$\tilde{s}_{k,j} = (k, (j+\omega)h),\ \omega \in [0, 1[\,. \tag{3.16}$$

• The discretized action space $\tilde{A}$ is defined, as proposed in Section 2.5, as

$$\tilde{A} = \{P_{min}, P_{min} + \ell, P_{min} + 2\ell \cdots, P_{min} + (N_a - 1)\ell = P_{max}\}\,,$$

where P_{min} and P_{max} are the peak powers, N_a is the number of actions considered in the discrete set, and ℓ is defined as

$$\ell = \frac{P_{max} - P_{min}}{N_a - 1}. \tag{3.17}$$

• For every $\tilde{s}_{k',j'}$, $\tilde{s}_{k,j}$, and $\tilde{a}_l$, the transition law is discretized as follows:

$$\tilde{Q}\left(\tilde{s}_{k',j'} | \tilde{s}_{k,j}, \tilde{a}_l\right) = Q\left(\{k'\} \times [j'h, (j'+1)h[\, | \tilde{s}_{k,j}, \tilde{a}_l\right), \tag{3.18}$$

where Q is given in Appendix A.1. In the sequel of this section, $\tilde{Q}$ is interpreted as a collection of matrices parametrized by $\tilde{a} \in \tilde{A}$. For all $\tilde{s}, \tilde{s}' \in \tilde{S}$ and for all $\tilde{a} \in \tilde{A}$ these matrices are expressed as follows

$$\tilde{Q}_{\tilde{a}}(\tilde{s}, \tilde{s}') = \tilde{Q}(\tilde{s}' | \tilde{s}, \tilde{a}). \tag{3.19}$$

With these notations, the finite LP is finally given by:

$$\begin{aligned}
\tilde{\eta}^{\star} = & \sup_{\mathbf{m}\in\mathbb{R}^{N_U}} \mathbf{R}\mathbf{m}^T \\
\text{s.t.} & \sum_{\tilde{s}\in\tilde{S}} m(\tilde{s},\tilde{a}) - Q_{\tilde{a}}\mathbf{m}^T = \mathbf{0},\ \forall \tilde{a}\in\tilde{A}_h \\
& \mathbf{1}\mathbf{m}^T = 1 \\
& \bar{\mathbf{P}}\mathbf{m}^T \leq P_A \\
& \mathbf{m} \geq \mathbf{0},
\end{aligned}$$

where $\mathbf{R}$, $\bar{\mathbf{P}}$ and $\mathbf{1}$ are the vectors of size N_U containing respectively $r(\tilde{u})$ for all $\tilde{u}\in\tilde{U}$, $\bar{p}(\tilde{u})$ for all $\tilde{u}\in\tilde{U}$ and only ones. Once the optimal $\mathbf{m}^{\star}$ is found, the corresponding policy consists in choosing randomly a power in $\tilde{A}$ with a distribution:

$$\varphi(\tilde{a}|s) = \frac{\mathbf{m}^{\star}(\tilde{s}_{k,j},\tilde{a})}{\sum_{\tilde{s}\in\tilde{S}} \mathbf{m}^{\star}(\tilde{s},\tilde{a})},\ \forall s\in S_{k,j}. \tag{3.20}$$

3.3.3 Numerical results

For the Type-II HARQ protocol, we also consider a HARQ protocol with a maximum number of transmission attempts $N_T = 5$ and an initial rate $R = 7$ bpcu. In these results, we consider without loss of generality, that the GNR α is exponentially distributed with average $\bar{\alpha} = 1$. When a power allocation is performed, we consider a partition of $\mathcal{S} = [0, R]$ into $N_I = 16$ parts of equal sizes, we consider a set $\tilde{A}$ from $P_{min} = -10dBW$ to $P_{max} = 10dBW$ with $N_a = 16$. In Figure 3.6 we compare three results. These three results are all compared at a given average SNR. Note that since on each slot a_n is independent of α_n, the average SNR can be computed as $P_A\bar{\alpha} = P_A$.

- The first result is depicted in Figure 3.6 as the blue curve with triangle marker. This curve corresponds to the result of the finite LP: $\tilde{\eta}^{\star}$ as a function of P_A.
- The second result is depicted in Figure 3.6 as the red curve with triangle marker. This curve corresponds to the result of the simulation of the IR-HARQ protocol using the power allocation $\varphi(\tilde{a}|s)$ computed thanks to the finite LP. These simulations has been realized by the Monte-Carlo method over 10^4 slots.
- The third result is depicted in Figure 3.6 as the violet curve with square marker. This curve corresponds to the IR-HARQ scheme with constant power allocation.

For the CC-HARQ protocol, we have done the same experiment except that $R = 3bpcu$. This result is given in Figure 3.7. For both the IR-HARQ and

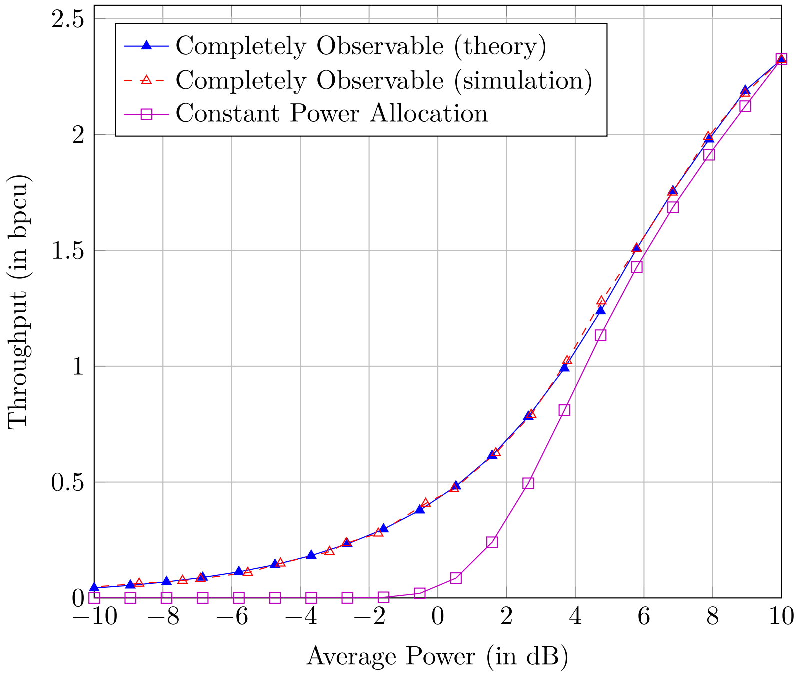

Figure 3.6 – *Comparison of the throughput of constant power allocation and variable power allocation as a function of Average SNR for an IR-HARQ protocol. Simulations parameters:* $N_T = 5$*,* $R = 7$ *bpcu,* $\bar{\alpha} = 1$*,* $P_{min} = -10dBW$*,* $P_{max} = 10dBW$*,* $N_I = 16$*, and* $N_a = 64$*.*

CC-HARQ cases, we can observe the same conclusion as the one for Type-I HARQ protocol: the proposed power allocation works well for medium to low SNR. In order to observe the impact of R, we present in Figures 3.8 and 3.9 the result obtained by maximizing over initial rates from $R = 0.1$ bpcu and $R = 10$ bpcu.

From these simulation results, we draw the exact same conclusion as for the Type-I HARQ: choosing the appropriate initial rate R gives in general better results than allocating power.

3.3.4 The Partially Observable problem

In this subsection, we consider a system based on the Type-II HARQ protocol. The only difference with the preceding section is that we now consider that the CQI only conveys the classical 1-bit feedback: ACK/NACK. This model is depicted in Figure 3.10 and is the one described in Chapter 1 except that we consider that the ACK/NACK bits are also used for power allocation.

Let $S_n = (K_n, I_n) \in \mathcal{S}$ be the random variable representing the state of the Type-II HARQ protocol at time nT. This state is defined in the preceding

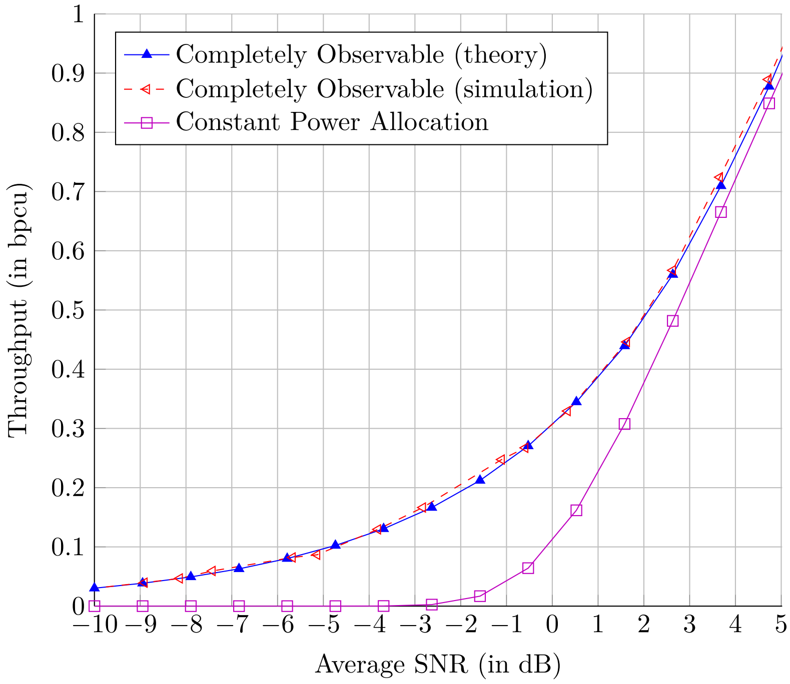

Figure 3.7 – *Comparison of the throughput of constant power allocation and variable power allocation as a function of Average SNR for a CC-HARQ protocol. Simulations parameters:* $N_T = 5$, $R = 3$ *bpcu,* $\bar{\alpha} = 1$, $P_{min} = -10dBW$, $P_{max} = 10dBW$, $N_I = 16$, *and* $N_a = 128$.

section. From the ACK/NACK bits, Tx can only infer K_n indeed, from the definition of K_n given in Section 1.4.1, one can observe that it is sufficient to "count" the NACK bits and to observe the ACK bits. On the other hand, the random variable I_n is non-observable from Tx by using only the ACK/NACK bits. Based only on this observation, we want to find the power allocation that maximizes the throughput of the Type-II HARQ protocol under the peak and average power constraints. This problem is the following:

$$\eta^\star(s_0) = \sup_{\pi \in \Pi_o} \quad \eta(s_0, \pi)$$
$$s.t. \quad \bar{P}(s_0, \pi) \leq P_A,$$

where Π_o is only the observable history (the history containing only the present and past values of K_n and the past values of A_n (the actions).

Since the corresponding completely observable problem has been shown to be a CMDP (see preceding subsection), it is logical to consider this problem as a PSI-MDP. This PSI-MDP is the following $(\mathcal{X}, \mathcal{Y}, \mathcal{A}, Q, r, \mathbf{c})$ where each component is defined as follows:

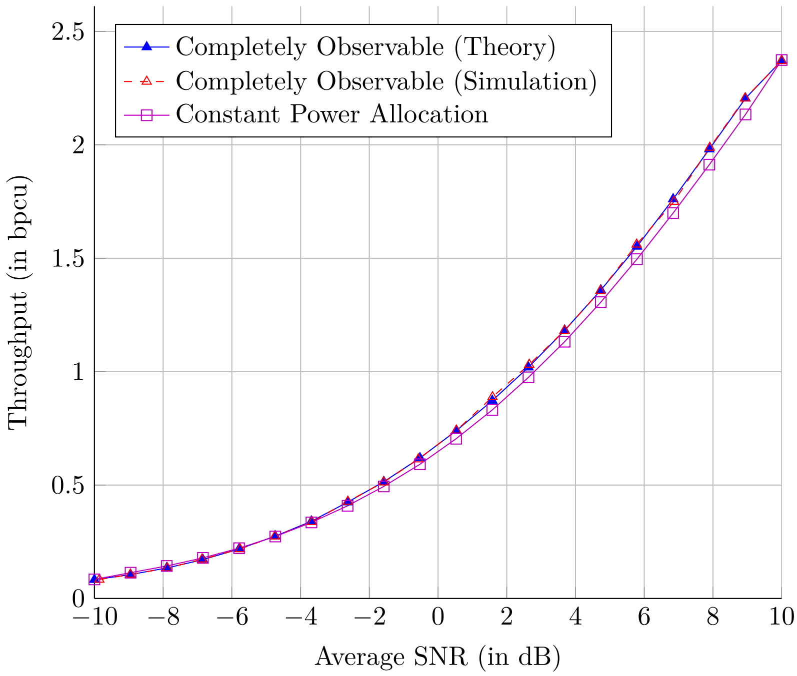

Figure 3.8 – *Comparison of the throughput of constant power allocation and variable power allocation as a function of Average SNR for an IR-HARQ protocol. Simulations parameters:* $N_T = 5$, $\bar{\alpha} = 1$, $P_{min} = -10dBW$, $P_{max} = 10dBW$, $N_I = 16$, *and* $N_a = 64$. *The curves presented, are obtained by taking the maximum over values of* R *going from* 0.5 *bpcu to* 10 *bpcu*

- $\mathcal{X} = [0, 1]$ is the non observable space. In our case, we define $\mathcal{X}$ as set of the normalized (by D_T) values of I_n: $X_n = I_n/D_T$,
- $\mathcal{Y} = \{0, 1, \ldots, N_T\}$ is the observable space. In this case, $Y_n = K_n$,
- $\mathcal{A} = [P_{min}, P_{max}]$ is the action space defined in the preceding subsection,
- Q is the transition law defined in the preceding subsection,
- r is the instantaneous reward for the throughput η defined in the preceding subsection,
- $\bar{p}$ is the instantaneous cost associated with average power $\bar{P}$ defined in the preceding subsection.

The observable history is defined as in Section 2.6.1,

$$h_{o,t} = (s_0, a_0, y_1, a_1, y_2, a_2 \ldots, y_{t-1}, a_{t-1}, y_t) .$$

The set of all the histories of size t is denoted by $\mathcal{H}_{o,t}$. Note that every history in $\mathcal{H}_{o,t}$ only considers the ACK/NACK through the y_j, $j \in [1, t]$.

Following the method presented in Section 2.6.2, we know that we can con-

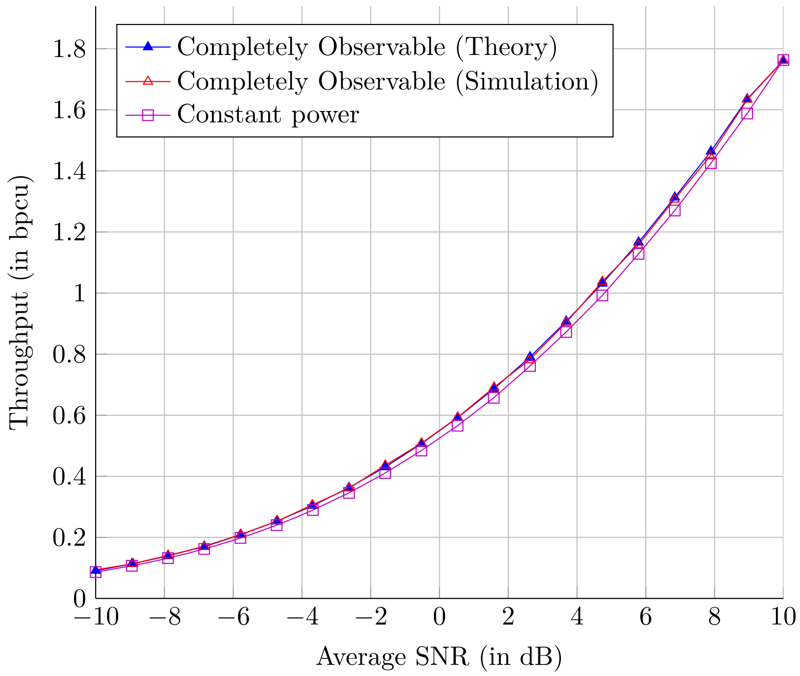

Figure 3.9 – *Comparison of the throughput of constant power allocation and variable power allocation as a function of Average SNR for an CC-HARQ protocol. Simulations parameters:* $N_T = 5$, $\bar{\alpha} = 1$, $P_{min} = -10dBW$, $P_{max} = 10dBW$, $N_I = 16$, *and* $N_a = 128$. *The curves presented, are obtained by taking the maximum over values of* R *going from* 0.1 *bpcu to* 5 *bpcu*

vert this PSI-MDP into a CMDP $(\mathcal{Z}, \mathcal{A}, Q_Z, r', \bar{p}')$ where each component is defined as follows.

