Adaptive and Efficient Resource Management for Emerging Wireless Networks

by

Ziaul Hasan Hashmi

B. Tech., Indian Institute of Technology, Kanpur, 2005

M. A. Sc., University of British Columbia, Vancouver, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES
(Electrical and Computer Engineering)

The University Of British Columbia
(Vancouver)

April 2013

Abstract

Recent unprecedented growth in the wireless market has forced network operators to find new techniques to reduce the operating costs, increase data rates, improve the spectrum utilization and reduce the energy consumption of various network elements. To overcome these challenges, opportunistic spectrum access through cognitive radios and relay-based cooperative communications have emerged as a new communication paradigm. However, to make these technologies practical, various resource management techniques must be optimized. Furthermore, we also need to explore the energy efficiency of these next generation wireless systems and identify key research issues and challenges in order to achieve sustainable "green" communication networks.

In this thesis, we design efficient resource allocation techniques for cognitive and cooperative networks and explore the energy efficiency of these systems. First, we study a capacity-maximizing power allocation problem in orthogonal frequency-division multiplexing (OFDM) based cognitive radio system while considering the allowed interference limits. We resort to an energy-aware capacity expression taking into account subcarrier availability and propose several suboptimal schemes that perform at a level close to that of optimal schemes. Second, we explore joint power and subcarrier allocation algorithms and fairness for OFDMA-based multiuser cooperative wireless systems. We challenge the traditional view of relaying algorithms by relaying only if it is beneficial. We state the problem in the form of a capacity-maximizing integer programming optimization problem and propose a heuristic solution. Next, we study the problem of relay selection in a cooperative network where regular mobile nodes could act as relays and cooperate if provided with incentives to do so. Using concepts from economics of asymmetric information, we propose incentive compatible schemes for relays and suggest a low-complexity heuristic relay selection scheme. Finally, we investigate the energy efficiency of the next generation wireless systems and present a short summary of methods and techniques to improve the energy efficiency of cellular networks. We also describe some important research issues and examine the major challenges to reduce the energy consumption of the cognitive and cooperative based emerging wireless networks. In concluding remarks, we also give some future research directions towards which the research in this thesis could lead.

Preface

The publications that has resulted because of the research conducted in this thesis are as follows:

- Ziaul Hasan, Gaurav Bansal, Ekram Hossain, and Vijay K. Bhargava,“Energy-Efficient Power Allocation in OFDM-Based Cognitive Radio Systems: A Risk-Return Model,” in *IEEE Transactions on Wireless Communications*, vol 8, no. 12, pp. 6078-6088, Dec. 2009 (appears in chapter 2).
- Ziaul Hasan, Ekram Hossain, Vijay K. Bhargava, “Resource Allocation for Multiuser OFDMA-based Amplify-and-Forward Relay Networks with Selective Relaying,” in *Proc. IEEE International Communications Conference (ICC) 2011*, Kyoto, Japan (appears in chapter 3).
- Olivier Duval, Ziaul Hasan, Ekram Hossain, and Vijay K. Bhargava, “Subcarrier Selection and Power Allocation for Amplify-and-Forward Relaying Over OFDM Links,” in *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1293-1297, April 2010 (related to results in chapter 3).
- Ziaul Hasan, Abbas Jamalipour, Vijay K. Bhargava, “Cooperative Communication and Relay Selection Under Asymmetric Information,” in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC), 2012*, pp.2373-2378, Paris, France (related to results in chapter 4).
- Ziaul Hasan, and Vijay K. Bhargava, “Relay Selection for OFDM Wireless Systems under Asymmetric Information: A Contract-Theory Based Approach,” submitted revision for publication in *IEEE Transactions on Wireless Communications*, Dec. 2012 (appears in chapter 4).
- Ziaul Hasan, Hamidreza Boostanimehr, and Vijay K. Bhargava, “Green Cellular Networks: A Survey, Some Research Issues and Challenges,” in *IEEE Communications Surveys & Tutorials*, vol.13, no.4, pp. 524-540, Fourth Quarter 2011 (appears in chapter 5).

In the research contributions made in Chapters 2, 3, 4, and 5, I am the primary author and researcher. I conducted the literature review and identified the research problems. Moreover, I independently formulated the research problems and carried out the mathematical analysis and simulations. I also wrote the associated manuscripts for publication. Chapter 2 is an extension of the work of my Masters thesis. Prof. Ekram Hossain co-authored the contributions made in Chapters 2 and 3 as he provided some important technical feedback during the formulation of the research problems in those chapters. He also provided editorial corrections while writing and revising the associated manuscripts for publication. Dr. Gaurav Bansal co-authored the contributions in Chapter 2 and he provided some editorial suggestions, comments and some corrections. While Dr. Olivier Duval is the primary author of the contributions associated with chapter 3, I identified and formulated the initial research problem and helped in the mathematical analysis and simulations. Dr. Duval helped with simulations, suggesting the heuristic solution and writing the associated publication. Prof. Abbas Jamalipour partially co-authored the contributions made in Chapter 4 since he provided some feedback during the formulation of the original research problem and gave important editorial suggestions. Hamidreza Boostanimehr co-authored the contributions in Chapter 5. He assisted with the literature survey and writing section 5.5 of the associated publication. He also helped in revising the paper and provided some critical feedback. My supervisor Prof. Vijay K. Bhargava co-authored the contributions made in chapters 2, 3, 4, and 5 of this thesis. I consulted him during the identification and formulation of the research problems. He provided editorial feedback for the associated manuscripts.

Table of Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, I would like to thank my supervisor, Prof. Vijay K. Bhargava, for his continuous guidance, encouragement and support throughout my graduate studies. I have been very fortunate to work at the Information Theory and Systems (ITS) laboratory under Prof. Bhargava's supervision. I would also like to thank Prof. Ekram Hossain for his mentorship, for contributing valuable insights to my thesis and for providing some very critical feedback. I am also indebted to Prof. Abbas Jamalipour for giving me an opportunity to work at his laboratory at University of Sydney which proved to be a very stimulating environment for me to carry out my research. A special thanks goes to Prof. Md. Jahangir Hossain for his excellent comments and feedback. I would also like to acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for supporting my graduate studies by awarding me the Alexander Graham Bell Canada Graduate Scholarship. Last but not the least, I would like to thank all my friends and family members for their continuous support that helped me overcome some difficult moments.

Chapter 1

Introduction

Recently, an unprecedented growth in the wireless communication market and a sudden proliferation of new high-end multimedia devices such as smartphones and tablets has led to an unanticipated increase in demand for radio spectrum and higher data rates. As more and more people are getting connected, information and communication technologies (ICTs) continue to spread throughout the world. According to a 2012 report by International Telecommunications Union (ITU), mobile-cellular subscriptions had reached almost 6 billion by the end of 2011, and most of this growth was seen in the developing world [1].

To meet this ever-growing demand of spectrum and data, several innovative technologies have been introduced to improve the overall spectral efficiency of the future wireless systems. Two such technologies, cognitive radio and cooperative systems, have emerged as a new communication paradigm for future wireless networks and have the capability to tackle issues of increasing spectral efficiency and improving overall data rates.

Moreover, since the growth in cellular communications also increases the overall energy consumption of wireless networks, addressing the energy efficiency of emerging wireless technologies is of immediate concern. Since cooperative and cognitive technologies are expected to be the backbone or the driving force behind the future wireless networks, making these technologies energy efficient is also an important area of research.

1.1 Background

In the following subsections, we will provide a brief introduction to cognitive radio and cooperative networks along with a short overview of a concept called *asymmetric information* that we will use in the latter part of our research.

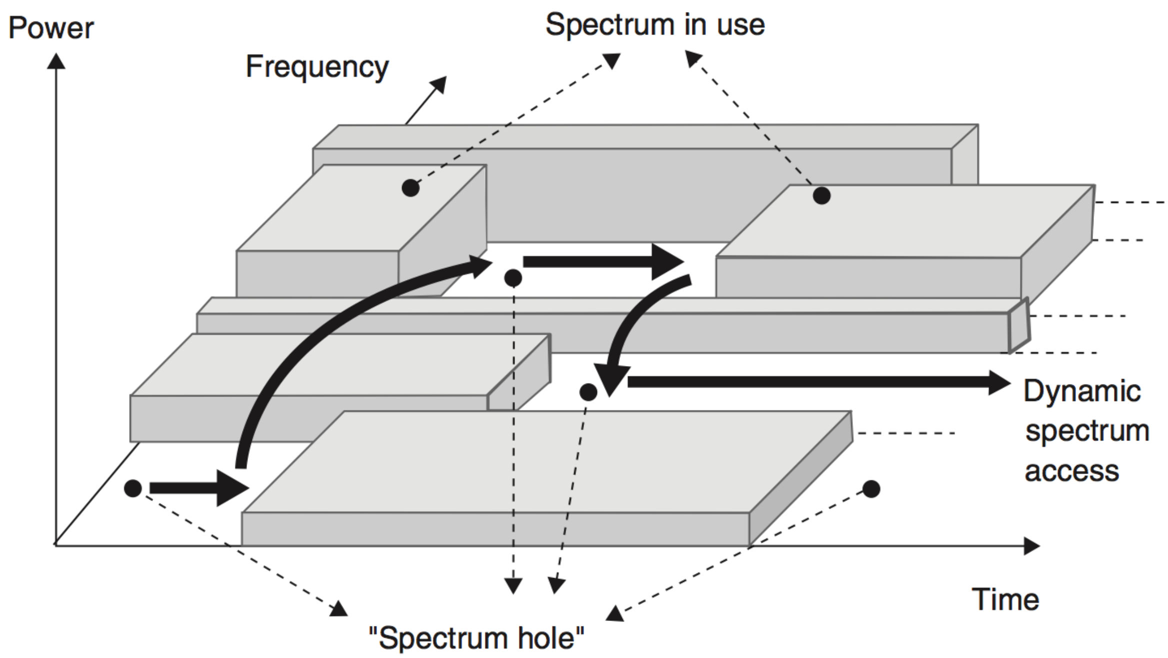


Figure 1.1: Spectrum holes generated during the normal usage of licensed frequency bands. Source:[2]

1.1.1 Cognitive Radio

Radio spectrum is a finite but a very valuable resource for wireless communications. Hence, efficient usage of this spectrum is of key importance for wireless system design. The legacy approach for the usage of spectrum so far has been to allocate specific regions of the radio spectrum for distinct services and also to assign the transmission license in distinct frequency bands exclusive to each operator. Although, such an approach avoids interference amongst different licensed users operating in the spectral vicinity and provides a good overall quality of service, unfortunately a lot of spectrum goes to waste when the licensed users are not using it to create "spectrum holes" as shown in Fig. 1.1. Moreover, initial investigations made by the spectrum-policy task force of the Federal Communications Commissions (FCC) have also indicated that most of the spectrum allocated to the licensed users is currently heavily underutilized [3]. Hence, new insights into the use of spectrum have challenged traditional approaches to spectrum management. This necessitates a new communication paradigm to harness the underutilized wireless spectrum by accessing it opportunistically. This new communication technology is referred to as dynamic spectrum access (DSA) or cognitive radio (CR). A CR system relies on opportunistic communication between unlicensed or secondary users (SUs) over temporarily unused spectral bands that are allocated to licensed or primary users (PUs). Hence, a CR system design involves smartly sensing unused spectrum and accordingly adapting to the transmission frequency, bandwidth, and time. Advancements in Software Defined Radio (SDR) technology has made a CR design possible by adjusting the transmission characteristics with the help of software on an embedded system [4]. CR networks have been seen as a promising solution to improve the current spectrum underutilization while simultaneously accommodating the increasing amount of services and applications in wireless networks [5, 6]. However, before CR networks are fully realized, there are three major research areas that must to be explored:

- *Spectrum sensing:* A CR network has to dynamically identify the parts of radio spectrum which are not used by primary users. This sensing, which is done in presence of noise, has to be extremely reliable and sensitive so that the opportunistic access does not cause any harmful interference to the PUs. A variety of techniques such as matched filter detection, energy detection, cyclostationary feature detection, and cooperative detection have beed studied in the literature [7, 8].

- *Spectrum management and sharing:* Once a CR network is effectively able to sense parts of the spectrum that are available for opportunistic access, the next crucial task is to determine which parts of spectrum to transmit and when. Spectrum management functions typically require a CR system to do the following: analyze and decide on the best spectrum available to meet the quality of service requirements, consider the activity or behaviour of primary users in that spectrum and to take into account sensing errors. Moreover, a CR system design must also incorporate techniques which effectively divide and share the spectrum among the secondary users [9].

- *Power Control:* Power control is used by CR systems to set the SNRs of the signals, which are sent in different frequency bands via opportunistic access, to such levels that this does not cause any interference to the nearby PUs. Hence, a CR system, hence has to ensure that the interference caused to these PUs remains below a certain threshold, which is also known as *interference temperature* [2].

1.1.2 Cooperative Communications

Conventional wireless systems are based on a point-to-point communication model that generally involves only two nodes, a source node and a destination node, which are responsible for the transmission of data. In such a communication model, fading and interference are major obstacles that limit the coverage and reliability of the wireless system and hence affect the overall spectral efficiency. Moreover, due to the limits on transmission power, as the transmission distance between the two nodes increases, it becomes increasingly difficult to maintain a desired quality of service. While spatial diversity techniques such as multiple-input multiple-output (MIMO) are fundamental

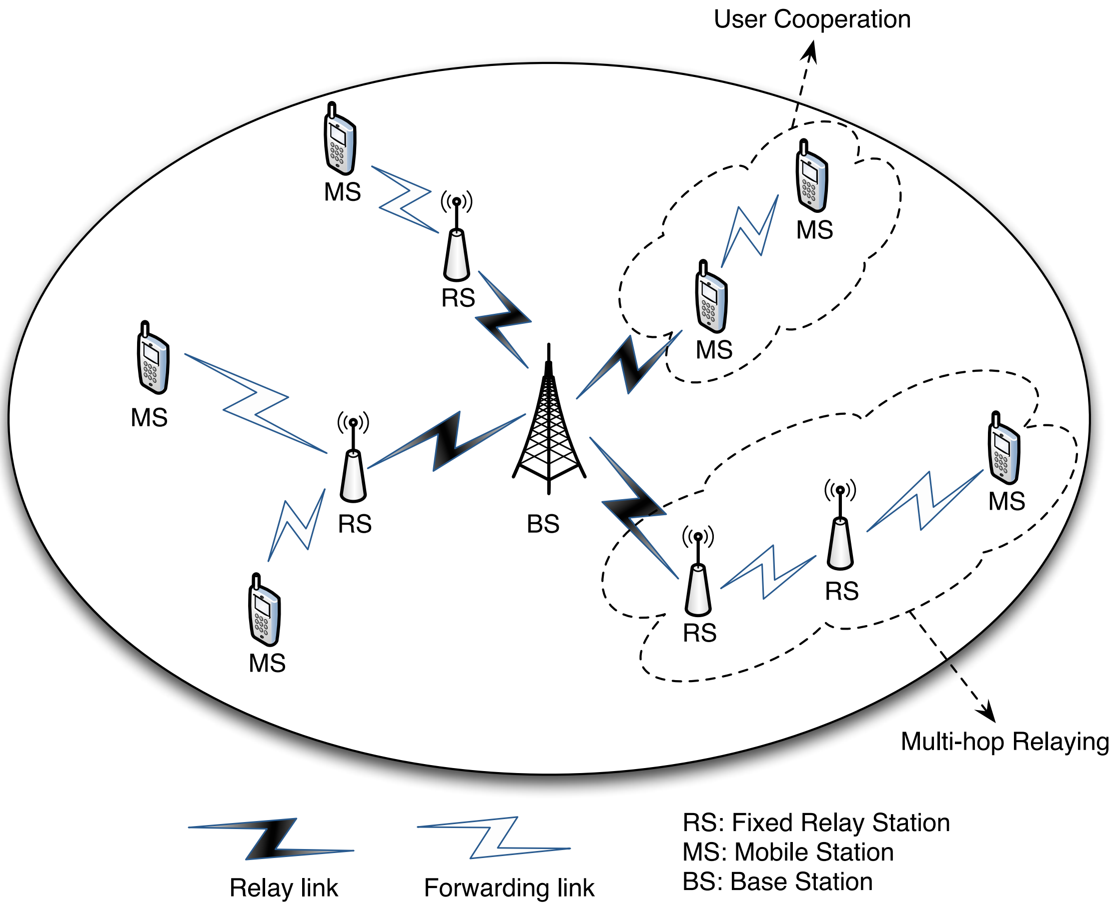


Figure 1.2: Downlink communication in a typical multi-hop relay network

in dealing with the fading effects, direct implementation of MIMO to each node might not be feasible due to cost, size and hardware restrictions. In order to overcome these challenges, cooperative communication techniques have evolved, which can provide potential gains in the overall system performance of wireless communication networks. In a cooperative network, there are some nodes that are used as relays to help transmit information between the source and the destination nodes. Cooperation in wireless networks is achieved either through the deployment of *fixed relays*, whose only purpose is to assist in information exchange between other nodes, or through *user cooperation* where regular mobile nodes can act as relays if needed and help facilitate communication for other nodes. A cooperative communication link could involve only one relay (single-hop) or a sequence of relays (multi-hop). A typical cooperative communication network is illustrated in Fig. 1.2. Relay nodes normally operate in a *half-duplex mode*, i.e., they cannot receive and transmit data in the same frequency band at the same time. Relaying is generally achieved by one of the two common signaling methods:

- *Amplify and Forward (AF):* In this signaling method, the relay first amplifies the received noisy signal by a certain gain and then retransmits it over the channel. AF relaying is simpler and cheaper to implement and performs better than direct transmission but also amplifies system noise during the retransmission phase.

- *Decode and Forward (DF):* DF relaying is a technique where the relay first decodes the received signal and then subsequently re-encodes and retransmits it. Compared to AF relaying, DF relaying is significantly more complex due to extra processing required yet it is more adaptable to channel conditions.

Relay-assisted cooperation is a promising technology that helps the following processes: load sharing across networks, reducing the transmission power, increasing the capacity, extending coverage, increasing energy efficiency and improving the spectral utilization [10]. Relay-based architectures have already been considered for next generation cellular standards such as 3GPP LTE-Advanced (LTE-A). However, cooperative networks still pose many challenges, especially in designing efficient resource allocation algorithms such as power allocation, multi-user communication and relay selection. User cooperation through proper incentives for relaying nodes is another challenging issue.

1.1.3 Asymmetric Information in Cognitive and Cooperative Networks

In economics, *information asymmetry* is a situation when one party has better information than the other. When the involved parties have differing goals, this imbalance in information can sometimes lead to poor transactions. Akerlof's seminal paper "The Market for Lemons" first brought these issues to the forefront of economic theory [11]. Examples of information asymmetry are widespread in everyday life, such as employees having better information about their efforts and abilities than their employers, customers knowing their personal preferences and taste better than retailers or service providers, governments knowing their own actions and expenditures better than the taxpayers. Information asymmetries are generally studied in the context of *principal-agent problems* and *contract theory* is a tool used to study how economic players can have contractual agreements in the presence of asymmetric information. More commonly, there are two economic agents in this model: the uninformed party, the principal that contracts the informed party which is the agent who possesses private information. The principal-agent model is a simplifying scheme that gives all the bargaining power to the principal, leaving the agent with only "yes or no" option. Hence, the objective of the uninformed party or the principal is to create incentive schemes that would tempt the agent to reveal its information or take desirable actions. Information asymmetry generally creates two kinds of incentive problems:

- *Hidden Information* (Adverse Selection): This is the scenario when the one party or agent has hidden characteristics or information that he/she may not reveal truthfully to the principal, that is the uninformed party. The uninformed party has imperfect information about the informed party's private information. The result of this asymmetry in information is adverse selection, which leads to undesirable results. A well-known example of adverse selection is insurance. If a company offers an insurance policy with premiums catered to an average-risk population, then this will attract more individuals from the high risk population and hence the company could possibly loose more money. A contract under these conditions creates incentives for agents so that they truthfully report their information. The principal instead offers a menu of contracts optimized for agents with different information, with the goal of separating these agents. This is called *screening* or self-selection. Signalling is another similar concept, where agents take initiative and signal their information to the principal.

- *Hidden Action* (Moral Hazard): In this scenario, the information asymmetry lies in the inability of principal or the uninformed party to directly observe or verify an agent's actions. Moral hazard is the situation under these circumstances where an agent or an informed party does not take full responsibility for its actions and hence has a tendency to act in such a way that may leave the uninformed party or the principal with the negative consequences of its own actions. Health insurance is a common example of moral hazard. An individual who is insured takes fewer precautions and indulges in activities that are dangerous to his heath because the insurer cannot observe his actions. A contract under these conditions creates incentives for agents to take the desired actions. The principal instead offers payments or transfers that depend on the outcome.

Unfortunately, due to inherent design features of cognitive and cooperative wireless networks, asymmetries of information are quite pervasive in both these systems. In cognitive radio, e.g., PUs are better informed with the channel gains on the secondary-primary link than the SUs hence they have better information about the interference generated by SUs. Moreover, some SUs may have better information on whether the PUs are currently occupying the channel or not. On the other hand, in cooperative networks, the relays are better informed about their power budgets and channel parameters than the source. Because of these asymmetries in information and the selfish nature of different wireless nodes, adverse selection and moral hazard problems naturally arise in resource allocation schemes for next generation wireless systems based on cognitive and cooperative technologies. Hence, *contract theory* is a viable tool to study and tackle such informational issues in advanced wireless networks.

1.2 Motivation, Scope and Objectives

The principal motivation for this research arises from the need to improve spectrum utilization and to increase the energy efficiency of the next generation wireless networks. Since cognitive and cooperative-based networks are prime candidates for future wireless communication systems, achieving the above-mentioned objectives for these networks require advanced radio resource management techniques. Orthogonal frequency-division multiplexing (OFDM) has been identified as key multi-carrier modulation technology that will be the driving force behind cognitive and cooperative wireless systems due to its great flexibility in dynamic allocation of unused spectrum to SUs in CR systems, ability to handle severe channel conditions and ease of designing adaptive and agile algorithms in relay-assisted networks. Since different OFDM subcarriers have different fading gains that are also time varying, we need to design adaptive resource allocation algorithms for wireless systems operating in specific scenarios. In this thesis, we study two kinds of resource allocation problems; one where the channel state information (CSI) of the system is available at the transmitter and another where the transmitter and other participating wireless nodes have asymmetric CSI.

First, we study resource allocation algorithms for cognitive wireless systems. Efficient and reliable subcarrier power allocation in OFDM-based CR networks is a challenging problem. When both SUs and PUs co-exist in the same geographic area or in close vicinity of each others' frequency bands, classical power allocation algorithms such as uniform power loading or water-filling algorithms generate unsatisfactory levels of mutual interference. Since there is a strict requirement on the interference generated to the PUs, these algorithms cannot be used for CR-OFDM systems. Moreover, one also needs to consider the effects of PU activity and spectrum sensing errors in opportunistic CR bands. Hence, we identify a problem of designing power allocation algorithms for OFDM-CR systems in order to maximize the transmission capacity of SUs while considering the above mentioned constraints and issues.

Next, we investigate resource allocation algorithms for cooperative wireless systems. Relay networks normally operate in a half-duplex mode where relay receives data in one time a lot and sends it to the destination in the second time slot. However, due to half-duplexity, the capacity is also reduced to half while providing diversity gain, which is particularly detrimental in scenarios where a direct path is stronger. Hence, relaying should be used only if it gives a gain in the capacity. Therefore, for OFDMA-based cooperative networks, we have to check on the subcarrier basis whether to relay or not. This motivates us to explore the joint problem of power and subcarrier allocation with selective relaying for OFDMA-based relay networks.

User cooperation, i.e., when regular mobile nodes assist as relays, is highly desirable for a cooperative network because it reduces the cost of adding infrastructure and network planning. However, other mobile nodes are selfish and would not cooperate at their own cost unless provided with some incentives to do so. This motivates us to study wireless networks with user cooperation. Relay selection is a difficult problem for such networks because the cost of cooperation for individual relays is hidden from nodes that require relaying due to asymmetric CSI. Studies in *contract theory* gives us some guidelines to solve this problem in this part of our research.

The last part of our research is about exploring the energy efficiency of cellular networks and investigating some techniques through which it can be improved. Since future wireless networks will be based on cognitive and cooperative systems, exploring the energy efficiency of such systems would lead directly to more efficient networks. Cognitive and cooperative wireless technologies are still in a nascent stage and including a "green" design in their development schemes can result in crucial gains in terms of reduction in energy consumption in the medium to long term.

To sum up, the main objective of this thesis is to develop various dynamic resource allocation schemes for OFDM-based cognitive and cooperative wireless networks thorough analytic modelling and numerical analysis. We look at these resource allocation problems in two different scenarios and provide some solutions to these challenging problems. In the first scenario specific nodes are well-informed of each other and in the second scenario some wireless nodes can be selfish users and need incentives. In addition, we also explore how cooperation and cognition could be used in next generation wireless systems in order to make these systems more energy efficient. The specific objectives of our research are described below.

1.2.1 Efficient Resource Allocation for OFDM-CR Systems

Since PU channels have to be utilized by SUs in a CR network without causing any degradation in service to the PUs, OFDM has been identified as a potential transmission technology for future CR systems [12]. This is mainly due to its flexibility in dynamically changing spectral environments and in allocating unused spectrum among SUs, which allows for simple adaptation of subcarriers to fast changing conditions in radio spectrum. Moreover, OFDM allows for multiuser diversity while overcoming frequency-selective fading which helps to enhance the overall spectrum utilization. A major challenge is to design efficient resource allocation algorithms (spectrum sharing and power allocation) that work well in OFDM-based CR networks. Subcarrier power allocation (or power loading) is a common technique to improve the system performance of OFDM systems by optimally allocating transmit power to the different subcarriers. If the channel state information is available at the transmitter, power loading can be used to optimize the error rate, the transmission capacity, or the transmit power.

The classical waterfilling approach [13], which states that the transmitter should avoid using subcarriers with poor channel conditions, has been considered as the optimal power allocation scheme for OFDM systems. A number of algorithms have been proposed in the literature for a single user case [14], [15], [16], [17], [18], [19]. A comprehensive survey on bit and power allocation algorithms for single user OFDM systems was presented in [20]. These algorithms, more generally known as bit and power loading schemes, are practical implementations of the classical waterfilling scheme. However, the traditional waterfilling approach is inefficient for OFDM-based cognitive radio networks due to the strict requirements on the interference generated to the primary users (PUs). Therefore, we need to design efficient resource allocation algorithms (spectrum sharing and power allocation) that also work well in OFDM-based CR networks which is the objective of this part of our research.

1.2.2 Resource Allocation for OFDMA Relay Networks with Selective Relaying

In recent years, resource allocation methods for OFDM without using cooperation have been studied thoroughly [21, 22]. Cooperative communication in the OFDM context has been shown to improve the achievable downlink communication rate between a base station and a destination node, using a single relay that performs relaying on all available subcarriers [23].

In cooperative relay networks, relaying is generally done in a half-duplex mode. By half-duplexity we mean that relay receives data in one time slot and sends it to the destination in the second time slot. In other words, relay doesn't send and receive during the same time slot. The common protocol used in such a system is second order diversity protocol, where source transmits in the first timeslot while both relay and destination listen and in the second timeslot, only the relay transmits and the destination listens [24]. Hence, the destination receives two copies of the original signal. With this protocol, although there is a gain in capacity due to diversity, the capacity is essentially reduced to half because of the half-duplexity. In the most of the relaying schemes proposed in the literature for cooperative OFDM systems, the source always sends data during first time slot and waits for the second time slot on all subcarriers. In this part of thesis, we challenge this scenario by letting the source choose which subcarriers to relay. We call this selective relaying. Specifically, for selective relaying in OFDM-based relay networks, we need to identify which subcarriers to use for relaying (i.e., subcarrier selection).

Several researchers have included selective relaying in their power allocation schemes. The research in [25] does select subchannels but the joint problem of subcarrier and power allocation in OFDMA was not treated. Subcarrier selection and bit-loading algorithms based on power gains, as presented in [26], do not consider the capacity loss when the base station does not transmit during the second timeslot. In [27], the authors used a selective subchannel relaying bit-loading scheme, but did not consider the problem of power allocation. This motivates us to tackle the problems of joint subcarrier and power allocation with selective relaying in OFDMA relay networks in this part of the thesis.

1.2.3 Relay Selection for OFDM Wireless Systems under Asymmetric Information

Cooperation via relays can be achieved either by installing fixed relays within the network coverage area or by making the other mobile nodes act as relays. The latter scenario, also known as user cooperation, is gaining attention, because of the minimal changes required in existing infrastructure and because it has been shown to not only increase the data rates but also to make the achievable rates less sensitive to channel variations [28, 29]. While user cooperation eliminates the cost of installing additional relay nodes, it increases the complexity of the overall system for several reasons. First, various dynamic resource allocation algorithms require near complete channel state information (CSI) from potential users assisting as relays. In the absence of this information, it is a challenge to design algorithms that dynamically select mobile users as potential relays [30]. Most relay selection algorithms for cooperative networks assume complete CSI [31, 32]. However, this information is private to mobile users and they may not be willing to share this information. This results in an asymmetry of available information between the source mobile user and the potential relays. Secondly, user cooperation poses a logistic challenge because the increased rate of one user comes at the expense of consumption of the limited resources of the relaying user (e.g. battery, power, bandwidth etc.). The potential relays are usually selfish nodes that could belong to different network entities/operators and hence may not be willing to cooperate without any additional incentives.

While relay selection with partial CSI have been explored by several authors e.g. [33, 34], no incentive-based mechanisms has been considered in these and other related works. To tackle this problem, game-theoretic models have generally been suggested for cooperative systems that are either reputation-based, resource exchange-based, or pricing based [35, references therein]. However, there are still many challenges in applying game theoretic solutions to cooperative systems including investigating the existence, uniqueness, computation and efficiency of the Nash Equilibrium, as well as addressing signalling overheads [35]. Moreover, to the best of our knowledge, the problem of relay selection under asymmetric information together with incentive-based mechanisms for multi-carrier systems such as OFDM, had not been studied. Therefore, the objective of this part of our research is to address this problem with simple pricing-based incentive mechanisms with minimal signalling overheads.

1.2.4 Green Cellular Networks

Energy efficiency in cellular networks is a growing concern for cellular operators not only to maintain profitability, but also to reduce the overall environment effects. The rising energy costs and carbon footprint of operating cellular networks have led to an emerging trend of addressing energy-efficiency amongst the network operators and regulatory bodies such as 3GPP and ITU. This is motivating the standardization authorities and network operators to continuously explore future technologies in order improve the entire network infrastructure. To this end, several world-wide initiatives such as "Energy Aware Radio and Network Technologies (EARTH)," "Towards Real Energy Efficiency Network Design (TREND)," "Cognitive Radio and Cooperative Strategies for Power Saving in Multi-Standard Wireless Devices (C2POWER)," "Optimizing Power Efficiency in Mobile Radio Networks (OPERA-NET)" are already under progress [36], [37], [38], [39].

So far, achieving high data rates has been the primary focus of research in emerging wireless systems, and not much consideration has been given to energy efficiency. However, many of these techniques significantly increase system complexity and energy consumption. Escalating energy costs and environmental concerns have already created an urgent need for more energy-efficient or "green" wireless communications. "Greening" wireless networks is a vast research discipline that seeks to cover all the layers of the protocol stack and various system architectures. In this process, it is important to identify the fundamental trade-offs linked with energy efficiency and the overall performance. Moreover, we need to design energy-efficient solutions for cooperative and cognitive networks, which will potentially drive the future generation of wireless communication. The objective of this part of our research is to study the energy efficiency of wireless systems in these contexts.

1.3 Thesis Outline

The rest of the thesis is organized as follows:

- In Chapter 2, we present a solution to a resource allocation problem which maximizes the cognitive radio (i.e., secondary) link capacity, while taking into account the PU activity or availability of the subcarriers (and hence the reliability of transmission by cognitive radios) and the limits on total interference generated to the PUs. We consider an energy-aware capacity expression by taking into account another factor called subcarrier availability. Optimizing such an expression saves valuable resources such as battery life by selectively allocating power to underutilized subcarriers. Based on a risk-return model, we formulate a convex optimization problem which incorporates a linear average rate loss function in the optimization objective to include the effect of subcarrier availability. Due to the complex structure of the optimal solution, we propose three suboptimal schemes, namely, the step-ladder, nulling and scaling schemes. We compare the performances of optimal and suboptimal algorithms with the performance of a classical waterfilling scheme. We conclude that waterfilling, which is unable to satisfy the interference criterion, performs the worst amongst all the schemes considered in this thesis.

- In Chapter 3, we address the problem of designing efficient resource allocation schemes for an OFDMA-based multiuser cooperative communication system that uses amplify and forward relaying. Using a two-phase relaying protocol, we assume that both the source and the relay have fixed power constraints and that the source employs a selective relaying mechanism. That is to say, the source adaptively decides on which frequency subcarriers relaying has to be performed. We first formulate the problem as a capacity maximizing integer programming optimization problem and then propose a heuristic solution that is composed of several steps. The first step is to suboptimally allocate subcarriers to different users based on proportional rate fairness. The next step is a two part iterative approach aimed at finding the relay decisions and power allocation at the source during the first phase and at the relay during the second phase. The last step is a waterfilling algorithm which allocates power at the source during the second phase to all the non-relaying subcarriers. Simulation results show that this selective relaying mechanism outperforms the "always relay" case while also providing an approximate proportional rate fairness.

- In Chapter 4, we use contract theory to tackle this problem of relay selection under asymmetric information in OFDM-based cooperative wireless systems that employs decode-and-forward (DF) relaying. We first design incentive compatible offers/contracts, consisting of a menu of payments and desired signal-to-noise-ratios (SNR)s at the destination. The source then broadcasts this menu to nearby mobile nodes. The nearby mobile nodes which are willing to relay, notify back the source with the contracts they agree to accept in each subcarrier. We show that when the source is under a budget constraint, the problem of relay selection in each subcarrier with the goal of maximizing capacity is a nonlinear non-separable knapsack problem. We propose a heuristic relay selection scheme to solve this problem. We compare the performance of our overall mechanism and the heuristic solution with a simple relay selection scheme. Selected numerical results show that our solution performs better and is close to optimal. The benefits of the overall mechanism introduced in this thesis is that it is simple to implement, needs limited interaction with potential relays and hence it requires minimal signalling overhead.

- In Chapter 5, we present techniques to enable green communications in future generation of wireless systems which will rely on cooperation and cognition to meet the increasing demand of high data rates. First, we present a brief survey of methods to improve the power efficiency of cellular networks. Next, we explore some research issues and challenges and suggest some techniques to enable an energy efficient or "green" cellular network. We focus on several important topics that are crucial towards reducing the energy consumption of the cellular networks. These topics include efficient base station redesign, heterogeneous network deployment, green communications via cognitive radio, cooperative relays to deliver green communications and energy efficient system design based on cooperation and cognition. Since base stations consume a maximum portion of the total energy used in a cellular system, we first provide a comprehensive survey on techniques to obtain energy savings in base stations. Next, we discuss how heterogenous network deployment based on micro, pico and femtocells can be used to achieve this goal. Since cognitive radio and cooperative relaying are undisputed future technologies, we propose a research vision to make these technologies more energy efficient. Lastly, we explore some broader perspectives in the realization of a "green" cellular network technology.

- Finally in chapter 6, we conclude this thesis by highlighting our contributions. We also briefly discuss ongoing work and suggest possible directions for future research.

Chapter 2

Energy Efficient Power Allocation in OFDM-Based Cognitive Radio Systems: A Risk-Return Model

2.1 Background

In this chapter, we specifically deal with the problem of power allocation for single user OFDM-CR system. Recently, authors in [40–42] observed that classical power allocation algorithms for OFDM systems such as uniform power loading or waterfilling are not optimal for CR networks because of the special properties of CR networks. Here, we consider the subcarrier power allocation problem for OFDM-based CR systems taking into account subcarrier availability, or in other terms, PU activity in the licensed bands. Including subcarrier availability in our capacity function and then optimizing this expression saves valuable resources such as battery life by selectively allocating lesser power to those bands which have higher PU activity. Hence, this is a more *energy-efficient* technique for power allocation in OFDM-CR networks. We approach this problem by first defining an average rate loss function and then introducing a risk-return model to incorporate subcarrier availability. This risk-return approach differs from traditional approaches in such as way that we could model the randomness in link capacity as a product of probability of sensing error/PU activity and average rate loss which is a function of allocated power in the corresponding subcarriers. Moreover, we also include the effect of the interference generated by both PUs and SU on each other.

There are two possible co-existence scenarios for the PUs. The first one proposed in [40] is an underlay scenario which considers the limit on the amount of interference created to a PU, which is occupying a particular set of subcarriers also used by the SU but is geographically located at a certain distance from the SU. This allows the SU to transmit within those set of subcarriers by keeping the transmit power level low enough to avoid unacceptable interference to the PUs that cannot be detected due to the large distance from SU. This requirement puts additional power constraint for OFDM-based CR system on each group of such subcarriers.

The second-coexistence scenario proposed in [41, 42] is for a downlink OFDM-CR network scenario in which PU bands are present in the close vicinity or adjacent to SU subcarriers, i.e., PU and SU are co-located in the same area with side-by-side bands. Since there is mutual interference between CR and PUs when both type of users co-exist in the side by side band, use of classical power loading schemes such as uniform loading and waterfilling may result in higher mutual interference in the PU bands. Optimal power allocation strategy in this CR system is to maximize the total transmission capacity of the SU while keeping the amount of interference generated to the PU bands within tolerable range.

Both these scenarios restrict the SU to keep its transmit power level low without causing any harmful interference to PUs. The authors in [40] modeled a CR network for the first co-existence scenario and proposed optimal and sub-optimal power loading schemes for this model. Similarly, in [41, 42], a power allocation algorithm was proposed for the second scenario. We extend our risk-return model to include these two PU co-existence scenarios and then propose a joint interference-limited model for OFDM-CR networks. We formulate a convex optimization problem for this model and then find the optimal solution for subcarrier power allocation. As we will notice, the optimal solution is no longer similar to waterfilling and has a very complex structure, so we suggest some viable sub-optimal schemes.

Key mathematical symbols that are used in this chapter are as follows: $[x]^+ = \max\{0, x\}$, $|S|$ gives the cardinality of set S, $\mathbb{Z}^+$ is the set of positive integers, $E\{X\}$ denotes the expectation of random variable X. The term ‘subcarrier’ has been used in this chapter for an orthogonal frequency band and the term ‘sub-band’ denotes a group a subcarriers occupied by a certain PU but opportunistically available to the SU in an underlay fashion.

The rest of the chapter is organized as follows. Section 2.2 presents the OFDM-based CR system considered in this chapter. The risk-return model and optimal power allocation among subcarriers for this model with a linear rate loss function is presented in Section 2.3. Section 2.4 presents the sub-optimal schemes for power allocation based on the risk-return model. Performance evaluation results are presented in Section 2.5. Section 2.6 states the conclusion.

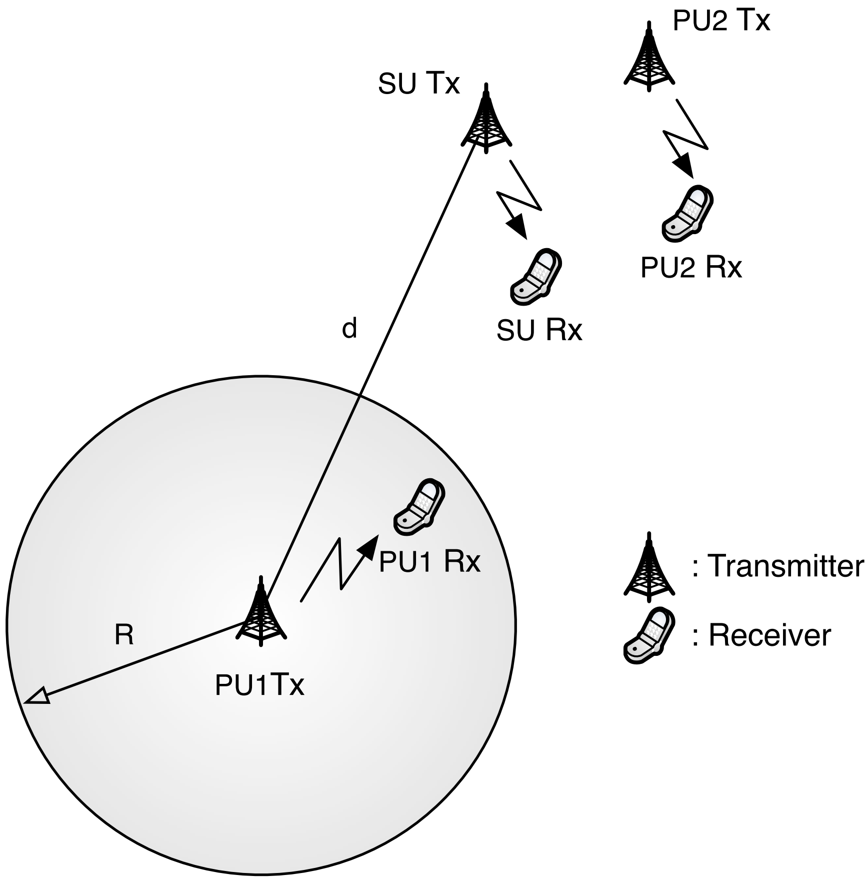


Figure 2.1: Cognitive radio system model.

2.2 System Model for Interference-Limited OFDM-CR System

2.2.1 OFDM-CR Model

We consider a typical cognitive radio wireless system as shown in Fig. 2.1 [40]. There is an SU transmitting data in underlay fashion in an opportunistically available sub-band licensed to a primary user PU1. This is the *first co-existence scenario* that we discussed earlier. A sub-band is said to be available if the interference caused to the PU1 receivers due to transmission using this subcarrier is within the acceptable range. PU1 is located at a certain distance geographically removed from SU, thus allowing the SU to transmit while keeping the interference level low enough. PU1 shields itself from SU interference by defining a protection area of radius $\mathcal{R}$ and adding a requirement on the SU to keep its interference power level at the margin of this area within a certain level $P^{(T)}$ [40]. Considering only distance-dependent path loss for simplicity, the SU's total

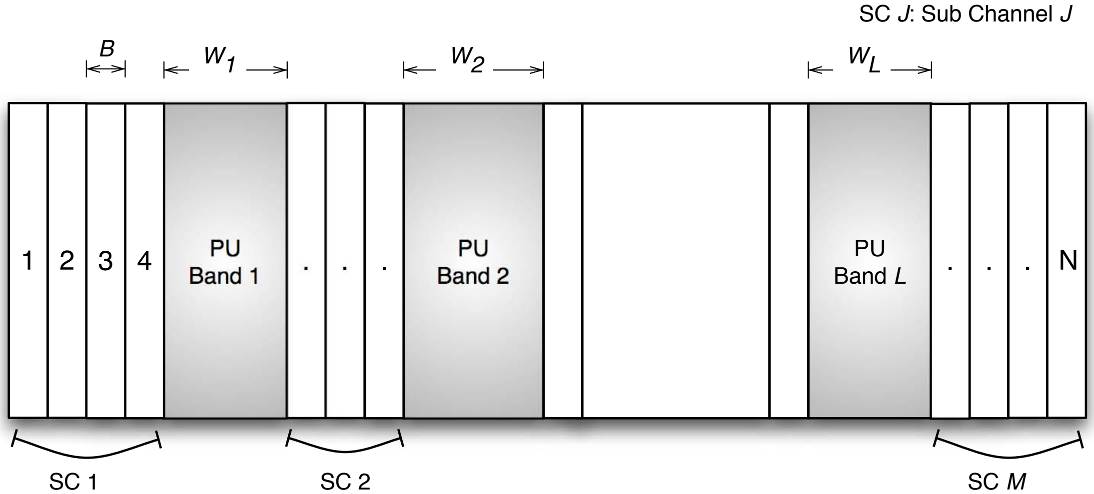


Figure 2.2: Opportunistic spectrum in an OFDM-based CR system with side by side PU bands.

transmission power S for this channel is constrained as follows:

$$S \leq P^{(T)}(d - \mathcal{R})^n \tag{2.1}$$

where d is the distance between SU transmitter and PU1 transmitter and n is the path-loss factor. We may not generally have any feedback channel between SU and PU transmitters. Therefore if we wish to include the effect of shadowing/fading in the above expression, we could keep an extra gap by reducing the overall power constraint on right side of (2.1) by a certain factor.

