Fall 2009

Cooperative retransmission protocols in fading channels : issues, solutions and applications

Igor Stanojev
New Jersey Institute of Technology

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

COOPERATIVE RETRANSMISSION PROTOCOLS IN FADING CHANNELS: ISSUES, SOLUTIONS AND APPLICATIONS

by
Igor Stanojev

Future wireless systems are expected to extensively rely on cooperation between terminals, mimicking MIMO scenarios when terminal dimensions limit implementation of multiple antenna technology. On this line, cooperative retransmission protocols are considered as particularly promising technology due to their opportunistic and flexible exploitation of both spatial and time diversity. In this dissertation, some of the major issues that hinder the practical implementation of this technology are identified and pertaining solutions are proposed and analyzed. Potentials of cooperative and cooperative retransmission protocols for a practical implementation of dynamic spectrum access paradigm are also recognized and investigated. Detailed contributions follow.

While conventionally regarded as energy efficient communications paradigms, both cooperative and retransmission concepts increase circuitry energy and may lead to energy overconsumption as in, e.g., sensor networks. In this context, advantages of cooperative retransmission protocols are reexamined in this dissertation and their limitation for short transmission ranges observed. An optimization effort is provided for extending an energy-efficient applicability of these protocols.

Underlying assumption of altruistic relaying has always been a major stumbling block for implementation of cooperative technologies. In this dissertation, provision is made to alleviate this assumption and opportunistic mechanisms are designed that incentivize relaying via a spectrum leasing approach. Mechanisms are provided for both cooperative and cooperative retransmission protocols, obtaining a meaningful upsurge of spectral efficiency for all involved nodes (source-destination link and the relays).

It is further recognized in this dissertation that the proposed relaying-incentivizing schemes have an additional and certainly not less important application, that is in dynamic spectrum access for property-rights cognitive-radio implementation. Provided solutions avoid commons-model cognitive-radio strict sensing requirements and regulatory and taxonomy issues of a property-rights model.

COOPERATIVE RETRANSMISSION PROTOCOLS IN FADING CHANNELS: ISSUES, SOLUTIONS AND APPLICATIONS

by
Igor Stanojev

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering

Department of Electrical and Computer Engineering

January 2010

APPROVAL PAGE

COOPERATIVE RETRANSMISSION PROTOCOLS IN FADING CHANNELS: ISSUES, SOLUTIONS AND APPLICATIONS

Igor Stanojev

Dr. Yeheskel Bar-Ness, Dissertation Advisor — Date
Distinguished Professor of Electrical and Computer Engineering, NJIT

Dr. Umberto Spagnolini, Committee Member — Date
Professor of Dipartimento di Elettronica e Informazione, Politecnico di Milano

---

Dr. Roy Yates, Committee Member — Date
Professor of Electrical and Computer Engineering, Rutgers University

---

Dr. Alexander M. Haimovich, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Osvaldo Simeone, Committee Member — Date
Assistant Professor of Electrical and Computer Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Igor Stanojev

Degree: Doctor of Philosophy

Date: January 2010

Undergraduate and Graduate Education:

- Doctor of Philosophy in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2010
- Doctor of Philosophy in Information Engineering,
  Politecnico di Milano, Milano, IT, 2010
- Master of Science in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2006
- Bachelor of Science in Electrical Engineering,
  University of Belgrade, Belgrade, Serbia and Montenegro, 2001

Major: Electrical Engineering

Publications:

I. Stanojev, O. Simeone, U. Spagnolini, Y. Bar-Ness and R. Pickholtz, "Cooperative ARQ via Auction-Based Spectrum Leasing," *IEEE Trans. Commun.*, second revision round.

I. Stanojev, O. Simeone, Y. Bar-Ness and D. Kim, "Energy Efficiency of Non-Collaborative and Collaborative Hybrid-ARQ Protocols," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 326-335, Jan. 2009.

O. Simeone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini and R. Pickholtz, "Spectrum Leasing to Cooperating Secondary Ad Hoc Networks," *IEEE Journ. Selected Areas Commun.*, vol. 26, no. 1, pp. 203-213, Jan 2008.

I. Stanojev, O. Simeone, Y. Bar-Ness and C. You, "Performance of Multi-Relay Collaborative Hybrid-ARQ Protocols over Fading Channels," *IEEE Commun. Lett.*, vol. 10, no. 6, pp. 522-524, July 2006.

I. Stanojev, G. Verticale and P. Giacomazzi, "A Negotiation Approach for Pricing the Wireless Access," *IEEE International Conference on Communications, ICC 2010*, to appear.

I. Stanojev and U. Spagnolini, "Interference-Avoidance-Based Distributed Synchronization in Cellular Wireless Networks," *in Proc. IEEE GLOBECOM 2009.*

I. Stanojev, O. Simeone, U. Spagnolini, Y. Bar-Ness and R. Pickholtz, "An Auction-Based Incentive Mechanism for Non-Altruistic Cooperative ARQ via Spectrum-Leasing," *in Proc. IEEE GLOBECOM 2009.*

I. Stanojev, O. Simeone, Y. Bar-Ness and M. Cho, "On the optimal number of hops in linear wireless ad hoc networks with hybrid ARQ," *in Proc. 6th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, WiOpt 2008.*

I. Stanojev, O. Simeone, Y. Bar-Ness and T. Yu, "Spectrum Leasing via Distributed Cooperation in Cognitive Radio," *in Proc. IEEE International Conference on Communications, ICC 2008.*

I. Stanojev, O. Simeone and Y. Bar-Ness, "Optimal Design of a Multi-Antenna Access Point with Decentralized Power Control Using Game Theory," *in Proc. IEEE Symposium on Dynamic Spectrum Access, DySPAN 2007.*

I. Stanojev, O. Simeone, Y. Bar-Ness and D. Kim, "On the Energy Efficiency of Hybrid ARQ Protocols in Fading Channels," *in Proc. IEEE International Conference on Communications, ICC 2007.*

I. Stanojev, O. Simeone and Y. Bar-Ness, "Performance Analysis of Collaborative Hybrid-ARQ Incremental Redundancy Protocols over Fading Channels," *in Proc. IEEE International Workshop on Signal Processing Advances for Wireless Communications, SPAWC 2006.*

I. Stanojev, O. Simeone and Y. Bar-Ness, "Performance Analysis of Collaborative Hybrid-ARQ Protocols over Fading Channels," *in Proc. IEEE Sarnoff Symp. 2006.*

To my family

ACKNOWLEDGMENT

It is unfair to numerous people that only my name is written on the front page of this work. Each line in this dissertation hides some precious research advice, careful but strong guidance, an encouraging word or a friendly moment.

I primarily want to express the gratitude to my advisors, Prof. Yeheskel Bar-Ness of NJIT and Prof. Umberto Spagnolini of Politecnico di Milano, whose contribution is not limited to research guidance and spans much more than this dissertation. I would particularly thank Prof. Bar-Ness for insisting and showing me the meaning of the phrase "hard-work" and the benefits of a disciplined approach to both research and life, for these helped me become a better and stronger person. Prof. Spagnolini showed me by example that each day, each moment, is an opportunity for learning and improving oneself. It was a true pleasure to work with such an enthusiastic person. To both of them, I am honored for their confidence in granting me the desire to be the first student to enroll in the joint NJIT - Politecnico di Milano PhD program.

Special thanks go to the committee members. Prof. Roy Yates of Rutgers University showed an immense insight and provided crucial remarks that both improved the quality of this dissertation and gave directions for future development. As a member of the Center for Wireless Communications and Signal Processing Research (CwCSPR), Prof. Alexander Haimovich was present during both my MS and PhD studies at NJIT, leaving a strong mark on my academic days. To Prof. Osvaldo Simeone of NJIT I owe the warmest thanks for his indispensable role in shaping me as a researcher. Without him nothing would be the same, and I am honored to call him a friend.

Funding of my dissertation was partially provided by the Ross Fellowship Memorial Fund, and the Elisha Yegal Bar-Ness Fellowship.

I wish that one day I can somehow repay Prof. Nirwan Ansari of NJIT. Whenever I had difficulties, whether academic, administrational or financial, he would always find

time, patience and precious advice, not rarely backing me up with his authority. I also appreciate the help of Office of Graduate Studies (NJIT), in particular Dr. Ronald Kane and Mrs. Clarisa Gonzalez-Lenahan, as well as the personnel of Office of International Students (NJIT), Mr. Scott Kline and Mr. Jeffrey Grundy. They would never roll their eyes when I would just bump into their office, with no appointment and out of walk-in hours.

Many thanks are due to Prof. Paolo Giacomazzi and Dr. Giacomo Verticale of Politecnico for their research contributions and for providing me the opportunity of tutoring young researchers, which now become my newly found passion.

Center for Wireless Communications and Signal Processing (CwCSPR) at NJIT, as well as Dipartimento di Elettronica e Informazione (DEI) at Politecnico were wonderful places to work in. To all the students from these labs, Nicola, Marco, Tariq, Bo, Cip, Vlad, Chen, Andrew, Jason, Laurie, MC, Jonathan, Stefano, Daniele, Domenico, Ilaria, Silvia, Teng, Razvan and many others, thanks for all the moments of joy, research debates and, most of all, for being my friends. A very special gratitude is reserved for Mrs. Marlene Toeroek, CwCSPRŠs power behind the throne, for her company and versatile help, whether in logistic support or providing a personal advice.

To Dr. Maryann McCoul, I owe more than the words can say. She has provided me with continued support, wisdom and, dare to say, friendship. Special appreciations are for Ozgur, the lady and now professor who will forever stay my dear friend. To Jelena, thanks for making me more complete and for showing me a new and enriched meaning to life. To John, Carol, Patty and Mickey, thanks for being my second family. Far away from my birthplace, I found a place where I feel home.

Finally, I am infinitely indebted to my sister Tanja (and Giunti) and Mom and Dad, to whom I dedicate this work. Their love and unconditional support carried me each day and each moment and helped me reach this far.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

Recent years have witnessed a steady growth in Internet access demand and, in particular, bandwidth consuming applications such as video streaming, gaming and data transfer. While this demand is readily supported by the highly mature wired (optical) communication technology, the customer requirements are turning to a "anywhere, anytime" and mobile access to information, thus emphasizing the need for efficient wireless (radio) communications. Wireless medium, however, imposes major communication challenges such as fading and shared access which, although the research in this field has achieved remarkable results, are yet to be addressed in a manner that would provide throughputs comparable to the wired media.

Among effective solutions for exploiting the channel and network structure for the purpose of improving the wireless transmission quality, *cooperative retransmission protocols* [1] have emerged as a promising technique. These protocols provide a synergy between two important communication paradigms, cooperation from neighboring terminals [2] [3] and opportunistic reallocation of spectrum resources to terminals with a satisfactory channel state [4]. In particular, a source-destination link applying these protocols might require assistance for possible retransmissions from one or more available neighboring terminals that were able to decode the original transmission [1]. Thus, the cooperative transmission is prescribed only if needed in an opportunistic fashion. Cooperative retransmission protocols significantly benefit from diversity gain, that is from transmission via multiple uncorrelated or loosely correlated channels. Namely. terminal cooperation enables spatial, while the time diversity is utilized through possible retransmissions. It is reemphasized that the opportunistic nature of these protocols guarantees a rational utilization of resources and diversity schemes.

This chapter is organized as follows. Cooperative retransmission protocols are described in Section 1.1. Issues related to these protocols, highlighted in Section 1.2, provide the basic motivation for this dissertation. A notion of dynamic spectrum access and cognitive radio is discussed in Section 1.3, where possible alternative applications of cooperative retransmission protocols are identified. The overview of the dissertation is provided in Section 1.4.

1.1 Cooperative Retransmission Protocols

In the following, cooperative communications and retransmission protocols are introduced separately. Two concepts are then merged to elaborate on cooperative retransmission protocols.

1.1.1 Cooperative Communications

Channel fading, which is a consequence of multipath signal propagation, presents a major challenge for implementation of broad capacity wireless networks. A crucial tool for mitigating fading while preserving the spectral resources (time and frequency) is spatial diversity, carried out through multiple-antenna placement at the source and/or the destination terminals. Built upon the concept of multiple collocated antennas, the Multiple-Input-Multiple-Output (MIMO) technology emerged and have by now provided efficient theoretical solutions [5]. For efficient operation of MIMO, however, antennas need to sufficiently spaced, typically one half of the wavelength, to enable transmission via spatially uncorrelated channels. This constraint often exceeds terminal dimensions, thus often limiting the multiple-antenna implementation to base stations and access points.

Cooperative networks with practical single-antenna stations provide an interesting alternative for exploiting spatial diversity, while avoiding the terminal size issue (see e.g., [2], [3]). Due to shared nature of the wireless medium, the information transmitted from one station can be overheard by any surrounding wireless station (Figure 1.1-a)). These

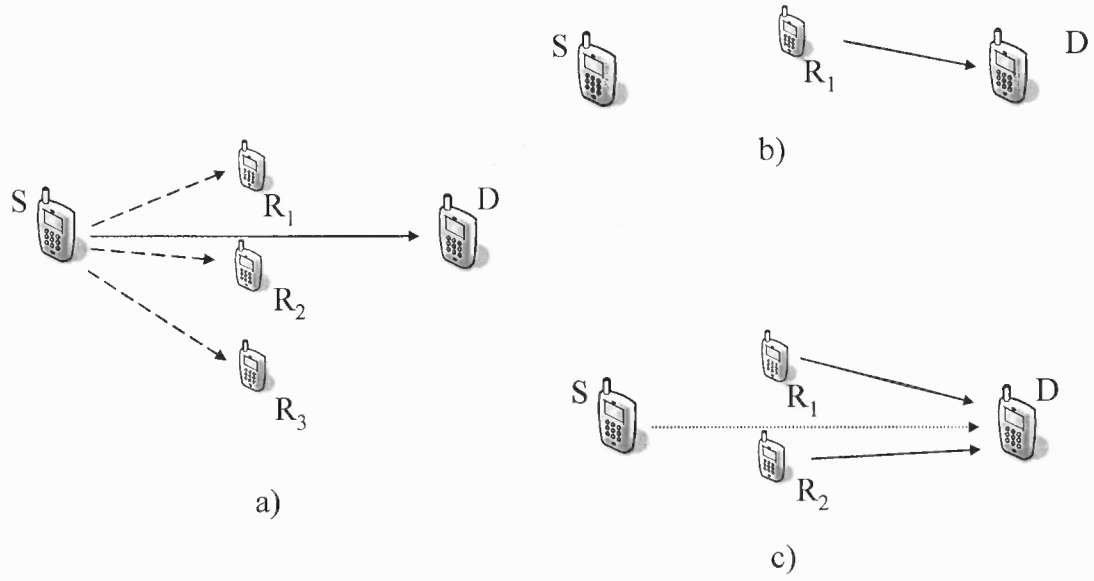

Figure 1.1 Possible cooperation scenarios: a) broadcast phase, b) cooperation phase - hopping (relay R_1 selected), c) cooperation phase - distributed space-time coded transmission (source is possibly involved).

surrounding stations can assist the current transmission acting as relays, by forwarding the received signal to the intended destination. Possible cooperation scenarios are illustrated in Figure 1.1: cooperation using a single relay (hopping) (Figure 1.1-b)) and relaying through a distributed antenna array formed by the relays and possibly source or the destination (Figure 1.1-c)). The former presents the most simple cooperation pattern, hopping, which is also a basic building block for *multihop routing*, whereby a packet is sequentially routed from source to destination through a series of hops (relays). In the latter case, multiple transmitters need to perform distributed beamforming or space-time (ST) coding. Notice in (Figure 1.1-c)) that the source can also participate as a transmitter during cooperation phase. In this dissertation, both scenarios will be considered.

Relaying can be also classified based on the received signal processing at the relays. In this dissertation, only Decode and Forward (DF) scheme is used, requiring that the relay successfully decodes the source codeword, before reencoding and forwarding it. If the error

is detected at the relay, the latter sustains from any retransmission and possibly listens the following transmission for another decoding attempt. It is worth noting that other relaying schemes, such as the Amplify and Forward (AF), are also used for relaying. In AF, the relays simply amplify whatever information they received from the source, and rebroadcast it. Notice that this scheme suffers from the noise enhancement, since together with the information, the noise at the relay receiving antenna is also amplified. Furthermore, it is noted that this scheme suffers from the power amplifier saturation. On the other hand, processing requirements for this scheme are relaxed as it requires no signal processing at the relays.

1.1.2 Retransmission (HARQ) Protocols

While deployed to increase the robustness of a wireless signal, techniques at the Physical (PHY) and Medium Access Control (MAC) layers, such as coding and error correction mechanisms, cannot guarantee error-free communication. Erroneous packet retransmission mechanism, Automatic Repeat Request (ARQ) was originally embedded in MAC layer solely to guarantee a (close-to) error-free packet transmission (it is noted that higher protocol layers, such as Transport and Application layers, also need to provide mechanisms for retransmission of erroneous data undetected at MAC layer).

With evolution of wireless technology, potential of ARQ mechanisms was recognized as one of the most promising tools for *cross-layer* design, whereby the functionalities at the MAC and PHY layers are jointly devised in order to protect communication from, and exploit the properties of, a wireless channel more efficiently. In particular, on fading channels, ARQ protocols provide the means for mitigating fading impairments through time diversity at the expense of delay [6]. Since ARQ protocols exploit additional time resource only when necessary (i.e., in the case of an error event at the destination), they are more efficient than basic time-diversity schemes, whereby the information is coded across a predetermined number of coherence intervals [7].

In this dissertation it is assumed that the system sustains from transmission of any following packets until the transmission of a current one is successfully completed. This protocol is also known as the Stop-and-Wait ARQ, according to the classification based on packet flow control patterns [8]. Other, more sophisticated protocols such as Go-Back-N and Selective-Repeat [8], are not considered here. After receiving and processing the packet transmitted from the source, the destination node typically checks the CRC (Cyclic Redundancy Check) header, added to the data bits at the transmitter, to determine whether the packet contains errors. If the packet is damaged, the destination sends a NACK (Not Acknowledge) message toward the source, signaling that an error occurred in the previous transmission, and that retransmission of the packet is required. If the source, within some predefined time, does not receive any message from the destination, it will assume that the transmission was unsuccessful and will retransmit the packet. This cycle proceeds until final successful reception, when the destination signals successful decoding sending an ACK (Acknowledge) message. It is noted that in order to prevent the system outage caused by numerous consecutive unsuccessful retransmission attempts (typically a consequence of a very hostile channel environment), a maximum number of retransmissions is usually predefined. If this number is reached, the retransmission is delayed for the time interval during which the channel is expected to change significantly. Alternatively, the packet is dropped and its retransmission, if required, is left for handling at the higher layers. This approach is commonly known as the *truncated* ARQ.

As mentioned above, transmission on noisy communication channel can lead to numerous unsuccessful retransmissions, thus reducing the system efficiency. To cope with this challenge (besides relying on truncated ARQ), ARQ protocols must be supported with a complementary mechanism that enhance the packet resilience toward channel conditions. In this context, Hybrid-ARQ (HARQ) is designed as an ARQ protocol upgraded with at least Forward Error Control (FEC) protection. Usually a low rate protection code is used in combination with interleaving to reduce the effect of fading and additive noise. When

no further enhancement is used, this merging of FEC and ARQ concepts is labeled as Type I Hybrid-ARQ Protocol (HARQ-TI). Notice that the performances of plain ARQ and HARQ-TI protocols are closely related, since the advantage of FEC technique can be simply parameterized by the coding gain. It is further noted that the information theoretic approach employed throughout this dissertation assumes coding. Plain ARQ protocol is thus not considered.

The motivation for proposing Type II HARQ protocols lies in the inability of HARQ-TI to significantly benefit with retransmissions. Type II HARQ protocols are protocols with memory, using buffers to preserve erroneous packets and combine them in a certain manner with other copies during the detection process, thus increasing reliability with each retransmission. Type II Hybrid-ARQ Chase Combining Protocol (HARQ-CC) [9], also referred to as the Packet Combining, does not introduce additional complexity on the source side, as the plain copies of the original packet are retransmitted upon receiving the NACK message. At the destination side, however, erroneous packets are buffered and MRC (Maximum Ratio Combining) combined with the most recently received packet. It is noted that a relatively inferior 'hard combining', performed through the bit-wise majority voting, can be implemented instead of MRC. In this dissertation only MRC is considered.

Type II Hybrid-ARQ Incremental Redundancy Protocol (HARQ-IR), often called Code Combining, stands for the most sophisticated HARQ protocol. Upon receiving the retransmission request, the source generates new parity bits (different with each retransmission attempt) and transmits them instead of the original packet. At the destination, received versions of packets are concatenated and processed according to the decoding rule. The effect is equivalent to resending the packet protected with lower coding rate with each attempt. Intuitively, performance of HARQ-IR is superior to that of HARQ-CC protocol, which in turn outperforms HARQ-TI, on the account of increased complexity at the receiving and / or transmitting side.

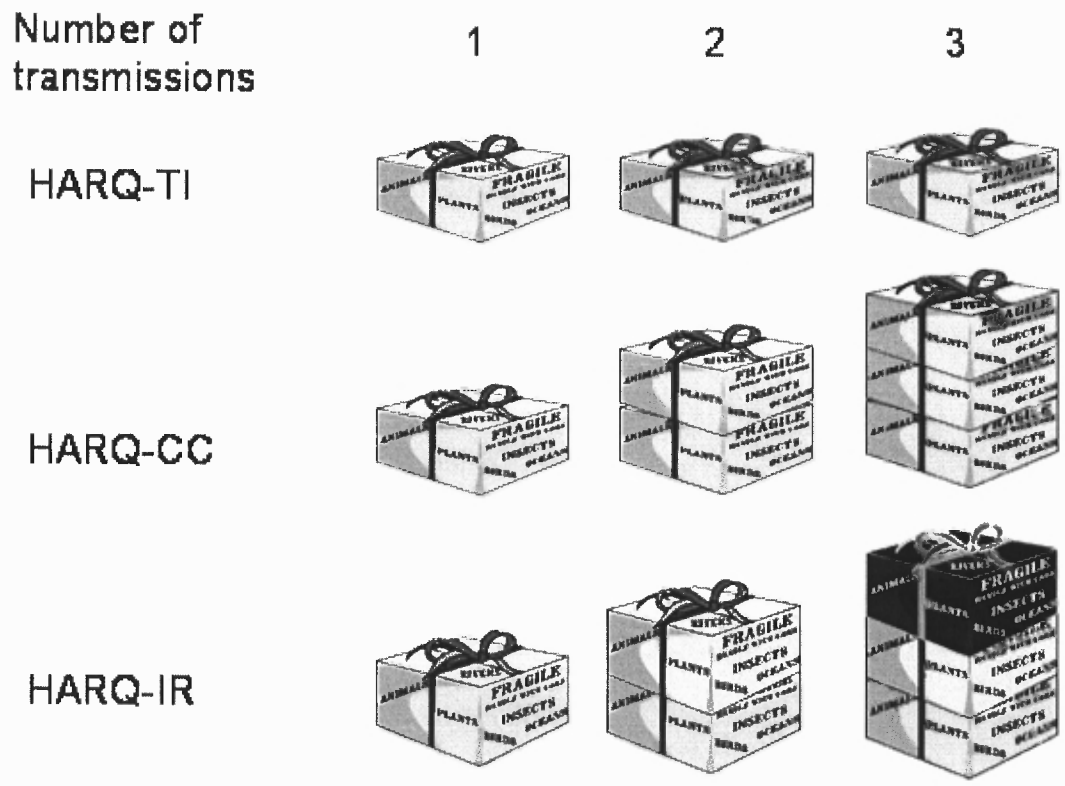

Figure 1.2 Packets available at destination after 1, 2 and 3 transmission attempts, for HARQ-TI, HARQ-CC and HARQ-IR protocols.

1.1.3 Cooperative Hybrid-ARQ Protocols

The idea of cooperative retransmission protocols was originally proposed in [1]. Conventional cooperative schemes commonly assign a dedicated resource (such as a time-slot) to relays to forward the information from the source [10]. In [1] it was recognized that this approach can lead to a possible waste of resources. For example, if the source can deliver its message to the destination without any assistance, the resource assigned to relaying (as well as the relaying power) is needlessly squandered. Similarly, if the channel from the source to a relay is faded, that relay should not be involved. Cooperative HARQ protocols [1] avoid these issues, as the relays are involved only if their assist is needed and if they have decoded the source's original transmission.

Cooperative retransmission protocols operate very similarly to cooperative paradigm illustrated in Figure 1.1. In order for cooperation phase to take part (Figure 1.1-b),-c)), the destination needs to send the NACK message. Otherwise, the destination has already decoded the source message (during the broadcast phase, Figure 1.1-a)) and cooperation and retransmission are not required. Additionally, it may be required for the relays that have successfully decoded the source transmission to signal their availability to the source and, if directed so, to switch from receiving to transmitting mode. Depending on HARQ type, the retransmitted codeword can be a copy of the original packet (HARQ-TI or HARQ-CC) or a new packet consisted of parity bits (HARQ-IR). The destination (and possibly, any remaining receiving relays), decode the data and, as for the non-cooperative HARQ, perform appropriate packet or code-combining with previously received codewords [11] if HARQ-CC or HARQ-IR are implemented, respectively. This procedure repeats until the CRC at the destination reveals successful detection and an ACK message is sent, or, in case of the truncated retransmission protocols, when a predefined maximum number of retransmissions is reached.

It is reemphasized that, besides employing cooperation and attaining spatial diversity, cooperative retransmission protocols also profit from time diversity through possible

retransmissions. Both spatial and time diversity are achieved opportunistically (only if the source cannot communicate with the destination by itself), preserving spectral resources.

1.2 Motivating Issues

Implementation of almost any novel technology is typically hindered by new challenges that emerge with it. Cooperative retransmission protocols are by no means exception to this rule. The most important issues that need to be considered are identified in the remainder of this section. Two of them, namely the energy consumption consideration and the altruistic relaying assumption, are the fundamental drives behind the work carried out in this dissertation. It is noted that all identified issues are not specific to cooperative retransmissions, but are inherited from cooperative and retransmission paradigms.

1.2.1 Energy Consideration

It was emphasized above that cooperative retransmission protocols utilize both spatial and time diversity, due to cooperation and retransmissions, respectively. This gives rise to both spectral- and energy-efficient nature of cooperative retransmission protocols [1]. While the spectral efficiency is undisputable, the energy-efficiency of cooperative retransmission protocols (and, in fact, any technology in modern networks) needs reassessment. Namely, the energy-efficiency issue has conventionally implied *transmission* energy only, which is justified for existing networks typified by relatively large transmission ranges. Many emerging wireless networks, such as ad-hoc and sensor networks, utilize densely spaced nodes so that the energy consumed by the circuitry other than the power amplifier (i.e., the processing electronics and transmit/receive circuitry) becomes of the same or even higher order of magnitude than transmission energy, as depicted in Figure 1.3. The need for a careful reconsideration of energy-efficiency as a metric becomes particularly important in case of wireless sensor networks, wherein the primary concern is indeed on battery life (notice that the spectral efficiency is not of primary importance in sensor networks).

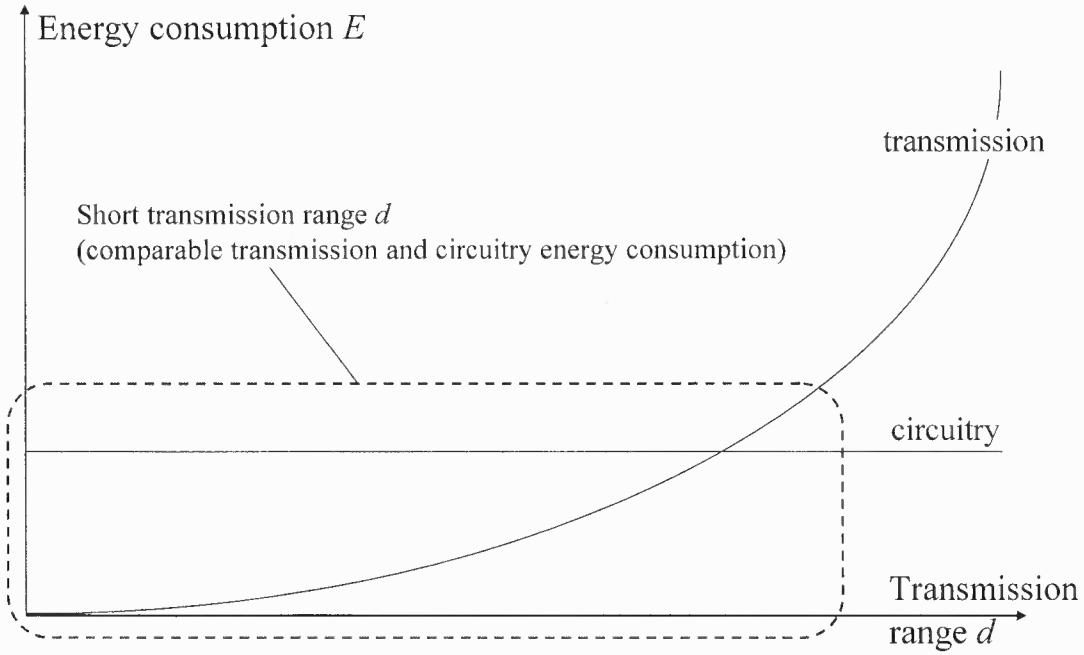

Figure 1.3 For small transmission ranges, circuitry energy is comparable or even dominating the transmission energy.

Both cooperation and retransmissions increase circuitry consumption. Cooperation implies multiple involved terminals (relays in addition to source and destination), each of which increases the overall circuitry consumption (while, in general, decreasing transmission energy [12] [13]). Similarly, each retransmission increases circuitry consumption (while decreasing transmission energy [1] [6] [14]). This trade-off is illustrated in Figure 1.4 (notice that Figure 1.4 is merely an illustration and that additional relays and transmissions generally have a quantitatively different influence (depending on channel gains, fading, etc.)). In case of cooperative retransmission protocols, where both patterns are employed, energy consumption issue becomes more complex and needs further investigations. This task will be carried out in Chapter 4.

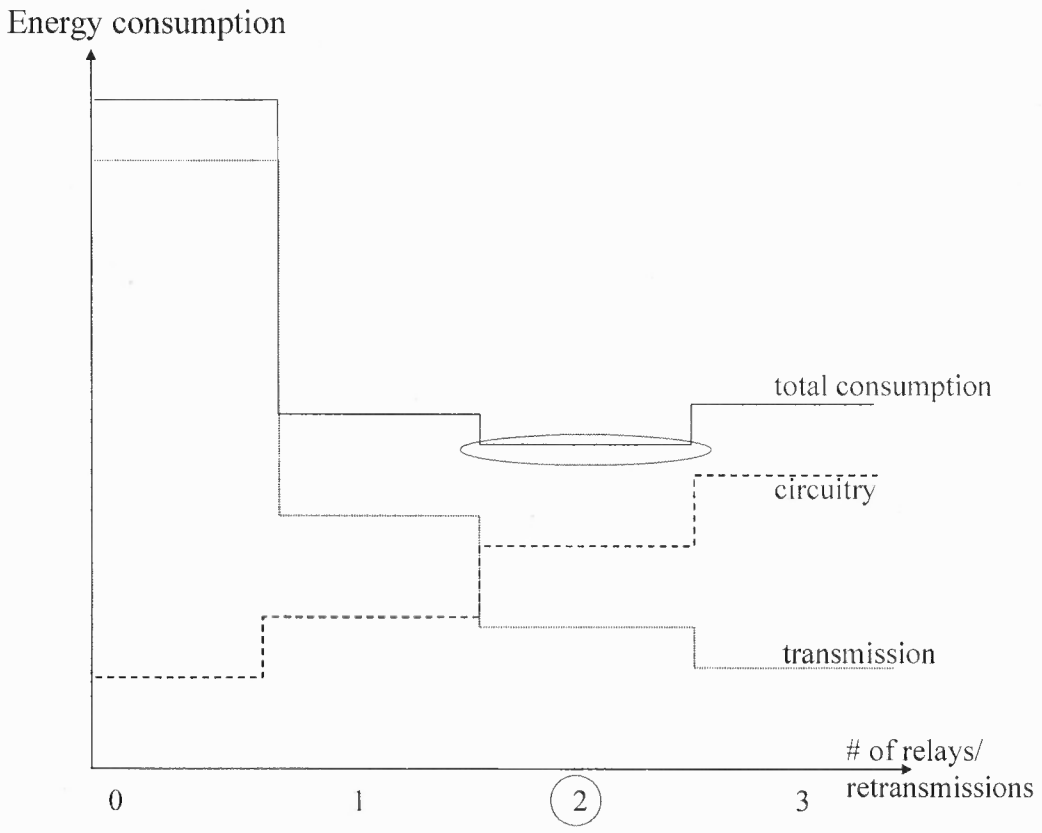

Figure 1.4 Cooperation and/or retransmissions can increase overall energy consumption

1.2.2 Altruistic Relaying Consideration

While cooperative paradigm presents a promising alternative for multiple-antenna technology, its implementation is by now limited to networks that can deploy dedicated relays, that is to the networks with infrastructure. In case of infrastructureless networks, which are becoming increasingly popular today, cooperation is not entirely justified as it implies an altruistic willingness of users' terminals to assist the communication they do not directly benefit from. Limitations of this assumption are emphasized in the case of terminals that are frequently used for relaying (e.g., due to convenient placement). Clearly, these terminals would be of a short operating life due to the battery drainage and, if designed intelligently, would refuse to relay.

Shortly, altruistic relaying assumption is the most critical drawback of cooperative technology and a satisfying solution is yet to be provided. Although an amount of research has been carried out on the topic [15]- [17], these works generally rely on mechanisms such as reputation, credit or node punishment, all of which require a long operational time horizon in order to enforce cooperation. In this context, a method is required that is more opportunistic and that provides a dynamic foundation for motivating the cooperation. Such a mechanism is proposed in Chapter 7 for purely cooperative networks. Although the scheme in Chapter 7 is by itself a very promising solution, in Chapter 8 it is shown that an even more practical mechanism can be engineered if resorting to cooperative retransmission protocols.

1.2.3 Additional Issues

Besides the two issues highlighted above, there are several other challenges related to cooperative HARQ protocols that need to be addressed before this technology can be fully deployed in the wireless standards. As with the energy-efficiency and relaying motivation, these additional challenges are not exclusive to cooperative retransmission protocols. For example, in the case of virtual antenna array (scenario in Figure 1.1-c),

design of distributed ST coding techniques is a concern. While signaling can ensure proper cooperation among multiple transmitting nodes, randomized ST coding can be a solution to avoid extensive messaging [18]. Terminal synchronization is required on both carrier (frequency) and symbol (timing) level, as well as on the block level in case of distributed ST block coding, for cooperation to provide full benefits [3]. It is noted in [3] that, given some form of network block synchronization (e.g., through periodic transmission of known synchronization prefixes), carrier and symbol synchronization can be built upon it [3]. Synchronization challenge further gives rise to signaling requirements and the consequent loss of resources, as well as designing the signaling protocol. To conclude, security and privacy are of major concern [19]. While these concerns require serious attention and have been under extensive research, they are out of scope of this dissertation.

1.3 Application: Dynamic Spectrum Access (Property-Rights Cognitive-Radio)

Spurred by the evidence that the current spectrum allocation granting exclusive use to licensed services is highly inefficient and that new wireless communication technologies allow effective spectrum sharing, *cognitive radio* emerged as a new paradigm for efficient spectrum utilization [20]. This principle has already inspired technological solutions and standardization efforts [21] [22]. The lively debate around this concept has by now broadened its scope to include substantially different technologies and solutions. The identifying feature, which seems to be common to different schools of thought on the subject, is the coexistence on the same spectral resource of both licensed (or primary) and unlicensed (or secondary) terminals and services [23]. Among the different debated positions, two main approaches to cognitive radio have emerged [23]- [25]:

- Commons model: according to this framework, primary terminals are oblivious to the presence of secondary users, thus behaving as if no secondary activity was present. Secondary users, instead, sense the radio environment in search of spectrum holes

(portions of the bandwidth where primary users are not active) and then exploit the detected transmission opportunities.

- Property-rights model (or spectrum leasing): here, primary users own the spectral resource and possibly decide to lease part of it to secondary users in exchange for appropriate remuneration.

While the stringent sensing requirements (namely, it is required that not only transmitted signal is detected, but the activity of receiving circuitry also) make the implementation of the commons model a challenging engineering problem [20], the property-rights model has been seldom analyzed in the communication literature on the grounds that its implementation is mostly a regulatory issue that hinges on the definition of a pricing model for spectrum leasing [24]. On this line, it is recognized here that the schemes proposed in Chapter 7 and Chapter 8, besides providing an attractive solution for the relaying motivation, can be conversely seen as a practical framework for the implementation of cognitive radio networks running according to the property-rights spectrum-leasing model. As will be elaborated later, the role of the primary node is played by the original source (and destination) and that of the secondary nodes by the relaying nodes. Moreover, retribution from secondary to primary nodes upon leasing is in the form of cooperation to the primary transmission. This enables on-the-air decisions and avoids the regulatory issues or money transactions that commonly hinder the implementation of the spectrum leasing concept.

1.4 Dissertation Outline

Dissertation preliminaries are introduced in Chapter 2. Therein, generic system model (block-fading channel) is provided along with the metrics used throughout this dissertation - channel capacity, outage probability, expected number of transmissions and throughput. Chi-square distribution, heavily exploited in this dissertation, is also briefly introduced.

The remainder of Chapter 2 is dedicated to a brif turorial of *game theory*, a mathematical concept that provides powerful tools for designing distributed networks with independent nodes.

Contributions of this dissertation are contained in Chapters 3-8. Particularly, Chapter 3 and Chapter 4 investigate cooperative retransmission protocols in a centralized scenario. Chapter 5 and Chapter 6 illustrate several applications of basic game-theoretic concepts to the communication scenario. Chapter 7 and Chapter 8 then discuss the cooperative and cooperative retransmission protocols, respectively, in a decentralized environment, applying the game-theoretic tools to elaborate on relaying motivation concepts. Material in these two chapters can be also considered as a promising framework for dynamic spectrum access based on the spectrum-leasing property-rights cognitive-radio model. Detailed descriptions of Chapters 3-8 follows.

In Chapter 3, cooperative and retransmission paradigms are contested. Relying on a simple cooperation model - linear multihop network, the following question is answered: given a delay-tolerant network, what should this delay be exploited for, hopping or (HARQ-TI and HARQ-CC) retransmissions? It is demonstrated that for practically encountered signal-to-noise ratio values, it is the combination of two patterns rather than exclusive one that guarantees the best performance in terms of achievable throughput.

Energy efficiency of truncated retransmission protocols in a single-user link (i.e., non-cooperative HARQ), or with the inclusion of a relay station (i.e., cooperative HARQ) is analyzed in Chapter 4 for HARQ-TI, HARQ-CC and HARQ-IR. The total energy consumption accounts for both the transmission energy and the energy consumed by the transmitting and receiving electronic circuitry of all involved terminals (source, destination and, possibly, the relay). Using the transmission time and transmission energy of each packet as optimization variables, the overall energy is minimized under an outage probability constraint. It is shown, for instance, that, if the circuitry energy consumption is not negligible, selection of the transmission energy is not only dictated by

the outage constraint, but is also significantly affected by the need to reduce the number of retransmissions. Results also demonstrate the performance limitations of cooperative HARQ protocols in terms of energy efficiency, when circuitry consumption is properly accounted for.

Chapter 5 illustrates a possible application of game-theory to communication networks. Research is motivated by the fact that intercell interference control is a crucial but hardly met requirement for a dense frequency reuse in cellular systems. Namely, due to large signalling overhead, multicell processing and resource scheduling are not practicable solutions, at least today. A decentralized transmission scheduling scheme where each base station exploits the knowledge of the intercell interference to locally allocate the resources is proposed. Each base station schedules the access to time or frequency resource so as to mitigate the generated interference and maximize its goodput. In this decentralized approach, the intercell signalling is replaced by the level of interference estimated locally and independently within each cell. Game-theoretic concept is used to study equilibria and each scheduler is model as a player that locally maximizes its objective (goodput) while interacting (or interfering) with others.

Chapter 6 can also be considered as an extensive introduction to the following chapters. Namely, it introduces Stackelberg game application to decentralized communication networks, and the concept of power control games. An uplink scenario with independent and rational terminals and an access point is considered. The optimal design of a multi-antenna access point in such a scenario is investigated by modelling the interaction between the access point on one side, and the distributed set of terminals on the other, as a Stackelberg game. As a game leader, the access point determines the network parameters (bandwidth and the number of receiving antennas) for the power control game played between the terminals (follower), so as to maximize the network utility per system resource (bandwidth and antennas). Two game models are considered, whereby the network utility is measured either in terms of power minimization or power efficiency maximization.

It is shown that a larger number of users motivates the provider (i.e., access point) to invest, as the overall performance enhancement well balances the costs. In certain decentralized scenarios, however, the system cannot efficiently cope with large amount of user. Furthermore, the trade-off between investing in different system resources, bandwidth and antennas, is observed.

In Chapter 7, provision is made to alleviate the assumption that available relays are willing to assist the ongoing transmission in an altruistic fashion. A scheme is proposed whereby a source can lease its spectrum to an ad hoc network of potentially relaying nodes in exchange for cooperation in the form of distributed space-time coding. On one hand, the source maximizes its achievable transmission rate, accounting for the possible contribution from cooperation. On the other hand, nodes in the ad hoc network compete among themselves for transmission within the leased time-slot following a distributed power control mechanism. The investigated model is conveniently cast in the framework of Stackelberg games. Analysis and numerical results show that the proposed mechanism achieves significant rate improvements for both the source-destination link and the motivated relaying terminals. It is further noted that the scheme can be also considered as a framework for property-rights cognitive radio implementation, based on trading secondary spectrum access for cooperation to the primary.

A novel distributed scheme that combines cooperative retransmission protocols with the spectrum leasing paradigm is proposed and analyzed in Chapter 8. The strategy harnesses the opportunistic gains of cooperative communications, while inherently providing a spectrum-rewarding incentive for the otherwise non-cooperative relays to assist the source's transmission. In this context, the scheme can be considered as extension of the proposal in Chapter 7, removing the extensive informational requirements and providing a fully decentralized solution. As in cooperative HARQ, the source might decide to hand over the possible retransmission slots to nearby stations that were able to decode the original transmission. In the proposed scheme, however, in exchange for the cooperation, the

relaying station is also awarded an opportunity to exploit the retransmission slot for its own traffic. Arbitration of relays' retransmissions is performed via an auction mechanism, with the source, the competing relays and the transmission slot acting as the auctioneer, the bidders and the bidding article, respectively. Auction theory (more generally, the theory of Bayesian games) is applied to analyze the scheme performance. Again, the setting here can be alternatively seen as a practical framework for implementation of property rights cognitive radio networks. Numerical results and analysis show that the proposed scheme enables an efficient dynamic resource allocation that provides relevant gains (e.g., transmission reliability) for both the original source (primary) and the cooperating nodes (secondary users).

CHAPTER 2

PRELIMINARIES

This chapter provides the most important concepts that are applied throughout this dissertation. It first details on the generic system model by describing the block fading Rayleigh model and introducing the metrics such as the channel capacity (achievable transmission rate) and the outage probability for three HARQ protocols and general expressions for expected number of transmissions and system throughput. The second part of the chapter is dedicated to the basics of game-theory, a crucial mathematical tool for design and analysis of distributed networks.

2.1 Generic Model and Metrics

2.1.1 Block Rayleigh Fading Model

A communication channel in which adjacent symbols (a block of symbols) are affected by the same fading value is commonly referred to as a *block-fading channel.* This channel model is applicable to a range of scenarios typified by low-speed (e.g., walking-speed and below) mobile terminals as in, e.g., sensor, indoor or personal communication system. With slowly moving terminals, the channel gain, albeit random, varies slowly enough with time that it can be assumed as constant along a block [26]. Block-fading channel also assumes, as is the case in this dissertation, statistically independent fading blocks (i.e., changing independently with each block), implying transmission intervals that are sufficiently separated in time as in, e.g., a time-division system, in frequency as in, e.g., a multicarrier system or both in time and in frequency, e.g., with slow time–frequency hopping. A constant (during a block) complex channel gain $h = \mathrm{Re}(h) + j\,\mathrm{Im}(h)$ is modeled as an independent zero-mean complex Gaussian random variable with variance $g_h = E[|h|^2]$, where $E[\cdot]$ denotes the expectation operator. While the channel gain can

be described by numerous distributions, e.g., Rayleigh, exponential and gamma, here the chi-square distribution is most frequently exploited. Thus, $|h|^2/g_h$ is a random variable with chi-square distribution with two degrees of freedom, $|h|^2 \sim \mathcal{X}_2^2$. Of particular interest is the cumulative density function of chi-square distribution with ν degrees of freedom, where ν is an even integer [27]:

$$\begin{aligned} F(x,\nu) &= \frac{\gamma(\nu/2, x/2)}{\Gamma(\nu/2)} \\ &= 1 - e^{-x/2} \sum_{i=0}^{\nu/2-1} \frac{(x/2)^i}{i!}, \end{aligned} \tag{2.1}$$

where $\gamma(\lambda,\mu) = (\lambda-1)!\left(1 - e^{-\mu}\sum_{i=0}^{\lambda-1}\frac{\mu^i}{i!}\right)$ is the incomplete Gamma function and $\Gamma(\lambda) = (\lambda-1)!$ is the ordinary Gamma function (note that these definitions of Gamma functions hold only for integer λ, i.e., for an even ν).

2.1.2 Channel Capacity

Channel *capacity* C, here also referred to as a (maximum) *achievable rate*, is defined as a maximum transmission rate for which one can drive the probability of erroneous communication arbitrarily close, but not necessarily exactly, to zero [28]. Denoting the transmitted and received signals as X and Y, respectively, capacity is given in terms of mutual information $I(X;Y)$

$$\begin{aligned} C &= \max_{P_X(\cdot), E[X^2]\leq P} I(X;Y) \\ &= \max_{P_X(\cdot), E[X^2]\leq P} H(Y) - H(Y|X), \end{aligned} \tag{2.2}$$

where H denotes the entropy [28], $P_X(\cdot)$ denotes the distribution of X and $E[X^2] \leq P$ is the power (variance) constraint of transmitted signal. For the Gaussian channel, relation between X and Y reads

$$Y = X + Z, \tag{2.3}$$

where $Z \sim \mathcal{N}(0, N_0)$ is a zero-mean complex additive Gaussian noise with single-sided spectral density N_0. Since $H(Y|X) = H(Z)$, recalling that the entropy of a zero-mean complex Gaussian variable with variance N_0 is $\log_2(2\pi e N_0)$ [28], the (2.2) becomes

$$C = \max_{P_X(\cdot), E[X^2] \leq P} H(Y) - \log_2(2\pi e N_0). \tag{2.4}$$

Resorting to the maximum entropy theorem [28], the above expression is maximized for $P_X(\cdot) \sim \mathcal{CN}(0, P)$ (thus, Y is also Gaussian $Y \sim \mathcal{CN}(P + N_0)$) and

$$\begin{aligned} C &= \log_2\left(2\pi e\left(P + N_0\right)\right) - \log_2(2\pi e N_0) \\ &= \log_2\left(1 + \frac{P}{N_0}\right), \end{aligned} \tag{2.5}$$

which is a famous result for capacity of the Gaussian channel (Shannon limit [26]). Shannon limit has been nearly achieved in the last decade. In fact, it is within a range of a tenth of a dB that the capacity achieving codes such as low-density parity-check (LDPC), turbo and repeat-accumulate (RA) codes can approach this limit [29]. Moreover, schemes exist for which the (relatively small) gap towards the limit can be presented simply via a constant power loss [30].

For the block-faded channel, (2.5) becomes (assuming matched filtering)

$$C = \log_2\left(1 + \frac{|h|^2 P}{N_0}\right). \tag{2.6}$$

For interference (or multiple access channel MAC) channel, denoting the ith transmitter-receiver channel as h_i, h_{ji} as the interference channel gain from the jth transmitter to the ith receiver (or to the common receiver in MAC channel), transmitting power of the ith node as P_i, interference for the ith receiver reads $I_i = \sum_{j \neq i} |h_{ji}|^2 P_j$ and, with the assumption that the latter is Gaussian, the achievable rate for the ith transmitter-receiver pair is

$$C_i = \log_2\left(1 + \frac{|h_i|^2 P_i}{N_0 + \Sigma_{j \neq i} |h_{ji}|^2 P_j}\right).$$

Retransmission Protocols Denoting the channel gain in ith (re)transmission as h^i, capacity achievable with HARQ-TI (recall that the erroneous packets received in previous retransmissions are discarded) after n retransmissions reads

$$C_{n,TI} = \max_{i=1,..n} \log_2 \left(1 + \left|h^i\right|^2 \frac{P}{N_0}\right). \tag{2.7}$$

For HARQ-CC, preserved retransmitted copies of the packet are soft combined at the receiver so that the rate reads [1]

$$C_{n,CC} = \log_2 \left(1 + \sum_{i=1}^{n} \left|h^i\right|^2 \frac{P}{N_0}\right). \tag{2.8}$$

Notice that the power summation in (2.8) is the consequence of soft (MRC) packet combining. For HARQ-IR, code-combining yields sum of information [11] [1]

$$C_{n,IR} = \sum_{i=1}^{n} \log_2 \left(1 + \left|h^i\right|^2 \frac{P}{N_0}\right) \tag{2.9}$$

As discussed in [31], the analysis of HARQ-IR protocols can be greatly simplified if an upper bound approximation, obtained by applying the Jensen inequality, is used

$$C_{n,IR} \leq n \log_2 \left(1 + \frac{1}{n} \sum_{i=1}^{n} \left|h^i\right|^2 \frac{P}{N_0}\right) \tag{2.10}$$

As demonstrated in [31] (and showed in this dissertation when encountered), this is a relatively tight upper bound for the capacity of HARQ-IR protocols.

For the case of cooperative retransmission protocols, capacity expressions become slightly more complicated and will be addressed as required.

