Universitat de les Illes Balears
Departament de Ciències Matemàtiques i Informàtica
Programa de Doctorat en Informàtica

Tesi Doctoral:

Cross-Layer Design for Quality of Service provisioning in AMC/ARQ-based wireless networks

Autor:

Jaume Ramis Bibiloni

DIRECTOR:

_____

Dr Guillem Femenias Nadal

DOCTORAND:

_____

Jaume Ramis Bibiloni

30 de Maig de 2012

Cross-Layer Design for Quality of Service provisioning in AMC/ARQ-based wireless networks

Autor: Jaume Ramis Bibiloni

Director: Guillem Femenias Nadal

Universitat de les Illes Balears

Cross-Layer Design for Quality of Service provisioning in AMC/ARQ-based wireless networks

Defended by
Jaume Ramis Bibiloni

A thesis submitted to *Departament de Ciències Matemàtiques i Informàtica* of the University of Balearic Islands in accordance with the requirements for the degree of
Doctor of Computer Science

Thesis Advisor
Dr. Guillem Femenias Nadal

May 2012

Cross-Layer Design for Quality of Service provisioning in AMC/ARQ-based wireless networks

Thesis Advisor: Guillem Femenias Nadal

Defended by: Jaume Ramis Bibiloni

El Dr. Guillem Femenias Nadal, Catedràtic d'Universitat del Departament de Ciències Matemàtiques i Informàtica de la Universitat de les Illes Balears

FA CONSTAR:

que la present memòria 'Cross-Layer Design for Quality of Service provisioning in AMC/ARQ-based wireless networks' presentada per en Jaume Ramis Bibiloni per optar al grau de Doctor en Informàtica, ha estat realitzada sota la seva direcció i reuneix la suficient matèria original per ser considerada com a tesi doctoral.

Palma, 30 de maig de 2012

Signat: Dr. Guillem Femenias Nadal

A la meva família, en especial a mon pare i a ma mare.
Gràcies de tot cor.

Vull agrair als meus companys de feina la seva ajuda i el seu suport. Sempre que els he necessitat m'han donat un cop de mà. No puc deixar sense anomenar ni a n'en Felip ni, per descomptat, a na Loren. Estic en deute amb ells pel seu assessorament, els seus consells i el seu inestimable ajut. I, sobretot, agraesc molt sincerament a n'en Guillem la seva generositat, dedicació i assistència, sense les quals aquest treball de recerca mai no hauria estat possible.
Gràcies a totes i a tots.

Summary

In the last years, the explosive development of wireless services and applications has produced an unprecedented revolution in wireless communications systems. The impact of wireless fading channels on the quality of service (QoS) provisioning for such heterogeneous mobile users is one of the most challenging issues for next-generation wireless networks.

In order to support the diverse QoS requirements of wireless applications, innovative techniques have been proposed at the physical (PHY) layer. Among them, the rate-adaptive modulation and coding (AMC) scheme has received significant research attention, resulting in its adoption in most state-of-the-art wireless communications standards.

With the aim of enhancing the link reliability to guarantee the QoS constraints of new applications, most modern communication systems use error control strategies at the data link control (DLC) layer. Among them, we can distinguish two basic approaches: the forward error correction (FEC) scheme, in which an error-correction code is used, and the automatic repeat request (ARQ) scheme, in which a code with good error-detection capability is used. In the latter case, when a received codeword is detected in error, the packet is retransmitted until it is correctly received (infinitely persistent ARQ) or until a preset number of retransmissions have taken place (truncated ARQ). In order to achieve the advantages of both strategies, most state-of-the-art wireless communications standards use combinations of ARQ and FEC. Moreover, recent proposals make use of one or multiple intermediate relay stations to forward data from a source node to the corresponding destination node. As a result, a performance improvement is obtained due to spatial diversity, which is generated by transmitting signals from different locations (source and relay(s)), thus providing independently faded versions of the signal at the receiver.

Cross-layer design in wireless networks, where one allows the stack protocol layers to interact and share information, has become increasingly popular over the past few years. In particular, many recent cross-layer proposals coincide in combining AMC at the PHY layer with an ARQ protocol at the DLC layer. These cross-layer designs improve the spectral efficiency by jointly exploiting the adaptability of AMC to the wireless channel conditions and the error-correcting capability of ARQ-based error control strategies.

The main goal of this dissertation is to provide a unified view of the cross-layer design, analysis and optimization of AMC/ARQ-based wireless systems to allow the joint optimization of both the PHY and the DLC layers. The adopted approach for tackling this problem will rely on the use of discrete time Markov

chains (DTMCs) to jointly consider the packet arrival, the queueing process and the PHY layer. To that end, a novel first-order two-dimensional Markov model of the PHY layer is developed, which takes into account the wireless channel characteristics and de AMC scheme. The availability of accurate models for the PHY layer characterization is one of the key motivations of this work in order to guarantee a correct analysis of the QoS metrics at the DLC layer. Using this model, the interactions between the PHY and DLC layers are analyzed either when infinitely persistent or truncated ARQ-based error control protocols are implemented. Additionally, a cooperative scheme is also proposed, in which the relay node is in charge of retransmitting the erroneously received packets at the destination. This DTMC-based model allows the analytic derivation of various system performance metrics, namely, throughput, average packet delay and packet loss rate (both due to buffer overflow and due to exceeding the maximum number of allowed retransmissions). For the sake of comparison with non Markov-based analytical tools, the infinitely persistent ARQ protocol is also analyzed through the effective bandwidth/capacity theory. Both analytical frameworks are compared, showing the superiority of the Markov-based approach, which more faithfully reproduces the real system behaviour at the cost of higher complexity in the analysis. The proposed analytical framework allows the formulation of cross-layer multidimensional design strategies, aiming at the maximization of the average throughput of the system while satisfying prescribed QoS requirements in the form of average packet loss rate and average delay. Finally, an explicit analysis of the impact on the system performance of the delay in the channel state information (CSI) feedback is also presented.

Resum

No hi ha cap dubte que en els darrers anys els serveis i les aplicacions en entorns de comunicacions sense fils han experimentat un desenvolupament sense precedents. Les demandes creixents en termes de taxa de transmissió i de mobilitat per part dels usuaris s'han d'afegir a la heterogeneïtat dels requeriments de qualitat de servei (QoS, de l'anglès *Quality of Service*) de les aplicacions. A tot això s'hi ha de sumar l'efecte dels esvaïments selectius, tant en temps com en freqüència, intrínsecs als canals sense fils. Aleshores, el desenvolupament d'aquest tipus de sistemes representa un vertader repte a l'hora de dissenyar solucions capaces de suportar les exigències de QoS, com per exemple taxa màxima d'error de paquet o bé retard màxim, tant per a les aplicacions mòbils actuals com per a les futures.

S'ha desenvolupat un gran nombre d'estratègies amb l'objectiu de fer front a l'impacte dels esvaïments sobre les prestacions d'aquests sistemes i aconseguir així millorar l'eficiència espectral i/o energètica de la capa física. Entre elles es poden destacar els esquemes de modulació i codificació adaptativa (AMC, de l'anglès *Adaptive Modulation and Coding*). La immensa majoria dels actuals estàndards de comunicacions sense fils han adoptat aquesta tècnica, que consisteix bàsicament en ajustar la modulació i la codificació a les condicions canviants del canal.

Per altra banda, els requeriments de QoS de les noves aplicacions tan sols es poden satisfer amb l'ajut d'esquemes que permeten millorar la fiabilitat de l'enllaç. Així, l'ús d'estratègies de control d'errors a la capa d'enllaç de dades (DLC, de l'anglès *Data Link Control*) s'ha convertit en una part fonamental dels sistemes de comunicacions actuals. Les tècniques de control d'errors es poden classificar bàsicament en dues categories. La primera la formen els esquemes que permeten la detecció i possible correcció d'errors (FEC, de l'anglès *Forward Error Correction*), mentre que la segona categoria correspon als esquemes que possibiliten la detecció d'errors i la sol·licitud de retransmissió (ARQ, de l'anglès *Automatic Repeat Request*). En aquest cas, quan un paquet és rebut amb errors, es va demanant la seva retransmissió fins que es rep correctament (persistència infinita) o bé fins que se supera el màxim nombre permès de retransmissions (truncat). La majoria dels estàndards de comunicacions sense fils moderns utilitzen combinacions dels esquemes FEC i ARQ per tal de gaudir dels avantatges d'ambdues estratègies. S'ha d'afegir, a més, que existeixen multitud de propostes recents en què es disposa d'un o varis nodes intermediaris entre la font i el destí, anomenats repetidors (en anglès *relays*), que assisteixen la comunicació. S'aconsegueix així millorar l'efectivitat de la transmissió explotant la diversitat espacial resultant de la incorporació de camins addicionals de comunicació entre la font i el destí.

D'acord amb la filosofia del disseny intercapes (en anglès *cross-layer*), en con-

traposició a l'optimització local a cada capa de la pila de protocols, es du a terme una optimització conjunta de determinats paràmetres corresponents a distintes capes, amb la finalitat de millorar les prestacions del sistema en la seva globalitat. En contret, en els darrers temps han aparegut un gran nombre de propostes de dissenys intercapes que combinen la utilització d'esquemes AMC a la capa física amb l'ús de protocols ARQ a la capa d'enllaç de dades. Aquests dissenys aconsegueixen millorar l'eficiència espectral traient profit de manera conjunta de l'adaptabilitat d'AMC a les condicions canviants del canal sense fils i de la capacitat detectora/correctora d'errors d'ARQ.

En aquest treball de recerca es desenvolupa un marc teòric per al disseny, l'anàlisi i l'optimització intercapes de sistemes sense fils que combinen esquemes AMC amb protocols ARQ. Aquesta proposta es basa en la utilització de cadenes discretes de Markov (DTMC, de l'anglès *Discrete Time Markov Chain*) per descriure de manera conjunta el comportament estadístic del procés d'arribades de paquets, el sistema de cues i la capa física. Per això, en primer lloc es desenvolupa un model de Markov bidimensional de primer ordre de la capa física que té en compte les característiques del canal sense fils i l'esquema AMC. Aquest és un pilar fonamental pel desenvolupament del present treball de recerca, ja que disposar d'un model capaç de capturar amb precisió el comportament de la capa física és un requisit indispensable per a garantir una anàlisi correcta dels paràmetres de QoS a la capa d'enllaç de dades. En base a aquest model, s'investiguen les interaccions entre aquestes dues capes quan s'implementen protocols ARQ, ja sigui amb persistència infinita o truncats. A més, es proposa un esquema cooperatiu en què un node repetidor és l'encarregat de dur a terme les retransmissions dels paquets que la font no aconsegueix transmetre correctament al destí. A partir de la modelització basada en DTMCs es poden obtenir les expressions analítiques dels paràmetres que caracteritzen el comportament de les prestacions del sistema, com són la taxa efectiva de transmissió de paquets (en anglès *throughput*), el retard mitjà i la taxa mitjana de pèrdua de paquets, ja sigui per desbordament de les cues o bé per superar el nombre màxim permès de retransmissions. Amb el propòsit de comparar aquesta metodologia amb eines analítiques no basades en l'ús de cadenes de Markov, el cas d'ARQ amb persistència infinita s'analitza també aplicant la teoria d'amplada de banda i capacitat efectives. Ambdues anàlisis són detalladament contrastades, confirmant la superioritat del primer mètode quant a fidelitat en l'estimació del comportament del sistema real, a costa d'un increment en la complexitat de l'anàlisi. El model analític proposat és utilitzat per a desenvolupar distintes estratègies de disseny intercapes amb l'objectiu de proporcionar garanties de QoS en termes de retard màxim i taxa màxima de pèrdua de paquets. Per acabar, s'investiguen els efectes que té el retard en l'estimació de l'estat del canal sobre els paràmetres de QoS del sistema.

Contents

List of Figures

List of Tables

Acronyms

2D-FSMC two-dimensional finite state Markov chain 19

3GPP 3rd Generation Partnership Project 148

4G Fourth Generation 1

ACF auto-correlation function 8

ACK positive acknowledgement 4

AFSMC amplitude-based finite state Markov chain 9

AMC adaptive modulation and coding 3

ARQ automatic repeat request 3

AWGN additive white Gaussian noise 33

BICM bit-interleaved coded modulation 4

BER bit error rate 23

CC Chase combining 5

CARQ cooperative ARQ 6

CRC cyclic redundancy check 4

CSI channel state information 4

CTP context tree pruning 23

D-BMAP discrete batch Markovian arrival process 20

DLC data link control 3

DTMC discrete time Markov chain 7

FEC forward error correction 3

FIFO first-in-first-out 16

FSMC finite state Markov chain 8

GBN-ARQ go-back-N ARQ 4

(H)ARQ (hybrid) automatic repeat request 6

HARQ hybrid automatic repeat request 5

HARQ-IR incremental redundancy 5

HARQ-CC Chase combining 5

HMM hidden Markov model 23

HSDPA High-Speed Downlink Packet Access 3

HSUPA High-Speed Uplink Packet Access 3

HSPA+ High-Speed Packet Access plus 148

IP Internet protocol 2

QoS quality of service 1

SNR signal-to-noise ratio 5

SPID subpacket identifier 70

SR-ARQ selective repeat ARQ 4

STBC space-time block code 104

SW-ARQ stop and wait ARQ 4

TCP transport control protocol 2

UMTS Universal Mobile Telecommunications System 1

WiMAX Worldwide Interoperability for Microwave Access 1

WLAN wireless local area network 1

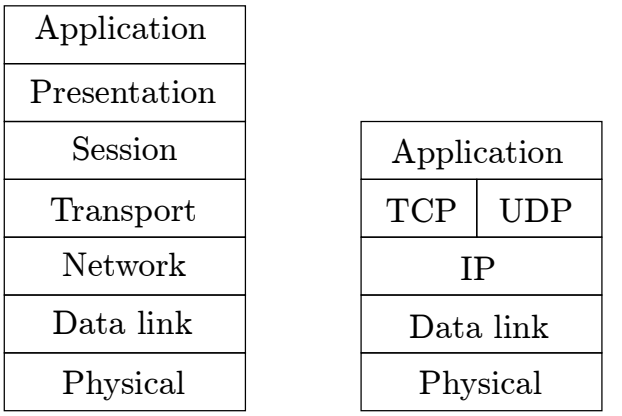
CHAPTER 1

Introduction

1.1 Context

In the last decade, fueled by the users' ever-increasing demands of throughput and mobility, an explosive development of wireless services and applications has taken place. Wireless technologies have allowed the distribution of high data-rate connections in localized hot-spots. In that sense, a substantial proliferation of IEEE 802.11 standard family-based wireless local area networks (WLANs) (IEEE, 1997) has been witnessed among residential users and commercial/business/academic environments. Moreover, the promise of *anytime-anywhere* connectivity has been fulfilled by cellular systems, such as Universal Mobile Telecommunications System (UMTS) and Long Term Evolution (LTE), which provide wide coverage areas, full mobility and roaming. The Fourth Generation (4G) cellular systems are expected to support reliable transmissions with peak data-rates ranging from 100 Mbps for high-mobility environments to 1 Gbps for low-mobility environments. Long Term Evolution Advanced (LTE-Advanced) and Mobile Worldwide Interoperability for Microwave Access (WiMAX), based on the IEEE 802.16e standard (IEEE, 2004), are emerging 4G candidates targeting these ambitious objectives (see e.g., (Abichar et al., 2010; Ghosh et al., 2010; Wang et al., 2008b) and references therein).

The unique characteristics of wireless networks along with the heterogeneous quality of service (QoS) requirements of wireless applications (e.g., Web access, interactive mobile multimedia applications, interactive gaming, video streaming, etc.) has urged the research community to efficiently manage the scarce wireless resources. In such networks, the time and frequency selectivity of wireless fading channels poses a great challenge when designing solutions to support QoS provisioning for current and envisaged heterogeneous mobile applications.



(a) OSI stack (b) TCP/IP stack

Figure 1.1: (a) The layered ISO/OSI architecture (b) The layered Internet architecture.

1.1.1 Cross-layer design

One of the greatest problems faced by researchers to accommodate the demanding services and applications for next-generation broadband wireless networks is the availability of strategies at different layers. The layered International Standardization Organization (ISO)/Open Systems Interconnection (OSI) architecture for networking (ITU, 1994; Wetteroth, 2001), on which the transport control protocol (TCP)/Internet protocol (IP) architecture is loosely based, is a successful example of the importance of a good architectural design. These architectures divide the overall networking task into layers, as illustrated in Fig. 1.1, and define a hierarchy of services to be provided by the individual layers (Dañobeitia et al., 2008). Each layer in the protocol stack is designed and operated independently. As stated by Haas (2001), Carneiro et al. (2004) or Rasmussen et al. (2006), the strict separation of functionalities of the design based on the conventional layered model may be inhibiting the effective implementation of guaranteed QoS provisioning strategies, forcing the network to operate in a suboptimal mode with respect to performance, QoS, and/or energy consumption. Hence, in order to meet the challenging demands of future wireless networks, it may be required to adopt new approaches where protocols can be designed allowing direct communication among nonadjacent layers (i.e. creating new interfaces between nonadjacent layers), sharing variables among layers, redefining the layer boundaries, designing protocols within a layer based on the details of another layer, jointly tuning of parameters across layers and so on. Such violations of a layered architecture have been termed as cross-layer design with respect to the reference architecture (see, e.g., (Kawadia and Kumar, 2005; Shakkottai et al., 2003; Srivastana and Motani, 2005) and references therein).

1.1.2 PHY/DLC cross-layer design

Following the spirit of cross-layer design, and in contrast to a local optimization of each of the layers, a joint optimization of the physical (PHY) and data link control (DLC) layers has been widely considered in order to improve the whole system performance. The impact of PHY layer techniques on supporting diverse DLC layer QoS requirements needs to be thoroughly analyzed. In this case, parameters and algorithms at both layers should be jointly considered. For example, when the queueing model is taken into account, the throughput at the DLC layer is lower than the one achieved at the PHY layer. The reasons for this loss are the buffer overflow due to finite-length queues, which may cause the drop of arriving packets, and the absence of packets waiting for transmission even if the channel is in good channel conditions.

In order to counteract the impact of wireless fading on the performance of the system, many communication strategies have been developed to improve the spectral and/or power efficiency of the PHY layer. Among them, adaptive modulation and coding (AMC) schemes, which denote the matching of the modulation and coding to the conditions of the radio link, have been intensely researched and widely adopted in most state-of-the-art wireless communications standards (IEEE 802.11, High-Speed Downlink Packet Access (HSDPA), High-Speed Uplink Packet Access (HSUPA), LTE, LTE-Advanced, IEEE 802.16, etc.). The unique nature of AMC to improve the performance of upper-layer protocols has spurred the development of cross-layer designs with the goal of integrating QoS provisioning protocols at higher network layers with AMC at the PHY layer. Specifically, many recent cross-layer design proposals between the PHY layer and the DLC layer coincide in combining AMC with an automatic repeat request (ARQ)-based error control protocol.

The use of error control strategies at the DLC layer has become an integral part in the design of modern communication systems to guarantee the QoS constraints of new applications. There are two basic approaches to sustain or enhance link reliability: the forward error correction (FEC) scheme and the ARQ scheme. In a FEC system an error-correction code is used. When a received codeword is detected in error and the number of errors is within the designed error-correcting capabilities of the code, the errors are corrected. In an ARQ system a code with good error-detection capability is used. If a codeword is erroneously received, the transmitter is instructed (through the return channel) to retransmit the same codeword. In order to exploit the advantages of both strategies, most state-of-the-art wireless communications standards (IEEE 802.11, HSDPA, HSUPA, LTE, LTE-Advanced, IEEE 802.16, etc.) use combinations of ARQ and FEC.

The cross-layer designs that combine AMC at the PHY layer with an ARQ protocol at the DLC layer aim at improving the spectral efficiency by jointly exploiting the adaptability of AMC to the wireless channel conditions and the error-correcting capability of ARQ (Dai and Letaief, 2008; Harsini et al., 2011; Ishizaki and Hwang, 2007; Kang et al., 2009; Le et al., 2006a,b, 2007; Liu et al., 2004, 2005a,b; Mardani et al., 2011; Poggioni et al., 2007, 2010; Shi and Yuan, 2008; Wang et al., 2007).

Adaptive modulation and coding (AMC)

AMC is a method to improve the spectral and/or power efficiency of a radio link for a given required quality (error probability) by matching the modulation and coding to the channel conditions (see, e.g., (Svensson, 2007) and references therein). Since any practical wireless channels perturb the transmitted waveforms, the detection process is never free from errors thus making mandatory the use of error correction/detection techniques. Depending on the severity of the fading, different degrees of information protection are required. AMC systems invariably require some channel state information (CSI) at the transmitter or, alternatively, the modulation and coding scheme to be used, so that the transmission can be adapted in accordance with the channel characteristics. Modulation and coding techniques that do not adapt to fading conditions require a fixed link margin to maintain acceptable performance when the channel quality is poor. Consequently, these systems are effectively designed for the worst-case channel conditions. Therefore, if the channel is not constant during the whole transmission, which is actually the case for most state-of-the-art wireless systems, adaptive schemes have a large performance advantage in comparison to fixed ones. Adaptation to the channel fading can increase average throughput, reduce required transmit power, and/or reduce the average probability of bit error by taking advantage of favorable channel conditions to operate at higher data rates or lower power, and reducing the data rate or increasing power as the channel degrades (Goldsmith, 2005).

Automatic repeat request (ARQ) strategies

There are basically two types of ARQ: stop and wait ARQ (SW-ARQ) and Continuous ARQ. With the first strategy, the transmitter sends a single packet and waits for a positive acknowledgement (ACK) or a negative acknowledgement (NACK) before sending more packets. This can be a very inefficient method due to the waste of time waiting for the receiver response when the round-trip delay is not negligible. On the contrary, with the continuous ARQ strategy, packets are sent continuously until a NACK is received. Two continuous ARQ techniques can be distinguished: go-back-N ARQ (GBN-ARQ), in which the transmitter resends the erroneously received packet plus all the packets that followed until the NACK was received, and selective repeat ARQ (SR-ARQ), in which only the erroneous packet is retransmitted. Both types of continuous ARQ offer greater throughput efficiency than SW-ARQ at the cost of greater memory and processing requirements. Among all ARQ schemes, SR-ARQ is reported to show the best throughput performance.

In Type-I hybrid FEC/ARQ schemes, both error detection and FEC bits are added to each packet prior to transmission (using, for instance, a concatenation of a bit-interleaved coded modulation (BICM) scheme and a cyclic redundancy check (CRC) code). When a received codeword is detected in error, two situations may arise. If the number of errors is within the designed error-correcting capabilities of the code, the errors are corrected. Otherwise, an uncorrectable error pattern is detected, the received coded data block is discarded and a retransmission is requested by the receiver, similar to standard ARQ. Retransmissions take place

at either the same or a different code rate until the packet is correctly received (infinitely persistent ARQ) or until a preset number of retransmissions have taken place (truncated ARQ). Although this method does not require a large buffer at the receiver, it is an inefficient method of implementing ARQ.

When using a more sophisticated form of hybrid FEC/ARQ, known as hybrid automatic repeat request (HARQ) (Cheng, 2006), if the receiver fails to decode a packet, any previously received signal is stored in a buffer and a retransmission request in the form of a NACK is fed back to the transmitter. In Type-I HARQ, upon reception of this NACK the transmitter sends the same coded packet again. At the receiver side, the optimal solution is to combine these multiple signals according to the maximal ratio combining (MRC) principle (Chase, 1973, 1985), an approach usually known as Chase combining. Alternatively, in the so-called Type-II or Type-III (when each packet is self-decodable) HARQ, also known as the incremental redundancy scheme, when the receiver fails to decode a packet, the decoded bits (or log-likelihood ratios) are stored in a buffer and a retransmission is requested. The transmitter, upon reception of a NACK, instead of resending the same coded packet, transmits additional redundant information in each retransmission by, for instance, varying the puncturing pattern. Obviously, incremental redundancy (IR) requires larger buffer size than Chase combining (CC); nevertheless, although the CC scheme can improve link performance by coherently combining multiple copies of the received signal, the IR strategy can also benefit from the combination of the different coded versions, which jointly form a lower-rate code with stronger error protection capabilities (Cheng, 2006). The performance comparison of Chase combining (HARQ-CC) and incremental redundancy (HARQ-IR) has been widely analyzed in the literature (Cheng, 2003, 2006; Frederiksen and Kolding, 2002; Frenger et al., 2001), showing that contrarily to the widely held belief that HARQ-IR is superior to HARQ-CC for all circumstances, this is not always the case, and on fading channels there are situations in which an HARQ-IR scheme actually performs poorer than a Chase combining system.

Cooperative ARQ

Signal fading arising from multipath propagation can be effectively mitigated through the use of spatial diversity, by means of multiple input multiple output (MIMO) techniques. Nevertheless, in practice, it is difficult to pack multiple antennas into small-size mobile nodes. To overcome this limitation, another form of spatial diversity, known as cooperative diversity, has recently emerged (see, e.g., (Nosratinia et al., 2004) and references therein). It is achieved by making use of one or multiple intermediate relay stations to forward data from a source node to the corresponding destination node. Performance improvement brought by cooperation is due to a better average signal-to-noise ratio (SNR) of the relay-destination link compared with that of the source-destination link. Moreover, it is also due to spatial diversity, which is generated by transmitting signals from different locations (source and relay(s)), thus providing independently faded versions of the signal at the receiver. The basic ideas underpinning cooperative communication can be traced back to the work of Cover and El Gamal on the information

theoretic properties of the relay channel (Cover and Gamal, 1979). The authors in (Sendonaris et al., 2003a,b) present the most important results regarding the capacity analysis of user cooperation, where two nodes cooperate by transmitting each bit over two successive bit intervals. Laneman et al. (2004) proposed different cooperative diversity schemes including fixed relaying, selection relaying and incremental relaying. Two fixed relaying protocols are examined, in which the relay either amplifies what it receives (amplify-and-forward) or fully decodes, re-encodes, and retransmits the source message (decode-and-forward), respectively. In the selection relaying protocol, if the measured source-relay channel gain falls below a certain threshold, the source simply continues its transmission to the destination, in the form of repetition or more powerful codes. On the contrary, if the channel gain lies above the threshold, the relay forwards what it receives from the source, using either amplify-and-forward or decode-and-forward, in an attempt to achieve diversity gain. Incremental relaying is a class of adaptive protocol based upon limited feedback from the destination terminal, which broadcasts a single bit of feedback to the source and the relay to indicate the success or failure of a transmission. In case of error, a retransmission is performed by the relay and the destination tries to combine the two transmissions. These incremental relaying protocols can be viewed as extensions of ARQ to the relay context: in ARQ, the source retransmits if the destination provides a negative acknowledgment via feedback; in incremental relaying, the relay retransmits in an attempt to exploit spatial diversity. The main difference between the cooperative communication methods and cooperative ARQ (CARQ) protocols is that in the latter the feedback from the destination terminal is exploited.

1.2 Problem formulation and adopted approach

As it has been previously mentioned, one of the most challenging issues for next-generation wireless networks is the provision of QoS guarantees when dealing with the high number of emergent multimedia applications. The service heterogeneity and the hostility and variability of mobile radio channels, have rendered unavoidable the development of efficient strategies to manage the scarce wireless resources. It has also been remarked that, with this objective, most state-of-the art wireless communications standards combine the use of AMC strategies at the PHY layer with an (hybrid) automatic repeat request ((H)ARQ)-based error control protocol at the DLC layer. Accordingly, many recent works focus on cross-layer designs aiming at improving the spectral efficiency by jointly exploiting the adaptability of AMC and the error-correcting capability of ARQ. How to optimally tune layer-specific design parameters to counteract the impact of wireless fading becomes a challenging task when different levels of QoS provisioning have to be satisfied. With this motivation, the main goal of this dissertation is to provide a unified view of the cross-layer design, analysis and optimization of AMC/(H)ARQ-based wireless systems to allow the joint optimization of both the PHY and the DLC layers. We specifically target the maximization of the average throughput while guaranteeing prescribed QoS constraints in terms of average packet loss rate and average delay. The adopted approach for tackling this problem will rely on the

use of Markov chains, a widely accepted theory by the research community, to model the system dynamics.

Bearing in mind this objective, the following considerations are of fundamental importance:

- In order to investigate the interactions between the PHY and DLC layers, accurate models for the wireless channel are required. Consequently, one of the key motivations of this work is the development of a Markov-based model able to faithfully reproduce the wireless flat-fading channel characteristics.

- The AMC scheme used in the PHY layer needs to be defined. In order to make the study realistic, two different AMC pools will be considered: the first one will be borrowed from the IEEE 802.11a standard (IEEE, 1997) and the second one from the IEEE 802.16e standard (IEEE, 2004). The reason for this selection is twofold: on the one hand, to demonstrate that the proposed analytical Markov-based model usability does not depend on the considered AMC scheme and, on the other hand, to enable the analysis of the HARQ schemes that, as it will be shown in Chapter 4, require the definition of different puncturing patterns corresponding to successive (re)transmissions, as those proposed by the 802.16e standard. Given an AMC pool, the transmission modes (TMs) will be selected according to a set of switching thresholds. How to determine these thresholds is a key issue that plays an important role in the system efficiency and, consequently, it needs to be thoroughly investigated. Additionally, the inaccuracies in the CSI used to adapt the transmission scheme to the channel characteristics should also be taken into account, and its effects on the system performance analyzed.

- Different (H)ARQ techniques ought to be considered, with both finite and infinite persistence. To that end, the existing (H)ARQ schemes need to be investigated and included in the formulation of the corresponding queueing process analysis. Furthermore, the performance improvement brought by cooperation when a relay node assists the communication between source and destination should be also studied.

- To solve the proposed optimization problem, the system behaviour has to be characterized through the analysis of the queueing process induced by the PHY layer (channel and AMC scheme), the packet arrival process and the (H)ARQ protocol. To that end, a discrete time Markov chain (DTMC) jointly describing the whole system dynamics has to be obtained, allowing the derivation of the system performance metrics (throughput, average packet delay and packet loss rate) by solving the global transition probability matrix of the system.

- The validity of the proposed cross-layer framework has to be verified. With this objective, analytical and Monte Carlo simulation results will be confronted. Furthermore, simulations should be also employed to analyze the performance of either infinitely persistent or truncated (H)ARQ schemes,

as well as to investigate the different (H)ARQ strategies and the benefits of cooperation.

1.3 Novelty and main contributions

Although the related work, the main references and the contributions of this dissertation are all provided and extensively described in each forthcoming chapter, the novelty with respect to previous research is briefly outlined in this section.

The most widely used flat fading channel characterization is a statistical model for the received signal amplitude that was proposed by Clarke (1968). This is an accurate *continuous* model but, due to its difficult mathematical tractability, discrete Markov models have often been adopted. Accordingly, most previous works on PHY/DLC cross-layer design rely on the use of first-order finite state Markov chains (FSMCs) to model the behaviour of the wireless channel. The auto-correlation function (ACF) corresponding to these analytical channel models can not fit the hypergeometric ACF of the simulated process (Tan and Beaulieu, 2000), which could impair the validity of the corresponding cross-layer designs. Moreover, the first-order FSMC models used in (Liu et al., 2004, 2005a,b) present several deficiencies that will be analyzed in Chapter 3, which have propagated to many other research works (see, e.g., (Ishizaki and Hwang, 2007; Le et al., 2006a,b; Poggioni et al., 2007; Wang et al., 2007)). Higher-order Markov models, which improve the first-order FSMC description at the cost of an increase of complexity, have been also examined (see, e.g., (Babich and Lombardi, 2000; Babich et al., 2000; Bergamo et al., 2002; Turin and van Nobelen, 1998)). To overcome the aforementioned limitations, a simple first-order two-dimensional FSMC model for the Rayleigh flat-fading channel is developed. This model, which is one of the main contributions of this research work, is able to improve the ACF fitting of the first-order FSMCs.

This characterization is then used to develop a novel methodology for the PHY layer modeling that incorporates the implementation of the AMC scheme. In contrast to the proposals of Liu et al. (2004, 2005a,b), the AMC scheme is designed independently from the channel model, leading to a better PHY layer characterization. The available TMs in the AMC pool are selected according to a set of switching thresholds that depend on the considered (H)ARQ protocol. Therefore, different PHY layer characterizations are obtained according to the selected error control protocol. Specifically, the PHY layer models corresponding to the infinitely persistent ARQ protocol and to the truncated (H)ARQ schemes are derived. Moreover, the generalization to the case of a cooperative system is also provided.

Furthermore, an analytical DTMC-based link-level queueing model of a point-to-point adaptive multi-rate wireless system using an (H)ARQ-based error control protocol is proposed. This approach integrates the aforementioned first-order two-dimensional FSMC PHY layer model, the packet generation, which is modeled as a Markovian arrival process that considers traffic burstiness, and the (H)ARQ scheme. Three error control protocols are analyzed: infinitely persistent ARQ, truncated (H)ARQ and CARQ. Traffic burstiness in an AMC/ARQ-based wire-

less system was considered by Le et al. (2006b). However, infinitely persistent "pure" ARQ-based error control schemes were considered, while the generalization to more sophisticated truncated-HARQ protocols was not addressed at all. Additionally, as most previous works on this topic, the main shortcoming of this proposal was its reliance on first-order amplitude-based finite state Markov chains (AFSMCs) to model the wireless fading channel. In the case of a truncated (H)ARQ protocol, there exist no works jointly considering the interactions between the AMC scheme, the (H)ARQ protocol, the traffic burstiness and the queues. Liu et al. (2004) proposed a scheme combining AMC with a truncated ARQ protocol in order to improve the overall system spectral efficiency. Nevertheless, queueing effects on the average packet delay were not taken into account. The authors in (Wang et al., 2007) focused on a cross-layer design combining AMC with truncated ARQ in which the queueing effects were considered. However, they assumed a memoryless packet arrival process. Moreover, in order to facilitate mathematical tractability of the queueing process, they relied on the rather unrealistic assumption of considering a time slotted system were only one frame was transmitted per slot, with each frame at the PHY layer containing at most one packet from the DLC layer. Kang et al. (2009) considered a joint design approach where HARQ-IR was associated with an AMC scheme, although the queueing process was not faced at all. The analysis of CARQ multi-rate wireless systems has been also widely explored (see, e.g., (Dai and Letaief, 2008; Harsini et al., 2011; Mardani et al., 2011; Shi and Yuan, 2008)). Nevertheless, previous studies neither consider the queueing process induced by both the AMC scheme and the CARQ protocol nor derive the packet delay in such systems. Besides, to the best of our knowledge, the impact of using outdated CSI on the performance of an AMC-based wireless system using an (H)ARQ-based error control protocol had not been addressed at all before being considered in this work.

The outcome of this analytical framework is a DTMC-based link-level queueing model that allows a unified analysis of AMC/(H)ARQ-based wireless systems, in which the most commonly used (H)ARQ schemes have been investigated. As a result, the formulation of cross-layer multidimensional design strategies, aiming at the maximization of the average throughput of the system while satisfying prescribed QoS requirements, has been provided. Simplified optimizations are also proposed, resulting in a lower complexity scheme at the cost of a performance loss. The analytical results allow a-priori design decisions to be made and enable dynamic reconfiguration strategies that take into account layer-specific design parameters. Hence, this cross-layer framework allows the fulfillment of the heterogeneous service-specific QoS requirements of wireless applications.

1.4 Organization of this research work

The structure of the thesis reflects the progression of the research project. It is divided into six chapters. A brief description of each one is provided below:

Chapter 1: *Introduction* - It provides the context of the thesis and an introductory description of the main topics of the research project. Furthermore, the

problem to be solved is formulated and the main contributions and related publications are outlined.

Chapter 2: *System model and Markov-based cross-layer approach* - The generic system block diagram is described in this chapter, where the Markov chain-based system model approach that will be used throughout the research work is introduced. More specifically, the packet arrival process is briefly described and the two-dimensional finite state Markov chain model for the Rayleigh flat-fading channel is developed. Finally, a general description of the cross-layer perspective that will be considered in the following chapters is provided.

Chapter 3: *Systems based on an infinitely persistent ARQ protocol* - A novel framework for the cross-layer design, analysis and optimization of wireless networks combining AMC at the PHY layer with an infinitely persistent Type-I hybrid FEC/ARQ protocol at the DLC layer is proposed. To that end, the PHY layer is modeled with a first-order two-dimensional Markov chain and the QoS performance at the DLC layer is investigated through the use of two different approaches: on the one hand, a DTMC that jointly describes the statistical behaviour of the arrival process, the queueing system and the PHY layer and, on the other hand, an effective bandwidth/capacity-based approach. Finally, constrained optimization problems are presented and discussed.

Chapter 4: *Systems based on a truncated (hybrid) ARQ protocol* - This chapter generalizes and extends the analytical Markov-based model that has been previously presented in Chapter 3 to the case of a truncated (H)ARQ scheme. Thus, Type-I hybrid FEC/ARQ, HARQ-CC and HARQ-IR schemes with finite persistence are analyzed and compared.

Chapter 5: *Systems based on cooperative ARQ protocols* - The effects of using CARQ in wireless relay systems using AMC at the PHY layer and an infinitely persistent type-I hybrid FEC/ARQ protocol at the DLC layer are analyzed in this chapter. To that end, and using the approach described in previous chapters, an equivalent cooperative PHY layer model is developed, allowing the analysis of the proposed CARQ scheme.

Chapter 6: *Analysis of an infinitely persistent ARQ protocol with outdated channel state information* - The aim of this chapter is the development of a methodology to extend the previous analytical framework, in which perfect CSI is assumed to be available at the receiver, to the case of a delayed CSI feedback channel. Therefore, imperfections in the AMC scheme caused by this delay are accounted for when an infinitely persistent Type-I hybrid FEC/ARQ protocol is considered. It is worth stressing that the proposed methodology could also be extended to any of the AMC/(H)ARQ cross-layer designs previously presented in this research work.

Chapter 7: *Conclusions* - It provides a summary of the main results and contributions of this project. The guidelines for future work are also outlined.

1.5 Related publications

The publications related to the present research work are listed below.

1.5.1 Book chapters

- J. Ramis and G. Femenias, *Chapter 17: Cross-layer design, analysis, and optimization of QoS-constrained AMC/ARQ-based wireless networks*, in Using Cross-Layer Techniques for Communication Systems, Habib F. Rashvand and Yousef S. Kavian, editors. IGI Global, pp.400-431, April 2012, ISBN: 978-1-4666-0960-0.

1.5.2 International Journal Publications

- G. Femenias, J. Ramis and L. Carrasco, *Using two-dimensional Markov models and the effective capacity approach for cross-layer design in AMC/ ARQ-based wireless networks*, IEEE Transactions on Vehicular Technology, vol.58, no.8, pp.4193-4203, October 2009, ISSN: 0018-9545.

- J. Ramis and G. Femenias, *Cross-layer design of adaptive multi-rate wireless networks using truncated HARQ*, IEEE Transactions on Vehicular Technology, vol.60, no.3, pp.944-954, March 2011, ISSN: 0018-9545.

- J. Ramis and G. Femenias, *Cross-layer optimization of AMC/CARQ-based wireless systems*, submitted to IEEE Transactions on Wireless Communications.

1.5.3 International Conference Contributions

- J. Ramis, L. Carrasco and G. Femenias, *A two-dimensional Markov model for cross-layer design in AMC/ARQ-based wireless networks*, IEEE Global Telecommunications Conference (GLOBECOM), New Orleans (USA), November - December 2008.

- L. Carrasco, J. Ramis and G. Femenias, *Multidimensional Markov models for the cross-layer design of multi-rate wireless systems using the effective capacity function*, IEEE 69th Vehicular Technology Conference (VTC) Spring, Barcelona (Spain), April 2009. **Best Paper Award**.

- J. Ramis, G. Femenias and L. Carrasco, *Cross-layer design of multi-rate wireless networks based on link layer truncated ARQ*, 2nd IEEE International Workshop on Cross Layer Design (IWCLD), IEEE Communications Society / EURASIP, Palma de Mallorca (Spain), June 2009.

- J. Ramis, G. Femenias, F. Riera-Palou and L. Carrasco, *Cross-Layer optimization of adaptive multi-rate wireless networks using truncated Chase combining HARQ*, IEEE Global Telecommunications Conference (GLOBECOM), Miami (USA), December 2010.

- J. Ramis, G. Femenias and L. Carrasco, *Cross-layer modeling of AMC/ARQ-based wireless networks with outdated CSI*, IEEE 8th International Symposium on Wireless Communication Systems (ISWCS), Aachen (Germany), November 2011.

- J. Ramis and G. Femenias, *Cross-layer modeling of wireless systems using AMC with Cooperative-ARQ error control*, European Wireless, Poznan (Poland), April 2012.

1.5.4 National Conference Contributions

- J. Ramis, L. Carrasco and G. Femenias, *Cross-layer design of multi-rate wireless systems using AMC with ARQ-based error control. A two dimensional Markov model approach*, VII Jornadas de Ingeniería Telemática (JITEL), pp. 1-8, Alcalá de Henares, September 2008.

- J. Ramis, G. Femenias and L. Carrasco, *Diseño intercapas en redes inalámbricas basadas en AMC y ARQ truncado*, VIII Jornadas de Ingeniería Telemática (JITEL), pp. 260-267, Cartagena, September 2009.

- J. Ramis, G. Femenias, F. Riera-Palou and L. Carrasco, *Throughput optimization in QoS constrained adaptive wireless networks using Chase combining HARQ*, IX Jornadas de Ingeniería Telemática (JITEL), pp. 145-152, Valladolid, September-October 2010. **Best Paper Award**.

1.6 Other Supporting Contributions

Although they do not constitute the main focus of this research work, the following additional contributions (in chronological order) can also be considered relevant. All of them are related to the topic of packet scheduling in wireless networks, and the three last papers have been developed using the approach introduced in this thesis.

- J. Ramis, L. Carrasco, G. Femenias and F. Riera-Palou, *Scheduling algorithms for 4G wireless networks*, The 12th IFIP International Conference on Personal Wireless Communications (PWC), pp. 264-276, Prague (Czech Republic), September 2007.

- L. Carrasco, G. Femenias and J. Ramis, *Cross-layer optimization of AMC/ARQ-based wireless networks with channel-aware multiuser scheduling protocols*, X Jornadas de Ingeniería Telemática (JITEL), pp. 174-181, Santander, September 2011. **Best Paper Award**.

- L. Carrasco, G. Femenias and J. Ramis, *Cross-layer design of multirate/multiuser wireless networks with channel-aware scheduling*, IEEE 8th International Symposium on Wireless Communication Systems (ISWCS), Aachen (Germany), November 2011.

- L. Carrasco, G. Femenias and J. Ramis, *Delay statistics and throughput performance for multi-rate wireless networks under multiuser diversity*, submitted to IEEE Transactions on Wireless Communications.

CHAPTER 2

System model and Markov-based cross-layer approach

In this chapter the generic system model and the assumptions that will form the basis for the development of the proposed cross-layer analytical framework are presented and briefly described. Afterwards, the Markov chain-based system model approach that will be used throughout the present research work is introduced. For this purpose, the discrete batch Markovian arrival process used to model the packet generation is briefly described, and a simple first-order two-dimensional FSMC model for the Rayleigh flat-fading channel is presented. To conclude, the cross-layer perspective for the design and optimization of wireless networks combining AMC at the PHY layer with an (H)ARQ protocol at the DLC layer is discussed.

2.1 Generic system block diagram

The point-to-point wireless packet communication system under consideration is illustrated in Fig. 2.1, showing the most important blocks involved in the transmission and reception processes. Following the (H)ARQ controller, and based on the CSI feedback from the receiver on a frame-by-frame basis, the AMC scheme selects a TM, which corresponds to a modulation and FEC convolutional code pair. Different code rates can be easily obtained through the combination of a fixed rate convolutional code followed by a puncturer. This technique, which will be further discussed in Chapter 4, consists in eliminating, according to a puncturing pattern, some of the bits at the output of the fixed rate convolutional encoder, obtaining in this way the set of allowed coding rates. At the receiver end, joint coherent demodulation and soft decoding are first employed to retrieve each coded packet. Then, after suitable depuncturing by conveniently padding dummy bits that do no affect the decoding metric, a Viterbi decoder is used to decode the received packet.

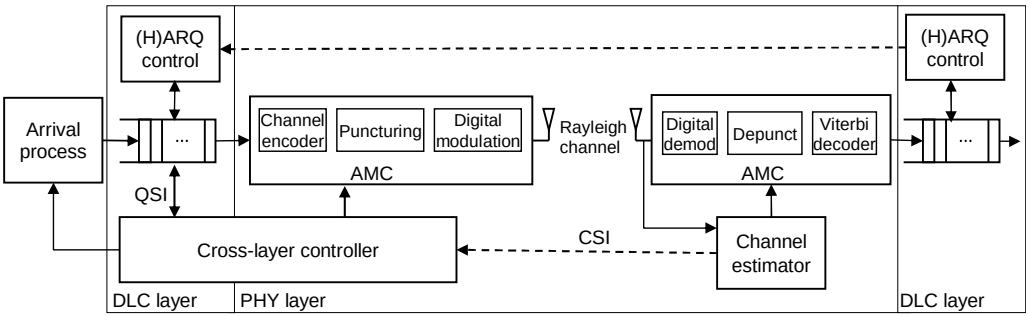


Figure 2.1: System model.

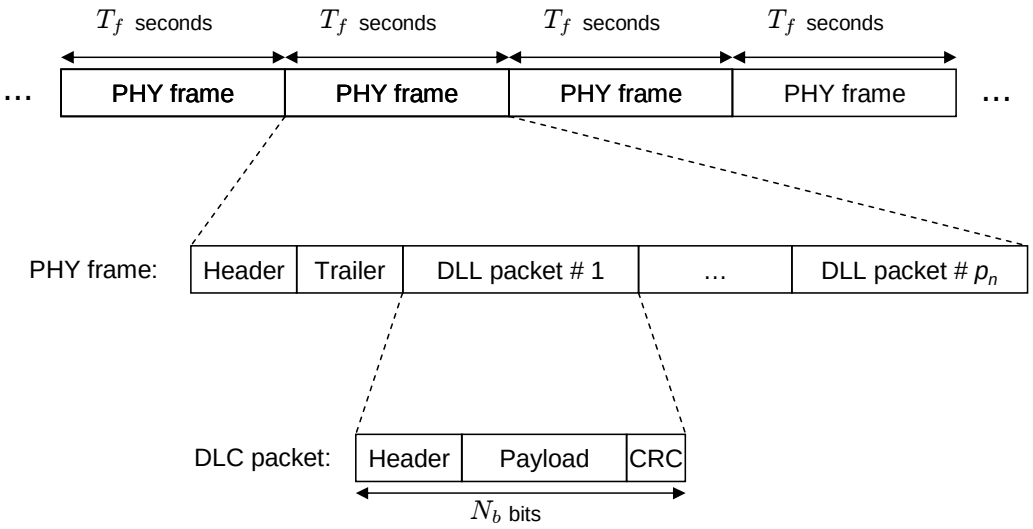


Figure 2.2: Frame and packet structures.

As shown in Fig. 2.2, the processing unit at the DLC layer is a packet of fixed size equal to N_b bits and the processing unit at the PHY layer is a frame made of a variable number of packets that depends on the TM selected by the AMC protocol. The AMC scheme is assumed to have a set $\boldsymbol{\mathcal{M}}_p = \{0, \ldots, M_p - 1\}$ of M_p *possible* TMs, each corresponding to a particular combination of modulation and coding strategies, including the case in which the transmitter does not transmit. As it will be explained in following chapters, some of these M_p *possible* TMs may be declared *useless* and thus, only a set $\boldsymbol{\mathcal{M}} = \{0, \ldots, M - 1\}$ of M *useful* TMs will be available to the AMC scheme.

The system is assumed to support QoS-guaranteed traffic characterized by a maximum average packet delay $D_{l\max}$ and/or a target link layer packet loss rate $P_{l\max}$.

2.1.1 Transmitter

At the transmitter side, the (H)ARQ controller manages a buffer (queue) that operates in a first-in-first-out (FIFO) mode and that is able to store up to $\overline{Q}$ packets.

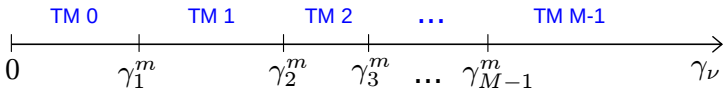


Figure 2.3: Transmission mode selection thresholds.

Encoded bits are mapped according to the selected modulation scheme, which is determined by the AMC controller at the receiver and fed back to the transmitter. According to (Liu et al., 2004), when implementing the AMC strategy, the entire SNR range is partitioned into a set of non-overlapping intervals defined by the partition

$$\mathbf{\Gamma}^m = \{[\,\gamma_0^m, \gamma_1^m\,), [\,\gamma_1^m, \gamma_2^m\,), \ldots, [\,\gamma_{M-1}^m, \gamma_M^m\,)\}\,, \tag{2.1}$$

with $\gamma_0^m = 0$ and $\gamma_M^m = \infty$. Then, as illustrated in Fig. 2.3, mode n is selected when $\gamma_\nu \in [\gamma_n^m, \gamma_{n+1}^m)$, where γ_ν denotes the instantaneous received SNR at time instant $t = \nu T_f$, where T_f is the frame period. The number of transmitted packets per frame depends on the TM n selected and it is given by $p_n = bR_n$, where R_n denotes the number of information bits per symbol used by TM n and b is a parameter that determines the number of transmitted packets per frame, to be set by the designer. For convenience, it will be considered that $p_0 < \cdots < p_{M_p-1}$, with $p_0 = 0$ (i.e., TM0 corresponds to the absence of transmission) and $p_{M_p-1} \triangleq \mathcal{C}_p$. It will be assumed that when the system selects a *useful* TM $n \in \boldsymbol{\mathcal{M}}$, it transmits c_n packets and, for convenience, it will also be considered that $c_0 < \cdots < c_{M-1}$, with $c_0 \geq 0$ and $c_{M-1} = \mathcal{C} \leq \mathcal{C}_p$.

2.1.2 Wireless channel

A Rayleigh block-fading channel model has been adopted (Biglieri et al., 2001) in which the channel gains corresponding to the νth frame transmission, h_ν, are characterized as a zero-mean circularly symmetric complex Gaussian random variable with unit power. That is, it is assumed that the frame duration T_f is much smaller than the coherence time of the channel. In this case, although the channel is assumed to remain invariant over at least one time frame interval, it is allowed to vary across successive frame intervals. These variations have been modeled using the Clarke's statistical Rayleigh fading process, characterized by a maximum normalized Doppler frequency $f_d T_f$, with f_d denoting the maximum Doppler frequency of the channel (Clarke, 1968).

2.1.3 Receiver

After detection and soft Viterbi decoding, an error detection process (using, for instance, a CRC code) is performed and the corresponding ACK/NACK message is fed back to the (H)ARQ controller. Given the short length of the ACK/NACK messages and the use of a high degree of FEC protection, an error-free and instantaneous (H)ARQ feedback channel can be safely assumed. Perfect CSI is considered to be available at the receiver side and, thus, a frame-by-frame TM selection process is performed at the receiver AMC controller and fed back to the

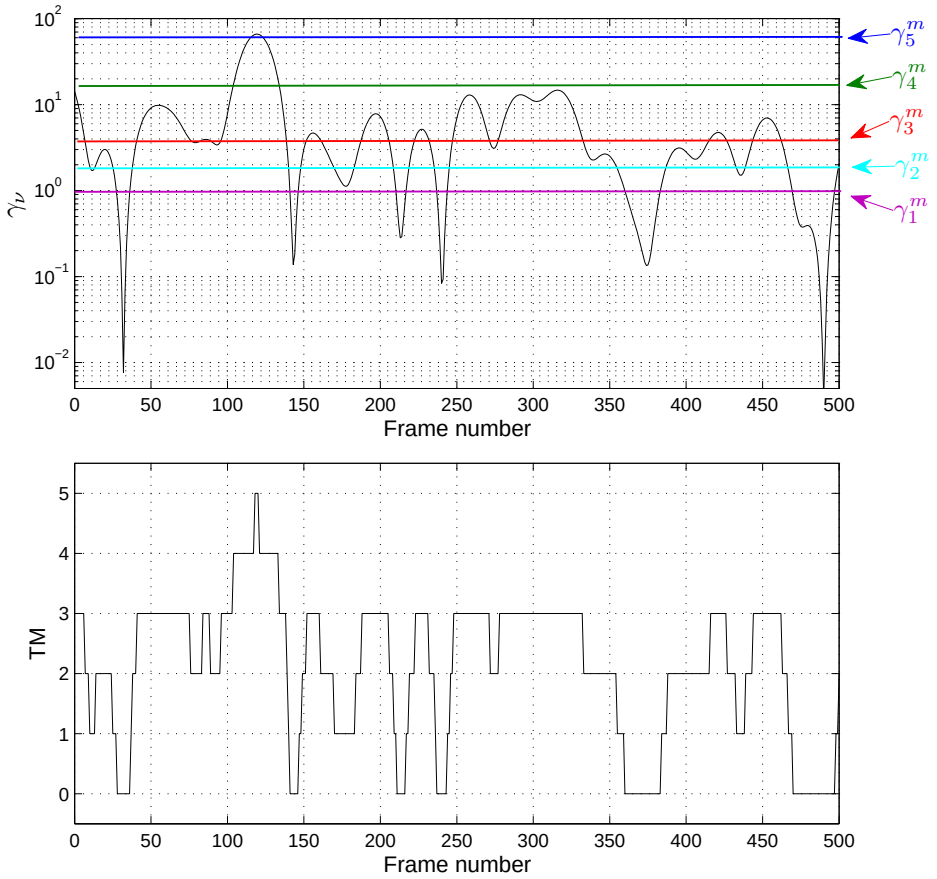


Figure 2.4: Example: instantaneous SNR and TM selection.

transmitter by using the feedback CSI channel without error and/or latency. Imperfections in the AMC scheme caused by the delay in the CSI feedback channel could be accounted for simply by using the methodology developed in Chapter 6.

Figure 2.4 shows an example of the time evolution of the instantaneous SNR and the selected TM at each frame interval. A wireless Rayleigh channel with $f_d T_f = 0.02$ and an average received SNR $\overline{\gamma} = 8$ dB has been considered. The frame period has a duration $T_f = 2$ ms. An AMC scheme with six available TMs has been supposed, with the following switching thresholds: $\gamma_0^m = 0, \gamma_1^m = 0.9906, \gamma_2^m = 1.9444, \gamma_3^m = 3.8598, \gamma_4^m = 16.7209, \gamma_5^m = 65.1990, \gamma_6^m = \infty$ (how these thresholds are obtained will be thoroughly explained in following chapters). Thus, as it can be observed in Fig. 2.4, the corresponding TMs are selected according to the partition

$$\begin{aligned}\mathbf{\Gamma}^m = &\{[0, 0.9906), [0.9906, 1.9444), [1.9444, 3.8598), [3.8598, 16.7209),\\ &[16.7209, 65.1990), [65.1990, \infty)\}.\end{aligned}$$

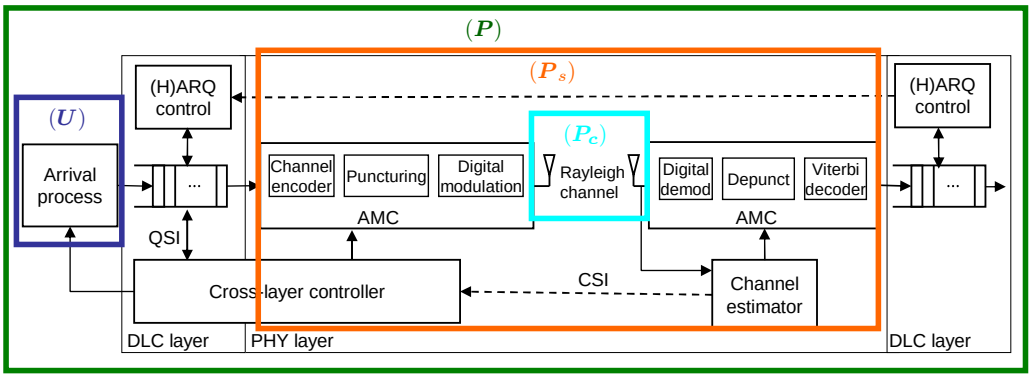


Figure 2.5: Markov chain-based system model approach.

2.2 Discrete time Markov chain model and analysis

Typically, PHY layer channel processes are expressed as general correlated stochastic processes. Moreover, DLC layer processes, such as queue arrival and departure, are described using Markovian models (Alfa, 2010). Hence, many cross-layer analysis techniques use Markov models to accurately describe the system dynamics and to simplify analysis. A key advantage of the Markov model-based system description is that it enables formulation and solution of the optimal cross-layer problem with standard optimization techniques. In this research work, an analytical link-level queueing model will be built up for the QoS-constrained AMC/(H)ARQ-based wireless system described in Section 2.1. This model will be obtained based on the use of DTMCs, through a comprehensive consideration of the following issues:

- the packet arrival process (blue rectangle in Fig. 2.5);
- the Rayleigh flat-fading channel model, which relies on the amplitude and the rate-of-change of the fading envelope (cyan rectangle in Fig. 2.5);
- the PHY layer, modeled taking into account the AMC scheme and the wireless channel statistics (orange rectangle in Fig. 2.5);
- the queueing process induced by the PHY layer (channel and AMC scheme), the packet arrival process and the (H)ARQ protocol (green rectangle in Fig. 2.5).

The PHY layer model is based on the use of an enhanced first-order two-dimensional finite state Markov chain (2D-FSMC) model for the wireless flat-fading channel. In contrast to previous proposals in the literature, this model relies on the amplitude and the rate-of-change of the fading envelope. Moreover, it incorporates the implementation of the AMC scheme used in the TM selection subsystem, which is designed independently from the channel model. Additionally, based on this PHY layer model, analytical link-level queueing models of a point-to-point adaptive multi-rate wireless system using either infinitely persistent or truncated (H)ARQ are proposed in forthcoming chapters. Using this approach,

analytical expressions for performance metrics such as throughput, average packet delay, queue length and packet loss rate, both due to buffer overflow and due to exceeding the maximum number of allowed retransmissions, are derived. The importance of each metric depends on the required QoS, which is service-specific (for voice traffic, the delay has clearly to be minimized, whereas for file transfer services it is more interesting to maximize the throughput). The resulting analytical link-level queueing model, which is based on the use of DTMCs, provides then a roadmap towards effective cross-layer multidimensional design strategies that are thoroughly explored. Simplified approaches of this problem are also proposed, resulting in simpler schemes at the cost of a performance loss. Particularly, it is shown that the analytical results not only allow a-priori design decisions to be made, but they also provide an insight that enables the derivation of dynamic reconfiguration strategies that take into account instantaneous CSI at the PHY layer and queue state information (QSI) at the DLC layer.

Unlike the PHY layer and the queueing process, which depend on the considered (H)ARQ protocol and, thus, will be carefully analyzed in the corresponding chapters, the packet arrival and the channel models are independent of the considered (H)ARQ scheme. For that reason, the following sections are devoted to the description of the packet generation model and the wireless fading channel model that will be used throughout this research work.

2.3 Discrete batch Markovian arrival process

As in (Kim and Krunz, 2000; Le et al., 2006b), it is assumed that the packet generation model adheres to a special case of the discrete batch Markovian arrival process (D-BMAP) (Blondia, 1993). The distribution of the number of customer arrivals per slot depends on a background state that can be characterized by a first-order Markov chain. The transition probability matrix of this Markov chain is

$$\boldsymbol{U} = \sum_{a=0}^{\infty} \boldsymbol{U}_a = \begin{bmatrix} u(0,0) & \cdots & u(\mathcal{A}-1,0) \\ \vdots & & \vdots \\ u(0,\mathcal{A}-1) & \cdots & u(\mathcal{A}-1,\mathcal{A}-1) \end{bmatrix}, \tag{2.2}$$

where $u(a_\mu, a_{\mu'})$ denotes the probability of a transition from phase a_μ to phase $a_{\mu'}$, with phase a_μ characterized by a batch arrival of size a_μ packets, as illustrated in Fig 2.6. The sub-stochastic[1] matrices $\boldsymbol{U}_a$, for all $a \in \{0, \ldots, \mathcal{A}-1\}$ are constructed by keeping only the $(a+1)$th row of $\boldsymbol{U}$ and setting all other rows to zero, and by definition of the D-BMAP, $\boldsymbol{U}_a = \boldsymbol{0}$ for all $a \geq \mathcal{A}$.

Owing to the Markovian property of the arrival process we have that $\boldsymbol{\omega} = \boldsymbol{\omega}\boldsymbol{U}$ and $\boldsymbol{\omega}\mathbf{1}_{\mathcal{A}} = 1$, where $\boldsymbol{\omega}$ denotes the D-BMAP steady-phase probability vector and $\mathbf{1}_{\mathcal{A}}$ is an all-ones column vector of length $\mathcal{A}$. Then, the average arrival rate λ can be calculated as

$$\lambda = \boldsymbol{\omega} \sum_{a=0}^{\mathcal{A}-1} a\, \boldsymbol{U}_a \mathbf{1}_{\mathcal{A}}. \tag{2.3}$$

[1] A sub-stochastic matrix is a square matrix with nonnegative entries so that every row adds up to at most 1.

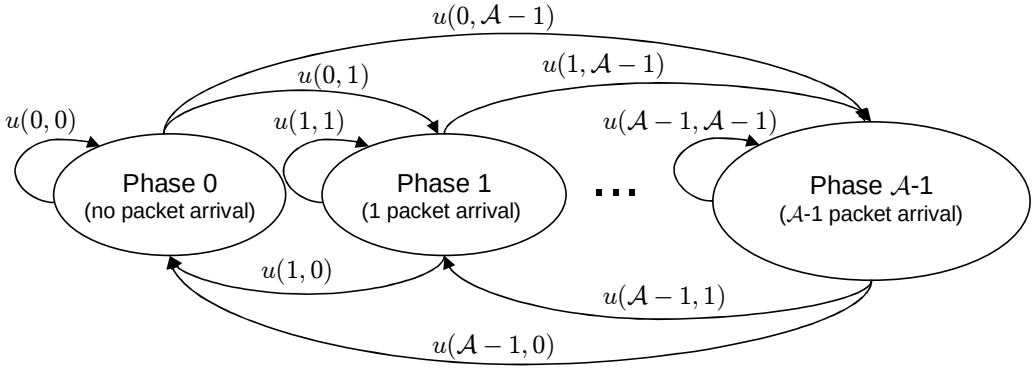


Figure 2.6: Transition state diagram of the discrete batch Markovian arrival process.

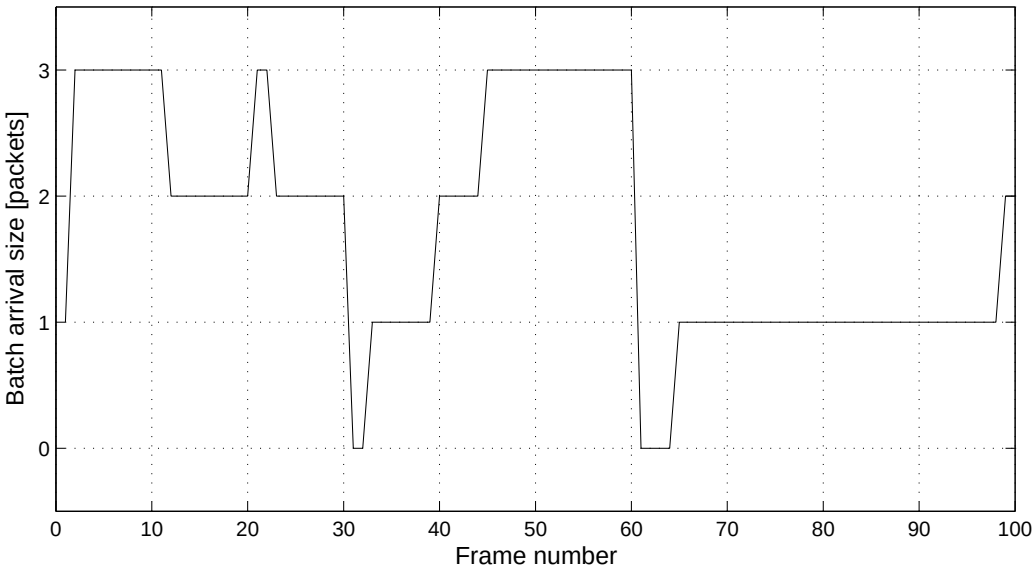


Figure 2.7: Example: batch arrival process.

The matrix $\boldsymbol{U}$ introduces memory in the packet arrival process. In fact, if the elements in the diagonal have a value greater than 0.5, then the probability of remaining on a given phase is larger than the probability o transitioning to another phase, thus leading to a high probability of generating packet batches of a given length. This behaviour can be observed in Fig. 2.7, which plots the time evolution of the batch arrival size of a D-BMAP characterized with the transition probability matrix

$$\boldsymbol{U} = \begin{bmatrix} 0.8 & 0.1 & 0.05 & 0.05 \\ 0.05 & 0.8 & 0.1 & 0.05 \\ 0.05 & 0.05 & 0.8 & 0.1 \\ 0.05 & 0.05 & 0.1 & 0.8 \end{bmatrix} \tag{2.4}$$

corresponding to an average arrival rate $\lambda = 1.62$ packets per frame.