- $\mathcal{Z} = \mathcal{P}(\mathcal{X}) \times \mathcal{Y}$ where $\mathcal{P}(\mathcal{X})$ is the set of every probability measures on $\mathcal{X}$ and $\mathcal{Y} = \{0, 1, \ldots, N_T\}$.
- $\mathcal{A}$ is the action space and is unchanged compared to the above definition.
- Q_Z is the transition matrix for Z_t, defined by equation (2.95).
- r' is defined as

$$\begin{cases} r'(z,a) = \int_{\mathcal{X}} r(x,y,a)b(dx) \\ \bar{p}'(z,a) = \int_{\mathcal{X}} \bar{p}(x,y,a)b(dx). \end{cases} \tag{3.21}$$

From the definitions of r and $\bar{p}$, we get:

$$r'(z,a) = \begin{cases} b \text{ if } y = 0 \\ 0 \text{ otherwise} \end{cases} \tag{3.22}$$

and

$$\bar{p}'(z,a) = \bar{p}(s,a) = a. \tag{3.23}$$

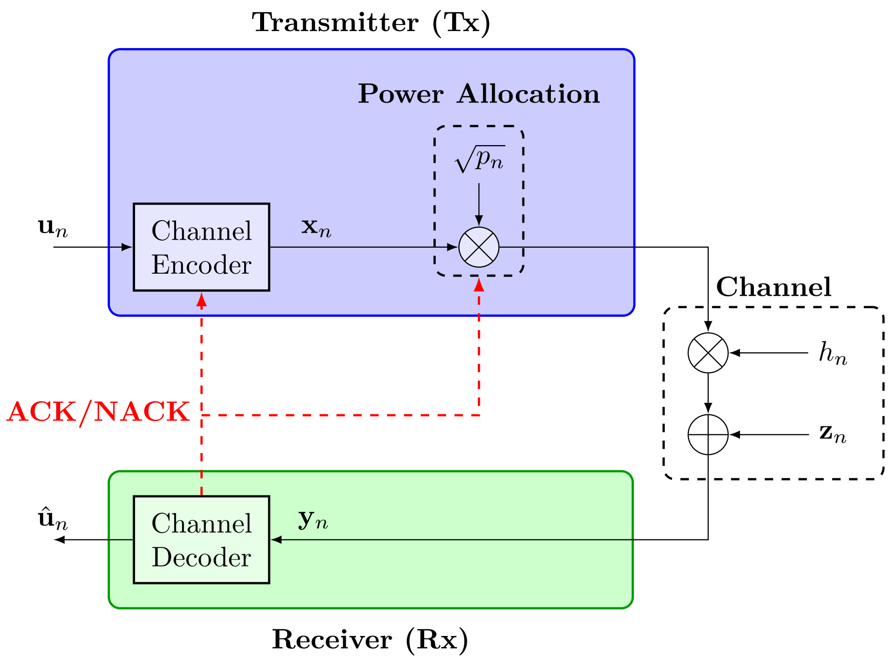

Figure 3.10 – *System model and block fading channel*

We now prove that this CMDP is solvable. In fact, from the preceding subsection, we have already proved Condition 2.2 except Condition 2.2(d) and Condition 2.2(e). To check 2.2(d) we observe that $\mathcal{Y}$ is finite and we use Remark 2.1. To check Condition 2.2(e), we observe that the state $(0, 0)$ corresponds to observing ACK and in consequence is completely observable. Moreover, we have shown in equation (3.13) that this state is accessible from every other state of $\mathcal{S}$. This allows us to use Remark 2.2, to verify that Condition 2.2(e) is checked.

So far, we have converted the initial PSI-MDP problem into a CMDP problem and we have proved that this new CMDP is solvable. To solve numerically this PSI-MDP, we have used a method that is similar to the method presented in Section 2.6.4. We have projected the beliefs (the elements of $\mathcal{P}(\mathcal{X})$) on the parametrized set of Beta(θ_1, θ_2) laws. The Beta laws belong to the exponential set of densities. For every $\theta_1 > 0$, $\theta_2 > 0$, if X is a random variable with distribution Beta(θ_1, θ_2), its pdf $f(x; \theta_1, \theta_2)$ is defined as follows:

$$f(x; \theta_1, \theta_2) = \frac{x^{\theta_1 - 1}(1 - x)^{\theta_2 - 1}}{\displaystyle\int_0^1 x^{\theta_1 - 1}(1 - x)^{\theta_2 - 1} dx}, x \in [0, 1]\,. \tag{3.24}$$

The choice for Beta laws is motivated by the two following facts: i) the set $\mathcal{X} = [0, 1]$, ii) using the method of moments (see [Kay, 1993]) instead of the

Kullback-Liebler divergence (used by [Zhou et al., 2010]), allows us to have a compact parametrization of the Beta laws. Indeed, the method of moments is based on the fact that we can compute the parameters (θ_1, θ_2) of a Beta distribution from the mean $\bar{m}$ and variance $\bar{v}$ of X by the following transformation

$$\begin{cases} \theta_1 = \bar{m}\left(\frac{\bar{m}(1-\bar{m})}{\bar{v}} - 1\right) \\ \theta_2 = (1-\bar{m})\left(\frac{\bar{m}(1-\bar{m})}{\bar{v}} - 1\right). \end{cases} \tag{3.25}$$

Furthermore, if X has a distribution $\text{Beta}(\theta_1, \theta_2)$, its mean $\bar{m}$ and variance $\bar{v}$ verify that:

$$(\bar{m}, \bar{v}) \in \Theta_{mv} = [0,1] \times [0, 0.25].$$

So instead of parametrizing the set of Beta laws by (θ_1, θ_2) as in Section 2.6.4, we will use the same set with $(\bar{m}, \bar{v})$. We have finally solved the CMDP problem $(\Theta_{mv} \times \mathcal{Y}, \mathcal{A}, Q'_Z, r', \hat{p}')$ using the method proposed in Section 2.5.

The result of this method is presented in Figure 3.11. These results has been computed by discretizing Θ_{mv} with a uniform grid of 100 possible values for $\bar{m}$, 100 possible values for $\bar{v}$, and only $N_a = 16$ possible powers for Tx. The results computed by the projected-CMDP are also compared with simulation results obtained by simulating the policy found by the method of Section 2.5 applied to the projected-CMDP. Similar results for the case of the CC-HARQ protocol are presented in Figure 3.12.

In both cases, we observe that the policies found by solving the projected-CMDP perform well; these policies outperform the results obtained by the classical Type-II HARQ methods and are close to the performance obtained with complete observability. The main advantage of these methods is that they require only 1-bit of feedback. This brings us to the following conclusion, for the case where the initial rate R cannot be optimized, we can find power allocation policies that perform close to the completely observable CMDP(that required multiple bits CQI) without changing anything to the HARQ protocol.

3.4 Conclusion

In this section, we solved the problem, proposed in Chapter 1, of finding the power allocation that maximizes the throughput of an HARQ protocol under peak and average power constraints. We have studied this power allocation problem for the Type-I HARQ protocol, the IR-HARQ protocol, and the CC-HARQ protocol. The approach proposed in this chapter consists in applying the CMDP framework presented in Chapter 2 to the controlled Markov chain representa-

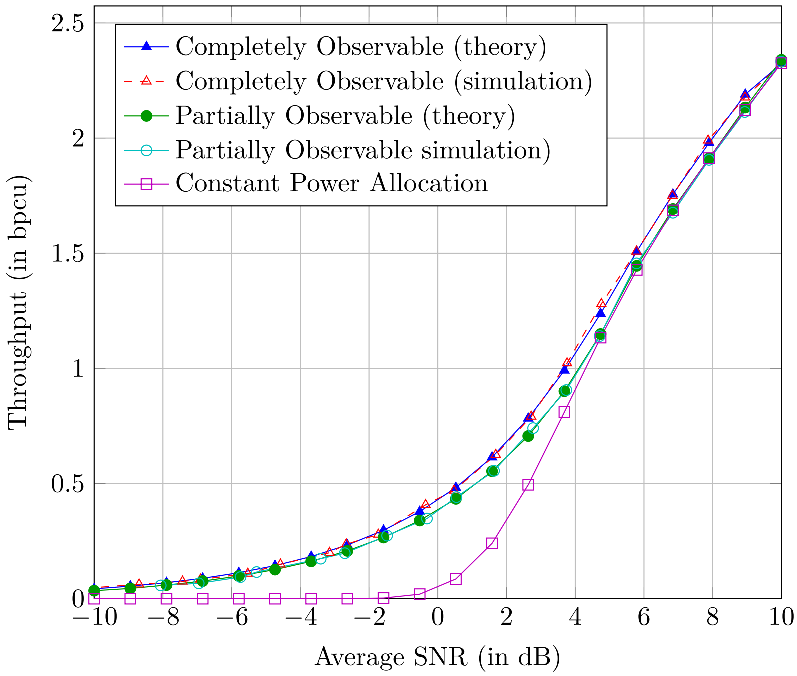

Figure 3.11 – *Comparison of the results presented in Figure 3.6 and the result obtained by solving the projected-CMDP for solving the PSI-MDP in the IR-HARQ case. The following simulations parameters has been considered:* $N_T = 5$, $R = 7$ *bpcu,* $\bar{\alpha} = 1$, $P_{min} = -10dB$, $P_{max} = 10dB$, $N_a = 16$, *and* Θ_{mv} *discretized with a uniform grid of* 100 *values for* $\bar{m}$ *and* 100 *values for* $\bar{v}$.

tion of the HARQ protocols proposed in Chapter 1. In any case, the policies found thanks to the discretization procedure presented in Section 2.5 have been compared with their simulations counterparts. From these simulations, we have shown that when the initial rate R can be optimized, the gain of performing power allocation is limited. This results tend to corroborate the affirmation done in [Szczecinski et al., 2011].

For the two Type-II HARQ protocols, we have considered two different situations: i) the state of the HARQ protocol is perfectly known by Tx or ii) only the number of attempts is known at Tx. The first case has been solved by applying the discretization procedure described in Section 2.5 for the corresponding CMDP. The second case is proposed as a PSI-MDP of the case i). We have provided simulation results showing that the degradation caused by the absence of multiple bits feedback is low.

In [Szczecinski et al., 2011], an adaptive IR-HARQ protocol is proposed. Their protocol adapts the size of the retransmissions according to some unquantized feedback. In [Szczecinski et al., 2011], a constrained dynamic programming approach is proposed to solve the adaptive IR-HARQ. In fact their method can

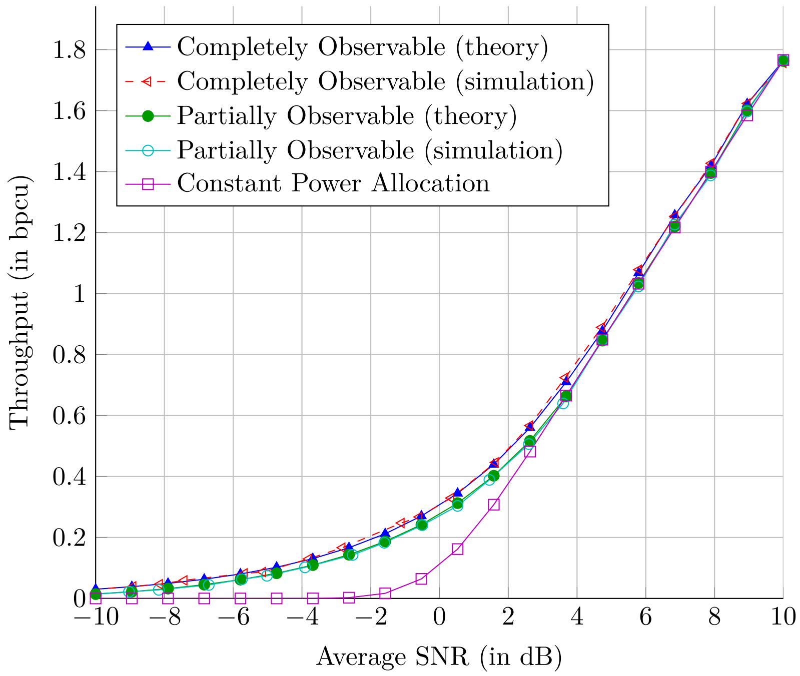

Figure 3.12 – *Comparison of the results presented in Figure 3.7 and the result obtained by solving the projected-CMDP for solving the PSI-MDP in the CC-HARQ case. The following simulations parameters has been considered:* $N_T = 5$, $R = 3$ *bpcu,* $\bar{\alpha} = 1$, $P_{min} = -10dB$, $P_{max} = 10dB$, $N_a = 16$, *and* Θ_{mv} *discretized with a uniform grid of* 100 *values for* $\bar{m}$ *and* 100 *values for* $\bar{v}$.

be viewed as a fractional programming between two finite horizon MDPs. With this approach, we could compute their results with $N_T = 5$. These results are given in Figure 3.13 and are compared with the power allocation results proposed in this chapter. The simulation results proposed in Figure 3.13 tend to show that rate adaptation is better suited for improving the performance of the IR-HARQ protocol. On the other hand to be really efficient, the adaptive IR-HARQ proposed in [Szczecinski et al., 2011] requires a large variety of coders and a theoretically unquantized feedback.

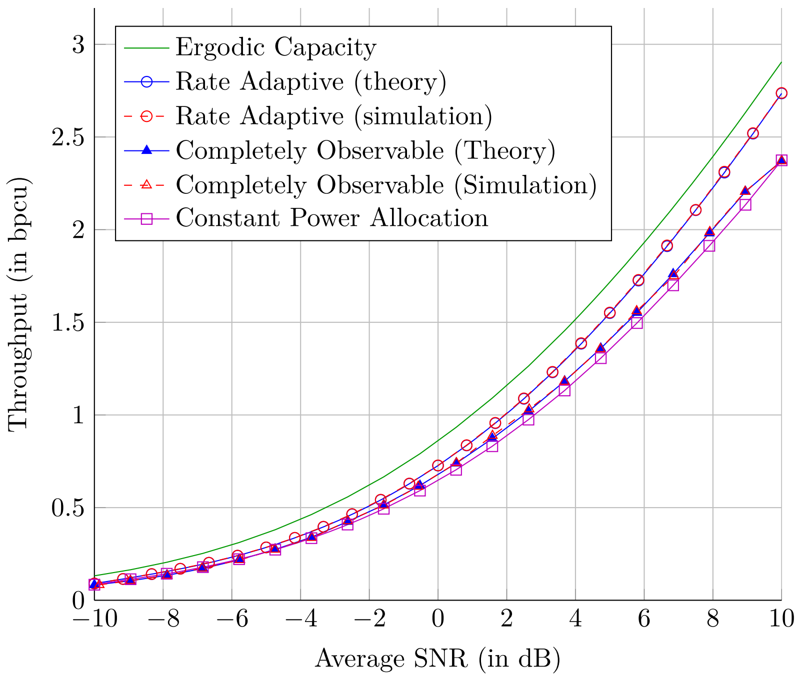

Figure 3.13 – *Comparison of the results presented in Figure 3.8 with the adaptive IR-HARQ protocol proposed in [Szczecinski et al., 2011]*

- Chapter 4 -

Resource Allocation for Cognitive Radio exploiting the HARQ protocol of the primary users

Contents

In this chapter, we show that the CMDP framework is appropriate for studying the power allocation of a secondary user that exploits the HARQ rounds of a primary user to manage its interference.

4.1 Introduction

Recently, the number of wireless services has dramatically increased generating a growth in the demand on radio spectrum. Additionally, because of the command-and-control regulation, every service is allocated to a dedicated bandwidth. In consequence, the radio spectrum has become a scarce resource. On the other hand, It has been shown that the radio spectrum is underutilized (see [FCC, 2002]). By allowing Secondary Users (SUs) to opportunistically access the bandwidth dedicated to licensed users (also referred to as Primary Users

(PUs)), *cognitive radio* has been shown to be a promising technique to improve the efficiency of wireless networks (see e.g. [Mitola III & Maguire, 1999], [Haykin, 2005], and references therein).

Historically, cognitive radio considered an Opportunistic Spectrum Access (OSA) framework. In this framework, the SUs probe the PUs bandwidth to detect *white-spaces*. The SUs use agile radios such as Software Defined Radio (SDR) to dynamically move from a white-space to another white-space targeting a *zero-interference* policy. This zero-interference constrain is relaxed in the *spectrum sharing* framework (see [Etkin et al., 2007], [Jovicic & Viswanath, 2009] and references therein) where the SUs can interfere the PUs as long as the degradation of the PUs performance remains within some Quality of Service (QoS) constraints. Spectrum sharing has been the subject of extensive research in recent years (see [Asghari & Aissa, 2010], [Bagayoko et al., 2010, 2011], [Etkin et al., 2007], [Jovicic & Viswanath, 2009], [Kang et al., 2009], [Makki & Eriksson, 2012; Makki et al., 2012, 2013], [Masmoudi et al., 2012], [Tajan et al., 2012],[Tannious & Nosratinia, 2010], [Zhang, 2008, 2010], and references therein).

In this chapter we are interested in a spectrum sharing context where the SU shares the channel with one PU implementing a HARQ protocol. This context is motivated by the fact that from an information theoretical perspective, it has been shown in [Eswaran et al., 2007] that, by listening to the PUs feedback bits, the SU can infer the throughput-loss of the PU. Furthermore using this information about the throughput-loss, the authors have shown that a non-negligible throughput can be achieved by the SU. This result is important because it means that, in order to be compliant with the SU, the PU only has to broadcast the feedback bits of its HARQ protocol.

Within the proposed context, the case of a PU implementing a Type-I HARQ protocol has been addressed in [Levorato et al., 2012]. In [Levorato et al., 2012], a CMDP approach is used to derive an optimal ON/OFF allocation while managing the throughput-loss of PU. In this chapter we propose to address the Type-II HARQ problem.

This chapter is organised as follows. In Section 4.2 we present the channel model, the protocol of the PU, the protocol of the SUand we finally propose an optimization problem. In Section 4.3 we show that the optimization problem proposed in Section 4.2 is equivalent to a CMDP, we prove the existence of a solution for this CMDP. In Section 4.3, we assume that the SU has access to the complete state of the PU HARQ protocol and uses this state to maximize its own throughput while managing the QoS loss of the PU. In Section 4.4, we propose to solve the problem in which the SU has only access to the PU feedback.

4.2 System model

In this section, we give more details on the system presented in Figure 4.1. We present the HARQ protocol that the PU is using in order to retrieve erroneous packets. We also present the power and rate allocation of the SU. The protocols of the PU and the SU are studied in terms of *throughput*.