2.1.3 Outage Probability

Recall that the capacity defined in Section 2.1.2 refers to a Gaussian (during a block) channel. Since this is in fact a random variable (depending on channel gain), the capacity of fading channel in Shannon sense (ergodic) does not exist [1] [26] [32]. Instead, one should resort to a non-ergodic metric *outage probability,* defined as the probability that the

(instantaneous Gaussian) channel capacity is lower than the rate r of the code used at the transmitter (thus, it is the probability of erroneous decoding at the destination),

$$p = \Pr\{C < r\}.$$

For the three HARQ protocols and capacities derived in Section 2.1.2, the outage probability reads

$$\begin{aligned} p_{n,TI} &= \Pr\left\{\max_{i=1,..n} |h^i|^2 < \frac{2^r-1}{P/N_0}\right\} && \text{(2.11a)} \\ &= F\left(\frac{2^r-1}{g_h P/N_0}, 2\right)^n \\ p_{n,CC} &= \Pr\left\{\sum_{i=1}^{n} |h^i|^2 < \frac{2^r-1}{g_h P/N_0}\right\} && \text{(2.11b)} \\ &= F\left(\frac{2^r-1}{P/N_0}, 2n\right) \\ p_{n,IR} &\geq \Pr\left\{\sum_{i=1}^{n} |h^i|^2 < n\frac{2^{r/n}-1}{g_h P/N_0}\right\} && \text{(2.11c)} \\ &= F\left(n\frac{2^{r/n}-1}{g_h P/N_0}, 2n\right), \end{aligned}$$

where the fact was used that the sum of squares of i independent Gaussian variables of a unit variance is a chi-square distributed random variable with i degrees of freedom (Section 2.1.1).

Expected Number of Transmissions and Throughput Expected number of transmissions, noted herein as $\bar{n}$ or $E[n]$, is a commonly used metric to quantify the HARQ performance. It is defined as the number of retransmissions, including the original transmission, required for successful decoding at the destination. It is thus strongly related to the outage probability and reads

$$\begin{aligned} \bar{n} &= \sum_{i=1}^{\infty} i\,(p_{i-1} - p_i) && \text{(2.12)} \\ &= 1 + \sum_{i=1}^{\infty} p_i, \end{aligned}$$

for non-truncated retransmission protocols and

$$\begin{aligned} \bar{n} &= \sum_{i=1}^{n-1} i\left(p_{i-1}-p_i\right)+np_n \\ &= 1+\sum_{i=1}^{n-1} p_i, \end{aligned} \tag{2.13}$$

for n-truncated (i.e., with n as a maximum delay) HARQ protocols. In (2.12) and (2.13), the fact was used that $p_{i-1}-p_i$ is the probability of successful decoding at the destination exactly at the ith transmission (with $p_0 = 1$) [11]. For the truncated protocols (2.13), the average number of retransmissions $\bar{n}$ is written as the sum of two terms, where the first accounts for the events in which successful decoding occurs in one of the first $n-1$ transmissions, and the second term corresponds to the complementary events in which successful decoding does not occur during the first $n-1$ transmissions (including the event of successful transmission exactly at the nth attempt, and the outage event).

The throughput for retransmission protocols is defined as the number of successfully transmitted bits per second (if r is also defined in bit/s) and, using renewal theory [11], reads

$$\begin{aligned} R &= r\frac{1-p_n}{\bar{n}} \\ &= r\frac{1-p_n}{1+\Sigma_{i=1}^{n-1}p_i}. \end{aligned} \tag{2.14}$$

2.2 Game-Theoretic Concepts

Lately, game theory [33] was recognized as a promising paradigm for modelling performance of wireless networks that involve multiple nodes not controlled by some central authority [34]. As these independent nodes (players in the game-theoretic jargon) have goals that are usually (but not necessarily) in conflict with each other, their selfish behavior might lead to extremely poor network performance. In particular, players are often defined as selfish and rational: the selfish player is interested solely in maximizing its own benefit, without concern for the collective good, while the rational player chooses only

those strategies that are best responses to his opponents' strategies. Game theory allows to predict the possible outcomes of interaction (game) between these competitive terminals, in terms of Nash Equilibria (NE) [33]. Therefore, it is a powerful tool for defining a set of rules to be enforced on the players that would lead to more desirable outcome. This section covers some of the game-theoretic fundamentals to be used throughout this dissertation.

2.2.1 Utility Function

A dominant approach to modeling independent players' interests is *utility theory*, quantifying a player's degree of preference across a set of available alternatives. Specifically, a utility function refers to the level of satisfaction the decision-taker receives as a result of its actions and is defined as a function that assigns a numerical value to the elements of the action set A ($u : A \rightarrow \mathcal{R}^1$) if for all $x, y \in A$, x is at least as preferred compared to y if and only if $u(x) \geq u(y)$ [33].

Acting optimally in an uncertain environment gets considerably complicated in the presence of two or more utility-maximizing players whose actions can affect each other's utilities. Non-cooperative game theory studies such a scenario (notice that the term "non-cooperative" does not necessarily imply conflicting players' interests, although it is the most appealing scenario). In this dissertation, cooperative (coalitional) game theory that assumes group-based decision-making or modeling, is not considered.

2.2.2 Normal Form

In game theory, the normal (or strategic) form is the most frequent game representation, wherein all possible game outcomes depend only on the players' combined actions. It is noted in [35] that even more complex game forms, such as Bayesian that accounts for environment randomness, or extensive-form games (that involve time dimension) can be reduced to a normal form. Specifically, a normal-form game is a tuple $\langle N, A, u \rangle$, where

- N is a finite set of n players

- A is actions domain, $A = A_1 \times \cdots \times A_n$, where A_i is a set of actions available to ith player. Moreover, action profile is denoted as $a = (a_1, ..., a_n) \in A$
- $u = (u_1, ..., u_n)$ where $u_i : A \rightarrow \mathcal{R}$ is a real-valued utility (payoff) function for ith player.

In this dissertation, the focus is entirely on *pure strategies*, referring to a game concept wherein a player selects a single action. Notice that pure strategy corresponds to an action. It is worth noting that another strategical concept exists, namely *mixed strategies*, wherein a strategy corresponds to randomization over an actions set with a given distribution. The latter concept is not considered in this work.

2.2.3 Games and Optimality

Optimality in game theory (similarly as in multi-objective optimization) is a rather vague notion. For example, a social optimum (e.g., weighted sum-utility) can imply a suboptimal performance for an individual player. Even from an individual player's point of view optimality is unclear as the payoff value depends not only on that player's strategy but also involves strategies chosen by other players. In the following the emphasis is on an individual perspective, yielding a fundamental game-theoretical concept, Nash equilibrium (NE).

Denote the strategy profile $s = (s_1, .., s_n)$ in the form $s = (s_i, s_{-1})$ where $s_{-i} = (s_1, ..., s_{i-1}, s_{i+1}, ..., s_n)$ is the set of strategies chosen by players other than player i. In case player i is aware of s_{-i}, its *best response* (not necessarily unique) that maximizes its utility is $s_i^* \in S_i$ such that $u_i(s_i^*, s_{-i}) \geq u_i(s_i, s_{-i})$ for any $s_i \in S_i$. However, it is unreasonable to assume a player that avails the knowledge of s_{-i} (if so, game-theoretic concepts would trivially reduce to a single-objective optimization problem). In this context, one needs an extension of the best response concept, which is a Nash equilibrium. In particular, a strategy profile s^* is a Nash equilibrium if s_i^*, for any $i = 1, .., n$ is a best response to s_{-i}^*

. In words, in Nash equilibrium any unilateral change of any player's strategy would not introduce additional payoff to that player.

2.2.4 Remarks

It is noted here that the above introduction to game-theoretic concepts is by no means complete. It is rather a brief summary of the concepts employed within this dissertation. Several other important notions, such as, e.g., Pareto dominance [33] [35], are not covered. Methods for determining existence, uniqueness and derivation of NE (e.g., the fixed point theorem and contraction analysis [33]), as well as the possible algorithms for reaching NE, are also not presented here, but introduced if and when necessary in this dissertation.

CHAPTER 3

THROUGHPUT CONSIDERATION: TO COOPERATE, RETRANSMIT OR BOTH?

In this chapter two fundamental technologies discussed throughout this presentation, the cooperative and the HARQ paradigm, are confronted. Using the multihop cooperation framework, it is investigated under which operational regimes (i.e., SNR values) each of these two paradigms proves superior in terms of achievable throughput. It will be shown that, unless operating in extreme low or high SNR regimes, it is not a specific paradigm, but their synergy that provides the optimal network performance.

The following section details on the underlying *linear network* model, a simplified framework but a powerful analytical tool for analysis of multihop networks. Existing research results based on this model are provided and the results of this chapter are placed in the corresponding perspective.

3.1 Background and the Chapter Overview

Characterized by low-powered single-antenna terminals and often lacking any supporting infrastructure, ad-hoc and sensor networks typically operate in the low signal-to-noise ratio (SNR) regime and, in general, need to rely on cooperation between stations in order to cover broader regions [36]. The simplest cooperative transmission approach is *multihopping* (a basic block is illustrated in Figure 1.1-b)), whereby a packet is sequentially routed from source to destination through a series of hops. In general, the optimal topology for a multihop network is the one with equidistantly placed terminals on the line between source and destination 3.1. This linear network, although rarely encountered in practice and considered to be rather optimistic, allows for a more tractable analysis and establishment of some of the fundamentals of multihop networks [37].

Multihop transmission for a linear ad-hoc network was investigated in [37], where it was demonstrated that multihop, both with or without spatial reuse[1], is advantageous (in terms of power efficiency) in the low-SNR regime, thanks to power gains or low interference (compared to the noise level), but fails for large values of SNR, due to interference limitations or throughput reduction (as a consequence of multiple transmissions). The approximate optimal number of hops in such a network, but without spatial reuse, is given in [38] by the same authors. The analysis of [37] and [38] is limited to Gaussian (unfaded) channels. In [39], a linear multihop network with no spatial reuse is analyzed under a quasi-static fading assumption: an upper bound on the probability of outage is found and exploited, along with the result of [38], to determine the optimal number of hops. Finally, in [40], the authors discuss the impact of various classes of per-hop memoryless retransmission protocols on linear multihop network performance. Namely, the main goal of [40] is to determine the statistics of the overall number of per-hop transmissions in the system, based on a Gilbert-Elliot model for the wireless channels.

In this chapter, the same model of a linear multihop network as in [37]- [39] is adopted. Specifically, a quasi-static fading without spatial reuse, as in [39], is considered and analysis therein extended by including HARQ retransmission protocols. It is noted that, with respect to [40], the setting herein differs in terms of a wireless channel model (quasi-static Rayleigh instead of Gilbert-Elliot of [40]); more importantly, unlike [40], the focus in this chapter is on the problem of optimal network design along the lines of [37]- [39].

The chapter answers the following question: given the maximum allowed delay and signal-to-noise ratio, what is the optimal number of hops that maximizes the end-to-end throughput? The preference between the multihop and HARQ technologies directly follows, as the difference between delay and the number of hops is the number of

[1]Spatial reuse refers to a multihop scheme whereby the terminals are allowed to transmit simultaneously, so as to increase the end-to-end throughput. For a network with terminals operating in a half duplex regime, neighboring terminals are not allowed to transmit at the same time.

transmission slots reserved for possible retransmissions. Analytical framework is provided for setting of the optimization problem, while the problem itself is solved using numerical methods. It is interesting to notice that the results obtained in this chapter qualitatively confirm the main conclusions of [37]- [39] (e.g., relative multihop gain in low- and high-SNR regimes), notwithstanding the differences in the underlying models, namely, the use of HARQ protocols in this chapter to cope with quasi-static fading channels. Most importantly, the results herein show that, in general, combination of both multihopping and HARQ retransmissions achieves an optimal network performance.

3.2 System Analysis

3.2.1 System Model

Consider a linear $k-$hop wireless network, consisting of the source N_1, destination N_{k+1} and $k-1$ relays, $N_2, .., N_k$, equidistantly placed on the line connecting N_1 and N_{k+1}, as depicted in Figure 3.1. The nodes operate in half-duplex, and only one node can transmit at a given slot, i.e., no spatial reuse is allowed. Thus, a packet originating from the source and intended for the destination is routed through each of the k hops in a separate time-slot and the successful transmission requires at least k time-slots. Furthermore, the overall delay tolerated by the network, measured in transmission slots, is $L \geq k$. The additional $L-k$ transmissions slots are used (if necessary) for retransmissions via HARQ-TI or HARQ-CC protocols on the hops that failed to support successful packet delivery. The channel gains $h_i^{(t)}$, where i is the number of the hop and t is the number of the retransmission attempt (including the original transmission, $t = 1$) on that hop, are modelled as independent proper complex Gaussian random variables with unit power. That is, the model assumes Rayleigh block-fading, whereby the channel gain is constant during a slot but changes independently with each slot, as discussed in Section 2.1.1. All nodes transmit with equal power, and the signal-to-noise ratio of the single-hop system ($k = 1$) is SNR. Denoting

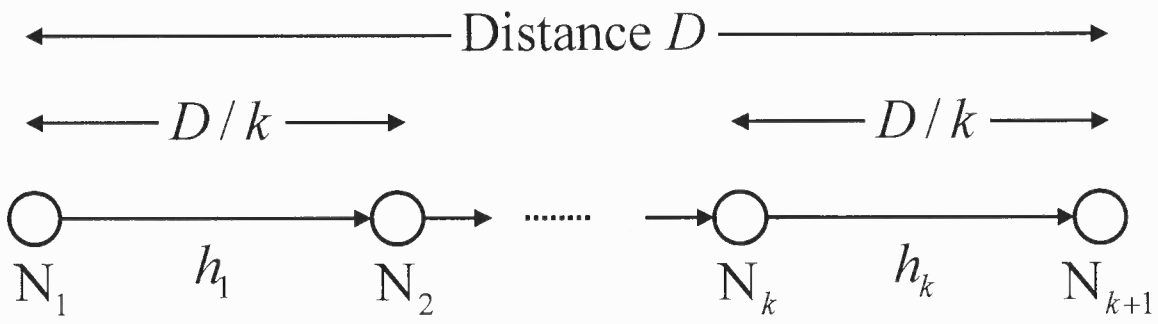

Figure 3.1 Linear wireless $k-$hop network.

the path-loss exponent as η, the signal-to-noise ratio on any of the hops of a k-hop network is then $SNR_k = SNR \cdot k^{\eta}$.

3.2.2 Analysis

As stated earlier, the goal here is to determine the optimal number of hops k, given the maximum delay L and SNR. The end-to-end throughput, defined as the average number of successfully transmitted bits per slot, is employed as the performance criterion. This section is largely dedicated to the derivation of the throughput and its relation to parameters k, L and SNR. Once this relation is established, the problem of solving for optimal k, i.e., $k_{opt}(L, SNR)$ is discussed in Section 3.2.3.

Using renewal theory, the overall throughput R can be shown to be (irrespective of the employed HARQ protocol) [11]

$$R = \frac{r\,(1 - p_L)}{\bar{l}}, \tag{3.1}$$

where r is the rate (in $bit/s/Hz$) of the original transmission (transmission rate); p_L is the probability of outage after L slots, i.e., the probability that after L slots the destination node N_{k+1} still did not successfully decode the packet; and $\bar{l}$ is the average number of slots per packet used for end-to-end transmission, $k \leq \bar{l} \leq L$. It is noted that a packet is dropped

(i.e., outage event occurs) if the maximum number of retransmissions L is reached and the packet is still not correctly decoded.

The average number of slots $\bar{l}$ in (3.1) can be written as

$$\begin{aligned} \bar{l} &= k + \sum_{l=k}^{L} (l-k)\,(p_{l-1} - p_l) + (L-k)p_L \\ &= k + \sum_{l=k}^{L-1} p_l, \end{aligned} \tag{3.2}$$

where the fact is exploited that $p_{l-1} - p_l$ is the probability of the successful end-to-end transmission achieved after exactly l slots [11]. Using the previous equation, the end-to-end throughput R in (3.1) can be conveniently expanded as

$$R = r\frac{1-p_L}{k + \sum_{l=k}^{L-1} p_l}. \tag{3.3}$$

In order to derive the probability of outage after l transmission slots p_l, the auxiliary probability $p_S(a_i)$ is defined as the probability the ith hop delivers successfully a packet from the node N_i to the node N_{i+1} after exactly $a_i \geq 1$ (integer) slots. Denoting $\sum_{\mathcal{A}_j}(\cdot)$ as the summation over all the tuples in the set $\mathcal{A}_j = \{(a_1, .., a_k) \mid a_i \in \mathcal{N}, a_1 + .. + a_k = j\}$, then $\sum_{\mathcal{A}_j} \prod_{i=1}^{k} p_S(a_i)$ is the probability of successful end-to-end transmission after a total delay of exactly j slots. Furthermore, since the outage after l transmission slots excludes all events that would lead to successful transmission within l slots (i.e., $j = k, .., l$), the probability of outage p_l can be written as

$$p_l = 1 - \sum_{j=k}^{l} \sum_{\mathcal{A}_j} \prod_{i=1}^{k} p_S(a_i). \tag{3.4}$$

Furthermore, the probability $p_S(a_i)$ can also be expressed in terms of the probability of unsuccessful transmission in the ith hop after a_i slots, $p_e(a_i)$ [11]

$$p_S(a_i) = p_e(a_i - 1) - p_e(a_i). \tag{3.5}$$

Notice that, in order to calculate the throughput R in (3.3), according to (3.4) and (3.5), it suffices to calculate the probability of unsuccessful transmission in the ith hop after a_i transmission slots, $p_e(a_i)$, which depends on employed HARQ protocols. As discussed in Section 2.1.3, the event of unsuccessful per-hop transmission after a_i transmission slots can be defined as the event wherein the rate achievable after a_i transmissions on the ith hop is smaller than the transmission rate r. The following two subsections, Section 3.2.2 and Section 3.2.2, are dedicated to the throughput derivation for HARQ-TI and HARQ-CC protocols, respectively.

HARQ-TI Since the HARQ-TI protocol is memoryless, i.e., the erroneously received packets are dropped, the probability $p_e(a_i)$ for HARQ-TI can be written as (recall Section 2.1.3)

$$\begin{aligned} p_{e,TI}(a_i) &= \Pr\left\{ r > \max_{t=1,..,a_i} \log_2\left(1+\left|h_i^{(t)}\right|^2 SNR_k\right)\right\} \\ &= \Pr\left\{ \max_{t=1,..,a_i} \left|h_i^{(t)}\right|^2 < \frac{2^r-1}{SNR_k}\right\} \\ &= \Pr\left\{ \left|h_i^{(t)}\right|^2 < \frac{2^r-1}{SNR_k}\right\}^{a_i}. \end{aligned} \tag{3.6}$$

Notice that (3.6) assumes that the terminals are using capacity-achieving Gaussian codebooks. Recalling that $h_i^{(t)}$ is a complex Gaussian random variable, it follows that $|h_i^{(t)}|^2$ is an exponentially distributed random variable, and, finally, $p_{e,TI}(a_i)$ becomes

$$p_{e,TI}(a_i) = \left(1-e^{-\mu(r,k,SNR)}\right)^{a_i}, \tag{3.7}$$

where

$$\mu(r,k,SNR) = \frac{2^r-1}{2SNR\cdot k^\eta}. \tag{3.8}$$

In the following, the dependence of μ on r, k and SNR is dropped for notation convenience.

Applying (3.7) to (3.5) leads to

$$p_{S,TI}(a_i) = e^{-\mu}\left(1-e^{-\mu}\right)^{a_i-1}. \tag{3.9}$$

Then, using (3.4), the probability of outage after l slots, p_l reads

$$\begin{aligned} p_{TI,l} &= 1-\sum_{j=k}^{l}\sum_{\mathcal{A}_j}\prod_{i=1}^{k} p_{S,TI}(a_i) \\ &= 1-\sum_{j=k}^{l}\sum_{\mathcal{A}_j} e^{-k\mu}\left(1-e^{-\mu}\right)^{j-k}, \end{aligned} \tag{3.10}$$

where the fact was used that for the set $\mathcal{A}_j$, $\sum_{i=1}^{k} a_i = j$. Furthermore, as shown in the Appendix A, the cardinality of set $\mathcal{A}_j$ is $\alpha_j = |\mathcal{A}_j| = \binom{j-1}{k-1}$. Then, (3.10) can be further simplified as

$$\begin{aligned} p_{TI,l} &= 1-\sum_{j=k}^{l}\alpha_j e^{-k\mu}\left(1-e^{-\mu}\right)^{j-k} \\ &= 1-e^{-k\mu}\sum_{j=0}^{l-k}\alpha_{j+k}\left(1-e^{-\mu}\right)^{j}. \end{aligned} \tag{3.11}$$

With (3.11), the denominator of (3.3), i.e., the average number of exploited transmission slots per packet $\bar{l}$ (3.2), for the HARQ-TI protocol, is easily shown to be

$$\begin{aligned} \bar{l}_{TI} &= k+\sum_{l=k}^{L-1} p_{TI,l} \\ &= L-e^{-k\mu}\sum_{l=k}^{L-1}\sum_{j=0}^{l-k}\alpha_{j+k}\left(1-e^{-\mu}\right)^{j} \\ &= L-e^{-k\mu}\sum_{l=0}^{L-k-1}\alpha_{l+k}(L-k-l)\left(1-e^{-\mu}\right)^{l}. \end{aligned} \tag{3.12}$$

Finally, applying (3.11) and (3.12) to (3.3), the throughput for HARQ-TI protocol reads

$$R_{\mathrm{TI}} = r\frac{e^{-k\mu}\sum_{l=0}^{L-k}\alpha_{l+k}(1-e^{-\mu})^{l}}{L-e^{-k\mu}\sum_{l=0}^{L-k-1}\alpha_{l+k}(L-l-k)(1-e^{-\mu})^{l}}. \tag{3.13}$$

HARQ-CC As discussed in Section 2.1.3, for HARQ-CC protocol, the previously received erroneous packets are preserved and soft-combined at the receiver with the

currently received packet, and the probability $p_e(a_i)$ for HARQ-CC protocol reads

$$\begin{aligned} p_{e,CC}(a_i) &= \Pr\left\{ r > \log_2\left(1 + \sum_{t=1}^{a_i} \left|h_i^{(t)}\right|^2 SNR_k\right)\right\} \\ &= \Pr\left\{ \sum_{t=1}^{a_i} \left|h_i^{(t)}\right|^2 < \frac{2^r - 1}{SNR_k}\right\} \\ &= F(2\mu, 2a_i), \end{aligned} \tag{3.14}$$

where $F(x, \nu)$ is a cumulative distribution function of a chi-square random variable taken at value x with ν degrees of freedom (recall Section 2.1.1), and μ was defined in (3.8). Notice that in (3.14) it was exploited that $\sum_{t=1}^{a_i} |h_i^{(t)}|^2$ is a chi-square random variable with $2a_i$ degrees of freedom. Recalling that a_i is an integer, the probability $p_{e,CC}(a_i)$ can also be written as [27]

$$\begin{aligned} p_{e,CC}(a_i) &= F(2\mu, 2a_i) = \frac{\gamma(a_i, \mu)}{\Gamma(a_i)} \\ &= 1 - e^{-\mu} \sum_{j=0}^{a_i - 1} \frac{\mu^j}{j!}, \end{aligned} \tag{3.15}$$

where $\gamma(a_i, \mu) = (a_i - 1)!\left(1 - e^{-\mu}\sum_{j=0}^{a_i-1}\frac{\mu^j}{j!}\right)$ is the incomplete Gamma function and $\Gamma(a_i) = (a_i - 1)!$ is the ordinary Gamma function (Section 2.1.1).

Using (3.15) in (3.5), it further follows that

$$p_{S,CC}(a_i) = e^{-\mu} \frac{\mu^{a_i - 1}}{(a_i - 1)!}. \tag{3.16}$$

Then, using (3.4), the outage probability becomes

$$\begin{aligned} p_{CC,l} &= 1 - \sum_{j=k}^{l} \sum_{\mathcal{A}_j} \prod_{i=1}^{k} p_{S,CC}(a_i) \\ &= 1 - \sum_{j=k}^{l} \sum_{\mathcal{A}_j} e^{-k\mu} \frac{\mu^{j-k}}{\prod_{i=1}^{k}(a_i - 1)!} \\ &= 1 - e^{-k\mu} \sum_{j=0}^{l-k} \beta_{j+k} \mu^j, \end{aligned} \tag{3.17}$$

where $\beta_l = \sum_{\mathcal{A}_j} \prod_{i=1}^{k} \frac{1}{(a_i-1)!}$. Furthermore, (3.17) can be exploited to determine the denominator of (3.3), i.e., the average number of slots per packet, $\bar{l}$ (3.2), for the HARQ-CC protocol

$$\begin{aligned} \bar{l}_{CC} &= k + \sum_{l=k}^{L-1} p_{CC,l} \\ &= L - e^{-k\mu} \sum_{l=k}^{L-1} \sum_{j=0}^{l-k} \beta_{j+k} \mu^j \\ &= L - e^{-k\mu} \sum_{l=0}^{L-k-1} \beta_{l+k} (L-k-l) \mu^l. \end{aligned} \tag{3.18}$$

Applying (3.17) and (3.18) to (3.3), the throughput R for HARQ-CC protocol finally reads

$$R_{CC} = r \frac{e^{-k\mu} \sum_{l=0}^{L-k} \beta_{l+k} \mu^l}{L - e^{-k\mu} \sum_{l=0}^{L-k-1} \beta_{l+k} (L-l-k) \mu^l}. \tag{3.19}$$

3.2.3 Discussion on System Design

As discussed in [11] [41] [42], in order for HARQ protocols to reach their full potential, an optimal choice of the rate r (i.e., transmission rate) for the given SNR, is mandatory. This is even more relevant for multihop systems that can exploit the increase in per-hop signal-to-noise ratio ($SNR_k = SNR \cdot k^{\eta}$) by increasing their transmission rate so as to compensate for the throughput reduction due to the k transmissions and the absence of spatial reuse. The optimization of rate R in (3.13) and (3.19) over transmission rate r can be stated as

$$r_{opt}(k, L, SNR) = \arg \max_{r \geq 0} R(r, k, L, SNR). \tag{3.20}$$

The solution to this problem demands, to the best of author's knowledge, numerical methods for global optimization; fortunately, (3.20) is a one-dimensional problem and, therefore, relatively easily solved. Interested reader is referred to [11] [41] [42], wherein

the authors present the analysis for the optimal rate r in single-hop networks using the HARQ protocols, under different assumptions and optimization goals.

Having obtained r_{opt} (3.20), the following and final step of the optimization process is to determine the optimal number of hops (relays),

$$k_{opt} = \arg \max_{k=1,2,..L} R(r_{opt}(k, L, SNR), k, L, SNR). \tag{3.21}$$

The problem (3.21) belongs to the integer programming class. It is noted, however, that one can approach it using an exhaustive search over $k \in \{1, 2, .., L\}$. Namely, since the maximum delay L is, in practice, rarely expected to be large, a brute force optimization (3.21) would in general be acceptable in terms of memory and processing demands.

3.3 Numerical Examples

As explained in the previous section, the optimization process exemplified by (3.20)-(3.21) is hardly tractable analytically. In order to get a further insight into the system behavior and understand the properties of optimal design, in this section numerical examples are provided.

Figure 3.2 aims at illustrating the optimization problem (3.20) by presenting the throughput R versus the transmission rate r, for a fixed delay $L = 14$, $1 \leq k \leq 12$, $SNR = -10$ dB and (as used throughout this section, unless explicitly mentioned otherwise) $\eta = 3$ and HARQ-CC protocol (3.19). It can be seen from this figure that, for given k, L and SNR, the throughput R has a quasi-concave shape as a function of rate r and a global maximum. In particular, as it also follows from (3.3), the throughput (for a given k) increases with the transmission rate r, but only up to a point (given by r_{opt} (3.20)) when the negative impact on the probability of outage p_l becomes dominant. Furthermore, note that the optimal transmission rate r_{opt} increases with an increase of the number of hops, due to the enlargement of effective SNR per hop, SNR_k.

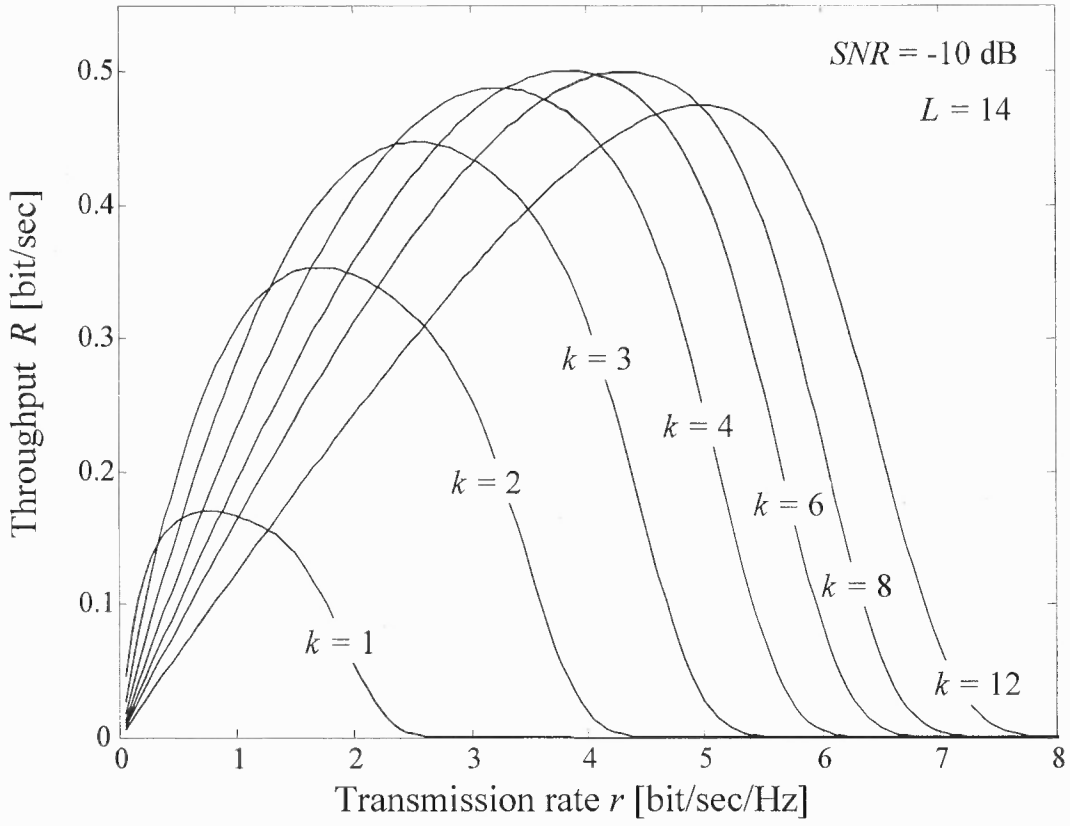

Figure 3.2 HARQ-CC: Throughput R versus transmission rate r and number of hops k, for the delay $L = 14$ and $SNR = -10$ dB.

Having obtained r_{opt} in (3.20), the next step is the optimization over the number of hops (3.21). Figure 3.2 already depicts one important property of this step: the optimized throughput (i.e., the throughput maximized over r) increases with k (i.e., with SNR_k) up to a point, given by k_{opt}, when the number of hops becomes a limiting factor for the throughput. To further shed light on properties of (3.21), Figure 3.3 shows the system throughput versus SNR for fixed $L = 4$, $1 \leq k \leq 3$ and optimized transmission rate r_{opt}. In fact, a larger amount of hops is mostly preferable in the low-SNR region, where the benefit of the effective SNR increase of k^{η} times is particularly important. However, performance of multihop schemes for larger SNR falls behind the single-hop scheme, as the rate becomes limited by the number of hops, and the retransmission protocols become preferable. Notice that these conclusions (obtained through numerical results) are similar to the results in [37], based on a Gaussian (unfaded) model. Furthermore, it is clear (and not shown) that the larger values of propagation-loss exponent η would exercise a positive influence on multihop scheme.

Based on Figure 3.3, Figure 3.4 aims at concluding the discussion on the optimization (3.21), by determining the optimal number of hops k_{opt}, for a given delay L and SNR. It shows that, as the SNR decreases, the system tends to increase the number of hops and, for extremely low SNR, the maximum delay is fully exploited for multihopping, $\lim_{SNR \to 0} k_{opt} = L$. Furthermore, note that for any SNR value (visible for $SNR > -30$ dB in Figure3.4), there exists an upper limit on the optimal number of relays that can improve the system performance, even if an infinite delay is allowed. It is also remarked that the delay values (i.e., L) in Figure 3.4 are extremely large and rarely encountered in practice. However, while the results would not be qualitatively altered for lower values of L, the choice of L in Figure 3.4 is convenient for description of delay-unconstrained system.

Figures 3.2-3.4 described the optimization process and the properties of optimal parameters r_{opt} and k_{opt}. The throughput of optimally designed system, i.e., the system

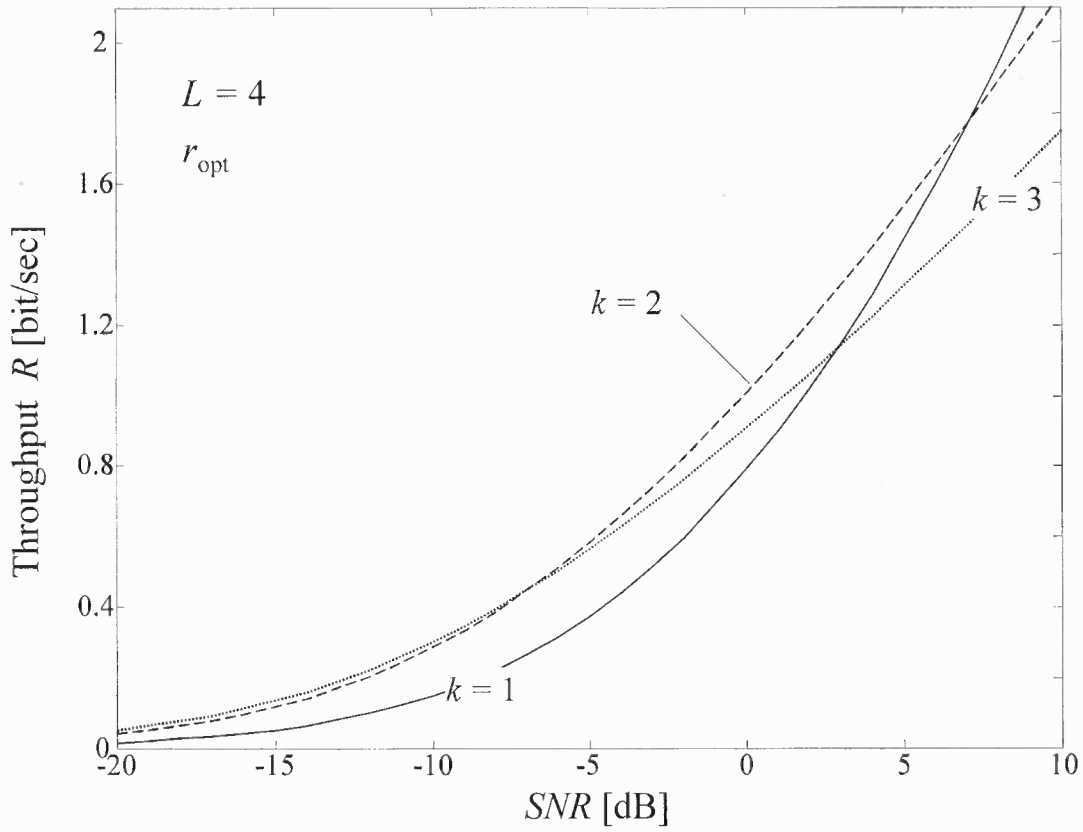

Figure 3.3 HARQ-CC: Throughput R versus SNR, for $L = 4$, $k = 1 \div 3$ and optimized transmission rate $r_{opt}(k, L, SNR)$.

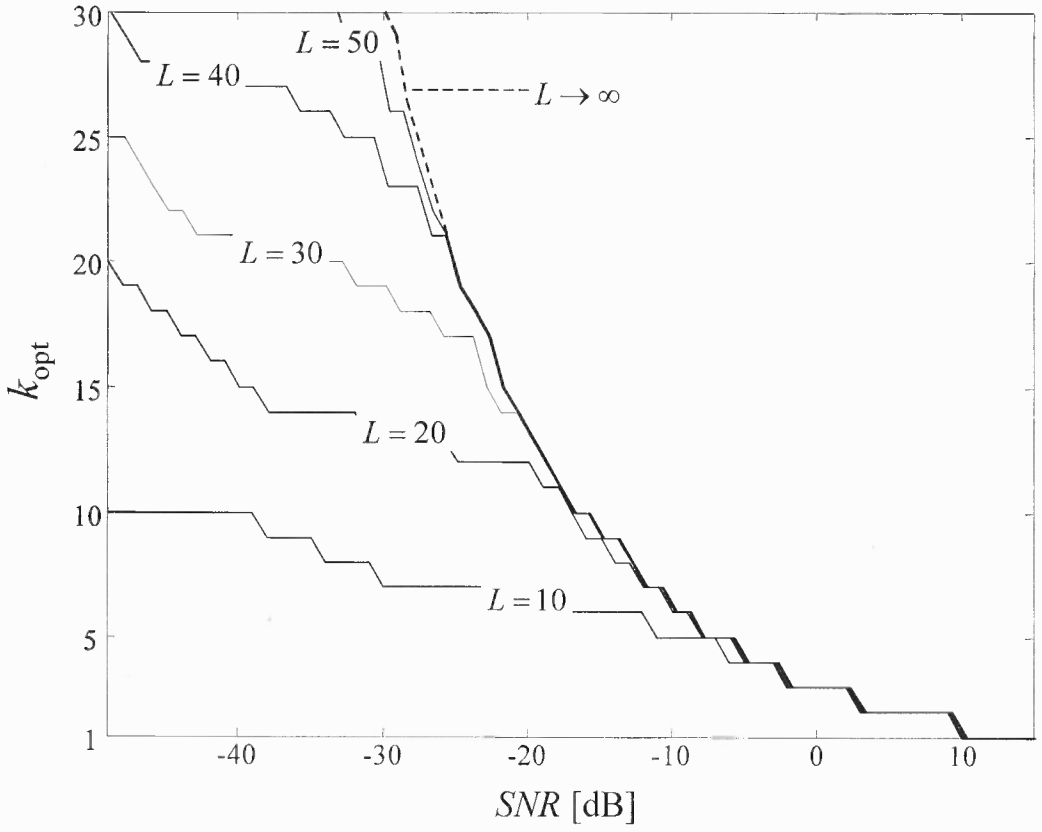

Figure 3.4 HARQ-CC: Optimized number of hops k_{opt} versus SNR and delay L.

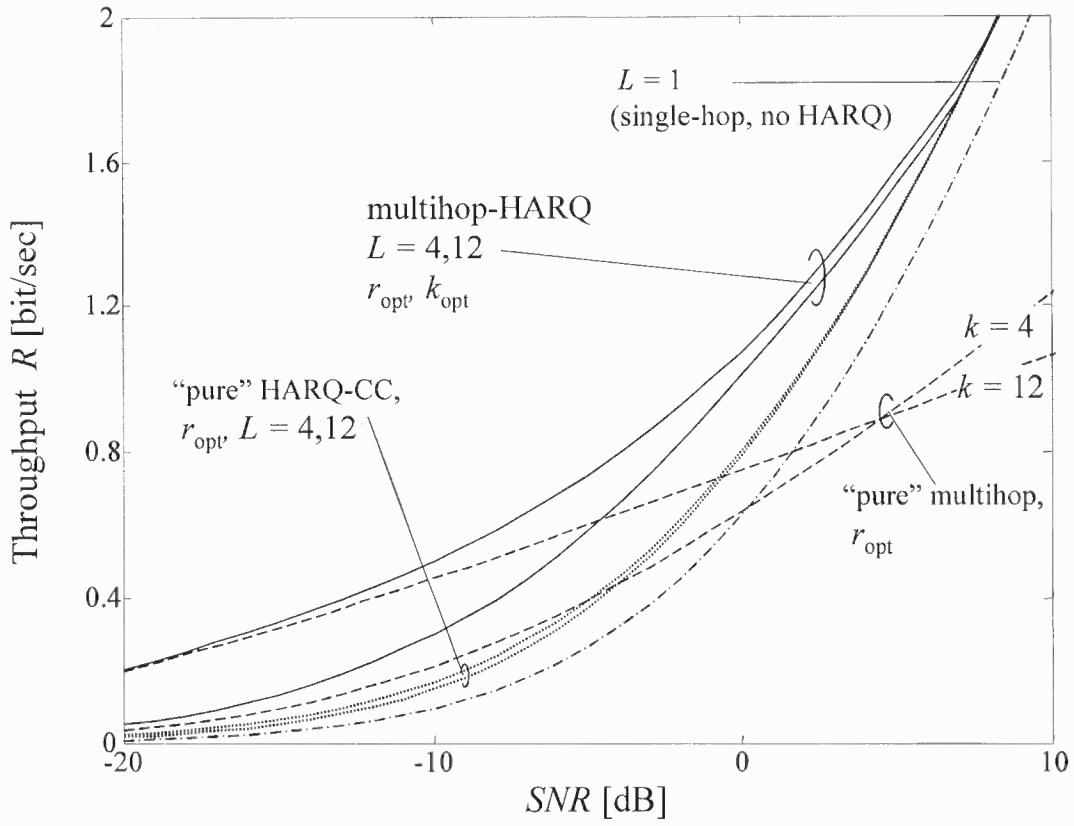

Figure 3.5 HARQ-CC: Throughput R versus SNR and delay L, for optimized transmission rate r_{opt}, and number of hops k_{opt}.

employing r_{opt} and k_{opt}, is shown in Figure 3.5, as a function of SNR and delay ($L = 4, 12$). For comparison, Figure 3.5 also shows the "pure" HARQ-CC (single-hop) and "pure" multihop (no retransmissions) schemes employing $L = 4, 12$ and $k = 4, 12$ slots, respectively, and a single hop system with no delay ($L = 1$). The figure confirms that the "pure" HARQ-CC and "pure" multihop schemes perform poorly in low and high-SNR region, respectively; however, by exploiting both approaches (multihop-HARQ), the system throughput will increase for a broad SNR region.

Finally, Figure 3.6 compares the throughput of optimized schemes employing the multihopping with HARQ-TI and HARQ-CC protocols. As expected, the scheme employing the memoryless HARQ-TI is outperformed by the scheme exploiting the

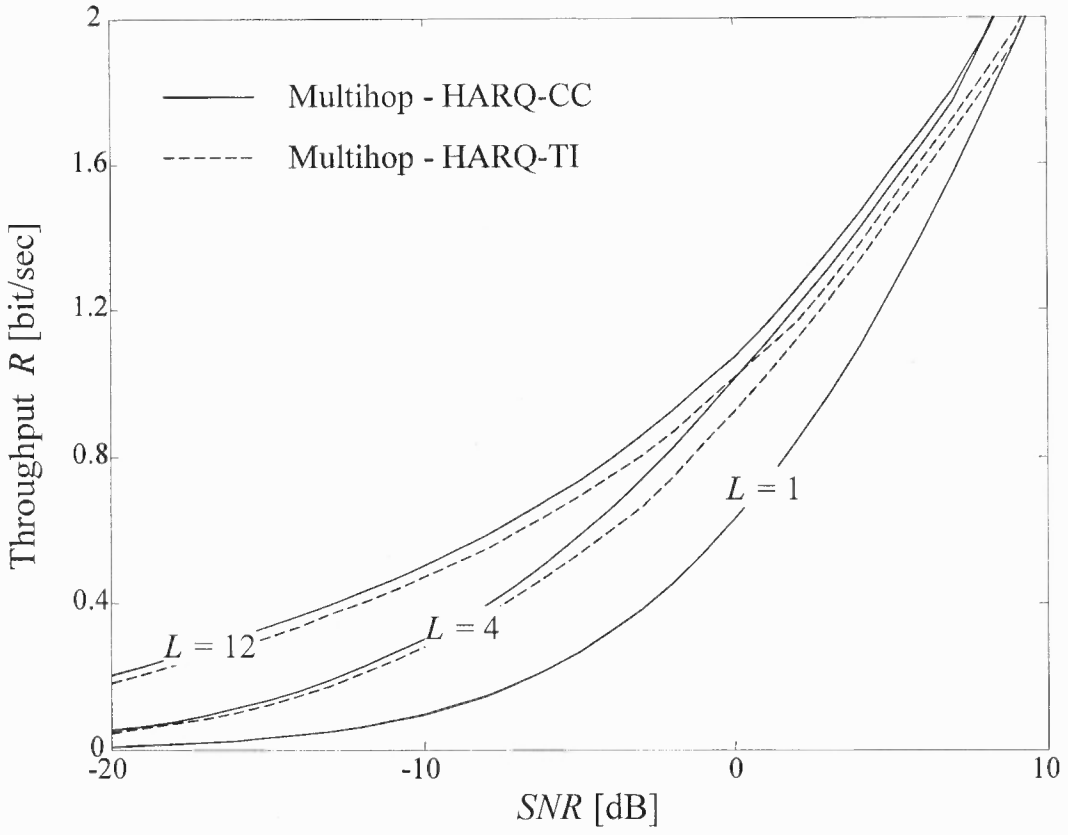

Figure 3.6 Throughput R versus SNR and delay L, for HARQ-TI and HARQ-CC, optimized transmission rate r_{opt} and number of hops k_{opt}.

HARQ-CC protocol. Notice that the difference between two schemes is relatively constant over a broad SNR region. This behavior is quite different than the behavior of "pure" HARQ schemes, wherein the difference between the two protocols is more emphasized in region where SNR is low. Namely, as the SNR decreases, the (optimized) network relies heavily on the multiple hops, rather than on retransmissions, and the memory of HARQ-CC protocol cannot achieve the full advantage. Optimization of the transmission rate r influences the system in a similar manner, as it prevents a need for a large number of retransmissions, which again reduces the advantage of HARQ-CC protocol.

3.4 Chapter Conclusion

In this chapter, the optimal linear network design, i.e., the number of employed hops/relays that maximizes end-to-end throughput, was investigated for a delay-tolerant (up to a given maximum delay) system employing both multihop and HARQ-TI/CC protocols. It was determined through analytical and numerical results that an optimally designed system exploits the delay primarily through multihopping and HARQ protocols in low and high-SNR region, respectively; the good performance of multihop in low-SNR region (and its poor performance in high-SNR region) confirms the analytical results of [38], that were limited to Gaussian (unfaded) channels. It was also observed that for relatively large values of SNR, even if infinite delay is allowed, there is an upper limit on the optimal number of relays that would result in further throughput increase. Most importantly, it was demonstrated that the system employing synergy of multihop and HARQ, if optimally designed, significantly benefits from the allowed delay, as compared to the systems employing only multihop or HARQ protocols.

CHAPTER 4

ENERGY CONSIDERATION: TO COOPERATE, RETRANSMIT, BOTH OR NEITHER?

Hybrid-ARQ, cooperative and cooperative Hybrid-ARQ protocols can decrease the transmission energy required for successful decoding at the destination, at the expense of an increased transmission delay. However, the impact on the *total* energy, including both the transmission power and *circuitry consumption* of all involved stations (i.e., the source, the destination, and, in the cooperative scenario, the relay), has not been investigated yet. This issue is of particular importance in modern wireless networks, where the distances between terminals (e.g., wireless sensors) are becoming smaller, and the energy exploited by the electronic circuits can become of the same or even larger order of magnitude than the transmission energy [43]. Since, in general, the circuitry consumption is proportional to the time that the terminals remain active and the number of involved terminals, the savings in transmissions energy due to the multiple HARQ transmissions and cooperation, respectively, can be neutralized or even surpassed by the increased circuitry consumption.

In this chapter, the minimum total energy per bit required for successful communication over fading channels is investigated for non-cooperative and cooperative truncated HARQ scenarios. For the cooperative case, scenario illustrated in Figure 2.2-c) is applied (space-time coded cooperation). Under a fixed outage probability constraint (as in [7]), the minimum total energy, based on the optimal transmission energy and transmission time (i.e., transmission rate) of a packet, is determined for HARQ-TI, HARQ-CC and HARQ IR protocols (in the latter case, a lower bound of the total energy is considered). The total consumption includes both the transmission energy and the energy consumed by the transmitting and receiving electronic circuitry of all involved terminals (source, destination and, possibly, the relay).

The remainder of the chapter is organized as follows. In Section 4.1, the system model and the performance analysis of standard (non-cooperative) HARQ protocols is provided. Section 4.2 extends the treatment to cooperative HARQ protocols. Numerical results in Section 4.3 corroborate the analysis and provide insight into the problem at hand. Section 4.4 summarizes the conclusions.

4.1 Non-Cooperative HARQ Protocols

In this section, the focus is on the energy efficiency of a single-link non-cooperative HARQ protocol. The treatment is then extended in Section 4.2 to the performance of cooperative HARQ.

4.1.1 System Overview

A single-link communication between a source and a destination over a block-fading Rayleigh channel is considered. Thus (recall Section 2.1.1), the fading channel is constant during each transmission and changes independently with each retransmission. The independent zero-mean unit-power complex Gaussian channel between the source and the destination at the ith (re)trans-mission of a given packet is denoted as $h_{SD}^{(i)}$. The single-sided thermal noise spectral density is N_0, while B (in Hertz) stands for the available bandwidth. The transmission power is the same for all retransmissions, including the original transmission. All signalling messages, such as ACK and NACK messages, are assumed to be significantly shorter than the user data packets and transmitted with perfect reliability and negligible overall energy consumption. Following [43], it is assumed here that each packet carries L bits and that transmission of each packet lasts T_{on} seconds (with a maximum value T, $T_{on} \leq T$, fixed by design constraints), where T_{on} is a design parameter of the system (Figure 4.1) (see also [44]). Notice that L/T_{on} can be interpreted as the system transmission rate r (in bit/sec) of the original transmission.

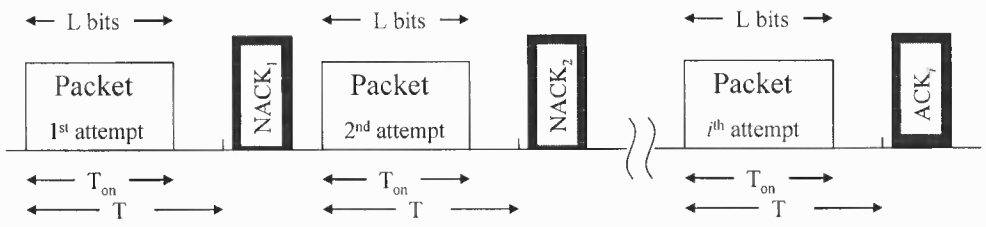

Figure 4.1 Illustration of the transmission protocol ($i \leq n$).