2.4 Improved Markov-based channel model

Often, the communication medium of a wireless system is characterized by a time-varying fading channel with memory. When designing these systems, efficient exploitation of the fading channel is supported by the availability of accurate and tractable channel models. This is specially true in designing radio resource management algorithms and selecting suitable transmission policies to match the instantaneous fading channel conditions. Because of its widespread applicability and acceptance, the most widely used flat fading channel model is a statistical model for the received signal amplitude that was proposed by Clarke (1968). In the case of isotropic two-dimensional scattering with an omnidirectional receiving antenna (ISORA), it can be shown (Gans, 1972; Tan and Beaulieu, 2000) that this fading model is a complex stationary process $X(t) = X_I(t) + jX_Q(t)$ whose quadrature components are zero mean mutually independent Gaussian processes with identical ACF

$$R_{X_I}(\tau) = R_{X_Q}(\tau) = \sigma^2 J_0\left(2\pi f_d \tau\right),$$

where σ^2 is the variance of each quadrature component and $J_0(\cdot)$ is the zero-order Bessel function of the first kind. The envelope of this fading process, $R(t) = |X(t)| = \sqrt{X_I^2(t) + X_Q^2(t)}$, is a stationary process whose first-order distribution is Rayleigh, with ACF given by

$$R_R(\tau) = \frac{\pi\sigma^2}{2} {}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}; 1; R_X(\tau)^2\right) \tag{2.5}$$

where ${}_2F_1(\cdot,\cdot;\cdot;\cdot)$ denotes the hypergeometric function (Abramowitz and Stegun, 1972).

Although this is an accurate *continuous* model for Rayleigh fading channels with memory, its main drawback is its difficult mathematical tractability in the design and performance evaluation of upper layer protocols (i.e. ARQ, medium access control (MAC), transport layer protocols, cross-layer design). *Discrete* Markov models for Rayleigh fading channels have been often adopted (i.e. (Chaskar et al., 1999; Femenias, 2005; Ishizaki and Hwang, 2007; Karmokar et al., 2006a,b; Le et al., 2006a,b; Liu et al., 2005a,b; Turin and Zorzi, 2002; Wang et al., 2007; Zhang and Du, 2007; Zorzi, 1997)). Starting from early works of Gilbert (1960) and Elliott (1963), an effective representation of flat fading channels with memory is given through the use of first-order FSMCs. Typically, a first-order FSMC is build by partitioning the range of received SNRs into a set of non-overlapping intervals defined by the partition

$$\mathbf{\Gamma}^c = \left\{[\gamma_0^c, \gamma_1^c), [\gamma_1^c, \gamma_2^c), \ldots, [\gamma_{K-1}^c, \gamma_K^c)\right\} \tag{2.6}$$

with $\gamma_0^c = 0$ and $\gamma_K^c = \infty$. Each state of the channel corresponds to one of these intervals. All the partition intervals (channel states) can then be represented by a codebook of nominal values of SNR, namely

$$\mathbf{\Psi}^c = \left\{\Psi_0^c, \Psi_1^c, \ldots, \Psi_{K-1}^c\right\}, \tag{2.7}$$

or by a codebook of related nominal channel quality measures (i.e. bit error rate (BER), packet error rate (PER)). The Gilbert-Elliott model is a two-state Markov channel where each state corresponds to a specific channel quality that is either noiseless or totally noisy. Wang and Moayeri (1995) proposed the extension from a two-state channel to a finite-state one, constructed by optimizing the system performance in the sense of maximizing the channel capacity. The partitioning schemes used by Wang and Moayeri were derived from the quantization analysis for pulse code modulation techniques but how these schemes are related to real channel conditions was not specified or validated with real or simulated channel data. Adopting the view of Wang and Moayeri, Zhang and Kassam (1999) proposed a first-order FSMC model where the number of states and partition intervals were designed according to a criterion based on the average duration of each state. Once again, the partition scheme proposed by Zhang and Kassam was not validated with real or simulated channel data. Another approach for the SNR partitioning that has been widely used in literature (see for example, (Babich and Lombardi, 2000; Bergamo et al., 2002; Tan and Beaulieu, 2000)) is to select the partition intervals in such a way that the steady state probabilities are all equal.

The first-order FSMC has gained a wide acceptance and seems to be finding a growing number of applications (see, e.g., (Ishizaki and Hwang, 2007; Karmokar et al., 2006a,b; Le et al., 2006a,b; Liu et al., 2005a,b; Wang et al., 2007; Zhang and Du, 2007)). Nevertheless, although a majority of the literature focuses on first-order FSMCs, presumably because of their simplicity and analytical tractability, Tan and Beaulieu (2000) show that first-order Markov processes having an exponentially decaying ACF can not fit the hypergeometric ACF of the simulated process. In fact, the significant mismatch of exponential and hypergeometric ACFs could compromise the design of higher layer protocols, whose performance may depend on a time scale for which this mismatch plays an important role.

Work has also been done to examine higher-order Markov models. Babich and Lombardi (2000) and Babich et al. (2000) proposed the context tree pruning (CTP) algorithm as a method to build higher-order Markov models to improve the first-order FSMC description. Turin and van Nobelen (1998) proposed the use of hidden Markov models (HMMs) of greater order to model the fading channel. An improved Markov model for Rayleigh fading was proposed by Bergamo et al. (2002), in which both the amplitude and the speed of variation of the fading envelope are jointly used to represent the Markov state. Obviously, as the order of the Markov models increases, so does the complexity and intractability of the model.

In this research work a simple first-order 2D-FSMC model for the Rayleigh flat-fading channel is proposed that is able to improve the ACF fitting of first-order FSMCs. This model uses a partitioning scheme that has been derived from the Max-Lloyd classical theory of optimal scalar quantization design (Lloyd, 1982; Max, 1960) and the Markov states are represented by the current fading envelope amplitude and the relative level (up, down) of previous samples of the envelope. In order to prove the validity and accuracy of our model we use the stochastic analysis proposed by Tan and Beaulieu (2000).

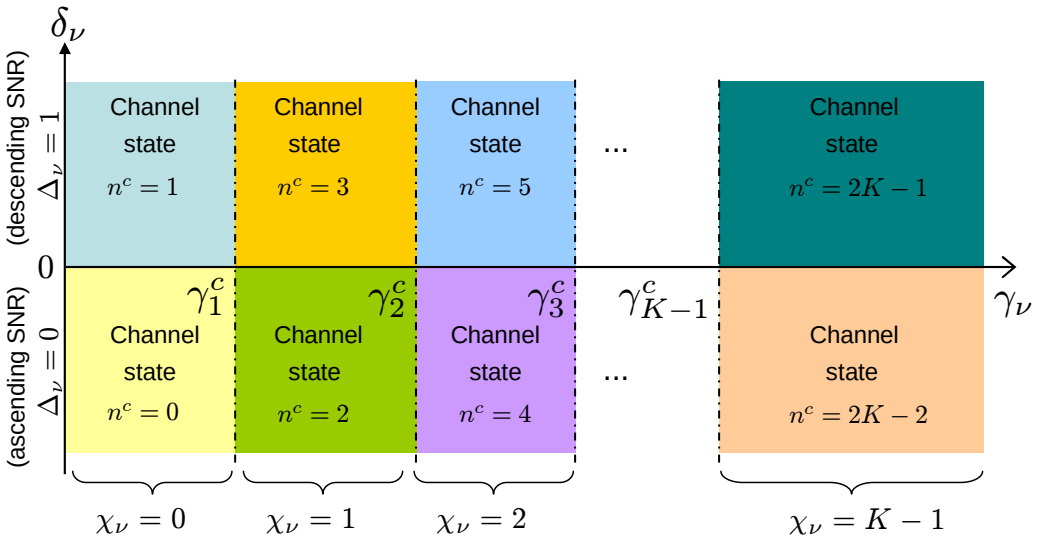


Figure 2.8: Two-dimensional channel partitioning.

2.4.1 Improved first-order two-dimensional Markov model

Let us consider the Rayleigh block-fading channel quantities γ_ν and $\delta_\nu = \gamma_{\nu-1} - \gamma_\nu$, representing the instantaneous SNR and the rate of change of the SNR. Let us also partition the ranges of γ_ν and δ_ν into sets of non-overlapping two-dimensional cells defined by the partitions $\mathbf{\Gamma}^c$, defined in (2.6), and

$$\mathbf{\Delta} = \{(-\infty, 0), [0, \infty)\}, \tag{2.8}$$

respectively. Thus, a first-order two-dimensional Markov channel model can be defined where each two-dimensional state is defined by the amplitude and the rate of change of the instantaneous SNR at the output of the fading channel. Each state of the channel corresponds to one of the rectangular cells represented in Fig. 2.8. That is, the Markov chain state of the channel at time instant $t = \nu T_f$ can be denoted as $\boldsymbol{\zeta}_\nu = (\chi_\nu, \Delta_\nu)$, $\nu = 0, 1, \ldots, \infty$, where $\chi_\nu \in \{0, \ldots, K-1\}$ with $\chi_\nu = k$ if and only if $\gamma_k^c \leq \gamma_\nu < \gamma_{k+1}^c$ and $\Delta_\nu \in \{0, 1\}$ is used to denote the *up* or *down* characteristic of the instantaneous SNR in time frame interval $t = (\nu - 1)T_f$ (if $\gamma_\nu < \gamma_{\nu-1}$ then $\delta_\nu > 0$, the instantaneous SNR is descending ($\Delta_\nu = 1$) and it can be tagged as *down*; on the contrary, if $\gamma_\nu \geq \gamma_{\nu-1}$ then $\delta_\nu \leq 0$, the instantaneous SNR is ascending ($\Delta_\nu = 0$) and it can be tagged as *up*). Equivalently, at any time instant $t = \nu T_f$ the channel state can be univocally characterized by an integer number $n_\nu^c = 2\chi_\nu + \Delta_\nu$ and obviously, $n_\nu^c \in \{0, \ldots, 2K-1\}$, with $\chi_\nu = \lfloor n_\nu^c/2 \rfloor$, where $\lfloor x \rfloor$ denotes the nearest integer less than or equal to x.

In our approach the partition $\mathbf{\Gamma}^c$ is designed assuming that the observable dummy output of our improved first-order two-dimensional Markov model at time instant $t = \nu T_f$ belongs to a codebook of nominal values of SNR given by $\mathbf{\Psi}^c$, defined in (2.7). The Max-Lloyd algorithm (Lloyd, 1982; Max, 1960), developed for the optimum design of non-uniform quantizers, is then used to determine the optimum partition and (dummy) codebook in the sense that it minimizes the mean square error between γ_ν and the (dummy) quantizer output.

Figure 2.9 illustrates the modeling of a wireless Rayleigh channel realization corresponding to $f_d T_f = 0.02$ and an average received SNR $\overline{\gamma} = 8$ dB, with a frame period $T_f = 2$ ms. Considering $K = 5$ channel states, the obtained partition $\mathbf{\Gamma}^c$ is

$$\begin{aligned}\mathbf{\Gamma}^c = &\{[0, 3.7837), [3.7837, 8.5343), [8.5343, 14.9385), [14.9385, 24.9534),\\ &[24.9534, \infty)\}.\end{aligned}$$

For example, let us consider time instant $t = 100T_f$, that is, frame number $\nu = 100$. It can be observed that the instantaneous SNR γ_{100} falls into the partition interval $[14.9385, 24.9534)$, with a downward trend. Hence, $\chi_{100} = 3$ and $\Delta_{100} = 1$, which means that the state of the channel can be denoted as $\boldsymbol{\zeta}_{100} = (3, 1)$. Equivalently, the channel state can be characterized by the integer number $n^c_{100} = 2 \cdot 3 + 1 = 7$. It is worth noting that the model introduced by Liu et al. (2005b) is an AFSMC model and, consequently, it does not take into consideration the *up/down* characteristic of the instantaneous SNR. Moreover, it defines the partition $\mathbf{\Gamma}^c$ using the AMC switching thresholds (i.e., $\mathbf{\Gamma}^c = \mathbf{\Gamma}^m$) and, therefore, it does not consider the statistics of the wireless channel. As a result, the channel state at a certain time instant corresponds to the selected TM, as shown in the example of Fig. 2.4.

Statistical characterization of the Markovian model

The joint probability density function (pdf) of the random variables γ_ν and $\gamma_{\nu-1}$ can be expressed as (Tan and Beaulieu, 1997, eq. (1))

$$p_{\gamma_\nu, \gamma_{\nu-1}}(x, y) = \frac{e^{-(x+y)/[\bar{\gamma}(1-\rho)]}}{\bar{\gamma}^2(1-\rho)} I_0\left(\frac{2\sqrt{\rho x y}}{\bar{\gamma}(1-\rho)}\right), \tag{2.9}$$

where $x \geq 0$, $y \geq 0$, $I_0(\cdot)$ is the modified zero-order Bessel function of the first kind, $\bar{\gamma}$ denotes the average SNR and

$$\rho = \frac{\mathrm{cov}(\gamma_\nu, \gamma_{\nu-1})}{\sqrt{\mathrm{var}(\gamma_\nu)\mathrm{var}(\gamma_{\nu-1})}} \tag{2.10}$$

denotes the normalized correlation between the random variables γ_ν and $\gamma_{\nu-1}$. Assuming the use of Clarke's statistical Rayleigh fading process to model the wireless flat-fading channel (Clarke, 1968), the normalized correlation coefficient can be calculated as $\rho = J_0(2\pi f_d T_f)$. Using (2.9), the channel will be in a state $n^c \in \{0, \ldots, 2K-1\}$ with probability

$$P^c(n^c) = \begin{cases} \int_{\gamma^c_{n^c/2}}^{\gamma^c_{n^c/2+1}} \int_0^x p_{\gamma_\nu, \gamma_{\nu-1}}(x, y)\, dydx, & n^c \text{ even} \\ \int_{\gamma^c_{(n^c-1)/2}}^{\gamma^c_{(n^c+1)/2}} \int_x^{+\infty} p_{\gamma_\nu, \gamma_{\nu-1}}(x, y)\, dydx, & n^c \text{ odd} \end{cases} \tag{2.11}$$

The channel transition probability matrix $\boldsymbol{P}_c = \left[P^c_{i,j}\right]_{0 \leq i,j \leq 2K-1}$ has stationary transition probabilities

$$\begin{aligned} P^c_{i,j} &= Pr\left\{n^c_\nu = j | n^c_{\nu-1} = i\right\} \\ &= \frac{\int_{x_I}^{x_S} \int_{y_I}^{y_S} \int_{z_I}^{z_S} p_{\gamma_\nu, \gamma_{\nu-1}, \gamma_{\nu-2}}(x, y, z)\, dzdydx}{P^c(i)}, \end{aligned} \tag{2.12}$$

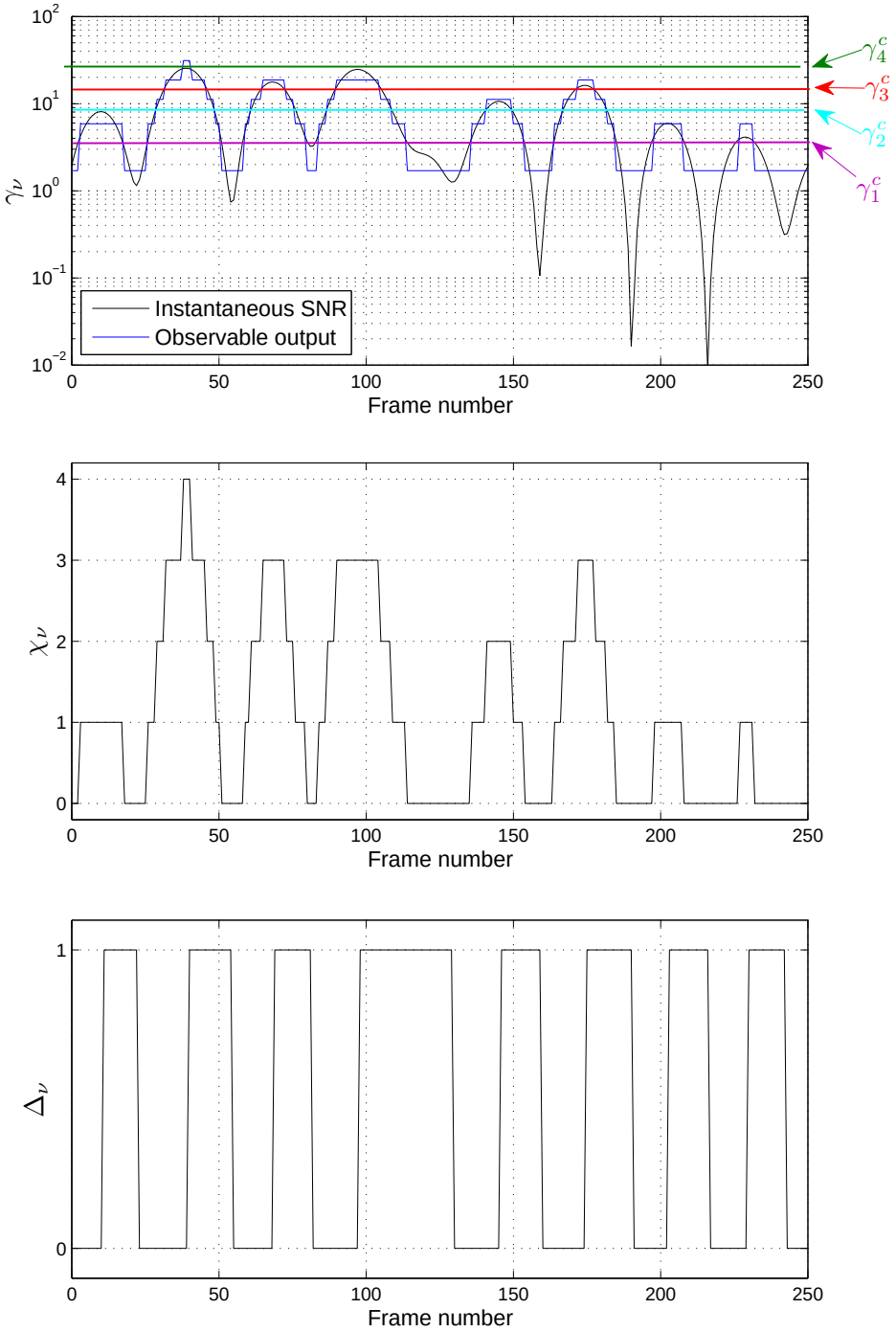


Figure 2.9: Example: modeling of the channel.

where the integration intervals are

$$[x_I, x_S] = \begin{cases} \left[\gamma^c_{\frac{j}{2}}, \gamma^c_{\frac{j}{2}+1}\right], \, j \text{ even} \\ \\ \left[\gamma^c_{\frac{j-1}{2}}, \gamma^c_{\frac{j+1}{2}}\right], \, j \text{ odd} \end{cases} \tag{2.13}$$

$$[y_I, y_S] = \begin{cases} \left[\gamma^c_{\frac{i}{2}}, \max\left\{\gamma^c_{\frac{i}{2}}, \min\left\{x, \gamma^c_{\frac{i}{2}+1}\right\}\right\}\right], \, i \text{ even } j \text{ even} \\ \\ \left[\min\left\{\gamma^c_{\frac{i}{2}+1}, \max\left\{x, \gamma^c_{\frac{i}{2}}\right\}\right\}, \gamma^c_{\frac{i}{2}+1}\right], \, i \text{ even } j \text{ odd} \\ \\ \left[\gamma^c_{\frac{i-1}{2}}, \max\left\{\gamma^c_{\frac{i-1}{2}}, \min\left\{x, \gamma^c_{\frac{i+1}{2}}\right\}\right\}\right], \, i \text{ odd } j \text{ even} \\ \\ \left[\min\left\{\gamma^c_{\frac{i+1}{2}}, \max\left\{x, \gamma^c_{\frac{i-1}{2}}\right\}\right\}, \gamma^c_{\frac{i+1}{2}}\right], \, i \text{ odd } j \text{ odd} \end{cases} \tag{2.14}$$

$$[z_I, z_S] = \begin{cases} [0, y], \, i \text{ even} \\ \\ [y, +\infty], \, i \text{ odd} \end{cases} \tag{2.15}$$

and $p_{\gamma_\nu, \gamma_{\nu-1}, \gamma_{\nu-2}}(x, y, z)$ denotes the joint pdf of the random variables γ_ν, $\gamma_{\nu-1}$ and $\gamma_{\nu-2}$ that, using (Chen and Tellambura, 2005, eqn. (2)) and making a change of variables, can be expressed as

$$\begin{aligned} p_{\gamma_\eta, \gamma_{\eta-1}, \gamma_{\eta-2}}(x, y, z) = & \frac{\det(\boldsymbol{W})}{\bar{\gamma}^3} \, e^{-(xw_{11}+yw_{22}+zw_{33})/\bar{\gamma}} \\ & \times \sum_{k=0}^{\infty} \varepsilon_k (-1)^k \cos(k\chi) I_k\left(\frac{2|w_{12}|}{\bar{\gamma}}\sqrt{xy}\right) \\ & \times I_k\left(\frac{2|w_{23}|}{\bar{\gamma}}\sqrt{yz}\right) I_k\left(\frac{2|w_{31}|}{\bar{\gamma}}\sqrt{zx}\right), \end{aligned} \tag{2.16}$$

where $x \geq 0$, $y \geq 0$, $z \geq 0$, $I_k(\cdot)$ is the modified order-k Bessel function of the first kind, ε_k is the Neumann factor ($\varepsilon_0 = 1$, $\varepsilon_k = 2$ for $k = 1, 2, \ldots$), $\boldsymbol{W}$ can be calculated as the inverse of $\boldsymbol{R}$, i.e., $\boldsymbol{W} = \boldsymbol{R}^{-1}$ with elements $w_{il} = |w_{il}|e^{j\chi_{il}}$, with $\boldsymbol{R}$ being the positive definite covariance matrix of the underlying Gaussian processes and $\chi = \chi_{12} + \chi_{23} + \chi_{31}$. Assuming the use of Clarke's statistical Rayleigh fading process to model the wireless flat-fading channel, the elements of $\boldsymbol{R}$ can be calculated as $R_{i,j} = J_0(2\pi|i-j|f_d T_f)$. The steady-state probability vector can be calculated because the transition probability matrix $\boldsymbol{P}_c$ and steady-state probability vector $\boldsymbol{\pi}_c = \left[\pi^c_0 \cdots \pi^c_{2K-1}\right]$ satisfy $\boldsymbol{\pi}_c \boldsymbol{P}_c = \boldsymbol{\pi}_c$ along with the normalization condition $\boldsymbol{\pi}_c \mathbf{1}_{2K} = 1$. The integrals involved in the calculation of $P^c(n^c)$ and $P^c_{i,j}$ can be computed either by Monte-Carlo numerical simulations or by using the series representation of the modified order-k Bessel function of the first kind (Abramowitz and Stegun, 1972, eq. (9.6.10)) (see also (Chen and Tellambura, 2005; Tan and Beaulieu, 1997)).

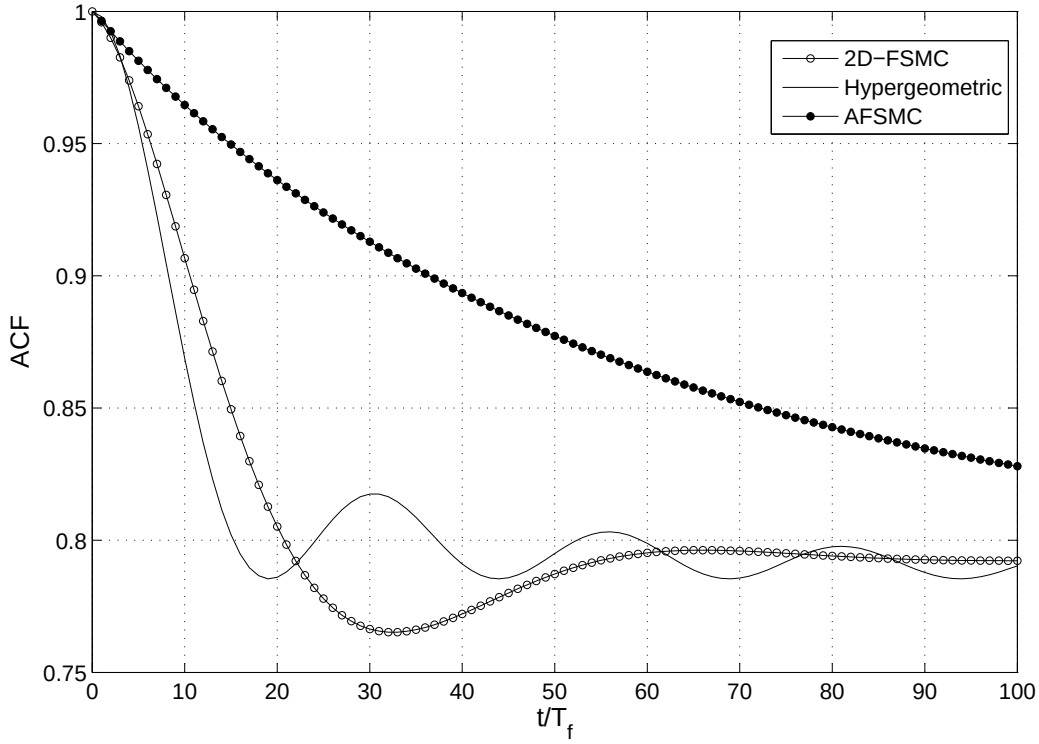


Figure 2.10: ACF comparison.

Model assessment

As suggested by Tan and Beaulieu (2000), the stochastic process theory can be used to compare our 2D-FSMC and the ISORA fading model, by contrasting the ACF for each Rayleigh process.

The ACF can be calculated using (Tan and Beaulieu, 2000, eq. (10))

$$\begin{aligned} R_R[m] &= \sum_{i=0}^{2K-1} \sum_{j=0}^{2K-1} \sqrt{\Psi^c_{\lfloor i/2 \rfloor} \Psi^c_{\lfloor j/2 \rfloor}} \Pr\left\{n^c_m = j, n^c_0 = i\right\} \\ &= \sum_{j=0}^{2K-1} \sqrt{\Psi^c_{\lfloor j/2 \rfloor}}\, \pi^c_j \sum_{i=0}^{2K-1} \sqrt{\Psi^c_{\lfloor i/2 \rfloor}}\, {P^c_{i,j}}^{(m)}, \end{aligned} \tag{2.17}$$

where ${P^c_{i,j}}^{(m)}$ denotes the (i,j)th element of the matrix $(\boldsymbol{P}_c)^m$.

As it has been previously remarked in Section 2.4, first-order AFSMCs, having an exponentially decaying ACF, can not fit the hypergeometric ACF of the statistical Rayleigh fading process used to model wireless flat-fading channels (Clarke, 1968; Tan and Beaulieu, 2000). Figure 2.10 plots the ACF obtained using an AFSMC and the proposed first-order 2D-FSMC, both with $K = 10$, to model a wireless fading channel characterized by a maximum Doppler frequency $f_d = 10$ Hz, with a frame duration $T_f = 2$ ms. It also shows the theoretical hypergeometric ACF. It is clearly observed that the ACF of the one-dimensional Markov model is far from the hypergeometric ACF of the statistical Rayleigh fad-

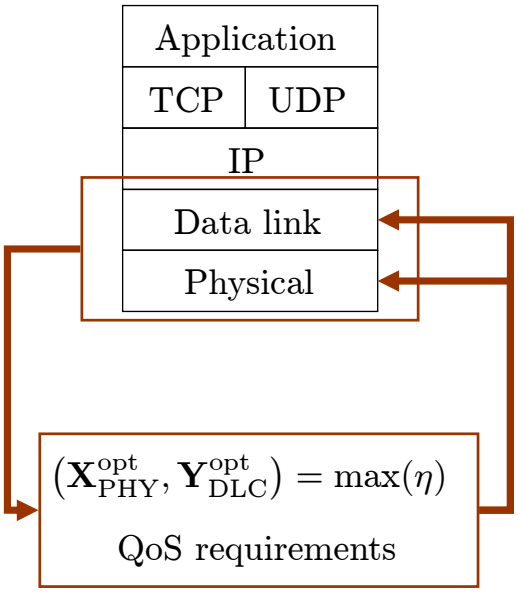


Figure 2.11: Cross-layer perspective.

ing process, whereas the improved first-order 2D-FSMC model provides a much better fit.

In Subsection 2.2 it has been mentioned that, in the following chapters, based on this channel model, a 2D-FSMC characterization of the PHY layer will be developed. Afterwards, using this PHY layer representation, different link-level queueing models will be proposed and analytical expressions for performance metrics such as throughput, average packet delay, queue length and packet loss rate will be derived. In Chapter 3, which is devoted to the analysis of an infinitely persistent ARQ protocol, with the aim of validating the developed 2D-FSMC PHY layer model, the analytical performance expressions for the proposed queueing model will be compared with those used by Le et al. (2006b), which were based on the PHY layer first-order AFSMC model introduced by Liu et al. (2005b). Results obtained using these models will be confronted with computer simulations obtained using Clarke's statistical Rayleigh fading process. Numerical results will show that, compared to the first-order AFSMC approach, our proposed 2D-FSMC model more faithfully reproduces the performance metrics of a *real* system based on Clarke's statistical Rayleigh fading model.

2.5 Cross-layer perspective

As it has been detailed in Chapter 1, one of the main objectives of this research work is the development of a novel analytical framework for the cross-layer modeling and design of wireless networks combining AMC at the PHY layer with an (H)ARQ protocol at the DLC layer, which will imply the exchange of layer-specific parameters. Particularly, as illustrated in Fig. 2.11, the cross-layer design will optimally determine specific design parameters at the PHY layer $\mathbf{X}_{\mathrm{PHY}}^{\mathrm{opt}}$ (e.g.,

AMC switching thresholds or set of AMC useful modes) and at the DLC layer $\mathbf{Y}_{\text{DLC}}^{\text{opt}}$ (e.g., average packet arrival rate or buffer size) that maximize the average throughput η of the system, while satisfying prescribed QoS requirements. These reconfiguration strategies will take into account instantaneous CSI at the PHY layer and QSI at the DLC layer.

2.6 Chapter summary

In this chapter, the generic system block diagram under investigation has been introduced and the main characteristics of the transmitter, the channel and the receiver have been described. The Markov-based approach that will be used throughout this research work has been presented, and the modeling of the packet generation process and of the wireless channel have been described. It has been demonstrated that, in comparison to the one-dimensional Markov models, the proposed first-order two-dimensional Markov model of the channel provides a much better fit to the hypergeometric ACF of the statistical Rayleigh fading process. Finally, the cross-layer perspective, which will be conceived as a QoS constrained optimization problem to be developed in the next chapters, has been described.

CHAPTER 3

Systems based on an infinitely persistent ARQ protocol

This chapter proposes a novel framework for the cross-layer design, analysis and optimization of wireless networks combining AMC at the PHY layer with an infinitely persistent Type-I hybrid FEC/ARQ protocol at the DLC layer. Most previous works rely on AFSMCs to model the PHY layer. It is shown that these models present several deficiencies that could compromise the design of higher-layer protocols. Thus, using the channel model developed in Section 2.4, a PHY layer first-order two-dimensional Markov model using both the amplitude and the rate-of-change of the fading envelope is presented. Based on this multidimensional PHY layer Markov model, the QoS performance at the DLC layer is investigated through the use of two different approaches. The first one relies on an analytical framework based on a DTMC that jointly describes the statistical behaviour of the arrival process, the queueing system and the PHY layer. The second one is based on the effective bandwidth and effective capacity theories. Both the DTMC-based and the effective bandwidth/capacity-based approaches are analyzed and compared in combination with our proposed PHY layer first-order two-dimensional Markov model in a cross-layer design aiming at maximizing the average throughput of the system under prescribed QoS requirements. To that end, multidimensional and simplified bidimensional constrained optimization problems are presented and different options for the AMC scheme are proposed. Numerical results show that the proposed analytical framework represents a significant improvement over previous models.

3.1 Introduction

As it has been mentioned in Chapter 1, many research efforts focus on cross-layer designs combining AMC schemes with an ARQ protocol (see, e.g., (Ishizaki

and Hwang, 2007; Le et al., 2006a,b; Liu et al., 2004, 2005a,b; Poggioni et al., 2007; Wang et al., 2007)). One of the main drawbacks of these research efforts, however, is their reliance on first-order AFSMCs to model the wireless fading channel. In Section 2.4 it has been pointed out that the exponentially decaying ACF of the first-order AFSMCs can not fit the hypergeometric ACF of the statistical Rayleigh fading process, which could compromise the design of higher layer protocols. Moreover, the analytical models used in (Liu et al., 2004, 2005a,b) present several deficiencies that have propagated to many other research works (see, e.g., (Ishizaki and Hwang, 2007; Le et al., 2006a,b; Poggioni et al., 2007; Wang et al., 2007)). First of all, the authors take for granted that, assuming slow fading conditions, transitions will happen only between adjacent states of the AFSMC. However, in many cases of practical interest, this assumption does not faithfully correspond to the real channel behaviour. Second, as it will be shown in Subsection 3.4.2, the assumptions used to perform the average PER calculation (see, e.g., (Liu et al., 2005a, eq. (5))) are of limited applicability in a wide range of channel states that must be confronted in solving the cross-layer optimization problem. Finally, the AMC threshold searching algorithm used in the TM selection and AFSMC design assumes that all *possible* TMs can always be considered as *useful* TMs. However, depending on the channel conditions and the QoS requirements, some of the *possible* TMs may be declared *useless* and thus, only a limited set of *useful* TMs will be available to the AMC scheme. In order to overcome the aforementioned problems, the enhanced first-order two-dimensional Markov model for the Rayleigh flat-fading channel that has been developed in Chapter 2 will be used in combination with the AMC scheme, which will be designed independently from the channel model, to build up a PHY layer model that relies on the amplitude and the rate-of-change of the fading envelope.

Once the PHY layer has been properly modeled, the DLC layer behaviour has to be introduced in the analysis. Basically, two approaches have been proposed in the literature. The first one, founded on the research work of Le et al. (2006a) and Le et al. (2006b), proposes an analytical framework using a DTMC that jointly describes the statistical behaviour of the arrival process, the queueing system and the PHY layer. Traffic burstiness was considered in this proposal, where the authors analyzed an infinitely persistent "pure" ARQ-based error control scheme. Performance metrics such as the average throughput, average delay or average packet loss probability can be determined through the calculation of the global Markovian transition probability matrix of the system. This model provides valuable radio link-level performance measures but, unfortunately, its complexity increases with the number of states of the arrival process, the number of states of the PHY layer Markov model or the maximum queue length of the system. The second approach follows a completely different methodology that relies on the well-known effective bandwidth theory (see, e.g., (Chang, 2000; Chang and Thomas, 1995)) and its closely related effective capacity theory (Wu and Negi, 2003). Previous works (see, e.g., (Ishizaki and Hwang, 2007; Tang and Zhanh, 2007; Zhang and Letaief, 2006)) have used this combined effective bandwidth/capacity approach to perform cross-layer design between the PHY layer and DLC layer but again, all of them are based on the use of AFSMCs to model the PHY layer. Compared with the analytical framework based on the finite-state

Table 3.1: 802.11a TMs with convolutionally coded modulation.

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
Modulation	BPSK	QPSK	QPSK	16QAM	64QAM
Code rate $R_c^{(n)}$	1/2	1/2	3/4	3/4	3/4
R_n (bits/symbol)	0.50	1.00	1.50	3.00	4.50
a_n	274.723	90.251	67.618	53.399	35.351
g_n	7.993	3.500	1.688	0.376	0.090
γ_{p_n} (dB)	-1.533	1.094	3.972	10.249	15.978

DTMCs, this approach is not only less complex but it is also easily scalable. As a consequence, it can be readily extended to multiuser shared-wireless media scenarios in order to develop QoS-constrained higher-layer protocols such as packet scheduling, admission control or adaptive resource-allocation schemes. However, it uses asymptotic approximations that for many cases of practical interest tend to overestimate/understimate the QoS performance metrics of the system while it remains to be seen whether these approximations are useful in order to perform the proposed cross-layer designs. Thus, both the DTMC-based and the effective bandwidth/capacity-based approaches will be analyzed and compared when combined with our proposed PHY layer first-order two-dimensional Markov model for cross-layer design in AMC/ARQ-based wireless networks.

3.2 Physical layer modeling

3.2.1 AMC pool and characterization of the packet error rate

Without loss of generality, convolutionally coded M-ary quadrature amplitude modulation (M-QAM) schemes adopted from the IEEE 802.11a standard (IEEE, 1997) will be used to define the AMC pool. All considered TMs (except the non-transmission mode) are listed in Table 3.1 (Liu et al., 2004).

Let us consider the channel quantity γ_ν denoting the instantaneous received SNR at time instant $t = \nu T_f$, where T_f is the frame period. For the assumed Rayleigh fading channel model, γ_ν is an exponentially distributed random variable with pdf

$$p_{\gamma_\nu}(\gamma) = (1/\overline{\gamma}) \exp(-\gamma/\overline{\gamma}), \quad \gamma \geq 0, \tag{3.1}$$

where $\overline{\gamma} = E\{\gamma_\nu\}$ is the average received SNR.

As it has been previously seen in Chapter 2, when implementing the AMC strategy, the entire SNR range is partitioned into a set of non-overlapping intervals defined by the partition $\mathbf{\Gamma}^m$ given by (2.1). Therefore, TM n will be selected when $\gamma_\nu \in [\gamma_n^m, \gamma_{n+1}^m)$ with probability

$$Pr(n) = \int_{\gamma_n^m}^{\gamma_{n+1}^m} p_{\gamma_\nu}(\gamma) d\gamma = e^{-\gamma_n^m/\overline{\gamma}} - e^{-\gamma_{n+1}^m/\overline{\gamma}}. \tag{3.2}$$

In the presence of additive white Gaussian noise (AWGN), the PER of these TMs can be approximated as

$$\mathrm{PER}_n(\gamma_\nu) \approx \begin{cases} 1 & , 0 \leq \gamma_\nu < \gamma_{p_n} \\ a_n e^{-g_n \gamma_\nu} & , \gamma_\nu \geq \gamma_{p_n} \end{cases} \tag{3.3}$$

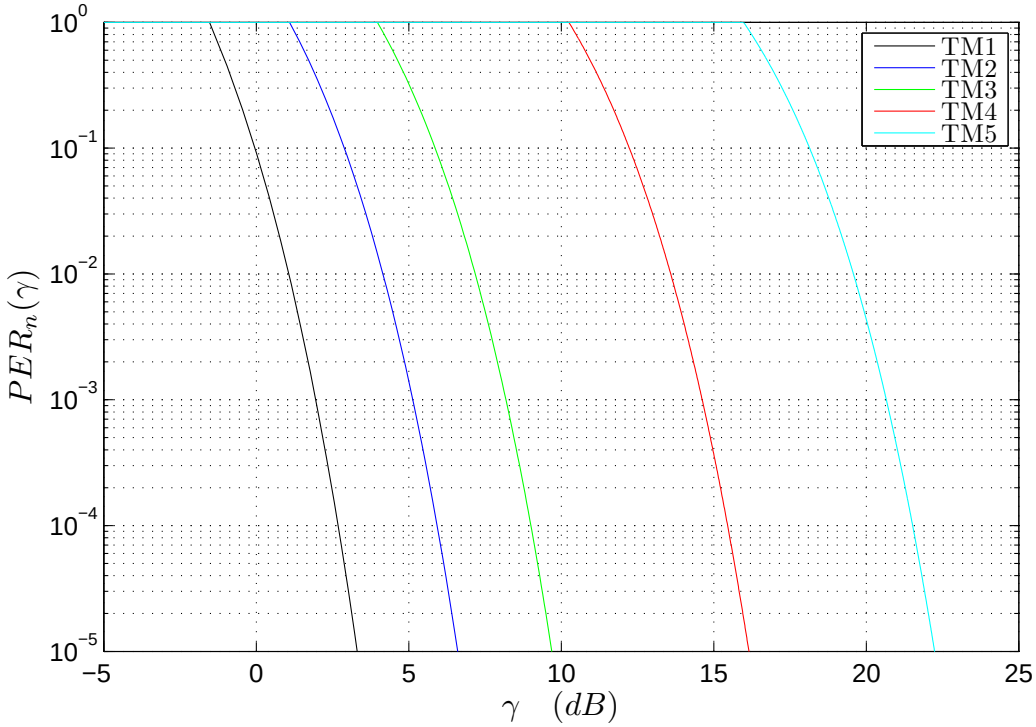


Figure 3.1: Instantaneous PER vs. SNR.

where n is the mode index and a_n, g_n and γ_{p_n}, listed in Table 3.1, are the fitting parameters for TMs with a packet length of 1080 bits (Liu et al., 2004). Figure 3.1 plots the instantaneous PER curves corresponding to these five TMs adopted from the IEEE 802.11a standard.

3.2.2 Physical layer two-dimensional Markov model

Given the AMC scheme, characterized by the partition $\mathbf{\Gamma}^m$ defined in (2.1), and the first-order two-dimensional Markov channel model, characterized by the partitions $\mathbf{\Gamma}^c$ and $\mathbf{\Delta}$ defined in (2.6) and (2.8), respectively, let us now partition the range of γ_ν into the set of N_{PHY} non-overlapping intervals defined by the partition

$$\mathbf{\Gamma}^{m,c} = \left\{ [\gamma_0^{m,c}, \gamma_1^{m,c}), [\gamma_1^{m,c}, \gamma_2^{m,c}), \ldots, \left[\gamma_{N_{\text{PHY}}-1}^{m,c}, \gamma_{N_{\text{PHY}}}^{m,c}\right) \right\} \tag{3.4}$$

with $\gamma_0^{m,c} = 0$, $\gamma_{N_{\text{PHY}}}^{m,c} = \infty$ and

$$\{\gamma_1^{m,c}, \ldots, \gamma_{N_{\text{PHY}}-1}^{m,c}\} = \text{sort}\left(\{\gamma_1^m, \ldots, \gamma_{M-1}^m\} \cup \{\gamma_1^c, \ldots, \gamma_{K-1}^c\}\right), \tag{3.5}$$

where each partition interval $\left[\gamma_k^{m,c}, \gamma_{k+1}^{m,c}\right)$ is characterized by a particular combination of TM and channel state. Figure 3.2 illustrates this process and shows how the partition $\mathbf{\Gamma}^{m,c}$ is obtained. As in Section 2.4, let us also consider the partition of $\delta_\nu = \gamma_{\nu-1} - \gamma_\nu$ into the set of non-overlapping intervals $\mathbf{\Delta} = \{(-\infty, 0), [0, \infty)\}$. Using this two-dimensional partitioning, a first-order two-dimensional Markov model for the PHY layer can be defined where each state corresponds to one of

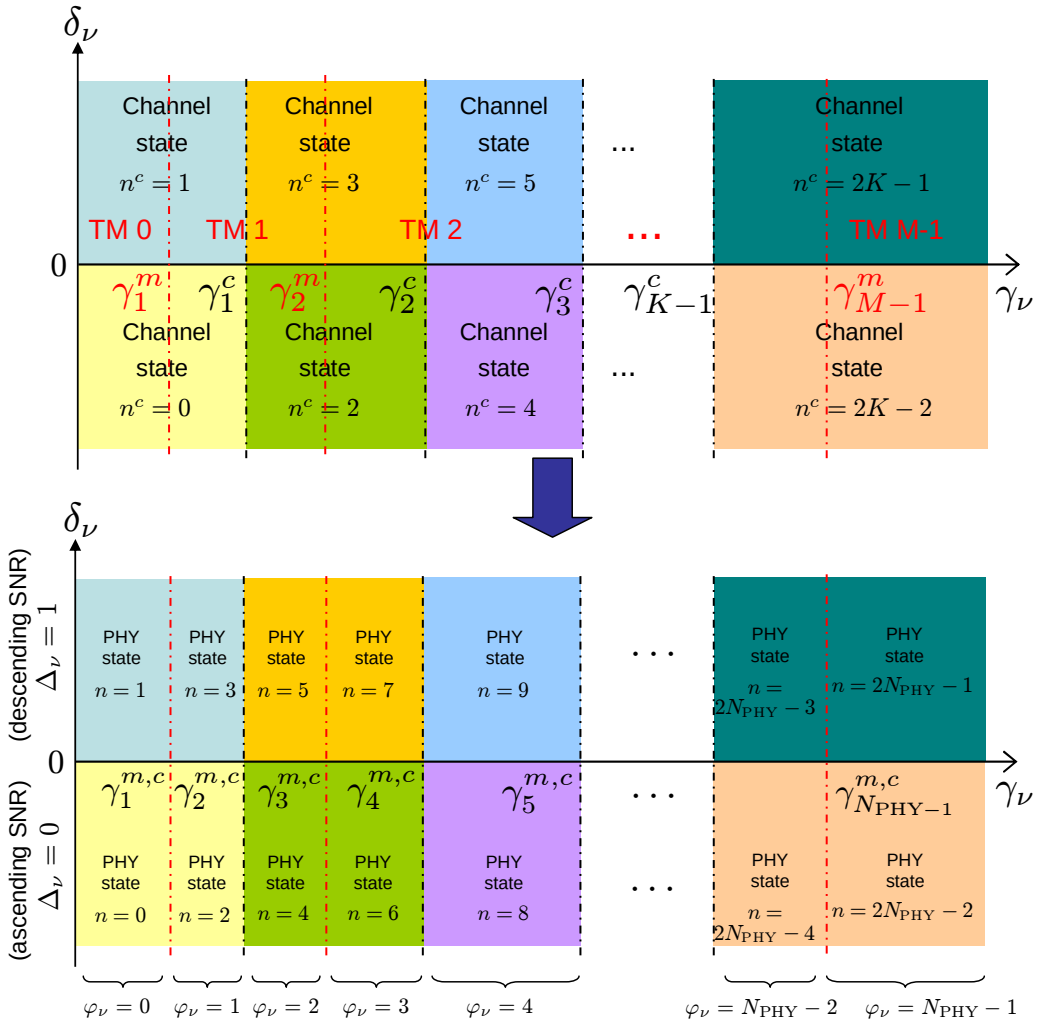


Figure 3.2: PHY layer partitions.

such two-dimensional rectangular-shaped cells, as illustrated in Fig. 3.2. Furthermore, the PHY layer Markov chain state at time instant $t = \nu T_f$ can be denoted as $\varsigma_\nu = (\varphi_\nu, \Delta_\nu)$, $\nu = 0, 1, \ldots, \infty$, where $\varphi_\nu \in \{0, \ldots, N_{\text{PHY}} - 1\}$ denotes the combination of TM and channel state in this frame interval and $\Delta_\nu \in \{0, 1\}$ is used to denote the *up* or *down* characteristic of the instantaneous SNR in time frame interval $t = (\nu - 1)T_f$. Equivalently, the PHY layer state ς_ν can be univocally characterized by an integer number $n_\nu = 2\varphi_\nu + \Delta_\nu$ and obviously, $n_\nu \in \{0, \ldots, 2N_{\text{PHY}} - 1\}$.

The PHY layer will be in a state $n \in \{0, \ldots, 2N_{\text{PHY}} - 1\}$ with probability

$$P^{\text{PHY}}(n) = \begin{cases} \int_{\gamma_{n/2}^{m,c}}^{\gamma_{n/2+1}^{m,c}} \int_0^x p_{\gamma_\nu, \gamma_{\nu-1}}(x, y)\, dydx, & n \text{ even} \\ \int_{\gamma_{(n-1)/2}^{m,c}}^{\gamma_{(n+1)/2}^{m,c}} \int_x^{+\infty} p_{\gamma_\nu, \gamma_{\nu-1}}(x, y)\, dydx, & n \text{ odd} \end{cases} \tag{3.6}$$

where $p_{\gamma_\nu, \gamma_{\nu-1}}$ is the joint pdf of the random variables γ_ν and $\gamma_{\nu-1}$ defined in (2.9), and each of these states will be characterized by an average PER that can

be calculated as

$$\overline{PER}_n^{\mathrm{PHY}} = \begin{cases} \dfrac{\int_{\gamma_{n/2}^{m,c}}^{\gamma_{n/2+1}^{m,c}} \int_0^x PER_{\beta_n}(x) p_{\gamma_\nu,\gamma_{\nu-1}}(x,y)dydx}{P^{\mathrm{PHY}}(n)}, & n \text{ even} \\ \dfrac{\int_{\gamma_{(n-1)/2}^{m,c}}^{\gamma_{(n+1)/2}^{m,c}} \int_x^{+\infty} PER_{\beta_n}(x) p_{\gamma_\nu,\gamma_{\nu-1}}(x,y)dydx}{P^{\mathrm{PHY}}(n)}, & n \text{ odd} \end{cases} \tag{3.7}$$

where β_n denotes the TM corresponding to the nth PHY layer state.

The PHY layer transition probability matrix

$$\boldsymbol{P}_s = \left[P_s(n_\mu, n_{\mu'})\right]_{n_\mu,n_{\mu'}=0}^{2N_{\mathrm{PHY}}-1}, \tag{3.8}$$

has stationary transition probabilities

$$\begin{aligned} P_s(n_\mu, n_{\mu'}) &= Pr\left\{n_\nu = n_{\mu'} | n_{\nu-1} = n_\mu\right\} \\ &= \frac{\int_{x_I}^{x_S}\int_{y_I}^{y_S}\int_{z_I}^{z_S} p_{\gamma_\nu,\gamma_{\nu-1},\gamma_{\nu-2}}(x,y,z)\,dzdydx}{P^{\mathrm{PHY}}(n_\mu)} \end{aligned} \tag{3.9}$$

where the integration intervals are

$$[x_I, x_S] = \begin{cases} \left[\gamma_{\frac{n_{\mu'}}{2}}^{m,c}, \gamma_{\frac{n_{\mu'}}{2}+1}^{m,c}\right], & n_{\mu'} \text{ even} \\ \left[\gamma_{\frac{n_{\mu'}-1}{2}}^{m,c}, \gamma_{\frac{n_{\mu'}+1}{2}}^{m,c}\right], & n_{\mu'} \text{ odd} \end{cases} \tag{3.10}$$

$$[y_I, y_S] = \begin{cases} \left[\gamma_{\frac{n_\mu}{2}}^{m,c}, \max\left\{\gamma_{\frac{n_\mu}{2}}^{m,c}, \min\left\{x, \gamma_{\frac{n_\mu}{2}+1}^{m,c}\right\}\right\}\right], & n_\mu \text{ even } n_{\mu'} \text{ even} \\ \left[\min\left\{\gamma_{\frac{n_\mu}{2}+1}^{m,c}, \max\left\{x, \gamma_{\frac{n_\mu}{2}}^{m,c}\right\}\right\}, \gamma_{\frac{n_\mu}{2}+1}^{m,c}\right], & n_\mu \text{ even } n_{\mu'} \text{ odd} \\ \left[\gamma_{\frac{n_\mu-1}{2}}^{m,c}, \max\left\{\gamma_{\frac{n_\mu-1}{2}}^{m,c}, \min\left\{x, \gamma_{\frac{n_\mu+1}{2}}^{m,c}\right\}\right\}\right], & n_\mu \text{ odd } n_{\mu'} \text{ even} \\ \left[\min\left\{\gamma_{\frac{n_\mu+1}{2}}^{m,c}, \max\left\{x, \gamma_{\frac{n_\mu-1}{2}}^{m,c}\right\}\right\}, \gamma_{\frac{n_\mu+1}{2}}^{m,c}\right], & n_\mu \text{ odd } n_{\mu'} \text{ odd} \end{cases} \tag{3.11}$$

$$[z_I, z_S] = \begin{cases} [0, y], & n_\mu \text{ even} \\ [y, +\infty], & n_\mu \text{ odd} \end{cases} \tag{3.12}$$

and $p_{\gamma_\nu,\gamma_{\nu-1},\gamma_{\nu-2}}(x,y,z)$ denotes the joint pdf of the random variables γ_ν, $\gamma_{\nu-1}$ and $\gamma_{\nu-2}$, which can be expressed as in (2.16).

The integrals involved in the calculation of $P^{\mathrm{PHY}}(n)$, $\overline{PER}_n^{\mathrm{PHY}}$ and $P_s(n_\mu, n_{\mu'})$ can be computed either by Monte-Carlo numerical simulations or by using the series representation of the modified order-k Bessel function of the first kind (Abramowitz and Stegun, 1972, eq. (9.6.10)) (see also (Chen and Tellambura, 2005; Tan and Beaulieu, 1997)).

3.3 Link-level queueing model and analysis

Once the PHY layer has been modeled, the DLC layer must also be introduced in the analytical model. According to Section 2.2, this research work pursues the development of an analytical link-level queueing model for QoS-constrained AMC/(H)ARQ-based wireless systems, which will be obtained based on the use of DTMCs. As it will be shown in Section 3.5, the real system behaviour will be faithfully reproduced by this Markov-based model but, unfortunately, its complexity increases with the number of states of the arrival process, the number of states of the PHY layer Markov model or the maximum queue length of the system. In order to attain a less complex analytical model, a second approach based on the effective bandwidth theory (see, e.g., (Chang, 2000; Chang and Thomas, 1995)) and its closely related effective capacity theory (Wu and Negi, 2003) will be also presented. Both approaches will be evaluated and compared when combined with our proposed PHY layer first-order two-dimensional Markov model for cross-layer design in QoS-constrained AMC/ARQ-based wireless systems.

3.3.1 Discrete time Markov chain-based model and analysis

Embedded Markov chain

Following the approach described in (Le et al., 2006b), the queueing process induced by both the infinitely persistent Type-I hybrid FEC/ARQ protocol and the AMC scheme can be formulated in discrete time with one time unit equal to one frame interval. The system states are observed at the beginning of each time unit. Let $\sigma_\nu = (q_\nu, a_\nu, n_\nu)$ denote the system state at time instant $t = \nu T_f$, where $q_\nu \in \{0, \ldots, \overline{Q}\}$ is the queue length at this time instant, $a_\nu \in \{0, \ldots, \mathcal{A}-1\}$ represents the phase of the D-BMAP and $n_\nu \in \{0, \ldots, 2N_{\text{PHY}} - 1\}$ corresponds to the PHY layer state. If we just look at the set of time instants $t = \nu T_f$, $\nu = 0, 1, \ldots, \infty$, the transitions between states σ_ν are Markovian. Therefore, an embedded Markov chain can be used to describe the underlying queueing process. The state space of this embedded finite state Markov chain is $\boldsymbol{\mathcal{S}} = \{\mathcal{S}_\mu\}_{\mu=1}^{N_s}$, where $N_s = 2(\overline{Q}+1)\mathcal{A}N_{\text{PHY}}$ and $\mathcal{S}_\mu = (q_\mu, a_\mu, n_\mu)$.

Figure 3.3 illustrates the queue transition process between states (q_μ, a_μ, n_μ) and $\left(q'_\mu, a'_\mu, n'_\mu\right)$, which correspond to two consecutive frame intervals. At the beginning of the first frame interval there are q_μ packets in the queue and, given that the number of packets that can be transmitted when the PHY layer is in state n_μ is $\mathcal{C}_{n_\mu}$, the number of packets that will be eventually transmitted is $\min\{\mathcal{C}_{n_\mu}, q_\mu\}$ (blue arrow). Among them, let us assume that h packets are successfully transmitted and leave the queue (green arrow), while $\min\{\mathcal{C}_{n_\mu}, q_\mu\} - h$ packets are received in error and, consequently, need to be kept in the queue (camel arrow). Obviously, the $q_\mu - \min\{\mathcal{C}_{n_\mu}, q_\mu\}$ packets that are not transmitted also remain in the queue (yellow arrow). Thus, the number of packets in the queue at the end of this frame interval will be equal to $q_\mu - h$ packets plus the arriving packets (purple arrow) that can be lodged in the queue. If there is no room for all the a_μ incoming packets, a buffer overflow will occur (orange arrow). Consequently, the number of packets in the buffer at the end of this frame interval will be equal

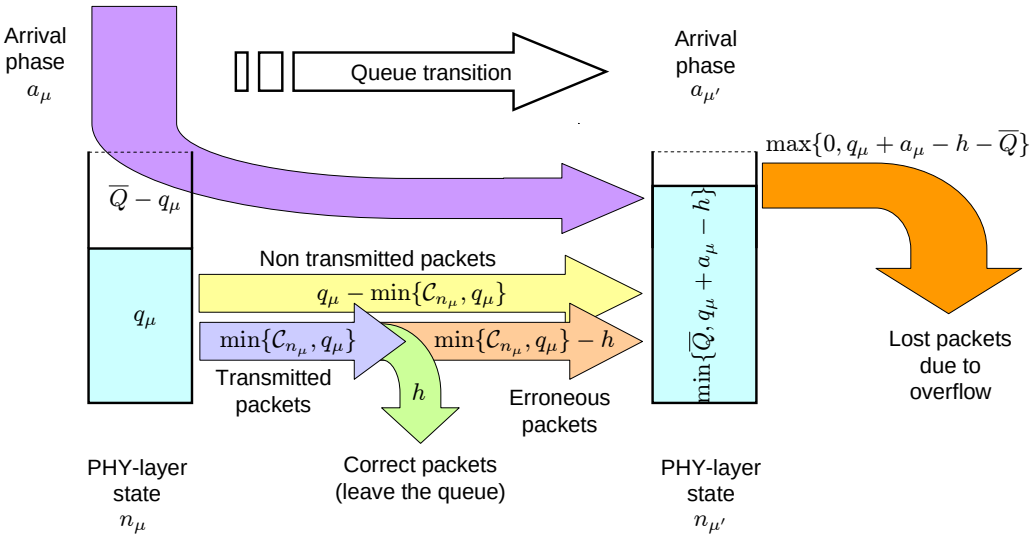


Figure 3.3: Queue transition for an infinitely persistent ARQ scheme.

to $\min\left\{\overline{Q}, q_\mu + a_\mu - h\right\}$, and the number of overflowed packets will be given by $\max\left\{0, q_\mu + a_\mu - h - \overline{Q}\right\}$.

Taking into account that an infinitely persistent Type-I hybrid FEC/ARQ protocol is used at the DLC layer, and assuming that the transmission outcomes (success or failure) of consecutive packets in a frame interval are independent, the probability that k packets are successfully transmitted (leave the queue) given $\mathcal{C}_{n_\mu}$ packets are transmitted when the PHY layer is in state $\mathcal{S}_\mu$ can be written as

$$p_{k,\mathcal{C}_{n_\mu}}^{(n_\mu)} = \binom{\mathcal{C}_{n_\mu}}{k} \left(\overline{PER}_{n_\mu}^{\text{PHY}}\right)^{\mathcal{C}_{n_\mu}-k} \left(1 - \overline{PER}_{n_\mu}^{\text{PHY}}\right)^{k}. \tag{3.13}$$

Considering that there are $q_\mu = q$ packets in the queue before transmission and that the PHY layer is in state n_μ, then $\min\left\{\mathcal{C}_{n_\mu}, q\right\}$ packets will be transmitted (blue arrow). Thus, in this situation, the probability that h packets are successfully transmitted (green arrow) and the PHY layer changes from state n_μ to state $n_{\mu'}$, can be expressed as

$$t_h^{(q)}\left(n_\mu, n_{\mu'}\right) = p_{h,\min\left\{\mathcal{C}_{n_\mu}, q\right\}}^{(n_\mu)} P_{n_\mu, n_{\mu'}}, \tag{3.14}$$

for $q \in \{0, \ldots, \overline{Q}\}$, $h \in \{0, \ldots, \min\{q, \mathcal{C}\}\}$ and $n_\mu, n_{\mu'} \in \{0, \ldots, 2N_{\text{PHY}} - 1\}$. The $2N_{\text{PHY}} \times 2N_{\text{PHY}}$ terms in (3.14) capturing all the cases where h packets are successfully transmitted given that there are q packets in the queue before transmission can be expressed in matrix form as

$$\boldsymbol{T}_h^{(q)} = \mathcal{D}\left(\boldsymbol{p}_h^q\right) \boldsymbol{P}_s, \tag{3.15}$$

for $q \in \{0, \ldots, \overline{Q}\}$ and $h \in \{0, \ldots, \min\{q, \mathcal{C}\}\}$, where

$$\boldsymbol{p}_h^q = \left(p_{h,\min\{\mathcal{C}_0, q\}}^{(0)} \quad \cdots \quad p_{h,\min\left\{\mathcal{C}_{2N_{\text{PHY}}-1}, q\right\}}^{(2N_{\text{PHY}}-1)}\right) \tag{3.16}$$

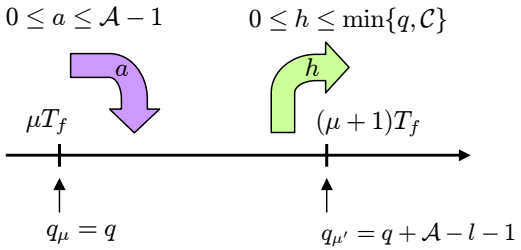


Figure 3.4: Arrival and departure of packets.

and $\mathcal{D}(\boldsymbol{x})$ is used to denote a diagonal matrix with the elements of the vector $\boldsymbol{x}$ in its main diagonal. Notice that for $q \geq \mathcal{C}_{2N_{\text{PHY}}-1} = \mathcal{C}$ the probabilities in these matrices do not depend on q and $\boldsymbol{T}_h^{(q)} = \boldsymbol{T}_h^{(\mathcal{C})}$.

As shown in Fig. 3.4, let $q_\mu = q$ and $q_{\mu'} = q + \mathcal{A} - l - 1$ be the number of packets in the queue in two consecutive frames. The maximum number of packet arrivals in one frame interval is equal to $\mathcal{A}-1$ and the maximum number of successfully transmitted packets in one frame interval is equal to $\min\{q,\mathcal{C}\}$; thus, it is quite obvious that $q_\mu - \min\{q,\mathcal{C}\} \leq q_{\mu'} \leq q_\mu + \mathcal{A} - 1$ or, equivalently, $0 \leq l \leq \mathcal{A} + \min\{q,\mathcal{C}\} - 1$. Furthermore, let $a_\mu = a$ and h be the number of arriving packets and the number of packets successfully transmitted in the first one of these frame intervals, respectively (purple and green arrows). In this case, $q_{\mu'} - q_\mu = \mathcal{A} - l - 1 = a - h$ or $h = l - \mathcal{A} + a + 1$. Thus, given that $0 \leq l - \mathcal{A} + a + 1 \leq \min\{q,\mathcal{C}\}$ or $\mathcal{A} - l - 1 \leq a \leq \mathcal{A} - l + \min\{q,\mathcal{C}\} - 1$, and $0 \leq a \leq \mathcal{A}-1$, the probability that the queueing system changes from a generic state $\mathcal{S}_\mu = (q, a_\mu, n_\mu) \in \boldsymbol{\mathcal{S}}$ to another generic state $\mathcal{S}_{\mu'} = (q+\mathcal{A}-l-1, a_{\mu'}, n_{\mu'}) \in \boldsymbol{\mathcal{S}}$ can be expressed as

$$A_{q,l}\left(\mathcal{S}_\mu, \mathcal{S}_{\mu'}\right) = \sum_{a=a_{\min}}^{a_{\max}} u_a(a_\mu, a_{\mu'})\; t_{l-\mathcal{A}+a+1}^{(q)}\left(n_\mu, n_{\mu'}\right) \tag{3.17}$$

where $a_{\min} = \max\{0, \mathcal{A}-l-1\}$ and $a_{\max} = \min\{\mathcal{A}-1, \mathcal{A}+\min\{q,\mathcal{C}\}-1-l\}$, for $q \in \{0,\ldots,\overline{Q}\}$, $l \in \{0,\ldots,\mathcal{A}+\min\{q,\mathcal{C}\}-1\}$, $a_\mu, a_{\mu'} \in \{0,\ldots,\mathcal{A}-1\}$ and $n_\mu, n_{\mu'} \in \{0,\ldots,2N_{\text{PHY}}-1\}$. The $2\mathcal{A}N_{\text{PHY}} \times 2\mathcal{A}N_{\text{PHY}}$ terms capturing all the cases where the queue length changes from q packets in one frame interval to $q+\mathcal{A}-l-1$ packets in the next frame interval can be expressed in matrix form as

$$\boldsymbol{A}_{q,l} = \begin{cases} \displaystyle\sum_{a=a_{\min}}^{a_{\max}} \boldsymbol{U}_a \otimes \boldsymbol{T}_{l-\mathcal{A}+a+1}^{(q)} & , l \in \{0,\ldots,l_{\max}(q)\} \\ \boldsymbol{0} & , \text{ otherwise} \end{cases} \tag{3.18}$$

for $q \in \{0,\ldots,\overline{Q}\}$, where $\otimes$ denotes the Kronecker product and $l_{\max}(q) = \mathcal{A} + \min\{q,\mathcal{C}\} - 1$. The resulting transition matrix of the Markov chain can then be written as

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{A}_{0,\mathcal{A}-1} & \cdots & \boldsymbol{A}_{0,\mathcal{A}-\overline{Q}} & \overline{\boldsymbol{A}}_{0,\mathcal{A}-\overline{Q}-1} \\ \vdots & & \vdots & \vdots \\ \boldsymbol{A}_{\overline{Q},\mathcal{A}+\overline{Q}-1} & \cdots & \boldsymbol{A}_{\overline{Q},\mathcal{A}} & \overline{\boldsymbol{A}}_{\overline{Q},\mathcal{A}-1} \end{bmatrix}, \tag{3.19}$$

where

$$\overline{\boldsymbol{A}}_{q,i} = \sum_{a=0}^{i} \boldsymbol{A}_{q,a}. \tag{3.20}$$

Notice that for $q \geq \mathcal{C}$ the transition probabilities in these matrix blocks do not depend on q and, therefore, for simplicity this index can be omitted, that is, $\boldsymbol{A}_{q,l} = \boldsymbol{A}_l$ and $\overline{\boldsymbol{A}}_{q,l} = \overline{\boldsymbol{A}}_l$ for all $q \geq \mathcal{C}$. As an example, assuming that $\overline{Q} \geq \mathcal{A}+\mathcal{C}$, the resulting transition matrix of the Markov chain can be written as

To derive the system performance measures, we need to obtain the steady-state probability vectors corresponding to each level of the transition matrix, which can be calculated using the fact that the transition probability matrix $\boldsymbol{P}$ and steady-state probability vector $\boldsymbol{\pi} = \begin{bmatrix}\boldsymbol{\pi}_0 & \boldsymbol{\pi}_1 & \cdots & \boldsymbol{\pi}_{\overline{Q}}\end{bmatrix}$ satisfy $\boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{\pi}$ along with the normalization condition $\sum_{i=0}^{\overline{Q}} \boldsymbol{\pi}_i \mathbf{1} = 1$, where $\mathbf{1}$ is a column vector of all ones with the appropriate length.