As shown in Fig. 2.1, there could be another primary user PU2 which is co-located in the same area as the SU but is using frequency bands which are adjacent or in between the available sub-bands. This is the *second co-existence scenario* that we mentioned earlier. To illustrate this in spectral domain, we consider a typical arrangement of the available subcarriers and PU bands as shown in Fig. 2.2 [42]. It is assumed that there are L frequency bands numbered from 1 to L with bandwidths $W_1, W_2, \cdots, W_L$, which are occupied by the co-located PU. We assume that the SU has the knowledge of position and relative distance of these PU bands to assess the interference created to these PU bands. In vicinity of these PU bands there are a total of N subcarriers available to the SU for opportunistic access using OFDM technology. These subcarriers are grouped into M sub-bands. Each sub-band is a group of subcarriers licensed to a PU which remains undetected in that sub-band and therefore allows the SU for spectrum underlay transmission. We further assume that the activities of the PUs in the corresponding sub-bands are uncorrelated. From now on we will use index i for subcarriers, index j for sub-bands and call PU$_j$ as the underlay scenario PU operating in jth sub-band.

Now let us define a mapping $\varphi : \mathbb{Z}^+ \to \mathbb{Z}^+$, where φ maps subcarrier index to the sub-band index as follows:

$$\varphi(i) = j \text{ if } i\text{th subcarrier belongs to sub-band } j. \tag{2.2}$$

We consider additive Gaussian noise with power spectral density N_0 as background noise and subcarrier bandwidth to be B Hz. The estimated channel gain between the transmitter-receiver pair of the SU for subcarrier i is denoted by h_i. We consider a slow channel fading such that channel gain remains constant during a power adaptation interval. We introduce another term, J_i which is the sum of the interference generated on ith subcarrier by the undetected PUs of the same sub-band and by the L side by side co-located PU bands. The interference created by these L PU bands can be estimated by integrating the power density spectrum of the PU signal across the ith subcarrier [42, 43]. If the power allocated to the subcarrier i is denoted by P_i then, the achievable rate r_i is given by

$$r_i = \eta \ln\left(1 + \frac{|h_i|^2 P_i}{N_0 B + J_i}\right) \tag{2.3}$$

where η is the fraction of the Shannon capacity which can be reliably guaranteed by the transmitter-receiver pair of the secondary user [44]. This factor η would generally depend on encoder-decoder pairs deployed by the SU but for simplicity we normalize η to unity. It has been inherently assumed in the above equation that interference introduced by the PUs to the SU is AWGN which is a reasonable approximation for sufficiently large number of PUs [42].

Now we discuss and define two different transmission power constraints on the SU's transmission band based on two possible co-existence scenarios of different PUs and a SU that we discussed so far: one in which a PU(s) is occupying the sub-band but is located at a certain distance allowing the SU to transmit within the same sub-band in an underlay fashion and the second scenario is the case when a PU is using an adjacent or side-by-side bands.

2.2.2 Sub-band Power Constraint

If a PU signal is detected in a sub-band, then all the subcarriers belonging to that sub-band are allocated zero power [40]. However, if the PU is not located close to the SU and remains undetected, then the SU can possibly transmit keeping the sum total of power allocated to all subcarriers for that sub-band within a limit as explained in the previous subsection by equation (2.1). Assuming T_j as the power constraint on jth sub-band, we can define T_j as follows:

$$T_j \triangleq \begin{cases} 0, & \text{if PU}_j \text{ is detected} \\ P_j^{(T)}(d_j - \mathcal{R}_j)^{n_j}, & \text{if PU}_j \text{ is not detected} \end{cases} \tag{2.4}$$

where $P_j^{(T)}$ is the interference power constraint set by PU$_j$ at the margin of the area whose radius is $\mathcal{R}_j$, d_j is the distance between SU's and PU$_j$'s transmitter and n_j is the corresponding path-loss factor. Therefore, if the total power allocated to subcarriers belonging to sub-band j is S_j, then $S_j \leq T_j$, where S_j is given by $S_j \triangleq \sum_{i \text{ s.t } \phi(i)=j} P_i$.

2.2.3 Interference Constraint Due to Co-located Primary User Bands

The signal in the SU subcarrier creates a spillover in the L frequency bands of the co-located PU. This spillover creates interference to the PU user and therefore SU must keep the interference within a prescribed threshold. The interference I_i^l introduced by the ith subcarrier on lth PU band can be expressed as [41–43]: $I_i^l = P_i \int_{d_{il}-W_l/2}^{d_{il}+W_l/2} \psi(f)df$, where $\psi(f)$ is a function in frequency f which depends on symbol duration and fading gain between CR transmitter and PU receiver, d_{il} represents the spectral distance between ith subcarrier and lth PU band. Therefore, those subcarriers which are right adjacent to the PU bands will cause higher amount of interference. Interference I_i^l can now be represented as $I_i^l = k_i^l P_i$, where parameter k_i^l depends on interrelated parameters between ith subcarrier and lth PU band [42]. If I_{th} is the total interference threshold prescribed by the L PU bands, then $\sum_{i=1}^{N}\sum_{l=1}^{L} I_i^l = \sum_{i=1}^{N}\sum_{l=1}^{L} k_i^l P_i \leq I_{th}$, which can be further simplified to $\sum_{i=1}^{N} K_i P_i \leq I_{th}$, where $K_i = \sum_{l=1}^{L} k_i^l$.

2.3 Energy Efficient Power allocation for OFDM-CR System

2.3.1 Risk-Return Model for Subcarrier Power Allocation

As we discussed earlier, including subcarrier availability in the capacity expression will give us energy aware resource allocation solutions. In order to incorporate the availability (or reliability) of subcarriers in our power allocation model, we refer to a general risk-return scenario. We treat the power allocated to a subcarrier as an investment in the band. A CR scenario could be thought as a risky environment compared to a non-CR scenario which could be treated as a risk-free environment. In a CR environment, loss of useful power can be represented as a rate loss whenever a PU reoccupies the channel or when there is an error in correctly sensing the channel.

Assuming the rate loss to be Δr_i for ith subcarrier, true rate R_i that can be achieved in the corresponding subcarrier is given by $r_i - \Delta r_i$. We define a real-valued increasing, concave and normalized average rate loss function $\mathcal{L}(P)$ for a power P invested in a subcarrier whenever a PU occupies a sub-band during the current time interval. This function $\mathcal{L}(P)$ may also involve cost of allocating resources (e.g., computational costs). We now define a joint probability α_j such that either the PU reoccupies jth sub-band during the current time interval or there is a sensing error in detecting the jth sub-band as available while it is actually not. Therefore, given this probability

$\alpha_{\varphi(i)}$ and the power invested P_i in the ith subcarrier, the expected rate loss or risk involved for that subcarrier is given by: $E\{\Delta r_i\} = \alpha_{\varphi(i)}\mathcal{L}(P_i)$. Hence, the expected capacity of the ith subcarrier in the cognitive environment is as follows:

$$E\{R_i\} = \ln\left(1 + \frac{|h_i|^2 P_i}{N_0 B + J_i}\right) - \alpha_{\varphi(i)}\mathcal{L}(P_i). \tag{2.5}$$

Therefore, the optimization problem can be written as follows: $\max_{P_1,P_2,\cdots,P_N} \sum_{i=1}^{N} E\{R_i\}$ with a constraint on power budget and interference generated to primary users. For $E\{R_i\}$ to be a concave function of power P_i, so that we could use standard convex optimization techniques, $\frac{\partial^2 E\{R_i\}}{\partial P_i^2} < 0$ for all subcarriers. This relation gives the following condition on $\mathcal{L}$ for the problem to be easily tractable using convex optimization approach:

$$\frac{\partial^2 \mathcal{L}}{\partial P_i^2} > -\frac{1}{\alpha_{\varphi(i)}\left(\frac{N_0 B + J_i}{|h_i|^2} + P_i\right)^2}, \quad \forall i \in \{1, 2, \cdots, N\}. \tag{2.6}$$

If we need $E\{R_i\}$ to be positive, increasing function of power within power budget P_{total} and interference limits I_{th} and $T_j, \forall j$, i.e. $\frac{\partial E\{R_i\}}{\partial P_i} \geq 0$, it gives

$$\frac{\partial \mathcal{L}}{\partial P_i} \leq \frac{1}{\alpha_{\varphi(i)}\left(\frac{N_0 B + J_i}{|h_i|^2} + P_i\right)}, \quad \forall i \in \{1, 2, \cdots, N\}. \tag{2.7}$$

It would be a reasonable assumption to make, that if zero power is invested, it gives no loss, i.e., $\mathcal{L}(0) = 0$. We will now investigate a practical choice for this loss function.

The energy efficient expected capacity expression (2.5) can also be viewed as a direct utility function with the second term as the cost of transmission. The cost of transmission is included in a utility function with price of spending power as binary variable in [45] and as an inverse factor in [46]. However a linear cost function in terms of power has been suggested as a more practical choice in [47–49]. The authors in [47] present a net utility function which is the difference of a strictly concave logarithmic rate function based on signal to interference noise ratio and a linear cost term on users' power. Again this approach is extended to multiple antenna channels in [48]. Similarly, energy aware utility regions were considered recently in [49] and the authors considered a utility function which is the difference of the capacity and a weighted power cost term where the weights corresponds to battery or QoS-requirement information.

The above discussion gives us the motivation to consider a linear late loss function. Therefore, in the next section, we will reformulate our optimization problem for a linear rate loss function which can also be viewed as a linear cost function. Moreover, a linear function satisfies (2.6) as in this case $\frac{\partial^2 \mathcal{L}}{\partial P_i^2} = 0$ for all subcarriers.

2.3.2 Optimal Power Allocation for an Interference-Limited OFDM-CR System with Linear Rate Loss

A linear rate loss function could be represented as follows: $\mathcal{L}(P) = \mathcal{C}P$, where $\mathcal{C}$ is normalized average cost per unit power for the secondary user to allocate resources and has units bits/s/W. Following a similar contention given by the authors in [48], we could say that the second term of the expected rate function penalizes the use of power in each sub-band. The term $\alpha_j\mathcal{C}$ can be viewed as a parameter specific to jth sub-band, and which enables to take into account if a particular opportunistic sub-band is rate-sensitive (small $\alpha_j\mathcal{C}$) or power-sensitive (large $\alpha_j\mathcal{C}$). A power-sensitive band would be the one with higher probability of sensing error or band occupancy causing additional transmission costs. Therefore data rates will be adapted with respect to the sensitivity to transmission conditions like channel quality, interferences etc. as well as to the transmission costs. This new criterion makes more sense in modern mobile networks in which battery lifetime remains an important parameter [48].

The cost parameter $\mathcal{C}$ is chosen such that it satisfies (2.7) for all subcarriers. In terms of the power allocation vector $\mathbf{P} = [P_1, P_2, \cdots, P_N]$, the optimization problem for this interference-limited model can be written as follows:

$$\mathbf{P}^* = \arg\max_{\mathbf{P}} \sum_{i=1}^{N} \left[\ln\left(1 + \frac{|h_i|^2 P_i}{N_0 B + J_i}\right) - \alpha_{\varphi(i)}\mathcal{C}P_i \right] \tag{2.8}$$

subject to

$$P_i \geq 0, \quad \forall i \in \{1, 2, \cdots, N\} \tag{2.9}$$

$$\sum_{j=1}^{M} S_j = \sum_{i=1}^{N} P_i \leq P_{total} \tag{2.10}$$

$$S_j \leq T_j, \quad \forall j \in \{1, 2, \cdots, M\} \tag{2.11}$$

and

$$\sum_{i=1}^{N} K_i P_i \leq I_{th}. \tag{2.12}$$

Proposition 1. There is a power allocation vector $\mathbf{P}$ which is the solution to the optimization problems defined in (2.8)-(2.12) and it is of the form

$$P_i^* = \left[w_i - \frac{N_0 B + J_i}{|h_i|^2} \right]^+, \forall i \in \{1, 2, \cdots, N\} \tag{2.13}$$

where w_i is the subcarrier threshold level, determined as follows:

Assume $A \triangleq \{j|S_j < T_j\}$, $B \triangleq \{j|S_j = T_j\}$.

1: Assign

$$
\begin{aligned}
w_i &= \frac{1}{\alpha_{\varphi(i)}\mathcal{C}}, \\
&\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in A && (2.14) \\
w_i &= \frac{1}{\gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C}}, \\
&\forall i \in \{1, 2, ..., N\} \text{ such that } \varphi(i) \in B && (2.15)
\end{aligned}
$$

where $\gamma_j \geq 0, \forall j \in B$ is determined by the following $|B|$ equations:

$$\sum_{i \text{ s.t. } \phi(i)=j} P_i^* = T_j, \quad \forall j \in B. \tag{2.16}$$

If the solution to above equations $P_i^* \geq 0$ exists and satisfies $\sum_{i=1}^{N} P_i^* < P_{total}$ and $\sum_{i=1}^{N} K_i P_i^* < I_{th}$, then it is optimal.

2: Assign

$$
\begin{aligned}
w_i &= \frac{1}{\lambda + \alpha_{\varphi(i)}\mathcal{C}}, \\
&\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in A && (2.17) \\
w_i &= \frac{1}{\lambda + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C}}, \\
&\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in B && (2.18)
\end{aligned}
$$

where $\lambda > 0$ and $\gamma_j > 0, \forall j \in B$ are determined by the following $|B| + 1$ equations:

$$\sum_{i=1}^{N} P_i^* = P_{total} \tag{2.19}$$

$$\sum_{i \text{ s.t } \phi(i)=j} P_i^* = T_j, \quad \forall j \in B. \tag{2.20}$$

If the solution to above equations $P_i^* \geq 0$ exists and satisfies $\sum_{i=1}^{N} K_i P_i^* < I_{th}$, then it is optimal.

3: Assign

$$
\begin{aligned}
w_i &= \frac{1}{\delta K_i + \alpha_{\varphi(i)}\mathcal{C}}, \\
&\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in A && (2.21) \\
w_i &= \frac{1}{\delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C}}, \\
&\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in B && (2.22)
\end{aligned}
$$

where $\delta > 0$ and $\gamma_j > 0, \forall j \in B$ are determined by the following $|B| + 1$ equations:

$$
\sum_{i=1}^{N} K_i P_i^* = I_{th} \tag{2.23}
$$

$$
\sum_{i \text{ s.t } \phi(i)=j} P_i^* = T_j, \quad \forall j \in B. \tag{2.24}
$$

If the solution to above equations $P_i^* \geq 0$ exists and satisfies $\sum_{i=1}^{N} P_i^* < P_{total}$, then it is optimal.

4: Assign

$$
\begin{aligned}
w_i &= \frac{1}{\lambda + \delta K_i + \alpha_{\varphi(i)}\mathcal{C}}, \\
&\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in A && (2.25) \\
w_i &= \frac{1}{\lambda + \delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C}}, \\
&\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in B && (2.26)
\end{aligned}
$$

where $\lambda > 0$, $\delta > 0$ and $\gamma_j > 0, \forall j \in B$ are determined by the following $|B| + 2$ equations

$$
\sum_{i=1}^{N} P_i^* = P_{total} \tag{2.27}
$$

$$
\sum_{i=1}^{N} K_i P_i^* = I_{th} \tag{2.28}
$$

$$
\sum_{i \text{ s.t. } \phi(i)=j} P_i^* = T_j, \quad \forall j \in B. \tag{2.29}
$$

If the solution to above equations $P_i^* \geq 0$ exists, then it is optimal.

Proof. See Appendix A. □

According to Proposition 1, the optimal power loading solution divides the sub-bands in two different sets: set A for which the total allocated power is smaller than the corresponding sub-band power constraint and set B, where the total allocated power is equal to the corresponding per sub-band power constraint. Also, Proposition 1 states that there could be four different scenarios under which power is allocated to different subcarriers. For cases 1 and 2, as long as we are strictly below the threshold limit I_{th} for some P_{total}, the power allocation is quite similar to sub-band waterfilling with a common water level for sub-bands in set A and different water levels for different sub-bands in set B. The water level (or threshold level) is case 1 is the highest compared to the other cases and this case is the easiest to solve. In cases 3 and 4, the power allocation is no longer waterfilling in any form and every subcarrier i has an individual subcarrier threshold level w_i.

As we could notice, obtaining the optimal solution is computationally intensive. One difficult challenge is to find the best partition of the sub-bands into sets A and B. With exhaustive search, we may require up to 2^M iterations and each iteration running on all four cases. In the worst case (as in case 4), we have to solve $M + 2$ set of equations to determine subcarrier water-levels w_i's. These equations are equivalent to multi-waterlevel multiconstrained waterfilling solutions as described in [50]. The authors in [50] showed that such a family of waterfilling solutions can be obtained with linear complexity in the order of number of subcarriers. Therefore, using the exhaustive search method, the optimal solution can be obtained with complexity of $O(2^M N)$. Due to this huge computational complexity, we require suboptimal or heuristic-based approaches to find power allocation which perform close to the optimal solution.

2.4 Heuristics-Based Suboptimal Power Allocation Schemes

Based on Proposition 1, we could intuitively observe the structure of the optimal solution. Those subcarriers closer in spectral distance to PU bands have higher K_i values hence cause higher amount of interference to the adjacent PU bands. Therefore, given a power allocation scheme which exceeds the interference threshold level I_{th}, power could be readjusted in a way such that subcarriers closer to PU bands have lesser power allocated to them in order to avoid any harmful interference introduced to the PU users. A nulling mechanism and a step ladder-based allocation algorithm for such a scenario were proposed in [42]. Also, those sub-bands with higher $\alpha_{\varphi(i)}$ value (or higher sensing error or PU activity) have a lower power level compared to the sub-bands with lower $\alpha_{\varphi(i)}$. Lastly, we have to make sure that total sub-band power is less than corresponding T_j values.

Based on these observations, we propose the following heuristic solution solvable in complexity of $O(MN)$, which is significantly lower than the optimal scheme:

1. **Set relative water levels:** We first set the water-levels based on $\alpha_{\varphi(i)}$ values for each sub-band. In this scheme we set the water level for jth sub-band to be $w-\kappa\alpha_j\mathcal{C}$, κ being a design parameter. This is a relative adjustment of water level and it reduces the constraint equation to the following piecewise linear equation in w, solvable by iterative algorithms:

$$\sum_{i=1}^{N}\left[w-\kappa\alpha_{\varphi(i)}\mathcal{C}-\frac{N_0B+J_i}{|h_i|^2}\right]^+ = P_{total}. \tag{2.30}$$

   Power allocation vector $\mathbf{P^{(1)}}$ is hence given by

$$P_i^{(1)} = \left[w-\kappa\alpha_{\varphi(i)}\mathcal{C}-\frac{N_0B+J_i}{|h_i|^2}\right]^+, \forall i\in\{1,\cdots,N\}. \tag{2.31}$$

   We pass $\mathbf{P^{(1)}}$ in step 2.

2. **Iterative readjustment of water levels based on T_j values (takes argument $\mathbf{P}$):** Based on power allocation vector $\mathbf{P}$, we divide the sub-bands in two sets E and F such that if (i) $S_j \leq T_j$, then $j \in E$ and if (ii) $S_j > T_j$, then $j \in F$. If F is empty we call $\mathbf{P}$ as $\mathbf{P^{(2)}}$ and proceed to step 3. Otherwise, we do waterfilling for sub-bands in set F with power budget T_j. We find $|F|$ number of w_j values such that

$$\sum_{i \text{ s.t } \varphi(i)=j}\left[w_j-\frac{N_0B+J_i}{|h_i|^2}\right]^+ = T_j\ , \forall j\in F. \tag{2.32}$$

   We then reset the power in ith subcarrier which belongs to jth sub-band ($j\in F$) as follows:

$$P_i = \left[w_{\varphi(i)}-\frac{N_0B+J_i}{|h_i|^2}\right]^+, \tag{2.33}$$
$$\forall i\in\{1,2,\cdots,N\} \text{ such that } \varphi(i)\in F.$$

   We then take the new power budget as $P'_{total} = P_{total} - \sum_{j\in F} T_j$ and re-distribute it to the sub-bands in set E by setting relative water-levels to sub-bands in E as in step 1. We call this new power allocation vector $\mathbf{P}$ and pass it back as an argument to step 2.

3. **Step ladder-based allocation, nulling, and scaling schemes:** If $\sum_{i=1}^{N} K_iP_i^{(2)} \leq I_{th}$, we stop here and $\mathbf{P^{(2)}}$ is our power allocation vector. Otherwise, we redistribute the power within the sub-bands which share boundaries with the PU bands. To achieve this, if we consider the step-ladder power profile [42], we assign lower power to subcarriers closer to PU bands (or higher K_i) with a step size which could either be constant or proportional to

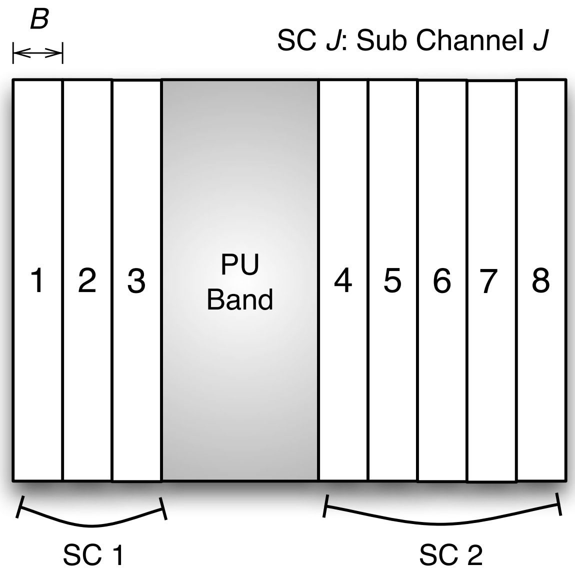


Figure 2.3: An example of a simple OFDM-CR system.

K_i values. To illustrate this mathematically, let us consider jth sub-band which is adjacent to some PU band. Power allocated to ith subcarrier, which belongs to this sub-band, is modified to $P_i^{(3)} = \frac{P}{K_i}$, $\forall i \leq N$ such that $\varphi(i) = j$, where P could be determined as follows (maintaining the total sub-band power): $\sum_{i \text{ s.t } \varphi(i)=j} P_i^{(3)} = \sum_{i \text{ s.t } \varphi(i)=j} P_i^{(2)}$. If we consider nulling mechanism (which is faster), we allocate zero power to subcarriers adjacent to PU band (one-nulling) or zero power to two subcarriers closest to PU band on each side (two-nulling). Lastly, there is another way we could obtain a power allocation which satisfies this interference criterion strictly by scaling down the power as $\mathbf{P^{(3)}} = (I_{th}/\sum_{i=1}^{N} K_i P_i^{(2)})\mathbf{P^{(3)}}$. We call this power allocation vector $\mathbf{P^{(3)}}$ as the heuristic solution.

Since there are total N subcarriers, each iteration in step 2 is solvable in $O(N)$ and there could be maximum M iterations in the worst case. Step 2 also has a higher complexity than step 1 and 3. Therefore, the worst case complexity of the algorithm will be $O(MN)$.

2.5 Performance Evaluation

2.5.1 Simulation Parameters and Assumptions

We consider a simple system consisting of two opportunistic sub-bands licensed to two primary users, namely PU_1, PU_2 as shown in Fig. 2.3. The two sub-bands consist of 3 and 5 subcarriers, respectively, with subcarrier bandwidth B to be 0.125 MHz each. We further assume that there is a PU band that lies between these two sub-bands. We consider a Rayleigh fading environment with average channel power gain, i.e. $E\{|h|^2\}$ equal to 1. Subcarriers 1 to 3 belong to PU_1 with an activity pattern/sensing error factor $\alpha_1 = 0.3$ and subcarriers 4 to 8 are licensed to PU_2 with $\alpha_2 = 0.8$. Total power budget is assumed to be $P_{total} = 10^{-5}$ W and noise power $N_0 = 10^{-11}$ W/Hz so that the average SNR per subcarrier is 0dB. The sub-band power constraints are $T_1 = 10^{-6}$ W and $T_2 = 5 \times 10^{-6}$ W. Total interference threshold limit is taken as $I_{th} = 2 \times 10^{-6}$ W. The values of $K_1, K_2, K_3, K_4, K_5, K_6, K_7, K_8$ are assumed to be 0.4, 0.78, 1.2, 1.3, 0.8, 0.5, 0.3 and 0.1, respectively. The values of K_i's are chosen randomly but correlate with the fact that subcarriers farther away from PU band cause less interference to the PU. The values of $J_1, J_2, J_3, J_4, J_5, J_6, J_7, J_8$ are taken to be 0.25×10^{-6}, 0.75×10^{-6}, 1.1×10^{-6}, 10^{-6}, 0.9×10^{-6}, 0.5×10^{-6}, 0.4×10^{-6} and 0.3×10^{-6} W/Hz, respectively. J_i's are also chosen randomly in accordance with the fact that the interference generated to the SU by the PU decreases with spectral distance from PU band. Assuming a linear rate loss function with normalized cost per unit power $\mathcal{C} = 3 \times 10^3$ bits/s/mW, we simulate the subcarrier power allocation for this system based on the optimal scheme discussed in Section 2.3 and the heuristic scheme discussed in Section 2.4. We set the value of the constant $\kappa = 4 \times 10^{-7}$ for suboptimal schemes. To simulate the optimal scheme, we use sequential quadratic programming [51] on the original optimization problem.

2.5.2 Simulation Results

In Fig. 2.4, we plot the achievable normalized expected capacity for the SU versus interference threshold I_{th} and power budget P_{total}. We observe that for a fixed P_{total}, as I_{th} is increased, the capacity increases and then it saturates. Lesser the power budget, quicker is the saturation which is because the allocation scheme is completely controlled by the power budget as we keep on relaxing the interference threshold. Same is true for a fixed I_{th}. Therefore, as expected (and from the symmetry of the graph), interference threshold and power budget play a very similar role. We also compare normalized average capacity with respect to reliability factors in Fig. 2.5. To get a clearer picture, we first relax the interference and the sub-band constraints and then plot the capacities by varying α_2. Note that the suboptimal schemes (nulling, step-ladder, and scaling) give equal capacity (Fig. 2.5). The optimal scheme gives higher capacity compared to waterfilling and suboptimal schemes. Also the expected capacity for different schemes decreases with α_2 and this decrease is the steepest (linear) for waterfilling.

In Fig. 2.6, we plot the average interference generated to adjacent PU bands by the SU for different power allocation schemes versus power budget P_{total} with a fixed thershold I_{th} and fixed sub-band power constraints. The inability of the waterfilling algorithm to maintain the total adjacent band interference is evident from this figure. This interference increases linearly with power budget for waterfilling because this scheme does not judiciously take this interference constraint in its power allocation scheme. The optimal and the scaling schemes perform the best where this interference gradually rises before it saturates to the allowed threshold I_{th} without crossing this limit. The nulling and the step-ladder schemes perform much better than the waterfilling schemes by keeping the interference limit much below in comparison to the water-filing scenario but finally saturating to a value slightly higher than allowed limit. This could be explained by the fact that in the last step of nulling and step-ladder we avoid causing additional interference to adjacent band PUs without specifically controlling it to a value below threshold limit. The step-ladder scheme generates slightly lower interference for high power budget values compared to the nulling mechanism but as we will see, it gives lower data rate compared to nulling.

Fig. 2.7 compares the total power allocated to the SU in the second sub-band versus P_{total} for different schemes under consideration. We observe that for waterfilling, total power in second sub-band increases linearly with power budget because waterfilling scheme does not consider sub-band power constraint. Both the suboptimal schemes and the optimal scheme never exceed the sub-band power constraint T_2. In the case of nulling, this value is slightly below the threshold level T_2 because in this case power in the subcarrier adjacent to PU band is set to zero which reduces the total allocated power. The scaling scheme, as expected, saturates to a total sub-band power much lower than T_2.

To compare the expected capacities (or expected transmission rates), we plot the expected achievable transmission capacity for different schemes in Fig. 2.8. The expected transmission capacity that can be achieved by the waterfilling algorithm is higher than the optimal and suboptimal schemes because the interference constraints are not satisfied. However, the performance of waterfilling algorithm in a way serves as a benchmark for comparison purposes. The expected capacity achieved by optimal scheme is greater than any other algorithm as long as all the interference and power constraints are met. We notice that the suboptimal schemes give much lower expected capacity compared to the optimal scheme. Interestingly, as we increase the power budget, one-nulling achieves a slightly higher expected capacity compared to the step-ladder algorithm. However, as we discussed earlier, one-nulling also generates additional interference at the same time. The reason behind this is that in the one-nulling scheme we inactivate only one subcarrier on either side of the PU band reducing only the interference caused by subcarriers immediately next to the PU bands. On the other hand, in the step-ladder approach we redistribute power to each subcarrier inversely proportional to the interference factors K_i. The scaling scheme achieves higher expected capacity than both nulling and step-ladder schemes until a certain power budget but it finally saturates to a lower expected capacity without violating the interference constraints.

From the simulation study, we conclude that although the nulling and step-ladder schemes do not have strict bound on interference generated to adjacent PU bands, they still perform better than waterfilling by closely satisfying the threshold limits. The scaling scheme satisfies all the constraints but gives a lower expected capacity for high power budget values. The optimal scheme performs the best by giving the highest expected capacity without violating any bounds but has much higher computational complexity than any of suboptimal schemes.

2.6 Conclusion

In this chapter, we have considered the problem of energy efficient resource allocation for maximizing the expected transmission rate for an OFDM-based cognitive radio system by taking into account the reliability of the available sub-bands (which depends on sensing error and PU activity), sub-band power constraints, and total allowed interference limit to the adjacent PU bands. We have introduced a risk-return model to incorporate channel reliability by defining an average rate loss function. The advantage of this model is that it not only takes into account channel reliability for power allocation but also the interference constraint limits. Due to the huge complexity of computation of the optimal power allocation solution, we have also proposed three viable suboptimal schemes, namely, the step-ladder, nulling, and scaling schemes, and compared their performances with respect to waterfilling and optimal schemes. Simulation results have revealed that the suboptimal schemes perform closer to the optimal scheme satisfying interference and power bounds but give lower expected capacities. The waterfilling scheme, unable to satisfy the constraints performs the poorest and hence impractical amongst all the schemes considered in this chapter. Also, the step-ladder scheme performs slightly better than the nulling mechanism in terms of interference generated to PU bands but also achieves lower capacity. The scaling scheme satisfies all the constraints but gives lower capacity for high power budgets.

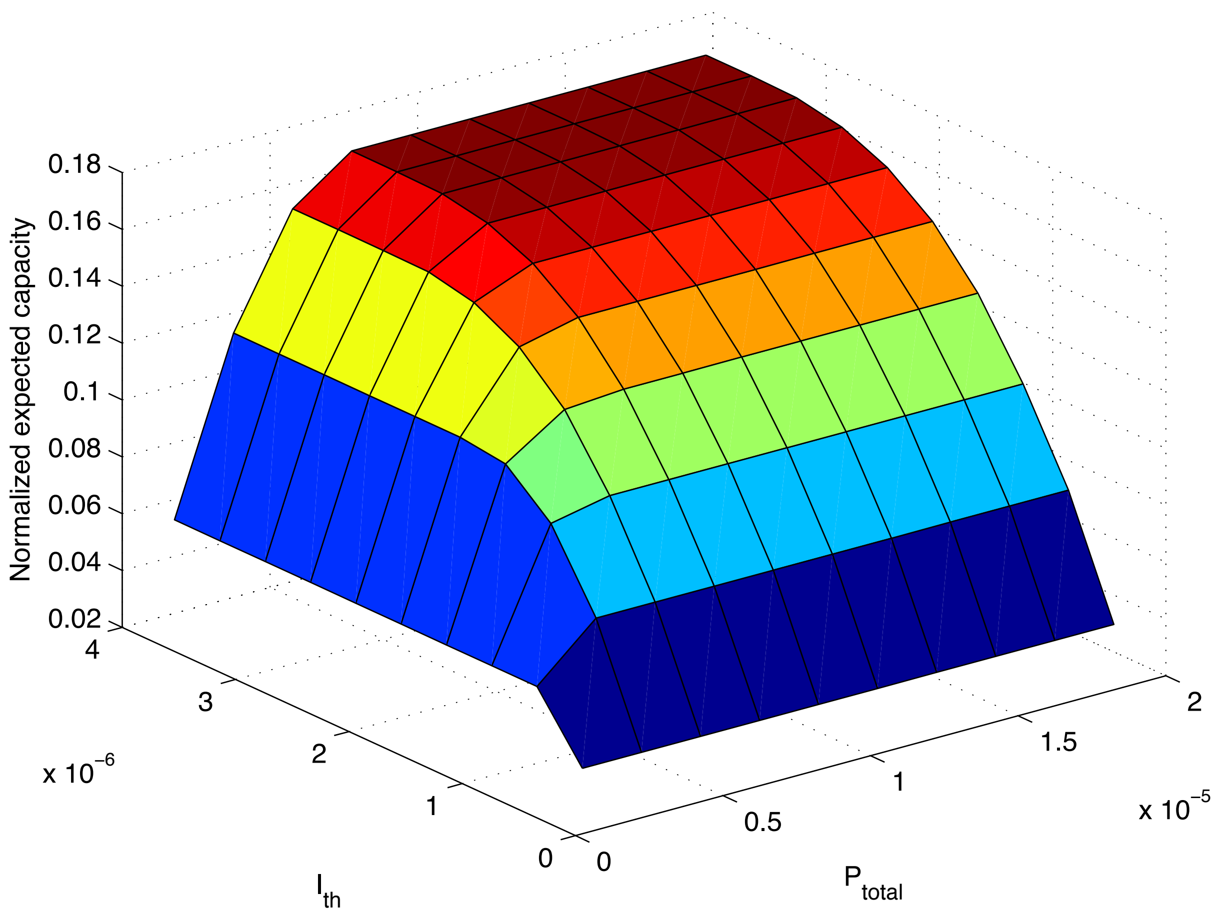


Figure 2.4: Expected normalized average capacity vs. P_{total} vs I_{th}.

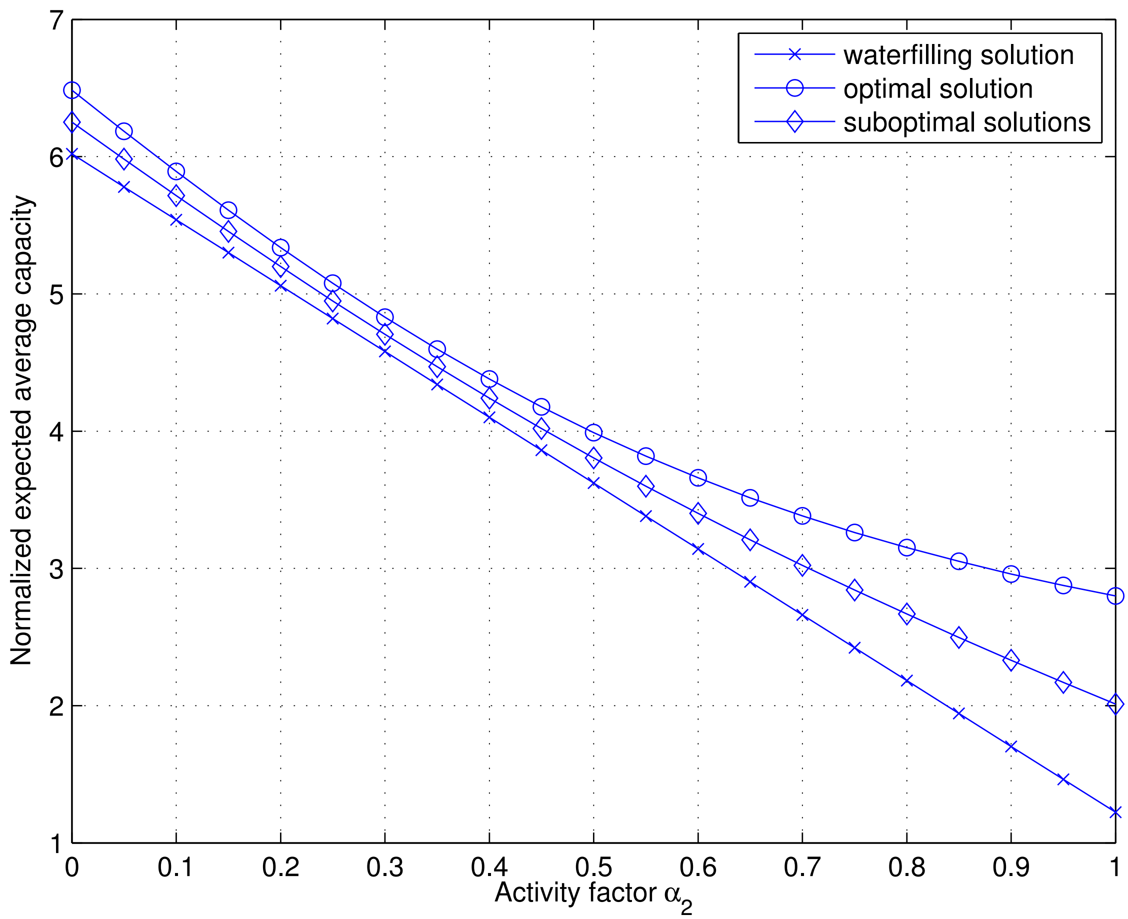


Figure 2.5: Activity α_2 vs. normalized average capacity for the secondary user under no constraints.

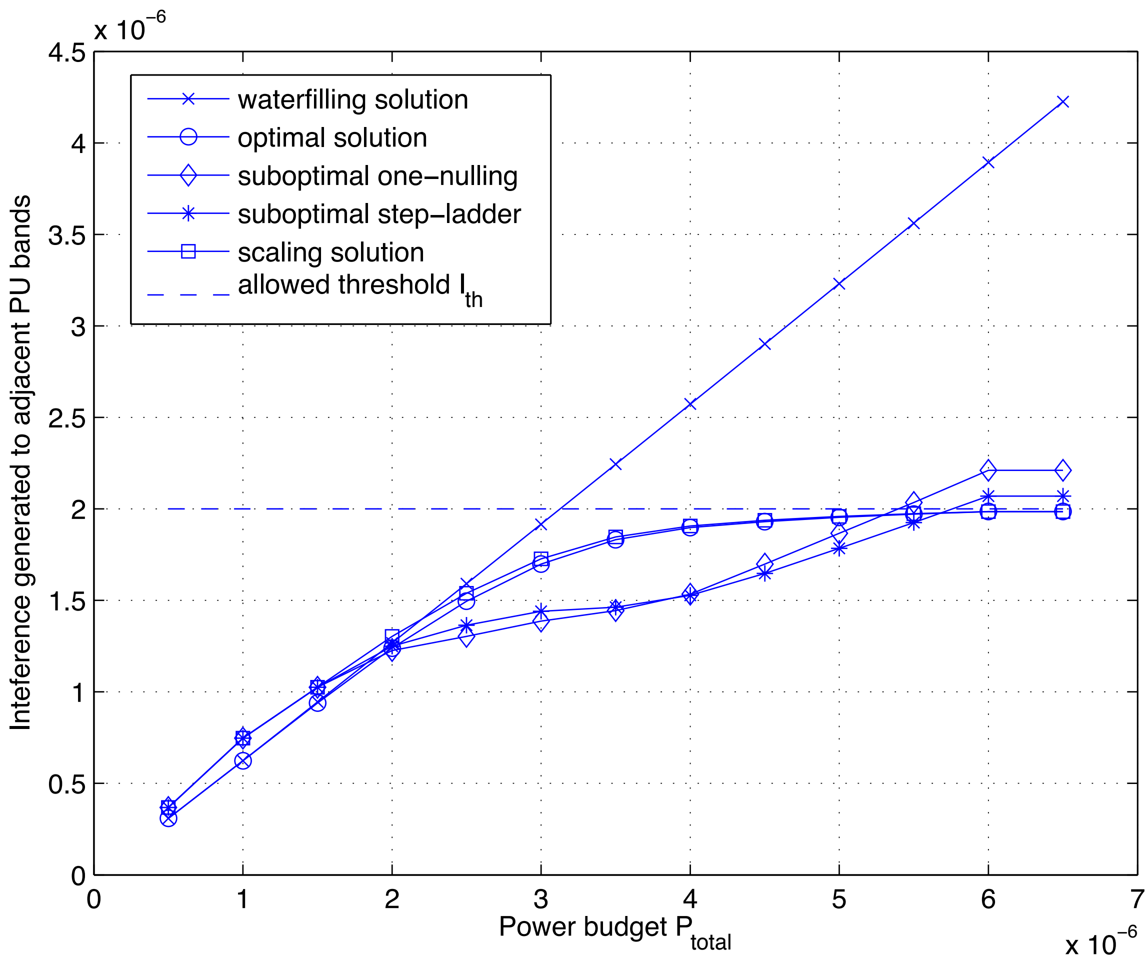


Figure 2.6: Power budget P_{total} vs. interference generated to adjacent PU bands by the secondary user.