The average transmission energy per bit received at the destination during a single (re)transmission is denoted as E_b, and the probability of unsuccessful decoding at the destination after i transmissions as $p_i(E_b, T_{on})$, where the dependence of the probability p_i on the energy E_b and transmission time T_{on} is emphasized. The QoS requirement to be met by the communication link is defined in terms of the maximum tolerated probability of outage p_{out} (i.e., the probability of unsuccessful detection of a packet). Then, by letting n be the *maximum allowed number of retransmissions of a packet*, including the original, the constraint imposed on the transmission energy E_b and transmission time T_{on} can be obtained from

$$p_n(E_b, T_{on}) \leq p_{out}. \tag{4.1}$$

The problem amounts to the following: given the maximum allowed number of transmissions n, find the optimal transmission energy E_b and the packet duration T_{on} that minimize the *total average* energy consumption per bit $E(E_b, T_{on})$, including both transmission and circuitry consumption, under condition (4.1), i.e.,

$$\begin{aligned} &\text{minimize } E(E_b, T_{on}) \\ &\text{subject to } p_n(E_b, T_{on}) \leq p_{out} \\ &\qquad 0 < T_{on} \leq T. \end{aligned} \tag{4.2}$$

The main goal of the analysis presented in this chapter is to determine the dependence of the total average energy E and the probability p_n on the parameters E_b and T_{on}, in order to provide the setting for the optimization in (4.2).

The total energy expenditure needs to incorporate the energies consumed by both the transmitting (source) and receiving (destination) sides. In particular, the contribution of transmitting node during any transmission can be modeled following [43], as

$$E_t = d^{\gamma} K_t E_b + P_{ct} \frac{T_{on}}{L}, \tag{4.3}$$

where the first term in the summation accounts for the energy consumed by the *power amplifier*, while the second term stands for the energy consumed by the *electronic circuitry in the transmit mode*. In particular, d is the transmission range, γ stands for the path loss exponent, K_t is a constant that depends on the physical characteristics of the link and power amplifier and P_{ct} measures the power (in Watts) consumed by the transmitting electronic circuitry (see [43] for details). It is noted that, in general, K_t depends on peak-to-average-power ratio of the transmitted signal, which can be in practice related to T_{on} through a constellation size. However, for the sake of simplicity, K_t is considered here as a constant, as in M-PSK or FSK modulation. Returning to the energy consumption model, the energy consumed by the *circuitry* on the *receiving* side (destination) during one transmission can be written as

$$E_r = P_{cr} \frac{T_{on}}{L}, \tag{4.4}$$

where P_{cr} stands for the power (in Watts) consumed by the receiving circuitry [43].

Notice that the block fading assumption considered in this work constitutes a major difference with respect to [43], in which the average probability of error is used as performance metric, thus implying an ergodic fading scenario. Furthermore, the circuitry consumption of the terminals in idle state is not addressed in this work. If such power, say P_{idle}, is introduced, the transmission and receiving energy consumption (4.3) and (4.4)

would read, respectively

$$E_t = d^{\gamma} K_t E_b + P_{ct}\frac{T_{on}}{L} + P_{idle}\left(\frac{T - T_{on}}{L}\right) \tag{4.5a}$$

$$= d^{\gamma} K_t E_b + (P_{ct} - P_{idle})\frac{T_{on}}{L} + P_{idle}\frac{T}{L},$$

$$E_r = (P_{cr} - P_{idle})\frac{T_{on}}{L} + P_{idle}\frac{T}{L}. \tag{4.5b}$$

Herein the extra terms due to P_{idle} are not considered as they are typically negligible as compared to the other contributions (see, e.g., [45]). However, the approach of Section 4.3.1 could be used to account for this consumption.

4.1.2 Performance Analysis

For the non-cooperative scenario, the total energy consumption per transmission is simply a summation of the energies spent by the transmitting and receiving side. Defining $P_c = P_{cr} + P_{ct}$ as the overall circuitry power consumption, and $\bar{\imath}$ as the average number of exploited transmissions, the total average energy consumption per bit $E(E_b, T_{on})$ (measured in Joules) reads

$$\begin{aligned} E(E_b, T_{on}) &= (E_t + E_r) \times \bar{\imath} \\ &= \left(d^{\gamma} K_t E_b + P_c\frac{T_{on}}{L}\right)\left(\sum_{i=1}^{n-1} i\,(p_{i-1} - p_i) + n p_{n-1}\right) \\ &= \left(d^{\gamma} K_t E_b + P_c\frac{T_{on}}{L}\right)\left(1 + \sum_{i=1}^{n-1} p_i\right), \end{aligned} \tag{4.6}$$

where the dependence of p_i on E_b and T_{on} is dropped for notation convenience and the fact that $p_{i-1} - p_i$ is the probability of successful decoding at the destination exactly at the ith transmission (with p_0 set as $p_0 = 1$) [11] is used. In the second line of (4.6), the term $\bar{\imath}$ is averaged over all the possible outcomes, i.e., i) the scenarios wherein successful decoding occurred in one of the first $n-1$ transmissions and ii) the scenario wherein the successful decoding did not occur during the first $n-1$ transmissions (thus, including event of successful transmission exactly at the nth attempt, and the outage event). In the

minimization of (4.6) according to (4.2), two trade-offs play a key role, as discussed in the following:

- While, in general, the required transmission energy E_b is a non-increasing function of the transmission time T_{on} (i.e., reducing the transmission rate reduces the transmission energy requirements), the circuitry consumption increases with T_{on} [43];
- A traditional approach to a system design, that aims at optimizing only the transmission energy, would entail the choice of the minimum transmission energy E_b that satisfies constraint (4.1) with equality, $p_n(E_b, T_{on}) = p_{out}$. However, this choice does not necessarily produce optimal results when minimizing the total energy. In fact, minimizing the energy E_b generally increases the average number of transmissions $\bar{\imath}$, thus leading to an increased circuitry consumption (4.6).

Recall that the optimization problem (4.2) does not place any explicit constraint on the values of average delay $\bar{\imath}$ and thus on the average throughput $L/T \cdot (1 - p_n(E_b, T_{on}))/\bar{\imath}$ [11], apart from the upper and lower bounds n and $L/T \cdot (1 - p_{out})/n$, respectively. Nevertheless, according to the aforementioned trade-offs, it will be shown that the impact of the circuitry energy forestalls the choice of extremely large delay $\bar{\imath}$ and thus of small average throughput. These trade-offs are discussed in details in Section 4.3 via numerical results.

As it can be seen from (4.6), in order to determine the total average energy consumption, the condition that the transmission energy per bit E_b and transmission time T_{on} must satisfy in order to guarantee (4.1) need to be computed, as well as the probability of unsuccessful decoding after i transmissions, p_i, which is a function of E_b and T_{on}. The rest of this section is devoted to this task. It is first address the reference case where no retransmissions occur ($n = 1$) in Section 4.1.1, and the analysis is then extended to HARQ protocols in the following subsections.

The reference case: one transmission ($n = 1$) The maximum rate (in bit/sec) achievable in one transmission reads

$$C_1 = B \times \log_2 \left(1 + |h_{SD}^{(1)}|^2 \frac{P_r}{N_0 B}\right), \tag{4.7}$$

where $P_r = LE_b/T_{on}$ is the received power. Notice that (4.7) assumes optimal Gaussian coding; more practical transmission schemes could be easily accommodated in the proposed framework by using the approximations in [30] (recall Section 2.1.2). A decoding error occurs when the instantaneous achievable rate (4.7) is smaller than the transmission rate L/T_{on} (see, e.g., [46]), so that

$$\begin{aligned} p_1 &= P\{C_1 < L/T_{on}\} \\ &= P\left\{|h_{SD}^{(1)}|^2 < \left(2^{\frac{L}{BT_{on}}} - 1\right) \frac{N_0 B T_{on}}{E_b L}\right\} \\ &= F_{\chi^2}\left[\mu_1(T_{on})/E_b, 2\right], \end{aligned} \tag{4.8}$$

where $F_{\chi^2}[x, \nu]$, $\nu = 1, 2, ...$ is the cumulative distribution function of a chi-square variable (recall Section 2.1.1) with ν degrees of freedom, taken at value x, while the coefficient

$$\mu_n(T_{on}) = \left(2^{\frac{L}{nBT_{on}}} - 1\right) N_0 \frac{nBT_{on}}{L} \tag{4.9}$$

is introduced for convenience of notation. Since no retransmissions are employed, equality can be imposed in condition (4.1), $p_1 = p_{out}$, resulting in the optimal energy per bit at the receiver

$$E_{b,opt}(T_{on}) = \frac{\mu_1(T_{on})}{F_{\chi^2}^{-1}[p_{out}, 2]}, \tag{4.10}$$

with $F_{\chi^2}^{-1}[y, \nu]$ denoting the inverse of $F_{\chi^2}[x, \nu]$ taken at value y. The total average energy consumption per bit reads from (4.6)

$$E\left(E_{b,opt}\left(T_{on}\right), T_{on}\right) = d^{\gamma} K_t \frac{\mu_1(T_{on})}{F_{\chi^2}^{-1}[p_{out}, 2]} + P_c \frac{T_{on}}{L}. \tag{4.11}$$

For this simple reference case, explicit solution of the optimization (4.2) is readily available. In particular, it can be easily shown that the function (4.11) is convex over T_{on} and, by setting its first derivative to zero, the optimal transmission time can be found to equal

$$T_{on,opt} = \frac{L}{B} \ln 2 \times \left(1 + \mathcal{W} \left(P_c \frac{F_{\chi^2}^{-1}[p_{out}, 2]}{eBd^{\gamma} K_t N_0} - \frac{1}{e} \right) \right)^{-1}, \tag{4.12}$$

where the Lambert W-function $\mathcal{W}$ is the inverse function of $f(W) = We^W$, and $e = 2.7182...$ is the base of a natural logarithm.

HARQ-TI With HARQ-TI, erroneous packets received in previous retransmissions are discarded, so that the probability of erroneous decoding after the ith retransmission reads

$$\begin{aligned} p_i &= P\{C_1 < L/T_{on}, ..., C_i < L/T_{on}\} \\ &= P\{C_k < L/T_{on}\}^i, \end{aligned} \tag{4.13a}$$

where C_k is the maximum rate achievable at any single transmission

$$C_k = B \times \log_2 \left(1 + |h_{SD}^{(k)}|^2 \frac{P_r}{N_0 B} \right). \tag{4.14}$$

Following a similar reasoning as in the previous Section, and dropping the dependence of μ_1 on T_{on} for convenience of notation, it is obtained that

$$p_i = F_{\chi^2}^i \left[\mu_1 / E_b, 2 \right], \tag{4.15}$$

and, imposing the QoS constraint (4.1):

$$E_b \geq \frac{\mu_1}{F_{\chi^2}^{-1} \left[p_{out}^{1/n}, 2 \right]}. \tag{4.16}$$

By substituting (4.15) into (4.6) and applying geometric sum equation the total average energy follows

$$E(E_b, T_{on}) = \left(d^{\gamma} K_t E_b + P_c \frac{T_{on}}{L} \right) \times \frac{1 - F_{\chi^2}^n \left[\frac{\mu_1}{E_b}, 2 \right]}{1 - F_{\chi^2} \left[\frac{\mu_1}{E_b}, 2 \right]}. \tag{4.17}$$

The total energy (4.17) along with the constraint (4.16) can then be used to solve optimization problem (4.2) in order to yield the optimal parameters T_{on} and E_b. This is a non-convex problem which will be solved in Section 4.3 using numerical simulations.

HARQ-CC For retransmission protocols that exploit memory (i.e., the previously received erroneous packets), the probability of unsuccessful decoding after i transmissions is generally given by

$$\begin{aligned} p_i &= P\{C_1 < L/T_{on}\} P\{C_2 < L/T_{on} | C_1 < L/T_{on}\} ... P\{C_i < L/T_{on} | C_{i-1} < L/T_{on}\} \\ &= P\{C_1 < L/T_{on}, C_2 < L/T_{on}, ..., C_i < L/T_{on}\}, \end{aligned} \tag{4.18}$$

where C_i is the maximum rate achievable with i transmissions (regardless of previous attempts). With HARQ-CC, the preserved retransmitted copies of the packet are soft combined at the receiver so that the rate C_i reads

$$C_i = B \times \log_2 \left(1 + \sum_{k=1}^{i} |h_{SD}^{(k)}|^2 \frac{P_r}{N_0 B} \right). \tag{4.19}$$

Notice that C_i, as introduced in (4.18), is not conditioned on previous (re)transmission attempts, so that the overall gain $\sum_{k=1}^{i} |h_{SD}^{(k)}|^2$ in (4.19) is simply a chi-square distributed random variable with $2i$ degrees of freedom (recall Section 2.1.1). Taking into account the fact that conditions $C_1 < L/T_{on}, ..., C_{i-1} < L/T_{on}$ are implied by $C_i < L/T_{on}$ (since $C_i \geq C_{i-1} \geq ... \geq C_1$), the probability of unsuccessful decoding after i transmissions

becomes

$$\begin{aligned} p_i &= P\left\{C_i < L/T_{on}\right\} \\ &= F_{\chi^2}\left[\mu_1/E_b, 2i\right]. \end{aligned} \tag{4.20}$$

The QoS constraint (4.1) can be now imposed by using (4.20) for $i = n$ to obtain the following bound on the required received energy

$$E_b \geq \frac{\mu_1}{F_{\chi^2}^{-1}[p_{out}, 2n]}, \tag{4.21}$$

while the total consumed energy per bit, from (4.20) and (4.6), is given by

$$E(E_b, T_{on}) = \left(d^{\gamma} K_t E_b + P_c \frac{T_{on}}{L}\right)\left(1 + \sum_{i=1}^{n-1} F_{\chi^2}\left[\frac{\mu_1}{E_b}, 2i\right]\right). \tag{4.22}$$

Equations (4.22) and (4.21) provide for this case the objective function and the QoS constraint, respectively, for optimization problem (4.2).

HARQ-IR For the HARQ-IR protocol, the achievable rate after i transmissions is equal to (see, e.g., [11])

$$C_i = B \times \sum_{k=1}^{i} \log_2\left(1 + |h_{SD}^{(k)}|^2 \frac{P_r}{N_0 B}\right). \tag{4.23}$$

As discussed in [11], the analysis of the outage probability for this protocol using (4.23) is rather complicated. Therefore, an upper bound approximation, which is obtained by applying the Jensen inequality to (4.23), is used:

$$C_i \leq B \times i \log_2\left(1 + \frac{1}{i} \sum_{k=1}^{i} |h_{SD}^{(k)}|^2 \frac{P_r}{N_0 B}\right). \tag{4.24}$$

The probability of unsuccessful decoding at the destination after the ith transmission can be now lower bounded as

$$p_i = P\left\{C_i < L/T_{on}\right\} \geq F_{\chi^2}\left[\mu_i/E_b, 2i\right]. \tag{4.25}$$

Notice that the only difference in the analytical expression for the probability p_i of HARQ-CC (4.20) and in the bound (4.25) for the HARQ-IR protocols is in the coefficient μ_i that changes with the number of transmission attempts i for HARQ-IR protocol. Therefore, the results of Sec 4.1.1 can be used to write the following bound on the required received energy per bit that satisfies (4.1)

$$E_b \geq \frac{\mu_n}{F_{\chi^2}^{-1}[p_{out}, 2n]}, \tag{4.26}$$

while, from (4.25) and (4.6), the total consumed energy per bit satisfies

$$E(E_b, T_{on}) \geq \left(d^{\gamma} K_t E_b + P_c \frac{T_{on}}{L}\right)\left(1 + \sum_{i=1}^{n-1} F_{\chi^2}\left[\frac{\mu_i}{E_b}, 2i\right]\right). \tag{4.27}$$

In Section 4.3 the optimization (4.2) will be considered with objective function given by the right-hand sides of the bound (4.27) and the QoS constraint (4.26). It is remarked that being based on (4.24), the results of the optimization will have to be interpreted as upper bounds on the performance of HARQ-IR. However, the tightness of the bound (4.24) was confirmed in [31] and corroborated in Section 4.3 via numerical simulations.

4.2 Cooperative HARQ Protocols

In this section, the results of Section 4.1 are extended to cooperative HARQ protocols.

4.2.1 System Overview

The principle of cooperative HARQ protocol of interest is shown in Figure 4.2. The relay station listens to the (re)transmissions of a given packet. In case the destination feeds back a NACK, if the relay has successfully decoded the transmission, it signals its availability and switches from the receiving to the transmitting mode. Then, for each of the following retransmissions, the source and relay form a distributed antenna array and cooperate by sending to the destination a space-time codeword [47]- [49], chosen according to the selected HARQ protocol (HARQ-TI, CC or IR) [1] [50].

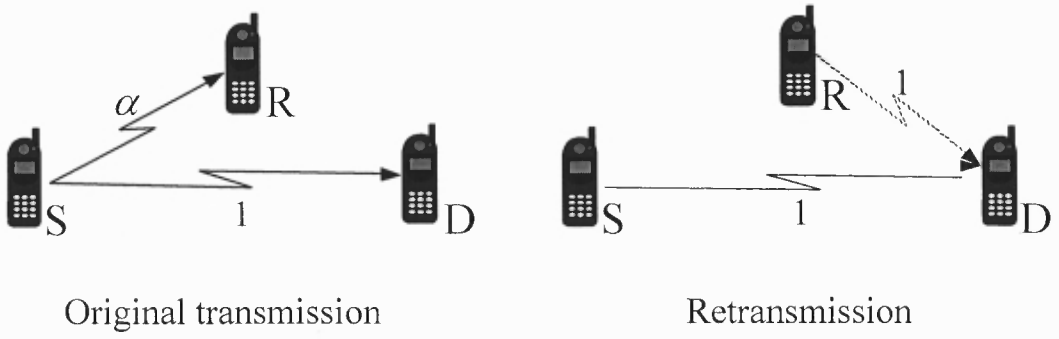

Figure 4.2 Illustration of the cooperative Hybrid-ARQ protocol.

The system model, described in Section 4.1.1, is herein extended to the cooperative scenario. The block Rayleigh fading channels between the source and the relay and between the relay and the destination, at the ith (re)transmission, are denoted as $h_{SR}^{(i)}$ and $h_{RD}^{(i)}$, respectively. The channels between any two nodes (source, destination and relay) are mutually independent. The average channel power gains are $E[|h_{RD}^{(i)}|^2] = 1$ and $E[|h_{SR}^{(i)}|^2] = \alpha$, with $\alpha > 1$ accounting for a scenario where the relay is relatively closer to the source than to the destination (recall from Section 4.1 that $E[|h_{SD}^{(i)}|^2] = 1$). The transmission power of both the source and the relay, and the transceiver specifications (4.3) and (4.4) for any of the nodes, are assumed to be the same.

4.2.2 Performance Analysis

The average total energy consumption of cooperative HARQ protocols needs to account for the activity of the relay station. For this purpose, the following notation conventions are introduced. Denote i as the index of the final transmission, that can be either a successful transmission, $i = 1, .., n$, or the unsuccessful nth transmission, $i = n$, when the maximum delay expires. Furthermore, denote j as the index of the transmission when the relay successfully decoded, $j = 1, 2, .., i-1$, or, in the case that the relay did not successfully decode prior to the ith (final) transmission, $j = i$ (by convention). Then, the relay contribution to the total energy expenditure is given by the energy spent by its

receiving circuitry E_r (from (4.4)) during the first $j$transmissions (while in the listening mode) and by the transmitting circuitry and power amplifier energy E_t (from (4.3)) during remaining $(i-j)$transmissions (while cooperating with the source), so that the total energy consumption reads

$$E_{ij} = (E_t + 2E_r) \times j + (2E_t + E_r) \times (i - j), \tag{4.28}$$

for $i = 1, 2, .., n$ and $j = 1, 2, .., i$.

Furthermore, a term q_k, $k = 0, .., n-1$, is introduced as the probability of unsuccessful decoding at the relay after k transmissions, and $p_{k|j}$, $k = 1, .., n$, as the probability of unsuccessful decoding at the destination after k transmissions, given that the relay decoded exactly at the jth transmission attempt (see definition above). As a convention, $p_{k|k}$ will be used for the case where the relay did not successfully decode prior to the kth transmission. Based on these definitions, the probability of successful decoding at the relay after exactly j transmissions is given as $q_{j-1} - q_j$ (with $q_0 = 1$), and at the destination after exactly i transmissions as $p_{i-1|j} - p_{i|j}$ [11] (according to the above, $p_{i-1|i-1} - p_{i|i}$ denotes the probability that the relay did not successfully decode after the ith transmission.). Now, the average total energy consumption can be written as

$$\begin{aligned} E(E_b, T_{on}) = &\sum_{i=1}^{n-1} \sum_{j=1}^{i-1} (q_{j-1} - q_j) \left(p_{i-1|j} - p_{i|j}\right) E_{ij} + \sum_{i=1}^{n-1} q_{i-1} \left(p_{i-1|i-1} - p_{i|i}\right) E_{ii} \\ &+ \sum_{j=1}^{n-1} (q_{j-1} - q_j)\, p_{n-1|j} E_{nj} + q_{n-1} p_{n-1|n-1} \left(E_{nn} - E_r\right), \end{aligned} \tag{4.29}$$

where: i) the first term describes the average energy contribution of the events where successful decoding at the destination occurred before the maximum number of transmission n was reached, and the relay was able to cooperate; ii) the second term presents the average energy of the events where successful decoding at the destination occurred before the maximum number of transmission n was reached, and the relay was unable to cooperate; iii) the third term describes the average energy contribution of the events where

the destination did not successfully decode before the maximum number of transmissions n was reached, with the cooperation from the relay; and iv) the fourth term stands for the average energy contribution of the event where the destination, as well as the relay, did not successfully decode before the maximum number of transmission n was reached. Notice that in the last term, receiving circuitry consumption E_r is subtracted from the energy consumption, as there is no benefit for the relay to try decoding at the last transmission attempt.

In order to evaluate (4.29), the probabilities q_j and $p_{i|j}$ need to be calculated for different cooperative HARQ schemes. The probability of outage, p_n, also needs to be determined for each of the cooperative HARQ schemes, so as to complete the setting for the optimization in (4.2). Though the analysis of the cooperative scenario is complicated by the presence of the relay, it is still based on the same concept discussed in the previous section for the non-cooperative case. Therefore, some of the results from Section 4.1 will be relied upon.

HARQ-TI For the HARQ-TI protocol, the results from Section 4.1.1 are used to directly write the probability of the unsuccessful decoding at the relay (see (4.15))

$$q_j = F_{\chi^2}^j \left[\mu_1 / \left(\alpha E_b\right), 2\right], \tag{4.30}$$

for $j = 1, .., n-1$. On the other hand, the maximum rates (in bit/sec) achievable at the destination after i transmissions, with the relay in receiving and transmitting mode, respectively, read

$$C_{i;\bar{R}} = B \times \log_2 \left(1 + |h_{SD}^{(i)}|^2 \frac{P_r}{N_0 B}\right) \tag{4.31a}$$

$$C_{i;R} = B \times \log_2 \left(1 + \left(|h_{SD}^{(i)}|^2 + |h_{RD}^{(i)}|^2\right) \frac{P_r}{N_0 B}\right), \tag{4.31b}$$

Notice that the summation $|h_{SD}^{(i)}|^2 + |h_{RD}^{(i)}|^2$ in (4.31b) describes the diversity effect of transmission from two antennas (of source and relay) through distributed space-time

coding. Using (4.31), $p_{i|j}$ becomes

$$\begin{aligned} p_{i|j} &= \prod_{k=1}^{j} P\left\{C_{k;\bar{R}} < L/T_{on}\right\} \prod_{l=j+1}^{i} P\left\{C_{l;R} < L/T_{on}\right\} \\ &= F_{\chi^2}^{j}\left[\mu_1/E_b, 2\right] F_{\chi^2}^{i-j}\left[\mu_1/E_b, 4\right]. \end{aligned} \tag{4.32}$$

The previous equation can be exploited with (4.30) to yield the outage probability

$$\begin{aligned} p_n &= \sum_{j=1}^{n-1} \left(q_{j-1} - q_j\right) p_{n|j} + q_{n-1} p_{n|n} \\ &= \sum_{j=1}^{n-1} F_{\chi^2}^{j-1}\left[\frac{\mu_1}{\alpha E_b}, 2\right]\left(1 - F_{\chi^2}\left[\frac{\mu_1}{\alpha E_b}, 2\right]\right) F_{\chi^2}^{j}\left[\frac{\mu_1}{E_b}, 2\right] F_{\chi^2}^{n-j}\left[\frac{\mu_1}{E_b}, 4\right] \\ &\quad + F_{\chi^2}^{n-1}\left[\frac{\mu_1}{\alpha E_b}, 2\right] F_{\chi^2}^{n}\left[\frac{\mu_1}{E_b}, 2\right]. \end{aligned} \tag{4.33}$$

Notice that in (4.33), the first term (summation) and the second term stand for the event when successful decoding at the relay occurred and did not occur, respectively, prior to the nth transmission. Using (4.30) and (4.32), the energy consumption can be expressed in terms of E_b and T_{on}, and can be applied, along with the previous equation, in the minimization (4.2) (see Section 4.3 for numerical examples and discussion).

HARQ-CC Exploiting results from Section 4.1.1, it is easily determined that for this case (see (4.20))

$$q_j = F_{\chi^2}\left[\mu_1/(\alpha E_b), 2j\right]. \tag{4.34}$$

With HARQ-CC, differently from HARQ-TI, the achievable rate at the destination at a given transmission is not only affected by whether or not the relay is transmitting, but also by the transmission attempt (if any) at which it successfully decoded. Accordingly, the maximum achievable rates (in bit/sec) at the destination with the relay in receiving and

transmitting mode, respectively, read

$$C_{i,i} = B \times \log_2 \left(1 + \sum_{k=1}^{i} |h_{SD}^{(k)}|^2 \frac{P_r}{N_0 B} \right) \tag{4.35a}$$

$$C_{i,j} = B \times \log_2 \left[1 + \left(\sum_{k=1}^{i} |h_{SD}^{(k)}|^2 + \sum_{k=j+1}^{i} |h_{RD}^{(k)}|^2 \right) \frac{P_r}{N_0 B} \right]. \tag{4.35b}$$

Notice that the rate in (4.35b) accounts for the j packets received at the destination with channel gains $|h_{SD}^{(k)}|^2$, where $k = 1, .., j$ (non-active relay), and the $(i-j)$ packets received with effective channel gains $|h_{SD}^{(k)}|^2 + |h_{RD}^{(k)}|^2$, where $k = j+1, .., i$ (due to cooperation with the relay). Since $C_{i,j} \geq C_{i-1,j} \geq ... \geq C_{1,j}$, the events $C_{1,j} < L/T_{on}, ..., C_{i-1,j} < L/T_{on}$ are subsumed by $C_{i,j} < L/T_{on}$, and $p_{i|j}$ is

$$\begin{aligned} p_{i|j} &= P\left\{ C_{i,j} < L/T_{on} \right\} \\ &= F_{\chi^2}\left[\mu_1 / E_b, 2(2i-j) \right]. \end{aligned} \tag{4.36}$$

Then, the outage probability becomes

$$\begin{aligned} p_n &= \sum_{j=1}^{n-1} \left(q_{j-1} - q_j \right) p_{n|j} + q_{n-1} p_{n|n} \\ &= \sum_{j=1}^{n-1} \left(F_{\chi^2}\left[\frac{\mu_1}{\alpha E_b}, 2(j-1) \right] - F_{\chi^2}\left[\frac{\mu_1}{\alpha E_b}, 2j \right] \right) F_{\chi^2}\left[\frac{\mu_1}{E_b}, 2(2n-j) \right] \\ &\quad + F_{\chi^2}\left[\frac{\mu_1}{\alpha E_b}, 2(n-1) \right] F_{\chi^2}\left[\frac{\mu_1}{E_b}, 2n \right]. \end{aligned} \tag{4.37}$$

Equations (4.36) and (4.34) can be applied in (4.29) to yield the expression for the overall energy consumption E, which can be used, along with the previous equation, to specify the optimization problem (4.2).

HARQ-IR From the discussion in Section 4.1.1, by using (4.24), the analysis of an upper bound on the performance of the HARQ-IR protocol can be simply derived from the corresponding equations for HARQ-CC (compare (4.25) and (4.20)). In particular, from (4.34) and (4.36), it follows that the probabilities q_i and $p_{i|j}$ for HARQ-IR, respectively,

satisfy

$$q_j \geq F_{\chi^2}\left[\mu_j/(\alpha E_b), 2j\right] \tag{4.38a}$$

$$p_{i|j} \geq F_{\chi^2}\left[\mu_i/E_b, 2(2i-j)\right], \tag{4.38b}$$

while a bound on the outage probability can be determined using (4.37)

$$\begin{aligned} p_n &= \sum_{j=1}^{n-1} (q_{j-1} - q_j)\, p_{n|j} + q_{n-1} p_{n|n} \\ &\geq \sum_{j=1}^{n-1} \left(F_{\chi^2}\left[\frac{\mu_{j-1}}{\alpha E_b}, 2(j-1)\right] - F_{\chi^2}\left[\frac{\mu_j}{\alpha E_b}, 2j\right] \right) F_{\chi^2}\left[\frac{\mu_n}{E_b}, 2(2n-j)\right] \\ &\quad + F_{\chi^2}\left[\frac{\mu_{n-1}}{\alpha E_b}, 2(n-1)\right] F_{\chi^2}\left[\frac{\mu_n}{E_b}, 2n\right]. \end{aligned} \tag{4.39}$$

As in Section 4.1.1, the optimization (4.2) for HARQ-IR will be considered based on the right-hand sides of (4.38)-(4.39). The resulting energy consumption will be in fact a lower bound (see Section 4.3 and [31] for further discussion on the tightness of the bound).

4.3 Numerical Results

Following the energy consumption model described in [43], the following setting is used here: $P_c = 210.8$ mW, divided as $P_{ct} = 98$ mW (transmit side) and $P_{cr} = 112.4$ mW (receive side), $K_t = 6.05 \times 10^9$, $\gamma = 3$, $N_0 = -171$ dBm/Hz and $B \cdot T = 1$; unless explicitly specified, $\alpha = 20$ dB and $B = 1$ Hz. For the lack of a proper model, and in order to be consistent with the existing literature on this subject, the same value of the circuitry power expenditure P_c (P_{ct} and P_{cr}) is assumed for all HARQ protocols. Nevertheless, in Section 4.3.1 a short study is presented to illustrate the conditions under which the more complicated schemes are truly advantageous.

Figure 4.3 details on the optimization problem (4.2) for non-cooperative HARQ-TI, HARQ-CC and HARQ-IR protocols, by showing the total energy consumption per bit E (from (4.6)) optimized over transmission energy E_b, for different values of normalized

transmission time BT_{on}/L, with maximum allowed number of transmission $n = 2$ and $n = 4$, $d = 1$m and $p_{out} = 10^{-2}$. As a reference, the system with one transmission only ($n = 1$), as discussed in Section 4.1.1, is also shown. For the HARQ-IR protocol, both the upper bound on the performance, achieved using (4.24), and the actual performance, simulated through Monte Carlo simulations (MC) are shown. It can be observed from the two curves for $n = 2$ and $n = 4$ that the upper bound matches fairly well with the simulated performance; henceforth, in the following the performance of HARQ-IR will be described through the upper bound (4.24). For any of the presented curves, the lower T_{on} region, wherein the overall consumption decreases with T_{on}, corresponds to the regime in which transmission energy is dominant, while for larger values of T_{on}, wherein energy consumption increases with T_{on}, the circuitry consumption is dominant. Circuitry-only consumption ($n \cdot P_c T_{on}/L$, $n = 1 \div 4$) is added in Figure 4.3 to further clarify this effect. This figure confirms that the most sophisticated protocol, HARQ-IR, performs at least as well as HARQ-CC, which in turn outperforms the simple HARQ-TI protocol. By allowing additional transmissions, overall energy efficiency is enhanced for all the HARQ protocols, and the optimal transmission time T_{on}, that minimizes the total energy consumption, decreases in order to compensate for the circuitry consumption increased due to multiple transmissions. Furthermore, notice the interesting behavior of HARQ-IR protocols: as the transmission time T_{on} increases, and the circuitry consumption becomes dominant, the system adjusts the optimal transmission energy E_b so as to reduce the average number of transmissions; circuitry-only consumption curves underline this effect. It is noted that this optimal choice of E_b corresponds to a larger energy than the minimum required to satisfy the QoS constraint (4.1). It is remarked that, though visible only for HARQ-IR, this behavior is common for all HARQ protocols, as further discussed below.

Figure 4.4 provides further insight into the total energy consumption optimization problem (4.2), by showing the average number of transmissions $\bar{\imath}$ and the probability p_n versus the normalized transmission time BT_{on}/L, for optimized transmission energy

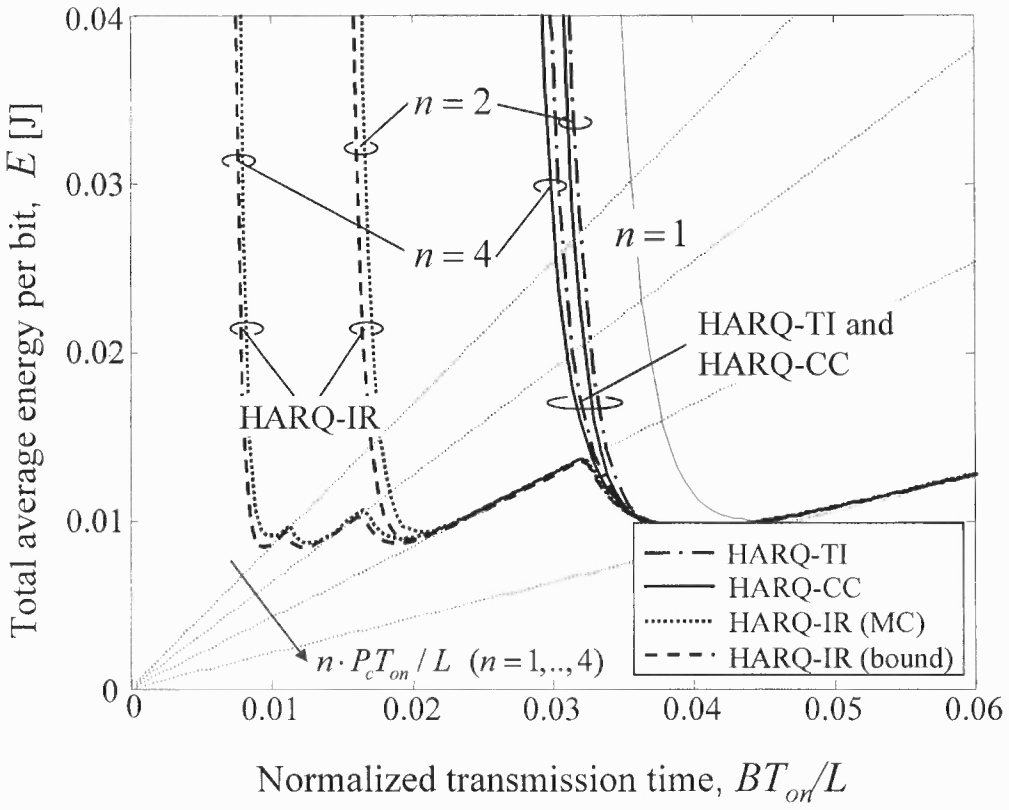

Figure 4.3 Non-cooperative protocols: total average energy per bit E, optimized over transmission energy E_b, versus the transmission time per bit BT_{on}/L ($d = 1m$, $p_{out} = 10^{-2}$).

$E_b(BT_{on}/L)$. The system parameters are chosen to be the same as for Figure 4.3, and for the sake of clarity, HARQ-TI protocol is not shown. It can be seen that the average number of (re)transmissions $\bar{\imath}$ decreases with T_{on}, and that the probability p_n falls below the required outage probability $p_{out} = 10^{-2}$ for all HARQ protocols. This confirms that for larger T_{on} (i.e., small transmission rates), the optimal transmission energy E_b increases beyond the required minimum (4.1), so as to reduce the (average) number of transmissions and thus limit the impact of circuitry consumption.

The overall energy consumption for non-cooperative and cooperative scenarios, minimized over both the transmission energy E_b and the transmission time T_{on} (recall (4.2)), versus the transmission range d, is shown in Figure 4.5 and Figure 4.6 for HARQ-CC and HARQ-IR protocols, respectively, for $n = 2$ and $n = 4$ and $p_{out} = 10^{-6}$. The performance of HARQ-TI protocol is not shown, but from Figure 4.3 it is easily inferred that its performance is similar and slightly worse than for HARQ-CC. For relatively small transmission distances and for any of the HARQ protocols, cooperation is clearly disadvantageous from an energy consumption standpoint. Namely, only for sufficiently large transmission ranges, where the circuity impact is relatively small, cooperation can yield reduction in transmission energy sufficient to compensate for the increased circuitry consumption (due to the receiving/transmitting activity of the relay). Increasing the number of retransmissions n further increases the threshold distance at which cooperation becomes advantageous (e.g., for HARQ-CC, this value is $d \approx 24$m for $n = 2$ and $d > 100$m for $n = 4$).

Finally, Figure 4.7 aims at assessing the performance advantage of HARQ protocols relative to the *time-diversity* schemes, whereby regardless of whether successful decoding at the destination could be achieved with less transmissions, the maximum number of transmissions is always exploited. In particular, this figure compares the minimized (over transmission energy E_b and transmission time T_{on}) total average consumed energy E , for a non-cooperative and cooperative HARQ-CC, with the time-diversity scheme that uses

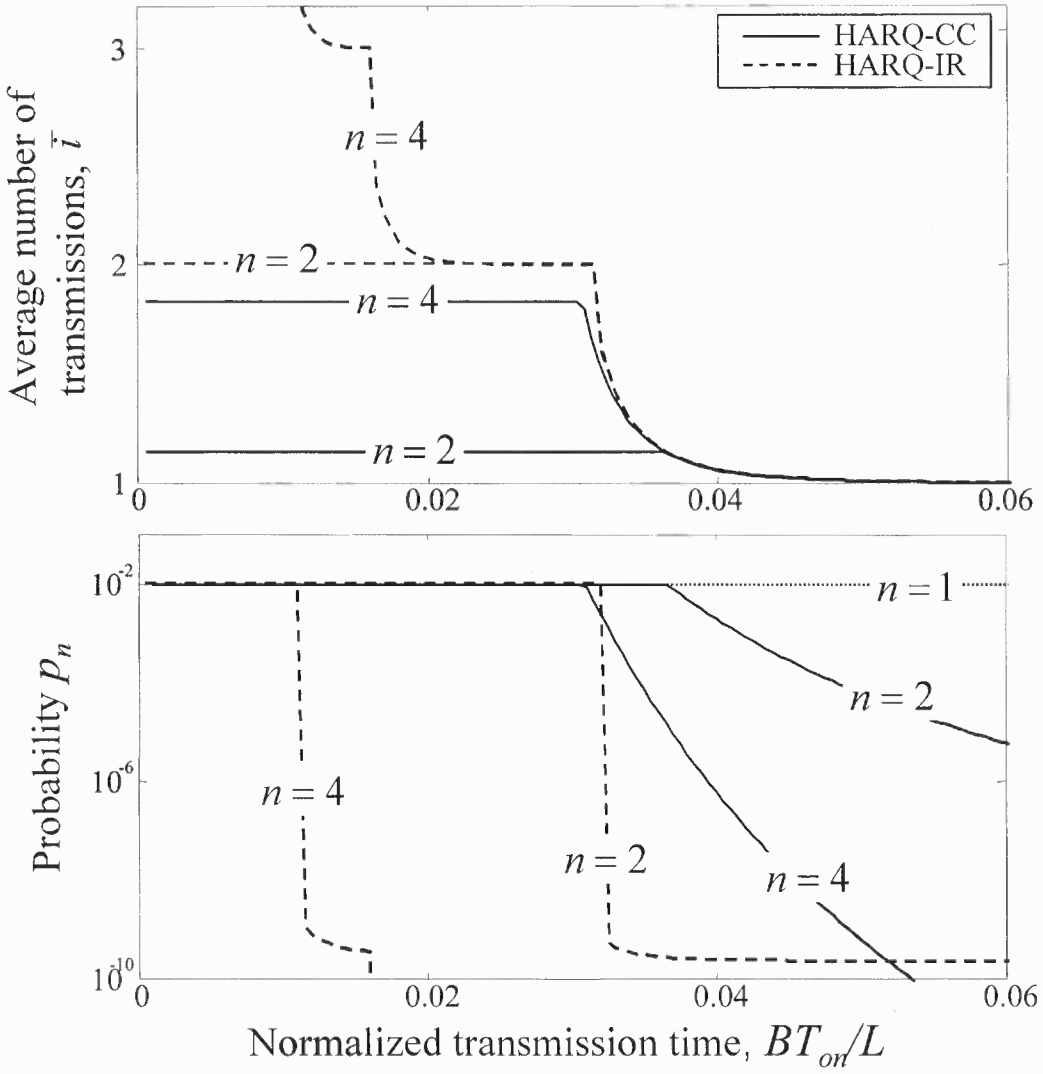

Figure 4.4 Non-cooperative protocols: average number of transmissions $\bar{t}$ and the outage probability p_n, after the optimization over transmission energy E_b, versus the transmission time per bit BT_{on}/L ($d = 1m$, $p_{out} = 10^{-2}$).

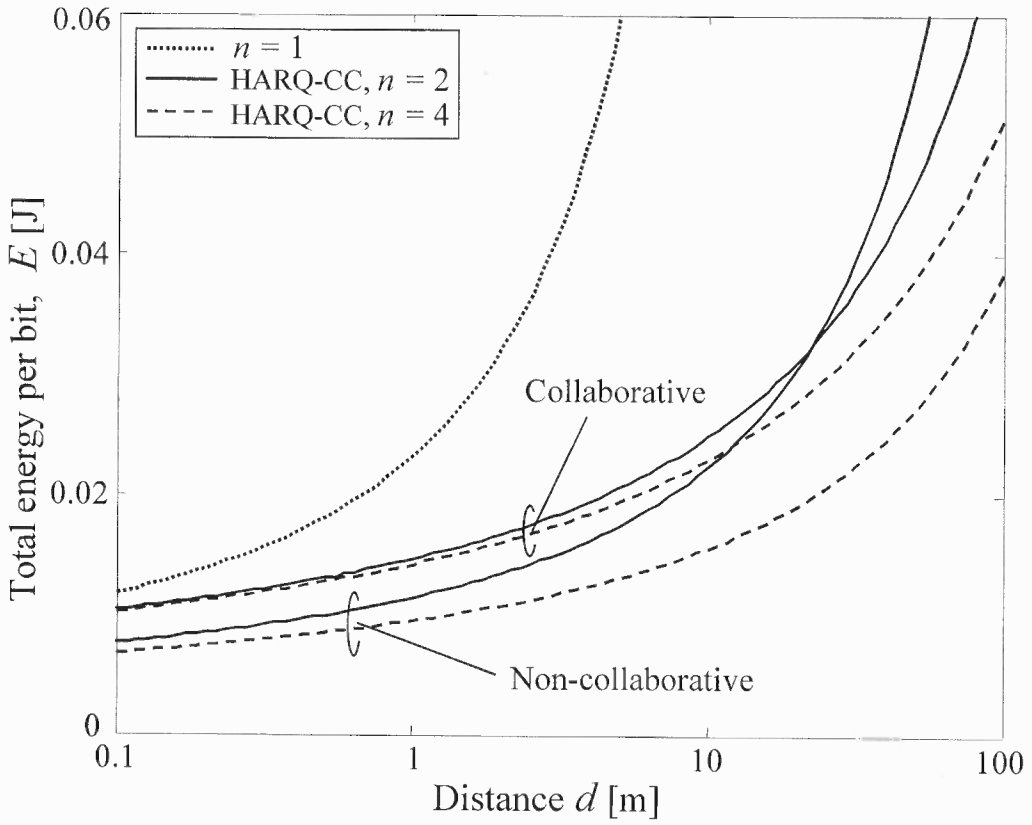

Figure 4.5 Non-cooperative and cooperative HARQ-CC protocol: minimum total energy per bit E versus the transmission range d ($p_{out} = 10^{-6}$).

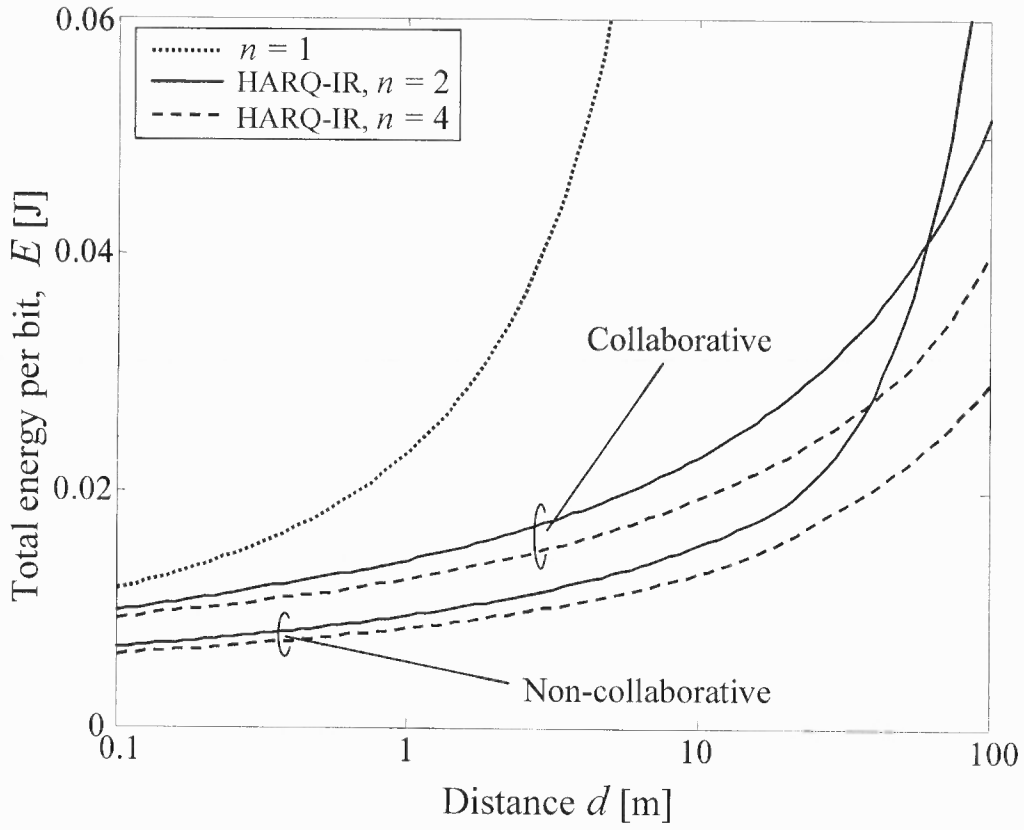

Figure 4.6 Non-cooperative and cooperative HARQ-IR protocol: minimum total energy per bit E versus the transmission range d ($p_{out} = 10^{-6}$).

soft packet combining at the receiver, similarly to HARQ-CC, versus the distance d, for $n = 2$, $p_{out} = 10^{-6}$ and two values of α, $\alpha = 0$dB and $\alpha = 20$dB. The time-diversity system is clearly inferior to HARQ systems, and even fails to outperform the system with $n = 1$ for small distances. Furthermore, due to the fact that all transmissions are exploited, increasing the number of transmissions can have an adverse effect on the total energy consumption in time-diversity systems (not shown here). Recall that, on the other hand, HARQ protocols overcome this problem and thus benefit with increased allowed number of retransmissions (i.e., delay). Finally, notice the slight increase of the energy consumption as the source-relay gain α decreases; with additional transmissions allowed, however, the influence of decreased α becomes negligible (not shown), as the relays avail more opportunities for decoding.

4.3.1 Processing energy of HARQ protocols

In this section, a coarse study of the influence of the energy consumed by baseband processing (*processing energy*) for different HARQ protocols is provided. It is noted that such contribution is typically not considered in related studies [43] [30] on the ground that it is generally negligible as compared to the consumption of power amplifier and other circuitry. However, here this analysis is of interest due to the different processing energy of the considered HARQ schemes. To elaborate, the focus herein is on the non-cooperative scheme and the processing energy per (re)transmission per bit consumed by both the source (coder) and the destination (decoder) is denoted as E_{pc} [1], which can take the values $E_{pc,TI}$, $E_{pc,CC}$ and $E_{pc,IR}$ for HARQ-TI, HARQ-CC and HARQ-IR protocols, respectively. The model and the analysis for minimizing the energy consumption are those presented in Section 4.1, with the only difference that in (4.6) it is now used

$$E_t + E_r = d^{\gamma} K_t E_b + P_c \frac{T_{on}}{L} + E_{pc}, \tag{4.40}$$

[1] Unlike the transmitting/receiving circuitry, the processing energy is here reasonably defined as a constant per transmission per bit, so as to mimic the approximately linear dependence of processing power versus the bit rate [51].