Packet loss rate and throughput

In this finite buffering ARQ-based error control system with infinite persistence, the packet loss rate P_l (measured in packets per second) is simply equal to the buffer overflow probability. As in (Le et al., 2006b), $\boldsymbol{V}_k$ denotes the stationary vector describing the probabilities that k packets are lost due to buffer overflow upon arrival of a burst of data packets. Assuming that a batch of a packets arrive at the link layer buffer (purple arrow in Fig. 3.3), if there are $q > \overline{Q} - a$ packets in the queue at the end of the previous frame interval and h packets are successfully transmitted (green arrow), then the number of packets that will be lost due to buffer overflow is $k = a - h - \overline{Q} + q$ (orange arrow). Therefore, $\boldsymbol{V}_k$ can be written as

$$\boldsymbol{V}_k = \sum_{a=1}^{\mathcal{A}-1} \sum_{q=\max\{0,\overline{Q}-a+1\}}^{\overline{Q}} \boldsymbol{\pi}_q \sum_{\substack{h=0 \\ a-h=\overline{Q}-q+k}}^{\mathcal{C}} \boldsymbol{U}_a \otimes \boldsymbol{T}_h^q. \tag{3.21}$$

Thus, the average number of lost packets due to buffer overflow can be obtained as

$$N_l = \sum_{k=1}^{\mathcal{A}-1} k\boldsymbol{V}_k \mathbf{1}. \tag{3.22}$$

Given the average number of arriving packets λ in one frame interval, the packet loss rate can then be calculated as

$$P_l = N_l/\lambda = (1/\lambda) \sum_{k=1}^{\mathcal{A}-1} k\boldsymbol{V}_k \mathbf{1}, \tag{3.23}$$

and the average throughput (measured in packets per frame) is given by

$$\eta = \lambda(1 - P_l). \tag{3.24}$$

Average queue length and average packet delay

Due to the assumption of infinite persistence in the ARQ-based error control system, the packet blocking probability is equal to the packet loss rate. Thus, using the well-known Little's formula (Kleinrock, 1975), the average delay for the embedded Markov chain can be calculated as

$$D_l = L_q/\lambda(1 - P_l) = L_q/\eta, \tag{3.25}$$

where L_q denotes the average number of packets in the queue, which can be obtained as

$$L_q = \sum_{q=1}^{\overline{Q}} q \ \boldsymbol{\pi}_q \mathbf{1}. \tag{3.26}$$

3.3.2 Effective bandwidth-based model and analysis

Although the proposed discrete time Markov chain-based approach provides an accurate estimation of the performance metrics at the DLC layer, valuable radio link-level performance measures are provided at the cost of an increase of complexity with the number of arrival process states, the number of PHY layer Markov model states or the maximum queue length of the system. For that reason, a second approach will be proposed to characterize the QoS provisioning performance at the DLC layer, which follows a completely different methodology that relies on the effective capacity concept (Wu and Negi, 2003). It turns out to be the dual problem of the effective bandwidth, which was extensively studied in the early 90's. The effective capacity and effective bandwidth allow the analysis of the so-called packet loss rate bound probability. Compared with the DTMC-based framework, this approach is not only less complex but in addition it could be readily extended to multiuser shared-wireless media scenarios.

Let us consider an arrival process $A(t)$, $t \geq 0$, representing the amount of data (measured in packets) generated over the time interval $[0, t)$. Let us also assume that the asymptotic log-moment generating function of $A(t)$, defined as

$$\Lambda(u) = \lim_{t\to\infty} (1/t) \log E[e^{uA(t)}], \tag{3.27}$$

exists for all $u \geq 0$. Then, the effective bandwidth function of $A(t)$ is defined by

$$E_B(u) = \frac{\Lambda(u)}{u}, \quad \forall \ u \geq 0. \tag{3.28}$$

Now, assuming a queue of infinite buffer size served by a channel of a constant service rate r such as an AWGN channel and using the theory of large deviations, it can be shown (Chang, 2000) that the probability that the queue size $Q(t)$ exceeds a certain threshold $\overline{Q}$ decays exponentially fast as the threshold increases i.e.,

$$\sup_t Pr\{Q(t) \geq \overline{Q}\} \approx \kappa(r) e^{-\psi(r)\overline{Q}}, \tag{3.29}$$

where $\kappa(r)$ denotes the probability that the buffer is not empty, which can be approximated by the ratio of the average arrival rate to the average service rate

(Chang, 2000) and ψ is a certain positive constant called the QoS exponent and can be obtained as the solution of $E_B(\psi) = r$. If the main QoS metric of interest is the delay $D_l(t)$ experienced by a source packet arriving at time t, then the probability of $D_l(t)$ exceeding a delay bound $\mathcal{D}_{l\max}$ satisfies (Wu and Negi, 2003)

$$\sup_t Pr\{D_l(t) \geq \mathcal{D}_{l\max}\} \approx \kappa(r) e^{-\psi_D(r)\mathcal{D}_{l\max}}, \tag{3.30}$$

where $\psi_D(r) = r\psi(r)$.

In this work, the arrival process adheres to a D-BMAP characterized by a transition matrix $\boldsymbol{U}$, then it can be shown that (Chang, 2000),

$$E_B(\psi) = \psi^{-1} \log(\delta_U(\psi)), \tag{3.31}$$

where δ_U is the Perron-Frobenius eigenvalue of the matrix $\boldsymbol{U}_{\varpi} = \mathcal{D}(\boldsymbol{\varpi})\boldsymbol{U}$, with $\boldsymbol{\varpi} \triangleq \left(e^{\lambda_0\psi}, \ldots, e^{\lambda_{\mathcal{A}-1}\psi}\right)$, where λ_μ denotes the number of packets per frame generated when the source is in state μ.

As proposed by Wu and Negi (2003), the concept of effective capacity, which is a dual problem of the original effective bandwidth, can be used to model the PHY layer. Let $r(t)$ denote the instantaneous capacity at time t. Then

$$S(t) = \int_0^t r(\tau) d\tau \tag{3.32}$$

is the service provided by the channel. If

$$\Lambda^{(C)}(-u) = \lim_{t\to\infty} (1/t) \log E[e^{-uS(t)}] \tag{3.33}$$

exists for all $u \geq 0$, then the effective bandwidth function of $r(t)$ can be defined as

$$E_C(u) = \frac{-\Lambda^{(C)}(-u)}{u}, \quad \forall\ u \geq 0. \tag{3.34}$$

Now, considering a queue of infinite buffer size supplied by a data source of constant data rate ϖ, then the effective bandwidth theory can be used to show that

$$\sup_t Pr\{Q(t) \geq \overline{Q}\} \approx \kappa(\varpi) e^{-\psi(\varpi)\overline{Q}}. \tag{3.35}$$

and, similar to (3.30)

$$\sup_t Pr\{D_l(t) \geq \mathcal{D}_{l\max}\} \approx \kappa(\varpi) e^{-\psi_D(\varpi)\mathcal{D}_{l\max}}. \tag{3.36}$$

Since the service process is a Markov modulated process characterized by a transition probability matrix $\boldsymbol{P}_s$, its effective capacity function can be obtained as

$$E_C(\psi) = -\psi^{-1} \log(\delta_P(-\psi)), \tag{3.37}$$

where δ_P is the Perron-Frobenius eigenvalue of the matrix $\boldsymbol{P}_\upsilon = \mathcal{D}(\boldsymbol{\upsilon})\boldsymbol{P}_s$, with $\boldsymbol{\upsilon} \triangleq \left(e^{-\hat{c}_0\psi}, \ldots, e^{-\hat{c}_{(2N_{PHY}-1)}\psi}\right)$, where $\hat{c}_n$ denotes the number of packets per frame leaving the queue when the PHY layer is in state n_μ, which using (3.13) can be calculated as

$$\hat{c}_{n_\mu} = \sum_{k=0}^{\mathcal{C}_{n_\mu}} k\ p_{k,\mathcal{C}_{n_\mu}}^{(n_\mu)}. \tag{3.38}$$

Packet loss rate and throughput

The effective bandwidth-based approach only provides statistical QoS guarantees for the average packet loss rate P_l in a system that incorporates an infinitely persistent Type-I hybrid FEC/ARQ protocol. That is, the probability loss constraint is guaranteed with a small violation probability ϵ. Then, the system throughput η can be calculated as

$$\eta = \lambda(1 - P_l) \tag{3.39}$$

and the packet loss probability P_l can be approximated by the tail probability

$$\sup_t Pr\{Q(t) \geq \overline{Q}\} \tag{3.40}$$

of the queue length at the DLC layer, which can be calculated by using (3.29) and (3.35).

3.4 Cross-layer optimization

3.4.1 Multidimensional approach

As shown in previous sections, given a maximum afforded queue length $\overline{Q}$, an average SNR $\bar{\gamma}$ and a normalized maximum Doppler frequency $f_d T_f$, performance measures of the system like, for instance, throughput, average packet delay or packet loss rate, are a function of the AMC transmission mode switching levels $\mathbf{\Gamma}^m \in \mathbb{R}_+^{M+1}$, where $\mathbb{R}_+$ denotes the set of non-negative real numbers, and the measured or estimated arrival packet rate $\lambda \in \mathbf{\Theta}$, where $\mathbf{\Theta}$ is the range of feasible arrival rate values. Moreover, if the system is to support QoS-guaranteed traffic, depending on the traffic type, this QoS will be characterized by different requirements. For instance, for data traffic a maximum packet loss rate $P_{l\max}$ should be guaranteed, whereas for delay-sensitive traffic in addition to a maximum packet loss rate $P_{l\max}$ a maximum average packet delay $D_{l\max}$ should also be satisfied. Therefore, if the objective of the link adaptation scheme is to maximize the average throughput when supporting QoS-guaranteed traffic characterized by a maximum packet loss rate $P_{l\max}$ and/or a maximum average packet delay $D_{l\max}$, the system needs to jointly optimize the selected protocol parameters at the PHY and DLC layers by solving the cross-layer optimization problem

$$\left(\mathbf{\Gamma}_{\text{opt}}^m, \lambda_{\text{opt}}\right) = \arg \max_{\mathbf{\Gamma}^m \in \mathbb{R}_+^{M+1}, \lambda \in \mathbf{\Theta}} \eta\left(\mathbf{\Gamma}^m, \lambda\right) \tag{3.41}$$

subject to the constraints

$$P_l\left(\mathbf{\Gamma}_{\text{opt}}^m, \lambda_{\text{opt}}\right) \leq P_{l\max} \quad \text{and/or} \quad D_l\left(\mathbf{\Gamma}_{\text{opt}}^m, \lambda_{\text{opt}}\right) \leq D_{l\max}. \tag{3.42}$$

Note that other cross-layer designs could also be considered, depending on which parameters are fixed and which are adjustable. For instance, the cross-layer strategy could involve the packet length or the buffer size in case these parameters were adaptive.

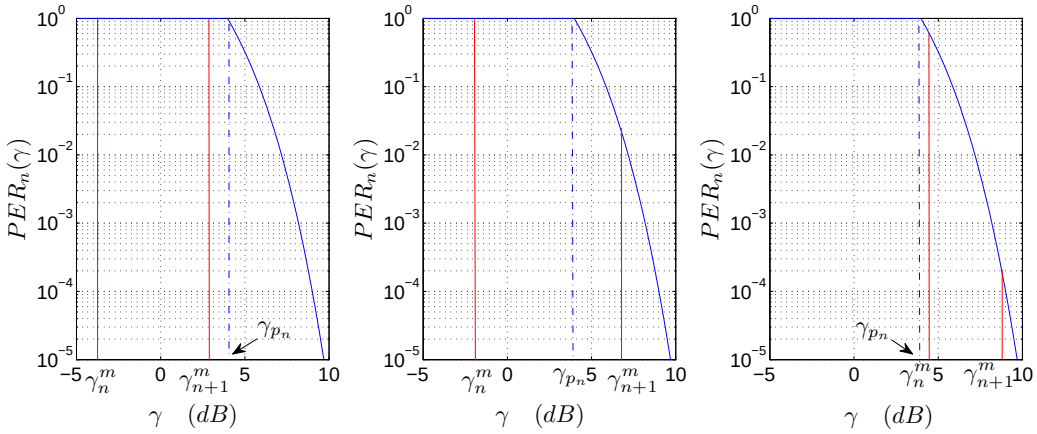


Figure 3.5: Average PER calculation.

The analytical expressions for η, P_l and D_l do not leave much room for developing efficient algorithms to solve our constrained optimization problem. However, considering that $\mathbf{\Gamma}^m$ and λ lie in a bounded space $\mathbb{R}_+^{M+1} \times \mathbf{\Theta}$, we could resort to a multidimensional exhaustive search to numerically solve the proposed cross-layer optimization problem.

3.4.2 Bidimensional simplification approach

In order to simplify this multidimensional optimization approach, let us define the AMC switching thresholds as follows: given γ_ν, the objective of AMC is to select the TM that maximizes the data rate while maintaining a PER less or equal than a prescribed value P_0. To this end, two possibilities are proposed, depending on whether the average PER or its instantaneous value is considered. The first one will be referred to as *ergodic* option, and the second one as *instantaneous* option.

Ergodic option for the AMC scheme

In this option the AMC scheme will be designed with the objective of maximizing the data rate while maintaining the average PER below a prescribed value P_0.

In (Liu et al., 2005a) it is taken for granted that, in practice, the partition boundary γ_n^m is always greater than the parameter γ_{p_n}. However, for high values of the average target PER P_0 this assumption is not always correct and can lead to misleading values of the partition $\mathbf{\Gamma}^m$. In order to avoid this problem, and contrarily to what was done by Liu et al. (2005a), three possibilities have to be considered regarding the relative value of γ_{p_n} with respect to γ_n^m and γ_{n+1}^m, as illustrated in Fig. 3.5. Therefore, the average PER of mode n must be calculated

as

$$\overline{\mathrm{PER}}_n = \frac{1}{Pr(n)} \int_{\gamma_n^m}^{\gamma_{n+1}^m} \mathrm{PER}_n(\gamma) p_{\gamma_\nu}(\gamma) d\gamma$$
$$\approx \begin{cases} 1 & , \gamma_{n+1}^m < \gamma_{p_n} \\ \frac{e^{-\gamma_n^m/\overline{\gamma}} - e^{-\gamma_{p_n}/\overline{\gamma}}}{Pr(n)} + \frac{a_n\left(e^{-b_n\gamma_{p_n}} - e^{-b_n\gamma_{n+1}^m}\right)}{b_n\overline{\gamma}Pr(n)} & , \gamma_n^m < \gamma_{p_n} \leq \gamma_{n+1}^m \\ \frac{a_n\left(e^{-b_n\gamma_n^m} - e^{-b_n\gamma_{n+1}^m}\right)}{b_n\overline{\gamma}Pr(n)} & , \gamma_{p_n} \leq \gamma_n^m \end{cases} \tag{3.43}$$

where $b_n = g_n + 1/\overline{\gamma}$. The average PER of the AMC scheme can then be calculated as

$$\overline{\mathrm{PER}} = \frac{\sum_{n=1}^{M-1} R_n Pr(n) \overline{\mathrm{PER}}_n}{\sum_{n=1}^{M-1} R_n Pr(n)} \tag{3.44}$$

and the partition boundaries $\mathbf{\Gamma}^m$ have to be determined in order to ensure that $\overline{\mathrm{PER}} \leq P_0$. With the aim of accomplishing this objective the following threshold searching algorithm is proposed:

1. Set $M = 0$, $\boldsymbol{\mathcal{M}} = \emptyset$, $n = M_p$, $\gamma_0^m = 0$, $\gamma_n^m = \infty$.

2. $n \leftarrow n - 1$
   If $n = 0$ go to step 3, otherwise search the unique $\gamma_n^m \in [0, \gamma_{n+1}^m)$ that satisfies $\overline{PER}_n = P_0$.

   - In case it exists, update $M \leftarrow M + 1$ and $\boldsymbol{\mathcal{M}} \leftarrow \{n\} \cup \{\boldsymbol{\mathcal{M}}\}$ and go to step 2.
   - In case it does not exist due to $\overline{PER}_n > P_0, \forall \gamma_n^m \in [0, \gamma_{n+1}^m)$, then: declare TM n as *useless*, update threshold subindices as $\gamma_{i-1}^m \leftarrow \gamma_i^m$, $i = n+1, \ldots, n+1+M$ and go to step 2.
   - In case it does not exist due to $\overline{PER}_n < P_0, \forall \gamma_n^m \in [0, \gamma_{n+1}^m)$, TM n is the last *useful* mode and, thus, update $M \leftarrow M+1$, $\boldsymbol{\mathcal{M}} \leftarrow \{n\} \cup \{\boldsymbol{\mathcal{M}}\}$ and threshold subindices $\gamma_{i-n}^m \leftarrow \gamma_i^m$, $i = n+1, \ldots, n+M$ and stop the searching algorithm.

3. TM 0 is the lowest used TM and, thus, update $M \leftarrow M + 1$ and $\boldsymbol{\mathcal{M}} \leftarrow \{0\} \cup \{\boldsymbol{\mathcal{M}}\}$ and stop the searching algorithm.

This threshold searching algorithm constitutes an important modification to that proposed by Liu et al. (2005a) and Wang et al. (2007), in which it was assumed that all *possible* TMs could always be considered as *useful* TMs. On the contrary, in the proposed threshold searching algorithm, depending on the channel conditions and the QoS requirements, some of the *possible* TMs may be declared *useless* and thus, only a limited set of *useful* TMs will be available to the AMC scheme.

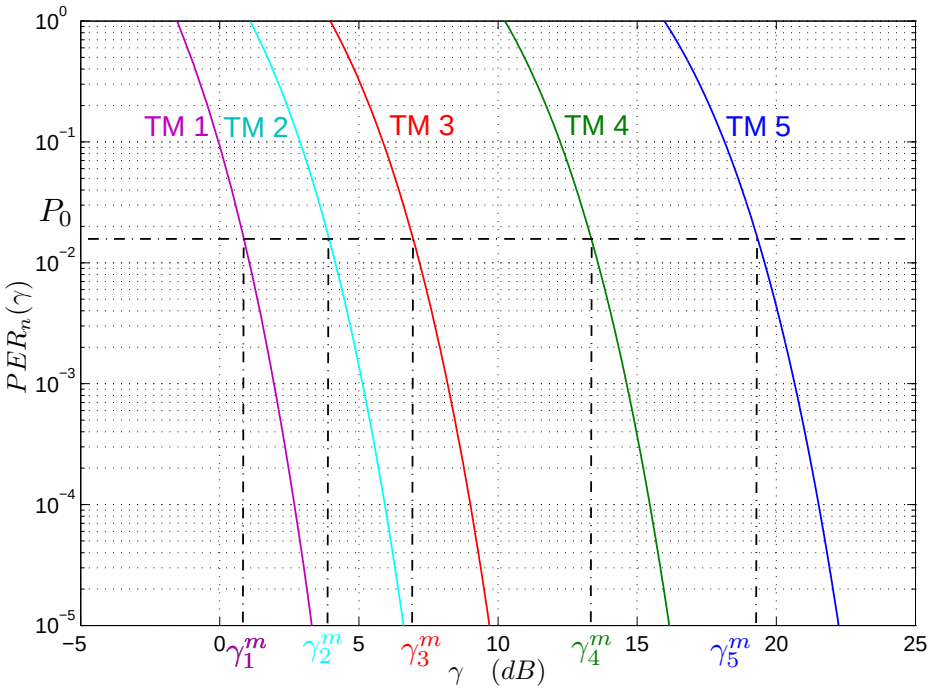


Figure 3.6: AMC switching thresholds calculation in the *Instantaneous* option.

Instantaneous option for the AMC scheme

In the *instantaneous* option the AMC scheme will be designed with the objective of maximizing the data rate while maintaining the instantaneous PER below a prescribed value P_0. Hence, the *instantaneous* option is more stringent than the *ergodic* one. As a result, more conservative AMC partitions will be obtained with this option, leading to more reliable transmissions.

According to (Liu et al., 2004), it is required that the instantaneous packet error rate $\mathrm{PER}_n(\gamma)$ falls below a prescribed value P_0. Thus, the partition boundaries $\mathbf{\Gamma}^m$ will be obtained as the minimum SNR required to achieve P_0, as illustrated in Fig. 3.6. Inverting the $\mathrm{PER}_n(\gamma)$ expression in (3.3) it can be shown that

$$\begin{cases} \gamma_0^m = 0 \\ \gamma_n^m = \frac{1}{g_n} ln\left(\frac{a_n}{P_0}\right) & , n = 1, \ldots, M-1 \\ \gamma_M^m = \infty \end{cases} \tag{3.45}$$

Simplified optimization problem

In both the *ergodic* and the *instantaneous* options the simplified constrained optimization problem can be formulated as

$$(P_{0\mathrm{opt}}, \lambda_{\mathrm{opt}}) = \arg \max_{P_0 \in \mathbb{R}_+, \lambda \in \mathbf{\Theta}} \eta\left(P_0, \lambda\right) \tag{3.46}$$

subject to the constraints

$$P_l\left(P_{0\mathrm{opt}}, \lambda_{\mathrm{opt}}\right) \leq P_{l\max} \quad \text{and/or} \quad D_l\left(P_{0\mathrm{opt}}, \lambda_{\mathrm{opt}}\right) \leq D_{l\max}. \tag{3.47}$$

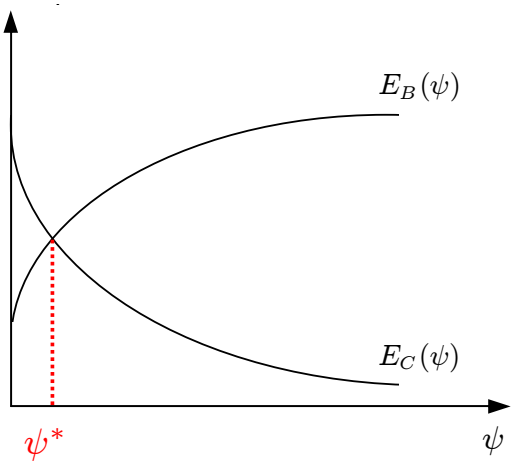


Figure 3.7: Solution of $E_B(\psi) - E_C(\psi) = 0$.

Results obtained in Subsection 3.3.1 can be directly applied to solve this optimization problem. However, when using the effective bandwidth-based approach, as pointed out in Subsection 3.3.2, it has to be taken into account that it only provides statistical QoS guarantees for the average packet loss rate P_l in a system that incorporates an infinitely persistent Type-I hybrid FEC/ARQ protocol. This implies that the probability loss constraint is guaranteed with a small violation probability ϵ. Then, according to Subsection 3.3.2, the system throughput η can be calculated as

$$\eta = \lambda(1 - P_l)$$

and the packet loss probability P_l can be approximated by the tail probability

$$\sup_t Pr\{Q(t) \geq \overline{Q}\}$$

of the queue length at the DLC layer.

As a consequence, in this case (3.46) must be rewritten as

$$\left(P_0^{\text{opt}}, \lambda^{\text{opt}}\right) = \arg \max_{P_0 \in \mathbf{\Phi}, \lambda \in \mathbf{\Theta}} \eta \approx \lambda(1 - \kappa e^{-\psi^* \overline{Q}}) \tag{3.48}$$

subject to the constraint

$$Pr\{P_l \leq P_{l\max}\} \approx \kappa e^{-\psi^* P_{l\max}} \leq \epsilon, \tag{3.49}$$

where ψ^* is the unique real solution of $E_B(\psi) - E_C(\psi) = 0$, as can be observed in Fig. 3.7, and κ is the relation between average arrival rate $\varpi_A \triangleq \lim_{\psi \to 0} E_B(\psi)$ and average service rate $\varpi_S \triangleq \lim_{\psi \to 0} E_C(\psi)$, as shown by Tang and Zhanh (2007). It is worth stating at this point that, except for low input data rates, the tail probability tends to overestimate the packet loss probability and it can only be used as long as $\varpi_A \leq \varpi_S$.

Furthermore, when delay-bound $\mathcal{D}_{l_{\max}}$ is the main QoS constraint, it could be accounted for by using eqs. (3.30) and (3.36). However, in this case the focus is on delay-bound violation probability instead of average packet delay.

Obviously, since P_0 and λ lie in a bounded space $\mathbb{R}_+ \times \boldsymbol{\Theta}$, we can resort to a 2-D exhaustive search to numerically solve the proposed cross-layer optimization problem, for both the *ergodic* and the *instantaneous* options. The bidimensional optimization approach has less degrees of freedom than the multidimensional strategy, resulting in a simpler scheme at the cost of a reduction in performance, as it will be illustrated in the following section.

Once the optimum values of $\boldsymbol{\Gamma}^m$ (either with the multidimensional approach or through the simplified bidimensional approach) and λ have been obtained, these values ought to be accordingly adjusted, with the help of the cross-layer controller, at the PHY and DLC layers, respectively. There exist a lot of strategies to manage the packet arrival rate and to adapt it to the system conditions. In fact, to offer QoS in a network, traffic entering the service provider network needs to be policed on the network boundary routers to make sure the traffic rate stays within the service limit. A Token Bucket can be used to enforce a traffic contract since it provides a precise algorithmic definition for conforming packets and non-conforming packets and therefore, it provides a bound on the average rate. In addition, traffic shapers smooth traffic and send it out at a constant rate. When tokens are exhausted, they buffer packets and send them out later, when tokens become available. Traffic policers, on the contrary, drop non-conforming packets in order to temporarily reduce congestion (Marcon et al., 2011; Mokdad and Ben-Othman, 2012; Vegesna, 2001).

3.5 Numerical Results

The objectives in this section are the following:

1. validation and analysis of the PHY/DLC modeling, consisting of:
    - the PHY layer first-order two-dimensional Markov model introduced in Subsection 3.2.2
    - the different queueing approaches for DLC layer analysis that have been presented in Section 3.3
2. study of the *ergodic* and *instantaneous* options for the AMC scheme developed in Section 3.4.2
3. analysis of the cross-layer design proposals of Section 3.4

To meet the aforementioned objectives, we will compare the analytical performance expressions of both the queueing Markov model-based approach (denoted as **2D-FSMC** in the following) and the effective capacity-based approach (denoted as **2D-FSMC-EbEc** in the following) introduced in Section 3.3, which are based on the PHY-layer first-order two-dimensional Markov model introduced in Section 3.2.2. They will be confronted with the analytical expressions used by Le

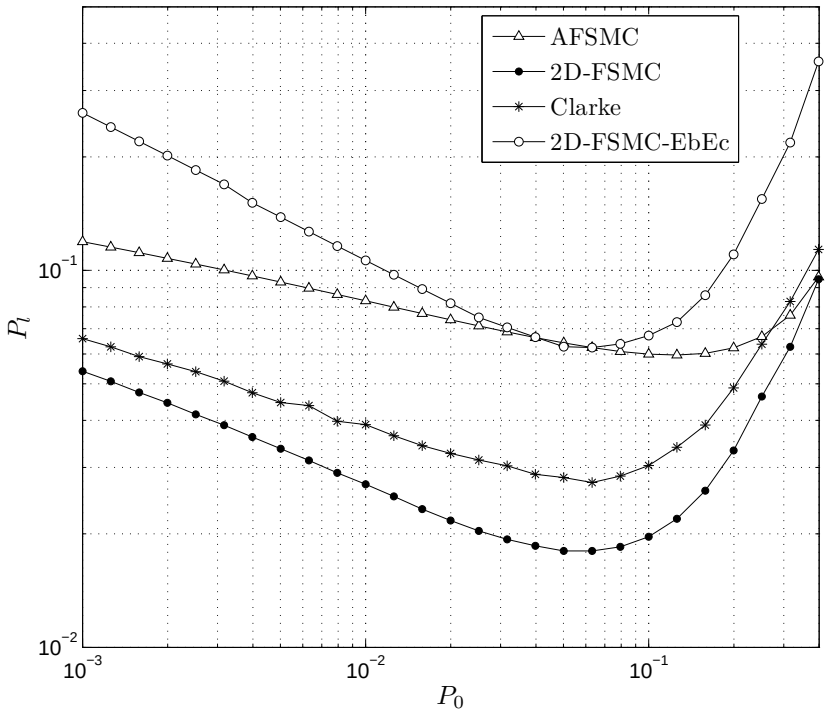


Figure 3.8: Average packet loss rate vs. target PER.

et al. (2006b) (denoted as **AFSMC** in the following), which were based on the PHY-layer first-order AFSMC model introduced by Liu et al. (2005b) and have been extensively used in the literature (see for example (Ishizaki and Hwang, 2007; Le et al., 2006a; Poggioni et al., 2007; Wang et al., 2007)). Moreover, results obtained using these models will be contrasted with computer simulations obtained using Clarke's statistical Rayleigh fading process to model the wireless flat-fading channel (Clarke, 1968) (denoted as **Clarke** in the following).

Unless otherwise specified, numerical results will be obtained for the following default parameters: a normalized maximum Doppler frequency $f_d T_f = 0.02$, an average received SNR $\overline{\gamma} = 8$ dB, a buffer size $\overline{Q} = 50$, a number of channel states $K = 10$, a target PER of $P_0 = 0.01$, a parameter $b = 2$ and a D-BMAP characterized with the transition probability matrix defined in (2.4). Moreover, the *ergodic* option for the AMC scheme will be considered as the default approach.

3.5.1 Validation and study of the PHY/DLC modeling

Figures 3.8 and 3.9 show, respectively, the dependence of the average packet loss rate P_l and the average throughput η on the target average PER P_0. In this scenario, where ARQ-based error control with infinite persistence at the DLC layer is considered, the packet loss rate is simply equal to the buffer overflow probability. Results reveal that an increase of P_0, implying the utilization of higher order TMs, causes a growth of the queueing service rate and thus a decrease of the buffer overflow probability and packet loss rate, while the average throughput raises accordingly. However, when the increase of the service rate cannot cope

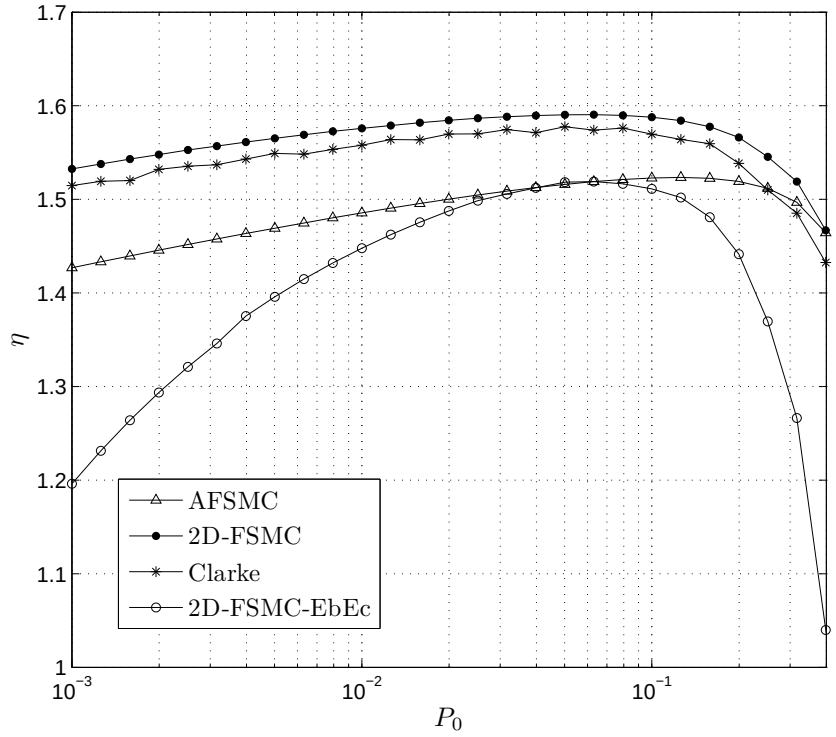


Figure 3.9: Average throughput vs. target PER.

with the huge number of required retransmissions, P_l rapidly takes off and η diminishes.

As expected, an analogous behaviour can be observed in Figures 3.10 and 3.11 for the average packet delay D_l and the average queue length L_q, respectively. It is worth noting that the 2D-FSMC-EbEc model does not allow to plot the D_l and L_q curves, as it has only been used to provide statistical QoS guarantees for the average packet loss rate P_l. It can be observed that the increase in the throughput obtained for higher P_0 values causes a decrease in D_l and L_q. Nevertheless, after reaching their minimum values, and similarly to the P_l and η behaviour, the average packet delay and queue length start to raise significantly.

As it can be observed, the behaviour of the *real* system based on Clarke's model is more faithfully reproduced by the proposed 2D-FSMC model, although this model slightly underestimates the average packet loss rate, while, as expected according to Section 3.4, the 2D-FSMC-EbEc approach overestimates P_l. Remarkably, note how the shape and location of the minimum of the curves obtained using Clarke's model coincide with those obtained using both 2D-FSMC and 2D-FSMC-EbEc proposed schemes, even for a small number of channel states K. On the contrary, the optimum throughput estimated with the AFSMC model is obtained for a P_0 value that is far from the one corresponding to the real system simulation. The accuracy in determining the location of the maximum of the average throughput is particularly important in order to ensure an optimal cross-layer design.

Figures 3.12, 3.13 and 3.14 depict the average throughput as a function of the

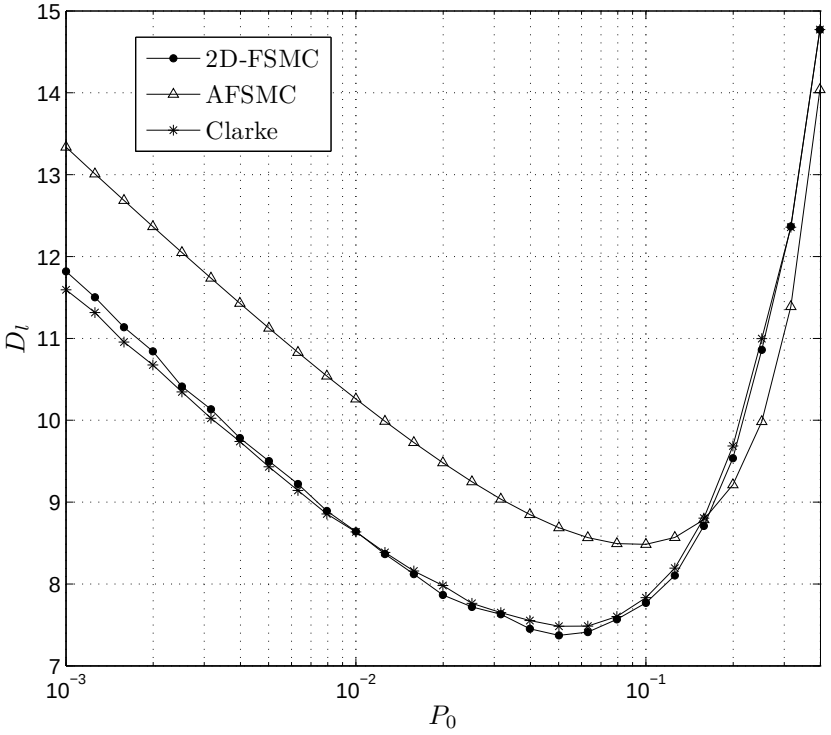


Figure 3.10: Average packet delay vs. target PER.

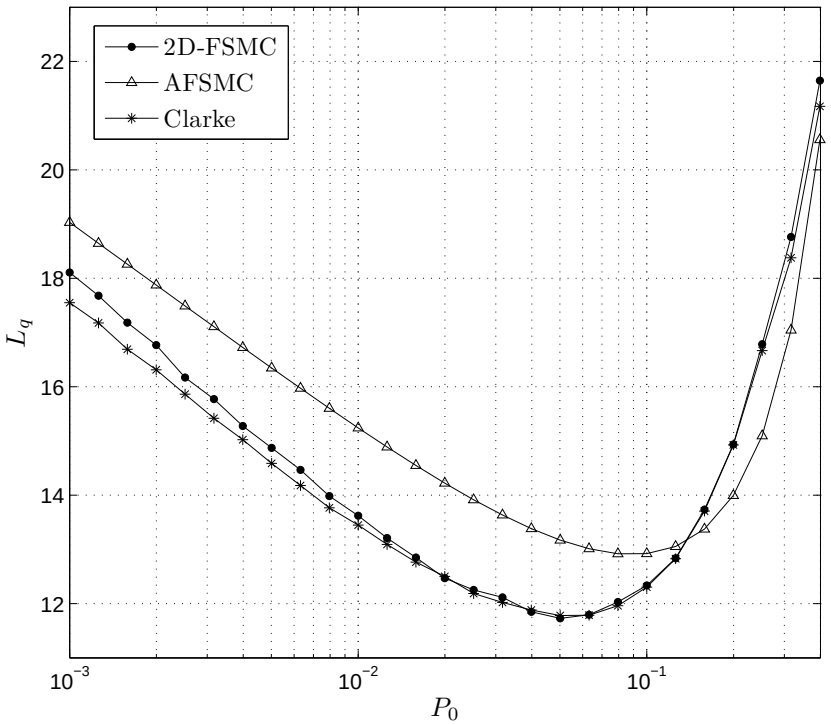


Figure 3.11: Average queue length vs. target PER.

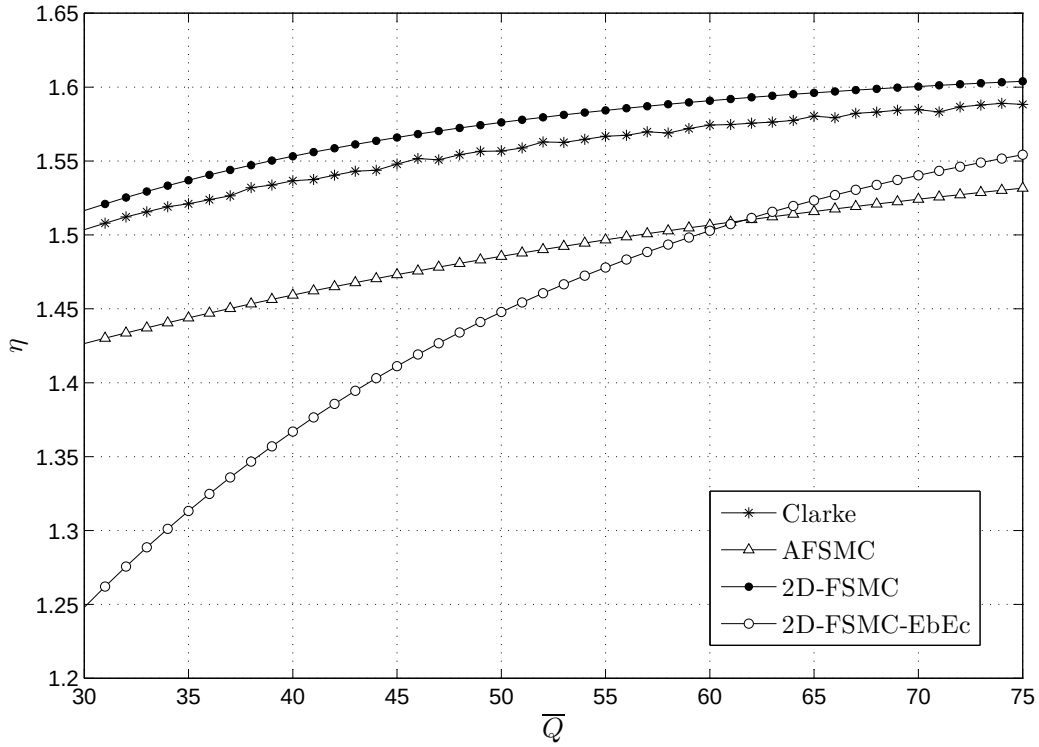


Figure 3.12: Average Throughput vs. Queue length.

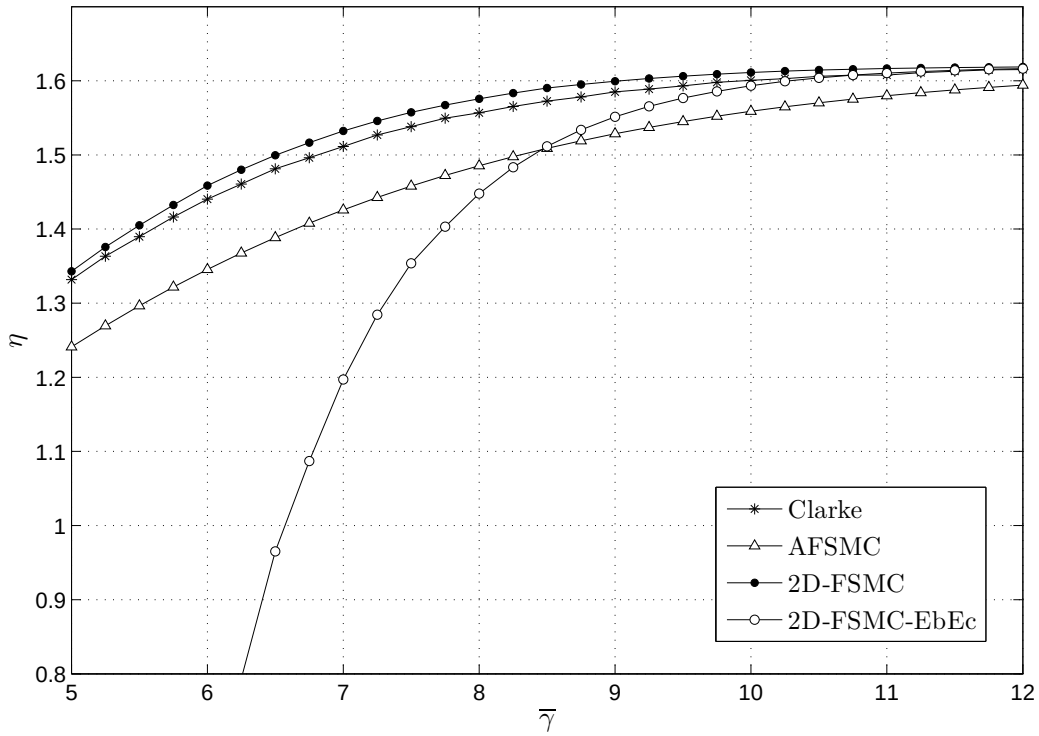


Figure 3.13: Average Throughput vs. Average SNR.

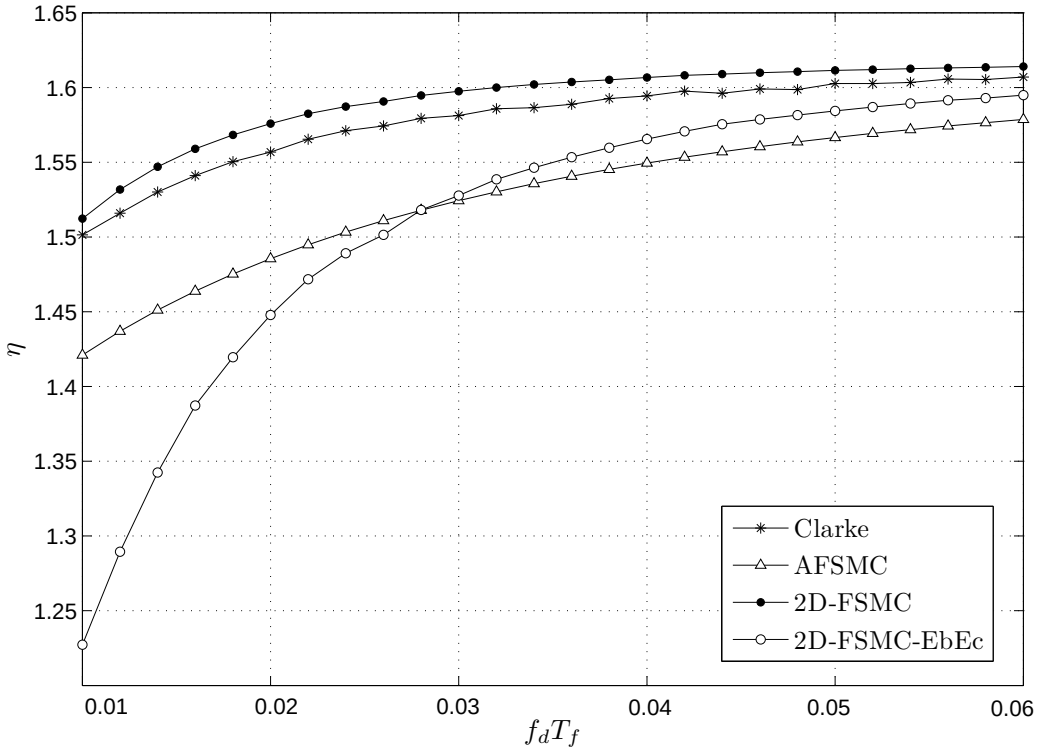


Figure 3.14: Average Throughput vs. maximum Doppler frequency.

queue length, the average SNR, and the maximum normalized Doppler frequency, respectively. From the shape of these plots, one can infer that, as expected, a bigger queue size implies a lower packet loss rate due to buffer overflow, and allows achieving higher throughput values. A greater average SNR corresponds to a better channel quality and, therefore, results in higher average transmission rate and implies a growth of η. On the contrary, a lower f_dT_f leads to a longer fading duration, which increases the average packet loss rate and, as a consequence, decreases the average throughput. From these figures it can be deducted that the 2D-FSMC model is able to accurately track the system parameters (throughput, packet loss rate) under a variety of environment and system conditions. Concerning the 2D-FSMC-EbEc model, Fig. 3.12 reveals that, as expected, its accuracy in predicting the performance metrics increases with the queue length. In addition, Fig. 3.13 shows that the accuracy of this model decreases when the average SNR is reduced; this is because as the average SNR decreases the output rate ϖ_S gradually approaches the input rate ϖ_A and, as stated in (Wu and Negi, 2003), when $\varpi_A \geq \varpi_S$ the queue tail approximation is no longer valid. Finally, Fig. 3.14 reveals that 2D-FSMC-EbEc model losses accuracy for quasi-static scenarios. For low values of the maximum Doppler frequency the channel remains in a particular state for a long time and therefore, in order to obtain the same accuracy, the queue length should be increased. In spite of these limitations the 2D-FSMC-EbEc model is a simple and powerful tool that could be used to extend this analysis to multiuser scenarios for a significant set of environments of practical interest.

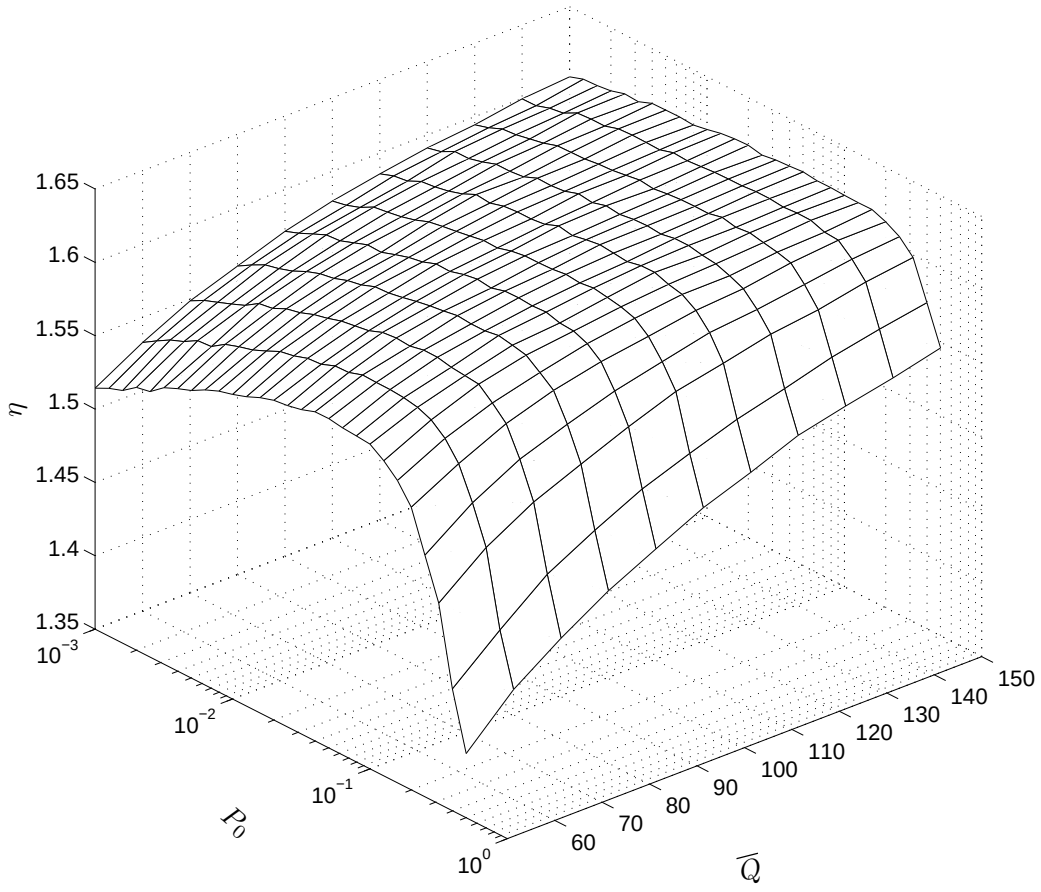


Figure 3.15: Average throughput versus target PER with different buffer sizes.

Figure 3.15 plots the dynamics of the average throughput η against the target PER P_0 and the buffer size $\overline{Q}$ obtained with the 2D-FSMC model. As expected, when $\overline{Q}$ gets larger, η increases. It is worth noting that for greater buffer sizes there is a smaller dependence of η on P_0. This behaviour stems from the fact that more spacious buffers can smooth the variations of P_l caused by the changes in the number of required retransmissions as a result of the different values of P_0. The optimum P_0 value, which corresponds to the minimum average packet loss rate P_l and, thus, to the maximum average throughput η, remains nearly invariant in spite of the variation of the buffer size $\overline{Q}$, as a result of the fact that the considered values of $\overline{Q}$ are much larger than the obtained average queue lengths L_q (for instance, see Fig. 3.11 for the case of $\overline{Q} = 50$).

The dependence of the average throughput η on both, the target PER P_0 and the average arrival rate λ, is jointly analyzed and illustrated in Fig. 3.16, which corresponds to the 2D-FSMC model. As it can be observed, η increases for higher λ values, until it stabilizes. Besides, the influence of P_0 on η is only noticeable for high values of λ, which correspond to the case in which a buffer overflow may occur.

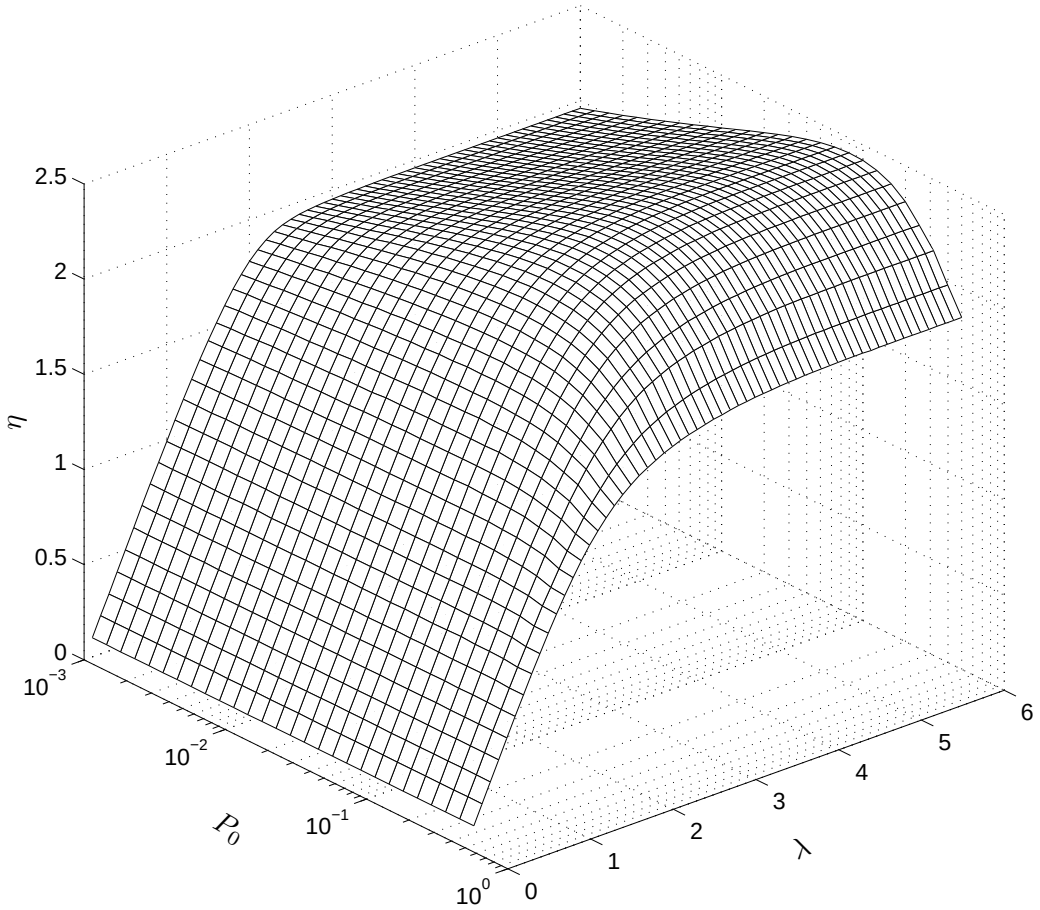


Figure 3.16: Average throughput versus target PER with different average arrival rates.

3.5.2 Comparing the *ergodic* and *instantaneous* approaches

In order to understand the behaviour of the system under the *ergodic* and *instantaneous* options for the AMC scheme, let us compare the TM selection probabilities for both cases, which are depicted in Fig. 3.17. As it can be observed, TM3, TM4 and TM5, which are high order TMs, are more probably selected when the *ergodic* option is utilized, whereas the lower order TMs, i.e TM0, TM1 and TM2, are more probably used when the *instantaneous* option is chosen. This behaviour is a consequence of the fact that the objective of the AMC scheme under the *ergodic* option is the maximization of the data rate while maintaining an average PER less or equal than a prescribed value P_0, while under the *instantaneous* option, it is required that the instantaneous PER falls below this prescribed value. Therefore, the *instantaneous* option is more stringent than the *ergodic* one and, as a result, gives birth to more conservative AMC partitions in terms of higher reliability transmissions. The graph also confirms that greater P_0 values entail an increase in the probability of selection of higher order TMs (i.e., TM3, TM4 and TM5), regardless of the considered option for the AMC scheme, whereas the probability of selecting TM0, TM1 and TM2 decreases.

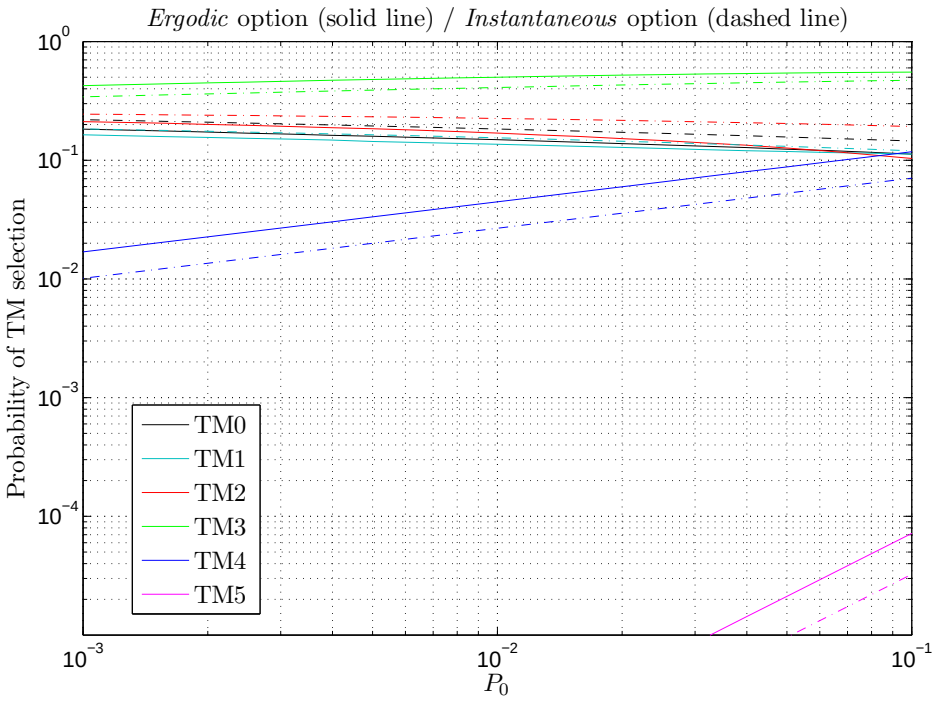


Figure 3.17: TM selection probabilities for the *ergodic* and *instantaneous* options for the AMC scheme.

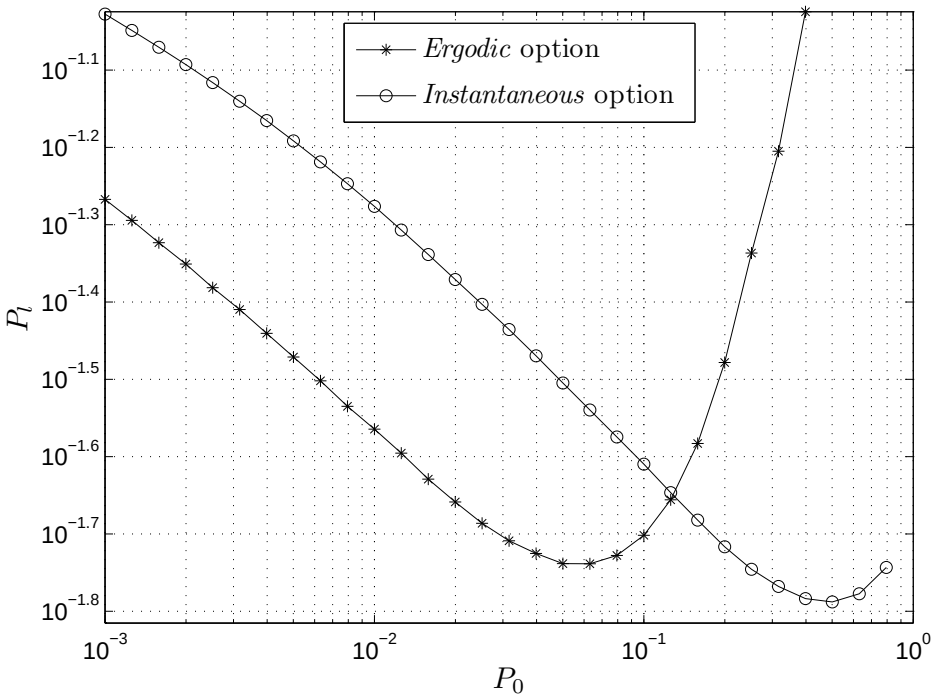


Figure 3.18: Average packet loss rate vs. target PER - *ergodic* and *instantaneous* options for the AMC scheme.

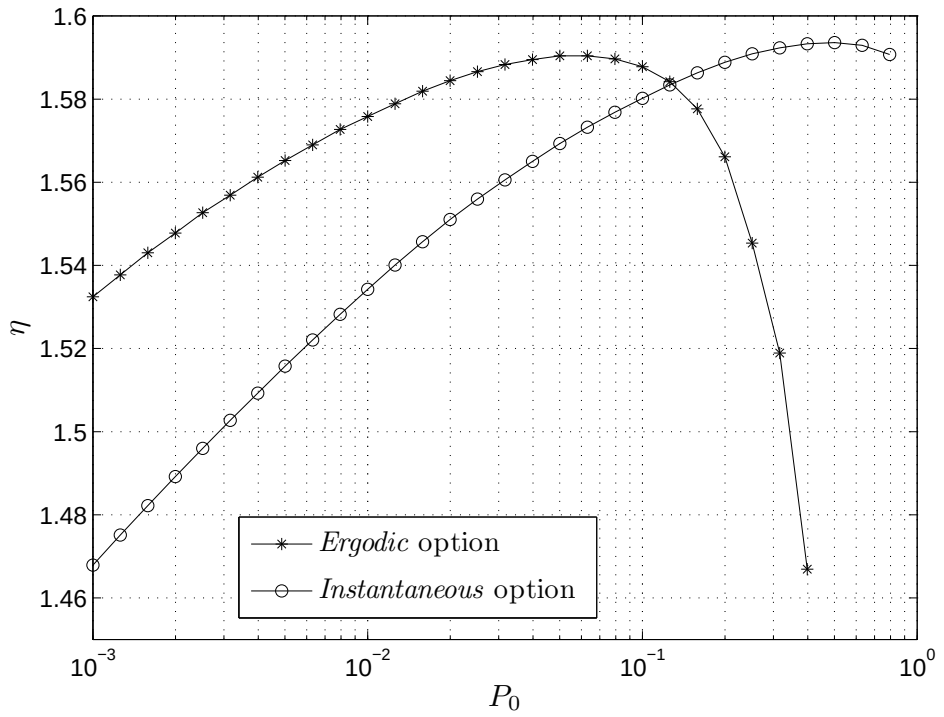


Figure 3.19: Average throughput vs. target PER - *ergodic* and *instantaneous* options for the AMC scheme.

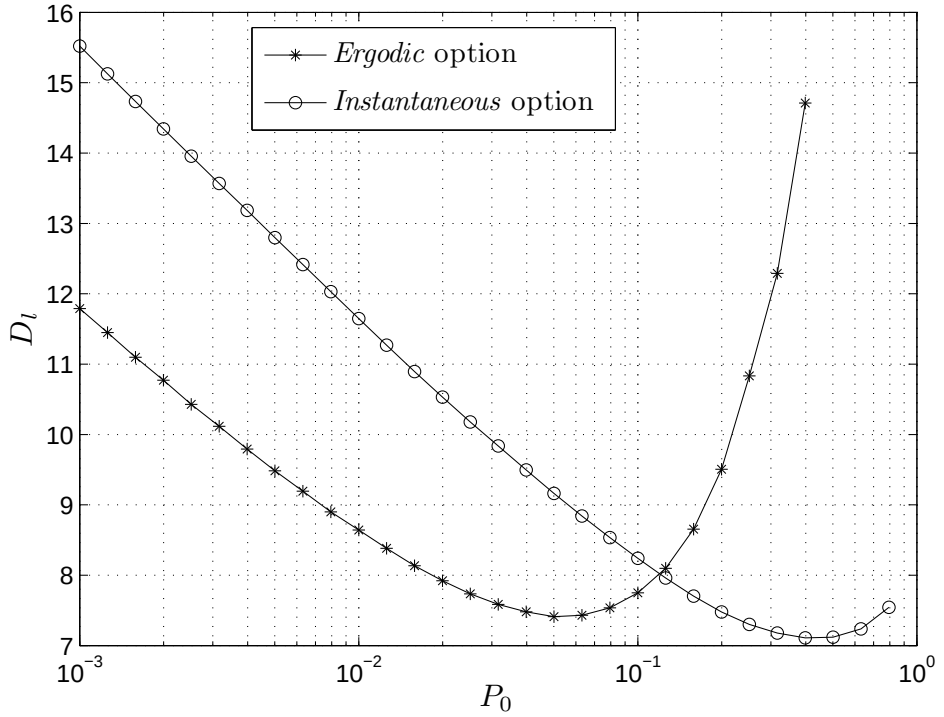


Figure 3.20: Average packet delay vs. target PER - *ergodic* and *instantaneous* options for the AMC scheme.

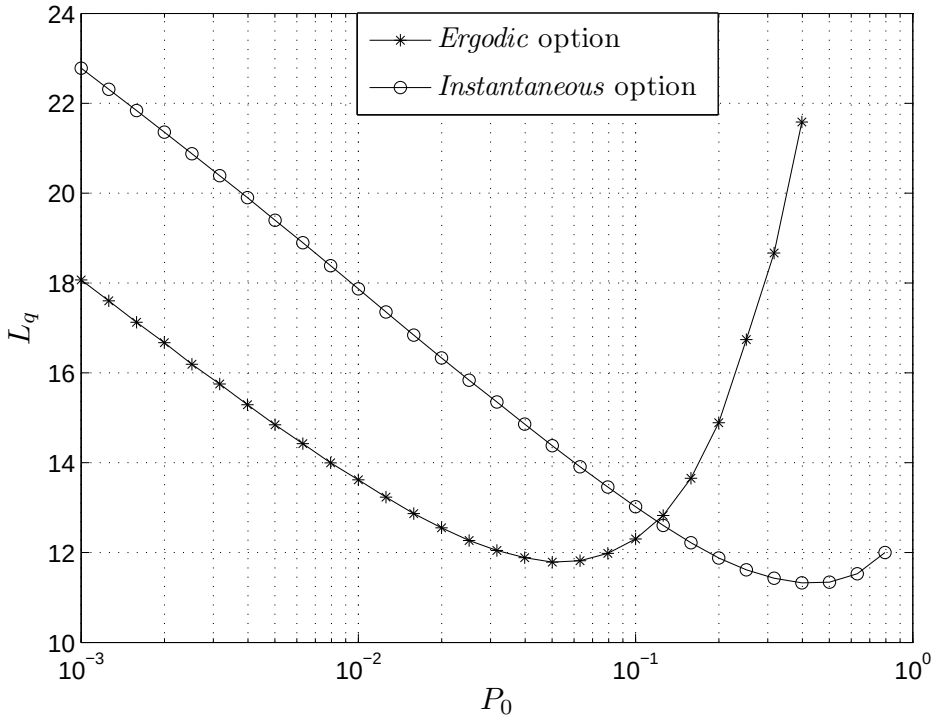


Figure 3.21: Average queue length vs. target PER - *ergodic* and *instantaneous* options for the AMC scheme.