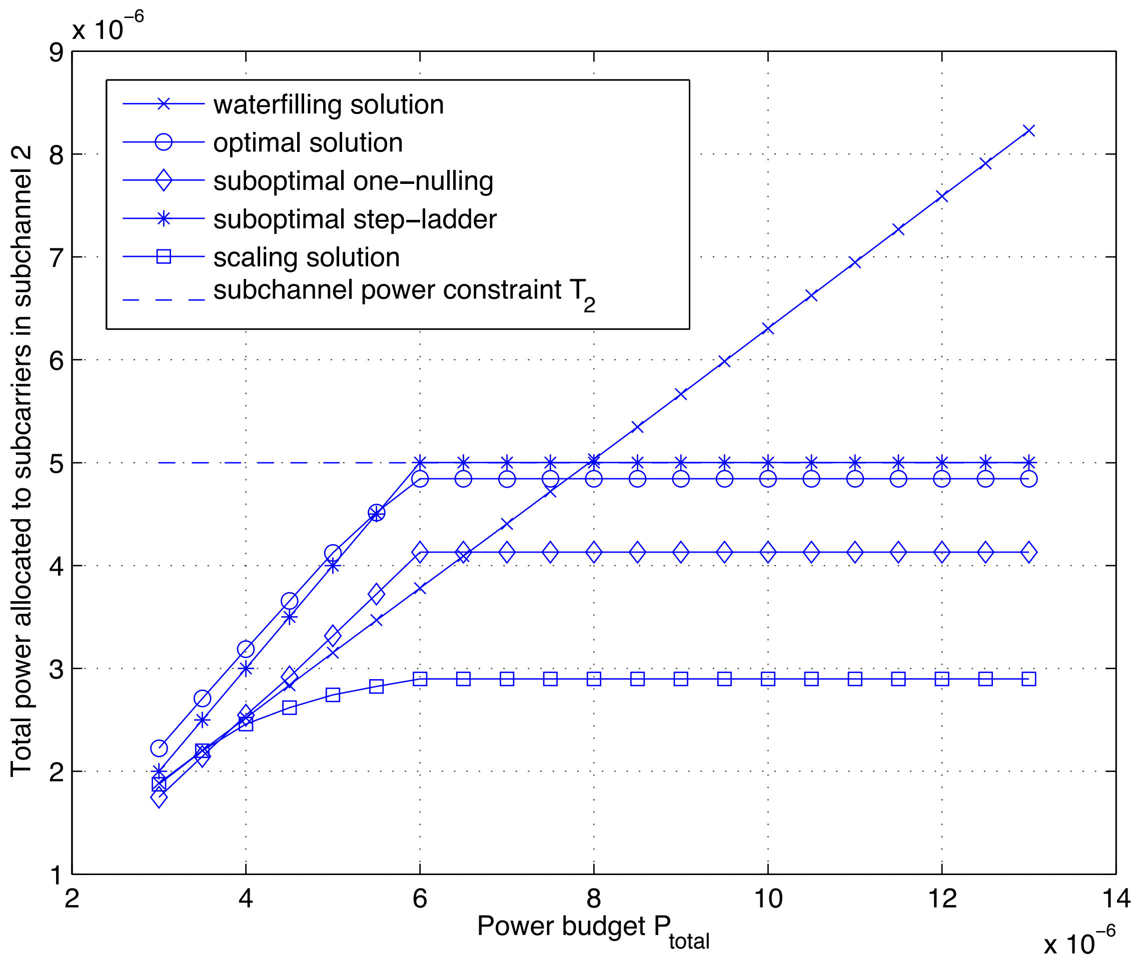


Figure 2.7: Power budget P_{total} vs. total power allocated to subcarriers in sub-band 2.

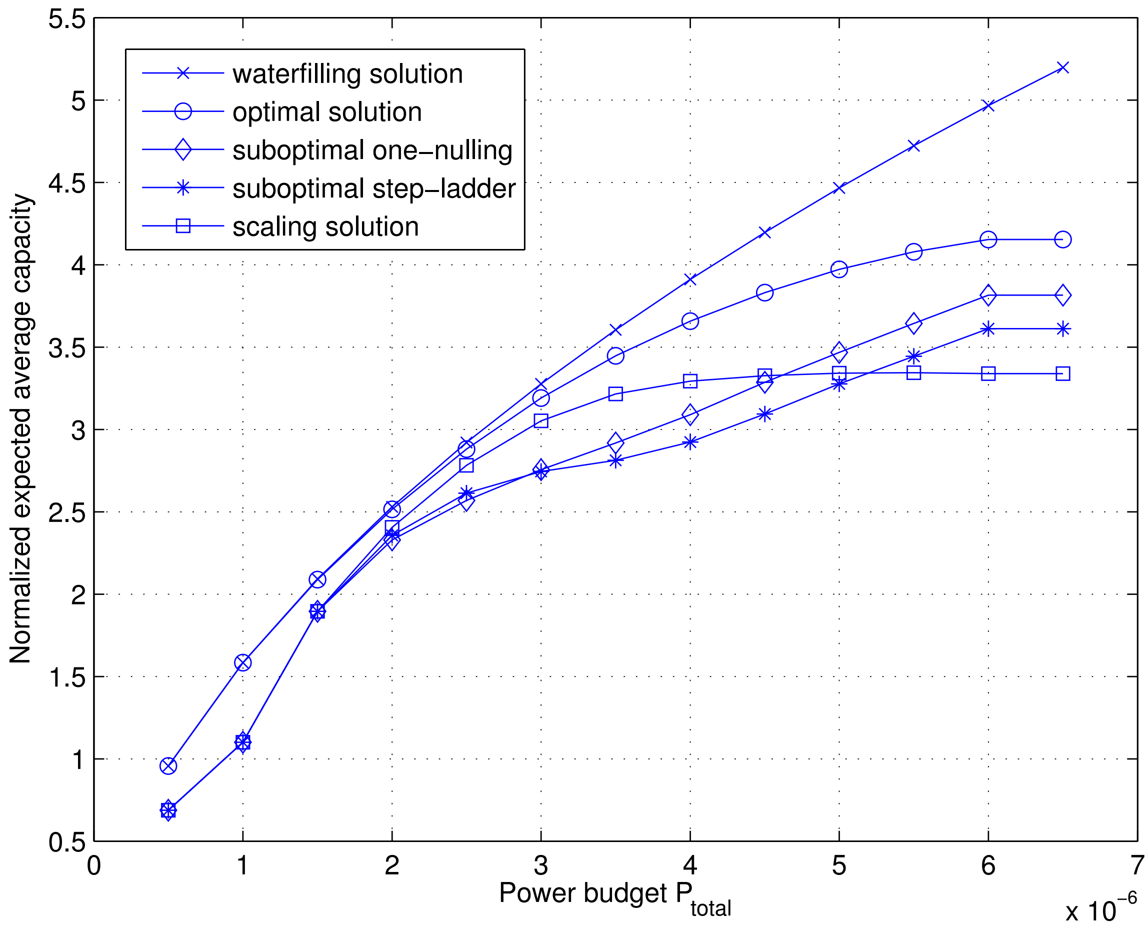


Figure 2.8: Power budget P_{total} vs. normalized average capacity for the secondary user.

Chapter 3

Resource Allocation for Multiuser OFDMA-based Amplify-and-Forward Relay Networks with Selective Relaying

3.1 Introduction

Cooperative communication via relays is considered an important technique to boost the overall system efficiency by extending the network coverage, increasing the system throughput and improving the link reliability. In relay-based cooperative wireless networks, a transmitter can send data to distant nodes through wireless relay nodes either by amplifying-and-forwarding the message (AF relaying) or by decoding-and-forwarding the message (DF relaying). The relay-based cooperative cellular architecture is pursued very actively by next generation broadband systems such as 3GPP Long Term Evolution (LTE) and IEEE 802.16j mobile WiMAX [52, 53]. In order to fully realize the advantages of cooperative communication systems, it is crucial to design efficient resource allocation algorithms.

With adaptive power allocation relay-assisted communication in systems employing orthogonal frequency-division multiplexing (OFDM) has been shown to considerably improve the achievable data rates between a base station and a destination node using a single relay [23]. Although, most of the earlier work on resource allocation for OFDM based relaying networks have been for single-user networks [23], a few authors have recently addressed this problem for multiuser OFDMA-based systems [54–56]. In [54, 55], the authors focus on designing resource allocation algorithms for a downlink OFDMA-based relay network utilizing regenerative or DF cooperation strategy while [56] mainly addresses resource allocation schemes for cooperative networks based on non-regenerative or AF relays.

In this chapter, we investigate resource allocation schemes for multiuser cooperative relay systems with selective relaying. Selective relaying is a technique through which the source adaptively chooses to relay on certain subcarriers depending on the potential gains [57, 58]. Here we choose AF relays over DF relays because of their lower cost and simpler design. Unlike the work presented in [56], in this chapter we do not ignore the direct links between source and destination. Also, we assume that source and relay have their individual power constraints because it is more practical. Based on these assumptions, we will present an iterative suboptimal scheme based on results in [57] to allocate subcarriers and power and to take relaying decisions. To ensure fairness amongst users, we introduce a proportional rate fairness criterion in our system [59]. Using simulation results based on the suboptimal schemes, we show that the algorithm not only converges to relay decisions and achievable capacity but it also gives much better performance compared to the "always relay" scenario.

The rest of the chapter is organized as follows. In Section 3.2, we introduce the system and formulate our problem. We present the heuristic solution to our problem in Section 3.3. In this section, we first present a suboptimal subcarrier allocation to achieve fairness amongst users and then present an iterative relay decision and power allocation scheme to maximize capacity. The numerical results are shown in Section 3.4. Finally, the conclusions are given in Section 3.5.

3.2 System Model and Problem Formulation

3.2.1 System Model and Assumptions

We take a downlink point-to-multipoint OFDMA system model with one source or base station, one non-regenerative (amplify and forward) relay station and multiple users. Suppose that there are M users sharing a total of K number of subcarriers which are available to the base station for transmission. We assume that the relaying is done in a half-duplex mode and we therefore use a two-phase AF protocol. Thus, in the first time slot, source transmits the data to various users while the relay and end users listen and on the second time slot, the relay amplifies and forwards the signal and only the end users listen. Such a protocol realizes a maximum order diversity and exhibits no receive collision [24]. We assume that the source employs selective relaying as proposed in [57]. Therefore, for each subcarrier on the second time slot, the source instructs the relay station to send the amplified signal only when there is an overall capacity gain via relaying (e.g., weak direct link). The source transmits data on the second time slot for those subcarriers where relaying is not used. Assuming that each subcarrier is assigned to one and only one user, we group the subcarriers assigned to user m in a set Ω_m. Therefore, $k \in \Omega_m$ iff subcarrier k is assigned to user m. Furthermore, since each subcarrier is assigned to only one user, sets Ω_m's are pairwise disjoint i.e., $\Omega_m \cap \Omega_n = \phi$ whenever $m \neq n$.

Since we do not relay on all subcarriers, we introduce a binary relay decision variable $\mu_{k,m}$ for the kth subcarrier and mth user. When $\mu_{k,m} = 1$, this means that relaying is used in the second time slot and the relay station amplifies and forwards the message sent by the base station to the mth user on kth subcarrier in the first time slot. If we do not use relay, then $\mu_{k,m}$ = 0, and the source transmits a different message on the second time slot.

$$\mu_{k,m} = \begin{cases} 0, & \text{no relaying on subcarrier } k \text{ for user } m \\ 1, & \text{relaying enabled on subcarrier } k \text{ for user } m. \end{cases}$$

Let us denote the channel gain on the kth subcarrier between source and relay as $h_{1,k}$, between relay and user m as $h_{2,k,m}$, and between source and user m as $h_{3,k,m}$. We assume all these channel gains to be slow-varying, which means that they would remain constant for both transmission and relaying time slots. We assume that the source sends data on kth subcarrier to the user m with power $P_{S,k,m}$ on the first time slot and with $\hat{P}_{S,k,m}$ on the second time slot. The relay amplifies the message by a factor $\alpha_{k,m}$ using power $P_{R,k,m}$ on the kth subcarrier for user m [23]:

$$\alpha_{k,m} = \sqrt{\frac{P_{R,k,m}}{P_{S,k,m}|h_{1,k}|^2 + \sigma_r^2}}. \tag{3.1}$$

Notice that $\hat{P}_{S,k,m}$ will be set to 0 when relay is used for that subcarrier and $P_{R,k,m}$ will be set to 0 if relay is not used. The noise variance within one OFDM subcarrier is taken as σ_r^2 at the relay receiver and σ_m^2 at user m. The SNR ratio $\rho_{k,m}$ on subcarrier k for user m is then [23]:

$$\begin{aligned} \rho_{k,m} &= \frac{P_{S,k,m}|\alpha_{k,m} h_{1,k} h_{2,k,m}|^2}{\sigma_m^2 + \sigma_r^2 |\alpha_{k,m} h_{2,k,m}|^2} + \frac{P_{S,k,m}|h_{3,k,m}|^2}{\sigma_m^2} \\ &= \frac{P_{S,k,m} a_k \cdot P_{R,k,m} b_{k,m}}{1 + P_{S,k,m} a_k + P_{R,k,m} b_{k,m}} + P_{S,k,m} c_{k,m} \end{aligned} \tag{3.2}$$

where $a_k = \frac{|h_{1,k}|^2}{\sigma_r^2}$, $b_{k,m} = \frac{|h_{2,k,m}|^2}{\sigma_m^2}$, and $c_{k,m} = \frac{|h_{3,k,m}|^2}{\sigma_m^2}$.

We assume that the base station has total information of all the channel gains and noise variance at the relay and destination receivers. The capacity for a relay communication in the kth subcarrier for mth user is given by

$$C_{R,k,m} = \frac{1}{2} \log_2(1 + \rho_{k,m}). \tag{3.3}$$

The $1/2$ factor is multiplied here because of the half duplex relaying process. Suppose if the relay is not being used, then the capacity for the non-relaying process will be the sum of capacities obtained during the first and second time slots and is therefore given by:

$$C_{S,k,m} = \frac{1}{2}(\log_2(1 + P_{S,k,m}c_{k,m}) + \log_2(1 + \hat{P}_{S,k,m}c_{k,m})). \tag{3.4}$$

Overall capacity on the kth subcarrier for the mth user should be chosen to be the maximum achievable of either $C_{R,k,m}$ or $C_{S,k,m}$, i.e., by $\max\{C_{R,k,m}, C_{S,k,m}\}$. However, if we use the relay decision variable $\mu_{k,m}$, we can combine (3.3) and (3.4) and we then obtain the following overall capacity expression using selective subcarrier relaying on the kth subcarrier by the mth user:

$$C_{k,m} = \mu_{k,m}C_{R,k,m} + (1 - \mu_{k,m})C_{S,k,m}. \tag{3.5}$$

Then, the total capacity of mth user will be given by:

$$C_m = \sum_{k\in\Omega_m} C_{k,m}. \tag{3.6}$$

3.2.2 Problem Formulation

Now given the total power constraints P_S at source for the first and the second time slot and P_R at the relay, our objective is to determine the subcarrier allocation sets Ω_m, relay decisions $\mu_{k,m}$, the power allocations at the source $P_{S,k,m}$, $\hat{P}_{S,k,m}$ and at the relay station $P_{R,k,m}$ for all subcarriers and all users such that we obtain the optimal global capacity $\sum_{m=1}^{M} C_m$. Hence, we seek a solution to the following optimization problem:

$$\max \sum_{m=1}^{M} \sum_{k\in\Omega_m} C_{k,m} \tag{3.7}$$

$$\mu_{k,m} = \{0, 1\} \quad \forall k, m \tag{3.8}$$

$$\sum_{k=1}^{K} \sum_{m=1}^{M} P_{S,k,m} \leq P_S \tag{3.9}$$

$$\sum_{k=1}^{K} \sum_{m=1}^{M} \hat{P}_{S,k,m} \leq P_S \tag{3.10}$$

$$\sum_{k=1}^{K} \sum_{m=1}^{M} P_{R,k,m} \leq P_R \tag{3.11}$$

$$\Omega_1 \cup \Omega_2 \cup ... \cup \Omega_M \subseteq \{1, 2, \cdots, K\} \tag{3.12}$$

$$\Omega_i \cap \Omega_j = \phi \text{ for } i \neq j \quad \forall i, j \in \{1, 2, \cdots, M\} \tag{3.13}$$

$$P_{S,k,m} \geq 0, \hat{P}_{S,k,m} \geq 0, P_{R,k,m} \geq 0. \tag{3.14}$$

In the optimization problem above, our objective is to maximize the global capacity under total power constraints. However, such an allocation may not be fair to all users since all the subcarriers can be given to users with higher channel gains and users with relatively poor channel quality might suffer. Due to this, we introduce the idea of proportional rate fairness amongst users by adding a set of constraints on the ratios of user data rates. Proportional rate fairness is beneficial for our system because we can inherently control the capacity ratios of users and on an average ensure that each user can meet the target data requirements, provided we have sufficient total available transmit power [59]. Mathematically, we can add the following constraint to our optimization problem:

$$C_1 : C_2 : C_3 : \cdots : C_M = \omega_1 : \omega_2 : \omega_3 : \cdots : \omega_M \tag{3.15}$$

where $\{\omega_m\}_{m=1}^{M}$ are the values of capacity ratios in order to guarantee proportional fairness of user data rates.

3.3 Heuristic Scheme for Subcarrier and Power Allocation with Relay Decisions

The optimization problem stated in (3.7)-(3.15) should ideally be solved jointly to obtain the optimal solution. However, since it is an integer programming problem, the complexity grows exponentially when the number of users and subcarriers increases. Moreover, the source has to re-compute the optimal power, subcarrier allocation, and relay decisions for both relay and source as the channel changes. Hence, fast and low-complexity suboptimal algorithms are more reasonable options to reduce the computational burden of the base stations.

In order to find a suboptimal solution, we divide our optimization problem in four subproblems. The first subproblem is to find the subcarrier allocation sets Ω_m and we discuss it in Section 3.3.1. Now, given the source powers in the first time-slot and subcarrier allocation sets, the second sub-problem in Section 3.3.2 is for the relay and the goal is to obtain relay decisions and relay powers in those subcarriers where relaying will be performed. The third sub-problem discussed in Section 3.3.3 is for the source. Here source determines source powers in the first time-slot assuming relay powers are given. Steps in Section 3.3.2 and Section 3.3.3 are iteratively repeated until the capacity and relay decisions converge. We start with equal source powers in the first time slot and we take the source power in the second time slot to be the same as the first slot in order to calculate capacities. The last step, as elaborated in Section 3.3.4, is for the source to allocate powers for the non-relaying subcarriers on the second time-slot.

3.3.1 Suboptimal Subcarrier Allocation

In this section, we propose an algorithm to allocate subcarriers to different users in suboptimal fashion based on results in [59, 60]. The problem is to allocate each of the K subcarriers to the M users with respect to the fairness criterion. To start with, we assume equal power has been allocated to each subcarrier by the source in each time slot and by the relay in the second time slot (corresponding to their power constraints), i.e.,

$$P_{S,k,m}, \hat{P}_{S,k,m} = \begin{cases} \frac{P_S}{K}, & \text{for } k \in \Omega_m \\ 0, & \text{otherwise} \end{cases} \tag{3.16}$$

and

$$P_{R,k,m} = \begin{cases} \frac{P_R}{K}, & \text{for } k \in \Omega_m \\ 0, & \text{otherwise} \end{cases} \tag{3.17}$$

for all $k = 1, 2, \cdots, K$ and $m = 1, 2, \cdots, M$. Now since power allocated in each subcarrier is fixed, subcarrier allocation is just an assignment problem. Based on these values we first calculate $C_{R,k,m}$ and $C_{S,k,m}$ for each subcarrier assuming it has been assigned to user m. Let us define the best possible rate $r_{k,m} = \max\{C_{R,k,m}, C_{S,k,m}\}$.

1. Initialization
   (a) Set $C_m = 0$, $\Omega_m = \phi$ for $m = 1, 2, \cdots, M$ and $\Delta = \{1, 2, \cdots, K\}$.
2. For $m = 1, 2, \cdots, M$
   (a) find k such that $r_{k,m} \geq r_{j,m}\ \forall j \in \Delta$
   (b) $\Omega_m = \Omega_m \cup \{k\}, \Delta = \Delta - \{k\}$ and $C_m = r_{k,m}$
3. While $\Delta \neq \phi$
   (a) find m satisfying $\frac{C_m}{\omega_m} \leq \frac{C_i}{\omega_i}$ for all $i \in \{1, 2, \cdots, M\}$
   (b) for the found m, find k satisfying $r_{k,m} \geq r_{j,m}\ \forall j \in \Delta$
   (c) for the found m and k, let $\Omega_m = \Omega_m \cup \{k\}$, $\Delta = \Delta - \{k\}$ and update $C_m = C_m + r_{k,m}$.

The idea behind this suboptimal subcarrier allocation is for each user to use those subcarriers first which can provide higher potential data rates with equal powers at source and relay. In the subsequent iterations, the user with the lowest proportional rate fairness has the option to choose which subcarrier to use. Suboptimality is introduced here because of the equal power distributions at the source and relay. Hence, only an approximate proportional rate fairness is achieved. Once we have determined a subcarrier allocation for different users, power allocation with relay decision can be done in the following two iterative steps at the source and relay and one non-iterative step at the source to allocate power in the second time slot [23, 57].

3.3.2 Subcarrier Selection and Power Allocation at the Relay

We assume that subcarrier allocation, i.e., sets Ω_m are given. In this first step, given a source powers for each subcarrier and in the first time slot $P_{S,k,m}$, we first calculate estimates of $\mu_{k,m}$ and then $P_{R,k,m}$. We begin with the following claim:

Proposition 1. *For all non-relaying subcarriers k (where the relay decision is 0) $\hat{P}_{S,k,m} \geq P_{S,k,m}$ for optimal allocation.*

Proof. Suppose $\mathbb{S}$ be the set of all subcarriers and $\mathbb{T}$ be the set of all non-relaying subcarriers. Now since the total power constraints P_S are the same for both time-slots and $\mathbb{T} \subseteq \mathbb{S}$, we can deduce that $\sum_{m=1}^{M} \sum_{k \in \mathbb{T} \cap \Omega_m} \hat{P}_{S,k,m} \geq \sum_{m=1}^{M} \sum_{k \in \mathbb{T} \cap \Omega_m} P_{S,k,m}$. We also know that the optimization objective, i.e., maximizing the log capacity is equivalent in both time slots over the non-relaying subcarriers and solution to which is water-filling [13]. Therefore, given the relay decisions and optimal $P_{S,k,m}$, since the total power available is more for the second time-slot, $\hat{P}_{S,k,m} \geq P_{S,k,m}$ for $k \in \mathbb{T}$. $\square$

To determine the relay decisions, we use a similar approach as in [57]. Suppose $k \in \Omega_m$, then we can say that in order to obtain gain in capacity via relaying in this subcarrier $C_{R,k,m} \geq C_{S,k,m}$, i.e.,

$$(1 + \rho_{k,m}) \geq (1 + P_{S,k,m} c_{k,m})(1 + \hat{P}_{S,k,m} c_{k,m}). \tag{3.18}$$

Using Claim 1, we can hence say that in order to improve capacity the following must be true:

$$(1 + \rho_{k,m}) \geq (1 + P_{S,k,m} c_{k,m})^2. \tag{3.19}$$

Substituting $\rho_{k,m}$ from (3.2), we can simplify this inequality as follows:

$$\vartheta_1 - \vartheta_2 P_{R,k,m} \leq 0 \tag{3.20}$$

where ϑ_1 and ϑ_2 are given by the following:

$$\vartheta_1 = P_{S,k,m} c_{k,m}(1 + P_{S,k,m} a_k)(1 + P_{S,k,m} c_{k,m})$$
$$\vartheta_2 = a_k b_{k,m} P_{S,k,m} - P_{S,k,m} b_{k,m} c_{k,m}(1 + P_{S,k,m} c_{k,m}).$$

Since $\vartheta_1 \geq 0$, in order to satisfy (3.20) or to obtain gain in capacity via relaying, first $\vartheta_2 > 0$ and then $P_{R,k,m} \geq \frac{\vartheta_1}{\vartheta_2}$. Substituting the values for ϑ_1 and ϑ_2, we can simplify these two conditions as follows:

$$a_k > c_{k,m}(1 + P_{S,k,m}c_{k,m}) \tag{3.21}$$

$$P_{R,k,m} \geq \gamma_{k,m} \tag{3.22}$$

$$\text{where } \gamma_{k,m} = \frac{c_{k,m}(1 + P_{S,k,m}a_k)(1 + P_{S,k,m}c_{k,m})}{b_{k,m}(a_k - c_{k,m}(1 + P_{S,k,m}c_{k,m}))}.$$

The condition in (3.21) roughly implies that the source-relay link must be better than the source-destination link and (3.22) implies that the minimum required relay power must be inversely proportional to channel quality of relay-destination link in that subcarrier.

For every subcarrier $k \in \Omega_m$ in each set Ω_m that does not satisfy the condition (3.21), we will first set $\mu_{k,m} = 0$, and $\gamma_{k,m} = 0$. Our problem is then to assign relay power to the remaining subcarriers which do satisfy it. From condition (3.22), it is obvious that for those subcarriers which satisfy (3.21), the minimum power required to obtain better capacity via relaying is $\gamma_{k,m}$. Therefore, as a first step, for all subcarriers on which we want to relay, we must select $\mu_{k,m} = 1$ and set $P_{R,k,m} \geq \gamma_{k,m}$ so that can we get better capacity through relaying. In order to maximize global capacity satisfying these constraints, we use a pre-processing heuristic. We first want to make sure that we can improve capacity in all subcarriers while satisfying global power constraints. We achieve this by successively setting $\mu_{k,m} = 0$ and $\gamma_{k,m} = 0$ for the subcarrier with the largest $\mu_{k,m}\gamma_{k,m}$ until $\sum_{m=1}^{M}\sum_{k\in\Omega_m}\mu_{k,m}\gamma_{k,m} \leq P_R$.

After this, we follow this optional step. We first calculate the relay power values $P_{R,k,m}$ that maximize the capacity. We then set $\mu_{k,m} = 0$ and $\gamma_{k,m} = 0$ for the subcarrier with the largest $\mu_{k,m}\gamma_{k,m}$ and recalculate $P_{R,k,m}$ values that maximize the capacity. We repeat this as long as this iterative process increases capacity. This iterative approach is guaranteed to give better capacity, however, with additional computational cost.

We determine the values of $P_{R,k,m}$ that maximize capacity using convex optimization techniques [61] to obtain the optimal power allocation for all subcarriers for which $\mu_{k,m} = 1$. The optimization sub-problem for the relay is formulated as follows:

$$\begin{aligned}
\text{minimize} \quad & -\frac{1}{2}\sum_{m=1}^{M}\sum_{k\in\Omega_m}\mu_{k,m}\log_2(1+\rho_{k,m}) \\
\text{subject to} \quad & P_{R,k,m} - \gamma_{k,m} \geq 0 \quad \forall k, m \\
& \sum_{m=1}^{M}\sum_{k=1}^{K} P_{R,k,m} - P_R = 0.
\end{aligned} \tag{3.23}$$

Using the KKT conditions for this convex optimization problem, the solution is given by:

$$P^*_{R,k,m} = \begin{cases} 0, \text{ when } \mu_{k,m} = 0, \text{ or } k \notin \Omega_m \\ \max\left\{\gamma_{k,m}, \frac{1+P_{S,k,m}a_k}{b_{k,m}(1+P_{S,k,m}a_k+P_{S,k,m}c_{k,m})}[\cdot]\right\}, \text{ otherwise} \end{cases}$$

where

$$[\cdot] = \frac{P_{S,k,m}a_k}{2}\sqrt{1+\frac{4b_{k,m}(1+P_{S,k,m}a_k+P_{S,k,m}c_{k,m})}{\nu\ln(2)P_{S,k,m}a_k(P_{S,k,m}a_k+1)}} - (2P_{S,k,m}c_{k,m}+P_{S,k,m}a_k+2). \tag{3.24}$$

The parameter ν is chosen such that the sum power constraint in (3.23) is fulfilled.

3.3.3 Power Allocation at the Source for the First Time Slot

Using the relay decision $\mu^*_{k,m}$ and power values at the relay $P^*_{R,k,m}$ calculated in the previous section, we now determine the source power allocation $P^*_{S,k,m}$ for all subcarriers $k \in \Omega_m$ in each set Ω_m. However, to reduce complexity, we do not add any restrictions on source power as it was taken in [57] in order to ensure the gain in capacity via relaying i.e., to satisfy (3.19). With simulations we will show that convergence in relay decisions is still achieved very quickly without affecting the performance.

The power allocation is performed in a similar way as in the previous section. Using the relay decisions $\mu_{k,m}$ and relay power $P_{R,k,m}$ values found in Section 3.3.2, we can obtain the optimal source power vector by solving the following optimization sub-problem:

$$\begin{aligned} \text{minimize} & -\frac{1}{2}\sum_{m=1}^{M}\sum_{k\in\Omega_m}\mu_{k,m}\log_2(1+\rho_{k,m}) \\ & \qquad + (1-\mu_{k,m})\log_2(1+P_{S,k,m}c_{k,m}) \\ \text{subject to} & \quad P_{S,k,m} \geq 0 \quad \forall k,m \\ & \sum_{m=1}^{M}\sum_{k=1}^{K}P_{S,k,m} - P_S = 0. \end{aligned} \tag{3.25}$$

When we solve the KKT conditions for this convex optimization problem, we get a cubic equation in $P_{S,k,m}$ for subcarriers where relaying is on. However, if we simplify these conditions ignoring the second term in equation (3.2) (i.e weak direct link for relaying subcarriers), we will obtain the following solution:

$$P^*_{S,k,m} = \begin{cases} \max\left\{0, \frac{1}{\nu \ln 2} - \frac{1}{c_{k,m}}\right\}, & \text{for } \mu_{k,m} = 0 \\ \max\left\{0, \frac{1}{a_k}\left([\cdot]\right)\right\}, & \text{for } \mu_{k,m} = 1 \\ 0, & \text{when } k \notin \Omega_m \end{cases} \tag{3.26}$$

where $[\cdot] = -1 + \frac{b_{k,m}P_{R,k,m}}{2}\left(-1 + \sqrt{1 + \frac{2a_k}{\nu \ln(2) b_{k,m} P_{R,k,m}}}\right)$. The parameter ν is chosen so that the sum power constraint in (3.25) is fulfilled.

3.3.4 Power Allocation at the Source for the Second Time Slot

Once we get convergence in capacity and relay decisions with the solution $P_{S,k,m}$ and $P_{R,k,m}$ using steps in sections 3.3.2 and 3.3.3 iteratively, we can estimate $\hat{P}_{S,k,m}$ separately as discussed in this section. The power allocation at the source for the second time slot is the solution to following sub-problem:

$$\begin{aligned} &\text{minimize} - \frac{1}{2}\sum_{m=1}^{M}\sum_{k\in\Omega_m}(1-\mu_{k,m})\log_2(1+\hat{P}_{S,k,m}c_{k,m}) \\ &\quad\text{subject to} \quad \hat{P}_{S,k,m} \geq 0 \quad \forall k, m \\ &\sum_{m=1}^{M}\sum_{k=1}^{K}\hat{P}_{S,k,m} - P_S = 0. \end{aligned} \tag{3.27}$$

This will be standard water-filling case [13] and using KKT conditions once again, we can write the solution as follows:

$$\hat{P}^*_{S,k,m} = \begin{cases} \max\left\{0, \frac{1}{\nu \ln 2} - \frac{1}{c_{k,m}}\right\}, & \mu_{k,m} = 0 \text{ and } k \in \Omega_m \\ 0, & \mu_{k,m} = 1 \text{ or } k \notin \Omega_m. \end{cases} \tag{3.28}$$

The parameter ν is chosen so that the sum power constraint in (3.27) is fulfilled.

3.4 Numerical Results

In this section, we evaluate the performance of our proposed resource allocation schemes for OFDMA based cooperative communication systems based on non-regenerative relay links. In our simulations, we assume frequency selective channels that undergo Rayleigh fading and path loss. We model the channel coefficients $h_{1,k}$, $h_{2,k,m}$ and $h_{3,k,m}$ with complex normal distribution as in [23]:

$$h_{1,k}, h_{2,k,m}, h_{3,k,m} \sim C\mathcal{N}\left(0, \frac{1}{L(1+d)^{\alpha}}\right) \tag{3.29}$$

where L is the number of channel taps in time domain, d denotes the distance between two nodes, and α is the path-loss exponent. In our simulations, we take number of channel taps L to be 4, path loss exponent α as 3 and we assume that the distance between source and users is 1000m with the relay placed in the middle. We define the average SNR between source and destination nodes as $\frac{P_S}{KL(1+d)^{\alpha}\sigma_m^2}$. If not specified, we choose this average SNR close to 0 dB, $P_S = P_R$, number of subcarriers $K = 64$ and noise variances $\sigma_m^2, \sigma_r^2 = 4.14 \times 10^{-10}$. We also take proportional rate ratios to be 1 (max-min fairness) unless stated otherwise.

Convergence

We first investigate the convergence of the proposed heuristic scheme by varying the number of iterations between Sections 3.3.2 and 3.3.3. To check the convergence of the algorithm, we look at both relaying decisions and average capacity as we vary the number of iterations. In Fig. 3.1, we plot the average normalized capacity obtained using identical channel gains with iterations. The number of subcarriers were K = 64 and the number of users were M = 6. Fig. 3.1 clearly shows that with just 3 iterations, we can achieve 99% of maximum achievable capacity. We look at the convergence of relaying decisions in Table 3.1. For this case, we choose K = 16 and M = 6. Table 3.1 shows a sample snapshot of relaying decisions as they vary with the number of iterations. Each row in the table shows the relaying decision vector per iteration. Clearly after just 3 iterations almost the all relaying decisions remain constant.

Table 3.1: Convergence of relaying decisions

Iteration Number	Relay Decisions μ per subcarrier (ordered with subcarrier index)
1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2	1 1 0 0 1 1 1 0 0 0 0 0 0 1 1 1
3	1 1 0 0 1 1 0 0 0 0 0 0 1 1 1 1
4	1 1 0 0 1 1 1 0 0 0 0 0 1 1 1 1
5	1 1 0 0 1 1 1 0 0 0 0 0 1 1 1 1
6	1 1 0 0 1 1 0 0 0 0 0 0 1 1 1 1
7	1 1 0 0 1 1 1 0 0 0 0 0 1 1 1 1
8	1 1 0 0 1 1 1 0 0 0 0 0 1 1 1 1
9	1 1 0 0 1 1 0 0 0 0 0 0 1 1 1 1

Capacity vs. number of users

We compare the average data rate per subcarrier as we vary the number of users for both selective and always relaying scenario in Fig. 3.2. As can be seen from Fig. 3.2, the proposed selective relaying scheme gives much better capacity compared to the "always relay" scenario. Also, as the number of users increases, the overall average capacity per subcarrier also increases for both the always and the selective relay schemes.

Fairness

In Fig. 3.3, we plot the average data rates of the users to show the distribution of data rates. We choose number of subcarriers $K = 64$ and number of users, $M = 6$. We set the desired proportional fairness as $C_1 : C_2 : C_3 : C_4 : C_5 : C_6 = 2 : 1 : 1 : 3 : 1 : 1$. The figure shows each users' data rates that can be achieved with proposed scheme with fairness and selective relaying, proposed scheme with maximum-achievable data rate and selective relaying, "always relay" case (without fairness), without any relaying (without fairness), and equal power allocation. It is obvious from the results that while other schemes cannot achieve fairness our proposed scheme with the fairness constraints can roughly guarantee the proportional rate at an expense of marginal loss in overall capacity.

3.5 Conclusion

We have proposed a suboptimal resource allocation algorithm that maximizes the capacity of a multiuser OFDMA-based cooperative communication system. We assume that the source employs a selective relaying mechanism to adaptively decide which subcarriers to use for relaying. We have shown that by carefully choosing the subcarriers to be used for relaying, we can improve the capacity. The heuristic solution that we have proposed is composed of several smaller subproblems. First, we suboptimally allocate subcarriers to different users to achieve an approximate proportional rate fairness amongst users. Next, we follow a two-step iterative scheme where first steps decides which subcarriers to use for relaying and allocates corresponding powers at the relay and the second step allocates powers at the source in the first hop. Using simulations, we have shown that convergence in terms of capacity and relaying decisions can be obtained in a few iterations. Finally, we allocate powers at the source for all non relaying subcarriers during the second-phase with a water-filling algorithm. Numerical results showed that this selective relaying mechanism performs better than the "always relay" case even in low SNR regions. In addition, fairness constraints were also approximately guaranteed. Further improvement can be made using subcarrier pairing, i.e., matching the best subcarriers in the first and second time slot.

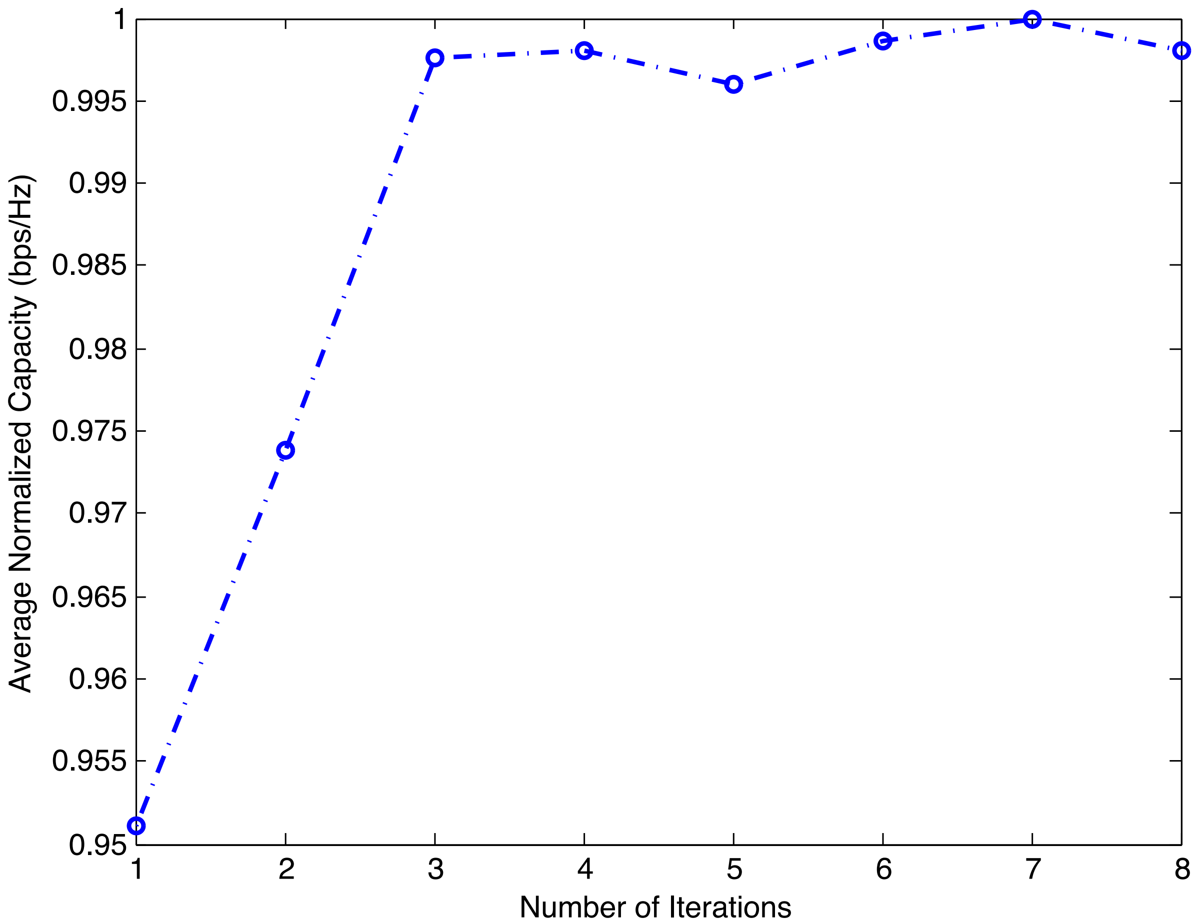


Figure 3.1: Convergence of data rates with number of iterations.

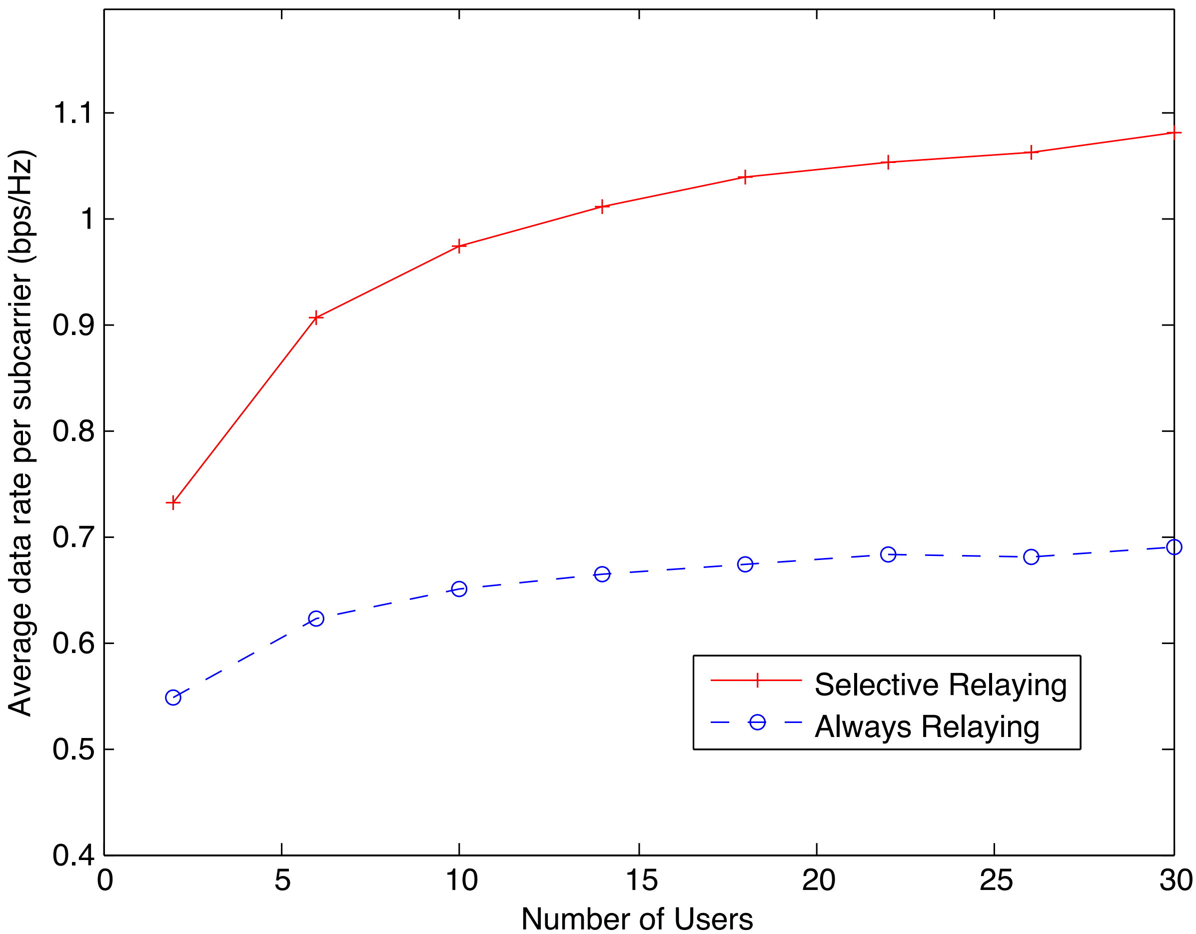


Figure 3.2: Average data rate per subcarrier vs. number of users.