4.2.1 Channel Model

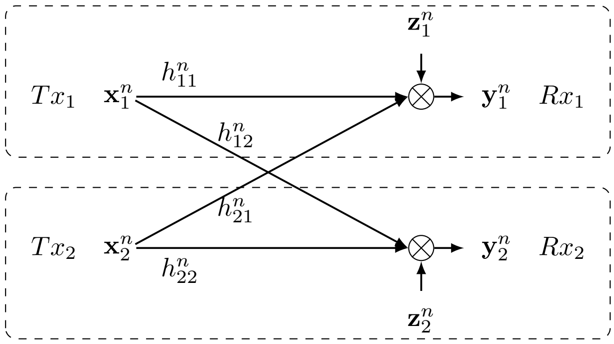

Figure 4.1 – *The Gaussian interference channel at the n^{th} slot.*

Similarly to [Jovicic & Viswanath, 2009], we consider the network illustrated in Figure 4.1. This network is composed of a PU (with transmitter Tx_1 and receiver Rx_1) and a SU (with transmitter Tx_2 and receiver Rx_2). The *PU* and the *SU* share a block-fading interference channel. In this context, the signals $\mathbf{y}_{1,n}$ and $\mathbf{y}_{2,n}$ respectively received in slot n by Rx_1 and Rx_2 are expressed as

$$\begin{cases} \mathbf{y}_1^n = h_{11,n}\mathbf{x}_{1,n} + h_{21,n}\mathbf{x}_{2,n} + \mathbf{z}_{1,n} \\ \mathbf{y}_2^n = h_{22,n}\mathbf{x}_{2,n} + h_{12,n}\mathbf{x}_{1,n} + \mathbf{z}_{2,n}, \end{cases} \tag{4.1}$$

where the slot is assumed to last L symbols (as in Chapter 1), $\mathbf{z}_{i,n} \in \mathbb{C}^L$ is a random vector of size L representing additive white Gaussian noise. The noise is a complex Gaussian random vector of size L. We consider, without loss of generality, that the vectors $\mathbf{z}_{i,n}$ have i.i.d. components of zero mean and unit variance. The input signals $\mathbf{x}_{1,n}$ and $\mathbf{x}_{2,n}$ are of size L and are assumed to be complex circular white Gaussian random vectors of zero mean and respective transmit powers $p_{1,n}$ and $p_{2,n}$.

The random variables defined as

$$\alpha_{ij,n} = |h_{ij,n}|^2 \tag{4.2}$$

are exponentially distributed with mean $\bar{\alpha}_{ij}$. The instantaneous Signal to Interference plus Noise Ratio (SINR) at Rx_i in slot n is denoted by $\beta_{i,n}$ and is defined as:

$$\beta_{,n} = \frac{p_{i,n}\alpha_{ii,n}}{1 + p_{j,n}\alpha_{ji,n}}. \tag{4.3}$$

4.2.2 The PU model

In this chapter, we consider that the PU implements an IR-HARQ protocol, with a maximum of $N_{T,1}$ attempts, with a constant power p_1, and with an initial rate of r_1 (see Section 1.3.3). Additionally to the classical assumptions on HARQ protocols given by Assumption 1.1, we suppose that the PU is compliant for the SU. In our situation, it means that the PU is aware of the possible presence of the SU and sends its feedback bits over a common broadcast channel so that the SU can hear it. On the other hand the PU does not manage the SU and consequently is oblivious to the presence (or absence) of the SU. For the sake of clarity, all assumptions that we have made about the PU are listed in Assumption 4.1.

Assumption 4.1.
(a) *the PU uses an IR-HARQ protocol,*
(b) *the parameters p_1, r_1 and $N_{T,1}$ are constant over time,*
(c) *the feedback channel is instantaneous and error-free,*
(d) *Tx_1 always has an information packet to transmit,*
(e) *the PU sends its feedback bits over a common broadcast channel.*
(f) *the PU is oblivious to the presence (or absence) of the SU, in particular Rx_1 processes the SU signal as noise.*

As it has been presented in Section 1.4.2, the state of the IR-HARQ protocol of the PU, at time nT, can be represented by two variables $s_n = (k_{1,n}, i_{1,n})$. The component $k_{1,n}$ can take one of the following $N_{T,1} + 1$ values

- **0:** Tx_1 starts the transmission of a new information packet after a successful decoding (ACK),
- **1:** Tx_1 has done 1 attempt and a NACK bit is received,
- **2:** Tx_1 has done 2 attempts and a NACK bit is received,

⋮

- $N_{T,1}$**:** Tx_1 has done $N_{T,1}$ attempts and NACK is received. This state corresponds to an outage event and, in consequence, to the start of the transmission of a new information packet.

The state component $i_{1,n}$ represents the ACMI at Rx_1. We only consider the throughput η_1 as main figure of merit for the performance of the PU. This throughput will be denoted as η_1 and is defined as

$$\eta_1(\pi_2) = \liminf_{t\to\infty} \frac{1}{t}\mathbb{E}_{\pi_2}\left[\sum_{n=0}^{t-1} R_1(s_n, p_{2,n})\right], \tag{4.4}$$

where $R_1(s_n)$ is defined as follows

$$R_1(s_n) = \begin{cases} r_1, & \text{if } s_n = (0,0) \\ 0, & \text{otherwise} \end{cases} \tag{4.5}$$

and where π_2 is a power allocation of the SU defined in its more general form as a random process $\pi_2 = \{p_{2,n}\}$. π_2 is written in equation (4.4) to stress the dependence of η_1 to the power allocation π_2 of the SU. This dependence, happens because of the Signal to Interference plus Noise Ratio (SINR), $\beta_{2,n}$ given in equation (4.3).

4.2.3 The SU model

We now describe the model of the SU, starting by describing the CSI assumptions on the SU. We suppose that the SU has partial CSIR: in this case it means that Rx_2 knows $h_{12,n}$ and $h_{22,n}$. We will also suppose SCSI at Tx_2 and Rx_2. In the Rayleigh channel case that is of interest here, SCSI at Tx_2 and Rx_2 means that both Tx_2 and Rx_2 know every $\overline{\alpha}_{ij}$. No CSIT is considered at Tx_2. For the rest of this chapter, we will make the following assumptions on the SUs system.

Assumptions 4.2.
(a) *Rx_2 processes the PU signal as noise.*
(b) *Tx_2 has an information packet to transmit.*
(c) *Tx_2 has no CSIT but has SCSI.*
(d) *Rx_2 has partial CSIR and complete SCSI.*
(e) *Both Tx_2 and Rx_2 know the IR-HARQ protocol of the PU, in particular, they know p_1, r_1 and $N_{T,1}$.*
(f) *Both Tx_2 and Rx_2 can decode the instantaneous and error-free feedback of the PU; they can both infer $k_{1,n}$.*

Assumption 4.2 implies that Rx_2 does not use Successive Interference Cancellation (SIC) techniques. Assumption 4.2(b) is classically made and indicates that the queue of Tx_2 is always non-empty so that it always has an information packet to transmit. Assumption 4.2(c) and Assumption 4.2(d) have been discussed at the beginning of this subsection. 4.2(e) often holds because parameters r_1, p_1 and $N_{T,1}$ are given in standards. Assumption 4.2(f) indicates that the SU can decode the feedback bits of the protocol of the PU. Assumption 4.2(f) is a direct consequence of Assumption 4.1(e).

4.2.4 Associated optimization problem

In this chapter, we assume that the SU potentially changes its transmission parameters (power and rate) in each slot. Theoretically it means that at each time nT, the SU chooses a power $p_{2,n}$ and a code-rate $r_{2,n}$ and uses these parameters to transmit data in slot n. Because of the no CSIT assumption, once $p_{2,n}$ has been chosen by Tx_2, the appropriate choice for $r_{2,n}$ is given by the following

$$r_{2,n} = r_2(p_{2,n}) = \max_{r \in \mathbb{R}^+} r\mathbb{P}\left[\log_2\left(1 + p_{2,n}\beta_{2,n}\right) \geq r\right]. \tag{4.6}$$

This choice for $r_{2,n}$ ensures to obtain the best instantaneous rate for a given level of interference $p_{2,n}$. In the end, it means that, at each slot, the SU only adapts its power $p_{2,n}$, choosing its rate by $r_{2,n} = r_2(p_{2,n})$.

The performance of the SU is also evaluated considering the throughput. The SU throughput is defined as

$$\eta_2(\pi_2) = \liminf_{t \to \infty} \frac{1}{t}\mathbb{E}_{\pi_2}\left[\sum_{n=0}^{t-1} r_2(p_{2,n})\right]. \tag{4.7}$$

In the same way, we also introduce here the SU average power as follows

$$\bar{P}_2(\pi_2) = \limsup_{t \to \infty} \frac{1}{t}\mathbb{E}_{\pi_2}\left[\sum_{n=0}^{t-1} p_{2,n}\right]. \tag{4.8}$$

In this thesis, we address the problem of finding the allocation π_2 that maximizes the throughput of the SU under peak power, average power and PU throughput requirement constraints. This general problem formulation can be stated as follows:

$$\begin{aligned}\eta_2^\star \quad = \quad & \sup_{\pi_2 \in \Pi} \; \eta_2(\pi_2) && (4.9)\\ \text{s.t.} \quad & \eta_1(\pi_2) \geq \eta_{1T} && (4.10)\\ & \bar{P}_2(\pi_2) \leq P_{2A} && (4.11)\\ & \forall n \in \mathbb{N},\; p_{2,n} \leq P_{2M}, && (4.12)\\ & \forall n \in \mathbb{N},\; p_{2,n} \geq 0 && (4.13)\end{aligned}$$

Equation (4.10) guarantees a minimal PUs throughput of η_{1T}. Equation (4.11), guarantees that the SU average power is below P_{2A}. Equation (4.12) guarantees that the SU peak power is below P_{2M}. The last equation, given by equation (4.13) guarantees that the SU peak power is positive.

4.3 The completely observable problem

In the preceding section, we have addressed the optimization problem (4.9) under Assumptions 4.1 and 4.2. To simplify the resolution of this optimization problem, we suppose in this section that, at every time nT, Tx_2 and Rx_2 have access to $i_{1,n}$, the ACMI at Rx_1. This access can be done by any means. In a realistic scenario, one can assume that a Rx_1 sends $i_{1,n}$ over a broadcast common control channel to be more compliant with the SU. In a more theoretical scenario, one can assume that Tx_2 and Rx_2 are genie aided. In any case, since we suppose that the SU has more information about the PU, the obtained results will be an upper bound on the achievable results considering Assumptions 4.1 and 4.2. The last advantage of considering that Tx_2 and Rx_2 have access to $i_{1,n}$ is that, using results of Section 1.4.2, the optimization problem (4.9) can be viewed as a CMDP. An *a posteriori* consequence of considering this CMDP is that the initial problem, considering Assumptions 4.1 and 4.2, can be thought as a PSI-MDP, this will be the object of Section 4.4.

4.3.1 The CMDP formulation

Similarly to what has been proposed in Chapter 3, we address the optimization problem given in equation (4.9) as a CMDP. The CMDP definition adapted to this problem is the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{W}, Q, R_2, R_1, P_2, \eta_{1T}, P_{2A})$ where each component is described as follows.

- The *state space*: $\mathcal{S} = \{0, 1, ..., N_{T,1}\} \times [0, r_1]$ is the set of all possible states of the PU IR-HARQ protocol (see Section 1.4.2 for more details). By construction this state is compact and then verifies Assumption 2.3(a).

- The *action space* $\mathcal{A} = [0, P_{2M}]$ is the space of all available powers for Tx_2. We suppose that for every $s \in \mathcal{S}$, $A(s) = \mathcal{A}$ so that $\mathcal{U} = \mathcal{S} \times \mathcal{A}$. By construction again $\mathcal{A}$ and $\mathcal{U}$ are compact which proves Assumption 2.3(a).

- In slot n, the gains $\alpha_{11,n}$, $\alpha_{12,n}$, $\alpha_{21,n}$ and $\alpha_{22,n}$ are unknown by Tx_2 and are considered as disturbances (see Assumption 4.2(c)). In consequence, we consider the following disturbance space: $\mathcal{W} = \mathbb{R}_+^4$.

- The *system function* $g(\cdot)$ is a deterministic function from $\mathcal{S} \times \mathcal{A} \times \mathcal{W}$ to $\mathcal{S}$ taking into account the evolution of the system from state $s_n \in \mathcal{S}$ in slot n to state $s_{n+1} \in \mathcal{S}$ in slot $n+1$ when the action $p_{2,n}$ is performed and when the disturbance is w_n. $g(\cdot)$ is globally defined as

$$g(s_n, p_{2,n}, w_n) = (k_{1,n+1}, i_{1,n+1}) = s_{n+1}. \tag{4.14}$$

Similarly to what has been done in Chapter 1, we present the function g in a table representation given in Table 4.1. This tabular can be interpreted as follows, if at time n, $s_n = (k_{1,n}, i_{1,n})$ with $0 < k_{1,n} < N_T - 1$, and if $i_{1,n} + \Delta(w_n, p_{2,n}) < D_T$, then s_{n+1} given by equation (4.14) will be $(k_n + 1, i_{1,n} + \Delta(w_n, p_{2,n}))$. We consider that the PU uses the channel code described in Section 1.2.3 so that

$$\Delta(w_n, p_{2,n}) = \log_2\left(1 + \frac{p_1 \alpha_{11,n}}{1 + p_{2,n}\alpha_{21,n}}\right), \tag{4.15}$$

where the $\alpha_{ij,n}$ are given by equation (4.2), p_1 is the *constant* primary power and note that we have assumed $\sigma_{z_1} = \sigma_{z_2} = 1$.

	$i_{1,n} + \Delta(w_n, p_{2,n}) < D_T$	$i_{1,n} + \Delta(w_n, p_{2,n}) \geq D_T$
$k_{1,n} = 0$	$(1, \Delta(w_n, p_{2,n}))$	$(0, 0)$
$0 < k_{1,n} < N_T - 1$	$(k_n + 1, i_{1,n} + \Delta(w_n, p_{2,n}))$	$(0, 0)$
$k_{1,n} = N_T - 1$	$(N_T, 0)$	$(0, 0)$
$k_{1,n} = N_T$	$(1, \Delta(w_n, p_{2,n}))$	$(0, 0)$

Table 4.1 – *State transition Table*

- The evolution of the system is again statistically represented by the *transition law* Q defined, for a given measurable subset $B \subset \mathcal{S}$ and a pair $(s, p_{2,n}) \in \mathcal{S} \times \mathcal{A}$ as follows:

$$Q(B|s, p_{2,n}) = \mathbb{P}\left[s^{n+1} = g(s_n, p_{2,n}, w_n) \in B | s_n = s, a_n = p_{2,n}\right]. \tag{4.16}$$

The expression of $Q(B|s,a)$ is provided in Appendix (D.1) and we show that $Q(B|s,a)$ verifies Assumption 2.3(d) in Appendix (D.2).

• In slot n, the definition of the primary instantaneous reward is given in Section 4.2 by equation (4.5). We remark here that R_1 is bounded and Lipschitz continuous so that it verifies Assumption 2.3(c).

• Similarly R_2^n can be written as function of $(s_n, p_{2,n}) \in \mathcal{S} \times \mathcal{A}$ as

$$R_2(s_n, p_{2,n}) = r_2(p_{2,n}). \tag{4.17}$$

The proof of Lipschitz continuity of R_2 is given in Appendix D.3. Consequently, R_2 is bounded and Lipschitz continuous so that it verifies Assumption 2.3(b).

• The instantaneous cost associated with the average power constraint is the following:

$$P_2(s_n, p_{2,n}) = p_{2,n}. \tag{4.18}$$

We remark here that P_2 is bounded and Lipschitz continuous so that it verifies Assumption 2.3(c).

We now briefly review, in our context, the concept of policy defined in Section 2.2.2.

Policy

Applying the definition given in Section 2.2.2 to the present case, the history up to time t is defined as follows,

$$h_t = \left(\left(k_{1,0}, i_{1,0}\right), p_{2,0}, \left(k_{1,1}, i_{1,1}\right), p_{2,1}, \ldots \left(k_{1,t}, i_{1,t}\right)\right) \tag{4.19}$$

The two main sets of policies that are of interest in this chapter are the following:

Π: The general set of policies, every $\pi_2 \in \Pi$ is defined as a sequence of stochastic kernels, $\pi_2 = \{\pi_{2,t}(dp_{2,t}|h_t)\}$,

Π_{RS}: The set of randomized stationary policies, every $\pi_2 \in \Pi$ is defined as a sequence of the form: $\pi_2 = \{\varphi(dp_{2,t}|k_{1,t}, i_{1,t})\}$. In the sequel, we will not distinguish between π_2 and φ.

Finally, the long-term reward and cost functions are given by equations (4.7), (4.4), and (4.8) (supposing without loss of generality that the initial state is $(0,0)$). In consequence, we have completely described the CMDP framework associated with the optimization problem (4.9) when $i_{1,n}$ is assumed to be known at Tx_2 and Rx_2. In the next subsection, we will show that we can discretize the state and action spaces to get approximate solutions of this CMDP.

4.3.2 Existence and computation of an optimal policy

In the preceding subsection, we have checked almost every assumptions of Assumption 2.3. The only one missing is Assumption 2.3(e). This assumption is that every policy $\varphi \in \Pi_{RS}$ induces a uniform ergodicity Markov chain on $\mathcal{S}$. To check Assumption 2.3(e), we will verify Condition 2.1(a): "There exists $s_0 \in \mathcal{S}$ and $\epsilon \in \mathbb{R}^+$ such that $Q(\{s_0\}\,|s, a) \geq 0$, with $\epsilon > 0$".

To check Condition 2.1(a), we remark that state $s_0 = (0, 0)$ is accessible from every other state. s_0 physically represents a successful decoding of Rx_1. The fact that s_0 is accessible from every other state means that from every state s, there is always a non-zero probability of a successful decoding at Rx_1 (as long as P_{2M} is finite). From any state $s = (k_1, i_1) \in \mathcal{S}$ and an action $p_2 \in \mathcal{A}$, using Table 3.1 (or equivalently the expression of Q given in D.3) we have

$$\begin{aligned} Q(\{s_0\}\,|k_1, i_1, p_2) &= \mathbb{P}\left[i_1 + \log_2(1 + \frac{p_1\alpha_{11}}{1 + p_2\alpha_{21}}) \geq r_1 | i_1, p_2\right] \\ &\geq \mathbb{P}\left[\log_2(1 + \frac{p_1\alpha_{11}}{1 + P_{2M}\alpha_{21}}) \geq r_1\right] > 0. \end{aligned} \tag{4.20}$$

where the first line is obtained by applying the definition of a successful decoding (see Section 1.2.3), and the second equation comes from the positiveness of i_1 and p_2 and from the finiteness of P_{2M}. This is sufficient to prove that Assumption 2.3(e) holds.