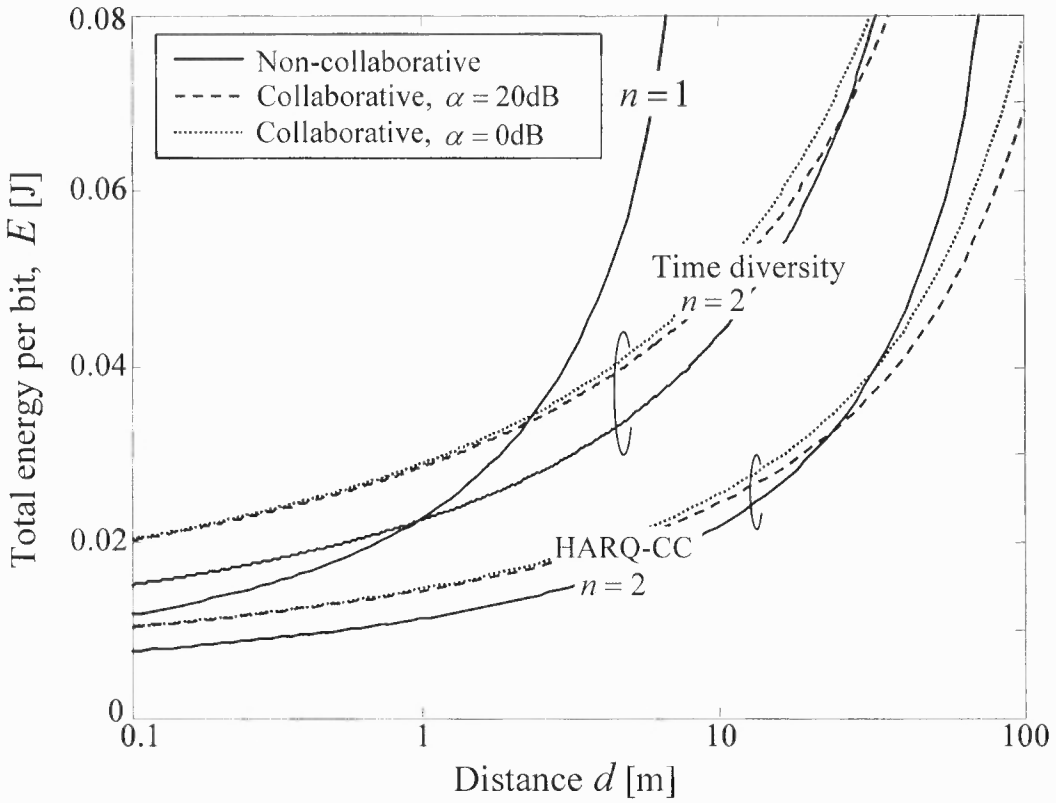

Figure 4.7 Non-cooperative and cooperative HARQ-CC protocols and time-diversity scheme: minimum total energy per bit E versus the transmission range d, for two values of α ($p_{out} = 10^{-6}$).

where E_{pc} depends on the employed protocol.

The following experiment is performed to assess under which conditions on the processing energy, HARQ-CC and HARQ-IR protocols preserve their benefits as compared to HARQ-TI. For given processing energy $E_{pc,TI}$, using the tools described in Section 4.1 and the numerical optimization described above, the optimal (minimized) energy per bit is found for HARQ-TI protocol. This value is denoted as the reference energy E_{TI}. For HARQ-CC and HARQ-IR protocols instead, the processing energies $E_{pc,CC}^{\max}$ and $E_{pc,IR}^{\max}$ that yield the overall energy consumption E_{TI} will be found. Thus, $E_{pc,CC}^{\max}$ and $E_{pc,IR}^{\max}$ stand for the available processing energies of HARQ-CC and HARQ-IR that would retain their advantage (in terms of energy consumption) comparing to HARQ-TI.

Figure 4.8 illustrates the energy levels E_{TI}, $E_{pc,CC}^{\max}$ and $E_{pc,IR}^{\max}$ versus the distance d, for $n = 2$, $p_{out} = 10^{-6}$, $E_{pc,TI} = 1.26 \cdot 10^{-9}$J (as specified for Viterbi decoder in IEEE 802.11a [52] [53]) and two values of the bandwidth, $B = 10$kHz and $B = 5$MHz. It can be seen from Figure 4.8 that for narrowband application ($B = 10$kHz), the available processing energies exceed by far the practically encountered processing energy levels (typically around several nJ [53]). On the other hand, for large bandwidths ($B = 5$MHz, as for maximum channel spacing in 802.15.4 [54]) and extremely small transmission ranges ($d < 1$m), the baseband energy can become a potential limitation for both HARQ-CC and HARQ-IR protocols. Finally, notice that throughout this section the normalized value $B = 1$Hz has been used. Larger bandwidths typically reduce the transmission energy (recall, e.g., (4.9)), which in turn reduces the optimal T_{on} and, therefore, the circuitry and the overall optimal energy value. The general conclusions of this section, however, are not affected by the choice of B.

4.4 Chapter Conclusion

In this chapter, the advantages of cooperative HARQ protocols have been reconsidered from the standpoint of energy efficiency, accounting for scenarios where the energy

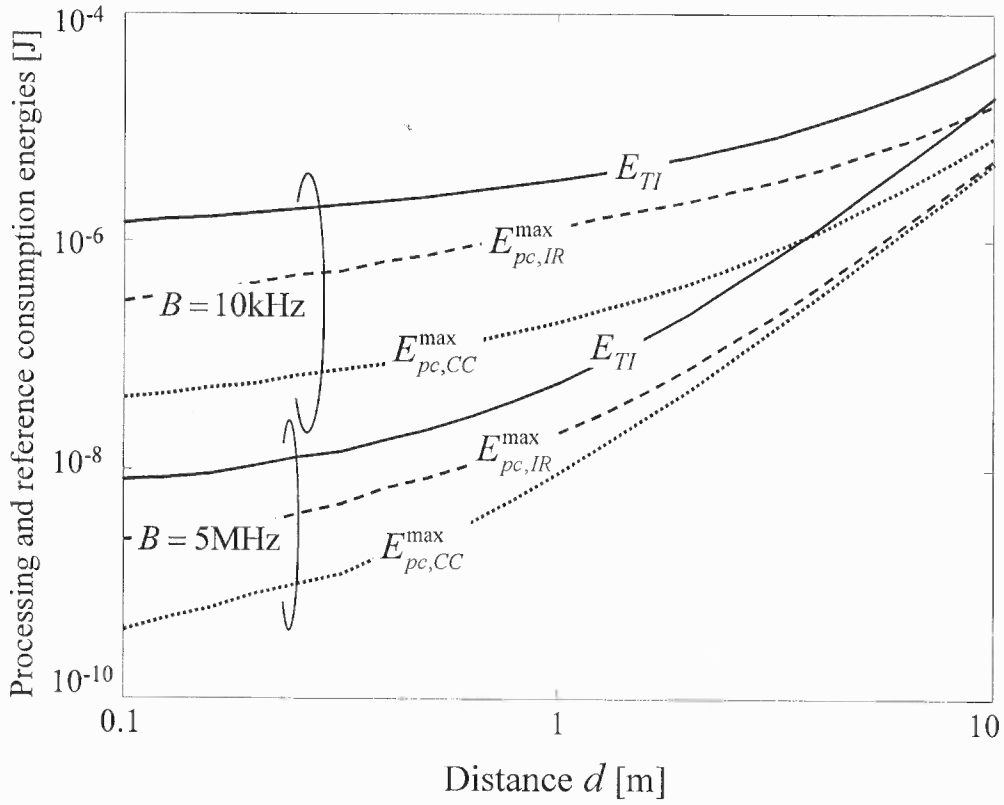

Figure 4.8 Minimum energy consumption E_{TI} spent by HARQ-TI for $E_{pc,TI} = 1.26$nJ, and maximum processing energies $E_{pc,CC}^{\max}$ and $E_{pc,IR}^{\max}$ that preserve the gains of the HARQ-CC and HARQ-IR protocols, respectively, over HARQ-TI, versus the transmission range d, for two values of bandwidth B ($p_{out} = 10^{-6}$, non-cooperative HARQ).

consumption due to the electronic circuitry is comparable to the transmission power. It has been demonstrated via numerical results that the total energy consumption can be reduced by optimizing over both transmission energy per bit and the packet transmission time (i.e., in practice the transmission rate of the original transmission). The results show that relevant gains in terms of energy efficiency can be achieved by resorting to HARQ protocols rather than conventional time-diversity schemes, where the number of retransmissions is fixed. Insights into optimal design choices are also provided. For instance, it is shown that the conventional optimization approach, whereby the transmission energy is selected to be the minimum value that satisfies the quality-of-service requirements, is suboptimal in the presence of circuitry energy consumption. Turning to cooperative HARQ protocols, the limitations of such schemes in terms of energy consumption have been identified in scenarios with small transmission ranges (of the order of a few tens of meters) and significant circuitry consumption. Finally, it was demonstrated that the processing energy requirements for sophisticated HARQ protocols, such as HARQ-CC and HARQ-IR, can pose practical limitations for large bandwidths and extremely small transmission ranges.

CHAPTER 5

POTENTIALS OF GAME THEORY: DISTRIBUTED TRANSMISSION SCHEDULING

The purpose of this chapter is to demonstrate an application of the game-theoretic tools (Section 2.2) to decentralized (distributed) communication networks. Focusing on the transmission scheduling problem in the networks with uncoordinated transmitters (here, cells), it is shown by using game theory that the transmitters can dynamically adjust their transmission power level and transmission resource (time-frequency) in a manner that efficiently exploits resource reuse or schedules the orthogonal transmission for low or high-interfering cells, respectively. Transmission decisions at the base stations rely solely on the measurements of interference created by other transmitters. This example qualitatively illustrates possible benefits from applying game theory, which is the major tool later to be used in Chapter 7 and Chapter 8 for analyzing dynamic spectrum access and relaying motivation schemes for distributed communications.

5.1 Background

Transmission scheduling [55] is an attractive alternative to access randomization [56] for the mitigation of the intercell interference in wireless cellular networks. Multicell scheduler is expected to address the resource requests from multiple base stations and to allocate the access in time and/or frequency in order to guarantee a close-to interference-free coexistence [55]. In this perspective, power control schemes can be considered as simple multicell schedulers where the optimization is restricted only to the adaptation of the power levels [57]. A straightforward solution to multicell resource management would require a central scheduler that communicates with base stations through a high-speed (optical) backbone. However, this approach requires a signalling overhead (as well as control

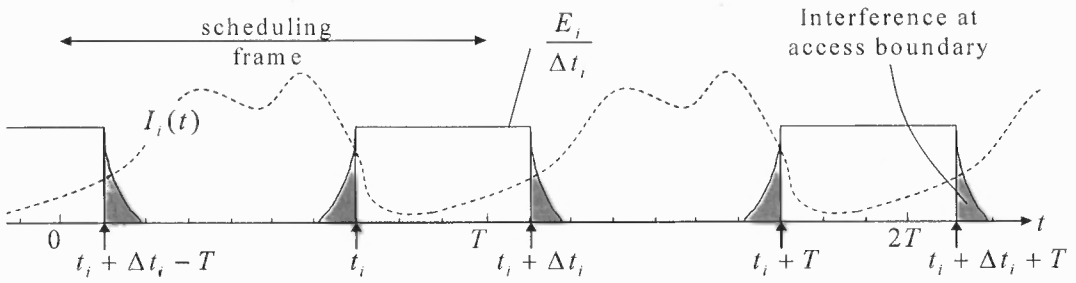

Figure 5.1 Illustration of interference-mitigating time schedule for ith cell according to the interference power profile $I_i(t)$.

signalling protocols for intercell control messaging) that cannot be accommodated in current systems [58]. Instead, resource scheduling is conventionally carried out locally by the base station for each cell without any signalling or exchanging transmission parameters.

Multicell scheduling without the explicit intercell signalling can rely only on the local measurement of the intercell interference and exploit it as an implicit intercell signalling. In this context, in this chapter a distributed intercell transmission scheduling scheme is proposed where each base station determines the *resource access* based only on the local estimate of the cell interference. The basic idea is illustrated in Figure 5.1 for time scheduling. According to the power profile $I_i(t)$ of the interference experienced by the ith cell, the services for the user of interest are scheduled within the interference minima. In particular, each base station schedules the use of the time/frequency resource so as to minimize the effect of interference and thus maximize the cell's goodput. Any change in the allocation of the resource for a single cell triggers the change in the interference measured by other cells that, in turn, exploit it as the new information to alter their transmission schedules. This loops back and induces the interference change to the first cell. It is noted that the scheme can be considered as an adds-on to the widely investigated power control paradigm [57], even if here it is assumed that each transmission is energy limited and adapted in time or frequency.

In order to gain insight on practical systems, this chapter will also refer to WiMAX or 3GPP-LTE as wireless protocol since it schedules both time and frequency with OFDM (orthogonal frequency division multiplex) multiple access (OFDMA) [56]. To simplify, it is further assumed that the scheduler decouples the resource assignment in time and frequency as for a dynamic access in decentralized TDMA and FDMA (time and frequency division multiple access, respectively) [59]. Since cells are not necessarily (time and frequency) synchronized, distributed scheduling is constrained not to fragment the time/frequency access as this would raise the intercell interference due to the interference leakage at the time/frequency access boundaries as sketched in Figure 5.1 (e.g., time dispersion due to multipath channel response for TDMA, or the frequency offset for FDMA) [60]. Moreover, avoiding the resource fragmentation can increase the resource utilization by, e.g., limiting the (intracell) signaling overhead and/or the number of training sequences (for time scheduling).

Game theory [33] has shown to be a natural framework to define interference-mitigating access methods in decentralized wireless networks [61] [62]. In particular, it has been successfully used in cognitive radio systems, where each (secondary) user accesses the spectrum according to the level of interference and thus it schedules the transmission over interference-free frequency intervals [63]. Motivated by these results, the game-theoretic tools are implemented here to investigate the equilibria existence, the convergence and the consequent performance of the proposed scheduling scheme. As discussed in Section 5.2, each scheduler acts as a player that maximizes local utility function (goodput) while implicitly interacting with others through intercell interference. For analytical tractability, in Section 5.3 the focus is on a simplified two-cells interference dominated model, even if similar conclusions hold (shown through numerical simulations in Section 5.4) for any number of cells in practical scenarios.

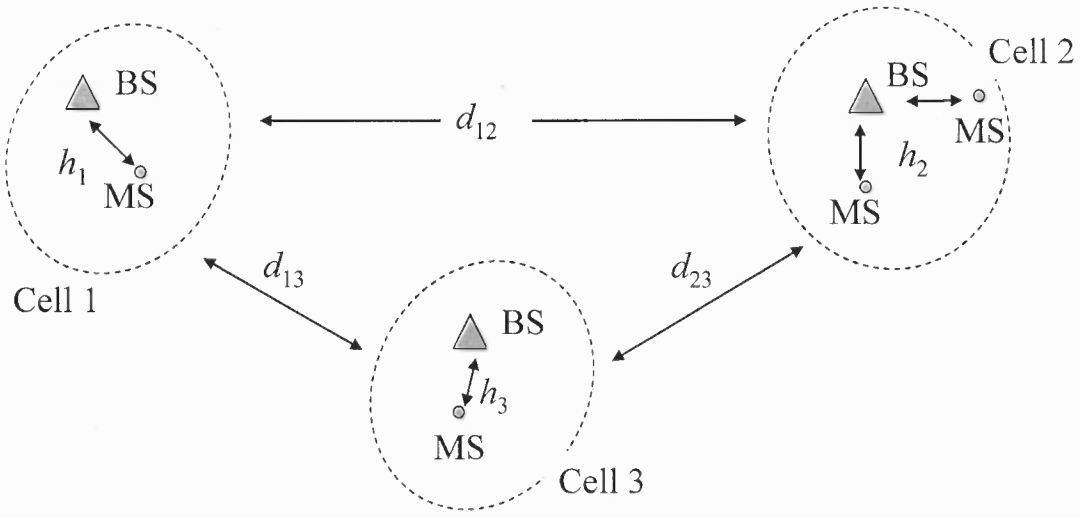

Figure 5.2 Illustration of a cellular system with $N = 3$ cells.

5.2 Problem Definition

5.2.1 Model Description and Scheduler Definition

Consider the system with N cells without a scheduler-dedicated backbone connection, as illustrated in Figure 5.2 (for $N = 3$). Transmissions between base station (BS) and mobile stations (MS) are controlled by the BS scheduler and performed over the channel with power gain h_i within the ith cell. The overall energy available for each intracell transmission is constrained to E_i and it is uniformly distributed over the accessed resource (Figure 5.1). For analytical convenience, the distance between any two cells (say d_{ij}) is considered significantly larger than the cell dimensions so that the intercell interference at BS and MS is identical (notice that interference is known to be asymmetric, extension to this case is straightforward). Interference power attenuation is reciprocal and reads $h_{ij} = h_{ji} = \alpha d_{ij}^{-\gamma}$, with path loss exponent γ and α as a scaling term that accounts for fading. All channel gains are assumed constant during the period of interest.

Cells share the time/frequency resource as for the OFDMA system. Here, a decoupled problem is considered where time and frequency access are considered separately. It is noted that scheduling for joint time and frequency resources could be

based on the same principles described here, without any conceptual modifications except for a more complex notation. To avoid the fragmentation of resource allocation, the access scheduled within any cell is constrained only to contiguous intervals. In the following the focus is on scheduling over time as for TDMA, extension to frequency allocation as for FDMA is straightforward.

In TDMA, all cells transmit periodically with the same nominal scheduling frame T using a set of time-slots according to a common transmission protocol (Figure 5.1), without any common synchronization reference (except the frame period T, usually set by the standard). Because of periodicity of the access, herein only one period $t \in [0, T]$ is taken into account. TDMA scheduling is now conveniently defined as the set $\{t_i, \Delta t_i\}$, where $t_i \leq T$ is the beginning time of the access to the frame and $\Delta t_i \leq T$ is the amount of occupied contiguous time intervals (or slots). Each slot is assumed to support the transmission of an uncoded word (to simplify) consisting of b differential phase shift keying symbols (any other modulation scheme can be easily accommodated [62]). Each of the b symbols is transmitted with power $E_i/(b\Delta t_i)$ and it is impaired by additive white Gaussian noise (AWGN) with power $\sigma_N^2 = N_0 B$, where N_0 is the single-sided spectral density and B is the available bandwidth. The intercell interference is time-varying due to the different scheduling strategies employed by each BS, even if the channels are assumed static

$$I_i(t) = \sum_{\substack{j \neq i \\ t \in \left(t_j, \langle t_j + \Delta t_j \rangle_T\right)}} \frac{h_{ji} E_j}{b \Delta t_j}, \tag{5.1}$$

where $\langle \cdot \rangle_T$ denotes the modulo T operation. The overall impairment is modelled as Gaussian with power $\sigma_i^2(t) = \sigma_N^2 + I_i(t)$.

Interference measurement experienced by the ith user $\sigma_i^2(t)$ across the frame interval is assumed perfectly accurate and used as information on the scheduling strategy adopted by the neighboring interfering cells. It is noted that impact of using different estimation techniques and of the estimation errors is not considered here. Local scheduling is

performed so as to fit the transmissions into one of interference minima (similarly to cognitive radio approach without any privileged user [63]) in order to maximize the cell's goodput

$$u_i(t_i, \Delta t_i; \boldsymbol{\sigma}_i^2(t)) = \sum_{t=t_i}^{(t_i+\Delta t_i)_T} p_{S,i}(t), \tag{5.2}$$

where $p_{S,i}(t)$ is the probability of successful transmission in ith cell during the tth slot that reads [64]

$$p_{S,i}(t) = \left(1 - \frac{1}{2}\exp\left(-\frac{h_i E_i/\Delta t_i}{b\sigma_i^2(t)}\right)\right)^b. \tag{5.3}$$

Notice that the only scheduling parameters in (5.2) that can be controlled by the ith BS are t_i and Δt_i, while the impairment $\boldsymbol{\sigma}_i^2(t)$ is out of the control of the scheduler and it is attained through measurements.

Distributed scheduling is based on the simple *best response rule*

$$\{t_i(k), \Delta t_i(k)\}_{opt} = \arg \max_{\{t_i(k), \Delta t_i(k)\}} u_i\left(t_i(k), \Delta t_i(k); \boldsymbol{\sigma}_i^2(t,k)\right) \tag{5.4}$$

that is updated periodically in each frame (k denotes the frame or iteration number) independently by every BS. According to (5.4), any change in the transmission scheduling of a single cell will alter the interference experienced by all the neighboring cells, thus highlighting the concept of interference-based intercell signalling.

5.2.2 Game Setup

Game theory is a natural framework to characterize the equilibria of the proposed scheme. The cells (or *players*) compete for access behaving in a *selfish* (player is interested solely in maximizing its own benefit, without any concern for the collective good) and *rational* (player chooses only the strategies that are the best responses to his opponents' strategies) manner, as discussed in Section 2.2. As defined in Section 5.2.1, the players adapt their transmission schedule $\{t_i, \Delta t_i\}$ (*strategies*) so as to maximize the goodput u_i in (5.2).

To formally comfort to a typical game-theoretic model where the utility of a player is explicitly impacted by the strategies of other players, the power of overall impairment $\boldsymbol{\sigma}_i^2(t)$ in $u_i(t_i, \Delta t_i; \boldsymbol{\sigma}_i^2(t))$ is replaced by the set of scheduling strategies $\mathbf{t}_{-i}, \boldsymbol{\Delta}\mathbf{t}_{-i}$, where $\mathbf{t}_{-i} = \{t_1, .., t_{i-1}, t_{i+1}, .., t_N\}$ and $\boldsymbol{\Delta}\mathbf{t}_{-i} = \{\Delta t_1, .., \Delta t_{i-1}, \Delta t_{i+1}, .., \Delta t_N\}$. To cast the analytical model and express the interference using $\mathbf{t}_{-i}$ and $\boldsymbol{\Delta}\mathbf{t}_{-i}$, it is assumed that the players are aware of interference channel gains h_{ij} and the powers used by other players (recall (5.1)). Notice that these assumptions are required only for analytical purposes; for the practical implementation, the scheme relies solely on the interference estimation.

The key concept of game theory is *(Nash) equilibrium* (Section 2.2): the state where any unilateral deviation in player's strategy would not produce any benefit. Herein, it is given by

$$\{t_i^*, \Delta t_i^*\} = \arg \max_{\{t_i, \Delta t_i\}} u_i(t_i, \Delta t_i; \mathbf{t}_{-i}^*, \boldsymbol{\Delta}\mathbf{t}_{-i}^*),\ i = 1, .., N, \tag{5.5}$$

with $\{t_i^*, \Delta t_i^*\}$ denoting the ith player's strategy in the equilibrium, and it is the focus of the remaining of the chapter.

5.3 Equilibrium Analysis (Two Cells)

This section provides analysis for the equilibria of the game described in Section 5.2.2 for a scenario where interference dominates ($I_i(t) \gg \sigma_N^2$) the overall performance $\sigma_i^2(t) \simeq I_i(t)$. A two-cells system is assumed with $E_1 = E_2 = E$ and channels $h_1 = h_2 = 1$, $h_{12} = h_{21} = h$ (symmetric system). It can be proved (not discussed here) that the similar equilibria to the ones discussed herein for $N = 2$ cells exist for any $N > 2$. In Section 5.4, besides corroborating the analysis for this simplified model, numerical results are provided for more practical scenarios, accounting for the noise impact and for an arbitrary number of cells.

Due to the cells' rationality (recall Section 5.2.2) and the low-noise assumption, there cannot be any unoccupied slots within the frame, so that $\Delta t_1 + \Delta t_2 \geq T$. Therefore,

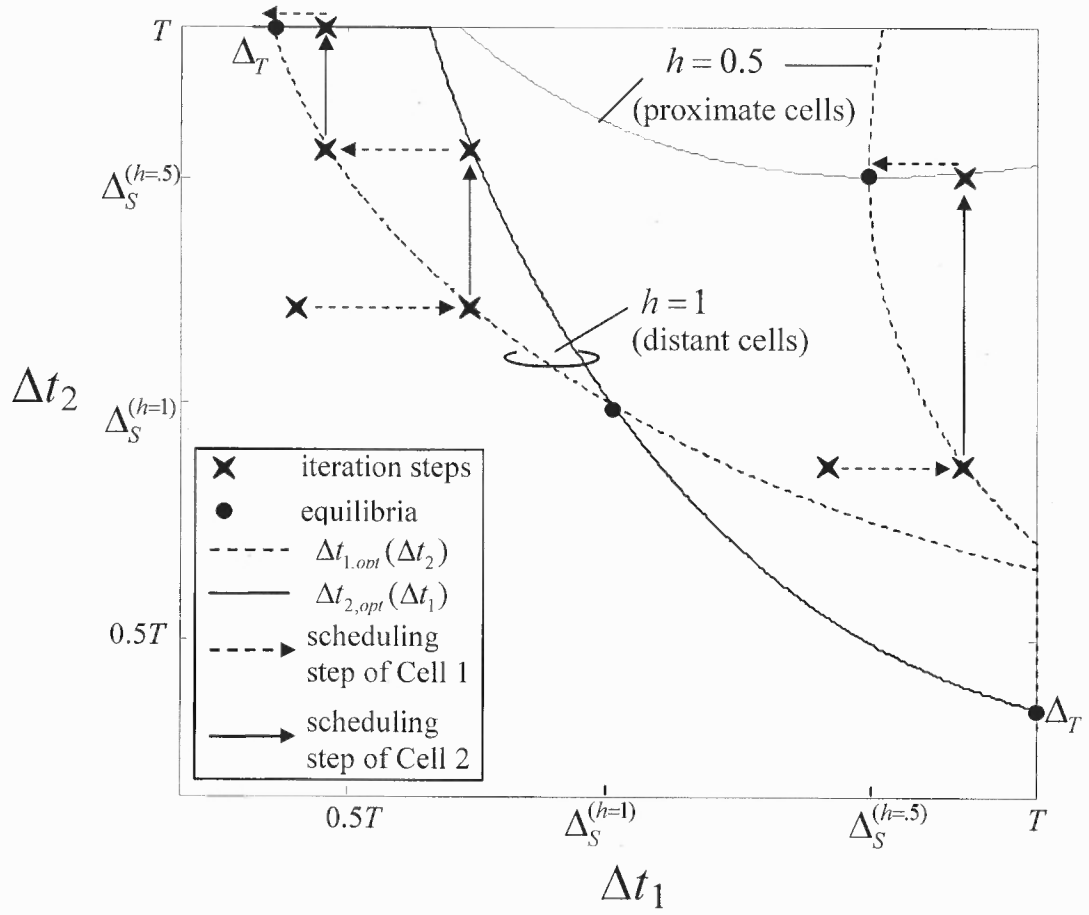

Figure 5.3 Optimal response $\Delta t_{1,opt}(\Delta t_2)$ and $\Delta t_{2,opt}(\Delta t_1)$, system equilibria and illustration of the best response algorithm convergence (for $h = 0.5$ and $h = 1$).

any choice of $\{t_1, \Delta t_1\}$ and $\{t_2, \Delta t_2\}$ results in the channel access with $\Delta t_1 + \Delta t_2 - T$ slots occupied simultaneously by two players (interfered slots), while the partitions of interference-free slots (as with $p_{S,i}(t) = 1$) occupied by the first or the second players is $T - \Delta t_2$ and $T - \Delta t_1$, respectively. It can be easily shown that there are exactly $T(\Delta t_1 + \Delta t_2 - T)$ such scenarios corresponding to different choices of the pair (t_1, t_2). Utility function (goodput) boils down to a dependence on strategies Δt_1 and Δt_2. For the first player, it reads

$$u_1(\Delta t_1; \Delta t_2) = (T - \Delta t_2) + (\Delta t_1 + \Delta t_2 - T)\left(1 - \frac{1}{2}e^{-\frac{\Delta t_2}{h\Delta t_1}}\right)^b. \tag{5.6}$$

It is the same for the second player, except for a switch of indexes in the notation due to the symmetry.

As the guideline for the analysis, Figure 5.3 shows the optimal responses $\Delta t_{i,opt}(\Delta t_j) = \arg\max_{\Delta t_i}(u_i(\Delta t_i; \Delta t_j))$, for any choice of Δt_j using (5.6) for $i, j = 1, 2$ and $h = 0.5$ or $h = 1$. Notice that the integer constraint on Δt_i is here neglected, as if the time slots were of infinitesimal duration. The system equilibria are marked in Figure 5.3 as the intersections of curves $\Delta t_{i,opt}(\Delta t_j)$ and can be obtained by jointly solving (if feasible) (5.5)

$$\Delta t_1^* = \arg\max_{\Delta t_1} u_1(\Delta t_1; \Delta t_2^*) \tag{5.7a}$$

$$\Delta t_2^* = \arg\max_{\Delta t_2} u_2(\Delta t_2; \Delta t_1^*). \tag{5.7b}$$

5.3.1 Symmetric Equilibrium

Due to the symmetry of the system, there exists a symmetric equilibrium with $\Delta t_1^* = \Delta t_2^* = \Delta_S$, as shown in Figure 5.3 for the channels $h = 0.5$ and $h = 1$, with the system (5.7) that boils down to $\Delta_S = \arg\max_{\Delta t_1} u_1(\Delta t_1; \Delta_S)$ (or, equivalently, $\Delta_S = \arg\max_{\Delta t_2} u_2(\Delta t_2; \Delta_S)$). The value of equilibrium Δ_S follows by solving $\partial u_1(\Delta t_1; \Delta t_2)/\partial \Delta t_1 = 0$ (numerical simulations, not shown here, confirm the quasi-concavity of $u_1(\Delta t_1; \Delta t_S)$ over Δt_1 for all reasonable values of h, i.e., for $h \in [0, \bar{h}]$,

where $\bar{h} \gg 1$):

$$1 - \frac{1}{2}e^{-\frac{\Delta t_2}{h\Delta t_1}} = (\Delta t_1 + \Delta t_2 - T)e^{-\frac{\Delta t_2}{h\Delta t_1}} \frac{b\Delta t_2}{2h\Delta t_1^2}. \tag{5.8}$$

Replacing $\Delta t_1 = \Delta t_2$ with Δ_S and recalling that $\Delta_S \leq T$ yields

$$\Delta_S = \min\left\{T, \frac{T}{2}\left(1 - \frac{h}{b}e^{\frac{1}{h}} + \frac{h}{2b}\right)^{-1}\right\}. \tag{5.9}$$

The analysis of (5.9) with respect to channel gain h shows that for $h \leq h_{\min}$, with $h_{\min}$ obtained as the smaller solution of $\Delta_S(h) = T$, i.e.,

$$h(e^{\frac{1}{h}} - \frac{1}{2}) = \frac{b}{2}, \tag{5.10}$$

the full frame reuse occurs. For example, using $b = 8$ symbols per slot it is $h_{\min} \simeq .44$. It is noted that for $h \geq h_{\max}$, where $h_{\max}$ is the larger of two solutions of (5.10), a full frame reuse occurs. Since $h_{\max} \gg 1$ (as revealed by numerical simulations, not shown here), it is reasonable to assume that the full reuse occurs only when $h \leq h_{\min}$.

5.3.2 Asymmetric Equilibrium

Figure 5.3 shows the existence of asymmetric equilibria for $h = 1$, where one of the players occupies the full frame ($\Delta t_1^* = T$ or $\Delta t_2^* = T$). Assuming that $\Delta t_2^* = T$, the goodput (5.6) of the first player boils down to

$$u_1(\Delta t_1) = \Delta t_1 \cdot \left(1 - \frac{1}{2}\exp\left(-\frac{T}{h\Delta t_1}\right)\right)^b. \tag{5.11}$$

Since $u_1(\Delta t_1) \geq 0$, $\lim_{\Delta t_1 \to 0} u_1(\Delta t_1) = 0$ and $\lim_{\Delta t_1 \to \infty} u_1(\Delta t_1) = 0$, the strategy in the equilibrium, denoted as Δ_T, is given by solving $\partial u_1(\Delta t_1)/\partial \Delta t_1 = 0$:

$$\Delta_T = \min\left\{T, \frac{T}{h}\left(-1/b - \mathcal{W}_{-1}\left(-2/b\exp\left(-1/b\right)\right)\right)^{-1}\right\}, \tag{5.12}$$

where $\mathcal{W}_k(x)$ is the kth branch of the multivalued Lambert W function[1] [65].

[1]Lambert W function is the solution $\mathcal{W}(x)$ of $\mathcal{W}e^{\mathcal{W}} = x$. If x is a real number, two real values for $\mathcal{W}(x)$ are possible for $-e^{-1} \leq x \leq 0$: the principal branch $\mathcal{W}_0(x)$ with $\mathcal{W}_0(x) \geq -1$, and the

The minimal value of the interference channel gain $h = h_T$ for which this equilibrium exists, appears when (recall Figure 5.3) the points $\Delta t_{1,opt}(\Delta t_2 = T)$ and $\Delta t_{2,opt}(\Delta t_1) = T$ (equivalently, $\Delta t_{2,opt}(\Delta t_1 = T)$ and $\Delta t_{1,opt}(\Delta t_2) = T$) coincide. Applying these conditions to (5.7) and by using the first derivative approach (as in (5.8)), this value is shown to to satisfy

$$2e^{\frac{\Delta_T}{h_T T}} - 1 = b\frac{\Delta_T^2}{h_T T^2}, \tag{5.13}$$

where Δ_T is given by (5.12). For example with $b = 8$, (5.13) yields quite a large value: $h_T = 0.82$.

5.3.3 Convergence and Equilibrium Stability

This section is concluded by discussing the issue of convergence and equilibrium stability. Assume that, when using the best response algorithm, the cells play (i.e., determine their strategy and act) sequentially. For the two-cells case, the algorithm boils down to $\Delta t_{1,opt}(k) = \arg\max_{\Delta t_1(k)} u_1(\Delta t_1(k); \Delta t_2(k-1))$ and $\Delta t_{2,opt}(k) = \arg\max_{\Delta t_2(k)} u_2(\Delta t_2(k); \Delta t_1(k))$ and it is illustrated in Figure 5.3. Notice that, if the only existing equilibrium is symmetric ($h = 0.5$ in Figure 5.3), it is also the converging point of the algorithm (starting from an arbitrary point $(\Delta t_1, \Delta t_2)$). For scenarios where asymmetric equilibria also exist ($h = 1$ in Figure 5.3), the converging point of the algorithm is always asymmetric. Symmetric equilibrium for $h = 1$ is 'unstable' as any change of strategy would trigger the schedulers to move the system towards one of the two stable (asymmetric) equilibria.

5.4 Numerical Results

This section provides some insight into the performance of the proposed scheme. Firstly, the analysis is corroborated for the simplified scenario in Section 5.3 and the impact of

---

second branch $\mathcal{W}_{-1}(x)$ with $\mathcal{W}_{-1}(x) \leq -1$. In the problem at hand, $W_{-1}(x)$ yields the desired solution, while $\mathcal{W}_0(x)$ would result in a complex value.

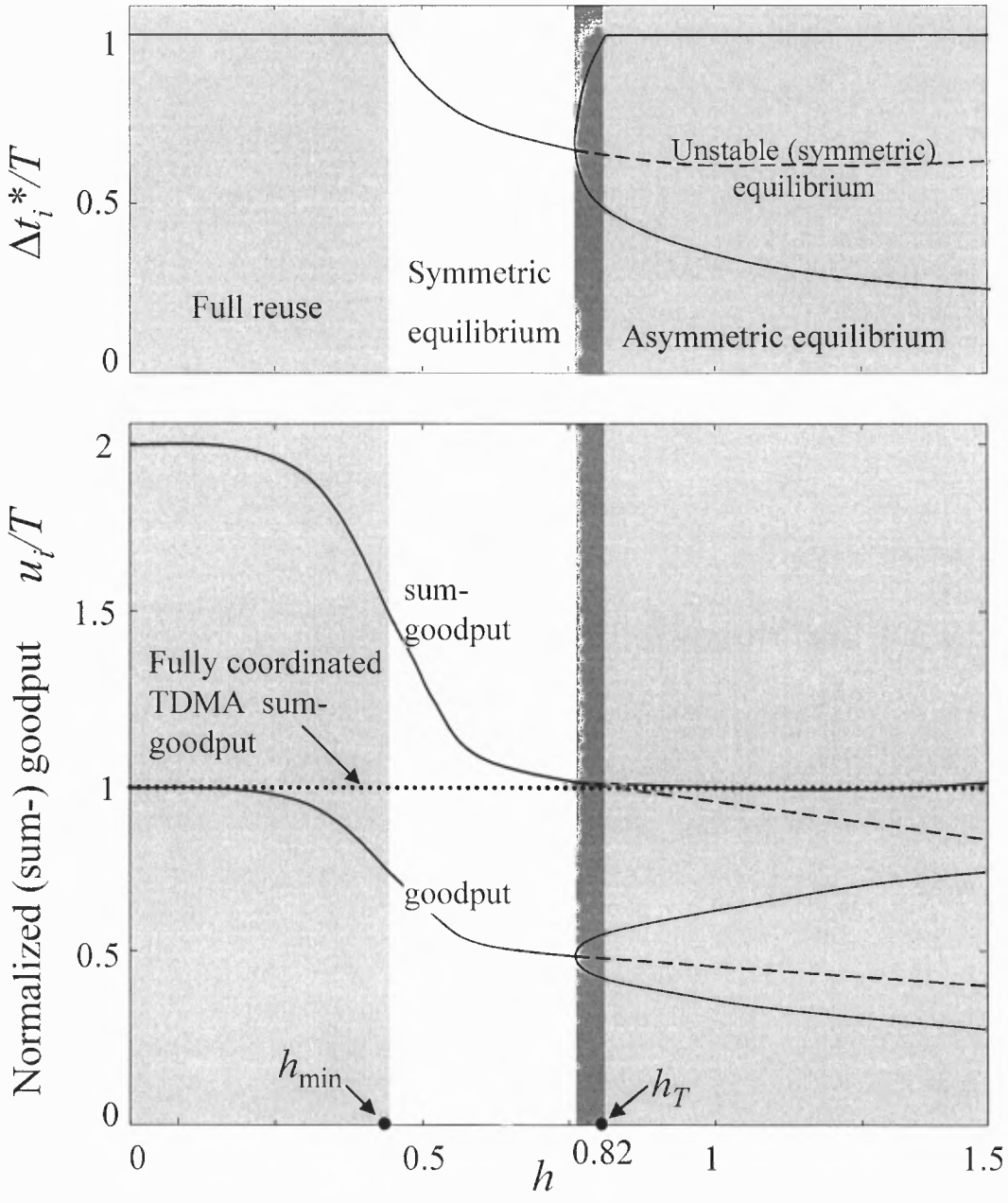

Figure 5.4 System equilibria and goodput versus h, for a symmetric interference-dominated two-cells scenario with $b = 8$.

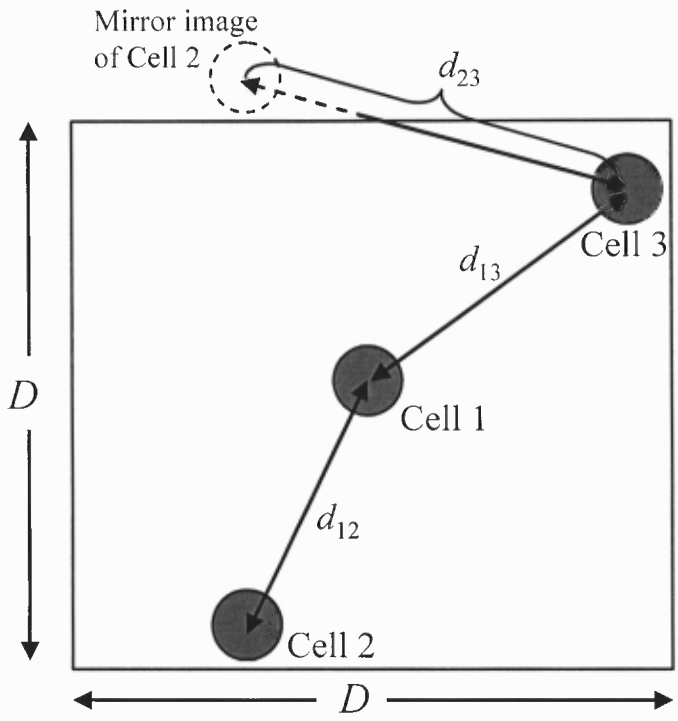

Figure 5.5 Illustration of $N = 3$ cells on a flat torus with both sides D.

interference is investigated. Then, a more practical scenario is studied, namely with i) an arbitrary number of cells and ii) accounting for the noise impact; the system performance in equilibrium is analyzed and an example is provided for the algorithm convergence to equilibria.

5.4.1 Corroborating Analysis

For the symmetric interference-dominated two-cells scenario analyzed in Section 5.3, Figure 5.4 shows all possible equilibria and the consequently obtained goodputs versus the interference channel h, for $b = 8$ (notice that this figure is, in fact, an extension of Figure 5.3). Figure 5.4 (upper figure) confirms the existence of equilibria addressed in Section 5.3 and shows that, for a relatively small range of interference channel h (dark shaded area), there also exist asymmetric equilibria with both Δt_1^*, $\Delta t_2^* < T$ (notice that the analytical

solution for this equilibrium using (5.7) is intractable and therefore not addressed in Section 5.3). Important values of h, namely $h_{\min}$ (5.10) and h_T (5.13), are also shown. For small interference ($h \leq h_{\min}$), cells exploit full frame with $\Delta t_1^* = \Delta t_2^* = T$. As h increases, Δt_i^* reduces and (in the case of symmetric equilibrium) the scheme resembles the TDMA with interfered frame portion around $0.2T \div 0.4T$. For approximately $h \leq 0.76$ the only existing equilibrium and, thus, the converging point of the scheduling algorithm (recall discussion in Section 5.3.3 and Figure 5.3) is symmetric. Raising the interference leads to asymmetric converging point and the symmetric equilibrium becomes unstable. As for the goodput (lower figure), it is shown that the proposed scheme significantly outperforms fully coordinated TDMA (with no reuse) for a relatively small interference channel h (sum-goodput is twice as large as in TDMA for $h \longrightarrow 0$). As interference increases and the asymmetric equilibrium becomes dominant, performance for two schemes is almost identical in terms of a sum-goodput. It is also interesting to note that for asymmetric equilibria, the sum-goodput is larger than the sum-goodput achievable in the (unstable) symmetric equilibrium for the same h. Interestingly, in the asymmetric equilibrium the cell that occupies a larger resource portion benefits with increase of the interference channel, as the other player is forced to occupy smaller portion (recall (5.12)). For the symmetric equilibrium, both cells suffer with increase of the interference channel h. It is also noted that a larger (smaller) number of symbols per slot b decreases (increases) the number of interfered slots, as the players are more (less) prone to interference, but does not qualitatively change the system equilibria (not shown here).

5.4.2 System Performance in Realistic Environment

Here, a more realistic scenario is considered, with frequency reuse pattern characterized by N cells randomly placed on a flat torus with both sides D (a flat torus is used instead of a square so as to avoid the 'privileged' cells at the square edges with a low interference level), as illustrated in Figure 5.5 for $N = 3$ cells. Interference channels are modeled

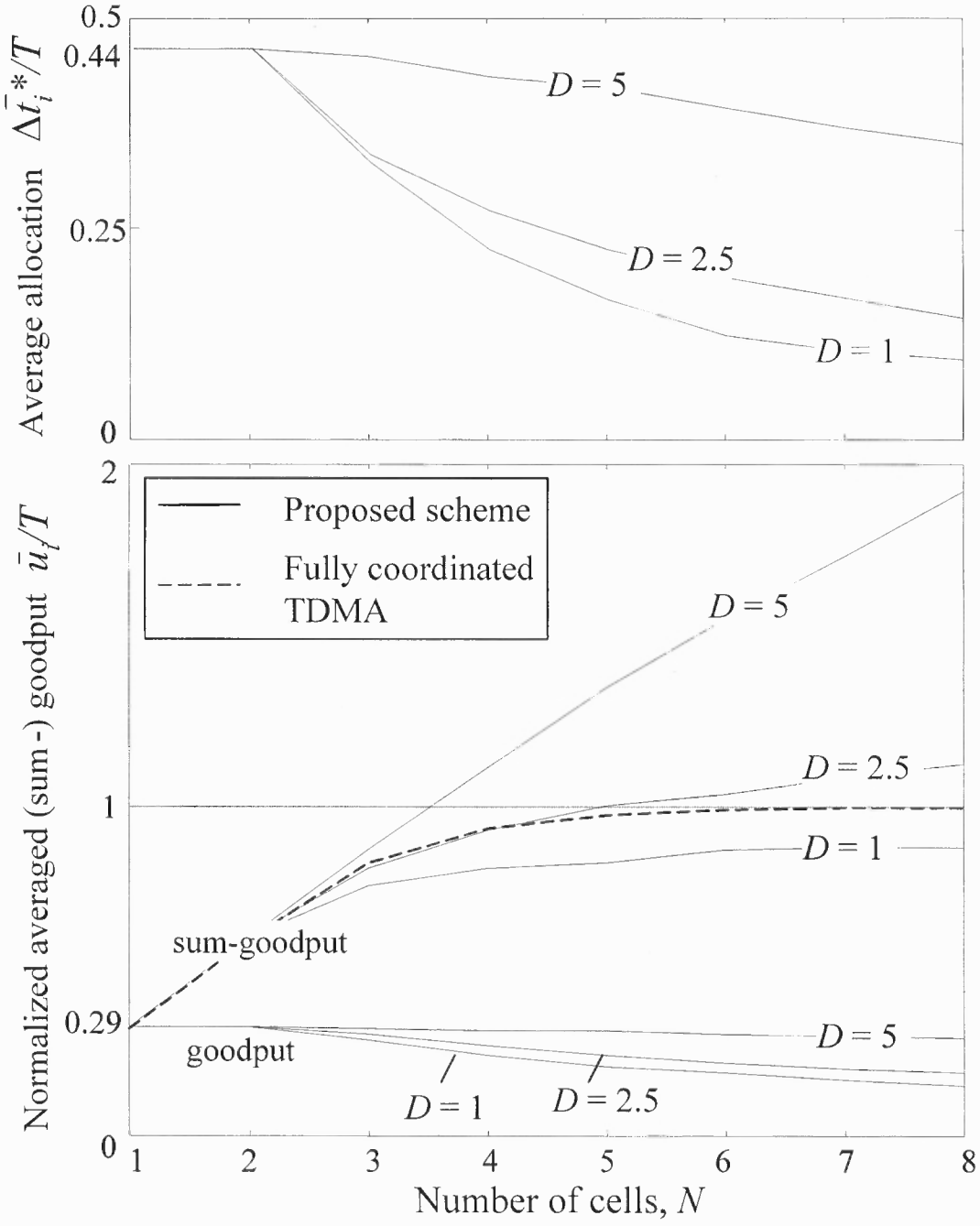

Figure 5.6 Average time allocation and goodput versus the number of cells, for different space (torus) dimensions D and $\sigma_N^2 = E/T$ and $b = 8$.

using the propagation model $h_{ij} = d_{ij}^{-\gamma}$ (no fading) with $\gamma = 3$, d_{ij} is the (minimum) distance between ith and jth cell, $h_i = 1$, $E_i = E$, $b = 8$ and *non-negligible noise* power $\sigma_N^2 = E/T$. Equilibria are achieved using the best response algorithm, as in (5.4) and Section 5.3.3. Figure 5.6 shows the time allocation $\Delta\bar{t}_i^*$ (upper figure) and corresponding (sum-)goodput $\bar{u}_i$ (lower figure) when averaging with respect to different cell placements, versus the number of cells. For the reference, the sum-goodput achievable with a fully coordinated TDMA is also shown. Notice that for $N \leq 2$ there is no interference, as the noise dominates the system performance and limits the strategies to $\Delta t_i^* \approx 0.44T$ (this value can be derived using analysis similar as in Section 5.3.2 or simply replacing T/h in (5.12) with E/σ_N^2). Due to the noise impact, the corresponding goodput is always lower than the allocated range, $u_i(\Delta t_i) < \Delta t_i$ (e.g., for $N = 1, 2$, $u_i \approx 0.29T < \Delta t_i^* \approx 0.44T$). As expected, the large number of cells, as well as the smaller space dimension D, reduce the cells' goodputs in equilibria. Increasing space dimension D reduces the interference level and allows multiple cells to reuse the time and outperform coordinated TDMA in terms of sum-goodput. Notice that due to the noise impact and constant energy per user, in fully coordinated TDMA (with the time fraction allocation that scales as $1/N$), the signal-to-noise ratio increases linearly with the number of cells N and the (normalized) sum-goodput approaches 1.

Figure 5.7 illustrates the dynamic response of the best-response algorithm in a realistic scenario with additionally activated/deactivated cells, thus changing network topology. In particular, Figure 5.7-(a) shows the randomly generated placement of $N = 8$ cells with $D = 2.5$ (as described in previous example and in Figure 5.5), Figure 5.7-(b) illustrates the cells' strategies versus time (in frames), whereas Figure 5.7-(c) highlights the cells' powers allocated in the last frame. At this point, it is first noted that the algorithm (5.4) can lead to significant discontinuities in the time access $\{t_i(k), \Delta t_i(k)\}$ over consecutive frames k, as the strategy at the kth frame $\{t_i(k), \Delta t_i(k)\}$ is optimized with no regard to the strategy used in the previous $((k-1)$th) frame (recall (5.4)). Since

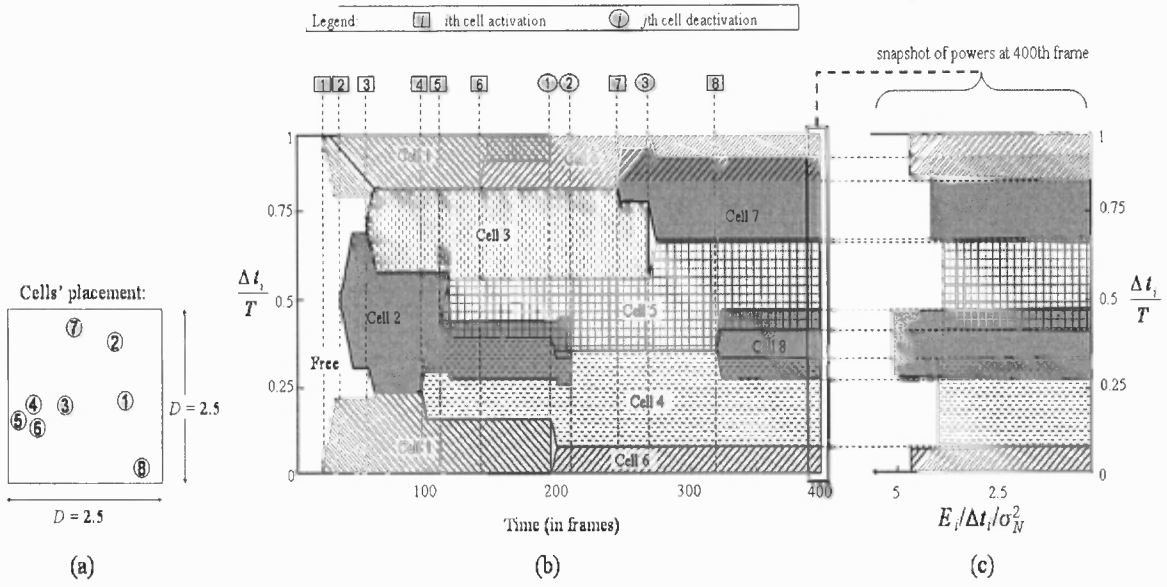

Figure 5.7 Cells' (random) placement (a), system response to the cells' activation/deactivation (b) and the snapshot of the cells' power distribution at the last 400th frame (c) for $D = 2.5$, $\sigma_N^2 = E/T$ and $b = 8$.

this issue can raise problems in signalling overhead, the algorithm employed in Figure 5.7 is modified so that the search of scheduling strategies in (5.4) becomes incremental $t_i(k) \in (t_i(k-1) \pm \delta)$, $\Delta t_i(k) \in (\Delta t_i(k-1) \pm \delta)$, where the incremental value is chosen as $\delta = 0.02T$. It can be noticed from Figure 5.7 that the system converges to equilibria relatively fast (in fact, it is limited only by the factor δ), in the order of 10 iterations (frames) and exhibits a TDMA-like performance (for the algorithm as given in (5.4), the rate of convergence would be approximately $2 \div 4$ frames (not shown here)). Notice that as N increases, the time reuse is exploited by relatively distant cells, while the cells within a close distance (i.e., with a strong interference, such as cells 4, 5 and 6 in Figure 5.7) maintain an interference-free scheduling.