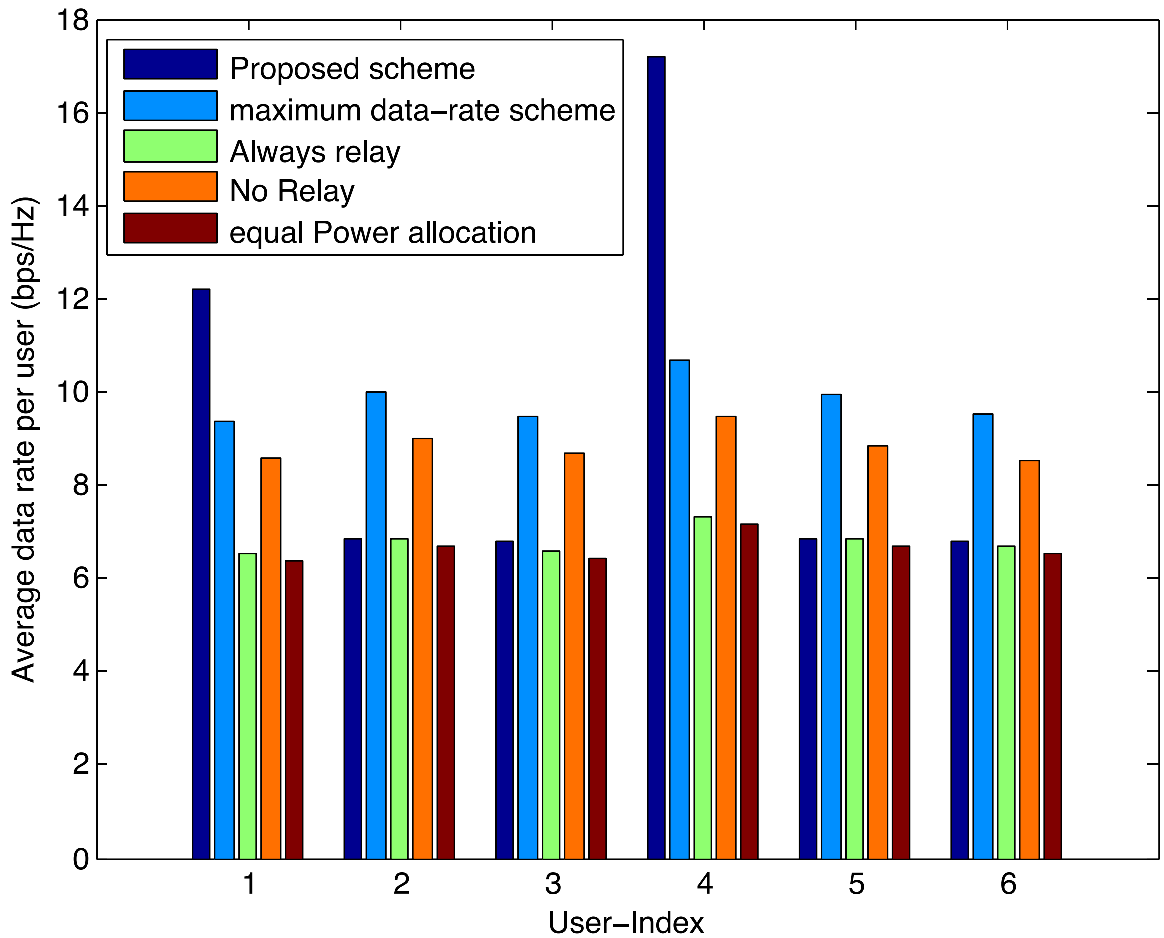


Figure 3.3: Average data rate per user vs. user (no. of users = 6, no. of subcarriers = 64).

Chapter 4

Relay Selection for OFDM Wireless Systems under Asymmetric Information: A Contract-Theory Based Approach

4.1 Introduction

Relay-assisted cooperation in wireless networks plays a key role in improving the overall efficiency of wireless networks by improving the system throughput, energy efficiency, spectrum usage, coverage, channel reliability and network cost reduction via spatial multiplexing and achieving diversity gains. Relay-based cellular network architectures have also been considered for next generation wireless systems such as 3GPP Long Term Evolution (LTE) and IEEE 802.16j mobile WiMAX [52, 53].

Cooperation with the help of relays can potentially assist a source node by forwarding its data to the destination either by amplify-and-forward (AF) or by decode-and-forward (DF) relay protocols. This cooperation can be achieved either by installing fixed relays or through user cooperation. User cooperation although improves performance of wireless systems, it requires incentives for the potential cooperating nodes to spend their energy acting as relays. Moreover, these potential relays are better informed than the source about their transmission costs, which depend on the exact channel conditions or CSI on their relay-destination links. This results in asymmetry of available information between the source and the relays. The potential relay nodes are usually selfish and hence may not be willing to cooperate without any additional incentives. In this chapter, we address the problem of relay selection under asymmetric information together with incentive-based mechanisms for OFDM-based cooperative systems. The objective of this part of the thesis is to propose simple pricing-based incentive mechanisms with minimal signalling overheads that can be employed in practical wireless systems.

Orthogonal Frequency Division Multiplexing (OFDM) has been adopted by many modern communication systems as standard multi-carrier modulation technology due to its ability to handle severe channel conditions without complex equalization filters. In a dual-hop OFDM-based cooperative communication system, choice of a relaying protocol, i.e., the interface between source-relay land the relay-destination links, is an important factor in defining the performance and complexity of such systems [62]. The authors in [62] present and analyze a comprehensive set of relaying protocols namely; time-domain versus frequency-domain processing, resource block-wise versus symbol-wise processing, AF versus DF, reordering/pairing of resource blocks, buffering over fading states, and optimization of time sharing. In this chapter, we consider a OFDM-based dual-hop cooperative communication system that uses DF relaying due to its obvious performance advantages over AF relaying. We assume no buffering and equal time sharing for the two hops because of the low complexity and simple analysis. Pairing of subcarriers has not been considered explicitly because under the assumptions that we discuss later, this information is not required for the proposed solution to the relay selection problem under asymmetric information.

We use a simple *principal-agent model* from microeconomics for source and relay, where source acts as the principal and relay is an agent [63]. In such a model, the bargaining power is kept with the principal, and the agent can either accept or decline an offer proposed by the principal. Such a model reduces the interaction needed between a source and a relay, and only the users willing to relay can participate. These offers or *contracts* proposed to relays are in the form of a menu of renumeration or monetary transfers for a specified service in each subcarrier, which in this chapter is taken as desired signal-to-noise ratio (SNR) at destination. Assuming that the source has limited information about the relay-agents (e.g. joint probability distribution of CSI), we use "contract theory" to design these set of *contracts* which are designed for different *types* of users (we will elaborate on "types" later on in this chapter). Contract theory is a field of economics that studies how economic players or agents create mutually agreeable contracts or arrangements in presence of asymmetric or incomplete information [64]. Contract-based solutions for spectrum sharing in wireless systems have recently been discussed in [65], [66], [67], and [68]. The challenge behind contract-based approach for wireless systems is due to the fact that the agents can lie to principal about their individual information in order to increase their utility and hence a contract in this situation tries to create incentives for agents to report their information truthfully.

Once the prospective relays confirm to the source which contracts (i.e. payments and SNRs) they are willing to accept for each subcarrier, the only problem that remains is to select appropriate relays in each subcarrier such that the overall capacity is maximized for the source under overall budget constraint. Without designing these contracts and having no information about exact channel conditions of the potential relays, it would be difficult for the source to optimally choose relays and offer them suitable renumeration. This is because relays can lie about their channel conditions and hence the source could possibly make inefficient payments with unsatisfactory performance. As we will see later in this chapter, this relay selection now becomes a nonlinear non-separable convex knapsack problem which cannot be solved with reasonable computational complexity and we therefore suggest a heuristic method to solve it. We then compare the performance of our overall mechanism and heuristic relay selection solution with a simple relay selection scheme and show that not only the proposed solution performs better but also it is near optimal under most general settings. The proposed scheme is very simple to implement in practical systems, requires almost no information about relays at source, limits the computational overheads only at the source and requires very limited interaction with potential relays.

A list of terms and definitions used in the chapter is given in Table 4.1. The organization of the chapter is as follows. We will first present our system model, problem description and solution approach in Section 4.2. The utility models for contract design will be discussed in Section 4.3. In Section 4.4.1, we will discuss a contract formulation under a complete or perfect information. In Section 4.4.2, we will show how to obtain an optimal contract design under asymmetric information. Relay section under a budget constraint will be discussed in Section 4.5. Numerical results will be presented in Section 4.6 and conclusions will be drawn in Section 4.7.

4.2 System Model

4.2.1 Problem Description

We consider a typical cooperative network scenario in which a particular mobile node acting as a source, wants to transmit a block of data to a destination node with the help of some nearby mobile nodes that can act as relays. Fig. 4.1 shows an example of such a cooperative network system. We assume that the source uses an OFDM-based multi-carrier system for the transmission technology with a total number of N subcarriers. We posit that there are M such mobile stations that could be the possible relay candidates. Based on its channel conditions, each relay node incurs a certain cost to provide a pre-specified SNR to the source on a particular subcarrier at the destination. Since there is no obligation for these mobile stations to forward its data towards the destination, the source mobile node must provide some incentives such as some monetary payment or credits, to these possible relay candidates. However, in such a system, it is practical to assume that the source is not only unaware of the number of possible relay candidates M but also the exact channel conditions on all the relay-destination links on each subcarrier. In absence of this information, the source does not know which relay nodes to choose and how much it should pay to each relaying node because relaying nodes are regular mobile users and are therefore selfish, and

Table 4.1: Terminology used

principal-agent model	microeconomic model where payoff to the principal (source) depends on an action taken by the relay-agent and the bargaining power is kept with the principal.
asymmetric information	one party (relay-agent) has more or better information than the other (source).
contract	a tuple consisting of a targeted SNR that a relay of a certain type on a particular subcarrier can provide on that subcarrier, and a corresponding payment that the source promises to make to that relay.
private information	relay agent's instantaneous channel gains between relay-destination link on all subcarriers.
type	relay agents private information on a particular subcarrier, i.e., channel gain on a relay-destination link on that subcarrier.
transfer	payment made to a relay node by a source in lieu of a targeted SNR at destination.
reservation utility	minimum utility the relay-agent will get by not accepting a contract.
incentive compatible	relay-agent chooses the contract designed for his type only.
individually rational	contract designed for each type gives the relay-agent at least as much utility as it would get by not accepting the offer.
revelation principle	Every equilibrium outcome with a mechanism is realized by a payoff-equivalent revelation mechanism that has an equilibrium where the relay-agents truthfully report their types.
single crossing property	indifference curves for two different types of relay agents cannot intersect more than once.
information rent	positive surplus that the relay receives by accepting a contract.

can potentially lie about their actual cost of transmission. The relay node's *private information*, i.e., instantaneous channel gains between relay-destination link on all subcarriers can be expressed as a vector $\mathbf{\Psi}_m = \{\theta_{1m}, \theta_{2m}, \cdots \theta_{Nm}\}$, where θ_{im} denotes the channel gain for the mth relay on the ith subcarrier for the corresponding relay-destination link. A relay node's channel gain on a relay-destination link on a particular subcarrier will henceforth be called its *type*[1] on that subcarrier and we will subsequently use the symbol θ without subscripts to indicate types in general. We assume these channel gains to be slow-varying, which means that they would remain constant for both transmission and relaying time slots.

Although the source is unaware of relay node's exact type or channel condition, we assume

[1]To avoid confusion, we would like to clarify that *type* of a relay is not to be confused with relay types in LTE systems which classifies relay nodes as type 1, 1a, 1b or 2[69]. In this chapter, usage of *type* is synonymous to the equivalent standard definitions in the theory of economics of asymmetric information [64].

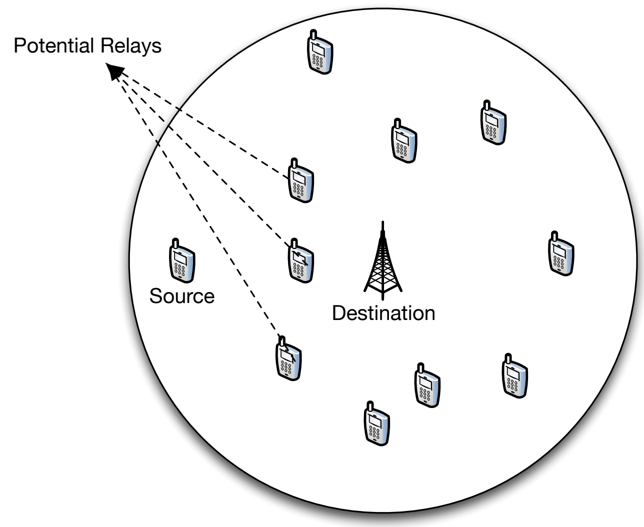


Figure 4.1: A typical cooperative wireless network system with user cooperation

that it has information about the joint distribution of the types and the set $\boldsymbol{\Theta} \subset \mathbb{R}^N_{\geq 0}$ from which these type vectors are drawn from. This is a reasonable assumption because source can learn about this distribution through the knowledge of fading environment parameters between relays and destination and these parameters could be provided to the source with limited feedback from the destination. However, as we will later observe that knowledge of this distribution only affects the optimality of the designed solution hence the source can begin transmission with just a priori belief about this distribution.

Moreover, we assume that the source has a maximum budget $\mathcal{T}$ in one time frame for the total payments or *transfers* that it can make to the relay nodes over all subcarriers. We further assume that the relays utilize space-time coded cooperative diversity for multi-relay transmission on the same subcarrier. Based on these assumptions, the problem can be described as follows: the source has to effectively choose a set of relays for every subcarrier in order to maximize its overall throughput in a given time frame, make optimal transfers to these relays without the knowledge of relays' private information, while making sure total transfers do not exceed the overall budget constraint $\mathcal{T}$.

4.2.2 Two Part Contract-based Solution

This overall problem of selecting relays on each subcarrier while providing transfers to different relays is quite difficult to solve because of the overall budget constraint, multi-dimensional information types and information asymmetry (difference in available information between the source and the potential relays). However with the help of contract theory, we will attempt to solve it by breaking the problem down into two parts:

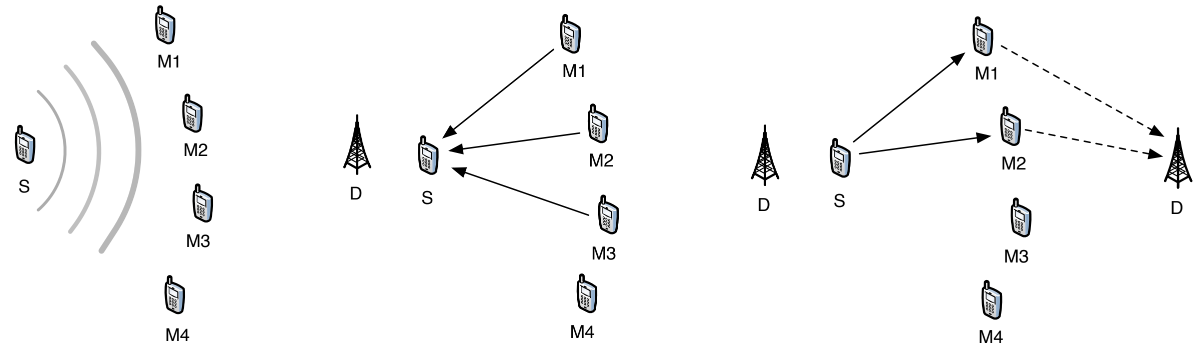
1. *Contract Design:* A *contract*, is defined as a tuple consisting of a targeted SNR at destination on a certain subcarrier that a relay can provide and a corresponding guaranteed transfer or monetary incentive that source promises to make. The source first designs a set of common contracts applicable to all subcarriers without using any specific budget constraints, and broadcasts them to all relay candidates using some established protocol. Since, the number of contracts broadcasted are independent of number of subcarriers, the signalling overhead will be small. The relays listen to the contracts offered by source then respond back identifying the contracts they are willing to accept for each subcarrier. This part also has less signalling overhead than regular communication because relays do not have to provide their actual channel feedback.

2. *Relay Selection:* Based on the accepted contracts by the relays on each subcarrier, source then chooses an optimum set of relays for each subcarrier considering its budget constraint, while maximizing its expected capacity. We assume that the source instructs the selected relays with space-time-codes for each subcarrier, and hence relays can transmit simultaneously on the same subcarrier. Each relay node hence acts as a "virtual antenna" and sends out the signal in a Multiple Input Multiple Output (MIMO) setting. All the relay transmissions hence occur in the same subcarrier and superimpose at the destination so that the overall SNR in each subcarrier is the summation of SNRs provided by all selected relays [9, 70].

An important assumption for this proposed two-part contract based solution is that the overall budget constraint $\mathcal{T}$ is sufficiently larger than the average cost of transmission for relay in each subcarrier. This assumption is reasonable because source expects the transmission to happen at least on a few subcarriers, hence source must set a budget constraint several times larger than the average cost of transmission for relay in each subcarrier. This assumption ensures that the monetary incentive or transfer for any contract pair would be sufficiently smaller that $\mathcal{T}$. Fig. 4.2 illustrates the complete representation of this mechanism in a three step process for a sample system.

4.3 Utility Models for Contract Design

As we explained in section 4.2.2, a *contract* is defined as a tuple consisting of a targeted SNR that a relay of a certain type on a particular subcarrier can provide on that subcarrier, and a corresponding payment that the source promises to make to that relay. For the purpose of contract design, we will focus our analysis for a general relay type θ (channel gain θ on the relay-destination link) on any subcarrier. To begin with, source first needs to design a set of common contract pairs $(\gamma(\theta), t(\theta))$ for a given range of types θ that are *incentive compatible* (IC) and *individually rational* (IR) on all



(a) Step 1: Source S broadcasts a set of contracts to all nearby relays M1, M2, M3, M4.

(b) Step 2: Relays M1, M2 and M3 accept certain contracts on certain subcarriers and respond to source indicating which contracts they are willing to accept.

(c) Step 3: Source selects relays M1 and M2, makes transfers, sends instructions and messages, and relays help transmit to destination D.

Figure 4.2: A contract based cooperative communication and relay selection mechanism

subcarriers. Here, $\gamma(\theta)$ is the SNR that the relay of type θ can provide at the destination and $t(\theta)$ is the transfer (e.g. monetary incentive) that the source node makes to the relay for a particular subcarrier. *Incentive compatible* means that the relay-agent chooses the contract designed for his type only. *Individually rational* means that the contract designed for each type gives the relay-agent at least as much utility as it would get by not accepting the offer or in other words, by not relaying. This minimum utility is also known as *reservation utility* and we will take it as 0 in the rest of the chapter. Using the *revelation principle*, we can focus our analysis on contract designs where agents declare their types truthfully or in other words, we can directly consider types while designing contracts [64]. We will now make a practical assumption that the relay nodes do not have a prior knowledge of source's budget $\mathcal{T}$ and the number of other relay candidates. In other words, the relay nodes have no way of knowing that whether the contract they will accept will be executed or not by the source. Assuming DF relaying with repetition coding, the SNR of the source-relay-destination link on a certain subcarrier will be given by $\min\{\gamma_{SR}, \gamma_{RD} + \gamma_{SD}\}$, where γ_{SR} is source-relay SNR and γ_{RD} is relay-destination SNR [9]. Ignoring the direct path, the source-relay link can be safely assumed to be stronger because source will only consider nearby mobile nodes who can decode the source information, as relay candidates and hence γ_{RD} will be the bottleneck [65]. Moreover, due to the assumption that source-relay link is always stronger, subcarrier pairing has also been ignored in the current problem formulation. Therefore from now on, we will approximate γ_{RD} as the net SNR for the overall link and will drop the subscript and call it just γ henceforth.

We now define a quasi-separable utility for the source with a contract pair $(\gamma(\theta), t(\theta))$ on each subcarrier as follows:

$$U(\gamma(\theta)) - t(\theta) \tag{4.1}$$

where $U(\cdot)$ is a concave function that gives the utility that the source gets with an SNR γ between a relay and destination on some subcarrier, when it is using only this particular relay, and this utility can be given by Shannon capacity formula:

$$U(\gamma) = \frac{1}{2}\log_2\left(1 + \min\{\gamma_{SR}, \gamma + \gamma_{SD}\}\right) \approx \frac{1}{2}\log_2\left(1+\gamma\right). \tag{4.2}$$

Here, the half factor is used to account for half duplexity of the relaying protocol and approximation is based on the argument in the previous paragraph. Notice that γ's and t's are functions of θ only in the sense that the contract pair is designed for a relay of type θ. However, a relay of θ is free to choose any other contract pair $(\gamma(\theta'), t(\theta'))$ but the relay is truthful off it accepts the pair $(\gamma(\theta), t(\theta))$. The overall utility of a relay candidate of type θ that announces its type truthfully can also be described by a quasi-separable function given by the difference between transfers and cost of transmission:

$$t(\theta) - \mathcal{C}(\gamma(\theta), \theta) \tag{4.3}$$

where $\mathcal{C}(\gamma, \theta)$ is the cost for relay of type θ to provide SNR γ at the destination on some subcarrier. This cost could be the summation of the cost of per unit power used for relaying in addition to fixed decoding costs. Ignoring the decoding costs, this cost can simply be given by:

$$\mathcal{C}(\gamma, \theta) = \frac{c\gamma}{\theta} \tag{4.4}$$

where c is a positive number denoting cost per unit power and $\frac{\gamma}{\theta}$ is the transmitted power by the relay. Conveniently, since $\frac{\partial^2 \mathcal{C}(\gamma,\theta)}{\partial\gamma\partial\theta} < 0$, the cost function satisfies a form of Spence-Mirrlees Condition or *single crossing property* [64]. What this means is that indifference curves (γ, t) (plot of contracts for which a relay gets constant utility) for two different types of relay agents cannot intersect more than once. Moreover, the economic significance of this condition is that the relay agents of higher types are willing to provide better SNR γ for a smaller increase in transfer t. In the following sections, we will see how to solve the problem of contract design for two specific cases: complete information scenario and incomplete information scenario.

4.4 Contract Formulation under Complete and Asymmetric Information

4.4.1 Contract Formulation under Complete Information: First-best scenario

This is the scenario when the source has the precise information of the relay types vector $\mathbf{\Psi}_m$, i.e., CSI between relay-destination link in each subcarrier. This is also known as the first-best scenario and will be used for benchmarking, because this is the ideal case for source, as it is in the best position to make the maximum use of the available potential relays. Therefore in this scenario, source only has to ensure that each relay agent is ready to accept the contract that he is about to offer, or in other words he only has to satisfy relay-agent's individual rationality condition in each subcarrier. The source's objective problem for a relay-agent type of θ in a certain subcarrier can now be written as:

$$\max_{\gamma(\theta),t(\theta)} U(\gamma(\theta)) - t(\theta) \tag{4.5}$$

$$\text{subject to} \qquad t(\theta) - \frac{c\gamma(\theta)}{\theta} \geq 0. \tag{4.6}$$

The source hence gives the relay-agent zero utility in order to maximize its own utility, i.e., the source extracts all the surplus from the relay-agent. Therefore, setting (4.6) to equality, substituting in (4.5) and then differentiating the objective w.r.t γ and finally equating it to zero, the optimal first-best contract $(\gamma(\theta), t(\theta))$ for a relay of type θ is given by $(\frac{\theta}{2c\ln 2} - 1, \frac{1}{2\ln 2} - \frac{c}{\theta})$.

4.4.2 Contract Formulation under Asymmetric Information: Second-best scenario

Theoretical analysis with continuous relay-agent types

In this section, we will analyze the structure of the solution using some standard theoretical analysis. In this case, we assume that the types of relays for all subcarriers are continuous and belong to a set $\Theta = [\underline{\theta}, \bar{\theta}]$ and has a joint probability distribution $f(\theta_1, \theta_2, \cdots \theta_N)$ (with $F(\theta_1, \theta_2, \cdots \theta_N)$ as cumulative density function), which is known to the source node. Let $\mathcal{P}(\hat{\theta}, \theta)$ be the profit or utility achieved by relay agent of type θ on a certain subcarrier who announces his type as $\hat{\theta}$. The profit is given by the following function:

$$\mathcal{P}(\hat{\theta}, \theta) = t(\hat{\theta}) - \mathcal{C}(\gamma(\hat{\theta}), \theta). \tag{4.7}$$

The contract $(\gamma(\theta), t(\theta))$ satisfies the incentive constraints (IC) if and only if being truthful gives a relay node at least as much utility as it gets by lying, i.e.,

$$\mathcal{P}(\theta, \theta) \geq \mathcal{P}(\hat{\theta}, \theta), \quad \forall(\theta, \hat{\theta}) \in \Theta^2 \qquad \text{(IC)}. \tag{4.8}$$

Hence, for the contract to be *incentive compatible*, the following first and second order conditions must hold:

$$\forall \theta \in \Theta, \begin{cases} \frac{\partial \mathcal{P}(\hat{\theta},\theta)}{\partial \hat{\theta}}|_{\hat{\theta}=\theta} = 0 & \text{(IC}_1\text{)} \\ \frac{\partial^2 \mathcal{P}(\hat{\theta},\theta)}{\partial \hat{\theta}^2}|_{\hat{\theta}=\theta} \leq 0 & \text{(IC}_2\text{)}. \end{cases} \tag{4.9}$$

Substituting for $\mathcal{P}(\hat{\theta}, \theta)$, these conditions can be simplified to

$$\forall \theta \in \Theta, \begin{cases} \frac{dt(\theta)}{d\theta} = \frac{c}{\theta} \frac{d\gamma(\theta)}{d\theta} & \text{(IC}_1\text{)} \\ \frac{d\gamma(\theta)}{d\theta} \geq 0 & \text{(IC}_2\text{)}. \end{cases} \tag{4.10}$$

This means that first both $\gamma(\theta)$ and $t(\theta)$ must be increasing in type θ (by IC_2) and secondly, IC_1 tells us how the increase in transfers w.r.t to the agent types are related to increase in deliverable SNR. Let $\rho(\theta)$ denotes the utility of the relay agent of type θ with the optimal truthful contract, i.e, a mechanism where relay chooses contract designed for his type only. Then, $\rho(\theta)$ can simply be given by $\mathcal{P}(\theta, \theta)$, i.e.,

$$\rho(\theta) = t(\theta) - \frac{c\gamma(\theta)}{\theta}. \tag{4.11}$$

Using IC_1, we can compute that

$$\frac{d\rho}{d\theta} = \frac{dt(\theta)}{d\theta} - \frac{c}{\theta} \frac{d\gamma(\theta)}{d\theta} + \frac{c\gamma(\theta)}{\theta^2} = \frac{c\gamma(\theta)}{\theta^2} \tag{4.12}$$

which is positive and implies that $\rho(\theta)$ is an increasing function of θ and hence the higher types benefit with higher returns. Assuming relay agent's *reservation utility* to be 0, its *individual rationality* (IR) condition can therefore simply be given by:

$$\rho(\underline{\theta}) = 0 \qquad \text{(IR)}. \tag{4.13}$$

This is because making transfers is costly to the source node and since higher type relay nodes have higher returns, source just has to give zero utility to the lowest type $\underline{\theta}$ to satisfy the IR condition.

Using equations (4.11), (4.12), and (4.13), we can hence write

$$t(\theta) = \frac{c\gamma(\theta)}{\theta} + \int_{\underline{\theta}}^{\theta} \frac{c\gamma(\tau)}{\tau^2} d\tau. \tag{4.14}$$

Now the source's objective is to maximize the expected utility which is given as:

$$\int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\bar{\theta}} \cdots \int_{\underline{\theta}}^{\bar{\theta}} \sum_{n=1}^{N} (U(\gamma(\theta_n)) - t(\theta_n)) f(\theta_1, \theta_2, \cdots, \theta_N) d\theta_1 d\theta_2 \cdots d\theta_N. \tag{4.15}$$

Proposition 2. *We can rewrite source's optimization problem as follows:*

$$\begin{aligned} \max_{\gamma(\theta)} \sum_{n=1}^{N} \int_{\underline{\theta}}^{\bar{\theta}} \left(U(\gamma(\theta)) - \frac{c\gamma(\theta)}{\theta} - \frac{c\gamma(\theta)}{\theta^2} \frac{1 - F_n(\theta)}{f_n(\theta)} \right) f_n(\theta) d\theta \\ \textit{subject to } IC_2 \textit{ or } \frac{d\gamma(\theta)}{d\theta} \geq 0 \ \textit{ (i.e. } \gamma \textit{ is increasing)} \\ \textit{and } \gamma \geq 0, \textit{ (SNR must be positive)} \end{aligned} \tag{4.16}$$

where $f_n(\theta)$ is the marginal probability distribution and $F_n(\theta)$ is corresponding cumulative distribution of types in the nth subcarrier.

Proof. Proof is provided in Appendix. □

The optimization problem in (4.16) can be interpreted as follows: the source has to maximize the expression in the brackets of the objective function subject to the constraint that $\gamma(\theta)$ is positive and increasing in θ, where the first two terms of the objective are same as in the source's optimization problem under complete information scenario and the last term measures the impact of incentive problem. In order to solve the optimization problem in (4.16), first we can just try to do pointwise maximization of the objective function at each θ. However, if pointwise maximization at each θ in (4.16) does not give us an increasing $\gamma(\theta)$ function, then we can use resort to optimal control theory.

Using this continuous case as a reference, we will now focus our analysis to the case where the types are considered to be discrete values rather than taken from a continuous set. This case is of more practical interest because the contracts can be easily broadcasted as finite number of values. In the next subsection, we will design incentive compatible contracts by approximating the continuous distribution by a discrete distribution with finite points. With the discussion above, we have a good idea about the structure of the optimal contract and we will notice some parallels when we discuss the discrete agent types case in the following subsection.

Solution with discrete relay-agent types

We will now see how to solve a more practical problem of designing contracts for relay agents with discrete types. This problem is more practical because the number of contracts are finite and can be transmitted to relay-agents in real-time. We quantize the set of types $\Theta = [\underline{\theta}, \bar{\theta}]$ with a quantization factor K such that the collection of types are reduced to a discrete set of K types, i.e., $\Theta = \{\delta_1, \delta_2 \cdots \delta_K\}$. Without loss of generality, we can assume that $\delta_1 < \delta_2 < \cdots < \delta_K$. We consider the quantization process to be uniform with equidistant values, i.e., $\delta_k = \underline{\theta} + \frac{k-1}{K}(\bar{\theta} - \underline{\theta})$, and if Θ is unbounded above, then $\bar{\theta}$ can be replaced by the upper limit of a desired confidence level. We chose quantization to be uniform mainly because of its ease of implementation and a closer representation of continuous distribution, however, in general a non-uniform quantization process can also be chosen depending upon how sensitive the cost function is to the variation in types. Using forward difference method, the probability that a relay-agent could be of type δ_k in nth subcarrier is given by $\pi_{kn} = P(\delta_k \le \theta_n < \delta_{k+1}) = F_n(\delta_{k+1}) - F_n(\delta_k)$ (δ_{K+1} can be replaced by $\bar{\theta}$) with $\sum_{k=1}^{K} \pi_{kn} = 1$. We assume that source is aware of this distribution on all subcarriers.

The objective of the source is to maximize its expected utility by designing an incentive compatible and individually rational optimal contract $(\gamma(\delta_k), t(\delta_k))$ (for simplicity, we will now refer it as (γ_k, t_k)) for each $\delta_k \in \Theta$, i.e.,

$$\max_{\gamma(\theta), t(\theta)} \sum_{n=1}^{N} \mathbf{E}_n[U(\gamma(\theta)) - t(\theta)] = \max_{\gamma_k, t_k \forall k} \sum_{n=1}^{N} \sum_{k=1}^{K} \pi_{kn}(U(\gamma_k) - t_k). \tag{4.17}$$

The *individual rationality* condition for this discrete scenario can now be given by:

$$t_k - \frac{c\gamma_k}{\delta_k} \ge 0, \quad \forall \delta_k \in \Theta \tag{4.18}$$

and the *incentive compatibility* condition is given by:

$$t_k - \frac{c\gamma_k}{\delta_k} \ge t_j - \frac{c\gamma_j}{\delta_k}, \quad \forall \delta_k, \delta_j \in \Theta. \tag{4.19}$$

Theorem 1. *For the optimal solution, the individual rationality condition for the lowest type is binding, i.e., $t_1 - \frac{c\gamma_1}{\delta_1} = 0$ and the others can be ignored.*

Proof. For any $\delta_k \in \Theta$ using the IC condition from equation (4.19) we can write that

$$t_k - \frac{c\gamma_k}{\delta_k} \ge t_1 - \frac{c\gamma_1}{\delta_k} \ge t_1 - \frac{c\gamma_1}{\delta_1} \tag{4.20}$$

since $\delta_k > \delta_1 \ge 0$ and $c\gamma \ge 0$. Therefore, if IR for δ_1 is inactive, so will be IR for δ_k. Hence, all the other IRs except for δ_1, can be ignored. Now, if IR for δ_1 is not binding, then all transfers t_k's can be reduced by the same amount, having no effect on IC and hence increasing source's utility. □

Theorem 2. *For the optimal solution,* $0 \leq \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_K$*, and all the downward adjacent ICs are binding and others can be ignored, i.e.,*

$$t_k - \frac{c\gamma_k}{\delta_k} = t_{k-1} - \frac{c\gamma_{k-1}}{\delta_k}, \quad \forall k \geq 2. \tag{4.21}$$

Proof. Proof is provided in Appendix. □

Now using theorems 1 and 2, the optimization problem in (4.17), (4.18), and (4.19) can be reduced to:

$$\begin{aligned} &\max_{\gamma_k \forall k} \sum_{n=1}^{N} \sum_{k=1}^{K} \pi_{kn}(U(\gamma_k) - t_k) \\ \text{s.t. } t_1 = \frac{c\gamma_1}{\delta_1}, t_k &= \frac{c\gamma_1}{\delta_1} + \sum_{i=2}^{k} \frac{c(\gamma_i - \gamma_{i-1})}{\delta_i} \\ &\text{and } 0 \leq \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_K. \end{aligned} \tag{4.22}$$

Proposition 3. *The optimization problem in (4.22) can be rewritten as follows:*

$$\max_{\gamma_k \forall k} \sum_{n=1}^{N} \sum_{k=1}^{K} \pi_{kn} g_n(\gamma_k) \ \ \textit{s.t.} \ \ 0 \leq \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_K, \tag{4.23}$$

where

$$g_n(\gamma_k) = \begin{cases} U(\gamma_k) - \frac{c\gamma_k}{\delta_k} - c\gamma_k \left(\frac{1}{\delta_k} - \frac{1}{\delta_{k+1}}\right)\left(\frac{1 - \sum_{i=1}^{k} \pi_{in}}{\pi_{kn}}\right), & \forall k < K \\ U(\gamma_k) - \frac{c\gamma_k}{\delta_K} \ \textit{if}\ k = K. \end{cases} \tag{4.24}$$

Proof. Proof is provided in Appendix. □

Now because of the concavity assumption on $U(\cdot)$, this problem can easily be transformed into a simple convex optimization problem and can be solved using standard methods. In fact, it can be shown that the inequality constraint of increasing γ's can also be neglected, and hence point-wise maximization of each $\sum_{n=1}^{N} g_n(\gamma_k)$ is sufficient here. Once we calculate contract pairs (γ_k, t_k) $(k \leq K)$, it can be easily verified that a relay that has a type θ in a certain subcarrier will automatically select contract k, if $\delta_k \leq \theta < \delta_{k+1}$, because this contract will maximize its utility.

The objective of the contract design so far was to design incentive compatible and individually rational offers that the interested relay-agents can accept without revealing their types or channel information directly to the source. This helps the source to segregate the relays in terms of their abilities to deliver certain SNRs at the destination and the price at which they are willing to do so. Hence, the source can select relays based on his budget constraint and we discuss this relay selection problem and its solution in next section in detail.

4.5 Relay Selection under a Budget constraint

In this section, we will discuss the relay selection procedure for the source under the budget constraint with either complete or incomplete information. While in perfect information scenario, the source is aware of the contracts acceptable by each relay in each subcarrier, in case of imperfect information, source broadcasts the contract menu to all relays and each relay responds with a desired contract pair for each subcarrier as discussed in section 4.4.2 (if a certain relay is unwilling to relay in a certain subcarrier, we assume it accepts null contract (0, 0)). In either case, the source knows a contract pair (γ_{mn}, t_{mn}) that is acceptable by relay m in the nth subcarrier. Under a budget constraint $\mathcal{T}$, the objective of the source now becomes to maximize its total capacity. Let $\mathcal{M} = \{1, 2, ..., M\}$ denote the set of all the relay agents who are willing to relay while providing a certain SNR at destination in each subcarrier for a certain price that is determined by the contract for that relay agent. The objective is to obtain a subset vector $\mathbf{S} = \{\mathcal{S}_n, \forall n \leq N \,|\mathcal{S}_n \subseteq \mathcal{M}\}$ (i.e. set of selected relays in each subcarrier),

$$\begin{aligned} &\max_{\mathbf{S}} \mathcal{C}(\mathbf{S}) \\ \text{s.t. } &\sum_{n=1}^{N} \sum_{\forall m \in \mathcal{S}_n} t_{mn} \leq \mathcal{T} \\ \text{where, } \mathcal{C}(\mathbf{S}) = &\sum_{n=1}^{N} \log_2 \left(1 + \sum_{\forall m \in \mathcal{S}_n} \gamma_{mn}\right). \end{aligned} \tag{4.25}$$

This problem is a nonlinear non-separable convex knapsack problem in its current form with $\mathcal{T}$ as knapsack size, t_{mn} as weights, γ_{mn} as values of items, and $\mathcal{C}(\mathbf{S})$ as the objective function. Because of the non-linearity and non-separability of the objective function, it is difficult to find the exact solution in the existing form of this problem [71]. The standard method to obtain the optimal solution is to use branch and bound algorithm, where at each step, a series of continuous subproblems are solved to obtain upper bounds by integer relaxation of the original problem. The branch and bound algorithm is discussed in detail in [72]. The authors in [73, 74] discuss methods to obtain these upper bounds for a general class of nonlinear non-separable knapsack problems. However, branch and bound method can still have the worst case complexity of exhaustive search and for M number of subcarriers and N number of potential relays, the worst case complexity could be as high as $O(2^{MN})$, which makes the branch and bound method practically infeasible due to exponential complexity. In fact, we verified the complexity to be exponential in most cases for a very simple system with simulations. Due to lack of space and practical importance, we will not go into further detail of obtaining the exact solution. Instead, in the next subsection, we will propose a heuristic solution based on the structure of our original problem described in (4.25).

4.5.1 Heuristic Solution

Here, we will discuss a few heuristics to solve the original problem by breaking it down in smaller problems that are standard 0-1 knapsack problems. We notice that if we could divide the overall constraint $\mathcal{T}$ by allocating a budget constraint $\mathcal{T}_n$ for subcarrier n, such that $\sum_{n=1}^{N} \mathcal{T}_n = \mathcal{T}$, then the sub-problem for relay selection in subcarrier n is just a standard knapsack problem, with $\mathcal{T}_n$ as knapsack size, t_{mn} as weights, and γ_{mn} as values of items. This problem can just be written as follows:

$$\max_{\mathcal{S}_n} \sum_{\forall m \in \mathcal{S}_n} \gamma_{mn} \quad \text{s.t.} \sum_{\forall m \in \mathcal{S}_n} t_{mn} \leq \mathcal{T}_n. \tag{4.26}$$

We also notice that the objective of the problem is to maximize the product $\Pi_{i=1}^{n}(1+\sum_{\forall m \in \mathcal{S}_n} \gamma_{mn})$, subject to the budget constraints. In order to maximize this product we need to maximize SNR $\sum_{\forall m \in \mathcal{S}_n} \gamma_{mn}$ for each subcarrier, while also making sure that none of the SNR's are too low, otherwise they will minimize the product. Based on this analogy, we will attempt to solve this problem using two forms of heuristics and we combine the results to obtain the final solution.

The first heuristic is to decompose $\mathcal{T}$ into $\mathcal{T}_n$ by using certain weights profile w_n for subcarrier n and hence, $\mathcal{T}_n$ could be given by $\mathcal{T}_n = \frac{w_n \mathcal{T}}{\sum_{n=1}^{N} w_n}$. The second heuristic will be to sequentially select relays for each subcarrier as long as we are under the budget constraint. We suggest three weight profiles for the first heuristic as follows:

1. *Equal subcarrier weights (ESW)*: The simplest possible way is to choose equal weights for all subcarriers:

$$w_n^{(1)} = 1, \quad \forall n. \tag{4.27}$$

A simple way to measure two different contracts relative to each other is to compare how much is the SNR per unit price for each contract. This metric, i.e., SNR per unit price, can be defined as the efficiency and can be used to compare the subcarriers relative to each other either by averaging the efficiencies of contracts in each subcarrier or by calculating net efficiency in each subcarrier. The following weight profiles are based on this observation:

2. *Average subcarrier efficiency weights (ASW)*: If the efficiency of a contract accepted by relay

Algorithm 1 Relay Selection Algorithm

for $i = 1$ to 4 **do**
  Set $\mathcal{S}_n^{(i)} = \phi, \forall n \leq N$
  if $i \leq 3$ **then** Obtain weights $w_n^{(i)}$ by (4.27), (4.28) or (4.29) for $i = 1, 2, 3$ respectively $\forall n \leq N$.
    Set $\mathcal{T}_n = w_n^{(i)} \mathcal{T} / \sum_{n=1}^{N} w_n^{(i)}, \forall n \leq N$
    for $n = 1$ to N **do**
      Set $\tau = \mathcal{T}_n$
      for $t = 0$ to $\mathcal{T}_n$ **do**
        $\Gamma[0, t] = 0$
      end for
      for $m = 1$ to M **do**
        for $t = 0$ to $\mathcal{T}_n$ **do**
          if $t_{mn} \leq t$ and $\Gamma[m-1, t-t_{mn}] + \gamma_{mn} > \Gamma[m-1, t]$ **then**
            $\Gamma[m, t] = \Gamma[m-1, t-t_{mn}] + \gamma_{mn}$, $s[m, t] = 1$
          else
            $\Gamma[m, t] = \Gamma[m-1, t]$, $s[m, t] = 0$
          end if
        end for
      end for
      for $m = M$ to 1 **do**
        if $s[m, \tau] = 1$ **then** $\mathcal{S}_n^{(i)} := \mathcal{S}_n^{(i)} \cup \{m\}$, $\tau = \tau - t_{mn}$
        end if
      end for
    end for
  else if $i = 4$ **then**
    while $\mathcal{T} \geq 0$ **do**
      for $n = 1$ to N **do**
        $m = \arg\max_{\forall m \notin \mathcal{S}_n^{(i)}} \{\gamma_{mn}/t_{mn}\}$, $\mathcal{T} = \mathcal{T} - t_{mn}$
        if $\mathcal{T} \geq 0$ **then** $\mathcal{S}_n^{(i)} := \mathcal{S}_n^{(i)} \cup \{m\}$
        end if
      end for
    end while
  end if
  Set $\mathbf{S}^{(i)} = \{\mathcal{S}_1^{(i)}, \mathcal{S}_2^{(i)} \cdots \mathcal{S}_N^{(i)}\}$
end for
return S as given by (4.30)

m in subcarrier n is defined by $e_{mn} = \frac{\gamma_{mn}}{t_{mn}}$, then the weights w_n of nth subcarrier are obtained by the following relation:

$$w_n^{(2)} = \frac{\sum_{m=1}^{M} e_{mn}}{M}, \quad \forall n. \tag{4.28}$$

3. *Net subcarrier efficiency weights (NSW)*: In this case, weights are obtained by calculating the net efficiency of a subcarrier, i.e., the ratio of maximum achievable SNR and the corresponding maximum transfer:

$$w_n^{(3)} = \frac{\sum_{m=1}^{M} \gamma_{mn}}{\sum_{m=1}^{M} t_{mn}}, \quad \forall n. \tag{4.29}$$

Using each of the three above mentioned weight profiles, we solve the original problem iteratively by solving N standard 0-1 type knapsack problems. These knapsack problems can be solved using dynamic programming by scaling and rounding transfers t_{mn} and budget $\mathcal{T}_n$ [75]. For each of the above mentioned weight profiles, we obtain solution subset vectors $\mathbf{S}^{(1)}$, $\mathbf{S}^{(2)}$ and $\mathbf{S}^{(3)}$ for each subcarrier. For the second heuristic solution, we perform an efficiency based relay selection as follows:

4. *Sequential Subcarrier Contract Pair Allocation (SSCPA)*: In this heuristic, we allocate a relay to each subcarrier sequentially by choosing a relay that provides greatest efficiency e_{mn} and has not been allocated in that subcarrier. We repeat this process until we run out of $\mathcal{T}$ and we obtain the corresponding solution subset vector and call it $\mathbf{S}^{(4)}$.