We finally suppose that $\eta_{1T} < \eta_1(\zeta_0)$ and $P_{2A} > 0$ to guarantee that the set of admissible policies is not-empty. The policy ζ_0 is the policy such that Tx_2 always transmits with power $p_2 = 0$. Using Theorem 2.1, it follows that the optimization problem given in equation (4.9) is solvable: there exists an optimal policy for equation (4.9) within the set Π_{RS}. It also proves that an optimal policy can be found using an infinite dimensional programming.

In the next subsection, we propose a discrete CMDP, that can be solved numerically and leads to an approximation of an optimal policy for the optimization problem (4.9).

4.3.3 Evaluation of the performances of the optimal policy

Similarly to what we proposed in Section 3.3.2, we propose different grids for $\mathcal{S}$ and $\mathcal{A}$ so we drop the indexes in the definition of the discrete CMDP. We now describe the discretization procedure proposed in Section 2.5. In particular we introduce the following finite CMDP:

$$\left(\tilde{S}, \tilde{A}, \tilde{Q}, \tilde{R}_2, \tilde{R}_1, \tilde{P}_2\right).$$

The proposed method is almost the same as the one for discretizing the CMDP in Section 3.3.2.

- The set $\tilde{S}$ is built from the following partition of the set $[0, r_1]$:

$$[0, r_1] = \bigcup_{j=0}^{N_I - 2} [jh,\ (j+1)h[\ \cup\ [(N_I - 1)h, r_1]\,, \tag{4.21}$$

where N_I represents the number of sets in the partition of $[0, r_1]$ and $h = r_1/N_I$ defines the length of each interval. The set $\tilde{S}$ is composed of elements of the following form:

$$\tilde{s}_{k,j} = (k, (j + \omega)h), \tag{4.22}$$

where ω belongs to $[0, 1[$.

- The set of actions is discretized as $\tilde{A} = \{0, \ell P_{2M}, \ldots, \ell(N_A - 1)P_{2M}\}$, where $\ell = 1/N_A$ and P_{2M} is the SU peak power constraint.
- For every $\tilde{s}_{k',j'}$, $\tilde{s}_{k,j}$, and $\tilde{a}_l$, the transition law is discretized as follows:

$$\tilde{Q}\left(\tilde{s}_{k',j'} | \tilde{s}_{k,j}, \tilde{a}_l\right) = Q\left(\{k'\} \times [j'h, (j'+1)h[\, | \tilde{s}_{k,j}, \tilde{a}_l\right). \tag{4.23}$$

$\tilde{Q}$ can be interpreted as a matrix of N_S lines and $N_S \times N_A$ columns.

- We finally define the reward functions on $\tilde{\mathcal{U}}$ as follows

$$\tilde{R}_2(\tilde{u}_k) = R_2(\tilde{u}_k) \tag{4.24}$$
$$\tilde{R}_1(\tilde{u}_k) = R_1(\tilde{u}_k) \tag{4.25}$$
$$\tilde{P}_2(\tilde{u}_k) = \bar{P}_2(\tilde{u}_k). \tag{4.26}$$

We finally build a finite linear programming based on the method proposed in Section 2.5. Let the notation $\mathbf{R_2}$, $\mathbf{R_1}$, $\mathbf{P_2}$ and $\mathbf{1}$ stand for the vectors of size N_U containing respectively $\tilde{R}_2(\tilde{u})$, $\tilde{R}_1(\tilde{u})$, $\tilde{P}_2(\tilde{u})$ and 1. Using these vectors, we can write the following finite linear programming:

$$\begin{aligned} \tilde{\eta}_2 = &\sup_{\mathbf{m} \in \mathbb{R}^{N_U}} \mathbf{R_2 m}^T \\ \text{s.t. } &\sum_{\tilde{s} \in \tilde{S}} m(\tilde{s}, \tilde{a}) - Q_{\tilde{a}} \mathbf{m}^T = \mathbf{0},\ \forall \tilde{a} \in \tilde{A} \\ &\mathbf{1 m}^T = 1 \\ &\mathbf{R_1 m}^T \geq \eta_{1T} \\ &\mathbf{P_2 m}^T \leq P_{2A} \\ &\mathbf{m} \geq \mathbf{0}. \end{aligned} \tag{4.27}$$

From a solution $\mathbf{m}^\star$ of the finite linear programming given in equation (4.27) we compute a policy for the optimization problem given in equation (4.9) as follows:

$$\tilde{\varphi}(da|s) = \sum_{\tilde{a}} \frac{\mathbf{m}^\star(\tilde{s}_j, \tilde{a})}{\sum_{\tilde{s}} \mathbf{m}^\star(\tilde{s}, \tilde{a})} \delta_{\tilde{a}}(da), \ \forall s \in S_j. \tag{4.28}$$

For some pairs (η_{1T}, P_{2A}) and some discretization parameters (N_S, N_A, ω), $\mathbf{m}^\star$ may not exist. This happens, in particular, if the linear programming (4.27) is not consistent. In this case, we have considered $\tilde{\varphi} = \zeta_0$. This choice is made because ζ_0 is admissible for the problem given in equation (4.9).

In the sequel, we will make the difference between $\eta_2(\tilde{\varphi})$ and $\tilde{\eta}_2 = \mathbf{R_2}\mathbf{m}^{\star T}$. $\eta_2(\tilde{\varphi})$ corresponds to the SU throughput of the policy $\tilde{\varphi}$. The policy $\tilde{\varphi}$ is a "continuous" policy built from $\mathbf{m}^\star$, solution of the problem (4.27). In a similar way we consider $\tilde{\eta}_1 = \mathbf{R_1}\mathbf{m}^{\star T}$ and $\eta_1(\tilde{\varphi})$ for the throughput of the PU, and $\tilde{P}_2 = \mathbf{P}_2\mathbf{m}^{\star T}$ and $\bar{P}_2(\tilde{\varphi})$ for the SU average power.

4.3.4 Influence of ω and N_S

In this subsection, we present some simulation results that highlight the impact of ω, N_S, and N_A. For all the results proposed in this section, we have considered a PU using an IR-HARQ protocol with 3 retransmissions (so that $N_{T,1} = 4$), $r_1 = 7.12$ bits per channel use (bpcu), and a power of $p_1 = 10dBW$. The channel parameters are $\bar{\alpha}_{11} = \bar{\alpha}_{22} = 1$ and $\bar{\alpha}_{12} = \bar{\alpha}_{21} = 0.5$. The secondary user has a peak power constraint $P_{2M} = 10dBW$, an average power constraint $P_{2A} = 5dBW$.

The parameter ω defines the position of the representative $\tilde{s}_i$ within the sets S_i. Taking $\omega = 0$ is equivalent to considering the PU worst case scenario. Indeed, when the discretized system is in state $\tilde{s}_i$, the true state s is in the set $S_i = \{k_i\} \times [j_i h, \ (j_i+1)h[$. Taking $\omega = 0$ is equivalent to considering that the state is in $j_i h$, which is the worst case scenario among every state of S_i. On the other hand, the case $\omega \to 1^-$ is equivalent to an optimistic guess on the real state s. In Figure 4.2, we show curves of the SU throughput as a function of η_{1T} for ω equal to 0, 0.5, and 0.999. The results presented in Figure 4.2 correspond either to the optimal value of the finite linear programming (dashed lines) or to the simulation of the policy given by the linear programming.

Figure 4.2 highlights that the optimal SU throughput computed with $\omega = 0$ underestimates the simulated SU throughput. On the other hand the optimal SU throughput computed with ω close to 1 overestimates the simulated SU throughput. In addition to these results, we provide the same curves for $\omega = 0.5$ where theoretical and simulated throughputs are close to each other.

In addition to these results, we compare in Figure 4.3 the simulated through-

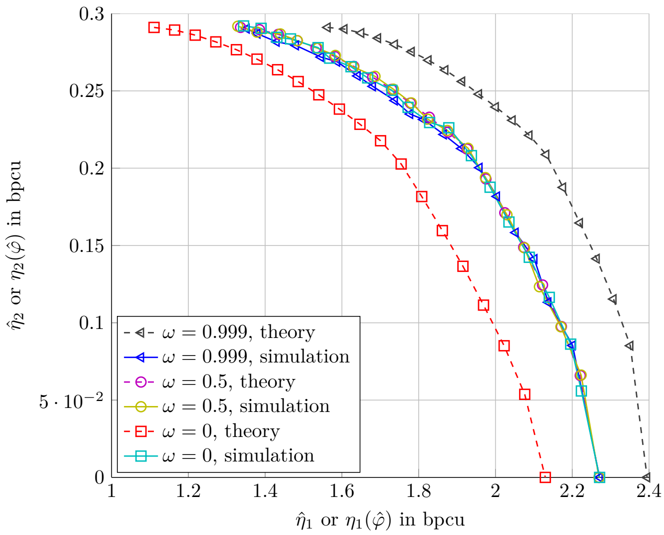

Figure 4.2 – *Comparison between throughput regions* $(\tilde{\eta}_1, \tilde{\eta}_2)$ *obtained by solving* (4.27) *and the simulated throughput region* $(\eta_1(\tilde{\varphi}), \eta_2(\tilde{\varphi}))$*. The comparison is done for* $\omega \in \{0, 0.5, 0.999\}$ *with* $N_I = 16$ *and* $N_A = 16$*. The simulated throughput region is computed using a Monte-Carlo method on* 10^6 *slots.*

put of the PU and the target PU throughput η_{1T}. The results presented in Figure 4.3 corroborate that the PU throughput is underestimated when $\omega = 0$, correctly estimated when $\omega = 0.5$, and underestimated when $\omega \to 1^-$.

We complete these two results by comparing the simulated values of the average power of the SU to the constraint value of $P_{2A} = 5dBW$. These comparisons are given in Figure 4.4 for $\omega = 0$, $\omega = 0.5$, and $\omega \to 1^-$. As expected, the results obtained when $\omega = 0.5$ are close to the constraint. Surprisingly, we observe that when $\omega = 0$, the constraint is sometimes not respected. On the other hand, when $\omega \to 1^-$ the simulated values of the average SU power remains within the constraint.

We will now analyse the influence of the parameter N_S. This parameter represents the number of states in $\tilde{S}$. Considering $N_S \to \infty$ is equivalent to considering $h \to 0$ therefore high values of N_S lead to better approximation of η_1, η_2, and $\bar{P}_2$. The limiting case is discussed in Section 4.5. Figure 4.5 highlights that the solution of the infinite linear programming can be approximated from below by increasing N_S and taking $\omega = 0$. On the contrary, Figure 4.6 indicates that the infinite linear programming can be approximated from above by increasing N_I and taking $\omega \to 1^-$.

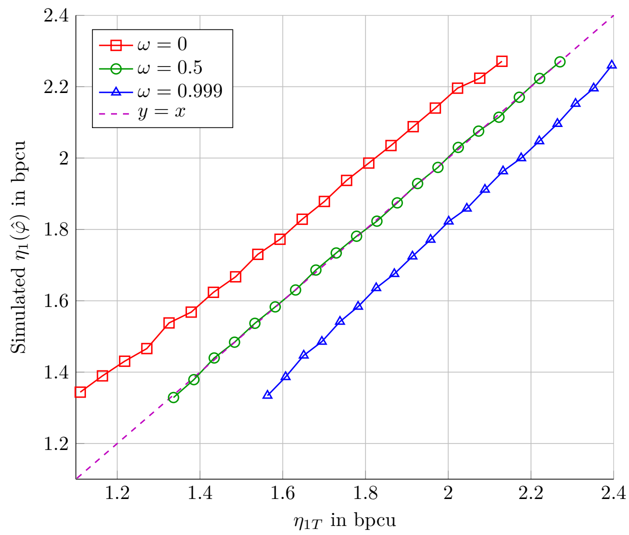

Figure 4.3 – *Simulated throughput of the PU versus* η_{1T} *for* $\omega \in \{0, 0.5, 0.999\}$. $\tilde{\varphi}$ *has been computed with* $N_I = 16$ *and* $N_A = 16$. *The simulated values of* $\eta_1(\tilde{\varphi})$ *are computed using Monte-Carlo methods on* 10^6 *slots.*

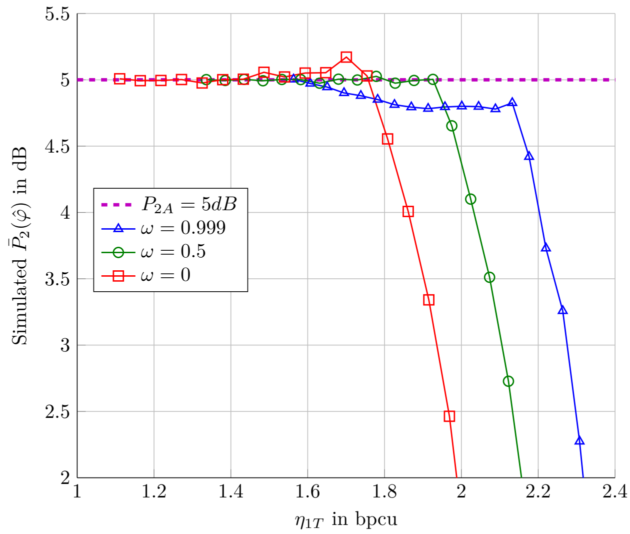

Figure 4.4 – *Comparison of the simulated average power* $\bar{P}_2(\tilde{\varphi})$ *and* P_{2A} *for different values of* η_{1T} *and* $\omega \in \{0, 0.5, 0.999\}$. $\tilde{\varphi}$ *has been computed with* $N_I = 16$ *and* $N_A = 16$. *The simulated values of* $\eta_1(\tilde{\varphi})$ *are computed using Monte-Carlo methods on* 10^6 *slots.*

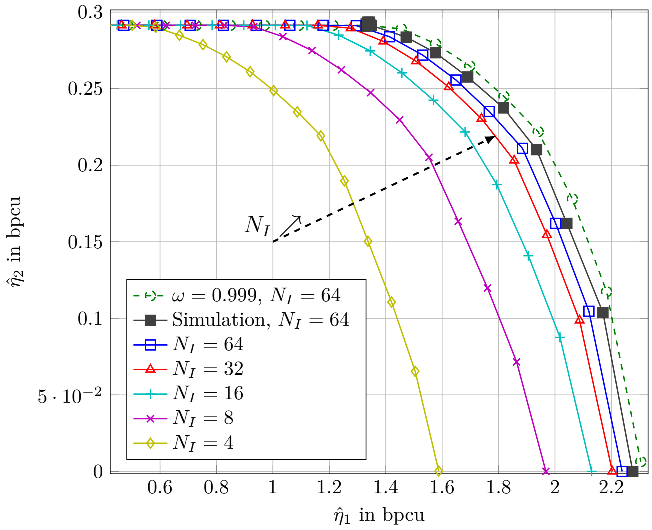

Figure 4.5 – *Throughput regions for different values of $N_I \in \{4, 8, 16, 32, 64\}$ with $N_A = 16$ and $\omega = 0$. We give two other curves as references. The first one is the simulated throughput region of $(\eta_1(\tilde{\varphi}), \eta_2(\tilde{\varphi}))$ where the $\tilde{\varphi}$ are computed with $N_I = 64$. The simulated values have been computed using Monte-Carlo method on 10^6 slots. The second reference is the throughput region obtained with $N_I = 64$, $N_A = 16$ and $\omega = 0.999$.*

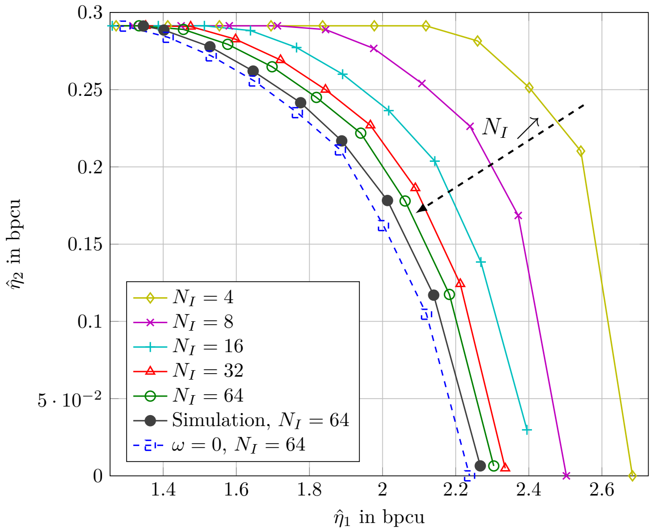

Figure 4.6 – *Throughput regions for different values of $N_I \in \{4, 8, 16, 32, 64\}$ with $N_A = 16$ and $\omega = 0.999$. We give two other curves as references. The first one is the simulated throughput region of $(\eta_1(\tilde{\varphi}), \eta_2(\tilde{\varphi}))$ where the $\tilde{\varphi}$ are computed with $N_I = 64$. The simulated values have been computed using Monte-Carlo method on 10^6 slots. The second reference is the throughput region obtained with $N_I = 64$, $N_A = 16$ and $\omega = 0$.*

4.4 The partially observable problem

In Section 4.3, we have shown that under Assumption 4.1, Assumption 4.2 and assuming that $i_{1,n}$ is known to Tx_2 and Rx_2, the optimization problem given in equation (4.9) is a CMDP that can be solved by discretizing the state and action spaces. In this section, we propose to tackle the initial problem which is given in equation(4.9) only under Assumption 4.1, Assumption 4.2.