5.5 Chapter Conclusion

In this chapter, benefits of using game theory in analysis and design of decentralized networks were illustrated on distributed transmission scheduling example. In particular,

the distributed intercell transmission scheduling scheme is proposed for non-fragmented channel access in cellular networks where independent schedulers rely only on sensing of the medium occupancy. Equilibria were studied using game theory for a symmetric interference-dominated scenario with two cells. The distributed scheme behaves similarly to conventional TDMA (or FDMA) but can efficiently adhere to full resource reuse when the interference is not detrimental. Numerical results confirm that the interference avoidance scheduling scheme is largely beneficial and could be easily adopted as distributed transmission scheduling for any cellular (or a hot-spot based) network without meaningful modifications of the existing control signalling.

Having qualitatively confirmed that game theory is an essential framework for engineering distributed networks, in the following chapter *Stackelberg game*, a game-theoretic model that can be used for an implicit control of non-cooperative nodes, is introduced. This model will then be used for the design of the dynamic spectrum access / relay motivation mechanism in Chapter 7.

CHAPTER 6

POTENTIALS OF GAME THEORY: IMPLICIT CONTROL OF NON-COOPERATIVE NODES

Before proceeding to the dynamic spectrum access and relaying motivation mechanisms described in Chapter 7, herein two basic prerequired concepts are introduced and investigated. Firstly, a multiple access channel (MAC) model where the transmitting terminals can adapt their power in order to improve the performance in some sense [66] is considered. Particularly, the focus is on a non-cooperative setting and, accordingly, on the *power control game* [62]. Secondly, a game-theoretic concept *Stackelberg game* [33] is considered as a framework for implementing and analyzing the mechanisms for implicit coordination of selfish and uncoordinated nodes.

6.1 Background

Power control is typically employed in uplink wireless channels in order to guarantee a sufficient strength of the user's signal while limiting its interfering effect on signals belonging to other users [66]. Optimal power control mechanisms require the access point (AP) to be able to control directly the power transmitted by mobile stations (MSs). This cannot be guaranteed in some wireless networks, such as in systems complying with the cognitive radio principle, where competitive behavior is expected to be predominant [20].

In this chapter, a system with decentralized power control (see, e.g., [62]) is considered. The fact is exploited that, although the MSs are not directly controlled by the AP, the power control game they participate in, along with its Nash Equilibrium (NE), is strongly dependent on the network parameters set by the AP (for example, available bandwidth and number of AP antennas). Therefore, the optimal system design requires the AP to set those parameters in a manner that provokes the most desirable power allocation

(NE) from the MSs [15]. This framework where one agent (set of MSs) acts subject to the strategy that the other agent (AP) chose (with the latter aware that his action is observed), is referred to as a Stackelberg game [33]. Moreover, the corresponding optimal pair of system parameters and power allocation is referred to as a Stackelberg Equilibrium (SE). A related work can be found in [15], where the provider (AP) acts as a Stackelberg leader whose goal is to encourage the cooperative transmission between terminals (follower), by optimizing the service prices and possible reimbursements.

Two network models are investigated here. The first assumes that the MSs' actions are dictated by the transmission power minimization under minimum capacity (transmission rate) constraints, while the second model is concerned with maximizing the power efficiency of the MSs. The service provider (AP) is consumer-oriented, and it aims at maximizing the users' preferences, while saving on investments such as bandwidth and network infrastructure (namely, AP antennas).

6.2 System Setup and Problem Def nition

6.2.1 System Setup

Consider a set $\mathcal{K}$ of K single-antenna MSs that are transmitting in the same time-frequency resource towards an AP with transmission powers P_i, $i = 1, .., K$, using asynchronous code-division access with processing gain $G \geq 1$. The set of all transmission powers is $\mathbf{P} = (P_1, P_2.., P_K)^T \in \mathcal{P}$, where $\mathcal{P}$ is the set of allowed MSs' powers, and the maximum transmission power per user is denoted as $P_{\max}$. The AP is equipped with N (receiving) antennas, and the independent identically distributed (iid) complex Gaussian channel gains between ith MS and jth AP antenna are denoted as h_{ij}. Using a vector notation, the set of channels between user i and N antennas is $\mathbf{h}_i = (h_{i1}, .., h_{iN})^T$, while the set of all channel gains is given by $N \times K$ matrix $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_K)$. The matched filtering (MF) at the AP is assumed with no interference cancellation. White Gaussian noise at any of the AP antennas is independent, with single-sided power spectral density N_0. Interference coming

from other users' signals is modelled as Gaussian noise. Assuming that the station are sending "Gaussian codewords" (Section 2.1.2) and, without loss of generality, that the used bandwidth is G Hz, the maximum achievable rate for the ith MS, C_i (in bit/sec), can be written as

$$C_i(\mathbf{P}, \mathbf{H}, N, G) = \log_2(1 + SINR_i), \tag{6.1}$$

where the Signal to Noise plus Interference Ratio for the ith MS, $SINR_i$, at the output of the MF is easily shown to be

$$SINR_i = \frac{P_i \left|\left|\mathbf{h}_i\right|\right|^2}{N_0 + \frac{1}{G}\sum_{\substack{k=1 \\ k \neq i}}^{K} \frac{\left|\mathbf{h}_i^H \mathbf{h}_k\right|^2}{\left|\left|\mathbf{h}_i\right|\right|^2} P_k}. \tag{6.2}$$

Notice that the dependence of the achievable rate C_i on the set of transmission powers $\mathbf{P}$, channel gain matrix $\mathbf{H}$ and the parameters set by the AP, N and G is emphasized in (6.1).

6.2.2 Problem Definition

There are two system entities to be distinguished, namely the set of MSs on one side and the AP on the other. The goal of the AP is the maximization of a long-term revenue (utility) function $U(N, G)$ that depends on both the network parameters (number of antennas N and processing gain G), that are under the direct control of the AP, and the behavior of the MSs that cannot be directly controlled by the AP. The revenue function $U(N, G)$ is defined as an average over the statistics of channel gains $\mathbf{H}$ in order to account for different (fading) scenarios.

The goal of each MS is to maximize its own (instantaneous) utility function $u_i(\mathbf{P}; N, G, \mathbf{H})$, $i = 1, .., K$, defined as to reflect MS's preferences, usually in terms of achievable transmission rate and/or consumed power. The degree of freedom of each MS, say ith, is its transmission power P_i, while the parameters N and G, and the channel matrix $\mathbf{H}$, are given. To emphasize this point, the notation $u_i(P_i, \mathbf{P}_{-i}; N, G, \mathbf{H})$ is used, where $\mathbf{P}_{-i}$ stands for the vector containing all but the ith element of $\mathbf{P}$ (i.e., it denotes the set of other

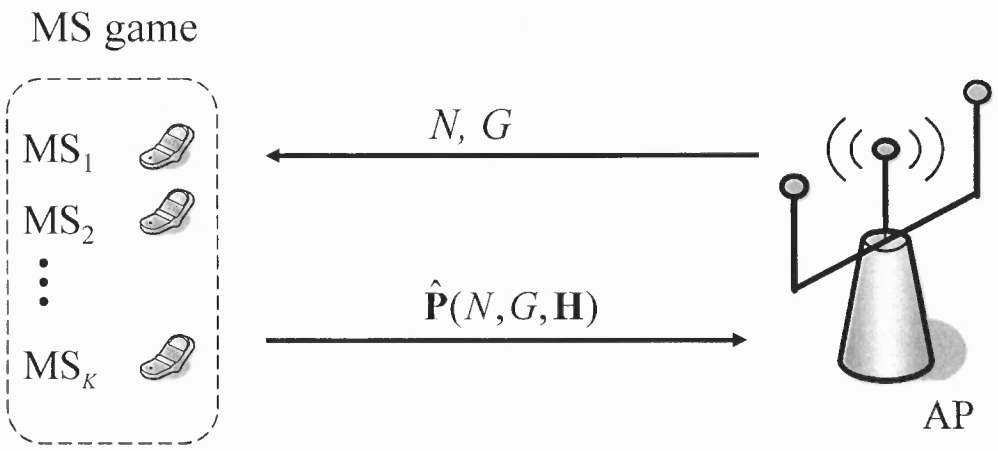

Figure 6.1 Overview of Stackelberg game between the AP and the MSs.

MSs' strategies). Furthermore, the MSs are independent and behave in a *selfish* and *rational* manner, with goals typically in direct conflict. The whole set of MSs can be presented as one entity that receives as input the network parameters set by the AP (N and G), and produces an output defined by a Nash Equilibrium (NE), $\hat{\mathbf{P}}\left(N,G,\mathbf{H}\right)=\left(\hat{P}_1,\hat{P}_2..,\hat{P}_K\right)^T$, of the non-cooperative game $\langle\mathcal{K},\mathcal{P},\{u_i\left(\cdot\right)\}\rangle$ played by MSs (see Figure 6.1).

The interaction between AP and the set of MSs described above can be studied in the framework of Stackelberg games. The AP represents the authority of the game (Stackelberg leader), playing the first move by setting the network parameters (N and G) towards the aim of increasing its revenue function $U(N,G)$. The MSs on the other side (Stackelberg follower) respond with the NE $\hat{\mathbf{P}}(N,G,\mathbf{H})$ of their non-cooperative game. In principle, this interchange of parameters and MS game outcomes continues until the Stackelberg Equilibrium (SE) is reached, i.e., until the AP finds the set of parameters (N and G) that, together with the corresponding NEs of the MS game, maximize its long-term (i.e., average over channel fading $\mathbf{H}$) revenue function $U(N,G)$.

6.3 Game Models

In the following, two game models are presented. In the first game, the MSs (follower) tackle the problem of minimizing the transmission power under minimum transmission rate constraint, while in the second they aim at the (unconstrained) maximization of power efficiency (bit/sec/W). For each game, the AP optimizes the network utility (in terms of collective MSs' preferences) per invested system resource, i.e., per antenna and bandwidth. Performance of the considered distributed models is assessed by comparison with the corresponding centralized scenarios.

6.3.1 Minimizing the Power under Capacity Constraints

MS Game For given network parameters N and G, the goal of the MS i is to minimize its own transmission power P_i under minimum transmission rate constraint, $C_{i,\min}$

$$\begin{aligned} &\text{minimize } P_i,\ i = 1, .., K, \\ &\text{subject to } C_i(\mathbf{P}, \mathbf{H}, N, G) \geq C_{i,\min}, \\ &\qquad P_i \in [0, P_{\max}], \end{aligned} \tag{6.3}$$

This problem can be formulated as the non-cooperative power control game (NPG) $\langle \mathcal{K}, \mathcal{P}, \{u_i(P_i, \mathbf{P}_{-i})\} \rangle$, where it is recalled that $\mathcal{K} = \{1, 2, .., K\}$ denotes the set of K players (MSs), the players' set of strategies $\mathcal{P}$ reads

$$\mathcal{P} = \{\mathbf{P} | P_i \in [0, P_{\max}],\ C_i(\mathbf{P}, \mathbf{H}, N, G) \geq C_{i,\min}, \forall i \in \mathcal{K}\}, \tag{6.4}$$

and the ith player's utility function is defined as

$$u_i(P_i, \mathbf{P}_{-i}) = -P_i,\ i \in \mathcal{K}. \tag{6.5}$$

Notice that the strategy sets for different users are coupled according to (6.4). Furthermore, the parameters set by the AP, i.e., N and G, and the channel gains $\mathbf{H}$, influence the game through its constraints and not through its utility $u_i(P_i, \mathbf{P}_{-i})$. To conclude on the game

setup, it is noticed that in a game theory framework a strictly concave utility function is preferred, so (6.5) is equivalently replaced with

$$u_i\left(P_i, \mathbf{P}_{-i}\right) = -\log_2 P_i,\ i \in \mathcal{K}, \tag{6.6}$$

where the base 2 of the log function is chosen purely for the sake of consistency with the definition of capacity (6.1).

Analysis of the game, namely the assessment, existence and uniqueness of NEs, is significantly simplified for the class of potential games [34], [68]. For a strategic game, say $\langle \mathcal{K}', \mathcal{P}', \{u_i'(\cdot)\}\rangle$, to be a potential game, there needs to exist a (potential) function $u_P' : \mathcal{P}' \rightarrow \mathbb{R}$ such that for all $i \in \mathcal{K}'$ and $(P_i', \mathbf{P}'_{-i})$, $(P_i'', \mathbf{P}'_{-i}) \in \mathcal{P}'$, it satisfies either $u_i'(P_i', \mathbf{P}'_{-i}) - u_i'(P_i'', \mathbf{P}'_{-i}) = u_P'(P_i', \mathbf{P}'_{-i}) - u_P'(P_i'', \mathbf{P}'_{-i})$, in which case it is called an *exact* potential game; or $u_i'(P_i', \mathbf{P}'_{-i}) - u_i'(P_i'', \mathbf{P}'_{-i}) > 0 \Leftrightarrow u_P'(P_i', \mathbf{P}'_{-i}) - u_P'(P_i'', \mathbf{P}'_{-i}) > 0$, in which case it is an *ordinal* potential game. The function $u_P'(\cdot)$ is called a *potential function.* For the scenario at hand, the NPG $\langle \mathcal{K}, \mathcal{P}, \{u_i(\cdot)\}\rangle$, is easily shown to be an (exact) potential game, $\langle \mathcal{K}, \mathcal{P}, u_P\rangle$, with the following potential function

$$u_P(\mathbf{P}) = -\sum_{i=1}^{K} \log_2 P_i. \tag{6.7}$$

Assuming the optimization problem (6.3) is feasible, the set of strategies $\mathcal{P}$ is compact. Furthermore, $u_P(\mathbf{P})$ is a continuous and strictly concave function on the interior of $\mathcal{P}$. It follows that a strategy $\mathbf{P}_{\text{opt}}$ that maximizes the potential $u_P(\mathbf{P})$, $\mathbf{P}_{\text{opt}} = \arg\max_{\mathbf{P}} U(\mathbf{P})$, is also a NE of the NPG $\langle \mathcal{K}, \mathcal{P}, \{u_i(\cdot)\}\rangle$ [68]. Furthermore, since the set $\mathcal{P}$ is also convex (in fact, it is a cone), following [68] the optimal $\mathbf{P}_{\text{opt}}$, and therefore the NE, $\hat{\mathbf{P}}(N, G, \mathbf{H}) = \mathbf{P}_{\text{opt}}$, is unique.

Both Gauss-Seidel and Jacobi algorithms, implementing best response, better response or the gradient projection rule, are guaranteed to reach the NE of the potential game at hand [68], [67]- [69]. Here the Gauss-Seidel algorithm is elaborated upon with the best response rule. The MSs play sequentially, and at the $(t+1)$th iteration the ith MS

updates its transmission power following:

$$P_i^{t+1} = \min\left(P_i^*, P_{\max}\right), \tag{6.8}$$

where P_i^* is the minimum power satisfying the constraint $C_i = C_{i,\min}$ (recall (6.1) and (6.2)):

$$P_i^* = \frac{\left(2^{C_{i,\min}} - 1\right)}{\left\|\mathbf{h}_i\right\|^2}\left(N_0 + \right. \tag{6.9}$$
$$\left. + \frac{1}{G}\sum_{k=1}^{i-1}\frac{\left|\mathbf{h}_i^H\mathbf{h}_k\right|^2}{\left\|\mathbf{h}_i\right\|^2}P_k^{t+1} + \frac{1}{G}\sum_{k=i+1}^{K}\frac{\left|\mathbf{h}_i^H\mathbf{h}_k\right|^2}{\left\|\mathbf{h}_i\right\|^2}P_k^t\right).$$

The converging point of the algorithm is the NE strategy set $\hat{\mathbf{P}}(N, G, \mathbf{H})$, where $\hat{\mathbf{P}}(N, G, \mathbf{H}) = \mathbf{P}_{\text{opt}}$.

AP Revenue Function The revenue function accounts for the preferences of the service provider, e.g., profit (if it is charging the users for the service while investing in equipment) or quality of service (measured in SINR ratios, achievable rates, the probability of error, etc.). Here a service provider is assumed that, following the users' interest, strives to minimize the total power expenditure. However, it is also interested in reducing the cost of the two primary resources: number of antennas and bandwidth. The following revenue function that measures the overall average network utility per system resource is proposed:

$$U(N, G) = \frac{-\sum_{i=1}^{K} E_{\mathbf{H}}\left[\log_2\left(\hat{P}_i\left(N, G, \mathbf{H}\right)\right)\right]}{NG}. \tag{6.10}$$

The expectation $E_{\mathbf{H}}[\cdot]$ is taken with respect to fading, since decentralized power control by the MSs is operated according to the instantaneous channel realization, while the system optimization is based on (long-term) channel statistics. Note that the revenue function in (6.10) depends on the NE of the MS game $\hat{\mathbf{P}}_i$, which in turn is a function of parameters N and G, set by the AP.

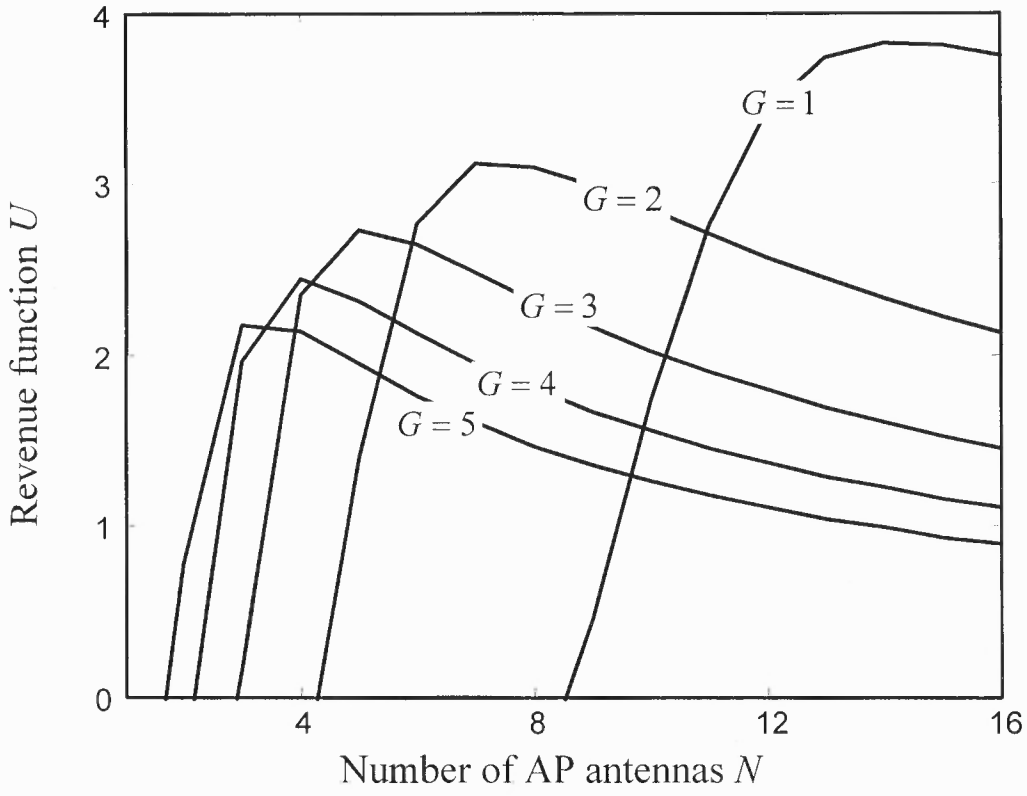

Figure 6.2 Revenue function of the AP, U, versus number of AP antennas N for different values of processing gain (bandwidth) G.

Centralized Scenario For reference, this subsection analyzes the case where the AP is able to control optimally not only the network parameters N and G, but also the MS transmission power, $\mathbf{P}(N, G, \mathbf{H})$, toward the goal of maximizing (6.10) (where the NE $\hat{\mathbf{P}}(N, G, \mathbf{H})$ is substituted with the variable $\mathbf{P}(N, G, \mathbf{H})$). From the discussion above, the decentralized solution of the power control (NE) for given N and G is the one that maximizes the potential (6.7). Comparing (6.7) with (6.10), it is easy to see that decentralized and centralized solution coincide in this case (see also [68]). Section 6.3.2 will discuss a scenario where this does not hold true.

System Performance The results in this section are obtained for the following parameters: $E\left[|h_{ij}|^2\right] = 1$, $C_{i,\min} = 1$ bit/sec, $P_{\max} = 2$ W and the average Signal

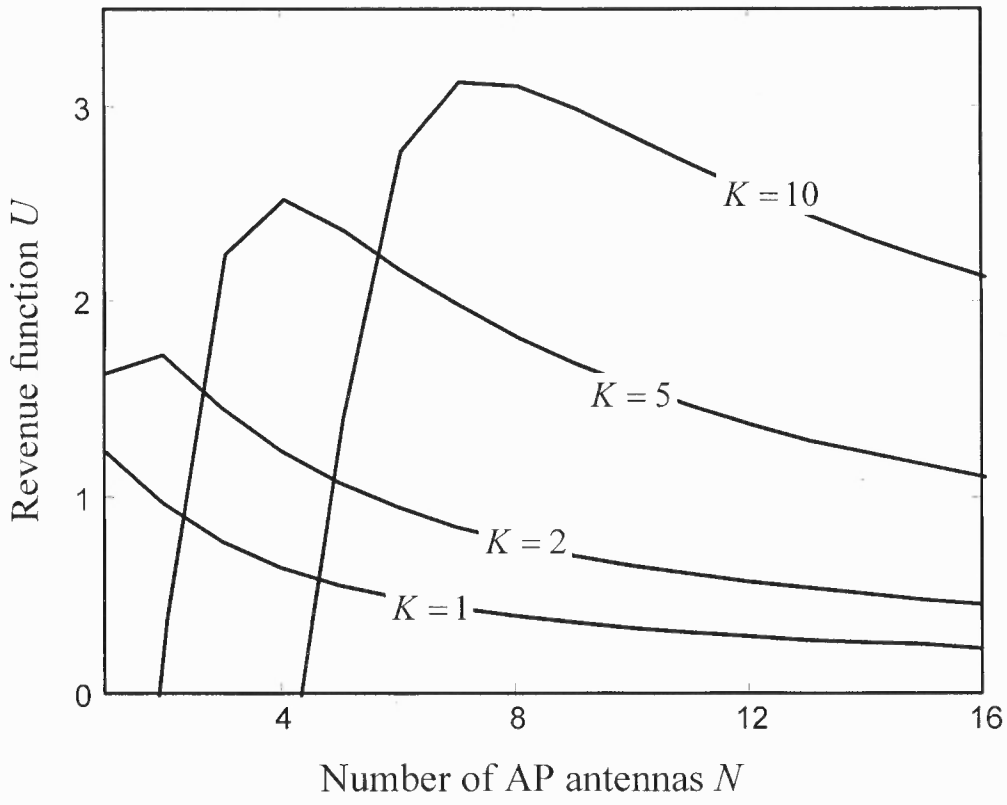

Figure 6.3 Revenue function of the AP, U, versus number of AP antennas N for different number of users K.

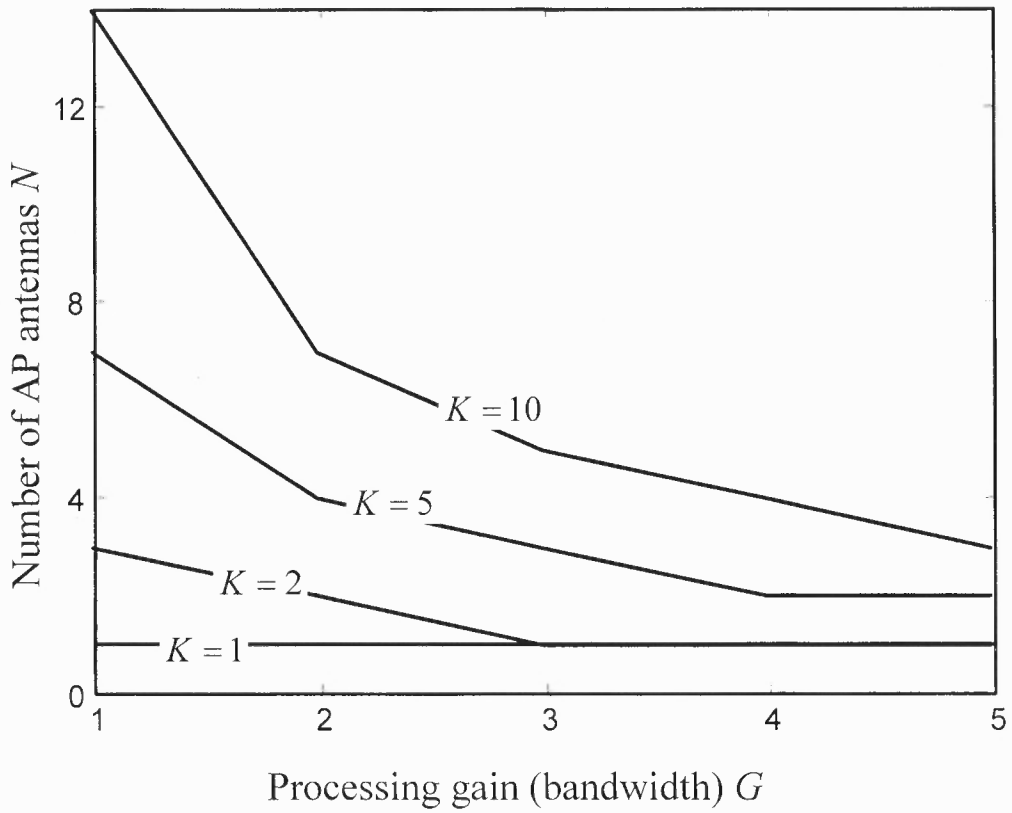

Figure 6.4 Stackelberg Equilibrium: dependence between number of antennas N and processing gain (bandwidth) G, with one parameter fixed and another optimally chosen by the AP, for different number of users K.

to Noise Ratio (defined as $SINR$ for $N = 1$, $K = 1$ and $P = P_{\max}$) is $SNR = 13$ dB. Figure 6.2 shows the revenue function $U(G, N)$ in (6.10) versus the number of antennas N for $K = 10$ MSs and different values of processing gain G. It can be observed that, for fixed G, the revenue increases with N up to a certain (optimal) point, after which the collective MS utility (6.7) (i.e., the numerator in (6.10)) becomes less then linearly proportional to N. In other words, U has a unique maximum, that is a SE, over N for fixed G. Moreover, increasing the processing gain G decreases the optimal value of N. While the reverse also holds, i.e., there is a unique maximum of revenue over G for fixed N, it is interesting to note that investing in antennas N has better effect on revenue function than buying more bandwidth (increasing G). The reason behind this can be explained by pointing out that the number of antennas has a two-fold effect on the SINR (6.2), i.e., power gain (in the numerator) and interference mitigation (in the denominator); on the other side, G results in interference mitigation only.

The revenue function versus the number of antennas N for $G = 2$ and different number of users K is presented in Figure 6.3. It is interesting to see that a larger number of users, though increasing the required network resources (i.e., antennas and, not shown here, bandwidth) at the optimal points, also increases the network revenue and is therefore desirable from the collective point of view. However, for large values of K, allowing additional users into the system has a negligible effect on the relative increase of the revenue function $U(G, N)$.

The optimal network parameter N (or G) set by the AP in the SE for fixed G (or N), is presented in Figure 6.4, for different number of users, K. The well known trade-off between bandwidth and spectral dimension is confirmed. Moreover, it is confirmed that increasing the number of user requires more resources.

6.3.2 Maximizing the Power Eff ciency

MS Game Instead of minimizing the power under the minimum transmission rate constraint as in Section 6.3.1, here the MSs' preference is the maximization of power efficiency:

$$\begin{aligned} &\text{maximize } \tfrac{C_i(\mathbf{P},\mathbf{H},N,G)}{P_i},\ i=1,..,K, \\ &\text{subject to } P_i \in [0, P_{\max}]. \end{aligned} \tag{6.11}$$

Under the assumption of selfish and rational MSs, problem (6.11) can be cast as a non-cooperative power control game (NPG) $\langle \mathcal{K}, \mathcal{P}, \{u_i(\cdot)\}\rangle$, where $\mathcal{K} = \{1, 2, .., K\}$ denotes the set of K players (MSs), the players' set of strategies $\mathcal{P}$ reads

$$\mathcal{P} = \{\mathbf{P} \mid P_i \in [0, P_{\max}],\ \forall i \in \mathcal{K}\},$$

and the ith player's utility function is defined as

$$u_i(P_i, \mathbf{P}_{-i}; N, G, \mathbf{H}) = \frac{C_i(\mathbf{P}, \mathbf{H}, N, G)}{P_i},\ i \in \mathcal{K}. \tag{6.12}$$

While this utility function strongly reflects the pragmatic preferences of the MSs, it needs a slight modification in order to avoid singularity at $P_i = 0$, while preserving quasi-concavity on $\mathcal{P}$:

$$u_i(P_i, \mathbf{P}_{-i}; N, G, \mathbf{H}) = \frac{C_i(\mathbf{P}, \mathbf{H}, N, G)}{P_i + P_c},\ i \in \mathcal{K}, \tag{6.13}$$

where P_c could be any conveniently chosen constant (for instance, it could account for the power consumed by electronic circuitry of MS [13]). Notice that the utility defined in (6.13) depends on AP parameters N and G, as well as the channel gains $\mathbf{H}$. A NPG with utility function as the one defined in (6.13) was investigated in [70].

In order to reach the NE, one can use the Jacobi algorithm, where all the users update their strategy in a parallel fashion using the Newton's method:

$$\mathbf{P}^{t+1} = \mathbf{P}^t + \alpha \left(d_1^t \frac{\partial u_1}{\partial P_1^t}, .., d_2^t \frac{\partial u_K}{\partial P_K^t} \right)^T, \tag{6.14}$$

where α is some conveniently chosen small number and d_i^t is chosen as $d_i^t = \left(\frac{\partial^2 u_i}{\left(\partial P_i^t\right)^2}\right)^{-1}$ [67]. The convergence point of the algorithm is the NE of the game, $\hat{\mathbf{P}}(N, G, \mathbf{H})$.

AP Revenue Function As in Section 6.3.1, the AP has preferences compatible with the MSs. Therefore, it aims at maximizing the (overall) power efficiency, averaged over fading, while accounting for the resource expenditure:

$$U(N, G) = \frac{1}{GN} \sum_{i=1}^{K} E_{\mathbf{H}} \left[\frac{C_i \left(\hat{\mathbf{P}}_i (N, G, \mathbf{H}), N, G, \mathbf{H} \right)}{\hat{P}_i + P_c} \right]. \tag{6.15}$$

Centralized Scenario For the centrally optimal solution, the problem boils down to maximizing the revenue function (6.15), by assuming that the AP can also control the set of the MSs' powers $\mathbf{P}(N, G, \mathbf{H})$. Therefore, the maximization is carried out with respect to G, N and $\mathbf{P}(N, G, \mathbf{H})$. This task can be performed numerically. As shown below, in this case the decentralized solution has degraded performance as compared to the centralized scenario.

System Performance Figure 6.5 shows the revenue function $U(G, N)$ versus the number of antennas N for different values of processing gain G, and parameters $E\left[|h_{ij}|^2\right] = 1$, $P_c = 0.1$ W, $P_{\max} = 2$ W and $SNR = 13$ dB. The conclusions are very similar to those for the power minimization problem. Furthermore, the dependence among N, G and K for the optimal (SE) solution is shown in Figure 6.6, revealing the similar system behavior to that of Figure 6.4.

Figure 6.7 shows the optimal revenue function U versus number of antennas N, for different number of users K and for both the distributed and centralized scenarios. As expected, centralized control allows to harness a larger revenue. However, as the number of antennas increase, the difference in performance between centralized and decentralized scheme reduces. This shows that with enough interference mitigation options, decentralized power control is not as harmful for the system performance.

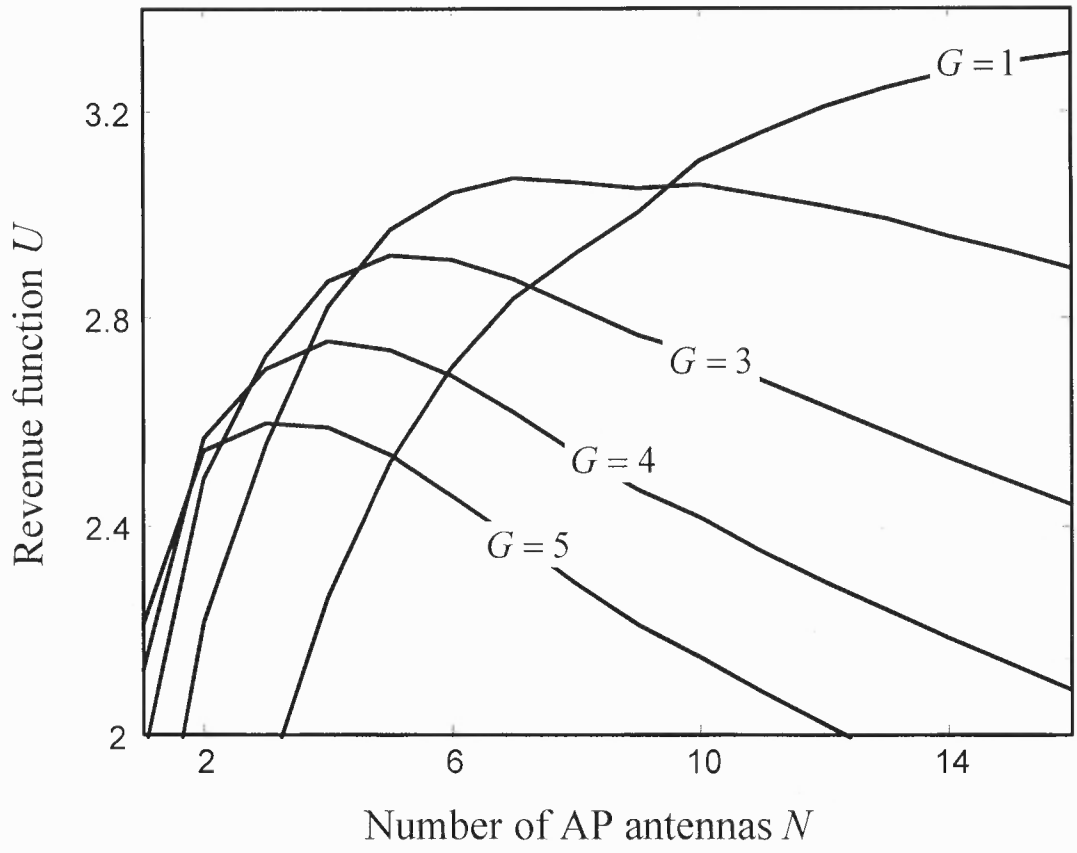

Figure 6.5 Revenue function U of the AP versus number of AP antennas N for different values of processing gain (bandwidth) G.

Moreover, it is clear from Figure 6.7 that, by increasing the number of users, the efficiency of the distributed scheme falls behind that of the optimal (centralized) scenario, thus confirming that large distributed systems pose the major challenge. Furthermore, it is very interesting to observe that, while the increased number of users is again desirable for the network (at least in centralized scenario), the relevant lack of efficiency for large K can diminish this gain in decentralized scenario.

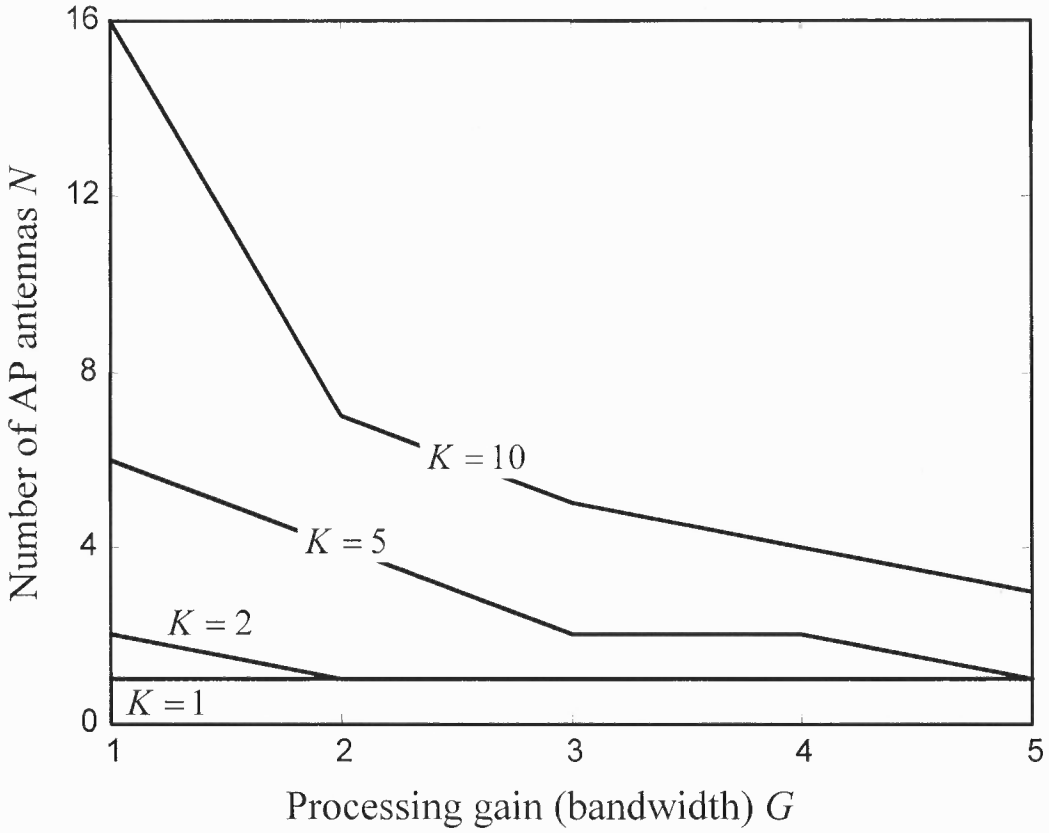

Figure 6.6 Stackelberg Equilibrium: dependence between number of antennas N and processing gain (bandwidth) G, with one parameter fixed and another optimally chosen by the AP, for different number of users K.

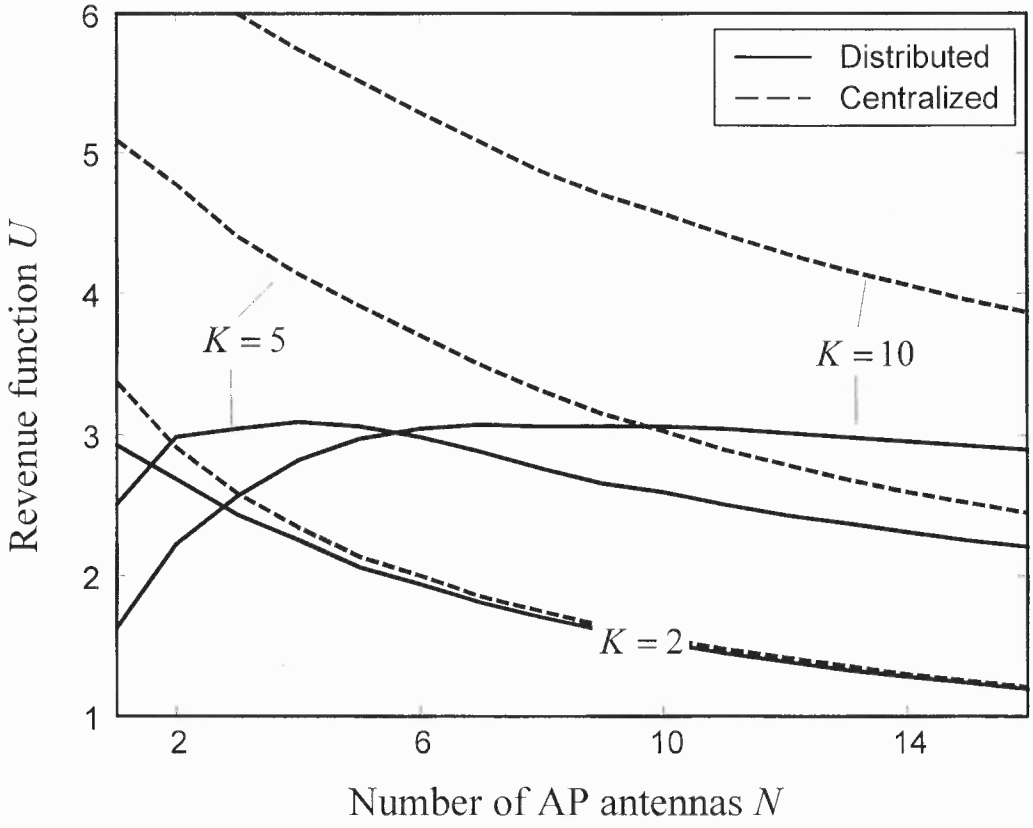

Figure 6.7 Revenue function U of the AP versus number of antennas N for different number of users K: comparison between centralized and distributed scenarios.

6.4 Chapter Conclusion

This chapter elaborated on two major concepts, game-theoretic Stackelberg game and communications power control game, that are extensively used in Chapter 7 for motivating the cooperation and dynamic spectrum access. In particular, the design of a multi-antenna access point was analyzed with decentralized power control in the uplink channel. The optimal solution, in terms of number of antennas and bandwidth, has been studied by modelling the problem as a Stackelberg game between the access point and competitive mobile stations. In this framework, it has been shown that a larger number of users motivates the provider (i.e., access point) to invest, as the overall performance enhancement well balances the costs. It was discussed, however, that in certain decentralized scenarios the system cannot efficiently cope with large amount of user. Furthermore, the well-known trade-off between system resources, bandwidth and antennas, was confirmed.

CHAPTER 7

DYNAMIC SPECTRUM ACCESS AND RELAYING MOTIVATION IN DECENTRALIZED NETWORKS VIA COOPERATION

In this chapter, the scheme for motivating the otherwise non-cooperative nodes in distributed networks to cooperate is provided. The scheme relies on two concepts introduced in Chapter 6, namely the Stackelberg game for modelling the nodes' decision processes and power control game for modelling the distributed communications environment. It will be shown that, by providing incentives for relaying, the proposed mechanism significantly increases the spectrum efficiency and provides the spectrum access for all involved nodes. As detailed later, the scheme can be also considered in the framework of property-rights cognitive radio, i.e., dynamic spectrum access.

7.1 Background

In distributed networks, the terminals are typically designed to behave in a selfish manner, which implies that willingness to cooperate cannot be assumed a priori. An approach is proposed that increases the efficiency of spectrum utilization by inherently providing incentives for the relays' assistance based on the spectrum leasing principle. Specifically, incentive for (otherwise non-cooperative) relaying station is given by the opportunity to exploit part of the retransmission slot for their data. In this chapter, the focus is on the baseline solution relying on the global system information at the source terminal, while the practical and fully decentralized scheme that exploits Cooperative retransmission is introduced in Chapter 8.

The main idea of the proposed spectrum leasing concept is as follows. On one hand, the goal of the source is the maximization of its own achievable rate towards the destination, by optimizing the amount of resources (fraction of time), if any, leased to

the relaying network and the amount of required (space-time coded) cooperation. On the other hand, different competing transmitters in the relaying network seek to increase their achievable rates towards the intended destinations under the constraints imposed by the resources leased by the source and by the overall cost of transmission power (including the power spent for cooperation). Moreover, they accept to cooperate with the source only if retributed with a large enough fraction of time. Given the competitive nature of relaying transmitters, the outcome of their interaction can be conveniently described by a non-cooperative power control game [61] [71] and, more specifically, by the corresponding Nash Equilibrium (NE) [72] (Section 2.2).

An appropriate analytical framework to study the spectrum leasing scenario at hand is that of Stackelberg games [72], which was already applied in Chapter 6. In such a hierarchical game model, one agent (the competitive relaying network) acts subject to the strategy chosen by the other agent (source), which in turns seeks maximization of its own utility (here the achievable rate). Source's strategy that yields the optimal solution and the corresponding power/cooperation response of the relaying network are jointly referred to as a Stackelberg equilibrium. The concept of a Stackelberg equilibrium can be further exploited to predefine a set of rules to be imposed on the players that would result in the most desirable interaction outcome.

It is further noted that the proposed solution can be seen as a practical framework for the implementation of cognitive radio networks running according to the property-rights model (spectrum leasing) [23]. In such networks (Section 1.3), primary (licensed) users may lease portions of the licensed spectrum to secondary (unlicensed) users in exchange for some form of retribution. Here, the role of the primary node is played by the original source and that of the secondary by the relaying nodes. Moreover, retribution from secondary to primary nodes is in the form of cooperation to the primary transmission. This enables on-the-air decisions and avoids the regulatory issues or money transactions that commonly hinder the implementation of the spectrum leasing concept.

Finally, several reputation- and/or credit (pricing) - based approaches have been adopted in order to stimulate cooperation among terminals [15]- [17]. These schemes are not opportunistic and generally require a long operational time horizon in order to enforce cooperation.

7.2 System Model

In the following, the details for the proposed game-theoretic model of spectrum leasing are given and the main system parameters are described.

7.2.1 Medium Access Control (MAC) Layer

Consider the system sketched in Figure 7.1, where a source S communicates with the intended destination D within a slot whose duration is normalized to one. In the same bandwidth, a set of possible relaying terminals (here often referred to as a relaying network) $\mathcal{R}$, composed of K transmitters $\{\mathrm{R}_i\}_{i=1}^{K}$ and K receivers $\{\mathrm{R}_{\mathrm{x}i}\}_{i=1}^{K}$, is active as well, seeking to exploit possible transmission opportunities. One-to-one communication in $\mathcal{R}$ is assumed, i.e., the data from the relay R_i is intended for the receiver $\mathrm{R}_{\mathrm{x}i}$ (interference channel). Furthermore, without loss of generality, the transmitting relays are sorted in descending order relative to the instantaneous channel power gain from S.

The source S is assumed to be able to grant the use of the bandwidth to a subset $\mathcal{R}(k) \subseteq \mathcal{R}$ of k transmitting relays in exchange for space-time coded cooperation so as to improve the quality of the communication link to its receiver D. In particular, if the source can benefit from cooperation (i.e., if it can achieve a larger rate than via direct transmission to the receiver D), then it performs transmission as shown in Figure 7.1-(a). A fraction of the slot dedicated to its transmission towards the relaying set $\mathcal{R}(k)$ is of duration $1-\alpha$ ($0 \leq \alpha < 1$). Selection of the nodes in $\mathcal{R}(k)$ doesn't require further signalling but is obtained automatically via rate adaptation. Namely, only the terminals R_i whose channels

from S are sufficiently good to support S' rate are activated (R_1 and R_2 in example depicted by Figure 7.1-(a)).

The remaining time α is decomposed into two subslots according to a parameter $0 \leq \beta \leq 1$. In the first subslot of duration $\alpha(1-\beta)$, the k active relays $R_i \in \mathcal{R}(k)$ are allowed to transmit their own data (Figure 7.1-(b)), and the transmissions scheme amounts to an interference channel [61] [71]. The last subslot is of duration $\alpha\beta$ and is used for cooperation: the set $\mathcal{R}(k)$ of active R_i form a distributed k-antenna array and cooperatively relay the source's codeword (decoded during the first subslot of duration $1-\alpha$) through distributed space-time coding towards D [48] (Figure 7.1-(c)).