The overall heuristic solution is hence chosen as the solution vector that gives highest capacity, i.e.,

$$\mathbf{S} = \arg \max_{\mathbf{S}^{(i)}|i=1,2,3,4} \{\mathcal{C}(\mathbf{S}^{(i)})\}. \tag{4.30}$$

Table 4.2 summarizes the overall heuristic solution and Algorithm 1 describes the entire relay selection heuristic. The net complexity of the algorithm can be calculated as follows: The SSCPA heuristic is just sorting and selecting MN contract pairs, hence the complexity can be $O(MN \log(MN))$. The three other heuristics based on weight profiles solve N knapsack problems, each upper bounded by pseudo polynomial complexity of $O(M\mathcal{T})$ (assuming $\mathcal{T}$ is rounded). Hence the overall complexity will be $O(MN\mathcal{T})$, which is much easier to handle than branch and bound algorithm for small to medium values of budget constraint.

4.6 Numerical Results

In this section, we will briefly present some of the numerical results. For simulations, we consider a system where the relay types θ are normalized, independent and uniformly distributed between 50 to 300 in each subcarrier. We consider uniform distribution for our simulations because uniform distribution is the maximum entropy probability distribution for any random variable contained in the distribution's support. It effectively means that source has no additional information about the types other than their support and is therefore a benchmark scenario. In order to generate a set

Table 4.2: List of heuristics

Heuristic Method	Description	Solution Subset Vector
ESW	Divide $\mathcal{T}$ based on weight profiles $w_n^{(1)} = 1, \ \forall n$ and solve N 0-1 knapsack problems	$\mathbf{S}^{(1)}$
ASW	Divide $\mathcal{T}$ based on weight profiles $w_n^{(2)} = \frac{\sum_{m=1}^{M} e_{mn}}{M}, \ \forall n$ and solve N 0-1 knapsack problems	$\mathbf{S}^{(2)}$
NSW	Divide $\mathcal{T}$ based on weight profiles $w_n^{(3)} = \frac{\sum_{m=1}^{M} \gamma_{mn}}{\sum_{m=1}^{M} t_{mn}}, \ \forall n$ and solve N 0-1 knapsack problems	$\mathbf{S}^{(3)}$
SSCPA	Sequential Allocation per subcarrier as per maximum efficiency	$\mathbf{S}^{(4)}$
Overall Heuristic	Combination of ESW, ASW, NSW and SSCPA	$\mathbf{S} = \arg\max_{\mathbf{S}^{(i)}\vert i=1,2,3,4} \{\mathcal{C}(\mathbf{S}^{(i)})\}.$

of first and second best contracts, we quantize the range of types with a quantization factor K to be 10. The number of subcarriers N is chosen to be 16 and the parameter c is taken to be 1. The simulation parameters and the corresponding first and second-best contracts are presented in Table 4.3. The first-best contracts are calculated when the source is completely aware of the relay-agent's discrete type (complete information). However, the first-best contracts are not incentive compatible (downward ICs do not hold) and it can be easily verified by plugging the parameters of Table 4.3 in relay-agent's overall utility. On the other hand, the second-best contracts are incentive compatible by virtue of design. The IC conditions can be verified with the corresponding second-best contracts, i.e., any relay of type θ s.t. $\delta_k \leq \theta < \delta_{k+1}$ will automatically pick the kth contract, because this contract will maximize its expected utility. Another noticeable difference between the first and second-best contract that can be seen from Table 4.3 is that under incomplete information, except for the lowest type, the source pays more to get a certain SNR than what it would have gotten for lesser price under complete information. Moreover, in case of incomplete information, the source asks for sub-efficient SNRs from all the relay types except from the highest type. This is in order to provide incentive for higher types to not to choose a lower types' contract, and the related concept is called *information rent* [76]. In simple words, information rent is the positive surplus that the relay receives and Table 4.3 clearly indicates that higher type relay gets more positive surplus for the contract designed for its type.

Next, we will evaluate the performance of four heuristics, namely ESW, ASW, NSW, SSCPA, that we suggested in section 4.5.1 with respect to each other and a few benchmarks that we will describe here. We compare the performance of these heuristics with the obvious simple solution,

Table 4.3: First and second best contracts

Relay types & distribution	First-best contract $(\gamma^{(1)}, t^{(1)})$ (γ in dB)	Second-best contract $(\gamma^{(2)}, t^{(2)})$ (γ in dB)	Information Rent $t^{(2)} - c\gamma^{(2)}/\theta$
$\delta_1 = 50, \pi_{1n} = 0.1$	(15.4490,0.7013)	(9.0401,0.1603)	0
$\delta_2 = 75, \pi_{2n} = 0.1$	(17.2510,0.7080)	(12.3131,0.2806)	0.0534
$\delta_3 = 100, \pi_{3n} = 0.1$	(18.5208,0.7113)	(14.6324,0.4008)	0.1102
$\delta_3 = 125, \pi_{4n} = 0.1$	(19.5021,0.7133)	(16.4428,0.5210)	0.1683
$\delta_3 = 150, \pi_{5n} = 0.1$	(20.3020,0.7147)	(17.9322,0.6412)	0.2271
$\delta_3 = 175, \pi_{6n} = 0.1$	(20.9773,0.7156)	(19.1990,0.7615)	0.2863
$\delta_3 = 200, \pi_{7n} = 0.1$	(21.5615,0.7163)	(20.3020,0.8817)	0.3457
$\delta_3 = 225, \pi_{8n} = 0.1$	(22.0764,0.7169)	(21.2794,1.0019)	0.4052
$\delta_3 = 250, \pi_{9n} = 0.1$	(22.5367,0.7173)	(22.1564,1.1221)	0.4649
$\delta_3 = 275, \pi_{10n} = 0.1$	(22.9528,0.7177)	(22.9528,1.2424)	0.5246

i.e., select the contracts that offer the best SNRs amongst all contracts for all subcarriers while satisfying the budget constraint. Moreover, we will also compare the performance of the proposed heuristics with respect to the solution of original problem in (4.25) with relaxed integer constraints as a benchmark. We will call these solutions as **"Best SNR contracts"** and **"Relaxed Solution"** respectively in the corresponding plots. In Figures 4.3, 4.4, and 4.5, we plot average capacity per subcarrier vs number of relay agents for three values of $\mathcal{T}$, i.e., 8, 16 and 24 respectively. The number of subcarriers N are fixed at 16 and quantization factor K is chosen to be 10. As we could notice from these plots, the heuristic SSCPA always performs better when there are fewer relay agents or when the budget constraint is large. The three other heuristics based on weight profiles, namely ESW, ASW and NSW have a very similar performance and perform better than SSCPA when number of relay agents are high and budget is not too big. The intuitive reasoning behind this observation is as follows. As the number of relay agents increase, there are more diversified contracts available per subcarrier to choose from. The ESW, ASW and NSW schemes by their inherent design try to maximize the product $\Pi_{i=1}^{n}(1 + \sum_{\forall m \in \mathcal{S}_n} \gamma_{mn})$ by splitting the budget in each subcarrier and hence maximizing the sum SNRs in every subcarrier. Under low to medium budget conditions and with high number of relay agents, these algorithms outperform SSCPA because the latter scheme could run out of budget before it could select contracts in each subcarrier. Moreover under such conditions, SSCPA performs poorly as the number of agents increase because sequential allocation may result in first choosing contracts that may need higher transfers and hence source may run out of budget too quickly without balancing sum SNRs well. This behavior can be seen numerically in Figures 4.3 and 4.4. Under high budget conditions, SSCPA scheme has more freedom to sequentially choose best and optimum contracts per subcarrier as long as the budget allows inherently improving the product $\Pi_{i=1}^{n}(1 + \sum_{\forall m \in \mathcal{S}_n} \gamma_{mn})$ while automatically balancing the SNRs per subcarrier. Fig. 4.5 demonstrates this adequately. Moreover, "Best SNR Contracts" solution not only has inferior performance compared to proposed heuristics in general, but the average capacity with this solution decreases as the number of relay agents increase. This is because the "Best SNR Contracts" solution just selects the contracts that offer the best SNRs amongst all subcarriers without actually balancing the SNRs amongst all subcarriers reducing the overall capacity. In addition to this, the gap between the envelope of proposed heuristics (overall heuristic solution) and "Relaxed Solution" decreases as $\mathcal{T}$ is increased. This gap reduces with budget because with higher budget more contracts can be chosen as whole per subcarrier hence reducing the difference in capacity obtained with the "Relaxed Solution". Notice that the optimal solution lies in between this gap, hence, smaller this gap is, better is the performance. Additionally, we notice that for the overall proposed heuristic, capacity tends to converge to a stable value as number of relay agents are increased. The convergence happens because of the diversification of independent relay types.

Fig. 4.6 compares the performance of the proposed heuristic, i.e., overall average capacity with respect to number of subcarriers for two values of budget constraints. The simulation parameters are provided under the figure. We could easily deduce from this figure that the performance of "Best SNR Contracts" saturates very quickly and is far inferior from the proposed heuristic solution because of lower overall average capacity. One reason why "Best SNR Contracts" has inferior performance is that in this solution the best SNRs may not be well-spread over all subcarriers and some of the subcarriers may be underused.

In Fig. 4.7, we plot the average capacity per subcarrier vs. quantization factor for the heuristic solution and for the "Best SNR Contracts" solution with two different budget constraints. Once again, the parameters are provided underneath the graph. It is interesting to observe that the average capacity per subcarrier for the proposed heuristic solution remains almost the same as we increase the quantization factor K, which means that it is not that advantageous to quantize the probability distribution to a very high factor in order to obtain better performance. For example, a quantization factor as low as 3 which essentially classifies the types as "good", "average", or "bad" can be sufficient. This further demonstrates numerically that the source needs to design and broadcast very few contracts to get fair performance which leads to less signalling overheads. However, this observation is valid for uniform distribution and results could vary for a non-uniform distribution where higher quantization factor may lead to some types being more probable than others.

Lastly, we analyze the performance of our system under both complete and incomplete information scenarios. In order to do that, we plot average capacity per subcarrier vs. the number of relay agents for two values of budget constraints under these two scenarios in Fig. 4.8. We use heuristic solution for relay selection to evaluate the capacity. For each budget constraint, we plot the case where we have complete information, i.e., source knows the efficient contracts for all relays, and for incomplete information we plot two cases; when first-best contracts are broadcasted, and when second-best contracts are broadcasted. As expected, on average, for each budget constraint, the performance gap due to information asymmetry between the complete information and the incomplete information with the second-best contracts is smaller than gap between complete information and the incomplete information with the first-best contracts. Moreover, the capacity obtained with the first-best contracts is almost always constant because the first-best contracts are not designed to be incentive compatible and only the smallest contract is selected by all relays, so increasing the number of agents has no effect on performance.

4.7 Conclusion

In this chapter, we study the problem of relay selection and incentive mechanisms in multi-carrier wireless systems under asymmetric information. In such networks, a source node is assumed to be ill-informed of possible potential relay nodes and their private information such as channel conditions on the relay-destination link. This information asymmetry makes it harder for the source to choose selfish relay nodes efficiently and hence to optimize its throughput. We address this classical problem by introducing a simple *principal-agent model* for source and relays. The advantage of this model is that it leaves the bargaining power completely to the principal which in our case is the source node, and this reduces signalling and computational overheads for the relays. We then divide the problem into two parts. In the first part, we use contract theory to design a common incentive compatible contracts for the relays (assuming that source has a joint distribution of private information or *types* of relays for all subcarriers), i.e., a relay of a certain type will choose contract designed for his type only. These contracts are broadcasted by the source to all relays and the interested relay-agents respond with the contracts they are willing to accept in each subcarrier. Once the source becomes aware of these contracts, the only problem that remains is to select appropriate relays in each subcarrier such that the overall capacity of source is maximized while the source is under a budget constraint. We formulate this problem of relay selection as a nonlinear non-separable knapsack problem and suggest a heuristic solution to solve it efficiently. The source then notifies the selected relays with instructions such as space-time codes and makes the required transfers. We have compared the performance of heuristic solution with a simple relay selection mechanism and have presented numerical results to show that our solution performs better under the most common settings. The overall mechanism introduced in this chapter is simple and has limited interaction between the source and the potential relays and participation of interested nodes is also voluntary. As part of future work, we will address more complex issues for the relay selection mechanism such as including the effect of direct links and considering selective relaying in the proposed system model (selective relaying is discussed in [77]), addressing energy efficiency with asymmetric information (energy efficiency for cooperative networks is explored [78]), and designing better heuristics and more efficient protocols.

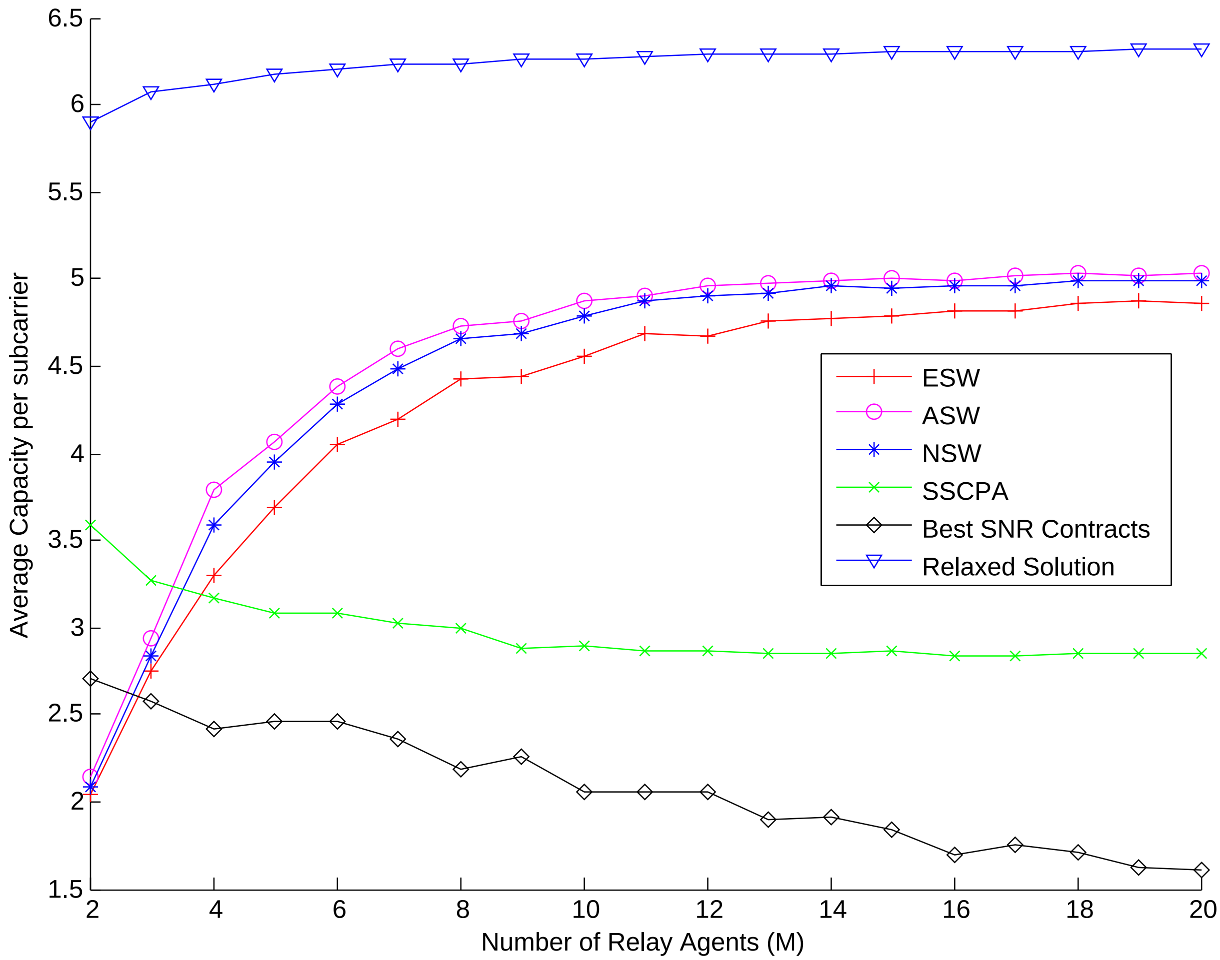


Figure 4.3: Comparison of different heuristics with $\mathcal{T} = 8$, $N = 16$, and $K = 10$

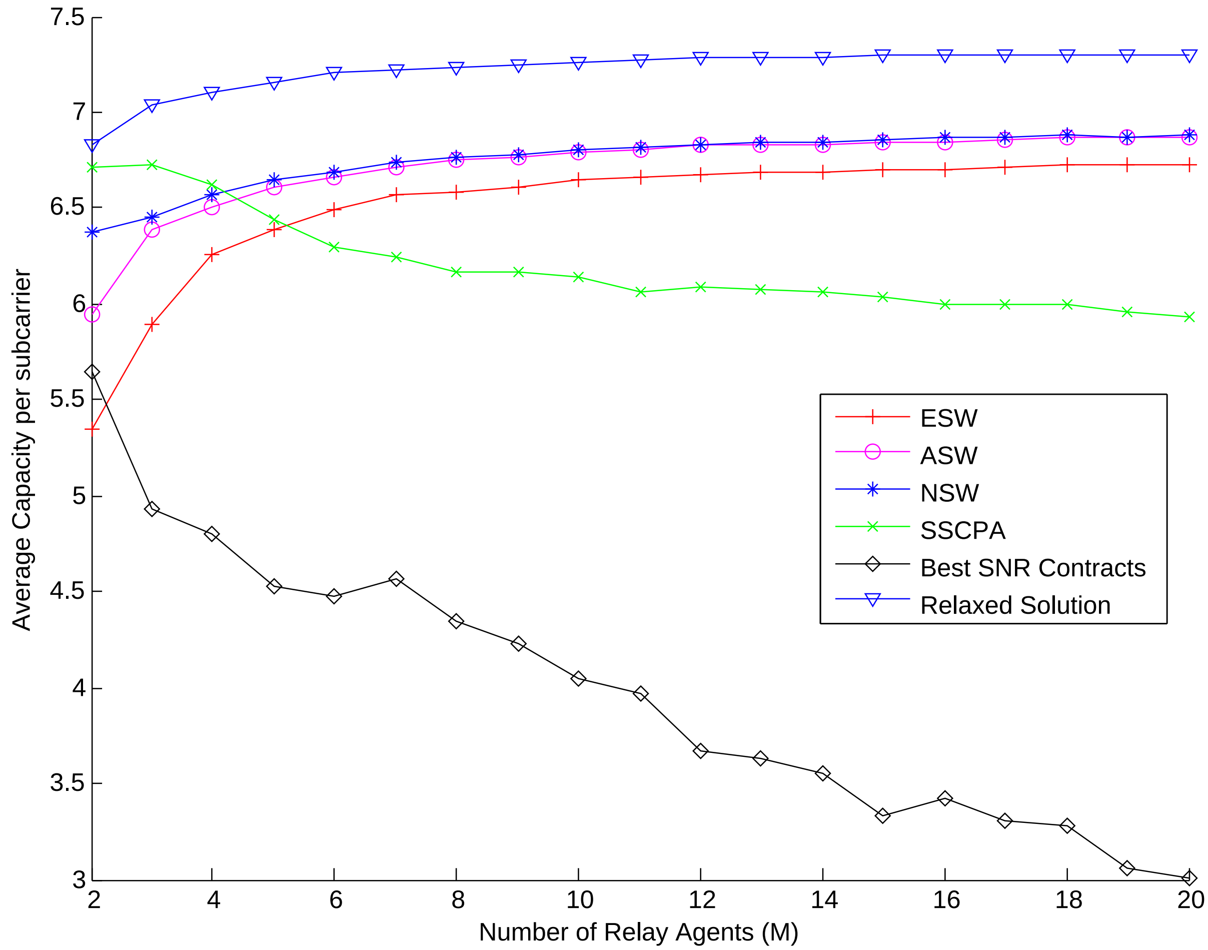


Figure 4.4: Comparison of different heuristics with $\mathcal{T} = 16$, $N = 16$, and $K = 10$

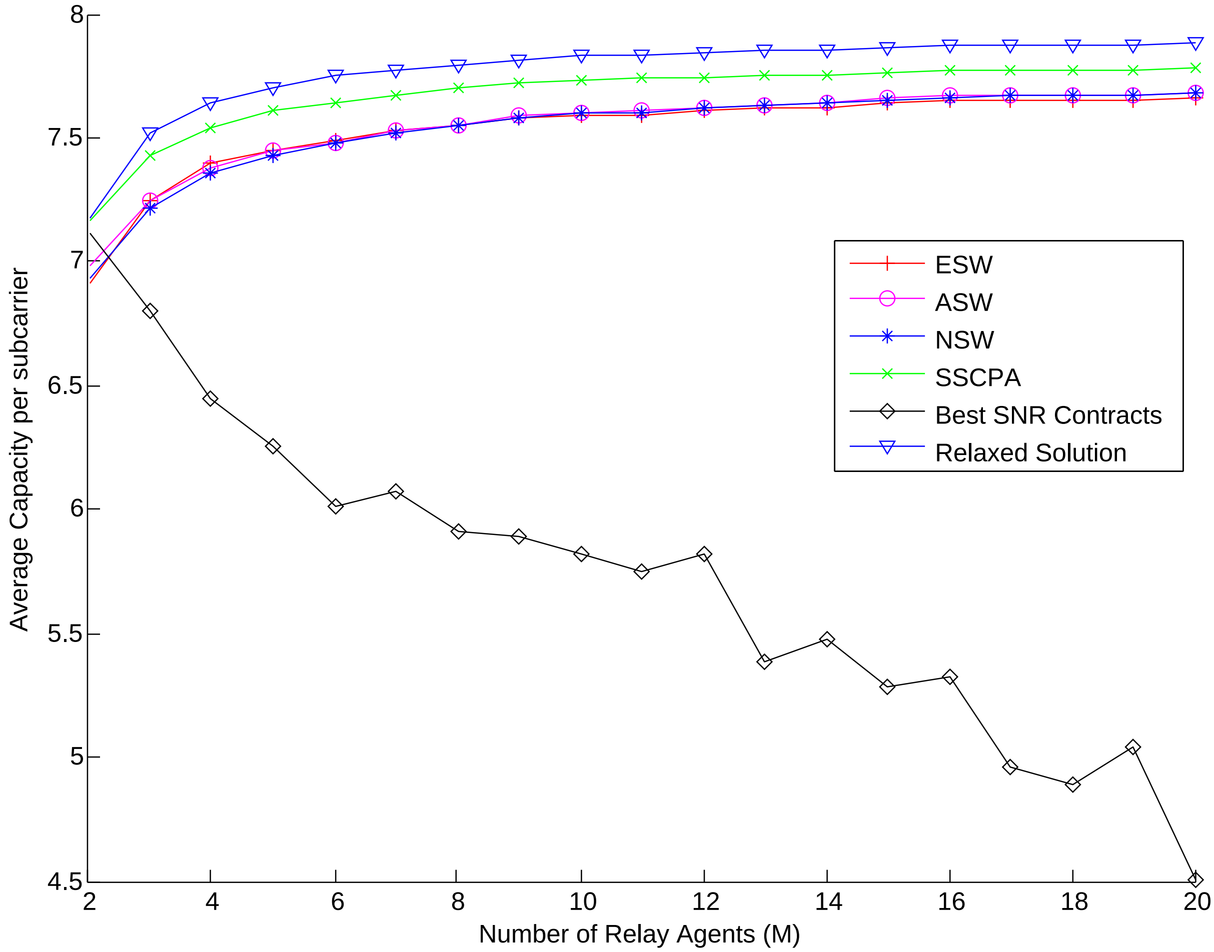


Figure 4.5: Comparison of different heuristics with $\mathcal{T} = 24$, $N = 16$, and $K = 10$

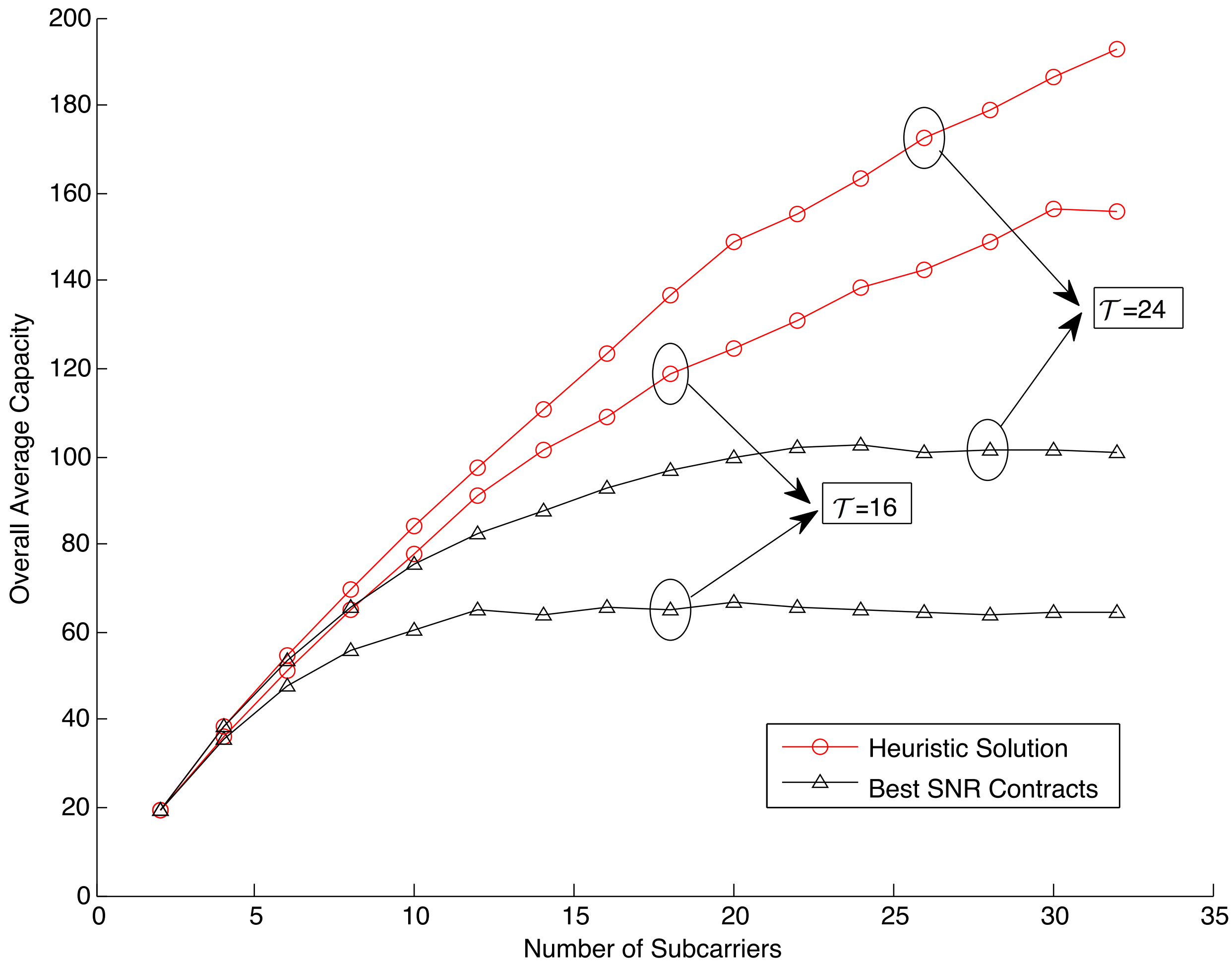


Figure 4.6: Capacity vs. number of subcarriers (N) for fixed $M = 10$, and $K = 10$ and two values of $\mathcal{T} = 16, 24$

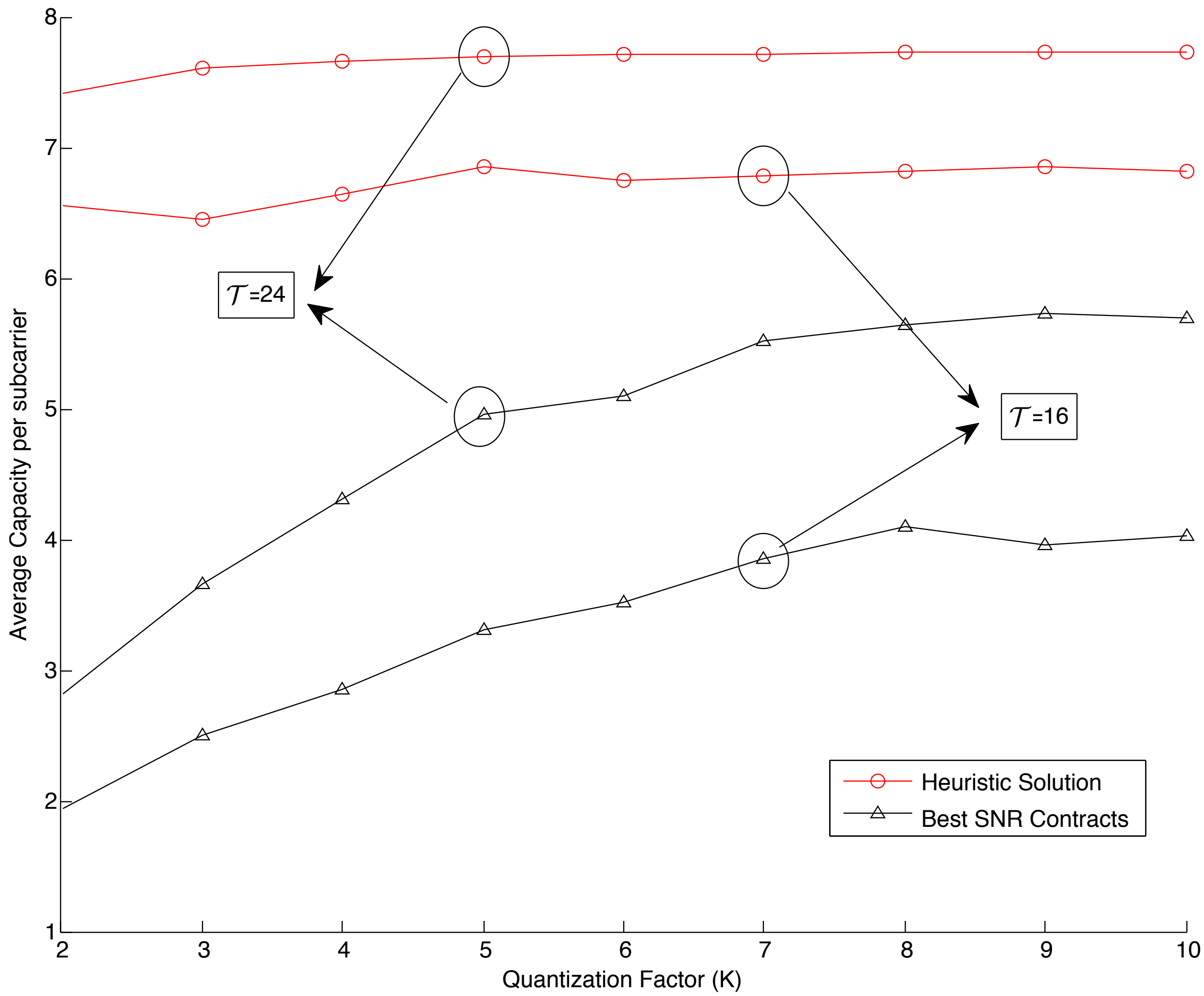


Figure 4.7: Average capacity vs. quantization Factor (K) for fixed $N = 16$, $M = 10$ and two values of $\mathcal{T} = 16, 24$

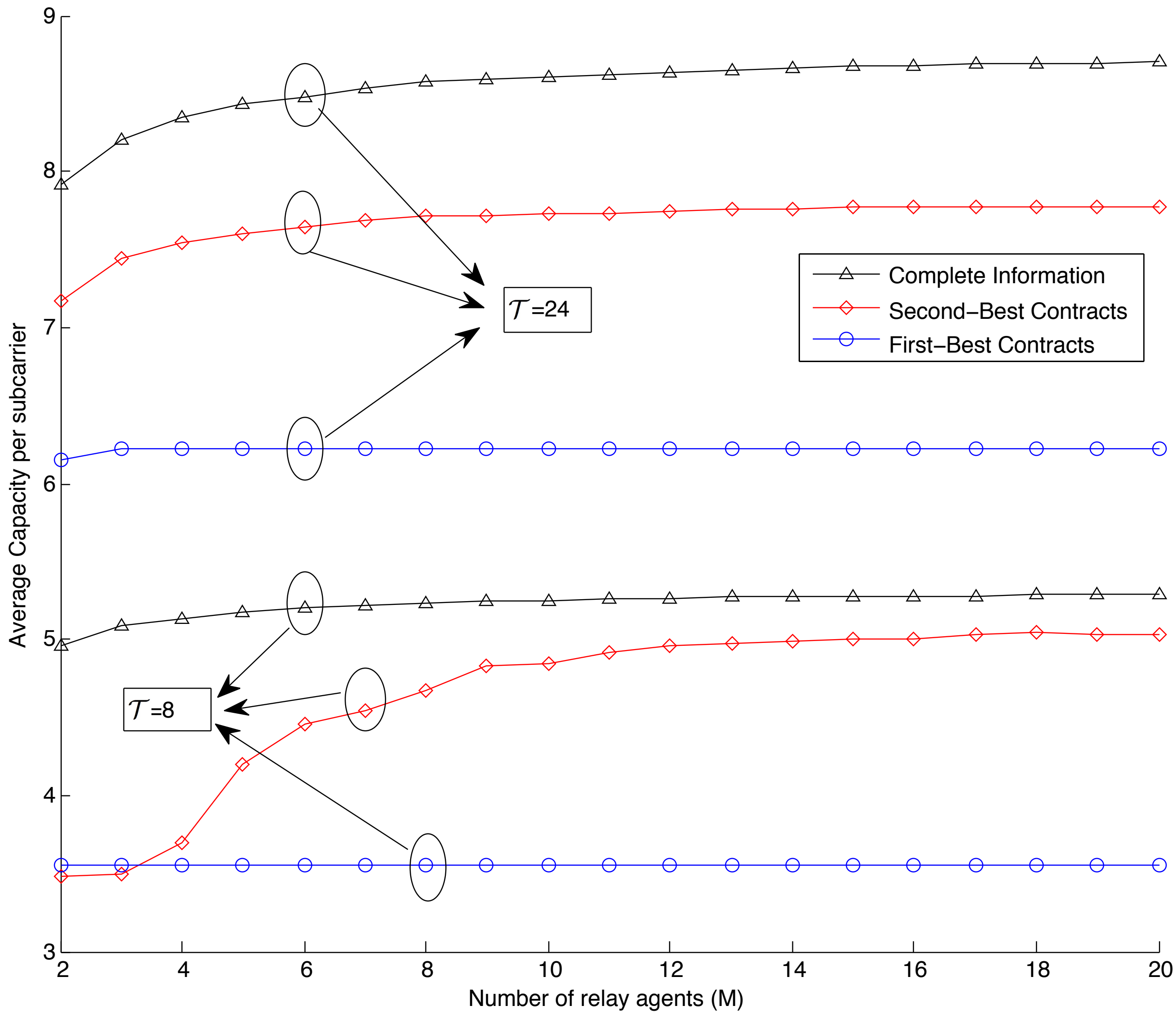


Figure 4.8: Comparison of first best and second-best contracts, and full information scenario with $N = 16$, $K = 10$, and two values of $\mathcal{T} = 8, 24$

Chapter 5

Green Cellular Networks: Research Issues and Challenges

5.1 Introduction

During the last decade, there has been tremendous growth in cellular networks market. The number of subscribers and the demand for cellular traffic has escalated astronomically. With the introduction of Android and iPhone devices, use of ebook readers such as iPad and Kindle and the success of social networking giants such as Facebook, the demand for cellular data traffic has also grown significantly in recent years. Hence, mobile operators find meeting these new demands in wireless cellular networks inevitable, while they have to keep their costs minimum.

Such unprecedented growth in cellular industry has pushed the limits of energy consumption in wireless networks. There are currently more than 4 million base stations (BSs) serving mobile users, each consuming an average of 25MWh per year. The number of BSs in developing regions are expected to almost double by 2012 as shown in Fig. 5.1. Information and Communication Technology (ICT) already represents around 2% of total carbon emissions (of which mobile networks represent about 0.2%), and this is expected to increase every year. In addition to the environmental aspects, energy costs also represent a significant portion of network operators' overall expenditures (OPEX). While the BSs connected to electrical grid may cost approximately 3000$ per year to operate, the off-grid BSs in remote areas generally run on diesel power generators and may cost ten times more.

The rising energy costs and carbon footprint of operating cellular networks have led to an emerging trend of addressing energy-efficiency amongst the network operators and regulatory bodies such as 3GPP and ITU [82, 83]. This trend has stimulated the interest of researchers in an innovative new research area called "green cellular networks". In this regard, the European Com-

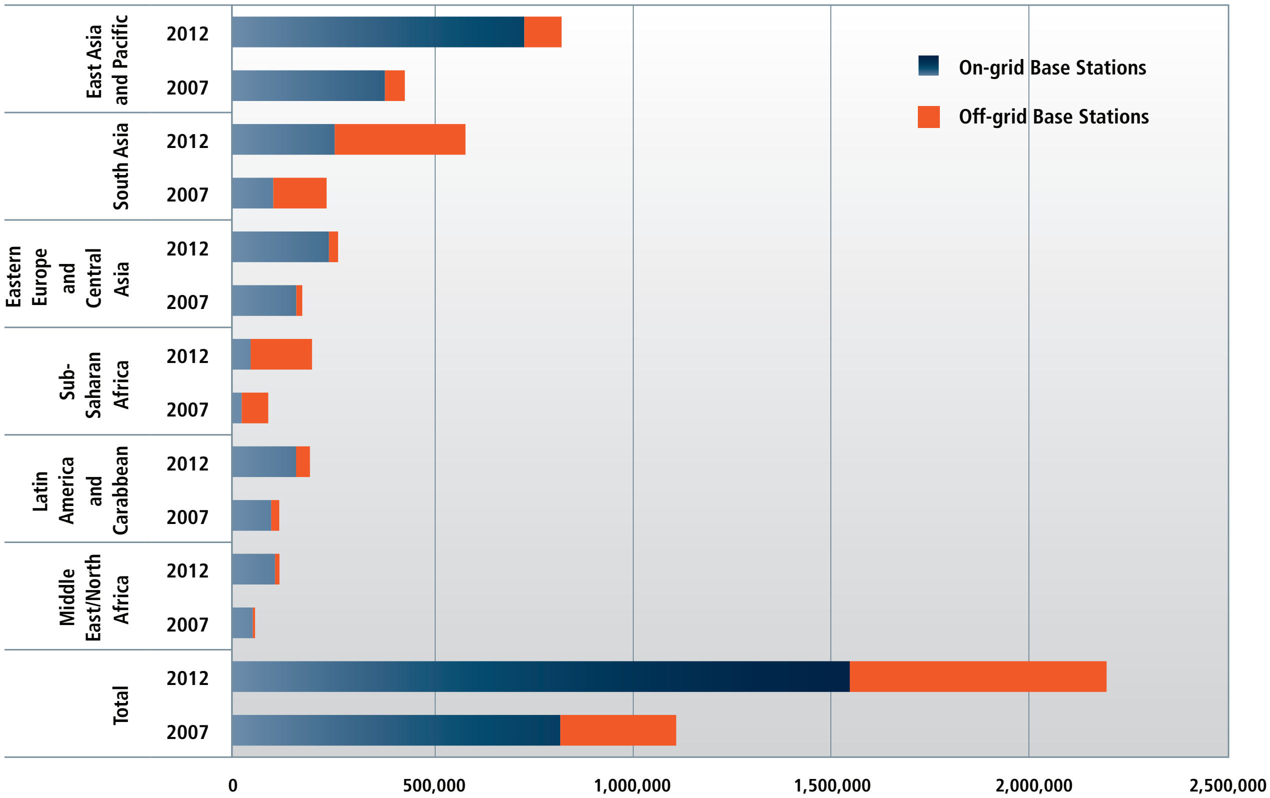
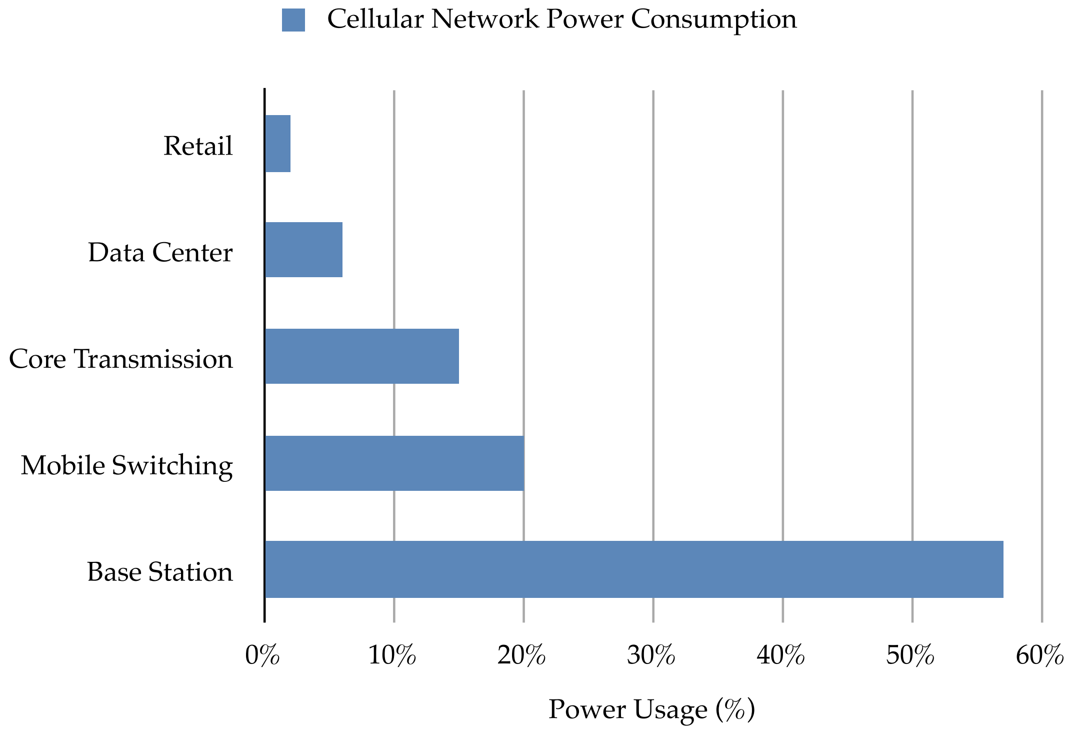
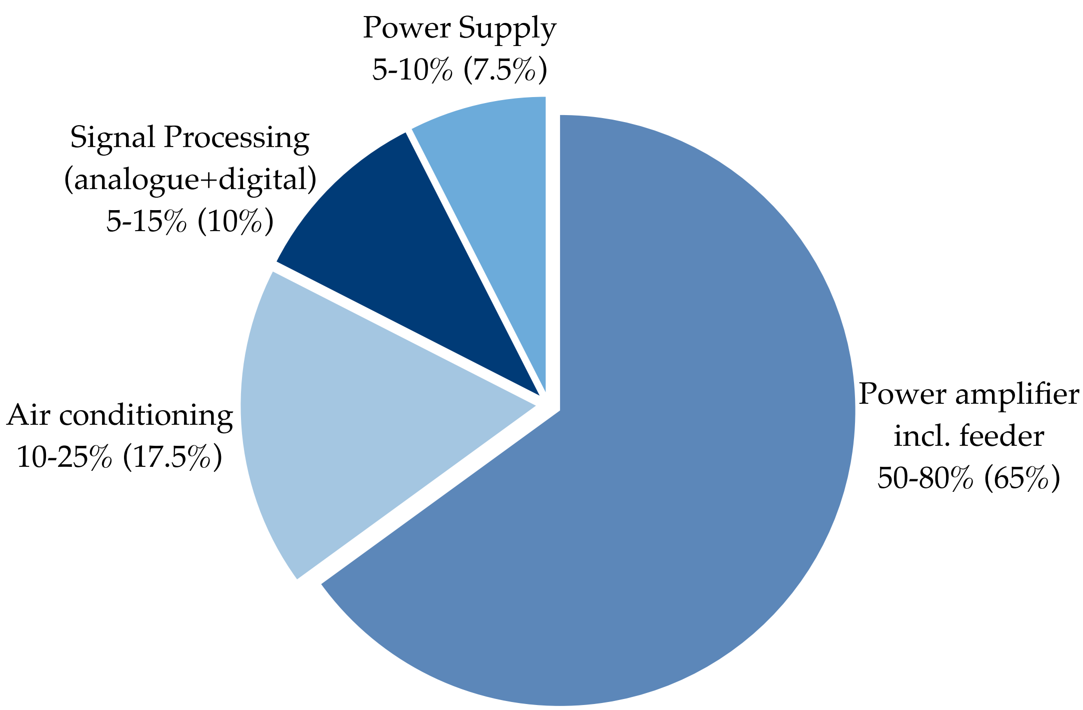


Figure 5.1: Growth in base stations in developing regions 2007-2012
Source: GSMA Research [79]

mission has recently started new projects within its seventh Framework Programme to address the energy efficiency of mobile communication systems, viz. "Energy Aware Radio and NeTwork TecHnologies (EARTH)", "Towards Real Energy-efficient Network Design (TREND)" and "Cognitive Radio and Cooperative strategies for Power saving in multi-standard wireless devices (C2POWER)" [36], [37], [38]. "Green radio" is a vast research discipline that needs to cover all the layers of the protocol stack and various system architectures and it is important to identify the fundamental trade-offs linked with energy efficiency and the overall performance [84]. Figures 5.2(a) and 5.2(b) show a breakdown of power consumption in a typical cellular network and gives us an insight into the possible research avenues for reducing energy consumption in wireless communications. In [84], the authors have identified four key trade-offs of energy efficiency with network performance; deployment efficiency (balancing deployment cost, throughput), spectrum efficiency (balancing achievable rate), bandwidth (balancing the bandwidth utilized) and delay (balancing average end-to-end service delay). To address the challenge of increasing power efficiency in future wireless networks and thereby to maintain profitability, it is crucial to consider various paradigm-shifting technologies, such as energy efficient wireless architectures and protocols, efficient BS redesign, smart grids, opportunistic network access or cognitive radio, cooperative relaying and



(a) Power consumption of a typical wireless cellular network [80](ref. therein)



(b) Power consumption distribution in radio base stations [81](ref. therein)

Figure 5.2: Breakdown of power consumption in a typical cellular network and corresponding base stations

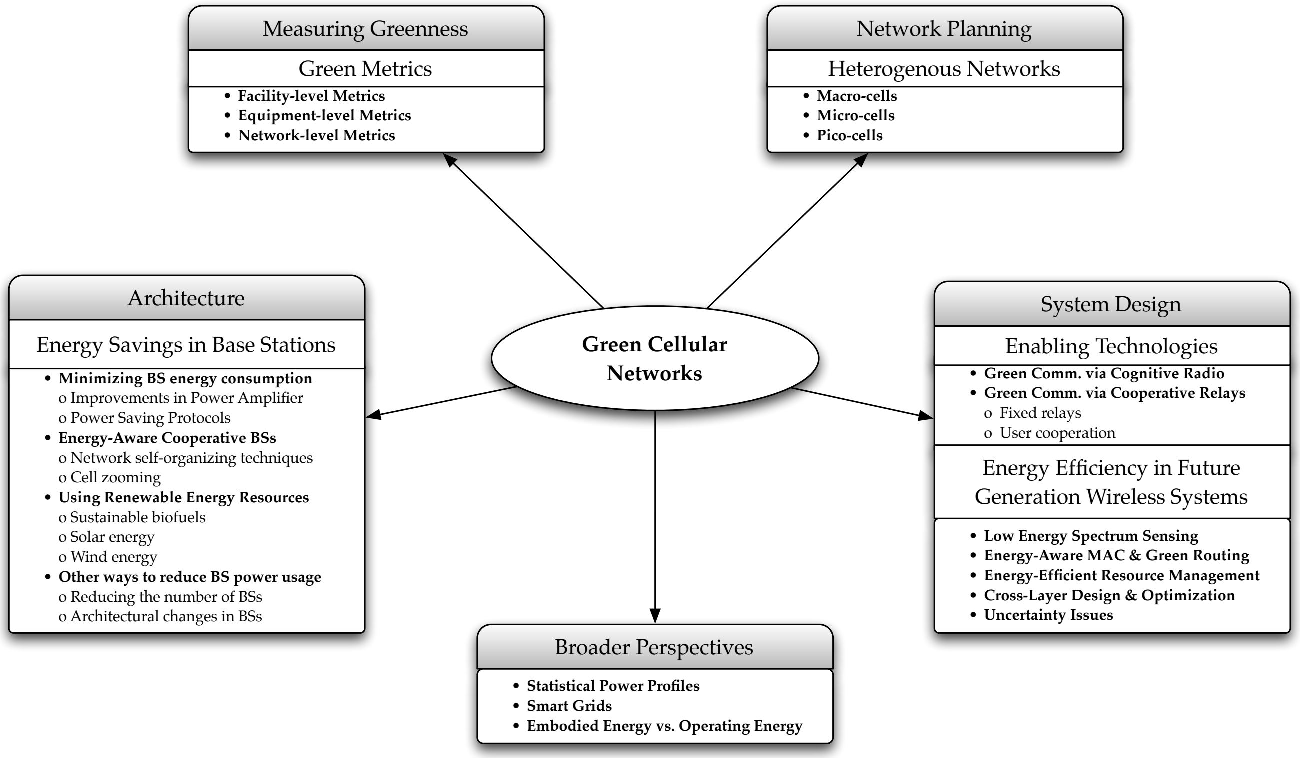


Figure 5.3: Technical roadmap for green cellular networks: a taxonomy graph

heterogenous network deployment based on smaller cells.