The only difference between the framework of Section 4.3 and the one of this section is that, in this section Tx_2 and Rx_2 do not know $i_{1,n}$. On the other hand, we still consider the PU to be compliant (cf. Assumption 4.1(e)). Hence, by counting the PU ACK and NACK bits, Tx_2 and Rx_2 can easily track $k_{1,n}$. This framework is then equivalent to the PSI-MDP model presented in Section 2.6.1 where $k_{1,n}$ is the fully observable part of the PU IR-HARQ protocol and $i_{1,n}$ is the hidden part of this system.

Let $\mathcal{X}$ and $\mathcal{Y}$ be defined as the respective sets: $\mathcal{X} = [0, r_1]$ and $\mathcal{Y} = \{0, 1, \ldots, N_{T,1}\}$, the PSI-MDP considered in this section is given by the tuple $(\mathcal{X}, \mathcal{Y}, \mathcal{A}, Q, R_2, R_1, P_2)$, where $\mathcal{A}$, Q, R_2, R_1, P_2 are these of Section 4.3.

To solve this PSI-MDP, wee apply the method proposed in Section 2.6.1: we first convert this PSI-MDP to the CMDP $(\mathcal{Z}, \mathcal{A}, Q_Z, R_2', R_1', P_2')$ where each component is defined as follows.

- $\mathcal{Z} = \mathcal{P}(\mathcal{X}) \times \mathcal{Y}$ where $\mathcal{P}(\mathcal{X})$ is the set of every probability measures on $\mathcal{X}$ and $\mathcal{Y} = \{0, 1, \ldots, N_T\}$.
- $\mathcal{A}$ is the action space and is unchanged compared to the above definition.
- Q_Z is the transition matrix for Z_t, defined by equation (2.95).
- R_2', P_2', and R_1' are defined, for every $z = (b_i, y) \in \mathcal{Z}$ and every $a \in \mathcal{A}$, as follows:

$$\begin{cases} R_2'(z,a) = \int_{\mathcal{X}} R_2(x,y,a) b_i(dx), \\ {R_1}'(z,a) = \int_{\mathcal{X}} R_1(x,y,a) b_i(dx) \\ {P_2}'(z,a) = \int_{\mathcal{X}} P_2(x,y,a) b_i(dx). \end{cases} \tag{4.29}$$

This CMDP is again projected on the Beta laws parametrized by their mean and variance. We have finally solved the CMDP problem

$$(\Theta_{mv} \times \mathcal{Y}, \mathcal{A}, Q_Z', R_2', R_1', P_2')$$

using the method proposed in Section 2.5.

In this subsection, we present some simulation results that highlight the PSI-MDP performances. We have considered a PU using an IR-HARQ protocol with 2 retransmissions (so that $N_{T,1} = 3$), $r_1 = 7.12$ bits per channel use (bpcu), and a power of $p_1 = 10dBW$. The channel parameters are $\bar{\alpha}_{11} = \bar{\alpha}_{22} = 1$ and $\bar{\alpha}_{12} =$

$\bar{\alpha}_{21} = 0.5$. The secondary user has a peak power constraint $P_{2M} = 10dBW$, an average power constraint $P_{2A} = 5dBW$. In Figure 4.7 we compare the results of Section 4.3 with $N_I = 32$ and $\omega = 0.5$ with the results obtained thanks to the PSI-MDP framework.

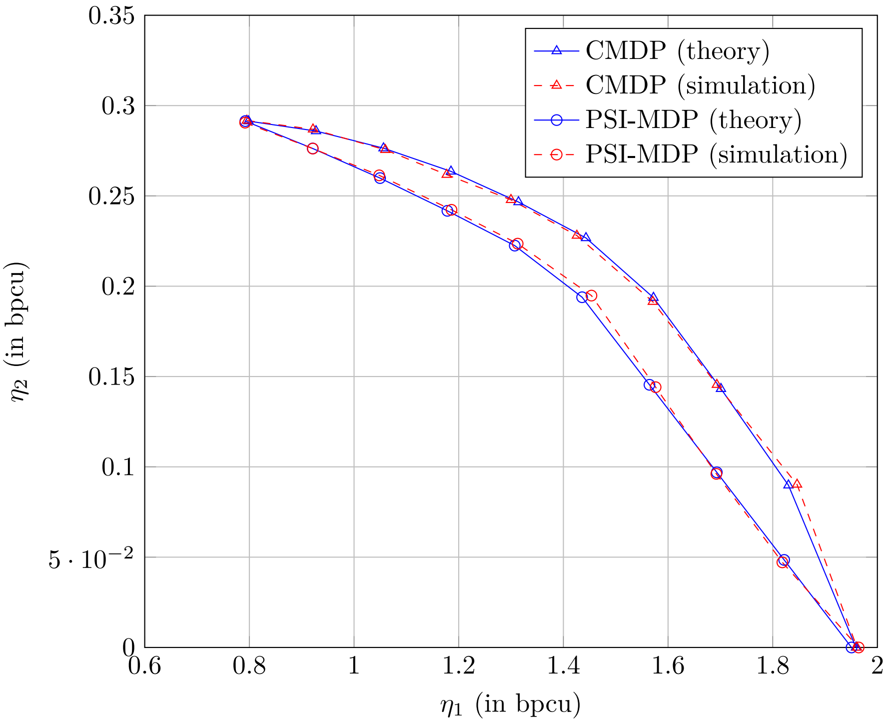

Figure 4.7 – *Throughput regions for the completely observable problem and the partially observable problem, for* N_A .

In Figure 4.7 , we observe that the loss due to the absence of knowledge of $i_{1,n}$ is small (less than 10% of throughput loss for the PU for a given SU throughput). However, the achieved SU throughput is still substantial and is performed by considering uniquely that the SU can hear the 1-bit feedback of the PU IR-HARQ protocol. Finally one can note that although the pdf of the ACMI has been projected on Beta laws, the simulated throughput is still very close to the one obtained using the projection.

4.5 Conclusion and Perpectives

In this chapter we have proposed to address the SU power allocation under a throughput-loss constraint for an IR-HARQ based PU. By assuming a compliant PU that broadcasts its feedback bits, we have shown that a significant SU throughput can be achieved while mitigating the throughput loss of the PU. We have proposed, in a first time, an upper bound on the SU achievable throughputs by considering that the SU has complete access to the state of the PU. In this case, we have shown that the SU power allocation reduces to a CMDP. This is a direct consequence of the IR-HARQ model proposed in Chapter 1. From this model, we have shown the existence of a solution and proposed a finite LP to approximate this solution. This approximation was simulated. We have in particular observed that in the case $\omega = 0.5$ only a few number of discrete states ($N_I = 16$) are required to obtain an accurate estimation of the optimal policy.

In a second time, we have considered the same SU power allocation except that the state of the PU is only accessible to the SU through the PU feedback bits. Although this only information is sufficient for ARQ for the Type-I HARQ protocol (see [Levorato et al., 2012]), it is not the case for the IR-HARQ protocol considered in our context. In the case of the IR-HARQ protocol, the feedback bits only provide a partial state information. In consequence, this problem has been considered as a PSI-MDP. Using the method proposed in Section 2.6.4, we have given approximations for the PSI-MDP solution. We have in particular observed that a non-negligible throughput can be obtained for the SU.

On the structure of optimal allocation policies

For practical considerations, one can think of some other (simpler) kind of policies. For instance policies with constant power. This kind of policy arises when the feedback of the PU cannot be heard by the SU. In this case, the SU does not have any information on the instantaneous state of the PU IR-HARQ protocol. Since the SU transmitter (Tx_2) has only SCSI, it will look for the maximal power (to maximize its own throughput) that guarantees the SU throughput target (η_{1T}). Figures 4.8 and 4.9 highlight that constant power allocations are the policies that maximize the SU throughput when $\eta_{1T} = 0$ and that there is a gain in using the PU feedback (this effect was pointed out in [Eswaran et al., 2007]).

One can also consider policies where the only admissible powers are P_{2M} or 0. These policies are called ON/OFF policies in our paper. These policies are somewhat at the opposite from the policies with constant power. Indeed Figure 4.9 shows that, when $P_{2M} = P_{2A}$, ON/OFF strategies are close to optimal strategies with $\mathcal{A} = [0, P_{2M}]$. This result is surprising and we do not have any proof of this optimality even if this fact has been observed in many different contexts (PU, SU and channel setting).

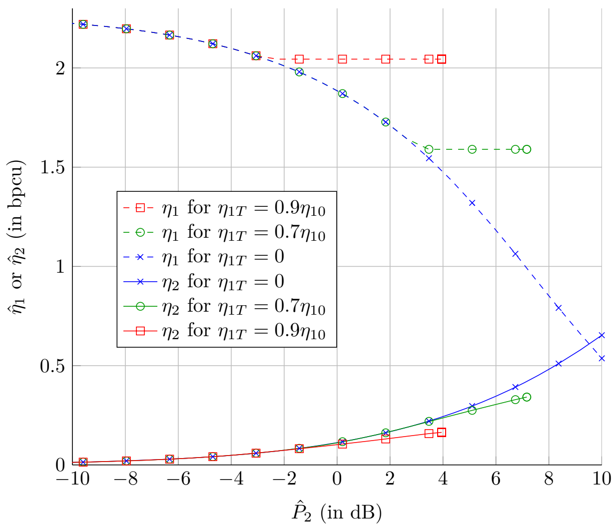

Figure 4.8 – *$\tilde{\eta}_1$ and $\tilde{\eta}_2$ versus $\tilde{P}_2$ for different values of η_{1T}.*

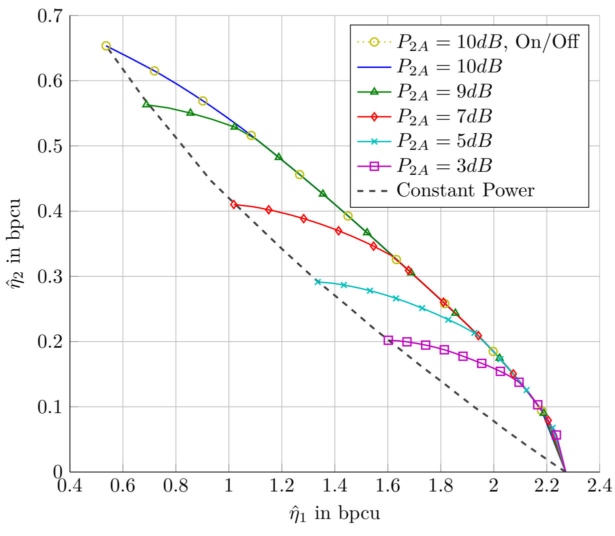

Figure 4.9 – *Throughput regions for different values of P_{2A}. The throughput region obtained when considering an On/Off allocation and a constant power allocation are give as references. These curves have been obtained taking $N_A = 32$, $N_I = 32$, and $\omega = 0.5$.*

On the admissibility of $\tilde{\varphi}$

As we have seen in the simulation results, it may happen that $\tilde{\varphi}$ lies outside the set Ω of admissible policies for the problem given by equation (4.9). This can be problematic in an HARQ context where we want to guarantee the QoS constraint for the PU. In this case, we propose to build a non-stationary policy $\tilde{\pi}$ such that,

(P1): if $\tilde{\varphi} \in \Omega$, then $\eta_2(\tilde{\pi}) = \eta_2(\tilde{\varphi})$, $\eta_1(\tilde{\pi}) = \eta_1(\tilde{\varphi})$ and $\bar{P}_2(\tilde{\pi}) = \bar{P}_2(\tilde{\varphi})$,

(P2): if $\tilde{\varphi} \notin \Omega$, then $\tilde{\pi} \in \Omega$.

To build $\tilde{\pi}$, we solve (4.27) with $\eta'_{1T} = \eta_{1T} + \epsilon_1$ and $P'_{2A} = P_{2A} - \epsilon_2$. The parameters $\epsilon_1 > 0$ and $\epsilon_2 > 0$ are chosen so that $\eta'_{1T} < \eta_{10}$ and $P'_{2A} > 0$. Tx_2 tracks the following values

$$\begin{cases} \eta_{1N} = \frac{1}{N} \sum_{n=0}^{t-1} R_1^n, \\ \bar{P}_{2N} = \frac{1}{N} \sum_{n=0}^{t-1} p_2^n. \end{cases}$$

Let Φ be the subset of $\mathbb{R}^2$ defined as $\Phi = \{(x, y) \in \mathbb{R}^2 : x \geq \eta_{1T} \text{ and } y \leq P_{2A}\}$. The policy $\tilde{\pi}$ is built as follows: if $(\eta_{1N}, \bar{P}_{2N}) \in \Phi$, Tx_2 uses $\tilde{\varphi}$. In every other cases, Tx_2 uses ζ_0. By assumption we have that $(\eta_{10}, 0) \in \Phi$. Using the results of [Ross, 1989], we can easily prove that the properties (P1) and (P2) hold. This method is heuristic and more evolved methods should be considered.

On Successive Interference Cancellation

In this chapter, the SU always considers the PU signal as noise. This means that the SU does not need to know the PU codebook and a fortiori the PU sent codeword. Although this is an advantage in terms of confidentiality (SU does not decode the PU messages), methods such that Backward Interference Cancellation (BIC) SIC can dramatically improve the performances of both systems. This was shown in [Michelusi et al., 2013a] in the Type-I HARQcontext. BIC and SIC techniques are extended in [Michelusi et al., 2013b] by introducing Chain Decoding (CD). These techniques have not yet been studied in a Type-II HARQ context but we can show their limit performances.

The limit of the BIC/SIC/CD schemes is attained when there is no interference from the PU to the SU. This limit means that all the interference is correctly removed for every block. This case corresponds to $\bar{\alpha}_{12} = 0$ in our context. In consequence, the case $\bar{\alpha}_{12} = 0$ is an upper bound on what is possible using the interference cancellation techniques of [Michelusi et al., 2013a] or [Michelusi et al., 2013b]. In Figure 4.10, this upper bound is compared to the throughput region for $\omega = 0.5$ of Figure 4.3. As it can be observed in Figure

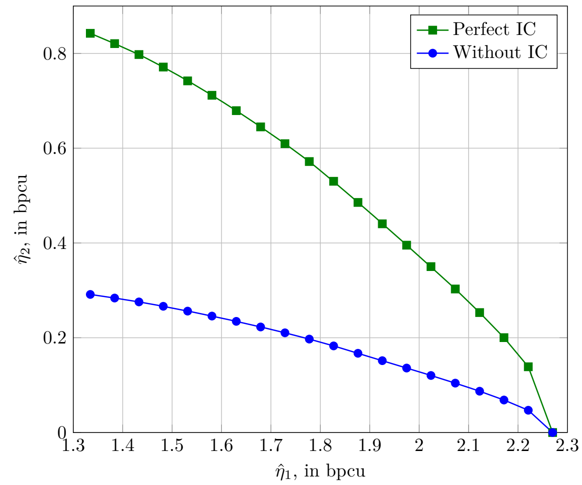

Figure 4.10 – *Comparison of the throughput regions with or without interference cancellation.*

4.10 the potential gain of the techniques proposed in [Michelusi et al., 2013a] and [Michelusi et al., 2013b] is substantial.

Conclusion

In this thesis we have proposed a CMDP approach to solve the problem of allocating resources for HARQ-based systems. This approach has been suggested because of the sequential evolution of a HARQ protocol. In this thesis, we have applied this approach, not only for the optimization of various HARQ protocols; but we have also considered this approach for problems involving HARQ protocols (even if these are not to be optimized).

In Chapter 1, we have proposed an introduction on the different types of HARQ protocols. In particular, we have presented the Type-I HARQ protocol, the CC-HARQ protocol and the IR-HARQ protocol. The throughput analysis of these three protocols has been proposed in the context of asymptotic Gaussian codebooks. This asymptotic Gaussian codebooks context has allowed us to derive controlled Markov models for representing the state evolution of these protocols depending on a power allocation. We have in particular derived a unified model to handle the CC-HARQ and the IR-HARQ protocols. Finally, the last contribution of Chapter 1 is to define a power allocation problem for optimizing the throughput of an HARQ protocol under peak and power constraints.

In Chapter 2, we have reviewed the existing literature about the general framework on CMDPs. This review not only presents the theoretical framework of CMDPs, but also the practical implementation aspects of CMDPs. In particular, we have given discretization procedures for dealing with approximating CMDPs with continuous state and action spaces. Since in the telecommunication area, the encountered problems are not always fully observable, we have presented a practical method for handling CMDP with partial state information (PSI-MDP). The main contribution of Chapter 2 has been to provide a practical condition, Condition 2.2, for ensuring that the CMDP corresponding to the PSI-MDP model is solvable.

In Chapter 3, we have analysed the throughput maximization under peak and power constraints. We have in particular considered two possible scenarios: i) the full state of the HARQ protocol is known by Tx before each slot, ii) this state is only partially observable. In the first case, we have that CMDPs are the appropriate tool for solving the allocation problems given in Chapter 1. The second proposed power allocation has been formulated as a PSI-MDP. The solution of this PSI-MDPhas shown that with only 1 bit of feedback, the power allocation obtained with the PSI-MDP case has almost the same performance as the one obtained with full state information (case i).

Finally, in Chapter 4, we have shown that with only one bit of feedback, we can evaluate and mitigate the throughput loss of the PU and obtain a non-negligible throughput for the SU. To show this result, we have formulated a completely observable problem in which the SU has access to the full state of the HARQ protocol of the PU. This completely observable problem has allowed us to upper-bound the achievable performance of the SU. From this completely observable problem, we have derived a PSI-MDP to handle the case where the SU can only listen to the feedback bits of the PU. In both cases, we have shown that non-negligible throughputs can be achieved by the SU while managing their impact on the performance of the IR-HARQ protocol of the PU.

To conclude this thesis, we want to point out that, due to their sequential nature, the HARQ protocols are in general more difficult to optimize than other communication systems. In the block-fading channel, there is often no exploitable closed-form expression for the throughput of the Type-II HARQ protocols. The framework proposed by the CMDPs is known to be well suited for sequential decision problems. In consequence, the framework of CMDPs is well suited for the resource allocation for HARQ protocols. On the other hand, there are still open problems left in this thesis, we now list some of them that are of notable interest.