7.2.2 Physical Layer

The channels between nodes are modeled as independent complex Gaussian random variables, invariant within each slot (Rayleigh ergodic block-fading channels). The following notation is used: h_S denotes the complex channel gain between source S and destination D; $h_{SR,i}$ the channel gain between S and relaying transmitter R_i; $h_{RD,i}$ between R_i and D; $h_{R,ij}$ between R_j and R_{xi} for any $i,j = 1,...,K$. Without loss of generality, relaying nodes are sorted according to their channels from S, i.e., $|h_{SR,1}|^2 \geq |h_{SR,2}|^2 \geq ... \geq |h_{SR,K}|^2$, so that, according to the discussion above, $\mathcal{R}(k) = \{1,2,..,k\}$. All the receivers have a perfect knowledge of the relevant channels, i.e., the R_i and D know the exact values of $h_{SR,i}$ and $h_{RD,i}$, respectively, for $i = 1,...,K$. Furthermore, the source is assumed to be aware of all the instantaneous channel power gains in the system (i.e., $|h_S|^2$, $|h_{SR,i}|^2$, $|h_{RD,i}|^2$ and $|h_{R,ij}|^2$), while the knowledge of the channel power gains $|h_{R,ij}|^2$ within the relaying network is required at the relaying terminals. Albeit ideal, the assumption of instantaneous Channel State Information (CSI) (this does not refer to the CSI at the receivers, as it can be easily facilitated using the training sequences) is very common in the literature on game-theoretic applications to wireless networks (see, e.g., [61] [71]) and provides an interesting framework for analysis. A scenario with CSI knowledge limited

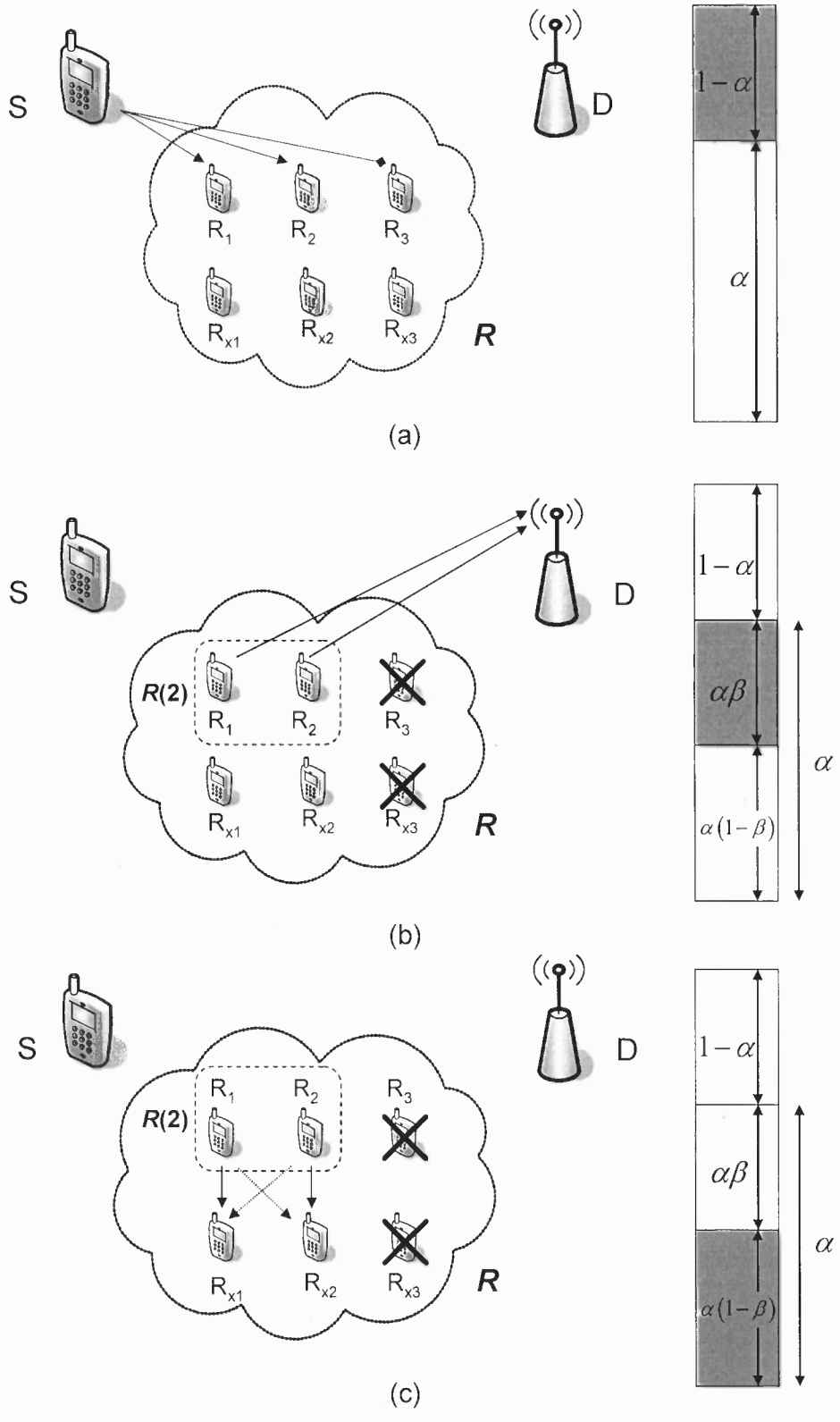

Figure 7.1 Cooperation-based spectrum leasing, for $K = 3$ relaying transmitters and receivers: (a) source's transmission; (b) space-time coded cooperation; (c) relays' own transmission.

only to the channel statistics is outside the scope of this work, and the related analysis and discussion can be found in [73], where *randomized* distributed space-time coding is considered. As for the synchronization issue on distributed space-time coding, the reader is referred to, e.g., [74].

The transmission power of the source is denoted as P_{S}. On the other hand, relays' transmit powers $\mathbf{P} = [P_1,..., P_k]^T$, $0 \leq P_i \leq P_{\max}$ are obtained as the outcome $\hat{\mathbf{P}}$ (NE) of the power control game played between relaying nodes in the subslot of duration $\alpha(1-\beta)$ (Figure 7.1-(b)), as detailed in Section 7.3.2. During the cooperating subslot (of duration $\alpha\beta$, Figure 7.1-(c)), the set of k activated relays $\mathcal{R}(k)$ is constrained to use the same powers $\hat{\mathbf{P}}$ that are the outcome of the power control game in the preceding subslot. Possible malicious behavior of the relays in the cooperative phase (i.e., using the power $P_i < \hat{P}_i$ or even refusing the cooperation, $P_i = 0$) is out of the scope. Finally, the single-sided spectral density of the independent white Gaussian noise at any of the receivers is N_0.

7.3 Game-Theoretic Analysis

In this section, the behavior of source and relaying network is described and analyzed. Discussion is provided for their interaction within a Stackelberg game framework.

7.3.1 Source

The source selects the slot allocation parameters (α, β) and the set of cooperating relaying nodes $\mathcal{R}(k)$ towards the aim of optimizing its transmission rate $C_{\mathrm{S}}(\alpha, \beta, k)$. As explained in Section 7.2.1, the set $\mathcal{R}(k)$ is selected as $\mathcal{R}(k) = \{\mathrm{R}_i | i = 1, .., k\}$, in order to simplify signalling (notice that on the negative side, this choice requires every relaying terminal to attempt decoding). Assuming decode-and-forward space-time coded cooperation from the set $\mathcal{R}(k)$ of k active relaying users [48], the achievable rate reads

$$C_{\mathrm{S}}(\alpha, \beta, k) = \begin{cases} \min\left\{(1-\alpha)\, R_{\mathrm{SR}}(k), \alpha\beta R_{\mathrm{RD}}(k, \beta)\right\}, & \alpha > 0 \\ R_{\mathrm{dir}}, & \alpha = 0 \end{cases} . \tag{7.1}$$

The first line in (7.1) stands for the rate that is the outcome of cooperation (recall that $\alpha > 0$ is the fraction of time slot dedicated for the relays' activity). It is the minimum between two terms: (*i*) the rate achievable in the first subslot (Figure 7.1-(a)) between the source S and the relaying transmitter R_k (recall that, due to ordering, R_k has the worst channel from S within the set $\mathcal{R}(k)$), which is easily shown to be $(1-\alpha)C_{\mathrm{SR}}(k)$, where (assuming random Gaussian codebooks)

$$C_{\mathrm{SR}}(k) = \log_2\left(1 + \frac{|h_{\mathrm{SR},k}|^2 P_{\mathrm{S}}}{N_0}\right); \tag{7.2}$$

(*ii*) the rate between the k active relaying transmitters $\mathcal{R}(k)$ and the destination D via space-time coding (subslot highlighted in Figure 7.1-(c)), $\alpha\beta C_{\mathrm{RD}}(k,\beta)$, with

$$C_{\mathrm{RD}}(k,\beta) = \log_2\left(1 + \sum_{i=1}^{k} \frac{|h_{\mathrm{RD},i}|^2 \hat{P}_i(k,\beta)}{N_0}\right), \tag{7.3}$$

where the dependence of the NE on the parameters (k,β) selected by the source is emphasized. Note that the rate (7.3) is obtained following the ideal information-theoretic assumption of orthogonal space-time coding able to harness the maximum degree of diversity from cooperation.

From (7.1), if the source decides not to employ the cooperation, i.e., $\alpha = 0$, then its rate is $C_{\mathrm{S}}(0,\beta,k) = C_{\mathrm{dir}}$, where

$$C_{\mathrm{dir}} = \log_2\left(1 + \frac{|h_{\mathrm{S}}|^2 P_{\mathrm{S}}}{N_0}\right) \tag{7.4}$$

is the rate achievable on the direct link between source S and destination D.

The source's optimization problem can now be summarized as

$$\begin{aligned} &\max_{\alpha,\beta,k} C_{\mathrm{S}}(\alpha,\beta,k) \\ &\text{s.t. } k \leq K,\ 0 \leq \alpha,\beta \leq 1. \end{aligned} \tag{7.5}$$

This problem can be interpreted as a Stackelberg game [72], whereby the source is the Stackelberg leader, that optimizes its strategy (α,β,k) in order to maximize its revenue

according to (7.5), aware that its decision will affect the strategy selected by the Stackelberg follower (the relaying set), i.e., the set of transmitting powers $\hat{P}_i\,(k,\beta)$.

7.3.2 Relaying Network

Any active relaying terminal R_i in the set $\mathcal{R}(k)$ attempts to maximize the rate towards its own receiver $\mathrm{R}_{\mathrm{x}i}$ (discounted by the overall cost of transmission power), acting in a rational and selfish way and being aware of the parameters (k,β) selected by the source. In particular, each relaying transmitter R_i chooses its transmitting power P_i according to the NE $\hat{P}_i\,(k,\beta)$ (it will be shown that it exists and is unique) of the non-cooperative power control game $\langle \mathcal{R}(k), \mathcal{P}(k), u_i\,(P_i,\mathbf{P}_{-i})\rangle$. The set of allowed (power) strategies $\mathcal{P}(k)$ reads

$$\mathcal{P}(k) = \left\{ \mathbf{P} = (P_1,..,P_k)^T \,|P_i \in [0, P_{\max}]\,,\ i = 1,..,k \right\}. \tag{7.6}$$

The utility function $u_i\,(P_i,\mathbf{P}_{-i})$ of the ith relaying node (player) is defined (similarly to, e.g., [75]) as the difference between the transmission rate $\alpha(1-\beta)C_i$ on the link between R_i and $\mathrm{R}_{\mathrm{x}i}$, where

$$C_i\,(P_i,\mathbf{P}_{-i}) = \log_2\left(1 + \frac{|h_{\mathrm{R},ii}|^2 P_i}{N_0 + \sum_{\substack{j=1\\ j\neq i}}^{k} |h_{\mathrm{R},ij}|^2 P_j}\right), \tag{7.7}$$

and the energy cost $c\cdot\alpha P_i$ (recall that α is the fraction of time where the active relaying nodes are transmitting), with c being the cost per unit transmission energy. Noticing from (7.7) that parameter α has no influence on the optimization process, utility simplifies to

$$u_i\,(P_i,\mathbf{P}_{-i}) = (1-\beta)C_i\,(P_i,\mathbf{P}_{-i}) - c\cdot P_i, \tag{7.8}$$

where $\mathbf{P}_{-i}$ is the vector that contains all the elements of $\mathbf{P}$ except the ith (i.e., it denotes the set of other players' strategies). Notice that the utility of each node (7.8) depends on k and parameter β, as well as on the power strategies of other activated users and the channel

realizations. However, the only degree of freedom, i.e., the strategy available to the ith relay for the optimization of (7.8), is its transmission power P_i.

As discussed in Section 2.2, a NE is a fixed point of the best responses of the nodes in $\mathcal{R}(k)$ [72]. Here, the best response of each user is obtained by setting the derivative of (7.8) with respect to P_i to zero, i.e., $\partial u_i(P_i,\mathbf{P}_{-i})/\partial P_i|_{\mathbf{P}=\hat{\mathbf{P}}} = 0$, for $i = 1, ..., k$. It is possible to show that the NE $\hat{\mathbf{P}}$ is the solution of the following set of k non-linear equations

$$\hat{P}_i = \left[\frac{1-\beta}{c} - \frac{N_0}{|h_{\mathrm{R},ii}|^2} - \sum_{j=1,j\neq i}^{k} \frac{|h_{\mathrm{R},ij}|^2}{|h_{\mathrm{R},ii}|^2}\hat{P}_j\right]_0^{P_{\max}}, \tag{7.9}$$

where the following notation is used

$$[x]_{x_\mathrm{m}}^{x_\mathrm{M}} = \begin{cases} x, \; x_\mathrm{m} \leq x \leq x_\mathrm{M} \\ x_\mathrm{m}, \; x < x_\mathrm{m} \\ x_\mathrm{M}, \; x > x_\mathrm{M} \end{cases}, \tag{7.10}$$

for any x, x_m, $x_M \in \mathcal{R}$. Therefore, the game has a unique NE if the system (7.9) has a unique solution. In particular, for given α and k, and $c = 0$, the game $\langle \mathcal{R}(k), \mathcal{P}(k), u_i\,(P_i,\mathbf{P}_{-i})\rangle$ has been discussed in the more general framework of wideband systems in [61] and [71], where it was shown that a NE exists and that it is unique if the matrix $\mathbf{H}$, defined as $[\mathbf{H}]_{ij} = |h_{\mathrm{S},ij}|^2$ is strictly diagonally dominant, i.e.,

$$\sum_{j=1,j\neq i}^{k} \frac{|h_{\mathrm{R},ij}|^2}{|h_{\mathrm{R},ii}|^2} < 1. \tag{7.11}$$

The condition (7.11) for uniqueness of the NE is intuitive since it simply imposes an upper bound on the interference: in fact, with negligible interference equations (7.9) become uncoupled and the solution clearly exists and is unique. In the following, it is assumed that (7.11) holds. Finally, notice in (7.9) that if β set by the source is too large, the result can be the denial of cooperation by the relays (by setting $\hat{P}_i = 0$).

7.3.3 Interaction between Source and Relaying Network

The interaction between the source and the relaying network is modeled as a Stackelberg game [72], whereby the source is considered as the game authority, i.e., the Stackelberg leader. The leader optimizes its strategy (α, β, k) in order to maximize its revenue (7.1), knowing that its decision will affect the strategy selected by the Stackelberg follower (the relaying network), namely the set of transmitting powers $\hat{\mathbf{P}}$. The latter is in fact determined by the NE of the relays' power control game described in the previous section. Maximization of the revenue of the source amounts to several trade-offs. For example, parameter β has two conflicting effects on cooperation: while increasing β entails more time for cooperation, it also renders cooperation from the relaying stations less likely since the cost induced by the transmitting power becomes dominant term in (7.8); furthermore, while a large value of k may limit the overall rate by reducing the term $(1-\alpha)C_{\mathrm{SR}}(k)$ in (7.1), at the same time it enhances the term $\alpha\beta C_{\mathrm{RD}}(k,\beta)$ in (7.1) thanks to cooperation.

Some analytical insight into the considered system is provided in the following. Since the parameter β appears only in the term $\alpha\beta C_{\mathrm{RD}}(k,\beta)$ of (7.1), it can be optimized independently by solving the following optimization problem (it can be proved that the optimization of (7.12) has a unique solution.)

$$\hat{\beta} = \arg \max_{\beta \in [0,1]} \beta C_{\mathrm{RD}}\left(k, \beta\right). \tag{7.12}$$

Moreover, for a given set $\mathcal{R}(k)$ and $\hat{\beta}$, the optimal fraction $\hat{\alpha}$ is given by making the two terms in the first line of (7.1) equal (so as to avoid performance bottlenecks), leading to:

$$\hat{\alpha} = \frac{1}{1 + \frac{\hat{\beta} C_{\mathrm{RD}}(k,\hat{\beta})}{C_{\mathrm{SR}}(k)}}, \tag{7.13}$$

and the optimized source rate (7.1) reads

$$C_{\mathrm{S}}(\hat{\alpha}, \hat{\beta}, k) = \frac{\hat{\beta} C_{\mathrm{RD}}(k,\hat{\beta}) C_{\mathrm{SR}}(k)}{C_{\mathrm{SR}}(k) + \hat{\beta} C_{\mathrm{RD}}(k,\hat{\beta})}. \tag{7.14}$$

Recall from Section 7.3.1 that the source decides to exploit the cooperation only if there exists some $k \leq K$ such that $C_{\mathrm{S}}(\hat{\alpha}, \hat{\beta}, k) > C_{\mathrm{dir}}$; otherwise it uses the direct link with achievable rate C_{dir}. Furthermore, it is noted that the optimization over the parameter β (7.12) and over the number of users k (as in (7.14)) requires numerical solving methods. The next section provides numerical results to corroborate these conclusions.

7.4 Numerical Results

In this section, a simple geometrical model is considered where the relays are all placed at approximately the same normalized distance $0 < d < 1$ from the source S and $1 - d$ from the destination D. Consequently, considering a path loss model, the average power gains of the channels read: $g_S = 1$, $g_{SR,i} = 1/d^{\gamma}$, and $g_{RD,i} = 1/(1-d)^{\gamma}$, where $\gamma = 2$ is the path loss coefficient. Moreover, in order to further reduce the number of system parameters and get better insight into the overall performance, it is set $g_{R,ij} = 1$ and $g_{R,ii} = g_R$ for $i, j = 1, ..., K$ and $i = j$. The source's power and the maximum relay's transmission power are $P_S = P_{max} = 1$, the signal-to-noise ratio is $SNR = P_S/N_0 = 0$dB, the cost per unit energy is $c = 0.1$ and, unless explicitly stated otherwise, the number of relaying transmitters is $K = 5$ and $g_R = 10$dB.

In keeping with the description of the optimization procedure given in the previous section, Figure 7.2 shows the optimal source rate $C_{\mathrm{S}}(\hat{\alpha}, \hat{\beta}, k)$ (7.14), averaged via Monte Carlo simulations over the Rayleigh fading realizations, for given subsets of relays $\mathcal{R}$ and versus the normalized distance d. It can be seen that for a relaying network placed at small distances it is better to activate (and thus cooperate with) a large number of relays given the large channel power gain from source to relaying network. Conversely, for large distances it is more convenient to cooperate only with the relays with the best instantaneous channel $|h_{SR,i}|^2$, exploiting multiuser diversity. The figure also compares the optimal source rate $C_{\mathrm{S},,2^K}$ achieved via exhaustive search over the 2^K subsets $\mathcal{R}$ (as if the source was *selecting* the relays) and the rate $R_{\mathrm{S}}(\hat{\alpha}, \hat{\beta}, k)$ obtained by restricting the search only to the K subsets

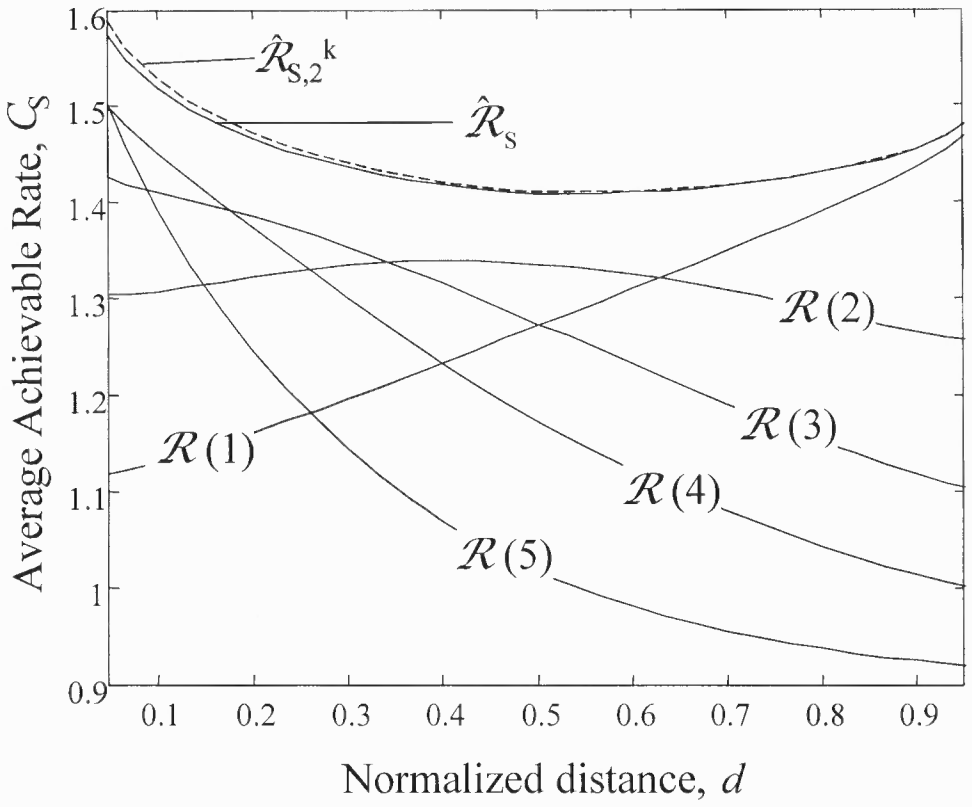

Figure 7.2 Source's rate $C_S(\hat{\alpha}, \hat{\beta}, k)$ averaged over fading, versus the normalized distance d, for given subsets of relays $\mathcal{R}(k)$, that contain the relaying transmitters with the k best power gains from the source. Also shown is the comparison between the rate $C_S(\hat{\alpha}, \hat{\beta}, k)$ and $C_{S,2^K}$ achieved via exhaustive search over the 2^K subsets $\mathcal{R}$ (as if the source was *selecting* the relays) ($P_S = P_{max} = 1$, $SNR = 0$dB, $K = 5$, $g_R = 10$dB)

$\mathcal{R}(k)$ with $k = 1, ..., K$. As it is clear, the reduction in order of complexity from 2^K to K entails almost no performance degradation.

Figure 7.3 shows the optimal parameters $\hat{\alpha}$ and $\hat{\beta}$, averaged over fading distribution, versus the normalized distance d for the same setting as in Figure 7.2. As it can be seen, the optimal value $\hat{\alpha}$ tends to decrease with distance d since, with increasing d, activation of relaying nodes becomes more demanding (i.e., it requires more time), leaving less time for the transmission of the relays. Moreover, it is seen that less interference between relaying channels (i.e., an increasing g_R) implies that the relays need to spend less power in order

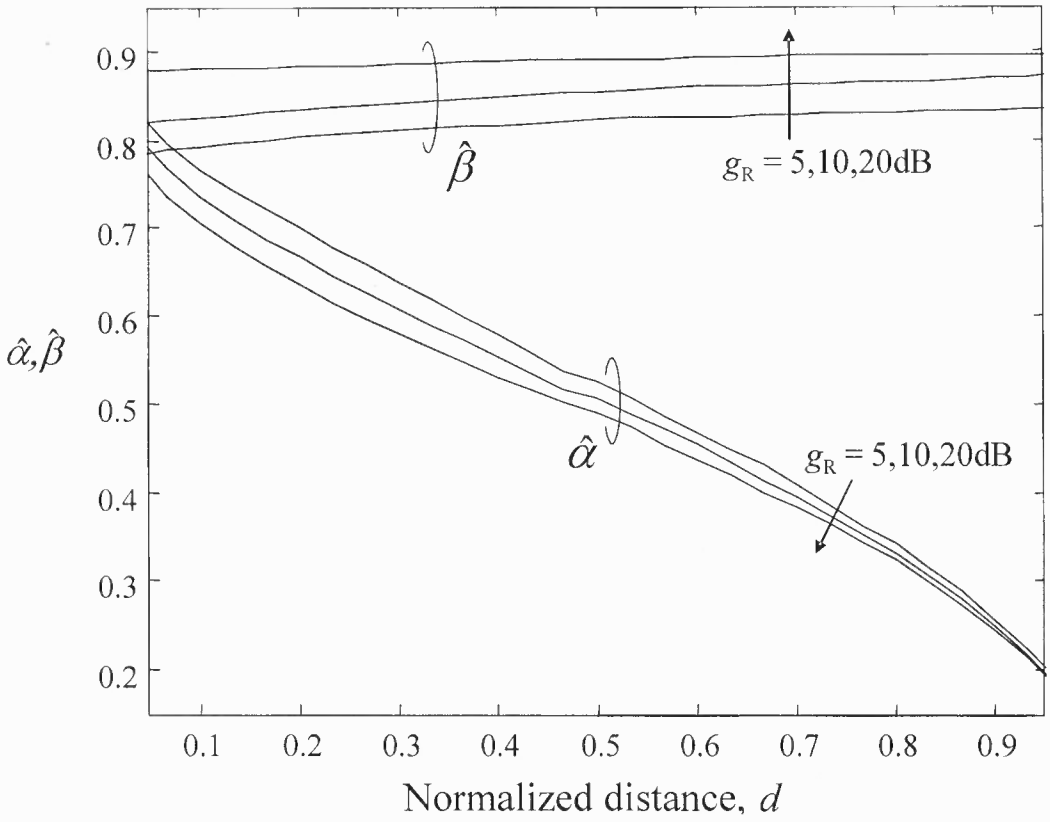

Figure 7.3 Optimal value of parameters α and β, averaged over fading, versus the normalized distance d ($P_S = P_{max} = 1$, $SNR = 0$dB, $K = 5$, $g_R = 10$dB).

to optimize their utility, and therefore are willing to use less power for cooperation, which leads to smaller leased times α (and larger β).

In order to get further insight into the system behavior, Figure 7.4 shows the cooperative rate $C_S(\alpha, \beta, k)$, averaged over different fading realizations, versus the normalized distance d, for α ranging from $1/8$ to $7/8$, with $\beta = 0.8$ and optimized $\mathcal{R}$. The rate on the direct link between S and D, C_{dir}, is also shown as a reference. As discussed above, as the distance d increases, the optimal $\hat{\alpha}$ decreases.

Finally, the average rate achieved by the (activated) relays, $\alpha(1-\beta)C_i$ is shown in Figure 7.5 as a function of the normalized distance d and the channel gain g_R. Since a larger distance entails a smaller optimal leased time α (see Figure 7.3), the relays'

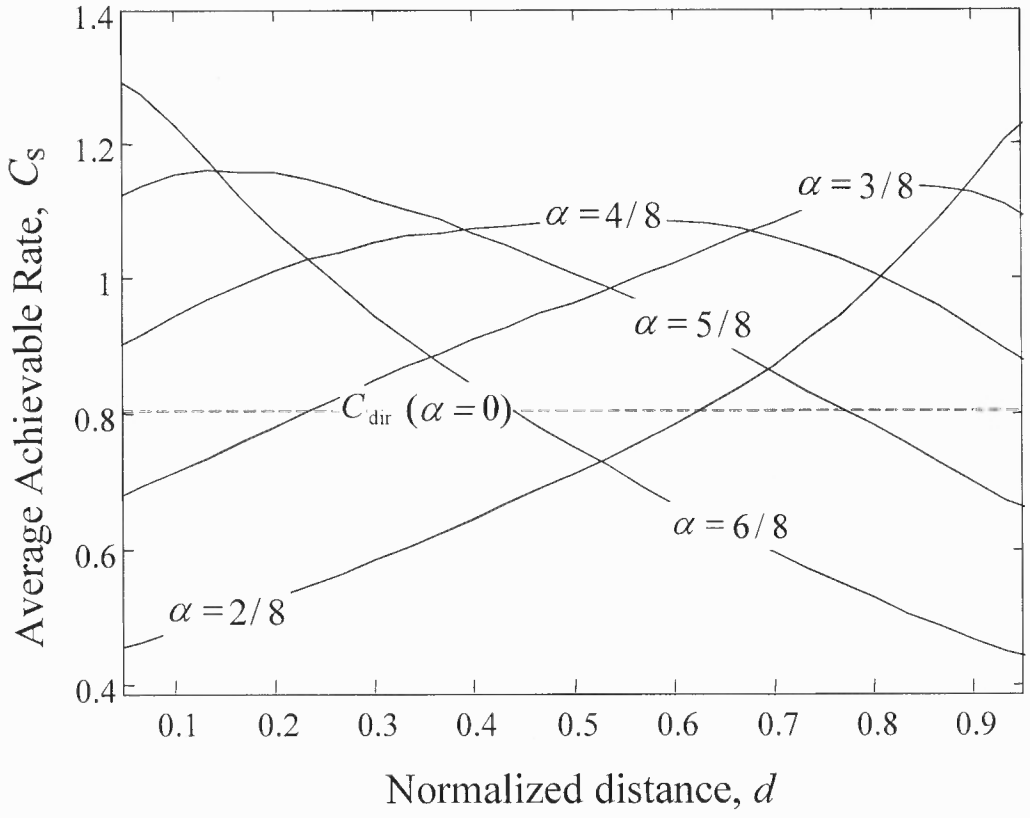

Figure 7.4 Source's rate $C_S(\alpha, \beta, k)$ averaged over fading, versus the normalized distance d between the source and the destination, for α ranging from $1/8$ to $7/8$, and $\beta = 0.8$. Dashed line refers to the rate achievable through direct transmission C_{dir} ($P_S = P_{max} = 1$, $SNR = 0$dB, $K = 5$, $g_R = 10$dB).

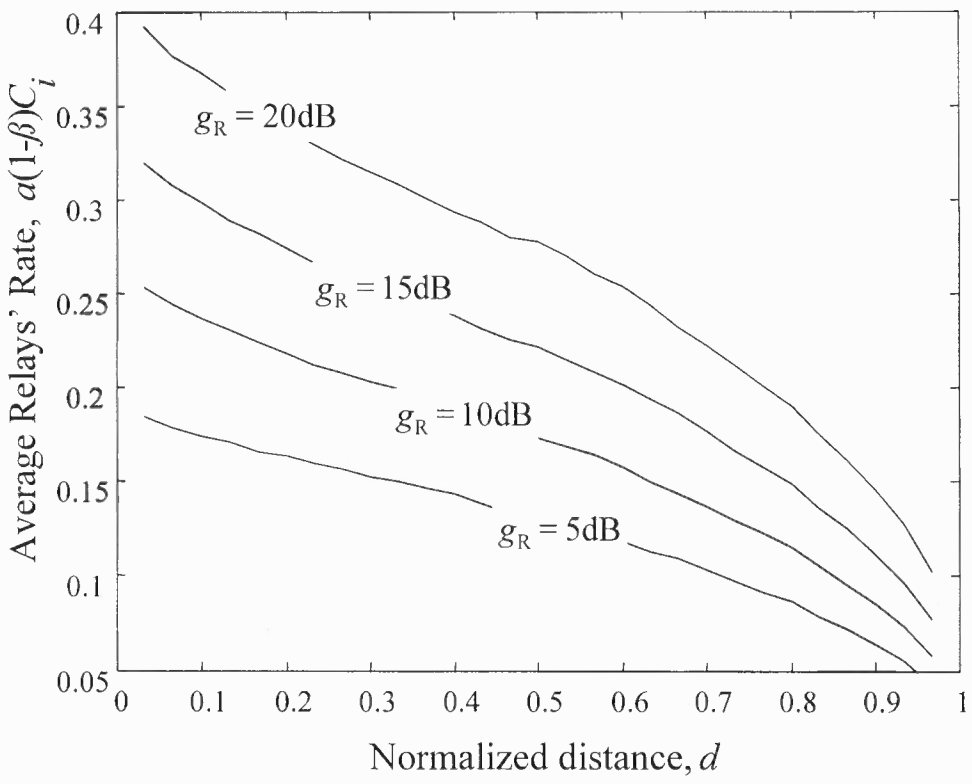

Figure 7.5 Relay's own rate $\alpha(1-\beta)C_i$, averaged over fading, versus the normalized distance d, in a symmetric scenario with different channel power gains g_R between the relaying transmitter-receiver pairs ($g_R = 5, 10, 15, 20$dB, $P_S = P_{max} = 1$, $SNR = 0$dB, $K = 5$).

rate decreases as the distance increases. Moreover, with increasing channel gain power g_R between relaying transmitter-receiver pairs (which entails better signal-to-interference ratios), as expected, the relaying nodes are able to achieve larger rates.

7.5 Chapter Conclusion

In this chapter, a distributed scheme was proposed that increases spectrum efficiency by hinging on cooperation-based spectrum leasing. The scheme inherently provides incentives for (otherwise non-cooperative) terminals to act as relays. Analysis has been carried out in the framework of Stackelberg games. Numerical results reveal significant benefits for

all involved terminals. Building upon the idea described in this chapter, in Chapter 8 a practical and fully decentralized scheme that exploits cooperative retransmission protocol will be proposed.

CHAPTER 8

DYNAMIC SPECTRUM ACCESS AND RELAYING MOTIVATION IN DECENTRALIZED NETWORKS VIA COOPERATIVE RETRANSMISSION PROTOCOLS

The idea of spectrum leasing via cooperation and relaying motivation was proposed in Chapter 7 for a system modelled as a Stackelberg game in which a source (primary user in cognitive jargon) attempts to maximize its throughput under the constraint that the behavior of the relaying (secondary) nodes satisfies a given equilibrium constraint. The solution therein requires a centralized decision process at the primary node based on global system information. The proposal in this chapter can be seen as a further elaboration on the ideas of Chapter 7 that offer a fully distributed solutions by exploiting cooperative retransmission protocols and auction theory.

8.1 Background and Chapter Overview

As discussed in Chapter 7, the use of cooperation from neighboring terminals [2] [3] assumes that relaying terminals agree to unconditionally assist communications they do not directly benefit from [76]. Such an unconditionally altruistic behavior is the default mode for dedicated relays stations (which are deployed with the exact goal of providing cooperation), but is arguably unrealistic for regular (e.g., users') mobile stations. In this chapter, a novel approach to cooperative HARQ is proposed that inherently provides an incentive for the relays' assistance based on spectrum leasing principle (similarly as in Chapter 7). Specifically, incentive for (otherwise non-cooperative) relaying station is given by the opportunity to exploit part of the retransmission slot for their own data. In other words, the source may lease a portion of the retransmission slots in exchange for cooperation. Relays compete for the retransmission slot through an auction mechanism,

by trying to make the best "retransmission offer" to the source (the source, the relays and the retransmission slot can be considered as the auctioneer, the bidders and the bidding article, respectively). In this process, the relays' goal is that of obtaining access to the retransmission slot in order to exploit it for transmission of their own data with a maximized transmission reliability, under the constraint of maintaining the "offer" made in the auction phase. As detailed later, the degrees of freedom for a relay when deciding on the retransmission offer are the fractions of the time slot and the energy to be invested in retransmission compared to transmission of own data. The proposed mechanism naturally leverage gains from opportunistic transmission and does not require centralized control with full system information. It is also noted that instead of the time-slots, the proposed scheme can similarly entail a number of subcarriers in an OFDM system.

It is remarked that, in addition to motivating the relays' cooperative behavior, the scheme proposed in this chapter can be alternatively considered (similarly as in the previous chapter) as a practical framework for the implementation of property-rights cognitive radio (spectrum leasing) [23]. Within this setting, the source acts as primary and the relays as secondary nodes. The proposed scheme prescribes the retribution for spectrum access from the secondary (relay) nodes to the primary (source) in the form of cooperation and in a fully distributed fashion. The strategy thus avoids the regulatory issues or money transactions of the standard implementation of the spectrum leasing concept.

There is an extensive literature on exploiting auction-theoretic frameworks for improving the efficiency of spectrum utilization. In [77], a spectrum sharing approach is proposed in which secondary users purchase channels from a primary user (i.e., a spectrum broker) through an auction process, with the payment metric based on received signal-to-noise ratio or received power. The proposed algorithm is shown to converge to the socially optimum equilibrium. In [78], a real-time auction framework that distributes spectrum among a large number of wireless users under an interference constraint is put forth and shown to result in a conflict-free spectrum allocation. An auction- and

game-theoretic framework that captures the interaction among spectrum broker, service providers, and end–users, in a multi–provider setting, is studied in [79]. Two auction mechanisms, the signal-to-noise and the power auction, are introduced in [80] to determine the (dedicated) relay selection and the relay's power allocation in a decentralized manner.

The chapter is organized as follows. In Section 8.2, the proposed scheme is described using the auction-theoretic framework and the main system parameters are provided. In Section 8.3, focusing on Vickrey (sealed-bid second-price) class of auctions, an extensive analysis of the scheme performance (namely, investigation of the system equilibrium and the system performance in terms of expected number of transmissions required for successful message delivery, for both the source and the relays) is provided. Feasibility of sealed-bid first-price auction integration into proposed scheme is discussed in Section 8.4. Numerical results are used in Section 8.5 to illustrate performance improvements achievable with the proposed scheme. In Section 8.6, possible directions for further investigations are highlighted. Concluding remarks are provided in Section 8.7.

8.2 System Model

In this section, the general overview of the proposed scheme (Section 8.2.1) is first presented. The relays' strategies and goals are formulated in Section 8.2.2, while the fundamental auction- (game-) theoretic concept of equilibrium (in particular, the dominant strategy equilibrium) for the problem at hand is introduced in Section 8.2.3. The physical layer parameters are given in n Section 8.2.4.

8.2.1 Model Overview

With reference to Figure 8.1-(a), consider a scenario with a source terminal S transmitting towards the access point AP, and a set of K (possibly) relaying terminals $\{\mathrm{R}_k\}_{k=1}^{K}$ that have their own data to transmit towards the AP. The source employs a retransmission protocol (ARQ) and, in case of a retransmission request (Negative Acknowledgement,

NACK message) from the AP, it is willing to lease the retransmission slot to one of the relays that have decoded its original transmission and can improve the quality of retransmission by taking it over (Figure 8.1-(b)). Without loss of generality, the relays available for retransmission are denoted as $\{\mathrm{R}_k\}_{k=1}^{n}$, where $n \leq K$ is their number. Simultaneously, the awarded relay is allowed to exploit the retransmission slot for its own data, under the constraint of maintaining the source's message retransmission quality agreed upon during the auction phase (to be detailed below). For analytical convenience, a memoryless retransmission protocol (Hybrid ARQ Type I) is assumed, although the scheme can in principle accommodate more sophisticated protocols (e.g., packet combining or incremental redundancy [1]).

Assuming quasi-static fading channels, the *transmission reliability*, or specifically, the probability of successful transmission, is adopted as the performance criterion for transmission quality. Such probabilities, as detailed in Section 8.2.4, are evaluated by source and relays based on the knowledge of the channel statistics of their own transmitting channels towards the AP, and on the measurement of a short training message broadcast by the AP along with the NACK. The following reliabilities are considered throughout this chapter:

- p_0, evaluated by the source, is the reliability of the source's message (re)transmission if performed by the source alone;
- p_k, evaluated by the relay R_k, is the reliability of the source's message retransmission if performed by the relay R_k;
- q_k, evaluated by the relay R_k, is the reliability of the relay R_k's message transmission, granted that R_k performs the source's message retransmission.

Relays may use the current reliabilities p_k as bids to be submitted to the source to enable the auction-based retransmission slot assignment. It is noted that in general any multiple-access scheme can be employed to ensure the collision-free submission of the

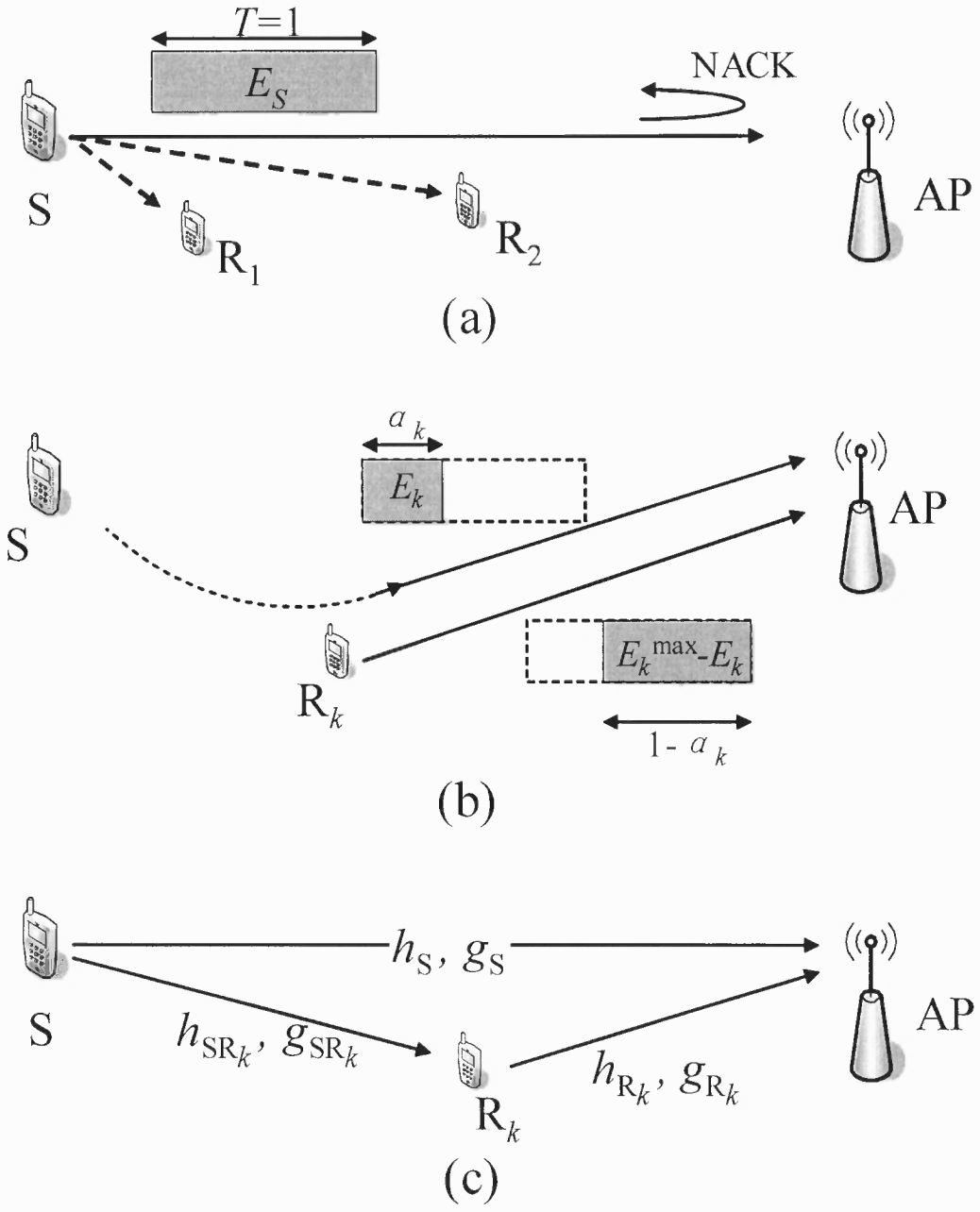

Figure 8.1 Proposed auction-based retransmission model: (a) original source transmission (broadcast) with NACK from AP ($K = 2$), (b) retransmission in case R_k wins the auction, (c) channels and respective average channel gains.

bids to the source (not further elaborated upon here). Having collected all the bids from the relays $\{\mathrm{R}_k\}_{k=1}^{n}$, the source decides to lease the retransmission slot to the relay that offered the highest p_k if the latter is larger than the source's *reserve price* (direct transmission probability) p_0.

8.2.2 Relays' Strategies and Goals

The duration of the (re)transmission slot is normalized to $T = 1$. To accommodate for transmission of two messages during the retransmission slot and evaluate the reliability p_k, the relay R_k plans to set a fraction $0 \leq \alpha_k \leq 1$ of the (leased) slot for the retransmission of the source's message, and reserve the remaining fraction of $1 - \alpha_k$ for transmission of its own data, as sketched in Figure 8.1-(b). Furthermore, denoting the R_k's transmission energy per slot as $E_k^{\max}$ [joule/channel symbol], the transmission energy invested for the source's retransmission (during the fraction α_k) is $0 \leq E_k \leq E_k^{\max}$ and thus, $E_k^{\max} - E_k$ is the energy left for transmission of its own data, during the remaining slot fraction of duration $1 - \alpha_k$. For the time being, it is sufficient to assume that the reliability of source's message retransmission p_k and reliability of transmission of its own data q_k are strictly increasing and decreasing functions, respectively, of α_k and E_k (the exact dependence is given in Section 8.3). It is further noted that in practice, time is quantized in, say, L (time) units or frames. Consequently, the slot partitioning in this work reflects into the allocation of $\alpha_k L$ and $(1 - \alpha_k)L$ units for the transmission of source's and relay's data, respectively. In addition, different power levels among slots imply system with varying power gain.

The relay R_k is interested in relaying the source's packet (i.e., in obtaining the retransmission slot) to attain the opportunity to transmit its own traffic as reliably as possible, with a minimum tolerated reliability $q_{k,\min}$. In principle, any utility function

reflecting these goals can be assigned to the relays. Here, the utility is chosen as

$$u_k(\{\alpha_i, E_i; t_i\}_{i=1}^n) = \begin{cases} q_k(\alpha_k, E_k; t_k) - q_{k,\min}, \text{ for } p_k(\alpha_k, E_k; t_k) = \max\left(p_0,\ p_i(\alpha_i, E_i; t_i)_{i=1}^n\right) \\ 0, \text{ otherwise,} \end{cases} \tag{8.1}$$

This says that, if relay k wins the auction (first condition in (8.1)), he accrues an utility equal to $q_k - q_{k,\min}$, whereas otherwise the utility is zero. Definition (8.1) makes explicit the dependence of the system parameters on the relays' resource allocation $\{\alpha_k, E_k\}$ and the *type* t_k. The type t_k summarizes the parameters characterizing R_k, such as $q_{k,\min}$ and $E_k^{\max}$ (other parameters of interest are introduced in Section 8.2.4). Notice that (8.1) reflects a trade-off for R_k between maximizing its transmission reliability q_k, which calls for small α_k and E_k, and the probability of being selected for transmission by providing the largest bid p_k, which calls for large α_k and E_k.

In the following subsection the important auction- (game-) theoretic concept of *dominant strategy equilibrium* is defined and applied to the setting described above.

8.2.3 Auction and Dominant Strategy Equilibrium

Relays select their transmission strategy α_k and E_k and, as a consequence, the bid p_k in a rational and selfish way, being interested in maximizing the utility (8.1). Such a scenario can be conveniently investigated in the framework of *auction theory* [33] [81]. Specifically, auction theory provides means to identify meaningful operational points corresponding to *equilibrium states* for the competitive decision processes. Identifying such equilibrium points can be used to predict the system behavior and to allow system design.

Following standard game-theoretic definitions (Section 2.2), an equilibrium point defines a set of relays' (players' in the game-theoretic jargon) strategies from which no relay has incentive (in some sense to be specified) to unilaterally deviate (i.e., if no other player does). Several equilibrium solutions may be defined that have different robustness properties with respect to the amount of information that a certain relay is assumed to

know regarding the other relays' types t_{-k} (subscript $-k$ denotes the complementary set, $t_{-k} = (t_i)_{i=1;i\neq k}^{n}$). Here, the focus is on the *dominant strategy equilibrium (DSE),* a concept that poses the strongest requirement in terms of robustness: DSE strategies are required to remain preferable to every relay *irrespective* of the amount of information available on the other relays' types [33]. DSE have thus two essential features: on one hand, they provide a reliable prediction of the system behavior due to the robustness property mentioned above; on the other hand, they can be implemented without the need for exchanging information regarding other relays' types. A formal definition follows.

A selection of strategies $\{(\alpha_k^*, E_k^*)_{t_k}\}_{k=1}^{n}$ (where $(\alpha_k, E_k)_{t_k}$ denotes the strategy played by relay R_k characterized by type t_k) is a dominant strategy equilibrium (DSE) if, for each type t_k of player k, for any $(\alpha_k, E_k)_{t_k}$ and any $(\alpha_{-k}, E_{-k})_{t_{-k}}$ for all the types t_{-k} of other players:

$$u_k(t_k, (\alpha_k^*, E_k^*)_{t_k}, t_{-k}, (\alpha_{-k}, E_{-k})_{t_{-k}}) \geq u_k(t_k, (\alpha_k, E_k)_{t_k}, t_{-k}, (\alpha_{-k}, E_{-k})_{t_{-k}}). \quad (8.2)$$

In other words, the DSE solution requires that strategy $(\alpha_k^*, E_k^*)_{t_k}$ for relay R_k (if its type is t_k) is the best response (In the sense of a *weak dominance*, as the inequality in (8.2) is not strict) against any realization of the opponent types t_{-k} and the corresponding strategies $(\alpha_{-k}, E_{-k})_{t_{-k}}$. Consequently, this strategy is played by a rational R_k even if the other players behave irrationally.

Finding a DSE solutions for a general class of auctions is prohibitive. However, for Vickrey auctions (also known as second-price auctions), solution can be typically found. This scheme will be discussed in details in Section 8.3; for a different utility function at the relays, a more traditional first-price auction will be analyzed in Section 8.4.

8.2.4 Physical Layer and Channel Model

In closing this section, the physical layer model is provided that details the specific instance of the reliability functions that will be considered for the rest of the chapter. The channels

over different links are modeled as independent complex Gaussian variables, invariant within the transmission slot (Rayleigh block fading). Delay between the (re)transmission slots is large enough to assume uncorrelated block fading. The following notation is employed to denote the instantaneous complex channel values within a (re)transmission slot (Figure 8.1-(c)): h_S between the source S and access point AP; h_{SR_k} between S and the relay R_k ($k = 1, ..., K$); and h_{R_k} between R_k and AP. The average channel power gains are $g_S = E[|h_S|^2]$, $g_{SR_k} = E[|h_{SR_k}|^2]$ and $g_{R_k} = E[|h_{R_k}|^2]$, where $E[\cdot]$ denotes the expectation operator. The source's energy per slot is E_S [joule/channel symbol] and the single-sided spectral density of the independent white Gaussian noise at any of the receivers is normalized to unity $N_0 = 1$. The target transmission rates (per each transmission) r_S [bit/s/Hz] for the source S and r_{R_k} [bit/s/Hz] for the relay R_k are considered fixed and set by the application.

To determine the transmission reliability for a given (re)transmission slot, each transmitting node exploits knowledge of the channel statistics towards the AP, along with outdated or noisy channel state information. Such information is obtained via a training sequence received by source and relay before transmission in the current block. The training sequence is embedded in a broadcast message from the AP and it can be, for instance, piggybacked in the ACK/NACK message (this assumes channel reciprocity as for time-division-duplex (TDD)). Channel variation during the interval between the estimation instant and the (re)transmission slot and/or channel estimation/ quantization noise are accounted for by a correlation parameter ρ, as in, e.g., [82] [83]. Notice that delay between the downlink channel estimation and the following (re)transmission slot needs to be considerably smaller than the delay between (re)transmissions, in order for block-fading to hold. The actual channel $h \in \{h_S, h_{R_1}, .., h_{R_K}\}$ during the (re)transmission slot is then obtained with respect to the estimated channel $\hat{h}$ as [82] [83]:

$$h = \rho\hat{h} + w, \tag{8.3}$$

where $\hat{h} \sim \mathcal{CN}(0, g)$, $g = E[|h|^2]$ and $w \sim \mathcal{CN}(0, \sigma^2)$ is the innovation term (due to the outdated knowledge or estimation/ quantization noise) with variance

$$\sigma^2 = (1 - \rho^2)g. \tag{8.4}$$

The normalized power channel gain $|h|^2/\sigma^2$, conditioned on the estimate $\hat{h}$, takes the distribution of a noncentral chi-square variable with two degrees of freedom and noncentrality parameter $|\rho\hat{h}|^2/\sigma^2$, $|h|^2/\sigma^2 \sim \chi_2^2(|\rho\hat{h}|^2/\sigma^2)$. From this distribution, the reliability is easily evaluated by assuming coding at the Shannon limit for a given target rate and considering the outage probability, as detailed in the following section.