Among all the promising energy saving techniques, cognitive radio and cooperative relaying, although already getting matured in many aspects, but still are in their infancy when it comes to the deployment issues in cellular networks. Therefore, it is crucial to promote the potentials of these techniques in cellular wireless networks. Moreover, it is necessary to be aware that still many energy concerns in cognitive and cooperative networks have remained as unanswered challenges, which raises the importance of further exploring these concerns.

In this paper, we provide a brief survey on some of the work that has already been done to achieve power efficiency in cellular networks, discuss some research issues and challenges and suggest some techniques to enable an energy efficient or "green" cellular network. We also put a special emphasis on cognitive and cooperative techniques, in order to bring attention to the benefits cellular systems can gain through employing such techniques, and also highlight the research avenues in making these techniques green. A taxonomy graph of our approach towards the design of green cellular networks is given in Fig. 5.3. As shown in the figure, we identify four important aspects of a green networking where we would like to focus: defining green metrics, bringing architectural changes in base stations, network planning, and efficient system design. In addition, some broader perspectives must also be considered. In the following sections we elaborate on each such aspect and discuss the related issues and challenges. We begin with a brief discussion on energy efficiency metrics in section 5.2. Since BSs consume the major chunk of input energy, we discuss the energy efficiency of BSs more at the component level in section 5.3. Here, we study how to minimize energy consumption of BS employing improvements in power amplifier, designing power saving protocols, implementing cooperative BS power management, using renewable energy resources and bringing some simple architectural changes. Section 5.4 addresses the energy efficiency from a network planning perspective where we discuss how different types of network deployments based on smaller cells can be used to increase the energy efficiency of a wireless system. Regarding the system design, we first explain the use of modern communication technologies such as cognitive radio and cooperative relays to enable green communication in cellular systems in section 5.5 and we expand this idea further in section 5.6 from a different perspective, where we discuss how the future wireless systems based on both cognitive and cooperative concepts can be made more energy efficient at the system level. Techniques such as low energy spectrum sensing, energy-aware medium access control and routing, efficient resource management, cross-layer design and addressing uncertainty issues have been examined in this context. Some broader perspectives have been discussed in 5.7 and conclusions are drawn in 5.8.

5.2 Measuring Greenness: The Metrics

Before starting any discussion on "green" networks, the first question naturally comes to mind is that what actually is "green"? How do we measure and define the degree of "greenness" in telecommunication networks? Although carbon footprint or CO_2 emissions would naturally be considered a measure of "greenness", but the share of carbon emissions for telecommunication networks is fairly low (less than 1%). However, please note that other motivations to obtain "green" wireless technology also include economic benefits (lower energy costs) and better practical usage (increased battery life in mobile devices), hence evaluation of energy savings or measuring energy efficiency seems to be a more apt choice for measuring "greenness". Thus, the notion of "green" technology in wireless systems can be made meaningful with a comprehensive evaluation of energy savings and performance in a practical system. This is where energy efficiency metrics play an important role. These metrics provide information in order to directly compare and assess the energy consumption of various components and the overall network. In addition, they also help us to set long term research goals of reducing energy consumption. With the increase in research activities pertaining to green communications and hence in number of diverse energy efficiency metrics, standards organizations such as European Technical Standards Institute (ETSI) and Alliance for Telecommunications Industry Solutions (ATIS) are currently making efforts to define energy efficiency metrics for wireless networks [85, 86].

Generally speaking, energy efficiency metrics of telecommunication systems can be classified into three main categories: *facility-level*, *equipment-level* and *network-level* metrics [87, 88]. Facility-level metrics relates to high-level systems where equipment is deployed (such as datacenters, ISP networks etc.), equipment level metrics are defined to evaluate performance of an individual equipment, and network level metrics assess the performance of equipments while also considering features and properties related to capacity and coverage of the network.

The Green Grid (TGG) association of IT professionals first proposed facility-level efficiency metrics called PUE (Power Usage Efficiency) and its reciprocal DCE (Data Center Efficiency) in [89] to evaluate the performance of power hogging datacenters. PUE which is defined as the ratio of total facility power consumption to total equipment power consumption, is although a good metric to quickly assess the performance of datacenters at a macro level, it fails to account for energy efficiency of individual equipments. Therefore, in order to quantify efficiency at the equipment level, ratio of energy consumption to some performance measure of a communication system would be more appropriate. However, grading the performance of a communication system is more challenging than it actually first appears, because the performance comes in a variety of different forms (spectral efficiency, number of calls supported in block of time, etc.) and each such performance measure affects this efficiency metric very differently. Some suggested metrics including power per user (ratio of total facility power to number of users) measured in [Watt/user], and energy consumption rating (ECR) which is the ratio of normalized energy consumption to effective full-duplex throughput and is measured in [Watt/Gbps] [90]. While power per user can be a useful metric for a network provider to evaluate economic tradeoffs, network planning etc., metrics such as ECR provide the manufacturers a better insight into performance of hardware components. However, even the busiest networks do not always operate on full load conditions, therefore it would be useful to complement metrics such as ECR to incorporate the dynamic network conditions such as energy consumption under full-load, half-load and idle cases. In this regard, other metrics such as ECRW (ECR-weighted), ECR-VL (energy efficiency metric over a variable-load cycle), ECR-EX (energy efficiency metric over extended-idle load cycle), telecommunications energy efficiency ratio (TEER) by ATIS, Telecommunication Equipment Energy Efficiency Rating (TEEER) by Verizons Networks and Building Systems consider total energy consumption as weighted sum of energy consumption of the equipment at different load conditions [90], [91], [92], [93]. As an example for TEEER, the total power consumption P_{total} is calculated by the following formula:

$$P_{\text{total}} = 0.35P_{\text{max}} + 0.4P_{50} + 0.25P_{\text{sleep}}, \tag{5.1}$$

where P_{max}, P_{50} and P_{sleep} are power consumption at full rate, half-rate and sleep mode, respectively, and the weights are obtained statistically. However, these metrics such as ECR, TEER, TEEER etc. are unable to capture all the properties of a system and research work is still active to suggest different types of metrics. Parker *et al.* recently proposed an absolute energy efficiency metric (measured in dBε) in [94], given by:

$$\text{dB}\varepsilon = 10\log_{10}\left(\frac{\text{Power/Bit Rate}}{kT\ln 2}\right), \tag{5.2}$$

where k is the Boltzmann constant and T is the absolute temperature of medium. The authors suggest that the inclusion of temperature aspect of the system is logical since classical thermodynamics is based on absolute temperature of the system under analysis. Using different examples, the authors contend that this metric is highly versatile and can be universally applied to any ICT system, subsystem and component.

While the energy efficiency metrics at the component and equipment level are fairly straightforward to define, it is more challenging to define metrics at a system or network level [88]. Just by including the area aspect of the network, a natural choice of a metric may at first seem to be [Watt/Gbps/km^2], but a careful analysis can explain that it can work counter to a “green” objective [95]. Using a simple example of a typical network scenario, it has been shown in [95] that due to the path loss, such a metric can only be valid when applied to networks with similar number of sites in a given area. In [86], ETSI proposes two network level metrics for GSM systems based on load conditions. In rural areas, which are generally under low load conditions, the objective is to reduce power consumption in a coverage region, hence the metric is given by:

$$PI_{\text{rural}} = \frac{\text{Total coverage area}}{\text{Power consumed at the site}}, \tag{5.3}$$

where PI_{rural} bears the unit of [km^2/Watt], and denotes the network performance indicator in rural areas. Urban areas on the hand have higher traffic demand than rural areas, hence capacity is considered instead of coverage area. A common metric under such full load conditions is therefore given by:

$$PI_{\text{urban}} = \frac{N_{\text{busyhour}}}{\text{Power consumed at the site}}, \tag{5.4}$$

where, N_{busyhour} is the number of users based on average busy hour traffic demand by users and average BS busy hour traffic, and PI_{urban} (users/Watt) is the network performance indicator in urban areas.

To summarize the discussion above, a non-exhaustive list of energy metrics is given in Table 5.1. Interested readers can find a more comprehensive taxonomy of green metrics in [87]. Due to the intrinsic difference and relevance of various communication systems and performance measures, it is doubtful that one single metric can suffice. However, in future, the “green” metrics must also consider deployment costs such as site construction and backhaul, and QoS requirements such as transmission delay etc. along with spectral efficiency in order to assess the true "greenness" of the system. Once a large consensus is reached on a small set of standard energy metrics in future, it will not only accelerate the research activities in green communications, but also help pave the way towards standardization.

Table 5.1: Some energy efficiency metrics

Metric	**Type**	**Units**	**Description**
PUE (Power Usage Efficiency)	Facility-Level	Ratio ($\geq$1)	Defined as ratio of total facility power consumption to total equipment power consumption.
DCE (Data Center Efficiency)	Facility-Level	Percentage	Defined as reciprocal of PUE.
Telecommunications Energy Efficiency Ratio (TEER)	Equipment-Level	Gbps/Watt	Ratio of useful work to power consumption
Telecommunications Equipment Energy Efficiency Rating (TEEER)	Equipment-Level	$-\log\left(\frac{\text{Gbps}}{\text{Watt}}\right)$	$-\log\left(\frac{P_{\text{total}}}{\text{Throughput}}\right)$, where P_{total} is given by equation (5.1)
Energy Consumption Rating (ECR)	Equipment-Level	Watt/Gbps	Ratio of energy consumption over effective system capacity
ECR-Weighted (ECRW)	Equipment-Level	Watt/Gbps	Calculated the same way as ECR except energy consumption is now calculated as $0.35E_f + 0.4E_h + 0.25E_i$, where each term corresponds to energy consumption in full load, half load and idle modes.
ECR-variable-load metric (ECR-VL)	Equipment-Level	Watt/Gbps	Average energy rating in a reference network described by an array of utilization weights [91].
ECR-extended-idle metric (ECR-EX)	Equipment-Level	Watt/Gbps	Average energy rating in a reference network, where extended energy savings capabilities are enabled [91].
Performance Indicator in rural areas (PI_{rural})	Network-Level	km^2/Watt	Ratio of total coverage area to power consumed at site as given by eq. (5.3)
Performance Indicator in urban areas (PI_{urban})	Network-Level	users/Watt	Ratio of number of subscribers to power consumed at the site as given by eq. (5.4)

5.3 Architecture: Energy Savings in Base Stations

Due to the rapidly growing demand for mobile communication technology, the number of worldwide cellular BSs has increased from a few hundred thousands to many millions within a last couple of years. Such a substantial jump in the number of BSs that power a cellular network accounts for the sudden increase in greenhouse gases and pollution, in addition to higher energy costs to operate them. With the advent of data intensive cellular standards, power-consumption for each BS can increase upto 1,400 watts and energy costs per BS can reach to $3,200 per annum with a carbon footprint of 11 tons of CO_2 [96]. The radio network itself adds up to 80% of an operator's entire energy consumption. Therefore, BS equipment manufacturers have begun to offer a number of eco and cost friendly solutions to reduce power demands of BSs and to support off-grid BSs with renewable energy resources. Nokia Siemens Networks Flexi Multiradio Base Station, Huawei Green Base Station and Flexenclosure E-site solutions are examples of such recent efforts [97], [98], [99]. In [100], the authors present various methods dealing with improved transmitter efficiency, system features, fresh air-cooling, renewable energy sources and energy saving during low traffic. A typical cellular network consists of three main elements; a core network that takes care of switching, BSs providing radio frequency interface, and the mobile terminals in order to make voice or data connections. As the number of BSs increases, it becomes crucial to address their energy consumption for a cellular network. In the next few subsections, we will discuss different ways to reduce energy consumption due to BSs.

5.3.1 Minimizing BS Energy Consumption

The energy consumption of a typical BS can be reduced by improving the BS hardware design and by including additional software and system features to balance between energy consumption and performance. In order to improve hardware design of a BS for energy consumption, we need to address the energy efficiency of the power amplifier (PA). A PA dominates the energy consumption of a BS and its energy efficiency depends on the frequency band, modulation and operating environment [100]. Some typical system features to improve BS energy efficiency are to shut down BS during low traffic or cell zooming [101, 102]. Besides hardware redesign and new system level features, there are various site level solutions that can be used in order to save energy. For example, outdoor sites can be used over wider level of temperatures, and thus less cooling would be required. Another solution is to use more fresh air-cooling rather than power consuming air conditioners for indoor sites. In addition, RF heads and modular BS design can be implemented to reduce power loss in feeder cables [100].

Improvements in Power Amplifier

There are three essential parts of a BS: radio, baseband and feeder. Out of these three, radio consumes more than 80% of a BS's energy requirement, of which power amplifier (PA) consumes almost 50% [103]. Shockingly, 80-90% of that is wasted as heat in the PA, and which in turn requires air-conditioners, adding even more to the energy costs. The total efficiency of a currently deployed amplifier, which is the ratio of AC power input to generated RF output power, is generally in anywhere in the range from 5% to 20% (depending on the standard viz. GSM, UMTS, CDMA and the equipment's condition) [104]. Modern BSs are terribly inefficient because of their need for PA linearity and high peak-to-average power ratios (PAPR). The modulation schemes that are used in communication standards such as WCDMA/HSPA and LTE are characterized by strongly varying signal envelopes with PAPR that exceeds 10dB. To obtain high linearity of the PAs in order to maintain the quality of radio signals, PAs have to operate well below saturation, resulting in poor power efficiency [81]. Depending on their technology (e.g Class-AB with digital pre-distortion) and implementation, the component level efficiency of modern amplifiers for CDMA and UMTS systems is in the order of approximately 30% to 40% [104]. Since these technologies have reached their limits, PAs based on special architectures such as digital pre-distorted Doherty-architectures and GaN (Aluminum Gallium nitride) based amplifiers seem to be more promising by pushing the power efficiency levels to over 50% [104]. Doherty PAs that consist of a carrier and a peak amplifier is advantageous by providing easy additional linearization using conventional methods such as feed-forward and envelope elimination and restoration (EER)[105]. Since GaN structures can work under higher temperature and higher voltage, they can potentially provide a higher power output. Additional improvements in efficiency can be obtained by shifting to switch-mode PAs from the traditional analog RF-amplifiers. Compared to standard analog PAs, switch-mode PAs tend to run cooler and draw less current. While amplifying a signal, a switch-mode amplifier turns its output transistors on and off at an ultrasonic rate. The switching transistors produce no current when they are switched off and produce no voltage when switched on, therefore generate very little power as heat resulting in a highly efficient power supply. It is expected that overall component-efficiency of these energy efficient devices could be around 70% [104].

One more significant setback in increasing power efficiency with PAs is that they perform better at maximum output power in order to maintain the required signal quality. However, during the low traffic load conditions (e.g night time), lot of energy is routinely wasted. Therefore, design of flexible PA architectures that would allow a better adaptation of the amplifier to the required output power needs to be addressed [104]. In addition to this, we need to investigate more efficient modulation schemes, because modulation also affects the PA efficiency. As an example, by focusing more on higher modulation schemes that require additional filtering in order to prioritize data over voice, linearity of PA is more desirable because of the non-constant envelope of the signal

[103]. Using different linearization techniques such as Cartesian feedback, digital pre-distortion and feed-forward along with different kind of DSP methods that reduces the requirement on the linear area of PA have also been suggested [100].

Power Saving Protocols

In the current cellular network architecture based on WCDMA/HSPA, BSs and mobile terminals are required to continuously transmit pilot signals. Newer standards such as LTE, LTE-Advanced and WiMAX have evolved to cater ever-growing high speed data traffic requirements. With such high data requirements, although BSs and mobile units (MU) employing newer hardware (such as multiple-input and multiple-output (MIMO) antennas) increase spectral efficiency allowing to transmit more data with the same power, power consumption is still a significant issue for future high speed data networks and they require energy conservation both in the hardware circuitry and protocols. A fairly intuitive way to save power is to switch off the transceivers whenever there is no need to transmit or receive. The LTE standard utilizes this concept by introducing power saving protocols such as discontinuous reception (DRX) and discontinuous transmission (DTX) modes for the mobile handset. DRX and DTX are methods to momentarily power down the devices to save power while remaining connected to the network with reduced throughput. Continuous transmission and reception in WCDMA/HSPA consumes significant amount of power even if the transmit powers are far below the maximum levels, and therefore power savings due to DRX and DTX is an attractive addition. IEEE 802.16e or Mobile WiMAX also has similar provisions for sleep mode mechanisms for mobile stations [106]. The device negotiates with the BS and the BS will not schedule the user for transmission or reception when the radio is off. There are three power-saving classes with different on/off cycles for the WiMAX standard.

Unfortunately, such power saving protocols for BSs have not been considered in the current wireless standards. The traffic per hour in a cell varies considerably over the time and BSs can regularly be under low load conditions, especially during the nighttime. In future wireless standards, energy saving potential of BSs needs to be exploited by designing protocols to enable sleep modes in BSs. The authors in [81] suggest making use of downlink DTX schemes for BSs by enabling micro-sleep modes (in the order of milliseconds) and deep-sleep modes (extended periods of time). Switching off inactive hardware of BSs during these sleep modes can potentially save a lot of power, especially under low load conditions.

5.3.2 Energy-Aware Cooperative BS Power Management

Traffic load in cellular networks have significant fluctuations in space and time due to a number of factors such as user mobility and behaviour. During daytime, traffic load is generally higher in office areas compared to residential areas, while it is the other way around during the night. Therefore, there will always be some cells under low load, while some others may be under heavy traffic load. Hence, a static cell size deployment is not optimal with fluctuating traffic conditions. For next generation cellular networks based on microcells and picocells and femtocells, such fluctuations can be very serious. While limited cell size adjustment called "cell-breathing" currently happens in currently deployed CDMA networks (a cell under heavy load or interference reduces its size through power control and the mobile user is handed off to the neighbouring cells), a more network-level power management is required where multiple BSs coordinate together. Since operating a BS consumes a considerable amount of energy, selectively letting BSs go to sleep based on their traffic load can lead to significant amount of energy savings. When some cells are switched off or in sleep mode, the radio coverage can be guaranteed by the remaining active cells by filling in the gaps created. Such concepts of self-organizing networks (SON) have been introduced in 3GPP standard (3GPP TS 32.521) to add network management and intelligence features so that the network is able to optimize, reconfigure and heal itself in order to reduce costs and improve network performance and flexibility [107]. The concept of SONs can be applied in order to achieve diverse objectives. For instance, in [108] different use cases for SONs are discussed, e.g., load balancing, cell outage management, management of relays and repeaters, etc. In the context of power efficiency, the performance of these self-organizing techniques were initially explored in [102, 109]. Using the numerical results, the authors here suggested that substantial amount of energy savings can be obtained (of the order of 20%, and above) by selectively reducing the number of active cells that are under low load conditions. On the other hand, a distributed algorithm is proposed in [110] in which BSs exchange information about their current level of power and take turns in reducing their powers. Recently, authors of [111, 112] introduced the notion of energy partitions which is the associations among powered-on and powered-off BSs, and use this notion as the basis of rearranging the energy configuration.

A similar but even more flexible concept called "Cell Zooming" was presented in [101]. Cell zooming is a technique through which BSs can adjust the cell size according to network or traffic situation, in order to balance the traffic load, while reducing the energy consumption. When a cell gets congested with increased number of users, it can zoom itself in, whereas the neighboring cells with less amount of traffic can zoom out to cover those users that cannot be served by the congested cell. Cells that are unable to zoom in may even go to sleep to reduce energy consumption, while the neighboring cells can zoom out and help serve the mobile users cooperatively. Another such proposal to dynamically adjust cell-size in a multi-layer cellular architecture was presented in [113].

Implementation

The framework for cell zooming can include a cell-zooming server (CS) (implemented in the gateway or distributed in the BSs) that senses the network state information such as traffic, channel quality etc [101] and hence makes decisions for for cell zooming. If there is a need for a cell to zoom in or out, it will coordinate with its neighboring cells by the assistance of a CS. Cells can zoom in or out by a variety of techniques such as physical adjustment, BS cooperation and relaying [101]. Physical adjustment can be either done by adjusting the transmit powers of BSs and also by adjusting antenna height and tilt for cells to zoom in or out. BS cooperation here means that multiple BSs cooperatively transmit or receive from MUs. For an MU, a cluster of BSs cooperating form a new cell, the size of which is sum of cell sizes of these BSs. Relaying can also be used for cell size adjustment in a way that relay stations can help transfer the traffic from a cell with heavy load to a cell in low load conditions [101, 113]. The authors in [113] propose dynamic self-organization of the cellular layers using techniques such as timed sleep mode, user location prediction and reverse channel sensing. In such networks, BS can also go to sleep mode where the energy consuming equipments such as air conditioner etc., can be switched off. The neighboring cells can then reconfigure to guarantee the coverage.

Benefits and Challenges

Self-organizing cellular networks can be useful in load balancing as well as energy conservation by deciding when to disperse load for load balancing and when to concentrate load for energy savings. The advantages of techniques such as cell zooming also include improved user experience such as better throughput and increased battery life. For e.g in [113], a two-layer cellular architecture achieves a power savings of up to 40% over the entire day. With techniques such as BS cooperation and relaying, inter-cell interference and fading effects can be mitigated and hence MUs can observe higher diversity gains and better coverage. However, sufficient challenges lay ahead to practically realize these networks such as radio frequency planning, configuring switching thresholds, avoiding coverage holes, tracing spatial and temporal traffic load fluctuations etc. [101, 113].

5.3.3 Using Renewable Energy Resources

In several remote locations of the world such as Africa and Northern Canada, electrical grids are not available or are unreliable. Cellular network operators in these off-grid sites constantly rely on diesel powered generators to run BSs which is not only expensive, but also generates CO_2 emissions. One such generator consumes an average of 1500 litres of diesel per month, resulting in a cost of approximately $30,000 per year to the network operator. Moreover, this fuel has to be physically brought to the site and sometimes it is even transported by helicopter in remote places, which adds further to this cost. In such places, renewable energy resources such as sustainable biofuels, solar and wind energy seem to be more viable options to reduce the overall network expenditure. Hence, adopting renewable energy resources could save cellular companies such recurrent costs, since they are capital intensive and cheaper to maintain. Also, since renewable energy is derived from resources that are regenerative, renewable energy resources do not generate greenhouse gases such as CO_2.

Recently, a program called "Green Power for Mobile" to use renewable energy resources for BSs has been started by 25 leading telecoms including MTN Uganda and Zain, united under the Global Systems for Mobile communications Association (GSMA) [114]. This program is meant to aid the mobile industry to deploy solar, wind, or sustainable biofuels technologies to power 118,000 new and existing off-grid BSs in developing countries by 2012. Powering that many BSs on renewable energy would save up to 2.5 billion litres of diesel per annum (0.35% of global diesel consumption of 700 billion litres per annum) and cut annual carbon emissions by up to 6.8 million tonnes.

Such BSs operating on renewable energy resources are expensive and network operators have been reluctant to adopt them because of fear of little commercial viability and lack of equipment expertise. However, according to a bi-annual recent report by GSMA, the implementation of green power technology represents a technically feasible and financially attractive solution with a pay-back period of less than three years at many sites [115].

5.3.4 Other Ways to Reduce BS Power Consumption

Since the energy consumption of the entire cellular network includes the summation of energy used by each BS, reducing the number of BSs has a direct impact on energy consumption of a cellular network. However, efficient network design and finding an optimal balance between cell size and BS capacity can be very challenging. Features such as 2-way and 4-way diversity, feeder less site, extended cell, low frequency band, 6-sector site and smart antenna can be used to minimize the number of BS sites [100].

Another way to improve power efficiency of a BS is to bring some architectural changes to the BS. Currently, the connection between the RF-transmitter and antenna is done by long coaxial cables that add almost 3dB to the losses in power transmission and therefore, low power RF-cables should be used and RF-amplifier has to be kept closer to the antenna [104]. This will improve the efficiency and reliability of the BS. In [116], the authors suggest an all-digital transmitter architecture for green BS that uses a combination of EER and pulse width modulation (PWM)/pulse position modulation (PPM) modulation.

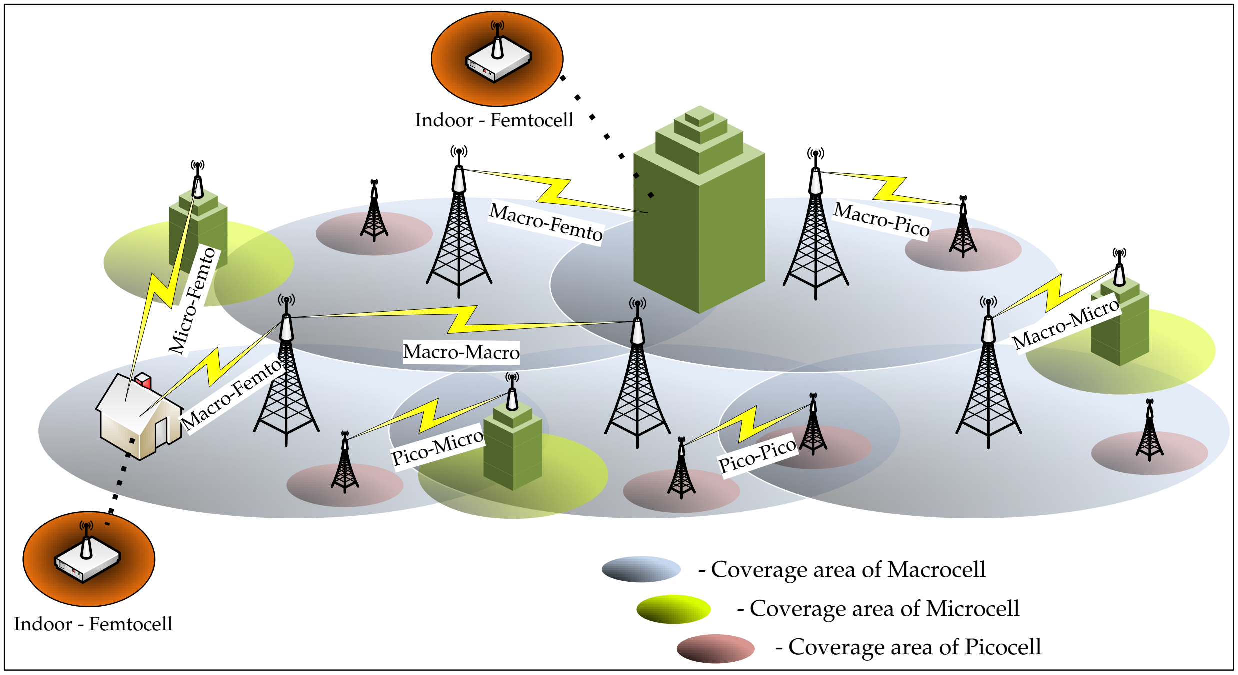


Figure 5.4: A typical heterogeneous network deployment

5.4 Network Planning: Heterogeneous Network Deployment

The exponential growth in demand for higher data rates and other services in wireless networks requires a more dense deployment of base stations within network cells. Whereas conventional macro-cellular network deployments are less efficient, it may not be economically feasible to modify the current network architectures. Macrocells are generally designed to provide large coverage and are not efficient in providing high data rates. One obvious way to make the cellular networks more power efficient in order to sustain high speed data-traffic is by decreasing the propagation distance between nodes, hence reducing the transmission power. Therefore, cellular network deployment solutions based on smaller cells such as micro, pico and femtocells are very promising in this context. A typical heterogeneous network deployment is shown in Fig. 5.4. A micro/picocell is a cell in a mobile phone network served by a low power cellular BS that covers a small area with dense traffic such as a shopping mall, residential areas, a hotel, or a train station. While a typical range of a micro/picocell is in the order of few hundred metres, femtocells are designed to serve much smaller areas such as private homes or indoor areas. The range of femtocells is typically only a few metres and they are generally wired to a private owners' cable broadband connection or a home digital subscriber line (DSL). Smaller cells because of their size are much more power efficient in providing broadband coverage. As an example, a typical femtocell might only have a 100mW PA, and draw 5W total compared to a 5KW that would be needed to support macrocell. An analysis by OFCOM (UK regulator) and Plextek concluded that femtocell deployment could have a 7:1 operational energy advantage ratio over the expansion of the macrocell network to provide approximately similar indoor coverage [117]. Simulations show that with only 20% of customers with picocells, a joint deployment of macrocell and picocell in a network can reduce the energy consumption of the network by up to 60% compared to a network with macro-cells only [104]. Another advantage of smaller cells is that they can use higher frequency bands suitable to provide high data rates and also offer localization of radio transmissions. However, deploying too many smaller cells within a macrocell may reduce the overall efficiency of the macrocell BS, since it will have to operate under low load conditions. Therefore, careful investigation of various deployment strategies should be done in order to find how to best deploy such smaller cells. In [118], Calin *et al.* provided insight into possible architectures/scenarios for joint deployments of macro and femtocells with an analysis framework for quantifying potential macro-offloading benefits in realistic network scenarios. Richter *et al.* in [119], investigate the impact of different deployment strategies on the power consumption of mobile communication network. Considering layouts with different number of micro BSs in a cell, in addition to macro sites, the authors introduce the concept of area power consumption as a system performance metric. Simulation results suggest that under full traffic load scenarios, the use of micro BSs has a rather moderate effect on the area power consumption of a cellular network and strongly depends on the offset power consumption of both the macro and micro sites [119]. In [120], the authors investigate the potential improvements of the same metric achievable in network layouts with different numbers of micro BSs together with macro sites for a given system performance targets under full load conditions.

As large-scale femtocell deployment can result in significant energy consumption, an energy saving procedure that allows femtocell BS to completely turn off its transmissions and processing when not involved in an active call was proposed in [121]. Depending on the voice traffic model, this mechanism can provide an average power saving of 37.5% and for a high traffic scenario, it can achieve five times reduction in the occurrence of mobility events, compared to a fixed pilot transmission [121]. A rather radial approach to create a link between fully centralized (cellular) and decentralized (ad hoc) networks in order to achieve more efficient network deployment is a paradigm shift towards self-organizing small-cell networks (SCNs) [122]. However, coverage and performance prediction, interference and mobility management together with security issues are some of the many issues that must be dealt while designing such networks.

5.5 Enabling Technologies: Cognitive Radio and Cooperative Relaying

Recently, the research on technologies such as cognitive radio and cooperative relaying has received a significant attention by both industry and academia. While cognitive radio is an intelligent and adaptive wireless communication system that enables us to utilize the radio spectrum in a more efficient manner, cooperative relays can provide a lot of improvement in throughput and coverage for futuristic wireless networks. However, developments in both these technologies also enable us to solve the problem of energy efficiency via smart radio transmission and distributed signal processing. In the following subsections, we will discuss how we can enable green communication in cellular systems using cognitive radio and cooperative relaying.

5.5.1 Green Communication via Cognitive Radio

Bandwidth efficiency has been always a crucial concern for wireless communication engineers, and there exist a rich literature on this matter, resulting in bandwidth efficient systems, but not always considering power efficiency. On the other hand, it has been realized that the allocated spectrum is highly underutilized [3], and this is where cognitive radio comes into the picture. The main purpose of cognitive radio is to collect information on the spectrum usage and to try to access the unused frequency bands intelligently, in order to compensate for this spectrum underutilization [123]. However, the question is why using spectrum more efficiently is important and how it can reduce power consumption? The answer lies under Shannon's capacity formula [124], where we can see the trade-off between the bandwidth and power. The capacity increases linearly with bandwidth, but only logarithmically with power. This means that in order to reduce power, we should seek for more bandwidth [125], or in other words, manage the spectrum optimally and dynamically, and this falls into the scope of cognitive radio. In fact, it has been shown in [126] that up to 50% of power can be saved if the operator dynamically manages its spectrum by activities such as dynamically moving users into particularly active bands from other bands, or the sharing of spectrum to allow channel bandwidths to be increased.

However, efficient spectrum usage is not the only concern of cognitive radio. Actually, in the original definition of cognitive radio by J. Mitola [127], every possible parameter measurable by a wireless node or network is taken into account (Cognition) so that the network intelligently modifies its functionality (Reconfigurability) to meet a certain objective. One of these objectives can be power saving. It has been shown in recent works that structures and techniques based on cognitive radio reduce the energy consumption, while maintaining the required quality-of-service (QoS), under various channel conditions [128, 129]. Nevertheless, due to the complexity of these proposed algorithms, still vendors find it unappealing to implement these techniques. Hence, a roadway to future would be striving for more feasible, less complex, and less expensive schemes within the scope of cognitive radio.

5.5.2 Cooperative Relays to Deliver Green Communication

In infra-structured wireless networks, extending coverage of a BS is an important issue. Considering well-known properties of the wireless channel, including large path losses, shadowing effects and different types of signal fading, covering very distant users via direct transmission becomes very expensive in terms of required power in order to establish a reliable connection. This high-power transmission requirement further translates into the high power consumption and also introduces high levels of interface at nearby users and BSs.

On the other hand, in recent years, cooperative communication techniques have been proposed to create a virtual MIMO systems, where installing large antennas on small devices such as MUs is not possible. Hence, using cooperative communication, well-known improvements of MIMO systems including coverage enlarging and capacity enhancement can be achieved [130]. Cooperative techniques also combat shadowing by covering coverage wholes [130]. In fact, early research has shown that relaying techniques extend the battery life [131], which is the first step towards energy efficient networks. In particular, multi-hop communication divides a direct path between mobile terminals and BS into several shorter links [132], in which wireless channel impairments such as path loss are less destructive, hence lower transmission power can be assigned to the BS and relays. Authors in [133] mentioned that two-hop communication consumes less energy than direct communication. And finally, it has been shown in [134] that using multi-hopping in CDMA cellular networks can reduce the average energy consumed per call.

Delivering green communication via cooperative techniques can be achieved by two different approaches. The first approach is to install fixed relays within the network coverage area in order to provide service to more users using less power. And the second approach is to exploit the users to act as relays. In this work, a relay is roughly defined as one of the network elements which can be fixed or mobile, much more sophisticated than a repeater, and it has capabilities such as storing and forwarding data, and cooperating in scheduling and routing procedures. While the second scenario eliminates the cost of installing relay nodes, it increases the complexity of the system, mostly because centralized or distributed algorithms must be designed to dynamically select relays among the users, as well as new user mobile terminals have to be designed such that they support relaying. In the two following sub-sections, we discuss these two scenarios.

Enabling Green Communication via Fixed Relays

Nonlinear signal attenuation or path loss is an interesting property of a wireless channel. This property helps to concentrate power on specific locations in a network, hence, leads to spatial reuse of various resources within a wireless network. A simple example in [135] shows that for an additive white Gaussian noise (AWGN) channel with a path loss exponent of 4, we can increase the number of BSs by a factor of 1.5 in an area unit, and reduce the transmitting power by a factor of 5, while achieving a same signal-to-noise ratio (SNR) level. In other words, a higher density of BSs leads to less energy consumption as well as a higher special reuse [135]. In fact, this is the key point which makes fixed relays a good candidate for delivering green communication as well as a general improvement of network performance. Installing new BSs in order to have a higher BS density can be very expensive. Therefore, we can install relays instead of new BSs, which is economically advantageous, and does not introduce much complexity to the network. First of all, relays need not be as high as BSs, because they are supposed to cover a smaller area with a lower power [130]. Secondly, relays can be wirelessly connected to a BS, instead of being attached to the backhaul of the network by wire using a complicated interface [130]. And finally, in cellular systems, unlike ad-hoc and peer-to-peer networks, complex routing algorithms are not necessary [130]. All these reasons make installing relays a potential solution to having more energy efficient cellular networks.

In a very recent work [135], the authors have discussed how it is possible to deliver a green communication structure in cellular networks, using fixed relays. In this paper, It is shown that relays provide a flexible way to improve the spatial reuse, are less complex than BSs and therefore cheaper to deploy, and the relays reduce the power in the system compared to systems based on direct transmission.

Green Communications in Cellular Networks via User Cooperation

User cooperation was first introduced in [29], and has been shown that not only it increases the data rate, but also the system is more robust, i.e., the achievable rates are less sensitive to channel variations. However, despite all these advantages, energy efficiency issues of user cooperation render this paradigm unappealing in wireless mobile networks. The reason is, increased rate of one user comes at the price of the energy consumed by another user acting as a relay. The limited battery life time of mobile users in a mobile network leads to selfish users who do not have incentive to cooperate. In fact, in a very recent work by Nokleby and Aazhang [136], this fundamental question has been posed: whether or not user cooperation is advantageous from the perspective of energy efficiency. In this paper, a game-theoretic approach is proposed to give users incentive to act as relays when they are idle, and it is shown that user cooperation has the potential of simultaneously improving both user's bits-per-energy efficiency under different channel conditions.

User cooperation in which selfish users find cooperation favourable to their energy concerns, has recently been considered, but has still not attracted much research. However, based on existing literature, this new approach can be a promising technique to increase the system performance in terms of energy efficiency in future wireless mobile networks.