Perspectives

• In this thesis, SCSI is required at the transmitter to perform resource allocation such as power or rate allocation considering HARQ protocols. It should be interesting to consider problems in which this SCSI is not available but can be inferred or learned along with the power allocation. Some learning algorithms exist and do not suppose any knowledge on the transition matrix (or kernel) for the MDP model. It is the case for example of some versions of the Q-learning algorithm. The major drawback of these techniques is that the constraints are in general difficult to handle. Extending these techniques to CMDPs is an interesting and challenging task.

• We have seen that the cornerstone of our approach is that the receiver accumulates some quantities: for example SNR or mutual information. Exploiting these quantity, we have built Markov models that are then used in CMDPs. In the context of resource allocation for BICM, the ACMI is often used to perform bit-loading and power allocation. It could be interesting to study if these problems can be interpreted in the CMDP framework.

• In this thesis, we have considered single user scenarios. Extending the proposed method to multiple users scenarios is a challenging task. However, stochastic games are a generalization of the CMDP framework for problems with multiple players. In consequence, we expect that stochastic games are an appropriate tool to generalize the results presented in this thesis to multi-user environments.

- Every optimization problem presented in this thesis can be re-written as follows:

$$\max_{\pi} \frac{\mathbb{E}_{\pi}\left[\sum_{i=1}^{T} R_i\right]}{\mathbb{E}_{\pi}\left[\sum_{i=1}^{T} D_i\right]}$$
$$s.t. \frac{\mathbb{E}_{\pi}\left[\sum_{i=1}^{T} C_i\right]}{\mathbb{E}_{\pi}\left[\sum_{i=1}^{T} D_i\right]} \leq \bar{C}.$$

This kind of problem is called fractional programming. The fractional programming approach for solving certain classes of CMDP has been proposed in [Neely, 2011]. Furthermore, we have already observed that this framework is appropriate for the design of rate adaptive HARQ systems. Generalizing this approach for optimization problems with continuous state space, should provide us new insights and new algorithms for solving rate or power allocation for HARQ based systems.

Appendices

- Appendix A -

Appendices of Chapter 1

A.1 Expression of $Q(B|s_n, p_n)$

$$Q(B_S|s_n, p_n) = \mathbb{P}[S_{n+1} \in B_S | S_n = s_n, P_n = p_n] \quad (A.1)$$

$$= \int_{\mathcal{S}} \mathbb{1}_B(s_{n+1}) Q(ds_{n+1}|s_n, p_n), \quad (A.2)$$

where $\mathbb{1}_B(s)$ is the function that is equal to 1 if $s_{n+1} \in B$ and 0 otherwise. In our case, the *transition kernel* has the following expression:

$$Q(ds_{n+1}|s_n, p_n) =$$
$$\begin{cases} \underline{\text{If } k_n < N_T - 1}: \\ \mathbb{P}[x_n + \Delta(\alpha_n, P_n) \geq D_T | P_n = p_n] \, \delta_{(0,0)}(ds_{n+1}) \\ \qquad + f_{\Delta|P}(x_{n+1} - x_n|p_n) \delta_{k_n+1}(dk_{n+1}) \ell(dx_{n+1}) \\ \underline{\text{If } k_n = N_T - 1}: \\ \mathbb{P}[x_n + \Delta(\alpha_n, P_n) \geq D_T | P_n = p_n] \, \delta_{(0,0)}(ds_{n+1}) \\ \qquad + \mathbb{P}[x_n + \Delta(\alpha_n, P_n) < D_T | P_n = p_n] \, \delta_{(N_T,0)}(ds_{n+1}) \\ \underline{\text{If } k_n = N_T:} \\ \mathbb{P}[x_n + \Delta(\alpha_n, P_n) \geq D_T | P_n = p_n] \, \delta_{(0,0)}(ds_{n+1}) \\ \qquad + f_{\Delta|P}(x_{n+1}|p_n) \delta_1(dk_{n+1}) \ell(dx_{n+1}) \end{cases} \quad (A.3)$$

Where, for Rayleigh channel and in the case of CC-HARQ protocol we have

$$\mathbb{P}[x + \Delta(\alpha_n, P_n) < D_T | P_n = p_n] = 1 - \exp\left[-\frac{D_T - x}{\bar{\alpha} p_n}\right], \quad (A.4)$$

and

$$f_{\Delta|P}(x|p_n) = \frac{1}{\bar{\alpha} p_n} \exp\left[-\frac{x}{\bar{\alpha} p_n}\right], x \geq 0. \quad (A.5)$$

For Rayleigh channel and in the case of IR-HARQ protocol we have

$$\mathbb{P}\left[x + \Delta(\alpha_n, P_n) < D_T | P_n = p_n\right] = 1 - \exp\left[-\frac{2^{D_T - x} - 1}{\bar{\alpha} p_n}\right], \tag{A.6}$$

and after some calculations we obtain

$$f_{\Delta|P}(x|p_n) = \frac{2^x \exp\left[-\frac{x}{\bar{\alpha} p_n}\right] \log(2)}{\bar{\alpha} p_n}, x \geq 0. \tag{A.7}$$

- Appendix B -

Appendices of Chapter 2

B.1 Stochastic Kernels

Definition B.1. [Bertsekas & Shreve, 1978, Chapter 7] Let $\mathcal{X}$ and $\mathcal{Y}$ be two separable metrizable spaces. A *stochastic kernel* $Q(dy|x)$ on $\mathcal{Y}$ given $\mathcal{X}$ is a collection of probability measures on $\mathcal{Y}$ parametrized by $x \in \mathcal{X}$.

Q is said to be *weak Feller* (or *weakly continuous*) if and only if for every measurable bounded function v, the function

$$x \mapsto \int_{\mathcal{Y}} v(y)Q(dy|x) \tag{B.1}$$

is bounded on $\mathcal{X}$.

Q is said to be *strong Feller* (or *strongly continuous*) if and only if for every measurable bounded function v, the function

$$x \mapsto \int_{\mathcal{Y}} v(y)Q(dy|x) \tag{B.2}$$

is bounded and continuous on $\mathcal{X}$.

B.2 Convergence of probability measures

Let $\mathcal{X}$ be a given Borel space. Let $\{m_n\}$ be a sequence of probability measures on the same space $\mathcal{X}$. Let m be probability measures on the same space $\mathcal{X}$. In this section we are interested in situations in which "$m_n \to m$" in some way see [Hernández-Lerma & Lasserre, 1996]. We first introduce the weak convergence concept.

Definition B.2. We say that m_n converges weakly to m, if and only if for every continuous and bounded function $v \in C(\mathcal{X})$:

$$\lim_{n\to\infty} \int_{\mathcal{X}} v(x)m_n(dx) = \int_{\mathcal{X}} v(x)m(dx). \tag{B.3}$$

The weak convergence is denoted as follows

$$m_n \xrightarrow{w.} m. \tag{B.4}$$

The second mode of convergence is the set-wise convergence.

Definition B.3. The set-wise convergence is defined as follows: for every $B \in \mathcal{B}(\mathcal{X})$,

$$m_n(B) \to m(B). \tag{B.5}$$

The last way of defining a convergence notion for m_n considers the TV-norm defined for every signed measure μ on $\mathcal{X}$ as follows:

$$||\mu||_{TV} = \sup_{B\in\mathcal{B}(\mathcal{S})} \mu(B) - \inf_{B\in\mathcal{B}(\mathcal{S})} \mu(B). \tag{B.6}$$

For two probability measures μ_1 and μ_2, $||\mu_1 - \mu_2||_{TV}$ is given by

$$||\mu_1 - \mu_2||_{TV} = 2 \sup_{B\in\mathcal{B}(\mathcal{X})} |\mu_1(B) - \mu_2(B)|\,. \tag{B.7}$$

We now introduce the convergence in TV-norm (sometimes referred to as the strong convergence).

Definition B.4. We say that m_n converges to m in the TV-norm if and only if

$$||m_n - m||_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{B.8}$$

B.3 Proof of the strong duality

This proof is based on the following fact. For every $\boldsymbol{\lambda}$, there exists a triplet $(\boldsymbol{\lambda}, u_{\boldsymbol{\lambda}}, \beta_{\boldsymbol{\lambda}})$ that is admissible for the dual problem (2.56). This result comes from the fact the Assumption 2.1 is sufficient for Theorem 2.2 of [Hernández-Lerma, 1989, Chapter 3.]. For a given $\boldsymbol{\lambda} \in \mathbb{R}^{n_c}$, the function

$$(s, a) \mapsto r(s, a) + \sum_{j=1}^{n_C} \lambda_j \left(V_j - c_j\left(s, a\right)\right)$$

is bounded and continuous. As a consequence, for every $\boldsymbol{\lambda} \in \mathbb{R}^{n_c}$, this theorem states that under Assumption 2.1, there exists $\beta_{\boldsymbol{\lambda}} \in \mathbb{R}$ and $u_{\boldsymbol{\lambda}} \in C(\mathcal{S} \times \mathcal{A})$ such that

$$\beta_{\boldsymbol{\lambda}} + u_{\boldsymbol{\lambda}}(s) = \max_{a\in A(s)} r(s,a) + \sum_{j=1}^{n_C} \lambda_j \left(V_j - c_j\left(s,a\right)\right) + \int_{\mathcal{S}} u_{\boldsymbol{\lambda}}(y) Q(dy|s,a). \quad \text{(B.9)}$$

Another consequence of the same theorem is that for every initial distribution ν_0,

$$\beta_{\boldsymbol{\lambda}} = \sup_{\pi\in\Pi} R(\nu_0,\pi) + \sum_{j=1}^{n_C} \lambda_j \left(V_j - C_j(\nu_0,\pi)\right) \quad \text{(B.10)}$$

Furthermore, because of the measurable selection theorem, there exists $\zeta_{\boldsymbol{\lambda}} \in \Pi_{DS}$ such that:

$$\beta_{\boldsymbol{\lambda}} + u_{\boldsymbol{\lambda}}(s) = r(s,\zeta_{\boldsymbol{\lambda}}(s)) + \sum_{j=1}^{n_C} \lambda_j \left(V_j - c_j\left(s,\zeta_{\boldsymbol{\lambda}}(s)\right)\right) + \int_{\mathcal{S}} u_{\boldsymbol{\lambda}}(y) Q(dy|s,\zeta_{\boldsymbol{\lambda}}(s)), \quad \text{(B.11)}$$

and

$$\beta_{\boldsymbol{\lambda}} = R(\zeta_{\boldsymbol{\lambda}}) + \sum_{j=1}^{n_C} \lambda_j \left(V_j - C_j\left(\zeta_{\boldsymbol{\lambda}}\right)\right). \quad \text{(B.12)}$$

We show here that the function $\boldsymbol{\lambda} \mapsto \beta_{\boldsymbol{\lambda}}$ is a continuous coercive function. This will be used to prove that this function attains its infimum. To do so, let $\boldsymbol{\epsilon} \in \mathbb{R}_+^{n_c}$, from equation (B.12) we obtain

$$\begin{aligned} |\beta_{\boldsymbol{\lambda}+\boldsymbol{\epsilon}} - \beta_{\boldsymbol{\lambda}}| &\leq \left|\sum_{j=1}^{n_C} \epsilon_j \left(V_j - C_j\left(\zeta_{\boldsymbol{\lambda}}\right)\right)\right| && \text{(B.13)} \\ &\leq ||\boldsymbol{\epsilon}||_\infty \sum_{j=1}^{n_c} |(V_j - C_j\left(\zeta_{\boldsymbol{\lambda}}\right))| && \text{(B.14)} \\ &\leq n_c\,||\boldsymbol{\epsilon}||_\infty \left(||\mathbf{V}||_\infty + ||\mathbf{c}_M||_\infty\right), && \text{(B.15)} \end{aligned}$$

where c_M is the vector $(c_{1,M}, c_{2,M}, \ldots, c_{n_c,M})$. From the boundedness of n_c, $||\mathbf{V}||_\infty$, and $||\mathbf{c}_M||_\infty$, we obtain that $\beta_{\boldsymbol{\lambda}}$ is Lipschitz continuous and so continuous.

Under Assumption 2.2, there exists $\varphi_0 \in \Pi_{RS}$ such that $\mathbf{C}(\varphi_0) < \mathbf{V}$. Using this φ_0 in (B.10) leads to

$$\lim_{||\boldsymbol{\lambda}||_\infty \to +\infty} \beta_{\boldsymbol{\lambda}} = +\infty. \quad \text{(B.16)}$$

There exists $\boldsymbol{\lambda}^\star$ such that

$$\beta_{\boldsymbol{\lambda}^\star} = \inf_{\boldsymbol{\lambda}>\mathbf{0}} \beta_{\boldsymbol{\lambda}}. \quad \text{(B.17)}$$

We finally prove that $\beta_{\boldsymbol{\lambda}^\star} = \beta^\star$. Since the triplet $(u_{\boldsymbol{\lambda}^\star}, \beta_{\boldsymbol{\lambda}^\star}, \boldsymbol{\lambda}^\star)$ is admissible for the dual LP (2.56), we have $\beta^\star \leq \beta_{\boldsymbol{\lambda}^\star}$. Let $(u, \beta, \boldsymbol{\lambda})$ be an admissible triplet for the dual LP (2.56). We have that for all $s \in \mathcal{S}$ and all $a \in A(s)$

$$\beta + u(s) \geq r(s,a) + \sum_{j=1}^{n_C} \lambda_j \left(V_j - c_j\left(s,a\right)\right) + \int_{\mathcal{S}} u(y) Q(dy|s,a). \tag{B.18}$$

For every $\varphi \in \Pi_{RS}$, integrating both sides by $\varphi(da|s)Q_\varphi^\infty(ds)$ gives

$$\beta \geq R(\varphi) + \sum_{j=1}^{n_C} \lambda_j \left(V_j - C_j(\varphi)\right). \tag{B.19}$$

Since it is true for every $\varphi \in \Pi_{RS}$, it is also true for $\zeta_{\boldsymbol{\lambda}}$ hence

$$\beta \geq R(\zeta_{\boldsymbol{\lambda}}) + \sum_{j=1}^{n_C} \lambda_j \left(V_j - C_j(\zeta_{\boldsymbol{\lambda}})\right) = \beta_{\boldsymbol{\lambda}}. \tag{B.20}$$

Taking now the infimum over all admissible triplet $(u, \beta, \boldsymbol{\lambda})$ gives that $\beta^\star \geq \beta_{\boldsymbol{\lambda}^\star}$.

B.4 Proof of the convergence of discrete approximation

To use properties shown for the unconstrained case in the constrained case, we will use Proposition 2.1 and Proposition 2.2. Proposition 2.1 shows that solving the infinite dimensional linear programming (2.50) is equivalent to solving the unconstrained MDP

$$(\mathcal{S}, \mathcal{A}, \mathcal{U}, Q, r + \sum \lambda_j (V_j - c_j))$$

and taking the infimum over all values of $\boldsymbol{\lambda}$. For every $\boldsymbol{\lambda}$, let $L(\boldsymbol{\lambda})$ be defined as

$$L(\boldsymbol{\lambda}) = \sup_{\varphi \in \Pi_{RS}} \left(R(\varphi) + \sum_{j=1}^{n_c} \lambda_j \left(V_j - C_j(\varphi)\right) \right). \tag{B.21}$$

If Assumption 2.2 holds, Proposition 2.2 shows that there exists λ_M such that $\boldsymbol{\lambda}^\star \leq \lambda_M$. This result implies that

$$R^\star = \inf_{\mathbf{0} \leq \boldsymbol{\lambda} \leq \lambda_M} L(\boldsymbol{\lambda}). \tag{B.22}$$

Consider now the discretized problem

$$(\mathcal{S}, \tilde{A}_h, \tilde{U}_h, \tilde{Q}_h, \tilde{r}_h + \sum \lambda_j (V_j - \tilde{c}_{j,h}))$$

and let $\tilde{L}_h$ be defined as

$$\tilde{L}_h(\boldsymbol{\lambda}) = \sup_{\varphi \in \tilde{\Pi}_{RS}} \left(\tilde{R}_h(\varphi) + \sum_{j=1}^{n_c} \lambda_j \left(V_j - \tilde{C}_{j,h}(\varphi) \right) \right). \tag{B.23}$$

$\tilde{L}_h(\boldsymbol{\lambda})$ is related to the optimal value of the discretized CMDP $(\mathcal{S}, \tilde{A}_h, \tilde{U}_h, \tilde{Q}_h, \tilde{r}_h, \tilde{\mathbf{c}}_h)$ by the following relationship:

$$\tilde{R}_h^\star = \inf_{\boldsymbol{\lambda} \geq \mathbf{0}} \tilde{L}_h(\boldsymbol{\lambda}). \tag{B.24}$$

Suppose that Assumption 2.2 holds for the CMDP $(\mathcal{S}, \tilde{A}_h, \tilde{U}_h, \tilde{Q}_h, \tilde{r}_h, \tilde{\mathbf{c}}_h)$. Proposition 2.2 shows that there exists $\tilde{\lambda}_M$ such that $\tilde{\boldsymbol{\lambda}}^\star \leq \tilde{\lambda}_M$. This result implies that

$$\tilde{R}_h^\star = \inf_{\mathbf{0} \leq \boldsymbol{\lambda} \leq \tilde{\lambda}_M} \tilde{L}_h(\boldsymbol{\lambda}). \tag{B.25}$$

From equations (B.22) and (B.25) we have

$$\begin{cases} R^\star = \inf_{\mathbf{0} \leq \boldsymbol{\lambda} \leq \lambda'_M} L(\boldsymbol{\lambda}). \\ \tilde{R}_h^\star = \inf_{\mathbf{0} \leq \boldsymbol{\lambda} \leq \lambda'_M} L(\boldsymbol{\lambda}). \end{cases} \tag{B.26}$$

where $\lambda'_M = \max\left(\lambda_M, \tilde{\lambda}_M\right)$. Also we have that for every $\boldsymbol{\lambda}$ such that $\boldsymbol{\lambda} \leq \lambda'_M$ the function $r + \sum \lambda_j (V_j - c_j)$ is $K(1 + n_c \lambda'_M)$-Lipschitz continuous. Using the result of [Chow, 1989] we have that there exist a constant K' such that for every $\boldsymbol{\lambda}$ such that $\boldsymbol{\lambda} \leq \lambda'_M$

$$L(\boldsymbol{\lambda}) - K'h \leq \tilde{L}_h(\boldsymbol{\lambda}) \leq L(\boldsymbol{\lambda}) + K'h \tag{B.27}$$

which, in turn implies

$$R^\star - K'h \leq \tilde{R}_h^\star \leq R^\star + K'h. \tag{B.28}$$

B.5 Proof of equation (2.91)

Let X and Y be some random variables. For the sake of readability we use the notation $f(x, y)$ instead of $f_{X,Y}(x, y)$ and similarly $f(x|y)$ instead of $f_{X|Y}(x|y)$. We now prove equation (2.91).