8.3 System Performance under Vickrey Auction Rules

In this section the Vickrey (sealed-bid second-price) auction is considered due to its convenient properties in resource allocation scenarios [81]. In particular, in Section 8.3.1 the preliminaries on Vickrey auction and motivation for this choice are provided. Applying the rules of this auction to the scheme at hand, the DSE is elaborated upon in Section 8.3.2 and the DSE's functional dependence on transmission parameters (defined in Section 8.2.4) are provided in Section 8.3.3 and Section 8.3.4. Taking DSE as the outcome of each auction, the system performance is then evaluated in terms of average number of slots required for reliable transmission of the source's and a relay's message, in Section 8.3.5 and Section 8.3.6, respectively. For analytical tractability, this task is performed by assuming all K relays collocated ($g_{SR_k} = g_{SR}$ and $g_{R_k} = g_R$) and identical ($q_{k,\min} = q_{\min}$, $E_k^{\max} = E^{\max}$ and $r_{R_k} = r_R$).

8.3.1 Background on Vickrey Auction

In sealed-bid second-price (Vickrey) auctions [84], the bidding item is awarded to the highest bidder at the price of the second highest bid (i.e., at the price of the highest losing bid). The most attractive property of Vickrey auction is its "truth telling nature": namely, a

dominant strategy for each bidder is to report to the auctioneer its evaluation of the bidding item *truthfully*. In particular, [84] defines *truthful bidding* as bidding with the "price at which a bidder would be on the margin of indifference as to whether he obtains the article or not,..., a highest amount he could afford to pay without incurring a net loss". To provide a brief intuition on the truthful bidding properties of Vickrey auctions, notice that if bidding less than the value of indifference, the bidder can only reduce his chance of winning while not affecting the price it would pay if he was the winner. On the other hand, if bidding with a value larger than that of indifference, the chance of winning increases but only if yielding an unprofitable outcome. As a consequence, *implementation of an optimal dominant strategy for Vickrey auctions at each bidder requires no information on the other bidders' strategies* or their evaluations of the bidding item, as this knowledge would not impact the truthful bidding strategy.

The Vickrey model generally results in an efficient goods allocation, as reported in [81] [84]- [86], almost identical to that of a classic English first-price ascending auction [85] [86]. Attractive properties of Vickrey auctions have also inspired related research within the wireless community. For example, [87] exploits Vickrey auction to determine the optimum partner selection in a self-configuring cooperative network. Vickrey auction was implemented in [88] to design a wireless network model that combats selfishness and enforces cooperation among nodes. In [89], an algorithm based on the Vickrey auction was applied to the problem of fair allocation of a wireless fading channel. As a final remark, notice that Vickrey auctions are vulnerable to malicious behavior of the auctioneer ("lying auctioneer") and the bidders (bidder collusion), and appropriate mechanisms should be applied for its protection (see, e.g., [90] for a discussion).

8.3.2 Dominant Strategy Equilibrium

Here the DSE solutions are investigated for the model at hand (as described in Section 8.2.3) when the source employs a Vickrey auction mechanism. It should be first noted that,

unlike in conventional auction theory, where both the auctioneer and the bidders have a common trade currency (money), herein the "profit" of the auctioneer and the bidders are based on clearly distinguished preferences (transmission reliability of the source's and a relay's message, respectively). Consequently, the problem needs to be formulated in the more general framework of *Bayesian games* (of which auction theory is a branch) [33] [91] [92]. This is formalized in Appendix B, along with a proof of the DSE existence. In the following, the details on the DSE are provided with a less rigorous but intuitive approach, relying on the auction framework (Section 8.3.1).

Let $\breve{k}$ denote the index of the winning relay and $\breve{p}_{\breve{k}}$ be the reliability it needs to provide to the source. For a Vickrey auction, these quantities read

$$\breve{k} = \begin{cases} \arg\max_k p_k(\alpha_k, E_k; t_k), & \text{if } \max_k p_k(\alpha_k, E_k; t_k) \geq p_0 \\ 0 & \text{otherwise,} \end{cases} \tag{8.5a}$$

$$\breve{p}_{\breve{k}} = \begin{cases} \max(p_0, \ \max_{k \neq \breve{k}} p_k(\alpha_k, E_k; t_k)) & \breve{k} = 1, ..., n \\ p_0 & \breve{k} = 0 \end{cases}, \tag{8.5b}$$

where (8.5a) simply states that the higher bidder is selected (as also assumed in (8.1)) and (8.5b) imposes that the winning relay only pays the second highest price. Notice that auction rules (8.5a)-(8.5b) also address the auction outcome $(\breve{k}, \breve{p}_{\breve{k}}) = (0, p_0)$ when none of the relays wins the auction. Also, in the case of multiple equal (highest) offers, the tie is broken by random allotment to one of the strongest bidders [84]. Finally, notice that the source can in practice choose a larger reserve price than p_0 in order to compensate for the practical cost of the cooperative scheme, such as signalization and delay.

As described in Section 8.3.1, the Vickrey auction admits a DSE with the strategies chosen so that the utility (profit) is on the margin of indifference as to whether the player wins the auction or not. Applying this principle to the utility (8.1), it is clear that in the DSE the following needs to be satisfied (a more formal proof is presented in Appendix B):

$$q_k(\alpha_k, E_k; t_k) = q_{k,\min}. \tag{8.6}$$

Given this constraint, maximization of the utility (8.1) is attained when the relay chooses the pair (α_k, E_k) so as to maximize $p_k(\alpha_k, E_k; t_k)$, under the constraint (8.6). Thus, the DSE prescribes each relay R_k to solve

$$\begin{aligned}(\alpha_k^*, E_k^*) &= \arg\max_{\alpha_k, E_k} p_k(\alpha_k, E_k; t_k), \\ \text{s.t. } & q_k(\alpha_k, E_k; t_k) = q_{k,\min}, \\ & 0 \leq \alpha_k \leq 1, \\ & 0 \leq E_k \leq E_k^{\max},\end{aligned} \tag{8.7}$$

and the bid submitted to the source is $p_k^* = p_k(\alpha_k^*, E_k^*; t_k)$. Optimization (8.7) can then be considered as local (relay-based) mapping f_k: $q_{k,\min} \rightarrow p_k^*$. Notice that, according to the monotonicity properties of the reliability functions, mapping $f_k : q_{k,\min} \rightarrow p_k^*$ is non-increasing in $q_{k,\min}$ (this will be more formally addressed in Section 8.3.3).

As specified in (8.5), at the end of the auction process, the winning relay $\mathrm{R}_{\breve{k}}$ is required to guarantee the source reliability $\breve{p}_{\breve{k}} = \max(p_0, \max_{k \neq \breve{k}} p_k^*)$. Therefore, the winning relay $\mathrm{R}_{\breve{k}}$ can re-adjust its transmission parameters $(\alpha_{\breve{k}}, E_{\breve{k}})$ by maximizing its reliability $q_{\breve{k}}$ while guaranteeing the required source reliability $\breve{p}_{\breve{k}}$, as (recall (8.1))

$$\begin{aligned}(\breve{\alpha}_{\breve{k}}, \breve{E}_{\breve{k}}) &= \arg\max_{\alpha_{\breve{k}}, E_{\breve{k}}} q_{\breve{k}}(\alpha_{\breve{k}}, E_{\breve{k}}; t_{\breve{k}}), \\ \text{s.t. } & p_{\breve{k}}(\alpha_{\breve{k}}, E_{\breve{k}}; t_{\breve{k}}) = \breve{p}_{\breve{k}}, \\ & 0 < \alpha_{\breve{k}} \leq 1, \\ & 0 < E_{\breve{k}} \leq E_{\breve{k}}^{\max},\end{aligned} \tag{8.8}$$

with the source reliability assigned to the winning relay $\mathrm{R}_{\breve{k}}$ being guaranteed by the constraint, $p_{\breve{k}}(\alpha_{\breve{k}}, E_{\breve{k}}; t_{\breve{k}}) = \breve{p}_{\breve{k}}$. The final reliability achieved by the winning relay $\mathrm{R}_{\breve{k}}$ is denoted as $\breve{q}_{\breve{k}} = q_{\breve{k}}(\alpha_{\breve{k}}, E_{\breve{k}}; t_{\breve{k}})$. Notice the similarity between (8.7) and (8.8), with difference being only in swapping constraints and objective reliabilities.

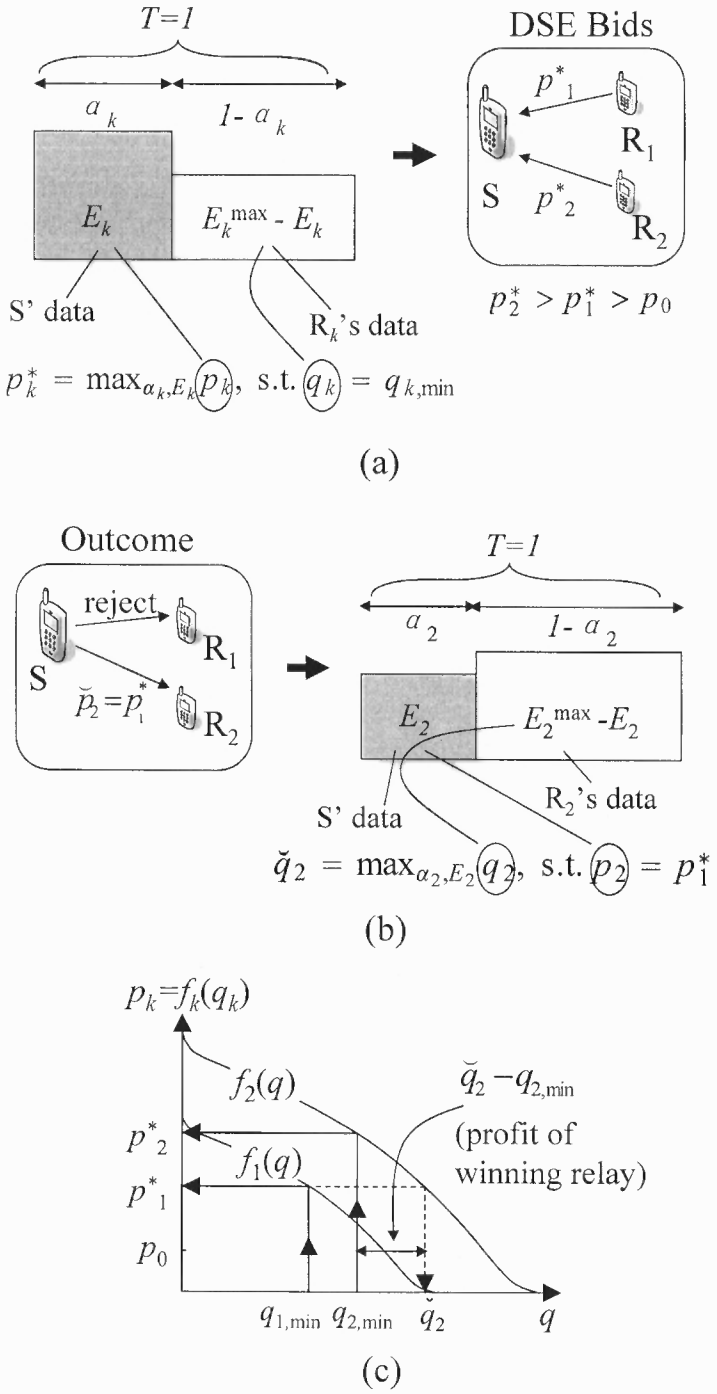

Figure 8.2 Proposed auction-based model (under the Vickrey auction rules): (a) upon reception of a NACK for the source' packet, relays submit their bids (DSE equilibrium), (b) winning relay's strategy readjustment and (c) summary of the auction process (DSE) with mapping $p_k^* = f_k(q_{k,\min})$ and profit of the winning relay $\breve{q}_{\breve{k}} - q_{\breve{k},\min}$ ($K = n = 2$ and $p_2^* > p_1^* > p_0$, with $\breve{k} = 2$ and $\breve{p}_{\breve{k}} = p_1^*$).

Figure 8.2 illustrates the auctioning process for $K = n = 2$, with winning relay being $\breve{k} = 2$ and $p_0 < p_1^* = \breve{p}_2 < p_2^*$. In particular, the bid by (8.7) and the winning relay's reliability obtained from (8.8) are illustrated in Figure 8.2-(a) and Figure 8.2-(b), respectively. The auction process can be summarized in Figure 8.2-(c), where bidding (8.7) reflects the mapping $q_{k,\min} \rightarrow p_k^*$, the value $\breve{p}_{\breve{k}}$ is the agreed source reliability and the mapping $\breve{p}_{\breve{k}} \rightarrow \breve{q}_{\breve{k}} \geq q_{\breve{k},\min}$ corresponds to optimization (8.8).

8.3.3 Solving for DSE (Problem (8.7))

The optimization problem (8.7) provides the bidding strategy in DSE as $p_k^* = f_k(q_{k,\min})$. To elaborate, the expressions for the reliabilities p_k and q_k of relay R_k are derived in the following (for simplicity of notation the dependence on α_k, E_k and t_k is dropped). Reliability (i.e., probability of successful transmission) is the probability that the channel can accommodate transmission rate, given the channel state information available at the transmitter that accounts for the outdated channel (8.3). Assuming coding at the Shannon limit for a given target rate r_S and r_{R_k}, reliabilities read

$$p_k = \Pr\left\{C_{S_k} \geq r_S | \hat{h}_{R_k}\right\} \tag{8.9a}$$

$$q_k = \Pr\left\{C_{R_k} \geq r_{R_k} | \hat{h}_{R_k}\right\}, \tag{8.9b}$$

where the rates achievable by the R_k during the two time intervals α_k and $1 - \alpha_k$, respectively, are

$$C_{S_k} = \alpha_k \log_2\left(1 + |h_{R_k}|^2 \frac{E_k}{\alpha_k}\right) \tag{8.10a}$$

$$C_{R_k} = (1 - \alpha_k) \log_2\left(1 + |h_{R_k}|^2 \frac{E_k^{\max} - E_k}{1 - \alpha_k}\right). \tag{8.10b}$$

Notice that the existing capacity-approaching codes can be easily accommodated in the framework (8.10a)-(8.10b) by scaling the energies for an appropriate gap [30] (Section 2.1.2). Probabilities (8.9) are evaluated using the outdated Rayleigh fading model (8.3).

Applying the constraint $q_k = q_{k,\min}$ of (8.7) to (8.9b) and (8.10b) yields

$$q_{k,\min} = \Pr\left\{|h_{R_k}|^2 \geq \frac{(1-\alpha_k)\left(2^{\frac{r_{R_k}}{1-\alpha_k}}-1\right)}{E_k^{\max}-E_k}\right\}$$
$$= 1 - F_{\chi^2}\left\{\frac{(1-\alpha_k)\left(2^{\frac{r_{R_k}}{1-\alpha_k}}-1\right)}{\sigma_{R_k}^2\left(E_k^{\max}-E_k\right)}, \frac{|\rho\hat{h}_{R_k}|^2}{\sigma_{R_k}^2}\right\}, \tag{8.11}$$

where $F_{\chi^2}\{x,\mu\}$ is the cumulative distribution function (cdf) of the noncentral chi-square distribution with two degrees of freedom and noncentrality parameter μ, taken at value x (Section 2.1.1). Revising (8.11) yields the following relationship between parameters α_k and E_k in (8.7) that satisfy $q_k = q_{k,\min}$

$$E_k(\alpha_k) = E_k^{\max} - \frac{(1-\alpha_k)\left(2^{\frac{r_{R_k}}{1-\alpha_k}}-1\right)}{\sigma_{R_k}^2 F_{\chi^2}^{-1}\left\{1-q_{k,\min}, \frac{|\rho\hat{h}_{R_k}|^2}{\sigma_{R_k}^2}\right\}}, \tag{8.12}$$
$$0 < \alpha_k \leq 1, \;\; 0 < E_k \leq E_k^{\max},$$

where $F_{\chi^2}^{-1}\{x,\mu\}$ is the cdf of the inverse noncentral chi-square distribution with two degrees of freedom and noncentrality parameter μ, taken at value x: $F_{\chi^2}^{-1}\{F_{\chi^2}\{x,\mu\},\mu\} = x$.

Similarly to (8.11), (8.9a) and (8.10a) yield to the source message reliability p_k, which is the objective in (8.7):

$$p_k = \Pr\left\{|h_{R_k}|^2 \geq \frac{\alpha_k\left(2^{rs/\alpha_k}-1\right)}{E_k(\alpha_k)}\right\} \tag{8.13}$$
$$= 1 - F_{\chi^2}\left\{\frac{\alpha_k\left(2^{rs/\alpha_k}-1\right)}{\sigma_{R_k}^2 E_k(\alpha_k)}, \frac{|\rho\hat{h}_{R_k}|^2}{\sigma_{R_k}^2}\right\},$$

where the relationship $E_k(\alpha_k)$ follows from (8.12). The optimization problem (8.7) (DSE) now boils down to trivial optimization for the time fraction

$$\alpha_k^* = \arg \max_{0<\alpha_k\leq 1} \frac{\alpha_k \left(2^{r_S/\alpha_k} - 1\right)}{\sigma_{R_k}^2 E_k(\alpha_k)}, \tag{8.14}$$

and the corresponding bidding value p_k^* offered to the source becomes

$$p_k^* = 1 - F_{\chi^2}\left\{ \frac{\alpha_k^* \left(2^{r_S/\alpha_k^*} - 1\right)}{\sigma_{R_k}^2 E_k(\alpha_k^*)}, \frac{|\rho \hat{h}_{R_k}|^2}{\sigma_{R_k}^2} \right\}. \tag{8.15}$$

Relationship (8.15) describes the mapping of a relay's minimum reliability $q_{k,\min}$ to the DSE bid p_k^* as $p_k^* = f_k(q_{k,\min})$ (notice that $q_{k,\min}$ is not shown explicitly in (8.15), but is an argument of $E_k(\alpha_k^*)$ from (8.12); furthermore, it is easy to see that p_k^* is non-increasing function of $q_{k,\min}$).

8.3.4 Auction Outcome in DSE (Problem (8.8))

The DSE outcome of the auction in terms of the reliability achieved by the source follows the mechanism (8.5)

$$\breve{p}_{\breve{k}} = \max\left\{ p_0, \max_{k\neq \arg\max_k p_k^*} p_k^* \right\}, \tag{8.16}$$

where $k = 1, .., n$. Reliability achieved by source alone (for the rate r_S) p_0 in (8.16) is given as

$$\begin{aligned} p_0 &= \Pr\left\{ C_S \geq r_S | \hat{h}_S \right\} \\ &= \Pr\left\{ |h_S|^2 \geq \frac{2^{r_S} - 1}{E_S} \right\} \\ &= 1 - F_{\chi^2}\left\{ \frac{2^{r_S} - 1}{\sigma_S^2 E_S}, \frac{|\rho \hat{h}_S|^2}{\sigma_S^2} \right\}. \end{aligned} \tag{8.17}$$

where $C_S = \log_2(1 + |h_S|^2 E_S)$ is the rate achievable on the channel between the source and the access point.

Assuming that the auction yields a relay's access, i.e., $\breve{k} \neq 0$, the second best offer that $\mathrm{R}_{\breve{k}}$ needs to provide to the source is given in (8.16). This reliability is further mapped into the improved reliability $\breve{q}_{\breve{k}}$ for $\mathrm{R}_{\breve{k}}$'s own transmission through (8.8). As indicated in Section 8.3.2, this optimization problem is identical to (8.7) (namely, parameters α_k and E_k in (8.7) become 'mirrored' as $1-\alpha_{\breve{k}}$ and $E_{\breve{k}}^{\max} - E_{\breve{k}}$ in (8.8)), with solutions that can be easily adopted from (8.14), (8.12) and (8.15):

$$\breve{\alpha}_{\breve{k}} = 1 - \arg\max_{0<\alpha_{\breve{k}}\leq 1} \frac{(1-\alpha_{\breve{k}})\left(2^{\frac{r_{R_{\breve{k}}}}{1-\alpha_{\breve{k}}}}-1\right)}{\sigma^2_{R_{\breve{k}}}\left(E_{\breve{k}}^{\max} - E_{\breve{k}}(\alpha_{\breve{k}})\right)}, \tag{8.18}$$

where

$$E_{\breve{k}}(\alpha_{\breve{k}}) = \frac{\alpha_{\breve{k}}\left(2^{rs/\alpha_{\breve{k}}}-1\right)}{\sigma^2_{R_{\breve{k}}} F^{-1}_{\chi^2}\left\{1-\breve{p}_{\breve{k}}, \frac{|\rho\hat{h}_{R_{\breve{k}}}|^2}{\sigma^2_{R_{\breve{k}}}}\right\}}, \tag{8.19}$$

$$0 < \alpha_{\breve{k}} \leq 1,\ 0 < E_{\breve{k}} \leq E_{\breve{k}}^{\max},$$

which leads to

$$\breve{q}_{\breve{k}} = 1 - F_{\chi^2}\left\{\frac{\left(1-\alpha^*_{\breve{k}}\right)\left(2^{\frac{r_{R_{\breve{k}}}}{1-\alpha_{\breve{k}}}}-1\right)}{\sigma^2_{R_{\breve{k}}}\left(E_{\breve{k}}^{\max} - E_{\breve{k}}(\alpha_{\breve{k}})\right)}, \frac{|\rho\hat{h}_{R_{\breve{k}}}|^2}{\sigma^2_{R_{\breve{k}}}}\right\}. \tag{8.20}$$

8.3.5 Average Number of Transmission Slots: Source

In order to ease analysis of the average number of transmission slots required for successful decoding of the source message at the AP, in this subsection collocated and identical relays are assumed. In this way, it is justified to consider the *number* of relays participating in auction rather than considering the *particular* relays. To amend the model, it is assumed that the source keeps on retransmitting a packet until $n \geq 1$ relays successfully decode this message and initiate the auction (or an ACK message is transmitted by AP). The auction for this source's packet is then closed for other $K-n$ relays (even if the auction outcome is the retransmission by the source). Any retransmission request for this packet triggers a

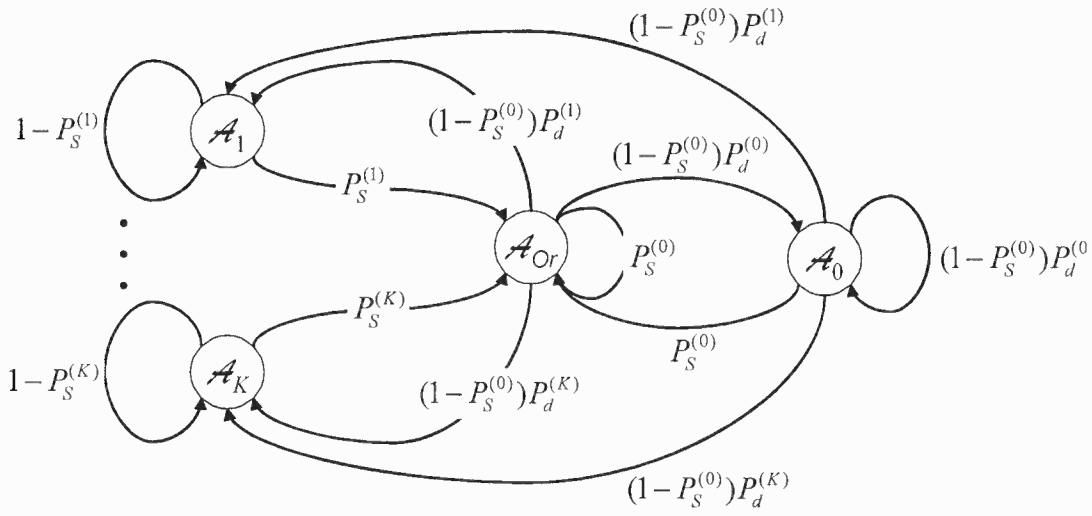

Figure 8.3 Markov process for the source' message transmission: state $\mathcal{A}_i-$ auctioned retransmission with i relays; state $\mathcal{A}_{Or}-$ original source transmission; probability $P_d^{(i)}-$ average probability that i relays decoded the source transmission; probability $P_S^{(i)}-$ average probability of successful decoding of the source data at AP with i relays participating in auction.

new auction with the same set of n bidders. As before, DSE is considered as an auction outcome.

The Markovian structure for the source's message delivery process with K involved relays is depicted in Figure 8.3. There are $K+2$ states, with state $\mathcal{A}_{Or}$ standing for the original transmission of the source, state $\mathcal{A}_0$ denoting the retransmission of the source when no relaying users decoded its message, while states $\mathcal{A}_n$ with $n = 1, ..., K$, stand for the retransmissions when n relays decoded the source's message and participate in the auction. It is easy to see that, apart from degenerate cases, the chain is irreducible and aperiodic and, thus, it has a steady-state distribution. For the states $\mathcal{A}_{Or}$ and $\mathcal{A}_n$, $n = 0, 1, .., K$, the steady-state probabilities are denoted as π_{Or} and π_n, respectively. To describe the transition probabilities, $P_S^{(n)}$ denotes the average probability of successful (re)transmission (average reliability) when n relays participate in auction, and $P_d^{(n)}$ is the probability that n relays decoded the source's transmission and are eligible to participate in the auction.

The average number of transmissions for the source $E[T_S]$ is the expected return time for the state $\mathcal{A}_{Or}$:

$$E[T_S] = \frac{1}{\pi_{Or}}. \tag{8.21}$$

If necessary, the throughput can be considered by applying the renewal-reward theorem [93] and reads $r_S \pi_{Or}$. To evaluate the steady-state probability π_{Or}, the system of equilibrium equations for Markov process (Figure 8.3) is taken into account:

$$\begin{aligned} 1 &= \sum_{n=0}^{K} \pi_n + \pi_{Or} \\ \pi_0 &= (\pi_{Or} + \pi_0)(1 - P_S^{(0)}) P_d^{(0)} \\ \pi_n &= (\pi_{Or} + \pi_0)(1 - P_S^{(0)}) P_d^{(n)} + \pi_n (1 - P_S^{(n)}), \\ \text{for } n &= 1, .., K. \end{aligned} \tag{8.22}$$

This set is easily solved for π_{Or} as

$$\pi_{Or} = \frac{1 - (1 - P_S^{(0)}) P_d^{(0)}}{1 + (1 - P_S^{(0)}) \sum_{n=1}^{K} \frac{P_d^{(n)}}{P_S^{(n)}}}. \tag{8.23}$$

Thus, to determine $E[T_S]$ one needs to solve for the probabilities $P_S^{(n)}$ and $P_d^{(n)}$, which is the focus below.

Calculation of $P_d^{(n)}$ The probability that n out of K relays correctly decoded the source's message is binomial

$$P_d^{(n)} = \binom{K}{n} P_{dec}^n (1 - P_{dec})^{K-n}, \tag{8.24}$$

where P_{dec} is the probability of successful decoding at any of the relays. This is the probability that the achievable rate $C_{SR_k} = \log_2(1 + |h_{SR_k}|^2 E_S)$ between the source and

any relay R_k is large enough to accommodate the transmission rate r_S used by the source:

$$\begin{aligned} P_{dec} &= \Pr\{C_{SR_k} \geq r_S\} \\ &= \Pr\left\{|h_{SR_k}|^2 \geq \frac{2^{r_S}-1}{E_S}\right\} \\ &= \exp\left(-\frac{2^{r_S}-1}{g_{SR}E_S}\right). \end{aligned} \tag{8.25}$$

The last equality follows from the Rayleigh fading channels assumption.

Calculation of $P_S^{(n)}$ The probability $P_S^{(n)}$ is the average reliability of the source's packet (re)transmission, given the number of decoding relays n. Recall that the probability p_0 in (8.17) accounts for the reliability when the source alone (re)transmits the message (i.e., $n = 0$), given the (outdated) channel realization $|\hat{h}_S|$. Furthermore, the probability $\breve{p}_{\breve{k}}$ in (8.16) is the reliability for the source's packet with $n \geq 1$ decoding relays, given the outdated channel realization $|\hat{h}_S|, |\hat{h}_{R_1}|, .., |\hat{h}_{R_n}|$. Thus, evaluation of $P_S^{(n)}$ accounts simply for (numerical) averaging of (8.17) or (8.16) over the distributions of $|\hat{h}_S|, |\hat{h}_{R_1}|, .., |\hat{h}_{R_n}|$, which are Rayleigh.

8.3.6 Average Number of Transmissions Slots: Relay

Notice that from the source's perspective, the relays exploit the *retransmission* slots. Nevertheless, these are referred to as the *transmission* slots, as there is no qualitative difference between transmission and retransmission slots. It is also emphasized that, unlike for the source, the number of relay's transmission slots accounts also for the slots when not transmitting at all, either due to the successfully delivered source's message, unsuccessful decoding at the relay, or not making the best offer. All assumptions stated in Section 8.3.5 are also valid here.

The tedious but quite straightforward analysis for the average number of transmission slots $E[T_R]$ required for the successful transmission of a single relay's data is summarized in Appendix C, exploiting the system's Markovian structure. Here the central expression is

given as

$$E[T_R] = \sum_{n=0}^{K} \frac{\pi_{n,R}\nu_{n,R}}{\sum_{i=0}^{K}\pi_{i,R}}, \tag{8.26}$$

which stands for the recurrence of states accounting for successful R_k's message transmission ($\nu_{n,R}$) weighted by the factor $\pi_{n,R}/\sum_{i=0}^{K}\pi_{i,R}$ over the possible number of decoding relays $n = 0, ..., K$. The steady state probabilities $\pi_{n,R}$ and the parameters $\nu_{n,R}$ are derived in details in Appendix C, applying the theory of Markov chains with rewards (Chapter 4 in [93]).

8.4 Note on Sealed-Bid First-Price Auction

This section considers the implementation of a sealed-bid first-price auction. Unlike Vickrey auction, in the first-price auction the winning relay $R_{\breve{k}}$ is required to provide the source with exactly the offered reliability $p^*_{\breve{k}}$ (rather than the one offered by the second highest bidder as in Vickrey auction) so that the reliability to be guaranteed by the winning relay becomes $\breve{p}_{\breve{k}} = \max_k p_k(\alpha_k, E_k; t_k)$ (compare with (8.5b) for Vickrey auction). With such settings, it can be seen that the best response of each user in terms of the transmission strategy (α_k, E_k) with respect to the utility (8.1) depends on other players' types and strategies, and a DSE generally does not exist [92]. To simplify analysis, one can consider the alternative relay utility functions

$$u'_k = \delta\left[\breve{k} - k\right] \cdot H(q_k - q_{k,\min}), \tag{8.27}$$

according to which the relays have no incentive to transmit with reliability larger than the minimum required $q_{k,\min}$ (notation: $\delta(x) = 0$ for $x \neq 0$ and $\delta(0) = 1$; $H(x) = 1$ for $x \geq 0$ and $H(x) = 0$, for $x < 0$). With such a utility, a DSE is easily shown to be given by (8.7). However, unlike for the Vickrey auction, here the transmission parameters (α_k, E_k) used by the winning bidder are directly those obtained from (8.7), rather than from (8.8). The analysis of this approach can be carried out by following Section 8.3.3,

directly applying (8.12), (8.14) and (8.15). Notice that, due to the relay's indifference to its transmission reliability q_k as long as $q_k \geq q_{k,\min}$ (implying that the winning relay transmits with exactly $q_{\check{k}} = q_{\check{k},\min}$) and the first-price auction setting, this approach yields an improved source's performance comparing to that of a Vickrey auction, as illustrated in Section 8.5 via numerical results.

8.5 Numerical Results

In this section, some insight are provided into the benefits of the proposed auction-based scheme. First, the optimization (8.7) is studied describing a relay's strategy in DSE and the impact of the relay's minimum reliability $q_{k,\min}$ is discussed. The overall system performance is then evaluated in terms of the expected number of transmissions for the source and the relays. Unless stated otherwise, the Vickrey auction is considered and the system parameters are set as: $N_0 = 1$, $E_S = E_k^{\max} = E_{AP} = 1$ [joule/channel symbol], $r_S = r_{R_k} = 1$ [bit/s/Hz], channel correlation $\rho = 0.5$. Furthermore, a simple geometrical model is assumed (as in Chapter 7) where all relays are placed at the same normalized distance $d \in (0, 1)$ from the source and $1 - d$ from the access point. Applying a path loss model, the channel power gains are $g_{SR} = g_S d^{-\gamma}$ and $g_R = g_S(1-d)^{-\gamma}$, where $\gamma = 3$ is the path loss exponent.

Figure 8.4 focuses on a single relay that has decoded the source's transmission and shows its dominant strategy pair (α_k^*, E_k^*) (upper figure) and the consequent bid p_k^* (lower figure) as provided by DSE (8.7), for different values of channel gain g_{R_k}. Here, the channel obtained during the AP's broadcast is assumed to take the mean value (of Rayleigh distribution) $|\hat{h}_{R_k}| = \sqrt{\pi g_{R_k}}/2$. As expected, smaller reliability constraint $q_{k,\min}$ enables the relay to use larger E_k^* and α_k^* and therefore to increase its bid p_k^*. Furthermore, as the relay's channel towards the access point g_{R_k} degrades, the capability of the relay to provide a large bid uniformly decreases.

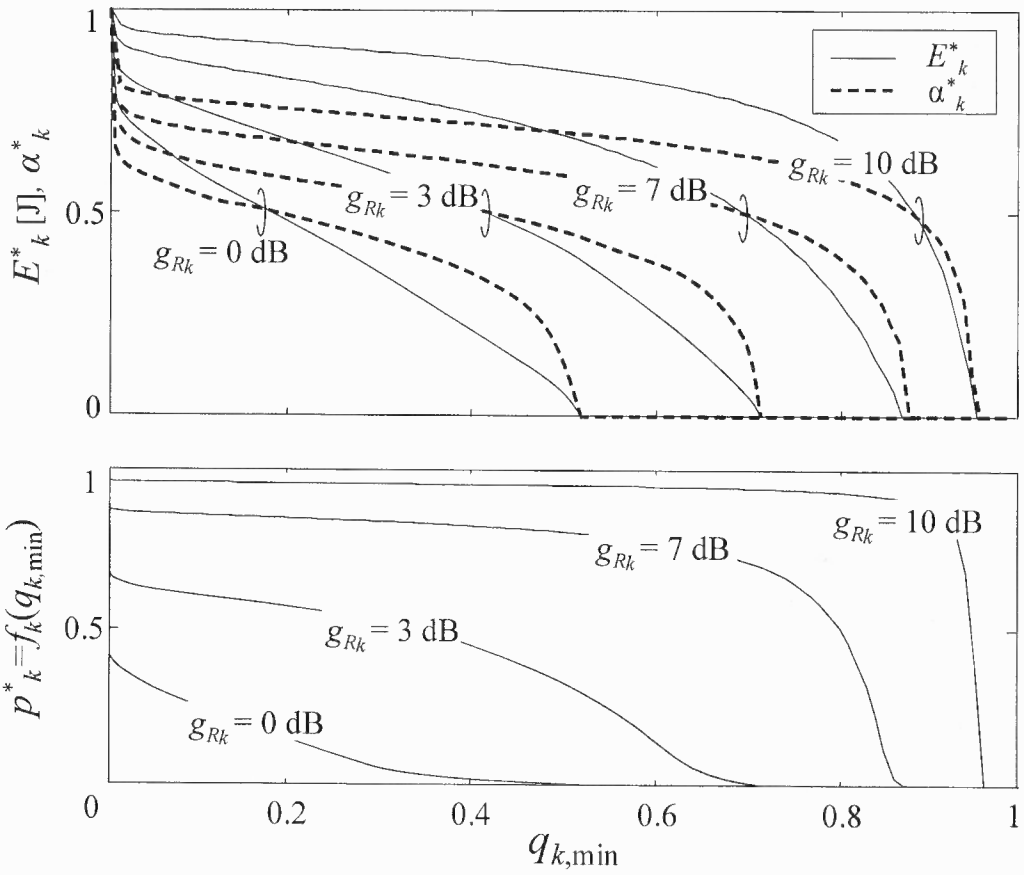

Figure 8.4 Optimization (8.7): DSE strategy pair (E_k^*,α_k^*) (upper figure) and the corresponding bid $p_k^* = f_k(q_{k,\min})$ (lower figure), for $g_{R_k} = 0, 3, 7, 10$ dB, and $|\hat{h}_{R_k}| = \sqrt{\pi g_R}/2$.

Influence of the outdated channel knowledge (Section 8.2.4) on the DSE (8.7) and the ability of relays to provide large bids is illustrated in Figure 8.5 that presents the mapping $p_k^* = f_k(q_{k,\min})$ for the estimates $|\hat{h}_{R_k}| = 1, 3$ and $g_{R_k} = 7$ dB, with different values of correlation parameter ρ. Notice that for $\rho \to 0$, the instantaneous channel $|\hat{h}_{R_k}|$ is not relevant as the distribution of the actual channel (recall (8.3) and (8.4)) remains Rayleigh. For $\rho \to 1$, on the contrary, there is no uncertainty on whether or not the relay can support its transmission and source message retransmission, and the bid becomes $p_k^* = 0$ or $p_k^* = 1$, depending on the value of $|\hat{h}_{R_k}|$. To support the source with large probabilities p_k^*, it is clearly beneficial to maintain large $|\hat{h}_{R_k}|$ with a relatively large correlation parameter ρ. Notice that the smaller values of g_{R_k} (not shown) have a detrimental effect on possibility to provide large bids p_k^* (recall (8.4) and (8.15)).

Further insight into the auction model and impact of the relay's minimum reliability, $q_{k,\min}$, is provided in Figure 8.6, that presents the average (over channel realizations) winning bid $E[p_{\breve{k}}]$ and the average provided reliability $E[\breve{p}_{\breve{k}}]$ from (8.17) (upper figure), as well as the average reliability achieved by the winning relay $E\mathbb{E}[\breve{q}_{\breve{k}}]$ from (8.20) (lower figure) versus $q_{\min}$ (same for all relays), for a model with $g_S = -4$ dB, $d = 0.5$ and the number of bidding relays $n = 1, 2, 4$. Notice that if only one relay participates in the auction $n = 1$, there is no benefit for the source, as its reserve price would be at the same time an auction outcome, $\breve{p} = p_0$. Larger values of $q_{\min}$ decrease the best offered $p_{\breve{k}}$ and the final agreed reliability $\breve{p}_{\breve{k}}$ in DSE, at the risk for the bidders of not making the offer larger than the source's reserve price. This risk is the reason for $\mathbb{E}[\breve{q}_{\breve{k}}]$ to decrease at certain point with $q_{\min}$ (namely, $\mathbb{E}[\breve{q}_{\breve{k}}]$ includes the scenarios when none of the relaying users made the offer larger than p_0, resulting in $\breve{q}_{\breve{k}} = 0$). Moreover, the larger number of relaying users participating in the auction, n, guarantees larger revenue (reliability) for the source, but not for the winning relay.

Figure 8.7 shows the average number of transmission slots for the source $E[T_S]$ (equation (8.21)) and the relaying user $E[T_R]$ (equation (8.26)) versus the placement of

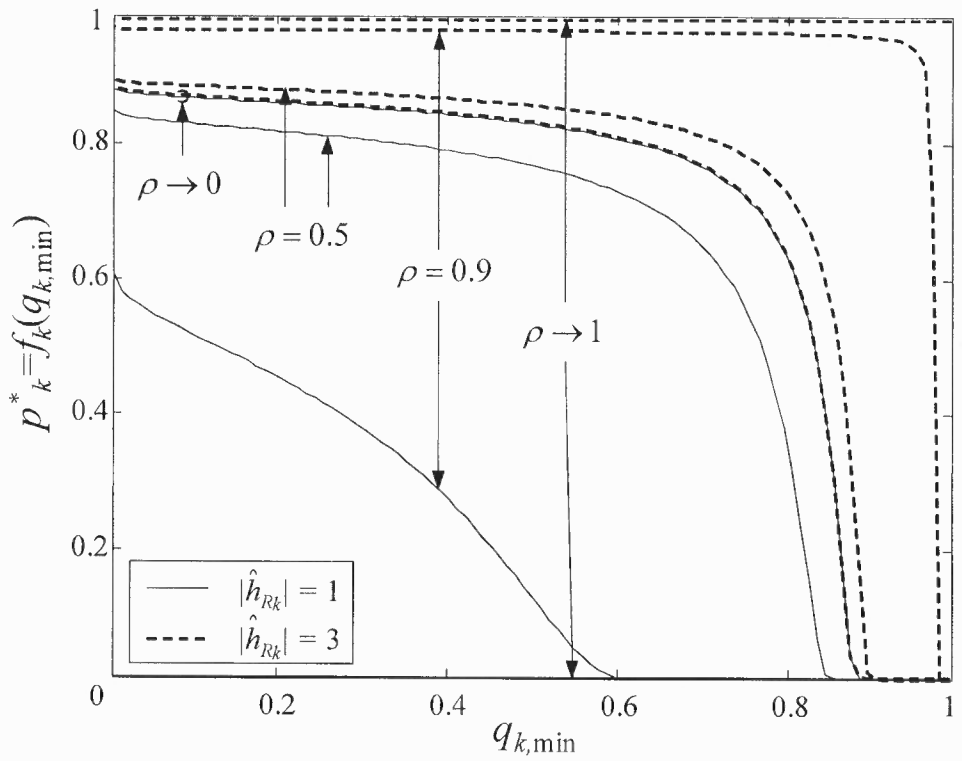

Figure 8.5 Optimization (8.7) (DSE): mapping $p_k^* = f_k(q_{k,\min})$ for $|\hat{h}_{R_k}| = 1$ (solid line) and 3 and $|\hat{h}_{R_k}| = 3$ (dashed line), for varying values of correlation parameter ρ and $g_{R_k} = 7$ dB.

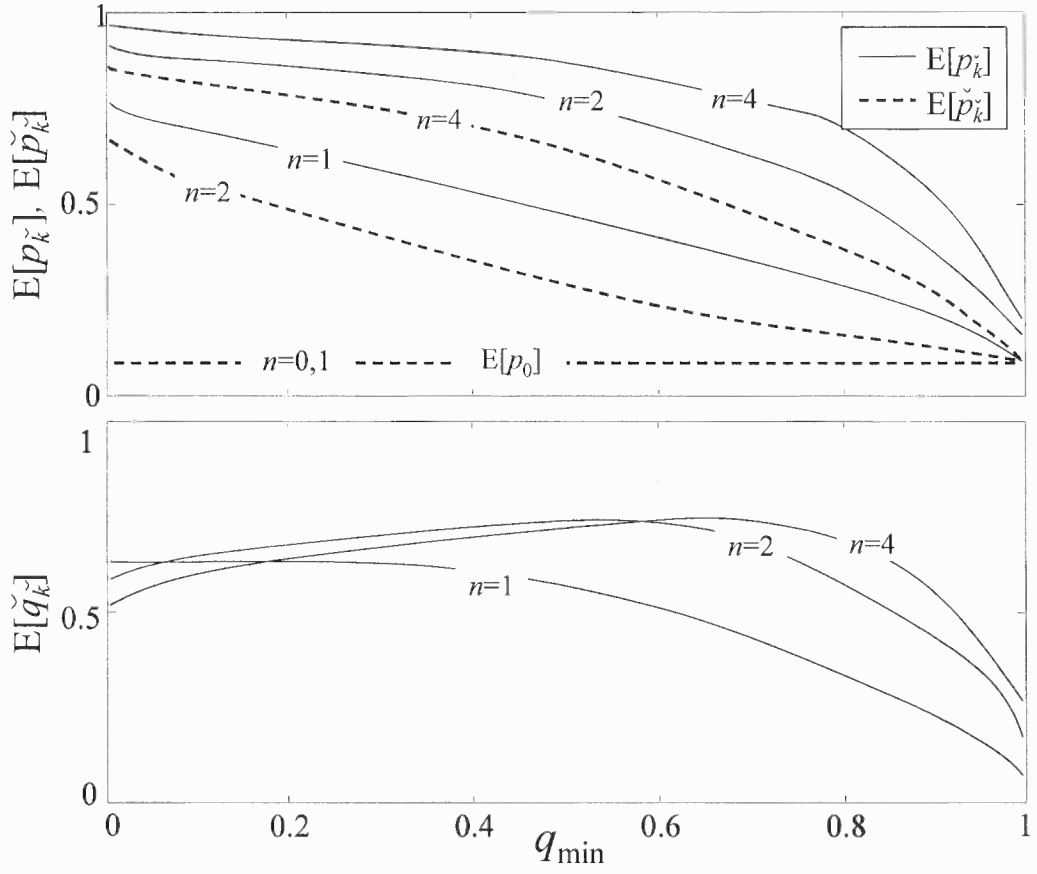

Figure 8.6 Average winning bid $E[p_{\check{k}}]$, average provided reliability $E[\check{p}_{\check{k}}]$ (8.17) (upper figure) and average achieved reliability $E[\check{q}_{\check{k}}]$ (8.20) (lower figure) in DSE versus $q_{\min}$ ($g_S = -4$ dB, $d = 0.5$ and $n = 1, 2, 4$).

relays d, for $g_S = -7$ dB (S-AP link in deep shadowing), $q_{\min} = 0, 0.2, 0.5$, and $K = 1, 2, 4, 8$. For clarity, for $K = 1$ and $E[T_R]$, only the curve with $q_{\min} = 0$ is shown, while $K = 8$ (for $E[T_R]$) is not shown at all. It is noted that $E[T_R]$, as defined in (8.26), is a per-user delay: if per-(relaying) network delay was considered, curves $E[T_R]$ would simply scale down with factor K. The figure shows that there is a significant source's performance increase due to the slot auction. For example, the average number of (re)transmissions (throughput) can be reduced (increased) more than five times here (from $E[T_S] \approx 17$ with $K = 1$ or $K = 0$, to $E[T_S] \approx 3$ with $K = 8$ and $q_{k,\min} = 0$). The benefits for the relays are also meaningful, as they can maintain transmission with $E[T_R] \approx 2$ (when $K = 1$). With larger K, performance of each relaying node decreases as the competition is beneficial only for the source. Moreover, the source performs better as the relays reduces its expectation to be paid-back ($q_{\min}$), but the opposite is not necessarily true for the relays (in fact, large $q_{\min}$ degrades relays' performance when they are further from the AP), as discussed above. Notice that the auction benefit for the source is reduced if the relays are placed far away from the access point ($d \to 0$) as the channels gains h_{R_k} are relatively small, or far away from the source ($d \to 1$) as the relays can hardly decode the source's transmission. Similar conclusions hold true for the relays, although for $d \to 1$ they can still achieve a relatively good performance (due to the strong channel towards the AP). In particular, notice that the placement d where the best source's and relays' performance is achieved (i.e., where $\mathbb{E}[T_S]$ or $\mathbb{E}[T_R]$ is minimized) increases with the number of users K. This is the consequence of *multiuser diversity*, i.e., the probability that at least one of the relays will decode (recall Section 8.3.5) increases with K, thus providing it with an opportunity of exploiting large channel $|h_{R_k^*}|$ for transmission. Recall that the similar behavior was also observed in Chapter 7.

Comparison of Vickrey and first-price sealed-bid auction (as discussed in Section 8.4) is illustrated in Figure 8.8, in terms of the average number of transmission slots for the source $E[T_S]$ and the relays $E[T_R]$, versus the placement d, for $g_S = -7$ dB, $q_{\min} = 0.3$ and

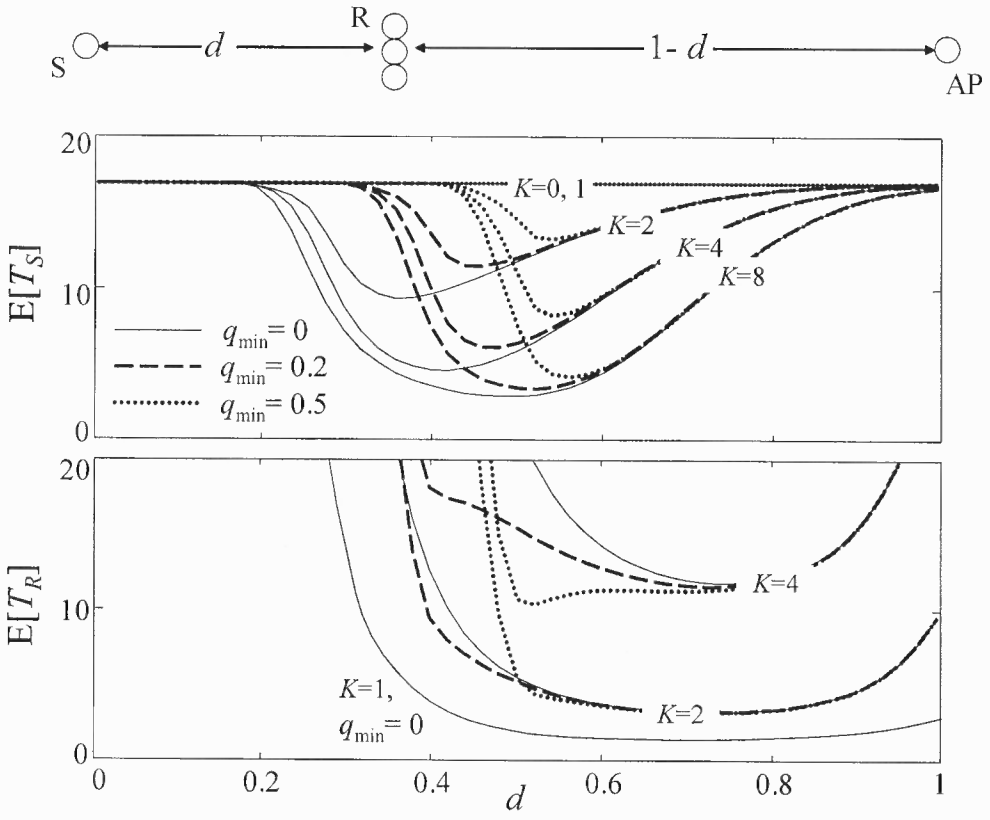

Figure 8.7 Expected number of transmission slots for the source and the relay, $E[T_S]$ (8.21) (upper figure) and $E[T_R]$ (8.26) (lower figure), respectively, versus the placement of the relays d, for $q_{\min} = 0, 0.2, 0.5$ ($g_S = -7$ dB and $K = 1, 2, 4, 8$).