The behaviour of these TM selection probabilities affect the QoS provisioning performance parameters, illustrated in Figs. 3.18, 3.19, 3.20 and 3.21, which have been obtained using the analytical expressions of the 2D-FSMC model. It is clearly observed that the curves corresponding to both approaches have an analogous behaviour, although with a left shift in the graphs corresponding to the *ergodic* option. This is because similar AMC switching thresholds can be obtained for both options, but for lower P_0 values with the *ergodic* approach. The reason is that, for a given target PER, higher order TMs are selected with the *ergodic* option in comparison to the *instantaneous* option and, thus, to obtain similar AMC switching thresholds with the latter scheme, a higher P_0 value is needed. As this approach is more stringent, it allows the achievement of slightly lower minimum values for the average packet loss rate, average packet delay and average queue length, whereas a higher maximum average throughput is obtained. As expected, when the *ergodic* option is considered, a lower optimum P_0 value is obtained in comparison to the one corresponding to the *instantaneous* approach.

3.5.3 Analysis of the cross-layer design

With the objective of analyzing the cross-layer design presented in Section 3.4, three optimization problems will be considered. In the first one, to validate and compare the proposed cross-layer framework when using the different PHY and DLC layer models (i.e., AFSMC, 2D-FSMC and 2D-FSMC-EbEc), Figs. 3.22, 3.23 and 3.24 depict the optimum average throughput for different values of the

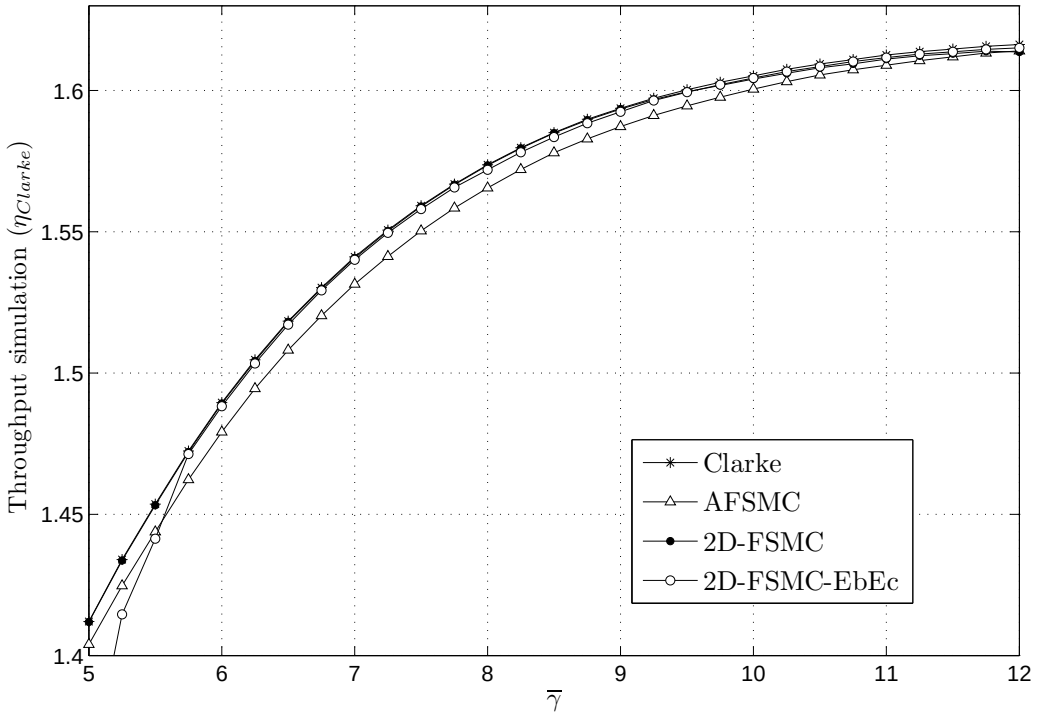


Figure 3.22: Simulated optimum average throughput vs. Average SNR.

average received SNR, maximum normalized Doppler frequency and buffer size. For clarity, we have set the average arrival rate using the D-BMAP characterized by the transition probability matrix (2.4) and only the parameter P_0 has been optimized. Additionally, no constraints on P_l and D_l have been set as otherwise throughput curves would only appear in the range of values fulfilling these constraints, impairing the comparison among the different techniques. These graphs plot η^{opt} obtained through Monte-Carlo simulation using the optimum values of P_0. As it can be observed, AFSMC, 2D-FSMC and 2D-FSMC-EbEc models are able to provide a good estimation/prediction of P_0^{opt} in terms of the average system throughput. However, it is interesting to note that the AFSMC model consistently provides the worst estimation. It should also be pointed out that although the 2D-FSMC-EbEc approach clearly overestimates the packet loss rate (see Fig. 3.8), it provides a better P_0^{opt} estimation than the AFSMC model.

A second optimization problem is presented in Fig. 3.25 with the aim of comparing the optimization process when the *ergodic* and the *instantaneous* options for the AMC scheme are considered. This graph plots the obtained QoS parameters η^{opt}, P_l, D_l and P_0^{opt} for different $\overline{\gamma}$ values, for both options. For the sake of clarity and simplicity in the interpretation of the obtained results, only the P_0 value has been optimized and no constraints have been set. A system with a buffer size $\overline{Q} = 10$ and a D-BMAP characterized by the transition probability matrix (2.4) has been considered. As expected, higher $\overline{\gamma}$ values result in the use of higher order TMs, which in turn imply a decrease in P_l and D_l and, consequently, an enhancement of the achieved throughput. The behaviour observed under the two

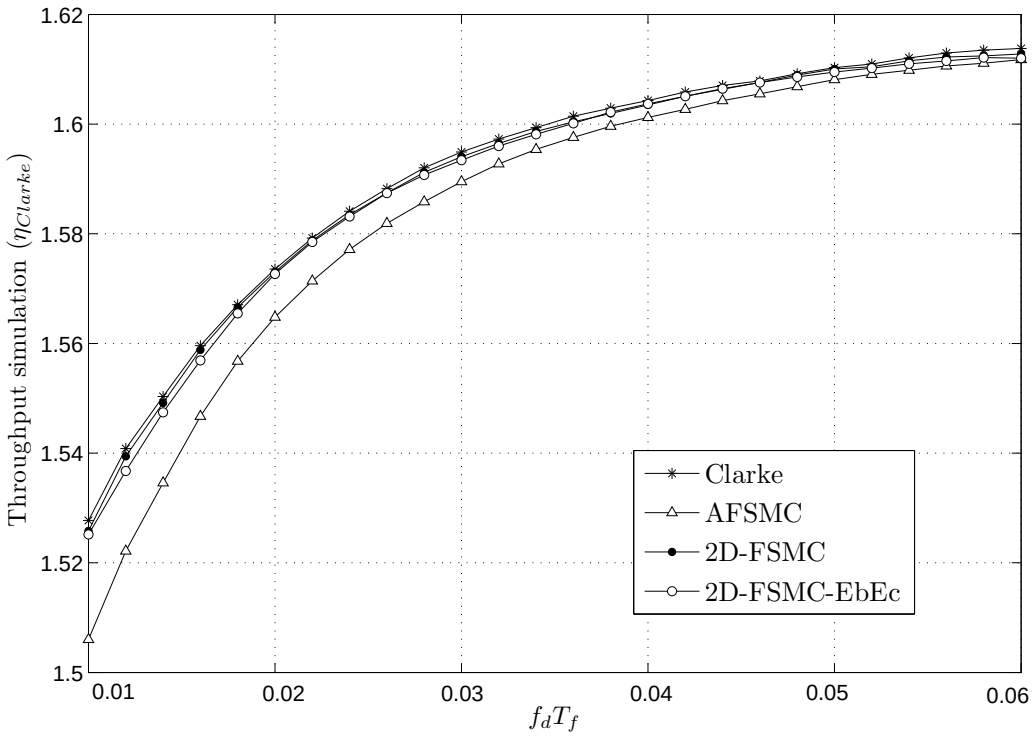


Figure 3.23: Simulated optimum average throughput vs. Maximum Doppler Frequency.

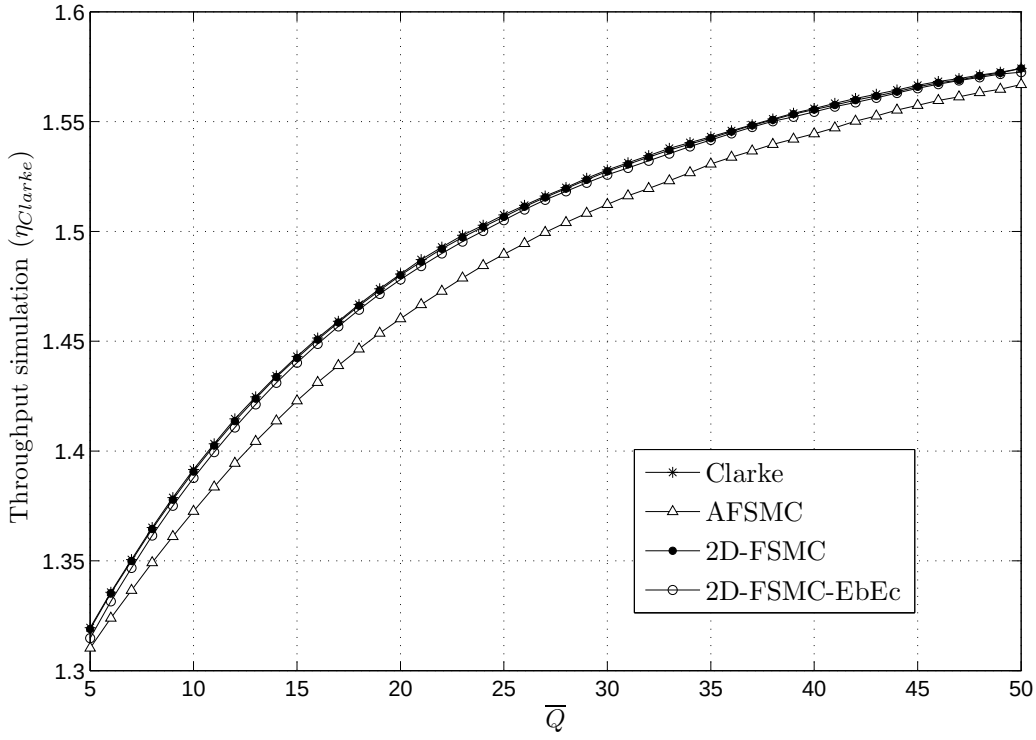


Figure 3.24: Simulated optimum average throughput vs. Queue length.

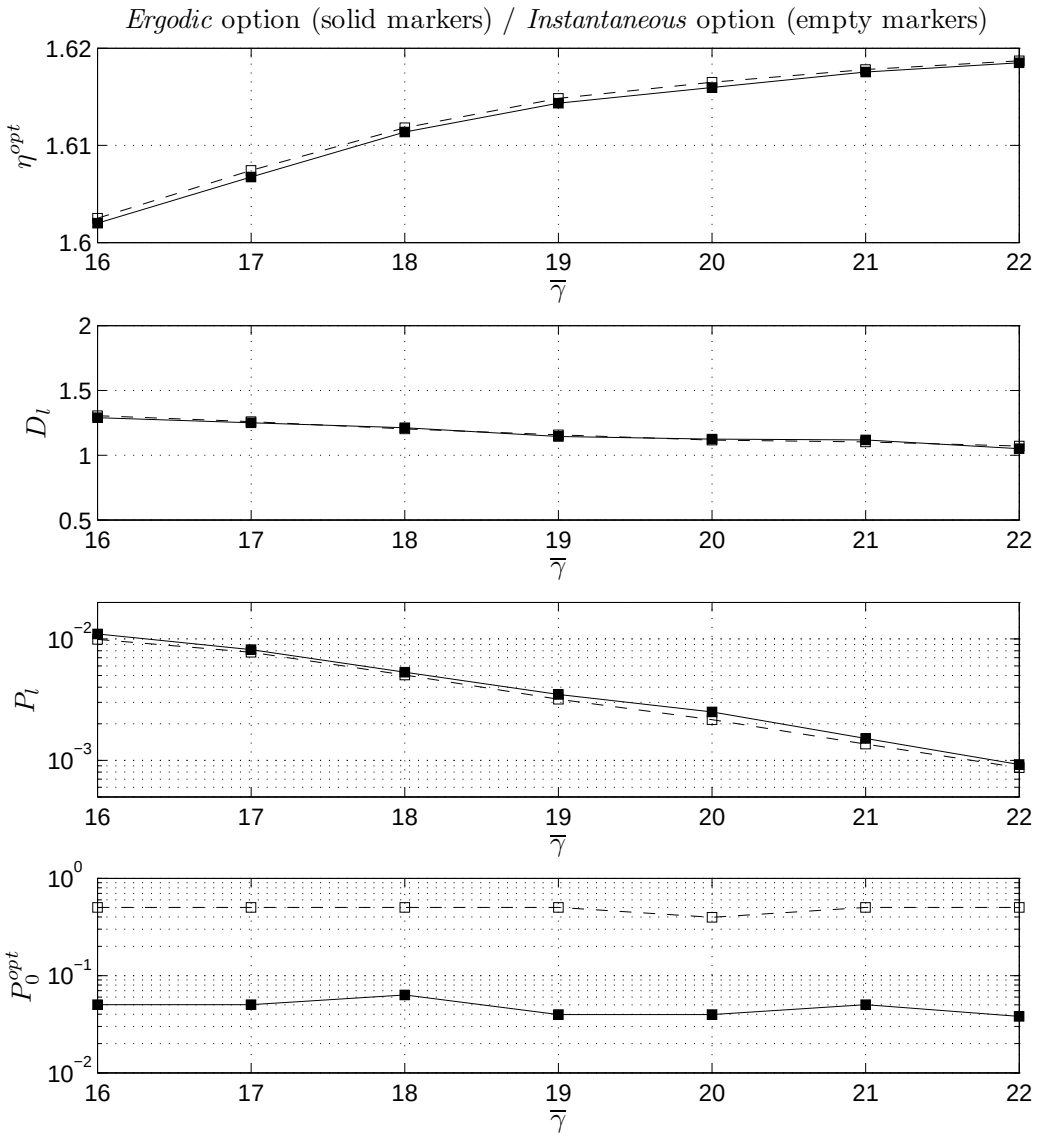


Figure 3.25: Optimization vs. Average SNR - *Ergodic* and *Instantaneous* options.

AMC options is aligned with that obtained in Subsection 3.5.2: a slightly better performance is achieved with the *instantaneous* option, which corresponds to P_0^{opt} values that are clearly higher than those corresponding to the *ergodic* option.

Finally, a constrained optimization problem is also proposed, in which QoS-guaranteed traffic characterized by a maximum average packet loss rate $P_{l\max} = 0.01$ packets/frame and a maximum average packet delay $D_{l\max} = 2$ frames has been considered. Traffic has been generated using a truncated Poisson process with a truncation length equal to 10 packets, i.e. $\mathbf{\Theta} = (0, 10]$. In this third optimization problem, the AMC scheme will be based on the *instantaneous* option. Results have been obtained using both, the multidimensional and the simplified bidimensional optimization problems, for a maximum queue length $\overline{Q} = 30$.

Figure 3.26 plots the dependence of the obtained QoS parameters η, P_l and D_l,

on the average SNR, as well as the optimum values for the arrival rate λ^{opt}. For low values of the average SNR, $D_{l\max}$ is the limiting constraint of the optimization problem, while the average packet loss rate P_l remains under $P_{l\max}$. On the contrary, for high $\overline{\gamma}$ values, the use of higher order TMs implies a decrease in D_l and P_l becomes the limiting QoS parameter while the packet delay falls below $D_{l\max}$. As expected, increased SNRs values entail an enhancement in the achieved throughput while ensuring the fulfillment of both QoS requirements. From the shape of these curves it is obvious that when the AMC switching thresholds are found using the multidimensional optimization approach, a higher maximum throughput is obtained in comparison to that achieved with the simplified bidimensional optimization. As it can be observed, although this enhancement becomes marginal for medium-to-high SNR values, it is significant for worst channel conditions. In fact, for the proposed constrained optimization problem with a SNR value $\overline{\gamma} = 6$ dB it is not possible to obtain a solution when the bidimensional approach is considered, while an average throughput $\eta = 0.1816$ packets/frame can be achieved with the multidimensional strategy.

The dependence of this cross-layer design on the maximum normalized Doppler frequency is depicted in Fig. 3.27 for an average received SNR $\overline{\gamma} = 15$ dB. For both the multidimensional and the bidimensional approaches, the maximum allowed delay $D_{l\max}$ is the limiting QoS parameter while the average packet loss rate P_l remains under $P_{l\max}$. These curves confirm that with higher $f_d T_f$ values, which correspond to shorter deep fading durations and, thus, to better channel conditions, the average packet loss rate and delay diminish and, consequently, the optimum sustainable arrival rate λ^{opt} increases while ensuring the fulfillment of both QoS requirements. As in the analysis of the dependence of this constrained optimization problem on the average received SNR, a slightly higher throughput can be obtained with the multidimensional approach than with the simplified bidimensional strategy.

3.6 Chapter summary

In this chapter an analytical framework for the cross-layer design, analysis and optimization of wireless systems combining AMC with an infinitely persistent Type-I hybrid FEC/ARQ protocol has been proposed. In order to achieve this goal, the PHY layer service process has been modeled as a simple first-order two-dimensional FSMC. The proposed PHY layer model has been compared with both Clarke's statistical Rayleigh fading model (Clarke, 1968) and the model introduced by Le et al. (2006b), which was based on the first-order AFSMC model introduced by Liu et al. (2005b). Numerical results have shown that the performance metrics of the *real* system based on Clarke's model are more faithfully reproduced by the proposed first-order two-dimensional FSMC model than by the first-order AFSMC approach. This demonstrates the implicit potential of using multidimensional channel fading Markov models in designing cross-layer strategies.

Based on this PHY layer multidimensional FSMC, two different approaches used to model the interactions between the PHY layer and the DLC layer queueing behaviour have been investigated. The first one is based on the use of a DTMC

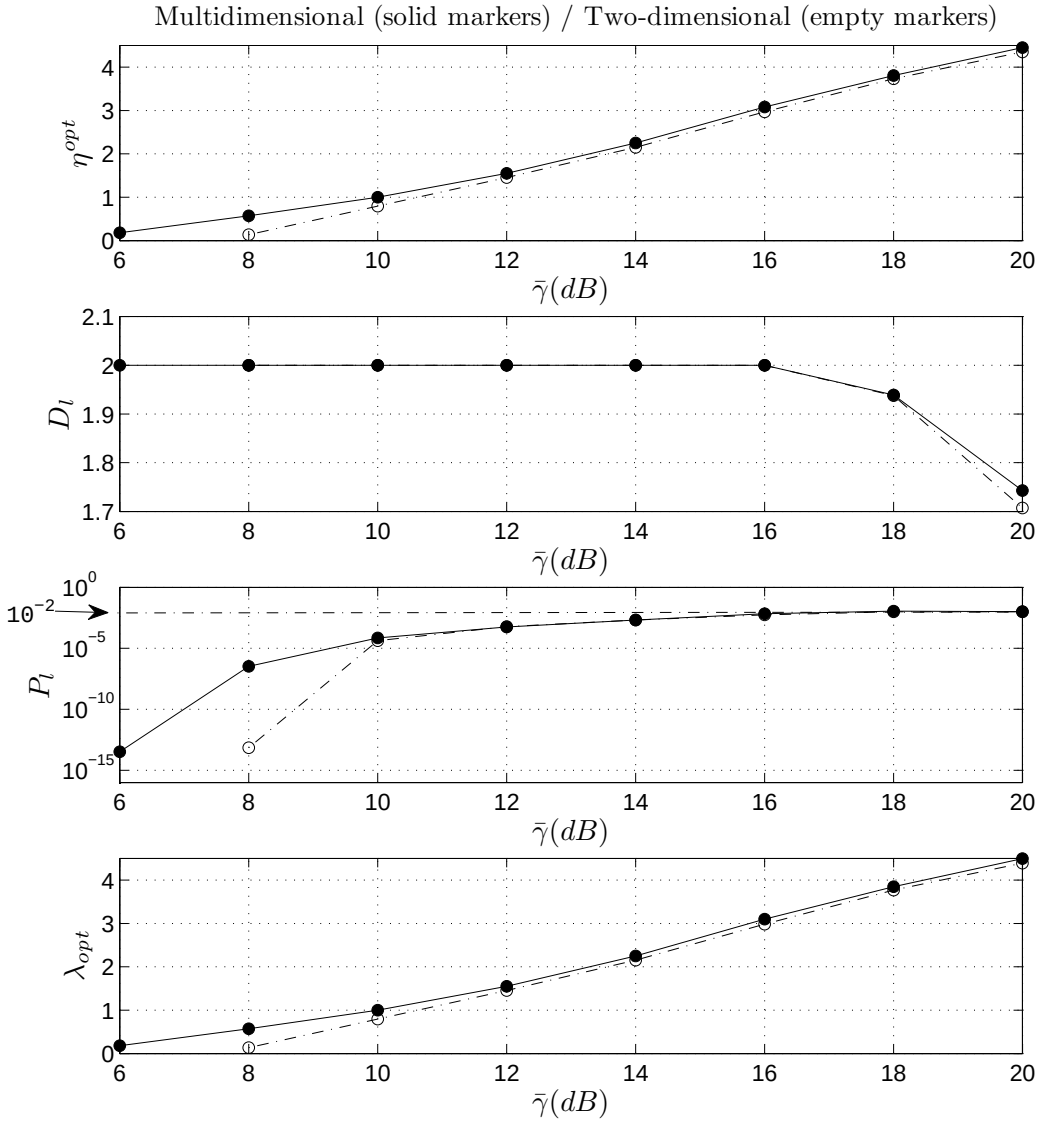


Figure 3.26: Optimization vs. Average SNR - multidimensional and bidimensional optimization.

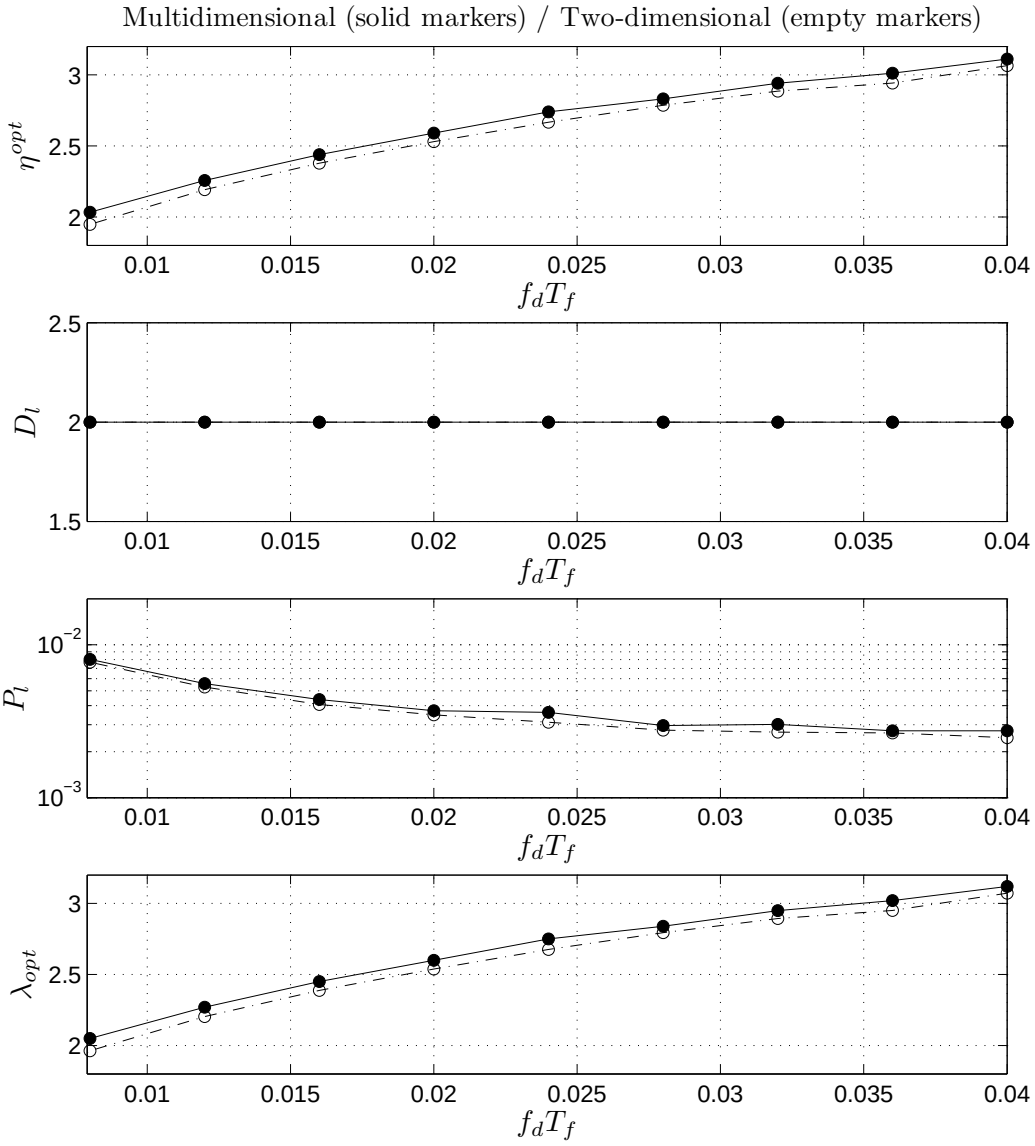


Figure 3.27: Optimization vs. Maximum normalized Doppler frequency - multidimensional and bidimensional optimization.

that jointly describes the statistical behaviour of the arrival process, the queueing system and the PHY layer. Performance metrics such as the system throughput, average delay or average packet loss probability have been accurately derived by solving the global transition probability matrix of the system. Unfortunately, valuable radio link-level performance measures are provided at the cost of an increase of complexity with the number of arrival process states, the number of PHY layer Markov model states or the maximum queue length of the system. The second approach, with lower complexity, follows a completely different methodology that relies on the effective bandwidth and effective capacity theories. To assess the validity of the proposed models, computer simulation results have been presented to show that the derived closed-form analytical expressions can be used to faithfully reproduce results based on Clarke's statistical Rayleigh fading model. However, it has been shown that for many cases of practical interest the asymptotic approximations used in the effective bandwidth/capacity-based approach seriously compromise its usefulness in the proposed cross-layer framework. Moreover, the analysis cannot be easily extended to systems in which an (H)ARQ scheme with finite persistence is used. For these reasons, from this point onwards, only the Markov-based approach will be considered in the analysis.

The proposed cross-layer analytical framework has allowed the formulation of multidimensional and simplified bidimensional optimization problems to maximize the system throughput under maximum packet loss rate and delay constraints. In comparison to the simplified bidimensional approach, a higher maximum throughput is obtained with the multidimensional optimization, although this enhancement is only noticeable in cases where the system experiences bad channel conditions. In the bidimensional simplification, two options have been considered for the AMC scheme: the *ergodic* and the *instantaneous* options. A nearly identical behaviour of the optimum system performance has been observed with both schemes, although the *instantaneous* approach slightly outperforms the *ergodic* one.

CHAPTER 4

Systems based on a truncated (hybrid) ARQ protocol

This chapter generalizes and extends the analytical Markov-based model presented in Chapter 3, where an infinitely persistent Type-I hybrid FEC/ARQ was considered, to the case of a truncated (H)ARQ scheme. Thus, it provides a unified analytical framework allowing the theoretical characterization of both the classical Type-I hybrid FEC/ARQ protocol and the more sophisticated Hybrid ARQ schemes based on Chase Combining and Incremental Redundancy. Following the same methodology used in Chapter 3, a first-order two-dimensional Markov model of the PHY layer is obtained. Afterwards, based on the use of DTMCs, the system model is developed and analytical performance measures are derived allowing the formulation of QoS constrained cross-layer optimization problems. Numerical results confirm the validity of the model an show that the HARQ schemes lead to a remarkable improvement on the maximum achieved throughput when compared to the classical type-I hybrid FEC/ARQ scheme.

4.1 Introduction

As shown in previous chapters, in order to provide heterogeneous QoS guarantees to applications, most wireless communications standards combine the error-correcting capability of (H)ARQ protocols at the DLC layer with the adaptation ability of AMC strategies at the PHY layer. The analytical framework proposed in Chapter 3 assumed the use of an infinitely persistent Type-I hybrid FEC/ARQ protocol. In practice, (H)ARQ protocols with finite persistence (also know as truncated (H)ARQ protocols) are used in order to enhance the average throughput while reducing delay jitter and/or limiting the transmitter and receiver buffer sizes.

Liu et al. (2004) proposed a scheme combining AMC at the PHY layer with a truncated ARQ protocol at the DLC layer in order to improve the overall sys-

tem spectral efficiency. However, for analytical convenience, the fading channel coefficients corresponding to the original and the retransmitted packets were assumed independent and identically distributed random variables and, furthermore, queueing effects on the average packet delay were not taken into account. In (Liu et al., 2005a), the same authors also proposed a cross-layer design combining finite buffer queueing at the DLC layer with AMC at the PHY layer and applied finite-state Markov chain analysis to derive analytical expressions for the packet loss rate and throughput. This paper, however, did not take into account the possible performance improvement from the ARQ protocol (i.e., unsuccessfully transmitted packets were dropped) and ignored correlation in the traffic arrival process. Wang et al. (2007) tried to solve some of the previous flaws by generalizing the cross-layer combining of queuing with AMC in (Liu et al., 2005a) and the cross-layer combining of queuing with truncated ARQ and AMC in (Liu et al., 2004). However, similarly to (Liu et al., 2004) and (Liu et al., 2005a), they assumed a memoryless packet arrival process. Moreover, in order to facilitate mathematical tractability of the queueing process, they relied on the rather unrealistic assumption of considering a time slotted system were only one frame was transmitted per slot, with each frame at the PHY layer containing at most one packet from the DLC layer. As it has been mentioned in Section 3.1, traffic burstiness was considered by Le et al. (2006b), where they provided an analytical framework for point-to-point wireless systems with infinite/finite buffer queueing and ARQ-based error control at the DLC layer, and AMC at the PHY layer. Nevertheless, infinitely persistent "pure" ARQ-based error control schemes were considered, while the generalization to more sophisticated truncated-HARQ protocols was not addressed at all. Kang et al. (2009) considered a joint design approach where IR-based HARQ was associated with an AMC design at the PHY layer, although the queueing process was not faced at all.

A general shortcoming of all these proposals is, as mentioned in Section 3.1, their reliance on first-order AFSMCs to model the wireless fading channel. In Section 2.4 it has been shown that the exponentially decaying ACF of the first-order AFSMCs can not fit the hypergeometric ACF of the statistical Rayleigh fading process, thus compromising the design of higher layer protocols. As for the infinitely persistent ARQ-based analysis presented in Chapter 3, a first-order two-dimensional FSMC model for the PHY layer will be developed, which is able to improve the ACF fitting of first-order AFSMCs.

In this chapter a new analytical link-level queueing model of a point-to-point adaptive multi-rate wireless system using truncated (H)ARQ is proposed. This contribution generalizes and extends the analytical tools presented in Chapter 3, where an infinitely persistent type-I hybrid FEC/ARQ was assumed. Using this approach, analytical expressions for performance metrics such as throughput, average packet delay, queue length and packet loss rate, both due to buffer overflow and due to exceeding the maximum number of allowed retransmissions, are derived. The analytical link-level queueing model, which is based on the use of DTMCs, is then applied to formulate a cross-layer design, conceived as a multi-dimensional constrained optimization problem, which can be used to exploit the joint impact on QoS performance measures of both AMC at the PHY layer and truncated (H)ARQ-based error control at the DLC layer. A bidimensional simpli-

Table 4.1: 802.16e TMs with convolutionally coded modulation.

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
Modulation	BPSK	QPSK	QPSK	16QAM	16QAM
Code rate $R_c^{(n)}$	1/2	2/3	5/6	2/3	5/6
R_n (bits/symbol)	1/2	4/3	5/3	8/3	10/3
$a_{n,0}$	4447.4	2068.5	514.7	850.9	142.9
$a_{n,1}$	2298.6	1344.3	3297.8	372.3	895.7
$a_{n,2}$	5944.9	1428.4	7247.9	3567.3	1057.1
$a_{n,3}$	5180.5	4374.3	8498.3	2335.5	5017.6
$g_{n,0}$	11.104	3.315	1.759	0.816	0.339
$g_{n,1}$	21.012	6.997	6.196	1.895	1.657
$g_{n,2}$	34.203	10.569	10.224	3.378	2.706
$g_{n,3}$	45.627	15.744	13.852	4.228	4.089
$\gamma_{p_{n,0}}$ (dB)	-1.212	3.623	5.502	9.172	11.660
$\gamma_{p_{n,1}}$ (dB)	-4.337	0.127	1.164	4.946	6.131
$\gamma_{p_{n,2}}$ (dB)	-5.950	-1.629	-0.608	3.841	4.105
$\gamma_{p_{n,3}}$ (dB)	-7.271	-2.737	-1.850	2.635	3.189

fication approach of this problem is also proposed, resulting in a simpler scheme at the cost of a reduction in performance.

4.2 Physical layer modeling

4.2.1 AMC pool and characterization of the packet error rate

We consider a truncated (H)ARQ protocol with a maximum allowed number of retransmissions set to N_r. Truncated type-I hybrid FEC/ARQ, HARQ-IR and HARQ-CC protocols will be analyzed. In all cases, packets will leave the buffer either after being successfully received by the mobile host or after N_r+1 failed attempts. Moreover, as it has been previously stated in Subsection 2.1.2, a Rayleigh block-fading channel model has been adopted (Biglieri et al., 2001) and it is assumed that the frame duration T_f is much smaller than the coherence time of the channel. In this case, although the channel is assumed to remain invariant over at least one time frame interval while it is allowed to vary across successive frame intervals, it is also highly probable that it will remain practically invariant over a large number of successive frame intervals.

The implementation of the HARQ-IR and HARQ-CC protocols require the definition of different puncturing patterns corresponding to successive (re)transmissions. Thus, without loss of generality, as the IEEE 802.16e standard (IEEE, 2004) allows these strategies, convolutionally coded M-QAM TMs adopted from this standard will be used in the AMC pool. Unlike the 802.11a AMC pool described in Table 3.1, which was taken from (Liu et al., 2004), the fitting parameters for the considered 802.16e TMs (except the non-transmission mode), which are listed in Table 4.1, will be derived in this Subsection. In this case, when using TM $n \in \boldsymbol{\mathcal{M}}$, a rate-1/2 convolutional encoder generates a sequence $\tilde{\boldsymbol{b}} = \{\tilde{b}_l\}_{l=1}^{2N_b}$ of encoded bits and, after puncturing, the system transmits $N_c^{(n)} = N_b/R_c^{(n)}$ coded bits per packet, where $R_c^{(n)}$ denotes the code rate obtained after puncturing when using TM n.

The different puncturing patterns for the HARQ-IR scheme in the IEEE 802.16e standard (IEEE, 2006) are shown in Table 4.2, where the puncturing

Table 4.2: 802.16e puncturing pattern definition for HARQ-IR.

		Code Rate			
		1/2	2/3	3/4	5/6
SPID = 0	X	1	10	101	10101
	Y	1	11	110	11010
SPID = 1	X	1	01	011	01011
	Y	1	11	101	10101
SPID = 2	X	1	10	110	10110
	Y	1	11	011	01011
SPID = 3	X	1	01	101	01101
	Y	1	11	110	10110

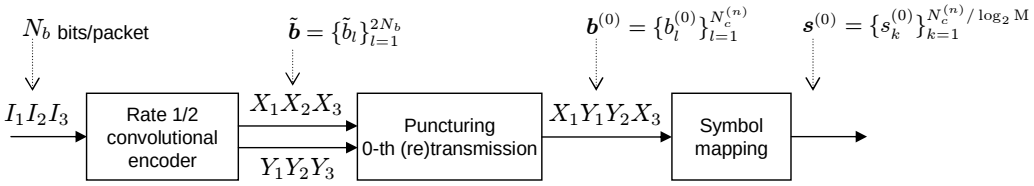


Figure 4.1: Example of the puncturing process for the first transmission $i = 0$ and $R_c^{(n)} = 3/4$.

pattern corresponding to subpacket identifier (SPID) $i + 1$ is obtained as the left cyclic shift of the puncturing pattern corresponding to SPID i. A 0 in the puncturing pattern means that the corresponding output bit of the fixed rate convolutional encoder will not be transmitted. Figure 4.1 illustrates an example of the puncturing process corresponding to the first transmission (SPID=0) and a code rate $R_c^{(n)} = 3/4$.

In type-I hybrid FEC/ARQ and in HARQ schemes based on the CC strategy the same puncturing pattern is applied for each (re)transmission (without loss of generality, the puncturing pattern corresponding to SPID=0 is applied to all (re)transmissions). In contrast, when using the IR strategy, a different puncturing pattern can be applied to successive (re)transmissions (Cheng, 2006). As it can be observed in Fig. 4.1, for a generic packet, the sequence of punctured coded bits corresponding to (re)transmission $i \in \{0, \ldots, N_r\}$ can be denoted as $\boldsymbol{b}^{(i)} = \{b_l^{(i)}\}_{l=1}^{N_c^{(n)}}$. These punctured coded bits are mapped onto a sequence of symbols $\boldsymbol{s}^{(i)} = \{s_k^{(i)}\}_{k=1}^{N_c^{(n)}/\log_2 \mathrm{M}}$ that are selected from the M-QAM constellation corresponding to TM n.

The received signal corresponding to transmitted symbol $s_k^{(i)}$, $k = 1, \ldots, \frac{N_c^{(n)}}{\log_2 \mathrm{M}}$, can be expressed as

$$r_k^{(i)} = s_k^{(i)} h_i + n_{i,k}, \tag{4.1}$$

where, by convenient abuse of notation, h_i denotes the channel gain of the frame period corresponding to the ith (re)transmission and $n_{i,k}$ is a zero-mean circularly symmetric complex Gaussian noise with variance $N_0/2$ per dimension. The instantaneous received SNR during the νth frame transmission is defined as $\gamma_\nu = E_s|h_\nu|^2/N_0$. The average received SNR is then given by $\overline{\gamma} = E\{\gamma_\nu\} = E_s/N_0$, where E_s denotes the average power of the received signal. As it has been previously seen in Chapter 2, when implementing the AMC strategy, the entire SNR

range is partitioned into a set of non-overlapping intervals defined by the partition $\mathbf{\Gamma}^m$ given in (2.1), and mode n is selected when $\gamma_\nu \in [\gamma_n^m, \gamma_{n+1}^m)$.

Chase combining

Assuming perfect CSI to be available at the receiver side, the HARQ-CC scheme combines multiple received signals according to the MRC principle. Thus, assuming without loss of generality that the 0th puncturing pattern has been used in all (re)transmissions, the combined signal after i (re)transmissions can be expressed as

$$\hat{r}_k^{(i)} = \sum_{j=0}^{i} r_k^{(j)} h_j^* = s_k^{(0)} \rho_i + \upsilon_k^{(i)}, \tag{4.2}$$

where $\rho_i = \sum_{j=0}^{i} |h_j|^2$ and $\upsilon_k^{(i)}$ is a zero-mean circularly symmetric complex Gaussian noise with variance $\sigma_{\upsilon,i}^2 = \rho_i N_0$. The logarithmic likelihood ratio (LLR) corresponding to bit $b_l^{(0)}$ mapped onto symbol $s_k^{(0)}$ on the ith (re)transmission can be expressed as

$$\begin{aligned} \lambda_l^{(i)} &= \log \frac{Pr\{b_l^{(0)} = 1 | \hat{r}_k^{(i)}, \rho_i\}}{Pr\{b_l^{(0)} = 0 | \hat{r}_k^{(i)}, \rho_i\}} \\ &= \log \frac{\sum_{\hat{s}_k \in S_{l,i}^{(1)}} \exp\left(-|\hat{r}_k^{(i)} - \hat{s}_k \rho_i|^2 / \sigma_{\upsilon,i}^2\right)}{\sum_{\hat{s}_k \in S_{l,i}^{(0)}} \exp\left(-|\hat{r}_k^{(i)} - \hat{s}_k \rho_i|^2 / \sigma_{\upsilon,i}^2\right)}, \end{aligned} \tag{4.3}$$

where $S_{l,i}^{(0)}$ and $S_{l,i}^{(1)}$ are, respectively, the sets of symbols $\hat{s}_k$ with the bit indexed by l, corresponding to the ith (re)transmission, equal to zero or one. These LLRs are subsequently depunctured to obtain the sequence $\tilde{\boldsymbol{\lambda}}^{(i)} = \{\tilde{\lambda}_l^{(i)}\}_{l=1}^{2N_b}$ that is then passed to the soft Viterbi decoder.

Incremental redundancy

For the HARQ-IR scheme, the LLR corresponding to bit $b_l^{(i)}$ mapped onto symbol $s_k^{(i)}$ on the ith (re)transmission is given by

$$\begin{aligned} \lambda_l^{(i)} &= \log \frac{Pr\{b_l^{(i)} = 1 | r_k^{(i)}, h_i\}}{Pr\{b_l^{(i)} = 0 | r_k^{(i)}, h_i\}} \\ &= \log \frac{\sum_{\hat{s}_k \in S_{l,i}^{(1)}} \exp\left(-|r_k^{(i)} - \hat{s}_k h_i|^2 / N_0\right)}{\sum_{\hat{s}_k \in S_{l,i}^{(0)}} \exp\left(-|r_k^{(i)} - \hat{s}_k h_i|^2 / N_0\right)}. \end{aligned} \tag{4.4}$$

After suitable depuncturing, and assuming an add-LLR soft-output packet combining scheme (Aïssa and Aniba, 2008), the sequence of combined LLRs is obtained as

$$\tilde{\boldsymbol{\lambda}}^{(i)} = \left\{\sum_{j=0}^{i} \tilde{\lambda}_l^{(j)}\right\}_{l=1}^{2N_b}, \tag{4.5}$$

and it is then sent to the soft Viterbi decoder.

Type-I hybrid FEC/ARQ

In the Type-I hybrid FEC/ARQ scheme, when an uncorrectable pattern is detected, the received packet is discarded and a retransmission is requested. Therefore, this strategy is equivalent to the HARQ-IR scheme with no addition of LLRs after depuncturing and assuming that the 0th puncturing pattern is used in all (re)transmissions. Consequently, the LLRs at the input of the soft Viterbi decoder can be expressed as $\tilde{\boldsymbol{\lambda}} = \{\tilde{\lambda}_l^{(0)}\}_{l=1}^{2N_b}$.

Modeling the packet error rate behaviour

When the Type-I hybrid FEC/ARQ scheme is adopted, the PHY layer instantaneous PER at the output of the soft Viterbi decoder will depend on the selected TM and on the instantaneous received SNR γ. Thus, the instantaneous PER can be obtained as the ratio between the erroneously transmitted packets using TM n and the overall number of transmitted packets. When using an HARQ scheme, as all the (re)transmissions are utilized to decode the packet, the instantaneous PER will also depend on the (re)transmission number. Consequently, the instantaneous PER can be obtained as the ratio between the erroneously transmitted packets after (re)transmission number i using TM n and the overall number of transmitted packets (obviously, only those packets that have been erroneously transmitted in (re)transmission numbers $0, \ldots, i-1$ will also be transmitted in the ith retransmission). The instantaneous PER curves obtained through Monte-Carlo simulation with a packet length of $N_b = 1080$ bits are shown in Fig. 4.2.

With the assumption of a slow block-fading channel model, the instantaneous PER at the output of the soft Viterbi decoder when TM n is selected under the Type-I hybrid FEC/ARQ scheme, for all the (re)transmissions of a packet, can be approximated as

$$\mathrm{PER}_n(\gamma) \approx \begin{cases} 1 & ,0 \leq \gamma < \gamma_{p_{n,0}} \\ a_{n,0} e^{-g_{n,0}\gamma} & ,\gamma \geq \gamma_{p_{n,0}} \end{cases} \tag{4.6}$$

When using more sophisticated HARQ protocols, the instantaneous PER corresponding to the nth TM over the ith (re)transmission can be safely approximated as

$$\mathrm{PER}_{n,i}(\gamma) \approx \begin{cases} 1 & ,0 \leq \gamma < \frac{\gamma_{p_{n,0}}}{i+1} \\ a_{n,0} e^{-g_{n,0}(i+1)\gamma} & ,\gamma \geq \frac{\gamma_{p_{n,0}}}{i+1} \end{cases} \tag{4.7}$$

for the HARQ-CC strategy, and

$$\mathrm{PER}_{n,i}(\gamma) \approx \begin{cases} 1 & ,0 \leq \gamma < \gamma_{p_{n,i}} \\ a_{n,i} e^{-g_{n,i}\gamma} & ,\gamma \geq \gamma_{p_{n,i}} \end{cases} \tag{4.8}$$

for the HARQ-IR strategy. The parameters $a_{n,i}$, $g_{n,i}$ and $\gamma_{p_{n,i}}$, listed in Table 4.1, are the fitting parameters for TM n and (re)transmission number i, which have been obtained by least-squares fitting the above approximate expressions for the PER to the curves obtained through Monte-Carlo simulation. It is clear from Fig. 4.2 that both curves are close to each other, which justifies the utilization of the exponential expression as an approximation of the instantaneous PER.

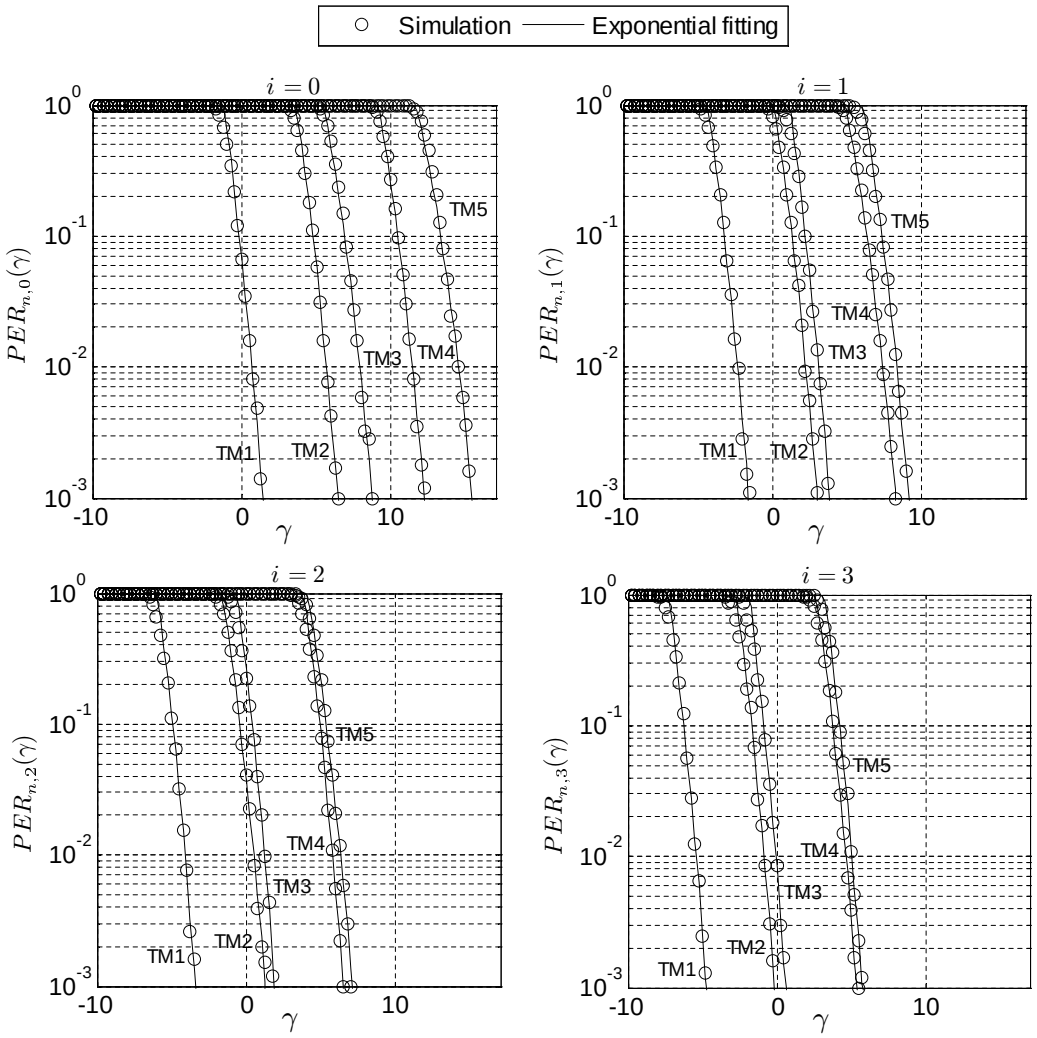


Figure 4.2: Instantaneous PER fitting curves.

Figure 4.3 depicts the PER curves corresponding to transmission modes TM1, TM3 and TM5 for (re)transmissions $i = 0, 1, 2, 3$ when both, HARQ-CC and HARQ-IR strategies are considered. It can be observed that the curves are the same for both techniques in the first transmission, whereas a better performance is obtained with the HARQ-IR strategy for successive retransmissions. It can be clearly appreciated that the enhancement in the achieved coding gain when the HARQ-IR technique is selected, increases for the higher order TMs. The reason for this behaviour is that, traditionally, the effects of the HARQ-CC scheme are modeled as allowing the SNRs of the involved packets to be added together. The advantage of the HARQ-IR scheme can then be modeled as providing additional coding gains on top of the combined SNR (Cheng, 2006). Thus, as lower order TMs are more robust, they do not benefit from the additional coding gains provided by the HARQ-IR scheme.

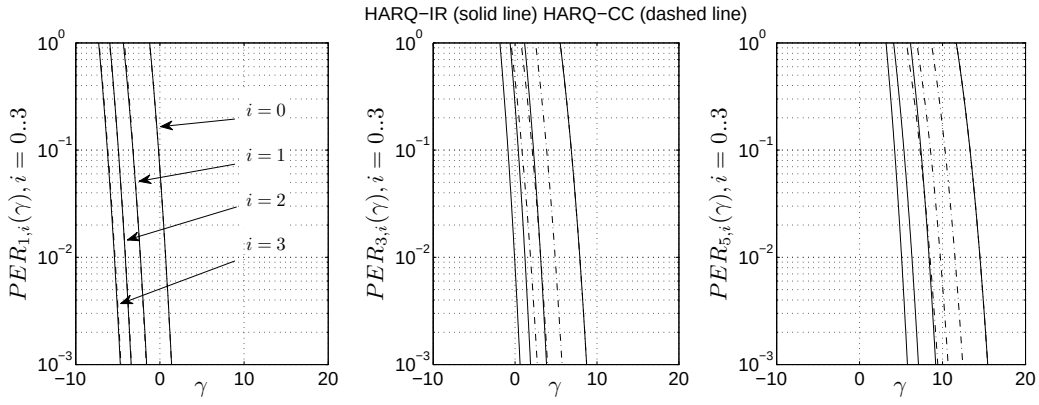


Figure 4.3: Instantaneous PER curves comparison for HARQ-CC and HARQ-IR for different TMs.

4.2.2 Physical layer two-dimensional Markov model

The procedure for the development of the Markov chain-based PHY layer model follows the same methodology proposed in Chapter 3, where an infinitely persistent Type-I hybrid FEC/ARQ protocol was analyzed. That is, it takes into consideration the modeling of the channel and the implementation of the AMC scheme. The channel model reproduces the amplitude and the rate-of-change of the fading envelope, characterized by the partitions $\mathbf{\Gamma}^c$ and $\mathbf{\Delta}$ defined in (2.6) and (2.8), respectively, and the AMC scheme is characterized by the partition $\mathbf{\Gamma}^m$ defined in (2.1). Therefore, the range of γ_ν is partitioned into the set of N_{PHY} non-overlapping intervals defined by the partition (3.4). Furthermore, it also considers the partition of $\delta_\nu = \gamma_{\nu-1} - \gamma_\nu$ into the set of non-overlapping intervals $\mathbf{\Delta} = \{(-\infty, 0), [0, \infty)\}$. Using this two-dimensional partitioning, a first-order two-dimensional Markov model for the PHY layer is defined where each state corresponds to one of such two-dimensional rectangular-shaped cells.

The PHY layer Markov chain state at time instant $t = \nu T_f$ can be denoted as $\boldsymbol{\varsigma}_\nu = (\varphi_\nu, \Delta_\nu)$, $\nu = 0, 1, \ldots, \infty$, where $\varphi_\nu \in \{0, \ldots, N_{\text{PHY}} - 1\}$ represents the combination of TM and channel state in this frame interval and $\Delta_\nu \in \{0, 1\}$ is used to denote the *up* or *down* characteristic of the instantaneous SNR in time frame interval $t = (\nu - 1)T_f$. Equivalently, the PHY layer state $\boldsymbol{\varsigma}_\nu$ can be univocally characterized by an integer number $n_\nu = 2\varphi_\nu + \Delta_\nu$ and obviously, $n_\nu \in \{0, \ldots, 2N_{\text{PHY}} - 1\}$.

The PHY layer will be in a state $n \in \{0, \ldots, 2N_{\text{PHY}} - 1\}$ with probability $P^{\text{PHY}}(n)$, which can be calculated using eqs. (3.6) and (2.9). Furthermore, the PHY-layer FSMC will be characterized by a transition probability matrix

$$\boldsymbol{P}_s = \left[P_s(n_\mu, n_{\mu'})\right]_{n_\mu, n_{\mu'}=0}^{2N_{\text{PHY}}-1}, \tag{4.9}$$

whose elements can be calculated using eqs. (3.9)-(3.12) and (2.16).

When the Type-I hybrid FEC/ARQ scheme is used, each of these states will be characterized by an average PER that can be calculated as in (3.7). When using more sophisticated HARQ protocols, these PHY layer states will be characterized

by a conditional average PER for the ith (re)transmission given by

$$\overline{PER}_{n,i}^{\text{PHY}} = \begin{cases} \dfrac{\int_{\gamma_{n/2}^{m,c}}^{\gamma_{n/2+1}^{m,c}} \int_0^x PER_{\beta_n,i|i-1}(x) p_{\gamma_\nu,\gamma_{\nu-1}}(x,y) dy dx}{P^{\text{PHY}}(n)}, & n \text{ even} \\ \dfrac{\int_{\gamma_{(n-1)/2}^{m,c}}^{\gamma_{(n+1)/2}^{m,c}} \int_x^{+\infty} PER_{\beta_n,i|i-1}(x) p_{\gamma_\nu,\gamma_{\nu-1}}(x,y) dy dx}{P^{\text{PHY}}(n)}, & n \text{ odd} \end{cases} \tag{4.10}$$

where β_n denotes the TM corresponding to the nth PHY layer state, $p_{\gamma_\nu,\gamma_{\nu-1}}(x,y)$ is the joint pdf of the random variables γ_ν and $\gamma_{\nu-1}$, and $PER_{n,i|i-1}(\cdot)$ represents the probability that the ith (re)transmission attempt fails conditioned on the previous transmissions being erroneous, that is,

$$PER_{n,i|i-1}(\gamma) = \frac{PER_{n,i}(\gamma)}{PER_{n,i-1}(\gamma)}. \tag{4.11}$$

4.3 Discrete time Markov chain-based link-level queueing model and analysis

4.3.1 Embedded Markov chain

As in Chapter 3, the queueing process induced by both the AMC scheme and the (H)ARQ protocol can be formulated in discrete time with one time unit equal to one frame interval. The system states are observed at the beginning of each time unit.

Let $\sigma_\nu = (\mathbf{q}_\nu, a_\nu, n_\nu)$ denote the system state at time instant $t = \nu T_f$, where $a_\nu \in \{0, \ldots, \mathcal{A}-1\}$ represents the phase of the D-BMAP and $n_\nu \in \{0, \ldots, 2N_{\text{PHY}}-1\}$ corresponds to the PHY layer state. Furthermore, $\mathbf{q}_\nu = (q_{\nu,0}, \ldots, q_{\nu,N_r})$ is the queue state at this time instant, with $q_{\nu,i}$ denoting the number of packets in the queue that have already been transmitted i times and $Q_\nu \triangleq \sum_{i=0}^{N_r} q_{\nu,i} \in \{0, \ldots, \overline{Q}\}$ denoting the total number of packets in the queue, as illustrated in Fig. 4.4. The buffer operates in a FIFO mode and, thus, the system will give priority to packets that have already been transmitted more times. Similarly to the infinitely persistent scheme, the underlying queueing process can be characterized through an embedded Markov chain. The state space of this embedded finite state Markov chain is $\boldsymbol{\mathcal{S}} = \{\mathcal{S}_n\}_{n=1}^{N_s}$ with size $N_s = 2N_{\text{PHY}}\mathcal{A}\sum_{k=0}^{\overline{Q}} \binom{k+N_r}{k}$.

The transition from state $\mathcal{S}_\mu = (\mathbf{q}_\mu, a_\mu, n_\mu)$ to state $\mathcal{S}_{\mu'} = (\mathbf{q}_{\mu'}, a_{\mu'}, n_{\mu'})$ is illustrated in Fig. 4.5. The queue state $\mathbf{q}_\nu = (q_{\nu,0}, \ldots, q_{\nu,N_r})$ contains the number of packets in the queue that have never been transmitted $q_{\nu,0}$, the number of packets in the queue that have already been transmitted one time $q_{\nu,1}$, an so on up to the number of packets in the queue that have already been transmitted N_r times q_{ν,N_r}, which is the maximum number of allowed retransmissions. Assuming that there are Q_μ packets in the queue at the beginning of the first frame interval and, given that the number of packets that can be transmitted when the PHY layer is in state n_μ is $\mathcal{C}_{n_\mu}$, the number of packets that will be actually transmitted is $\min\{\mathcal{C}_{n_\mu}, Q_\mu\}$ (blue arrows in Fig. 4.5). Transmission will start with packets that have been transmitted N_r times. Obviously, these transmitted packets will leave the queue either due to a successful transmission (green

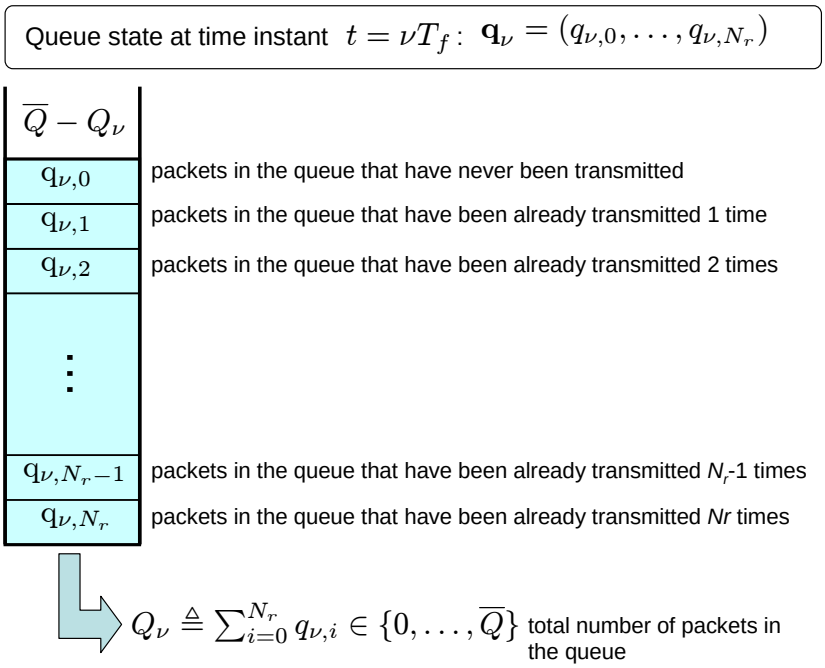


Figure 4.4: Queue state for a truncated (H)ARQ scheme.

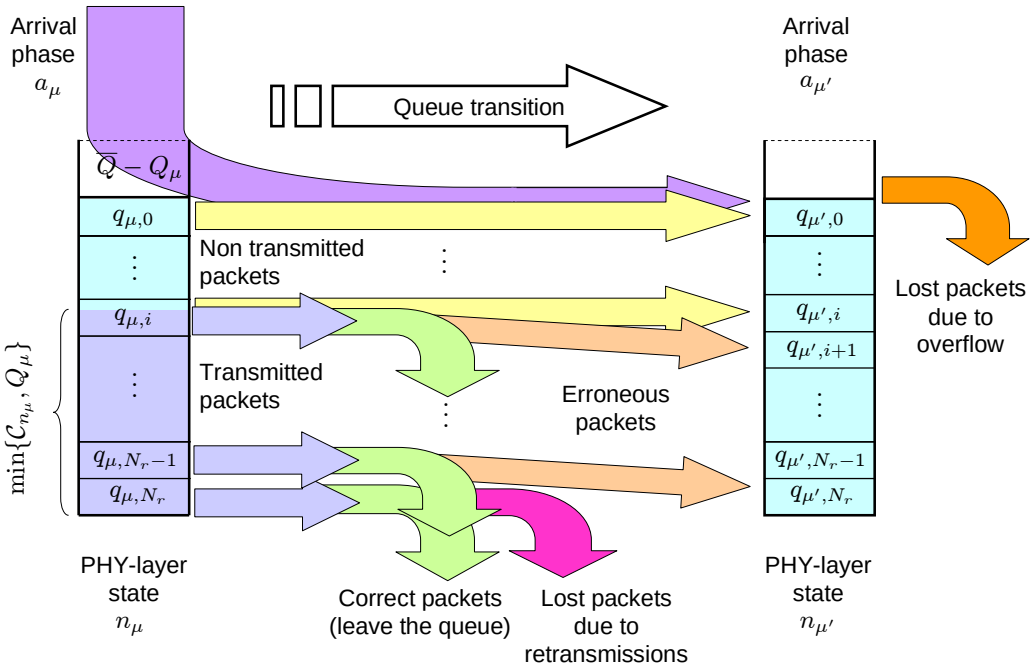


Figure 4.5: Queue transition for a truncated (H)ARQ scheme.

arrow) or, in case of an erroneous transmission, due to exceeding the maximum number of allowed retransmissions (pink arrow). In case that the transmission capacity, $\min\{\mathcal{C}_{n_\mu}, Q_\mu\}$, allows the transmission of all q_{μ,N_r} packets and more packets can be also transmitted, the next selected packets for transmission will be chosen among the packets that have already been transmitted $N_r - 1$ times. Those among them that are correctly transmitted will leave the queue (green arrow), while those that are erroneously transmitted will contribute to form, in the next time interval, the set of packets that have already been transmitted N_r times (camel arrow). An analogous behaviour is observed for the packets that have been transmitted $N_r - 2$ times, and so on until the $\min\{\mathcal{C}_{n_\mu}, Q_\mu\}$ packets have been transmitted. Obviously, the $Q_\mu - \min\{\mathcal{C}_{n_\mu}, Q_\mu\}$ packets that are not transmitted will also remain in the queue (yellow arrow). Moreover, the newly arrived packets a_μ (purple arrow) that can be stored in the buffer, will be included in the set of packets that have never been transmitted $q_{\mu',0}$. In case there is not enough room in the queue for all the new packets, those that will be dropped (orange arrow) will increase the packet loss rate due to buffer overflow.

The transition from state $\mathcal{S}_\mu = (\mathbf{q}_\mu, a_\mu, n_\mu)$ to state $\mathcal{S}_{\mu'} = (\mathbf{q}_{\mu'}, a_{\mu'}, n_{\mu'})$ has a probability that can be written as

$$P_{\mathcal{S}_\mu,\mathcal{S}_{\mu'}} = u\left(a_\mu, a_{\mu'}\right) P_s\left(n_\mu, n_{\mu'}\right) P_{\mathbf{q}_\mu,\mathbf{q}_{\mu'}|a_\mu,n_\mu}, \tag{4.12}$$

where $u\left(a_\mu, a_{\mu'}\right)$ denotes the transition probability between D-BMAP phases a_μ and $a_{\mu'}$, which can be obtained from matrix $\boldsymbol{U}$, $P_s\left(n_\mu, n_{\mu'}\right)$ is the PHY layer transition probability between states n_μ and $n_{\mu'}$, which can be derived from matrix $\boldsymbol{P}_s$, and $P_{\mathbf{q}_\mu,\mathbf{q}_{\mu'}|a_\mu,n_\mu} = \prod_{i=0}^{N_r} P_{q_{\mu,i},q_{\mu',i}|a_\mu,n_\mu}$ corresponds to the queue transition probability from state $\mathbf{q}_\mu$ to state $\mathbf{q}_{\mu'}$ when the D-BMAP is in phase a_μ (purple arrow) and the PHY layer state is n_μ.