5.6 Design: Addressing Energy Efficiency in Future Generation Wireless Systems

In previous sections, we discussed that how cognitive radio and cooperative communication are becoming key technologies to address the power efficiency of a cellular network. As we mentioned earlier, European Union has already started C2POWER project with objectives to reduce power consumption of mobile terminals using cognitive and cooperative technologies by up to 50%. In this section, we will mainly discuss techniques to enable green communication in future generation of wireless systems that will rely on cooperation and cognition to meet the increasing demand for high data rates. So far, achieving high data rate has been the primary focus of research in cooperative and cognitive radio systems, without much consideration of energy efficiency. However, many of these techniques significantly increase system complexity and energy consumption. For instance, in the context of green communication via cognitive radio, authors of [137] mention that there are two fundamental but entangled aspects: how to use cognitive radio for energy efficiency purposes, and how to make the cognitive radio operate in an energy efficient manner. Escalating energy costs and environmental concerns have already created an urgent need for more energy-efficient "green" wireless communication. Hence, we need to be proactive in designing energy-efficient solutions for cooperative and cognitive networks, which will potentially drive the future generation of wireless communication. As an example, if cognitive and cooperative techniques are expected to give 50% of power savings, then an additional 50% improvement in the energy efficiency of these techniques will further increase the net savings by 25%.

In the next few subsections, we will discuss an approach to obtain energy efficiency of cellular networks on an algorithmic and protocol design level, instead of energy-efficient circuitry design for communication devices.

5.6.1 Low-Energy Spectrum Sensing

The use of cognitive radio technology requires frequent sensing of the radio spectrum and processing of the sensor data which would require additional power. Therefore, it is necessary to design energy-efficient sensing schemes so that improvement in data rate due to opportunistically acquired spectrum does not lead to significant increase in the energy consumption. Low-complexity spectrum sensing techniques such as energy detection require high sensing time to accurately detect a primary signal and even fail to detect the signal at low SNR due to presence of noise-uncertainty [138]. Therefore, detectors exploiting the cyclostationarity of the primary signals have been studied in the literature that perform better at low SNR. However, they are highly complex and need significant processing power. Therefore, design of low-complexity cyclostationary detectors needs to be investigated. Cooperative spectrum sensing improves the sensing performance by using the spatial diversity between various sensors [138]. However, cooperative sensing would also increase the signaling overhead and thus, energy consumption. By taking into consideration the power consumed for sensing, processing and transmitting sensing data, we need to find conditions under which cooperative sensing is more energy efficient in order to achieve a certain sensing performance. New strategies should be designed to select the sensors to participate in cooperative sensing that could reduce the power consumed without severe loss in the sensing performance. Also, optimal location of the sensors should be determined that would make the sensing system energy efficient in presence of a single and multiple primary users.

Cluster-based sensing architecture has been shown to achieve higher energy efficiency and hence cluster-based designs to reduce power consumption should be considered in research [139]. Sequential detection techniques also needs to be explored to improve the energy efficiency of the system [140]. Compressive sensing has recently been proposed to reduce the complexity of wideband sensing by sampling at a rate significantly lower than Nyquist rate, taking advantage of the sparse nature of the radio spectrum usage [141]. Therefore, efficient cooperative compressive spectrum sensing schemes is also a possible research area.

5.6.2 Energy-Aware Medium Access Control and Green Routing

Medium access control (MAC) in cooperative and cognitive wireless systems introduces a number of new challenges unseen in traditional wireless systems. For example, coordinating medium access in presence of multiple relays with different channel qualities requires a much more agile and adaptive MAC in cooperative systems. In cognitive radio systems, sensing accuracy, duration and time varying availability of primary user channels are some of the factors affecting the MAC design. The need for optimizing energy consumption further adds another dimension that can be conflicting to the goal of achieving better system performance, user satisfaction and QoS. Many of the cooperative and cognitive wireless systems will rely on multihop communication between a transmitter and its intended receiver. In addition to MAC design, proper routing schemes will thus be necessary to achieve desired end-to-end QoS.

Although, a number of MAC and routing schemes specialized for cooperative and cognitive networks exist in the literature [142, 143], little research has been done regarding the energy efficiency of such systems. For instance, a significant volume of research exists on joint routing and spectrum allocation with objectives of throughput maximization in multi-hop cognitive and cooperative systems. In [144], decentralized and localized algorithms for joint dynamic routing, relay assignment, and spectrum allocation in a distributed and dynamic environment are proposed and analyzed. However, most of the research on joint routing and spectrum allocation does not take into account power efficiency constraints directly. Nevertheless, throughput maximization via routing-driven spectrum allocation can be interpreted as power efficiency, since more throughput is achieved using the same amount of power.

As another example of MAC and routing schemes specialized for cooperative and cognitive networks, in [145], Alonso-Zarate *et al.* proposed persistent relay carrier sensing multiple access (PRCSMA) MAC protocol employing distributed cooperative automatic retransmission request (C-ARQ) scheme (users who overheard the message can act as spontaneous relays for retransmission) in IEEE 802.11 wireless networks and in [146], they recently evaluated the energy consumption of this protocol. In particular, Alonso-Zarate *et al.* described the conditions under which a C-ARQ scheme with PRCSMA outperforms non-cooperative ARQ schemes in terms of energy efficiency. On the other hand, some energy-aware MAC and routing mechanisms [147, 148] exist primarily for wireless sensor networks. However, sensor networks are very different than cooperative and cognitive networks in system dynamics and performance objectives. Therefore for cellular networks, objective should be to investigate novel energy-efficient MAC and routing schemes design for cooperative and cognitive wireless networks. In addition, we need to focus on optimizing energy consumption while delivering desired system performance, user satisfaction and QoS.

Hybrid-ARQ (HARQ) are another set of ARQ type protocols that use Forward-Error-Correction (FEC) coding and can be typically employed at the MAC layer to improve QoS and robustness for delay insensitive applications. There are three important subclasses of HARQ protocols namely: HARQ-IT (Type I, in which erroneous data packets are retransmitted for memoryless detection), HARQ-CC (chase combining, where packets in error are preserved for soft combining), and HARQ-IR (Incremental Redundancy, where every retransmission contains different information bits than previous one). HARQ protocols can potentially reduce the transmission energy required for decoding at the destination for delay insensitive systems and the total energy consumption for both the transmission power and the energy consumed in the electronic circuitry of all involved terminals (source, destination and, even relays) has been studied in [149]. Hence, future MAC protocols for cognitive and cooperative systems that employ HARQ has potential to reduce energy costs of such systems.

In cooperative systems, the medium time is accessed for both direct and relayed transmissions. Addition of relayed transmission means more power consumption in the network. Future research on developing a MAC protocol that will be able to suitably quantify potential performance gain in QoS against any additional energy consumption and coordinate medium access among direct and relayed transmissions, will be important. Also, while doing so, focus should be on low-complexity schemes so that the energy savings acquired are not wasted in an increased need for processing power. For routing in a multihop cooperative system, we need to employ new protocols that can intelligently use the most energy-efficient path given the relays that are selected by the resource allocation and MAC schemes. In order to facilitate the operation of our targeted mechanisms, we must explore analytical models that can quantify trade-offs between energy savings and end-to-end QoS performance from selecting alternate routing paths. In this regard, there has already been a paradigm shift from early flooding-based and hierarchical protocols to geographic and self-organizing coordinate-based routing solutions and Internet Engineering Task Force (IETF) Routing Over Low power and Lossy networks (ROLL) working group is in process of standardization of Routing Protocol for Low power and lossy networks (RPL) [150, 151].

For cognitive networks, energy efficiency in the MAC can be increased significantly if the access mechanism is designed to avoid collisions between primary and secondary users. Existing random access based protocols must be modified to achieve this objective in a distributed cognitive MAC with as low system complexity as possible. Statistical information of available channels can be used for QoS provisioning such as in [152] but we should also consider energy efficiency as a trade-off. Furthermore, we must focus on developing analytical models to relate important parameters of these random access methods to resulting energy consumption and QoS performance. This will enable the system engineers to choose optimal parameter values to minimize energy consumption while satisfying desired QoS performance. In a multihop cognitive radio network, due to the presence of primary user spectrum, more energy efficient routes can now be selected which would not be available without cognitive technology. Routing algorithms should be designed such that they can utilize these additional routes to minimize energy consumption.

5.6.3 Energy-Efficient Resource Management with Applications in Heterogeneous Networks

Energy consumption in wireless networks is closely related to their radio resource management schemes. Recently, power-efficient resource management for wireless networks based on cooperative and cognitive architectures has been discussed in [153, 154]. However, current research that addresses energy efficient resource management for these systems under a variety of network objectives and constraints is not yet fully developed.

For cooperative systems, relaying mechanisms that minimize energy consumption while satisfying certain QoS performance criterion should be investigated. We also need to explore distributed schemes based on economic models with energy as a cost in the overall utility function. More specifically, we need to find answers to three fundamental questions: "where to place relays", "whom to relay" and "when to relay". In order to answer the first question "where to place relays", we first need to obtain the optimal relay geometry, in terms of energy consumption, within a cell with different number of relays and then we must also optimize the number of relays. The second question is related to the design of optimal relay selection criterion. This relay selection criterion should be based on both fairness and energy efficiency. The authors in [155, 156] have proposed an energy efficient distributed relay selection criterion with finite-state Markov channels and adaptive modulation and coding in a single user cooperative system. Such ideas should be explored for multi-user scenario. To solve the last problem of "when to relay", design of resource allocation strategies for single and multiuser wireless systems should be studied such that relaying is selectively enabled so as to reduce overall power consumption [157, 158].

To improve energy efficiency of cognitive radio systems, energy consumed per bit can be taken as performance metric [159]. We also need to investigate low power consumption based scheduling mechanisms in presence of multiple cognitive users. In [160], the authors have proposed some energy-efficient and low complexity scheduling mechanisms for uplink cognitive cellular networks and have shown that round-robin scheduling is more energy efficient than opportunistic scheduling while providing the same average capacity and BER. Using mathematical tools based on dynamic programming and optimal control, we need to design resource allocation schemes for cognitive radio systems such that overall power consumption is minimized over a period of time while providing satisfactory performance.

Lastly, we also need to investigate into the design of energy-aware heterogeneous networks, where the macrocell (high-power node) and femtocell (low-power node) coexist for co-channel deployment. Femtocell must cognitively adapt to its surrounding environment and transmit in such a way in order not to create cross-tier interference to main cellular network [161]. Using cooperative relaying, the network coverage of these femtocells can be improved without causing huge interference to the macrocell system. For such heterogeneous network architecture, employing power-efficient resource management techniques such as selective relaying, energy efficient modulation etc. can be very attractive. Further research on optimal cell sizes and femtocell BS locations taking into consideration the energy spent for the system backhaul and signaling overhead, can save a lot of power. This way we can further reduce the energy consumption of macrocell BSs and user handsets to achieve a given system performance.

5.6.4 Cross-Layer Design and Optimization

Cellular wireless communication systems ought to support different kinds of applications, including voice and data applications. Each application has a different energy limit, required bit rate, bit error rate, delay constraints, outage probability, etc. Traditionally, these requirements have been tried to be achieved within the scope of a layered structure, called protocol stack. In fact, there has been a significant amount of research tackling these problems within each layer, assuming each layer operates independently of the other layers. Examples of these efforts in the link layer are employing MIMO techniques, channel coding, power control and adaptive resource allocation techniques. In the MAC layer, different channelization or random access schemes, along with scheduling and power control can be mentioned. Moreover, regarding the network layer, a rich literature can be found on the energy-constrained and delay-constrained routing. And finally, adjusting QoS requirements adaptively is a venue to meet the users demand in the application layer [162].

The rationale behind using the protocol stack is that it helps the designer to break the design problem into several simpler problems, namely layer modules. This paradigm also makes the evaluation of the proposed algorithms easier. However, limiting each layer to be independent of others and sub-optimality of this modularized paradigm lead to a poor performance, especially when the resources such as energy are scarce [162]. Therefore, cross-layer design can be a very useful tool to minimize energy consumed across the entire protocol stack. In cross-layer design, we try to escape from the limitations that the traditional waterfall-like concept of protocol stack imposes. In the new paradigm, we want to not only consider the interdependencies between different layers, but also take advantage of them. In particular, for energy saving purposes, it is necessary to consider the invariably changing operation conditions in cellular networks. Due to the mobility of the users, and also characteristics of wireless channel along with the nature of modern applications, propagation environment and application requirements are time varying. Thus, more holistic control algorithms from cross-layer perspective must be designed which adapts the system to these dynamics at run time. Going in this direction, greener cellular communication systems can be delivered compared to the existing ones [163]. However, if not carefully designed cross-layer design might itself lead to increased complexity and energy consumption. Hence, we should explore the cross-layer alternatives to schemes proposed for an individual layer (PHY, MAC etc.) and analyze important tradeoffs in energy consumption and system performance. Multiple relays in cooperative communication and spectrum sensing mechanism in cognitive radio networks introduce new challenges in cross-layer design for these networks [164, 165].

One of the objectives should be to devise cross-layer schemes that will allow joint optimization of some or all of the following parameters: assigning the subcarriers, rates, and power (physical layer attributes), channel access mechanisms (MAC layer attribute), routing (network layer attribute), and rate (transport layer attribute) while taking into account system related errors (e.g. sensing errors in cognitive radios) and other errors that contribute to cross layer issues. Further, in order to save energy in an ad-hoc wireless network, more packets should be transmitted when channel quality is good and at the same time collision losses due to network congestion must be reduced so that the packets need not to be retransmitted. In this regard, we need to explore new mechanisms for cognitive radio and cooperative ad-hoc networks by which channel traffic can be measured either by single-bit or multi-bit signalling overhead or by loss or delay in the network.

Energy-efficient cross-layer schemes that will optimize resources in various layers while considering the channel quality as well as the network traffic, should be investigated.

An example, where cross-layer design would be crucial, is in the use of cooperative relaying to improve spectrum diversity in cognitive radio networks [166]. Large gains in efficiency and fairness of resource sharing can be obtained by cooperation among cognitive radio nodes. Specifically, some cognitive radio users with low traffic demand can help improve spectrum efficiency by acting as relays for the cognitive radio users that have high traffic demand but low available bandwidth. For such a network, cross-layer design is important as while performing resource allocation (relay, power etc.), transmission demand of each cognitive radio user has to be taken into account.

5.6.5 Addressing Uncertainty Issues

Most research in the field of cooperative and cognitive radio systems is mainly based on the assumption of perfect channel state information (CSI), which is often unrealistic in practice. Presence of non-Gaussian noise, quantization effects, fast varying environment, delay in CSI feedback systems and hardware limitations are the main factors that cause errors in CSI. The performance of cognitive radio sensing system is drastically impaired, when various wireless channels e.g., detecting, reporting, and inter-user channels have uncertainty [167]. For cooperative systems, the optimal relay selection and robust resource allocation with imperfect CSI has also remained largely unexplored. For cognitive radio systems, it is also important to take into account the effects of imperfect sensing [168]. Spectrum sensing is further complicated due to uncertainty in interference from other secondary networks [169]. Providing robustness in conjunction with energy efficient solutions to such scenarios is, therefore, a task of significant practical interest. Also, the robustness of the efficient scheduling schemes for MAC and cross-layer optimization needs further investigation taking uncertainty in the channel congestion into account. In order to maintain energy savings under imprecise conditions, we must investigate the robustness of our proposed energy-efficient schemes and compare the performance with the existing schemes in practical scenario with uncertain environment. Hence, depending on the QoS targets, robust algorithms for energy efficient resource optimization considering uncertainty in CSI, should be explored.

5.7 Some Broader Perspectives

The most important issue in developing networks which are energy-aware is to model the consumption of the wireless interfaces [170]. Usually, the wireless interface consumes energy with the same rate in receive, transmit or idle states. In turn, the less the wireless interface is operating, the less energy is consumed. Based on the preceding argument, the best strategy to minimize the energy consumption is to shut down the wireless interface, or to go to energy saving mode as much as possible. In order to achieve this, algorithms needed to determine when it is suitable to switch to energy saving mode or turning off the transceivers. We already have discussed strategies with the aforementioned concept in this paper. For instance, we have mentioned Discontinuous Reception (DRX) and Discontinuous Transmission (DTX) modes in LTE standard, and sleep mode mechanism in IEEE 802.16e, both for mobile terminals. We also have talked about enabling sleep mode for BSs. However, these methods are based on instantaneous observations. On the other hand, the traffic pattern is dramatically different in different times of the day or in different geographical locations. In a broader perspective, there can be a data-base in BS and mobile terminals, in which the traffic pattern during different times of the day is saved. Based on this obtained statistics, dynamic algorithms can be designed in order to switch the BS or mobile terminal to a different power profile appropriate for that time of the day. In a recent paper by Dufkova *et al.* [171], it has been shown that if such predictions on users are available, savings from 25% up to 50% can be achieved, depending on the time of the day. However, their results are based on off-line optimizations and represent an upper bound on the energy savings possible.

From another perspective, BSs distributed over a certain geographical area are connected to a power grid. In recent years, smart grid has emerged to coordinate the power generators, transmission systems and appliances utilizing two-way communication lines between all these different entities. These two-way communication lines can be dedicated point-to-point wireless channels or IP-based connections [172]. On the other hand, BSs in general, are power hungry elements. Hence, looking at BSs as power consumers or appliances, and absorbing them in a smart grid can exceedingly increase the power efficiency without adversely affecting the QoS and capacity. This can be done by adding measurement sensors which can update the status of BSs, and then transmit them to the other BSs and smart grid control system. Here, in addition to cooperation with each other, BSs also cooperate with the power system to manage the energy consumption.

Contrary to the ideas mentioned by far, Humar *et al.* in [173] suggest a different way of thinking in energy efficiency modeling. Almost all the research on making cellular communication green results in larger number of BSs with lower level of powers, since the objective is to reduce the *operating energy*. However, authors of [173] noticed that in all the cases, the new BSs are more sophisticated equipments, and producing these sophisticated equipments requires more energy compared to conventional ones. This energy which is associated with all the processes of producing an equipment is called *embodied energy*. According to this paper, embodied energy accounts for a significant proportion of energy consumed by the BS, and taking this energy into account along with operating energy in modeling cellular network's energy consumption results in solutions which disagree with increasing the number of BSs and lowering their power.

Table 5.2: Energy savings obtained by some of the discussed techniques

Description	**Reported savings**
Improvements in Power Amplifier	- up to 50% with doherty architecture and GaN-based amplifiers - up to 70% with switch-mode power amplifiers
Network self-organizing techniques	between 20-40% BS power savings
Renewable Energy Resources in off-grid sites	up to 0.35% of global diesel consumption
Heterogeneous network deployment	up to 60% savings compared to a network with macro-cells
Dynamic spectrum management	up to 50%

5.8 Conclusion

This paper addresses the energy efficiency of cellular communication systems, which is becoming a major concern for network operators to not only reduce the operational costs, but also to reduce their environmental effects. We began our discussion with green metrics or energy efficiency metrics. Here, we presented a brief survey of current efforts for the standardization of the metrics and the challenges that lay ahead. Regarding architecture, since BSs represent a major chunk of energy consumed in a cellular network, we then presented an exhaustive survey of methods that have been currently adopted or will be adopted in future in order to obtain energy savings from BSs. In particular, we discussed the recent improvements in power amplifier technology that can be used to bring energy savings in BSs. Improvements in the power amplifier will not only decrease the power consumption of the hardware system, but will also make the BS less dependant on air-conditioning. We also discussed the power saving protocols such as sleep modes, that have been suggested for next generation wireless standards. Such power saving protocols at the BS side still need to be explored in future wireless systems. Next, we discussed energy-aware cooperative BS power management, where certain BSs can be turned off depending on the load. A recent concept called “Cell zooming” appears to be a promising solution in this regard. Another way to significantly reduce the power consumption of BSs, in particular, those at the off-grid sites, is by using renewable energy resources such as solar and wind energy in place of diesel generators. Lastly, we discussed how minimizing the number of BSs with a better network design and bringing minor architectural changes can be beneficial in achieving energy efficiency.

Heterogeneous network deployment based on smaller cells such as micro, pico and femtocells is another significant technique that can possibly reduce the power consumption of a cellular network. However, as some of the recent research suggests, careful network design is required as deploying too many smaller cells may in fact reduce the power efficiency of the central BS. Also, when a large number of BSs with small cell sizes are deployed, the embodied energy consumption will dominate and lead to an increase in total energy consumption [173]. We also discussed how emerging technologies such as cognitive radio and cooperative relaying can be useful for obtaining "green" network technology. In this regard, we discussed research challenges to address energy efficiency in cognitive and cooperative networks including low-energy spectrum sensing, energy-aware MAC and routing, efficient resource management, cross-layer optimization, and uncertainty issues. Finally, we explored some broader perspectives such as statistical power profiles, smart grid technology and embodied energy to achieve energy efficient cellular network. Table 5.2 lists the energy savings reported by authors, that can be obtained by some of the techniques discussed in the paper.

In summary, research on energy efficient or "green" cellular network is quite broad and a number of research issues and challenges lay ahead. Nevertheless, it is in favor of both the network operators and the society to swiftly address these challenges to minimize the environmental and financial impact of such a fast growing and widely adopted technology. This article attempts to briefly explore the current technology with respect to some aspects related to green communications and we discuss future research that may prove beneficial in pursuing this vision.

Chapter 6

Conclusions and Future Research Directions

6.1 Conclusions

In this thesis, we have developed various resource management schemes for OFDM-based cooperative and cognitive systems and have also explored energy efficiency of next generation wireless systems. We have made the following four major contributions in this dissertation.

First, we presented a solution to the problem of energy efficient resource allocation for maximizing the expected transmission rate for an OFDM-CR system by taking into account the "reliability" of the opportunistic frequency bands (which depends on sensing errors and PU activity). In addition to reliability, we also included sub-band power constraints and total allowed interference limit to the adjacent PU bands in our problem formulation. We introduced a risk-return model that considers a linear average rate loss function in the optimization objective, in order to integrate the channel reliability. Because of the high complexity of the computation of the optimal solution, we suggested three suboptimal schemes; step-ladder, nulling and scaling schemes. Comparing the performance of these suboptimal schemes to waterfilling and optimal schemes, we demonstrated that suboptimal schemes perform on a level close to that of the optimal scheme while also satisfying the desired constraints. We also studied the pros and cons of each of these suboptimal schemes. We concluded that classical waterfilling cannot be used for practical OFDM-CR systems because it fails to keep interference limits of desired levels.

Second, we studied the problem of designing efficient resource allocation schemes for multiuser OFDMA-based cooperative network with selective relaying. Selective relaying is a concept that challenges the traditional view of the "always relay" scenario, in which source adaptively decides on whether relaying should be performed on each subcarrier or not. Assuming both the source and the relay have fixed power constraints and that the system uses a two-phase relaying protocol with amplify and forward relaying, we formulated the problem as a capacity maximizing integer programming optimization problem and then proposed a heuristic solution. The heuristic solution is composed of several steps; the first is to suboptimally allocate subcarriers to different users based on proportional rate fairness, next step is a two part iterative approach aimed at finding relay decisions and power allocation at the source during the first phase and at the relay during the second phase, and the last step is a waterfilling algorithm to allocate power at the source during the second phase to all the non-relaying subcarriers. Numerical results demonstrate that this selective relaying mechanism outperforms the "always relay" case even in low SNR regions, while also providing an approximate proportional rate fairness.

Third, we considered a relay network that relies on user cooperation. We discussed how in such a system we require incentive mechanisms to make communication possible because of the asymmetric CSI on the relay-destination link between the source and the relay. Specifically, we studied the problem of relay selection with incentive mechanisms in OFDM-based cooperative wireless system that employs DF relaying. Due to the information asymmetry, it is much more difficult for the source to choose the best relay nodes while maximizing its throughput. With the help of contract theory, we addressed this problem by modelling the source and the relays in a simple *principal-agent* type system and solved the problem by splitting it in two parts. In the first part we designed incentive compatible offers/contracts for the relays, consisting of a menu of payments and desired signal-to-noise-ratios (SNR)s at the destination. The source then broadcasts this menu to nearby mobile nodes and the relays respond. The second part of the problem is to choose optimal relays while the source is under a budget. This problem is shown to be a nonlinear non-separable knapsack problem and we suggested a heuristic solution to solve it efficiently. Selected numerical results show that our solution performs better than a simple relay selection mechanism and that this performance is also close to optimal.

Finally, we explored the energy efficient design of emerging cellular systems. We discussed some research issues and challenges and suggested some techniques to enable the creation of an energy efficient or "green" cellular network. We identified four important aspects of a green networking system design: defining green metrics, bringing architectural changes in base stations, network planning, and efficient system design. We began with a brief discussion on energy efficiency metrics. Since BSs consume the major chunk of input energy, we discussed the energy efficiency of BSs more at the component level. Here, we studied how to minimize energy consumption of BS by employing improvements in the power amplifier, designing power saving protocols, implementing cooperative BS power management, using renewable energy resources and bringing some simple architectural changes. Regarding energy efficiency from a network planning perspective, we discussed how different types of network deployments based on smaller cells, otherwise known as heterogenous networks can be used to increase the energy efficiency of a wireless system. Regarding the system design, we put a special emphasis on cognitive and cooperative techniques, in order to bring attention to the benefits cellular systems can attain by employing such techniques, and we also highlighted new research avenues in making these techniques green. In this context we have examined techniques such as low energy spectrum sensing, energy-aware medium access control and routing, efficient resource management, cross-layer design and addressing uncertainty issues. Towards the end of this study we discussed some broader perspectives and possible future trends in realizing a "green" cellular network technology.

6.2 Future Research Directions

The research conducted in this thesis has resulted in many interesting and challenging problems which leads us to the following possible research directions.

6.2.1 Power Allocation Schemes for OFDM-CR Systems

In future, we can extend the rate-maximizing power allocation schemes for OFDM-CR systems based on partial channel information. A related analysis for general OFDM systems was presented in [174]. Another important point to note is that in the work presented in chapter 2, we took only one overall interference threshold for all the adjacent PU bands which is a limitation. More than one constraint could exist due to the presence of multiple primary users in the adjacent bands. So in future, we could consider a general scenario such as this. Furthermore, we can extend the work to find power and rate adaptation algorithms for both socio-optimal and non-cooperative multiuser CR network scenario.

6.2.2 Subcarrier Pairing and Power Allocation Schemes for OFDM-based Cooperative Networks

In our research on cooperative systems with selective relaying in Chapter 3, we so far assumed that during the second time slot the source does not reallocate power to those subcarriers on which relaying is not performed. However, a higher capacity can be achieved by reallocating power on such subcarriers because more power can be allocated per subcarrier during the second time slot. Therefore we can modify our problem in such a way that the source reallocates power on the second time slot. Subcarrier pairing is another technique which completely utilizes the spatial diversity and through which we can further boost overall data rate by pairing different subcarriers during the second time slot for a half duplex cooperative relay system. We can therefore formulate a problem of subcarrier pairing in a multiuser OFDMA relay network and extend our model for selective relaying with subcarrier pairing for a multiuser system.

6.2.3 Joint Relay Selection and Resource Allocation Algorithms for Multiuser Multi-relay System

When there is more than one relay in a system, selecting relays to optimize the data rate can be very challenging. Relay selection and resource allocation are very important to maximize the gains achieved by using cooperative relay networks over conventional networks. Relay selection will depend on the SNR between different relays nodes, mobiles, and base stations and on the number of available subcarriers at each relay. A multi-relay scenario was recently proposed in [175]. However, selective relaying was again not considered in this work. Also, to the best of our knowledge, a multiuser, multi-relay has never been addressed in the literature. So one possible direction of research could be to devise simplified low-complexity schemes in order to solve this complex problem for multiuser multi-relay system together with selective relaying.

6.2.4 Adverse Selection and Moral Hazard Problems in Cognitive Radios and Cooperative Networks

In Chapter 4, we studied one particular scenario of information asymmetry, i.e., adverse selection in cooperative wireless networks. However, many different scenarios of information asymmetry could exist both in cognitive radio and in cooperative networks. For example, we can also model the user cooperation in relay network as a moral hazard problem instead of an adverse selection problem. Specifically, the action of a relay can be defined as the amount of power the relay uses to transmit the source's data. Although, the source cannot observe how much power relay has used for retransmission, it can estimate the received SNR and it knows the probability of receiving that SNR given the action taken by relay. Similarly we can also formulate adverse selection and moral hazard problems in cognitive radios. One example of information asymmetry in cognitive radio is that the SUs are not perfectly aware of channel gains on the secondary-primary link. Hence SUs are not aware of the interference that they cause to the primary users. Once again, contract theory-based modeling can be used to solve these problems.

Bibliography

[1] ITU-D Telecommunication Development Sector Report "Measuring the Information Society: Executive Summary", 2012. *Available:* http://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-ICTOI-2012-SUM-PDF-E.pdf → pages 1

[2] Ian F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty,"NeXt Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey," *C*omputer Networks (Elsevier) Journal, vol. 50, pp. 2127-2159, Sep. 2006. → pages 2, 3

[3] Federal Communications Commission, "Spectrum Policy Task Force," Rep. ET Docket no. 02-135, Nov. 2002. → pages 2, 94

[4] Hseyin Arslan, *Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems*, Springer, 2007. → pages 3

[5] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE Journal of Selected Areas in Communications*, vol. 23, no. 2, pp. 201-220, Feb 2005. → pages 3

[6] *Cognitive Wireless Communication Networks*, Eds. E. Hossain and V. K. Bhargava, Springer, 2007. → pages 3

[7] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," *P*roc. of Asilomar Conf. on Signals, Systems, and Computers, pp. 772-776, Pacific Grove, CA, Nov. 2004. → pages 3

[8] Z. Li, F. R. Yu, and M. Huang, "A Distributed Consensus-Based Cooperative Spectrum Sensing in Cognitive Radios," *IEEE Trans. Vehicular Technology*, vol. 59, no. 1, pp. 383-393, Jan. 2010. → pages 3

[9] A. F. Molisch, *Wireless Communications*, second edition, Wiley, 2011. → pages 3, 55, 56

[10] *Cooperative Cellular Wireless Networks*, Eds. E. Hossain, D. I. Kim, and V. K. Bhargava, Cambridge University Press, 2011. → pages 5

[11] George A. Akerlof, "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, vol. 84, no. 3, pp. 480-500, Aug 1970. → pages 5

[12] T. Weiss and F. K. Jondral, "Spectrum pooling: An innovative strategy for the enhancement of spectrum efficiency," *IEEE Communication Magazine*, vol. 43, no. 3, pp. S8-S14, Mar 2004. → pages 8

[13] D. Tse and P. Vishwanath, *Fundamentals of Wireless Communications*, Cambridge University Press, 2005. → pages 8, 41, 44

[14] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "Practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. on Commun.*, vol. 43, no. 2-4, pp. 773-775, 1995. → pages 8

[15] A. Leke and J. M. Cioffi, "A maximum rate loading algorithm for discrete multitone modulation systems," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, vol. 3, pp. 1514-1518, Phoenix, Ariz, USA, Nov. 1997. → pages 8

[16] R. F. H. Fischer and J. B. Huber, "A new loading algorithm for discrete multitone transmission," in *Proc. of IEEE GLOBECOM'96*, vol. 1, pp. 724-728, London, UK, Nov. 1996. → pages 8

[17] J. Jang, K. B. Lee, and Y. H. Lee, "Transmit power and bit allocation for OFDM systems in a fading channel," in *Proc. of IEEE GLOBECOM'03*, vol. 2, pp. 858-862, 2003. → pages 8

[18] B. S. Krongold, K. Ramchandran, and D. L. Jones, "Computationally efficient optimal power allocation algorithm for multicarrier communication systems," *IEEE Trans. on Commun.*, vol. 48, pp. 23-27, January 2000. → pages 8

[19] N. Papandreou and T. Antonakopoulos, "A new computationally efficient discrete bit-loading algorithm for DMT applications," *IEEE Trans. on Commun.*, vol. 53, no. 5, pp. 785-789, May 2005. → pages 8

[20] N. Papandreou and T. Antonakopoulos, "Bit and power allocation in constrained multicarrier systems: The single-user case," *EURASIP Journal on Advances in Signal Processing*, Issue 1, vol 2008. → pages 9

[21] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, Oct. 1999, pp. 1474-1758. → pages 9

[22] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. IEEE 51st Vehicular Technology Conference (VTC 2000-Spring)*, Tokyo, vol. 2, 2000, pp. 1085-1089. → pages 9

[23] I. Hammerstrom and A. Wittneben, "On the optimal power allocation for nonregenerative OFDM relay links," in *Proc. IEEE ICC'06*, vol. 10, pp. 4463–4468, Jun. 2006. → pages 9, 35, 37, 41, 45

[24] R. U. Nabar, H. Bolcskei, and F. W. Kneubhler, "Fading relay channels: Performance limits and space-time signal design," in *IEEE Journal on Selected Areas in Communications*, vol.22, no. 6, pp.1099-1109, June 2004. → pages 9, 36

[25] M. Pischella and J.-C. Belfiore, "Power control in distributed cooperative OFDMA cellular networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1900–1906. → pages 9

[26] B. Gui and L. J. Cimini, Jr., "Bit loading algorithms for cooperative OFDM systems," *EURASIP J. Wirel. Commun. Netw.*, vol. 2008, no. 2, pp. 1–9, 2008. → pages 9

[27] M. Yi, Y. Na, and R. Tafazolli, "Bit and power loading for OFDM-based three-node relaying communications," in *IEEE Transactions on Signal Processing*, vol. 56, no. 7, Jul. 2008, pp. 3236–3247. → pages 9

[28] W. Zhuang, and M. Ismail, "Cooperation in wireless communication networks," *IEEE Wireless Commun.*, vol. 19, no. 2, pp. 10-20, 2012. → pages 10

[29] 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. → pages 10, 96

[30] A. Ibrahim, A. Sadek, Weifeng Su, and K. Liu, "Cooperative communications with relay-selection: when to cooperate and whom to cooperate with?," *IEEE Trans. on Wireless Commun.*, vol. 7, no. 7, pp. 2814-2827, 2008. → pages 10

[31] I Krikidis, T Charalambous, and J. S Thompson, "Buffer-Aided Relay Selection for Cooperative Diversity Systems without Delay Constraints," *IEEE Trans. on Wireless Commun.*, vol.11, no.5, pp.1957-1967, May 2012. → pages 10

[32] E. Beres, and R. Adve, "Selection cooperation in multi-source cooperative networks," *IEEE Trans. on Wireless Commun.*, vol. 7, pp. 118-127, Jan. 2008. → pages 10

[33] I. Ahmed, A. Nasri, D.S. Michalopoulos, R. Schober, and R. K. Mallik, "Relay Subset Selection and Fair Power Allocation for Best and Partial Relay Selection in Generic Noise and Interference," *IEEE Trans. on Wireless Commun.*, vol.11, no.5, pp.1828-1839, May 2012. → pages 10

[34] I. Krikidis, J. Thompson, S. McLaughlin, and N. Goertz, "Amplify-and-forward with partial relay selection," *IEEE Commun. Letters*, vol.12, no.4, pp.235-237, April 2008. → pages 10

[35] Dejun Yang, Xi Fang, and Guoliang Xue, "Game theory in cooperative communications," *IEEE Wireless Commun.*, vol.19, no.2, pp.44-49, April 2012. → pages 10

[36] "Energy Aware Radio and NeTwork TecHnologies (EARTH)". https://www.ict-earth.eu/ → pages 11, 79

[37] "Towards Real Energy-efficient Network Design (TREND)". http://www.fp7-trend.eu/ → pages 11, 79

[38] "Cognitive Radio and Cooperative strategies for Power saving in multi-standard wireless devices (C2POWER)". http://www.ict-c2power.eu/ → pages 11, 79

[39] "Optimizing Power Efficiency in Mobile Radio Networks (OPERA-NET)". http://opera-net.org/default.aspx → pages 11

[40] P. Wang, M. Zhao, L. Xiao, S. Zhou, and J. Wang, "Power allocation in OFDM-based cognitive radio systems," in *Proc. of IEEE GLOBECOM'07*, pp. 4061-4065, Nov. 2007. → pages 14, 15, 16, 18

[41] G. Bansal, M. J. Hossain, and V. K. Bhargava, "Adaptive power loading for OFDM-based cognitive radio systems," in *Proc. of IEEE International Communications Conference (ICC'07)*, pp. 5137-5142, June 2007. → pages 15, 19

[42] G. Bansal, M. J. Hossain, and V. K. Bhargava, "Optimal and suboptimal power allocation schemes for OFDM-based cognitive radio systems," *IEEE Trans. on Wireless Comm.*, vol. 7, no. 11, pp. 4710-4718, Nov. 2008. → pages 14, 15, 17, 18, 19, 24, 25

[43] T. Weiss, J. Hillenbrand, A. Krohn, and F. K. Jondral, "Mutual interference in OFDM-based spectrum pooling systems," in *Proc. of IEEE VTC'04*, vol. 4, pp. 1873-1877, May 2004. → pages 18, 19

[44] J. Acharya and R. D. Yates, "A framework for dynamic spectrum sharing between cognitive radios," in *Proc. of IEEE ICC'07*, pp. 5166-5171, June 2007. → pages 18

[45] D. Acemoglu, A. Ozdaglar, and R. Srikant, "The marginal user principle for resource allocation in wireless networks, in *IEEE Conf. on Dec. and Control*, vol. 2, pp. 1544-1549, 2004. → pages 20

[46] F. Meshkati, M. Chiang, H. V. Poor, and S. C. Schwartz, "A game theoretic approach to energy-efficient power control in multicarrier CDMA systems, in *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 6, pp. 1115-1129, June 2006. → pages 20

[47] S. Gunturi and F. Paganini, "Game theoretic approach to power control cellular CDMA," *Proc. of IEEE VTC'03*, vol. 4, pp. 2362-2366, Oct. 2003. → pages 20

[48] J. Duplicy, J. Louveaux, and L. Vandendorpe "Utility-based MIMO uplink beamforming," *Proc. of IEEE Workshop SAMSP*, pp. 254-257, 2006. → pages 20, 21

[49] Eduard A. Jorswieck and Holger Boche, "Energy-Aware Utility Regions: Multiple Access Pareto Boundary," *Proc. of ICC'09*, June 2009. → pages 20

[50] D. P. Palomar, and J. R. Fonollasa, "Practical algorithms for a family of waterfilling solutions," *IEEE Trans. on Signal Processing*, vol. 53, no. 2, pp. 686-695, Feb. 2005. → pages 24

[51] Dimitri P. Bertsekas *Nonlinear Programming*, Athena Scientific, 2003. → pages 27

[52] S. W. Peters *et. al*, "Relay architectures for 3GPP LTE-Advanced," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 618787, 14 pages, 2009. → pages 35, 50

[53] IEEE 802.16's Relay Task Group: http://www.ieee802.org/16/relay/ → pages 35, 50

[54] Y. Pan, A. Nix, and M. Beach, "Resource allocation techniques for OFDMA-based decode-and-forward relaying networks," in *Proc. of IEEE VTC, Spring 2008*, pp.1717–1721, 11-14 May 2008. → pages 35

[55] R. Kwak and J. M. Cioffi, "The subchannel-allocation for OFDMA relaying downlink systems with total power constraint," in *Proc. of IEEE VTC, Fall 2008*, pp.1–5, 21-24 Sept. 2008. → pages 35

[56] J. Yuan and Q. Wang, "Adaptive resource allocation schemes for multiuser OFDMA nonregenerative relay networks," in *Proc. of IEEE ICC 2010)*, pp.1–5, 23-27 May 2010. → pages 35, 36

[57] O. Duval, Z. Hasan, E. Hossain, F. Gagnon, and V. K. Bhargava, "Subcarrier selection and power allocation for amplify-and-forward relaying over OFDM links," *IEEE Trans. on Wireless Commun.*, vol.9, no.4, pp.1293–1297, April 2010. → pages 36, 41, 43

[58] H. Boostanimehr, O. Duval, V. K. Bhargava, F. Gagnon, "Selective subcarrier pairing and power allocation for decode-and-forward OFDM relay systems," in *Proc. of IEEE ICC 2010)*, pp.1–5, 23-27 May 2010. → pages 36

[59] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans. on Wireless Commun.*, vol.4, no.6, pp. 2726–2737, Nov. 2005. → pages 36, 39, 40

[60] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. IEEE 51st Vehicular Technology Conference (VTC 2000-Spring)*, Tokyo, vol. 2, 2000, pp. 1085-089. → pages 40

[61] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. → pages 42, 125

[62] T. Riihonen, R. Wichman, and S. Werner, "Evaluation of OFDM(A) Relaying Protocols: Capacity Analysis in Infrastructure Framework," *IEEE Trans. on Vehicular Tech.*, vol.61, no.1, pp.360-374, Jan. 2012. → pages 51

[63] Herbert Gintis, *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction*, Princeton University Press, 2009. → pages 51

[64] Bernard Salanié, *The Economics of Contracts: A Primer*, The MIT Press, March 2005. → pages 51, 53, 56, 57

[65] L. Duan, L. Gao, and J. Huang, "Contract-based cooperative spectrum sharing," *Proc. of IEEE DySPAN*, pp.399-407, May 2011. → pages 51, 56

[66] D. M. Kalathil, and R. Jain, "Spectrum Sharing through Contracts," *Proc. of IEEE DySPAN*, pp. 1-9, 2010. → pages 51

[67] D. M. Kalathil, and R. Jain, "A contracts-based approach for spectrum sharing among cognitive radios," *Proc. of WiOpt*, pp. 91-97, 2010. → pages 51

[68] L. Gao, X. Wang, Y. Xu, and Q. Zhang, "Spectrum Trading in Cognitive Radio Networks: A Contract-Theoretic Modeling Approach," *IEEE Journal on Selected Areas in Commun.*, vol. 29, no. 4, pp. 843-855, 2011. → pages 51

[69] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for WiMax and LTE-advanced mobile systems," *IEEE Communications Magazine*, vol.47, no.10, pp.100-105, October 2009. → pages 53

[70] J. N. Laneman, and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, vol.49, no.10, pp. 2415-2425, Oct. 2003. → pages 55

[71] K. M. Bretthauer, and B. Shetty, "The nonlinear knapsack problem-algorithms and applications," *European Journal of Operational Research*, vol. 138, no. 3, pp. 459-472, 2002. → pages 63

[72] F. S. Hillier, *Introduction to Operations Research*, McGraw Hill, 2004. → pages 63

[73] T. C. Sharkey, H. E. Romeijn, and J. Geunes, "A class of nonlinear nonseparable continuous knapsack and multiple-choice knapsack problems," *Math. Program.*, vol. 126, no. 1, pp. 69-96, Mar. 2009. → pages 63

[74] H. E. Romeijn, J. Geunes, and K. Taaffe, "On a nonseparable convex maximization problem with continuous knapsack constraints," *Operations Research Letters*, vol. 35, no. 2, pp. 172-180, Mar. 2007. → pages 63

[75] B. H. Korte, and J. Vygen, *Combinatorial optimization: Theory and Algorithms*, Springer Verlag, 2008. → pages 66

[76] Patrick Bolton, *Contract Theory*, The MIT Press, December 10, 2004. → pages 67

[77] Z. Hasan, E. Hossain, and V. K. Bhargava, "Resource Allocation for Multiuser OFDMA-Based Amplify-and-Forward Relay Networks with Selective Relaying," *IEEE International Conference on Communications (ICC), 2011*, pp.1-6, 5-9 June 2011. → pages 70

[78] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green Cellular Networks: A Survey, Some Research Issues and Challenges," *IEEE Communications Surveys & Tutorials*, vol.13, no.4, pp. 524-540, Fourth Quarter 2011. → pages 71

[79] Green Power for Mobile, GSMA, "Community Power - Using Mobile to Extend the Grid". *Available:* http://www.gsmworld.com/documents/gpfm _community_power11_white_paper_lores.pdf → pages 79

[80] C. Han et al., "Green radio: radio techniques to enable energy-efficient wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 46-54, 2011. → pages 80

[81] L. M Correia, et. al, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Communications Magazine*, vol.48, no.11, pp.66-72, November 2010. → pages 80, 87, 88