$b_{x,t}(x_t)$ is the a posteriori distribution of X_t given $H_t = (H_{t-1}, A_{t-1}, Y_t)$. More formally, $b_{x,t}(x_t)$ is defined as

$$b_{x,t}(x_t) = f\left(x_t | h_{t-1}, a_{t-1}, y_t\right). \tag{B.29}$$

By definition of the conditioning on y_t, we have

$$b_{x,t}(x_t) = \frac{f\left(x_t, y_t|h_{t-1}, a_{t-1}\right)}{f\left(y_t|h_{t-1}, a_{t-1}\right)}. \tag{B.30}$$

By marginalizing over X_{t-1} we rewrite $f\left(x_t, y_t|h_{t-1}, a_{t-1}\right)$ as follows:

$$\begin{aligned} f\left(x_t, y_t|h_{t-1}, a_{t-1}\right) &= \int_{\mathcal{X}} f\left(x_t, y_t, x_{t-1}|h_{t-1}, a_{t-1}\right) dx_{t-1} \qquad \text{(B.31)}\\ &= \int_{\mathcal{X}} f\left(x_t, y_t|x_{t-1}, h_{t-1}, a_{t-1}\right) f(x_{t-1}|h_{t-1}, a_{t-1}) dx_{t-1} \qquad \text{(B.32)} \end{aligned}$$

The same reasoning can be done to write $f\left(y_t|h_{t-1}, a_{t-1}\right)$ as follows:

$$f\left(y_t|h_{t-1}, a_{t-1}\right) = \int_{\mathcal{X}} f\left(y_t|x_{t-1}, h_{t-1}, a_{t-1}\right) f\left(x_{t-1}|h_{t-1}, a_{t-1}\right) dx_{t-1} \tag{B.33}$$

Since X_{t-1} is non-observable and since A_{t-1} depends on the observable history H_{t-1} only, A_{t-1} and X_{t-1} are independent. This implies that $f\left(x_{t-1}|h_{t-1}, a_{t-1}\right) = f\left(x_{t-1}|h_{t-1}\right)$. Using the definitions of $b_{x,t-1}$ and q we have $b_{x,t-1} = f\left(x_{t-1}|h_{t-1}\right)$ and $q\left(x_t, y_t|x_{t-1}, y_{t-1}, a_{t-1}\right) = f\left(x_t, y_t|x_{t-1}, h_{t-1}, a_{t-1}\right)$. Finally, $b_{x,t}$ is obtained from $b_{x,t-1}$, y_{t-1}, a_{t-1} and y_t thanks to the following relationship:

$$b_{x,t}(x_t) = \frac{\int_{\mathcal{X}} q\left(x_t, y_t|x_{t-1}, y_{t-1}, a_{t-1}\right) b_{x,t-1}(x_{t-1}) dx_{t-1}}{\int_{\mathcal{X}}\int_{\mathcal{X}} q\left(x'_t, y_t|y_{t-1}, a_{t-1}, x_{t-1}\right) b_{x,t-1}\left(x_{t-1}\right) dx_{t-1} dx'_t}. \tag{B.34}$$

B.6 Proof of Theorem 2.6

To prove Theorem 2.6, we will show that Condition 2.2(e) implies that for every $z \in \mathcal{Z}$ and every $a \in \mathcal{A}$, Q_Z verifies that for every $B \in \mathcal{B}(\mathcal{P}(\mathcal{X}))$ and every $C \in \mathcal{B}(\mathcal{Y})$,

$$Q_Z(B \times C|z, a) \geq \nu_Y(C \cap C_0)\delta_{m_0}(B). \tag{B.35}$$

This will prove that Condition 2.1(b) holds and in consequence that Assumption 2.1(e) holds.

We now show equation (B.35). From equation (2.95) and using C_0 defined in Condition $2.2(e) - i)$, we have that for every $B \in \mathcal{B}(\mathcal{P}(\mathcal{X}))$ and every $C \in \mathcal{B}(\mathcal{Y})$:

$$\begin{aligned} Q_Z(B \times C|z_{t-1}, a_{t-1}) &\geq Q_Z(B \times (C \cap C_0)|z_{t-1}, a_{t-1})\\ &= \int_{C \cap C_0} \mathbb{1}\left(H_b(z_{t-1}, a_{t-1}, y_t) \in B\right) Q'_Y\left(dy_t|z_{t-1}, a_{t-1}\right) \end{aligned}$$

Because $C \cap C_0 \subset C_0$ and because of the Condition $2.2(e) - ii)$, we have that for every $y \in C \cap C_0$, $H_b(z, a, y) = m_0$ independently of z and a. Therefore, we have

$$\begin{aligned} Q_Z(B \times C | z_{t-1}, a_{t-1}) &\geq \delta_{m_0}(B) \int_{C \cap C_0} Q'_Y \left(dy_t | z_{t-1}, a_{t-1} \right) \\ &\geq \delta_{m_0}(B) \nu_Y (C \cap C_0), \end{aligned} \tag{B.36}$$

where the second inequality is obtained by using Condition 2.2(*e*) in equation (2.93). This in turn implies that Q_Z verifies Assumption 2.1(e).

B.7 Proof of Remark 2.2

Let $s_0 \in \mathcal{S}$ be a state such that for every $s \in \mathcal{S}$ and every $a \in \mathcal{A}$, $Q(\{s_0\} | s, a) > \epsilon$ and $Q_X(dx'|x, y, a, y_0) = \delta_{x_0}(dx')$. From the definition of Q_Z we have

$$\begin{aligned} Q_Z(B \times C | z_{t-1}, a_{t-1}) &\geq Q_Z(B \times (C \cap \{s_0\}) | z_{t-1}, a_{t-1}) \\ &= \int_{C \cap \{y_0\}} \mathbb{1} \left(H_b(z_{t-1}, a_{t-1}, y_t) \in B \right) Q'_Y \left(dy_t | z_{t-1}, a_{t-1} \right) \\ &= \delta_{y_0}(C) \mathbb{1} \left(H_b(z_{t-1}, a_{t-1}, y_t) \in B \right) Q'_Y \left(\{y_0\} | z_{t-1}, a_{t-1} \right) \end{aligned}$$

Moreover using the disintegration of Q we have that

$$Q(\{s_0\} | s, a) = Q_X(\{x_0\} | s, a, y_0) Q_Y(\{y_0\} | s, a),$$

using now the assumptions $Q_X(dx'|x, y, a, y_0) = \delta_{x_0}(dx')$ and $Q(\{s_0\} | s, a) > \epsilon$ proves that $Q_Y(\{y_0\} | s, a) > \epsilon$ which in turn proves that $Q'_Y(\{y_0\} | s, a) > \epsilon$. This finally proves that

$$Q_Z(B \times C | z_{t-1}, a_{t-1}) \geq \epsilon \delta_{y_0}(C) \delta_{m_0}(B),$$

where $m_0 = \delta_{x_0}$. To complete the proof note that $\delta_{y_0}(\mathcal{Y}) = 1$.

- Appendix C -

Appendices of Chapter 3

C.1 Lipschitz continuity of Q, the Type-I HARQ case

In this appendix, we show that in the case of Rayleigh fading channel, Assumption 2.3(d) holds.

For two probability measures μ_1 and μ_2 on $\mathcal{S}$ (which is a discrete space), it is shown in [Hernández-Lerma, 1989] that

$$||\mu_1 - \mu_2||_{TV} = \sum_{k \in \mathcal{S}} |\mu_1(k) - \mu_2(k)| \,. \tag{C.1}$$

We now apply this result to $\mu_1 = Q(\cdot|s,a)$ and $\mu_2 = Q(\cdot|s',a')$ expressed thanks to equations (3.1) and (3.2).

In the case $s \neq s'$ we have

$$||Q(\cdot|s,a) - Q(\cdot|s',a')||_{TV} \leq 2 \leq 2\,||(s,a) - (s',a')||_{\infty} \,. \tag{C.2}$$

The first inequality comes from triangular inequality and the second inequality comes from the fact that if $s \neq s'$, $||(s,a) - (s',a')||_{\infty} \geq 1$.

In the case $s \neq s'$ we have by direct calculations that

$$||Q(\cdot|s,a) - Q(\cdot|s',a')||_{TV} = 2\,|p_o(a) - p_o(a')| \tag{C.3}$$

By direct calculations on the expression of p_o in the case of Rayleigh fading (this expression is given in equation (3.2)), one can easily show that this function is K-Lipschitz continuous for some K. This proves that

$$||Q(\cdot|s,a) - Q(\cdot|s',a')||_{TV} = 2K\,|(s,a) - (s',a')| \,. \tag{C.4}$$

In consequence, Assumption 2.3(d) holds taking $K' = \max(2, 2K)$.

C.2 Lipschitz continuity of Q, the Type-II HARQ case

We focus here on the IR-HARQ case, but this proof is the same (with obvious changes) for the case of CC-HARQ. Remark first that the definition of the total variation norm given in (2.29) implies that $||p_1 - p_2||_{TV} \leq 2$. This implies that if $u = ((k,i),a)$ and $u' = ((k',i'),a')$ with $k \neq k'$, we have

$$||u - u'||_\infty \geq 1 \geq \frac{1}{2} ||Q(\cdot|u) - Q(\cdot|u')||_{TV} \,. \tag{C.5}$$

We consider first the case $k = k'$. Using the definition of Q of equation (A.3) and considering arbitrarily that $i' \geq i$, we bound $||Q(\cdot|u) - Q(\cdot|u')||_{TV}$ using the triangle inequality as follows:

$$\begin{aligned} ||Q(\cdot|u) - Q(\cdot|u')||_{TV} \leq \left|F_{\Delta|A}(R - i'|a') - F_{\Delta|A}(R - i|a)\right| \\ + \int_i^{i'} f_{\Delta|A}(x - i|a)dx \\ + \int_{i'}^{R} \left|f_{\Delta|A}(x - i|a) - f_{\Delta|A}(x - i'|a)\right| dx, \end{aligned} \tag{C.6}$$

where $F_{\Delta|A}$ and $f_{\Delta|A}$ are defined in equations (A.6) and (A.7). Because we have assumed that $\mathcal{A} = [P_{min}P_{max}]$ with $P_{min} > \epsilon$ for some $\epsilon > 0$, it follows from direct calculations from (A.6) and (A.7), that the function $(x,a) \mapsto F_{\Delta|A}(x|a)$ is K-Lipschitz for some K, the function $(x,a) \mapsto f_{\Delta|A}(x,a)$ is bounded by some constant K' and is K''-Lipschitz. From these observations it follows that

$$||Q(\cdot|u) - Q(\cdot|u')||_{TV} \leq (K + K' + r_1 K'')\, ||u - u'||_\infty \,.$$

To conclude the proof, it suffices to take $K_4 = \max(2, K + K' + r_1 K'')$.

- Appendix D -

Appendices of Chapter 4

D.1 Expression of $Q(B|s_n, p_n)$

$$
\begin{aligned}
Q(B|s_n,p_n) &= \mathbb{P}\left[S_{n+1} \in B | S_n = s_n, P_n = p_n\right] && \text{(D.1)}\\
&= \int_{\mathcal{S}} \mathbb{1}_B(s_{n+1}) Q(ds_{n+1}|s_n,p_n), && \text{(D.2)}
\end{aligned}
$$

where $\mathbb{1}_B(s_{n+1})$ is the function that is equal to 1 if $s_{n+1} \in B$ and 0 otherwise. In our case, the *transition kernel* has the following expression:

$$
Q(ds_{n+1}|s_n,p_n) = \begin{cases}
\underline{\text{If } k_n < N_T - 1} : \\
\mathbb{P}\left[i_n + \Delta\left(w_n, p_{2,n}\right) \geq R | p_{2,n}\right] \delta_{(0,0)}(ds_{n+1}) \\
\qquad\qquad + f_{\Delta|P}(i_{n+1} - i_n | p_{2,n}) \delta_{k_n+1}(dk_{n+1}) \ell(di_{n+1}) \\
\underline{\text{If } k_n = N_T - 1} : \\
\mathbb{P}\left[i_n + \Delta\left(w_n, p_{2,n}\right) \geq R | p_{2,n}\right] \delta_{(0,0)}(ds_{n+1}) \\
\qquad\qquad + \mathbb{P}\left[i_n + \Delta\left(w_n, p_{2,n}\right) < D_T | p_{2,n}\right] \delta_{(N_T,0)}(ds_{n+1}) \\
\underline{\text{If } k_n = N_T :} \\
\mathbb{P}\left[i_n + \Delta\left(w_n, p_{2,n}\right) \geq R | p_{2,n}\right] \delta_{(0,0)}(ds_{n+1}) \\
\qquad\qquad + f_{\Delta|P}(i_{n+1} | p_{2,n}) \delta_1(dk_{n+1}) \ell(di_{n+1})
\end{cases} \tag{D.3}
$$

By definition, we have $\mathbb{P}\left[i_n + \Delta\left(w_n, p_{2,n}\right) \geq R | p_{2,n}\right] = 1 - F_{\Delta|P}(R - i_n | p_{2,n})$, where, for Rayleigh channels, we have that α_{ij} is exponentially distributed with mean $\bar{\alpha}_{ij}$. After tedious computations we get, $F_{\Delta|P}(\delta|p_2)$:

$$F_{\Delta|P}(x|p_2) = \mathbb{P}\left[\log_2\left(1+\frac{\alpha_{11}p_1}{1+\alpha_{21}p_2}\right) \leq x|p_2\right]$$
$$= \begin{cases} \dfrac{\left(1-\exp\left[-\frac{2^x-1}{p_1\bar{\alpha}_{11}}\right]\right)p_1\bar{\alpha}_{11}+(2^x-1)\,p_2\bar{\alpha}_{21}}{p_1\bar{\alpha}_{11}+(2^x-1)\,p_2\bar{\alpha}_{21}}, & \text{if } x \geq 0 \\ 0, \text{otherwise}, \end{cases} \tag{D.4}$$

where the time indexes has been dropped. Let also $f_{\Delta|P}(x|p_2)$ be the corresponding to pdf,

$$f_{\Delta|P}(x|p_2) = \frac{\partial F_{\Delta|P}}{\partial x}(x,p_2)$$
$$= \begin{cases} \dfrac{2^x\exp\left[-\frac{2^x-1}{p_1\bar{\alpha}_{11}}\right]\log(2)\left(p_1\bar{\alpha}_{11}\left(1+p_2\bar{\alpha}_{21}\right)+(2^x-1)\,p_2\bar{\alpha}_{11}\right)}{\left(p_1\bar{\alpha}_{11}+(2^x-1)\,p_2\bar{\alpha}_{12}\right)^2}, \text{if } x \geq 0 \\ 0, \text{ otherwise}. \end{cases} \tag{D.5}$$

D.2 Proof of Lipschitz continuity $Q(\cdot|s_n, p_n)$

The proof of this Lipschitz continuity property is exactly the same as the one proposed in Appendix C.2 except that we consider the functions $F_{\Delta|P}(x|p_2)$ and $f_{\Delta|P}(x,p_2)$ from equations (D.4) and (D.5).

D.3 Proof of Lipschitz continuity of r_2

In this appendix we prove that the function defined as

$$r_{2,n} = r_2(p_{2,n}) = \max_{r\in\mathbb{R}^+} r\mathbb{P}\left[\log_2\left(1+p_{2,n}\beta_{2,n}\right) \geq r\right]. \tag{D.6}$$

is Lipschitz continuous.

We first study the function defined as follows

$$v(r,p_2) = r\mathbb{P}\left[\log_2\left(1+p_2\beta_2\right) \geq r\right]. \tag{D.7}$$

After some tedious calculations, $v(r,p_2)$ can be expressed as follows:

$$v(r,p_2) = r\frac{p_2\bar{\alpha}_{22}\exp\left[-\frac{2^r-1}{p_2\bar{\alpha}_{22}}\right]}{p_2\bar{\alpha}_{22}+(2^r-1)\,p_1\bar{\alpha}_{12}}. \tag{D.8}$$

By direct computation of the derivatives of $v(r, p_2)$, we show that this function is bounded and Lipschitz continuous for some K. For every $p_2, p_2' \in \mathcal{A}$, we have

$$|r_2(p_2) - r_2(p_2')| = |v(\bar{r}(p_2), p_2) - v(\bar{r}(p_2'), p_2')| \,, \tag{D.9}$$

where $\bar{r} = \arg\max_{r \in \mathbb{R}^+} r\mathbb{P}\left[\log_2\left(1 + p_2\beta_2\right) \geq r\right]$. We suppose, without loss of generality that $p_2 \geq p_2'$, remarking that, from its definition, the function $r_2(p_2)$ is an increasing in p_2, we get

$$\begin{aligned} r_2(p_2) - r_2(p_2') &= v(\bar{r}(p_2), p_2) - v(\bar{r}(p_2'), p_2'), \\ &\leq v(\bar{r}(p_2'), p_2) - v(\bar{r}(p_2'), p_2'), \\ &\leq K\left(p_2 - p_2'\right), \end{aligned} \tag{D.10}$$

where the second inequality comes from the fact that $r_2(p_2) \geq v(\bar{r}(p_2'), p_2)$ and the third inequality comes from the fact that the function $v(r, p)$ defined by equation (D.7) is K-Lipschitz.