Let us consider that the system state is $\mathcal{S}_\mu$ and that $Q_{\mu,i} = \sum_{l=i}^{N_r} q_{\mu,l}$ represents the number of packets in the queue that have already been transmitted i or more times (obviously, $Q_{\mu,0} = Q_\mu$). Additionally, let $\tau_\mu = \min\{Q_\mu, \mathcal{C}_{n_\mu}\}$ denote the total number of transmitted packets (blue arrows), $\tau_{\mu,i} = \min\{q_{\mu,i}, \tau_\mu - Q_{\mu,i+1}\}$ the number of transmitted packets among those in the queue that have already been transmitted i times and $\epsilon_{\mu,i}$ the number of packets erroneously transmitted among those in the queue that have already been transmitted i times (camel and pink arrows). Using this notation, the feasible queue transitions are only those satisfying

$$q_{\mu',N_r} = \begin{cases} q_{\mu,N_r} - \tau_\mu & , \tau_\mu \leq q_{\mu,N_r} \\ \epsilon_{\mu,N_r-1} & , \tau_\mu > q_{\mu,N_r} \end{cases} \tag{4.13}$$

$$q_{\mu',i} = \begin{cases} q_{\mu,i} & , \tau_\mu \leq Q_{\mu,i+1} \\ Q_{\mu,i} - \tau_\mu & , Q_{\mu,i+1} < \tau_\mu \leq Q_{\mu,i} \\ \epsilon_{\mu,i-1} & , \tau_\mu > Q_{\mu,i} \end{cases} \tag{4.14}$$

for $i \in \{1, \ldots, N_r - 1\}$, and

$$q_{\mu',0} = \begin{cases} \min\left\{\overline{Q} - Q_{\mu',1}, q_{\mu,0} + a_\mu\right\} & , \tau_\mu \leq Q_{\mu,1} \\ \min\left\{\overline{Q} - Q_{\mu',1}, Q_\mu + a_\mu - \tau_\mu\right\} & , Q_{\mu,1} < \tau_\mu \leq Q_\mu. \end{cases} \tag{4.15}$$

Consequently, by defining

$$\mathcal{P}_y^x(z) \triangleq \binom{x}{y} z^y (1-z)^{x-y}, \tag{4.16}$$

and assuming a slow block-fading channel, the queue transition probabilities, given the arrival phase is a_μ (purple arrow) and the PHY layer state is n_μ, can be safely approximated as

$$P_{q_{\mu,N_r},q_{\mu',N_r}|a_\mu,n_\mu} = \begin{cases} 1 & , \begin{array}{l} q_{\mu',N_r} = q_{\mu,N_r} - \tau_\mu \\ \tau_\mu \leq q_{\mu,N_r} \end{array} \\ \mathcal{P}_{q_{\mu',N_r}}^{\tau_{\mu,N_r-1}}\left(\overline{PER}_{n_\mu,N_r-1}^{\mathrm{PHY}}\right) & , \begin{array}{l} \tau_\mu > q_{\mu,N_r} \\ \tau_{\mu,N_r-1} \geq q_{\mu',N_r} \end{array} \\ 0 & , \text{ otherwise} \end{cases} \tag{4.17}$$

$$P_{q_{\mu,i},q_{\mu',i}|a_\mu,n_\mu} = \begin{cases} 1 & , \begin{array}{l} q_{\mu,i} = q_{\mu',i} \\ \tau_\mu \leq Q_{\mu,i+1} \end{array} \\ 1 & , \begin{array}{l} Q_{\mu,i+1} < \tau_\mu \leq Q_{\mu,i} \\ q_{\mu',i} = Q_{\mu,i} - \tau_\mu \end{array} \\ \mathcal{P}_{q_{\mu',i}}^{\tau_{\mu,i-1}}\left(\overline{PER}_{n_\mu,i-1}^{\mathrm{PHY}}\right) & , \begin{array}{l} \tau_\mu > Q_{\mu,i} \\ \tau_{\mu,i-1} \geq q_{\mu',i} \end{array} \\ 0 & , \text{ otherwise} \end{cases} \tag{4.18}$$

for $i \in \{1, \ldots, N_r - 1\}$, and

$$P_{q_{\mu,0},q_{\mu',0}|a_\mu,n_\mu} = \begin{cases} 1 & , \begin{array}{l} \tau_\mu \leq Q_{\mu,1} \\ q_{\mu',0} = \min\left\{\overline{Q} - Q_{\mu',1}, q_{\mu,0} + a_\mu\right\} \end{array} \\ 1 & , \begin{array}{l} Q_{\mu,1} < \tau_\mu \leq Q_\mu \\ q_{\mu',0} = \min\left\{\overline{Q} - Q_{\mu',1}, Q_\mu + a_\mu - \tau_\mu\right\} \end{array} \\ 0 & , \text{ otherwise.} \end{cases} \tag{4.19}$$

Notice that for the Type-I hybrid FEC/ARQ strategy, as the conditional average PER only depends on the PHY layer state n_μ, $\overline{PER}_{n_\mu,i}^{\mathrm{PHY}} = \overline{PER}_{n_\mu}^{\mathrm{PHY}}$ for all $i \in \{0, \ldots, N_r\}$.

As in the infinitely persistent case, the transition probability matrix $\boldsymbol{P}$ and steady-state probability vector $\boldsymbol{\pi} = \left[\pi_{\mathcal{S}_1} \cdots \pi_{\mathcal{S}_{N_s}}\right]$ satisfy $\boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{\pi}$ along with the normalization condition $\boldsymbol{\pi}\mathbf{1}_{N_s} = 1$. The complexity involved in calculating the steady-state probabilities increases exponentially with the number of phases in the arrival process, the number of states in the PHY layer Markov model and the number of queue states in the system. Consequently, for very large buffer sizes, efficient strategies should be devised to tackle this problem. This is an issue that has been left for further research.

4.3.2 Packet loss rate and throughput

The number of lost packets due to buffer overflow when the system changes from state $\mathcal{S}_\mu$ to state $\mathcal{S}_{\mu'}$ (orange arrow) corresponds to the arriving packets that can not be stored in the queue due to lack of space. The available space in the queue at the end of this frame interval for the packets that have never been transmitted can be obtained as the difference between the buffer size, $\overline{Q}$, and the number of packets that will have been already transmitted 1 or more times at time instant $t = \mu' T_f$, $Q_{\mu',1}$, that is, $\overline{Q} - Q_{\mu',1}$. This available space will be occupied by the packets that have never been transmitted at the end of this frame interval, which will correspond to $q_{\mu,0}$ if no packets among them are transmitted, and to $q_{\mu,0} - \tau_{\mu,0} = Q_\mu - \tau_\mu$, otherwise. Moreover, the incoming packets a_μ should also be stored in the buffer. Thus, the number of packets that overflow from the buffer can be obtained as

$$N_{l_{BO}|\mathcal{S}_\mu,\mathcal{S}_{\mu'}} = \begin{cases} \max\{0, Q_{\mu',1} + q_{\mu,0} + a_\mu - \overline{Q}\} & , \tau_\mu \leq Q_{\mu,1} \\ \max\{0, Q_{\mu',1} + Q_\mu - \tau_\mu + a_\mu - \overline{Q}\} & , \tau_\mu > Q_{\mu,1}. \end{cases} \tag{4.20}$$

Therefore, the average number of lost packets due to buffer overflow can be calculated as

$$\bar{N}_{l_{BO}} = \sum_{\mu=1}^{N_s} \sum_{\mu'=1}^{N_s} \pi_{\mathcal{S}_\mu} P_{\mathcal{S}_\mu,\mathcal{S}_{\mu'}} N_{l_{BO}|\mathcal{S}_\mu,\mathcal{S}_{\mu'}} \tag{4.21}$$

and the packet loss rate $P_{l_{BO}}$ (measured in packets per frame) can then be obtained as

$$P_{l_{BO}} = \bar{N}_{l_{BO}}/\lambda. \tag{4.22}$$

The number of lost packets due to exceeding the maximum number of allowed retransmissions when the system state is $\mathcal{S}_\mu$ corresponds to those packets that are erroneously transmitted after having been already transmitted N_r times, and can be calculated as

$$N_{l_{\text{ARQ}}|\mathcal{S}_\mu} = \begin{cases} \sum_{l=0}^{\tau_{\mu,N_r}} l \; \mathcal{P}_l^{\tau_{\mu,N_r}} \left(\overline{PER}_{n_\mu}^{\text{PHY}}\right) & \text{(Type-I hybrid FEC/ARQ)} \\ \sum_{l=0}^{\tau_{\mu,N_r}} l \; \mathcal{P}_l^{\tau_{\mu,N_r}} \left(\overline{PER}_{n_\mu,N_r}^{\text{PHY}}\right) & \text{(CC or IR)} \end{cases} \tag{4.23}$$

with $\tau_{\mu,N_r} = \min\{q_{\mu,N_r}, p_n\}$, and the corresponding average number of lost packets can be obtained as

$$\bar{N}_{l_{\text{ARQ}}} = \sum_{\mu=1}^{N_s} \pi_{\mathcal{S}_\mu} N_{l_{\text{ARQ}}|\mathcal{S}_\mu}. \tag{4.24}$$

Accordingly, the probability of packet loss due to exceeding N_r retransmissions can be expressed as

$$P_{l_{\text{ARQ}}} = \bar{N}_{l_{\text{ARQ}}}/\lambda. \tag{4.25}$$

In our finite buffering truncated ARQ-based error control system, the packet loss rate P_l (measured in packets per frame) can be expressed as $P_l = P_{l_{\text{BO}}} + P_{l_{\text{ARQ}}}$, and given the packet loss rate P_l, the average throughput can be calculated as

$$\eta = \lambda(1 - P_l). \tag{4.26}$$

4.3.3 Average queue length and average packet delay

Using the Little's formula (Kleinrock and Lam), and taking into consideration that the effective packet arrival rate is equal to $\lambda(1 - P_{l_{\mathrm{BO}}})$, the average delay for our embedded Markov chain can be calculated as

$$D_l = L_q/\lambda(1 - P_{l_{\mathrm{BO}}}), \tag{4.27}$$

where L_q denotes the average number of packets in the queue, which is given by

$$L_q = \sum_{\mu=1}^{N_s} \pi_{\mathcal{S}_\mu} Q_\mu. \tag{4.28}$$

4.4 Cross-layer optimization

4.4.1 Multidimensional approach

As shown in previous sections, given a maximum afforded queue length $\overline{Q}$, an average SNR $\bar{\gamma}$ and a normalized maximum Doppler frequency $f_d T_f$, performance measures of the system like, for instance, throughput, average packet delay or packet loss rate, are a function of the AMC transmission mode switching levels $\mathbf{\Gamma}^m \in \mathbb{R}_+^{M+1}$, where $\mathbb{R}_+$ denotes the set of non-negative real numbers, and the measured or estimated arrival packet rate $\lambda \in \mathbf{\Theta}$, where $\mathbf{\Theta}$ is the range of feasible arrival rate values. Therefore, if the objective of the link adaptation scheme is to maximize the average throughput when supporting QoS-guaranteed traffic characterized by a maximum packet loss rate $P_{l\max}$ and a maximum average packet delay $D_{l\max}$, the system needs to jointly optimize the selected protocol parameters at the PHY and DLC layers by solving the cross-layer optimization problem

$$\left(\mathbf{\Gamma}^m_{\mathrm{opt}}, \lambda_{\mathrm{opt}}\right) = \arg \max_{\mathbf{\Gamma}^m \in \mathbb{R}_+^{M+1}, \lambda \in \mathbf{\Theta}} \eta\left(\mathbf{\Gamma}^m, \lambda\right) \tag{4.29}$$

subject to the constraints

$$P_l\left(\mathbf{\Gamma}^m_{\mathrm{opt}}, \lambda_{\mathrm{opt}}\right) \leq P_{l\max}, \quad D_l\left(\mathbf{\Gamma}^m_{\mathrm{opt}}, \lambda_{\mathrm{opt}}\right) \leq D_{l\max}. \tag{4.30}$$

The analytical expressions for η, P_l and D_l do not leave much room for developing efficient algorithms to solve our constrained optimization problem. However, considering that $\mathbf{\Gamma}^m$ and λ lie in a bounded space $\mathbb{R}_+^{M+1} \times \mathbf{\Theta}$, we could resort to a multidimensional exhaustive search to numerically solve the proposed cross-layer optimization problem.

4.4.2 Bidimensional simplification approach

As it has been previously studied in Subsection 3.4.2, two possibilities can be considered for this simplification: the *ergodic* option and the *instantaneous* option. As shown in Section 3.5.2, a nearly identical behaviour of the optimum system performance is observed for both options (only a slight enhancement is obtained

with the *instantaneous* approach). For that reason, from this point onwards, only the *instantaneous* option will be considered in the analysis.

Thus, in order to simplify the multidimensional optimization approach, let us define the AMC switching thresholds as the instantaneous SNR values for which the value of $PER_{n,N_r} = P_0$, that is,

$$\gamma_n^m = \begin{cases} \frac{1}{g_{n,0}} ln\left(\frac{a_{n,0}}{P_0}\right) & \text{(Type-I hybrid FEC/ARQ)} \\ \frac{1}{(N_r+1)g_{n,0}} ln\left(\frac{a_{n,0}}{P_0}\right) & \text{(CC)} \\ \frac{1}{g_{n,N_r}} ln\left(\frac{a_{n,N_r}}{P_0}\right) & \text{(IR)} \end{cases} \tag{4.31}$$

for $n = 1, \ldots, M-1$, with $\gamma_0^m = 0$, and $\gamma_M^m = \infty$. In this case, the simplified constrained optimization problem can be formulated as

$$(P_{0\text{opt}}, \lambda_{\text{opt}}) = \arg \max_{P_0 \in \mathbb{R}_+, \lambda \in \mathbf{\Theta}} \eta\left(P_0, \lambda\right) \tag{4.32}$$

subject to the constraints

$$P_l\left(P_{0\text{opt}}, \lambda_{\text{opt}}\right) \leq P_{l\max}, \quad D_l\left(P_{0\text{opt}}, \lambda_{\text{opt}}\right) \leq D_{l\max}. \tag{4.33}$$

Obviously, because P_0 and λ lie in a bounded space $\mathbb{R}_+ \times \mathbf{\Theta}$, we can resort to a 2-D exhaustive search to numerically solve the proposed cross-layer optimization problem. The bidimensional optimization approach has less degrees of freedom than the multidimensional strategy, resulting in a simpler scheme at the cost of a reduction in performance, as it will be illustrated in the following section.

Obviously, IR requires larger buffer size than CC at the receiver side. Nevertheless, although the CC scheme can improve link performance by coherently combining multiple copies of the received signal, the IR strategy can also benefit from the combination of the different coded versions, which jointly form a lower-rate code with stronger error protection capabilities (Cheng, 2006).

The TM selection probabilities when considering both the HARQ-CC and the HARQ-IR schemes, are shown in Figs. 4.6 and 4.7, respectively. These graphs reveal that the probability of choosing higher order TMs increases with each retransmission. This is because, as it has been previously mentioned in Section 4.1, the CC scheme improves link performance by coherently combining multiple copies of the received signal, and the IR strategy benefits from these and also certain coding gains. Thus, as more (re)transmissions have taken place, a less robust TM can be selected. The comparison of the curves corresponding to both HARQ strategies demonstrates that when the HARQ-CC protocol is utilized, the probability of selecting the most robust TMs (TM1, TM2) augments, while the highest order TM (i.e., TM5) is more probably selected when the HARQ-IR is used. This behaviour is in agreement with the fact that, in comparison to the CC scheme, the IR strategy takes advantage of stronger error protection capabilities. As expected, it can be observed that, in general, greater P_0 values entail an increase in the probability of selecting higher order TMs, no matter which HARQ protocol has been considered.

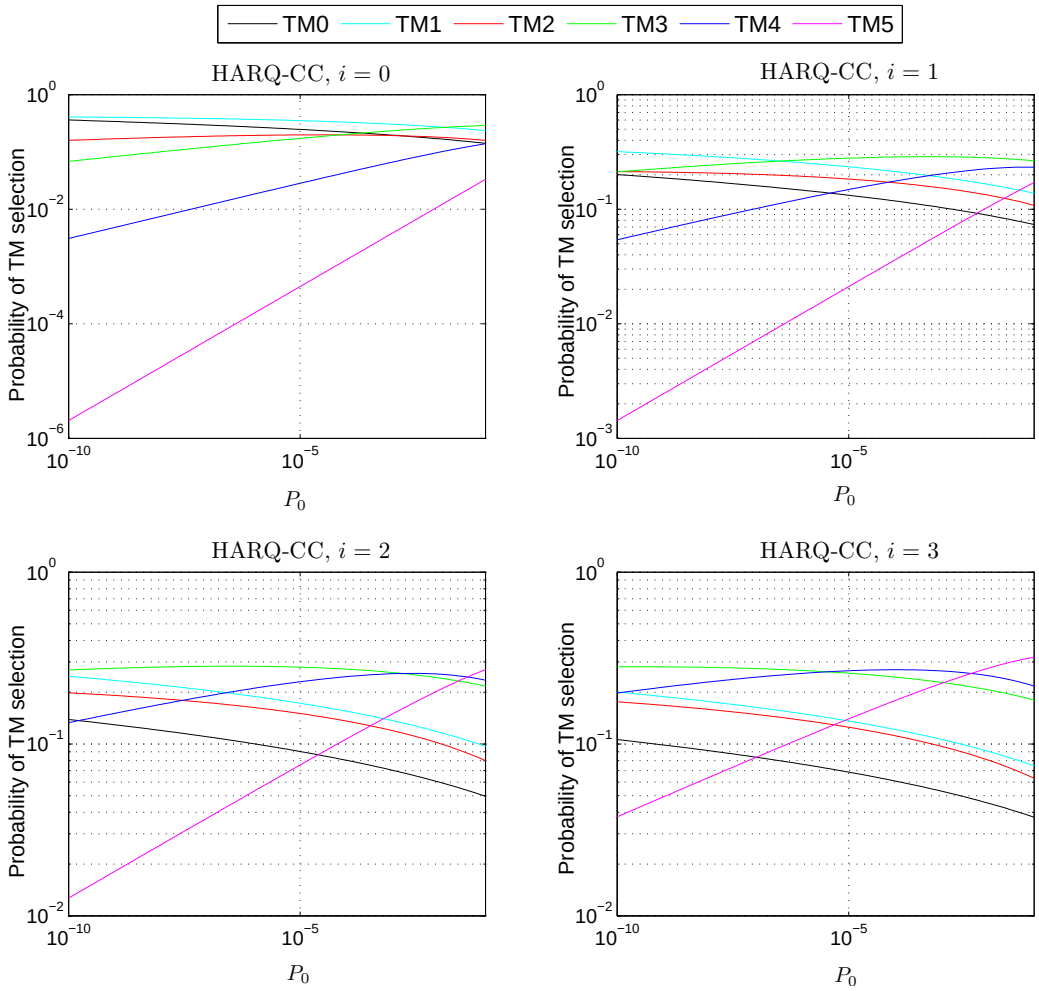


Figure 4.6: TM selection probabilities for the AMC scheme with HARQ-CC.

As in the case of the infinitely persistent ARQ protocol, after having obtained the optimum values of $\mathbf{\Gamma}^m$ (either using the multidimensional or the simplified bidimensional approaches) and λ, these values should be accordingly adjusted with the help of the cross-layer controller at the PHY and DLC layers, respectively.

4.5 Numerical results

In order to validate the analytical 2D-FSMC model introduced in Section 4.3 and to analyze the cross-layer design proposal of Section 4.4, this section will be organized as follows:

1. Firstly, the obtained analytical performance expressions for the truncated Type-I hybrid FEC/ARQ protocol will be contrasted with computer simulations obtained using Clarke's statistical Rayleigh fading process to model the wireless flat-fading channel (Clarke, 1968). Results will be compared with

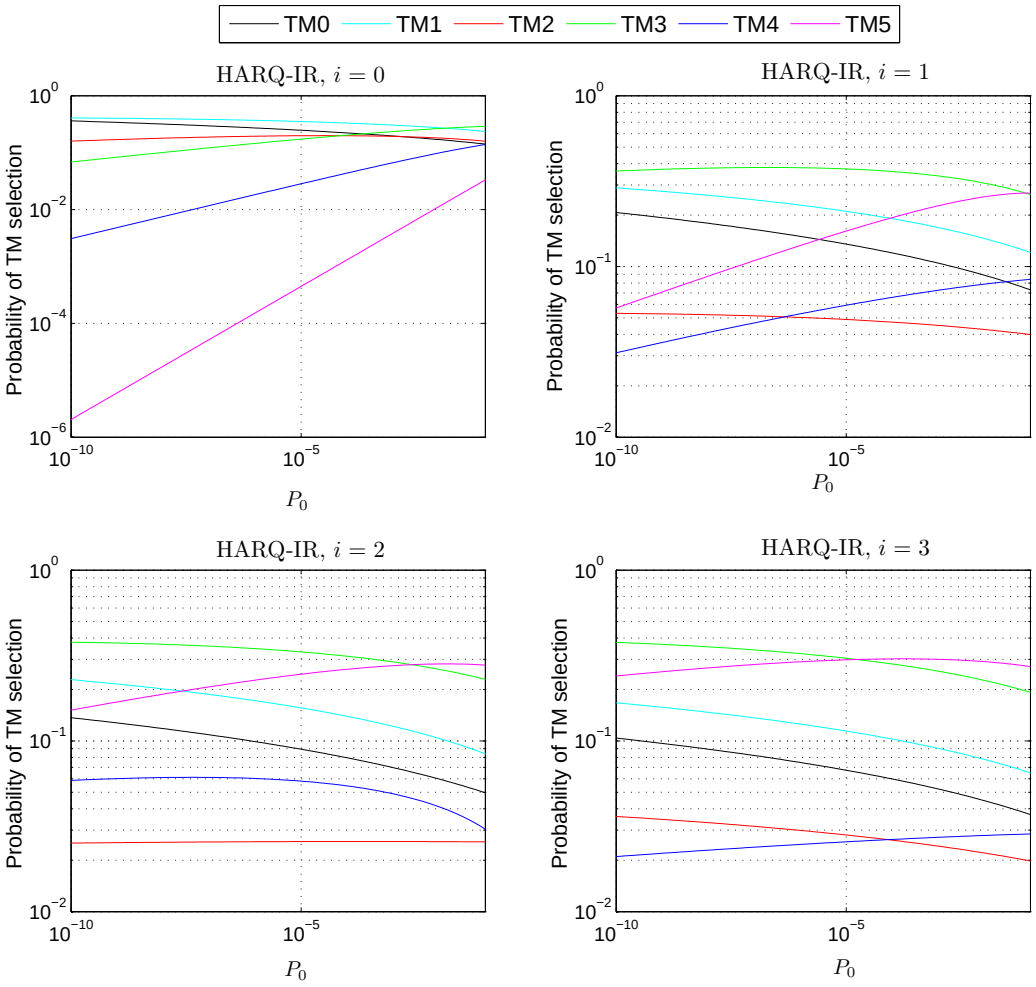


Figure 4.7: TM selection probabilities for the AMC scheme with HARQ-IR.

those corresponding to the infinitely persistent Type-I hybrid FEC/ARQ protocol developed in Chapter 3. Without loss of generality, the 802.11a AMC pool will be initially used for this basic ARQ scheme. It is worth noting that, in the next subsection, the 802.16e TMs will be considered, demonstrating that the proposed analytical Markov-based model usability does not depend on the considered AMC scheme. Moreover, although only the *instantaneous* option has been taken into consideration in the analysis of Subsection 4.4.2, both schemes developed in Subsection 3.4.2 will be directly applied to the truncated Type-I hybrid FEC/ARQ case, with the aim of confirming the behaviour that has been previously observed in Subsection 3.5.2, where the infinitely persistent case was considered.

2. Secondly, the validation of both, the HARQ-IR and the HARQ-CC protocols will be presented. In this case, the 802.16e TMs will be used, as they allow the implementation of the HARQ schemes through the definition of different puncturing patterns corresponding to successive (re)transmissions, as shown

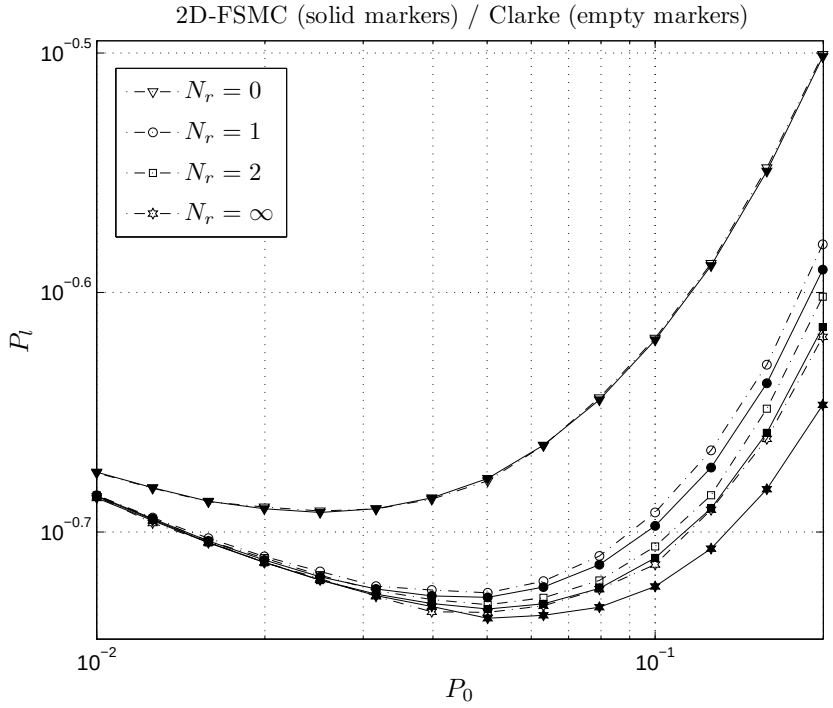


Figure 4.8: Average packet loss rate vs. target PER.

in Subsection 4.2.1. Thus, from this point onwards, this AMC pool will be used.

3. Afterwards, the performance of both schemes, the Type-I hybrid FEC/ARQ and the HARQ, will be compared.

4. Finally, the cross-layer design proposals of Section 4.4 will be analyzed.

4.5.1 Validation and study of the Type-I hybrid FEC/ARQ protocol

In order to verify the validity of the proposed cross-layer framework for the case of a truncated Type-I hybrid FEC/ARQ protocol and for the sake of comparison with the previously developed infinitely persistent scheme, analytical results obtained with the 2D-FSMC model will be confronted with computer simulation results obtained using Clarke's statistical Rayleigh model. The following parameters have been considered: a normalized maximum Doppler frequency $f_d T_f = 0.02$, an average received SNR $\overline{\gamma} = 8$ dB, a buffer size $\overline{Q} = 5$, a number of channel states $K = 10$, a parameter $b = 2$ and a D-BMAP characterized with the transition probability matrix (2.4). Note that, without loss of generality, the 802.11a AMC pool will be used and the *ergodic* option will be considered.

Figure 4.8 plots the dependence of the average packet loss rate P_l on the target average PER P_0, while Figures 4.9 and 4.10 show the dependence of its two components, $P_{l_{BO}}$ and $P_{l_{ARQ}}$, on P_0. Comparing these graphics it is clear that, for the P_0 values of interest, the packet loss due to buffer overflow $P_{l_{BO}}$ is clearly the

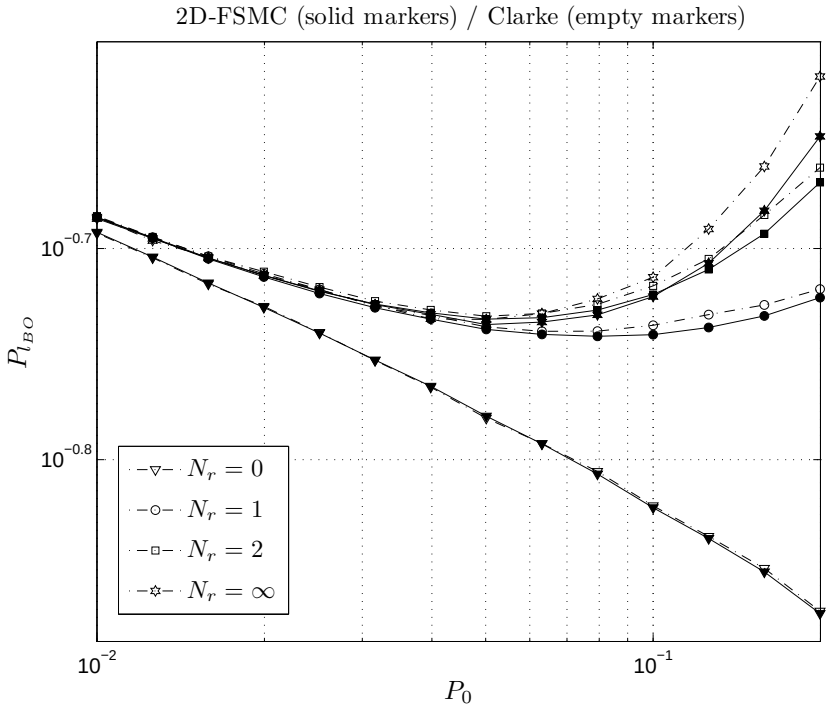


Figure 4.9: Average packet loss rate due to buffer overflow vs. target PER.

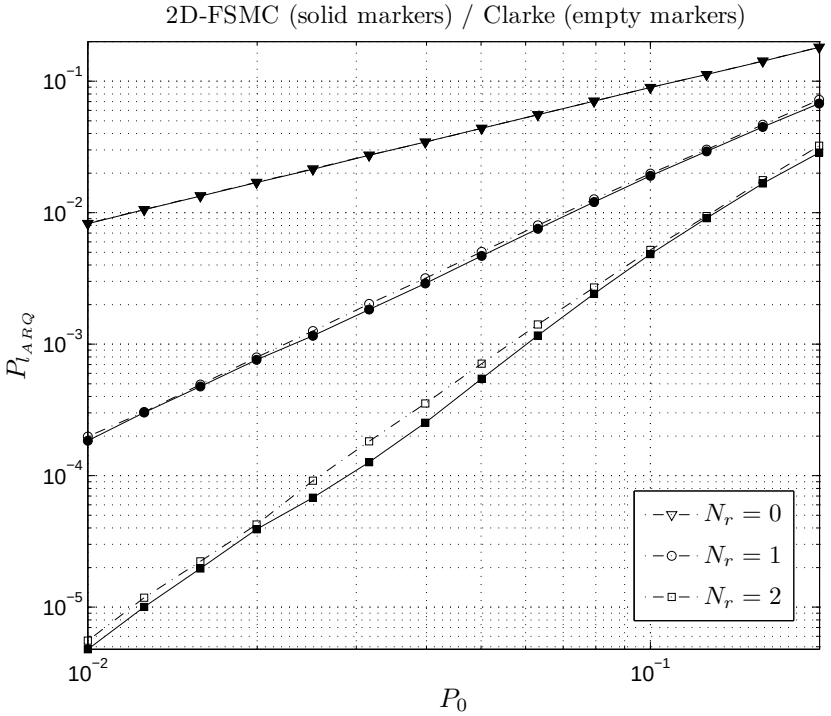


Figure 4.10: Average packet loss rate due to exceeding the maximum number of allowed retransmissions vs. target PER.

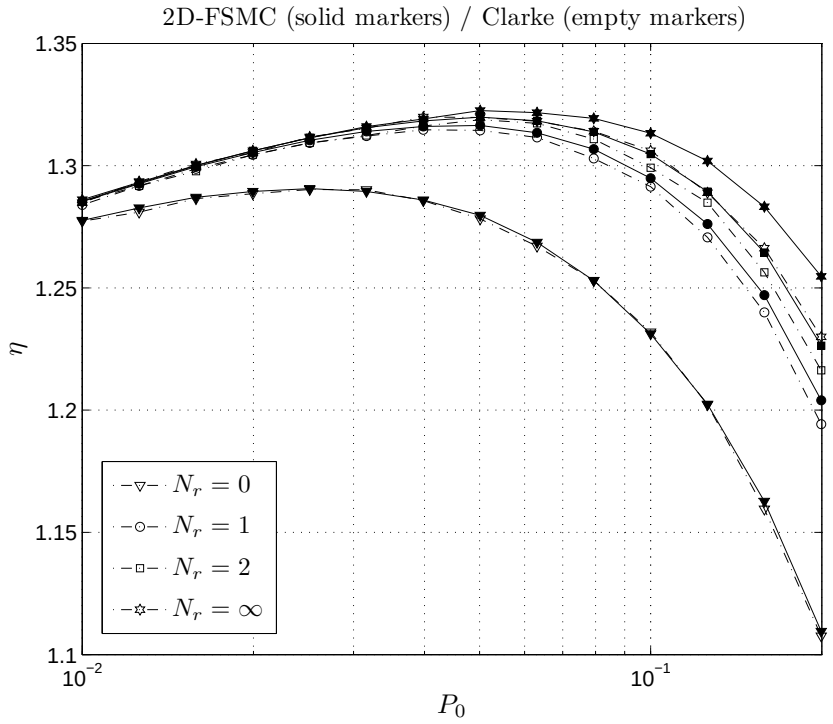


Figure 4.11: Average throughput vs. target PER.

dominant packet loss factor. The Pl_{BO} graph reveals that an increase of P_0, which implies the utilization of higher order TMs, causes a growth of the queueing service rate and, thus, a decrease in the buffer overflow probability Pl_{BO}. However, except in the case in which retransmissions are not allowed, when the increase of the service rate cannot cope with the huge number of required retransmissions, Pl_{BO} rapidly takes off. Obviously, the use of higher order TMs caused by an increase of P_0 also implies a higher packet loss rate due to exceeding the maximum number of allowed retransmissions Pl_{ARQ}. As shown in the graph depicting Pl_{ARQ}, higher N_r values imply a decrease in the packet loss due to excessive retransmissions, however, if more retransmissions are allowed more packets remain in the queue and Pl_{BO} raises accordingly.

As expected, Figures 4.11, 4.12 and 4.13 show, respectively, an increase in throughput η, but also in delay D_p and queue length L_q with the number of allowed retransmissions. However the most significant gain is obtained going from $N_r = 0$ to $N_r = 1$ and the use of higher N_r values is not worthwhile. These results are aligned with those obtained in (Liu et al., 2004). As it can be observed, in all cases the behaviour of the simulation of a FIFO queuing system under a truncated Type-I hybrid FEC/ARQ protocol and with a PHY layer based on Clarke's model is faithfully reproduced by the presented analytical PHY-DLC layer 2D-FSMC model. Once again, the shape and location of the minimum of the curves obtained by simulation (Clarke's model) coincide with those obtained using the proposed analytical model (2D-FSMC). The accuracy in determining the location of the maximum of the throughput and the minimum of the average

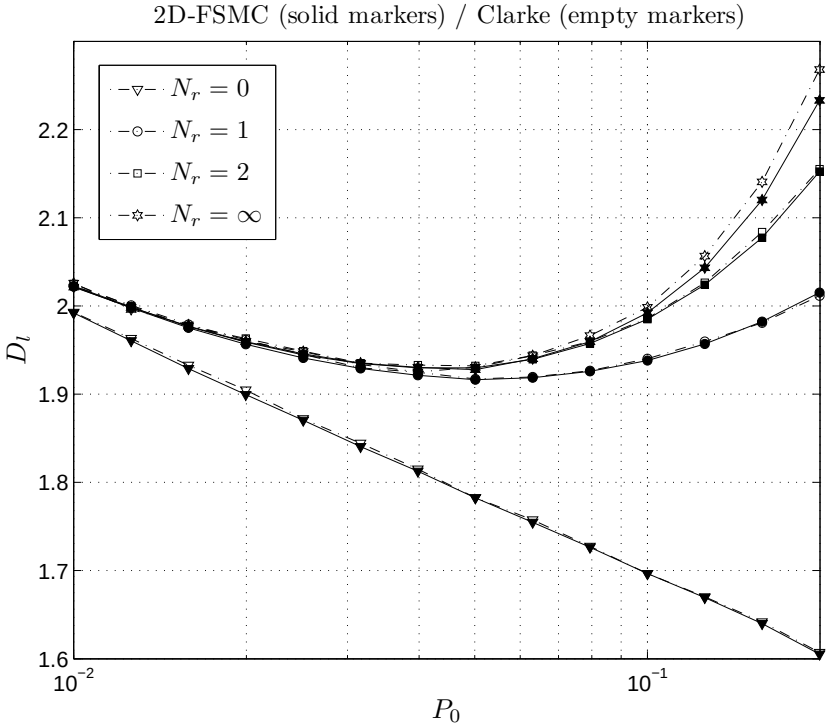


Figure 4.12: Average packet delay vs. target PER.

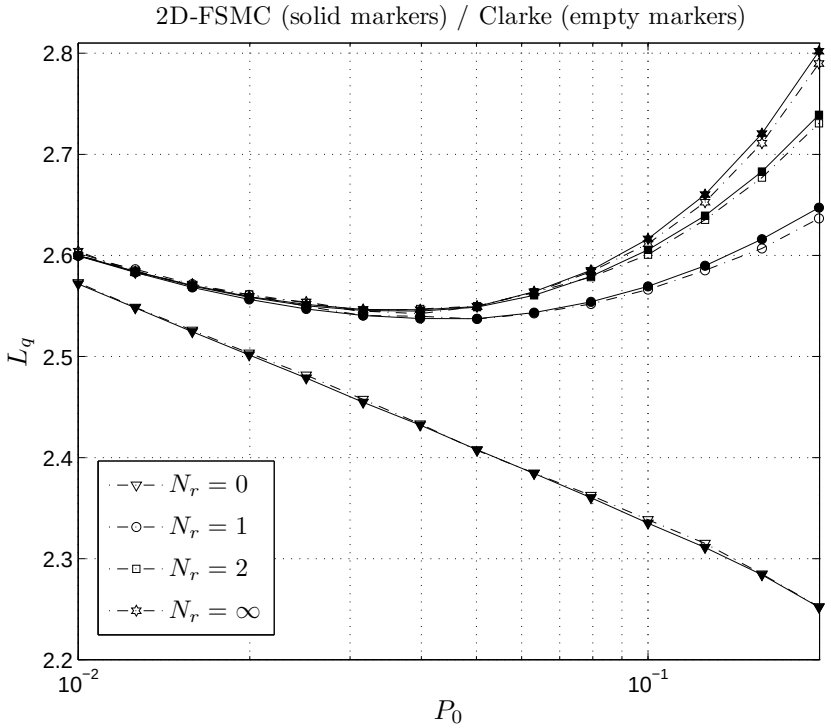


Figure 4.13: Average queue length vs. target PER.

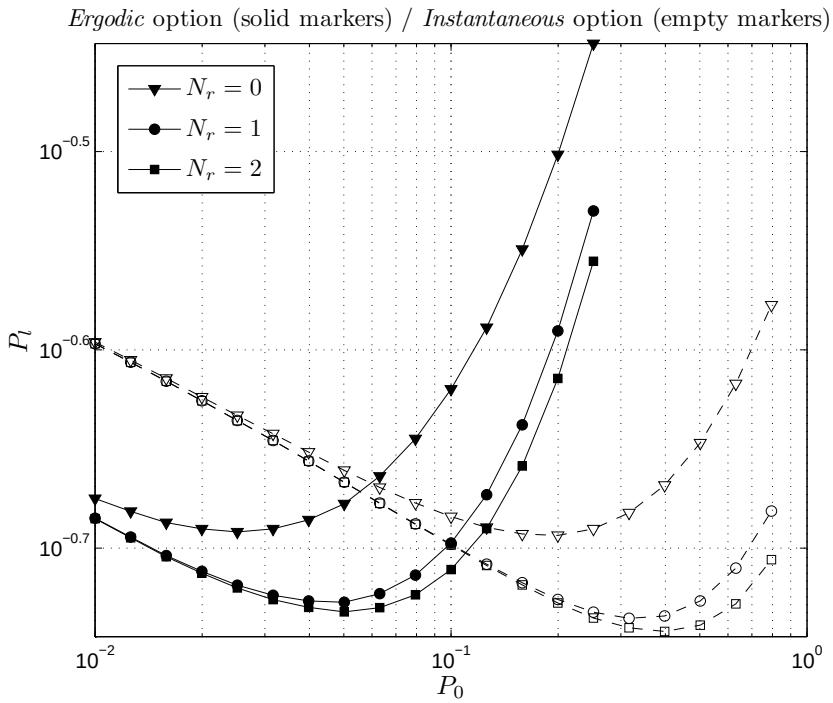


Figure 4.14: Average packet loss rate vs. target PER - *ergodic* and *instantaneous* options for the AMC scheme.

packet loss rate or the average packet delay is particularly important to ensure an optimal cross-layer design.

In order to analyze the influence on the behaviour of the system of the *ergodic* and *instantaneous* options for the AMC scheme, Figs. 4.14, 4.15, 4.16, 4.17, 4.18 and 4.19 show the curves of the QoS provisioning performance parameters obtained using the analytical expressions of the 2D-FSMC model for both AMC options, that is, the average packet loss rate P_l and its components $P_{l_{\mathrm{BO}}}$ and $P_{l_{\mathrm{ARQ}}}$, the average throughput η, the average packet delay D_p and the average queue length L_q. As in the case of an infinitely persistent Type-I hybrid FEC/ARQ protocol, both schemes have an analogous behaviour, although with a left shift in the graphs corresponding to the *ergodic* option. Therefore, the optimum throughput is obtained for higher P_0 values when the *instantaneous* option is considered. Because this approach is more stringent, it achieves a slightly better performance. Altogether, these results are aligned with those corresponding to the infinitely persistent case, thus confirming the similar optimum behaviour of both approaches.

4.5.2 Validation and study of the HARQ-IR and HARQ-CC protocols

Once again, results obtained with the proposed analytical Markov-based model and using Clarke's statistical Rayleigh model will be compared. Without loss of generality, and as it has been previously indicated, from this point onwards the

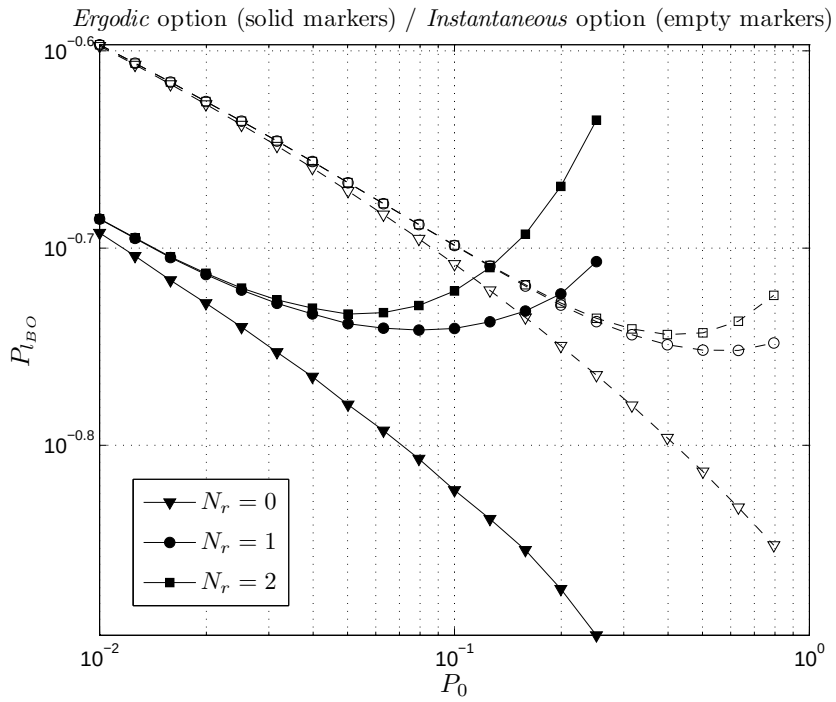


Figure 4.15: Average packet loss rate due to buffer overflow vs. target PER - *ergodic* and *instantaneous* options for the AMC scheme.

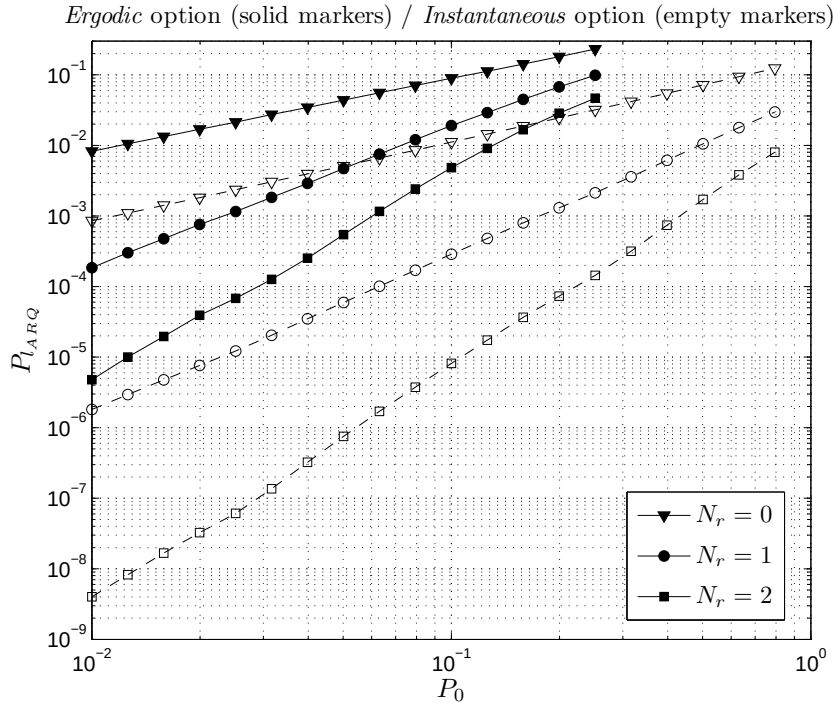


Figure 4.16: Average packet loss rate due to exceeding the maximum number of allowed retransmissions vs. target PER - *ergodic* and *instantaneous* options for the AMC scheme.

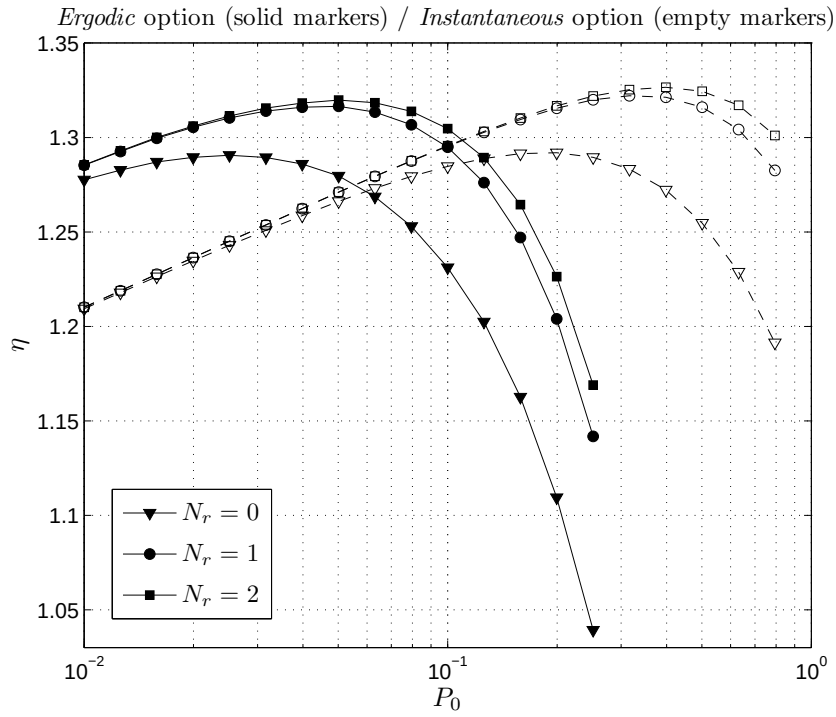


Figure 4.17: Average throughput vs. target PER - *ergodic* and *instantaneous* options for the AMC scheme.

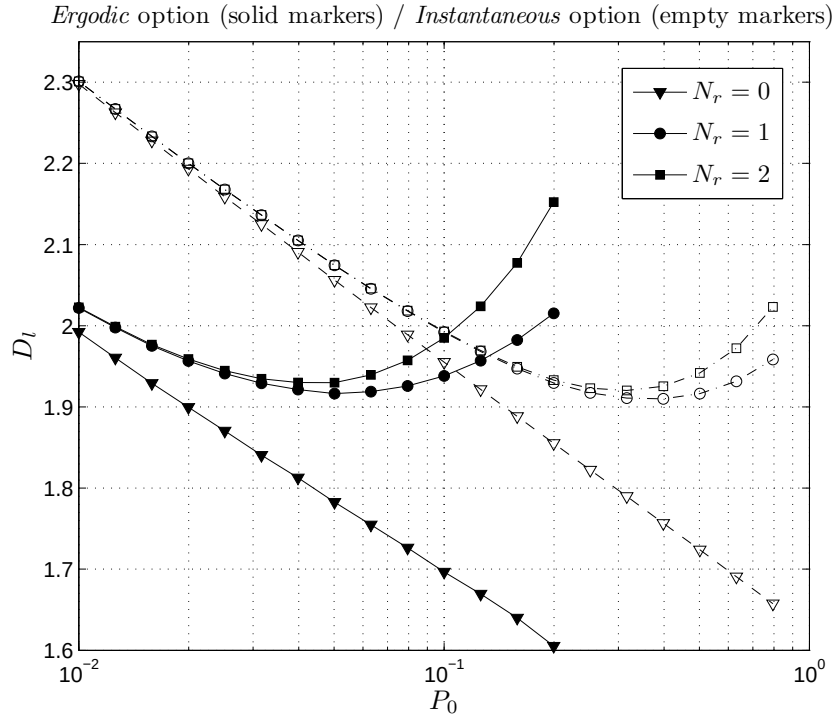


Figure 4.18: Average packet delay vs. target PER - *ergodic* and *instantaneous* options for the AMC scheme.

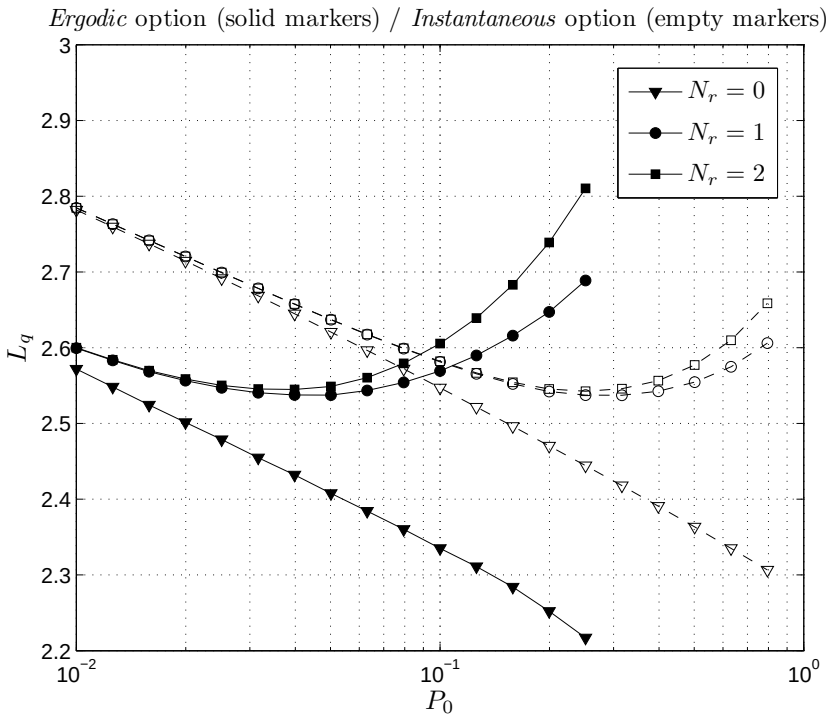


Figure 4.19: Average queue length vs. target PER - *ergodic* and *instantaneous* options for the AMC scheme.

considered TMs are borrowed from the IEEE 802.16e standard, and the AMC scheme will be designed according to the *instantaneous* option. The considered parameters are: a normalized maximum Doppler frequency $f_d T_f = 0.02$, an average received SNR $\overline{\gamma} = 8$ dB, a buffer size $\overline{Q} = 8$, a number of channel states $K = 5$, a parameter $b = 6$ and a D-BMAP characterized by the transition probability matrix (2.4).

Figure 4.20 depicts the dependence of the average throughput η on the target instantaneous PER P_0, for the HARQ-CC and the HARQ-IR schemes. The presented curves have been obtained using both, the proposed analytical model and the real system simulation. It can be clearly observed that as the number of allowed retransmissions is increased, the maximum achieved throughput raises; however, the most significant gain is obtained in going from $N_r = 0$ to $N_r = 1$, and the advantage of using higher N_r values becomes marginal. This figure demonstrates that higher P_0 values lead to a growth of the queueing service rate and, consequently, to a larger throughput. However, after reaching the optimum, increased P_0 values imply the use of TMs that are not robust enough, which in turn leads to a reduction in η. In agreement with Frederiksen and Kolding (2002), it is worth noting that HARQ-IR and HARQ-CC strategies present a similar performance when a link adaptation algorithm is used. Since both strategies provide almost the same performance, only results corresponding to the IR scheme will be displayed from this point onwards.

The dependence of the average packet loss rate P_l, as well as that of its two

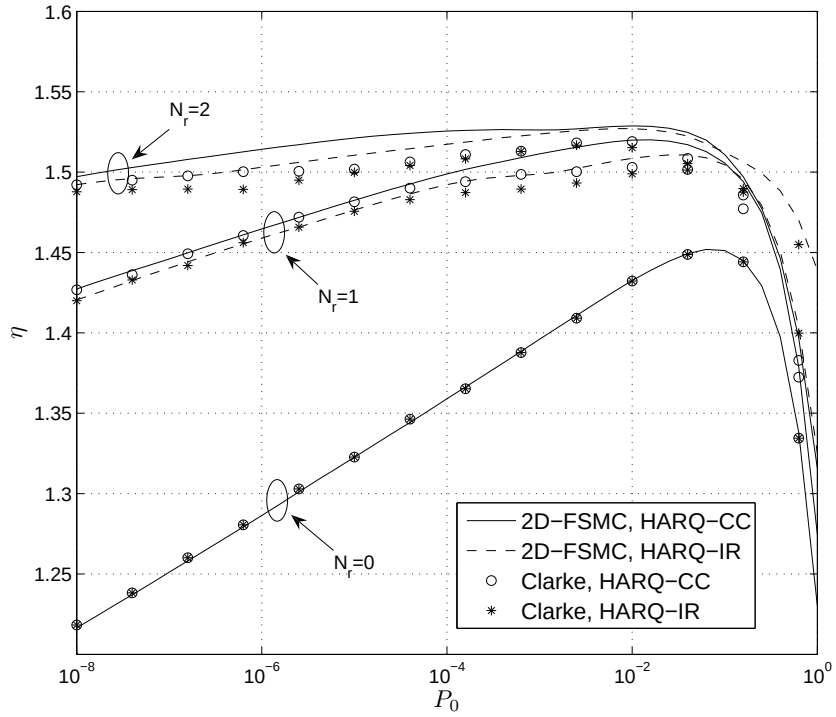


Figure 4.20: Average throughput vs. target PER, HARQ-CC and HARQ-IR.

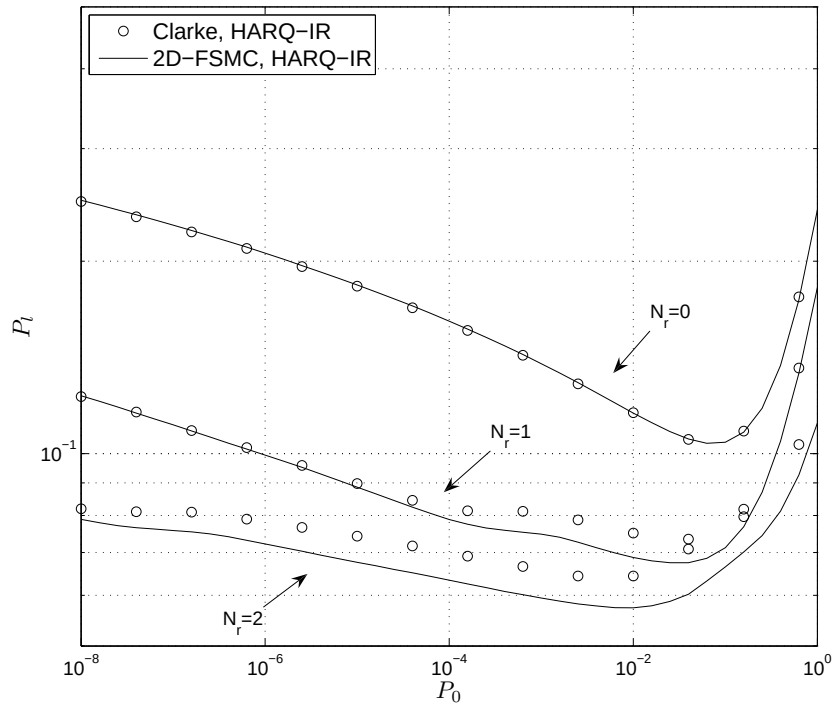


Figure 4.21: Total average packet loss rate vs. target PER, HARQ-IR.

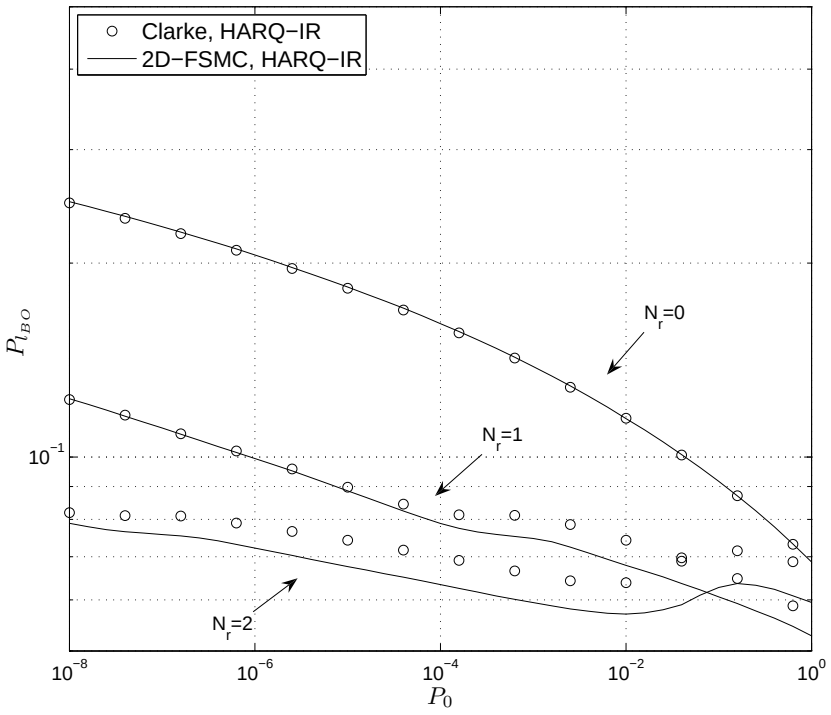


Figure 4.22: Average packet loss rate due to buffer overflow vs. target PER, HARQ-IR.

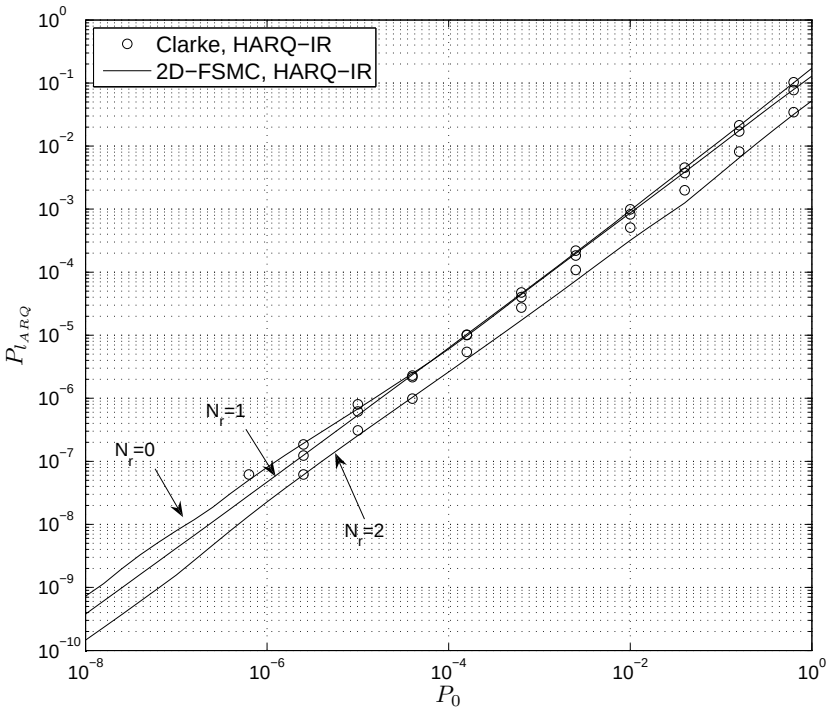


Figure 4.23: Average packet loss rate due to exceeding N_r vs. target PER, HARQ-IR.

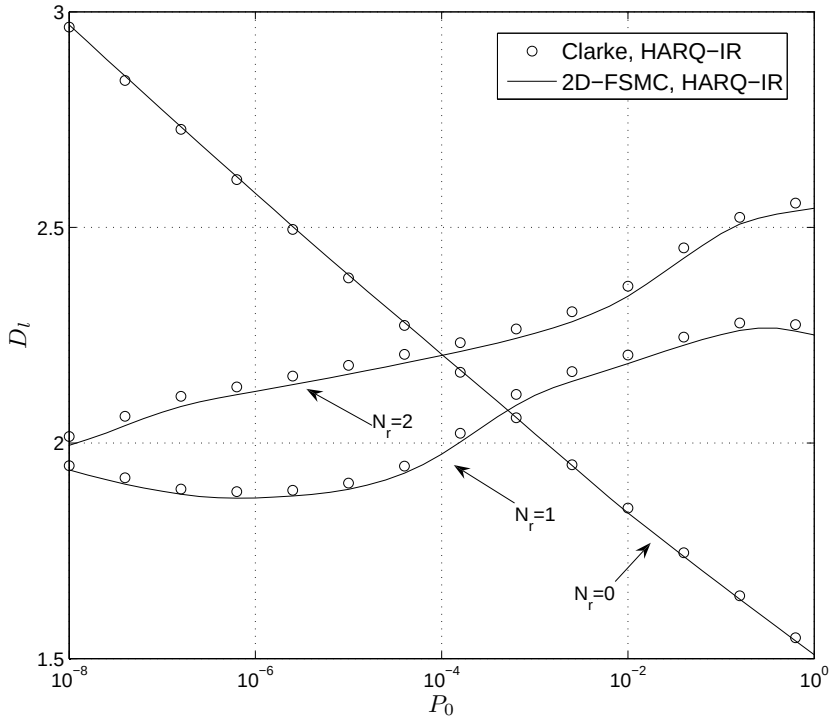


Figure 4.24: Average packet delay vs. target PER, HARQ-IR.

components $P_{l_{ARQ}}$ and $P_{l_{BO}}$, on the target instantaneous PER P_0, is shown in Figs. 4.21, 4.22 and 4.23. The analysis of these graphs demonstrates that higher P_0 values bring along the utilization of less robust (e.g., higher order) TMs, causing a growth of the packet loss rate due to exceeding the maximum number of allowed retransmissions. Furthermore, higher P_0 values also lead to an increase of the queueing service rate and, consequently, to a decrease in the buffer overflow probability $P_{l_{BO}}$. Nevertheless, it can be observed that when the number of allowed retransmissions N_r grows, the increase of the service rate cannot cope with the huge number of required retransmissions, resulting in more packets remaining in the queue and causing $P_{l_{BO}}$ to raise accordingly. As expected, when more retransmissions are allowed, lower $P_{l_{ARQ}}$ values are obtained. Comparing Figs. 4.22 and 4.23 it is clear that the packet loss due to buffer overflow $P_{l_{BO}}$ is the dominant packet loss factor except for high P_0 values. Thus, as it is shown in Fig. 4.21, the influence on the average packet loss rate of the packet losses due to exceeding the maximum number of allowed retransmissions is only noticeable for very high values of the target instantaneous PER P_0.

Figures 4.24 and 4.25 reveal that, for the case in which no retransmissions are allowed, an increase in P_0, implying the transmission of more packets per frame, entail a reduction in the queue length and also in the packet delay. However, when the retransmission of erroneously received packets is allowed, although more packets are transmitted per frame when P_0 increases, a higher percentage of them will need to be retransmitted and, consequently, more packets will remain in the queue, thus also increasing the average packet delay. Obviously, the higher the

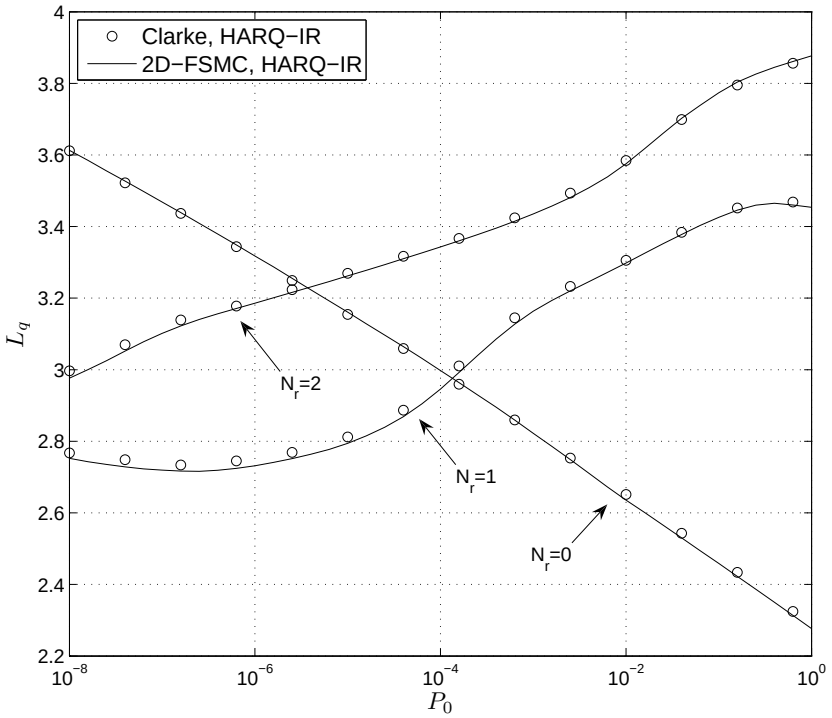


Figure 4.25: Average queue length vs. target PER, HARQ-IR.

number of allowed retransmissions, the more remarkable the growth of D_l and L_q is.