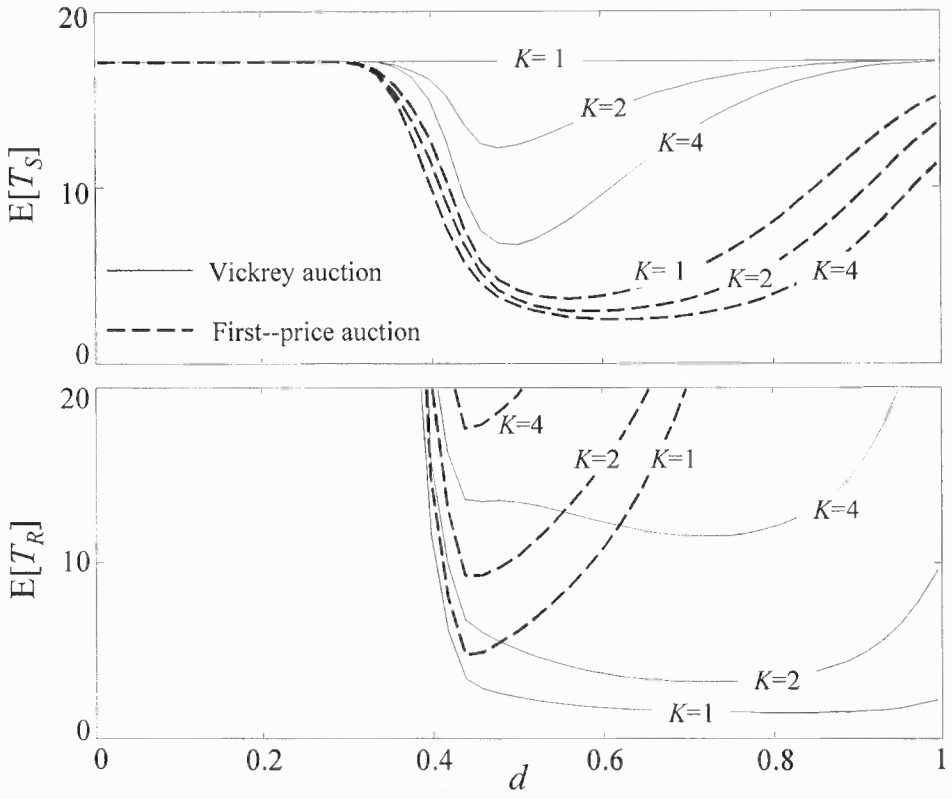

Figure 8.8 Expected number of transmission slots for the source and the relay, $E[T_S]$ (8.21) (upper figure) and $E[T_R]$ (8.26) (lower figure), respectively, versus the placement of the relays d, for Vickrey and first-price auction (as discussed in Section 8.4) ($g_S = -7$ dB, $q_{\min} = 0.2$ and $K = 1, 2, 4$).

$K = 1, 2, 4$. Notice that the Vickrey auction increases the benefits for the relays, while the first-price auction is preferred for the source. Nevertheless, it is emphasized that this comparison is not entirely fair, as in the first-price auction (recall utility (8.27)) relays have no incentive to achieve reliability larger than $q_{\min}$ (which would minimize $E[T_R]$) but only to gain the spectrum access for transmission of their data with reliability $q_{\min}$.

The largest benefit of the proposed model (and, in general, relay models) is when the source is in a deep fade and, therefore, finds the cooperation from the relays indispensable. To describe the performance of such a scenario, Figure 8.9 shows the average number of transmission slots for the source $E[T_S]$ and the relays $E[T_R]$, when g_S is in the range of

$g_S = -15$ dB $\div$ 5 dB, while the channel gains g_{SR} and g_R are fixed, $g_{SR} = g_R = 9$ dB, $q_{\min} = 0$ and $K = 1, 2, 4, 8$. This figure is also exploited for a brief investigation of the impact of the outdated channel knowledge, by considering $\rho = 0.1$ and $\rho = 0.5$. From the figure, it can be seen that for an improvement of the source's performance of about 2 dB, only two relays are required, while with $K = 8$ the source transmits with $\mathbb{E}[T_S] \approx 2$ even in the deep fade when $g_S = -13$ dB. The relaying users achieve the best performance when the source is in deep fade. Interestingly, increasing correlation ρ has a negative impact on the source's performance (at least for the $g_S < g_{RS}$, as in this example; recall Figure 8.5 and related discussion) because in this model the rate r_S for the source is fixed and the unavailability of any relays to cooperate with a source in deep fade does not change over time (for large ρ). As for the relays, since the channel g_R is relative large ($g_R = 9$ dB), the relays gain with the accurate channel knowledge.

8.6 Directions for Future Research and Open Issues

The proposed scheme opens the door to several possible future investigations. In addition to the application to cognitive radio, auction-based incentives can be further considered as a candidate for motivating node cooperation in multihop routing. A multi-auctioneer scenario where more than one source is considered or where a relay and a source can exchange their roles, is also an open issue to be investigated. To address the practical costs of the scheme, the impact of the signaling overhead, delay and energy consumed by receivers should be evaluated according to a signalling protocol. Auction vulnerability and its protection from the bidders' collusion are standard issues in auction-based mechanisms that need to be considered for this application. Moreover, the scheme is sensitive to other types of malicious behavior: for instance, due to statistical means of bidding, a winning relay can retransmit source's message with a smaller reliability than agreed, hoping that the consequent decoding failure will be (wrongly) prescribed to a bad channel realization.

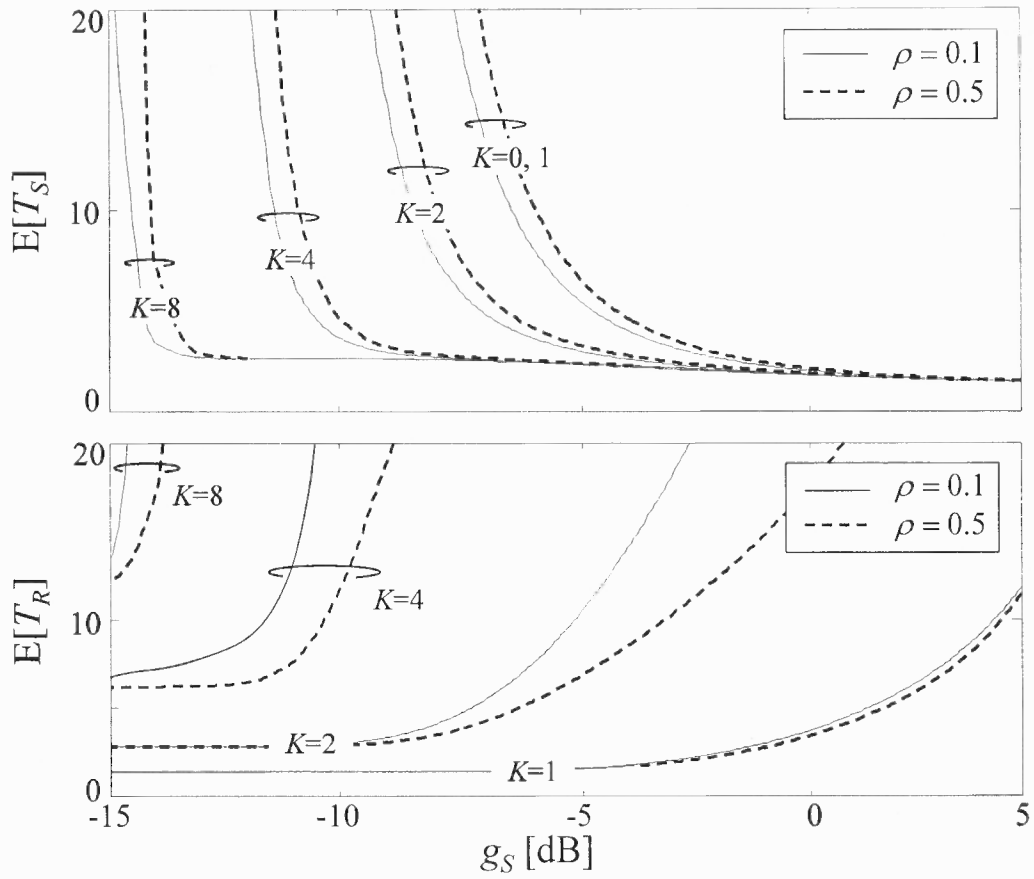

Figure 8.9 Average number of transmission slots for the source and the relay, $E[T_S]$ 8.21) (upper figure) and $E[T_R]$ (8.26) (lower figure), respectively, versus g_S, for $\rho = 0.1, 0.5$ ($g_{SR} = g_R = 9$ dB, $q_{\min} = 0$, $K = 1, 2, 4, 8$).

8.7 Concluding remarks

In this chapter, a novel scheme that leverages cooperative ARQ to opportunistically re-assign retransmission slots and exploits spectrum leasing to incentivize cooperation has been proposed. The strategy is fully decentralized and implemented through an auction-based mechanism. The results in this chapter (largely based on Vickrey class of auctions), besides motivating a novel cooperative ARQ scheme, can be seen as providing a promising solution for the implementation of property-rights cognitive radio networks. In fact, by exploiting the proposed spectrum leasing strategy, dynamic and opportunistic secondary spectrum access can be obtained via only local interactions at the medium access and physical layers. This is in contrast with other solutions that involve either long-term spectral allocations and/or pertain to the whole communication protocol stack. Numerical results show that, with the proposed solution, secondary (relaying) users can achieve excellent performance in the licensed spectrum, while simultaneously significantly improving performance of the primary (source) user.

CHAPTER 9

CONCLUSIONS

9.1 Contributions

Future wireless systems are expected to extensively rely on cooperation between terminals, mimicking MIMO scenarios when terminal dimensions limit implementation of multiple antenna technology. On this line, cooperative retransmission protocols are considered as particularly promising technology due to their opportunistic and flexible exploitation of both spatial and time diversity. In this dissertation, some of the major issues that hinder the practical implementation of this technology are identified and pertaining solutions are proposed and analyzed. Potentials of cooperative and cooperative retransmission protocols for a practical implementation of dynamic spectrum access paradigm have been also recognized and investigated. Detailed contributions follow:

- While conventionally regarded as energy efficient communications paradigms, both cooperative and retransmission concepts increase circuitry energy and may lead to energy overconsumption in, e.g., sensor networks. In this context, advantages of cooperative HARQ protocols have been reexamined and their limitation for short transmission ranges observed. An optimization effort was provided for extending an energy-efficient applicability of these protocols. Interesting optimization-related facts were also observed. For example. increasing the transmission energy can in fact reduce the overall energy consumption.

- Inherent assumption of altruistic relaying has always been a major stumbling block for implementation of cooperative technologies. In this dissertation, provision is made to alleviate this assumption and opportunistic mechanisms were designed that incentivize relaying via a spectrum leasing approach. Mechanisms were provided for both cooperative (Chapter 7) and cooperative retransmission protocols (Chapter 8).

It was noted that the scheme presented in Chapter 8 is not constrained by commonly encountered game-theoretic limitations - extensive informational requirements and is applicable to a wider and a more practicable range of scenarios. Both solution provide a meaningful upsurge of spectral efficiency for all involved nodes (source-destination link and the relays).

- It was recognized that the relaying-incentivizing schemes in Chapter 7 and Chapter 8 have an additional and certainly not less important application, that is in dynamic spectrum access for property-rights cognitive radio implementation. This solution avoids commons-model cognitive-radio strict sensing requirements problems and regulatory/taxonomy issues.

Besides the above list, additional contributions were made in the supporting research. In particular,

- Potentials of game-theoretic tools for networks with limited or no signalling were demonstrated in Chapter 5. It was shown that the interference resulting from transmissions of surrounding entities (networks, nodes or cells) can be exploited as an implicit signaling that enables sufficient information for a self-organizing network (or set of networks) to efficiently allocate the resource (i.e., avoid interference and/or allow for efficient reuse) based entirely on local transmission scheduling.

- It was shown in Chapter 6 that even after the deployment of simple, independent and selfish terminals, there exist sufficient means for a network to exercise an implicit control over them. Namely, it is on network to determine the networking parameters (e.g., bandwidth and level of macrodiversity) that can influence the nodes' behavior and thus intelligently perform their coordination. It is also the network's task to determine a satisfying trade-off between investing into resources and attaining a socially desirable behavior of independent terminals.

9.2 Future Work

It is the author's opinion that this dissertation opens the door for a plethora of possible future investigations. In particular, an interested reader/researcher is advised to focus on distributed networks and further investigate possibilities for dynamic spectrum access using cooperative (retransmission) technologies. Here several interesting directions are proposed.

- Noticing that schemes proposed in Chapter 7 and Chapter 8 consider only a single primary / source-destination link, it is a natural extension to investigate a multi-primary scenario. This, however, should not be only a formal extension. For example, consider multiple primary nodes participating in multiple access channel (MAC) transmission as in, e.g., uplink CDMA. Assume a power control scheme and notice that the benefits of employing incentivized secondary help are not solely applied to the primary making this choice (and the secondary node) but also to other primary nodes. As it is anticipated that the secondary avails good channel to access point (and towards the primary), the relaying is likely to require a smaller invested transmission power and thus yield decreased interference for other primary transmissions. Although a non-competitive modeling of primary nodes already promises an interesting scenario, competitive primary transmissions would most likely highly reward research efforts. Both non-competitive and competitive scenarios are currently under investigation.

- In the context of dynamic spectrum access, notice that the relaying nodes need not necessarily be modeled as secondary. In fact, they can assume roles of software-defined radios that are willing to opportunistically turn to a provider offering an alternative communication technology. In other words, the proposed scheme would serve as a convenient framework for modeling vertical handover (i.e., a Media-Independent-Handover) in PHY and MAC layers.

- A scheme where a relay and a source can exchange their roles (i.e., a scenario where nodes are of 'equal rights') is also an open issue to be investigated and would provide a more complete picture of the relaying-motivating proposals.

- Even though the two relaying-motivating (dynamic spectrum access) schemes introduced in Chapter 7 and Chapter 8 are founded on conceptually different approaches and using different metrics, their goals are ultimately very similar. It would be therefore very interesting to find a common thread that would enable their comparison.

- Multihop routing is an interesting paradigm for covering broad regions in ad-hoc, wireless or vehicular networks. Synthesis of the proposed relaying-motivating schemes with multihop routing appears as an indeed promising research topic.

- Recall that power-pricing was exploited in Chapter 7 for the power control game. In general, pricing is used in game-theoretic (or utility-theoretic) literature to improve performance of independent nodes. This opportunity is not heavily explored in this dissertation and, while it cannot bring conceptually novel conclusions, it is likely to yield an even improved scheme performance.

- Another aspect that was largely out of scope of this dissertation but that deserves to be addressed in more details refers to algorithms for reaching equilibria and their converging properties. While this is by no doubts a demanding task, it could stand for the most significant contribution to the material provided in this dissertation.

APPENDIX A

CARDINALITY OF SET A_J

Recall that the set $\mathcal{A}_j$ is defined as

$$\mathcal{A}_j = \{(a_1, .., a_k) \mid a_i = 1, 2, ..; a_1 + .. + a_k = j\}. \tag{A.1}$$

The cardinality of such a set is equal to the number of solutions $(a_1, .., a_k)$ of the following equality:

$$a_1 + ... + a_k = j, \tag{A.2}$$

where $a_1, .., a_k$ are positive integers. Notice that this problem is equivalent to placing $k-1$ objects of one type between j sequentially placed objects of another type. The number of solutions of such a problem and, therefore, the cardinality of set $\mathcal{A}_j$, is then

$$|\mathcal{A}_j| = \binom{j-1}{k-1}. \tag{A.3}$$

APPENDIX B

MODEL FORMULATION AS BAYESIAN GAME

Here a more formal mathematical formulation of the auction-theoretic model given in Section 8.3.2 is provided and the expression (8.7) for its DSE is proved. As the relays reasonably do not avail of the complete information on other relays (e.g., their number, current channel states, power and reliability constraints), the scenario can be conveniently modeled in the framework of theory of Bayesian games (games with incomplete information setting) [91] [92]. Noticing that the source needs not to be considered as a player in the game (namely, its role is only to inform the relays on the auction rules, its reserve price p_0 and the winning relay), the game at hand is 'played' by relays only and described by the tuple [92] $\Gamma = \left(\{\mathrm{R}_k\}_{k=1}^{n}, \{t_k\}_{k=1}^{n}, \{(\alpha_k, E_k)\}_{k=1}^{n}, (\breve{k}, \breve{p}_{\breve{k}}), \mathcal{M}_{p_0}, w_{k=1,..,n}, u_{k=1,..,n}\right)$, where

- $\{\mathrm{R}_k\}_{k=1}^{n}$ denotes the set of bidding players, i.e., the relays that have successfully decoded the source's transmission;
- $\{t_k\}$ is the set of possible types t_k that characterize the relay (player) R_k. The type is a collection of parameters defined in Section 8.2.4 $t_k = \left(q_{k,\min}, E_k^{\max}, C_{0,R_k}, g_{R_k}, \hat{h}_{R_k}\right)$;
- $\{(\alpha_k, E_k)\}$ is the strategy space for kth relay. For a given type t_k, a strategy pair (α_k, E_k) uniquely identifies the reliability p_k offered to the source as $p_k(\alpha_k, E_k; t_k)$ (details are provided in Section 8.3);
- $(\breve{k}, \breve{p}_{\breve{k}})$ is the outcome of the auction, identifying the winner $\breve{k}$ and the reliability $\breve{p}_{\breve{k}}$ to be provided by the winner. The auction may also result in no retransmitting user selected, which corresponds to the outcome $(0, p_0)$;

- $\mathcal{M}_{p_0}$ is the mechanism that maps the strategies $(\alpha_k, E_k)_{k=1}^n$ into the outcome $(\breve{k}, \breve{p}_{\breve{k}})$, depending on the reserve price p_0. For a Vickrey auction, it is given by (8.5).

- $w_k(t_{-k}|t_k)$ is conditional distribution describing the belief of the relay R_k about the types of all the other users $t_{-k} = (t_i)_{i=1;i\neq k}^n$. It will be shown that, as a consequence of the Vickrey auction setting (8.5), this will not play any role in the solution concept of interest here;

- $u_k(\{\alpha_i, E_i; t_i\}_{i=1}^n)$ is the utility reflecting the relay R_k's preferences and goals. It is given by (8.1) or, more conveniently:

$$u_k(\{\alpha_i, E_i; t_i\}_{i=1}^n) = (q_k(\alpha_k, E_k; t_k) - q_{k,\min}) \cdot \delta\left(\breve{k} - k\right). \tag{B.1}$$

The DSE of game Γ is given by the following statement. The game Γ admits a dominant strategy equilibrium (DSE) with strategy chosen by the relay R_k being given by (8.7). A sketch of the proof of this statement is provided, as it basically follows from the "truth telling" properties of Vickrey auctions discussed above. Firstly, from (8.1) notice that for any given relay's message reliability $q_k(\alpha_k, E_k; t_k) = q$ and type t_k, a weakly dominant strategy (as in (8.2)) is the one that maximizes $p_k(\alpha_k, E_k; t_k)$, $p = \max_{\alpha_k, E_k} p_k(\alpha_k, E_k; t_k)$ subject to $q_k(\alpha_k, E_k; t_k) = q$. The corresponding mapping is denoted as $f_k : q \rightarrow p$ and notice that, given the monotonicity properties of the reliability functions, this mapping is non-increasing in q. With this conclusion, one can argue that in a dominant strategy equilibrium it must hold that $q_k = q_{k,\min}$, applying arguments similar to those described in Section 8.3.1 and [84] for the conventional Vickrey auction (recall that herein, a Vickrey auction setting is specified by the mapping mechanism $\mathcal{M}_{p_0}$ as given in (8.5)). Namely, choosing $q_k < q_{k,\min}$ implies $p_k = f_k(q_k) \geq f_k(q_{k,\min})$ which increases the chance of winning the auction but only if yielding a non-positive utility (8.1) (recall (8.5)). On the other hand, the strategy that produces $q_k > q_{k,\min}$ implies $p_k = f_k(q_k) \leq f_k(q_{k,\min})$, reducing the chance of winning while not affecting the actual (second-price) reliability to

be provided to the source if winning the auction (recall (8.5)). Thus, the strategy for the relay R_k in a dominant strategy equilibrium (α_k^*, E_k^*) is provided through optimization (8.7) and the bid submitted to the source is $p_k^* = p_k(\alpha_k^*, E_k^*; t_k)$.

As expected, a weakly dominant strategy for the game Γ follows the intuition behind the solution of conventional Vickrey auctions - choosing the strategy (α_k^*, E_k^*) that yields $q_k(\alpha_k^*, E_k^*; t_k) = q_{k,\min}$ is in fact bidding with value on the margin of indifference (where the player's utility is zero), as in Section 8.3.2. It is reemphasized that, as in the Vickrey auction, the informational requirements are significantly reduced, as the dominant strategy does not require any information at a relay regarding other relays' types (i.e., w_k becomes superfluous).

APPENDIX C

EVALUATION OF $E[T_R]$

Here, the analysis is given for the average number of transmission slots $\mathrm{E}[T_R]$ required for successful delivery of a relay's, say R_k, message. In Figure C.1, the Markov process for the relay's message transmission is illustrated. There are $3K + 2$ states in the depicted structure, each specifying whether the transmission (if any) of R_k's data in the current slot was successful and the transmission to take place in the following slot:

- $\mathcal{A}_0$ - no successful transmission of R_k's data; the source (re)transmits in the following slot.
- $\mathcal{A}_{0,R}$ - successful transmission of R_k's data; the source transmits in the following slot.
- $\mathcal{A}_{\bar{n}}$ - no successful transmission of R_k's data; auctioned transmission with n bidders, with no participation of R_k, takes place in the following slot.
- $\mathcal{A}_n$ - no successful transmission of R_k's data; auctioned transmission with n bidders, with participation of R_k, takes place in the following slot.
- $\mathcal{A}_{n,R}$ - successful transmission of R_k's data; auctioned transmission with n bidders, with participation of R_k, takes place in the following slot.

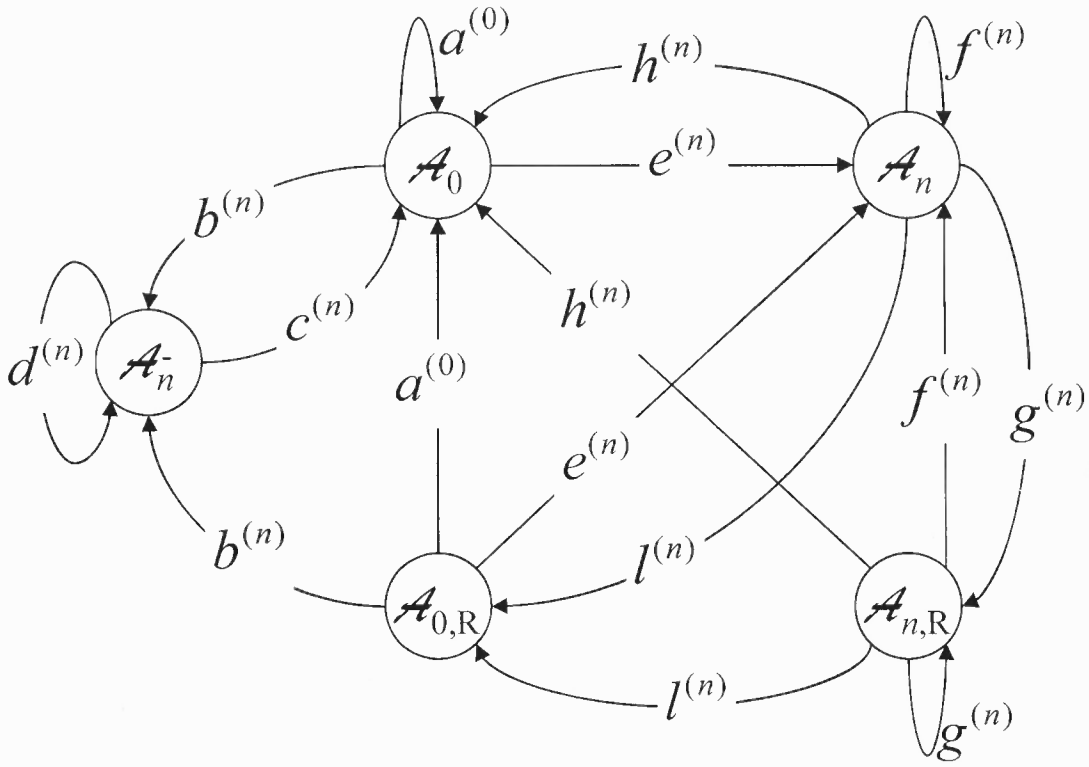

Figure C.1 Markov process for the relay's message transmission: state $\mathcal{A}_0$ - no successful transmission of R_k's data, the source (re)transmits in the following slot; state $\mathcal{A}_{0,R}$ - successful transmission of R_k's data, the source transmits in the following slot; state $\mathcal{A}_{\bar{n}}$ - no successful transmission of R_k's data, auctioned transmission with n bidders, with no participation of R_k, takes place in the following slot; state $\mathcal{A}_n$ - no successful transmission of R_k's data; auctioned transmission with n bidders, with participation of R_k, takes place in the following slot; state $\mathcal{A}_{n,R}$ - successful transmission of R_k's data; auctioned transmission with n bidders, with participation of R_k, takes place in the following slot ($a^{(0)}, b^{(n)}, .., l^{(n)}$ are transition probabilities).

Transition probabilities in Figure C.1 are given as:

$$
\begin{aligned}
a^{(0)} &= P_S^{(0)} + (1 - P_S^{(0)})P_d^{(0)}, \\
b^{(n)} &= \frac{K-n}{K}(1 - P_S^{(0)})P_d^{(n)}, \\
c^{(n)} &= P_S^{(n)}, \\
d^{(n)} &= 1 - P_S^{(n)}, \\
e^{(n)} &= \frac{n}{K}(1 - P_S^{(0)})P_d^{(n)}, \\
f^{(n)} &= \frac{n-1}{n}\left(1 - P_S^{(n)}\right) + \frac{1}{n}\mathrm{E}\left[1 - \max\left(\breve{p}_{\breve{k}|n}, \breve{q}_{\breve{k}|n}\right)\right], \\
g^{(n)} &= \frac{1}{n}\mathrm{E}\left[\left(\breve{q}_{\breve{k}|n} - \breve{p}_{\breve{k}|n}\right)^{+}\right], \\
h^{(n)} &= \frac{n-1}{n}P_S^{(n)} + \frac{1}{n}\mathrm{E}\left[\left(\breve{p}_{\breve{k}|n} - \breve{q}_{\breve{k}|n}\right)^{+}\right], \\
l^{(n)} &= \frac{1}{n}\mathrm{E}\left[\min\left(\breve{p}_{\breve{k}|n}, \breve{q}_{\breve{k}|n}\right)\right],
\end{aligned}
\tag{C.1}
$$

where $(x)^{+} = \max(0, x)$, the probabilities $P_S^{(n)}$ and $P_d^{(n)}$ are provided in Section 8.3.5 and the expectation $\mathrm{E}[\cdot]$ can be numerically calculated by averaging over the distribution of $|\hat{h}_S|$ and $|\hat{h}_{R_k}|$ (which are Rayleigh). Notice that in (C.1) it is emphasized that $\breve{p}_{\breve{k}}$ and $\breve{q}_{\breve{k}}$, given by (8.16) and (8.20) respectively, depend on the number of bidders n.

The expected number of slots $\mathrm{E}[T_R]$ is equivalent to the average recurrence of the states denoting the successful R_k's transmission. Denoting the average number of steps required to reach any of these states when starting from the state $\mathcal{A}_{n,R}$, for $n = 0, .., K$ (i.e., including the state $\mathcal{A}_{0,R}$), as $\nu_{n,R}$, $\mathrm{E}[T_R]$ reads:

$$
\mathrm{E}[T_R] = \sum_{n=0}^{K} \frac{\pi_{n,R}\nu_{n,R}}{\sum_{i=0}^{K}\pi_{i,R}}, \tag{C.2}
$$

where $\pi_{n,R}$ is the steady state probability of the state $\mathcal{A}_{n,R}$, while the term $\pi_{n,R}/\sum_{i=0}^{K}\pi_{i,R}$ is the normalized averaging parameter. In the following, a straightforward approach is provided to determine $\pi_{n,R}$ and $\nu_{n,R}$, $n = 0, .., K$, required to solve (C.2).

To determine $\nu_{n,R}$, the theory of Markov chains with rewards is applied and, e.g., [93] (Chapter 4) is followed to write the following system of equations:

$$\begin{aligned}
\nu_0 &= 1+\nu_0 a^{(0)}+\sum_{n=1}^{K}\nu_{\bar{n}}b^{(n)}+\sum_{n=1}^{K}\nu_n e^{(n)} \qquad \text{(C.3)}\\
\nu_{0,R} &= 1+\nu_0 a^{(0)}+\sum_{n=1}^{K}\nu_{\bar{n}}b^{(n)}+\sum_{n=1}^{K}\nu_n e^{(n)}\\
\nu_{\bar{n}} &= 1+\nu_{\bar{n}} d^{(n)}+\nu_0 c^{(n)},\ n=1,..,K\\
\nu_n &= 1+\nu_0 h^{(n)}+\nu_n f^{(n)},\ n=1,..,K,\\
\nu_{n,R} &= 1+\nu_0 h^{(n)}+\nu_n f^{(n)},\ n=1,..,K.
\end{aligned}$$

The set (C.3) yields the following solution:

$$\nu_{0,R} = \frac{1+\sum_{n=1}^{K}\frac{b^{(n)}}{1-d^{(n)}}+\sum_{n=1}^{K}\frac{e^{(n)}}{1-f^{(n)}}}{1-\sum_{n=1}^{K}\frac{b^{(n)}c^{(n)}}{1-d^{(n)}}-\sum_{n=1}^{K}\frac{e^{(n)}h^{(n)}}{1-f^{(n)}}-a^{(0)}}, \qquad \text{(C.4a)}$$

$$\nu_{n,R} = \frac{1+h^{(n)}\nu_{0,R}}{1-f^{(n)}},\ n=1,..,K. \qquad \text{(C.4b)}$$

For the steady state probabilities, applying a similar approach as in Section 8.3.5, the following relation is reached

$$\pi_{n,R} = \frac{g^{(n)}e^{(n)}}{(1-g^{(n)}-f^{(n)})\sum_{i=1}^{K}\left(\frac{l^{(i)}e^{(i)}(1-g^{(i)})}{1-g^{(i)}-f^{(i)}}+\frac{l^{(i)}b^{(i)}}{1-d^{(i)}}\right)}\pi_{0,R}. \qquad \text{(C.5)}$$

Applying (C.4) and (C.5) in (C.2), the solution for $\mathrm{E}[T_R]$ follows.

REFERENCES

[1] B. Zhao and M. C. Valenti, "Practical Relay Networks: A Generalization of Hybrid-ARQ," *IEEE Journ. Selected Areas Commun.*, vol. 23, no. 1, pp. 7-18, Jan. 2005.

[2] A. Sendonaris, E. Erkip and B. Aazhang, "User Cooperation Diversity. Part I: System Description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927-1938, Nov. 2003.

[3] J. N. Laneman, D. N. C. Tse and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[4] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic Transmission Scheduling with Resource-Sharing Constraints in Wireless Networks," *IEEE Journ. Selected Areas Commun.*, vol. 19, no. 10, pp. 2053-2064, Oct. 2001.

[5] A. Goldsmith, S. A. Jafar, N. Jindal and S. Vishwanath, "Capacity Limits of MIMO Channels," *IEEE Journ. Selected Areas Commun.*, vol. 21, no. 5, pp. 684-702, June 2003.

[6] H. El-Gamal, G. Caire and M. O. Damen, "The MIMO ARQ Channel: Diversity-Multiplexing-Delay Tradeoff," *IEEE Trans. Inf. Theory,* vol. 52, no. 8, pp. 3601-3621, Aug. 2006.

[7] P. Wu and N. Jindal, "Analysis of Fixed Outage Transmission Schemes: A Finer Look at the Full Multiplexing Point," in *Proc. IEEE ICC*, May 2008.

[8] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*, 1st ed. Prentice Hall, 1995.

[9] D. Chase, "Code Combining - A Maximum-Likelihood Decoding Approach for Combining an Arbitrary Number of Noisy Packets," *IEEE Trans. Commun.*, vol. COM-33, pp. 385-393, May 1985.

[10] M. Janani, A. Hedayat, T. E. Hunter and A. Nosratinia, "Coded Cooperation in Wireless Communications: Space-Time Transmission and Iterative Decoding," *IEEE Trans. Signal Processing*, vol. 50, pp. 362-371, Feb. 2004.

[11] G. Caire and D. Tuninetti, "The Throughput of Hybrid-ARQ Protocols for the Gaussian Collision Channel," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 1971-1988, July 2001.

[12] A. Paulraj, R. Nabar and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge Univ. Press, 2003.

[13] S. Cui, A. J. Goldsmith and A. Bahai, "Energy-Efficiency of MIMO and Cooperative Techniques in Sensor Networks," *IEEE Journ. Selected Areas Commun.*, vol. 22, no. 6, pp. 1089-1098, Aug. 2004.

[14] C. E. Jones, K. M. Sivalingam, P. Agrawal, and J. C. Chen, "A Survey of Energy Efficient Network Protocols for Wireless Networks," *ACM Wireless Networks*, vol. 7, no. 4, pp. 343-358, July 2001.

[15] O. Ileri, S. C. Mau and N. B. Mandayam, "Pricing for Enabling Forwarding in Self-Configuring Ad Hoc Networks," *IEEE Journ. Selected Areas Commun.*, vol. 23, no. 1, pp. 151-162, Jan. 2005.

[16] F. Milan, J. J. Jaramillo and R. Srikant, "Achieving Cooperation in Multihop Wireless Networks of Selfish Nodes," in *Proc. Workshop on Game Theory for Networks*, Oct. 2006.

[17] S. Zhong, J. Chen and Y. R. Yang, "Sprite: A Simple, Cheat-Proof, Credit-Based System for Mobile Ad-Hoc Networks," in *Proc. IEEE INFOCOM*, pp. 1987- 1997, Mar. 2003.

[18] A. Scaglione, Dl. L. Goeckel and J. N. Laneman, "Cooperative Communications in Mobile Ad Hoc Networks," *IEEE Signal Processing Mag.*, vol. 23, no. 5, pp. 18-29, Sep. 2006.

[19] C. Politis, T. Oda, S. Dixit, A. Schieder, K.-Y. Lach, M. Smirnov, S. Uskela and R. Tafazolli, "Cooperative Networks for the Future Wireless World," *IEEE Commun. Mag.*, vol.42, no.9, pp. 70-79, Sep. 2004.

[20] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE Journ. Selected Areas Commun.*, vol. 23, no. 2, pp. 201-220, Feb. 2005.

[21] http://www.ieee802.org/21/

[22] C. Cordeiro, K. Challapali, D. Birru and Sai Shankar N, "IEEE 802.22: The First Worldwide Wireless Standard Based on Cognitive Radio," in *Proc. IEEE DySPAN*, pp. 328-337, 2005.

[23] J. O. Neel, *Analysis and Design of Cognitive Radio Networks and Distributed Radio Resource Management Algorithms*, Ph.D. dissertation, Virginia Polytechnic Institute, September 2006.

[24] J. M. Peha, "Approaches to Spectrum Sharing," *IEEE Commun. Mag.*, vol 43, no. 2, pp. 10-12, Feb. 2005.

[25] G. Faulhaber and D. Farber, "Spectrum Management: Property Rights, Markets and the Commons," in *Proc. Telecommunications Policy Research Conference*, Oct. 2003.

[26] E. Biglieri, G. Caire and G. Taricco, "Limiting Performance of Block-Fading Channels with Multiple Antennas," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1273-1289, May 2001.

[27] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, 1972.

[28] G. Kramer, "Topics in Multi-User Information Theory," *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 4-5, pp. 265-444, 2007.

[29] M. Fossorier and S. Olcer, "Capacity Approaching Codes, Iterative Decoding Algorithms and Their Applications," *IEEE Commun. Mag.*, vol. 41, no. 8, pp. 100-140, Aug. 2003.

[30] A. J. Goldsmith and S. G. Chua, "Variable-Rate Variable-Power MQAM for Fading Channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218-1230, Oct. 1997.

[31] I. Stanojev, O. Simeone and Y. Bar-Ness, "Performance Analysis of Collaborative Hybrid-ARQ Incremental Redundancy Protocols over Fading Channels," in *Proc. of IEEE SPAWC*, July 2006.

[32] L. Ozarow, S. Shamai and A. D. Wyner, "Information Theoretic Considerations for Cellular Mobile Radio," *IEEE Trans. Veh. Tech.*, vol. 43, pp. 359-378, May 1994.

[33] D. Fudenberg and J. Tirole, *Game Theory*. The MIT Press, 2002.

[34] A. B. MacKenzie and L. A. DaSilva, *Game Theory for Wireless Engineers*. Morgan & Claypool, 2006.

[35] K. Leyton-Brown and Y. Shoham, *Essentials of Game Theory: A Concise, Multidisciplinary Introduction*, Morgan & Claypool Publishers, 2008.

[36] A. J. Goldsmith and S. B. Wicker, "Design Challenges for Energy-Constrained Ad Hoc Wireless Networks," *IEEE Wireless Commun. Mag.*, vol. 9, no. 4, pp. 8-27, Aug. 2002.

[37] M. Sikora, J. N. Laneman, M. Haenggi, D. J. Costello and T. E. Fuja, "Bandwidth- and Power-Efficient Routing in Linear Wireless Networks," *IEEE Trans. Inf. Theory,* vol. 52, no. 6, pp. 2624-2633, Jun. 2006.

[38] M. Sikora, J. N. Laneman, M. Haenggi, D. J. Costello and T. E. Fuja, "On the Optimum Number of Hops in Linear Wireless Networks," in *Proc. IEEE ITW*, pp. 165-169, 2004.

[39] Ö. Oyman and S. Sandhu, "A Shannon-Theoretic Perspective on Fading Multihop Networks," pp. 525-530, in *Proc. IEEE CISS*, 2006.

[40] T. Issariyakul and E. Hossain, "Performance Modeling and Analysis of a Class of ARQ Protocols in Multi-Hop Wireless Networks," *IEEE Trans. Wireless Commun.,* vol. 5, no. 12, pp. 3460-3468, Dec. 2006.

[41] I. Bettesh and S. Shamai, "Optimal Power and Rate Control for Minimal Average Delay: The Single-User Case," *IEEE Trans. Inf. Theory,* vol. 52, no. 9, pp. 4115-4141, Sep. 2006.

[42] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer Combining of Adaptive Modulation and Coding with Truncated ARQ over Wireless Links," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1746-1755, Sep. 2004.

[43] S. Cui, A. J. Goldsmith and A. Bahai, "Energy-Constrained Modulation Optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349-2360, Sep. 2005.

[44] E. Uysal-Biyikoglu, B. Prabhakar and A. El Gamal, "Energy-Efficient Packet Transmission over a Wireless Link," *IEEE Trans. Networking*, vol. 10, no. 4, pp. 487-499, Aug. 2002.

[45] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, "Energy-Aware Wireless Microsensor Networks," *IEEE Signal Processsing Mag.*, vol. 19, pp. 40-50, Mar. 2002.

[46] L. Zheng and D. N. C. Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple Antenna Channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073-1096, May 2003.

[47] T. E. Hunter, S. Sanaye and A. Nosratinia, "Outage Analysis of Coded Cooperation," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 375-391, Feb. 2006.

[48] J. N. Laneman and G. W. Wornell, "Distributed Space-Time Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415-2425, Oct. 2003.

[49] A. Høst-Madsen and J. Zhang, "Capacity Bounds and Power Allocation for Wireless Relay Channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2020-2040, June 2005.

[50] I. Stanojev, O. Simeone, Y. Bar-Ness and C. You, "Performance of Multi-Relay Collaborative Hybrid-ARQ Protocols over Fading Channels," *IEEE Commun. Lett.*, vol. 10, no. 7, pp. 522-524, July 2006.

[51] I. Kang, A. N. Willson, "Low-Power Viterbi Decoder for CDMA Mobile Terminals," *IEEE Journ. Solid-State Circuits*, vol. 33, no. 3, pp. 473-482, Mar. 1998.

[52] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11a, 1999.

[53] C.C. Lin, Y. H. Shih, H. C. Chang and C. Y. Lee, "Design of a Power-Reduction Viterbi Decoder for WLAN Applications," *IEEE Trans. Circuits Syst.*, vol. 52, no. 6, pp. 1148-1156, June 2005.

[54] *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)*, IEEE Std. 802.15.4, 2003.

[55] X. Liu, E. K. P. Chong and N. B. Shroff, "Opportunistic Transmission Scheduling with Resource-Sharing Constraints in Wireless Networks," *IEEE Journ. Selected Areas Commun.*, vol. 19, no. 10, pp. 2053-2064, Oct. 2001.

[56] H. Liu, and G. Li, *OFDM-Based Broadband Wireless Networks*, John Wiley & Sons, Inc, 2005.

[57] S. G. Kiani, G. E. Øien and D. Gesbert, "Maximizing Multi-Cell Capacity Using Distributed Power Allocation and Scheduling", in *Proc. IEEE WCNC,* Mar. 2007.

[58] K. Chawla and X. Qiu, "Quasi-Static Resource Allocation with Interference Avoidance for Fixed Wireless Systems," *IEEE Journ. Selected Areas Commun.*, vol. 17, no. 3, pp. 493-504, Mar. 1999.

[59] Y. Akaiwa and H. Andoh, "Channel Segregation - A Self-Organized Dynamic Channel Allocation Method: Application to TDMA/FDMA Microcellular System," *IEEE Journ. Selected Areas Commun.*, vol. 11, no. 6, pp. 949-954, Aug. 1993.

[60] D. D. Lin, R. A. Pacheco, T. J. Lim, and D. Hatzinakos, "Joint Estimation of Channel Response, Frequency Offset and Phase Noise in OFDM," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3542-3554, Sep. 2006.

[61] R. Etkin, A. Parekh and D. Tse, "Spectrum Sharing for Unlicensed Bands," in *Proc. Allerton Conference on Communication, Control and Computing, 2005.*

[62] C. U. Saraydar, N. B. Mandayam and D. J. Goodman, "Efficient Power Control Via Pricing in Wireless Data Networks," *IEEE Trans. Commun.*, vol. 50, pp. 291-303, Feb. 2002.

[63] N. Nie and C. Comaniciu, "Adaptive Channel Allocation Spectrum Etiquette for Cognitive Radio Networks," in *Proc. IEEE DySPAN*, November 2005.

[64] J. G. Proakis, *Digital Communications*, McGraw-Hill International Editions, 1995.

[65] R. M. Corless, G. H. Gonnet, D. E. Hare, D. Jeffrey and D. E. Knuth, "On the Lambert W Function," *Adv. in Computational Math.*, vol. 5, 1996.

[66] R. D. Yates, "A Framework for Uplink Power Control in Cellular Radio Systems", *IEEE Journ. Selected Areas Commun.*, vol. 13, pp. 1341-1347, Sep. 1995.

[67] D. P. Bertsekas and J. N. Tsitsiklis, *Nonlinear Programming*. Athena Scientific, 1997.

[68] G. Scutari, S. Barabarossa and D. Palomar, "Potential Games: A Framework for Vector Power Control Problems with Coupled Constraints," in *Proc. ICASSP* 2006.

[69] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computations: Numerical Methods*. Athena Scientific, 1999.

[70] A. Augustín, O. Mu noz and J. Vidal, "A Game Theoretic Approach for Cooperative MIMO Schemes with Cellular Reuse of the Relay Slot," in *Proc. ICASSP* 2004.

[71] G. Scutari, D. P. Palomar and S. Barbarossa, "Optimal Linear Precoding/Multiplexing for Wideband Multipoint-to-Multipoint Systems based on Game Theory-Part I: Nash Equilibria," *IEEE Trans. on Signal Processing*, vol. 56, no. 3, pp. 1230-1249, Mar. 2008.

[72] M. J. Osborne and A. Rubenstein, *A Course in Game Theory*, MIT Press, 1994.

[73] O. Simeone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini and R. Pickholtz, "Spectrum Leasing to Cooperating Secondary Ad Hoc Networks," *IEEE Journ. Selected Areas Commun.*, vol. 26, no. 1, pp. 203-213, Jan. 2008.

[74] X. Li, "Space-Time Coded Multi-Transmission Among Distributed Transmitters Without Perfect Synchronization," *IEEE Signal Processing Letters,* vol. 11, no. 12, pp. 948-951, Dec. 2004.

[75] J. Huang, R. A. Berry and M. L. Honig, "Distributed Interference Compensation for Wireless Networks," *IEEE Journ. Selected Areas Commun.*, vol. 24, no. 5, pp. 1074-1084, May 2006.

[76] L. Lai and H. El Gamal, "On Cooperation in Energy Efficient Wireless Networks: The Role of Altruistic Nodes," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1868-1878, May 2008.

[77] J. Huang, R. Berry, and M. L. Honig, "Auction-Based Spectrum Sharing," *ACM/Springer Mobile Networks and Apps.*, vol. 11, no. 3, pp. 405-418, June 2006.

[78] S. Gandhi, C. Buragohain, L. Cao, H. Zheng, and S. Suri, "A General Framework for Wireless Spectrum Auctions," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 22-33, Apr. 2007.

[79] S. Sengupta, M. Chatterjee, and S. Ganguly, "An Economic Framework for Spectrum Allocation and Service Pricing with Competitive Wireless Service Providers," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 89-98, Apr. 2007.

[80] J. Huang, Z. Han, M. Chiang and H. V. Poor, "Auction-Based Resource Allocation for Cooperative Communications," *IEEE Journ. Selected Areas Commun.*, vol. 26, no. 7, pp. 1226-1237, Sep. 2008.

[81] P. Klemperer, "Auction Theory: A Guide to the Literature," *J. Economics Surveys,* vol. 13, no. 3, pp. 227-286, July 1999.

[82] P. Xia, S. Zhou and G. B.Giannakisz, "Adaptive MIMO-OFDM Based on Partial Channel State Information," *IEEE Trans. Signal Processing*, vol.52, no.1, pp. 202-213, Jan. 2004.

[83] S. Ye, R. Blum, and L. Cimini, "Adaptive Modulation for Variable Rate OFDM Systems with Imperfect Channel Information," in *Proc. IEEE Vehicular Technology Conference*, 2002.

[84] W. Vickrey, "Counterspeculations, Auctions, and Competitive Sealed Tenders," *Journal of Finance*, vol. 16, no. 1, pp. 8-37, 1961.

[85] D. Lucking-Reiley, "Vickrey Auctions in Practice: From Nineteenth-Century Philately to Twenty-First-Century E-Commerce," *The Journal of Economic Perspectives*, vol. 14, no. 3, pp. 183-192, Summer 2000.

[86] M. Rothkopf, T. Teisberg, and E. Kahn, "Why are Vickrey Auctions Rare?" *Journal of Political Economy*, vol. 98, no. 1, pp. 94-109, Feb. 1990.

[87] A. Mukherjee and H. M. Kwon, "Robust Auction-Theoretic Partner Selection in Cooperative Diversity Wireless Networks," in *Proc. Allerton Conference on Signals, Systems, and Computers*, pp. 443-447, Nov. 2007.

[88] C. Demir and C. Comaniciu, "An Auction Based AODV Protocol for Mobile Ad Hoc Networks with Selfish Nodes," in *Proc. IEEE International Conference on Communications*, pp. 3351-3356, June 2007.

[89] J. Sun, E. Modiano and L. Zheng, "Wireless Channel Allocation Using an Auction Algorithm," *IEEE Journ. Selected Areas Commun.*, vol. 24, no. 5, pp. 1085-1096, May 2006.

[90] T. Sandholm, "Issues in Computational Vickrey Auctions," *International Journal of Electronic Commerce*, vol. 4 , no. 3, pp. 107-129, Mar. 2000.

[91] S. Adlakha, R. Johari, and A. Goldsmith, "Competition in Wireless Systems via Bayesian Interference Games," 2007. [Online]. Available at arXiv.org.

[92] S. Izmalkov, "Bayesian-Nash Games", [Online, MIT Open Courseware]. Available at ocw.mit.edu.

[93] R. Gallager, *Discrete Stochastic Processes*. Kluwer, 1996.

[94] I. Stanojev, O. Simeone and Y. Bar-Ness, "Optimal Design of a Multi-Antenna Access Point with Decentralized Power Control Using Game Theory," in *Proc. IEEE DySPAN* 2007.

[95] I. Stanojev, O. Simeone, Y. Bar-Ness and D. Kim, "Energy Efficiency of Non-Collaborative and Collaborative Hybrid-ARQ Protocols," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 326-335, Jan. 2009.

[96] V. Tarokh, H. Jafarkhani and A. R. Calderbank, "Space Time Block Codes from Orthogonal Designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456-1467, July 1999.

[97] S. M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE Journ. Selected Areas Commun.*, vol. 16, pp. 1451-1458, Oct. 1998.

[98] M. Dianati, X. Ling, K. Naik and X. Shen, "A Node-Cooperative ARQ Scheme for Wireless Ad Hoc Networks," *IEEE Trans. Veh. Technol.*, vol. 55, no. 3, pp. 1032-1044, May 2006.

[99] M. Levorato, S. Tomasin and M. Zorzi, "Cooperative Spatial Multiplexing for Ad Hoc Networks with Hybrid ARQ: System Design and Performance Analysis," *IEEE Trans. Commun.,* vol. 56, no. 9, pp. 1545-1555. Sep. 2008.

[100] D. Zhang, O. Ileri and N. B. Mandayam, "Bandwidth Exchange as an Incentive for Relaying," in *Proc. IEEE Conference on Information Sciences and Systems*, pp. 749-754, Mar. 2008.

[101] S. Verdu, *Multiuser Detection.* Cambridge University Press, 1998.

[102] M. Hayajneh and C. T. Abdallah, "Distributed Joint Rate and Power Control Game-Theoretic Algorithms for Wireless Data," *IEEE Commun. Lett.*, vol. 8, no. 8, pp. 511-513, Aug. 2004.

[103] C. A. St Jean and B. Jabbari, "Game-Theoretic Power Control in DS-CDMA Wireless Networks with Successive Interference Cancellation," *IEEE Electr. Lett.*, vol. 42, no. 3, Feb. 2006.