Bibliography

Aberdeen, D. (2003). *A (Revised) Survey of Approximate Methods for Solving Partially Observable Markov Decision Processes*. Technical report, rep., National ICT Australia. 54

Altman, E. (1999). *Constrained Markov decision processes*. Stochastic modeling. Chapman & Hall/CRC. 36, 37, 39, 40, 44, 48, 52

Anderson, E. & Nash, P. (1987). *Linear programming in infinite-dimensional spaces: theory and applicationsa*. Wiley-Interscience series in discrete mathematics and optimization. Wiley. 43, 46, 47

Aoun, M. E. (2012). *Optimisation des techniques de codage et de retransmission pour les systemèmes radio avec voie de retour*. PhD thesis, Télécom Bretagne. 16

Arapostathis, A., Borkar, V. S., Fernández-Gaucherand, E., Ghosh, M. K., & Marcus, S. I. (1993). Discrete-time controlled markov processes with average cost criterion: a survey. *SIAM J. Control Optim.*, 31(2), 282–344. 54, 58

Asghari, V. & Aissa, S. (2010). Adaptive rate and power transmission in spectrum-sharing systems. *IEEE Trans. on Wireless Commu.*, 9(10), 3272–3280. 90

Bagayoko, A., Fijalkow, I., & Tortelier, P. (2011). Power control of spectrum-sharing in fading environment with partial channel state information. *IEEE Trans. Signal Process.*, 59(5), 2244–2256. 90

Bagayoko, A., Tortelier, P., & Fijalkow, I. (2010). Spectrum-sharing power control with outage performance requirements and direct links csi only. In *Proc. IEEE 21st Int Personal Indoor and Mobile Radio Communications (PIMRC) Symp* (pp. 651–655). 90

Bellman, R. (1952). On the theory of dynamic programming. *Proceedings of the National Academy of Sciences of the United States of America*, 38(8), 716. 33

Bellman, R. (1957). *Dynamic Programming*. Princeton, NJ, USA: Princeton University Press, 1 edition. 33

Bertsekas, D. & Shreve, S. (1978). *Stochastic optimal control: the discrete time case*. Mathematics in science and engineering. Academic Press. 35, 37, 54, 57, 119

Bertsekas, D. P. (2001). *Dynamic Programming and Optimal Control, Two Volume Set*. Athena Scientific, 2nd edition. 33

Brennan, D. G. (1959). Linear diversity combining techniques. *Proceedings of the IRE*, 47(6), 1075–1102. 19

Buckingham, D. & Valenti, M. (2008). The information-outage probability of finite-length codes over awgn channels. In *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on* (pp. 390–395). 8

Caire, G. & Tuninetti, D. (2001). The throughput of hybrid-arq protocols for the gaussian collision channel. *IEEE Trans. Inform. Theory*, 47(5), 1971–1988. 8, 11, 12, 13, 15

Chaitanya, T. & Larsson, E. (2011). Outage-optimal power allocation for hybrid arq with incremental redundancy. *IEEE Transactions on Wireless Communications*, 10(7), 2069–2074. 23

Chaitanya, T. & Larsson, E. (2013). Optimal power allocation for hybrid arq with chase combining in i.i.d. rayleigh fading channels. *IEEE Transactions on Communications*, 61(5), 1835–1846. 23

Cheng, J. (2006). Coding performance of hybrid arq schemes. *IEEE Trans. Commun.*, 54(6), 1017–1029. 8, 12, 20, 22

Cheng, J.-F., Wang, Y.-P., & Parkvall, S. (2003). Adaptive incremental redundancy [wcdma systems]. In *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, volume 2 (pp. 737–741 Vol.2). 2, 8, 66

Chow, C. (1989). *Multigrid algorithms and complexity results for discrete-time stochastic control and related fixed-point problems*. PhD thesis, Massachusetts Institute of Technology. xi, 49, 50, 51, 52, 62, 123

Cover, T. M. & Thomas, J. A. (2006). *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience. 12

Djonin, D. & Krishnamurthy, V. (2007). Mimo transmission control in fading channels; a constrained markov decision process formulation with monotone randomized policies. *Signal Processing, IEEE Transactions on*, 55(10), 5069–5083. 33

Eswaran, K., Gastpar, M., & Ramchandran, K. (2007). Bits through arqs: Spectrum sharing with a primary packet system. In *Proceedings of the 2007 IEEE International Symposium on Information Theory (ISIT 2007)* Nice, France. 2, 4, 90, 106

Etkin, R., Parekh, A., & Tse, D. (2007). Spectrum sharing for unlicensed bands. *Selected Areas in Communications, IEEE Journal on*, 25(3), 517–528. 90

FCC (2002). *Report of the Spectrum Efficiency Working Group.* Technical report, FCC. 2, 89

Feinberg, E. & Shwartz, A. (2002). *Handbook of Markov Decision Processes: Methods and Applications*. International series in operations research & management science. Kluwer Academic Publishers. 33

Goldsmith, A. (2005). *Wireless Communications*. New York, NY, USA: Cambridge University Press. 11

Haykin, S. (2005). Cognitive radio: brain-empowered wireless communications. *IEEE J. Select. Areas Commun.*, 23(2), 201–220. 90

Hernández-Lerma, O. (1989). *Adaptive Markov control processes*. Applied mathematical sciences. Springer-Verlag. 36, 37, 38, 39, 40, 54, 57, 58, 120, 127

Hernández-Lerma, O., González-Hernández, J., & López-Martínez, R. R. (2003). Constrained average cost markov control processes in borel spaces. *SIAM J. Control Optim.*, 42(2), 442–468. 39, 43, 44, 47

Hernández-Lerma, O. & Lasserre, J.-B. (1996). *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Number vol. 1 in Applications of Mathematics Stochastic Modelling and Applied Probability. Springer Verlag. 37, 39, 42, 43, 44, 45, 47, 59, 119

Hernández-Lerma, O. & Lasserre, J.-B. (1999). *Further Topics on Discrete-time Markov Control Processes*. Number vol. 2 in Applications of Mathematics Series. Springer Verlag. 37, 39

Isom, J. D., Meyn, S. P., & Braatz, R. D. (2008). Piecewise linear dynamic programming for constrained pomdps. In *AAAI* (pp. 291–296). 54

Jovicic, A. & Viswanath, P. (2009). Cognitive radio: An information-theoretic perspective. *IEEE Trans. Inform. Theory*, 55(9), 3945–3958. 90, 91

Kang, X., Liang, Y.-C., Garg, H., & Zhang, L. (2009). Sensing-based spectrum sharing in cognitive radio networks. *Vehicular Technology, IEEE Transactions on*, 58(8), 4649–4654. 90

Karmokar, A., Djonin, D., & Bhargava, V. (2006). POMDP-based coding rate adaptation for type-i hybrid arq systems over fading channels with memory. *IEEE Trans. Wireless Commun.*, 5(12), 3512 –3523. 33

Kay, S. M. (1993). *Fundamentals of statistical signal processing. [Volume 1]. , estimation theory Texte imprimé.* Prentice Hall signal processing series. Englewood Cliffs (N.J.): Prentice Hall PTR. 84

Kim, D., Lee, J., Kim, K.-E., & Poupart, P. (2011). Point-based value iteration for constrained pomdps. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Three* (pp. 1968–1974).: AAAI Press. 54

Knopp, R. & Humblet, P. (2000). On coding for block fading channels. *IEEE Trans. Inform. Theory*, 46(1), 189–205. 11

Kurano, M., Nakagami, J.-I., & Huang, Y. (2000). Constrained markov decision processes with compact state and action spaces: the average case. *Optimization*, 48(2), 255–269. 39, 40, 44, 46

Le Duc, A. (2010). *Performance Closed-form Derivations and Analysis of Hybrid ARQ Retransmission Schemes in a Cross-layer Context.* PhD thesis, Télécom ParisTech. 13, 15, 19

Levorato, M., Mitra, U., & Zorzi, M. (2009). Cognitive interference management in retransmission-based wireless networks. In *Proc. 47th Annual Allerton Conf. Communication, Control, and Computing Allerton 2009* (pp. 94–101). 29, 33

Levorato, M., Mitra, U., & Zorzi, M. (2012). Cognitive interference management in retransmission-based wireless networks. *IEEE Trans. Inform. Theory*, 58(5), 3023–3046. 33, 90, 106

Li, Y. & Ryan, W. (2007). Mutual-information-based adaptive bit-loading algorithms for ldpc-coded ofdm. *Wireless Communications, IEEE Transactions on*, 6(5), 1670–1680. 8, 66

Lin, S. & Costello, D. J. (2004). *Error control coding*, volume 123. Prentice-hall Englewood Cliffs. 14, 16

Makki, B. & Eriksson, T. (2012). On the average rate of harq-based quasi-static spectrum sharing networks. *Wireless Communications, IEEE Transactions on*, 11(1), 65–77. 90

Makki, B., Graell i Amat, A., & Eriksson, T. (2012). Harq feedback in spectrum sharing networks. *Communications Letters, IEEE*, 16(9), 1337–1340. 90

Makki, B., Graell i Amat, A., & Eriksson, T. (2013). Green communication via power-optimized harq protocols. *Vehicular Technology, IEEE Transactions on*, PP(99), 1–1. 90

Manne, A. S. (1960). Linear programming and sequential decisions. *Management Science*, 6(3), pp. 259–267. 33

Marcille, S. (2013). *Allocation de ressources pour les réseaux ad hoc mobiles basés sur les protocoles HARQ*. PhD thesis, Télécom ParisTech. 8, 15, 16

Marcille, S., Ciblat, P., & Le Martret, C. (2012a). Optimal resource allocation in harq-based ofdma wireless networks. In *MILITARY COMMUNICATIONS CONFERENCE, 2012 - MILCOM 2012* (pp. 1–6). 24

Marcille, S., Ciblat, P., & Le Martret, C. J. (2011). Performance computation of cross-layer hybrid arq schemes at ip layer in the presence of corrupted acknowledgments. In *Cross Layer Design (IWCLD), 2011 Third International Workshop on* (pp. 1–5).: IEEE. 16

Marcille, S., Ciblat, P., & Le Martret, C. J. (2012b). A cross-layer harq scheme robust to imperfect feedback. In *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on* (pp. 143–147).: IEEE. 16

Masmoudi, R., Belmega, E., Fijalkow, I., & Sellami, N. (2012). A closed-form solution to the power minimization problem over two orthogonal frequency bands under qos and cognitive radio interference constraints. In *Dynamic Spectrum Access Networks (DYSPAN), 2012 IEEE International Symposium on* (pp. 212–222). 90

Meyn, S. & Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. New York, NY, USA: Cambridge University Press, 2nd edition. 27, 40

Michelusi, N., Popovski, P., Simeone, O., Levorato, M., & Zorzi, M. (2013a). Cognitive access policies under a primary arq process via forward-backward interference cancellation. *IEEE J. Select. Areas Commun.*, PP(99), 1–13. 33, 108, 109

Michelusi, N., Popovski, P., & Zorzi, M. (2013b). Cognitive access policies under a primary arq process via chain decoding. In *Information Theory and Applications Workshop (ITA), 2013* (pp. 1–8). 33, 108, 109

Mitola III, J. & Maguire, G.Q., J. (1999). Cognitive radio: making software radios more personal. *IEEE Pers. Commun.*, 6(4), 13 –18. 2, 90

Neely, M. (2011). Online fractional programming for markov decision systems. In *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on* (pp. 353–360). 113

Negi, R. & Cioffi, J. (2002). Delay-constrained capacity with causal feedback. *IEEE Transactions on Information Theory*, 48(9), 2478–2494. 33

Pfletschinger, S. & Navarro, M. (2010). Versatile link adaptation based on mutual information. In *Communications (ICC), 2010 IEEE International Conference on* (pp. 1–6). 8, 66

Puterman, M. L. (1994). *Markov decision processes: discrete stochastic dynamic programming*. Wiley series in probability and statistics. Wiley-Interscience. 33

Ross, K. W. (1989). Randomized and past-dependent policies for markov decision processes with multiple constraints. *Operations Research*, 37(3), pp. 474–477. 108

Ross, S. M. (2006). *Introduction to probability models*. Access Online via Elsevier. 14

Roy, N., Gordon, G. J., & Thrun, S. (2005). Finding approximate pomdp solutions through belief compression. *J. Artif. Intell. Res.(JAIR)*, 23, 1–40. 54

Sesia, S., Caire, G., & Vivier, G. (2004). Incremental redundancy hybrid arq schemes based on low-density parity-check codes. *IEEE Trans. Commun.*, 52(8), 1311–1321. 8, 11, 15

Sesia, S., Toufik, I., & Baker, M. (2011). *LTE - The UMTS Long Term Evolution: From Theory to Practice*. Wiley. 1

Stiglmayr, S., Bossert, M., & Costa, E. (2007). Adaptive coding and modulation in OFDM systems using BICM and rate-compatible punctured codes. In *Proc. European Wireless Conf.* Paris. 8, 66

Stiglmayry, S., Klotz, J., & Bossert, M. (2008). Adaptive coding and modulation in mimo ofdma systems. In *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on* (pp. 1–5). 8

Stupia, I., Giannetti, F., Lottici, V., & Vandendorpe, L. (2009). A novel link performance prediction method for coded mimo-ofdm systems. In *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE* (pp. 1–5). 66

Szczecinski, L., Duhamel, P., & Rahman, M. (2011). Adaptive incremental redundancy for harq transmission with outdated csi. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE* (pp. 1–6). xii, 11, 29, 67, 86, 87, 88

Tajan, R., Poulliat, C., & Fijalkow, I. (2012). Opportunistic secondary spectrum sharing protocols for primary implementing an IR type Hybrid-ARQ protocol (regular paper). In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Kyoto, 25/03/12-30/03/12* (pp. (electronic medium)). http://www.ieee.org/: IEEE. 90

Tannious, R. A. & Nosratinia, A. (2010). Cognitive radio protocols based on exploiting hybrid arq retransmissions. *IEEE Trans. Wireless Commun.*, 9(9), 2833–2841. 90

Theocharous, G. & Mahadevan, S. (2002). Approximate planning with hierarchical partially observable markov decision process models for robot navigation. In *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, volume 2 (pp. 1347–1352). 33

Tse, D. & Viswanath, P. (2005). *Fundamentals of Wireless Communication*. Cam. 10

Tuninetti, D. (2011). On the benefits of partial channel state information for repetition protocols in block fading channels. *IEEE Transactions on Information Theory*, 57(8), 5036–5053. 11, 29, 33, 67

Wan, L., Tsai, S., & Almgren, M. (2006). A fading-insensitive performance metric for a unified link quality model. In *Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE*, volume 4 (pp. 2110–2114). 66

Yu, H. & Bertsekas, D. P. (2004). Discretized approximations for pomdp with average cost. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence* (pp. 619–627).: AUAI Press. 54, 62

Zhang, R. (2008). Optimal power control over fading cognitive radio channel by exploiting primary user csi. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE* (pp. 1 –5). 90

Zhang, R. (2010). On active learning and supervised transmission of spectrum sharing based cognitive radios by exploiting hidden primary radio feedback. *IEEE Trans. Commun.*, 58(10), 2960–2970. 90

Zhou, E., Fu, M. C., & Marcus, S. I. (2010). Solving continuous-state pomdps via density projection. *Automatic Control, IEEE Transactions on*, 55(5), 1101–1116. 54, 55, 59, 60, 85

Zorzi, M. & Rao, R. R. (1996). On the use of renewal theory in the analysis of arq protocols. *IEEE Transactions on Communications*, 44(9), 1077–1081. 14, 15, 17, 23

Résumé

Dans les standards actuels tels que HSDPA ou LTE, des protocoles de retransmissions (ARQ : Automatic Repeat reQuest) sont utilisés conjointement au codage de canal afin de palier aux erreurs dues à l'absence ou la mauvaise de connaissance de canal à la transmission. On garantit ainsi la fiabilité du lien physique pour les couches OSI supérieures (du moins un taux d'erreur paquet faible). De tels protocoles sont appelés protocoles de retransmission hybrides (HARQ). L'objectif de cette thèse est de proposer des outils permettant l'analyse et l'optimisation des systèmes de communication en présences de protocoles HARQ avec une emphase particulière sur les systèmes cognitifs. La radio cognitive est une approche permettant à des utilisateurs non-licenciés de communiquer dans les mêmes bandes de fréquences que des utilisateurs licenciés afin d'augmenter l'efficacité spectrale des réseaux sans fil. Les utilisateurs secondaires doivent néanmoins limiter les interférences générées sur les signaux des utilisateurs primaires. Dans ce contexte, nous étudierons les débits atteignables par un utilisateur secondaire utilisant l'observation du protocole HARQ de l'utilisateur primaire afin de contrôler son interférence.

Abstract

Automatic Repeat Request protocols (ARQ) are widely implemented in current mobile wireless standards such as HSDPA and LTE. In general, ARQ protocols are combined with channel coding to overcome errors caused by the lack of channel knowledge at the transmitter side. These protocols are called Hybrid ARQ protocols (HARQ). HARQ protocols ensure a good reliability (at least a small packet error rate) of the physical layer for the OSI upper layers. The purpose of this thesis is to provide tools for the analysis and the optimization of HARQ communication systems with an emphasis on cognitive systems. Cognitive Radio (CR) is an approach aiming to increase the spectral efficiency of wireless networks. In a CR context, unlicensed users are allowed to communicate within the same frequency bands and at the same time as licensed users. Secondary users must however limit the amount of interference generated on the primary users signals. In this thesis, we consider a scenario in which the secondary user interferes a primary user employing a HARQ protocol. When the secondary user knows the state of the primary HARQ protocol, we show that a joint power and rate allocation can be performed to limit the interference.