Results presented in Figs. 4.20-4.25 reveal that, in all cases, the behaviour of a real FIFO queuing system using a truncated (H)ARQ protocol is faithfully reproduced by the proposed analytical Markov model. The shape and location of the minimum of the curves obtained by simulation (Clarke's model) coincide with those obtained using our analytical model, which is particularly important to ensure an optimal cross-layer design.

4.5.3 Comparison of the Type-I hybrid FEC/ARQ and the HARQ protocols

With the purpose of comparing the behaviour of the Type-I hybrid FEC/ARQ and the HARQ protocols, analytical results corresponding to both schemes are confronted and analyzed in this subsection. The same default parameters as in Subsection 4.5.2 are considered.

Figures 4.26 to 4.28 depict the dependence of the average packet loss rate P_l, as well as its two components $P_{l_{BO}}$ and $P_{l_{ARQ}}$, on the target instantaneous PER P_0. The packet loss rate due to buffer overflow when the Type-I hybrid FEC/ARQ scheme is used clearly exceeds the packet losses corresponding to the HARQ-IR strategy. The opposite behaviour is observed for the packet losses due to exceeding the maximum number of allowed retransmissions. However, as $P_{l_{BO}}$ is the dominant factor for the P_0 values of interest, a lower minimum P_l is achieved with the

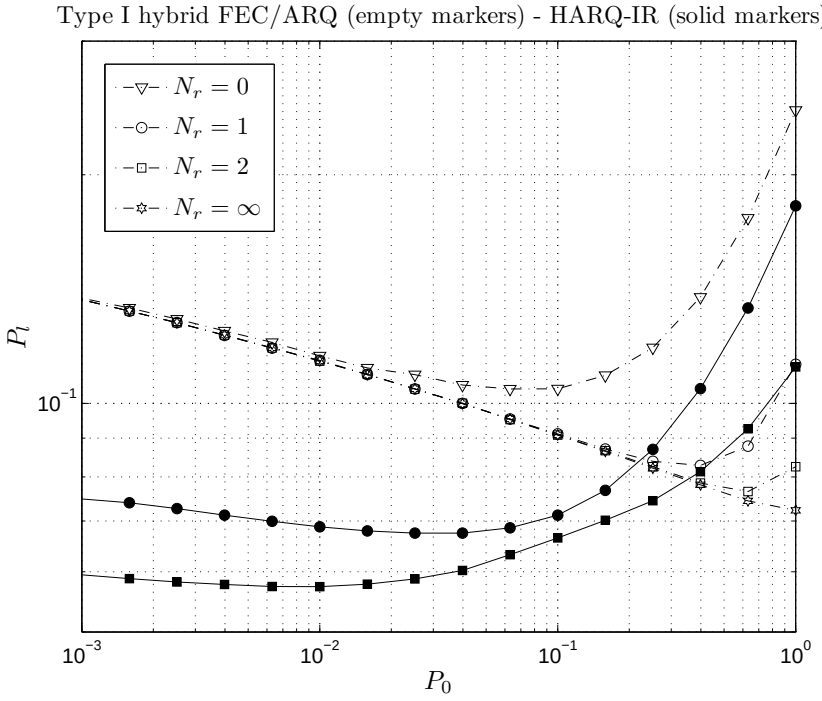


Figure 4.26: Average packet loss rate vs. target PER comparison of the Type-I hybrid FEC/ARQ and the HARQ protocols.

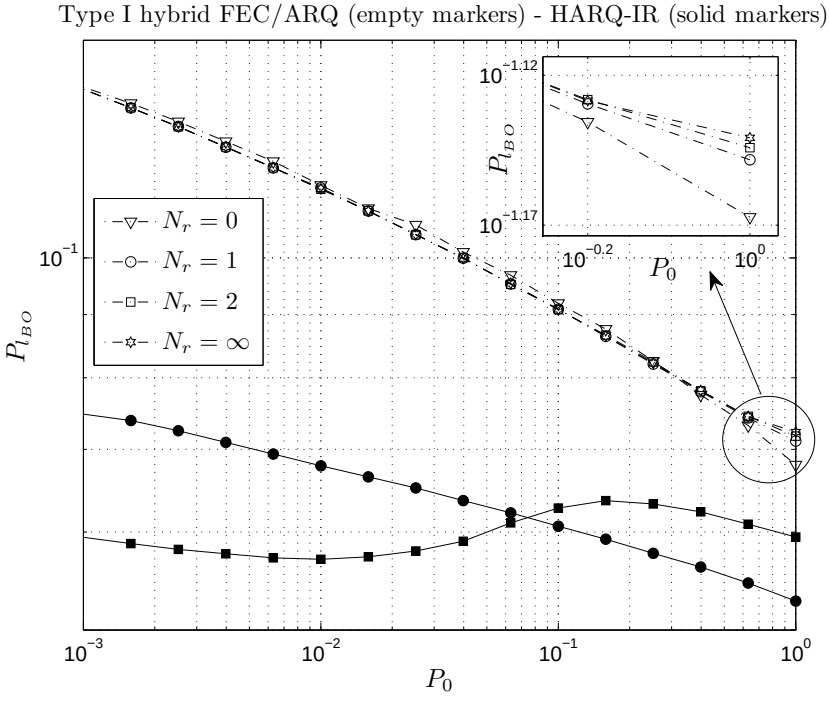


Figure 4.27: Average packet loss rate due to buffer overflow vs. target PER comparison of the Type-I hybrid FEC/ARQ and the HARQ protocols.

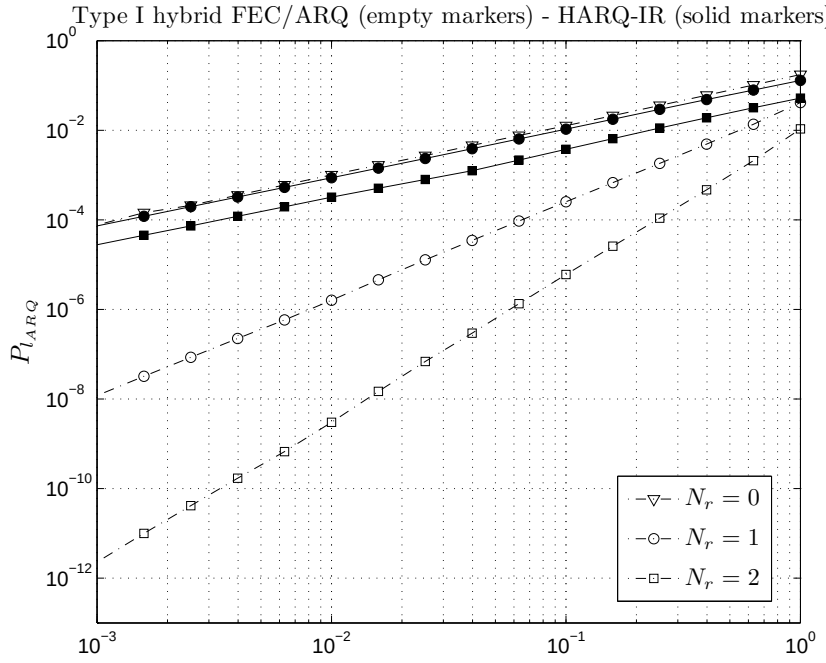


Figure 4.28: Average packet loss rate due to exceeding the maximum number of allowed retransmissions vs. target PER comparison of the Type-I hybrid FEC/ARQ and the HARQ protocols.

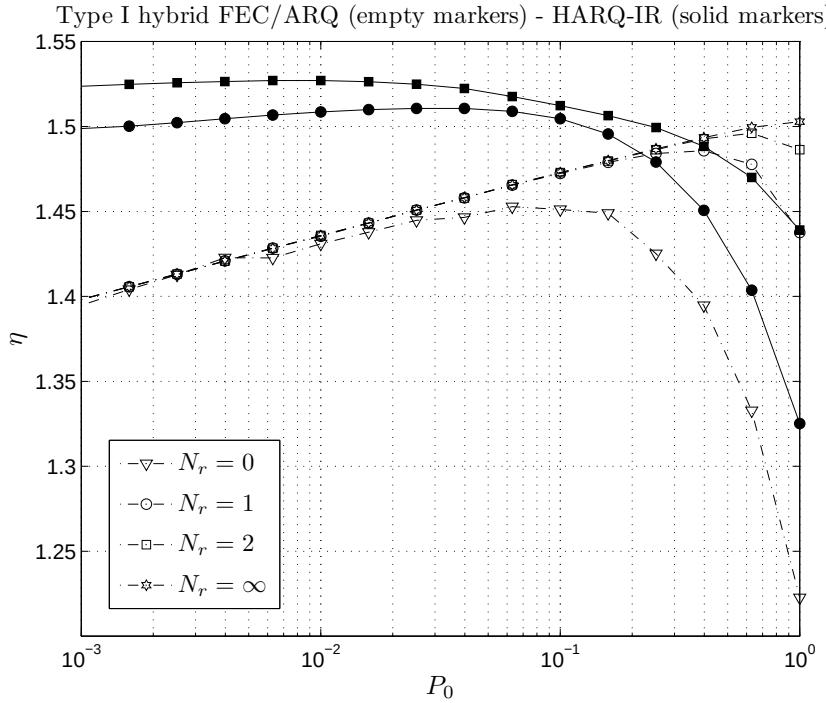


Figure 4.29: Average throughput vs. target PER comparison of the Type-I hybrid FEC/ARQ and the HARQ protocols.

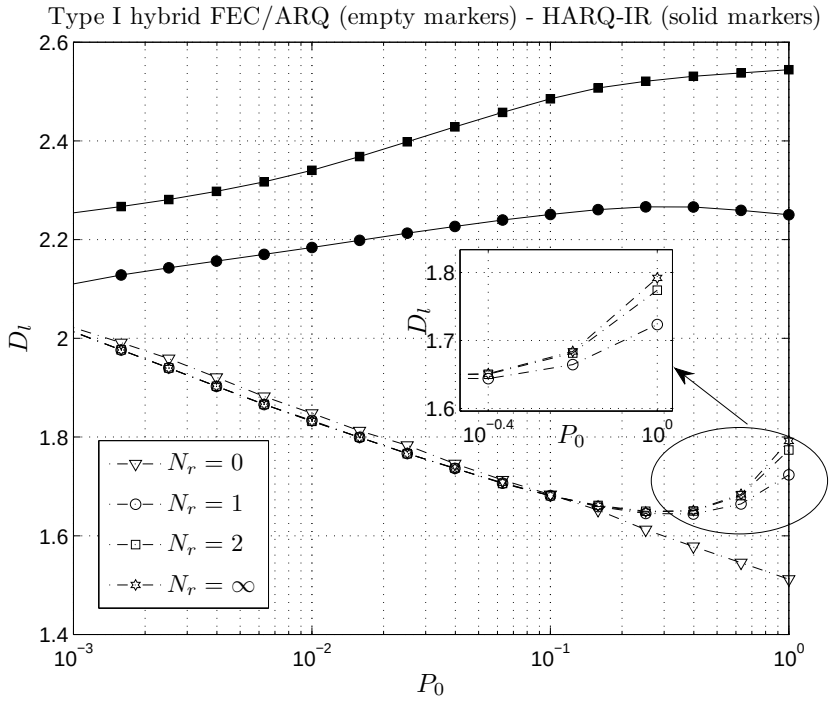


Figure 4.30: Average packet delay vs. target PER comparison of the Type-I hybrid FEC/ARQ and the HARQ protocols.

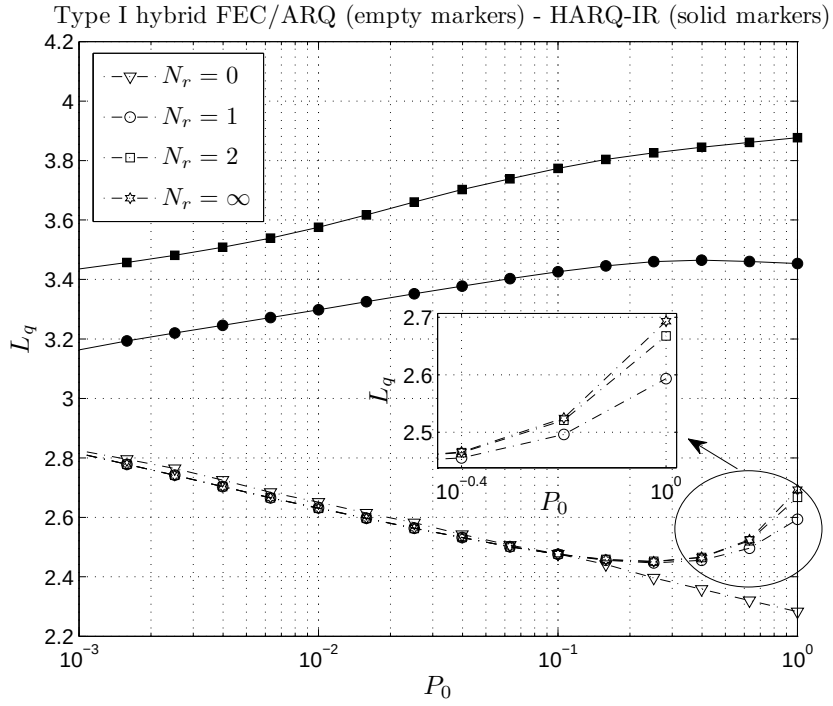


Figure 4.31: Average queue length vs. target PER comparison of the Type-I hybrid FEC/ARQ and the HARQ protocols.

HARQ-IR protocol. Consequently, as it is shown in Fig. 4.29, plotting the dependence of the average throughput on P_0, an enhanced optimum value of η is reached when the HARQ-IR strategy is utilized. It has to be highlighted that the maximum achieved throughput with the classical Type I FEC/ARQ scheme, even with infinite persistence, is clearly lower than the optimum throughput corresponding to the HARQ-IR protocol, even when only one retransmission is allowed.

The behaviour of the average packet delay and queue length is represented in figs. 4.30 and 4.31, respectively. They demonstrate that the buffer occupancy is higher with the HARQ-IR protocol, which in turn implies that, although a lower packet loss rate due to buffer overflow is achieved with this scheme, a longer delay is also obtained.

4.5.4 Analysis of the cross-layer design

With the objective of analyzing the cross-layer design proposed in Section 4.4, QoS-guaranteed traffic characterized by a maximum average packet loss rate $P_{l\max} = 0.1$ packets/frame and a maximum average packet delay $D_{l\max} = 4$ frames has been considered. Traffic has been generated using a truncated Poisson process with a truncation length equal to 4 packets, i.e. $\mathbf{\Theta} = (0, 4]$. Figure 4.32 plots the dependence of the obtained QoS parameters η and D_l, on the maximum normalized Doppler frequency, for an average received SNR $\overline{\gamma} = 6$ dB, whereas in Fig. 4.33 the dependence of this cross-layer design on the average SNR with a normalized maximum Doppler frequency $f_d T_f = 0.02$ is depicted. These figures illustrate results obtained for both, the multidimensional and the simplified bidimensional optimization problems. Although not shown in the figures, the constraint on the maximum average packet loss rate has been satisfied with equality in all simulations. It can be inferred from the shape of the curves that for higher $f_d T_f$ values, corresponding to shorter fading durations, the optimum throughput η increases while ensuring the fulfillment of both QoS requirements. It can also be observed that for higher $\overline{\gamma}$ values, which lead to the use of higher order TMs and imply a decrease in P_l, a better throughput can be achieved.

Figures 4.32 and 4.33 reveal, in short, a significant improvement in the maximum achieved throughput when using the HARQ-IR error control protocol in comparison to the classical type-I hybrid FEC/ARQ scheme. In fact, it can be observed that the IR scheme with $N_r = 1$ reaches a much higher optimum throughput than the classical type-I hybrid FEC/ARQ scheme, even with $N_r = 2$. As expected, when the AMC switching thresholds are found using the multidimensional optimization approach, a higher maximum throughput is obtained in comparison to the bidimensional optimization strategy, thus confirming the complexity-performance trade off. As already mentioned, similar results have been obtained with the HARQ-CC due to the use of an AMC scheme at the PHY layer. It is worth noting that this gain largely outperforms the improvement obtained in Subsection 3.5.3 for the infinitely persistent Type-I hybrid FEC/ARQ protocol with the 802.11a AMC pool.

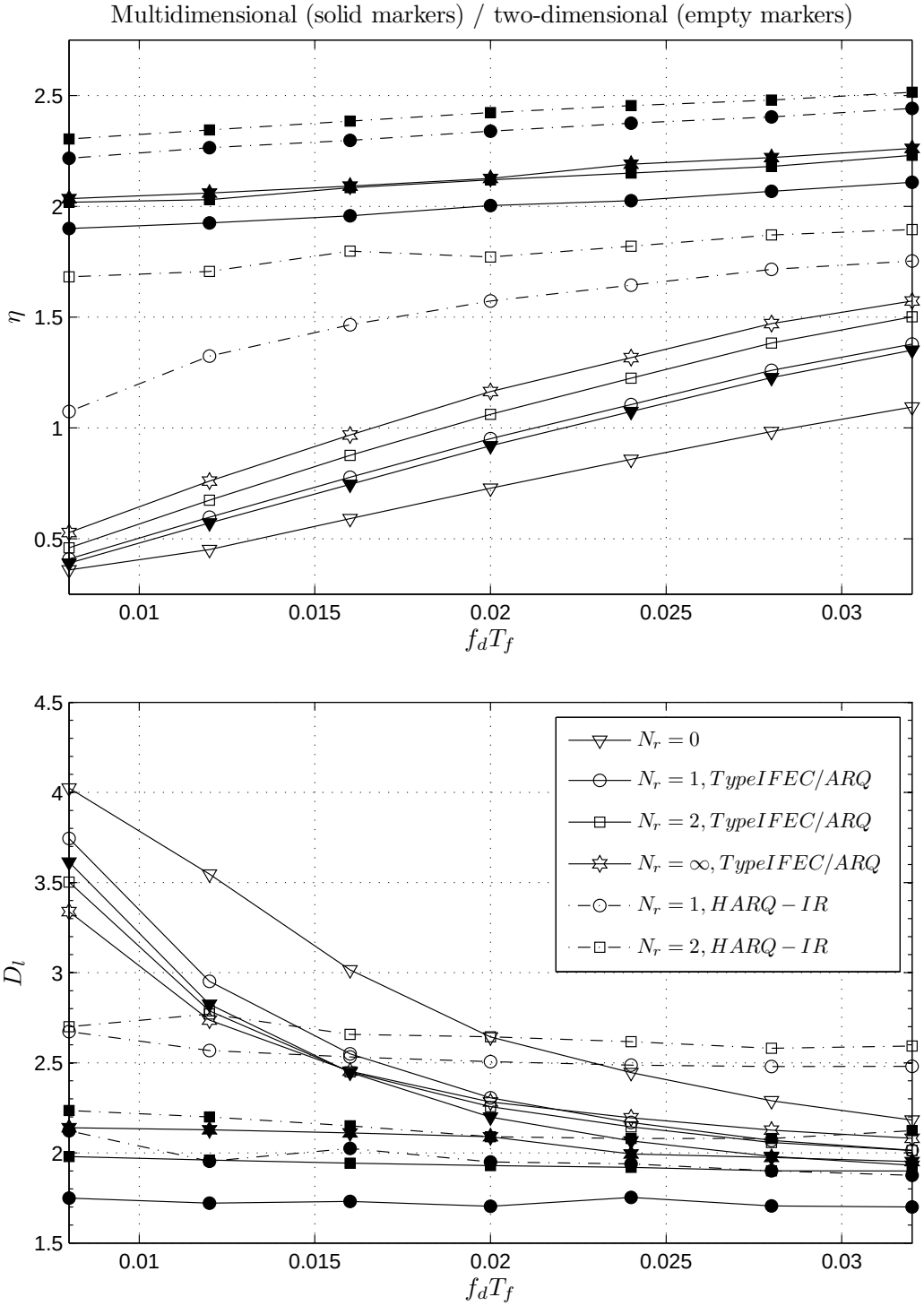


Figure 4.32: Cross-layer optimization vs. maximum normalized Doppler frequency.

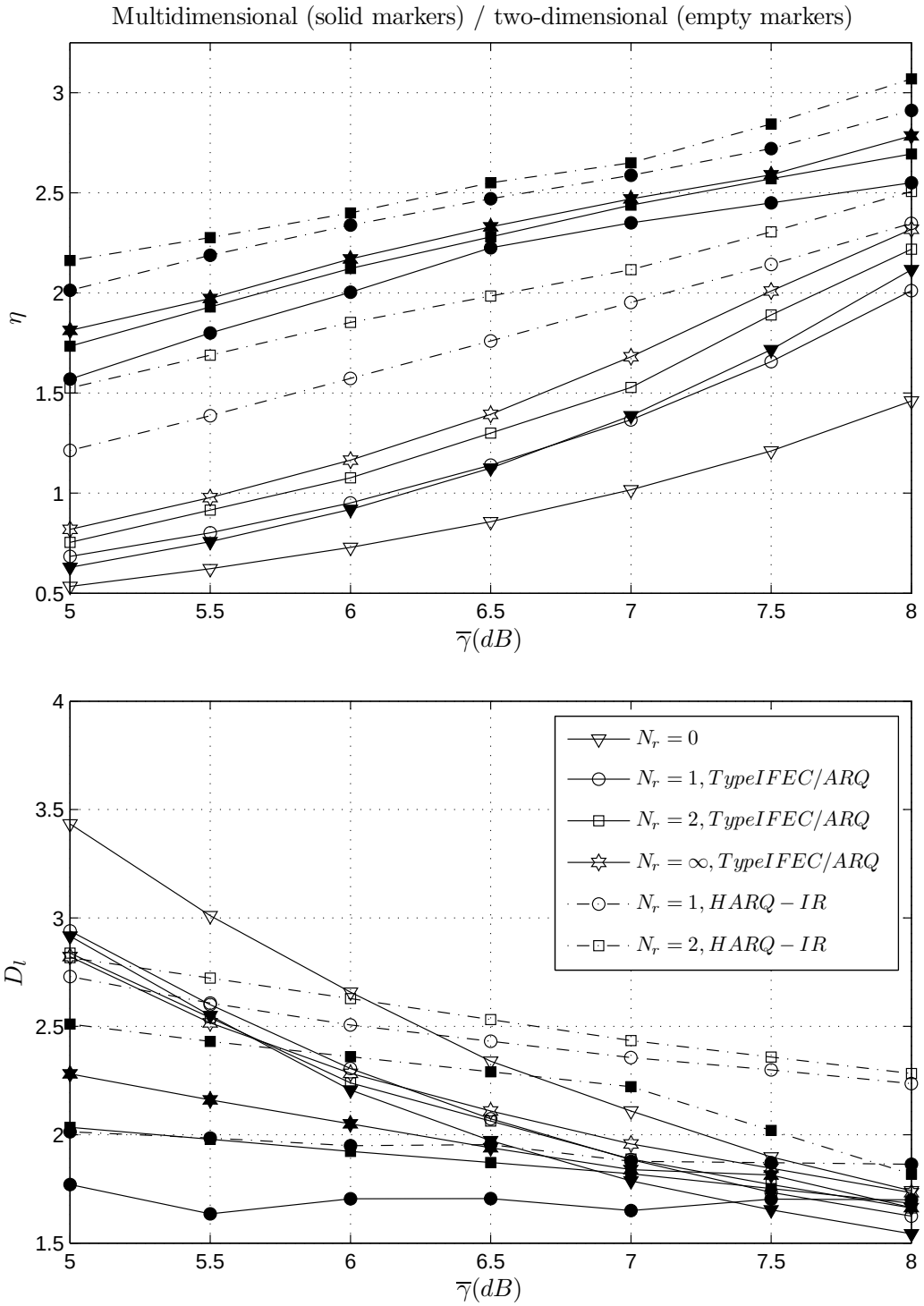


Figure 4.33: Cross-layer optimization vs. average SNR.

4.6 Chapter summary

A cross-layer DTMC-based queuing model jointly exploiting the error-correcting capability of truncated (H)ARQ protocols at the DLC layer and the adaptation ability of AMC strategies at the PHY layer has been developed. This DTMC model is based on a comprehensive consideration of packet multimedia traffic sources modeled as a D-BMAP, finite queue-length systems, truncated (H)ARQ protocols, AMC strategies, and a wireless channel first-order two-dimensional Markov model using both the amplitude and the rate of change of the fading envelope. The stationary state probability distribution of this multidimensional DTMC has been used to obtain closed-form analytical expressions for performance metrics such as throughput, average packet delay, queue length and packet loss rate, both due to buffer overflow and due to exceeding the maximum number of allowed retransmissions. Furthermore, this cross-layer analytical framework has been exploited to formulate multidimensional and simplified bidimensional constrained optimization problems aiming at maximizing the system throughput under maximum packet loss rate and delay constraints.

To assess the validity of the proposed analytical cross-layer framework, computer simulation results have been presented to show that the derived closed-form analytical expressions can be used to faithfully reproduce results based on Clarke's statistical Rayleigh fading model. Numerical results have shown that both HARQ-CC and HARQ-IR schemes lead to a remarkable improvement on the maximum achieved throughput when compared to the classical type-I hybrid FEC/ARQ scheme. The amount of this improvement largely depends on the channel conditions and the maximum number of allowed ARQ retransmissions. As in (Frederiksen and Kolding, 2002), it has also been shown that HARQ-IR protocols provide a small throughput gain when compared to HARQ-CC schemes due to the use of PHY layer AMC algorithms. Numerical results have also confirmed the benefits of the multidimensional optimization in attaining the maximum throughput in comparison to the simplified bidimensional approach.

CHAPTER 5

Systems based on cooperative ARQ protocols

This chapter aims at analyzing the effects of using CARQ in wireless relay systems using AMC at the PHY layer and an infinitely persistent type-I hybrid FEC/ARQ protocol at the DLC layer. To that end, and following the methodology described in previous chapters, a cross-layer multidimensional DTMC-based queuing model is developed to jointly exploit the capabilities of CARQ and AMC. Based on the stationary state probability distribution of the proposed multidimensional DTMC, closed-form analytical expressions for performance metrics such as average throughput and packet delay, as well as average queue lengths and packet loss rates at both the source and the relay nodes, are derived, and the impact of cooperative relaying on these performance parameters is analyzed. Numerical results confirm the validity of the model and reveal that the proposed CARQ protocol consistently outperforms the classical type-I hybrid FEC/ARQ scheme and provides a noticeable improvement in correlated fading channel scenarios.

5.1 Introduction

The performance analysis of CARQ protocols has been previously covered in the literature (e.g., (Boujemaa, 2009; Byun and Kim, 2011; Cerutti et al., 2008; Dianati et al., 2006; Ren and Letaief, 2009; Yu et al., 2006)). In (Dianati et al., 2006; Ren and Letaief, 2009; Yu et al., 2006) different decode-and-forward CARQ protocols for the type-I hybrid FEC/ARQ scheme were proposed. Dianati et al. (2006) proposed a CARQ scheme for wireless *ad hoc* networks and developed an analytical model to obtain the throughput, average delay (defined as the time required to complete the transmission of all the fragments of a single packet, without considering the queueing delay) and delay jitter by using a two-state Markov process for wireless fading channels. Results indicate that a small number of cooperative nodes provide significant performance gain, specially when the average quality of

the channel between the sender and the receiver nodes is poor. In (Yu et al., 2006) three different CARQ protocols were proposed for a single relay-assisted network, which combine incremental relaying and selection relaying (Laneman et al., 2004). All three protocols attain full (second) order diversity, which is superior to traditional ARQ protocol, where only first order diversity is achieved. A time-slot allocation scheme to increase the effective capacity subject to the queue-overflow statistical QoS requirement for a multi-relay cooperative wireless network was proposed in (Ren and Letaief, 2009). This is obtained through the transformation of spatial diversity into temporal diversity by appropriately allocating a time slot of optimal length to each relay. HARQ protocols, for both the CC and the IR strategies, have also been considered to implement cooperation. Cerutti et al. (2008) presented a queueing model to analytically derive the throughput, buffer occupancy and delay (defined as the elapsed time between the data frame generation at the source (S) and the correct reception at the destination (D)) for single-source and single-relay CARQ protocols with hybrid retransmission techniques. The work of Cerutti *et al.* is considered the first delay model of cooperative HARQ (Boujemaa, 2009), however, in order to keep the analysis simple, some unrealistic assumptions were made. For instance, if the relay transmission to the destination fails, it withdraws the packet and the source transmits again the packet. For that reason, the relay (R) does not need to keep a copy of the data from S more than one time frame. In addition, the source transmitter buffer capacity is unbounded. Boujemaa (2009) developed a cooperative HARQ with opportunistic relaying protocol and derived the theoretical expressions of the expected delay. In contrast to (Cerutti et al., 2008), it is assumed in this work that only the first transmission is made by the source while the remaining transmissions are conducted by the relay with the highest relay-destination instantaneous SNR. Two unified frameworks for protocol analysis were presented in (Byun and Kim, 2011), one for decode-and-forward cooperative HARQ protocols and the other for amplify-and-forward cooperative HARQ protocols, providing actual performance comparison and evaluation with respect to channel and environment among different protocols.

The analysis of CARQ multi-rate wireless systems has been also widely treated. In (Dai and Letaief, 2008) the authors propose a cross-layer design combining truncated ARQ at the DLC layer and cooperative diversity at the PHY layer, where the packet length and the modulation level are jointly optimized in order to maximize the throughput at the DLC layer. The relays are not fixed and only the ones that correctly detect the packet are selected for possible retransmission, with both the source node and the relays utilizing a suitable orthogonal space-time block code (STBC) to retransmit the packet. AMC and cooperative diversity at the PHY layer are combined with HARQ at the DLC layer in the cross-layer design for a single-source and single-relay *ad hoc* network proposed in (Shi and Yuan, 2008). This is a hybrid scheme in which the system switches from direct transmission to cooperative transmission (via the relay) when the received packet is erroneously detected at the destination and the source-destination channel is in deep fade. The PER and spectral efficiency are analytically derived and compared for direct transmission, cooperative transmission and the proposed hybrid transmission scheme. Results show that the latter scheme outperforms the direct transmission and it is as good as the cooperative transmission scheme at a much

lower cooperation cost. The analysis of the spectral efficiency gain achieved by a joint design of discrete-rate AMC with CARQ while satisfying the packet loss rate constraint in a single-source single-relay network is the main contribution of Mardani et al. (2011). The authors propose a cross-layer approach to design an optimized rate adaptation policy where the source and relay nodes select their transmission rates in each frame based on the received CSI feedback from the destination node. They also devise a fixed rate selection algorithm for the scenario where providing per-frame feedback is not feasible but the channel statistics are available at the destination node. In (Harsini et al., 2011) a similar cross-layer approach for AMC design is developed, where a HARQ protocol is considered. Results show that the proposed cooperative HARQ protocol can effectively mitigate the effect of channel correlation thus outperforming the HARQ protocol in terms of throughput and packet loss rate.

To the best of our knowledge, previous studies neither consider the queueing process induced by both the AMC scheme and the CARQ protocol nor derive the packet delay in such systems. To that end, this chapter aims at investigating the effects of using CARQ for wireless relay systems using AMC at the PHY layer and an infinitely persistent type-I hybrid FEC/ARQ protocol at the DLC layer. Hence, a cooperative PHY layer model is developed and incorporated in the analysis of the queueing process induced by both the source and relay nodes. Following the methodology described in previous chapters, the proposed Markov chain-based model is then used to obtain closed-form analytical expressions for performance metrics such as average throughput and packet delay, as well as average queue lengths and packet loss rates at both the source and the relay nodes. The analytical expressions will then be used to formulate cross-layer multidimensional and simplified tridimensional constrained optimization problems aiming at maximizing the system throughput under prescribed QoS requirements in terms of maximum allowable average packet loss rate and delay.

5.2 System model and assumptions

Unlike previously considered ARQ schemes, cooperative retransmission techniques consist of at least three nodes. For that reason, the considered system slightly departs from the general system model presented in Chapter 2 and, therefore, will be thoroughly described in this section.

The point-to-point wireless packet communication system under consideration is illustrated in Fig. 5.1, showing the most important blocks involved in the transmission and reception processes. It is composed of source (S), relay (R) and destination (D) nodes. At the beginning of the frame interval, node R transmits packets according to the TM selected by the AMC controller at the relay. This transmission spans a fraction of the frame period and the remaining fraction, if any, is allocated to node S, which uses the TM selected by the AMC controller at the source. Figure 5.2 illustrates two examples of a transmission in the considered cooperative system. In both cases there are $q^R = 3$ packets in the queue of the relay and $q^S = 5$ packets in the queue of the source, and the selected TM at node S allows the transmission of $c_{n^S} = 6$ packets per frame. Figure 5.2(a) corresponds

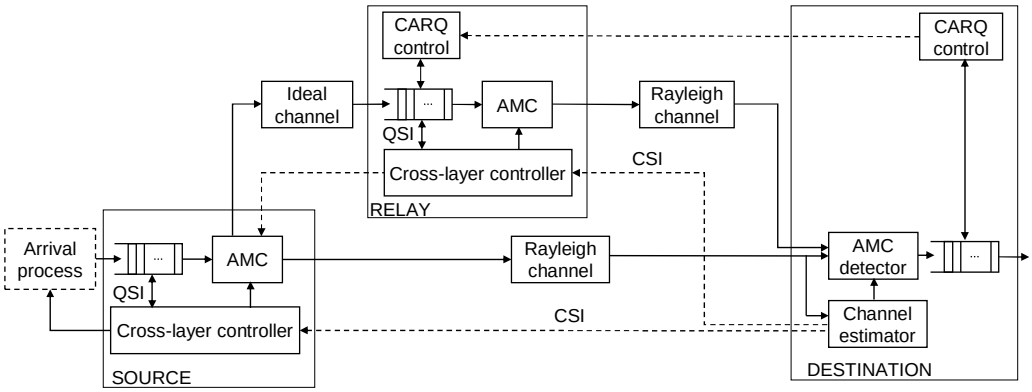


Figure 5.1: System model.

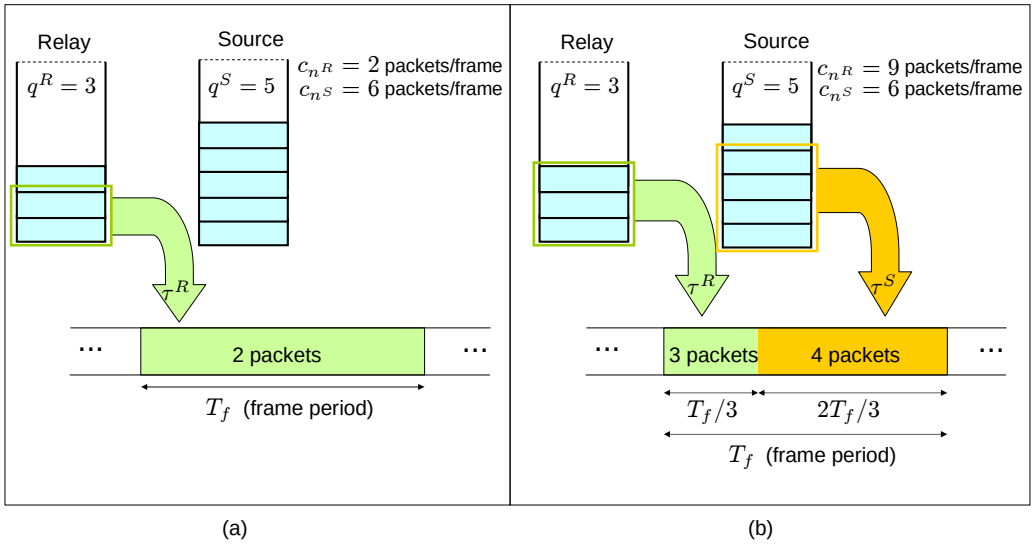


Figure 5.2: Cooperative transmission examples.

to an scenario in which node R can transmit $c_{nR} = 2$ packets in a frame period T_f. Thus, only $\tau^R = 2$ packets of this node will be transmitted in this frame interval and no remaining time will be left for the source. In the situation illustrated in Fig. 5.2(b), node R can transmit $c_{nR} = 9$ packets per frame and, consequently, $\tau^R = 3$ packets of the relay buffer will be selected for transmission, spanning 1/3 of the frame duration. Therefore, the remaining time, that is, $2T_f/3$, will be used by the source to transmit $\tau^S = 4$ packets.

The transmitter at the source node manages a buffer (queue) that operates in a FIFO mode and can store up to $\overline{Q}_S$ packets. We consider a CARQ protocol in which the data frames transmitted by the source node are received by both the relay and the destination nodes. We assume that the S-R link is robust enough such that the relay can always correctly decode the source messages. This is a realistic assumption in wireless networks with a large number of potential relays, possibly with S-R line-of-sight operation, so that the S node can select a relay

node with a good channel condition (Wang et al., 2008a). The study of such a relay selection algorithm is beyond the scope of this research work. Consequently, packets will be removed from the source buffer after being transmitted, whether they are correctly received by the mobile host or not (it is the relay that will be in charge of retransmitting the erroneously received packets at the destination).

As it has been previously seen in Chapter 2, a Rayleigh block-fading channel has been adopted. Thus, the channel gains corresponding to the νth frame transmission, h_ν^L, with $L = S$ denoting the S-D link and $L = R$ denoting the R-D link, are characterized as a zero-mean circularly symmetric complex Gaussian random variable with unit power.

After detection and soft Viterbi decoding at the destination node, an error detection process (using, for instance, a CRC code) is performed, and the corresponding ACK/NACK message is fed back to the CARQ controller at the relay node. Given the short length of the ACK/NACK messages and the use of a high degree of FEC protection, an error-free CARQ feedback channel can be safely assumed. When implementing the AMC strategy, the entire SNR range is partitioned into a set of non-overlapping intervals defined by the partition

$$\mathbf{\Gamma}^{m_L} = \{[\gamma_0^{m_L}, \gamma_1^{m_L}), \ldots, [\gamma_{M_L-1}^{m_L}, \gamma_{M_L}^{m_L})\} \tag{5.1}$$

with $L = S$, denoting the S-D link. M_S represents the number of *useful* TMs for the S-D link, and mode n is selected when $\gamma_\nu^S \in [\gamma_n^{m_S}, \gamma_{n+1}^{m_S})$, where γ_ν^S denotes the instantaneous received SNR corresponding to the S-D link at time instant $t = \nu T_f$.

As it has already been mentioned, packets are supposed to be correctly received and stored in the relay buffer, which has a maximum size of $\overline{Q}_R$ packets. According to the ACK/NACK messages sent back from node D to node R, those packets that have been correctly received at the destination will be removed from the relay buffer. On the contrary, in case they are detected in error, they will be kept in the queue and retransmitted until they are correctly received. Similarly to the AMC strategy followed over the S-D link, a partition for the TM selection is defined as in (5.1), with $L = R$, which denotes the R-D link. M_R is the number of *useful* TMs for the R-D link, and mode n is selected when $\gamma_\nu^R \in [\gamma_n^{m_R}, \gamma_{n+1}^{m_R})$, with γ_ν^R being the instantaneous received SNR corresponding to the R-D link at time instant $t = \nu T_f$.

Perfect CSI is assumed to be available at the receiver side of both, the S-D and the R-D links and, therefore, a frame-by-frame TM selection process for each link is performed at the receiver AMC controller and fed back to the corresponding transmitters without error and/or latency.

5.3 Physical layer modeling

5.3.1 AMC pool and characterization of the packet error rate

As it has been previously mentioned, an infinitely persistent Type-I hybrid FEC/ ARQ protocol will be considered, and without loss of generality, convolutionally

coded M-QAM schemes adopted from the IEEE 802.11a standard (IEEE, 1997) will be used in the AMC pool, whose TMs are listed in Table 3.1 (Liu et al., 2004). Therefore, in the presence of AWGN, the PER of these TMs can be approximated as in (3.3).

5.3.2 Physical layer two-dimensional Markov model

Channel modeling

Following the procedure developed in Section 2.4, let us consider the Rayleigh block-fading channel quantities γ_ν^L and $\delta_\nu^L = \gamma_{\nu-1}^L - \gamma_\nu^L$, with $L = S$ denoting the S-D link and $L = R$ the R-D link. Let us also partition the ranges of γ_ν^L and δ_ν^L into sets of non-overlapping two-dimensional cells defined by the partitions

$$\mathbf{\Gamma}^{c_L} = \left\{ [\gamma_0^{c_L}, \gamma_1^{c_L}), [\gamma_1^{c_L}, \gamma_2^{c_L}), \ldots, \left[\gamma_{K_L-1}^{c_L}, \gamma_{K_L}^{c_L}\right) \right\}, \tag{5.2}$$

with $\gamma_0^{c_L} = 0$ and $\gamma_{K_L}^{c_L} = \infty$, and $\mathbf{\Delta} = \{(-\infty, 0), [0, \infty)\}$, respectively. The partitions $\mathbf{\Gamma}^{c_L}$ are designed using the methodology introduced in Section 2.4, for both the S-D link and the R-D link. Thus, a first-order two-dimensional Markov channel model can be defined where each state of the channel corresponds to one of such cells. That is, the Markov chain state of the S-D and R-D channels at time instant $t = \nu T_f$ can be denoted as $\boldsymbol{\zeta}_\nu^L = \left(\chi_\nu^L, \Delta_\nu^L\right)$, $\nu = 0, 1, \ldots, \infty$, where $\chi_\nu^L = k$ if and only if $\gamma_k^{c_L} \leq \gamma_\nu^L < \gamma_{k+1}^{c_L}$, and $\Delta_\nu^L = 0$ (or $\Delta_\nu^L = 1$) if and only if $\delta_\nu^L < 0$ (or $\delta_\nu^L \geq 0$).

Physical layer modeling

The PHY layer model of the proposed cooperative system will be developed using a procedure analogous to that followed in Subsections 3.2.2 and 4.2.2, where an infinitely persistent Type-I hybrid FEC/ARQ protocol and an HARQ protocol were considered, respectively.

Taking into account both, the TM selection process used by the AMC scheme, defined by the partition $\mathbf{\Gamma}^{m_L}$, and the first-order two dimensional Markov channel model, characterized by the partitions $\mathbf{\Gamma}^{c_L}$ and $\mathbf{\Delta}$, let us now partition the range of γ_ν^L into the set of non-overlapping intervals defined by the partition

$$\mathbf{\Gamma}^{mc_L} = \{[\gamma_0^{mc_L}, \gamma_1^{mc_L}), \ldots, [\gamma_{N_{\mathrm{PHY}_L}-1}^{mc_L}, \gamma_{N_{\mathrm{PHY}_L}}^{mc_L})\}, \tag{5.3}$$

where $\gamma_0^{mc_L} = 0$, $\gamma_{N_{\mathrm{PHY}_L}}^{mc_L} = \infty$,

$$\begin{aligned} &\{\gamma_1^{mc_L}, \ldots, \gamma_{N_{\mathrm{PHY}_L}-1}^{mc_L}\} \\ &\quad = \mathrm{sort}\left(\{\gamma_1^{m_L}, \ldots, \gamma_{M_L-1}^{m_L}\} \cup \{\gamma_1^{c_L}, \ldots, \gamma_{K_L-1}^{c_L}\}\right), \end{aligned} \tag{5.4}$$

and each partition interval $[\gamma_k^{mc_L}, \gamma_{k+1}^{mc_L})$ is characterized by a particular combination of TM and channel state. As in Subsection 3.2.2, let us also consider the partition of δ_ν^L into the set of non-overlapping intervals $\mathbf{\Delta} = \{(-\infty, 0), [0, \infty)\}$. Using this two-dimensional partitioning, a first-order two-dimensional Markov model

for the PHY layer of both the S-D link and the R-D link can be defined where each state corresponds to one of such two-dimensional rectangular-shaped cells. Furthermore, the PHY layer Markov chain state at time instant $t = \nu T_f$ can be denoted as $\boldsymbol{\varsigma}_\nu^L = \left(\varphi_\nu^L, \Delta_\nu^L\right)$, $\nu = 0, 1, \ldots, \infty$, where $\varphi_\nu^L \in \{0, \ldots, N_{\mathrm{PHY}_L} - 1\}$ denotes the combination of TM and channel state in this frame interval and $\Delta_\nu^L \in \{0, 1\}$ is used to denote the *up* or *down* characteristic of the instantaneous SNR in time frame interval $t = (\nu - 1)T_f$.

At any time instant $t = \nu T_f$ the PHY layer states of the S-D and R-D links can be univocally characterized by an integer number $n_\nu^L = 2\varphi_\nu^L + \Delta_\nu^L$ and obviously, $n_\nu^L \in \{0, \ldots, 2N_{\mathrm{PHY}_L} - 1\}$. The PHY layer will be in state $n_L \in \{0, \ldots, 2N_{\mathrm{PHY}_L} - 1\}$ with a steady-state probability $P^{\mathrm{PHY}_L}(n_L)$, which can be calculated using eqs. (3.6) and (2.9), and each of these states will be characterized by a conditional average packet error rate $\overline{PER}_{n_L}^{\mathrm{PHY}_L}$ that can be obtained as in (3.7). Furthermore, the PHY layer FSMC of the R-D and S-D links will be characterized by a transition probability matrix

$$\boldsymbol{P}_s^L = \left[P_s^L(n_\mu^L, n_{\mu'}^L)\right]_{n_\mu^L, n_{\mu'}^L = 0}^{2N_{\mathrm{PHY}_L} - 1}, \tag{5.5}$$

whose elements can be calculated using eqs. (3.9)-(3.12) and (2.16). In this research work, the steady-state probabilities, the conditional average packet error rates and the state-transition probabilities have all been computed either numerically or by simulation.

Cooperative physical layer model

Once the PHY layers of the S-D and R-D links have been modeled, we can proceed to model the whole PHY layer of the proposed cooperative system. To that end, we define the equivalent cooperative PHY layer state at time instant $t = \nu T_f$ as $\boldsymbol{\varsigma}_\nu = \left(\varphi_\nu^S, \Delta_\nu^S, \varphi_\nu^R, \Delta_\nu^R\right)$, $\nu = 0, 1, \ldots, \infty$, where $\varphi_\nu^S \in \{0, \ldots, N_{\mathrm{PHY}_S} - 1\}$ and $\varphi_\nu^R \in \{0, \ldots, N_{\mathrm{PHY}_R} - 1\}$ denote the combination of TM and channel state in this frame interval corresponding to the S-D and R-D links, respectively, while $\Delta_\nu^S \in \{0, 1\}$ and $\Delta_\nu^R \in \{0, 1\}$ denote the *up* or *down* characteristic of the instantaneous SNR in time frame interval $t = (\nu - 1)T_f$ of the S-D and R-D links, respectively. Therefore, the cooperative PHY layer state at any time instant $t = \nu T_f$ can be univocally characterized by $\boldsymbol{n}_\nu = \left(n_\nu^S, n_\nu^R\right)$. As the S-D and R-D PHY layer states are independent, the cooperative PHY layer will be in one of the $4N_{\mathrm{PHY}_S}N_{\mathrm{PHY}_R}$ possible states $\boldsymbol{n} = (n_S, n_R)$ with a steady-state probability

$$P^{\mathrm{PHY}}(\boldsymbol{n}) = P^{\mathrm{PHY}_S}(n_S)P^{\mathrm{PHY}_R}(n_R). \tag{5.6}$$

Additionally, the cooperative PHY layer FSMC will be characterized by a transition probability matrix

$$\boldsymbol{P}_s = \left[P_s(\boldsymbol{n}_\mu, \boldsymbol{n}_{\mu'})\right]_{\boldsymbol{n}_\mu, \boldsymbol{n}_{\mu'} = 0}^{4N_{\mathrm{PHY}_S}N_{\mathrm{PHY}_R} - 1}, \tag{5.7}$$

where

$$P_s(\boldsymbol{n}_\mu, \boldsymbol{n}_{\mu'}) = P_s^S(n_\mu^S, n_{\mu'}^S)P_s^R(n_\mu^R, n_{\mu'}^R). \tag{5.8}$$

5.4 Discrete time Markov chain-based link-level queueing model and analysis

5.4.1 Embedded Markov chain

The queueing process induced by both the CARQ protocol and the AMC scheme can be formulated in discrete time with one time unit equal to one frame interval. The system states are observed at the beginning of each time unit. Let $\sigma_\nu = (a_\nu, \boldsymbol{n}_\nu, \boldsymbol{q}_\nu)$ denote the system state at time instant $t = \nu T_f$, where $a_\nu \in \{0, \ldots, \mathcal{A}-1\}$ represents the phase of the D-BMAP, $\boldsymbol{n}_\nu = \left(n_\nu^S, n_\nu^R\right)$, with $n_\nu^L \in \{0, \ldots, 2N_{\mathrm{PHY}_L} - 1\}$, represents the cooperative PHY layer state, and $\boldsymbol{q}_\nu = \left(q_\nu^S, q_\nu^R\right)$ denotes the queue state, with $q_\nu^S \in \{0, \ldots \overline{Q}_S\}$ and $q_\nu^R \in \{0, \ldots \overline{Q}_R\}$ corresponding to the number of packets in the queues of the source and the relay nodes at this time instant, respectively. If we just look at the set of time instants $t = \nu T_f$, $\nu = 0, 1, \ldots, \infty$, the transitions between states are Markovian. Therefore, an embedded Markov chain can be used to describe the underlying queueing process. The state space of this embedded finite state Markov chain is $\boldsymbol{\mathcal{S}} = \{\mathcal{S}_n\}_{n=1}^{N_s}$ with size $N_s = \mathcal{A} 4 N_{\mathrm{PHY}_S} N_{\mathrm{PHY}_R} (\overline{Q}_S + 1)(\overline{Q}_R + 1)$.

The transition probability from state $\mathcal{S}_\mu = (a_\mu, \boldsymbol{n}_\mu, \boldsymbol{q}_\mu)$ to $\mathcal{S}_{\mu'} = (a_{\mu'}, \boldsymbol{n}_{\mu'}, \boldsymbol{q}_{\mu'})$ can be written as

$$P_{\mathcal{S}_\mu, \mathcal{S}_{\mu'}} = u\left(a_\mu, a_{\mu'}\right) P_s\left(\boldsymbol{n}_\mu, \boldsymbol{n}_{\mu'}\right) P_{\boldsymbol{q}_\mu, \boldsymbol{q}_{\mu'} | a_\mu, \boldsymbol{n}_\mu}, \tag{5.9}$$

where $u\left(a_\mu, a_{\mu'}\right)$ denotes the transition probability between D-BMAP phases a_μ and $a_{\mu'}$, which can be obtained from matrix $\boldsymbol{U}$, $P_s\left(\boldsymbol{n}_\mu, \boldsymbol{n}_{\mu'}\right)$ is the cooperative PHY layer transition probability between states $\boldsymbol{n}_\mu$ and $\boldsymbol{n}_{\mu'}$, which can be derived from matrix $\boldsymbol{P}_s$, and $P_{\boldsymbol{q}_\mu, \boldsymbol{q}_{\mu'} | a_\mu, \boldsymbol{n}_\mu}$ corresponds to the queue transition probability from state $\boldsymbol{q}_\mu$ to state $\boldsymbol{q}_{\mu'}$ when the D-BMAP is in phase a_μ and the cooperative PHY layer state is $\boldsymbol{n}_\mu$.

Let us assume that the system is in state $\mathcal{S}_\mu$. At the beginning of the frame interval the relay node transmits (see the examples of Fig. 5.2)

$$\tau_\mu^R = \min\left\{q_\mu^R, c_{n_\mu^R}\right\} \tag{5.10}$$

packets, where $c_{n_\mu^R}$ denotes the number of packets that can be transmitted by node R in a whole frame interval, when the PHY layer state of the R-D link is n_μ^R. This transmission spans a fraction of the frame period and the remaining fraction will be allocated to the source node, which will transmit

$$\tau_\mu^S = \begin{cases} \min\left\{q_\mu^S, c_{n_\mu^S}\right\} & , \tau_\mu^R = 0 \\ \min\left\{q_\mu^S, \left\lfloor c_{n_\mu^S}\left(1 - \frac{\tau_\mu^R}{c_{n_\mu^R}}\right)\right\rfloor\right\} & , \tau_\mu^R > 0 \end{cases} \tag{5.11}$$

packets, where $\lfloor x \rfloor$ denotes the nearest integer less than or equal to x, with $c_{n_\mu^S}$ being the number of packets that node S can transmit in a whole frame interval when the S-D link PHY layer state is n_μ^S.

Three main cases ought to be considered in order to obtain the transition probabilities, which correspond to the following situations:

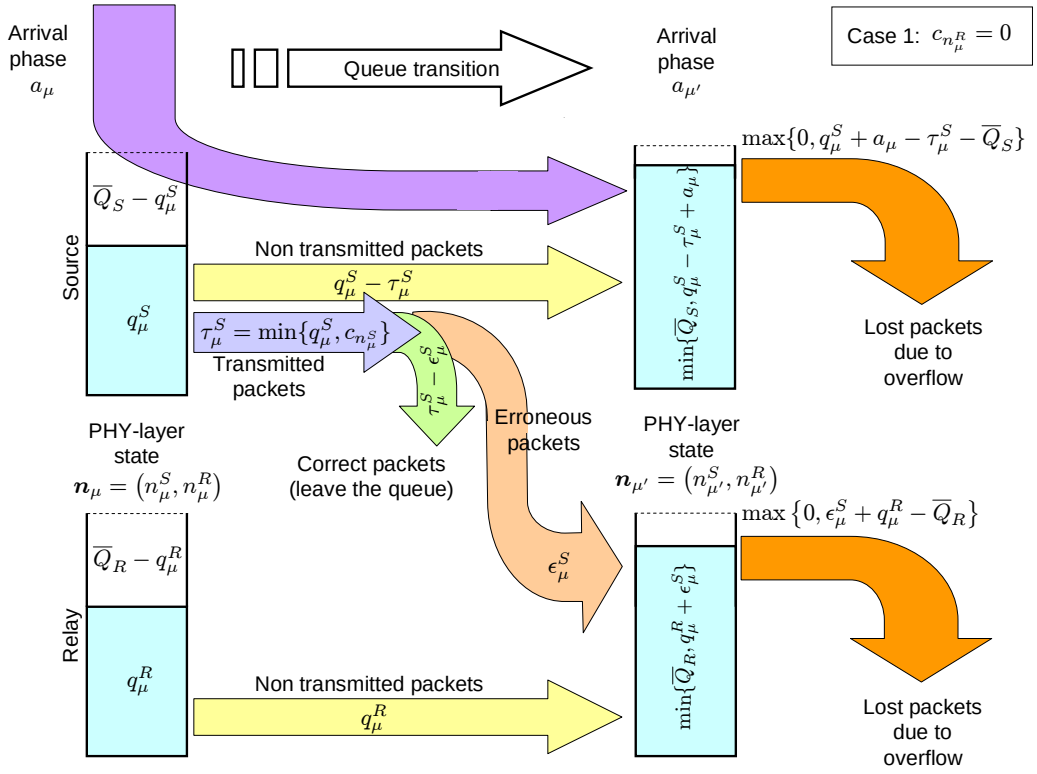


Figure 5.3: Queue transition: *case 1*.

- *Case 1 (Fig. 5.3):* no packets from the relay are transmitted due to bad R-D channel conditions ($c_{n_\mu^R} = 0$).

- *Case 2 (Fig. 5.4):* only packets from the relay can be transmitted ($0 < c_{n_\mu^R} \leq q_\mu^R$).

- *Case 3 (Fig. 5.4):* the system can transmit packets from both, the relay and the source nodes ($c_{n_\mu^R} > q_\mu^R$).

In cases *1* and *3* a certain number of packets from the source are selected for transmission (blue arrow) and, depending on how many of these packets are erroneously transmitted, there exists the possibility of a buffer overflow at the relay. It is obvious that in both cases $q_{\mu'}^R \leq q_\mu^R + \tau_\mu^S$. Moreover, if $q_{\mu'}^R \leq \overline{Q}_R - 1$ holds, no packets will be dropped. On the contrary, if $q_{\mu'}^R = \overline{Q}_R$, the queue of the relay will be full of packets and a buffer overflow may have occurred (orange arrow).

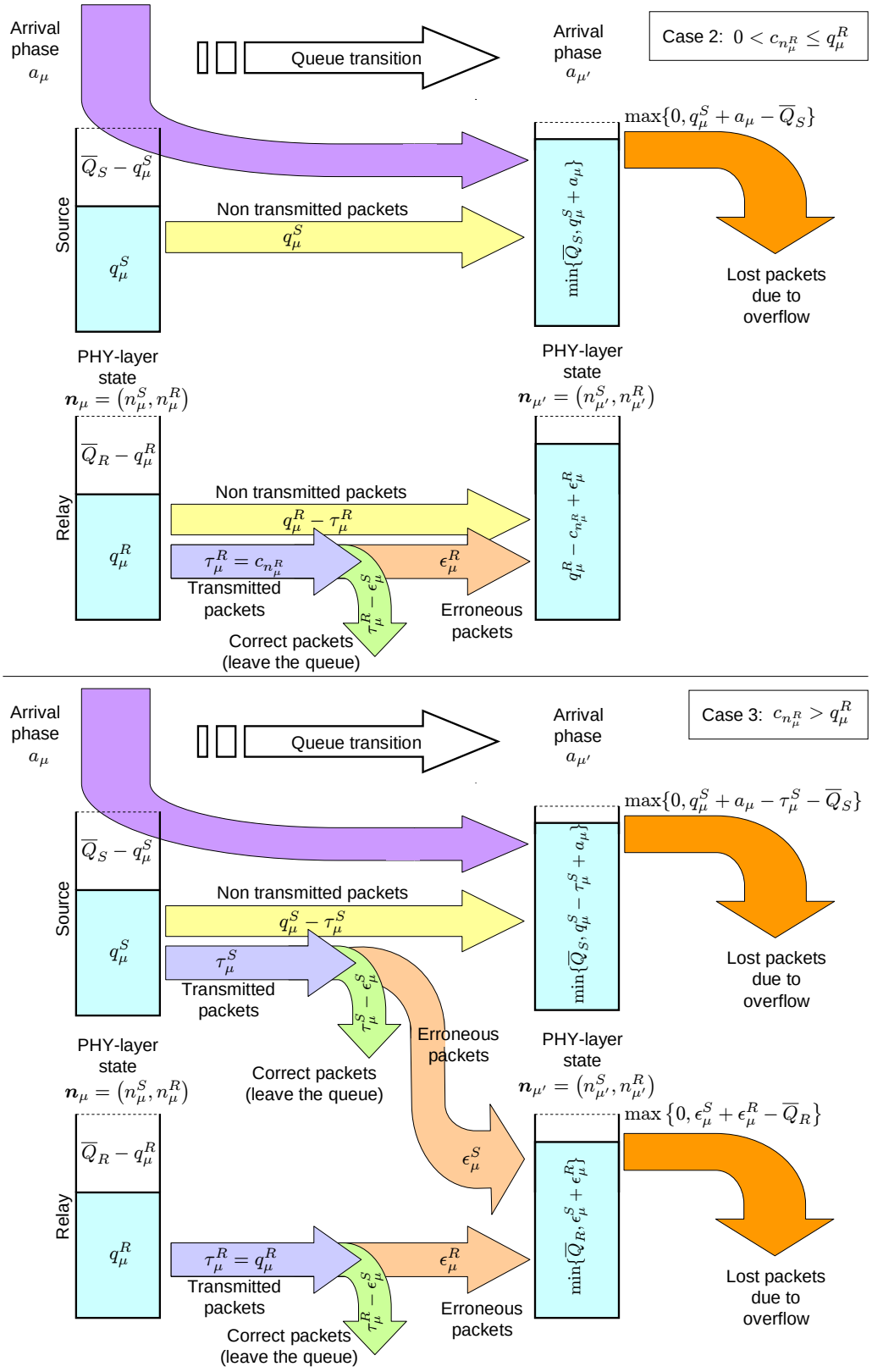


Figure 5.4: Queue transition: *case 2* and *case 3*.

Consequently, using (4.16), the state transition probabilities can be derived as

$$
P_{q_\mu,\boldsymbol{q}_{\mu'}|a_\mu,\boldsymbol{n}_\mu}
= \begin{cases}
\mathcal{P}_{\epsilon_\mu^S}^{\tau_\mu^S}\left(\overline{PER}_{n_\mu^S}^{\mathrm{PHY}_S}\right) & , \begin{array}{l} c_{n_\mu^R}=0 \\ q_{\mu'}^S=\min\left\{\overline{Q}_S, q_\mu^S-\tau_\mu^S+a_\mu\right\} \\ q_{\mu'}^R\le\min\left\{\overline{Q}_R-1, q_\mu^R+\tau_\mu^S\right\} \\ \epsilon_\mu^S=q_{\mu'}^R-q_\mu^R \end{array} \\
\sum_{\epsilon_\mu^S=\epsilon_{\min}^S}^{\epsilon_{\max}^S}\mathcal{P}_{\epsilon_\mu^S}^{\tau_\mu^S}\left(\overline{PER}_{n_\mu^S}^{\mathrm{PHY}_S}\right) & , \begin{array}{l} c_{n_\mu^R}=0 \\ q_{\mu'}^S=\min\left\{\overline{Q}_S, q_\mu^S-\tau_\mu^S+a_\mu\right\} \\ q_{\mu'}^R=\overline{Q}_R\le q_\mu^R+\tau_\mu^S \\ \epsilon_{\min}^S=\overline{Q}_R-q_\mu^R, \epsilon_{\max}^S=\tau_\mu^S \end{array} \\
\mathcal{P}_{\epsilon_\mu^R}^{c_{n_\mu^R}}\left(\overline{PER}_{n_\mu^R}^{\mathrm{PHY}_R}\right) & , \begin{array}{l} 0<c_{n_\mu^R}\le q_\mu^R \\ q_{\mu'}^S=\min\left\{\overline{Q}_S, q_\mu^S+a_\mu\right\} \\ q_{\mu'}^R=q_\mu^R-c_{n_\mu^R}+\epsilon_\mu^R, \forall\epsilon_\mu^R\in\left\{0,\ldots,c_{n_\mu^R}\right\} \end{array} \\
\sum_{\epsilon_\mu^R=\epsilon_{\min}^R}^{\epsilon_{\max}^R}\mathcal{P}_{\epsilon_\mu^R}^{q_\mu^R}\left(\overline{PER}_{n_\mu^R}^{\mathrm{PHY}_R}\right) \times\mathcal{P}_{q_{\mu'}^R-\epsilon_\mu^R}^{\tau_\mu^S}\left(\overline{PER}_{n_\mu^S}^{\mathrm{PHY}_S}\right) & , \begin{array}{l} c_{n_\mu^R}>q_\mu^R \\ q_{\mu'}^S=\min\left\{\overline{Q}_S, q_\mu^S-\tau_\mu^S+a_\mu\right\} \\ q_{\mu'}^R\le\min\left\{\overline{Q}_R-1, q_\mu^R+\tau_\mu^S\right\} \\ \epsilon_{\min}^R=\max\left\{0, q_{\mu'}^R-\tau_\mu^S\right\}, \epsilon_{\max}^R=\min\left\{q_\mu^R, q_{\mu'}^R\right\} \end{array} \\
\sum_{\epsilon_\mu^R=\epsilon_{\min}^R}^{\epsilon_{\max}^R}\sum_{\epsilon_\mu^S=\epsilon_{\min}^S}^{\epsilon_{\max}^S}\mathcal{P}_{\epsilon_\mu^R}^{q_\mu^R}\left(\overline{PER}_{n_\mu^R}^{\mathrm{PHY}_R}\right) \times\mathcal{P}_{\epsilon_\mu^S}^{\tau_\mu^S}\left(\overline{PER}_{n_\mu^S}^{\mathrm{PHY}_S}\right) & , \begin{array}{l} c_{n_\mu^R}>q_\mu^R \\ q_{\mu'}^S=\min\left\{\overline{Q}_S, q_\mu^S-\tau_\mu^S+a_\mu\right\} \\ q_{\mu'}^R=\overline{Q}_R\le q_\mu^R+\tau_\mu^S \\ \epsilon_{\min}^R=\max\left\{0, \overline{Q}_R-\tau_\mu^S\right\}, \epsilon_{\max}^R=q_\mu^R \\ \epsilon_{\min}^S=\overline{Q}_R-\epsilon_\mu^R, \epsilon_{\max}^S=\tau_\mu^S \end{array} \\
0 & , \text{ otherwise}
\end{cases}
\tag{5.12}
$$

The first two lines of (5.12) correspond to case *1*, with the first line describing the situation where the relay buffer is not full, and the second line coinciding with a full buffer (an overflow may have occurred). Next line corresponds to case *2* and the two last lines to case *3*. In this latter case, and similarly to case *1*, lines four and five represent the situations with a not full and a completely full relay queue, respectively.

To derive the system performance measures, the steady-state probability vector needs to be obtained, which can be calculated because the transition probability matrix $\boldsymbol{P}$ and steady-state probability vector $\boldsymbol{\pi}=\left[\pi_{\mathcal{S}_1}\cdots\pi_{\mathcal{S}_{N_s}}\right]$ satisfy $\boldsymbol{\pi P}=\boldsymbol{\pi}$ along with the normalization condition $\boldsymbol{\pi}\mathbf{1}_{N_s}=1$.