[82] 3GPP TR 32.826, Telecommunication management; Study on Energy Savings Management (ESM), (Release 10), Mar 2010. *Available:* http://www.3gpp.org/ftp/Specs/html-info/32826.htm → pages 78

[83] ITU-T Focus Group on Future Networks (FG FN), FG-FN OD-66, Draft Deliverable on "Overview of Energy Saving of Networks", October 2010. *Available:* http://www.itu.int/dms_pub/itu-t/oth/3A/05/T3A050000660001MSWE.doc → pages 78

[84] Chen Yan, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 30-37, 2011. → pages 79

[85] Alliance for Telecommunications Industry Solutions, "ATIS Report on Wireless Network Energy Efficiency", ATIS Exploratory Group on Green (EGG), Jan. 2010. → pages 83

[86] European Telecommunications Standards Institute, Environmental Engineering (EE) Energy Efficiency of Wireless Access Network Equipment, ETSI TS 102 706, v1.1.1, Aug. 2009 → pages 83, 84

[87] A. P. Bianzino, A. K. Raju, and D. Rossi, "Apple-to-Apple: A framework analysis for energy-efficiency in networks," *Proc. of SIGMETRICS, 2nd GreenMetrics workshop*, 2010. → pages 83, 84

[88] T. Chen, H. Kim, and Y. Yang, "Energy efficiency metrics for green wireless communications," *2010 International Conference on Wireless Communications and Signal Processing (WCSP)*, pp.1-6, 21-23 Oct. 2010. → pages 83, 84

[89] C. Belady, et al., "Green Grid Data Center Power Efficiency Metrics: PUE and DCIE," *The Green Grid*, 2008. → pages 83

[90] The Energy Consumption Rating Initiative, "Energy Efficiency for Network Equipment: Two Steps Beyond Greenwashing", White Paper, August 10 2008. *Available:* http://www.ecrinitiative.org/pdfs/ECR_-_TSBG_1_0.pdf → pages 83

[91] The Energy Consumption Rating Initiative, "Network and Telecom Equipment-Energy and Performance Assessment", Draft 3.0.1, December 14, 2010. *Available:* http://www.ecrinitiative.org → pages 83, 85

[92] Transmittal of ATIS Energy Efficiencies Specifications. *Available:* http://www.itu.int/md/dologin_md.asp?lang=en&id=T09-FG.ICT-IL-0003!!MSW-E. → pages 83

[93] Verizon NEBS Compliance: Energy Efficiency Requirements for Telecommunications Equipment, Issue 4, August 2009. *Available:* http://www.verizonnebs.com/TPRs/VZ-TPR-9205.pdf → pages 83

[94] M. Parker and S. Walker, "Roadmapping ICT: An Absolute Energy Efficiency Metric," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 3, no. 8, August 2011. → pages 84

[95] Asif D. Gandhi, and Mark E. Newbury, “Evaluation of the energy efficiency metrics for wireless networks”, *Bell Labs Technical Journal*, vol. 16, issue 1, pages 207-215, June 2011. → pages 84

[96] Telecommunication Predictions 2010, Technology, Media & Telecommunications Industry Group, Deloitte. *Available:* http://www.deloitte.com/assets/Dcom-UnitedStates/Local Assets/Documents/TMT_us_tmt/us_tmt_telecompredictions2010.pdf → pages 86

[97] “Multiradio Base Station makes network evolution easier and greener than ever”, Press Release Feb 5 2009, Nokia Siemens Networks. *Available:* http://www.nokiasiemensnetworks.com/sites/default/files/document/ NokiaSiemensNetworks_2009_02_05_enFlexiMultiradioBTS.pdf → pages 86

[98] “The green CDMA base station”, Huawei Communicate, Issue 45, 2009. *Available:* http://www.huawei.com/ilink/en/download/HW_082746 → pages 86

[99] Flexenclosure E-site Case Study, *Available:* http://www.flexenclosure.com/Esite.aspx → pages 86

[100] J. T Louhi, “Energy efficiency of modern cellular base stations,” *29th International Telecommunications Energy Conference (INTELEC), 2007*, pp.475-476, Sept. 30 2007-Oct. 4 2007. → pages 86, 88, 91

[101] Zhisheng Niu, Yiqun Wu, Jie Gong, and Zexi Yang, “Cell zooming for cost-efficient green cellular networks,” *IEEE Communications Magazine*, vol.48, no.11, pp.74-79, November 2010. → pages 86, 89, 90

[102] M. A Marsan, and M. Meo, “Energy efficient wireless Internet access with cooperative cellular networks,” *Comput. Netw.* (2010), → pages 86, 89

[103] Ashwin Amanna, “Green Communications: Annotated Review and Research Vision”, *Virginia Tech.*. → pages 87, 88

[104] H. Claussen, L. T. W Ho, and F. Pivit, “Effects of joint macrocell and residential picocell deployment on the network energy efficiency,” *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2008*, pp.1-6, 15-18 Sept. 2008. → pages 87, 91, 93

[105] James Sean Kim et.al., “Advanced Power Amplifier Design Using Doherty Configurations”, *Global Seminars*, Ansoft Corporation. → pages 87

[106] M. J. Chang, Z. Abichar, and Chau-Yun Hsu, “WiMAX or LTE: Who will Lead the Broadband Mobile Internet?,” *IT Professional* , vol.12, no.3, pp.26-32, May-June 2010. → pages 88

[107] 3G Americas, “The benefits of son in LTE: Self-optimizing and self-organizing networks”, *White Paper*, Dec 2009. → pages 89

[108] L.C. Schmelz, et. al, “Self-organisation in Wireless Networks Use Cases and their Interrelation”, *22nd WWRF*, May 2009. → pages 89

[109] M. A Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo “Optimal Energy Savings in Cellular Access Networks,” *IEEE International Conference on Communications Workshops, 2009*, pp.1-5, 14-18 June 2009. → pages 89

[110] I. Viering, et. al., “A Distributed Power Saving Algorithm for Cellular Networks”, *Proc. of IWSOS*, Dec 2009. → pages 89

[111] K. Samdanis, D. Kutscher, and M. Brunner, “Self-organized energy efficient cellular networks,” *Proc. of IEEE PIMRC’10*, pp. 1665-1670, 2010. → pages 89

[112] K. Samdanis, D. Kutscher, and M. Brunner, “Dynamic energy-aware network re-configuration for cellular urban infrastructures,” *IEEE GLOBECOM GreenComm3 Workshops 2010*, pp. 1448-1452, 2010. → pages 89

[113] S. Bhaumik, G. Narlikar, and S. Chattopadhyay, “Breathe to stay cool: adjusting cell sizes to reduce energy consumption,” *Proc. of the First ACM SIGCOMM Workshop on Green Networking*, Aug-Sep. 2010. → pages 89, 90

[114] “Green Power for Mobile ”, http://www.gsmworld.com/our-work/mobile_planet/green_power_for_mobile/ → pages 91

[115] “Bi-annual Report November 2010”, Green Power for Mobile, GSMA. *Available:* http://www.gsm.org/documents/GPM_Bi-Annual_Report_June_10.pdf → pages 91

[116] Vandana Bassoo, Kevin Tom, A. K. Mustafa, Ellie Cijvat, Henrik Sjoland, and Mike Faulkner, “A Potential Transmitter Architecture for Future Generation Green Wireless Base Station,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 821846, 8 pages, 2009. → pages 91

[117] “Understanding the Environmental Impact of Communication Systems”, Plextek, Final Report, 27 April 2009. *Available:* http://stakeholders.ofcom.org.uk/binaries/research/technology-research/environ.pdf. → pages 93

[118] D. Calin, H. Claussen, and H. Uzunalioglu, “On femto deployment architectures and macrocell offloading benefits in joint macro-femto deployments,” *IEEE Communications Magazine*, vol. 48, no. 1, pp. 26-32, 2010. → pages 93

[119] F. Richter, A. J. Fehske, and G. P Fettweis, “Energy Efficiency Aspects of Base Station Deployment Strategies for Cellular Networks,” *2009 IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*, pp.1-5, 20-23 Sept. 2009. → pages 93

[120] A. J Fehske, F. Richter, and G. P Fettweis, “Energy Efficiency Improvements through Micro Sites in Cellular Mobile Radio Networks,” *2nd International Workshop on Green Communications, GLOBECOM 2009*, pp.1-5, Nov. 30 2009-Dec. 4 2009. → pages 93

[121] I. Ashraf, L.T.W. Ho, H. Claussen, "Improving Energy Efficiency of Femtocell Base Stations Via User Activity Detection," *Proc. of IEEE WCNC'10*, pp.1-5, 18-21 April 2010. → pages 93

[122] J. Hoydis, M. Kobayashi, and M. Debbah, "Green Small-Cell Networks," *IEEE Vehicular Technology Magazine*, vol. 6, no. 1, pp. 37-43, 2011. → pages 93

[123] S. Haykins, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Commun.*, vol. 23, no. 2, pp. 201-220, Feb. 2005. → pages 94

[124] C. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, pp. 10-21, January 1949. → pages 94

[125] D. Grace, Jingxin Chen, Tao Jiang, and P. D. Mitchell, "Using cognitive radio to deliver Green communications," *CROWNCOM '09*, pp.1-6, 22-24 June 2009. → pages 94

[126] O. Holland, V. Friderikos, and A.H. Aghvami, "Green Spectrum Management for Mobile Operators", *IEEE GLOBECOM GreenComm3 Workshops 2010*, pp. 1458-1463, 2010. → pages 94

[127] J. Mitola III, and G. Q. Maguire, Jr., "Cognitive radio: making software radios more personal," *IEEE Personal Communications Magazine*, vol. 6, no. 4, pp. 13-18, Aug. 1999. → pages 94

[128] An He, et. al., "System power consumption minimization for multichannel communications using cognitive radio," *IEEE International Conference on Microwaves, Communications, Antennas and Electronics Systems (COMCAS), 2009*, pp.1-5, 9-11 Nov. 2009. → pages 94

[129] An He, et. al., "Minimizing Energy Consumption Using Cognitive Radio," *IEEE International Performance, Computing and Communications Conference (IPCCC), 2008*, pp.372-377, 7-9 Dec. 2008. → pages 94

[130] R. Pabst, et. al, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Communications Magazine*, vol. 42, no. 9, pp. 80-89, Sept. 2004. → pages 95, 96

[131] J. N Laneman, and G. W Wornell, "Energy-efficient antenna sharing and relaying for wireless networks," *IEEE Wireless Communications and Networking Conference (WCNC), 2000*, pp.7-12 vol.1, 2000. → pages 95

[132] X. J. Li, B. C. Seet, and P. H. J. Chong, "Multihop cellular networks: Technology and economics," *Computer Networks*, vol. 52, pp. 1825-1837, June 2008. → pages 95

[133] J. Y. Song, H. Lee, and D. H. Cho, "Power Consumption Reduction by Multi-hop Transmission in Cellular Networks," *Proc. IEEE Vehicular Technology Conf.*, vol. 5, pp. 3120-3124, Sept. 2004. → pages 95

[134] A. Radwan, H. S Hassanein, "NXG04-3: Does Multi-hop Communication Extend the Battery Life of Mobile Terminals?," *IEEE Global Telecommunications Conference (GLOBECOM), 2006*, pp.1-5, Nov. 27 2006-Dec. 1 2006. → pages 95

[135] R. Rost, and G. Fettweis, “Green communication in cellular network with fixed relay nodes,” *Cooperative Cellular Wireless Communications* to be published by Cambridge university press, Sept. 2010. → pages 96

[136] M. Nokleby, and B. Aazhang, “User Cooperation for Energy-Efficient Cellular Communications,” *IEEE International Conference on Communications (ICC) 2010*, pp.1-5, 23-27 May 2010. → pages 96

[137] G. Gur and F. Alagoz, “Green wireless communications via cognitive dimension: an overview,” *IEEE Network Magazine*, vol. 25, no. 2, pp. 50-56, 2011. → pages 97

[138] R. Tandra, A. Sahai, and S. M. Mishra, “What is a spectrum hole and what does it take to recognize one?,” *Proc. of the IEEE, vol. 97*, pp. 824-848, Apr. 2009. → pages 98

[139] J. Wei, and X. Zhang, “Energy-Efficient Distributed Spectrum Sensing for Wireless Cognitive Radio Networks,” *Proc. of the IEEE INFOCOM'10*, pp. 1-6, Mar 2010. → pages 98

[140] S. J. Kim, and G. B. Giannakis, “Rate-Optimal and Reduced-Complexity Sequential Sensing Algorithms for Cognitive OFDM Radios,” *Proc. of the IEEE CISS'09*, pp. 141-146, Mar. 2009. → pages 98

[141] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, “Compressive Wide-band Spectrum Sensing,” *Proc. of the IEEE ICASSP'09*, pp. 2337-2340, Apr. 2009. → pages 98

[142] C. Cormio, and K. R. Chowdhury, “A survey on MAC protocols for cognitive radio networks,” *Ad Hoc Networks*, vol. 7, pp. 1315-1329, Sep. 2009. → pages 98

[143] H. Adam, W. Elmenreich, C. Bettstetter, and S. M. Senouci, “CoRe-MAC: a MAC-protocol for cooperative relaying in wireless networks,” *Proc. IEEE Globecom'09*, pp. 1-6, Dec. 2009. → pages 98

[144] L. Ding, T. Melodia, S. N. Batalama, and J. D. Matyjas, ”Distributed Routing, Relay Selection, and Spectrum Allocation in Cognitive and Cooperative Ad Hoc Networks,” *Proc. of IEEE SECON'10*, pp.1-9, 2010 → pages 99

[145] J. Alonso-Zarate, E. Kartsakli, C. Verikoukis, and L. Alonso, “Persistent RCSMA: A MAC Protocol for a Distributed Cooperative ARQ Scheme in Wireless Networks,” *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1-14, 2008. → pages 99

[146] J. Alonso-Zarate, et. al.,“Energy-Efficiency Evaluation of a Medium Access Control Protocol for Cooperative ARQ”, *IEEE International Communications Conference'11*, June 2011. → pages 99

[147] B. Yahya, and J. B. Othman, “Energy efficient and QoS aware medium access control for wireless sensor networks,”, *Concurrency and Computation: Practice and Experience*, vol. 22, Jul. 2010. → pages 99

[148] S. C. Ergen, and P. Varaiya, “Energy efficient routing with delay guarantee for sensor networks,” *Wireless Networks*, vol. 13, pp. 679-690, Oct. 2007. → pages 99

[149] I. Stanojev, O. Simeone, Y. Bar-Ness, and D. H. Kim, “Energy efficiency of non-collaborative and collaborative Hybrid-ARQ protocols,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 1, pp. 326-335, 2009. → pages 99

[150] *RPL: IPv6 Routing Protocol for Low power and Lossy Networks*, ROLL IETF Internet-Draft, March 2011, draft-ietf-roll-rpl-19 [work in progress]. → pages 100

[151] T. Watteyne, A. Molinaro, M. G. Richichi, and M. Dohler, “From MANET To IETF ROLL Standardization: A Paradigm Shift in WSN Routing Protocols,” *Accepted to Publication in IEEE Communications Surveys & Tutorials*. → pages 100

[152] A. Alshamrani, X.S. Shen, and Liang-Liang Xie, “QoS Provisioning for Heterogeneous Services in Cooperative Cognitive Radio Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 819-830, April 2011. → pages 100

[153] Z. Hasan, G. Bansal, E. Hossain and V. Bhargava, “Energy-efficient power allocation in OFDM- based cognitive radio systems: A risk-return model,” *IEEE Trans. Wireless Commun.*, vol.8, pp. 6078-6088, Dec. 2009. → pages 100

[154] V. A. Le, R. A. Pitaval, S. Blostein, T. Riihonen, and R. Wichman, “Green cooperative communication using threshold-based relay selection protocols,” *Proc. of IEEE ICGCS’10*, pp.521-526, Jun. 2010. → pages 100

[155] Y. Wei, F. R. Yu, M. Song, and V.C.M. Leung, “Energy Efficient Distributed Relay Selection in Wireless Cooperative Networks with Finite State Markov Channels,” *IEEE GLOBECOM’09*, pp.1-6, Nov-Dec. 2009. → pages 101

[156] Y. Wei, F.R. Yu, and M. Song, “Distributed Optimal Relay Selection in Wireless Cooperative Networks with Finite State Markov Channels,” *IEEE Trans. Vehicular Technology*, vol. 59, no. 5, pp. 2149-2158, June 2010. → pages 101

[157] O. Duval, Z. Hasan, E. Hossain, F. Gagnon, and V. K. Bhargava, “Subcarrier selection and power allocation for amplify-and-forward relaying over OFDM links,” *IEEE Trans. on Wireless Commun.*, vol.9, no.4, pp.1293-1297, April 2010. → pages 101

[158] Z. Hasan, E. Hossain, and V. K. Bhargava, “Resource Allocation for Multiuser OFDMA-based Amplify-and-Forward Relay Networks with Selective Relaying”, *IEEE International Communications Conference’11*, June 2011. → pages 101

[159] Song Gao, Lijun Qian, and D. Vaman, “Distributed energy efficient spectrum access in cognitive radio wireless ad hoc networks,” *IEEE Trans. on Wireless Commun.*, vol. 8, no. 10, pp. 5202-5213, 2009. → pages 101

[160] H. Ding, J. Ge, D. B. da Costa, and Z. Jiang, “Energy-Efficient and Low-Complexity Schemes for Uplink Cognitive Cellular Networks,” *IEEE Communications Letters*, vol. 14, no. 12, pp. 1101-1103, 2010. → pages 101

[161] Z. Bharucha, A. Saul, G. Auer, and H. Haas, "Dynamic Resource Partitioning for Downlink Femto-to-Macro-Cell Interference Avoidance," *EURASIP J. Wireless Commun. and Networking*, vol. 2010, Article ID 143413, 2010. → pages 101

[162] A. J. Goldsmith, and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Wireless Communications*, vol.9, no.4, pp. 8-27, Aug. 2002. → pages 102

[163] A. Dejonghe, et. al., "Green Reconfigurable Radio Systems," *IEEE Signal Processing Magazine*, vol.24, no.3, pp.90-101, May 2007. → pages 102

[164] A. Vosoughi, and Y. Jia, "Maximizing Throughput in Cooperative Networks via Cross-layer Adaptive Designs," *Proc. of IEEE Sarnoff Symposium'10*, pp. 1-6, Apr. 2010. → pages 102

[165] C. Luo, F. R. Yu, H. Ji, and V.C.M. Leung, "Cross-Layer Design for TCP Performance Improvement in Cognitive Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp.2485-2495, Jun 2010. → pages 102

[166] Q. Zhang, J. Jia, and J. Zhang, "Cooperative Relay to Improve Diversity in Cognitive Radio Networks," *IEEE Commun. Mag.*, vol. 47, pp. 111-117, Feb. 2009. → pages 103

[167] Y. Wang, C. Feng, Z. Zeng, and C. Guo, "A robust and energy efficient cooperative spectrum sensing scheme in cognitive radio networks," *Proc. IEEE ICACT'09*, vol. 1, pp. 640-645, Feb. 2009. → pages 103

[168] L. Ruan and V. K. N. Lau, "Power Control and Performance Analysis of Cognitive Radio Systems under Dynamic Spectrum Activity and Imperfect Knowledge of System State," *IEEE Trans Wireless Commun.*, vol. 8, pp. 4616-4622, Sep. 2009. → pages 103

[169] A. Ghasemi, and E. S. Sousa, "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs," *IEEE Commun. Mag.*, vol. 46, pp. 32-39, Apr. 2008. → pages 103

[170] G. Anastasi, M. Conti, E. Gregori, and A. Passarella. "Performance comparison of power saving strategies for mobile web access". *Performance Evaluation*, vol. 53, 273-294, 2003. → pages 103

[171] K. Dufkova, M. Bjelica, B. Moon, L. Kencl, and J-Y. Le Boudec, "Energy Savings for Cellular Network with Evaluation of Impact on Data Traffic Performance", *Proc. of European Wireless Conference*, Apr. 2010. → pages 104

[172] Y. Kim, M. Thottan, V. Kolesnikov, and W. Lee, "A secure decentralized data-centric information infrastructure for smart grid," *IEEE Commun. Mag.*, vol. 48, Nov. 2010. → pages 104

[173] I. Humar, X. Ge, L. Xiang, M. Jo, M. Chen, and J. Zhang, "Rethinking energy efficiency models of cellular networks with embodied energy," *IEEE Network Magazine*, vol. 25, no. 2, pp. 40-49, 2011. → pages 104, 106

[174] Y. Yao and G. B. Giannakis, "Rate-maximizing power allocation in OFDM based on partial channel knowledge," *IEEE Trans. on Wireless Comm.*, vol. 4, no. 3, pp. 1073-1083, 2005. → pages 109

[175] Wenbing Dang, Meixia Tao, Hua Mu, and Jianwei Huang, "Subcarrier-Pair Based Resource Allocation for Cooperative Multi-Relay OFDM systems", in *IEEE Transactions on Wireless Commnunications*, vol. 9, no. 5, pp. 1640-1649, May 2010. → pages 110

Appendix A

Proof of Proposition 1

We can rewrite the optimization problem as follows:

$$\mathbf{P}^* = \arg\min_{\mathbf{P}} \sum_{i=1}^{N} \left[\alpha_{\varphi(i)} \mathcal{C} P_i - \ln \left(1 + \frac{|h_i|^2 P_i}{N_0 B + J_i} \right) \right] \tag{A.1}$$

subject to

$$P_i \geq 0, \quad \forall i \in \{1, 2, \cdots, N\} \tag{A.2}$$

$$\sum_{i=1}^{N} P_i \leq P_{total} \tag{A.3}$$

$$\sum_{i=1}^{N} K_i P_i \leq I_{th} \tag{A.4}$$

and

$$S_j \leq T_j, \quad \forall j \in \{1, 2, \cdots, M\} \tag{A.5}$$

where S_j is given by

$$S_j = \sum_{i \text{ s.t. } \phi(i)=j} P_i. \tag{A.6}$$

We now use convex optimization theory [61] to solve this optimization problem. Using Lagrange multipliers μ_i for the inequality constraints in (A.2), λ for the inequality constraint in (A.3), δ for the inequality constraint in (A.4) and γ_j for the inequality constraint in (A.5), the Karush-Kuhn-Tucker (KKT) conditions will include equations (A.2), (A.3), (A.4), (A.5) with optimum solution P_i^* and the following equations:

$$\mu_i \geq 0, \qquad \forall i \in \{1, 2, \cdots, N\} \tag{A.7}$$

$$\lambda \geq 0 \tag{A.8}$$

$$\delta \geq 0 \tag{A.9}$$

$$\gamma_j \geq 0, \qquad \forall j \in \{1, 2, \cdots, M\} \tag{A.10}$$

$$\mu_i P_i^* = 0, \qquad \forall i \in \{1, 2, \cdots, N\} \tag{A.11}$$

$$\lambda(\sum_{i=1}^{N} P_i^* - P_{total}) = 0 \tag{A.12}$$

$$\delta(\sum_{i=1}^{N} K_i P_i^* - I_{th}) = 0 \tag{A.13}$$

$$\gamma_j(S_j - T_j) = 0, \qquad \forall j \in \{1, 2, \cdots, M\} \tag{A.14}$$

and

$$-\frac{1}{\frac{N_0 B + J_i}{|h_i|^2} + P_i^*} + \alpha_{\varphi(i)}\mathcal{C} - \mu_i + \lambda + \delta K_i + \gamma_{\varphi(i)} = 0, \qquad \forall i \in \{1, 2, \cdots, N\}. \tag{A.15}$$

We can rewrite (A.15) as follows:

$$\mu_i = \lambda + \delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C} - \frac{1}{\frac{N_0 B + J_i}{|h_i|^2} + P_i^*}. \tag{A.16}$$

Substituting (A.16) into (A.7) and (A.11), we can eliminate μ_i, which yields

$$\lambda + \delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C} \geq \frac{1}{\frac{N_0 B + J_i}{|h_i|^2} + P_i^*}, \qquad \forall i \in \{1, 2, \cdots, N\} \tag{A.17}$$

$$\left(\lambda + \delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C} - \frac{1}{\frac{N_0 B + J_i}{|h_i|^2} + P_i^*}\right) P_i^* = 0, \qquad \forall i \in \{1, 2, \cdots, N\}. \tag{A.18}$$

Now if $\lambda + \delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C} < \frac{|h_i|^2}{N_0 B + J_i}$, then (A.17) can hold only if $P_i^* > 0$, which from equation (A.18) gives,

$$P_i^* = \frac{1}{\lambda + \delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C}} - \frac{N_0 B + J_i}{|h_i|^2}. \tag{A.19}$$

On the other hand, if $\lambda + \delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C} \geq \frac{|h_i|^2}{N_0 B + J_i}$ then $P_i^* > 0$ is not possible as it would violate (A.18). So $P_i^* = 0$ is the only solution in this case. Combining these results, the solution can be rewritten as follows:

$$P_i^* = \begin{cases} \frac{1}{\lambda+\delta K_i+\gamma_{\varphi(i)}+\alpha_{\varphi(i)}\mathcal{C}} - \frac{N_0 B + J_i}{|h_i|^2}, \\ \qquad \text{if } \lambda + \delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C} < \frac{|h_i|^2}{N_0 B + J_i} \\ 0, \\ \qquad \text{if } \lambda + \delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C} \geq \frac{|h_i|^2}{N_0 B + J_i} \end{cases} \tag{A.20}$$

which is equivalent to

$$P_i^* = \left[w_i - \frac{N_0 B + J_i}{|h_i|^2} \right]^+, \forall i \in \{1, 2, \cdots, N\} \tag{A.21}$$

where $w_i = \frac{1}{\lambda+\delta K_i+\gamma_{\varphi(i)}+\alpha_{\varphi(i)}\mathcal{C}}$ is defined as subcarrier threshold level. Next, in order to determine the values of λ, δ, and γ_j we first divide the sub-bands in two sets A and B such that if (i) $S_j < T_j$, then $j \in A$ and if (ii) $S_j = T_j$, then $j \in B$. Depending upon the constraints, either of these sets could be empty.

Now for any $j \in A$, since $S_j < T_j$, (A.14) suggests $\gamma_j = 0$. Similarly, for $j \in B$, since $S_j = T_j$, (A.10) and (A.14) lead to $\gamma_j \geq 0$. We observe from (A.12) and (A.13), any value $\lambda > 0$ or $\delta > 0$ demands the equality sign for the constraints in (A.3) and (A.4), respectively. Also, we notice that each P_i^* is maximum when either $\lambda = 0$ or $\delta = 0$ compared to the case when $\lambda > 0$, or $\delta > 0$, respectively. Therefore, if we are unable to satisfy equality for the constraints in (A.3) and (A.4) for any positive values for λ or δ, then we should set largest possible P_i^* (larger P_i^* makes each inequality tighter, since $K_i > 0$) which suggests either $\lambda = 0$ or $\delta = 0$ or both depending on the constraint equations. This leads to the following four possible cases:

Case 1: $\lambda = 0$ and $\delta = 0$

Let us assign

$$w_i = \frac{1}{\alpha_{\varphi(i)}\mathcal{C}}, \tag{A.22}$$

$$\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in A$$

$$w_i = \frac{1}{\gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C}}, \tag{A.23}$$

$$\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in B$$

where $\gamma_j \geq 0, \forall j \in B$ is determined by the following $|B|$ equations:

$$\sum_{i \text{ s.t } \phi(i)=j} P_i^* = T_j, \quad \forall j \in B \tag{A.24}$$

If the solution to the above equations $P_i^* \geq 0$ exists and satisfies $\sum_{i=1}^{N} P_i^* \leq P_{total}$ and $\sum_{i=1}^{N} K_i P_i^* \leq I_{th}$, then this suggests $\lambda = 0$, $\delta = 0$ and P_i^* is the optimal solution.

Case 2: $\lambda > 0$ and $\delta = 0$

Let us assign

$$w_i = \frac{1}{\lambda + \alpha_{\varphi(i)}\mathcal{C}}, \tag{A.25}$$

$$\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in A$$

$$w_i = \frac{1}{\lambda + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C}}, \tag{A.26}$$

$$\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in B$$

where $\lambda > 0$ and $\gamma_j \geq 0, \forall j \in B$ are determined by the following $|B| + 1$ equations:

$$\sum_{i=1}^{N} P_i^* = P_{total} \tag{A.27}$$

$$\sum_{i \text{ s.t } \phi(i)=j} P_i^* = T_j, \quad \forall j \in B. \tag{A.28}$$

If the solution to the above equations $P_i^* \geq 0$ exists and satisfies $\sum_{i=1}^{N} K_i P_i^* \leq I_{th}$, then this suggests $\delta = 0$ and P_i^* is the optimal solution.

Case 3: $\lambda = 0$ and $\delta > 0$

Let us assign

$$
\begin{aligned}
w_i &= \frac{1}{\delta K_i + \alpha_{\varphi(i)}\mathcal{C}}, && \text{(A.29)}\\
&\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in A\\
w_i &= \frac{1}{\delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C}}, && \text{(A.30)}\\
&\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in B
\end{aligned}
$$

where $\delta > 0$ and $\gamma_j \geq 0, \forall j \in B$ are determined by the following $|B| + 1$ equations

$$\sum_{i=1}^{N} K_i P_i^* = I_{th} \tag{A.31}$$

$$\sum_{i \text{ s.t } \phi(i)=j} P_i^* = T_j, \quad \forall j \in B. \tag{A.32}$$

If the solution to the above equations $P_i^* \geq 0$ exists and satisfies $\sum_{i=1}^{N} P_i^* \leq P_{total}$, then this suggests $\lambda = 0$, and P_i^* is the optimal solution.

Case 4: $\lambda > 0$ and $\delta > 0$

Let us assign

$$
\begin{aligned}
w_i &= \frac{1}{\lambda + \delta K_i + \alpha_{\varphi(i)}\mathcal{C}}, && \text{(A.33)}\\
&\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in A\\
w_i &= \frac{1}{\lambda + \delta K_i + \gamma_{\varphi(i)} + \alpha_{\varphi(i)}\mathcal{C}}, && \text{(A.34)}\\
&\forall i \in \{1, 2, \cdots, N\} \text{ such that } \varphi(i) \in B
\end{aligned}
$$

where $\lambda > 0$, $\delta > 0$ and $\gamma_j \geq 0, \forall j \in B$ are determined by the following $|B| + 2$ equations:

$$\sum_{i=1}^{N} P_i^* = P_{total} \tag{A.35}$$

$$\sum_{i=1}^{N} K_i P_i^* = I_{th} \tag{A.36}$$

$$\sum_{i \text{ s.t. } \phi(i)=j} P_i^* = T_j, \quad \forall j \in B. \tag{A.37}$$

If the solution to above equations $P_i^* \geq 0$ exists then that is the optimal solution.

Assuming P_{total}, I_{th} and T_j to be positive, if there is a solution which falls in the category of any of the four cases, it would satisfy all the KKT conditions. Due to strict concavity of the objective function there will be a unique solution. This completes the proof of Proposition 1.

Appendix B

Proof of Proposition 2

Lets assume the marginal distribution of type θ_n to be $f_n(\theta_n)$ and the corresponding cumulative distribution to be $F_n(\theta_n)$ (s.t. $F_n'(\theta) = f_n(\theta)$), where $f_n(\theta_n)$ is given by

$$f_n(\theta_n) = \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\bar{\theta}} \cdots \int_{\underline{\theta}}^{\bar{\theta}} f(\theta_1, \theta_2, \cdots, \theta_N) d\theta_1 d\theta_2 \cdots \theta_{n-1} \theta_{n+1} \cdots d\theta_N. \tag{B.1}$$

Using the expression (B.1), we can simplify (4.15) and subsequently substitute (4.14) as follows,

$$\begin{aligned}
& \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\bar{\theta}} \cdots \int_{\underline{\theta}}^{\bar{\theta}} \sum_{n=1}^{N} (U(\gamma(\theta_n)) - t(\theta_n)) f(\theta_1, \theta_2, \cdots, \theta_N) d\theta_1 d\theta_2 \cdots d\theta_N \\
= \;& \sum_{n=1}^{N} \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\bar{\theta}} \cdots \int_{\underline{\theta}}^{\bar{\theta}} (U(\gamma(\theta_n)) - t(\theta_n)) f(\theta_1, \theta_2, \cdots, \theta_N) d\theta_1 d\theta_2 \cdots d\theta_N \\
= \;& \sum_{n=1}^{N} \int_{\underline{\theta}}^{\bar{\theta}} (U(\gamma(\theta_n)) - t(\theta_n)) f_n(\theta_n) d\theta_n = \sum_{n=1}^{N} \int_{\underline{\theta}}^{\bar{\theta}} (U(\gamma(\theta)) - t(\theta)) f_n(\theta) d\theta \\
= \;& \sum_{n=1}^{N} \int_{\underline{\theta}}^{\bar{\theta}} \left(U(\gamma(\theta)) - \frac{c\gamma(\theta)}{\theta} - \int_{\underline{\theta}}^{\theta} \frac{c\gamma(\tau)}{\tau^2} d\tau \right) f_n(\theta) d\theta
\end{aligned} \tag{B.2}$$

We can use integration by parts to rewrite the last term of (B.2) as follows

$$\begin{aligned}
\int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} \frac{c\gamma(\tau)}{\tau^2} f_n(\theta) d\tau d\theta &= \int_{\underline{\theta}}^{\bar{\theta}} F_n'(\theta) \int_{\underline{\theta}}^{\theta} \frac{c\gamma(\tau)}{\tau^2} d\tau d\theta \\
&= F_n(\theta) \int_{\underline{\theta}}^{\theta} \frac{c\gamma(\tau)}{\tau^2} d\tau \bigg|_{\underline{\theta}}^{\bar{\theta}} - \int_{\underline{\theta}}^{\bar{\theta}} \frac{c\gamma(\theta)}{\theta^2} F_n(\theta) d\theta \\
&= \int_{\underline{\theta}}^{\bar{\theta}} \frac{c\gamma(\theta)}{\theta^2} (1 - F_n(\theta)) d\theta \\
&= \int_{\underline{\theta}}^{\bar{\theta}} \frac{c\gamma(\theta)}{\theta^2} \frac{1 - F_n(\theta)}{f_n(\theta)} f_n(\theta) d\theta \qquad \text{(B.3)}
\end{aligned}$$

Substituting (B.3) in (B.2), we can reduce the corresponding optimization problem to (4.16). Q.E.D.

Appendix C

Proof of Theorem 2

Let us consider two relay-agent types δ_k and δ_j such that $\delta_k > \delta_j$. The mutual IC conditions for these relay types are

$$t_k - \frac{c\gamma_k}{\delta_k} \geq t_j - \frac{c\gamma_j}{\delta_k}, \tag{C.1}$$
$$t_j - \frac{c\gamma_j}{\delta_j} \geq t_k - \frac{c\gamma_k}{\delta_j}. \tag{C.2}$$

Adding these two IC conditions gives

$$\frac{c\gamma_k}{\delta_k} + \frac{c\gamma_j}{\delta_j} \leq \frac{c\gamma_j}{\delta_k} + \frac{c\gamma_k}{\delta_j}$$
$$\text{or, } \delta_j(\gamma_k - \gamma_j) \leq \delta_k(\gamma_k - \gamma_j). \tag{C.3}$$

Since $\delta_k > \delta_j$, we can say that $\gamma_k \geq \gamma_j$. And because γ's are all assumed to be positive, we have

$$0 \leq \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_K. \tag{C.4}$$

So, we proved the first result and now we must also prove that for the optimal solution, ICs reduce to (4.21) when compounded with (C.4). This proof is a slightly more intricate but we will first prove that when the downward adjacent ICs are binding, i.e., (4.21) is true, all the other ICs are automatically satisfied (sufficiency) and we will then prove that when one or more downward adjacent ICs is not binding, the contract cannot be optimal and a better contract can be obtained by binding the downward adjacent ICs recursively by reducing the transfers, hence increasing the source's utility (necessity).

C.1 Proof of Sufficiency

Suppose the adjacent downward ICs are binding, then we can rewrite (4.21) as follows:

$$(t_k - t_{k-1}) = \frac{c(\gamma_k - \gamma_{k-1})}{\delta_k}, \quad \forall k \geq 2. \tag{C.5}$$

Lets consider some j such that $j > k$, then using (C.4) and (C.5), we can write:

$$\begin{aligned}
(t_j - t_k) &= (t_j - t_{j-1}) + (t_{j-1} - t_{j-2}) + \cdots + (t_{k+1} - t_k) \\
&= \frac{c(\gamma_j - \gamma_{j-1})}{\delta_j} + \frac{c(\gamma_{j-1} - \gamma_{j-2})}{\delta_{j-1}} + \cdots \frac{c(\gamma_{k+1} - \gamma_k)}{\delta_{k+1}} \\
&\leq \frac{c(\gamma_j - \gamma_{j-1})}{\delta_k} + \frac{c(\gamma_{j-1} - \gamma_{j-2})}{\delta_k} + \cdots \frac{c(\gamma_{k+1} - \gamma_k)}{\delta_k} \\
&= \frac{c(\gamma_j - \gamma_k)}{\delta_k}.
\end{aligned} \tag{C.6}$$

Now, lets take some j such that $j < k$, then using (C.4) and (C.5), we can write:

$$\begin{aligned}
(t_k - t_j) &= (t_k - t_{k-1}) + (t_{k-1} - t_{k-2}) + \cdots + (t_{j+1} - t_j) \\
&= \frac{c(\gamma_k - \gamma_{k-1})}{\delta_k} + \frac{c(\gamma_{k-1} - \gamma_{k-2})}{\delta_{k-1}} + \cdots \frac{c(\gamma_{j+1} - \gamma_j)}{\delta_{j+1}} \\
&\geq \frac{c(\gamma_k - \gamma_{k-1})}{\delta_k} + \frac{c(\gamma_{k-1} - \gamma_{k-2})}{\delta_k} + \cdots \frac{c(\gamma_{j+1} - \gamma_j)}{\delta_k} \\
&= \frac{c(\gamma_k - \gamma_j)}{\delta_k}.
\end{aligned} \tag{C.7}$$

Combining equations (C.6) and (C.7), we get:

$$(t_k - t_j) \geq \frac{c(\gamma_k - \gamma_j)}{\delta_k} \text{ or } t_k - \frac{c\gamma_k}{\delta_k} \geq t_j - \frac{c\gamma_j}{\delta_k}, \quad \forall j \neq k. \tag{C.8}$$

Hence all the ICs are automatically satisfied when adjacent downward ICs are binding.

C.2 Proof of Necessity

Suppose one or more downward adjacent ICs are not binding for optimal solution (and IRs are satisfied because it is a solution). Lets take one such type δ_k for which adjacent downward IC is inactive, so using IR for type δ_{k-1} we have:

$$t_k - \frac{c\gamma_k}{\delta_k} > t_{k-1} - \frac{c\gamma_{k-1}}{\delta_k} \geq t_{k-1} - \frac{c\gamma_{k-1}}{\delta_{k-1}} \geq 0 \quad \forall j \geq k, j \leq k \tag{C.9}$$

What this means is that if we reduce all t_j's, $\forall j \geq k$ with equal amount through which the adjacent downward IC for type δ_k becomes active, it will not affect any of the IRs and the existing relation (whether binding or not binding) between adjacent downward ICs for every other type. We can iteratively repeat the process (starting with the lowest type for which adjacent downward IC is inactive) till all the adjacent downward ICs are binding. In this process, we have only reduced the transfers to bind all the downward adjacent ICs, which in turn automatically guarantees that all the other ICs are satisfied from the sufficiency conditions proved in Part 1. This therefore increases source's profit because transfers are expensive for the source. Hence, the original contract cannot be optimal because we found a better contract, which is a contradiction.

Appendix D

Proof of Proposition 3

Lets look at the second term $\sum_{n=1}^{N}\sum_{k=1}^{K}\pi_{kn}t_k$ of the objective of optimization problem in (4.22). Substituting the equality constraints from (4.22), we write the individual terms of the summation as follows:

$$\begin{aligned}
\sum_{n=1}^{N}\pi_{1n}t_1 &= \sum_{n=1}^{N}\pi_{1n}\left(\frac{c\gamma_1}{\delta_1}\right) \\
\sum_{n=1}^{N}\pi_{2n}t_2 &= \sum_{n=1}^{N}\pi_{2n}\left(\frac{c\gamma_1}{\delta_1}+\frac{c\gamma_2}{\delta_2}-\frac{c\gamma_1}{\delta_2}\right) \\
\sum_{n=1}^{N}\pi_{3n}t_3 &= \sum_{n=1}^{N}\pi_{3n}\left(\frac{c\gamma_1}{\delta_1}+\frac{c\gamma_2}{\delta_2}-\frac{c\gamma_1}{\delta_2}+\frac{c\gamma_3}{\delta_3}-\frac{c\gamma_2}{\delta_3}\right) \\
&\vdots \\
\sum_{n=1}^{N}\pi_{Kn}t_K &= \sum_{n=1}^{N}\pi_{Kn}\left(\frac{c\gamma_1}{\delta_1}+\frac{c\gamma_2}{\delta_2}-\frac{c\gamma_1}{\delta_2}+\frac{c\gamma_3}{\delta_3}-\frac{c\gamma_2}{\delta_3}\cdots+\frac{c\gamma_K}{\delta_K}-\frac{c\gamma_{K-1}}{\delta_K}\right).
\end{aligned}$$

Now adding all the equations above, collecting the terms for each γ_k, and using the fact that $\sum_{i=1}^{K} \pi_{in} = 1$ we can derive the following :

$$\begin{aligned}
\sum_{n=1}^{N}\sum_{k=1}^{K} \pi_{kn} t_k &= c\gamma_1 \left(\frac{1}{\delta_1} \sum_{n=1}^{N}\sum_{i=1}^{K} \pi_{in} - \frac{1}{\delta_2} \sum_{n=1}^{N}\sum_{i=2}^{K} \pi_{in} \right) + c\gamma_2 \left(\frac{1}{\delta_2} \sum_{n=1}^{N}\sum_{i=2}^{K} \pi_{in} - \frac{1}{\delta_3} \sum_{n=1}^{N}\sum_{i=3}^{K} \pi_{in} \right) \\
&\quad + \cdots c\gamma_k \left(\frac{1}{\delta_k} \sum_{n=1}^{N}\sum_{i=k}^{K} \pi_{in} - \frac{1}{\delta_{k+1}} \sum_{n=1}^{N}\sum_{i=k+1}^{K} \pi_{in} \right) + \cdots + \sum_{n=1}^{N} \pi_{Kn} \frac{c\gamma_K}{\delta_K} \\
&= \sum_{k=1}^{K-1} c\gamma_k \left(\frac{1}{\delta_k} \sum_{n=1}^{N}\sum_{i=k}^{K} \pi_{in} - \frac{1}{\delta_{k+1}} \sum_{n=1}^{N}\sum_{i=k+1}^{K} \pi_{in} \right) + \sum_{n=1}^{N} \pi_{Kn} \frac{c\gamma_K}{\delta_K} \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K-1} \pi_{kn} c\gamma_k \left(\frac{1}{\delta_k} + \left(\frac{1 - \sum_{i=1}^{k} \pi_{in}}{\pi_{kn}} \right) \left(\frac{1}{\delta_k} - \frac{1}{\delta_{k+1}} \right) \right) + \sum_{n=1}^{N} \pi_{Kn} \frac{c\gamma_K}{\delta_K} \qquad \text{(D.1)}
\end{aligned}$$

Substituting back in (4.22), we get (4.24). Q.E.D.