5.4.2 Packet loss rate and throughput

A source buffer overflow may occur in all three aforementioned cases, while no packets will be dropped from the buffer of the relay in *case 2*. The number of lost packets due to buffer overflow (orange arrow) when the system state is $\mathcal{S}_\mu$ is

given by

$$N_{l_{BO}|\mathcal{S}_\mu}^S = \begin{cases} \max\{0, q_\mu^S + a_\mu - \tau_\mu^S - \overline{Q}_S\} & , c_{n_\mu^R} = 0 \\ \max\{0, q_\mu^S + a_\mu - \overline{Q}_S\} & , 0 < c_{n_\mu^R} \leq q_\mu^R \\ \max\{0, q_\mu^S + a_\mu - \tau_\mu^S - \overline{Q}_S\} & , c_{n_\mu^R} > q_\mu^R \end{cases} \tag{5.13}$$

for the source node, and by

$$N_{l_{BO}|\mathcal{S}_\mu}^R = \begin{cases} \sum\limits_{\epsilon_\mu^S=\epsilon_{\min}^S}^{\epsilon_{\max}^S} \max\left\{0, \epsilon_\mu^S + q_\mu^R - \overline{Q}_R\right\} \\ \quad \times \mathcal{P}_{\epsilon_\mu^S}^{\tau_\mu^S}\left(\overline{PER}_{n_\mu^S}^{\mathrm{PHY}_S}\right) & \begin{array}{l} c_{n_\mu^R} = 0 \\ , \ \tau_\mu^S > \overline{Q}_R - q_\mu^R \\ \epsilon_{\min}^S = \overline{Q}_R - q_\mu^R + 1, \epsilon_{\max}^S = \tau_\mu^S \end{array} \\ \sum\limits_{\epsilon_\mu^R=\epsilon_{\min}^R}^{\epsilon_{\max}^R} \sum\limits_{\epsilon_\mu^S=\epsilon_{\min}^S}^{\epsilon_{\max}^S} \max\left\{0, \epsilon_\mu^S + \epsilon_\mu^R - \overline{Q}_R\right\} \\ \quad \times \mathcal{P}_{\epsilon_\mu^R}^{q_\mu^R}\left(\overline{PER}_{n_\mu^R}^{\mathrm{PHY}_R}\right) \mathcal{P}_{\epsilon_\mu^S}^{\tau_\mu^S}\left(\overline{PER}_{n_\mu^S}^{\mathrm{PHY}_S}\right) & \begin{array}{l} c_{n_\mu^R} > q_\mu^R \\ , \ \tau_\mu^S > \overline{Q}_R - q_\mu^R \\ \epsilon_{\min}^R = \max\left\{0, \overline{Q}_R - \tau_\mu^S + 1\right\}, \epsilon_{\max}^R = q_\mu^R \\ \epsilon_{\min}^S = \overline{Q}_R - \epsilon_\mu^R + 1, \epsilon_{\max}^S = \tau_\mu^S \end{array} \end{cases} \tag{5.14}$$

for the relay. Therefore, the average number of lost packets due to buffer overflow can be calculated as

$$\bar{N}_{l_{BO}}^L = \sum_{\mu=1}^{N_s} \pi_{\mathcal{S}_\mu} N_{l_{BO}|\mathcal{S}_\mu}^L \tag{5.15}$$

with $L = S$ denoting the S-D link and $L = R$ denoting the R-D link, and the packet loss rate $P_{l_{BO}}^L$ (measured in packets per frame) can then be obtained as

$$P_{l_{BO}}^L = \bar{N}_{l_{BO}}^L / \lambda. \tag{5.16}$$

In the proposed CARQ-based error control system, the packet loss rate P_l can be expressed as

$$P_l = P_{l_{BO}}^S + P_{l_{BO}}^R, \tag{5.17}$$

and given the packet loss rate P_l, the average throughput can be calculated as

$$\eta = \lambda(1 - P_l). \tag{5.18}$$

5.4.3 Average queue length and average packet delay

The average transmission delay is an important measure to characterize the system performance, which is defined as the average delay a packet received by the destination has experienced since its transmission from the source. By capitalizing on Little's theorem (Kleinrock and Lam), and taking into consideration that the effective packet arrival rate at the source is equal to

$$\lambda_S = \lambda(1 - P_{l_{BO}}^S), \tag{5.19}$$

the average delay corresponding to packets correctly transmitted by node S can be calculated as

$$D_l^S = \mathcal{L}_q^S / \lambda(1 - P_{l_{BO}}^S), \tag{5.20}$$

where $\mathcal{L}_q^S$ denotes the average number of packets in the queue of the source. Taking into account that the packets that have been erroneously transmitted by node S will be queued in the buffer of node R in order to be retransmitted, the effective packet arrival rate at the relay can be calculated as

$$\lambda_R = \lambda(1 - P_{l_{BO}}^S)\overline{PER}^S(1 - P_{l_{BO}}^R), \tag{5.21}$$

where $\overline{PER}^S$ corresponds to the average PER of the AMC scheme of the S-D link, and can be obtained using eqs. (3.2, 3.3, 3.43, 3.44). Therefore, the average delay experienced by packets transmitted through node R is equal to the average waiting time in the relay buffer plus one additional time interval accounting for the packet transmission by the source itself. In this regard, we can determine the average waiting time as

$$D_l^R = \frac{\mathcal{L}_q^R}{\lambda(1 - P_{l_{BO}}^S)\overline{PER}^S(1 - P_{l_{BO}}^R)} + T_f, \tag{5.22}$$

where $\mathcal{L}_q^R$ denotes the average number of packets in the queue of the relay. The average queue lengths can be calculated from the steady-state probabilities as

$$\mathcal{L}_q^L = \sum_{\mu=1}^{N_s} \pi_{\mathcal{S}_\mu} q_\mu^L. \tag{5.23}$$

The average number of correctly transmitted packets through links S-D and R-D, denoted as N_t^L, can be obtained as

$$\begin{aligned} N_t^S &= \lambda(1 - P_{l_{BO}}^S)(1 - \overline{PER}^S) \\ N_t^R &= \lambda(1 - P_{l_{BO}}^S)\overline{PER}^S(1 - P_{l_{BO}}^R), \end{aligned} \tag{5.24}$$

The overall average packet delay for our embedded Markov chain can then be calculated as

$$\begin{aligned} D_l &= \frac{N_t^S D_l^S + N_t^R(D_l^S + D_l^R)}{N_t^S + N_t^R} \\ &= D_l^S + \frac{\overline{PER}^S(1 - P_{l_{BO}}^R)}{1 - \overline{PER}^S P_{l_{BO}}^R} D_l^R. \end{aligned} \tag{5.25}$$

5.5 Cross-layer optimization

5.5.1 Multidimensional approach

As shown in previous sections, given the maximum buffer sizes $\overline{Q}_S$ and $\overline{Q}_R$, the average SNRs $\bar{\gamma}_S$ and $\bar{\gamma}_R$, and a normalized maximum Doppler frequency $f_d T_f$, performance measures of the system like, for instance, throughput, average packet delay or packet loss rate, are a function of the AMC transmission mode switching levels $\mathbf{\Gamma}^{m_L} \in \mathbb{R}_+^{M_L+1}$, where $\mathbb{R}_+$ denotes the set of non-negative real numbers, and the measured or estimated arrival packet rate $\lambda \in \mathbf{\Theta}$, where $\mathbf{\Theta}$ is the range

of feasible arrival rate values. Therefore, if the objective of the link adaptation scheme is to maximize the average throughput when supporting QoS-guaranteed traffic characterized by a maximum packet loss rate $P_{l\max}$ and a maximum average packet delay $D_{l\max}$, the system needs to jointly optimize the selected protocol parameters at the PHY and DLC layers by solving the cross-layer optimization problem

$$\left(\mathbf{\Gamma}^{m_S}_{\text{opt}}, \mathbf{\Gamma}^{m_R}_{\text{opt}}, \lambda_{\text{opt}}\right) = \arg \max_{\mathbf{\Gamma}^{m_S}\in\mathbb{R}_+^{M_S+1}, \mathbf{\Gamma}^{m_R}\in\mathbb{R}_+^{M_R+1}, \lambda\in\mathbf{\Theta}} \eta\left(\mathbf{\Gamma}^{m_S}, \mathbf{\Gamma}^{m_R}, \lambda\right) \tag{5.26}$$

subject to the constraints

$$\begin{aligned} &P_l\left(\mathbf{\Gamma}^{m_S}, \mathbf{\Gamma}^{m_R}, \lambda\right) \leq P_{l\max}, \\ &D_l\left(\mathbf{\Gamma}^{m_S}, \mathbf{\Gamma}^{m_R}, \lambda\right) \leq D_{l\max}. \end{aligned} \tag{5.27}$$

The analytical expressions for η, P_l and D_l do not leave much room for developing efficient algorithms to solve this constrained optimization problem. However, considering that $\mathbf{\Gamma}^{m_S}$, $\mathbf{\Gamma}^{m_R}$ and λ lie in a bounded space $\mathbb{R}_+^{M_S+1} \times \mathbb{R}_+^{M_R+1} \times \mathbf{\Theta}$, the proposed cross-layer optimization problem can be numerically solved through a multidimensional exhaustive search.

5.5.2 Tridimensional simplification approach

In order to simplify this multidimensional optimization approach, a simplified tridimensional optimization is proposed. Although two possibilities can be considered for this simplification (the *ergodic* option and the *instantaneous* option), and because as shown in Section 3.5.2 a similar optimum performance is observed for both schemes and just a slight enhancement is obtained with the *instantaneous* option, only the latter approach will be considered.

Assuming the use of TM n, the instantaneous PER at the output of the soft Viterbi decoder can be approximated as in (3.3). In most previous works on this topic (see e.g., (Liu et al., 2004)), the AMC switching thresholds have been obtained as the instantaneous SNR values for which $PER_n = P_0 \leq 1$, that is,

$$\gamma_n^m = \begin{cases} \frac{1}{g_n}\ln\left(\frac{a_n}{P_0}\right) & , P_0 < 1 \\ \gamma_{p_n} & , P_0 = 1 \end{cases} \tag{5.28}$$

for $n = 1, \ldots, M-1$, with $\gamma_0^m = 0$, and $\gamma_M^m = \infty$. This methodology limits the minimum achievable threshold values to γ_{p_n}. In order to increase the dynamic range of the AMC achievable switching thresholds, a parameter $V_0 \in \mathbb{R}_+$ is defined, and the γ_n^m values will be obtained using the methodology summarized in the following threshold searching algorithm, and illustrated with an example in Fig. 5.5:

1. Set $M = 0$, $\boldsymbol{\mathcal{M}} = \emptyset$, $n = M_p$, $\gamma_0^m = 0$, $\gamma_n^m = \infty$.
2. $n \leftarrow n - 1$
   If $n = 0$ go to step 3, otherwise search the unique $\gamma_n^m \in [0, \gamma_{n+1}^m)$ that satisfies $\gamma_n^m = \frac{1}{g_n} ln\left(\frac{a_n}{V_0}\right)$.

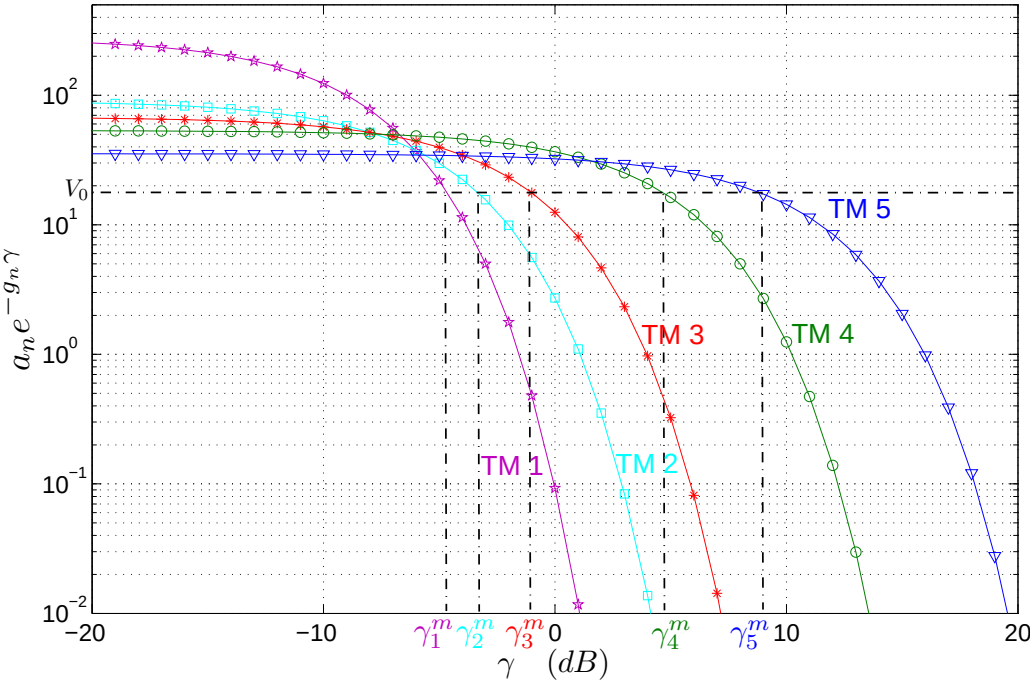


Figure 5.5: AMC switching thresholds searching algorithm.

- In case $0 \leq \gamma_n^m \leq \gamma_{n+1}^m$ update $M \leftarrow M+1$ and $\boldsymbol{\mathcal{M}} \leftarrow \{n\} \cup \{\boldsymbol{\mathcal{M}}\}$ and go to step 2.
- In case it does not exist and $V_0 < a_n$, then: declare TM n as *useless*, update threshold subindices as $\gamma_{i-1}^m \leftarrow \gamma_i^m$, $i = n+1, \ldots, n+1+M$ and go to step 2.
- In case it does not exist and $V_0 > a_n$, TM n is the last *useful* mode and, thus, update $M \leftarrow M+1$, $\boldsymbol{\mathcal{M}} \leftarrow \{n\} \cup \{\boldsymbol{\mathcal{M}}\}$ and threshold subindices $\gamma_{i-n}^m \leftarrow \gamma_i^m$, $i = n+1, \ldots, n+M$ and stop the searching algorithm.

3. TM 0 is the lowest used TM and, thus, update $M \leftarrow M+1$ and $\boldsymbol{\mathcal{M}} \leftarrow \{0\} \cup \{\boldsymbol{\mathcal{M}}\}$ and stop the searching algorithm.

This threshold searching algorithm must be applied to both, the S-D and the R-D links to obtain their AMC switching thresholds. Using this approach, the simplified constrained optimization problem can be formulated as

$$\left(V_{0\,\text{opt}}^S, V_{0\,\text{opt}}^R, \lambda_{\text{opt}}\right) = \arg \max_{V_0^S \in \mathbb{R}_+, V_0^R \in \mathbb{R}_+, \lambda \in \boldsymbol{\Theta}} \eta\left(V_0^S, V_0^R, \lambda\right) \tag{5.29}$$

subject to the constraints

$$\begin{aligned} P_l\left(V_{0\,\text{opt}}^S, V_{0\,\text{opt}}^R, \lambda_{\text{opt}}\right) &\leq P_{l\max}, \\ D_l\left(V_{0\,\text{opt}}^S, V_{0\,\text{opt}}^R, \lambda_{\text{opt}}\right) &\leq D_{l\max}. \end{aligned} \tag{5.30}$$

Obviously, because V_0^S, V_0^R and λ lie in a bounded space $\mathbb{R}_+ \times \mathbb{R}_+ \times \boldsymbol{\Theta}$, we can resort to a 3-D exhaustive search to numerically solve the proposed cross-layer

optimization problem. The tridimensional optimization approach has less degrees of freedom than the multidimensional strategy, resulting in a simpler scheme at the cost of a reduction in performance.

5.6 Numerical results

In this section, selected numerical results are presented to illustrate the performance of the proposed adaptive multi-rate CARQ wireless system, with the following main objectives:

1. the validation and analysis of the 2D-FSMC PHY/DLC modeling, which is based on the cooperative PHY layer model introduced in Subsection 5.3.2
2. the analysis of the cross-layer design proposals of Section 5.5

In order to verify the validity and accuracy of the approximations used in the Markovian approach, presented results, which are obtained by evaluating the analytical expressions derived in previous sections, are also confronted with computer simulation results obtained using Clarke's statistical Rayleigh model (Clarke, 1968). Finally, the cross-layer design framework is used to obtain PHY-DLC system parameters that maximize the overall system throughput under the constraint of a QoS-guaranteed service. Furthermore, to evaluate the potential benefits of the CARQ protocols, the performance of cooperative ARQ-based schemes is compared with their conventional non-cooperative ARQ counterparts over S-D channels.

Unless otherwise specified, both analytical and simulation results are obtained using the following default parameters: a normalized maximum Doppler frequency $f_d T_f = 0.02$, an average received SNR of the S-D channel $\bar{\gamma}_S = 8$ dB, source and relay buffer sizes $\overline{Q}_S = 15$ and $\overline{Q}_R = 10$, respectively, a number of channel states $K = 5$, a parameter $b = 2$ and a D-BMAP with $\mathcal{A} = 10$ phases and characterized by the transition probability matrix

$$\boldsymbol{U} = \begin{bmatrix} u_0 & u_1 & u_2 & \cdots & u_2 \\ u_2 & u_0 & u_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & u_2 \\ u_2 & \cdots & u_2 & u_0 & u_1 \\ u_2 & \cdots & u_2 & u_2 & u_3 \end{bmatrix}, \tag{5.31}$$

with $u_0 = 0.25$, $u_1 = 0.7$, $u_2 = 0.0063$ and $u_3 = 0.9437$, corresponding to an average arrival rate $\lambda = 7.84$ packets per frame. For both the S and R nodes, the five AMC modes listed in Table 3.1, adopted from the IEEE 802.11a standard, are used in the adaptive modulation and coding pool.

Without loss of generality and for the sake of simplicity, the S, R and D nodes are assumed to lie on a straight line, where the S-D distance is normalized to unity, and the S-R and R-D distances are set to d and $1-d$, respectively. In this case, the average SNR of the S-D and R-D channels are $\bar{\gamma}_S$ and $\bar{\gamma}_R = \bar{\gamma}_S(1-d)^{-\alpha}$ (Mardani et al., 2011), respectively, where α denotes the path loss exponent, which is set to 4 in the following experiments.

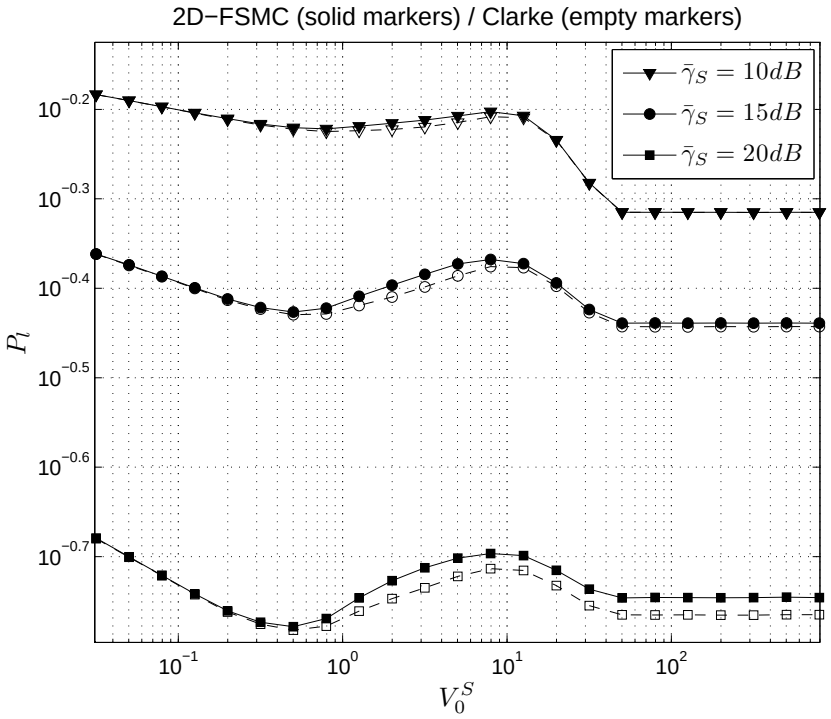


Figure 5.6: Average packet loss rate P_l vs. V_0^S.

5.6.1 Performance results for the AMC/CARQ scheme

In order to analyze the behaviour of the proposed cooperative system in terms of V_0^S, Figs. 5.6 to 5.12 reveal the dependence on this parameter of all the performance measures, that is, average packet loss rate P_l (Fig. 5.6) and its components $P_{l_{BO}}^S$ (Fig. 5.7) and $P_{l_{BO}}^R$ (Fig. 5.8), average throughput η (Fig. 5.9), average packet delay D_l (Fig. 5.10), and average queue lengths at both, the source node $\mathcal{L}_q^S$ (Fig. 5.11) and the relay node $\mathcal{L}_q^R$ (Fig. 5.12). V_0^R has been set to its optimum value, which corresponds to ${V_0^R}_{\text{opt}} = 10^{-0.5}$ for $\bar{\gamma}_S = 10$ dB and 15 dB. Notice that an optimum V_0^R value for $\bar{\gamma}_S = 20$ dB does not exist, in other words, the throughput does not depend on V_0^R for such a high SNR, as it will be discussed when commenting Fig. 5.16. The presented curves correspond to different $\bar{\gamma}_S$ values with $d = 0.5$, and have been obtained using both, the proposed analytical model and the real system simulation. It can be clearly observed that higher V_0^S values bring along the utilization of higher order TMs at the source, causing a growth of the queueing service rate and, consequently, a decrease in $\mathcal{L}_q^S$ and $P_{l_{BO}}^S$. However, after reaching a (relative) minimum, increased V_0^S values imply the use of TMs that are not robust enough, which in turn implies that more packets need to be retransmitted thus leading to a growth of the number of packets in the buffer of the relay $\mathcal{L}_q^R$. Consequently, the remaining time after node R has transmitted its packets decreases, the effective rate of packets leaving the queue of node S diminishes and, therefore, the average source queue length $\mathcal{L}_q^S$ and the average number of lost packets due to the source buffer overflow $P_{l_{BO}}^S$ grow. Nevertheless,

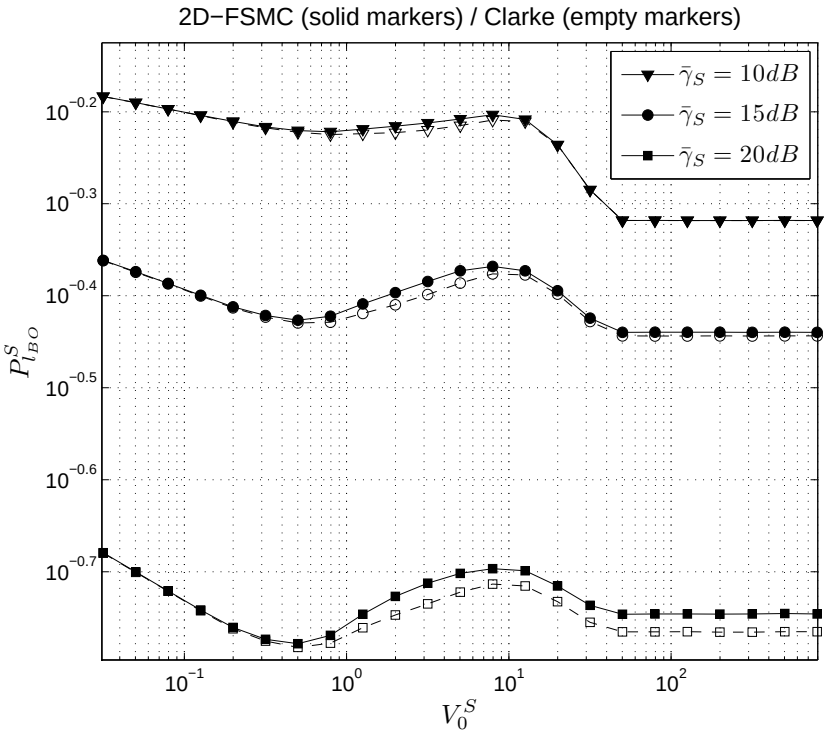


Figure 5.7: Average packet loss rate $P^S_{l_{BO}}$ vs. V_0^S.

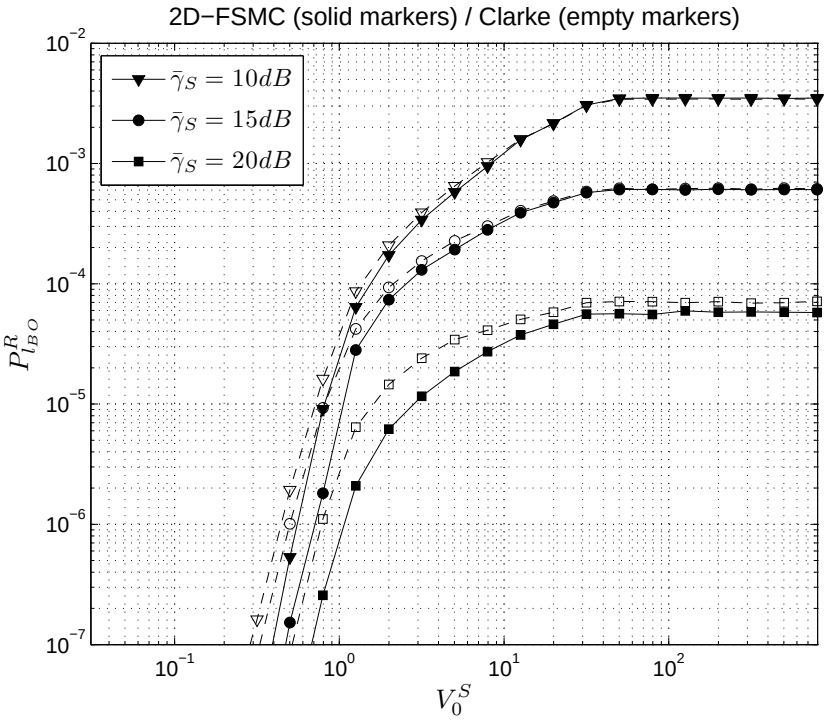


Figure 5.8: Average packet loss rate $P^R_{l_{BO}}$ vs. V_0^S.

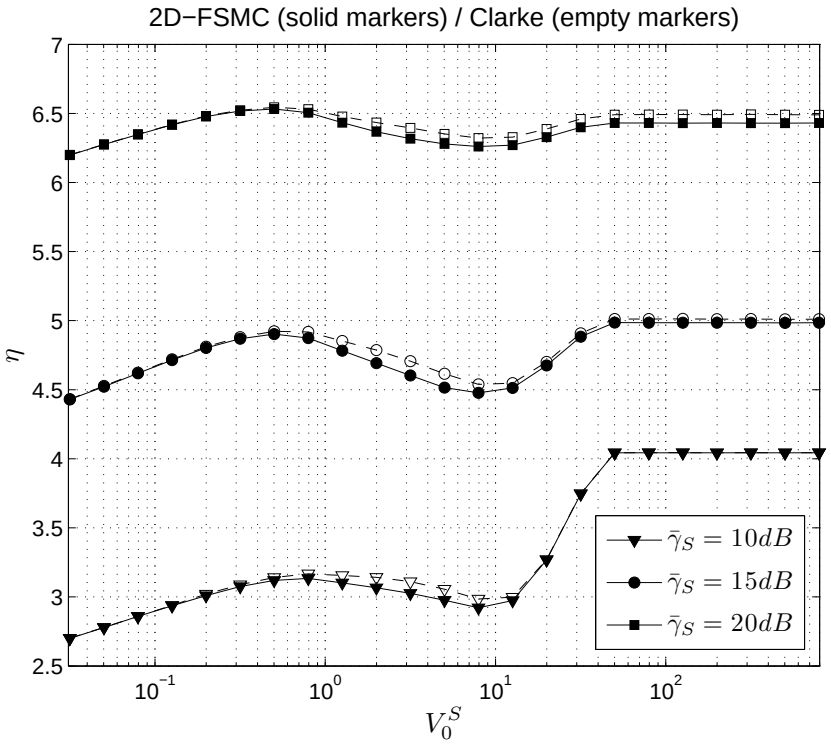


Figure 5.9: Average throughput η vs. V_0^S.

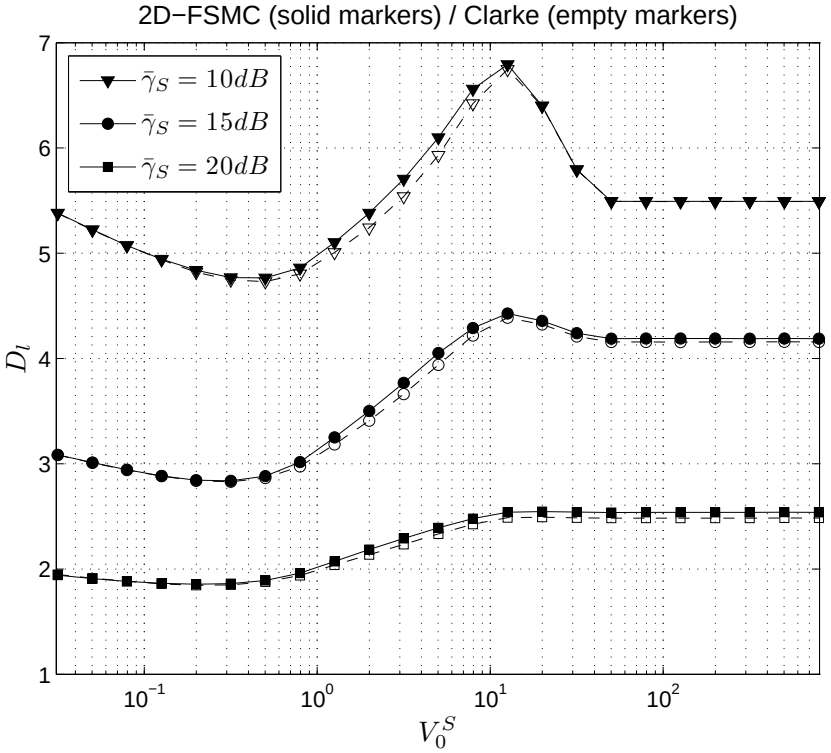


Figure 5.10: Average packet delay D_l vs. V_0^S.

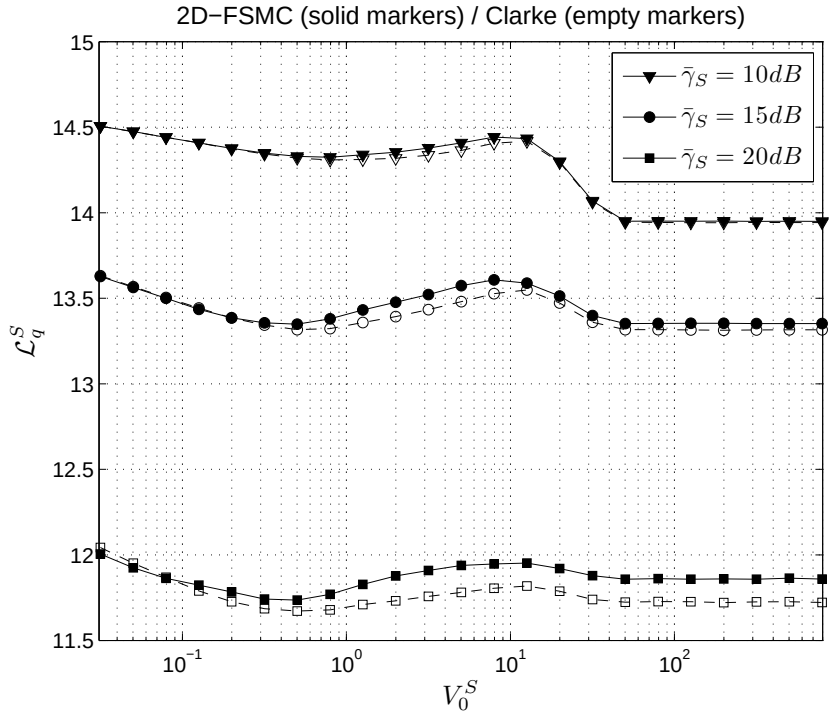


Figure 5.11: Average queue length $\mathcal{L}_q^S$ vs. V_0^S.

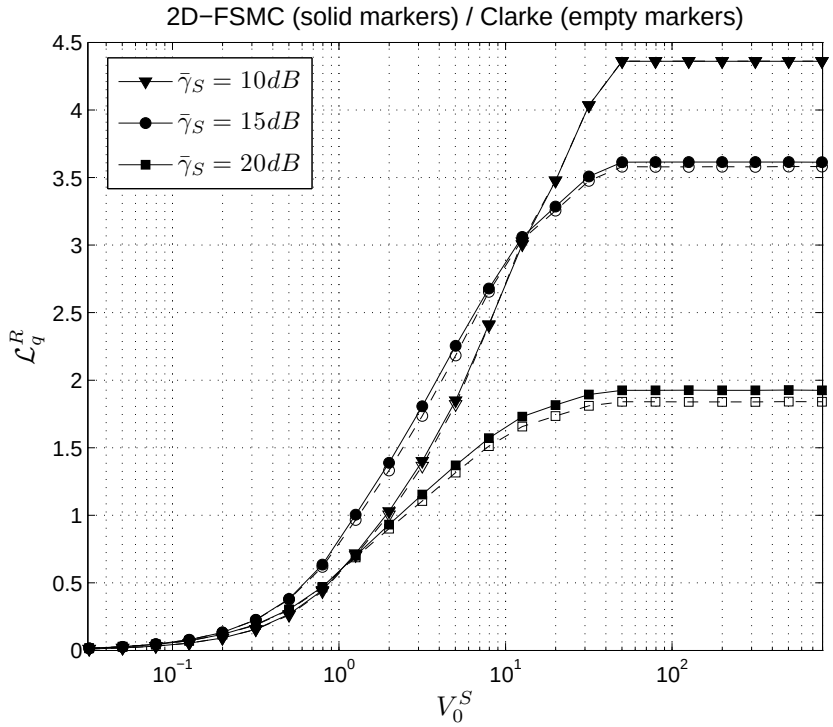


Figure 5.12: Average queue length $\mathcal{L}_q^R$ vs. V_0^S.

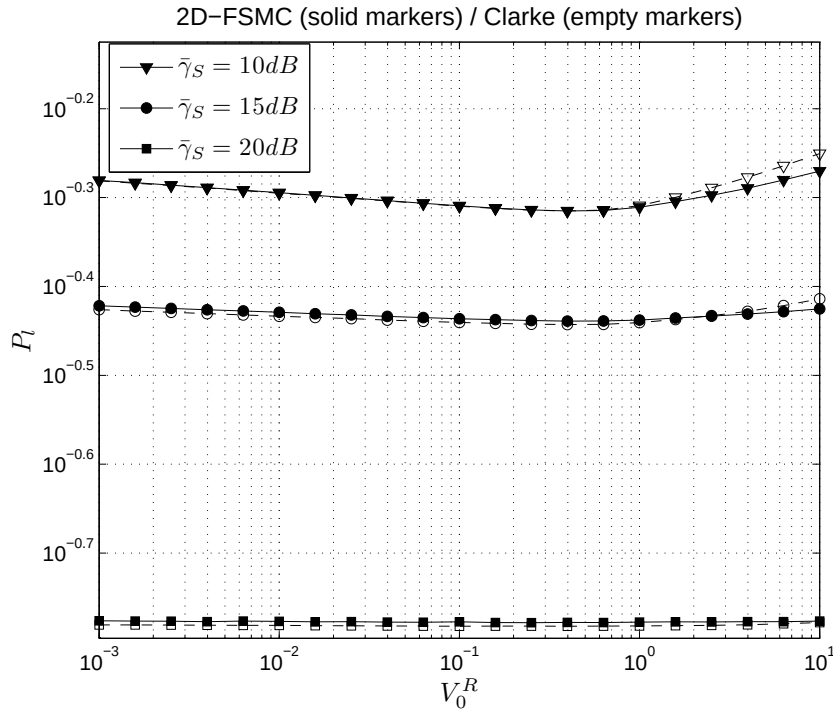


Figure 5.13: Average packet loss rate P_l vs. V_0^R.

as V_0^S keeps increasing and the source transmitter uses higher order TMs, more packets per frame leave its queue, $\mathcal{L}_q^S$ diminishes and, thus, $P_{l_{BO}}^S$ decreases again, until it reaches the situation in which the only *useful* TM is the highest possible TM (i.e., TM 5) and the curves flatten. As higher V_0^S values bring along the utilization of less robust (i.e., higher order) TMs at the source, it can be observed an increase in the number of packets that need to be retransmitted and, consequently, an increase in $\mathcal{L}_q^R$ and $P_{l_{BO}}^R$, until the only *useful* TM is TM 5 and both curves flatten. As Fig. 5.6 shows, and comparing figures corresponding to both $P_{l_{BO}}^S$ and $P_{l_{BO}}^R$, it is clear that the packet loss due to buffer overflow at the source $P_{l_{BO}}^S$ is the dominant packet loss factor. As Fig. 5.9 reveals, the average throughput η is in agreement with the behaviour of both $P_{l_{BO}}^S$ and $P_{l_{BO}}^R$. As expected, higher SNR values imply an enhancement in the system behaviour, i.e. greater throughput and reduced delay. It is worth noting that the optimum throughput corresponding to $\bar{\gamma}_S = 10$ dB and 15 dB is obtained when only the highest TM is used at the source, implying that a huge amount of packets are erroneously received and, thus, the relay is responsible of its correct retransmission to the destination. Nevertheless, for $\bar{\gamma}_S = 20$ dB, the optimum throughput corresponds to a V_0^S value that implies the availability at the source of not only TM 5 but also more robust TMs, which means that most packets will be correctly transmitted from the source to the destination without the aid of the relay.

Figures 5.13 to 5.19 plot, respectively, the dependence on V_0^R of P_l, $P_{l_{BO}}^S$, $P_{l_{BO}}^R$, η, D_l, $\mathcal{L}_q^S$ and $\mathcal{L}_q^R$, when V_0^S is set to its optimum value, which corresponds to ${V_0^S}_{\text{opt}} = 10^{1.9}$ for $\bar{\gamma}_S = 10$ dB and 15 dB, while ${V_0^S}_{\text{opt}} = 10^{-0.45}$ for $\bar{\gamma}_S = 20$ dB.

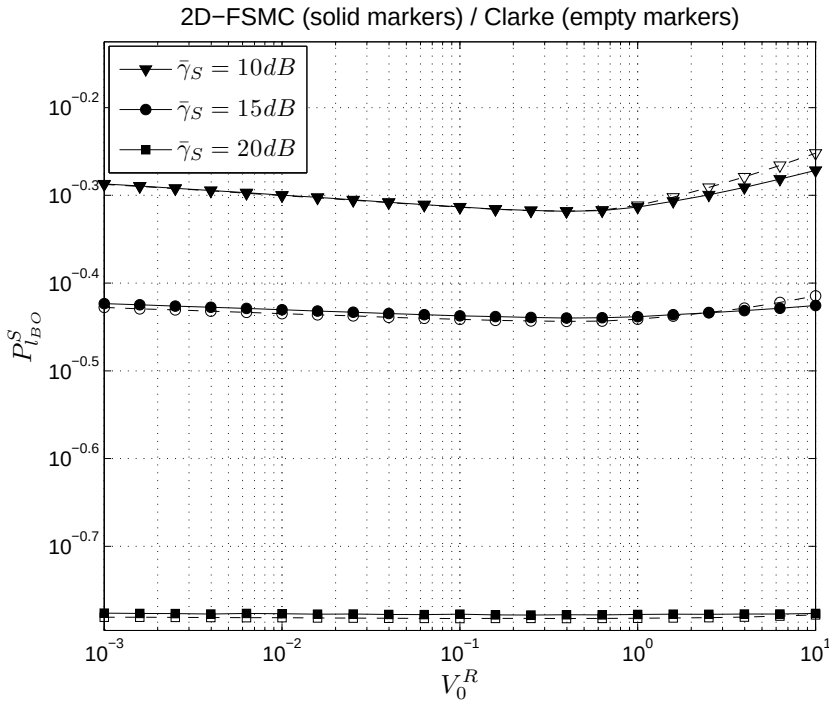


Figure 5.14: Average packet loss rate $P^S_{l_{BO}}$ vs. V_0^R.

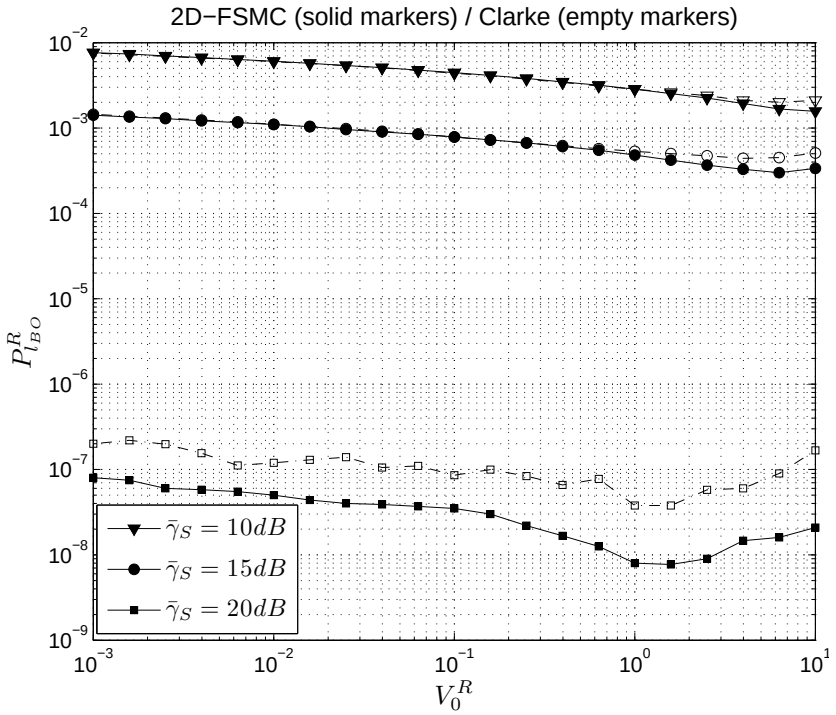


Figure 5.15: Average packet loss rate $P^R_{l_{BO}}$ vs. V_0^R.

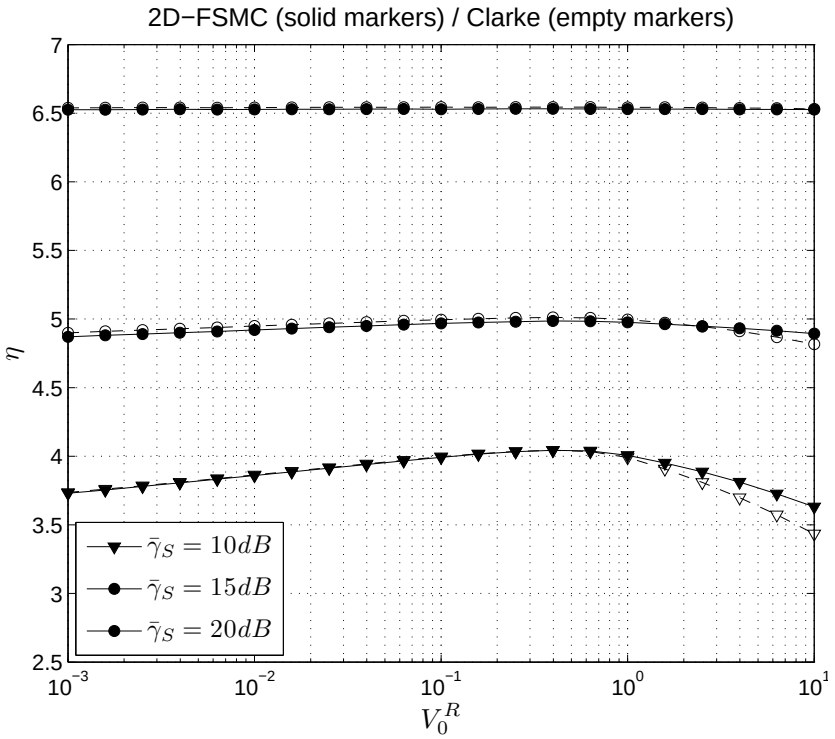


Figure 5.16: Average throughput η vs. V_0^R.

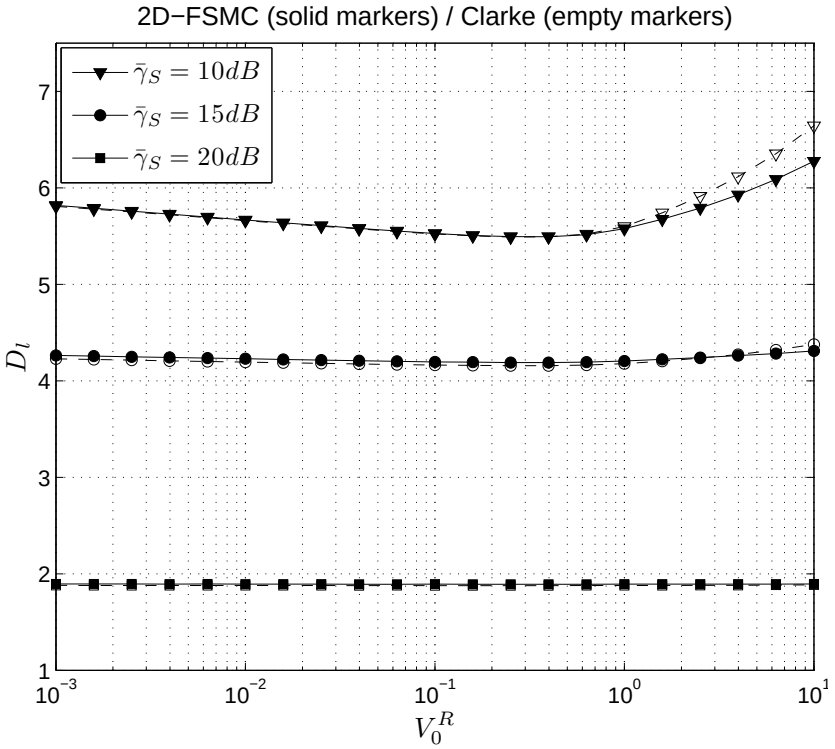


Figure 5.17: Average packet delay D_l vs. V_0^R.

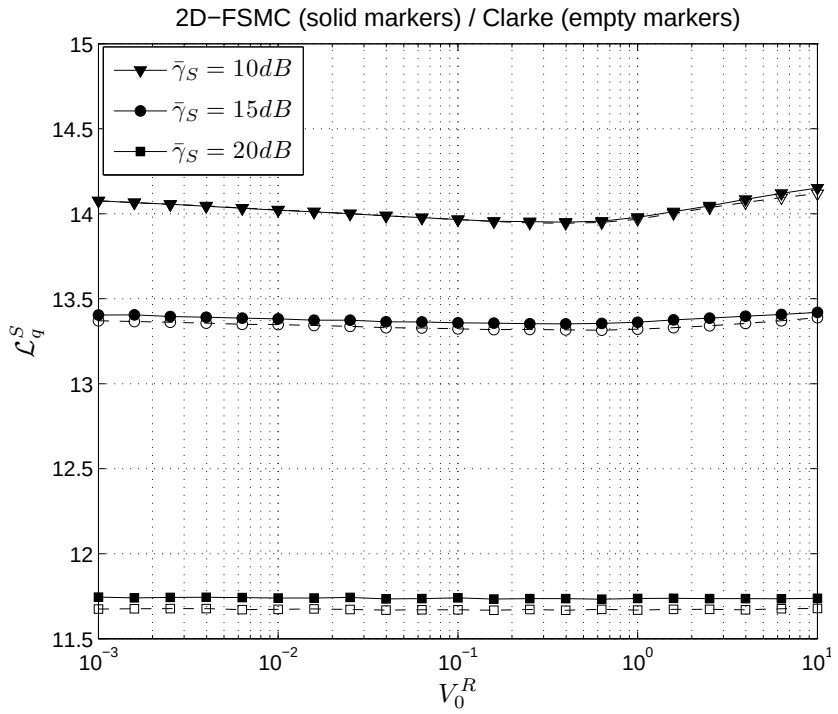


Figure 5.18: Average queue length $\mathcal{L}_q^S$ vs. V_0^R.

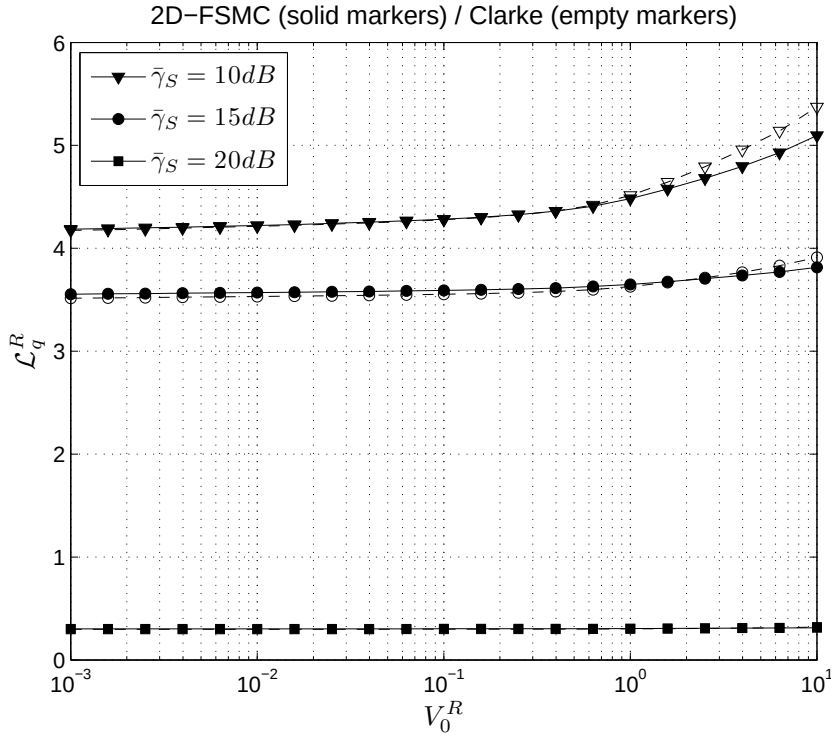


Figure 5.19: Average queue length $\mathcal{L}_q^R$ vs. V_0^R.

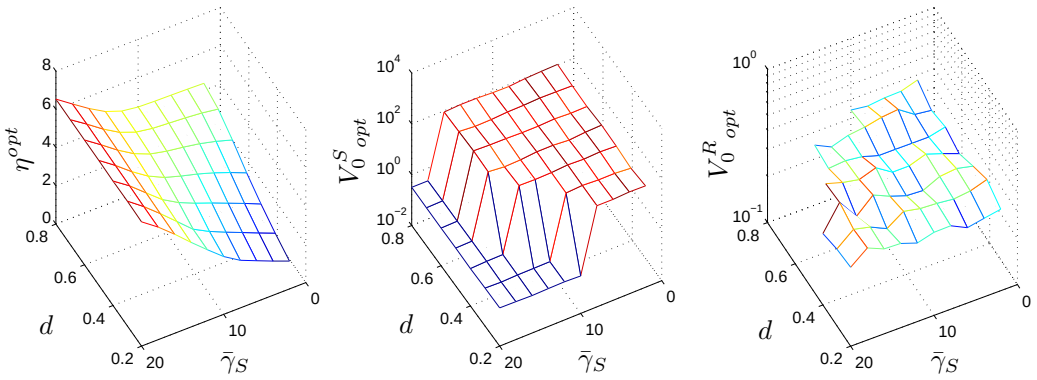


Figure 5.20: Throughput optimization vs. $\bar{\gamma}_S$ and d.

Greater V_0^R values lead to the use of higher order TMs at node R, which in turn imply a faster service rate and, consequently, a reduction in $P_{l_{BO}}^R$. Note, however, that these less robust TMs entail a growth of the number of required retransmissions and also of the average relay buffer size $\mathcal{L}_q^R$. As a result, when the increase of the service rate cannot cope with the huge number of required retransmissions, $P_{l_{BO}}^R$ raises accordingly. It can be clearly observed that higher V_0^R values also reduce $\mathcal{L}_q^S$ and $P_{l_{BO}}^S$ whenever the selected TMs by the relay are robust enough and not too many retransmissions are required, thus leaving sufficient time for the source to transmit its packets. However, as higher order TMs are being selected by node R, more packets need to be retransmitted, the remaining time for node S transmission is shorter and $\mathcal{L}_q^S$ and $P_{l_{BO}}^S$ take off. Moreover, it has to be pointed out that higher SNR values, which imply that the source plays a more important role on the correct transmission of the packets, entail a lower dependence of the performance of the system on V_0^R. Figures 5.16 and 5.17 confirm that higher SNR values imply an improvement in the average throughput η and delay D_l.

Results presented in Figs. 5.6 to 5.19 reveal that, in all cases, the behaviour of a real FIFO queuing system using a CARQ protocol is faithfully reproduced by the proposed analytical Markov model.

5.6.2 Cross-layer design. Cooperative versus conventional schemes

With the objective of analyzing the cross-layer design proposed in Section 5.5, two optimization problems are considered. In the first one, the dependence of the optimum average throughput on the average received SNR and the distance between nodes S and R, for the tridimensional optimization approach, has been obtained and presented in Fig. 5.20, where the optimum values $V_{0\ \mathrm{opt}}^S$ and $V_{0\ \mathrm{opt}}^R$ have been also depicted. For clarity and completeness of this figure, we have set the average arrival rate using the D-BMAP characterized by the transition probability matrix $\boldsymbol{U}$ and only the parameters V_0^S and V_0^R have been optimized. Additionally, no constraints on P_l and D_l have been set as otherwise throughput curves would only appear in the range of values fulfilling these constraints, making

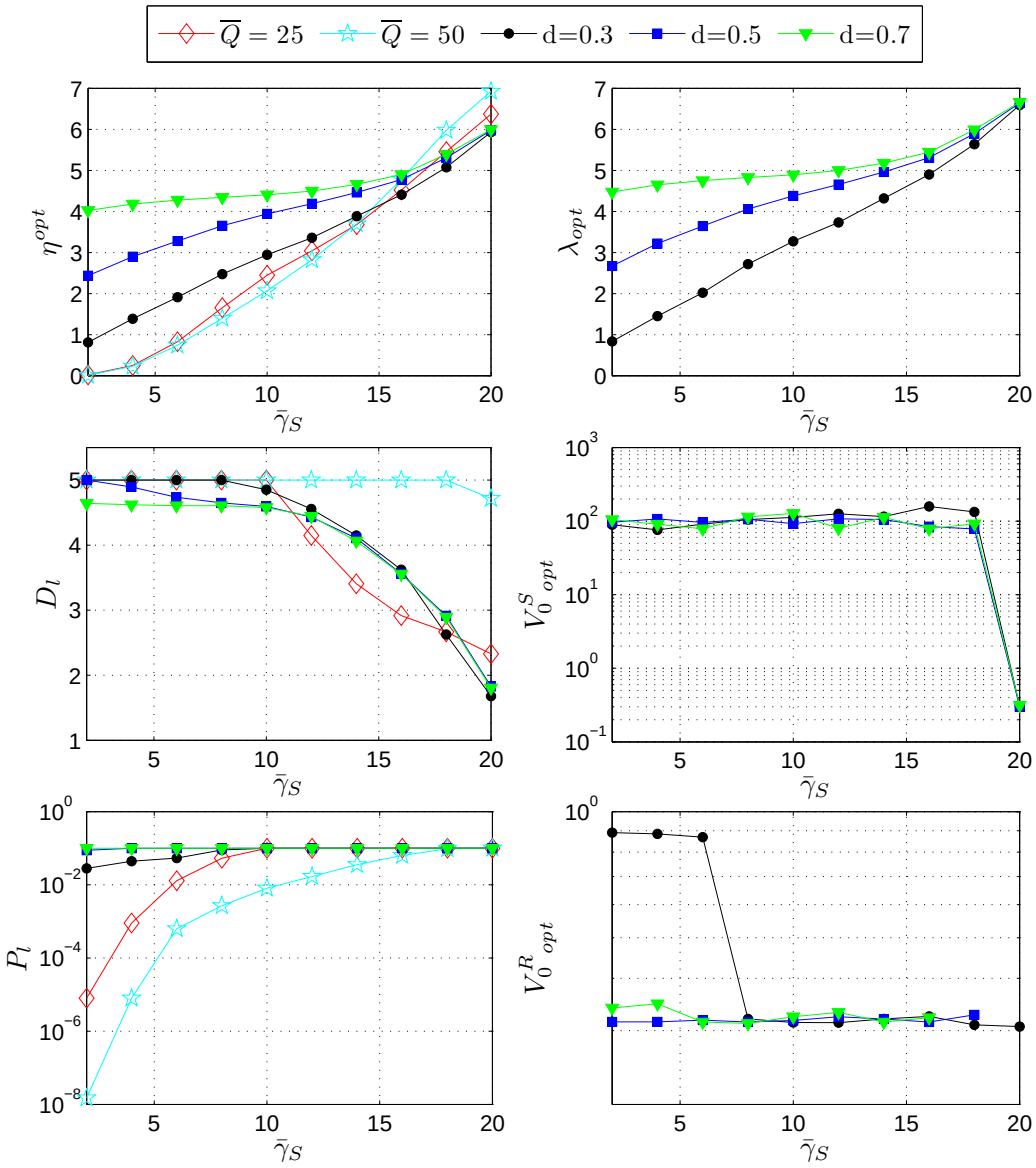


Figure 5.21: Constrained throughput optimization vs. $\bar{\gamma}_S$ for CARQ and Type-I hybrid FEC/ARQ.

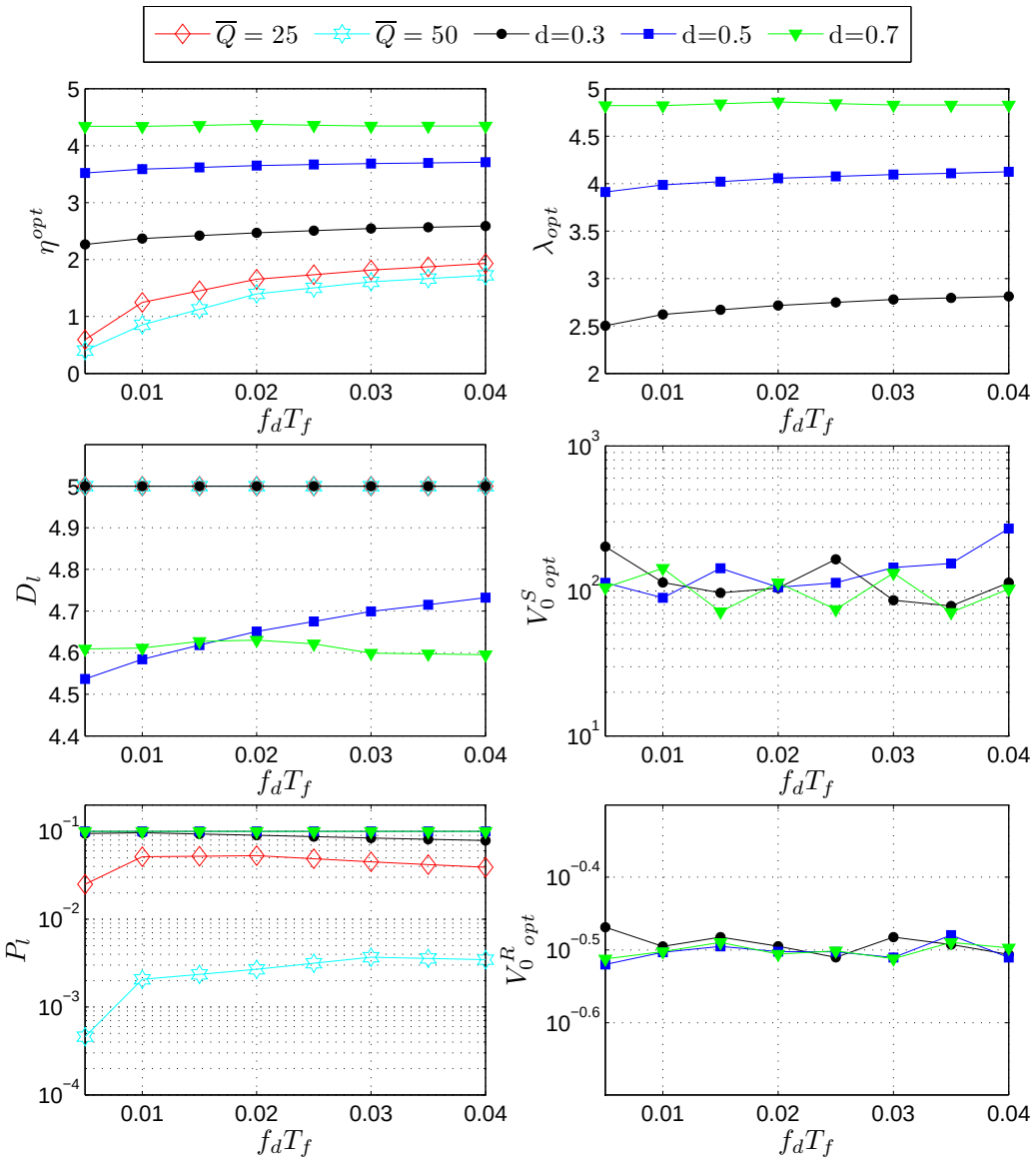


Figure 5.22: Constrained throughput optimization vs. $f_d T_f$ for CARQ and Type-I hybrid FEC/ARQ.

more difficult the interpretation of the obtained results. It can be inferred from Fig. 5.20 that higher SNR values, which correspond to better channel conditions, entail an enhancement in the optimum achieved throughput. It is also clear that an improvement in the average throughput is also obtained for a shorter distance between the relay and the destination (greater d), although this improvement becomes marginal for high SNR values. The graphs depicting ${V_0^S}_{\text{opt}}$ and ${V_0^R}_{\text{opt}}$ are in agreement with the behaviour that has been previously observed in Subsection 5.6.1: for low $\bar{\gamma}_S$ a high ${V_0^S}_{\text{opt}}$ value is obtained, which corresponds to the situation in which only the highest TM is used at the source, while for better channel conditions a lower ${V_0^S}_{\text{opt}}$ is needed, which implies the availability at the source of more robust TMs (the relay will reduce its activity). Because a longer R-D distance (lower d) corresponds to a worst R-D channel and, consequently, to a reduced implication of the relay in the transmission, the step down in ${V_0^S}_{\text{opt}}$ is observed for lower $\bar{\gamma}_S$ values when d diminishes. On the contrary, no matter how high the SNR is, the ${V_0^R}_{\text{opt}}$ value will always fall below ${V_0^S}_{\text{opt}}$. This is because the relay can never trust in any other link to retransmit the erroneously received packets at the destination and, thus, it needs all the available TMs in order to improve the reliability of the transmission. Moreover, as a lower dependence of the throughput on V_0^R is observed for a better R-D channel (shorter R-D distance or, equivalently, greater d), no optimum V_0^R can be obtained for high d values, specially for increased SNRs.

In the second optimization problem, QoS-guaranteed traffic characterized by a maximum average packet loss rate $P_{l\max} = 0.1$ packets/frame and a maximum average packet delay $D_{l\max} = 5$ frames has been considered. Traffic has been generated using a truncated Poisson process with a truncation length equal to 10 packets, i.e. $\boldsymbol{\Theta} = (0, 10]$.

Figure 5.21 plots the dependence on the average SNR of the optimum average throughput and the corresponding average delay and packet loss rate when the CARQ (with $d = 0.3$, $d = 0.5$ and $d = 0.7$) and the classical Type-I hybrid FEC/ARQ protocols (with maximum buffer sizes $Q = 25$ and $Q = 50$ packets) are considered. The constrained optimization for the latter protocol has been developed as proposed in Subsection 3.4.2 with the *Instantaneous* option for the AMC scheme. Moreover, the optimum values ${V_0^S}_{\text{opt}}$, ${V_0^R}_{\text{opt}}$ and λ_{opt} are provided. It can be observed that, with the Type-I hybrid FEC/ARQ protocol, for lower SNR values the maximum allowed delay $D_{l\max}$ is the limiting QoS parameter, while the average packet loss rate P_l remains under $P_{l\max}$. However, for higher SNR values, the obtained packet loss rate equals $P_{l\max}$ while the packet delay falls below its maximum permitted value. The same behaviour is also observed for the CARQ protocol when $d = 0.3$. On the contrary, for higher d values, $P_{l\max}$ is always the limiting constraint. These graphs confirm that a shorter distance between nodes R and D implies an improvement in the achieved throughput, although this improvement becomes marginal for high SNR values. As expected, with greater SNRs the optimum sustainable arrival rate increases while ensuring the fulfillment of both QoS requirements. From the shape of these curves it is obvious that, in comparison to the Type-I hybrid FEC/ARQ, either for a maximum buffer size $\overline{Q} = 25$ or $\overline{Q} = 50$ packets, the CARQ protocol allows a better performance in terms of throughput for low-to-medium SNR values, although for higher SNRs

the basic FEC/ARQ protocol slightly outperforms the CARQ scheme. Hence, for good channel conditions it is better to directly transmit packets from the source to the destination without any support from the relay.

In order to clearly illustrate how the channel correlation affects the performance of the system, Fig. 5.22 plots the dependence of the obtained η_{opt}, the corresponding D_l and P_l, and the optimum values ${V_0^S}_{\text{opt}}$, ${V_0^R}_{\text{opt}}$ and λ_{opt}, on the maximum normalized Doppler frequency. Either with the Type-I hybrid FEC/ARQ protocol or the CARQ protocol when $d = 0.3$, the maximum allowed delay $D_{l\max}$ is the limiting QoS parameter, while $P_{l\max}$ is always the limiting constraint with the CARQ protocol for higher d values. As expected, a shorter distance between nodes R and D entails an improvement in the achieved throughput. It can be clearly observed that the throughput performance of the Type-I hybrid FEC/ARQ scheme depends on the level of channel correlation, and the throughput loss is severe for a highly correlated channel (small $f_d T_f$). However, the throughput of the CARQ scheme is virtually independent of the Doppler frequency, specially for short R-D distances, and is always greater than that achieved with the other schemes. This observation indicates that the CARQ is an effective solution to overcome the serious effects of channel correlation.

The curves depicting ${V_0^S}_{\text{opt}}$ and ${V_0^R}_{\text{opt}}$ in both Fig. 5.21 and Fig. 5.22 are in agreement with the behaviour that has been observed in previous plots: a high ${V_0^S}_{\text{opt}}$ value is obtained for low-to-medium $\bar{\gamma}_S$ and for better channel conditions a lower ${V_0^S}_{\text{opt}}$ is needed, while the ${V_0^R}_{\text{opt}}$ value always remains below ${V_0^S}_{\text{opt}}$.

5.7 Chapter summary

In this chapter, generalizing and extending the analytical tools previously presented in this research work, a cross-layer DTMC-based queueing model for systems that combine an adaptive multi-rate PHY layer with a CARQ protocol at the DLC layer has been developed. The stationary state probability distribution of this multidimensional DTMC has been used to obtain closed-form analytical expressions for performance metrics such as throughput, average packet delay, queue lengths and packet loss rate due to buffer overflow. Furthermore, this cross-layer analytical framework allowed the formulation of multidimensional and simplified tridimensional optimization problems to maximize the system throughput under maximum packet loss rate and delay constraints.

To assess the validity of the proposed model, computer simulation results have been presented to show that the derived closed-form analytical expressions can be used to faithfully reproduce results based on Clarke's statistical Rayleigh fading model. Numerical results have shown that the proposed CARQ scheme can lead to a remarkable improvement on the maximum achieved throughput when compared to the classical type-I hybrid FEC/ARQ scheme, although the amount of this improvement largely depends on the channel conditions and on the distance between the source and the relay. It has been also demonstrated that the CARQ protocol can effectively mitigate the effect of correlation in the direct source to destination channel by exploiting cooperative diversity through

a relay node. Therefore, it provides a noticeable performance gain in correlated fading channel scenarios.

CHAPTER 6

Analysis of an infinitely persistent ARQ protocol with outdated channel state information

As it has been mentioned, state-of-the-art wireless communications standards combine (H)ARQ protocols at the DLC layer with AMC strategies at the PHY layer. Most previous work on cross-layer modeling of these networks assumes the availability of ideal CSI. However, in practical systems the CSI used to perform link adaptation may be outdated. This chapter presents a cross-layer multidimensional DTMC-based queuing model that jointly exploits the capabilities of ARQ and AMC, by explicitly accounting for imperfections in the AMC scheme caused by the delay in the CSI feedback channel. The proposed model is used to obtain closed-form analytical expressions for the system performance measures and an explicit analysis of the impact of CSI sensing delay on these parameters is presented. Although an infinitely persistent Type I hybrid FEC/ARQ protocol is considered, the analysis could be extended to any of the ARQ-based protocols introduced earlier.

6.1 Introduction

In Chapters 3-5, different cross-layer designs involving the PHY and DLC layers have been developed. In order to compensate for the impact of wireless fading, all these proposals coincide in combining AMC with an (H)ARQ protocol, aiming at improving the spectral efficiency by jointly exploiting the adaptability of AMC to the wireless channel conditions and the error-correcting capability of (H)ARQ. All results are based on the assumption of availability of ideal CSI at both the transmitter and receiver sides. However, the CSI used to perform link adaptation in practical systems may be outdated due to the transmission delay and the processing imperfections both at the transmitter and the receiver. Adaptive designs assuming perfect CSI are useful only when these imperfections are limited. For

instance, an adaptive system with a delayed error-free feedback CSI channel can tolerate a normalized time delay up to about $f_d\tau \leq 10^{-2}$ without a noticeable degradation in the average system performance (Alouini and Goldsmith, 2000).

The impact of using outdated CSI on the performance of an AMC-based wireless system using an (H)ARQ-based error control protocol rests to be done. In contrast to previous chapters, where the system performance was optimized assuming that perfect CSI was available at both transmitter and receiver, CSI imperfections will be explicitly accounted for in this chapter, where a type-I hybrid FEC/ARQ protocol is considered. Particularly, the AMC TM selection thresholds will be optimized with constraints on a prescribed instantaneous PER. For this purpose, an AMC threshold searching algorithm will be presented that, depending on the normalized CSI sensing delay, the channel conditions and the QoS requirements, discerns *useful* TMs from *useless* ones. That is, only a limited set of *useful* TMs will be available to the AMC scheme.

Analogously to previous chapters, a multidimensional DTMC-based analysis will be carried out and, based on the stationary state probability distribution, closed-form analytical expressions for performance metrics such as average throughput, queue length, packet delay and packet loss rate will be derived. A detailed analysis of the impact of using outdated CSI on these performance parameters will be presented.

6.2 System model and assumptions

The same generic point-to-point wireless packet communication system described in Section 2.1 and illustrated in Fig. 2.1 will be considered, as well as the frame and packet structures shown in Fig. 2.2. An infinitely persistent Type I hybrid FEC/ARQ protocol will be analyzed and, without loss of generality, convolutionally coded M-QAM schemes adopted from the IEEE 802.11a standard (IEEE, 1997) will be used in the AMC pool, listed in Table 3.1.

A Rayleigh block-fading channel model has been adopted (Biglieri et al., 2001), assuming that the frame duration T_f is much smaller than the channel coherence time. CSI is supposed to be available at the receiver side and, thus, a frame-by-frame TM selection process is performed at the receiver AMC controller. CSI and feedback information are sent back to the transmitter by using the feedback CSI channel, with a CSI sensing delay τ.

6.3 Physical layer modeling

6.3.1 Characterization of the packet error rate and AMC scheme

Characterization of the packet error rate

Let us consider the channel quantity γ_ν denoting the instantaneous received SNR at time instant $t = \nu T_f$, where T_f is the frame period. For the considered Rayleigh fading channel model, γ_ν is an exponentially distributed random variable with pdf defined by (3.1).

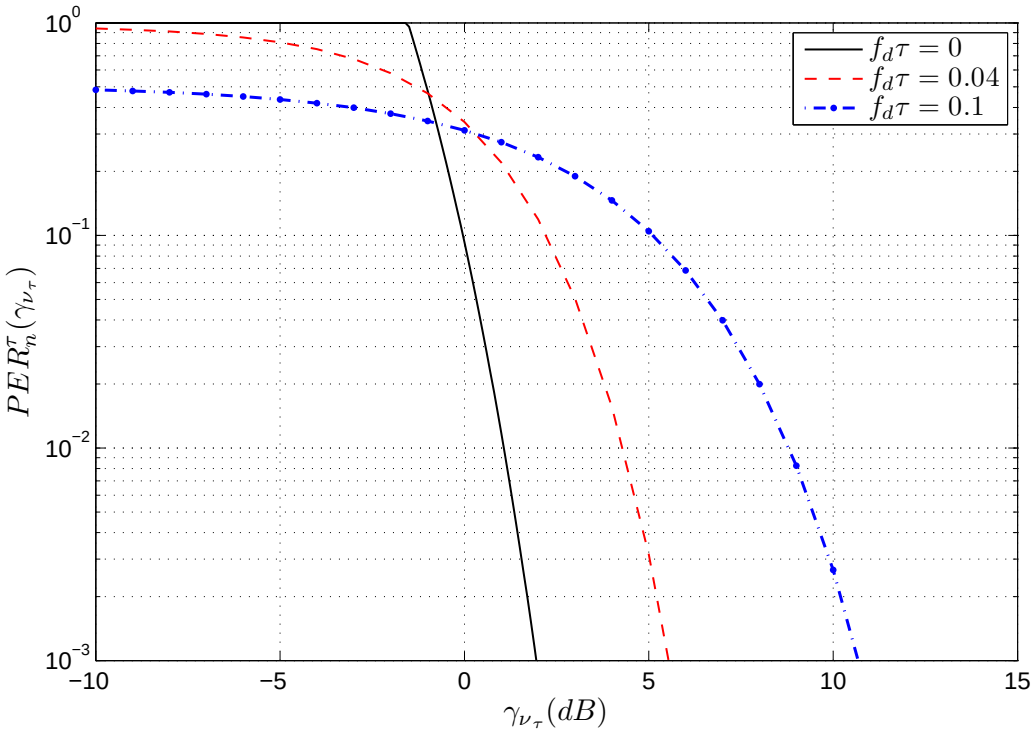


Figure 6.1: Instantaneous PER for TM1 vs. instantaneous SNR for different normalized CSI sensing delays.

Assuming the use of TM n, the instantaneous PER at the output of the soft Viterbi decoder can be approximated as in (3.3).

For an adaptive system with a CSI sensing delay τ, the receiver decides which is the best TM to use in transmission frame ν, based on an outdated instantaneous SNR γ_{ν_τ}, *measured* at time instant $t = \nu T_f - \tau$. In fact, the receiver can predict the *instantaneous* PER for TM n as

$$\mathrm{PER}_n^\tau(\gamma_{\nu_\tau}) = \int_0^\infty \mathrm{PER}_n(\gamma_\nu) p_{\gamma_\nu|\gamma_{\nu_\tau}}(\gamma_\nu, \gamma_{\nu_\tau}) d\gamma_\nu \tag{6.1}$$

where the pdf of γ_ν conditioned on γ_{ν_τ} is given by (Alouini and Goldsmith, 2000, eq. (41))

$$p_{\gamma_\nu|\gamma_{\nu_\tau}}(\gamma, \gamma_\tau) = \left(\frac{1}{\overline{\gamma}(1-\rho)}\right) e^{-\left(\frac{\gamma+\rho\gamma_\tau}{(1-\rho)\overline{\gamma}}\right)} I_0\left(\frac{2\sqrt{\rho\gamma\gamma_\tau}}{(1-\rho)\overline{\gamma}}\right), \tag{6.2}$$

with $\rho = J_0^2(2\pi f_d \tau)$, where $I_0(\cdot)$ is the modified zero-order Bessel function of the first kind and $J_0(\cdot)$ is the zero-order Bessel function of the first kind.

Figure 6.1 depicts the instantaneous PER curves $\mathrm{PER}_n^\tau(\gamma_{\nu_\tau})$ corresponding to different normalized CSI sensing delays $f_d\tau$ for TM1. It can be clearly observed that as $f_d\tau$ increases, the instantaneous PER curves flatten and shift to the right. This behaviour is the result of the averaging of the $\mathrm{PER}_n(\gamma_\nu)$ values involved in the integration process described in equation (6.1). The higher the normalized CSI sensing delay, the greater the variability of γ_ν with respect to γ_{ν_τ}, thus implying a flattening of the $\mathrm{PER}_n^\tau(\gamma_{\nu_\tau})$ curves. Although not shown in this work, an

analogous behaviour can be observed for all TMs. As a result, according to the AMC scheme that will be presented next, and in order to maintain a sufficiently low instantaneous PER, higher SNR values will be required to select a given TM, which will in turn lead to a more conservative AMC scheme.

Adaptive modulation and coding

Given γ_{ν_τ}, the objective of the AMC scheme is to select the TM that maximizes the data rate while maintaining an instantaneous PER below a prescribed value P_0. According to (Liu et al., 2004), the entire SNR range is partitioned into a set of non-overlapping intervals defined by the partition (2.1) , and mode n will be selected when $\gamma_{\nu_\tau} \in [\gamma_n^m, \gamma_{n+1}^m)$ with probability

$$Pr(n) = \int_{\gamma_n^m}^{\gamma_{n+1}^m} p_{\gamma_{\nu_\tau}}(\gamma) d\gamma = e^{-\gamma_n^m/\overline{\gamma}} - e^{-\gamma_{n+1}^m/\overline{\gamma}}. \tag{6.3}$$

Thus, the partition boundary γ_n^m will be obtained as

$$\gamma_n^m = \{\gamma : \mathrm{PER}_n^\tau(\gamma) = P_0\}, \forall n. \tag{6.4}$$

In trying to solve (6.4) it can happen that, depending on the normalized CSI sensing delay $f_d\tau$, the channel conditions and the QoS requirements, some of the possible TMs may be declared useless and thus, only a limited set of useful TMs will be available to the AMC scheme. As an example, it can be observed in Fig. 6.2, plotting the instantaneous PER vs. the instantaneous SNR for a normalized CSI sensing delay $f_d\tau = 0.1$, that for the given P_0 value TM0 will be declared useless.

In order to achieve the objective $\mathrm{PER}_n^\tau(\gamma) = P_0$ we propose the following threshold searching algorithm:

1. Set $M = 0$, $\boldsymbol{\mathcal{M}} = \emptyset$, $n = M_p$, $\gamma_0^m = 0$, $\gamma_n^m = \infty$.

2. $n \leftarrow n - 1$
   If $n = 0$ go to step 3, otherwise search the unique $\gamma_n^m \in [0, \gamma_{n+1}^m]$ that satisfies $\mathrm{PER}_n^\tau(\gamma_n^m) = P_0$.

   - In case it exists, update $M \leftarrow M + 1$ and $\boldsymbol{\mathcal{M}} \leftarrow \{n\} \cup \{\boldsymbol{\mathcal{M}}\}$ and go to step 2.
   - In case it does not exist due to $\mathrm{PER}_n^\tau(\gamma_n^m) < P_0, \forall \gamma_n^m \in [0, \gamma_{n+1}^m]$, TM n is the last *useful* mode and, thus, update $M \leftarrow M+1$, $\boldsymbol{\mathcal{M}} \leftarrow \{n\} \cup \{\boldsymbol{\mathcal{M}}\}$ and threshold subindices $\gamma_{i-n}^m \leftarrow \gamma_i^m$, $i = \{n+1, \ldots, n+M\}$ and stop the searching algorithm.

3. TM 0 is the lowest used TM and, thus, update $M \leftarrow M + 1 = M_p$ and $\boldsymbol{\mathcal{M}} \leftarrow \{n\} \cup \{\boldsymbol{\mathcal{M}}\}$ and stop the searching algorithm.

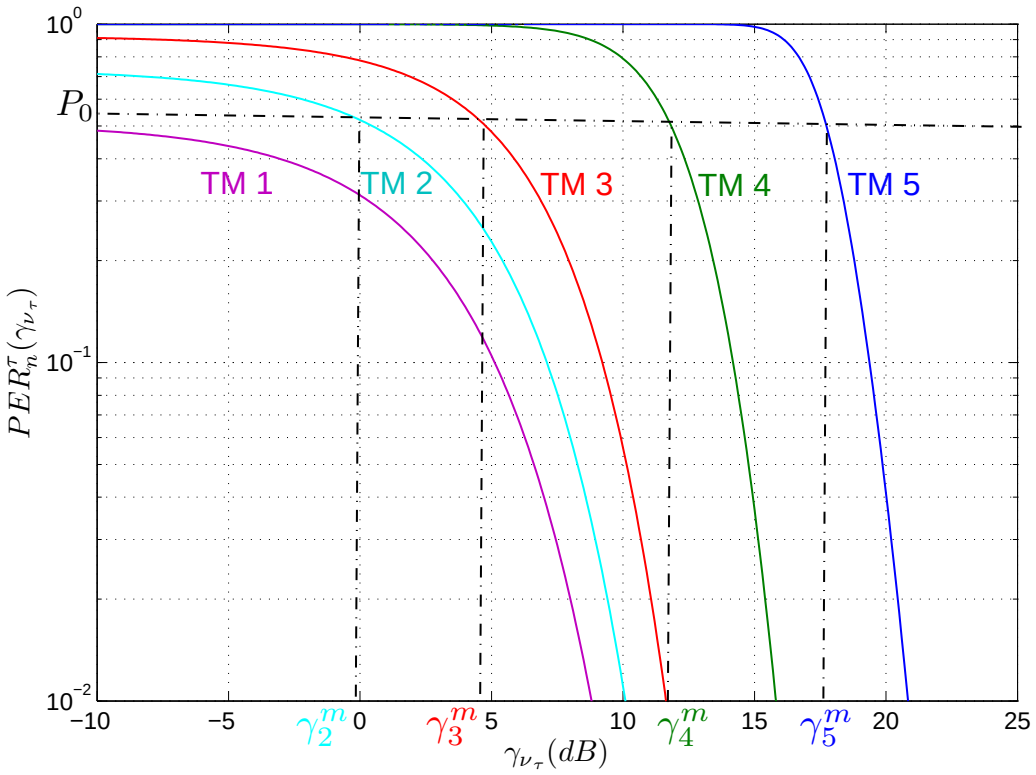


Figure 6.2: Instantaneous PER vs. instantaneous SNR for a normalized CSI sensing delay $f_d\tau = 0.1$.

6.3.2 Physical layer two-dimensional Markov model

Following the same methodology described in previous chapters, the Markov chain-based PHY layer model will take into consideration the first-order two-dimensional Markov channel model, characterized by the partitions $\mathbf{\Gamma}^c$ and $\mathbf{\Delta}$ defined in (2.6) and (2.8) respectively, and the TM selection process used by the AMC scheme, characterized by the partition $\mathbf{\Gamma}^m$ defined in (2.1). Then, the range of γ_ν will be partitioned into the set of N_{PHY} non-overlapping intervals defined by the partition (3.4), where each partition interval $\left[\gamma_k^{m,c}, \gamma_{k+1}^{m,c}\right)$ is characterized by a particular combination of TM and channel state. Furthermore, let us also consider the partition of $\delta_\nu = \gamma_{\nu-1} - \gamma_\nu$ into the set of non-overlapping intervals $\mathbf{\Delta} = \{(-\infty, 0), [0, \infty)\}$. Using this two-dimensional partitioning, a first-order two-dimensional Markov model for the PHY layer can be defined where each state corresponds to one of such two-dimensional rectangular-shaped cells. Therefore, the PHY layer Markov chain state at time instant $t = \nu T_f$ can be denoted as $\varsigma_\nu = (\varphi_\nu, \Delta_\nu)$, $\nu = 0, 1, \ldots, \infty$, where $\varphi_\nu \in \{0, \ldots, N_{\text{PHY}} - 1\}$ denotes the combination of TM and channel state in this frame interval and $\Delta_\nu \in \{0, 1\}$ is used to denote the *up* or *down* characteristic of the instantaneous SNR in time frame interval $t = (\nu - 1)T_f$.

At any time instant $t = \nu T_f$ the PHY layer state can be univocally characterized by an integer number $n_\nu = 2\varphi_\nu + \Delta_\nu$ and obviously, $n_\nu \in \{0, \ldots, 2N_{\text{PHY}} - 1\}$.

The PHY layer will be in a state $n \in \{0, \ldots, 2N_{\text{PHY}}-1\}$ with a steady-state probability $P^{\text{PHY}}(n)$, which can be calculated using eqs. (3.6) and (2.9), and each of these states will be characterized by a conditional average PER given by

$$\overline{PER}_n^{\text{PHY},\tau} = \begin{cases} \dfrac{\int_{\gamma_{n/2}^{m,c}}^{\gamma_{n/2+1}^{m,c}} \int_0^x PER_{\beta_n}^{\tau}(x) p_{\gamma_\nu,\gamma_{\nu-1}}(x,y)dydx}{P^{\text{PHY}}(n)}, & n \text{ even} \\ \dfrac{\int_{\gamma_{(n-1)/2}^{m,c}}^{\gamma_{(n+1)/2}^{m,c}} \int_x^{+\infty} PER_{\beta_n}^{\tau}(x) p_{\gamma_\nu,\gamma_{\nu-1}}(x,y)dydx}{P^{\text{PHY}}(n)}, & n \text{ odd} \end{cases} \tag{6.5}$$

where $PER_{\beta_n}^{\tau}$ denotes the *instantaneous* PER of TM β_n corresponding to the nth PHY layer state when the CSI sensing delay is τ, and $p_{\gamma_\nu,\gamma_{\nu-1}}(x,y)$ is the joint pdf of the random variables γ_ν and $\gamma_{\nu-1}$, given by (2.9).

Furthermore, the PHY layer FSMC will be characterized by a transition probability matrix

$$\boldsymbol{P}_s = \left[P_s(n_\mu, n_{\mu'})\right]_{n_\mu,n_{\mu'}=0}^{2N_{\text{PHY}}-1},$$

whose elements can be calculated using eqs. (3.9)-(3.12) and (2.16).

6.4 Discrete time Markov chain-based link-level queueing model and analysis

After having modeled the PHY layer, the DLC layer is introduced in the model. The analytical framework developed in Subsection 3.3.1 can be directly applied to the case of an infinitely persistent Type I hybrid FEC/ARQ protocol with outdated CSI, although the PHY layer has to be modeled as in Subsection 6.3.2 instead of as in Subsection 3.2.2. As a result, the only difference in the analytical expressions corresponding to the cases of perfect CSI and outdated CSI can be found in (3.13), which corresponds to the probability that k packets are successfully transmitted (leave the queue) given $\mathcal{C}_{n_\mu}$ packets are transmitted when the PHY layer is in state $\mathcal{S}_\mu$, and that now can be written as

$$p_{k,\mathcal{C}_{n_\mu}}^{(n_\mu)} = \binom{\mathcal{C}_{n_\mu}}{k} \left(\overline{PER}_{n_\mu}^{\text{PHY},\tau}\right)^{\mathcal{C}_{n_\mu}-k} \left(1 - \overline{PER}_{n_\mu}^{\text{PHY},\tau}\right)^{k}. \tag{6.6}$$

6.5 Numerical results

To verify the validity of the proposed cross-layer framework, analytical results obtained with our proposed 2D-FSMC model will be confronted with computer simulation results obtained using Clarke's statistical Rayleigh model. Unless otherwise specified, numerical results correspond to the following default parameters: normalized maximum Doppler frequency $f_d T_f = 0.02$, average received SNR $\overline{\gamma} = 8$ dB, buffer size $\overline{Q} = 50$, number of channel states $K = 10$, parameter $b = 2$ and a D-BMAP that has been characterized with the transition probability matrix (2.4).

The dependence of the average packet loss rate P_l, throughput η, packet delay D_l and queue length L_q, on the target instantaneous PER P_0, is illustrated in

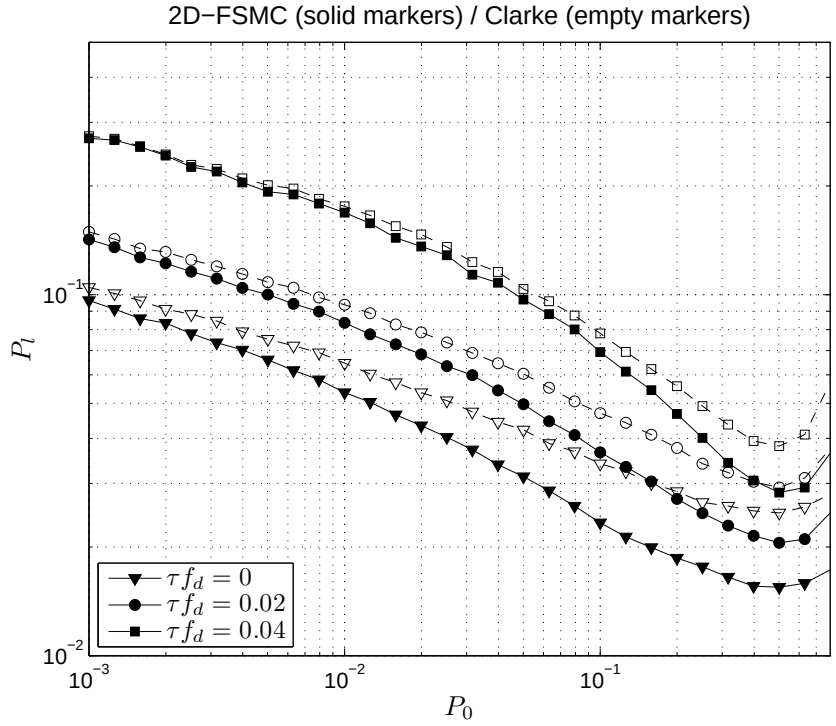


Figure 6.3: Average packet loss rate vs. target PER.

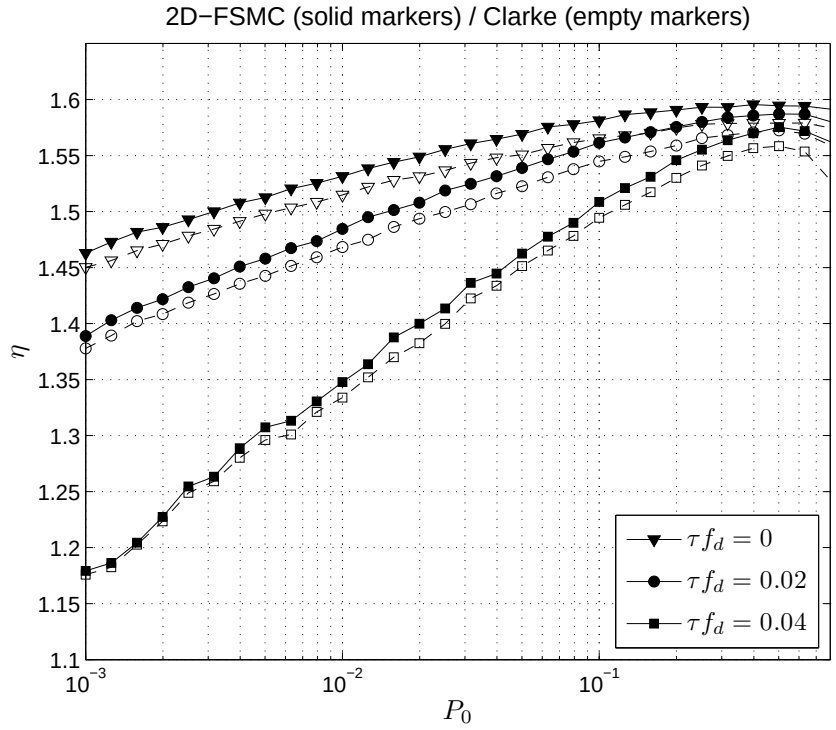


Figure 6.4: Average throughput vs. target PER.

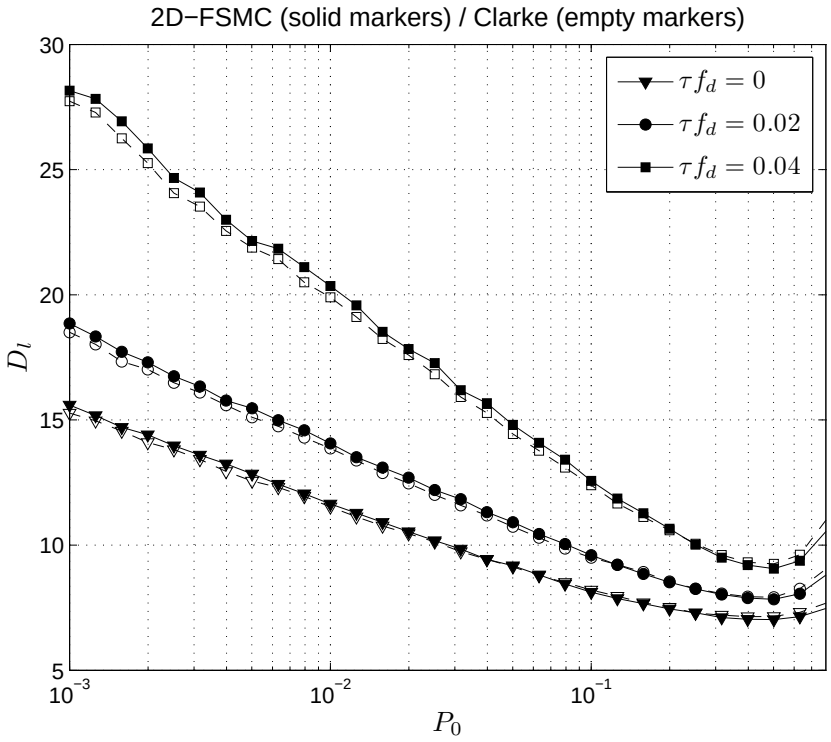


Figure 6.5: Average packet delay vs. target PER.

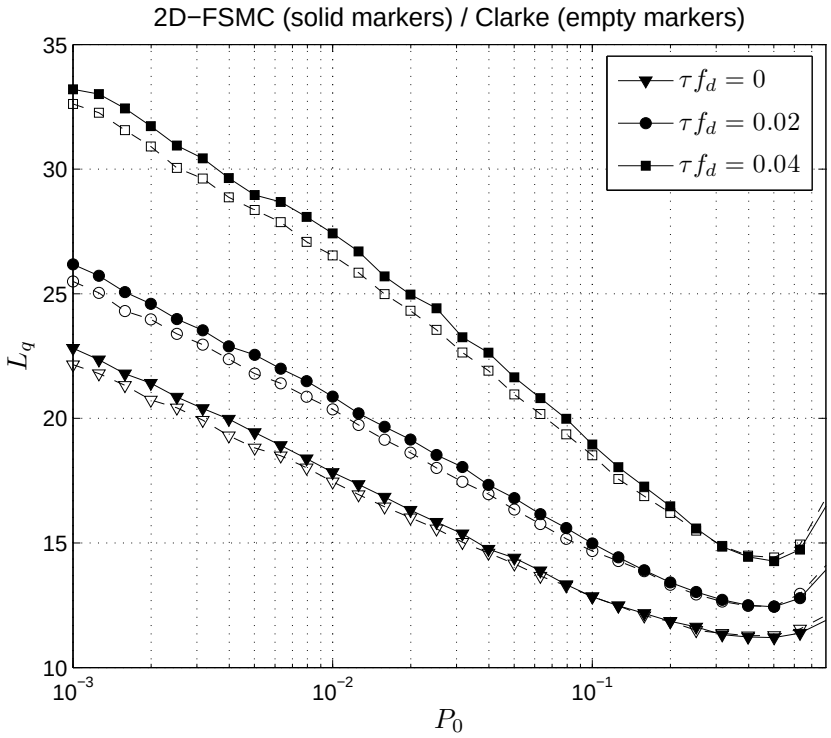


Figure 6.6: Average queue length vs. target PER.

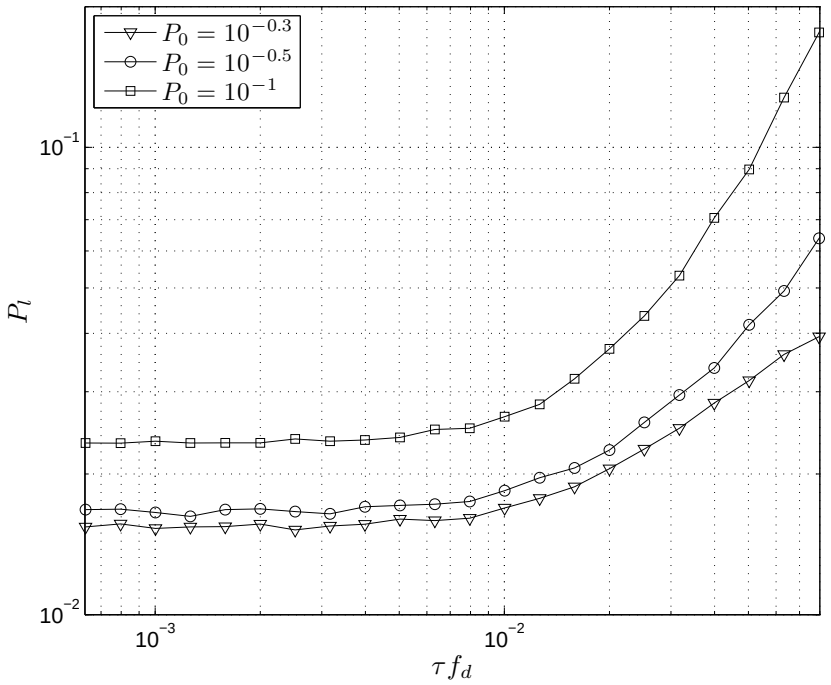


Figure 6.7: Average packet loss rate vs. normalized CSI sensing delay.

Figs. 6.3-6.6, respectively. Different values of the normalized CSI sensing delay $f_d\tau$ have been considered. The analysis of these graphs reveals that higher P_0 values imply the utilization of higher order TMs, which leads to an increment of the queueing service rate and, consequently, to a decrease in the buffer size L_q and in the overflow probability P_l, thus causing an increase in the throughput η. Nevertheless, it can be appreciated that when the growth of the service rate for high P_0 values cannot cope with the huge number of required retransmissions, the amount of packets that remain in the queue L_q augments and P_l takes off while η diminishes. As expected, an analogous behaviour can be observed for the average packet delay D_l. These graphs show that higher CSI sensing delays entail a degradation of the system performance. It can also be observed that the throughput drop when P_0 moves from its optimum value raises for higher $f_d\tau$ values. This higher rate of decrease in η is due to the fact that at the receiver side the system performs AMC with ideal CSI, while as a result of the delay in the CSI feedback channel, involving variability of the instantaneous SNR, the selected TM can be outdated at the transmitter side. Note that in all cases, and for any value of the normalized CSI sensing delay, the behaviour of the simulation of the infinitely persistent Type-I hybrid FEC/ARQ protocol with a PHY layer based on Clarke's model, is faithfully reproduced by the presented analytical PHY/DLC 2D-FSMC model. The shape of the curves and the location of the optimum values obtained by simulation (Clarke's) coincide with those obtained using the proposed analytical model (2D-FSMC), which is particularly important to ensure an optimal cross-layer design.

Figures 6.7-6.10 illustrate the sensitivity of the performance parameters to

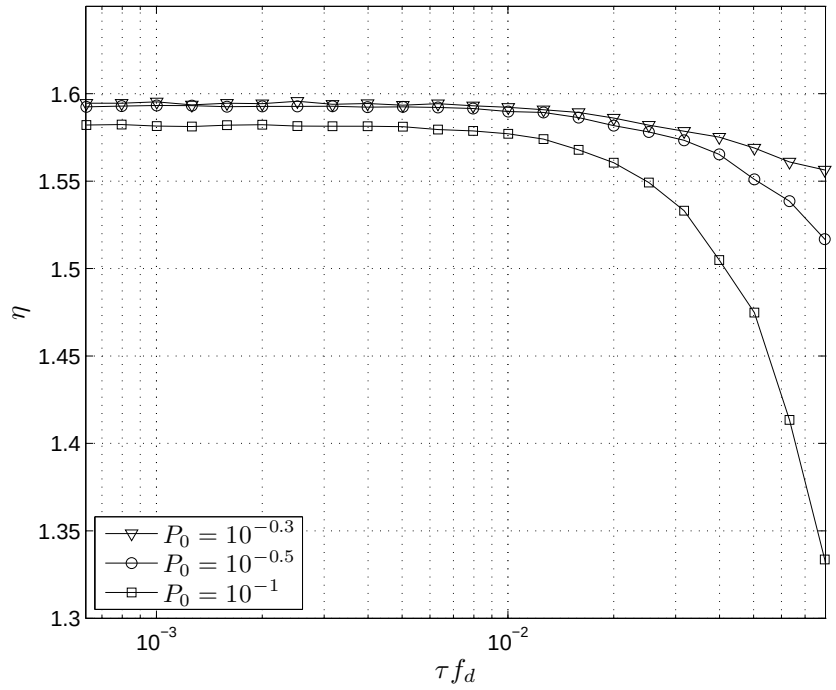


Figure 6.8: Average throughput vs. normalized CSI sensing delay.

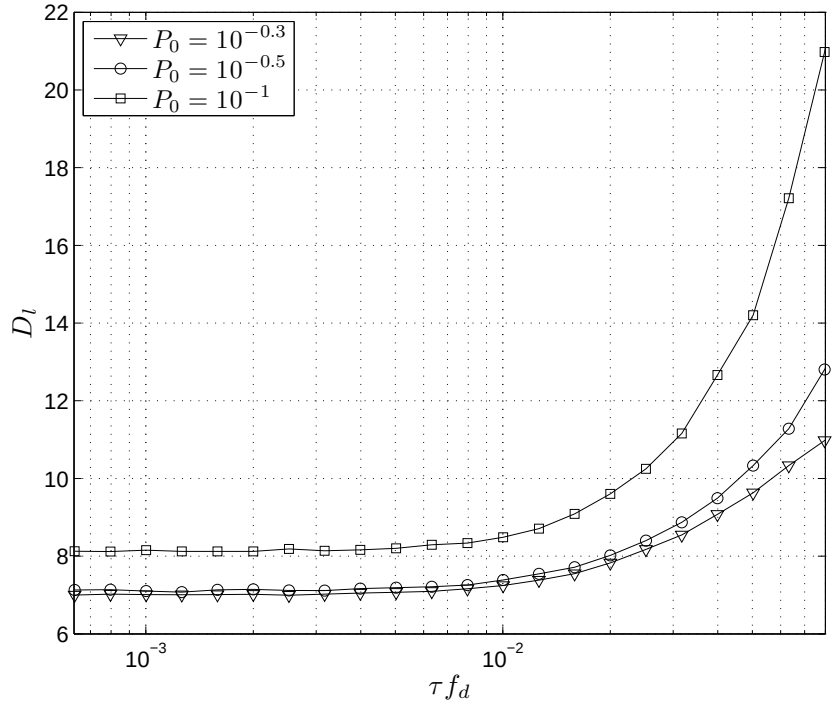


Figure 6.9: Average packet delay vs. normalized CSI sensing delay.

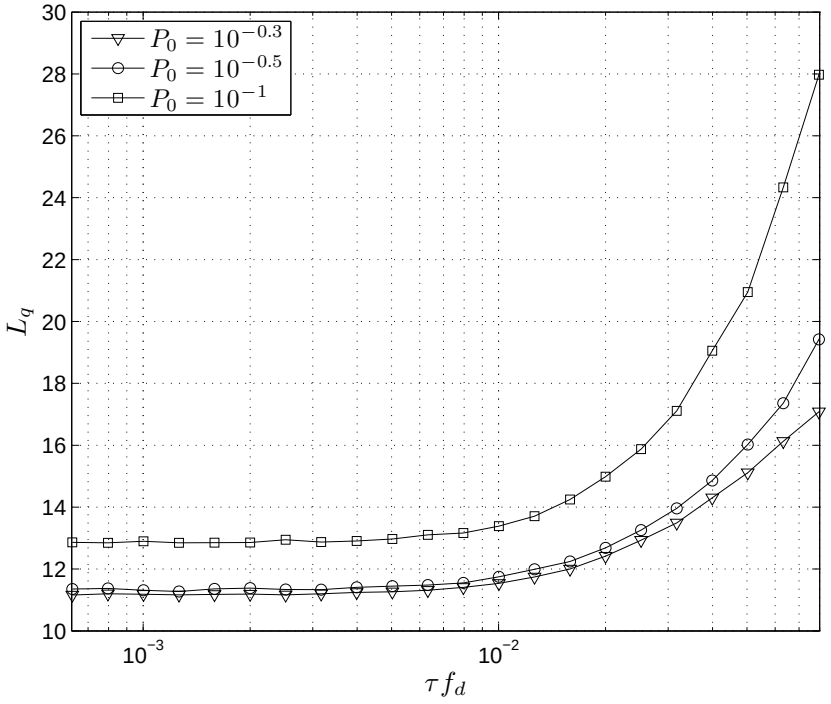


Figure 6.10: Average queue length vs. normalized CSI sensing delay.

the variation of the normalized CSI sensing delay $f_d\tau$ for different values of the target instantaneous PER P_0. It can be observed that, irrespective of the values of P_0, a normalized CSI sensing delay up to about 10^{-2} can be tolerated without a noticeable degradation in the system performance. Furthermore, results clearly show that for normalized CSI sensing delays greater than 10^{-2}, systems with a target instantaneous PER far from its optimum value ($P_0^{opt} \approx 10^{-0.3}$) suffer a greater performance degradation. These results are aligned with those obtained in (Alouini and Goldsmith, 2000).

6.6 Chapter summary

In this chapter, a methodology to account for imperfections in the AMC scheme caused by the delay in the CSI feedback channel, has been developed. Thus, using the previously proposed first-order two-dimensional FSMC model-based approach for AMC/(H)ARQ-based wireless systems, the impact of CSI sensing delay on the average packet loss rate, packet delay, throughput and queue length has been analyzed. To account for this CSI imperfection, an AMC threshold searching algorithm with the capability of discriminating between *useful* and *useless* TMs, has been proposed.

Numerical results have demonstrated that, irrespective of the target instantaneous PER, a normalized CSI sensing delay up to about 10^{-2} can be tolerated without a noticeable degradation in the system performance. For normalized CSI sensing delays greater than 10^{-2}, systems with a target instantaneous PER far

from its optimum value suffer a greater performance degradation, thus revealing the importance of optimal cross-layer designs.

Although an infinitely persistent Type I hybrid FEC/ARQ protocol has been considered, the proposed methodology could be extended to any of the AMC/(H)ARQ cross-layer designs that have been presented in this research work.

CHAPTER 7

Conclusions

The results of this research work suggest that in AMC/(H)ARQ-based systems, improved performance under QoS constraints can be obtained by efficiently jointly adjusting specific design parameters at the PHY layer and at the DLC layer. Although each chapter includes a stand-alone summary regarding specific contributions, in what follows the summary of conclusions of this dissertation is provided.

7.1 Conclusions and main contributions

In this research work an analytical framework for the cross-layer design, analysis and optimization of AMC/(H)ARQ-based wireless systems has been developed. A novel PHY layer model has been proposed, based on the use of an enhanced first-order two-dimensional FSMC model for the wireless flat-fading channel that is able to improve the ACF fitting of first-order FSMCs. In contrast to previous proposals in the literature, this model relies on the amplitude and the rate-of-change of the fading envelope. Moreover, it incorporates the implementation of the AMC scheme used in the TM selection subsystem, which is designed independently from the channel model.

Based on the aforementioned PHY layer multidimensional FSMC, the interactions between the PHY layer and the DLC layer queueing behaviour have been investigated for both, infinitely persistent and truncated (H)ARQ protocols. To that end, a DTMC has been used that jointly describes the statistical behaviour of the arrival process, the queueing system and the PHY layer. Performance metrics such as system throughput, average delay or average packet loss rate, both due to buffer overflow and due to exceeding the maximum number of allowed retransmissions, have been derived by solving the global transition probability matrix of the system. Moreover, the analysis has been extended to the case of a cooperative system where a relay node assists the source-destination link. Valuable radio link level performance measures have been provided at the cost of an increase in com-

plexity depending on the number of arrival process phases, the number of PHY layer Markov model states or the maximum queue length of the system. For the sake of comparison with non Markov-based analytical tools, the Type-I hybrid FEC/ARQ protocol with infinite persistence has also been analyzed through the effective bandwidth/capacity theory, which accurately characterizes the close interactions between the PHY layer and the QoS provisioning performance at the DLC layer. This approach follows a completely different methodology and, compared with the DTMC-based framework, it is not only less complex but in addition it could be readily extended to multiuser shared-wireless media scenarios. Nevertheless, it has been shown that for many cases of practical interest the asymptotic approximations used in this approach seriously compromise its usefulness in the proposed cross-layer framework.

The resulting analytical link-level queueing model provides then a roadmap towards effective cross-layer multidimensional design strategies that have been thoroughly explored. A simplification approach of this problem has been also proposed, resulting in a simpler scheme at the cost of a performance loss. Particularly, it has been shown that the analytical results not only allow a-priori design decisions to be made, but they also provide an insight that enables the derivation of dynamic reconfiguration strategies that take into account instantaneous CSI at the PHY layer and QSI at the DLC layer. Particularly, the cross-layer design optimally determines specific design parameters at the PHY (e.g., AMC switching thresholds and set of AMC useful modes), and DLC layers (e.g., average packet arrival rate and buffer size), which maximize the average throughput of the system, while satisfying prescribed QoS requirements. Obviously, the importance of each metric depends on the required QoS, which is service-specific (for voice traffic, the delay has clearly to be minimized, whereas for file transfer services it is more interesting to maximize the throughput).

Analytical and Monte Carlo simulation results have confirmed the validity and suitability of the proposed cross-layer framework. The PHY layer model has been compared with both Clarke's statistical Rayleigh fading model (Clarke, 1968) and the model proposed in (Le et al., 2006b), which was based on the first-order AFSMC model introduced by Liu et al. (2005b). Numerical examples have shown that, compared to the first-order AFSMC approach, the proposed first-order two-dimensional FSMC model reproduces more reliably the performance metrics of the *real* system based on Clarke's statistical Rayleigh fading model. To assess the validity of the DLC analysis and cross-layer designs developed throughout this research work, computer simulation results have been presented showing that the derived closed-form analytical expressions can be used to faithfully reproduce results based on Clarke's model. Moreover, the effects of the channel quality and design parameters on the network performance have been illustrated. A remarkable conclusion that can be derived from the examination of the results is that the shape and location of the optimal values of the curves obtained using Clarke's model coincide with those obtained with the different DLC layer performance analysis, even for a small number of channel states. This demonstrates the implicit potential of using multidimensional channel fading Markov models in designing cross-layer strategies. Numerical results have also been used to compare the performance of either infinitely persistent or truncated classical Type-I hybrid

FEC/ARQ schemes with the truncated HARQ-CC and HARQ-IR strategies. It has been demonstrated that both HARQ protocols consistently outperform the classical Type-I hybrid FEC/ARQ scheme. It has been shown that HARQ-IR protocols provide a small throughput gain when compared to HARQ-CC schemes due to the use of PHY layer AMC algorithms. Moreover, results have also confirmed the benefits of the cross-layer multidimensional optimization in attaining the maximum throughput in comparison to the simplified two-dimensional approach. Additionally, a CARQ mechanism has been exhaustively investigated. In comparison with the non-cooperative counterparts, CARQ-based architectures have shown to perform significantly better thanks to the exploitation of the additional spatial diversity. Finally, imperfections in the AMC scheme caused by the delay in the CSI feedback channel have also been accounted for in Chapter 6. Numerical results have revealed that a normalized CSI sensing delay up to about 10^{-2} can be tolerated without a noticeable degradation in the system performance.

To recap, the following list outlines the main outcomes of this research work:

- An improved first-order two-dimensional Markov model for the Rayleigh flat fading channel, which relies on the amplitude and the rate-of-change of the fading envelope.
- An enhanced PHY layer first-order two-dimensional Markov model that incorporates the aforementioned proposed channel model and the implementation of the AMC scheme used in the TM selection subsystem, which efficiently describes the PHY layer channel dynamics.
- A DTMC-based analysis of the DLC layer QoS performance, considering the packet arrival process, the proposed PHY layer model and the queueing system for both infinitely persistent and truncated (H)ARQ protocols, as well as CARQ.
- The comparison of the DTMC and the effective bandwidth/capacity theory approaches for the infinitely persistent Type I hybrid FEC/ARQ protocol.
- A cross-layer design for throughput maximization in AMC/(H)ARQ-based wireless networks with average packet loss rate and delay constraints.
- The multidimensional and simplified bi/tri-dimensional approaches for the cross-layer design.
- *Ergodic* and *instantaneous* options for the AMC schemes used in the TM selection process when the simplified cross-layer design is considered.
- The validation of the proposed PHY layer model and cross-layer designs by confronting analytical and Monte Carlo simulation results.
- An exhaustive analysis of the effects of the channel quality and design parameters on the system performance.
- An in-depth study and comparison of the Type I hybrid FEC/ARQ, HARQ-CC and HARQ-IR, as well as CARQ protocols.

- The analysis of the impact of CSI sensing delay on the performance parameters.

7.2 Future work

One of the main drawbacks of cross-layer optimization approaches based on FSMCs is that the complexity involved in calculating the steady-state probabilities increases with the number of phases in the arrival process, the number of states in the PHY layer Markov model, and the number of queue states in the system. Consequently, for large buffer sizes, efficient strategies should be devised to make the problem solution computationally feasible. This is an issue that has been left for further research.

Future efforts in the context of cooperation should be devoted to the extension of the proposed CARQ scheme to environments where more than one relay is available and, therefore, a relay selection strategy must be devised. Additionally, packet combining strategies at the receiver constitute another promising avenue for future work.

Besides the aforementioned proposals, which basically deepens specific issues of the already presented work, the proposed cross-layer framework could be expanded to encompass other aspects such as:

MIMO. The use of multiple antennas at transmitter and receiver, popularly known as MIMO, is a powerful technology that achieves very high-speed wireless links while providing QoS guarantees and improved coverage in non-line-of-sight (NLOS) environments. It constitutes a significant research and engineering challenge that has been taken into account in recent 3rd Generation Partnership Project (3GPP), such as High-Speed Packet Access plus (HSPA+), LTE and LTE-advanced, as well as in the IEEE 802.16e and IEEE 802.11n standards. The performance improvements resulting from the use of MIMO systems are due to array gain, diversity gain, spatial multiplexing gain, and interference reduction (Paulraj et al., 2004). This promising PHY layer technique and its impact on the QoS provisioning performance at higher layers should be considered by utilizing the analytical framework that has been developed in this research work.

Scheduling. Next generation wireless networks require not only the development of high-performance PHY-layer technologies, but also powerful resource management strategies to provide high throughput and efficient use of resources. Scheduling algorithms are key components for QoS provisioning in wireless systems. As it has been previously mentioned, due to its scalability, the effective bandwidth/capacity-based approach that has been introduced in Subsection 3.3.2, can be extended to multiuser scenarios. Based on this theoretical framework the analysis of different packet scheduling techniques for the resource management in wireless networks with heterogeneous QoS requirements remains an open issue.

TCP/IP. Many existing applications are based on TCP/IP, the Internet end-to-end data transport protocol. At the core of TCP a congestion control

scheme proposed by Jacobson (1988) is found that is based on a dynamic window. When packet losses are detected, the window size is reduced and hence, the offered rate of the data connection. The size is gradually increased when successful transmissions occur. Traffic control and bandwidth allocation schemes in wire-line data networks are usually implemented under the assumption that data packet losses are due to buffer overflows and congestion at the routers. However, the QoS provision over wireless networks is more challenging due to harsh channel transmission characteristics, host mobility, and bandwidth fluctuations. Due to errors attributable to the wireless channel, when there is a wireless link in the end-to-end path of a TCP connection, TCP performance suffers from significant throughput degradation due to underutilization of available wireless bandwidth. This occurs because TCP cuts back its window in response to losses on the wireless link, and the subsequent growth of the window is typically sluggish, compared to variations in the quality of the wireless link. In order to analyze this problem and to investigate a suitable DLC layer error-recovery mechanism that could 'hide' the fluctuations of the wireless medium to TCP, the cross-layer analytical framework that has been proposed in this research work should be extended to the analysis of TCP performance over wireless links combining an AMC scheme at the PHY layer with an ARQ-based protocol at the DLC layer.

Bibliography

Z. Abichar, J.M. Chang, and Ch.Y. Hsu. WiMAX or LTE: Who will lead the broadband Mobile Internet? *IT Professional*, 12(3):26 –32, 2010.

M. Abramowitz and I.A. Stegun, editors. *Handbook of mathematical functions with formulas, graphs and mathematical tables*. Dover Publications, Inc., ninth edition, 1972.

S. Aïssa and G. Aniba. BER analysis of M-QAM with packet combining over Space-Time Block Coded MIMO fading channels. *IEEE Transactions on Wireless Communications*, 7(3):799–805, 2008.

A. S. Alfa, editor. *Queueing theory for telecommunications - discrete time modeling of a single node system*. Springer, first edition, 2010.

M.S. Alouini and A.J. Goldsmith. Adaptive modulation over Nakagami fading channels. *Wireless Personal Communications*, 13(1-2):119–143, 2000.

F. Babich and G. Lombardi. A Markov model for the mobile propagation channel. *IEEE Transactions on Vehicular Technology*, 49(1):63–73, 2000.

F. Babich, O. Kelly, and G. Lombardi. Generalized Markov modeling for flat fading channels. *IEEE Transactions on Communications*, 48(4):547–551, 2000.

P. Bergamo, D. Maniezzo, A. Giovanardi, G. Mazzini, and M. Zorzi. Improved Markov model for Rayleigh fading envelope. *IEE Electronics Letters*, 38(10): 477–478, 2002.

E. Biglieri, G. Caire, and G. Taricco. Limiting performance of block fading channels with multiple antennas. *IEEE Transactions on Information Theory*, 47(4): 1273–1289, 2001.

C. Blondia. A discrete time batch Markovian arrival process as B-ISDN traffic model. *Belgian Journal of Operations Research, Statistics and Computer Science*, 32:3–23, 1993.

H. Boujemaa. Delay analysis of cooperative truncated HARQ with opportunistic relaying. *IEEE Transactions on Vehicular Technology*, 58(9):4795 –4804, 2009.

I. Byun and K.S. Kim. Cooperative hybrid-ARQ protocols: Unified frameworks for protocol analysis. *ETRI Journal*, 33(5):759–769, 2011.

G. Carneiro, J. Ruela, and M. Ricardo. Cross-layer design in 4G wireless terminals. *IEEE Wireless Communications*, 11(2):7–13, 2004.

I. Cerutti, A. Fumagalli, and P. Gupta. Delay models of single-source single-relay cooperative ARQ protocols in slotted radio networks with Poisson frame arrivals. *IEEE/ACM Transactions on Networking*, 16(2):371 –382, 2008.

C.S. Chang, editor. *Performance guarantees in communication networks*. Springer-Verlag, first edition, 2000.

C.S. Chang and J. A. Thomas. Effective bandwidth in high-speed digital networks. *IEEE Journal on Selected Areas in Communications*, 13(6):1091–1100, 1995.

D. Chase. A combined coding and modulation approach for communications over dispersive channels. *IEEE Transactions on Communications*, 21(3):159–174, 1973.

D. Chase. Code combining: A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets. *IEEE Transactions on Communications*, 33(5):385–393, 1985.

H. M. Chaskar, T. V. Lakshman, and U. Madhow. TCP over wireless with link level error control analysis and design methodology. *IEEE/ACM Transactions on Networking*, 7(5):605–615, 1999.

Y. Chen and Ch. Tellambura. Infinite series representations of the trivariate and quadrivariate Rayleigh distribution and treir applications. *IEEE Transactions on Communications*, 53(12):2092–2101, 2005.

J.F. Cheng. On the coding gain of Incremental Redundancy over Chase Combining. *Proceedings IEEE GLOBECOM*, 1:107–112, 2003.

J.F. Cheng. Coding performance of Hybrid ARQ schemes. *IEEE Transactions on Communications*, 54(6):1017–1029, 2006.

R. H. Clarke. A statistical theory of mobile radio reception. *Bell System Technical Journal*, 47(6):957–1000, 1968.

T. M. Cover and A. A. E. Gamal. Capacity theorems for the Relay channel. *IEEE Transactions on Information Theory*, 25(5):572–584, 1979.

L. Dai and K. Letaief. Throughput maximization of ad-hoc wireless networks using adaptive cooperative diversity and truncated ARQ. *IEEE Transactions on Communications*, 56(11):1907 –1918, 2008.

B. Dañobeitia, J. L. Ferrer-Gomila, and G. Femenias. Cross-layer architecture design in wireless networks. *Proceedings JITEL*, 1:305–312, 2008.

M. Dianati, X. Ling, K. Naik, and X. Shen. A node-cooperative ARQ scheme for wireless ad hoc networks. *IEEE Transactions on Vehicular Technology*, 55(3): 1032 –1044, 2006.

E. O. Elliott. Estimates of error rates for codes on burst-noise channels. *Bell System Technical Journal*, 42:1977–1997, 1963.

G. Femenias. SR ARQ for adaptive modulation systems combined with selection transmit diversity. *IEEE Transactions on Communications*, 53(6):998–1006, 2005.

F. Frederiksen and T.E. Kolding. Performance and modeling of WCDMA/HSDPA transmission/H-ARQ schemes. *Proceedings VTC Fall*, 1:472–476, 2002.

P. Frenger, S. Parkvall, and E. Dahlman. Performance comparison of HARQ with Chase Combining and Incremental Redundancy for HSDPA. *Proceedings VTC Fall*, 3:1829–1833, 2001.

M. J. Gans. A power-spectral theory of propagation in the mobile radio environment. *IEEE Transactions on Vehicular Technology*, 21(1):27–38, 1972.

A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas. LTE-Advanced: next-generation wireless broadband technology. *IEEE Wireless Communications*, 17(3):10 –22, 2010.

E. N. Gilbert. Capacity of a burst-noise channel. *Bell System Technical Journal*, 39:1253–1265, 1960.

A. Goldsmith. *Wireless Communications*. Cambridge University Press, first edition, 2005.

Z. J. Haas. Design methodologies for adaptive and multimedia networks (guest editorial). *IEEE Communications Magazine*, 39(11):106–107, 2001.

J.S. Harsini, F. Lahouti, M. Levorato, and M. Zorzi. Analysis of non-cooperative and cooperative type II hybrid ARQ protocols with AMC over correlated fading channels. *IEEE Transactions on Wireless Communications*, 10(3):877 –889, 2011.

IEEE. *802.11: Standard for Wireless LAN Medium Access Control and Physical Layer Specifications*. IEEE, New York, 1997.

IEEE. *802.16: Standard for Local and Metropolitan area networks*. IEEE, New York, 2004.

IEEE. *802.16 Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*. IEEE, New York, 2006.

F. Ishizaki and G.U. Hwang. Cross-layer design and analysis of wireless networks using the effective bandwidth function. *IEEE Transactions on Wireless Communications*, 6(9):3214–3219, 2007.

ITU. *Information technology - Open Systems Interconnection - Basic reference model: the basic model, ITU-T Recommendation X.200*. ITU, 1994.

V. Jacobson. Congestion avoidance and control. *Proceedings ACM SIGCOMM*, 18(4):314–329, 1988.

Ch.G. Kang, S.H. Park, and J.W. Kim. Design of Adaptive Modulation and Coding scheme for truncated Hybrid ARQ. *Wireless Personal Communications*, 53(2):269–280, 2009.

A.K. Karmokar, D.V. Djonin, and V.K. Bhargava. Optimal and suboptimal packet scheduling over correlated time varying flat fading channels. *IEEE Transactions on Wireless Communications*, 5(2):446–456, 2006a.

A.K. Karmokar, D.V. Djonin, and V.K. Bhargava. POMDP-based coding rate adaptation for type-I hybrid ARQ systems over fading channels with memory. *IEEE Transactions on Wireless Communications*, 5(12):3512–3523, 2006b.

V. Kawadia and P. R. Kumar. A cautionary perspective on cross-layer design. *IEEE Wireless Communications*, 12(1):3–11, 2005.

J. G. Kim and M. M. Krunz. Delay analysis of selective repeat ARQ for a Markovian source over wireless channel. *IEEE Transactions on Vehicular Technology*, 49(5):1968–1981, 2000.

L. Kleinrock. *Queuing Systems*, volume I. Wiley, New York, 1975.

L. Kleinrock and S. S. Lam. Packet switching in a multiaccess broadcast channel: Performance evaluation. *IEEE Transactions on Communications*, 23(4).

J. N. Laneman, D. N. C. Tse, and G. W. Wornell. Cooperative diversity in wireless networks: efficient protocols and outage behavior. *IEEE Transactions on Information Theory*, 50(12):3062–3080, 2004.

L.B. Le, E. Hossain, and A.S. Alfa. Service differentiation in multirate wireless networks with weighted Round-Robin scheduling and ARQ-based error control. *IEEE Transactions on Communications*, 54(2):208–215, 2006a.

L.B. Le, E. Hossain, and A.S. Alfa. Radio link level performance evaluation in wireless networks using multi-rate transmission with ARQ-based error control. *IEEE Transactions on Wireless Communications*, 5(10):2647–2653, 2006b.

L.B. Le, E. Hossain, and T. Le-Ngoc. Interaction between radio link level truncated ARQ, and TCP in multi-rate wireless networks: a cross-layer performance analysis. *IET Communications*, 1(5):821–830, 2007.

Q. Liu, S. Zhou, and G. B. Giannakis. Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links. *IEEE Transactions on Wireless Communications*, 3(5):1746–1755, 2004.

Q. Liu, S. Zhou, and G. B. Giannakis. Queuing with adaptive modulation and coding over wireless links: cross-layer analysis and design. *IEEE Transactions on Wireless Communications*, 4(3):1142–1153, 2005a.

Q. Liu, S. Zhou, and G. B. Giannakis. Cross-layer scheduling with prescribed QoS guarantees in adaptive wireless networks. *IEEE Journal on Selected Areas in Communications*, 23(5):1056–1066, 2005b.

S.P. Lloyd. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

M. Marcon, M. Dischinger, K.P. Gummadi, and A. Vahdat. The local and global effects of traffic shaping in the Internet. *Proceedings COMSNETS*, pages 1–10, 2011.

M. Mardani, J. Seifali Harsini, F. Lahouti, and B. Eliasi. Link-adaptive and QoS-provisioning cooperative ARQ - applications to relay-assisted land mobile satellite communications. *IEEE Transactions on Vehicular Technology*, 60(7): 3192 –3206, 2011.

J. Max. Quantization for minimum distortion. *IRE Transactions on Information Theory*, 6:7–12, 1960.

L. Mokdad and J. Ben-Othman. Adaptive traffic shaping for WiMAX networks. *Proceedings ComComAp*, pages 187 –192, 2012.

A. Nosratinia, T.E. Hunter, and A. Hedayat. Cooperative communication in wireless networks. *IEEE Communications Magazine*, 42(10):74 – 80, 2004.

A.J. Paulraj, D.A. Gore, R.U. Nabar, and H. Bölcskei. An overview of MIMO communications - a key to gigabit wireless. *Proceedings of the IEEE*, 92(2): 198–218, 2004.

M. Poggioni, L. Rugini, and P. Banelli. Analyzing performance of multi-user scheduling jointly with AMC and ARQ. *Proceedings GLOBECOM*, pages 3483–3488, 2007.

M. Poggioni, L. Rugini, and P. Banelli. QoS analysis of a scheduling policy for heterogeneous users employing AMC jointly with ARQ. *IEEE Transactions on Communications*, 58(9):2639–2652, 2010.

L. K. Rasmussen, E. Uhlemann, and F. Brännström. Concatenated systems and cross-layer design. *Proceedings Australian Communication Theory Workshop*, pages 80 – 86, 2006.

S. Ren and K.B. Letaief. Maximizing the effective capacity for wireless cooperative relay networks with QoS guarantees. *IEEE Transactions on Communications*, 57(7):2148 –2159, 2009.

A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity. Part I. System description. *IEEE Transactions on Communications*, 51(11):1927 – 1938, 2003a.

A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity. Part II. Implementation aspects and performance analysis. *IEEE Transactions on Communications*, 51(11):1939 – 1948, 2003b.

S. Shakkottai, T. S. Rappaport, and P. C. Karlsson. Cross-layer design for wireless networks. *IEEE Communications Magazine*, 41(10), 2003.

F. Shi and D. Yuan. Cross-layer combination of cooperative HARQ with AMC in wireless ad-hoc networks. *Proceedings ICCS*, pages 896 –900, 2008.

V. Srivastana and M. Motani. Cross-layer design: a survey and the road ahead. *IEEE Communications Magazine*, 43(12):112–119, 2005.

A. Svensson. An introduction to adaptive QAM modulation schemes for known and predicted channels. *IEEE Transactions on Communications*, 95(12):2322–2336, 2007.

Ch.C. Tan and N.C. Beaulieu. Infinite series representations of the bivariate Rayleigh and Nakagami-*m* distributions. *IEEE Transactions on Communications*, 45(10):1159–1161, 1997.

Ch.C. Tan and N.C. Beaulieu. On first-order Markov modeling for the Rayleigh fading channel. *IEEE Transactions on Communications*, 48(12):2032–2040, 2000.

J. Tang and X. Zhanh. Cross-layer modeling for quality of service guarantees over wireless links. *IEEE Transactions on Wireless Communications*, 6(12): 4504–4512, 2007.

W. Turin and R. van Nobelen. Hidden Markov modeling of flat fading channels. *IEEE Journal on Selected Areas in Communications*, 16(9):1809–1817, 1998.

W. Turin and M. Zorzi. Performance analysis of delay-constrained communications over slow Rayleigh fading channels. *IEEE Transactions on Wireless Communications*, 1(4):801–807, 2002.

S. Vegesna. *IP Quality of Service.* Cisco Press, 2001.

Ch. Wang, Y. Fan, I. Krikidis, J.S. Thompson, and H.V. Poor. Superposition-coded concurrent decode-and-forward relaying. *Proceedings ISIT*, pages 2390 –2394, 2008a.

F. Wang, A. Ghosh, C. Sankaran, P. Fleming, F. Hsieh, and S. Benes. Mobile WiMAX systems: performance and evolution. *IEEE Communications Magazine*, 46(10):41–49, 2008b.

H.S. Wang and N. Moayeri. Finite state Markov channel – a useful model for radio communication channels. *IEEE Transactions on Vehicular Technology*, 44(1):163–171, 1995.

X. Wang, Q. Liu, and G.B. Giannakis. Analizing and optimizing adaptive modulation coding jointly with ARQ for QoS-guaranteed traffic. *IEEE Transactions on Vehicular Technology*, 56(2):710–720, 2007.

D. Wetteroth. *OSI reference model for telecommunications.* McGraw Hill, 2001.

D. Wu and R. Negi. Effective capacity: a wireless link model for support of quality of service. *IEEE Transactions on Wireless Communications*, 2(4):630–643, 2003.

G. Yu, Z. Zhang, and P. Qiu. Cooperative ARQ in wireless networks: Protocols description and performance analysis. *Proceedings ICC*, 8:3608 –3614, 2006.

Q. Zhang and S.A. Kassam. Finite state Markov model for Rayleigh fading channels. *IEEE Transactions on Communications*, 47(11):1688–1692, 1999.

X. Zhang and Q. Du. Cross-layer modeling for QoS-driven multimedia multicast/broadcast over fading channels in mobile wireless networks. *IEEE Communications Magazine*, 45(8):62–70, 2007.

Y.J. Zhang and K.B. Letaief. Cross-layer adaptive resource management for wireless packet networks with OFDM signaling. *IEEE Transactions on Wireless Communications*, 5(11):3244–3254, 2006.

M. Zorzi. On the analytical computation of the interference statistics with applications to the performance evaluation of mobile radio systems. *IEEE Transactions on Communications*, 45(1):103–109, 1997.